

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 731 437**

51 Int. Cl.:

C12N 15/90 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.11.2015 PCT/US2015/062023**

87 Fecha y número de publicación internacional: **26.05.2016 WO16081923**

96 Fecha de presentación y número de la solicitud europea: **20.11.2015 E 15804289 (5)**

97 Fecha y número de publicación de la concesión europea: **20.03.2019 EP 3221457**

54 Título: **Métodos y composiciones para la modificación genética dirigida mediante el uso de pares de ARN guías**

30 Prioridad:

21.11.2014 US 201462083005 P
19.06.2015 US 201562182314 P
28.08.2015 US 201562211421 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
15.11.2019

73 Titular/es:

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

72 Inventor/es:

MURPHY, ANDREW J.;
FRENDEWEY, DAVID;
LAI, KA-MAN VENUS;
AUERBACH, WOJTEK;
DROGUETT, GUSTAVO;
GAGLIARDI, ANTHONY;
VALENZUELA, DAVID M.;
VORONINA, VERA;
MACDONALD, LYNN y
YANCOPOULOS, GEORGE D.

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 731 437 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 mediante el uso de pares de ARN guías

5 Referencia cruzada a las solicitudes relacionadas

Esta solicitud reivindica el beneficio de la solicitud de patente de Estados Unidos núm. 62/083,005, presentada el 21 de noviembre de 2014, la solicitud de patente de Estados Unidos núm. 62/182,314, presentada el 19 de junio de 2015, y la solicitud de patente de Estados Unidos núm. 62/211,421, presentada el 28 de agosto de 2015.

10

Referencia a una lista de secuencias presentada

Como un archivo de texto por medio del sistema electrónico efs web

15

La Lista de Secuencias escrita en el archivo 472225SEQLIST.txt es de 32,7 kb, y se creó el 20 de noviembre de 2015.

Antecedentes de la invención

20

Aunque se han hecho progresos en la transformación de varios loci genómicos, aún quedan muchos tipos de loci genómicos que no pueden transformarse de manera eficiente o modificaciones genómicas que no pueden lograrse adecuadamente o de manera eficiente con estrategias de transformación convencionales. Por ejemplo, surgen dificultades cuando se intenta crear delecciones genómicas dirigidas grandes u otras modificaciones genéticas dirigidas grandes, particularmente en células y organismos eucariotas.

25

En particular, es difícil producir de manera eficiente células o animales que son homocigóticos o heterocigóticos combinados (por ejemplo, hemicigóticos) para una delección genómica dirigida grande u otra modificación genómica cuando se usan estrategias de transformación convencionales. Por ejemplo, aunque pueden obtenerse ratones de generación F0 heterocigóticos para una delección genómica dirigida grande por medio de estrategias de transformación convencionales, se requiere el cruzamiento posterior de estos ratones heterocigóticos para producir ratones de generación F2 que son homocigóticos para la delección. Estas etapas de cruzamiento adicionales son costosas y consumen tiempo. Yoshimi y otros 2014, Nature Communications, Vol. 5, páginas 1-17; Fujii Wataru y otros 2014, Biochemical and Biophysical Research Communication, vol. 445, núm. 4 páginas 791-794 e información complementaria páginas 1-6 y S.J.Gratz y otros 2014, Genetics, vol.196, núm.4 páginas 961-971 describen métodos de modificación bialélica de un genoma dentro de una célula. La patente US 2005/144655 describe un método para identificar una inserción dirigida de un inserto de ácido nucleico en una célula diploide. El documento WO2014/104878 describe un método para modificar un genoma dentro de una célula que es heterocigótica para un primer alelo en donde la célula se modifica para convertirla en homocigótica para el primer alelo y en donde la secuencia de reconocimiento de ARN de CRISPR es centromérica al locus correspondiente al primer alelo. S.M.Byrne y otros 2014, Nucleic Acids Research, vol. 43, núm. 3, páginas e21-e21, describe que el uso de dos ARN guías simples conduce al reemplazo homocigótico del gen de THY1 humano con su homólogo de ratón. Yang y otros 2003, Nature Biotechnology, vol.21, páginas 447-451 y Susan M. Byrne y otros 2014, Methods in Enzymology, Vol 546, páginas 119-138, describen la determinación de la variación del número de copias por FISH.

30

35

40

Resumen

45

La invención es como se define en las reivindicaciones.

En un primer aspecto, la invención proporciona un método para hacer una modificación bialélica a un locus genómico diana en un genoma dentro de una célula, que comprende:

50

(I) introducir en una población de células:

(a) una proteína Cas;

(b) un primer ARN guía que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana;

55

(c) un segundo ARN guía que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana; y

(d) un vector de transformación que comprende un inserto de ácido nucleico flanqueado por un brazo de homología 5' que se hibrida con una secuencia diana 5' dentro del locus genómico diana y un brazo de homología 3' que se hibrida con una secuencia diana 3' dentro del locus genómico diana, en donde si la célula es un embrión en etapa de una célula el vector de transformación tiene una longitud de no más de 5 kb;

60

en donde el genoma comprende un par de primer y segundo cromosomas homólogos que comprenden el locus genómico diana, opcionalmente en donde la primera y segunda secuencias de reconocimiento de ARN de CRISPR flanquean la totalidad o parte de una secuencia codificante de un gen; y

en donde la proteína Cas escinde al menos una de la primera y segunda secuencias de reconocimiento de ARN de CRISPR para generar al menos una ruptura bicatenaria en cada uno del primer y segundo cromosomas homólogos, opcionalmente en donde la introducción del primer y segundo ARN guías da como resultado el aumento de la eficiencia de modificación bialélica en comparación con la introducción del primer ARN guía o el segundo ARN guía solos; y

65

(II) identificar una célula que comprende un locus genómico diana modificado que comprende una delección y/o una inserción, en donde la identificación comprende realizar un ensayo cuantitativo de modificación de alelo y un ensayo de retención,

en donde el ensayo de modificación de alelo comprende:

- 5 (a) un ensayo de ganancia de alelo para determinar el número de copias de una plantilla de ADN de inserto de ácido nucleico en una muestra de ADN genómico de la célula; y/o
- 10 (b) un ensayo de pérdida de alelo para determinar el número de copias en la muestra de ADN genómico de una plantilla de ADN dentro de una región del locus genómico diana reconocida para la delección, y en donde el ensayo de retención determina el número de copias en la muestra de ADN genómico de una plantilla de ADN de la secuencia diana 5' que se pretende retener en el locus genómico diana modificado y/o una plantilla de ADN de la secuencia diana 3' que se pretende retener en el locus genómico diana modificado,
- 15 en donde la célula no se produce mediante el uso de un proceso que involucra la modificación de la identidad genética de la línea germinal de seres humanos o que involucra el uso de un embrión humano para fines industriales o comerciales, y
- en donde el método no es un método para el tratamiento del cuerpo humano o animal mediante terapia.

Se proporcionan métodos y composiciones para modificar un genoma dentro de una célula. En un aspecto, la descripción proporciona métodos para hacer una modificación a un genoma dentro de una célula, que comprenden poner en contacto el genoma con: (a) una primera proteína Cas; (b) un primer ARN de CRISPR que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR dentro de un locus genómico diana; (c) un segundo ARN de CRISPR que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana; (d) un ARNtracr; y (e) un vector de transformación que comprende un inserto de ácido nucleico flanqueado por un brazo de homología 5' que se hibrida con una secuencia diana 5' y un brazo de homología 3' que se hibrida con una secuencia diana 3', siempre que si la célula es un embrión en etapa de una célula el vector de transformación tiene una longitud de no más de 5 kb; en donde el genoma comprende un par de primer y segundo cromosomas homólogos que comprenden el locus genómico diana; y en donde la primera proteína Cas escinde al menos una de la primera y segunda secuencias de reconocimiento de ARN de CRISPR para generar al menos una ruptura bicatenaria en al menos uno del primer y segundo cromosomas homólogos. En un aspecto, la descripción proporciona métodos para hacer una modificación bialélica a un genoma dentro de una célula, que comprenden poner en contacto el genoma con: (a) una primera proteína Cas; (b) un primer ARN de CRISPR que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR dentro de un locus genómico diana; (c) un segundo ARN de CRISPR que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana; (d) un ARNtracr; y (e) un vector de transformación que comprende un inserto de ácido nucleico flanqueado por un brazo de homología 5' que se hibrida con una secuencia diana 5' y un brazo de homología 3' que se hibrida con una secuencia diana 3', siempre que si la célula es un embrión en etapa de una célula el vector de transformación tiene una longitud de no más de 5 kb; en donde el genoma comprende un par de primer y segundo cromosomas homólogos que comprenden el locus genómico diana; y en donde la primera proteína Cas escinde al menos una de la primera y segunda secuencias de reconocimiento de ARN de CRISPR para generar al menos una ruptura bicatenaria en al menos uno del primer y segundo cromosomas homólogos.

Los métodos pueden comprender además identificar una célula que comprende el genoma modificado. En algunos métodos, el inserto de ácido nucleico comprende un casete de selección adyacente a un primer brazo de homología que se hibrida con una primera secuencia diana, en donde el primer brazo de homología es el brazo de homología 5' y la primera secuencia diana es la secuencia diana 5', o en donde el primer brazo de homología es el brazo de homología 3' y la primera secuencia diana es la secuencia diana 3', en donde la identificación comprende: (a) obtener ADN de la célula; (b) exponer el ADN de la célula a una sonda que se une dentro de la primera secuencia diana, una sonda que se une dentro del inserto de ácido nucleico, y una sonda que se une dentro de un gen de referencia que tiene un número de copias conocido, en donde cada sonda genera una señal detectable tras la unión; (c) detectar las señales de la unión de cada una de las sondas; y (d) comparar la señal de la sonda del gen de referencia con la señal de la sonda de la primera secuencia diana para determinar un número de copias de la primera secuencia diana, y comparar la señal de la sonda del gen de referencia con la señal de la sonda del inserto de ácido nucleico para determinar un número de copias del inserto de ácido nucleico, en donde un número de copias del inserto de ácido nucleico de uno o dos y un número de copias de la primera secuencia diana de dos indica la inserción dirigida del inserto de ácido nucleico en el locus genómico diana, y en donde un número de copias del inserto de ácido nucleico de uno o más y un número de copias de la primera secuencia diana de tres o más indica una inserción aleatoria del inserto de ácido nucleico en un locus genómico distinto al locus genómico diana.

En algunos métodos, la primera proteína Cas escinde al menos una de la primera y segunda secuencias de reconocimiento de ARN de CRISPR en cada uno del primer y segundo cromosomas homólogos para generar al menos una ruptura bicatenaria en cada uno del primer y segundo cromosomas homólogos. En algunos métodos, la primera proteína Cas escinde la primera y segunda secuencias de reconocimiento de ARN de CRISPR en al menos uno del primer y segundo cromosomas homólogos para generar al menos dos rupturas bicatenarias en al menos uno del primer y segundo cromosomas homólogos.

Algunos métodos comprenden además poner en contacto el genoma con: un tercer ARN de CRISPR que se hibrida con una tercera secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana; y un cuarto ARN de CRISPR que se hibrida con una cuarta secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico

diana. Opcionalmente, la primera secuencia de reconocimiento de ARN de CRISPR y la tercera secuencia de reconocimiento de ARN de CRISPR se encuentran separadas por aproximadamente 25 pb a aproximadamente 50 pb, aproximadamente 50 pb a aproximadamente 100 pb, aproximadamente 100 pb a aproximadamente 150 pb, aproximadamente 150 pb a aproximadamente 200 pb, aproximadamente 200 pb a aproximadamente 250 pb, aproximadamente 250 pb a aproximadamente 300 pb, aproximadamente 300 pb a aproximadamente 350 pb, aproximadamente 350 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 450 pb, aproximadamente 450 pb a aproximadamente 500 pb, aproximadamente 500 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 700 pb, aproximadamente 700 pb a aproximadamente 800 pb, aproximadamente 800 pb a aproximadamente 900 pb, aproximadamente 900 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 6 kb, aproximadamente 6 kb a aproximadamente 7 kb, aproximadamente 7 kb a aproximadamente 8 kb, aproximadamente 8 kb a aproximadamente 9 kb, aproximadamente 9 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 30 kb, aproximadamente 30 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 50 kb, aproximadamente 50 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 70 kb, aproximadamente 70 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 90 kb, o aproximadamente 90 kb a aproximadamente 100 kb. Opcionalmente, la segunda secuencia de reconocimiento de ARN de CRISPR y la cuarta secuencia de reconocimiento de ARN de CRISPR se encuentran separadas por aproximadamente 25 pb a aproximadamente 50 pb, aproximadamente 50 pb a aproximadamente 100 pb, aproximadamente 100 pb a aproximadamente 150 pb, aproximadamente 150 pb a aproximadamente 200 pb, aproximadamente 200 pb a aproximadamente 250 pb, aproximadamente 250 pb a aproximadamente 300 pb, aproximadamente 300 pb a aproximadamente 350 pb, aproximadamente 350 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 450 pb, aproximadamente 450 pb a aproximadamente 500 pb, aproximadamente 500 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 700 pb, aproximadamente 700 pb a aproximadamente 800 pb, aproximadamente 800 pb a aproximadamente 900 pb, aproximadamente 900 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 6 kb, aproximadamente 6 kb a aproximadamente 7 kb, aproximadamente 7 kb a aproximadamente 8 kb, aproximadamente 8 kb a aproximadamente 9 kb, aproximadamente 9 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 30 kb, aproximadamente 30 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 50 kb, aproximadamente 50 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 70 kb, aproximadamente 70 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 90 kb, o aproximadamente 90 kb a aproximadamente 100 kb. Opcionalmente, la primera y tercera secuencias de reconocimiento de ARN de CRISPR son un primer par de secuencias de reconocimiento de ARN de CRISPR, y la segunda y cuarta secuencias de reconocimiento de ARN de CRISPR son un segundo par de secuencias de reconocimiento de ARN de CRISPR, en donde el primer par y el segundo par se encuentran separados por aproximadamente 25 pb a aproximadamente 50 pb, aproximadamente 50 pb a aproximadamente 100 pb, aproximadamente 100 pb a aproximadamente 150 pb, aproximadamente 150 pb a aproximadamente 200 pb, aproximadamente 200 pb a aproximadamente 250 pb, aproximadamente 250 pb a aproximadamente 300 pb, aproximadamente 300 pb a aproximadamente 350 pb, aproximadamente 350 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 450 pb, aproximadamente 450 pb a aproximadamente 500 pb, aproximadamente 500 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 700 pb, aproximadamente 700 pb a aproximadamente 800 pb, aproximadamente 800 pb a aproximadamente 900 pb, aproximadamente 900 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 150 kb, aproximadamente 150 kb a aproximadamente 200 kb, aproximadamente 200 kb a aproximadamente 300 kb, aproximadamente 300 kb a aproximadamente 400 kb, aproximadamente 400 kb a aproximadamente 500 kb, aproximadamente 500 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 1,5 Mb, aproximadamente 1,5 Mb a aproximadamente 2 Mb, aproximadamente 2 Mb a aproximadamente 2,5 Mb, aproximadamente 2,5 Mb a aproximadamente 3 Mb, aproximadamente 3 Mb a aproximadamente 4 Mb, aproximadamente 4 Mb a aproximadamente 5 Mb, aproximadamente 5 Mb a aproximadamente 10 Mb, aproximadamente 10 Mb a aproximadamente 20 Mb, aproximadamente 20 Mb a aproximadamente 30 Mb, aproximadamente 30 Mb a aproximadamente 40 Mb, aproximadamente 40 Mb a aproximadamente 50 Mb, aproximadamente 50 Mb a aproximadamente 60 Mb, aproximadamente 60 Mb a aproximadamente 70 Mb, aproximadamente 70 Mb a aproximadamente 80 Mb, aproximadamente 80 Mb a aproximadamente 90 Mb, o aproximadamente 90 Mb a aproximadamente 100 Mb.

En algunos métodos, la primera proteína Cas escinde al menos dos de la primera, segunda, tercera y cuarta secuencias de reconocimiento de ARN de CRISPR para generar al menos dos rupturas bicatenarias en al menos uno del primer y segundo cromosomas homólogos. En algunos métodos, la primera proteína Cas escinde al menos dos de la primera, segunda, tercera y cuarta secuencias de reconocimiento de ARN de CRISPR para generar al menos dos rupturas bicatenarias en el primer y segundo cromosomas homólogos.

En algunos métodos, el inserto de ácido nucleico se inserta entre las secuencias diana 5' y 3'. Opcionalmente, las secuencias diana 5' y 3' están dentro del locus genómico diana. Opcionalmente, la célula no es un embrión en etapa de una célula, y el vector de transformación es un vector de transformación grande (LTVEC) que es de al menos 10 kb.

5 En algunos métodos, poner en contacto el genoma con el primer y segundo ARN de CRISPR da como resultado el aumento de la eficiencia de modificación bialélica en comparación con poner en contacto el genoma con el primer ARN de CRISPR o el segundo ARN de CRISPR solos. En algunos métodos, la célula es diploide, y la modificación bialélica da como resultado homocigosidad o heterocigosidad combinada en el locus genómico diana. Opcionalmente, la heterocigosidad combinada es hemicigosidad. En algunos métodos, la modificación bialélica comprende una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo. En algunos métodos, la modificación bialélica comprende la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer y segundo cromosomas homólogos. En algunos métodos, la modificación bialélica comprende además la inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer y segundo cromosomas homólogos. En algunos métodos, la modificación bialélica comprende: (1) la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer y segundo cromosomas homólogos; y (2) la inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer cromosoma homólogo pero no en el segundo cromosoma homólogo. En algunos métodos, la modificación bialélica comprende: (1) la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo; y (2) la interrupción de un locus entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el segundo cromosoma homólogo. En algunos de los métodos, la modificación bialélica comprende: (1) la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo; (2) una inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer cromosoma homólogo; y (3) la interrupción de un locus entre las secuencias diana 5' y 3' en el segundo cromosoma homólogo. En algunos métodos, la modificación bialélica comprende: (1) la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo; y (2) una inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer cromosoma homólogo, en donde la secuencia del inserto de ácido nucleico es homóloga u ortóloga a la secuencia eliminada.

En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se encuentran separadas por aproximadamente 1 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 150 kb, aproximadamente 150 kb a aproximadamente 200 kb, aproximadamente 200 kb a aproximadamente 300 kb, aproximadamente 300 kb a aproximadamente 400 kb, aproximadamente 400 kb a aproximadamente 500 kb, aproximadamente 500 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 1,5 Mb, aproximadamente 1,5 Mb a aproximadamente 2 Mb, aproximadamente 2 Mb a aproximadamente 2,5 Mb, o aproximadamente 2,5 Mb a aproximadamente 3 Mb. En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se encuentran separadas por al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, 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 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, 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. En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se encuentran separadas por aproximadamente 25 pb a aproximadamente 50 pb, aproximadamente 50 pb a aproximadamente 100 pb, aproximadamente 100 pb a aproximadamente 150 pb, aproximadamente 150 pb a aproximadamente 200 pb, aproximadamente 200 pb a aproximadamente 250 pb, aproximadamente 250 pb a aproximadamente 300 pb, aproximadamente 300 pb a aproximadamente 350 pb, aproximadamente 350 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 450 pb, aproximadamente 450 pb a aproximadamente 500 pb, aproximadamente 500 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 700 pb, aproximadamente 700 pb a aproximadamente 800 pb, aproximadamente 800 pb a aproximadamente 900 pb, o aproximadamente 900 pb a aproximadamente 1 kb. En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se encuentran separadas por menos de 25 pb, menos de 50 pb, menos de 100 pb, menos de 150 pb, menos de 200 pb, menos de 250 pb, menos de 300 pb, menos de 350 pb, menos de 400 pb, menos de 450 pb, menos de 500 pb, menos de 600 pb, menos de 700 pb, menos de 800 pb, menos de 900 pb, menos de 1 kb, menos de 2 kb, menos de 3 kb, menos de 4 kb, menos de 5 kb, o menos de 10 kb.

En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se ubican cada una al menos 50 pb, al menos 100 pb, al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 600 pb, al menos 700 pb, al menos 800 pb, al menos 900 pb, al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, al menos 6 kb, al menos 7 kb, al menos 8 kb, al menos 9 kb, 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, o al menos 100 kb de las secuencias diana 5' y 3'. En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se ubican cada una entre aproximadamente 50 pb a aproximadamente 100 pb, aproximadamente 200 pb a aproximadamente 300 pb, aproximadamente 300 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 500 pb, aproximadamente 500 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 700 pb, aproximadamente 700 pb a aproximadamente 800 pb, aproximadamente 800 pb a

aproximadamente 900 pb, aproximadamente 900 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 30 kb, aproximadamente 30 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 50 kb, o aproximadamente 50 kb a aproximadamente 100 kb de las secuencias diana 5' y 3'. En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se ubican cada una más de 50 pb, más de 100 pb, más de 200 pb, más de 300 pb, más de 400 pb, más de 500 pb, más de 600 pb, más de 700 pb, más de 800 pb, más de 900 pb, más de 1 kb, más de 2 kb, más de 3 kb, más de 4 kb, más de 5 kb, más de 6 kb, más de 7 kb, más de 8 kb, más de 9 kb, más de 10 kb, más de 20 kb, más de 30 kb, más de 40 kb, más de 50 kb, más de 60 kb, más de 70 kb, más de 80 kb, más de 90 kb, o más de 100 kb de las secuencias diana 5' y 3'.

En algunos métodos, el ácido nucleico eliminado 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 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. En algunos métodos, el ácido nucleico eliminado es de 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, 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. Opcionalmente, el ácido nucleico eliminado es de 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, o al menos 2 Mb.

En algunos métodos, el vector de transformación está en forma lineal. Opcionalmente, el vector de transformación es monocatenario o bicatenario. En algunos métodos, la célula no es un embrión en etapa de una célula, y el vector de transformación es un vector de transformación grande (LTVEC) que es de al menos 10 kb. En algunos métodos, la célula no es un embrión en etapa de una célula, y el vector de transformación es un vector de transformación grande (LTVEC), en donde la suma total de los brazos de homología 5' y 3' del LTVEC es al menos 10 kb. Opcionalmente, el LTVEC es 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, 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. Opcionalmente, la suma total de los brazos de homología 5' y 3' del LTVEC es 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.

En algunos métodos, la célula huésped es una célula eucariota. Opcionalmente, la célula eucariota es una célula de mamífero, una célula humana, una célula no humana, una célula de roedor, una célula de ratón, y una célula de rata. Opcionalmente, la célula eucariota es una célula pluripotente, una célula no pluripotente, una célula pluripotente no humana, una célula pluripotente humana, una célula pluripotente de roedor, una célula pluripotente de ratón, una célula pluripotente de rata, una célula madre embrionaria (ES) de ratón, una célula ES de rata, una célula ES humana, una célula madre adulta humana, una célula progenitora humana restringida por el desarrollo, o una célula madre pluripotente inducida (iPS) humana. Opcionalmente, la célula eucariota es un embrión en etapa de una célula. Opcionalmente, la célula eucariota es un embrión en etapa de una célula, y el vector de transformación es de entre aproximadamente 50 nucleótidos a aproximadamente 5 kb de longitud. Opcionalmente, la célula eucariota es un embrión en etapa de una célula, y el vector de transformación es ADN monocatenario y es de entre aproximadamente 60 a aproximadamente 200 nucleótidos de longitud.

En algunos métodos, la primera proteína Cas es Cas9. En algunos métodos, la primera proteína Cas tiene actividad nucleasa en ambas cadenas de ADN bicatenario.

- En algunos métodos, la primera proteína Cas es una nickasa. Algunos métodos comprenden además poner en contacto el genoma con: (f) una segunda proteína Cas que es una nickasa; (g) un tercer ARN de CRISPR que se hibrida con una tercera secuencia de reconocimiento de ARN de CRISPR; y (h) un cuarto ARN de CRISPR que se hibrida con una cuarta secuencia de reconocimiento de ARN de CRISPR; en donde la primera proteína Cas escinde una primera cadena de ADN genómico dentro de la primera secuencia de reconocimiento de ARN de CRISPR y dentro de la segunda secuencia de reconocimiento de ARN de CRISPR, y la segunda proteína Cas escinde una segunda cadena de ADN genómico dentro de la tercera secuencia de reconocimiento de ARN de CRISPR y dentro de la cuarta secuencia de reconocimiento de ARN de CRISPR.
- En algunos métodos, el primer ARN de CRISPR y el ARNtracr se fusionan entre sí como un primer ARN guía (ARNg), y/o el segundo ARN de CRISPR y el ARNtracr se fusionan entre sí como un segundo ARNg. En algunos métodos, el primer ARN de CRISPR y el ARNtracr son moléculas de ARN separadas, y/o el segundo ARN de CRISPR y el ARNtracr son moléculas de ARN separadas.
- En algunos métodos, el contacto comprende introducir la primera proteína Cas, el primer y segundo ARN de CRISPR, y el ARNtracr en la célula. En algunos métodos, (a) la primera proteína Cas se introduce en la célula en forma de una proteína, un ARN mensajero (ARNm) que codifica la primera proteína Cas, o un ADN que codifica la primera proteína Cas; (b) el primer ARN de CRISPR se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el primer ARN de CRISPR; (c) el segundo ARN de CRISPR se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el segundo ARN de CRISPR; y/o (d) el ARNtracr se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el ARNtracr. En algunos métodos, la primera proteína Cas, el primer ARN de CRISPR, y el ARNtracr se introducen en la célula como un primer complejo de ARN y proteína, y/o la primera proteína Cas, el segundo ARN de CRISPR, y el ARNtracr se introducen en la célula como un segundo complejo de ARN y proteína. En algunos métodos, (a) el ADN que codifica la primera proteína Cas se une operativamente a un primer promotor en un primer constructo de expresión; (b) el ADN que codifica el primer ARN de CRISPR se une operativamente a un segundo promotor en un segundo constructo de expresión; (c) el ADN que codifica el segundo ARN de CRISPR se une operativamente a un tercer promotor en un tercer constructo de expresión; y/o (d) el ADN que codifica el ARNtracr se une operativamente a un cuarto promotor en un cuarto constructo de expresión; en donde el primer, segundo, tercer, y cuarto promotores son activos en la célula. Opcionalmente, el primer, segundo, tercer, y/o cuarto constructos de expresión son componentes de una sola molécula de ácido nucleico. En algunos métodos, (a) el ADN que codifica la primera proteína Cas se une operativamente a un primer promotor en un primer constructo de expresión; (b) los ADN que codifican el primer ARN de CRISPR y el ARNtracr se fusionan entre sí en un ADN que codifica un primer ARN guía (ARNg) y se unen operativamente a un segundo promotor en un segundo constructo de expresión; y/o (c) los ADN que codifican el segundo ARN de CRISPR y el ARNtracr se fusionan entre sí en un ADN que codifica un segundo ARNg y se unen operativamente a un tercer promotor en un tercer constructo de expresión; en donde el primer, segundo, y tercer promotores son activos en la célula. Opcionalmente, el primer, segundo, y/o tercer constructos de expresión son componentes de una sola molécula de ácido nucleico.
- En algunos métodos, la célula se ha modificado para disminuir la unión de extremos no homólogos (NHEJ) y/o para aumentar la conversión génica o la reparación dirigida por homología (HDR). Opcionalmente, la célula se ha modificado para disminuir la expresión o la actividad de ADN-PK y/o para disminuir la expresión o la actividad de PARP1. Opcionalmente, la célula se ha modificado para disminuir la expresión o la actividad de la ligasa IV. Opcionalmente, la disminución de la expresión o la actividad es inducible, reversible, temporalmente específica, y/o espacialmente específica.
- En algunos métodos, (1) la célula no es un embrión en etapa de una célula, y el vector de transformación es un vector de transformación grande, en donde los brazos de homología 5' y 3' tienen una suma total de al menos 10 kb; (2) la primera y segunda secuencias de reconocimiento de ARN de CRISPR se ubican cada una más de 200 pb, más de 300 pb, más de 400 pb, más de 500 pb, más de 600 pb, más de 700 pb, más de 800 pb, más de 900 pb, más de 1 kb, más de 2 kb, más de 3 kb, más de 4 kb, más de 5 kb, más de 6 kb, más de 7 kb, más de 8 kb, más de 9 kb, más de 10 kb, más de 20 kb, más de 30 kb, más de 40 kb, más de 50 kb, más de 60 kb, más de 70 kb, más de 80 kb, más de 90 kb, o más de 100 kb de las secuencias diana 5' y 3'; (3) la primera proteína Cas escinde la primera y segunda secuencias de reconocimiento de ARN de CRISPR en al menos uno del primer y segundo cromosomas homólogos para generar al menos dos rupturas bicatenarias en al menos uno del primer y segundo cromosomas homólogos; y (4) la modificación bialélica comprende la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo y una inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer cromosoma homólogo, en donde la secuencia del inserto de ácido nucleico es homóloga u ortóloga a la secuencia eliminada.
- La descripción proporciona además métodos para producir un animal no humano de generación F0, que comprenden: (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 bialélica. Algunos métodos comprenden: (a) poner en contacto el genoma en una célula ES no humana con: (i) una primera proteína Cas; (ii) un primer ARN de CRISPR que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR dentro de un locus genómico diana; (iii) un segundo ARN de CRISPR que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana; (iv) un ARNtracr; y (v) un vector de transformación

5 que comprende un inserto de ácido nucleico flanqueado por un brazo de homología 5' y un brazo de homología 3', en donde el genoma comprende un par de primer y segundo cromosomas homólogos que comprenden el locus genómico diana; y en donde la primera proteína Cas escinde al menos una de la primera y segunda secuencias de reconocimiento de ARN de CRISPR para generar al menos una ruptura bicatenaria en al menos uno del primer y segundo cromosomas homólogos; (b) identificar una célula ES no humana que comprende la modificación bialélica; (c) introducir la célula ES no humana que comprende la modificación bialélica en un embrión huésped no humano; y (d) 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 bialélica.

10 En algunos métodos, la primera proteína Cas escinde al menos una de la primera y segunda secuencias de reconocimiento de ARN de CRISPR en cada uno del primer y segundo cromosomas homólogos para generar al menos una ruptura bicatenaria en cada uno del primer y segundo cromosomas homólogos. En algunos métodos, la primera proteína Cas escinde la primera y segunda secuencias de reconocimiento de ARN de CRISPR en al menos uno del primer y segundo cromosomas homólogos para generar al menos dos rupturas bicatenarias en al menos uno del primer y
15 segundo cromosomas homólogos.

20 En algunos métodos, el animal no humano es un ratón, la célula ES no humana es una célula ES de ratón, y el embrión huésped no humano es un embrión huésped de ratón. En algunos métodos, el animal no humano es una rata, la célula ES no humana es una célula ES de rata, y el embrión huésped no humano es un embrión huésped de rata.

25 En algunos métodos, la modificación bialélica da como resultado homocigosidad o heterocigosidad combinada en el locus genómico diana. Opcionalmente, la heterocigosidad combinada es hemocigosidad.

30 En algunos métodos, el primer ARN de CRISPR y el ARNtracr se fusionan entre sí como un primer ARN guía (ARNg), y/o el segundo ARN de CRISPR y el ARNtracr se fusionan entre sí como un segundo ARNg. En algunos métodos, el primer ARN de CRISPR y el ARNtracr son moléculas de ARN separadas, y/o el segundo ARN de CRISPR y el ARNtracr son moléculas de ARN separadas.

35 En algunos métodos, el contacto comprende introducir la primera proteína Cas, el primer y segundo ARN de CRISPR, y el ARNtracr en la célula. En algunos métodos, (a) la primera proteína Cas se introduce en la célula en forma de una proteína, un ARN mensajero (ARNm) que codifica la primera proteína Cas, o un ADN que codifica la primera proteína Cas; (b) el primer ARN de CRISPR se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el primer ARN de CRISPR; (c) el segundo ARN de CRISPR se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el segundo ARN de CRISPR; y/o (d) el ARNtracr se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el ARNtracr. En algunos métodos, la primera proteína Cas, el primer ARN de CRISPR, y el ARNtracr se introducen en la célula como un primer complejo de ARN y proteína, y/o la primera proteína Cas, el segundo ARN de CRISPR, y el ARNtracr se introducen en la célula como un segundo complejo de ARN y proteína. En algunos métodos, (a) el ADN que codifica la primera proteína Cas se une operativamente a un primer promotor en un primer constructo de expresión; (b) el ADN que codifica el primer ARN de CRISPR se une operativamente a un segundo promotor en un segundo constructo de expresión; (c) el ADN que codifica el segundo ARN de CRISPR se une operativamente a un tercer promotor en un tercer constructo de expresión; y/o (d) el ADN que codifica el ARNtracr se une operativamente a un cuarto promotor en un cuarto constructo de expresión; en donde el primer, segundo, tercer, y cuarto promotores son activos en la célula. Opcionalmente, el primer, segundo, tercer, y/o cuarto constructos de expresión son componentes de una sola molécula de ácido nucleico. En algunos métodos, (a) el ADN que codifica la primera proteína Cas se une operativamente a un primer promotor en un primer constructo de expresión; (b) los ADN que codifican el primer ARN de CRISPR y el ARNtracr se fusionan entre sí en un ADN que codifica un primer ARN guía (ARNg) y se unen operativamente a un segundo promotor en un segundo constructo de expresión; y/o (c) los ADN que codifican el segundo ARN de CRISPR y el ARNtracr se fusionan entre sí en un ADN que codifica un segundo ARNg y se unen operativamente a un tercer promotor en un tercer constructo de expresión; en donde el primer, segundo, y tercer promotores son activos en la célula. Opcionalmente, el primer, segundo, y/o tercer constructos de expresión son componentes de una sola molécula de ácido nucleico.

45 En algunos métodos, la primera proteína Cas es Cas9. En algunos métodos, la primera proteína Cas tiene actividad nucleasa en ambas cadenas de ADN bicatenario.

50 En algunos métodos, la primera proteína Cas es una nickasa. Algunos métodos comprenden además poner en contacto el genoma con: (f) una segunda proteína Cas que es una nickasa; (g) un tercer ARN de CRISPR que se hibrida con una tercera secuencia de reconocimiento de ARN de CRISPR; y (h) un cuarto ARN de CRISPR que se hibrida con una cuarta secuencia de reconocimiento de ARN de CRISPR; en donde la primera proteína Cas escinde una primera cadena de ADN genómico dentro de la primera secuencia de reconocimiento de ARN de CRISPR y dentro de la segunda secuencia de reconocimiento de ARN de CRISPR, y la segunda proteína Cas escinde una segunda cadena de ADN genómico dentro de la tercera secuencia de reconocimiento de ARN de CRISPR y dentro de la cuarta secuencia de reconocimiento de ARN de CRISPR.

60 En algunos métodos, la célula se ha modificado para disminuir la unión de extremos no homólogos (NHEJ) y/o para aumentar la conversión génica o la reparación dirigida por homología (HDR). Opcionalmente, la célula se ha modificado

para disminuir la expresión o la actividad de ADN-PK y/o para disminuir la expresión o la actividad de PARP1. Opcionalmente, la célula se ha modificado para disminuir la expresión o la actividad de la ligasa IV. Opcionalmente, la disminución de la expresión o la actividad es inducible, reversible, temporalmente específica, y/o espacialmente específica.

La descripción proporciona además métodos para producir un animal no humano de generación F0, que comprenden implantar en una madre sustituta un embrión genéticamente modificado en etapa de una célula que se produjo por cualquiera de los métodos anteriores; en donde la madre sustituta produce el animal no humano de generación F0 que comprende la modificación bialélica.

La descripción proporciona además métodos para modificar un genoma dentro de una célula que es heterocigótica para un primer alelo, que comprenden poner en contacto el genoma con: (a) una primera proteína Cas; (b) un ARNtracr; (c) un primer ARN de CRISPR que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR dentro de un segundo alelo, en donde el primer alelo está en un primer cromosoma homólogo y el segundo alelo está en un locus correspondiente en un segundo cromosoma homólogo; y (d) un segundo ARN de CRISPR que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR dentro del segundo alelo; en donde la primera proteína Cas escinde al menos una de la primera y segunda secuencias de reconocimiento de ARN de CRISPR para generar al menos una ruptura bicatenaria y secuencias de extremos que experimentan recombinación, en donde la recombinación es entre el primer y segundo alelos para formar un genoma modificado que es homocigótico para el primer alelo. Algunos métodos comprenden además identificar una célula que es homocigótica para el primer alelo.

En algunos métodos, la primera proteína Cas escinde la primera secuencia de reconocimiento de ARN de CRISPR y la segunda secuencia de reconocimiento de ARN de CRISPR. En algunos métodos, la primera proteína Cas escinde la primera secuencia de reconocimiento de ARN de CRISPR y la segunda secuencia de reconocimiento de ARN de CRISPR para generar al menos dos rupturas bicatenarias y secuencias de extremos que experimentan recombinación. En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se ubican dentro del segundo alelo pero no en el primer alelo. En algunos métodos, la proteína Cas y el primer ARN de CRISPR no aparecen juntos en la naturaleza.

En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se encuentran separadas por aproximadamente 1 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 150 kb, aproximadamente 150 kb a aproximadamente 200 kb, aproximadamente 200 kb a aproximadamente 300 kb, aproximadamente 300 kb a aproximadamente 400 kb, aproximadamente 400 kb a aproximadamente 500 kb, aproximadamente 500 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 1,5 Mb, aproximadamente 1,5 Mb a aproximadamente 2 Mb, aproximadamente 2 Mb a aproximadamente 2,5 Mb, o aproximadamente 2,5 Mb a aproximadamente 3 Mb. En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se encuentran separadas por al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, 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 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, 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. En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se encuentran separadas por aproximadamente 25 pb a aproximadamente 50 pb, aproximadamente 50 pb a aproximadamente 100 pb, aproximadamente 100 pb a aproximadamente 150 pb, aproximadamente 150 pb a aproximadamente 200 pb, aproximadamente 200 pb a aproximadamente 250 pb, aproximadamente 250 pb a aproximadamente 300 pb, aproximadamente 300 pb a aproximadamente 350 pb, aproximadamente 350 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 450 pb, aproximadamente 450 pb a aproximadamente 500 pb, aproximadamente 500 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 700 pb, aproximadamente 700 pb a aproximadamente 800 pb, aproximadamente 800 pb a aproximadamente 900 pb, o aproximadamente 900 pb a aproximadamente 1 kb. En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se encuentran separadas por menos de 25 pb, menos de 50 pb, menos de 100 pb, menos de 150 pb, menos de 200 pb, menos de 250 pb, menos de 300 pb, menos de 350 pb, menos de 400 pb, menos de 450 pb, menos de 500 pb, menos de 600 pb, menos de 700 pb, menos de 800 pb, menos de 900 pb, menos de 1 kb, menos de 2 kb, menos de 3 kb, menos de 4 kb, menos de 5 kb, o menos de 10 kb.

En algunos métodos, las diferencias de secuencias entre el primer alelo y el segundo alelo abarcan aproximadamente 100 pb a aproximadamente 200 pb, aproximadamente 200 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 800 pb, aproximadamente 800 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 150 kb, aproximadamente 150 kb a aproximadamente 200 kb, aproximadamente 200 kb a aproximadamente 300 kb, aproximadamente 300 kb a aproximadamente 400 kb,

aproximadamente 400 kb a aproximadamente 500 kb, aproximadamente 500 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 1,5 Mb, aproximadamente 1,5 Mb a aproximadamente 2 Mb, aproximadamente 2 Mb a aproximadamente 2,5 Mb, o aproximadamente 2,5 Mb a aproximadamente 3 Mb. En algunos métodos, las diferencias de secuencias entre el primer alelo y el segundo alelo abarcan al menos 100 pb, al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 600 pb, al menos 700 pb, al menos 800 pb, al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, al menos 6 kb, al menos 7 kb, al menos 8 kb, al menos 9 kb, al menos 10 kb, 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, 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.

En algunos métodos, el primer alelo comprende una modificación dirigida y el segundo alelo es un alelo de tipo silvestre. En algunos métodos, el primer alelo es un alelo de tipo silvestre, y el segundo alelo comprende una mutación que causa enfermedad.

En algunos métodos, la recombinación comprende la conversión génica. En algunos métodos, la recombinación comprende la pérdida de heterocigosidad (LOH).

En algunos métodos, la célula huésped es una célula eucariota. Opcionalmente, la célula eucariota es una célula de mamífero, una célula humana, una célula no humana, una célula de roedor, una célula de ratón, o una célula de rata. Opcionalmente, la célula eucariota es una célula pluripotente, una célula no pluripotente, una célula pluripotente no humana, una célula pluripotente humana, una célula pluripotente de roedor, una célula pluripotente de ratón, una célula pluripotente de rata, una célula madre embrionaria (ES) de ratón, una célula ES de rata, una célula ES humana, una célula madre adulta humana, una célula progenitora humana restringida por el desarrollo, o una célula madre pluripotente inducida (iPS) humana.

En algunos métodos, la primera proteína Cas es Cas9. En algunos métodos, la primera proteína Cas tiene actividad nucleasa en ambas cadenas de ADN bicatenario.

En algunos métodos, la primera proteína Cas es una nickasa. Algunos métodos comprenden además poner en contacto el genoma con: (f) una segunda proteína Cas que es una nickasa; (g) un tercer ARN de CRISPR que se hibrida con una tercera secuencia de reconocimiento de ARN de CRISPR; y (h) un cuarto ARN de CRISPR que se hibrida con una cuarta secuencia de reconocimiento de ARN de CRISPR; en donde la primera proteína Cas escinde una primera cadena de ADN genómico dentro de la primera secuencia de reconocimiento de ARN de CRISPR y dentro de la segunda secuencia de reconocimiento de ARN de CRISPR, y la segunda proteína Cas escinde una segunda cadena de ADN genómico dentro de la tercera secuencia de reconocimiento de ARN de CRISPR y dentro de la cuarta secuencia de reconocimiento de ARN de CRISPR.

En algunos métodos, el primer ARN de CRISPR y el ARNtracr se fusionan entre sí como un primer ARN guía (ARNg), y/o el segundo ARN de CRISPR y el ARNtracr se fusionan entre sí como un segundo ARNg. En algunos métodos, el primer ARN de CRISPR y el ARNtracr son moléculas de ARN separadas, y/o el segundo ARN de CRISPR y el ARNtracr son moléculas de ARN separadas.

En algunos métodos, el contacto comprende introducir la primera proteína Cas, el primer y segundo ARN de CRISPR, y el ARNtracr en la célula. En algunos métodos, (a) la primera proteína Cas se introduce en la célula en forma de una proteína, un ARN mensajero (ARNm) que codifica la primera proteína Cas, o un ADN que codifica la primera proteína Cas; (b) el primer ARN de CRISPR se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el primer ARN de CRISPR; (c) el segundo ARN de CRISPR se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el segundo ARN de CRISPR; y/o (d) el ARNtracr se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el ARNtracr. En algunos métodos, la primera proteína Cas, el primer ARN de CRISPR, y el ARNtracr se introducen en la célula como un primer complejo de ARN y proteína, y/o la primera proteína Cas, el segundo ARN de CRISPR, y el ARNtracr se introducen en la célula como un segundo complejo de ARN y proteína. En algunos métodos, (a) el ADN que codifica la primera proteína Cas se une operativamente a un primer promotor en un primer constructo de expresión; (b) el ADN que codifica el primer ARN de CRISPR se une operativamente a un segundo promotor en un segundo constructo de expresión; (c) el ADN que codifica el segundo ARN de CRISPR se une operativamente a un tercer promotor en un tercer constructo de expresión; y/o (d) el ADN que codifica el ARNtracr se une operativamente a un cuarto promotor en un cuarto constructo de expresión; en donde el primer, segundo, tercer, y cuarto promotores son activos en la célula. Opcionalmente, el primer, segundo, tercer, y/o cuarto constructos de expresión son componentes de una sola molécula de ácido nucleico. En algunos métodos, (a) el ADN que codifica la primera proteína Cas se une operativamente a un primer promotor en un primer constructo de expresión; (b) los ADN que codifican el primer ARN de CRISPR y el ARNtracr se fusionan entre sí en un ADN que codifica un primer ARN guía (ARNg) y se unen operativamente a un segundo promotor en un segundo constructo de expresión; y/o (c) los ADN que codifican el segundo ARN de CRISPR y el ARNtracr se fusionan entre sí en un ADN que codifica un segundo ARNg y se unen operativamente a un tercer promotor en un tercer constructo de expresión; en donde el primer, segundo, y tercer promotores son activos en la célula. Opcionalmente, el primer, segundo, y/o tercer constructos de expresión son componentes de una sola molécula de ácido nucleico.

En algunos métodos, la célula se ha modificado para disminuir la unión de extremos no homólogos (NHEJ) y/o para aumentar la conversión génica o la reparación dirigida por homología (HDR). Opcionalmente, la célula se ha modificado para disminuir la expresión o la actividad de ADN-PK y/o para disminuir la expresión o la actividad de PARP1. Opcionalmente, la célula se ha modificado para disminuir la expresión o la actividad de la ligasa IV. Opcionalmente, la disminución de la expresión o la actividad es inducible, reversible, temporalmente específica, y/o espacialmente específica.

La descripción proporciona además métodos para modificar un genoma dentro de una célula que es heterocigótica para un primer alelo, que comprenden poner en contacto el genoma con: (a) una primera proteína Cas; (b) un ARNtracr; y (c) un primer ARN de CRISPR que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR sin especificidad por alelo, en donde el primer alelo está en un primer cromosoma homólogo y la secuencia de reconocimiento de ARN de CRISPR es centromérica al locus correspondiente al primer alelo en un segundo cromosoma homólogo; y en donde la primera proteína Cas escinde la primera secuencia de reconocimiento de ARN de CRISPR para generar una ruptura bicatenaria y la célula se modifica para convertirla en homocigótica para el primer alelo. Algunos métodos comprenden además identificar una célula que es homocigótica para el primer alelo. En algunos métodos, la proteína Cas y el primer ARN de CRISPR no aparecen juntos en la naturaleza.

Tales métodos pueden comprender además poner en contacto el genoma con un segundo ARN de CRISPR que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR sin especificidad por alelo centromérica al locus correspondiente al primer alelo en un segundo cromosoma homólogo, en donde la primera proteína Cas escinde al menos una de la primera y segunda secuencias de reconocimiento de ARN de CRISPR para generar al menos una ruptura bicatenaria. En algunos métodos, la primera proteína Cas escinde la primera secuencia de reconocimiento de ARN de CRISPR y la segunda secuencia de reconocimiento de ARN de CRISPR.

En algunos métodos, la pérdida de heterocigosidad ocurre en la orientación telomérica de la ruptura bicatenaria.

En algunos métodos, la primera y segunda secuencias de reconocimiento de ARN de CRISPR se ubican en el segundo cromosoma homólogo pero no en el primer cromosoma homólogo. En algunos métodos, el primer sitio de reconocimiento de ARN de CRISPR está de aproximadamente 100 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 10 Mb, aproximadamente 10 Mb a aproximadamente 20 Mb, aproximadamente 20 Mb a aproximadamente 30 Mb, aproximadamente 30 Mb a aproximadamente 40 Mb, aproximadamente 40 Mb a aproximadamente 50 Mb, aproximadamente 50 Mb a aproximadamente 60 Mb, aproximadamente 60 Mb a aproximadamente 70 Mb, aproximadamente 70 Mb a aproximadamente 80 Mb, aproximadamente 80 Mb a aproximadamente 90 Mb, o aproximadamente 90 Mb a aproximadamente 100 Mb del centrómero. En algunos métodos, el primer alelo está de aproximadamente 100 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 10 Mb, aproximadamente 10 Mb a aproximadamente 20 Mb, aproximadamente 20 Mb a aproximadamente 30 Mb, aproximadamente 30 Mb a aproximadamente 40 Mb, aproximadamente 40 Mb a aproximadamente 50 Mb, aproximadamente 50 Mb a aproximadamente 60 Mb, aproximadamente 60 Mb a aproximadamente 70 Mb, aproximadamente 70 Mb a aproximadamente 80 Mb, aproximadamente 80 Mb a aproximadamente 90 Mb, o aproximadamente 90 Mb a aproximadamente 100 Mb del primer sitio de reconocimiento de ARN de CRISPR. En algunos métodos, la región del segundo cromosoma homólogo que se reemplaza por pérdida de heterocigosidad es de aproximadamente 100 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 10 Mb, aproximadamente 10 Mb a aproximadamente 20 Mb, aproximadamente 20 Mb a aproximadamente 30 Mb, aproximadamente 30 Mb a aproximadamente 40 Mb, aproximadamente 40 Mb a aproximadamente 50 Mb, aproximadamente 50 Mb a aproximadamente 60 Mb, aproximadamente 60 Mb a aproximadamente 70 Mb, aproximadamente 70 Mb a aproximadamente 80 Mb, o aproximadamente 80 Mb a aproximadamente 90 Mb, o aproximadamente 90 Mb a aproximadamente 100 Mb.

En algunos métodos, la célula huésped es una célula eucariota. Opcionalmente, la célula eucariota es una célula de mamífero, una célula humana, una célula no humana, una célula de roedor, una célula de ratón, una célula de rata, una célula pluripotente, una célula no pluripotente, una célula pluripotente no humana, una célula pluripotente humana, una célula pluripotente de roedor, una célula pluripotente de ratón, una célula pluripotente de rata, una célula madre embrionaria (ES) de ratón, una célula ES de rata, una célula ES humana, una célula madre adulta humana, una célula progenitora humana restringida por el desarrollo, una célula madre pluripotente inducida (iPS) humana, o un embrión en etapa de una célula.

En algunos métodos, la primera proteína Cas es Cas9. En algunos métodos, la primera proteína Cas tiene actividad nucleasa en ambas cadenas de ADN bicatenario. En algunos métodos, la primera proteína Cas es una nickasa. Opcionalmente, la primera proteína Cas es una nickasa, y en donde el método comprende además poner en contacto el genoma con una segunda proteína Cas que es una nickasa, un tercer ARN de CRISPR que se hibrida con una tercera secuencia de reconocimiento de ARN de CRISPR, y un cuarto ARN de CRISPR que se hibrida con una cuarta secuencia

de reconocimiento de ARN de CRISPR, en donde la primera proteína Cas escinde una primera cadena de ADN genómico dentro de la primera secuencia de reconocimiento de ARN de CRISPR y dentro de la segunda secuencia de reconocimiento de ARN de CRISPR, y la segunda proteína Cas escinde una segunda cadena de ADN genómico dentro de la tercera secuencia de reconocimiento de ARN de CRISPR y dentro de la cuarta secuencia de reconocimiento de ARN de CRISPR.

En algunos métodos, el primer ARN de CRISPR y el ARNtracr se fusionan entre sí como un primer ARN guía (ARNg), y/o el segundo ARN de CRISPR y el ARNtracr se fusionan entre sí como un segundo ARNg. En algunos métodos, el primer ARN de CRISPR y el ARNtracr son moléculas de ARN separadas, y/o el segundo ARN de CRISPR y el ARNtracr son moléculas de ARN separadas.

En algunos métodos, el contacto comprende introducir la primera proteína Cas, el primer y segundo ARN de CRISPR, y el ARNtracr en la célula. En algunos métodos, la primera proteína Cas se introduce en la célula en forma de una proteína, un ARN mensajero (ARNm) que codifica la primera proteína Cas, o un ADN que codifica la primera proteína Cas. Opcionalmente, el ADN que codifica la primera proteína Cas se une operativamente a un primer promotor en un primer constructo de expresión, en donde el primer promotor es activo en la célula. En algunos métodos, el primer ARN de CRISPR se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el primer ARN de CRISPR. Opcionalmente, el ADN que codifica el primer ARN de CRISPR se une operativamente a un segundo promotor en un segundo constructo de expresión, en donde el segundo promotor es activo en la célula. En algunos métodos, el segundo ARN de CRISPR se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el segundo ARN de CRISPR. Opcionalmente, el ADN que codifica el segundo ARN de CRISPR se une operativamente a un tercer promotor en un tercer constructo de expresión, en donde el tercer promotor es activo en la célula. En algunos métodos, el ARNtracr se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el ARNtracr. Opcionalmente, el ADN que codifica el ARNtracr se une operativamente a un cuarto promotor en un cuarto constructo de expresión, en donde el cuarto promotor es activo en la célula. Opcionalmente, el primer, segundo, tercer, y/o cuarto constructos de expresión son componentes de una sola molécula de ácido nucleico.

Opcionalmente, el ADN que codifica la primera proteína Cas se une operativamente a un primer promotor en un primer constructo de expresión; los ADN que codifican el primer ARN de CRISPR y el ARNtracr se fusionan entre sí en un ADN que codifica un primer ARN guía (ARNg) y se unen operativamente a un segundo promotor en un segundo constructo de expresión; y/o los ADN que codifican el segundo ARN de CRISPR y el ARNtracr se fusionan entre sí en un ADN que codifica un segundo ARNg y se unen operativamente a un tercer promotor en un tercer constructo de expresión; en donde el primer, segundo, y tercer promotores son activos en la célula. Opcionalmente, el primer, segundo, y/o tercer constructos de expresión son componentes de una sola molécula de ácido nucleico.

Opcionalmente, la primera proteína Cas, el primer ARN de CRISPR, y el ARNtracr se introducen en la célula como un primer complejo de ARN y proteína, y/o la primera proteína Cas, el segundo ARN de CRISPR, y el ARNtracr se introducen en la célula como un segundo complejo de ARN y proteína.

En algunos métodos, la célula se ha modificado para disminuir la unión de extremos no homólogos (NHEJ) y/o para aumentar la conversión génica o la reparación dirigida por homología (HDR). Opcionalmente, la célula se ha modificado para disminuir la expresión y/o la actividad de uno o más de los siguientes: ADN-PK, PARP1, y ligasa IV. Opcionalmente, la disminución de la expresión o la actividad es inducible, reversible, temporalmente específica, y/o espacialmente específica.

En algunos métodos, el primer alelo comprende una mutación. Opcionalmente, la mutación es una modificación dirigida. En algunos métodos, el primer alelo es un alelo de tipo silvestre, y el locus correspondiente en el segundo cromosoma homólogo comprende una mutación.

La descripción proporciona además métodos que identifican la inserción dirigida de un inserto de ácido nucleico en un locus genómico diana en una célula diploide que no es un embrión en etapa de una célula, que comprenden: (a) obtener ADN de la célula, en donde la célula se ha puesto en contacto con un vector de transformación grande (LTVEC) que comprende el inserto de ácido nucleico flanqueado por un primer brazo de homología que se hibrida con una primera secuencia diana y un segundo brazo de homología que se hibrida con una segunda secuencia diana, en donde el inserto de ácido nucleico comprende un casete de selección adyacente al primer brazo de homología; (b) exponer el ADN de la célula a una sonda que se une dentro de la primera secuencia diana, una sonda que se une dentro del inserto de ácido nucleico, y una sonda que se une dentro de un gen de referencia que tiene un número de copias conocido, en donde cada sonda genera una señal detectable tras la unión; (c) detectar las señales de la unión de cada una de las sondas; y (d) comparar la señal de la sonda del gen de referencia con la señal de la sonda de la primera secuencia diana para determinar un número de copias de la primera secuencia diana, y comparar la señal de la sonda del gen de referencia con la señal de la sonda del inserto de ácido nucleico para determinar un número de copias del inserto de ácido nucleico, en donde un número de copias del inserto de ácido nucleico de uno o dos y un número de copias de la primera secuencia diana de dos indica la inserción dirigida del inserto de ácido nucleico en el locus genómico diana, y en donde un número de copias del inserto de ácido nucleico de uno o más y un número de copias de la primera secuencia diana de tres o más indica una inserción aleatoria del inserto de ácido nucleico en un locus genómico distinto al locus genómico diana.

- En algunos métodos, la señal de la unión de la sonda de la primera secuencia diana se usa para determinar un valor de ciclo umbral (Ct) de la primera secuencia diana, la señal de la unión de la sonda del gen de referencia se usa para determinar un valor de ciclo umbral (Ct) del gen de referencia, y el número de copias de la primera secuencia diana se determina mediante la comparación del valor de Ct de la primera secuencia diana y el valor de Ct del gen de referencia.
- En algunos métodos, la señal de la unión de la sonda del inserto de ácido nucleico se usa para determinar un valor de ciclo umbral (Ct) del inserto de ácido nucleico, y el número de copias del inserto de ácido nucleico se determina mediante la comparación del valor de Ct de la primera secuencia diana y el valor de Ct del gen de referencia.
- En algunos métodos, el casete de selección comprende un gen de resistencia a fármaco.
- En algunos métodos, el inserto de ácido nucleico es de al menos 5 kb, al menos 10 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 150 kb, al menos 200 kb, al menos 250 kb, al menos 300 kb, al menos 350 kb, al menos 400 kb, al menos 450 kb, o al menos 500 kb. En algunos métodos, la distancia entre las secuencias a las que se unen las sondas en la primera secuencia diana y el casete de selección es de no más de 100 nucleótidos, 200 nucleótidos, 300 nucleótidos, 400 nucleótidos, 500 nucleótidos, 600 nucleótidos, 700 nucleótidos, 800 nucleótidos, 900 nucleótidos, 1 kb, 1,5 kb, 2 kb, 2,5 kb, 3 kb, 3,5 kb, 4 kb, 4,5 kb, o 5 kb.
- Algunos métodos comprenden además determinar el número de copias de la segunda secuencia diana. Opcionalmente, la etapa (b) comprende además exponer el ADN de la célula a una sonda que se une a la segunda secuencia diana, la etapa (c) comprende además detectar la señal de la unión de la sonda de la segunda secuencia diana, y la etapa (d) comprende además comparar la señal de la sonda del gen de referencia con la señal de la sonda de la segunda secuencia diana para determinar un número de copias de la segunda secuencia diana.
- Algunos métodos comprenden además determinar el número de copias de una o más secuencias adicionales dentro del inserto de ácido nucleico. Opcionalmente, la etapa (b) comprende además exponer el ADN de la célula a una o más sondas adicionales que se unen al inserto de ácido nucleico, la etapa (c) comprende además detectar la señal de la unión de la una o más sondas adicionales, y la etapa (d) comprende además comparar la señal de la sonda del gen de referencia con la señal de la una o más sondas del inserto de ácido nucleico para determinar los números de copias de la una o más secuencias adicionales dentro del inserto de ácido nucleico. Opcionalmente, la una o más secuencias adicionales dentro del inserto de ácido nucleico comprenden una secuencia adyacente a la segunda secuencia diana.
- En algunos métodos, el LTVEC se diseña para eliminar una secuencia endógena del locus genómico diana, o la célula se ha puesto en contacto además con una proteína Cas, un primer ARN de CRISPR que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR dentro de un locus genómico diana, un segundo ARN de CRISPR que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana, y un ARNtracr. Opcionalmente, tales métodos comprenden además determinar el número de copias de las secuencias endógenas en el locus genómico diana. Opcionalmente, la etapa (b) comprende además exponer el ADN de la célula a una sonda que se une a la secuencia endógena en el locus genómico diana, la etapa (c) comprende además detectar la señal de la unión de la sonda de la secuencia endógena, y la etapa (d) comprende además comparar la señal de la sonda del gen de referencia con la señal de la sonda de la secuencia endógena para determinar un número de copias de la secuencia endógena.
- Breve descripción de las figuras
- La Figura 1 muestra una representación esquemática de la delección simultánea del ectodominio Lrp5 de ratón y su reemplazo con una versión humana correspondiente LRP5 mediante el uso de un LTVEC y uno o dos ARNg para la región 5' (A, B, B2), la región media (C, D), y la región 3' (E2, E, F). El LTVEC se muestra en la porción superior de la figura, y el locus del gen Lrp5 de ratón se muestra en la porción inferior de la figura. Las posiciones de los sitios de escisión de Cas9 guiados por los ocho ARN guías se indican con las flechas verticales debajo de la secuencia del gen de ratón. Las flechas horizontales representan cebadores de PCR para las secuencias de ratón y humano.
- La Figura 2A muestra una representación esquemática general de la delección simultánea de un gen de ratón y su reemplazo con una versión humana correspondiente mediante el uso de un LTVEC y dos ARN guías (ARN guías A y B). El LTVEC se muestra en la porción superior de la Figura 2A, y el locus del gen de ratón se muestra en la porción inferior de la Figura 2A. Las posiciones de los sitios de escisión de Cas9 guiados por los dos ARN guías se indican con las flechas debajo de la secuencia del gen de ratón.
- Las Figuras 2B-E muestran las modificaciones bialélicas únicas (tipos de alelos) que ocurren a una frecuencia mayor cuando se usan dos ARN guías. Las líneas gruesas con rayas diagonales indican el gen de ratón, las líneas de puntos indican delecciones en el gen de ratón, y las líneas negras gruesas indican la inserción del gen humano. La Figura 2B muestra alelos homocigóticos colapsados (delección grande inducida por CRISPR). La Figura 2C muestra alelos con transformación homocigótica. La Figura 2D muestra alelos con transformación hemocigótica. La Figura 2E muestra alelos heterocigóticos combinados.

La Figura 3A y 3B muestran ensayos de PCR que confirman los genotipos de clones seleccionados. La Figura 3A muestra los resultados de ensayos de PCR de largo alcance de clones de células ES seleccionados mediante el uso de los cebadores m-lr-f y m-5'-r, que establecen la unión entre el inserto humano y las secuencias fuera de las homólogas al brazo de homología 5', para demostrar de este modo la transformación correcta. La Figura 3B muestra los resultados de los ensayos de PCR 5' Del J, 5' Ins J, Del A + F, y Del A + E2. 5' Del J representa los productos de PCR mediante el uso de los cebadores m-5'-f y m-5'-r, que amplifican la secuencia de tipo silvestre que rodea el sitio de escisión del ARNg A para establecer la retención o la pérdida de esta secuencia. 5' Ins J representa los productos de PCR mediante el uso de los cebadores m-5'-f y h-5'-r, que establecen una unión entre el inserto humano y el genoma de ratón. Este ensayo proporcionará un resultado positivo en clones con integración tanto dirigida como aleatoria. Del A + F representa el tamaño de amplicón esperado (359 pb) y las bandas reales para la delección grande mediada por la escisión por el par de ARNg A y F en los clones BO-F10 y AW-A8. Del A + E2 representa la misma idea para el clon BA-A7. NT indica sin plantilla, +/- indica control de tipo silvestre de células ES híbridas VGF1 progenitoras, H/+ indica genotipo humanizado heterocigótico, H/ Δ indica genotipo humanizado hemocigótico, H/H indica genotipo humanizado homocigótico, y Δ/Δ indica genotipo con delección homocigótica.

La Figura 4A-C muestra el análisis de hibridación fluorescente *in situ* (FISH) de los clones de células ES de ratón AW-D9 (Figura 4A) y BA-D5 (Figura 4C), que se transformaron con el LTVEC de humanización de Lrp5 combinado con Cas9 y dos ARNg, y el clon BS-C4 (Figura 4B), que se transformó con el LTVEC solo. Las flechas indican las posiciones de las señales de hibridación en la banda B del cromosoma 19. Una señal roja indica la hibridación solamente con la sonda de ratón (flecha discontinua, Figura 4B). Una señal de color amarillo mixto indica la hibridación tanto con la sonda de ratón roja como con la sonda humana verde. Una banda B de cromosoma 19 que tiene una señal roja (flecha discontinua) y la otra banda B de cromosoma 19 que tiene una señal amarilla (flecha sólida) confirmaron la transformación del locus correcto y el genotipo heterocigótico para el clon BS-C4 (Figura 4B). Las bandas B de ambos cromosomas 19 que tienen una señal amarilla (flechas sólidas, Figuras 4A y 4C) confirmaron la transformación del locus correcto y los genotipos homocigóticos para los clones AW-D9 y BS-C4.

La Figura 5 muestra una representación esquemática del cromosoma 19 con ensayos diseñados para examinar los eventos de conversión génica o recombinación mitótica mediados por dos ARN guías mediante el análisis de la pérdida de heterocigosidad (LOH) en células ES híbridas VGF1. Las posiciones aproximadas de los ensayos de número de copias cromosómicas (CCN) por qPCR TaqMan® se muestran con flechas. Las posiciones aproximadas de los ensayos de PCR de polimorfismo de variantes estructurales (SV) se muestran con cheurones con sus distancias (en Mb) del locus Lrp5 proporcionadas anteriormente. Las posiciones aproximadas de los ensayos de discriminación alélica TaqMan® de variantes de nucleótido simple (SNV) se muestran con puntas de flecha con sus distancias (en Mb) del locus de Lrp5 proporcionadas debajo. Las posiciones de los sitios de reconocimiento de ARNg para F, E2, D, B2, y A se muestran con flechas diagonales encima de la representación del gen Lrp5.

La Figura 6 muestra una representación esquemática de la delección simultánea de la región desde el exón 2 hasta el codón de terminación del gen C5 (Hc) de ratón y su reemplazo con una versión humana correspondiente C5 mediante el uso de un LTVEC y uno o dos ARNg para la región 5' (A, B), la región media (C, D), y la región 3' (E, E2). El LTVEC se muestra en la porción superior de la figura, y el locus del gen C5 (Hc) de ratón se muestra en la porción inferior de la figura. Las posiciones de los sitios de escisión de Cas9 guiados por los seis ARN guías se indican con las flechas debajo de la secuencia del gen de ratón.

Las Figuras 7A y 7B muestran el análisis de hibridación fluorescente *in situ* (FISH) de los clones de células ES de ratón Q-E9 (Figura 7A) y O-E3 (Figura 7B), que se transformaron con el LTVEC de humanización de Hc combinado con Cas9 y dos ARNg. Las flechas indican las posiciones de las señales de hibridación en la banda B de cromosoma 2. Una señal roja indica la hibridación solamente con la sonda de ratón (flecha discontinua, Figura 7A). Una señal de color amarillo mixto indica la hibridación tanto con la sonda de ratón roja como con la sonda humana verde (flecha sólida). Una banda B de cromosoma 2 que tiene una señal roja (flecha discontinua) y la otra banda B de cromosoma 2 que tiene una señal amarilla (flecha sólida) confirmaron la transformación del locus correcto y el genotipo heterocigótico para el clon Q-E9 (Figura 7A). Las bandas B de ambos cromosomas 2 que tienen una señal amarilla (flechas sólidas, Figura 7B) confirmaron la transformación del locus correcto y el genotipo homocigótico para el clon O-E3.

La Figura 8 muestra una representación esquemática de la delección simultánea del gen Ror1 de ratón y su reemplazo con una versión humana correspondiente ROR1 mediante el uso de un LTVEC y uno o dos ARNg para la región 5' (A, B), la región media (D, C), y la región 3' (E, F). El LTVEC se muestra en la porción superior de la figura, y el locus del gen Ror1 de ratón se muestra en la porción inferior de la figura. Las posiciones de los sitios de escisión de Cas9 guiados por los seis ARN guías se indican con las flechas debajo de la secuencia del gen de ratón.

La Figura 9 muestra una representación esquemática de la delección simultánea del gen Trpa1 de ratón y su reemplazo con una versión humana correspondiente TRPA1 mediante el uso de un LTVEC y uno o dos ARNg para la región 5' (A, A2, B), la región media (C, D), y la región 3' (E2, E, F). El LTVEC se muestra en la porción superior de la figura, y el locus del gen Trpa1 de ratón se muestra en la porción inferior de la figura. Las posiciones de los sitios de escisión de Cas9 guiados por los ocho ARN guías se indican con las flechas debajo de la secuencia del gen de ratón.

La Figura 10A-E muestra los resultados de los ensayos de variación estructural (SV) de los clones BR-B4, BP-G7, BO-G11, BO-F10, BO-A8, y BC-H9, mediante el uso de ADN de VGF1 (F1H4), 129, y B6 como controles. Los ensayos se

realizaron a las siguientes distancias teloméricas al locus de Lrp5: 13,7 Mb (Figura 10A), 20,0 Mb (Figura 10B), 36,9 Mb (Figura 10C), 48,3 Mb (Figura 10D), y 56,7 Mb (Figura 10E). Las posiciones de los productos de PCR para los alelos B6 y 129 se muestran con las flechas.

5 La Figura 11A-C muestra gráficos de discriminación alélica para las 0,32 Mb centroméricas de Lrp5 (Figura 11A), 1,2 Mb teloméricas de Lrp5 (Figura 11B), y 57,2 Mb teloméricas de Lrp5 (Figura 11C). Los valores en cada eje representan la intensidad de fluorescencia relativa. Los gráficos representan cuatro réplicas para cada muestra, que se muestran como puntos sólidos (alelo B6), puntos abiertos (alelo 129), y puntos con líneas diagonales (ambos alelos B6/129).

10 La Figura 12A-C es una representación esquemática que muestra un posible mecanismo de recombinación mitótica durante la fase G2 del ciclo celular que puede producir eventos homocigóticos y conversión génica de amplio alcance detectados por la pérdida de heterocigosidad. La Figura 12A muestra cromosomas homólogos replicados que muestran las dos cromátidas en una célula ES 129/B6 híbrida heterocigótica para una humanización dirigida en el homólogo 129. Las flechas de dos puntas indican las rupturas bicatenarias potenciales generadas por la escisión por Cas9 dirigida por dos ARNg que promueven el intercambio recíproco por recombinación homóloga entre las cromátidas en cromosomas homólogos, mostrado como un entrecruzamiento en el lado centromérico del alelo diana, lo que da como resultado las cromátidas híbridas que se muestran en la Figura 12B. La Figura 12C muestra que después de la mitosis y la división celular, son posibles cuatro tipos de segregación de los cromosomas a las células hijas. Dos con retención de la heterocigosidad, un heterocigótico del tipo progenitor (Hum/+, porción superior izquierda) y un heterocigótico por intercambio equivalente (Hum/+, porción superior derecha), no pueden distinguirse por los ensayos LOH. Otros dos muestran pérdida de heterocigosidad, un homocigótico humanizado (Hum/Hum, por ejemplo el clon BO-A8, porción inferior izquierda) con pérdida de alelos B6 teloméricos y un homocigótico de tipo silvestre (+/+, porción inferior derecha) con pérdida de alelos 129 teloméricos. Este último tipo se perderá debido a que no retiene el casete de resistencia a fármaco del alelo humanizado.

25 La Figura 13 muestra una representación esquemática de la delección simultánea de la región desde el exón 2 hasta el codón de terminación del gen C5 (Hc) de ratón y su reemplazo con una versión humana correspondiente C5 mediante el uso de un vector de transformación con tamaños de los brazos de homología de 35 kb y 31 kb (LTVEC) o un vector de transformación con tamaños de los brazos de homología de 5 kb cada uno (STVEC) y uno o dos ARNg para la región 5' (A, B), la región media (C, D), y la región 3' (E, E2). Los dos vectores de transformación se muestran en la porción superior de la figura, y el locus del gen C5 (Hc) de ratón se muestra en la porción inferior de la figura. Las posiciones de los sitios de escisión de Cas9 guiados por los seis ARN guías se indican con las flechas verticales debajo de la secuencia del gen de ratón, y los cebadores usados para el tamizaje se indican con flechas horizontales. Las posiciones de los ensayos de ganancia de alelos (GOA) que cuantifican el número de copias del inserto y los ensayos de pérdida de alelos (LOA) que cuantifican la secuencia de ratón reconocida para la delección se indican con los triángulos.

30 La Figura 14 muestra una representación esquemática de la delección simultánea de los primeros cinco exones del gen Cmah de ratón y su reemplazo con un reportero lacZ y un casete de selección de resistencia a higromicina mediante el uso de un LTVEC y dos ARNg de la región 5' (A, B). El LTVEC se muestra en la porción superior de la figura, y el locus del gen Cmah de ratón se muestra en la porción inferior de la figura. Las posiciones de los sitios de escisión de Cas9 guiados por los dos ARN guías se indican con las flechas verticales debajo de la secuencia del gen de ratón, y las posiciones de los ensayos GOA que cuantifican el número de copias del inserto y los ensayos LOA que cuantifican la secuencia de ratón reconocida para la delección se indican con los triángulos.

45 La Figura 15 muestra una representación esquemática de los eventos de escisión y el producto de supresión producido (SEQ ID NO: 112) cuando el locus del gen Cmah de ratón (SEQ ID NO: 109) se reconoce con dos ARNg para la región 5' (A y B; SEQ ID NOS: 107 y 108, respectivamente). Las secuencias de ARNg hibridadas al locus del gen Cmah están en negritas, las proteínas Cas9 se representan mediante los óvalos punteados, los sitios de escisión de Cas9 se indican con las flechas verticales, y los motivos adyacentes al protoespaciador (PAM) están en recuadros. Las posiciones aproximadas del cebador directo, la sonda, y el cebador inverso del ensayo LOA TaqMan® se indican con las barras horizontales y las flechas en la parte superior de la figura. Los fragmentos 5' y 3' producidos después de la escisión y supresión son las SEQ ID NOS: 110 y 111, respectivamente.

50 Las Figuras 16A-E muestran posibles mecanismos que explican los resultados observados, que incluyen la pérdida de heterocigosidad (LOH), en experimentos de humanización asistida por CRISPR/Cas9 en células ES de ratón híbridas F1 que tienen un complemento cromosómico haploide derivado de la cepa de ratón 129S6/SvEvTac y un complemento cromosómico haploide derivado de la cepa de ratón C57BL/6NTac (B6). La Figura 16A muestra el intercambio recíproco de cromátidas por entrecruzamiento mitótico donde ocurre una modificación heterocigótica en el cromosoma 129 antes de la replicación del genoma o después de la replicación del genoma seguido de la conversión génica entre cromátidas hermanas. La Figura 16B muestra el intercambio recíproco de cromátidas por entrecruzamiento mitótico donde una sola cromátida del 129 se modifica después de la replicación del genoma. La Figura 16C muestra el intercambio recíproco de cromátidas por entrecruzamiento mitótico donde no ha ocurrido transformación por LTVEC, pero ha ocurrido escisión por Cas9 en el cromosoma 129 o el B6 (se muestra la escisión de B6). La Figura 16D muestra la copia de cromátidas mediante la replicación inducida por ruptura donde ocurre una modificación heterocigótica en el cromosoma 129 antes de la replicación del genoma o después de la replicación del genoma seguido de la conversión génica entre cromátidas hermanas. La Figura 16E muestra la copia de cromátidas mediante la replicación inducida por ruptura donde una sola

cromátida 129 se modifica después de la replicación del genoma. La Figura 16F muestra la copia de cromátidas mediante la replicación inducida por ruptura donde no ha ocurrido transformación por LTVEC, pero ha ocurrido escisión por Cas9 en el cromosoma 129 o el B6 (se muestra la escisión de B6).

5 La Figura 17A-C muestra estrategias de tamizaje de las modificaciones dirigidas. La Figura 17A muestra una estrategia de tamizaje de modificación de alelo (MOA) estándar para detectar la transformación heterocigótica por un vector de transformación grande (LTVEC) en la que una secuencia endógena en un cromosoma de ratón se elimina y se reemplaza con un inserto Neo-SDC. La estrategia usa las sondas TaqMan® mTU y mTD contra regiones corriente arriba y corriente abajo de la secuencia endógena reconocida para la delección. La Figura 17B muestra el uso de ensayos de retención TaqMan® (sondas retU y retD) en combinación con ensayos de modificación de alelo (MOA) (sondas mTGU, mTM, y mTGD para el ensayo de pérdida de alelo (LOA), y sondas hTU y hTD para el ensayo de ganancia de alelo (GOA)) para tamizar la humanización asistida por CRISPR/Cas9. La Figura 17C muestra el uso de ensayos de retención TaqMan® (sondas retU y retD) en combinación con ensayos de pérdida de alelo (LOA) (sondas mTGU, mTM, y mTGD)) para tamizar las delecciones asistidas por CRISPR/Cas9 mediante el uso de pares de ARN guías (gU y gD).

15 La Figura 18 muestra representaciones esquemáticas (no están a escala) de una región de aproximadamente 900 kb de un locus de cadena pesada de inmunoglobulina de ratón con segmentos génicos de región variable reemplazados con sus equivalentes humanos (triángulos) y un vector de transformación con un inserto P_{gk}-Neo (promotor de la fosfoglicerato quinasa I unido operativamente al gen de neomicina fosfotransferasa) flanqueado por sitios loxP. Se usan dos ARNg para escindir el locus de cadena pesada de inmunoglobulina de ratón en el extremo 5' y se usan dos ARNg para escindir el locus en el extremo 3', y el vector de transformación elimina y reemplaza el locus de cadena pesada de inmunoglobulina de ratón con el inserto P_{gk}-Neo. Las posiciones de los sitios de escisión de Cas9 guiados por los cuatro ARN guías se indican con las flechas verticales debajo del locus diana. Las líneas horizontales circuladas representan las sondas TaqMan® para los ensayos de modificación de alelo (MOA) (hlgH31, hlgH1, mlgHA1, mlgHA7, y hlgH9) y los ensayos de retención (5' IgH Brazo 1, 5' IgH Brazo 2, mlgM-398, y mlgM-1045).

Definiciones

30 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 químicamente o bioquímicamente modificados o derivatizados. Los términos incluyen además polímeros que se han modificado, tales como polipéptidos que tienen cadenas principales peptídicas modificadas.

35 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. Estos incluyen ADN o ARN monocatenario, bicatenario, y de múltiples cadenas, ADN genómico, ADNc, híbridos ADN-ARN, y polímeros que comprenden bases purínicas, bases pirimidínicas, u otras bases nucleotídicas naturales, químicamente modificadas, bioquímicamente modificadas, no naturales, o derivatizadas.

40 "Optimización de 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 mayor frecuencia o con la máxima frecuencia en los genes de la célula huésped mientras se mantiene la secuencia de aminoácidos nativa. Por ejemplo, un ácido nucleico que codifica una proteína Cas puede modificarse para sustituir los codones que tienen una mayor frecuencia de uso en una célula procarionota o eucariota dada, que incluye una célula bacteriana, una célula de levadura, una célula humana, una célula no humana, una célula de mamífero, una célula de roedor, una célula de ratón, una célula de rata, una célula de hámster, o cualquier otra célula huésped, en comparación con la secuencia de ácido nucleico de origen natural. Las tablas de uso de codones están fácilmente disponibles, por ejemplo, en la "Base de datos de uso de codones." Estas tablas pueden adaptarse en una serie de maneras. Ver Nakamura y otros (2000) *Nucleic Acids Research* 28:292. Se dispone además de algoritmos informáticos para la optimización de codones de una secuencia particular para la expresión en un huésped particular (ver, por ejemplo, Gene Forge).

55 "Unión operativa" o estar "unido operativamente" incluye la yuxtaposición de dos o más componentes (por ejemplo, un promotor y otro elemento de secuencia) de manera que ambos componentes funcionen normalmente y permitan la posibilidad de que al menos uno de los componentes pueda mediar una función que se ejerce sobre al menos uno de los otros componentes. Por ejemplo, un promotor puede unirse operativamente a una secuencia codificante si el promotor controla el nivel de transcripción de la secuencia codificante en respuesta a la presencia o ausencia de uno o más factores reguladores de la transcripción.

60 "Complementariedad" de ácidos nucleicos significa que una secuencia de nucleótidos en una cadena de ácido nucleico, debido a la orientación de sus grupos de nucleobases, forma enlaces de hidrógeno con otra secuencia en una cadena de ácido nucleico opuesta. Las bases complementarias en el ADN son típicamente A con T y C con G. En el ARN, son típicamente C con G y U con A. La complementariedad puede ser perfecta o sustancial/suficiente. La complementariedad perfecta entre dos ácidos nucleicos significa que los dos ácidos nucleicos pueden formar un dúplex en el que cada base en el dúplex se enlaza a una base complementaria mediante apareamiento de Watson-Crick. Complementariedad "sustancial" o "suficiente" significa que una secuencia en una cadena no es completamente y/o perfectamente

complementaria a una secuencia en una cadena opuesta, sino que ocurre el enlace entre bases en las dos cadenas suficiente para formar un complejo híbrido estable en un conjunto de condiciones de hibridación (por ejemplo, concentración de sales y temperatura). Tales condiciones pueden predecirse mediante el uso de las secuencias y cálculos matemáticos estándar para predecir la T_m (temperatura de fusión) de las cadenas hibridadas, o por determinación empírica de la T_m mediante el uso de métodos de rutina. La T_m incluye la temperatura a la cual se desnaturaliza el 50 % de una población de complejos de hibridación formados entre dos cadenas de ácidos nucleicos. A una temperatura por debajo de la T_m , se favorece la formación de un complejo de hibridación, mientras que a una temperatura por encima de la T_m , se favorece la fusión o separación de las cadenas en el complejo de hibridación. La T_m puede calcularse para un ácido nucleico que tiene un contenido de G+C conocido 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 de T_m conocidos tienen en cuenta las características estructurales del ácido nucleico.

"Condición de hibridación" incluye el entorno acumulado en el que una cadena de ácido nucleico se enlaza a una segunda cadena de ácido nucleico mediante interacciones entre cadenas 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*, 2.sup.nd ed., pp. 1.90-1.91, 9.47-9.51, 1 1.47-11.57 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque son posibles los errores de apareamiento entre bases. Las condiciones adecuadas 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. Mientras mayor sea el grado de complementación entre dos secuencias de nucleótidos, mayor será el valor de la temperatura de fusión (T_m) de los híbridos de los ácidos nucleicos que tienen esas secuencias. Para las hibridaciones entre ácidos nucleicos con tramos cortos de complementariedad (por ejemplo complementariedad de 35 o menos, 30 o menos, 25 o menos, 22 o menos, 20 o menos, o 18 o menos nucleótidos) la posición de los errores de apareamiento cobra importancia (ver Sambrook y otros, más arriba, 11.7-11.8). Típicamente, la longitud de un ácido nucleico hibridable es de al menos aproximadamente 10 nucleótidos. Las longitudes mínimas ilustrativas de 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 sales 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.

No es necesario que la secuencia del polinucleótido sea 100 % complementaria a la de su ácido nucleico diana para que sea específicamente hibridable. Por otra parte, un polinucleótido puede hibridarse con uno o más segmentos de manera que los segmentos intermedios o adyacentes no estén involucrados en el evento de hibridación (por ejemplo, una estructura de lazo o una estructura en 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 con una región diana dentro de la secuencia de ácido nucleico diana a la que se encuentra orientado. Por ejemplo, un ARNg en el que 18 de 20 nucleótidos son complementarios a una región diana, y por lo tanto se hibridaría específicamente, representaría un 90 % complementariedad. En este ejemplo, los nucleótidos no complementarios restantes pueden agruparse o intercalarse con nucleótidos complementarios y no es necesario que sean contiguos entre sí o a nucleótidos complementarios.

El porcentaje de complementariedad entre tramos particulares de secuencias de ácidos nucleicos dentro de los ácidos nucleicos puede determinarse de manera rutinaria mediante el uso de programas BLAST (herramientas de búsqueda de alineamiento local básico) 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 (Paquete de análisis de secuencias Wisconsin, Versión 8 para Unix, Genetics Computer Group, University Research Park, Madison Wis.), mediante el uso de ajustes predeterminados, 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. Se reconoce a lo largo de la descripción que algunos componentes pueden tener variantes y fragmentos activos. Tales componentes incluyen, por ejemplo, proteínas Cas, ARN de CRISPR, ARNtracr, y ARN guías. La actividad biológica de cada uno de estos componentes se describe en otra parte en la presente descripción.

"Identidad de secuencia" o "identidad" en el contexto de dos secuencias de polinucleótidos o polipéptidos hace referencia a los residuos en las dos secuencias que son iguales cuando se alinean para su correspondencia máxima en una ventana de comparación especificada. Cuando el porcentaje de identidad de secuencia se usa en referencia a proteínas se reconoce que las posiciones de los residuos que no son idénticas difieren frecuentemente por sustituciones de aminoácidos conservativas, 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 conservativa de la sustitución. Se dice que las secuencias que difieren por tales sustituciones conservativas tienen "similitud de secuencia" o "similitud". Los medios para hacer este ajuste son bien conocidos para los expertos en la técnica. Típicamente, esto involucra la calificación de una sustitución conservativa como un error de apareamiento parcial en lugar de total, lo que aumenta de este modo el porcentaje de identidad de secuencia. Por lo tanto, por ejemplo, cuando a un aminoácido idéntico se le da una puntuación de 1 y a una sustitución no conservativa se le da una puntuación de cero, a una sustitución conservativa se le da una puntuación entre cero y 1. La puntuación de las sustituciones conservativas se calcula, por ejemplo, como se implementa en el programa PC/GENE (Intelligenetics, Mountain View, California).

El "porcentaje de identidad de secuencia" incluye el valor determinado mediante la comparación de dos secuencias óptimamente alineadas en una ventana de comparación, en donde la porción de la secuencia de polinucleótido 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 deleciones) para el alineamiento óptimo de las dos secuencias. El porcentaje se calcula mediante la determinación del número de posiciones en las que la base de ácido nucleico o el residuo de aminoácido idéntico aparece 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 producir el porcentaje de identidad de secuencia.

A menos que se indique de otra manera, los valores de identidad/similitud de secuencia incluyen el valor obtenido mediante el uso del programa GAP Versión 10 que usa los siguientes parámetros: % identidad y % similitud de una secuencia de nucleótidos mediante el uso de Peso de GAP de 50 y Peso de Longitud de 3, y la matriz de puntuación nwsgapdna.cmp; % identidad y % similitud de una secuencia de aminoácidos mediante el uso de Peso de GAP de 8 y Peso de Longitud de 2, y la matriz de puntuación BLOSUM62; o cualquier programa equivalente de este. "Programa equivalente" incluye cualquier programa de comparación de secuencias que, para dos secuencias cualesquiera en cuestión, genera un alineamiento que tiene coincidencias de nucleótidos o residuos de aminoácidos idénticos y un porcentaje de identidad de secuencia idéntico en comparación con el alineamiento correspondiente generado por el programa GAP Versión 10.

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

Composiciones o métodos "que comprenden" o "que incluyen" uno o más elementos mencionados pueden incluir otros elementos no mencionados específicamente. 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 números enteros dentro de o que definen el intervalo, y todos los subintervalos definidos por los números enteros dentro del intervalo.

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

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

Descripción detallada

I. Ejemplo general

Se proporcionan métodos y composiciones para modificar un genoma dentro de una célula. Los métodos y composiciones emplean sistemas CRISPR/Cas que usan dos ARN guías (ARNg) que reconocen sitios diferentes dentro de un solo locus genómico diana. Por ejemplo, los métodos y composiciones pueden emplear sistemas CRISPR/Cas que usan los dos ARN guías (ARNg) para crear pares de rupturas bicatenarias en sitios diferentes dentro de un solo locus genómico diana. Alternativamente, los métodos y composiciones pueden emplear sistemas CRISPR/Cas que usan los dos ARN guías (ARNg) para crear pares de rupturas monocatenarias en sitios diferentes dentro de un solo locus genómico diana. En algunos métodos, dos o más ARN guías (por ejemplo, tres o cuatro) pueden usarse, por ejemplo, para crear dos o más rupturas monocatenarias o rupturas bicatenarias en sitios diferentes dentro de un solo locus genómico diana.

Algunos métodos promueven modificaciones genéticas bialélicas y comprenden el colapso del genoma, de manera que una secuencia de ácido nucleico grande se elimina de un cromosoma entre dos sitios de escisión. Otros métodos promueven modificaciones genéticas bialélicas y comprenden la deleción simultánea de una secuencia de ácido nucleico dentro de la célula y su reemplazo con una secuencia de ácido nucleico exógena. Como se describe en más detalle a continuación, estos métodos que usan dos ARNg aumentan la eficiencia de generación de células o animales con modificaciones genéticas dirigidas bialélicas al promover la generación de tales células y animales en una sola etapa de

transformación. Por consiguiente, el número de animales y crías necesarios para generar un animal con una modificación genética dirigida bialélica se reduce.

Otros métodos comprenden la conversión génica o la pérdida de heterocigosidad, de manera que un genoma que es heterocigótico para un alelo se modifica para convertirlo en homocigótico para el alelo por medio de la escisión en sitios determinados por dos ARNg en el alelo correspondiente en un cromosoma homólogo correspondiente. Como se describe en más detalle a continuación, el uso de dos ARNg en estos métodos aumenta la frecuencia de conversión génica y hace posible la conversión génica en tramos grandes de ADN cromosómico.

II. Sistemas CRISPR/Cas

Los métodos y composiciones descritos en la presente descripción pueden utilizar sistemas de repeticiones palindrómicas cortas agrupadas y regularmente intercaladas (CRISPR)/proteína asociada a CRISPR (Cas) o componentes de tales sistemas para modificar un genoma dentro de una célula. Los sistemas CRISPR/Cas incluyen transcritos y otros elementos involucrados en la expresión de, o el control de la actividad de, los genes Cas. Un sistema CRISPR/Cas puede ser un sistema tipo I, tipo II, o tipo III. Los métodos y composiciones descritos en la presente descripción emplean sistemas CRISPR/Cas mediante la utilización de complejos CRISPR (que comprenden un ARN guía (ARNg) en complejo con una proteína Cas) para la escisión dirigida al sitio de ácidos nucleicos.

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 algo que indique la intervención de la mano del hombre, tal como la alteración o mutación de uno o más componentes del sistema respecto a su estado de origen natural, que esté al menos sustancialmente libre de al menos uno de otro componente con el que se asocia de manera natural en la naturaleza, o que se asocie con al menos uno de otro componente con el que no se asocia de manera natural. Por ejemplo, algunos sistemas CRISPR/Cas emplean complejos CRISPR de origen no natural que comprenden un ARNg y una proteína Cas que no aparecen juntos en la naturaleza.

A. Endonucleasas Cas guiadas por ARN

Las proteínas Cas generalmente comprenden al menos un dominio de reconocimiento o unión al ARN. Tales dominios pueden interactuar con los ARN guías (ARNg, descritos en más detalle a continuación). Las proteínas Cas pueden comprender además dominios nucleasa (por ejemplo, dominios DNasa o RNasa), dominios de unión al ADN, dominios helicasa, dominios de interacción proteína-proteína, dominios de dimerización, y otros dominios. Un dominio nucleasa posee actividad catalítica para la escisión de ácidos nucleicos. La escisión incluye la ruptura de los enlaces covalentes de una molécula de ácido nucleico. La escisión puede producir extremos romos o extremos escalonados, y puede ser monocatenaria o bicatenaria.

Los ejemplos de proteínas Cas incluyen Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e (CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9 (Csn1 o Csx12), Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (CasA), Cse2 (CasB), Cse3 (CasE), Cse4 (CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, y Cu1966, y homólogos o versiones modificadas de estas.

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 típicamente comparten cuatro motivos principales 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.*, *Staphylococcus aureus*, *Nocardiopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycoloides*, *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 arabicum*, *Ammonifex degensii*, *Caldicelulosiruptor becscii*, *Candidatus Desulforudis*, *Clostridium botulinum*, *Clostridium difficile*, *Finegoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsonii*, *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*. Ejemplos adicionales de los miembros de la familia de 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 esta. La secuencia de aminoácidos de una proteína Cas9 de *S. pyogenes* puede encontrarse, por ejemplo, en la base de datos SwissProt con el número de acceso Q99ZW2.

Las proteínas Cas pueden ser proteínas de tipo silvestre (es decir, las que aparecen en la naturaleza), proteínas Cas modificadas (es decir, variantes de proteínas Cas), o fragmentos de proteínas Cas de tipo silvestre o modificadas. Las proteínas Cas también pueden ser variantes o fragmentos activos de proteínas Cas de tipo silvestre o modificadas. Las variantes o fragmentos activos pueden comprender al menos 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o 100 % de la secuencia de aminoácidos de una proteína Cas de tipo silvestre o modificada.

97 %, 98 %, 99 % o más identidad de secuencia con 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 tanto retienen la actividad de inducción de mellas o de inducción de rupturas bicatenarias. Los ensayos de actividad de inducción de mellas o de inducción de rupturas bicatenarias se conocen y generalmente miden la actividad total y la especificidad de la proteína Cas sobre ADN sustratos que contienen el sitio de escisión.

Las proteínas Cas pueden modificarse para aumentar o disminuir su afinidad de unión al ácido nucleico, su especificidad de unión al ácido nucleico, y/o su 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 nucleasa de la proteína Cas pueden modificarse, eliminarse, o inactivarse, o una proteína Cas puede truncarse para eliminar dominios que no son esenciales para la función de la proteína o para optimizar (por ejemplo, mejorar o reducir) la actividad de la proteína Cas.

Algunas proteínas Cas comprenden al menos dos dominios nucleasa, tales como dominios DNasa. Por ejemplo, una proteína Cas9 puede comprender un dominio nucleasa similar a RuvC y un dominio nucleasa similar a HNH. Cada uno de los dominios RuvC y HNH puede cortar una cadena diferente de ADN bicatenario para hacer una ruptura bicatenaria en el ADN. Ver, por ejemplo, Jinek y otros (2012) Science 337:816-821.

Uno o ambos dominios nucleasa pueden eliminarse o mutarse de modo que ya no sean funcionales o que tengan actividad nucleasa reducida. Si uno de los dominios nucleasa se elimina o se muta, la proteína Cas resultante (por ejemplo, Cas9) puede referirse como una nickasa y puede generar una ruptura monocatenaria en una secuencia de reconocimiento de ARN de CRISPR dentro de un ADN bicatenario pero no una ruptura bicatenaria (es decir, puede escindir la cadena complementaria o la cadena no complementaria, pero no ambas). Si ambos dominios nucleasa se eliminan o se mutan, la proteína Cas resultante (por ejemplo, Cas9) tendrá una capacidad reducida de escindir ambas cadenas de un ADN bicatenario. 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*. Del mismo modo, H939A (histidina a alanina en la posición de aminoácido 839) o H840A (histidina a alanina en la posición de aminoácido 840) en el dominio HNH de Cas9 de *S. pyogenes* pueden convertir la 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. Tales mutaciones pueden generarse mediante el uso de métodos bien conocidos tales como mutagénesis dirigida al sitio, mutagénesis mediada por PCR, o síntesis génica total. Los ejemplos de otras mutaciones que crean nickasas pueden encontrarse, por ejemplo, en los documentos WO/2013/176772A1 y WO/2013/142578A1.

Las proteínas Cas también pueden ser proteínas de fusión. Por ejemplo, una proteína Cas puede fusionarse a un dominio de escisión, un dominio de modificación epigenética, un dominio de activación de la transcripción, o un dominio represor de la transcripción. Ver el documento WO 2014/089290. Las proteínas Cas pueden fusionarse además a un polipéptido heterólogo que proporciona mayor o menor estabilidad. El dominio o polipéptido heterólogo fusionado puede ubicarse en el extremo N-terminal, el extremo C-terminal, 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 localización subcelular. Tales secuencias pueden incluir, por ejemplo, una señal de localización nuclear (NLS) tal como la NLS de SV40 para la orientación al núcleo, una señal de localización mitocondrial para la orientación a la mitocondria, una señal de retención en ER, y similares. Ver, por ejemplo, Lange y otros (2007) J. Biol. Chem. 282:5101-5105. Por ejemplo, las proteínas Cas pueden fusionarse a una o más señales de localización nuclear (por ejemplo, dos o tres señales de localización nuclear). Tales señales de localización subcelular pueden ubicarse en el extremo N-terminal, el extremo C-terminal, o en cualquier parte dentro de la proteína Cas. Una 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 de HIV-1, el motivo de penetración celular TLM del virus de la hepatitis B humana, MPG, Pep-1, VP22, un péptido de penetración celular del virus Herpes simplex, o una secuencia peptídica de poliarginina. Ver, por ejemplo, el documento WO 2014/089290. El dominio de penetración celular puede ubicarse en el extremo N-terminal, el extremo C-terminal, o en cualquier parte dentro de la proteína Cas.

Las proteínas Cas pueden comprender además un polipéptido heterólogo para facilitar la detección o la purificación, tal como una proteína fluorescente, una etiqueta de purificación, o una etiqueta tipo epítopo. Los ejemplos de proteínas fluorescentes incluyen 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, Citrine, Venus, YPet, PhiYFP, ZsYellow1), proteínas fluorescentes azules (por ejemplo eBFP, eBFP2, Azurite, mKalamal, GFPuv, Sapphire, T-sapphire), proteínas fluorescentes cian (por ejemplo eCFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), proteínas fluorescentes rojas (mKate, mKate2, mPlum, monómero DsRed, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monómero, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), proteínas fluorescentes naranjas (mOrange, mKO, Kusabira-Orange, Kusabira-Orange 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, tiorredoxina (TRX), poli(NANP), etiqueta para la

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

5 Las proteínas Cas pueden proporcionarse en cualquier forma. Por ejemplo, una proteína Cas puede proporcionarse en forma de una proteína, tal como una proteína Cas en complejo con un ARNg. Alternativamente, una proteína Cas puede proporcionarse 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 tener los codones optimizados para la traducción eficiente a proteína en una célula u organismo particular. Por ejemplo, el ácido nucleico que codifica la proteína Cas puede modificarse para sustituir los codones que tienen una mayor frecuencia de uso en una célula bacteriana, una célula de levadura, una célula humana, una célula no humana, una célula de mamífero, una célula de roedor, una célula de ratón, una célula de rata, o cualquier otra célula huésped de interés, en comparación con la secuencia de polinucleótido de origen natural. Cuando un ácido nucleico que codifica la proteína Cas se introduce en la célula, la proteína Cas puede expresarse de manera transitoria, de manera condicional, o de manera constitutiva en la célula.

Los ácidos nucleicos que codifican las proteínas Cas pueden integrarse de manera estable en el genoma de la célula y unirse operativamente a un promotor activo en la célula. Alternativamente, los ácidos nucleicos que codifican las proteínas Cas pueden unirse operativamente a un promotor en un constructo de expresión. Los constructos de expresión incluyen cualquiera de los constructos de ácido nucleico capaces de dirigir la expresión de un gen u otra secuencia de ácido nucleico de interés (por ejemplo, un gen Cas) y que pueden transferir una secuencia de ácido nucleico de interés de este tipo a una célula diana. Por ejemplo, el ácido nucleico que codifica la proteína Cas puede estar en un vector de transformación que comprende un inserto de ácido nucleico y/o un vector que comprende un ADN que codifica un ARNg. Alternativamente, puede estar en un vector o plásmido que se encuentra separado del vector de transformación que comprende el inserto de ácido nucleico y/o separado del vector que comprende el ADN que codifica el ARNg. Los promotores que pueden usarse en un constructo de expresión incluyen, por ejemplo, promotores activos en una célula pluripotente de rata, eucariota, de mamífero, de mamífero no humano, humana, de roedor, de ratón, o de hámster. Pueden usarse además promotores activos en un embrión en etapa de una célula. Tales promotores pueden ser, por ejemplo, promotores condicionales, promotores inducibles, promotores constitutivos, o promotores específicos de tejido. Ejemplos de otros promotores se describen en otra parte en la presente descripción.

B. ARN guías (ARNg)

Un "ARN guía" o "ARNg" incluye una molécula de ARN que se une a una proteína Cas y dirige a la proteína Cas hacia un ubicación específica dentro de un ADN diana. Los ARN guías pueden comprender dos segmentos: un "segmento de orientación al ADN" y un "segmento de unión a proteína." "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 orientador." Otros ARNg son una sola molécula de ARN (un solo polinucleótido de ARN), que también puede llamarse una "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/2013/142578A1, y WO 2014/131833A1. Los términos "ARN guía" y "ARNg" son incluyentes, e incluyen tanto los ARNg de dos moléculas como los ARNg de molécula simple.

Un ARNg de dos moléculas ilustrativo comprende una molécula tipo ARNcr ("ARN de CRISPR" o "ARN orientador" o "ARNcr" o "repetición de ARNcr") y una molécula tipo ARNtracr correspondiente ("ARN de CRISPR transactivador" o "ARN activador" o "ARNtracr" o "andamio"). Un ARNcr comprende tanto el segmento de orientación al ADN (monocatenario) del ARNg como un tramo de nucleótidos que forma una mitad del dúplex de ARNbc del segmento de unión a proteína 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 proteína del ARNg. Un tramo de nucleótidos de un ARNcr es complementario a y se hibrida con un tramo de nucleótidos de un ARNtracr para formar el dúplex de ARNbc del dominio de unión a proteína del ARNg. Como tal, puede decirse que cada ARNcr tiene un ARNtracr correspondiente. Los ARNtracr pueden estar en cualquier forma (por ejemplo, ARNtracr de longitud completa o ARNtracr parciales activos) y ser de longitudes variables. Las formas de ARNtracr pueden incluir transcritos primarios o formas procesadas. Por ejemplo, en *S. pyogenes*, las diferentes formas de los ARNtracr incluyen versiones de 171 nucleótidos, 89 nucleótidos, 75 nucleótidos, y 65 nucleótidos. Ver, por ejemplo, Deltcheva y otros (2011) Nature 471:602-607 y el documento WO 2014/093661.

El ARNcr y el ARNtracr correspondiente se hibridan para formar un ARNg. El ARNcr proporciona además el segmento monocatenario de orientación al ADN que se hibrida con una secuencia de reconocimiento de ARN de CRISPR. Si se usa para la modificación dentro de una célula, la secuencia exacta de una molécula de ARNcr o ARNtracr dada puede diseñarse de modo que sea específica para 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 orientación al ADN (ARNcr) de un ARNg dado comprende una secuencia de nucleótidos que es complementaria a una secuencia en un ADN diana. El segmento de orientación al ADN de un ARNg interactúa con un ADN diana de una manera específica de la secuencia por medio de hibridación (es decir, apareamiento de bases). Como tal, la secuencia de nucleótidos del segmento de orientación al ADN puede variar y determina la ubicación dentro del ADN diana con la que van a interactuar el ARNg y el ADN diana. El segmento de orientación al ADN de un ARNg en cuestión puede modificarse de modo que se hibride con cualquier secuencia deseada dentro de un ADN diana. Los ARNcr de origen natural difieren en dependencia del sistema de Cas9 y el organismo pero frecuentemente contienen un segmento de orientación de entre 21 a 72 nucleótidos de longitud, flanqueado por dos repeticiones directas (DR) de una longitud de entre 21 a 46 nucleótidos (ver, por ejemplo, el documento WO2014/131833). En el caso de *S. pyogenes*, las DR son de 36 nucleótidos de longitud y el segmento de orientación es de 30 nucleótidos de longitud. La DR ubicada en 3' es complementaria a y se hibrida con el ARNtracr correspondiente, el que a su vez se une a la proteína Cas9.

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

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

Los ARNtracr pueden estar en cualquier forma (por ejemplo, ARNtracr de longitud completa o ARNtracr parciales activos) y ser de longitudes variables. Pueden incluir transcritos primarios o formas procesadas. Por ejemplo, los ARNtracr (como parte de un ARN guía simple o como una molécula separada como parte de un ARNg de dos moléculas) puede 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) Nature 471:602-607; documento WO 2014/093661. Los ejemplos de ARNtracr dentro de los ARN guías simples (ARNgs) incluyen los segmentos de ARNtracr encontrados dentro de las versiones +48, +54, +67, y +85 de los ARNgs, donde "+n" indica que hasta el nucleótido +n del ARNtracr de tipo silvestre se incluye en el ARNgs. Ver el documento US 8,697,359.

El porcentaje de complementariedad entre la secuencia de orientación al ADN y la secuencia de reconocimiento de ARN de CRISPR dentro del ADN diana puede ser al menos 60 % (por ejemplo, al menos 65 %, al menos 70 %, al menos 75

5 %, 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 de orientación al ADN y la secuencia de reconocimiento de ARN de CRISPR dentro del ADN diana es al menos 60 % en aproximadamente 20 nucleótidos contiguos. En un ejemplo, el porcentaje de complementariedad entre la secuencia de orientación al ADN y la secuencia de reconocimiento de ARN de CRISPR dentro del ADN diana es 100 % en los 14 nucleótidos contiguos en el extremo 5' de la secuencia de reconocimiento de ARN de CRISPR dentro de la cadena complementaria del ADN diana y tan bajo como 0 % en el resto. En un caso de este tipo, puede considerarse que la secuencia de orientación al ADN es de 14 nucleótidos de longitud. En otro ejemplo, el porcentaje de complementariedad entre la secuencia de orientación al ADN y la secuencia de reconocimiento de ARN de CRISPR dentro del ADN diana es 100 % en los siete nucleótidos contiguos en el extremo 5' de la secuencia de reconocimiento de ARN de CRISPR dentro de la cadena complementaria del ADN diana y tan bajo como 0 % en el resto. En un caso de este tipo, puede considerarse que la secuencia de orientación al ADN es de 7 nucleótidos de longitud.

15 El segmento de unión a proteína de un ARNg puede comprender dos tramos de nucleótidos que son complementarios entre sí. Los nucleótidos complementarios del segmento de unión a proteína se hibridan para formar un dúplex de ARN bicatenario (ARNbc). El segmento de unión a proteína de un ARNg en cuestión interactúa con una proteína Cas, y el ARNg dirige la proteína Cas unida a una secuencia de nucleótidos específica dentro del ADN diana por medio del segmento de orientación al ADN.

20 Los ARN guías pueden incluir modificaciones o secuencias que proporcionan características convenientes adicionales (por ejemplo, estabilidad modificada o regulada; orientación subcelular; detección con una etiqueta fluorescente; un sitio de unión de una proteína o un complejo de proteínas; y similares). Los ejemplos de tales modificaciones incluyen, por ejemplo, una caperuza 5' (por ejemplo, una caperuza de 7-metilguanilato (m7G)); una cola de poliadenina 3' (es decir, una cola poli(A) 3'); una secuencia riboswitch (por ejemplo, para permitir la estabilidad regulada y/o la accesibilidad regulada por proteínas y/o complejos de proteínas); una secuencia de control de la 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, mitocondria, cloroplastos, y similares); una modificación o secuencia que proporciona la detección (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 de proteínas (por ejemplo, proteínas que actúan sobre el ADN, que incluyen activadores de la transcripción, represores de la transcripción, ADN metiltransferasas, ADN desmetilasas, histona acetiltransferasas, histona desacetilasas, y similares); y combinaciones de estas.

35 Un ARNg puede comprender una secuencia de ácido nucleico que codifica un ARNcr y un ARNtracr. Por ejemplo, un ARNg puede comprender: (a) un ARN quimérico que tiene la secuencia de ácido nucleico 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' (SEQ ID NO: 1); o (b) un ARN quimérico que tiene la secuencia de ácido nucleico 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCG-3' (SEQ ID NO: 2).

40 En algunos casos, el ARNcr comprende 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAU-3' (SEQ ID NO: 3); 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG (SEQ ID NO: 4); o 5'-GAGUCCGAGCAGAAGAAGUUUA-3' (SEQ ID NO: 5).

45 En algunos casos, el ARNtracr comprende, 5'-AAGGCUAGUCCG-3' (SEQ ID NO: 6) o 5'-AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' (SEQ ID NO: 7).

50 Los ARN guías pueden proporcionarse en cualquier forma. Por ejemplo, el ARNg puede proporcionarse en forma de ARN, ya sea como dos moléculas (ARNcr y ARNtracr separados) o como una molécula (ARNgs), y opcionalmente en forma de un complejo con una proteína Cas. El ARNg puede proporcionarse además en forma de ADN que codifica el ARN. El ADN que codifica el ARNg puede codificar una sola molécula de ARN (ARNgs) o moléculas de ARN separadas (por ejemplo, ARNcr 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.

55 Cuando un ADN que codifica un ARNg se introduce en la célula, el ARNg puede expresarse de manera transitoria, de manera condicional, o de manera constitutiva en la célula. Los ADN que codifican los ARNg pueden integrarse de manera estable en el genoma de la célula y unirse operativamente a un promotor activo en la célula. Alternativamente, los ADN que codifican los ARNg pueden unirse operativamente a un promotor en un constructo de expresión. Por ejemplo, el ADN que codifica el ARNg puede estar en un vector de transformación que comprende un inserto de ácido nucleico y/o un vector que comprende un ácido nucleico que codifica una proteína Cas. Alternativamente, puede estar en un vector o un plásmido que se encuentra separado del vector de transformación que comprende el inserto de ácido nucleico y/o separado del vector que comprende el ácido nucleico que codifica la proteína Cas. Los promotores que pueden usarse en tales constructos de expresión incluyen promotores activos, por ejemplo, en una célula pluripotente de rata, eucariota, de mamífero, de mamífero no humano, humana, de roedor, de ratón, o de hámster. Pueden usarse además promotores activos en un embrión en etapa de una célula. 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, un promotor de polimerasa III U6 de rata, o un promotor de polimerasa III U6 de ratón. Ejemplos de otros promotores se describen en otra parte en la presente descripción.

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

C. Secuencias de reconocimiento de ARN de CRISPR

El término "secuencia de reconocimiento de ARN de CRISPR" incluye las secuencias de ácidos nucleicos presentes en un ADN diana a las que se unirá un segmento de orientación al ADN de un ARNg, siempre que existan condiciones suficientes para la unión. Por ejemplo, las secuencias de reconocimiento de ARN de CRISPR incluyen las secuencias para las que se diseña un ARN guía de modo que tengan complementariedad, donde la hibridación entre una secuencia de reconocimiento de ARN de CRISPR y una secuencia de orientación al ADN promueve la formación de un complejo CRISPR. No se requiere necesariamente de complementariedad completa, siempre que exista complementariedad suficiente para causar la hibridación y promover la formación de un complejo CRISPR. Las secuencias de reconocimiento de ARN de CRISPR incluyen además sitios de escisión de proteínas Cas, descritos en más detalle a continuación. Una secuencia de reconocimiento de ARN de CRISPR puede comprender cualquier polinucleótido, que puede ubicarse, por ejemplo, en el núcleo o el citoplasma de una célula o dentro de un organelo de una célula, tal como una mitocondria o un cloroplasto.

La secuencia de reconocimiento de ARN de CRISPR dentro de un ADN diana puede reconocerse por (es decir, unirse a, o hibridarse con, o ser complementaria a) una proteína Cas o un ARNg. Las condiciones de unión de ADN/ARN adecuadas incluyen las condiciones fisiológicas presentes normalmente en una célula. Otras condiciones de unión de ADN/ARN adecuadas (por ejemplo, condiciones en un sistema libre de células) se conocen la técnica (ver, por ejemplo, Molecular Cloning: A Laboratory Manual, 3ra Ed. (Sambrook y otros, Harbor Laboratory Press 2001)). La cadena del ADN diana que es complementaria a y se hibrida con la proteína Cas o ARNg puede llamarse la "cadena complementaria," y la cadena del ADN diana que es complementaria a la "cadena complementaria" (y que por lo tanto no es complementaria a la proteína Cas o ARNg) puede llamarse "cadena no complementaria" o "cadena plantilla."

La proteína Cas puede escindir el ácido nucleico en un sitio dentro de o fuera de la secuencia de ácido nucleico presente en el ADN diana a la que se unirá el segmento de orientación al ADN de un ARNg. El "sitio de escisión" incluye la posición de un ácido nucleico en la que una proteína Cas produce una ruptura monocatenaria o una ruptura bicatenaria. Por ejemplo, la formación de un complejo CRISPR (que comprende un ARNg hibridado a una secuencia de reconocimiento de ARN de CRISPR y en complejo con una proteína Cas) puede dar como resultado 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 diana a la que se unirá un segmento de orientación al ADN de un ARNg. Si el sitio de escisión está fuera de la secuencia de ácido nucleico a la que se unirá el segmento de orientación al ADN del ARNg, el sitio de escisión aún se considera dentro de la "secuencia de reconocimiento de ARN de CRISPR." El sitio de escisión puede estar solamente en una 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 (lo que produce extremos romos) o pueden estar en sitios diferentes en cada cadena (lo que produce extremos escalonados). Los extremos escalonados pueden producirse, por ejemplo, mediante el uso de dos proteínas Cas, cada una de las cuales produce una ruptura monocatenaria en un sitio de escisión diferente en cada cadena, lo que produce de este modo una ruptura bicatenaria. Por ejemplo, una primera nickasa puede crear una ruptura monocatenaria en la primera cadena de ADN bicatenario (ADNbc), y una segunda nickasa puede crear una ruptura monocatenaria en la segunda cadena de ADNbc de manera que se crean secuencias salientes. En algunos casos, la secuencia de reconocimiento de ARN de CRISPR de la nickasa en la primera cadena se encuentra separada de la secuencia de reconocimiento de ARN de CRISPR de la nickasa en la segunda cadena por al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500, o 1000 pares de bases.

La escisión sitio específica del ADN diana por Cas9 puede ocurrir en ubicaciones determinadas por (i) la complementariedad de apareamiento de bases entre el ARNg y el ADN diana y (ii) un motivo corto, llamado el motivo adyacente al protoespaciador (PAM), en el ADN diana. El PAM puede flanquear la secuencia de reconocimiento de ARN de CRISPR. Opcionalmente, la secuencia de reconocimiento de ARN de CRISPR puede estar flanqueada en el extremo 3' por el PAM. Por ejemplo, el sitio de escisión de Cas9 puede estar aproximadamente 1 a aproximadamente 10 o 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 la Cas9 de *S. pyogenes* o una Cas9 estrechamente relacionada), la secuencia PAM de la cadena no complementaria puede ser 5'-N₁GG-3', donde N₁ es cualquier nucleótido de ADN y está inmediatamente 3' de la secuencia de reconocimiento de ARN de CRISPR de la cadena no complementaria del ADN diana. Como tal, la secuencia PAM de la cadena complementaria sería 5'-CCN₂-3', donde N₂ es cualquier nucleótido de ADN y está inmediatamente 5' de la secuencia de reconocimiento de ARN de CRISPR de la cadena complementaria del ADN diana. En algunos de tales casos, N₁ y N₂ pueden ser complementarios y el par de bases N₁-N₂ puede ser cualquier par de bases (por ejemplo, N₁=C y N₂=G; N₁=G y N₂=C; N₁=A y N₂=T; o N₁=T, y N₂=A).

Los ejemplos de secuencias de reconocimiento de ARN de CRISPR incluyen una secuencia de ADN complementaria al segmento de orientación al ADN de un ARNg, o una secuencia de ADN de este tipo además de una secuencia PAM. Por ejemplo, el motivo diana puede ser una secuencia de ADN de 20 nucleótidos que precede inmediatamente a un motivo NGG reconocido por una proteína Cas, tal como GN₁₉NGG (SEQ ID NO: 8) o N₂₀NGG (SEQ ID NO: 9) (ver, por ejemplo, el documento WO 2014/165825). La guanina en el extremo 5' puede facilitar la transcripción por la ARN polimerasa en las células. Otros ejemplos de secuencias de reconocimiento de ARN de CRISPR pueden incluir dos nucleótidos de guanina en el extremo 5' (por ejemplo, GGN₂₀NGG; SEQ ID NO: 10) para facilitar la transcripción eficiente por la polimerasa T7 *in vitro*. Ver, por ejemplo, el documento WO 2014/065596. Otras secuencias de reconocimiento de ARN de CRISPR pueden tener entre 4-22 nucleótidos de longitud de las SEQ ID NOS: 8-10, que incluyen el 5' G o GG y el 3' GG o NGG. Aún otras secuencias de reconocimiento de ARN de CRISPR pueden tener entre 14 y 20 nucleótidos de longitud de las SEQ ID NOS: 8-10. Ejemplos específicos de secuencias de reconocimiento de ARN de CRISPR incluyen secuencias de ADN complementarias a ácidos nucleicos que comprenden cualquiera de las SEQ ID NOS: 11-38.

La secuencia de reconocimiento de ARN de CRISPR puede ser cualquier secuencia de ácido nucleico endógena o exógena a una célula. La secuencia de reconocimiento de ARN de CRISPR puede ser una secuencia que codifica un producto génico (por ejemplo, una proteína) o una secuencia no codificante (por ejemplo, una secuencia reguladora) o puede incluir ambas. En algunos casos, la secuencia de reconocimiento de ARN de CRISPR puede estar dentro de un gen o ácido nucleico asociado a enfermedad y/o dentro de un gen o ácido nucleico asociado a ruta de señalización. Un gen o ácido nucleico asociado a enfermedad incluye cualquier gen o ácido nucleico que produce productos de transcripción o traducción a un nivel anormal o en una forma anormal en células derivadas de tejidos afectados por enfermedad en comparación con tejidos o células de un control sin enfermedad. Por ejemplo, un gen asociado a enfermedad puede poseer una o mutaciones o variaciones genéticas que son directamente responsables de la etiología de una enfermedad o están en desequilibrio de ligamiento con uno o más genes que son responsables de la etiología de una enfermedad. Los productos transcritos o traducidos pueden ser conocidos o desconocidos, y pueden estar a un nivel normal o anormal. Los ejemplos de genes y ácidos nucleicos asociados a enfermedad están disponibles del Instituto de Medicina Genética McKusick-Nathans, Universidad Johns Hopkins (Baltimore, MD) y del Centro Nacional para la Información Biotecnológica, Biblioteca Nacional de Medicina (Bethesda, MD), disponible en la red mundial. Para ejemplos adicionales de genes y ácidos nucleicos asociados a enfermedad, ver la patente de Estados Unidos núm. 8,697,359.

Las mutaciones en los genes que causan enfermedad pueden ser mutaciones recesivas o mutaciones dominantes. Los organismos diploides (es decir, organismos que tienen dos copias de cada cromosoma) típicamente portan dos copias de cada gen. Si las dos copias en un individuo son idénticas, el individuo es homocigótico para el gen. Si las copias están en alelos diferentes, el individuo es heterocigótico para el gen. El término genotipo incluye si un individuo porta mutaciones en un solo gen (o genes), y el término fenotipo incluye las consecuencias físicas y funcionales de ese genotipo. Las mutaciones recesivas incluyen mutaciones en las que ambos alelos deben ser mutantes para que se observe un fenotipo mutante (es decir, el organismo debe ser homocigótico para el alelo mutante para mostrar el fenotipo mutante). Las mutaciones recesivas pueden, por ejemplo, inactivar un gen afectado y conducir a una pérdida de función. Por ejemplo, una mutación recesiva puede eliminar la totalidad o parte de un gen de un cromosoma, interrumpir la expresión de un gen, o alterar la estructura de la proteína codificada, lo que altera de este modo su función. Por el contrario, las mutaciones dominantes incluyen mutaciones en las que el fenotipo mutante se observa en un organismo que es heterocigótico para la mutación (es decir, el organismo porta un alelo mutante y un alelo de tipo silvestre). Una mutación dominante puede, por ejemplo, conducir a una ganancia de función. Por ejemplo, una mutación dominante puede aumentar la actividad de un producto génico dado, confiere una nueva actividad al producto génico, o conduce a su expresión espacial y temporal inadecuada. Una mutación dominante puede asociarse además con una pérdida de función. En algunos casos, si dos copias de un gen se requieren para la función normal, eliminar una sola copia puede causar un fenotipo mutante. Tales genes son haploinsuficientes. En otros casos, las mutaciones en un alelo pueden conducir a un cambio estructural en la proteína que interfiere con la función de la proteína de tipo silvestre codificada por el otro alelo. Tales mutaciones son mutaciones negativas dominantes. Algunos alelos pueden asociarse tanto con un fenotipo recesivo como con uno dominante.

Algunas secuencias de reconocimiento de ARN de CRISPR están dentro de un gen o ácido nucleico que comprende una mutación. La mutación puede ser, por ejemplo, una mutación dominante o una mutación recesiva. En algunos casos, la mutación dominante está dentro de una célula que es heterocigótica para la mutación dominante (es decir, la célula comprende un alelo de tipo silvestre y un alelo mutante que comprende la mutación dominante). En algunos de tales casos, la secuencia de reconocimiento de ARN de CRISPR puede estar dentro del alelo mutante pero no en el alelo de tipo silvestre. Alternativamente, la secuencia de reconocimiento de ARN de CRISPR puede estar dentro del alelo de tipo silvestre pero no en el alelo mutante.

III. Vectores de transformación e insertos de ácido nucleico

Los métodos y composiciones descritos en la presente descripción pueden utilizar además vectores de transformación que comprenden insertos de ácido nucleico y brazos de homología para modificar un genoma dentro de una célula. En tales métodos, el inserto de ácido nucleico se integra en un locus genómico diana determinado por los brazos de homología a través de un evento de recombinación homóloga. Los métodos proporcionados en la presente descripción pueden beneficiarse de los agentes tipo nucleasa (por ejemplo, proteínas Cas) en combinación con el evento de recombinación homóloga. Tales métodos emplean la mella o ruptura bicatenaria creada por el agente tipo nucleasa en un

sitio de escisión de nucleasa en combinación con recombinación homóloga para facilitar la integración dirigida del inserto de ácido nucleico en el locus genómico diana.

A. Vectores de transformación e insertos de ácido nucleico para células distintas a los embriones en etapa de una célula

(1) Inserto de ácido nucleico

Uno o más insertos de ácido nucleico separados pueden emplearse en los métodos descritos en la presente descripción, y pueden introducirse en una célula por medio de vectores de transformación separados o en el mismo vector de transformación. Los insertos de ácido nucleico incluyen segmentos de ADN a integrar en loci genómicos diana. La integración de un inserto de ácido nucleico en un locus diana puede dar como resultado la adición de una secuencia de ácido nucleico de interés al locus diana, la delección de una secuencia de ácido nucleico de interés en el locus diana, y/o el reemplazo de una secuencia de ácido nucleico de interés en el locus diana (es decir, delección e inserción).

El inserto de ácido nucleico o el ácido nucleico correspondiente en el locus diana que se reemplaza puede ser una región codificante, un intrón, un exón, una región no traducida, una región reguladora, un promotor, un potenciador, o cualquier combinación de estos. Por otra parte, el inserto de ácido nucleico o el ácido nucleico correspondiente en el locus diana que se reemplaza puede ser de cualquier longitud deseada, que incluye, 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 nucleótidos de longitud, 1,5 kb a 2 kb de nucleótidos de longitud, 2 kb a 2,5 kb de nucleótidos de longitud, 2,5 kb a 3 kb de nucleótidos de longitud, 3 kb a 5 kb de nucleótidos de longitud, 5 kb a 8 kb de nucleótidos de longitud, 8 kb a 10 kb de nucleótidos de longitud o más. En otros casos, la longitud 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, 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. Aún 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. Algunos insertos de ácido nucleico pueden ser incluso menores. A modo de ejemplo, un inserto de aproximadamente 4 nucleótidos a aproximadamente 12 nucleótidos de longitud puede insertarse para crear un sitio de enzima de restricción.

En algunos vectores de transformación, el inserto de ácido nucleico 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 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. 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.

En algunos casos, el reemplazo del ácido nucleico en el locus diana da como resultado la delección de una secuencia de ácido nucleico en el intervalo 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 una 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 está 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 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.

En otros casos, el inserto de ácido nucleico o el ácido nucleico correspondiente en el locus diana que se reemplaza puede ser de al menos 10 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 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 inserto de ácido nucleico puede comprender ADN genómico o cualquier otro tipo de ADN. Por ejemplo, el inserto de ácido nucleico puede ser de un procarionte, un eucariote, 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 doméstico, un mamífero de uso agrícola, o cualquier otro organismo de interés.

El inserto de ácido nucleico y/o el ácido nucleico en el locus diana pueden comprender una secuencia codificante o una secuencia no codificante, tal como un elemento regulador (por ejemplo, un promotor, un potenciador, o un elemento de unión a represor de la transcripción). Por ejemplo, el inserto de ácido nucleico puede comprender un alelo sustituido de al menos un exón de un gen endógeno, o un alelo sustituido del gen endógeno completo (es decir, "sustitución por intercambio de genes").

Por ejemplo, el inserto de ácido nucleico puede ser homólogo u ortólogo a una secuencia que se reconoce para la delección en el locus genómico diana. El inserto de ácido nucleico homólogo u ortólogo puede reemplazar la secuencia que se reconoce para la delección en el locus genómico de interés. Una secuencia homóloga incluye una secuencia de ácido nucleico que es idéntica o sustancialmente similar a una secuencia de referencia conocida, de manera 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. 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 humana con una secuencia de ácido nucleico humana homóloga u ortóloga (es decir, el inserto de ácido nucleico se inserta en el lugar de la secuencia de ADN no humana correspondiente en su locus genómico endógeno).

El inserto de ácido nucleico puede comprender además 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 accionadora en orientación con sentido con respecto a la transcripción de un gen diana; (b) un casete de selección de fármaco (DSC) en orientación con 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 separador de exones y un módulo tipo trampa de genes 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 accionadora y el DSC; y (ii) contiene la NSI en orientación con 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. Tales marcadores de selección incluyen, pero no se limitan, a neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa (puro^r), blastidina S desaminasa (bsr^r), xantina/guanina fosforribosil transferasa (gpt), o timidina quinasa del virus herpes simplex (HSV-k), o una combinación de estos. El polinucleótido que codifica el marcador de selección puede unirse operativamente a un promotor activo en una célula a transformar. Los ejemplos de promotores se describen en otra parte en la presente descripción.

En algunos vectores de transformación, el inserto de ácido nucleico comprende un gen reportero. Los ejemplos de genes reporteros son los 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), DsRed, ZsGreen, MmGFP, mPlum, mCherry, tdTomato, mStrawberry, J-Red, mOrange, mKO, mCitrine, Venus, YPet, Emerald, CyPet, Cerulean, T-Sapphire, fosfatasa alcalina, y una combinación de estas. Tales genes reporteros pueden unirse operativamente a un promotor activo en una célula a transformar. Los ejemplos de promotores se describen en otra parte en 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 de selección y genes reporteros que pueden incluirse se analizan en detalle en otra parte en la presente descripción.

En algunos vectores de transformación, el inserto de ácido nucleico comprende un ácido nucleico flanqueado con secuencias diana de recombinación sitio específica. Aunque el inserto de ácido nucleico completo puede estar flanqueado por tales secuencias diana de recombinación sitio específica, cualquier región o polinucleótido de interés individual dentro del inserto de ácido nucleico también puede estar flanqueado por tales sitios. Las secuencias diana de recombinación sitio específica, 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 estas. En un ejemplo, los sitios de recombinación sitio específica flanquean un polinucleótido que codifica un marcador de selección y/o un gen reportero contenido dentro del inserto de ácido nucleico. Después de la integración del inserto de ácido nucleico en un locus diana, las secuencias entre los sitios de recombinación sitio específica pueden eliminarse.

(2) Vectores de transformación

Los vectores de transformación pueden emplearse para introducir el inserto de ácido nucleico en un locus genómico diana y comprenden el inserto de ácido nucleico y 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 de ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN). Para facilitar la referencia, los brazos de homología se refieren en la presente descripción 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 respecto al inserto de ácido nucleico dentro del vector de transformación. Los brazos de homología 5' y 3' corresponden a regiones dentro del locus diana, que se refieren en la presente descripción como "secuencia diana 5'" y "secuencia diana 3'," respectivamente. Algunos vectores de transformación comprenden brazos de homología 5' y 3' sin inserto de ácido nucleico. Tales vectores de transformación pueden funcionar para eliminar la secuencia entre las secuencias diana 5' y 3' sin insertar un inserto de ácido nucleico.

Un brazo de homología y una secuencia diana "corresponden" o son "correspondientes" entre sí cuando las dos regiones comparten entre sí un nivel de identidad de secuencia suficiente 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 diana dada y el brazo de homología correspondiente que se encuentra en el vector de transformación puede ser cualquier grado de identidad de secuencia que permita que ocurra la recombinación homóloga. Por ejemplo, la cantidad de identidad de secuencia compartida por el brazo de homología del vector de transformación (o un fragmento de este) y la secuencia diana (o un fragmento de esta) 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. Por otra parte, una región correspondiente de homología entre el brazo de homología y la secuencia diana correspondiente puede ser de cualquier longitud que sea suficiente para promover la recombinación homóloga en el sitio de reconocimiento escindido. Por ejemplo, un brazo de homología dado y/o la secuencia diana correspondiente pueden comprender regiones correspondientes de homología que son de al menos aproximadamente 5-10 kb, 5-15 kb, 5-20 kb, 5-25 kb, 5-30 kb, 5-35 kb, 5-40 kb, 5-45 kb, 5-50 kb, 5-55 kb, 5-60 kb, 5-65 kb, 5-70 kb, 5-75 kb, 5-80 kb, 5-85 kb, 5-90 kb, 5-95 kb, 5-100 kb, 100-200 kb, o 200-300 kb de longitud o más (tal como se describe en los vectores LTVEC descritos en otra parte en la presente descripción) de manera que el brazo de homología tenga homología suficiente para experimentar recombinación homóloga con las secuencias diana correspondientes dentro del genoma de la célula.

Los brazos de homología pueden corresponder a un locus que es nativo para una célula (por ejemplo, el locus diana), o alternativamente pueden corresponder a una región de un segmento de ADN heterólogo o exógeno que se integró en el genoma de la célula, que incluye, por ejemplo, transgenes, casetes de expresión, o regiones de ADN heterólogas o exógenas. Alternativamente, los brazos de homología del vector de transformación pueden corresponder a una región de un cromosoma artificial de levadura (YAC), un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano, o cualquier otra región modificada genéticamente contenida en una célula huésped adecuada. Más aún, los brazos de homología del vector de transformación pueden corresponder a o derivarse de una región de una biblioteca de BAC, una biblioteca de cósmidos, o una biblioteca de fagos P1. En ciertos casos, los brazos de homología del 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, títí, mono rhesus), un mamífero doméstico, un mamífero de 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 mediante el uso de un método convencional o que puede transformarse solamente de manera incorrecta o solamente con una eficiencia significativamente baja en ausencia de una mella o ruptura bicatenaria inducida por un agente tipo nucleasa (por ejemplo, una proteína Cas). En algunos casos, los brazos de homología se derivan de ADN sintético.

En algunos vectores de transformación, los brazos de homología 5' y 3' corresponden a un genoma diana. Alternativamente, los brazos de homología pueden ser de un genoma relacionado. Por ejemplo, el genoma diana es un genoma de ratón de una primera cepa, y los brazos de transformación son de un genoma de ratón de una segunda cepa, en donde la primera cepa y la segunda cepa son diferentes. En ciertos casos, los brazos de homología son del genoma del mismo animal o son del genoma de la misma cepa, por ejemplo, el genoma diana es un genoma de ratón de una primera cepa, y los brazos de 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 diana correspondiente, que incluye, por ejemplo, al menos 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 en más detalle a continuación, 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 vectores de transformación para ayudar a la modificación de un locus diana. Tales agentes tipo nucleasa pueden promover la recombinación homóloga entre el vector de transformación y el locus diana. Cuando se emplean agentes tipo nucleasa en combinación con un vector de transformación, el vector de transformación puede comprender brazos de homología 5' y 3' correspondientes a secuencias diana 5' y 3' ubicadas en proximidad suficiente a un sitio de escisión de nucleasa a fin de promover la ocurrencia de un evento de recombinación homóloga entre las secuencias diana y los brazos de homología tras una mella o ruptura bicatenaria en el sitio de escisión de nucleasa. El término "sitio de escisión de nucleasa" incluye una secuencia de ADN en la que se crea una mella o ruptura bicatenaria por un agente tipo nucleasa (por ejemplo, un sitio de escisión de Cas9). Las secuencias diana dentro del locus diana que corresponden a los brazos de homología 5' y 3' del vector de transformación se "ubican en proximidad suficiente" a un sitio de escisión de nucleasa si la distancia es tal que promueve la ocurrencia de un evento de recombinación homóloga entre las secuencias diana 5' y 3' y los brazos de homología tras una mella o ruptura bicatenaria en el sitio de reconocimiento. Por lo tanto, en casos específicos, las secuencias diana correspondientes a los brazos de homología 5' y/o 3' del vector de transformación están dentro de al menos 1 nucleótido de un sitio de reconocimiento dado o están dentro de al menos 10 nucleótidos a aproximadamente 14 kb de un sitio de reconocimiento dado. En algunos casos, el sitio de escisión de nucleasa está inmediatamente adyacente a al menos una o ambas de las secuencias diana.

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

El uso combinado del vector de transformación (que incluye, por ejemplo, un vector de transformación grande) con un agente tipo nucleasa puede dar como resultado un aumento de la eficiencia de transformación en comparación con el uso del vector de transformación solo. Por ejemplo, cuando un vector de transformación se usa junto con un agente tipo nucleasa, la eficiencia de transformación del vector de transformación puede aumentarse en al menos dos veces, al menos tres veces, al menos 4 veces, o al menos 10 veces en comparación con el uso del vector de transformación solo.

50 (3) Vectores de transformación grandes

Algunos vectores de transformación son "vectores de transformación grandes" o "LTVEC," que incluyen vectores de transformación que comprenden brazos de homología que corresponden a y se derivan de secuencias de ácidos nucleicos mayores que las usadas típicamente por otros enfoques destinados a realizar la recombinación homóloga en células. Los LTVEC incluyen además vectores de transformación que comprenden insertos de ácido nucleico que tienen secuencias de ácidos nucleicos mayores que las usadas típicamente por otros enfoques destinados a realizar la recombinación homóloga en células. Por ejemplo, los LTVEC hacen posible la modificación de loci grandes que no pueden adaptarse a los vectores de transformación basados en plásmidos tradicionales debido a sus limitaciones de tamaño. Por ejemplo, el locus diana puede ser (es decir, los brazos de homología 5' y 3' pueden corresponder a) un locus de la célula que no puede transformarse mediante el uso de un método convencional o que puede transformarse solamente de manera incorrecta o solamente con una eficiencia significativamente baja en ausencia de una mella o ruptura bicatenaria inducida por un agente tipo nucleasa (por ejemplo, una proteína Cas).

Los ejemplos de LTVEC incluyen vectores derivados de un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano, o un cromosoma artificial de levadura (YAC). Los ejemplos no limitantes de LTVEC y métodos para producirlos se describen, por ejemplo, en la patente de Estados Unidos núm. 6,586,251; la patente de Estados Unidos núm.

6,596,541; la patente de Estados Unidos núm. 7,105,348; y el documento WO 2002/036789(PCT/US01/45375). Los LTVEC pueden estar en forma lineal o en forma circular.

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

15 En algunos casos, un LTVEC 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.

35 En algunos LTVEC, la suma total del brazo de homología 5' y el brazo de homología 3' es al menos 10 kb. En otros LTVEC, el brazo de homología 5' está en el intervalo de aproximadamente 5 kb a aproximadamente 100 kb y/o el brazo de homología 3' está en el intervalo de aproximadamente 5 kb a aproximadamente 100 kb. Cada brazo de homología puede ser, por ejemplo, de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 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. La suma total de los brazos de homología 5' y 3' puede ser, por ejemplo, de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 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 de al menos 5 kb, al menos 10 kb, al menos 15 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 110 kb, al menos 120 kb, al menos 130 kb, al menos 140 kb, al menos 150 kb, al menos 160 kb, al menos 170 kb, al menos 180 kb, al menos 190 kb, o al menos 200 kb. Del mismo modo, la suma total de los brazos de homología 5' y 3' puede ser de al menos 5 kb, al menos 10 kb, al menos 15 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 110 kb, al menos 120 kb, al menos 130 kb, al menos 140 kb, al menos 150 kb, al menos 160 kb, al menos 170 kb, al menos 180 kb, al menos 190 kb, o al menos 200 kb.

En algunos casos, el LTVEC y el inserto de ácido nucleico se diseñan de modo que permitan una delección en el locus diana de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 150 kb, o de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 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 al menos 10 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 150 kb, al menos 200 kb, al menos 250 kb, al menos 300 kb, al menos 350 kb, al menos 400 kb, al menos 450 kb, o al menos 500 kb o más.

En otros casos, el LTVEC y el inserto de ácido nucleico se diseñan de modo que permitan una inserción en el locus diana 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 al menos 10 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 150 kb, al menos 200 kb, al menos 250 kb, al menos 300 kb, al menos 350 kb, al menos 400 kb, al menos 450 kb, o al menos 500 kb o más.

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

B. Vectores de transformación e insertos de ácido nucleico para embriones en etapa de una célula

Los vectores de transformación para el uso en embriones en etapa de una célula son de no más de 5 kb de longitud y pueden ser de ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN), pueden ser monocatenarios o bicatenarios, y pueden estar en forma circular o en forma lineal. Un vector de transformación ilustrativo para el uso en embriones en etapa de una célula es de entre aproximadamente 50 nucleótidos a aproximadamente 5 kb de longitud. Por ejemplo, un vector de transformación para el uso en embriones en etapa de una célula puede ser de entre aproximadamente 50 a aproximadamente 100, aproximadamente 100 a aproximadamente 200, aproximadamente 200 a aproximadamente 300, aproximadamente 300 a aproximadamente 400, aproximadamente 400 a aproximadamente 500, aproximadamente 500 a aproximadamente 600, aproximadamente 600 a aproximadamente 700, aproximadamente 700 a aproximadamente 800, aproximadamente 800 a aproximadamente 900, o aproximadamente 900 a aproximadamente 1000 nucleótidos de longitud. Alternativamente, un vector de transformación para el uso en embriones en etapa de una célula puede ser de entre aproximadamente 1 kb a aproximadamente 1,5 kb, aproximadamente 1,5 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 2,5 kb, aproximadamente 2,5 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 3,5 kb, aproximadamente 3,5 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 4,5 kb, o aproximadamente 4,5 kb a aproximadamente 5 kb de longitud. Alternativamente, un vector de transformación para el uso en embriones en etapa de una célula puede ser, por ejemplo, de no más de 5 kb, 4,5 kb, 4 kb, 3,5 kb, 3 kb, 2,5 kb, 2 kb, 1,5 kb, 1 kb, 900 nucleótidos, 800 nucleótidos, 700 nucleótidos, 600 nucleótidos, 500 nucleótidos, 400 nucleótidos, 300 nucleótidos, 200 nucleótidos, 100 nucleótidos, o 50 nucleótidos de longitud. En el caso de donantes de ADN monocatenario, los vectores de transformación ilustrativos pueden ser de entre aproximadamente 60 nucleótidos y aproximadamente 200 nucleótidos (por ejemplo, aproximadamente 60 nucleótidos a aproximadamente 80 nucleótidos, aproximadamente 80 nucleótidos a aproximadamente 100 nucleótidos, aproximadamente 100 nucleótidos a aproximadamente 120 nucleótidos, aproximadamente 120 nucleótidos a aproximadamente 140 nucleótidos, aproximadamente 140 nucleótidos a aproximadamente 160 nucleótidos, aproximadamente 160 nucleótidos a aproximadamente 180 nucleótidos, o aproximadamente 180 nucleótidos a aproximadamente 200 nucleótidos).

Tales vectores de transformación comprenden brazos de homología 5' y 3' correspondientes a regiones dentro del locus diana (secuencia diana 5' y secuencia diana 3', respectivamente). Opcionalmente, el vector de transformación comprende un inserto de ácido nucleico (por ejemplo, un segmento de ADN a integrar en un locus genómico diana) flanqueado por los brazos de homología 5' y 3'. La integración de un inserto de ácido nucleico en un locus diana puede dar como resultado la adición de una secuencia de ácido nucleico de interés al locus diana, la delección de una secuencia de ácido nucleico de interés en el locus diana, o el reemplazo de una secuencia de ácido nucleico de interés en el locus diana (es decir, delección e inserción).

Una región correspondiente de homología entre el brazo de homología y la secuencia diana correspondiente puede ser de cualquier longitud que sea suficiente para promover la recombinación homóloga. Los brazos de homología ilustrativos para el uso en embriones en etapa de una célula son de entre aproximadamente 20 nucleótidos a aproximadamente 2,5

kb de longitud (por ejemplo, aproximadamente 30 nucleótidos a aproximadamente 100 nucleótidos de longitud). Por ejemplo, un brazo de homología dado y/o la secuencia diana correspondiente pueden comprender regiones correspondientes de homología que son de entre aproximadamente 20 a aproximadamente 30, aproximadamente 30 a aproximadamente 40, aproximadamente 40 a aproximadamente 50, aproximadamente 50 a aproximadamente 60, aproximadamente 60 a aproximadamente 70, aproximadamente 70 a aproximadamente 80, aproximadamente 80 a aproximadamente 90, aproximadamente 90 a aproximadamente 100, aproximadamente 100 a aproximadamente 150, aproximadamente 150 a aproximadamente 200, aproximadamente 200 a aproximadamente 250, aproximadamente 250 a aproximadamente 300, aproximadamente 300 a aproximadamente 350, aproximadamente 350 a aproximadamente 400, aproximadamente 400 a aproximadamente 450, o aproximadamente 450 a aproximadamente 500 nucleótidos de longitud, de manera que los brazos de homología tengan homología suficiente para experimentar recombinación homóloga con las secuencias diana correspondientes dentro del genoma de la célula. Alternativamente, un brazo de homología dado y/o la secuencia diana correspondiente pueden comprender regiones correspondientes de homología que son de entre aproximadamente 0,5 kb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 1,5 kb, aproximadamente 1,5 kb a aproximadamente 2 kb, o aproximadamente 2 kb a aproximadamente 2,5 kb de longitud. En el caso de donantes de ADN monocatenario, los brazos de homología ilustrativos pueden ser de entre aproximadamente 30 nucleótidos y aproximadamente 60 nucleótidos (por ejemplo, aproximadamente 30 a aproximadamente 40 nucleótidos, aproximadamente 40 nucleótidos a aproximadamente 50 nucleótidos, o aproximadamente 50 nucleótidos a aproximadamente 60 nucleótidos).

Como se describió anteriormente, los brazos de homología pueden corresponder a un locus que es nativo para una célula (por ejemplo, el locus diana), o alternativamente pueden corresponder a una región de un segmento de ADN heterólogo o exógeno que se integró en el genoma de la célula. Como se describió anteriormente, las secuencias diana 5' y 3' se ubican preferentemente en proximidad suficiente al sitio de escisión de Cas a fin de promover la ocurrencia de un evento de recombinación homóloga entre las secuencias diana y los brazos de homología tras una ruptura monocatenaria (mella) o una ruptura bicatenaria en el sitio de escisión de Cas.

El inserto de ácido nucleico o el ácido nucleico correspondiente en el locus diana que se elimina y/o reemplaza puede ser de varias longitudes. Un inserto de ácido nucleico ilustrativo o ácido nucleico correspondiente en el locus diana que se elimina y/o reemplaza es de entre aproximadamente 10 nucleótidos a aproximadamente 5 kb de longitud. Por ejemplo, un inserto de ácido nucleico o un ácido nucleico correspondiente en el locus diana que se elimina y/o reemplaza puede ser de entre aproximadamente 1 a aproximadamente 10, aproximadamente 10 a aproximadamente 20, aproximadamente 20 a aproximadamente 30, aproximadamente 30 a aproximadamente 40, aproximadamente 40 a aproximadamente 50, aproximadamente 50 a aproximadamente 60, aproximadamente 60 a aproximadamente 70, aproximadamente 70 a aproximadamente 80, aproximadamente 80 a aproximadamente 90, aproximadamente 90 a aproximadamente 100, aproximadamente 100 a aproximadamente 110, aproximadamente 110 a aproximadamente 120, aproximadamente 120 a aproximadamente 130, aproximadamente 130 a aproximadamente 140, aproximadamente 140 a aproximadamente 150, aproximadamente 150 a aproximadamente 160, aproximadamente 160 a aproximadamente 170, aproximadamente 170 a aproximadamente 180, aproximadamente 180 a aproximadamente 190, aproximadamente 190 a aproximadamente 200, aproximadamente 200 a aproximadamente 300, aproximadamente 300 a aproximadamente 400, aproximadamente 400 a aproximadamente 500, aproximadamente 500 a aproximadamente 600, aproximadamente 600 a aproximadamente 700, aproximadamente 700 a aproximadamente 800, aproximadamente 800 a aproximadamente 900, o aproximadamente 900 a aproximadamente 1000 nucleótidos de longitud. A modo de ejemplo, un inserto de aproximadamente 4 nucleótidos a aproximadamente 12 nucleótidos de longitud puede insertarse para crear un sitio de enzima de restricción. Del mismo modo, un inserto de ácido nucleico o un ácido nucleico correspondiente en el locus diana que se elimina y/o reemplaza puede ser de entre aproximadamente 1 kb a aproximadamente 1,5 kb, aproximadamente 1,5 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 2,5 kb, aproximadamente 2,5 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 3,5 kb, aproximadamente 3,5 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 4,5 kb, o aproximadamente 4,5 kb a aproximadamente 5 kb de longitud. Un ácido nucleico que se elimina de un locus genómico diana puede ser además de entre aproximadamente 5 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 30 kb, aproximadamente 30 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 50 kb, aproximadamente 50 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 70 kb, aproximadamente 70 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 90 kb, aproximadamente 90 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 200 kb, aproximadamente kb a aproximadamente 300 kb, aproximadamente 300 kb a aproximadamente 400 kb, aproximadamente 400 kb a aproximadamente 500 kb, aproximadamente 500 kb a aproximadamente 600 kb, aproximadamente 600 kb a aproximadamente 700 kb, aproximadamente 700 kb a aproximadamente 800 kb, aproximadamente 800 kb a aproximadamente 900 kb, aproximadamente 900 kb a aproximadamente 1 Mb o más. Alternativamente, un ácido nucleico que se elimina de un locus genómico diana puede ser de entre aproximadamente 1 Mb a aproximadamente 1,5 Mb, aproximadamente 1,5 Mb a aproximadamente 2 Mb, aproximadamente 2 Mb a aproximadamente 2,5 Mb, aproximadamente 2,5 Mb a aproximadamente 3 Mb, aproximadamente 3 Mb a aproximadamente 4 Mb, aproximadamente 4 Mb a aproximadamente 5 Mb, aproximadamente 5 Mb a aproximadamente 10 Mb, aproximadamente 10 Mb a aproximadamente 20 Mb, aproximadamente 20 Mb a aproximadamente 30 Mb, aproximadamente 30 Mb a aproximadamente 40 Mb, aproximadamente 40 Mb a aproximadamente 50 Mb, aproximadamente 50 Mb a aproximadamente 60 Mb, aproximadamente 60 Mb a aproximadamente 70 Mb, aproximadamente 70 Mb a aproximadamente 80 Mb, aproximadamente 80 Mb a aproximadamente 90 Mb, o aproximadamente 90 Mb a aproximadamente 100 Mb.

Como se describió anteriormente, el inserto de ácido nucleico puede comprender ADN genómico o cualquier otro tipo de ADN, el inserto de ácido nucleico o el ácido nucleico correspondiente en el locus diana que se elimina y/o reemplaza puede ser una región codificante o una región no codificante, y el inserto de ácido nucleico puede ser homólogo u ortólogo a una secuencia que se reconoce para la delección en el locus genómico diana. El inserto de ácido nucleico puede comprender además un alelo condicional, un polinucleótido que codifica un marcador de selección, un gen reportero, uno o más casetes de expresión, uno o más casetes de delección, o un inserto de ácido nucleico que comprende un ácido nucleico flanqueado con secuencias diana de recombinación sitio específica como se describió anteriormente.

C. Promotores

Varias secuencias de ácidos nucleicos descritas en la presente descripción pueden unirse operativamente a promotores. Tales promotores pueden ser activos, por ejemplo, en una célula pluripotente de rata, eucariota, de mamífero, de mamífero no humano, humana, de roedor, de ratón, o de hámster. También puede usarse un promotor activo en un embrión en etapa de una célula. Un promotor puede ser, por ejemplo, un promotor activo de manera constitutiva, un promotor condicional, un promotor inducible, un promotor restringido temporalmente (por ejemplo, un promotor regulado por el desarrollo), o un promotor restringido espacialmente (por ejemplo, un promotor específico de célula o específico de tejido). Los ejemplos de promotores pueden encontrarse, por ejemplo, en el documento WO 2013/176772.

Los ejemplos de promotores inducibles incluyen, por ejemplo, promotores regulados químicamente y promotores regulados físicamente. Los promotores regulados químicamente incluyen, por ejemplo, promotores regulados por alcohol (por ejemplo, un promotor del gen de la alcohol deshidrogenasa (alcA)), promotores regulados por tetraciclina (por ejemplo, un promotor sensible a la tetraciclina, una secuencia operadora 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 estrógeno, 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 la luz (por ejemplo, un promotor inducible por la luz o un promotor reprimible por la luz).

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

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

Un promotor puede seleccionarse además en base al tipo celular. Por ejemplo, varios promotores conocidos son útiles en una célula eucariota, una célula de mamífero, una célula de mamífero no humano, una célula pluripotente, una célula pluripotente no humana, una célula pluripotente humana, una célula ES humana, una célula madre adulta humana, una célula progenitora humana restringida por el desarrollo, una célula iPS humana, una célula humana, una célula de roedor, una célula de rata, una célula de ratón, una célula de hámster, un fibroblasto, o una célula CHO.

IV. Métodos para modificar genomas y producir animales no humanos genéticamente modificados

A. Métodos para modificar un genoma

Se proporcionan varios métodos para modificar un genoma dentro de una célula a través del uso de dos ARN guías para reconocer regiones diferentes dentro de un solo locus genómico diana. Se proporcionan además métodos que usan dos o más ARN guías (por ejemplo, tres ARN guías o cuatro ARN guías) para reconocer regiones diferentes dentro de un solo locus genómico diana. Los métodos pueden ocurrir *in vitro*, *ex vivo*, o *in vivo*. Tales métodos promueven la creación de modificaciones genéticas bialélicas y pueden comprender el colapso del genoma u otras modificaciones dirigidas tales como la delección simultánea de una secuencia de ácido nucleico dentro del genoma y su reemplazo con una secuencia de ácido nucleico exógena.

La modificación génica dirigida por recombinación homóloga entre un vector de transformación y un locus diana puede ser muy ineficiente, especialmente en tipos celulares distintos a las células madre embrionarias de roedor. El uso de un vector de transformación en combinación con una ruptura bicatenaria del ADN dirigida por nucleasa en el locus diana puede mejorar enormemente la eficiencia de transformación heterocigótica para modificaciones simples, tales como delecciones o inserciones pequeñas.

La combinación de un vector de transformación con un sistema CRISPR/nucleasa Cas9 guiado por un ARN guía (ARNg) también puede aumentar la eficiencia de transformación heterocigótica para modificaciones génicas muy grandes y de baja eficiencia, tales como la delección de un gen de ratón y el reemplazo simultáneo con su equivalente humano

(humanización). Tales modificaciones pueden involucrar deleciones e inserciones muy grandes (por ejemplo, >50 kb) (ver la transformación de Lrp5, C5(Hc), Ror1, y Trpa1 en el Ejemplo 1).

5 Durante la reparación dirigida por homología de una o más rupturas bicatenarias generadas por una nucleasa tal como Cas9 en un locus genómico diana, la una o más rupturas se procesan primero para crear un saliente monocatenario en 3' por resección del extremo 5'. Después Rad51 se polimeriza en el ADN monocatenario en busca de una secuencia homóloga, ocurre la invasión de la cadena en el ADN dúplex plantilla homólogo no dañado (por ejemplo, el vector de transformación), y se forma una estructura de lazo D intermedia para facilitar la reparación de la una o más rupturas bicatenarias mediante el uso del ADN homólogo no dañado (por ejemplo, el vector de transformación) como plantilla. Después las secuencias cromosómicas se reemplazan con el inserto de ácido nucleico del vector de transformación mediante un evento de entrecruzamiento doble que involucra las regiones de homología flanqueantes. Varios factores afectan la ejecución adecuada de este proceso, tales como el tamaño del inserto de ácido nucleico, la longitud de las regiones homólogas a los brazos de homología del vector de transformación, y las posiciones de las regiones homólogas a los brazos de homología del vector de transformación (por ejemplo, en relación con la una o más rupturas bicatenarias).

15 A medida que el inserto de ácido nucleico o la secuencia eliminada en el locus genómico diana aumenta, el proceso de resección se vuelve más impredecible, la estabilidad de la estructura de lazo D intermedia disminuye y se vuelve más impredecible, y el éxito del proceso de recombinación en general disminuye y se vuelve más impredecible. Por ejemplo, a medida que el tamaño de la modificación dirigida aumenta, el riesgo de recombinación interna aumenta, particularmente cuando existe similitud entre la secuencia que se reemplaza y la secuencia que se inserta. Cuando ocurre dicha recombinación interna, el intercambio por recombinación homóloga tiene lugar internamente a la región diana prevista, y el inserto de ácido nucleico completo no se incorpora en el locus genómico diana. Además, el pensamiento convencional es que la eficiencia de las inserciones mediadas por HR disminuye a medida que la distancia entre la ruptura bicatenaria y el sitio de mutación o inserción aumenta (por ejemplo, más de 100 pb o 200 pb). Ver Beumer y otros (2013) Genes\Genomes\Genetics 3:657-664; Elliott y otros (1998) Mol. Cell. Biol. 18:93-101; y Byrne y otros (2015) Nucleic Acids Research 43(3):e21.

30 El logro de una modificación génica dirigida que crea una deleción grande en un locus genómico diana e inserta simultáneamente un pedazo grande de ADN extraño requiere la formación de una estructura omega doble como un intermediario de recombinación. Mientras mayor sea la modificación, menor será la estabilidad de la estructura. En los tipos celulares distintos a los embriones en etapa de una célula, pueden usarse LTVEC que tienen una suma total de 10 kb o más de homología total. Los LTVEC con brazos de homología que tienen una suma total de 10 kb o más de homología total aumentan la estabilidad del intermediario de recombinación omega doble para facilitar una deleción grande simultánea mediada por nucleasa y su reemplazo con un inserto de ácido nucleico grande y hacen posible además no solo que las rupturas bicatenarias adyacentes a las regiones de homología mejoren la eficiencia de transformación sino que también hacen posible que aumente la eficiencia de transformación de aquellas alejadas de las regiones de homología.

40 Para las modificaciones génicas que involucran humanizaciones muy grandes, la combinación de un vector de transformación con un sistema CRISPR/nucleasa Cas9 guiado por dos ARNg puede mejorar además la eficiencia de transformación más allá de la lograda con un ARNg (ver las humanizaciones de Lrp5, C5 (Hc), Ror1, y Trpa1 en el Ejemplo 1). El uso de dos ARNg produce resultados inesperados en este sentido. En comparación con la transformación con un ARNg, que produce modificaciones bialélicas a una frecuencia baja o no las produce en absoluto, la transformación con dos ARNg da como resultado la creación de células con transformación homocigótica, células con deleción homocigótica, y células con transformación heterocigótica combinada (que incluyen células con transformación hemicigótica) a una tasa significativamente aumentada.

50 El método para crear tres tipos de alelos, con transformación homocigótica, con deleción homocigótica, y con transformación heterocigótica combinada (particularmente con transformación hemicigótica), en un solo experimento de transformación proporciona nuevas posibilidades y eficiencias mejoradas para las modificaciones génicas dirigidas. Para una modificación génica simple, tal como la deleción dirigida de un gen en células ES de ratón y su reemplazo con una secuencia que codifica una proteína que reporta la expresión génica (por ejemplo, β -galactosidasa o una proteína fluorescente), la combinación de un vector de transformación con un sistema CRISPR/Cas9 guiado por dos ARNg mejora la producción de células ES con transformación heterocigótica, que pueden usarse después para producir ratones de generación F0 completamente derivados de células ES mediante el método VelociMouse®. Ver Poueymirou y otros (2007) Nat. Biotech. 25:91-99. Estos ratones son útiles para estudiar la expresión génica específica de tejido con el alelo reportero sustituido. Los clones de células ES con transformación homocigótica producidos en el mismo experimento pueden convertirse a VelociMice con una deleción génica con transformación homocigótica, que puede estudiarse en cuanto a las consecuencias fenotípicas de la desactivación génica así como la expresión génica del reportero. La producción de VelociMice a partir de células ES que tienen una deleción homocigótica inducida por CRISPR del gen diana hace posible la comprobación del fenotipo de desactivación génica observado en los ratones con transformación homocigótica y puede revelar las diferencias fenotípicas entre una deleción génica limpia y una deleción acompañada por la inserción del reportero y un casete de selección por fármaco. Los clones de células ES heterocigóticas combinadas (y particularmente hemicigóticas) que portan el alelo transformado por deleción e inserción y la deleción inducida por CRISPR hacen posible la producción de VelociMice con las mismas oportunidades para el estudio que los derivados de los clones con

transformación homocigótica y con delección homocigótica. Además, estos ratones pueden cruzarse para establecer líneas de ratones mutantes de delección tanto dirigida como simple a partir de un solo clon de células ES.

5 Estas ventajas tienen valor añadido cuando se extienden al caso de una humanización. Un uso importante de la humanización de un gen de ratón es crear un modelo animal en el cual probar un producto terapéutico específico para seres humanos. Para que una humanización sea un modelo eficaz, el gen de ratón debe eliminarse o inactivarse para evitar las interacciones entre los productos génicos de ratón y humano que pudieran afectar la función biológica o la interacción adecuada con el fármaco. Al mismo tiempo, el gen humano debe ser capaz de sustituir las funciones biológicas de su equivalente de ratón. Estos requisitos pueden probarse mediante la combinación de una nucleasa Cas9 guiada por dos ARNg con un vector de transformación diseñado para hacer una delección simultánea de un gen de ratón y su reemplazo con el gen humano. Los VelociMice derivados de células ES que tienen una humanización dirigida homocigótica pueden compararse con los VelociMice derivados de células ES que tienen una delección homocigótica inducida por CRISPR del gen de ratón. Si la delección de desactivación causa un fenotipo mutante observable y los ratones humanizados no expresan este fenotipo sino que en lugar de esto son normales, entonces el gen humano es capaz de sustituir las funciones biológicas del gen de ratón. Los ratones humanizados homocigóticos o los portadores de una combinación heterocigótica combinada (por ejemplo, hemicigótica) de un alelo humanizado y un alelo con delección inducida por CRISPR pueden usarse como modelos animales para estudiar el mecanismo de acción y la eficacia del producto terapéutico específico para seres humanos. Los VelociMice heterocigóticos combinados (por ejemplo, hemicigóticos) pueden usarse además para generar líneas de ratones tanto humanizados como con desactivación génica por delección mediante cruzamiento convencional. Por lo tanto, a partir de un solo experimento de transformación génica que combina un sistema CRISPR de dos ARNg con un vector de transformación, se producen ratones genéticamente modificados que crean valiosos modelos del gen de ratón para la prueba preclínica de un producto terapéutico y líneas con desactivación génica para estudiar la función biológica del homólogo de ratón de la diana farmacológica humana.

25 (1) Métodos para generar, promover, o aumentar la frecuencia de modificaciones genéticas bialélicas

En la presente descripción se proporcionan métodos para hacer modificaciones bialélicas a un genoma dentro de una célula o para promover o aumentar la frecuencia de modificaciones bialélicas de un genoma dentro de una célula. Tales métodos pueden dar como resultado, por ejemplo, el colapso de un genoma para eliminar una sección grande de ADN genómico entre dos secuencias de ADN genómico que posteriormente se recombinan. Tales métodos también pueden dar como resultado la inserción de un inserto de ácido nucleico o la delección de una sección grande de ADN genómico y su reemplazo con un inserto de ácido nucleico.

35 Los métodos proporcionados en la presente descripción para modificar un genoma dentro de una célula comprenden poner en contacto el genoma con una primera proteína Cas, un primer ARN de CRISPR que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR dentro de un locus genómico diana, un segundo ARN de CRISPR que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana, y un ARNtracr. Opcionalmente, el genoma puede ponerse en contacto además con ARN de CRISPR adicionales que se hibridan con secuencias de reconocimiento de ARN de CRISPR dentro del locus genómico diana, tales como un tercer ARN de CRISPR que se hibrida con una tercera secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana y/o un cuarto ARN de CRISPR que se hibrida con una cuarta secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana. Las modificaciones bialélicas pueden generarse al poner en contacto el genoma con una primera proteína Cas, un primer ARN de CRISPR que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR dentro de un locus genómico diana, un segundo ARN de CRISPR que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana, y un ARNtracr. Como se describe en más detalle a continuación, la proteína Cas, los ARN de CRISPR, y el ARNtracr pueden introducirse en la célula en cualquier forma y por cualquier medio. Del mismo modo, la totalidad o parte de la proteína Cas, los ARN de CRISPR, y el ARNtracr puede introducirse simultáneamente o secuencialmente en cualquier combinación. El contacto del genoma puede ocurrir directamente (es decir, un componente se pone en contacto directamente con el propio genoma) o indirectamente (es decir, un componente interactúa con otro componente que se pone en contacto directamente con el genoma).

55 El genoma puede comprender un par de primer y segundo cromosomas homólogos que comprenden el locus genómico diana. La primera proteína Cas puede escindir uno o ambos de estos cromosomas dentro de una o ambas de la primera y segunda secuencias de reconocimiento de ARN de CRISPR (es decir, en un primer sitio de escisión dentro de la primera secuencia de reconocimiento de ARN de CRISPR y/o en un segundo sitio de escisión dentro de la segunda secuencia de reconocimiento de ARN de CRISPR). Si se usa además un tercer y/o cuarto ARN de CRISPR, la primera proteína Cas puede escindir uno o ambos de estos cromosomas dentro de una o ambas de la tercera y/o cuarta secuencias de reconocimiento de ARN de CRISPR (es decir, en un tercer sitio de escisión dentro de la tercera secuencia de reconocimiento de ARN de CRISPR y/o en un cuarto sitio de escisión dentro de la cuarta secuencia de reconocimiento de ARN de CRISPR). Después los eventos de escisión pueden generar al menos una ruptura bicatenaria en uno o ambos de los cromosomas. Los eventos de escisión pueden generar además al menos dos rupturas bicatenarias en uno o ambos de los cromosomas. Si se usan nickasas Cas, los eventos de escisión pueden generar al menos una ruptura monocatenaria en uno o ambos de los cromosomas, o al menos dos rupturas monocatenarias en uno o ambos de los cromosomas. Si se usa un tercer y/o cuarto ARN de CRISPR, los eventos de escisión pueden generar al menos tres de cuatro rupturas monocatenarias o bicatenarias en uno o ambos de los cromosomas. Después las secuencias de extremos

generadas por las rupturas bicatenarias pueden experimentar recombinación, o después las secuencias de extremos generadas por las rupturas monocatenarias pueden experimentar recombinación. Después puede identificarse una célula que tiene el genoma modificado que comprende la modificación bialélica.

5 Por ejemplo, la primera proteína Cas puede escindir el genoma en un primer sitio de escisión dentro de la primera secuencia de reconocimiento de ARN de CRISPR en el primer y segundo cromosomas homólogos y en un segundo sitio de escisión dentro de la segunda secuencia de reconocimiento de ARN de CRISPR en al menos el primer cromosoma homólogo, lo que genera de este modo secuencias de extremos en el primer y segundo cromosomas homólogos. Después las secuencias de extremos pueden experimentar recombinación para formar un genoma con una modificación bialélica que comprende una modificación dirigida. La modificación dirigida puede comprender una delección entre la primera y
10 segunda secuencias de reconocimiento de ARN de CRISPR en al menos el primer cromosoma.

La primera y segunda secuencias de reconocimiento de ARN de CRISPR pueden estar en cualquier parte dentro de un locus genómico diana. La primera y segunda secuencias de reconocimiento de ARN de CRISPR pueden flanquear cualquier región genómica de interés. Por ejemplo, la primera y segunda secuencias de reconocimiento de ARN de CRISPR pueden flanquear la totalidad o parte de una secuencia codificante de un gen, tal como el locus Lrp5, el locus C5 (Hc), el locus Ror1, o el locus Trpa1. La primera y segunda secuencias de reconocimiento de ARN de CRISPR pueden flanquear además la totalidad o parte de una secuencia codificante del gen Cmah. Alternativamente, la primera y segunda secuencias de reconocimiento de ARN de CRISPR pueden flanquear una secuencia no codificante, tal como un elemento regulador (por ejemplo, un promotor), o tanto secuencias codificantes como no codificantes. La tercera y cuarta secuencias de reconocimiento de ARN de CRISPR pueden estar, por ejemplo, en cualquier parte dentro de la región genómica de interés que está flanqueada por la primera y segunda secuencias de reconocimiento de ARN de CRISPR.

A modo de ejemplo, la tercera secuencia de reconocimiento de ARN de CRISPR puede estar adyacente a la primera secuencia de reconocimiento de ARN de CRISPR, y la cuarta secuencia de reconocimiento de ARN de CRISPR puede estar adyacente a la segunda secuencia de reconocimiento de ARN de CRISPR. Por lo tanto, la primera y tercera secuencias de reconocimiento de ARN de CRISPR pueden ser un primer par de secuencias de reconocimiento de ARN de CRISPR, y la segunda y cuarta secuencias de reconocimiento de ARN de CRISPR pueden ser un segundo par de secuencias de reconocimiento de ARN de CRISPR. Por ejemplo, la primera y tercera secuencias de reconocimiento de ARN de CRISPR (y/o la segunda y cuarta secuencias de reconocimiento de ARN de CRISPR) pueden encontrarse separadas por aproximadamente 25 pb a aproximadamente 50 pb, aproximadamente 50 pb a aproximadamente 100 pb, aproximadamente 100 pb a aproximadamente 150 pb, aproximadamente 150 pb a aproximadamente 200 pb, aproximadamente 200 pb a aproximadamente 250 pb, aproximadamente 250 pb a aproximadamente 300 pb, aproximadamente 300 pb a aproximadamente 350 pb, aproximadamente 350 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 450 pb, aproximadamente 450 pb a aproximadamente 500 pb, aproximadamente 500 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 700 pb, aproximadamente 700 pb a aproximadamente 800 pb, aproximadamente 800 pb a aproximadamente 900 pb, aproximadamente 900 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 1,5 kb, aproximadamente 1,5 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 2,5 kb, aproximadamente 2,5 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 3,5 kb, aproximadamente 3,5 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 4,5 kb, aproximadamente 4,5 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 6 kb, aproximadamente 6 kb a aproximadamente 7 kb, aproximadamente 7 kb a aproximadamente 8 kb, aproximadamente 8 kb a aproximadamente 9 kb, o aproximadamente 9 kb a aproximadamente 10 kb. A modo de ejemplo, la primera y tercera secuencias de reconocimiento de ARN de CRISPR (y/o la segunda y cuarta secuencias de reconocimiento de ARN de CRISPR) pueden encontrarse separadas por aproximadamente 100 pb a aproximadamente 1 kb. Alternativamente, la primera y tercera secuencias de reconocimiento de ARN de CRISPR (y/o la segunda y cuarta secuencias de reconocimiento de ARN de CRISPR) pueden solaparse.

El primer par de secuencias de reconocimiento de ARN de CRISPR puede ubicarse cerca del extremo 5' del locus genómico diana y el segundo par puede ubicarse cerca del extremo 3' del locus genómico diana. Alternativamente, el primer y segundo pares pueden ubicarse ambos cerca del extremo 5' del locus genómico diana o pueden ubicarse ambos cerca del extremo 3' del locus diana. Alternativamente, uno o ambos de los pares pueden ubicarse internamente dentro del locus genómico diana. Por ejemplo, la primera secuencia de reconocimiento de ARN de CRISPR o el primer par de secuencias de reconocimiento de ARN de CRISPR puede estar menos de 25 pb, menos de 50 pb, menos de 100 pb, menos de 150 pb, menos de 200 pb, menos de 250 pb, menos de 300 pb, menos de 350 pb, menos de 400 pb, menos de 450 pb, menos de 500 pb, menos de 600 pb, menos de 700 pb, menos de 800 pb, menos de 900 pb, menos de 1 kb, menos de 2 kb, menos de 3 kb, menos de 4 kb, menos de 5 kb, o menos de 10 kb del extremo 5' del locus genómico diana. Del mismo modo, la segunda secuencia de reconocimiento de ARN de CRISPR o el primer par de secuencias de reconocimiento de ARN de CRISPR puede estar menos de 25 pb, menos de 50 pb, menos de 100 pb, menos de 150 pb, menos de 200 pb, menos de 250 pb, menos de 300 pb, menos de 350 pb, menos de 400 pb, menos de 450 pb, menos de 500 pb, menos de 600 pb, menos de 700 pb, menos de 800 pb, menos de 900 pb, menos de 1 kb, menos de 2 kb, menos de 3 kb, menos de 4 kb, menos de 5 kb, o menos de 10 kb del extremo 3' del locus genómico diana.

Alternativamente, la primera secuencia de reconocimiento de ARN de CRISPR o el primer par de secuencias de reconocimiento de ARN de CRISPR puede estar, por ejemplo, al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, 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 del extremo 5' del locus genómico diana. Del mismo modo, la segunda secuencia de reconocimiento de ARN de CRISPR o el primer par de secuencias de reconocimiento de ARN de CRISPR puede estar, por ejemplo, al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, 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 del extremo 3' del locus genómico diana.

Alternativamente, la primera secuencia de reconocimiento de ARN de CRISPR o el primer par de secuencias de reconocimiento de ARN de CRISPR puede estar, por ejemplo, aproximadamente 25 pb a aproximadamente 50 pb, aproximadamente 50 pb a aproximadamente 100 pb, aproximadamente 100 pb a aproximadamente 150 pb, aproximadamente 150 pb a aproximadamente 200 pb, aproximadamente 200 pb a aproximadamente 250 pb, aproximadamente 250 pb a aproximadamente 300 pb, aproximadamente 300 pb a aproximadamente 350 pb, aproximadamente 350 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 450 pb, aproximadamente 450 pb a aproximadamente 500 pb, aproximadamente 500 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 700 pb, aproximadamente 700 pb a aproximadamente 800 pb, aproximadamente 800 pb a aproximadamente 900 pb, aproximadamente 900 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 6 kb, aproximadamente 6 kb a aproximadamente 7 kb, aproximadamente 7 kb a aproximadamente 8 kb, aproximadamente 8 kb a aproximadamente 9 kb, aproximadamente 9 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 30 kb, aproximadamente 30 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 50 kb, aproximadamente 50 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 70 kb, aproximadamente 70 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 90 kb, o aproximadamente 90 kb a aproximadamente 100 kb del extremo 5' del locus genómico diana. Del mismo modo, la segunda secuencia de reconocimiento de ARN de CRISPR o el primer par de secuencias de reconocimiento de ARN de CRISPR puede estar, por ejemplo, aproximadamente 25 pb a aproximadamente 50 pb, aproximadamente 50 pb a aproximadamente 100 pb, aproximadamente 100 pb a aproximadamente 150 pb, aproximadamente 150 pb a aproximadamente 200 pb, aproximadamente 200 pb a aproximadamente 250 pb, aproximadamente 250 pb a aproximadamente 300 pb, aproximadamente 300 pb a aproximadamente 350 pb, aproximadamente 350 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 450 pb, aproximadamente 450 pb a aproximadamente 500 pb, aproximadamente 500 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 700 pb, aproximadamente 700 pb a aproximadamente 800 pb, aproximadamente 800 pb a aproximadamente 900 pb, aproximadamente 900 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 6 kb, aproximadamente 6 kb a aproximadamente 7 kb, aproximadamente 7 kb a aproximadamente 8 kb, aproximadamente 8 kb a aproximadamente 9 kb, aproximadamente 9 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 30 kb, aproximadamente 30 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 50 kb, aproximadamente 50 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 70 kb, aproximadamente 70 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 90 kb, o aproximadamente 90 kb a aproximadamente 100 kb del extremo 3' del locus genómico diana.

El primer y segundo sitios de escisión o la primera y segunda secuencias de reconocimiento de ARN de CRISPR pueden encontrarse separados, por ejemplo, por aproximadamente 1 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 150 kb, aproximadamente 150 kb a aproximadamente 200 kb, aproximadamente 200 kb a aproximadamente 300 kb, aproximadamente 300 kb a aproximadamente 400 kb, aproximadamente 400 kb a aproximadamente 500 kb, aproximadamente 500 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 1,5 Mb, aproximadamente 1,5 Mb a aproximadamente 2 Mb, aproximadamente 2 Mb a aproximadamente 2,5 Mb, o aproximadamente 2,5 Mb a aproximadamente 3 Mb. El primer y segundo sitios de escisión o la primera y segunda secuencias de reconocimiento de ARN de CRISPR pueden encontrarse separados además, por ejemplo, por aproximadamente 3 Mb a aproximadamente 4 Mb, aproximadamente 4 Mb a aproximadamente 5 Mb, aproximadamente 5 Mb a aproximadamente 10 Mb, aproximadamente 10 Mb a aproximadamente 20 Mb, aproximadamente 20 Mb a aproximadamente 30 Mb, aproximadamente 30 Mb a aproximadamente 40 Mb, aproximadamente 40 Mb a aproximadamente 50 Mb, aproximadamente 50 Mb a aproximadamente 60 Mb, aproximadamente 60 Mb a aproximadamente 70 Mb, aproximadamente 70 Mb a aproximadamente 80 Mb, aproximadamente 80 Mb a aproximadamente 90 Mb, o aproximadamente 90 Mb a aproximadamente 100 Mb. El primer y segundo sitios de escisión o la primera y segunda secuencias de reconocimiento de ARN de CRISPR pueden encontrarse separados además, por ejemplo, por aproximadamente 25 pb a aproximadamente 50 pb, aproximadamente 50 pb a aproximadamente 100 pb, aproximadamente 100 pb a aproximadamente 150 pb, aproximadamente 150 pb a aproximadamente 200 pb, aproximadamente 200 pb a aproximadamente 250 pb, aproximadamente 250 pb a aproximadamente 300 pb, aproximadamente 300 pb a aproximadamente 350 pb, aproximadamente 350 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 450 pb, aproximadamente 450 pb a aproximadamente 500 pb, aproximadamente 500 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 700 pb,

aproximadamente 700 pb a aproximadamente 800 pb, aproximadamente 800 pb a aproximadamente 900 pb, o aproximadamente 900 pb a aproximadamente 1 kb. Del mismo modo, el primer par de secuencias de reconocimiento de ARN de CRISPR puede encontrarse separado del segundo par secuencias de reconocimiento de ARN de CRISPR, por ejemplo, por aproximadamente 25 pb a aproximadamente 50 pb, aproximadamente 50 pb a aproximadamente 100 pb, aproximadamente 100 pb a aproximadamente 150 pb, aproximadamente 150 pb a aproximadamente 200 pb, aproximadamente 200 pb a aproximadamente 250 pb, aproximadamente 250 pb a aproximadamente 300 pb, aproximadamente 300 pb a aproximadamente 350 pb, aproximadamente 350 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 450 pb, aproximadamente 450 pb a aproximadamente 500 pb, aproximadamente 500 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 700 pb, aproximadamente 700 pb a aproximadamente 800 pb, aproximadamente 800 pb a aproximadamente 900 pb, aproximadamente 900 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 150 kb, aproximadamente 150 kb a aproximadamente 200 kb, aproximadamente 200 kb a aproximadamente 300 kb, aproximadamente 300 kb a aproximadamente 400 kb, aproximadamente 400 kb a aproximadamente 500 kb, aproximadamente 500 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 1,5 Mb, aproximadamente 1,5 Mb a aproximadamente 2 Mb, aproximadamente 2 Mb a aproximadamente 2,5 Mb, aproximadamente 2,5 Mb a aproximadamente 3 Mb, aproximadamente 3 Mb a aproximadamente 4 Mb, aproximadamente 4 Mb a aproximadamente 5 Mb, aproximadamente 5 Mb a aproximadamente 10 Mb, aproximadamente 10 Mb a aproximadamente 20 Mb, aproximadamente 20 Mb a aproximadamente 30 Mb, aproximadamente 30 Mb a aproximadamente 40 Mb, aproximadamente 40 Mb a aproximadamente 50 Mb, aproximadamente 50 Mb a aproximadamente 60 Mb, aproximadamente 60 Mb a aproximadamente 70 Mb, aproximadamente 70 Mb a aproximadamente 80 Mb, aproximadamente 80 Mb a aproximadamente 90 Mb, o aproximadamente 90 Mb a aproximadamente 100 Mb.

Alternativamente, el primer y segundo sitios de escisión o la primera y segunda secuencias de reconocimiento de ARN de CRISPR pueden encontrarse separados, por ejemplo, por al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, 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 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, 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. Del mismo modo, el primer par de secuencias de reconocimiento de ARN de CRISPR puede encontrarse separado del segundo par secuencias de reconocimiento de ARN de CRISPR, por ejemplo, por al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, 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 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, 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.

Alternativamente, el primer y segundo sitios de escisión o la primera y segunda secuencias de reconocimiento de ARN de CRISPR pueden encontrarse separados por menos de 25 pb, menos de 50 pb, menos de 100 pb, menos de 150 pb, menos de 200 pb, menos de 250 pb, menos de 300 pb, menos de 350 pb, menos de 400 pb, menos de 450 pb, menos de 500 pb, menos de 600 pb, menos de 700 pb, menos de 800 pb, menos de 900 pb, menos de 1 kb, menos de 2 kb, menos de 3 kb, menos de 4 kb, menos de 5 kb, o menos de 10 kb. Del mismo modo, el primer par de secuencias de reconocimiento de ARN de CRISPR puede encontrarse separado del segundo par secuencias de reconocimiento de ARN de CRISPR, por ejemplo, por menos de 25 pb, menos de 50 pb, menos de 100 pb, menos de 150 pb, menos de 200 pb, menos de 250 pb, menos de 300 pb, menos de 350 pb, menos de 400 pb, menos de 450 pb, menos de 500 pb, menos de 600 pb, menos de 700 pb, menos de 800 pb, menos de 900 pb, menos de 1 kb, menos de 2 kb, menos de 3 kb, menos de 4 kb, menos de 5 kb, o menos de 10 kb.

Las secuencias de extremos creadas por escisión del genoma en el primer y/o segundo sitios de escisión pueden ser extremos romos o extremos escalonados, y la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR puede incluir la totalidad o parte de la secuencia de ácido nucleico entre y que incluye la primera y segunda secuencias de reconocimiento de ARN de CRISPR. Del mismo modo, las secuencias de extremos creadas por escisión del genoma en el tercer y/o cuarto sitios de escisión pueden ser extremos romos o extremos escalonados. Por ejemplo, la delección puede incluir solamente una porción de la secuencia de ácido nucleico entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR y/o solamente una porción de la primera secuencia de reconocimiento de ARN de CRISPR y/o la segunda secuencia de reconocimiento de ARN de CRISPR. Alternativamente, la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR puede incluir la totalidad de la secuencia de ácido nucleico entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR. Del mismo modo, la delección puede incluir la primera secuencia de reconocimiento de ARN de CRISPR y/o la segunda secuencia de reconocimiento de ARN de CRISPR, o porciones de estas. En algunos casos, la delección comprende además secuencias ubicadas fuera de la primera y segunda secuencias de reconocimiento de ARN de CRISPR (es decir, secuencias que no incluyen y no están entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR).

La delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR puede ser de cualquier longitud. Por ejemplo, el ácido nucleico eliminado 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, 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 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, el ácido nucleico eliminado puede ser, por ejemplo, de al menos 10 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 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, al menos 200 kb, al menos 250 kb, al menos 300 kb, al menos 350 kb, al menos 400 kb, al menos 450 kb, o al menos 500 kb o más. En algunos casos, el ácido nucleico eliminado puede ser de 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 10 Mb, al menos 20 Mb, al menos 30 Mb, al menos 40 Mb, al menos 50 Mb, al menos 60 Mb, al menos 70 Mb, al menos 80 Mb, al menos 90 Mb, o al menos 100 Mb (por ejemplo, la mayor parte de un cromosoma).

La delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR puede ser una delección precisa en donde el ácido nucleico eliminado consiste solamente en la secuencia de ácido nucleico entre el primer y segundo sitios de escisión de proteína Cas de manera que no existan delecciones o inserciones adicionales en el locus genómico diana modificado. La delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR también puede ser una delección imprecisa que se extiende más allá del primer y segundo sitios de escisión de proteína Cas, consistente con la reparación imprecisa por unión de extremos no homólogos (NHEJ), lo que da como resultado delecciones y/o inserciones adicionales en el locus genómico modificado. Por ejemplo, la delección puede extenderse aproximadamente 1 pb, aproximadamente 2 pb, aproximadamente 3pb, aproximadamente 4 pb, aproximadamente 5 pb, aproximadamente 10 pb, aproximadamente 20 pb, aproximadamente 30 pb, aproximadamente 40 pb, aproximadamente 50 pb, aproximadamente 100 pb, aproximadamente 200 pb, aproximadamente 300 pb, aproximadamente 400 pb, aproximadamente 500 pb, o más allá del primer y segundo sitios de escisión de proteína Cas. Del mismo modo, el locus genómico modificado puede comprender inserciones adicionales consistentes con la reparación imprecisa por NHEJ, tales como inserciones de aproximadamente 1 pb, aproximadamente 2 pb, aproximadamente 3 pb, aproximadamente 4 pb, aproximadamente 5 pb, aproximadamente 10 pb, aproximadamente 20 pb, aproximadamente 30 pb, aproximadamente 40 pb, aproximadamente 50 pb, aproximadamente 100 pb, aproximadamente 200 pb, aproximadamente 300 pb, aproximadamente 400 pb, aproximadamente 500 pb, o más.

El contacto puede ocurrir en ausencia de una secuencia donante exógena o la presencia de una secuencia donante exógena, siempre que si la célula es un embrión en etapa de una célula, la secuencia donante exógena tiene una longitud de no más de 5 kb. Las moléculas o secuencias exógenas incluyen moléculas o secuencias que normalmente no están presentes en una célula. La presencia normal incluye la presencia con respecto a la etapa del desarrollo particular y las condiciones del entorno de la célula. Una molécula o secuencia exógena, por ejemplo, puede incluir una versión mutada de una secuencia endógena correspondiente dentro de la célula, tal como una versión humanizada de la secuencia endógena. Por el contrario, las moléculas o secuencias endógenas incluyen moléculas o secuencias que normalmente están presentes en una célula particular en una etapa del desarrollo particular en condiciones del entorno particulares.

La secuencia donante exógena puede estar dentro de un vector de transformación y puede comprender un inserto de ácido nucleico flanqueado por brazos de homología 5' y 3' que corresponden a secuencias diana 5' y 3' dentro del genoma, siempre que si la célula es un embrión en etapa de una célula, el vector de transformación tiene una longitud de no más de 5 kb. En tipos celulares distintos a los embriones en etapa de una célula, el vector de transformación puede ser más largo. En tipos celulares distintos a los embriones en etapa de una célula, el vector de transformación puede ser, por ejemplo, un vector de transformación grande (LTVEC) como se describe en la presente descripción, y puede ser de al menos 10 kb. Por lo tanto, en algunos métodos, el genoma se pone en contacto además con un vector de transformación, y el inserto de ácido nucleico se inserta entre las secuencias diana 5' y 3'.

Alternativamente, la secuencia donante exógena puede comprender brazos de homología 5' y 3' sin inserto de ácido nucleico. Tales vectores de transformación sin inserto de ácido nucleico pueden facilitar delecciones precisas entre las secuencias diana 5' y 3' dentro del genoma. Tales delecciones precisas pueden ser, por ejemplo, de al menos 10 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 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, 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, o al menos 2 Mb o más.

En algunos de tales métodos, los brazos de homología 5' y 3' corresponden a secuencias diana 5' y 3' en el locus genómico diana que comprende la primera secuencia de reconocimiento de ARN de CRISPR del primer ARN de CRISPR y/o la segunda secuencia de reconocimiento de ARN de CRISPR del segundo ARN de CRISPR. La primera y segunda secuencias de reconocimiento de ARN de CRISPR o el primer y segundo sitios de escisión pueden estar adyacentes a la secuencia diana 5', adyacentes a la secuencia diana 3', o no adyacentes ni a la secuencia diana 5' ni la secuencia diana 3'. Alternativamente, la primera secuencia de reconocimiento de ARN de CRISPR o el primer sitio de escisión puede estar adyacente a la secuencia diana 5', y la segunda secuencia de reconocimiento de ARN de CRISPR o el segundo sitio de escisión puede estar adyacente a la secuencia diana 3'. Alternativamente, la primera secuencia de reconocimiento de ARN de CRISPR o el primer sitio de escisión puede estar adyacente a la secuencia diana 5' o la secuencia diana 3', y la segunda secuencia de reconocimiento de ARN de CRISPR o el segundo sitio de escisión puede no estar adyacente ni a la secuencia diana 5' ni la secuencia diana 3'.

Por ejemplo, la primera y/o segunda secuencias de reconocimiento de ARN de CRISPR pueden ubicarse entre las secuencias diana 5' y 3' o pueden estar adyacentes a o en proximidad con la secuencia diana 5' y/o la secuencia diana 3', tal como dentro de 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 20 kb, 30 kb, 40 kb, 50 kb, 60 kb, 70 kb, 80 kb, 90 kb, 100 kb, 110 kb, 120 kb, 130 kb, 140 kb, 150 kb, 160 kb, 170 kb, 180 kb, 190 kb, 200 kb, 250 kb, 300 kb, 350 kb, 400 kb, 450 kb, o 500 kb de las secuencias diana 5' y/o 3'. Del mismo modo, el primer y/o segundo sitios de escisión pueden ubicarse entre las secuencias diana 5' y 3' o pueden estar adyacentes a o en proximidad con la secuencia diana 5' y/o la secuencia diana 3', tal como dentro de 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 20 kb, 30 kb, 40 kb, 50 kb, 60 kb, 70 kb, 80 kb, 90 kb, 100 kb, 110 kb, 120 kb, 130 kb, 140 kb, 150 kb, 160 kb, 170 kb, 180 kb, 190 kb, 200 kb, 250 kb, 300 kb, 350 kb, 400 kb, 450 kb, o 500 kb de las secuencias diana 5' y/o 3'. Por ejemplo, la primera secuencia de reconocimiento de ARN de CRISPR o el primer sitio de escisión puede estar dentro de 50 pb, 100 pb, 200 pb, 300 pb, 400 pb, 500 pb, 600 pb, 700 pb, 800 pb, 900 pb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 20 kb, 30 kb, 40 kb, 50 kb, 60 kb, 70 kb, 80 kb, 90 kb, 100 kb, 110 kb, 120 kb, 130 kb, 140 kb, 150 kb, 160 kb, 170 kb, 180 kb, 190 kb, 200 kb, 250 kb, 300 kb, 350 kb, 400 kb, 450 kb, o 500 kb de la secuencia diana 5' o ambas secuencias diana 5' y 3'. Del mismo modo, la segunda secuencia de reconocimiento de ARN de CRISPR o el segundo sitio de escisión puede estar dentro de 50 pb, 100 pb, 200 pb, 300 pb, 400 pb, 500 pb, 600 pb, 700 pb, 800 pb, 900 pb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 20 kb, 30 kb, 40 kb, 50 kb, 60 kb, 70 kb, 80 kb, 90 kb, 100 kb, 110 kb, 120 kb, 130 kb, 140 kb, 150 kb, 160 kb, 170 kb, 180 kb, 190 kb, 200 kb, 250 kb, 300 kb, 350 kb, 400 kb, 450 kb, o 500 kb de la secuencia diana 3' o ambas secuencias diana 5' y 3'.

Alternativamente, la primera y/o segunda secuencias de reconocimiento de ARN de CRISPR pueden ubicarse al menos 50 pb, al menos 100 pb, al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 600 pb, al menos 700 pb, al menos 800 pb, al menos 900 pb, al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, al menos 6 kb, al menos 7 kb, al menos 8 kb, al menos 9 kb, 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 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, 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 de las secuencias diana 5' y/o 3'. Del mismo modo, el primer y/o segundo sitios de escisión pueden ubicarse al menos 50 pb, al menos 100 pb, al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 600 pb, al menos 700 pb, al menos 800 pb, al menos 900 pb, al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, al menos 6 kb, al menos 7 kb, al menos 8 kb, al menos 9 kb, 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 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, 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 de las secuencias diana 5' y/o 3'. Por ejemplo, la primera secuencia de reconocimiento de ARN de CRISPR o el primer sitio de escisión puede ubicarse al menos 50 pb, al menos 100 pb, al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 600 pb, al menos 700 pb, al menos 800 pb, al menos 900 pb, al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, al menos 6 kb, al menos 7 kb, al menos 8 kb, al menos 9 kb, 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 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, 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 de la secuencia diana 5' o de ambas secuencias diana 5' y 3'. Del mismo modo, la segunda secuencia de reconocimiento de ARN de CRISPR del segundo sitio de escisión puede ubicarse al menos 50 pb, al menos 100 pb, al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 600 pb, al menos 700 pb, al menos 800 pb, al menos 900 pb, al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, al menos 6 kb, al menos 7 kb, al menos 8 kb, al menos 9 kb, 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 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, 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 de la secuencia diana 3' o de ambas secuencias diana 5' y 3'.

Por ejemplo, la primera y/o segunda secuencias de reconocimiento de ARN de CRISPR pueden ubicarse entre aproximadamente 50 pb a aproximadamente 100 pb, aproximadamente 200 pb a aproximadamente 300 pb,

Los métodos descritos en la presente descripción promueven y aumentan la frecuencia de modificaciones bialélicas. En particular, al poner en contacto el genoma con el primer y segundo ARN de CRISPR, la eficiencia de producción de modificaciones bialélicas puede aumentarse en comparación con poner en contacto el genoma con el primer ARN de CRISPR o el segundo ARN de CRISPR solos. La eficiencia de producción de modificaciones bialélicas puede aumentarse además al poner en contacto el genoma con el primer, segundo, y tercer ARN de CRISPR, o el primer, segundo, tercer, y cuarto ARN de CRISPR. Las modificaciones bialélicas incluyen eventos en los que se hace la misma modificación al mismo locus en cromosomas homólogos correspondientes (por ejemplo, en una célula diploide), o en los que se hacen modificaciones diferentes al mismo locus en cromosomas homólogos correspondientes. Los cromosomas homólogos incluyen cromosomas que tienen los mismos genes en los mismos loci pero posiblemente en alelos diferentes (por ejemplo, cromosomas que se aparean durante la meiosis). El término alelo incluye cualquiera de una o más formas alternativas de una secuencia genética. En una célula u organismo diploide, los dos alelos de una secuencia dada típicamente ocupan loci correspondientes en un par de cromosomas homólogos.

Una modificación bialélica puede dar como resultado homocigosidad para una modificación dirigida o heterocigosidad combinada (por ejemplo, hemicigosidad) para la modificación dirigida. Un solo experimento de transformación con una población de células puede producir células que son homocigóticas para una modificación dirigida (por ejemplo, la humanización de un locus), células que son heterocigóticas combinadas para esa modificación dirigida (que incluyen células que son hemicigóticas para la modificación dirigida), y células con colapso homocigótico entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR (es decir, una secuencia de ácido nucleico grande se elimina entre dos secuencias de reconocimiento de ARN de CRISPR). La homocigosidad incluye situaciones en las que ambos alelos de un locus diana (es decir, los alelos correspondientes en ambos cromosomas homólogos) tienen la modificación dirigida. La heterocigosidad combinada incluye situaciones en las que ambos alelos del locus diana (es decir, los alelos en ambos cromosomas homólogos) se han modificado, pero se han modificado de maneras diferentes (por ejemplo, una modificación dirigida en un alelo e inactivación o interrupción del otro alelo). La interrupción de la secuencia de ácido nucleico endógena puede ser el resultado, por ejemplo, cuando una ruptura bicatenaria creada por la proteína Cas se repara por reparación del ADN mediada por unión de extremos no homólogos (NHEJ), que genera un alelo mutante que comprende una inserción o una delección de una secuencia de ácido nucleico y de este modo causa la interrupción de ese locus genómico. Los ejemplos de interrupción incluyen la alteración de un elemento regulador (por ejemplo, un promotor o potenciador), una mutación con cambio de sentido, una mutación sin sentido, una mutación de desplazamiento del marco de lectura, una mutación de truncamiento, una mutación nula, o una inserción o delección de un pequeño número de nucleótidos (por ejemplo, que causa una mutación de desplazamiento del marco de lectura). La interrupción puede dar como resultado la inactivación (es decir, la pérdida de función) o pérdida del alelo.

Por ejemplo, una modificación bialélica puede dar como resultado la heterocigosidad combinada si la célula tiene un alelo con la modificación dirigida y otro alelo que no es capaz de expresarse o de otra manera no es funcional. La heterocigosidad combinada incluye la hemicigosidad. La hemicigosidad incluye situaciones en las que solo un alelo (es decir, un alelo en uno de los dos cromosomas homólogos) del locus diana está presente. Por ejemplo, una modificación bialélica puede dar como resultado la hemicigosidad para una modificación dirigida si la modificación dirigida ocurre en un alelo con una pérdida o delección correspondiente del otro alelo.

En un ejemplo específico, la modificación bialélica puede comprender una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el par de primer y segundo cromosomas homólogos. Las delecciones pueden ocurrir simultáneamente, o la delección puede ocurrir inicialmente en el primer cromosoma homólogo, después la célula logra la homocigosidad mediante el uso del primer cromosoma homólogo como secuencia donante para reparar una o más rupturas bicatenarias en el segundo cromosoma homólogo por medio de recombinación homóloga, tal como por conversión génica. La secuencia de ácido nucleico eliminada en el primer y segundo cromosomas homólogos puede ser igual, parcialmente solapada, o diferente. Alternativamente, la modificación bialélica puede comprender una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo y pérdida del alelo o locus correspondiente en el segundo cromosoma homólogo. Alternativamente, la modificación bialélica puede comprender una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo e inactivación o interrupción del alelo o locus correspondiente entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el segundo cromosoma homólogo.

Si se usa una secuencia donante, la modificación bialélica puede comprender una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR así como una inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el par de primer y segundo cromosomas homólogos, lo que da como resultado de este modo un genoma modificado homocigótico. La delección y la inserción pueden ocurrir simultáneamente en ambos cromosomas, o la delección y la inserción pueden ocurrir inicialmente en el primer cromosoma homólogo, después la célula logra la homocigosidad mediante el uso del primer cromosoma homólogo como secuencia donante para reparar la(s) ruptura(s) bicatenaria(s) en el segundo cromosoma homólogo por medio de recombinación homóloga, tal como por conversión génica. Por ejemplo, sin desear limitarse por ninguna teoría particular, la inserción del inserto de ácido nucleico podría ocurrir en el primer cromosoma homólogo (con o sin escisión por la proteína Cas), y el segundo cromosoma homólogo puede modificarse después mediante un evento de conversión génica estimulado por la escisión por la proteína Cas en el segundo cromosoma homólogo.

Alternativamente, la modificación bialélica puede dar como resultado un genoma modificado heterocigótico combinado. Por ejemplo, la modificación dirigida puede comprender una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer y segundo cromosomas homólogos y una inserción del inserto de ácido nucleico en el primer cromosoma homólogo pero no en el segundo cromosoma homólogo. Alternativamente, la modificación dirigida puede comprender una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR así como la inserción del inserto de ácido nucleico en el primer cromosoma homólogo y la inactivación o interrupción del alelo o locus correspondiente en el segundo cromosoma homólogo. Alternativamente, la modificación bialélica puede dar como resultado un genoma modificado hemicigótico en el que la modificación dirigida puede comprender una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR así como la inserción del inserto de ácido nucleico en el primer cromosoma homólogo y la pérdida o delección del alelo o locus correspondiente en el segundo cromosoma homólogo.

Las modificaciones genéticas dirigidas homocigóticas y heterocigóticas combinadas (particularmente hemicigóticas) son ventajosas porque el proceso para producir animales genéticamente modificados con estas modificaciones (descritos en más detalle a continuación) puede ser más eficiente y consumir menos tiempo. En muchas situaciones, tales como la eliminación de un gen para estudiar el efecto de su ausencia, la simple heterocigosidad de una modificación genética dirigida (es decir, modificación en un alelo y sin cambio en el otro alelo) no es suficiente. Con las estrategias de transformación convencionales, podrían obtenerse animales de generación FO que son heterocigóticos para una delección genómica dirigida grande, pero se requiere el entrecruzamiento posterior de estos animales heterocigóticos para producir animales de generación F1 que son homocigóticos para la delección. Estas etapas de cruzamiento adicionales son costosas y consumen tiempo. La capacidad de crear animales genéticamente modificados de generación FO que son homocigóticos o heterocigóticos combinados (particularmente hemicigóticos) para una modificación genética dirigida da como resultado ganancias de eficiencia significativas y ahorro de tiempo debido a que se requieren menos etapas de cruzamiento.

(2) Conversión génica o pérdida de heterocigosidad

En algunos métodos, el genoma a modificar está dentro de una célula que es heterocigótica para un primer alelo, y el gen se modifica para convertirlo en homocigótico para el primer alelo. El término heterocigótico incluye situaciones en las que un genoma comprende alelos diferentes en uno o más loci cromosómicos correspondientes (por ejemplo, alelos diferentes en loci correspondientes en cromosomas homólogos). El término homocigótico incluye situaciones en las que un genoma comprende el mismo alelo en loci cromosómicos correspondientes (por ejemplo, en cromosomas homólogos correspondientes). En algunos de tales métodos, la célula puede lograr la homocigosidad mediante el uso del primer alelo como secuencia donante para reparar una ruptura bicatenaria en un segundo alelo correspondiente por medio de recombinación homóloga, tal como conversión génica. Típicamente, la extensión de la conversión génica se limita a algunos cientos de pares de bases. Ver, por ejemplo, Kasperek & Humphrey (2011) *Seminars in Cell & Dev. Biol.* 22:886-897. Sin embargo, el uso de pares de ARN guías que dirigen la escisión en sitios de escisión diferentes dentro de un solo locus puede promover y mejorar las capacidades de conversión génica en tramos más largos.

Tales métodos pueden ser útiles en varios contextos. El primer alelo puede comprender una mutación. En algunos métodos, por ejemplo, el primer alelo contiene una modificación genética dirigida deseada. El logro de la homocigosidad para esa modificación genética dirigida puede dar como resultado ahorros de tiempo y costos significativos si, por ejemplo, el objetivo es crear un animal no humano que sea homocigótico para esa modificación. En otros métodos, el primer alelo es un alelo de tipo silvestre de un gen que corresponde a un segundo alelo del gen que causa enfermedad. Alternativamente, el segundo alelo puede comprender cualquier mutación. Los métodos pueden usarse entonces para lograr el objetivo final de la terapia génica de reemplazar el alelo que causa enfermedad con el alelo de tipo silvestre en su locus cromosómico natural.

En algunos de tales métodos para modificar un genoma que es heterocigótico para un primer alelo para convertirlo en homocigótico para el primer alelo, el genoma se pone en contacto con una proteína Cas, un ARNtracr, y un primer ARN de CRISPR que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR dentro de un segundo alelo, y un segundo ARN de CRISPR que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR dentro del segundo alelo, en donde el primer alelo está en un primer cromosoma homólogo y el segundo alelo está en un locus correspondiente en un segundo cromosoma homólogo (es decir, el primer alelo y el segundo alelo pueden ser alelos correspondientes en un par de primer y segundo cromosomas homólogos). Opcionalmente, el genoma puede ponerse en contacto con ARN de CRISPR adicionales (por ejemplo, un tercer ARN de CRISPR, o un tercer y cuarto ARN de CRISPR) que se hibridan con secuencias de reconocimiento de ARN de CRISPR dentro del segundo alelo. La proteína Cas puede escindir una o ambas de la primera y segunda secuencias de reconocimiento de ARN de CRISPR (es decir, en un primer sitio de escisión dentro de la primera secuencia de reconocimiento de ARN de CRISPR y/o en un segundo sitio de escisión dentro de la segunda secuencia de reconocimiento de ARN de CRISPR). La escisión del genoma en el primer y/o segundo sitios de escisión puede crear extremos romos en el ADN genómico o puede crear extremos escalonados. Los sitios de escisión pueden repararse después a través de la recombinación entre el primer y segundo alelos, lo que da como resultado un genoma modificado que es homocigótico para el primer alelo. Después puede identificarse una célula que tiene el genoma modificado.

En algunos métodos, la primera y/o segunda secuencias de reconocimiento de ARN de CRISPR se ubican dentro del segundo alelo pero no dentro del primer alelo. El primer y/o segundo alelos pueden ser alelos de tipo silvestre o pueden comprender modificaciones dirigidas u otras desviaciones de un alelo de tipo silvestre. Por ejemplo, el primer alelo puede comprender una modificación dirigida deseada y el segundo alelo puede ser un alelo de tipo silvestre. Alternativamente, el primer alelo puede ser un alelo de tipo silvestre, y el segundo alelo puede comprender una modificación no deseada, tal como una mutación que causa enfermedad. En algunos de tales métodos, ocurre la reparación génica dirigida o la corrección génica dirigida de manera que la mutación que causa enfermedad en el segundo alelo se corrige por medio de recombinación mediante el uso del primer alelo como secuencia donante.

El primer y segundo sitios de escisión o la primera y segunda secuencias de reconocimiento de ARN de CRISPR pueden encontrarse separados, por ejemplo, por aproximadamente 1 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 150 kb, aproximadamente 150 kb a aproximadamente 200 kb, aproximadamente 200 kb a aproximadamente 300 kb, aproximadamente 300 kb a aproximadamente 400 kb, aproximadamente 400 kb a aproximadamente 500 kb, aproximadamente 500 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 1,5 Mb, aproximadamente 1,5 Mb a aproximadamente 2 Mb, aproximadamente 2 Mb a aproximadamente 2,5 Mb, o aproximadamente 2,5 Mb a aproximadamente 3 Mb.

Alternativamente, el primer y segundo sitios de escisión o la primera y segunda secuencias de reconocimiento de ARN de CRISPR pueden encontrarse separados, por ejemplo, por al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, 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 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, 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.

En algunos métodos, las diferencias de secuencias entre el primer alelo y el segundo alelo abarcan aproximadamente 100 pb a aproximadamente 200 pb, aproximadamente 200 pb a aproximadamente 400 pb, aproximadamente 400 pb a aproximadamente 600 pb, aproximadamente 600 pb a aproximadamente 800 pb, aproximadamente 800 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 80 kb, aproximadamente 80 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 150 kb, aproximadamente 150 kb a aproximadamente 200 kb, aproximadamente 200 kb a aproximadamente 300 kb, aproximadamente 300 kb a aproximadamente 400 kb, aproximadamente 400 kb a aproximadamente 500 kb, aproximadamente 500 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 1,5 Mb, aproximadamente 1,5 Mb a aproximadamente 2 Mb, aproximadamente 2 Mb a aproximadamente 2,5 Mb, o aproximadamente 2,5 Mb a aproximadamente 3 Mb.

En otros métodos, las diferencias entre el primer alelo y el segundo alelo abarcan al menos 100 pb, al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 600 pb, al menos 700 pb, al menos 800 pb, al menos 800 pb, al menos 1 kb, al menos 2 kb, al menos 3 kb, al menos 4 kb, al menos 5 kb, al menos 6 kb, al menos 7 kb, al menos 8 kb, al menos 9 kb, al menos 10 kb, 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, 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.

En otros de tales métodos para modificar un genoma en una célula que es heterocigótica para un primer alelo para convertirla en homocigótica para el primer alelo, el genoma se pone en contacto con una proteína Cas, un ARNtracr, y un primer ARN de CRISPR sin especificidad por alelo que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR. El primer alelo está en un primer cromosoma homólogo, y la secuencia de reconocimiento de ARN de CRISPR es centromérica (es decir, más cercana al centrómero) al locus correspondiente al primer alelo en un segundo cromosoma homólogo. La proteína Cas puede escindir la primera secuencia de reconocimiento de ARN de CRISPR para generar una ruptura bicatenaria. La recombinación puede ocurrir después para modificar la célula para convertirla en homocigótica para el primer alelo.

Opcionalmente, la célula es heterocigótica para uno o más alelos adicionales, la primera secuencia de reconocimiento de ARN de CRISPR es centromérica a los loci correspondientes al uno o más alelos adicionales en el segundo cromosoma homólogo, y la recombinación modifica la célula para convertirla en homocigótica para el uno o más alelos adicionales.

Opcionalmente, el método puede comprender además poner en contacto el genoma con un segundo ARN de CRISPR sin especificidad por alelo que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR centromérica al locus correspondiente al primer alelo en el segundo cromosoma homólogo, en donde la proteína Cas escinde al menos una de la primera y segunda secuencias de reconocimiento de ARN de CRISPR para generar al menos una ruptura bicatenaria. Opcionalmente, el método puede comprender además poner en contacto el genoma con ARN de CRISPR

sin especificidad por alelo adicionales (por ejemplo, un tercer ARN de CRISPR, o un tercer y un cuarto ARN de CRISPR) que se hibridan con secuencias de reconocimiento de ARN de CRISPR centroméricas al locus correspondiente al primer alelo en un segundo cromosoma homólogo. Después puede identificarse una célula que tiene el genoma modificado.

5 En algunos métodos, la primera (o segunda, tercera, o cuarta) secuencias de reconocimiento de ARN de CRISPR se ubican en el segundo cromosoma homólogo pero no en el primer cromosoma homólogo. El primer (o segundo, tercer, o cuarto) sitio de reconocimiento de ARN de CRISPR puede estar de aproximadamente 100 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 10 Mb, aproximadamente 10 Mb a aproximadamente 20 Mb, aproximadamente 20 Mb a aproximadamente 30 Mb, aproximadamente 30 Mb a aproximadamente 40 Mb, aproximadamente 40 Mb a aproximadamente 50 Mb, aproximadamente 50 Mb a aproximadamente 60 Mb, aproximadamente 60 Mb a aproximadamente 70 Mb, aproximadamente 70 Mb a aproximadamente 80 Mb, aproximadamente 80 Mb a aproximadamente 90 Mb, o aproximadamente 90 Mb a aproximadamente 100 Mb del centrómero.

15 El primer alelo y/o el uno o más alelos adicionales pueden comprender una mutación tal como una modificación dirigida. Alternativamente, el primer alelo y/o el uno o más alelos adicionales pueden ser un alelo de tipo silvestre, y los loci correspondientes en el segundo cromosoma homólogo pueden comprender mutaciones tales como una mutación que causa enfermedad. El primer alelo puede estar de aproximadamente 100 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 10 Mb, aproximadamente 10 Mb a aproximadamente 20 Mb, aproximadamente 20 Mb a aproximadamente 30 Mb, aproximadamente 30 Mb a aproximadamente 40 Mb, aproximadamente 40 Mb a aproximadamente 50 Mb, aproximadamente 50 Mb a aproximadamente 60 Mb, aproximadamente 60 Mb a aproximadamente 70 Mb, aproximadamente 70 Mb a aproximadamente 80 Mb, aproximadamente 80 Mb a aproximadamente 90 Mb, o aproximadamente 90 Mb a aproximadamente 100 Mb del primer sitio de reconocimiento de ARN de CRISPR. Alternativamente, el primer alelo puede estar al menos 100 pb, al menos 1 kb, al menos 10 kb, al menos 100 kb, al menos 1 Mb, al menos 10 Mb, al menos 20 Mb, al menos 30 Mb, al menos 40 Mb, al menos 50 Mb, al menos 60 Mb, al menos 70 Mb, al menos 80 Mb, al menos 90 Mb, o al menos 100 Mb o más del primer sitio de reconocimiento de ARN de CRISPR.

30 La proteína Cas puede ser Cas9. Esta puede tener actividad nucleasa en ambas cadenas del ADN bicatenario, o puede ser una nickasa. En algunos métodos, la proteína Cas y el primer ARN de CRISPR no aparecen juntos en la naturaleza.

35 La recombinación puede comprender la pérdida de heterocigosidad telomérica (es decir, hacia el telómero) de la ruptura bicatenaria (por ejemplo, una conversión génica o pérdida de heterocigosidad polar o direccional). La región del segundo cromosoma homólogo que se reemplaza por pérdida de heterocigosidad puede ser de aproximadamente 100 pb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 10 kb, aproximadamente 10 kb a aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 1 Mb, aproximadamente 1 Mb a aproximadamente 10 Mb, aproximadamente 10 Mb a aproximadamente 20 Mb, aproximadamente 20 Mb a aproximadamente 30 Mb, aproximadamente 30 Mb a aproximadamente 40 Mb, aproximadamente 40 Mb a aproximadamente 50 Mb, aproximadamente 50 Mb a aproximadamente 60 Mb, aproximadamente 60 Mb a aproximadamente 70 Mb, aproximadamente 70 Mb a aproximadamente 80 Mb, aproximadamente 80 Mb a aproximadamente 90 Mb, o aproximadamente 90 Mb a aproximadamente 100 Mb. Alternativamente, la región del segundo cromosoma homólogo que se reemplaza puede ser de al menos 100 pb, al menos 1 kb, al menos 10 kb, al menos 100 kb, al menos 1 Mb, al menos 10 Mb, al menos 20 Mb, al menos 30 Mb, al menos 40 Mb, al menos 50 Mb, al menos 60 Mb, al menos 70 Mb, al menos 80 Mb, al menos 90 Mb, o al menos 100 Mb o más. Por ejemplo, puede reemplazarse la mayor parte del cromosoma.

B. Métodos para producir un animal no humano genéticamente modificado

50 Los animales no humanos genéticamente modificados pueden generarse con el empleo de los varios 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 mediante el uso de los métodos descritos anteriormente; (2) seleccionar la célula pluripotente genéticamente modificada; (3) introducir la célula pluripotente genéticamente modificada en un embrión huésped; y (4) implantar el embrión huésped que comprende la célula pluripotente genéticamente modificada en una madre sustituta. Se genera una progenie a partir de la célula pluripotente genéticamente modificada. La célula donante puede introducirse en un embrión huésped en cualquier etapa, tal como la etapa de blastocisto o la etapa de premórula (es decir, la etapa de 4 células o la etapa de 8 células). Se genera progenie que es capaz de transmitir la modificación genética a través de la línea germinal. La célula pluripotente puede ser una célula ES (por ejemplo, una célula ES de ratón o una célula ES de rata) como se analizó en otra parte en la presente descripción. Ver, por ejemplo, la patente de Estados Unidos núm. 7,294,754.

60 Alternativamente, el método para producir un animal no humano genéticamente modificado puede comprender: (1) modificar el genoma de un embrión en etapa de una célula mediante el uso de los métodos descritos anteriormente; (2) seleccionar el embrión genéticamente modificado; y (3) implantar el embrión genéticamente modificado en una madre sustituta. Se genera progenie que es capaz de transmitir la modificación genética a través de la línea germinal.

Pueden usarse además técnicas de transferencia nuclear para generar los animales mamíferos no humanos. En resumen, los métodos de transferencia nuclear pueden incluir las etapas de: (1) enucleo un oocito o proporcionar un oocito enucleado; (2) aislar o proporcionar una célula o núcleo donante a combinar con el oocito enucleado; (3) insertar la célula o núcleo en el oocito enucleado para formar una célula reconstituida; (4) implantar la célula reconstituida en el útero de un animal para formar un embrión; y (5) permitir que el embrión se desarrolle. En tales métodos, los oocitos generalmente se obtienen de animales fallecidos, aunque también pueden aislarse de oviductos y/u ovarios de animales vivos. Los oocitos pueden madurarse en una variedad de medios conocidos para los expertos en la técnica antes de la enucleación. La enucleación del oocito puede realizarse en un número de maneras bien conocidas para los expertos en la técnica. La inserción de la célula o núcleo donante en el oocito enucleado para formar una célula reconstituida puede ser por microinyección de una célula donante debajo de la zona pelúcida antes de la fusión. La fusión puede inducirse mediante la aplicación de un pulso eléctrico de DC a través del plano de contacto/fusión (electrofusión), por exposición de las células a sustancias químicas que promueven la fusión, tales como polietilenglicol, o por medio de un virus inactivado, tal como el virus Sendai. Una célula reconstituida puede activarse por medios eléctricos y/o no eléctricos antes, durante, y/o después de la fusión del donante nuclear y el oocito receptor. Los métodos de activación incluyen pulsos eléctricos, choque inducido químicamente, penetración por esperma, aumento de los niveles de cationes divalentes en el oocito, y reducción de la fosforilación de proteínas celulares (igual que por medio de inhibidores de quinasas) en el oocito. Las células reconstituidas activadas, o embriones, pueden cultivarse en medio bien conocido para los expertos en la técnica y después transferirse al útero de un animal. Ver, por ejemplo, los documentos US20080092249, WO/1999/005266A2, US20040177390, WO/2008/017234A1, y la patente de Estados Unidos núm. 7,612,250.

Algunos métodos para producir un animal no humano genéticamente modificado comprenden métodos para producir un animal no humano de generación FO. Tales métodos pueden comprender poner en contacto el genoma en una célula ES no humana con una proteína Cas, un primer ARN de CRISPR que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR, un segundo ARN de CRISPR que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR, y un ARNtracr. La proteína Cas puede escindir el genoma dentro de la primera y segunda secuencias de reconocimiento de ARN de CRISPR para generar secuencias de extremos. Las secuencias de extremos pueden experimentar recombinación para formar un genoma con una modificación dirigida, y la modificación dirigida puede comprender una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR.

Los métodos pueden comprender además: (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. Después la madre sustituta puede producir el animal no humano de generación FO 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 FO. Los animales que portan el locus genómico genéticamente modificado pueden identificarse por medio de un ensayo de modificación de alelo (MOA) como se describe en la presente descripción.

Los varios métodos proporcionados en la presente descripción permiten la generación de un animal FO no humano genéticamente modificado en donde las células del animal FO genéticamente modificado comprenden la modificación dirigida. Se reconoce que en dependencia del método usado para generar el animal FO, el número de células dentro del animal FO que tienen la secuencia de nucleótidos de interés y carecen del casete de recombinasa y/o el casete de selección variará. La introducción de las células ES donantes en un embrión en etapa de premórula de un organismo correspondiente (por ejemplo, un embrión de ratón en etapa de 8 células) por medio de por ejemplo, el método VELOCIMOUSE® permite que un mayor porcentaje de la población de células del animal FO comprenda células que tienen la secuencia de nucleótidos de interés que comprende la modificación genética dirigida. En casos específicos, al menos 50 %, 60 %, 65 %, 70 %, 75 %, 85 %, 86 %, 87 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de la contribución celular del animal FO no humano comprende una población de células que tiene la modificación dirigida. En otros casos, al menos una o más de las células germinales del animal FO tienen la modificación dirigida.

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

En algunos casos, el animal FO generado por los métodos y composiciones descritos en la presente descripción puede cruzarse con un animal de tipo silvestre para generar una generación F1 que es heterocigótica para la modificación dirigida. Después los animales de la generación F1 pueden cruzarse entre sí para generar un animal F2 homocigótico para la modificación dirigida. La progenie F1 puede genotipificarse mediante el uso de cebadores y/o sondas específicos para determinar si la modificación genética dirigida está presente.

C. Genomas y loci genómicos diana

Un genoma o locus genómico diana modificado mediante los métodos descritos en la presente descripción puede incluir cualquier segmento o región de ADN dentro de una célula. El genoma o locus genómico diana puede ser nativo para 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. Tales segmentos de ADN heterólogos o exógenos pueden incluir transgenes, casetes de expresión, polinucleótido que codifica marcadores de selección, o regiones heterólogas o exógenas de ADN genómico.

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

10 D. Formas de Cas9 y ARN guía

En algunos métodos, el contacto del genoma comprende introducir una o más proteínas Cas, uno o más ARN de CRISPR, y uno o más ARNtracr en la célula. La introducción puede llevarse a cabo por cualquier medio, y uno o más de los componentes (por ejemplo, dos de los componentes, o la totalidad de los componentes) pueden introducirse en la célula simultáneamente o secuencialmente en cualquier combinación.

15 Un ARN de CRISPR y un ARNtracr pueden fusionarse entre sí como un ARN guía (ARNg) a introducir en una célula. Alternativamente, un ARN de CRISPR y el ARNtracr pueden ser moléculas de ARN distintas. Un ARN de CRISPR puede introducirse en la célula en forma de un ARN o en forma de un ADN que codifica el ARN de CRISPR. Del mismo modo, un ARNtracr puede introducirse en la célula en forma de un ARN o en forma de un ADN que codifica el ARNtracr, y un ARNg puede introducirse en la célula en forma de un ARN o en forma de un ADN que codifica el ARNg.

20 Una proteína Cas puede introducirse en la célula en forma de una proteína, un ARN mensajero (ARNm) que codifica la proteína Cas, o un ADN que codifica la proteína Cas. En algunos métodos, una proteína Cas, un ARN de CRISPR, y un ARNtracr pueden introducirse en la célula como un complejo de ARN y proteína. Del mismo modo, una proteína Cas y un ARNg pueden introducirse en la célula como un complejo de ARN y proteína. La proteína Cas puede ser una proteína Cas permeable a la célula (por ejemplo, una proteína Cas con un dominio de penetración celular).

25 Un ADN que codifica una proteína Cas, un ARN de CRISPR, o un ARNtracr puede unirse operativamente a un promotor activo en la célula. Tales ADN pueden estar en uno o más constructos de expresión. En algunos métodos, uno o más de tales constructos de expresión pueden ser componentes de una sola molécula de ácido nucleico. Por ejemplo, los ADN que codifican una o más proteínas Cas, los ADN que codifican uno o más ARN de CRISPR, y los ADN que codifican uno o más ARNtracr pueden ser todos componentes de una sola molécula de ácido nucleico. Alternativamente, pueden encontrarse separados en cualquier combinación entre dos, tres, cuatro, o más moléculas de ácidos nucleicos.

30 De manera similar, un ADN que codifica una proteína Cas o un ADN que codifica un ARNg puede unirse operativamente a un promotor activo en la célula. Tales ADN también pueden estar en uno o más constructos de expresión. En algunos métodos, uno o más de tales constructos de expresión pueden ser componentes de una sola molécula de ácido nucleico. Por ejemplo, los ADN que codifican una o más proteínas Cas y los ADN que codifican uno o más ARNg pueden ser todos componentes de una sola molécula de ácido nucleico. Alternativamente, pueden encontrarse separados en cualquier combinación entre dos, tres, cuatro, o más moléculas de ácidos nucleicos.

35 En algunos métodos, la proteína Cas y el ARN de CRISPR y/o el ARNtracr no aparecen juntos en la naturaleza. En algunos métodos, por ejemplo, la proteína Cas y el primer ARN de CRISPR no aparecen juntos en la naturaleza, la proteína Cas y el segundo ARN de CRISPR no aparecen juntos en la naturaleza, y/o la proteína Cas y el ARNtracr no aparecen juntos en la naturaleza.

40 En algunos métodos, la proteína Cas es una proteína Cas9. La proteína Cas puede fusionarse a un polipéptido heterólogo, tal como una señal de localización nuclear (NLS). La proteína Cas puede tener actividad de escisión total y crear rupturas bicatenarias dentro del ADN genómico (por ejemplo, una ruptura bicatenaria con extremos romos), o puede ser una nickasa que puede escindir una sola cadena de ADN genómico.

45 En algunos métodos, se emplean pares de nickasas. Por ejemplo, el genoma puede ponerse en contacto con una primera y segunda nickasas que escinden cadenas de ADN opuestas, de manera que el genoma se modifica a través de la doble formación de mellas. La primera nickasa puede escindir una primera cadena de ADN genómico (es decir, la cadena complementaria), y la segunda nickasa puede escindir una segunda cadena de ADN genómico (es decir, la cadena no complementaria). Alternativamente, ambas nickasas pueden escindir la misma cadena. La primera y segunda nickasas pueden crearse, por ejemplo, mediante la mutación de un residuo catalítico en el dominio RuvC (por ejemplo, la mutación D10A descrita en otra parte en la presente descripción) de la primera nickasa y la mutación de un residuo catalítico en el dominio HNH (por ejemplo, la mutación H840A descrita en otra parte en la presente descripción) de la segunda nickasa. Alternativamente, la primera nickasa puede usarse para crear ambas mellas.

50 En algunos de tales métodos, la doble formación de mellas puede emplearse para crear una o más rupturas bicatenarias que tienen extremos escalonados. Por ejemplo, la doble formación de mellas se emplea para crear extremos escalonados en el primer y segundo sitios de escisión. La primera nickasa puede escindir la primera cadena de ADN dentro de la primera y segunda secuencias de reconocimiento de ARN de CRISPR con las que se hibridan el primer y segundo ARN de CRISPR, y la segunda nickasa puede escindir la segunda cadena de ADN dentro de la tercera y cuarta secuencias

diana de reconocimiento de ARN de CRISPR con las que se hibridan el tercer y cuarto ARN de CRISPR. Alternativamente, la primera nickasa puede usarse para formar mellas en la primera, segunda, tercera, y cuarta secuencias de reconocimiento de ARN de CRISPR. La primera y tercera secuencias diana de reconocimiento de ARN de CRISPR pueden posicionarse para crear un primer sitio de escisión de manera que las mellas creadas por la primera y segunda nickasas en la primera y segunda cadenas de ADN creen una ruptura bicatenaria (es decir, el primer sitio de escisión comprende las mellas dentro de la primera y tercera secuencias de reconocimiento de ARN de CRISPR). Del mismo modo, la segunda y cuarta secuencias de reconocimiento de ARN de CRISPR pueden posicionarse para crear un segundo sitio de escisión de manera que las mellas creadas por la primera y segunda nickasas en la primera y segunda cadenas de ADN creen una ruptura bicatenaria (es decir, el segundo sitio de escisión comprende las mellas dentro de la segunda y cuarta secuencias de reconocimiento de ARN de CRISPR). En algunos casos, las mellas dentro de la primera y tercera secuencias de reconocimiento de ARN de CRISPR y/o la segunda y cuarta secuencias de reconocimiento de ARN de CRISPR pueden ser mellas desplazadas. La ventana de desplazamiento puede ser, por ejemplo, de al menos aproximadamente 5 pb, 10 pb, 20 pb, 30 pb, 40 pb, 50 pb, 60 pb, 70 pb, 80 pb, 90 pb, 100 pb o más. Ver Ran y otros (2013) Cell 154:1380-1389; Mali y otros (2013) Nat. Biotech.31:833-838; y Shen y otros (2014) Nat. Methods 11:399-404.

E. Métodos para introducir ácidos nucleicos y proteínas en células

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

Los métodos proporcionados en la presente descripción pueden comprender introducir en una célula uno o más constructos de polinucleótidos o 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 manera tal que la secuencia acceda al interior de la célula. Los métodos proporcionados en la presente descripción no dependen de un método particular para introducir un ácido nucleico o proteína en la célula, solo que el ácido nucleico o proteína acceda al interior de al menos una célula. Los métodos para introducir ácidos nucleicos y proteínas en varios tipos celulares 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. Por ejemplo, una célula empleada en los métodos descritos en la presente descripción puede tener un gen que codifica Cas preexistente incorporado de manera estable en su genoma (es decir, una célula con Cas lista). "Incorporado de manera estable" o "introducido de manera estable" incluye la introducción de un polinucleótido en la célula de manera que la secuencia de nucleótidos se integre en el genoma de la célula y sea capaz de heredarse por su progenie. Puede usarse cualquier protocolo para la incorporación estable de los constructos de ADN o de los varios 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 no limitantes incluyen métodos de transfección químicos que usan 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 Kriegler, M (1991). Transfer and Expression: A Laboratory Manual. Nueva York: W. H. Freeman and Company. pp. 96-97); dendrímeros; o polímeros catiónicos tales como DEAE-dextrana o polietilénimina. Los métodos no químicos incluyen electroporación; sonoporación, y transfección óptica. La transfección basada en partículas incluye el uso de una pistola de genes, o transfección asistida por imanes (Bertram (2006) Current Pharmaceutical Biotechnology 7, 277-28). Pueden usarse además métodos virales de transfección.

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

La introducción de ácidos nucleicos o proteínas en una célula (por ejemplo, un embrión en etapa de una célula) también puede llevarse a cabo por microinyección. En embriones en etapa de una célula, la microinyección puede ser en el pronúcleo materno y/o paterno o en el citoplasma. Si la microinyección es solamente en un pronúcleo, se prefiere el pronúcleo paterno debido a su mayor tamaño. La microinyección de un ARNm es preferentemente en el citoplasma (por ejemplo, para suministrar el ARNm directamente a la maquinaria de traducción), mientras que la microinyección de una proteína Cas o una molécula de ácido nucleico que codifica una proteína Cas o que codifica un ARN se prefiere en el núcleo/pronúcleo. Alternativamente, la microinyección puede llevarse a cabo por inyección tanto en el núcleo/pronúcleo como en el citoplasma: una aguja puede introducirse primero en el núcleo/pronúcleo y puede inyectarse una primera cantidad, y mientras se retira la aguja del embrión en etapa de una célula puede inyectarse una segunda cantidad en el citoplasma. Si una proteína Cas se inyecta en el citoplasma, la proteína Cas comprende preferentemente una señal de localización nuclear para garantizar su suministro al núcleo/pronúcleo. Los métodos para llevar a cabo la microinyección se conocen bien. Ver, por ejemplo, Nagy y otros (Nagy A, Gertsenstein M, Vintersten K, Behringer R., 2003, Manipulating

the Mouse Embryo. Cold Spring Harbor, Nueva York: Cold Spring Harbor Laboratory Press); Meyer y otros (2010) *Proc Natl Acad Sci USA* 107:15022-15026 y Meyer y otros (2012) *Proc Natl Acad Sci USA* 109:9354-9359.

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 período de tiempo, al menos cuatro veces durante un período de tiempo, al menos cinco veces durante un período de tiempo, al menos seis veces durante un período de tiempo, al menos siete veces durante un período de tiempos, al menos diez veces durante un período de tiempo, al menos once veces, al menos doce veces durante un período de tiempo, al menos trece veces durante un período de tiempo, al menos catorce veces durante un período de tiempo, al menos quince veces durante un período de tiempo, al menos dieciséis veces durante un período de tiempo, al menos diecisiete veces durante un período de tiempo, al menos dieciocho veces durante un período de tiempo, al menos diecinueve veces durante un período de tiempo, o al menos veinte veces durante un período de tiempo.

Cuando se introducen en la célula agentes tipo nucleasa y vectores de transformación (por ejemplo, LTVEC para células distintas a los embriones en etapa de una célula), estos pueden introducirse simultáneamente. Alternativamente, el agente tipo nucleasa puede introducirse por separado del vector de transformación. Por ejemplo, el agente tipo nucleasa puede introducirse antes de la introducción del vector de transformación, o puede introducirse después de la introducción del vector de transformación.

F. Mecanismos de recombinación y métodos para alterar la prevalencia de unión de extremos no homólogos, conversión génica, o recombinación homóloga

La recombinación incluye cualquier proceso de intercambio de información genética entre dos polinucleótidos y puede ocurrir por cualquier mecanismo. La recombinación en respuesta a rupturas bicatenarias (DSB) ocurre principalmente a través de dos rutas de reparación del ADN conservadas: la unión de extremos no homólogos (NHEJ) y la recombinación homóloga (HR). Ver Kasperek & Humphrey (2011) *Seminars in Cell & Dev. Biol.* 22:886-897. La NHEJ incluye la reparación de rupturas bicatenarias en un ácido nucleico por ligación directa entre sí de los extremos de las rupturas sin la necesidad de una plantilla homóloga. La ligación de secuencias no contiguas por NHEJ frecuentemente puede dar como resultado deleciones, inserciones, o translocaciones cerca del sitio de la ruptura bicatenaria. La recombinación puede ocurrir además por medio de reparación dirigida por homología (HDR) o recombinación homóloga (HR). La HDR o HR incluye una forma de reparación de ácidos nucleicos que puede requerir homología de secuencias de nucleótidos, usa una molécula "donante" para la reparación por plantilla de una molécula "diana" (es decir, la que experimentó la ruptura bicatenaria), y conduce a la transferencia de información genética del donante a la diana. Sin desear limitarse por ninguna teoría particular, tal transferencia puede involucrar la corrección de errores de apareamiento del heterodúplex de ADN que se forma entre la diana con ruptura y el donante, y/o la hibridación de cadenas dependiente de síntesis, en la que el donante se usa para resintetizar la información genética que se convertirá en parte de la diana, y/o procesos relacionados. En algunos casos, el polinucleótido donante, una porción del polinucleótido donante, una copia del polinucleótido donante, o una porción de una copia del polinucleótido donante se integra en el ADN diana.

En el contexto de la modificación del genoma de una célula que es heterocigótica para un alelo para convertirla en homocigótica para ese alelo, la recombinación puede incluir cualquier medio por el cual las células homocigóticas se derivan de células heterocigóticas. Tales medios pueden incluir, por ejemplo, pérdida de heterocigosidad (LOH), conversión génica, o eventos de entrecruzamiento que ocurren mediante cualquier mecanismo de recombinación conocido. Sin desear limitarse por la teoría, la LOH puede ocurrir, por ejemplo, por medio de recombinación mitótica, con o sin conversión génica, o por medio de pérdida y duplicación de cromosomas. Ver, por ejemplo, Lefebvre y otros (2001) *Nat. Genet.* 27:257-258. La conversión génica en este contexto puede incluir la transferencia unidireccional de material genético de una secuencia donante a un aceptor altamente homólogo (es decir, el intercambio no recíproco de información genética de una molécula a su homóloga). La conversión génica incluye cualquier medio para copiar un alelo mediante cualquier mecanismo de recombinación conocido. Por ejemplo, la conversión génica puede involucrar la transferencia no recíproca de información genética de una secuencia intacta a una región homóloga que contiene una ruptura bicatenaria, y puede ocurrir entre cromátidas hermanas, cromosomas homólogos, o secuencias homólogas en la misma cromátida o en cromosomas diferentes. Ver, por ejemplo, Chen y otros (2007) *Nat. Rev. Genet.* 8:762-775. En casos específicos, la conversión génica es el resultado directo de la recombinación homóloga como resultado de la copia de información genética a partir de un cromosoma homólogo. Esto puede conducir a la pérdida de heterocigosidad (LOH) localizada cuando las secuencias homólogas no son idénticas.

A modo de ejemplo, la LOH podría ocurrir a través de intercambio recíproco de cromátidas por entrecruzamiento mitótico, o por copia de cromátidas mediante replicación inducida por ruptura. En cualquier caso, podría ocurrir una modificación heterocigótica en la que un cromosoma se transforma antes de la replicación del genoma. Alternativamente, podría transformarse una sola cromátida después de la replicación del genoma, seguido de conversión génica entre cromátidas.

En cualquiera de los métodos descritos en la presente descripción, la célula puede ser una célula que se ha modificado para aumentar o disminuir la actividad de NHEJ. Del mismo modo, la célula puede ser una célula que se ha modificado para aumentar la conversión génica o la actividad de HDR. Tales modificaciones pueden comprender modificaciones en la expresión o la actividad de genes involucrados en la regulación de NHEJ, conversión génica, y/o HDR. Por ejemplo, la

disminución de la actividad de NHEJ y/o el aumento de la actividad de HDR pueden promover el colapso bialélico de regiones genómicas entre secuencias de reconocimiento de ARN de CRISPR correspondientes a dos ARNg. Sin desear limitarse por ninguna teoría particular, un mecanismo mediante el cual puede ocurrir el colapso genómico bialélico es por reparación mediada por NHEJ o reparación mediada por HDR dentro de un primer alelo y la creación de un segundo alelo idéntico por medio de mecanismos de HDR, tales como la conversión génica (ver el Ejemplo 1). Por lo tanto, promover las rutas mediadas por HDR (por ejemplo, mediante la disminución de la actividad de NHEJ o mediante el aumento de la actividad de HDR también puede promover el colapso bialélico de regiones genómicas. De manera similar, sin desear limitarse por ninguna teoría particular, la conversión de una célula heterocigótica a una célula homocigótica mediante el uso de pares de ARN guías que reconocen un solo locus puede promoverse si la actividad de NHEJ disminuye y la actividad de HDR (por ejemplo, la actividad de conversión génica) aumenta en proporción.

Pueden usarse inhibidores para aumentar o disminuir la actividad de NHEJ o para aumentar o disminuir la actividad de HDR. Tales inhibidores pueden ser, por ejemplo, moléculas pequeñas o ácidos nucleicos inhibidores tales como ácidos nucleicos de interferencia cortos (por ejemplo, ARN de interferencia corto (ARNic), ARN bicatenario (ARNbc), microARN (miARN), y ARN en horquilla corto (ARNhc)) u oligonucleótidos antisentido específicos para un transcrito génico. Los inhibidores pueden dirigirse a enzimas involucradas en NHEJ o HDR o a su regulación corriente arriba por modificación postraduccional por medio de, por ejemplo, fosforilación, ubiquitinación, y sumoilación.

En células de mamífero, la NHEJ es el mecanismo de reparación de DSB predominante y está activo a lo largo del ciclo celular. En vertebrados, la ruta de NHEJ "canónica" o "clásica" (C-NHEJ) requiere varios factores centrales, que incluyen ADN-PK, Ku70-80, Artemis, ligasa IV (Lig4), XRCC4, CLF, y Pol μ para reparar una DSB. Ver Kasperek & Humphrey (2011) *Seminars in Cell & Dev. Biol.* 22:886-897. Durante la NHEJ, los extremos de ADN se unen a la proteína Ku protectora de extremos muy abundante, que funciona como una estación de acoplamiento para la incorporación de los otros componentes de la NHEJ.

Por lo tanto, en algunos de los métodos descritos en la presente descripción, la célula se ha modificado para reducir o eliminar o para aumentar la expresión o la actividad de factores involucrados en la C-NHEJ. Por ejemplo, en algunos métodos, la célula se ha modificado para reducir o eliminar la expresión o la actividad de ADN-PK, Ku70-80, Artemis, ligasa IV (Lig4), XRCC4, CLF, y/o Pol μ . En métodos específicos, la célula se ha modificado para reducir o eliminar la expresión o la actividad de ADN-PK o para aumentar la expresión o la actividad de ADN-PK (por ejemplo, la expresión o la actividad de las ADN-PKc; secuencia de UniProt ilustrativa designada P97313). Los ejemplos de inhibidores de ADN-PKc incluyen, por ejemplo, NU7026, y NU7441. Ver, por ejemplo, la patente de Estados Unidos núm. 6,974,867. En métodos específicos, la célula se ha modificado para reducir o eliminar la expresión o la actividad de ligasa IV o para aumentar la expresión o la actividad de ligasa IV. Un ejemplo de un inhibidor de ligasa IV es SCR7.

Pueden usarse además inhibidores dirigidos a proteínas de puntos de regulación del ciclo celular como ATM (por ejemplo, KU55933), CHK1/CHK2 (por ejemplo, KLD1162 o CHIR-124) y ATR (por ejemplo, VE 821) para mejorar sinérgicamente los efectos de inhibidores de la reparación del ADN específicos o para prevenir efectos secundarios no intencionales como la detención del ciclo celular y/o la apoptosis (ver Ciccía y otros (2010) *Mol Cell* 40:179).

La interrupción de C-NHEJ puede aumentar los niveles de unión anormal mediada por rutas de NHEJ "alternativas" (A-NHEJ) y también puede aumentar la reparación por HR. Las rutas A-NHEJ muestran un sesgo hacia las uniones mediadas por microhomología y siguen cinéticas más lentas que la C-NHEJ. Se ha propuesto la participación de varios factores, que incluyen el complejo MRN (MRE11, RAD50, NBS1), CtIP, XRCC1, PARP, Lig1, y Lig3. Ver Kasperek & Humphrey (2011) *Seminars in Cell & Dev. Biol.* 22:886-897 y Claybon y otros (2010) *Nucleic Acids Res.* 38(21):7538-7545.

Por lo tanto, en algunos de los métodos descritos en la presente descripción, la célula se ha modificado para reducir o eliminar o para aumentar la expresión o la actividad de factores involucrados en la A-NHEJ. Por ejemplo, en algunos métodos, la célula se ha modificado para reducir o eliminar la expresión o la actividad de MRE11, RAD50, NBS1, CtIP, XRCC1, PARP (por ejemplo, PARP1), Lig1, y/o Lig3. En otros métodos, la célula se ha modificado para aumentar la expresión o la actividad de MRE11, RAD50, NBS1, CtIP, XRCC1, PARP (por ejemplo, PARP1), Lig1, y/o Lig3. En métodos específicos, la célula se ha modificado para reducir o eliminar la expresión o la actividad de PARP1 o para aumentar la expresión o la actividad de PARP1 (secuencia de UniProt ilustrativa designada P11103). Los ejemplos de inhibidores de PARP (por ejemplo, NU1025, Iniparib, Olaparib) incluyen nicotinamidas; isoquinolinonas y dihidroisoquinolinonas; bencimidazoles e indoles; ftalazin-1(2H)-onas y quinazolinonas; isoindolinonas y análogos y derivados de estas; fenantridinas y fenantridinonas; benzopironas y análogos y derivados de estas; derivados insaturados del ácido hidroxímico y análogos y derivados de este; piridazinas, que incluyen piridazinas fusionadas y análogos y derivados de estas; y/u otros compuestos tales como cafeína, teofilina, y timidina, y análogos y derivados de estas. Ver, por ejemplo, la patente de Estados Unidos núm. 8,071,579.

La C-NHEJ exhibe además una relación competitiva con HR de manera que la interrupción de la C-NHEJ también puede conducir al aumento de la reparación por HR. Tal competencia entre NHEJ y HR puede explotarse ya que la interrupción de NHEJ puede conducir a la transformación génica mejorada a través de la reducción de la integración aleatoria y posiblemente el aumento de la integración dirigida por recombinación homóloga.

Existen varias formas de reparación por recombinación homóloga, que incluyen hibridación de cadenas simples, conversión génica, entrecruzamientos, y replicación inducida por ruptura. La hibridación de cadenas simples es una forma minoritaria de reparación por HR en la que secuencias monocatenarias homólogas en cualquier lado de una DSB reseca se hibridan, lo que da como resultado la reconstitución del cromosoma. La hibridación de cadenas simples genera 5 delecciones de tamaño variable, en dependencia de la distancia que separa las dos regiones de homología de secuencia. La conversión génica incluye el intercambio no recíproco de información genética de una molécula a su homóloga, como resultado directo de la HR como resultado de la copia de información genética a partir de un cromosoma homólogo. Esto puede conducir a la LOH localizada cuando las secuencias homólogas no son idénticas. Normalmente, la extensión de la 10 conversión génica se limita a algunos cientos de pares de bases. Sin embargo, se ha informado la conversión génica de tramos largos en algunos fondos genéticos, que incluyen la deficiencia de RAD51C. Ver Nagaraju y otros (2006) Mol. Cell. Biol. 26:8075-8086. Los entrecruzamientos pueden ocurrir, por ejemplo, entre cromosomas homólogos, y tienen el potencial de conducir a translocaciones recíprocas si ocurren en G1 o a translocaciones no recíprocas y LOH que se extiende desde el sitio de ruptura hasta el telómero distal si ocurren en G2. La replicación inducida por ruptura es una variante de HR en la que después de la invasión de la cadena, la replicación del ADN continúa hasta el extremo del 15 cromosoma. Por lo tanto, existen muchos mecanismos mediante los cuales la HR puede promover la LOH.

Por lo tanto, en algunos de los métodos descritos en la presente descripción, la célula se ha modificado para reducir o eliminar o para aumentar la expresión o la actividad de factores involucrados en la HR. Por ejemplo, en algunos métodos, la célula se ha modificado para aumentar la expresión o la actividad de RAD51, RAD52, RAD54, RAD55, RAD51C, 20 BRCA1, y/o BRCA2. En otros métodos, la célula se ha modificado para reducir o eliminar la expresión o la actividad de RAD51, RAD52, RAD54, RAD55, RAD51C, BRCA1, y/o BRCA2.

En algunos métodos, la expresión o la actividad de otras proteínas involucradas en la regulación de NHEJ y/o HR pueden alterarse. Por ejemplo, en algunos métodos, la célula se ha modificado para reducir o eliminar la expresión o la actividad 25 de Chk2, para reducir o eliminar la expresión o la actividad de Clspn, para reducir o eliminar la expresión o la actividad de Setd2, para aumentar la expresión o la actividad de Kat2a, y/o para aumentar la expresión o la actividad de Rad51. En otros métodos, la célula se ha modificado para aumentar la expresión o la actividad de Chk2, para aumentar la expresión o la actividad de Clspn, para aumentar la expresión o la actividad de Setd2, para reducir o eliminar la expresión o la actividad de Kat2a, y/o para reducir o eliminar la expresión o la actividad de Rad51. 30

Chk2 (también conocida como Chek2 y Rad53; el homólogo de *S. pombe* es Cds1) es una proteína quinasa serina/treonina requerida para la detención del ciclo celular mediada por punto de regulación, activación de la reparación del ADN, y apoptosis en respuesta a la presencia de rupturas bicatenarias en el ADN. Ver Blaikley y otros (2014) Nucleic Acids Research 42:5644-5656. Clspn (también conocida como Claspina; el homólogo de *S. pombe* es Mrc1) es una 35 proteína requerida para la detención del ciclo celular mediada por punto de regulación en respuesta al daño del ADN. Se ha informado que la delección de homólogos de Chk2 o Clspn en *S. pombe* da como resultado un fenotipo hiperrecombinante que exhibe niveles significativamente elevados de conversión génica inducida por ruptura en comparación con el tipo silvestre. Específicamente, se informó que los niveles de conversión génica se encontraban significativamente aumentados, mientras que se informó que los niveles de unión de extremos no homólogos (NHEJ), conversión de cromátidas hermanas (SCC), y pérdida de heterocigosidad (LOH) se encontraban disminuidos. Ver Blaikley y otros (2014) Nucleic Acids Research 42:5644-5656. 40

Kat2a (también conocida como Gcn5 y Gcn512) es una histona acetiltransferasa ubicua que promueve la activación de la transcripción y se ha informado su asociación con la reparación de rupturas bicatenarias. La acetilación de la lisina 36 de 45 la histona H3 (H3K36) dependiente de Kat2a aumenta la accesibilidad a la cromatina, aumenta la resección, y promueve la recombinación homóloga mientras que suprime la unión de extremos no homólogos. Ver Pai y otros (2014) Nat. Commun. 5:4091. Setd2 (también conocida como Kiaa1732, Kmt3a, y Set2) es una histona metiltransferasa que trimetila específicamente la lisina 36 de la histona H3 (H3K36me3) mediante el uso de lisina 36 desmetilada (H3K36me2) como sustrato. La metilación H3K36 dependiente de Setd2 reduce la accesibilidad a la cromatina, reduce la resección, y 50 promueve la NHEJ. Ver Pai y otros (2014) Nat. Commun. 5:4091.

Rad 51 (también conocida como Reca, Rad51A, y homólogo 1 de la proteína Rad51 de reparación del ADN) es una proteína que funciona con Rad52 y otras proteínas para efectuar el intercambio de cadenas durante la recombinación homóloga, lo que forma un heterodúplex de ADN que se resuelve por reparación de errores de apareamiento para producir 55 un tramo de conversión génica. En células de mamífero, se ha informado que la sobreexpresión de Rad51 y Rad52 aumenta la frecuencia de recombinación homóloga y conversión génica. Ver Yanez & Porter (1999) Gene Ther. 6:1282-1290 y Lambert & Lopez (2000) EMBO J. 19:3090-3099.

Las modificaciones en la expresión o la actividad de los genes involucrados en la regulación de la NHEJ, la conversión 60 génica, y/o la reparación dirigida por homología pueden ser espacialmente o temporalmente específicas y también pueden ser inducibles o temporales y reversibles. Por ejemplo, pueden construirse varias formas de casetes para permitir la delección en tipos de células o tejidos específicos, en etapas del desarrollo específicas, o tras la inducción. Tales casetes pueden emplear un sistema de recombinasa en el que el casete está flanqueado en ambos lados por sitios de reconocimiento de recombinasa y puede eliminarse mediante el uso de una recombinasa expresada en el tipo celular deseado, expresada en la etapa del desarrollo deseada, o expresada o activada tras la inducción. Tales casetes pueden 65 construirse además de modo que incluyan una serie de pares de sitios de reconocimiento de recombinasa diferentes que

se colocan de manera que puedan generarse alelos nulos, condicionales, o una combinación de condicionales/nulos, como se describe en el documento US 2011/0104799. La regulación de genes de recombinasa puede controlarse de varias maneras, tal como mediante la unión operativa de un gen de recombinasa a un promotor (u otro elemento regulador) específico de célula, específico de tejido, o regulado por el desarrollo, o mediante la unión operativa de un gen de recombinasa a una 3'-UTR que comprende un sitio de reconocimiento de un miARN que es activo solamente en tipos celulares, tipos de tejidos, o etapas del desarrollo particulares. Una recombinasa puede regularse además, por ejemplo, mediante el empleo de una proteína de fusión que coloca la recombinasa bajo el control de un efector o metabolito (por ejemplo, CreER^{T2}, cuya actividad es controlada positivamente por tamoxifeno), o mediante la colocación del gen de recombinasa bajo el control de un promotor inducible (por ejemplo, uno cuya actividad es controlada por doxiciclina y TetR o variantes de TetR). Los ejemplos de varias formas de casetes y medios para regular genes de recombinasa se proporcionan, por ejemplo, en los documentos US 8,518,392; US 8,354,389; y US 8,697,851.

G. Células y animales

Varias composiciones y métodos proporcionados en la presente descripción emplean células, tales como células de un animal. Tales células pueden ser de un animal no humano. Tales células pueden ser células eucariotas, que incluyen, por ejemplo, células fúngicas (por ejemplo, de levadura), células vegetales, células animales, células de mamífero, 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, una célula pluripotente, tal como una célula pluripotente no humana (por ejemplo, una célula madre embrionaria (ES) de ratón o una célula ES de rata) o una célula pluripotente humana, o una célula no pluripotente. Las células totipotentes incluyen células indiferenciadas que pueden dar lugar a cualquier tipo celular, y las células pluripotentes incluyen células indiferenciadas que poseen la capacidad de desarrollarse a más de un tipo celular diferenciado. Tales células pluripotentes y/o totipotentes pueden ser, por ejemplo, células madre embrionarias (ES) o células similares a ES, tales como células madre pluripotentes inducidas (iPS). Las células madre embrionarias incluyen células totipotentes o pluripotentes de origen embrionario 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 a células de cualquiera de las tres capas germinales de los vertebrados (endodermo, ectodermo, y mesodermo).

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

Las células eucariotas incluyen además células primarias. Las células primarias incluyen células o cultivos de células que se han aislado directamente de un organismo, un órgano, o un tejido. Las células primarias incluyen células que no son transformadas ni inmortalizadas. Estas incluyen cualquier célula obtenida de un organismo, un órgano, o un tejido que no se sometió previamente a pases en cultivo de tejidos o se ha sometido previamente a pases en cultivo de tejidos pero no es capaz de someterse a pases indefinidamente en cultivo de tejidos. Tales células pueden aislarse por 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 de músculo liso. Por ejemplo, las células primarias pueden derivarse de tejidos conectivos, tejidos musculares, tejidos del sistema nervioso, o tejidos epiteliales.

Las células eucariotas incluyen además células inmortalizadas. Las células inmortalizadas incluyen células de un organismo multicelular que normalmente no proliferarían indefinidamente pero, debido a mutación o alteración, han evadido la senescencia celular normal y en lugar de esto pueden seguir experimentando división. Tales mutaciones o alteraciones pueden ocurrir de manera natural o inducirse de manera intencional. Los ejemplos de células inmortalizadas incluyen células de ovario de hámster chino (CHO), células embrionarias de riñón humano (por ejemplo, células HEK 293), y células fibroblásticas embrionarias 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 el cultivo o para expresar genes o proteínas recombinantes.

Las células eucariotas pueden incluir además embriones en etapa de una célula (es decir, oocitos fertilizados o cigotos). Tales embriones en etapa de una célula pueden ser de cualquier fondo genético (por ejemplo, BALB/c, C57BL/6, 129, o una combinación de estos), pueden ser frescos o congelados, y pueden derivarse de cruzamiento natural o fertilización *in vitro*.

El término "animal," en referencia a células, células pluripotentes y/o totipotentes, células ES, células donantes, y/o embriones huésped, incluyen mamíferos, peces, y aves. Los mamíferos incluyen, por ejemplo, seres humanos, primates no humanos, monos, simios, gatos, perros, caballos, toros, ciervo, bisonte, oveja, roedores (por ejemplo, ratones, ratas, hámsteres, cobayas), ganado (por ejemplo, especies bovinas tales como vacas, buey, etc.; especies ovinas tales como oveja, cabras, etc.; y especies porcinas tales como cerdos y verracos). Las aves incluyen, por ejemplo, pollos, pavos,

avestruz, gansos, patos, etc. Los animales domésticos y los animales de uso agrícola también se incluyen. El término "animal no humano" excluye a los seres humanos.

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/Kal_wN, C57BL/6, C57BL/6J, C57BL/6ByJ, C57BL/6NJ, C57BL/10, C57BL/10ScSn, C57BL/10Cr, y C57BL/Ola. Las células pluripotentes y/o totipotentes de ratón también pueden ser de una mezcla de una cepa 129 mencionada anteriormente y una cepa C57BL/6 mencionada anteriormente (por ejemplo, 50 % 129 y 50 % C57BL/6). Del mismo modo, las células pluripotentes y/o totipotentes de ratón pueden ser de una mezcla de cepas 129 mencionadas anteriormente o una mezcla de cepas BL/6 mencionadas anteriormente (por ejemplo, la cepa 129S6 (129/SvEvTac)) Un ejemplo específico de una célula ES de ratón es una célula ES de ratón VGF1. Ver, por ejemplo, Auerbach y otros (2000) *Biotechniques* 29, 1024-1028, 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 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, la célula pluripotente y/o totipotente de rata puede ser de una cepa DA o de una cepa ACI. La cepa de rata ACI se caracteriza por tener agutí negro, con patas y vientre blanco y un haplotipo RT1^{av1}. 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 agutí y un haplotipo RT1^{av1}. 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 rata son de una cepa de rata endogámica. Ver, por ejemplo, el documento U.S. 2014/0235933 A1, presentado el 20 de febrero de 2014.

Los ejemplos de células pluripotentes humanas incluyen células ES humanas, células madre adultas humanas, células progenitoras humanas restringidas por el desarrollo, y células madre pluripotentes inducidas (iPS) humanas, tales como células iPS humanas activadas y células iPS humanas vírgenes. 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), factores de transcripción de la familia Myc (por ejemplo, c-Myc, 1-Myc, n-Myc), factores de transcripción de la familia similar a 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 pueden generarse además, por ejemplo, mediante el uso de miARN, moléculas pequeñas que imitan las acciones de los factores de transcripción, o especificadores de linaje. Las células iPS humanas se caracterizan por su capacidad de diferenciarse a cualquier célula de las tres capas germinales de los vertebrados, por ejemplo, el endodermo, el ectodermo, o el mesodermo. Las células iPS humanas se caracterizan además por su capacidad de propagarse indefinidamente en condiciones de cultivo *in vitro* adecuadas. Ver, por ejemplo, Takahashi y Yamanaka (2006) *Cell* 126:663-676. Las células ES humanas activadas y las células iPS humanas activadas incluyen células que expresan características similares a las de las células epiblasticas después de la implantación y están comprometidas para la especificación del linaje y la diferenciación. Las células ES humanas vírgenes y las células iPS humanas vírgenes incluyen células que expresan características similares a las de las células ES de la masa celular interna de un embrión antes de la implantación y no están comprometidas para 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 huésped pueden referirse como "células donantes." La célula pluripotente y/o totipotente genéticamente modificada puede ser de la misma cepa que el embrión huésped o de una cepa diferente. Del mismo modo, la madre sustituta puede ser de la misma cepa que la célula pluripotente y/o totipotente genéticamente modificada y/o el embrión huésped, o la madre sustituta puede ser de una cepa diferente a la de la célula pluripotente y/o totipotente genéticamente modificada y/o el embrión huésped.

Una variedad de embriones huésped puede emplearse en los métodos y composiciones descritos en la presente descripción. Por ejemplo, las células pluripotentes y/o totipotentes que tienen la modificación genética dirigida pueden introducirse en un embrión en etapa de 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, la etapa de 4 células, la etapa de 8 células, la etapa de 16 células, la etapa de 32 células, o la etapa de 64 células. El embrión huésped puede ser además un blastocisto o puede ser un embrión preblastocisto, un embrión en etapa de premórula, un embrión en etapa de mórula, un embrión en etapa de mórula descompactada, o un embrión en etapa de mórula compactada. Cuando se emplea un embrión de ratón, la etapa del embrión huésped puede ser una Etapa 1 de Theiler (TS1), una TS2, una TS3, una TS4, una TS5, y una TS6, con referencia 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 de TS1, TS2, TS3, y TS4. En algunos casos, el embrión huésped comprende una zona pelúcida, y la célula donante es una célula ES que se introduce en el embrión huésped a través de un agujero en la zona pelúcida. En otros casos, el embrión huésped es un embrión sin zona. Aún en otros casos, el embrión huésped en etapa de mórula se encuentra agregado.

5

H. Métodos para identificar células con genomas modificados

Algunos de los métodos anteriores comprenden además identificar una célula que tiene un genoma modificado. Varios métodos pueden usarse para identificar las células que tienen una modificación dirigida, tal como una delección o una inserción. Tales métodos pueden comprender identificar una célula que tiene la modificación dirigida en un locus diana (por ejemplo, entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR). El tamizaje puede hacerse para identificar tales células con loci genómicos modificados.

10

La etapa de tamizaje puede comprender un ensayo cuantitativo para evaluar la modificación de alelo (MOA) de un cromosoma progenitor. Por ejemplo, el ensayo cuantitativo puede llevarse a cabo por medio de una PCR cuantitativa, tal como una PCR en tiempo real (qPCR). La PCR en tiempo real puede utilizar un primer conjunto de cebadores que reconoce el locus diana y un segundo conjunto de cebadores que reconoce un locus de referencia no transformado. El conjunto de cebadores puede comprender una sonda fluorescente que reconoce la secuencia amplificada.

15

La etapa de tamizaje puede comprender además un ensayo de retención, que es un ensayo usado para distinguir entre las inserciones dirigidas correctas de un inserto de ácido nucleico en un locus genómico diana de inserciones transgénicas aleatorias del inserto de ácido nucleico en ubicaciones genómicas fuera del locus genómico diana. Los ensayos convencionales para tamizar modificaciones dirigidas, tales como PCR de largo alcance o transferencia southern, vinculan el vector de transformación insertado con el locus diana. Debido a sus tamaños grandes de los brazos de homología, sin embargo, los LTVEC no permiten el tamizaje mediante tales ensayos convencionales. Para tamizar la transformación por LTVEC, pueden usarse ensayos de modificación de alelo (MOA) que incluyen ensayos de pérdida de alelo (LOA) y ganancia de alelo (GOA) (ver, por ejemplo, el documento US 2014/0178879 y Frenthewey y otros (2010) *Methods Enzymol.* 476:295-307). El ensayo de pérdida de alelo (LOA) invierte la lógica del tamizaje convencional y cuantifica el número de copias del locus nativo al que se dirigió la mutación. En un clon de células transformado correctamente, 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 encuentra interrumpido por la modificación dirigida. El mismo principio puede aplicarse a la inversa 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 transformado correctamente ha perdido una copia del gen diana nativo y ganado una copia del gen de resistencia a fármaco u otro marcador insertado.

20

25

30

35

A modo de ejemplo, la reacción en cadena de la polimerasa (qPCR) cuantitativa puede usarse 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 diana o entre cero, una, y dos copias del inserto de ácido nucleico puede usarse para desarrollar un ensayo MOA. Por ejemplo, TaqMan® puede usarse para cuantificar el número de copias de una plantilla de ADN en una muestra de ADN genómico, especialmente por 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(los) diana. Por lo tanto, se realizan dos amplificaciones TaqMan® (cada una con su sonda respectiva). Una sonda TaqMan® determina el "Ct" (ciclo umbral) del gen de referencia, mientras que la otra sonda determina el Ct de la región del(los) gen(es) o locus(los) diana que se reemplaza(n) por transformación exitosa (es decir, un ensayo LOA). El Ct es una cantidad que refleja la cantidad de ADN inicial para cada una de las sondas 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 plantilla para una reacción TaqMan® dará como resultado un aumento de aproximadamente una unidad de Ct. Las reacciones TaqMan® en células donde un alelo del(los) gen(es) o locus(los) diana se ha reemplazado por recombinación homóloga dará como resultado un aumento de un Ct para la reacción TaqMan® de la diana sin un aumento en el Ct del gen de referencia en comparación con el ADN de células no transformadas. Para un ensayo GOA, puede usarse otra sonda TaqMan® para determinar el Ct del inserto de ácido nucleico que reemplaza el(los) gen(es) o locus(los) diana por transformación exitosa.

40

45

50

Debido a que los pares de ARNg pueden crear delecciones grandes mediadas por Cas en un locus genómico diana, puede ser útil aumentar los ensayos LOA y GOA estándar para comprobar la transformación correcta por LTVEC (es decir, en células distintas a los embriones en etapa de una célula). Por ejemplo, los ensayos LOA y GOA solos pueden no distinguir los clones de células transformados correctamente de los clones en los que una delección grande inducida por Cas del locus genómico diana coincide con la integración aleatoria de un LTVEC en otra parte en el genoma, particularmente si el ensayo GOA emplea una sonda contra un casete de selección dentro del inserto de LTVEC. Debido a que la presión de selección en la célula transformada se basa en el casete de selección, la integración transgénica aleatoria del LTVEC en otra parte en el genoma generalmente incluirá el casete de selección y las regiones adyacentes del LTVEC pero excluirá 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 alrededor de 5 kb o más de longitud con un casete de selección adyacente al brazo de homología 3', generalmente el brazo de homología 3' pero no el brazo de homología 5' se integrará transgénicamente con el casete de selección. Alternativamente, si el casete de selección adyacente al brazo de homología 5', generalmente el brazo de homología 5' pero no el brazo de homología 3' se integrará transgénicamente

55

60

65

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, una delección heterocigótica en el locus genómico diana combinada con una integración transgénica aleatoria del LTVEC proporcionará la misma lectura que una integración dirigida heterocigótica del LTVEC en el locus genómico diana. Para comprobar la transformación correcta por el LTVEC, pueden usarse ensayos de retención, solos o junto con ensayos LOA y/o GOA.

Los ensayos de retención determinan los números de copias de una plantilla de ADN en la secuencia diana 5' (correspondiente al brazo de homología 5' del LTVEC) y/o la secuencia diana 3' (correspondiente al brazo de homología 3' del LTVEC). En particular, es útil la determinación del número de copias de una plantilla de ADN en la secuencia diana correspondiente al brazo de homología que es adyacente al casete de selección. En células diploides, los números de copias mayores que dos generalmente indican la integración transgénica del LTVEC aleatoriamente fuera del locus genómico diana en lugar de en el locus genómico diana, lo que no es conveniente. Los clones transformados correctamente retendrán un número de copias de dos. Además, los números de copias de menos de dos en tales ensayos de retención generalmente indican delecciones grandes mediadas por Cas que se extienden más allá de la región reconocida para la delección, lo que tampoco es conveniente.

En un ensayo de retención ilustrativo para identificar una inserción dirigida de un inserto de ácido nucleico en un locus genómico diana en una célula diploide, el ADN se obtiene primero de una célula que se ha puesto en contacto con un vector de transformación grande (LTVEC) que comprende el inserto de ácido nucleico flanqueado por un primer brazo de homología que se hibrida con una primera secuencia diana y un segundo brazo de homología que se hibrida con una segunda secuencia diana, en donde el inserto de ácido nucleico comprende un casete de selección adyacente al primer brazo de homología. Opcionalmente, el casete de selección puede comprender un gen de resistencia a fármaco. Después el ADN se expone a una sonda que se une dentro de la primera secuencia diana, una sonda que se une dentro del inserto de ácido nucleico, y una sonda que se une dentro de un gen de referencia que tiene un número de copias conocido, en donde cada sonda genera una señal detectable tras la unión. Después se detectan las señales de la unión de cada una de las sondas. La señal de la sonda del gen de referencia se compara con la señal de la sonda de la primera secuencia diana para determinar un número de copias de la primera secuencia diana, y la señal de la sonda del gen de referencia se compara con la señal de la sonda del inserto de ácido nucleico para determinar un número de copias del inserto de ácido nucleico. Un número de copias del inserto de ácido nucleico de uno o dos y un número de copias de la primera secuencia diana de dos generalmente indica la inserción dirigida del inserto de ácido nucleico en el locus genómico diana, y un número de copias del inserto de ácido nucleico de uno o más y un número de copias de la primera secuencia diana de tres o más generalmente indica una inserción aleatoria del inserto de ácido nucleico en un locus genómico distinto al locus genómico diana.

La señal de la unión de la sonda de la primera secuencia diana puede usarse para determinar un valor de ciclo umbral (Ct) de la primera secuencia diana, la señal de la unión de la sonda del gen de referencia puede usarse para determinar un valor de ciclo umbral (Ct) del gen de referencia, y el número de copias de la primera secuencia diana puede determinarse mediante la comparación del valor de Ct de la primera secuencia diana y el valor de Ct del gen de referencia. Del mismo modo, la señal de la unión de la sonda del inserto de ácido nucleico puede usarse para determinar un valor de ciclo umbral (Ct) del inserto de ácido nucleico, y el número de copias del inserto de ácido nucleico puede determinarse mediante la comparación del valor de Ct de la primera secuencia diana y el valor de Ct del gen de referencia.

El inserto de ácido nucleico en el LTVEC puede ser, por ejemplo, de al menos 5 kb, al menos 10 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 150 kb, al menos 200 kb, al menos 250 kb, al menos 300 kb, al menos 350 kb, al menos 400 kb, al menos 450 kb, o al menos 500 kb. La distancia entre las secuencias a las que se unen las sondas en la primera secuencia diana y el casete de selección puede ser, por ejemplo, de no más de 100 nucleótidos, 200 nucleótidos, 300 nucleótidos, 400 nucleótidos, 500 nucleótidos, 600 nucleótidos, 700 nucleótidos, 800 nucleótidos, 900 nucleótidos, 1 kb, 1,5 kb, 2 kb, 2,5 kb, 3 kb, 3,5 kb, 4 kb, 4,5 kb, o 5 kb.

Tales métodos pueden comprender además ensayos de retención adicionales para determinar el número de copias de la segunda secuencia diana. Por ejemplo, tales métodos pueden comprender además exponer el ADN de la célula a una sonda que se une a la segunda secuencia diana, detectar la señal de la unión de la sonda de la segunda secuencia diana, y comparar la señal de la sonda del gen de referencia con la señal de la sonda de la segunda secuencia diana para determinar un número de copias de la segunda secuencia diana.

Del mismo modo, tales métodos pueden comprender además ensayos GOA adicionales para determinar el número de copias de una o más secuencias adicionales dentro del inserto de ácido nucleico. Por ejemplo, tales métodos pueden comprender además exponer el ADN de la célula a una o más sondas adicionales que se unen al inserto de ácido nucleico, detectar la señal de la unión de la una o más sondas adicionales, y comparar la señal de la sonda del gen de referencia con la señal de la una o más sondas del inserto de ácido nucleico adicionales para determinar los números de copias de la una o más secuencias adicionales dentro del inserto de ácido nucleico.

Del mismo modo, cuando el LTVEC se diseña para eliminar una secuencia endógena del locus genómico diana o cuando se usan pares de ARNg (por ejemplo, para crear pares de rupturas bicatenarias en sitios diferentes dentro de un solo locus genómico diana y eliminar la secuencia endógena intermedia), tales métodos pueden comprender además un

5 ensayo LOA para determinar el número de copias de las secuencias endógenas en el locus genómico diana. Por ejemplo, tales métodos pueden comprender además exponer el ADN de la célula a una sonda que se une a la secuencia endógena en el locus genómico diana, detectar la señal de la unión de la sonda de la secuencia endógena, y comparar la señal de la sonda del gen de referencia con la señal de la sonda de la secuencia endógena para determinar un número de copias de la secuencia endógena.

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

15 Para las modificaciones genéticas dirigidas generadas sin el uso de LTVEC, pueden usarse ensayos convencionales para tamizar las modificaciones dirigidas, tales como PCR de largo alcance, transferencia southern, o secuenciación de Sanger. Tales ensayos se usan típicamente para obtener evidencias de una unión entre el vector de transformación insertado y el locus genómico diana. Por ejemplo, para un ensayo de PCR de largo alcance, un cebador puede reconocer una secuencia dentro del ADN insertado mientras el otro reconoce una secuencia del locus diana más allá de los extremos de los brazos de homología del vector de transformación.

20 Si diferentes versiones de una secuencia se asocian con un número de acceso en momentos diferentes, se entiende que es la versión asociada con el número de acceso en la fecha de presentación efectiva de esta solicitud. La fecha de presentación efectiva quiere decir la de antes de la fecha de presentación actual o la fecha de presentación de una solicitud de prioridad que hace referencia al número de acceso si es aplicable. Del mismo modo, si diferentes versiones de una publicación, sitio web o similar se publican en momentos diferentes, se entiende que es la versión de publicación más reciente en la fecha de presentación efectiva de la solicitud a menos que se indique de otra manera. Cualquier característica, etapa, elemento, modalidad, o aspecto de la invención puede usarse en combinación con cualquier otro a menos que se indique específicamente de otra manera.

Tabla 1. Descripción de las secuencias.

SEQ ID NO	Tipo	Descripción
1	ARN	ARNg
2	ARN	ARNg
3	ARN	ARNcr
4	ARN	ARNcr
5	ARN	ARNcr
6	ARN	ARNtracr
7	ARN	ARNtracr
8	ADN	Secuencia de reconocimiento de ARN de CRISPR
9	ADN	Secuencia de reconocimiento de ARN de CRISPR
10	ADN	Secuencia de reconocimiento de ARN de CRISPR
11	ADN	Segmento de orientación al ADN del ARNg A de C5 (Hc) (100 pb del punto extremo del locus diana)
12	ADN	Segmento de orientación al ADN del ARNg B de C5 (Hc) (500 pb del punto extremo del locus diana)
13	ADN	Segmento de orientación al ADN del ARNg C de C5 (Hc) (38 200 y 37 500 pb de los puntos extremos del locus diana)
14	ADN	Segmento de orientación al ADN del ARNg D de C5 (Hc) (43 500 y 32 200 pb de los puntos extremos del locus diana)
15	ADN	Segmento de orientación al ADN del ARNg E de C5 (Hc) (500 pb del punto extremo del locus diana)
16	ADN	Segmento de orientación al ADN del ARNg E2 de C5 (Hc) (100 pb del punto extremo del locus diana)
17	ADN	Segmento de orientación al ADN del ARNg A de Lrp5 (50 pb del punto extremo del locus diana)

ES 2 731 437 T3

	18	ADN	Segmento de orientación al ADN del ARNg B de Lrp5 (500 pb del punto extremo del locus diana)
5	19	ADN	Segmento de orientación al ADN del ARNg B2 de Lrp5 (1000 pb del punto extremo del locus diana)
	20	ADN	Segmento de orientación al ADN del ARNg C de Lrp5 (29 900 y 38 430 pb de los puntos extremos del locus diana)
10	21	ADN	Segmento de orientación al ADN del ARNg D de Lrp5 (29 950 y 38 380 pb de los puntos extremos del locus diana)
	22	ADN	Segmento de orientación al ADN del ARNg E2 de Lrp5 (1000 pb del punto extremo del locus diana)
15	23	ADN	Segmento de orientación al ADN del ARNg E de Lrp5 (500 pb del punto extremo del locus diana)
	24	ADN	Segmento de orientación al ADN del ARNg F de Lrp5 (50 pb del punto extremo del locus diana)
20	25	ADN	Segmento de orientación al ADN del ARNg A de Ror1 (200 pb del punto extremo del locus diana)
	26	ADN	Segmento de orientación al ADN del ARNg B de Ror1 (1000 pb del punto extremo del locus diana)
25	27	ADN	Segmento de orientación al ADN del ARNg D de Ror1 (54 300 y 55 500 pb de los puntos extremos del locus diana)
	28	ADN	Segmento de orientación al ADN del ARNg C de Ror1 (54 500 y 55 300 pb de los puntos extremos del locus diana)
30	29	ADN	Segmento de orientación al ADN del ARNg E de Ror1 (1000 pb del punto extremo del locus diana)
	30	ADN	Segmento de orientación al ADN del ARNg F de Ror1 (200 pb del punto extremo del locus diana)
35	31	ADN	Segmento de orientación al ADN del ARNg A de Trpa1 (100 pb del punto extremo del locus diana)
	32	ADN	Segmento de orientación al ADN del ARNg A2 de Trpa1 (500 pb del punto extremo del locus diana)
40	33	ADN	Segmento de orientación al ADN del ARNg B de Trpa1 (1000 pb del punto extremo del locus diana)
	34	ADN	Segmento de orientación al ADN del ARNg C de Trpa1 (25 600 y 19 740 pb de los puntos extremos del locus diana)
45	35	ADN	Segmento de orientación al ADN del ARNg D de Trpa1 (26 970 y 18 370 pb de los puntos extremos del locus diana)
	36	ADN	Segmento de orientación al ADN del ARNg E2 de Trpa1 (1000 pb del punto extremo del locus diana)
50	37	ADN	Segmento de orientación al ADN del ARNg E de Trpa1 (500 pb del punto extremo del locus diana)
	38	ADN	Segmento de orientación al ADN del ARNg F de Trpa1 (100 pb del punto extremo del locus diana)
55	39	ADN	cebador directo de 190045
	40	ADN	cebador directo de 190061
60	41	ADN	cebador directo de 190068
	42	ADN	cebador directo de 190030
	43	ADN	cebador directo de 190033 (igual al cebador directo de SV 48,3 en la Figura 5)
65	44	ADN	cebador directo de 190013
	45	ADN	cebador inverso de 190045

ES 2 731 437 T3

	46	ADN	cebador inverso de 190061
	47	ADN	cebador inverso de 190068
5	48	ADN	cebador inverso de 190030
	49	ADN	cebador inverso de 190033 (igual al cebador inverso de SV 48,3 en la Figura 5)
	50	ADN	cebador inverso de 190013
10	51	ADN	Sonda C2 (B6) - SNV 0,32 en la Figura 5
	52	ADN	Sonda T3 (B6) - SNV 1,2 en la Figura 5
	53	ADN	Sonda T6 (B6) - SNV 11,1 en la Figura 5
15	54	ADN	Sonda T7 (B6) - SNV 13,2 en la Figura 5
	55	ADN	Sonda T8 (B6) - SNV 17,5 en la Figura 5
	56	ADN	Sonda T9 (B6) - SNV 25,8 en la Figura 5
20	57	ADN	Sonda T10 (B6) - SNV 33,0 en la Figura 5
	58	ADN	Sonda T11 (B6) - SNV 38,3 en la Figura 5
	59	ADN	Sonda T13 (B6) - SNV 49,6 en la Figura 5
25	60	ADN	Sonda T14 (B6) - SNV 57,2 en la Figura 5
	61	ADN	Sonda C2 (129) - SNV 0,32 en la Figura 5
	62	ADN	Sonda T3 (129) - SNV 1,2 en la Figura 5
	63	ADN	Sonda T6 (129) - SNV 11,1 en la Figura 5
30	64	ADN	Sonda T7 (129) - SNV 13,2 en la Figura 5
	65	ADN	Sonda T8 (129) - SNV 17,5 en la Figura 5
	66	ADN	Sonda T9 (129) - SNV 25,8 en la Figura 5
35	67	ADN	Sonda T10 (129) - SNV 33,0 en la Figura 5
	68	ADN	Sonda T11 (129) - SNV 38,3 en la Figura 5
	69	ADN	Sonda T13 (129) - SNV 49,6 en la Figura 5
40	70	ADN	Sonda T14 (129) - SNV 57,2 en la Figura 5
	71	ADN	Cebador directo C2 - SNV 0,32 en la Figura 5
	72	ADN	Cebador directo T3 - SNV 1,2 en la Figura 5
45	73	ADN	Cebador directo T6 - SNV 11,1 en la Figura 5
	74	ADN	Cebador directo T7 - SNV 13,2 en la Figura 5
	75	ADN	Cebador directo T8 - SNV 17,5 en la Figura 5
50	76	ADN	Cebador directo T9 - SNV 25,8 en la Figura 5
	77	ADN	Cebador directo T10 - SNV 33,0 en la Figura 5
	78	ADN	Cebador directo T11 - SNV 38,3 en la Figura 5
55	79	ADN	Cebador directo T13 - SNV 49,6 en la Figura 5
	80	ADN	Cebador directo T14 - SNV 57,2 en la Figura 5
	81	ADN	Cebador inverso C2 - SNV 0,32 en la Figura 5
	82	ADN	Cebador inverso T3 - SNV 1,2 en la Figura 5
60	83	ADN	Cebador inverso T6 - SNV 11,1 en la Figura 5
	84	ADN	Cebador inverso T7 - SNV 13,2 en la Figura 5
	85	ADN	Cebador inverso T8 - SNV 17,5 en la Figura 5
65	86	ADN	Cebador inverso T9 - SNV 25,8 en la Figura 5

ES 2 731 437 T3

	87	ADN	Cebador inverso T10 - SNV 33,0 en la Figura 5
	88	ADN	Cebador inverso T11 - SNV 38,3 en la Figura 5
5	89	ADN	Cebador inverso T13 - SNV 49,6 en la Figura 5
	90	ADN	Cebador inverso T14 - SNV 57,2 en la Figura 5
	91	ADN	Cebador directo de SV 13,7 en la Figura 5
10	92	ADN	Cebador inverso de SV 13,7 en la Figura 5
	93	ADN	Cebador directo de SV 20,0 en la Figura 5
	94	ADN	Cebador inverso de SV 20,0 en la Figura 5
15	95	ADN	Cebador directo de SV 36,9 en la Figura 5
	96	ADN	Cebador inverso de SV 36,9 en la Figura 5
	97	ADN	Cebador directo de SV 56,7 en la Figura 5
20	98	ADN	Cebador inverso de SV 56,7 en la Figura 5
	99	ADN	cebador m-lr-f en la Figura 1
	100	ADN	cebador m-5'-f en la Figura 1
25	101	ADN	cebador m-A en la Figura 1
	102	ADN	cebador h-lr-r en la Figura 1
	103	ADN	cebador m-5'-r en la Figura 1
	104	ADN	cebador h-5'-r en la Figura 1
30	105	ADN	cebador m-F en la Figura 1
	106	ADN	cebador m-E2 en la Figura 1
	107	ARN	Segmento de orientación al ADN del ARNg A de Cmah
35	108	ARN	Segmento de orientación al ADN del ARNg B de Cmah
	109	ADN	Locus Cmah
	110	ADN	Locus Cmah corriente arriba del corte de ARNg A
40	111	ADN	Locus Cmah corriente abajo del corte de ARNg B
	112	ARN	Secuencia del locus Cmah suprimida por los ARNg A y B
	113	ADN	Cebador m-5'-F de C5 en la Figura 13
45	114	ADN	Cebador m-5'-R de C5 en la Figura 13
	115	ADN	Cebador h-5'-R de C5 en la Figura 13
	116	ADN	Cebador directo TaqMan® de Cmah en la Figura 15
50	117	ADN	Sonda TaqMan® de Cmah en la Figura 15
	118	ADN	Cebador inverso TaqMan® de Cmah en la Figura 15
	119	ADN	Cebador directo de 7064retU
55	120	ADN	Cebador inverso de 7064retU
	121	ADN	Sonda TaqMan® de 7064retU
	122	ADN	Cebador directo de 7064retD
	123	ADN	Cebador inverso de 7064retD
60	124	ADN	Sonda TaqMan® de 7064retD
	125	ADN	Cebador directo de 7140retU
	126	ADN	Cebador inverso de 7140retU
65	127	ADN	Sonda TaqMan® de 7140retU

ES 2 731 437 T3

	128	ADN	Cebador directo de 7140retD
	129	ADN	Cebador inverso de 7140retD
5	130	ADN	Sonda TaqMan® de 7140retD
	131	ADN	Cebador directo de LOA para mADAM6-2
	132	ADN	Cebador inverso de LOA para mADAM6-2
10	133	ADN	Sonda de LOA para mADAM6-2
	134	ADN	Cebador directo de LOA para hlgH31
	135	ADN	Cebador inverso de LOA para hlgH31
15	136	ADN	Sonda de LOA para hlgH31
	137	ADN	Cebador directo de LOA para hlgH9
	138	ADN	Cebador inverso de LOA para hlgH9
20	139	ADN	Sonda de LOA para hlgH9
	140	ADN	Cebador directo de LOA para hlgH1
	141	ADN	Cebador inverso de LOA para hlgH1
25	142	ADN	Sonda de LOA para hlgH1
	143	ADN	Cebador directo de GOA para Neo
	144	ADN	Cebador inverso de GOA para Neo
30	145	ADN	Sonda de GOA para Neo
	146	ADN	Cebador directo de ensayo de retención de 5' IgH Brazo1
	147	ADN	Cebador inverso de ensayo de retención de 5' IgH Brazo1
35	148	ADN	Sonda de ensayo de retención de 5' IgH Brazo1
	149	ADN	Cebador directo de ensayo de retención de mlgM398
	150	ADN	Cebador inverso de ensayo de retención de mlgM398
40	151	ADN	Sonda de ensayo de retención de mlgM398
	152	ADN	Cebador directo de ensayo de retención de mlgM1045
	153	ADN	Cebador inverso de ensayo de retención de mlgM1045
45	154	ADN	Sonda de ensayo de retención de mlgM1045
	155	ADN	Cebador directo de ensayo de retención de 3' IgH Brazo2
	156	ADN	Cebador inverso de ensayo de retención de 3' IgH Brazo2
50	157	ADN	Sonda de ensayo de retención de 3' IgH Brazo2
	158	ADN	Cebador directo de mlgHp2 progenitor
	159	ADN	Cebador inverso de mlgHp2 progenitor
55	160	ADN	Sonda de mlgHp2 progenitor
	161	ADN	Cebador directo de mlgKd2 progenitor
	162	ADN	Cebador inverso de mlgKd2 progenitor
60	163	ADN	Sonda de mlgKd2 progenitor
	164	ADN	Cebador directo de hlgK5 progenitor
	165	ADN	Cebador inverso de hlgK5 progenitor
65	166	ADN	Sonda de hlgK5 progenitor
	167	ADN	Secuencia de orientación al ADN de 3' ARNg-1_I
	168	ADN	Secuencia de orientación al ADN de 3' ARNg_II

169	ADN	Secuencia de orientación al ADN de 5' ARNg_I
170	ADN	Secuencia de orientación al ADN de 5' ARNg_II
171	ADN	Cebador directo de ensayo de retención de 5' IgH Brazo2
172	ADN	Cebador inverso de ensayo de retención de 5' IgH Brazo2
173	ADN	Sonda de ensayo de retención de 5' IgH Brazo2
174	ADN	Cebador directo de ensayo de retención de 3' IgH Brazo1
175	ADN	Cebador inverso de ensayo de retención de 3' IgH Brazo1
176	ADN	Sonda de ensayo de retención de 3' IgH Brazo1

15 **EJEMPLOS**

Ejemplo 1. Transformación mediada por CRISPR/Cas9 mediante el uso de un ARN guía o dos ARN guías.

20 **Materiales y métodos**

Cultivo de células ES, tamizaje, y electroporación

25 Los experimentos descritos en la presente descripción se realizaron con VGF1, nuestra línea de células ES XY híbrida C57BL6NTac/129S6SvEvF1 (Poueymirou y otros (2007) Nat. Biotechnol. 25:91-99; Valenzuela y otros (2003) Nat. Biotechnol. 21:652-659). Las células ES se cultivaron como se describió previamente (Matise y otros (2000) en Joyner, A.L. ed. Gene Targeting: a practical approach, pp. 100-132, Oxford University Press, Nueva York).

30 Las electroporaciones (EP) se realizaron con 7,5 millones de células en una cubeta de 2 mm de ranura en un volumen final de 0,12 ml. Las condiciones eléctricas para la EP fueron 700 V, 400 ohm de resistencia, y 25 microF de capacitancia mediante el uso de un sistema de electroporación BTX ECM 630 (Harvard Apparatus, Holliston, MA). La cantidad de LTVEC por EP fue 0,0015 mg, de plásmido que expresa Cas9 fue 0,005 mg y de plásmido que expresa los ARNGs fue 0,010 mg. Algunas EP se realizaron con la adición de 100 ng de un plásmido que confiere resistencia a puomicina para permitir la selección de clones sin seleccionar la resistencia a neomicina expresada por el LTVEC. Después de la EP, las células se sembraron en dos placas gelatinizadas de 15 cm y el medio se cambió diariamente. El medio de selección que contiene ya sea G-418 sulfato a 100 ug/ml o puomicina a 0,0015 mg/ml se inició 48 horas después de la EP y continuó hasta 10 días después de la EP. Las colonias se seleccionaron en PBS y se añadieron a una placa de 96 pocillos que contiene tripsina al 0,05 % y se permitió la disociación durante 15 minutos, se neutralizó con medio y se usó para el aislamiento de ADN para el tamizaje.

40 El método de modificación de alelo (Frendewey y otros (2010) Methods Enzymol. 476:295-307) se usó para identificar los clones de células ES transformados correctamente y para determinar los genotipos del alelo de ratón.

Diseño de secuencias guías

45 Aproximadamente 200 pb de ADN que rodean la posición de 50 pb, 100 pb, 500 pb, o 1 kb dentro de la porción eliminada de Lrp5 u otros genes diana, tanto corriente arriba como corriente abajo, se introdujeron en la herramienta de diseño de CRISPR (crispr.mit.edu) para obtener las secuencias de ARNg posibles. Después las secuencias de ARNg potenciales se filtraron para garantizar que solo permitirán el corte del ADN endógeno y no del inserto de humanización en el LTVEC.

50 **Clonación de ARN guías simples**

55 Los ARNGs se clonaron como dúplex de oligos (IDT) en pMB_sgRNA (promotor U6) en sitios Bsmbl fusionados al andamio de 77 pb para la expresión continua del ARN, o se adquirieron como plásmidos de expresión validados de GeneCopoeia (guías A, B, B2, E2, E, y F de LRP5). Los plásmidos producidos internamente se confirmaron por PCR y secuenciación de Sanger.

Plantilla de ADN para la confirmación del genotipo

60 El ADN se purificó a partir de la célula ES, clones derivados de células ES que se habían sometido a electroporación con un vector de transformación y un plásmido que expresa Cas9 y un plásmido que expresa uno de varios ARN guías (ARNg) o dos plásmidos que expresan combinaciones de ARNg diferentes. Los clones identificados por ensayos de PCR cuantitativa de modificación de alelo (es decir, pérdida de alelo o ganancia de alelo) que tienen una delección dirigida del locus diana de ratón e inserción del vector de transformación o que tienen delecciones inducidas por Cas9/ARNg se seleccionaron para ensayos de PCR convencionales complementarios.

65 **Diseño de oligonucleótidos**

Se diseñaron dos ensayos de PCR para cada combinación de ARNg. La primera PCR fue un ensayo de delección para detectar el colapso entre las secuencias de reconocimiento de ARN de CRISPR de combinaciones de ARNg diferentes. El segundo ensayo de PCR, que es un ensayo 5', incluyó dos ensayos de PCR. El primero fue un ensayo de 5' humano para alelos humanizados y se diseñó a través de la unión entre ratón y humano. El segundo fue un ensayo de 5' de ratón para alelos endógenos de ratón y se diseñó a través de la unión 5' de la delección dirigida.

Reacción de PCR y clonación TOPO

Se usó TaKaRa LA Taq ADN polimerasa (# de Cat. RR002M) para amplificar la plantilla de ADN de células ES. Cada mezcla de reacción del ensayo de PCR se analizó con un control negativo de agua. Las mezclas de ensayo contenían lo siguiente: 0,005 ml de plantilla de ADN de células ES; tampón IX LA PCR Buffer II (Mg²⁺ plus); 0,01 mM mezcla de dNTP; 0,0075 mM Oligo directo (cada una); 0,0075 mM Oligo inverso (cada una); 5000 unidades/ml LA Taq polimerasa; y ddH₂O hasta 0,025 ml.

El programa de termociclado de la PCR consistió en 94 °C durante un minuto; seguido de 35 ciclos de 94 °C durante 30 segundos, gradiente de hibridación a 60 °C durante 30 segundos, y 68 °C durante un minuto por kb amplificada; seguido de polimerización a 72 °C durante 10 minutos.

Los productos de PCR se fraccionaron por electroforesis en un gel de agarosa al 2 % con una escalera de ADN de 1 kb plus de Invitrogen (# de Cat. 10787-018) y/o escalera de ADN de 50 pb de Invitrogen (# de Cat. 10416-014). Los productos de PCR restantes se clonaron en el vector pCR4-TOPO según las instrucciones del kit de clonación TOPO TA de Invitrogen (# de Cat. K4575-02) para su secuenciación. Las reacciones de clonación se transformaron químicamente en células One Shot Top10 y estas se sembraron en placas de agar con X-gal a 0,06 mg/ml y kanamicina a 0,025 mg/ml.

Secuenciación

Las colonias blancas se inocularon en LB que contiene kanamicina a 0,025 mg/ml y se incubaron durante toda la noche con agitación a 37 °C. Cada colonia representó un amplicón de una población de productos ensayados. Se extrajo el ADN de cada cultivo bacteriano mediante el uso del kit de purificación de plásmidos a escala minipreparativa de QIAGEN (# de Cat. 12123). La secuencia de ADN de los insertos se determinó en una mezcla de reacción de secuenciación que incluía 0,002 ml de producto de PCR clonado en TOPO, solución potenciadora PCRx 1x (solución madre 10x) (Cat. X11495-017), 0,0075 mM oligo (M13F o M13R), y ddH₂O hasta 0,015 ml.

Análisis de la secuenciación

Se descartaron las secuencias indeterminadas y la secuencia del vector pCR4-TOPO de los resultados de la secuenciación, para aislar la secuencia de los insertos de PCR. Después los fragmentos secuenciados se alinearon con una referencia y se analizaron las variaciones.

Secuenciación de clones colapsados

Los productos de PCR de los clones positivos colapsados se clonaron en el vector pCR4-TOPO según las instrucciones del fabricante (Invitrogen # de Cat. K4575-02), después se transformaron químicamente en células One Shot Top10 y estas se sembraron en placas de agar con X-gal a 0,060 mg/ml y kanamicina a 0,025 mg/ml. Se extrajo el ADN de los cultivos bacterianos mediante el uso del kit de purificación de plásmidos a escala minipreparativa de QIAGEN (# de Cat. 12123). Después los resultados de la secuenciación de los insertos se alinearon con una referencia de colapso predicha y se analizaron las variaciones de indel. Se predijo que Cas9 escinde a 3 pares de bases del PAM en la secuencia reconocida por el ARNg. La secuencia dentro de la escisión predicha se eliminó de la referencia y el resto se usó para el alineamiento con los resultados.

Ensayos de discriminación alélica TaqMan® de variantes de nucleótido simple (SNV)

La reacción de discriminación alélica TaqMan® fue de 0,008 ml que contiene ADN genómico, sondas/cebadores específicos para cada polimorfismo, y mezcla maestra de PCR de expresión génica TaqMan®. Las sondas se encargaron a Life Technologies (Thermo) y los cebadores a IDT. La sonda del alelo 129 se etiquetó con colorante VIC; la sonda del alelo B6 se etiquetó con colorante FAM. Cada ensayo alélico TaqMan® se realizó por cuadruplicado en una placa de 384 pocillos y se ejecutó en una plataforma ViiA 7 de Applied BioSystems. El programa de ciclos de PCR de SNV fue el siguiente: 95 °C durante 10 minutos seguido de 40 ciclos de lo siguiente: 95 °C durante 15 segundos, 60 °C durante 60 segundos, y 60 °C durante 30 segundos. El análisis de la corrida y la evaluación de los resultados se hicieron mediante el uso del software de ViiA 7 v1.1.

Análisis FISH

Clones de células ES seleccionados se analizaron en Cell Line Genetics (Madison, Wisconsin) o en el Instituto Van Andel (Grand Rapids, Michigan) mediante el uso de hibridación fluorescente in situ (FISH) por sus procedimientos estándar. Se

proporcionaron los BAC de ratón y humano como sondas para el análisis con 2 colores.

Mejora del colapso del genoma y/o la humanización de los loci diana

5 Para efectuar una deleción precisa, en una etapa de la totalidad o parte de un gen de roedor y opcionalmente su reemplazo simultáneo con la totalidad o parte de su homólogo humano, las siguientes moléculas de ácidos nucleicos se introdujeron por electroporación en células ES de roedor: (1) un LTVEC; (2) un plásmido o ARNm que codifica una endonucleasa Cas9; y (3) uno o más plásmidos que codifican uno o más ARN guías simples de CRISPR (ARNg) o los ARNg como tal. En cada experimento, el LTVEC se linealizó. En algunos experimentos, el LTVEC comprendía la totalidad o parte de un gen humano que codifica el producto génico (proteína o ARN) flanqueado por brazos de homología de ADN de roedor diseñados para dirigir un evento de recombinación homóloga que elimina el gen de roedor e inserta el gen humano. En otros experimentos, el LTVEC se diseñó para transformar un locus separado tal como el locus Ch25h. En cualquier caso, el LTVEC también portaba un casete de selección por fármaco que dirige la expresión de una enzima (por ejemplo, neomicina fosfotransferasa) que imparte resistencia a un fármaco antibiótico (por ejemplo, G418).

15 Las células ES que captaron el LTVEC y lo incorporaron en sus genomas fueron capaces de crecer y formar colonias en una placa de cultivo de tejidos en un medio de crecimiento que contiene el fármaco antibiótico. Debido a que se introdujeron 500 a 1000 veces más moléculas de ácido nucleico que codifican CRISPR/Cas9 y que codifican ARNg que moléculas de LTVEC, la mayor parte de las colonias resistentes a fármaco que contienen LTVEC también contenían, al menos de manera transitoria, los componentes de CRISPR/Cas9. Se seleccionaron colonias resistentes a fármaco y se tamizaron por el método de modificación de alelo (Valenzuela y otros (2003) Nat. Biotech. 21:652-660; Frenthewey y otros (2010) Methods Enzymol. 476:295-307) para identificar los clones que tenían el alelo humanizado transformado correctamente. Además, se usaron ensayos de PCR en tiempo real que reconocen secuencias en los brazos de homología del LTVEC, referidos como ensayos de retención, para comprobar la transformación correcta del LTVEC en el genoma de ratón. La determinación del número de copias de estos ensayos de retención proporcionó más claridad para ayudar a distinguir los clones de ES transformados correctamente, que retuvieron un número de copias de dos, de los clones en los que una deleción grande inducida por Cas9 del locus diana de ratón coincide con la integración aleatoria del LTVEC en otra parte en el genoma, en cuyo caso los ensayos de retención tuvieron un número de copias de tres (o más). La capacidad de los pares de ARNg de crear deleciones grandes mediadas por Cas9 en el locus diana de ratón significa que podría añadirse ensayos de retención a los ensayos LOA y GOA estándar como se describió previamente para proporcionar más claridad y comprobar la transformación correcta. Por lo tanto, los ensayos de retención se diseñaron y se usaron junto con los ensayos LOA y GOA.

35 En cada experimento, se usaron uno o dos ARNg. Los ARNg usados individualmente dirigieron la escisión por Cas9 cerca del extremo 5' del locus diana (es decir, la deleción dirigida del gen de ratón), el medio del locus diana, o el extremo 3' del locus diana. Cuando se usaron dos ARNg, un ARNg dirigió la escisión por Cas9 cerca del extremo 5' del locus diana y el otro ARNg dirigió la escisión por Cas9 en el medio del locus diana o cerca del extremo 3' del locus diana.

40 Locus Lrp5

En un conjunto de experimentos, el LTVEC se diseñó para crear una deleción de 68 kb de la porción del gen Lrp5 de ratón (proteína 5 relacionada con el receptor de lipoproteínas de baja densidad) que codifica el ectodominio y un reemplazo simultáneo con un fragmento de 91 kb de la secuencia homóloga del gen LRP5 humano (Figura 1). El LTVEC comprendía el fragmento de 91 kb del gen LRP5 humano flanqueado por brazos de homología que contienen 7 kb y 33 kb de ADN genómico derivado de partes del locus Lrp5 de ratón que flanquean la secuencia de 68 kb del gen Lrp5 de ratón prevista para la deleción. En experimentos separados, el LTVEC de humanización de Lrp5 se combinó con un plásmido que codifica Cas9 y un segundo plásmido que codifica uno de ocho ARNg (A, B, B2, C, D, E2, E, F) diseñados para crear rupturas bicatenarias dentro de la región del gen Lrp5 de ratón reconocida para la deleción. Los ARNg se diseñaron de modo que eviten el reconocimiento de cualquier secuencia en la porción insertada del gen LRP5 humano. En otros experimentos, se combinó el LTVEC y el plásmido que codifica Cas9 con plásmidos que codifican dos ARNg diferentes que reconocen sitios diferentes dentro de la región del gen Lrp5 de ratón reconocida para la deleción.

Los clones de células ES resistentes a fármaco se tamizaron en cuanto a las humanizaciones dirigidas por ensayos de modificación de alelo (Valenzuela y otros (2003) Nat. Biotechnol. 21:652-659; Frenthewey y otros (2010) Methods Enzymol. 476:295-307) para secuencias dentro de la deleción y para secuencias dentro del casete de selección por fármaco y el inserto de gen humano. Los clones se calificaron como transformados correctamente si tenían pérdida de una de las dos secuencias del gen endógeno de ratón y ganaban una copia del inserto humano, y además retenían dos copias de secuencias de retención (ubicadas en el brazo de homología del LTVEC). Los dos ensayos de retención para este tamizaje fueron ensayos TaqMan® que usan los siguientes cebadores y sondas: cebador directo de 7064retU CCTCCTGAGCTTTCTTTGCAG (SEQ ID NO: 119); cebador inverso de 7064retU CCTAGACAACACAGACTGTATCA (SEQ ID NO: 120); sonda TaqMan® de 7064retU TTCTGCCCTTGAAAAGGAGAGGC (SEQ ID NO: 121); cebador directo de 7064retD CCTCTGAGGCCACCTGAA (SEQ ID NO: 122); cebador inverso de 7064retD CCCTGACAAGTTCTGCCTTCTAC (SEQ ID NO: 123); sonda TaqMan® de 7064retD TGCCCAAGCCTCTGCAGCTTT (SEQ ID NO: 124).

Los resultados de la humanización asistida por CRISPR/Cas9 del gen Lrp5 se resumen en la Tabla 2. Cuando se introdujo el LTVEC solo en células ES, 1,9 % de los clones resistentes a fármaco tamizados portaban un alelo humanizado heterocigótico transformado correctamente (ver la columna Het. Transf. en la Tabla 2, que incluye clones en los que el alelo que no es diana no se mutó en absoluto o tenía una mutación inducida por CRISPR pequeña tal como una delección pequeña causada por NHEJ). Por el contrario, la combinación del LTVEC con endonucleasas Cas9 guiadas por siete de los ocho ARNg probados (A, B, B2, C, D, E2, E y F; ver la Tabla 1) produjo mutaciones heterocigóticas monoalélicas transformadas correctamente a eficiencias en el intervalo de 2,1 a 7,8 %. Para la escisión por Cas9 guiada por B2 y D, además de la transformación monoalélica, se detectó humanización homocigótica bialélica a una frecuencia de 1,0-2,1 %. Nunca se ha observado transformación bialélica con un LTVEC por sí solo, incluso para alelos con delección simple y pequeña. Las células ES con humanización homocigótica de Lrp5 pueden convertirse directamente mediante el método VELOCIMOUSE® (Poueymirou y otros (2007) Nat. Biotech. 25:91-99) en ratones completamente derivados de células ES listos para estudios fenotípicos y de eficacia de fármacos.

Los ensayos MOA concebidos para detectar mutaciones de NHEJ inducidas por ARNg/Cas9 en o cerca de los sitios de escisión predichos demostraron actividad de mutación para todos los ARNg probados (no se muestran los datos). La proporción de mutaciones inducidas por ARNg ya sean monoalélicas o bialélicas detectadas entre todos los clones ensayados varió por locus y posición. No hubo una correlación fuerte entre la actividad de mutación de los ARNg y la transformación con LTVEC, pero las eficiencias de transformación más bajas se asociaron frecuentemente con los ARNg que tenían las frecuencias de mutación más bajas.

La combinación de dos ARNg que reconocen extremos diferentes de la región del gen Lrp5 reconocida para la delección aumentó la eficiencia de transformación de humanización total, predominantemente mediante el aumento de la frecuencia de eventos de transformación homocigóticos para tres de las cinco combinaciones probadas (Tabla 2). Debido a que la combinación de los ARNg tiene el potencial de crear delecciones grandes entre los sitios de escisión de Cas9 programados por los ARNg, también se observaron clones de células ES hemocigóticas que portaban una humanización dirigida en un alelo de Lrp5 y una delección grande inducida por CRISPR en el otro alelo (combinación de ARNg A + F, Tabla 2). Además, para dos de las combinaciones de ARNg (A + F y A + E2), se identificaron clones de células ES con un genotipo único: delecciones grandes mediadas por CRISPR en ambos alelos de Lrp5.

Tabla 2. Resultados del tamizaje de la humanización asistida por CRISPR/Cas9 del ectodominio de Lrp5 mediante el uso de ARNg individuales y ARNg combinados.

ARNg	Distancia del sitio de ARNg de los extremos 5'/3' de la delección dirigida (pb)	Eficiencia de transformación por tipo de alelo				
		Het. Transf. (% de Ef.)	Hemi. Transf. (% de Ef.)	Homo. Transf. (% de Ef.)	Total Transf. (% de Ef.)	Homo. Del. (% de Ef.)
A	50 (5')	7,8			7,8	
B	500 (5')	4,2			4,2	
B2	1000 (5')	6,2		1,0	7,2	
C	29 900 (5')/ 38 430 (3')	4,1			4,1	
D	29 950 (5')/ 38 380 (3')	5,2		2,1	7,3	
E2	1000 (3')	2,1			2,1	
E	500 (3')	0,0			0,0	
F	50 (3')	4,2			4,2	
A+F	A: 50 (5')	6,6	2,9	2,2	11,7	2,9
	F: 50 (3')					
B + E	B: 500 (5')	2,5			2,5	
	E: 500 (3')					
B2 + E2	B2: 1000 (5')	4,2		2,1	6,3	
	E2: 1000 (3')					
A + E	A: 50 (5')	4,6		6,2	10,8	
	E: 500 (3')					
A + E2	A: 50 (5')	2,0		4,0	6,0	4,0
	E2: 1000 (3')					

Ninguna	N/A	1,9		1,9	
---------	-----	-----	--	-----	--

5 Como se demuestra en la Tabla 2, se observó un aumento significativo en el porcentaje de clones que tenían transformación bialélica cuando se usaron dos ARNg que reconocen un solo locus en lugar de un ARNg (ver la Figura 2A), lo que indica que el uso de combinaciones de ARNg promueve las modificaciones bialélicas. La Figura 2A muestra una representación esquemática general de la delección simultánea de un gen de ratón y su reemplazo con una versión humana correspondiente mediante el uso de un LTVEC y dos ARN guías (A y B). Los tipos de alelos mutantes únicos que se observan a una frecuencia mucho mayor cuando se usan dos ARNg incluyen alelos con colapso homocigótico (Figura 2B; Δ/Δ), alelos con transformación homocigótica (Figura 2C; Hum/Hum), alelos con transformación hemocigótica (Figura 2D; (Hum/ Δ)), y otros alelos con transformación heterocigótica combinada (por ejemplo, un alelo tiene una humanización dirigida por LTVEC y el otro alelo tiene una mutación inducida por CRISPR tal como una delección pequeña) (Figura 2E).

15 Se realizaron varios ensayos de PCR para respaldar y confirmar los genotipos en base a los ensayos MOA. Los cebadores se muestran en la Figura 1 y pueden encontrarse en la Tabla 1. El LTVEC de Lrp5 tenía un brazo de homología 5' que era lo suficientemente corto (6,9 kb) para demostrar la transformación mediante una PCR ensayada para detectar una conexión física entre el inserto humano y la secuencia genómica de ratón adyacente (Figura 1). Se observó el producto de PCR esperado de 7,5 kb con ADN de los clones calificados como heterocigóticos, hemocigóticos, u homocigóticos pero no con el ADN de la línea de células ES progenitoras o de clones calificados como con delecciones bialélicas grandes (Figura 3A), lo que confirma por lo tanto las interpretaciones de transformación hechas por el tamizaje con MOA (es decir, LOA y GOA) y respalda las delecciones bialélicas grandes inferidas. El ensayo de PCR 5'-Del-J, que examinó las secuencias en las uniones de la delección y la inserción (Figura 3B), produjo un producto de 330 pb con el ADN de la línea de células ES progenitoras y de la mayoría de los clones humanizados heterocigóticos (no se muestran los datos). Para el clon heterocigótico AW-C3, el ensayo 5'-Del-J produjo un producto más pequeño que lo esperado (Figura 3B), lo que sugiere que la escisión por ARNg A/Cas9 indujo una mutación de delección pequeña en el alelo que no es diana, que también se detectó por un ensayo MOA para la escisión con ARNg A (no se muestran los datos). Como se esperaba, el ensayo 5'-Del-J fue negativo para los clones con alelos con delección hemocigótica, homocigótica, y bialélica. La PCR 5'-Ins-J (Figura 3B), que examinó las secuencias en la unión entre el extremo 5' del inserto de ADN humano y la secuencia flanqueante de ratón adyacente, produjo un producto de 478 pb en clones heterocigóticos, hemocigóticos, y homocigóticos, dado que estos tienen al menos un alelo con humanización dirigida. El ensayo 5'-Ins-J PCR no produjo producto para clones con delecciones bialélicas grandes (Figura 3B). Para confirmar las delecciones grandes en clones con delección hemocigótica y bialélica, se realizaron PCR con cebadores que reconocen secuencias fuera de los sitios diana de los pares de ARNg. La PCR Del(A + F), que ensayó una delección entre los sitios de los ARNg A y F (Figura 1), produjo un solo producto de aproximadamente 360 pb con el ADN de los clones AW-A8 y BO-F10 (Figura 3B), lo que confirmó que al menos uno de los alelos de Lrp5 tenía una delección grande. Del mismo modo, la PCR Del(A + E2), que ensayó una delección grande entre los sitios de los ARNg A y E2, produjo un solo producto de aproximadamente 250 pb con el ADN del clon BA-A7. Las PCR de delección, junto con los ensayos de unión, LOA, y GOA, respaldan el genotipo de delección bialélica grande. Los resultados de los ensayos que se muestran en la Figura 3A y 3B son ejemplos representativos de ensayos similares que se realizaron además de la hibridación fluorescente *in situ* (FISH; Figura 4A-C) para confirmar los genotipos bialélicos resumidos en la Tabla 2.

La hibridación fluorescente *in situ* (FISH) se usó para confirmar la humanización dirigida homocigótica del gen Lrp5. Los clones de células ES calificados mediante ensayos de PCR cuantitativos y convencionales como con transformación homocigótica de experimentos de transformación en los que el LTVEC de humanización de Lrp5 (Figura 1) se combinó con Cas9 y dos ARNg (A más F o A más E2) se enviaron a un servicio comercial de citología para FISH y análisis de cariotipo. Un cromosoma artificial bacteriano (BAC) que porta el gen Lrp5 de ratón se etiquetó con un marcador fluorescente rojo y se usó como sonda para identificar loci Lrp5 endógenos, y un BAC que porta el gen LRP5 humano se etiquetó con un marcador fluorescente verde y se usó como sonda para identificar las cromátidas transformadas con el inserto humano. Las sondas BAC etiquetadas se hibridaron con frotis de metafase de los clones transformados y se visualizaron por microscopía de fluorescencia. Los cromosomas en los frotis se visualizaron por tinción con DAPI (4',6-diamidino-2-fenilindol), y los cariotipos separados de cada clon se determinaron por tinción Giemsa. Un resultado típico se muestra en la Figura 4A para el clon AW-D9, que se encontró que tiene un cariotipo normal 40XY (no se muestra). La fotografía compuesta en la Figura 4A muestra que la señal roja de la sonda de BAC de ratón y la señal verde de la sonda de BAC humano se colocalizan en la banda citológica B en ambas copias del cromosoma 19 de ratón, la ubicación conocida del gen Lrp5. La fotografía compuesta en la Figura 4C muestra la misma transformación homocigótica para otro clon (BA-D5). Estos resultados confirman que el fragmento de 91 kb del gen LRP5 humano en el LTVEC de humanización (Figura 1) se insertó correctamente en el locus Lrp5 de ratón previsto en ambos cromosomas 19 homólogos en los clones AW-D9 y BA-D5. Por el contrario, la fotografía compuesta en la Figura 4B muestra que la señal roja de la sonda de BAC de ratón y la señal verde de la sonda de BAC humano se colocalizan en la banda citológica B en una sola copia del cromosoma 19 de ratón (flecha sólida), mientras que solo la señal roja de la sonda de BAC de ratón se localiza en la banda citológica B en la otra copia del cromosoma 19 de ratón. Estos resultados confirman que el fragmento de 91 kb del gen LRP5 humano en el LTVEC de humanización (Figura 1) se insertó correctamente en el locus Lrp5 de ratón previsto solamente en una copia del cromosoma 19 (transformación heterocigótica). También indican (junto con otros controles que no se muestran) que la sonda de BAC humano no se hibrida de manera cruzada con el locus Lrp5 de ratón sino que solo reconoce el inserto LRP5 humano.

La presencia en ciertos clones de mutaciones de indel inducidas por CRISPR idénticas formadas en ambos alelos por reparación por unión de extremos no homólogos evidente sugirió la ocurrencia de eventos de conversión génica en células híbridas F1H4 (que se componen de 50 % de la cepa 129SvS6 y 50 % de la cepa C57BL/6N). Para ganar conocimientos del mecanismo que fundamenta la transformación bialélica mejorada cuando se usan dos ARNg, se tamizaron siete clones que tenían humanizaciones dirigidas homocigóticas o deleciones homocigóticas grandes inducidas por CRISPR seguido de la transformación con el LTVEC y cualquiera de las combinaciones de ARNg A más F o A más E2.

La Figura 5 muestra ejemplos de ensayos diseñados para examinar eventos de conversión génica mediados por dos ARN guías. Específicamente, la posibilidad de conversión génica se examinó mediante el análisis de la pérdida de heterocigosidad (LOH) en células ES híbridas F1H4 (que se componen de 50 % de la cepa 129 SvS6 y 50 % de la cepa C57BL/6N). La conversión génica puede demostrarse por la pérdida de heterocigosidad en polimorfismos conocidos entre 129SvS6 (129) y C57BL/6N (B6), y por lo tanto los ensayos de PCR se diseñaron para diferenciar entre estos dos tipos de alelo. Los polimorfismos de variantes estructurales (SV) se ensayaron por PCR convencionales diseñadas para detectar las diferencias entre los alelos 129 y B6. Aunque solo uno de los ensayos de SV usados a continuación se muestra en la Figura 5, el concepto es el mismo para cada uno. Los cebadores se diseñaron en base a las variaciones estructurales (SV) entre las cepas B6 y 129 de ratón y se muestran en la Tabla 1. Las condiciones de diseño de los cebadores se restringieron para identificar SV -25 pb y producir productos de PCR -300 pb; estas condiciones se seleccionaron de manera que cualquier cambio sería visible por electroforesis en gel.

Antes de ejecutar las PCR de los clones, los ensayos se validaron y optimizaron contra ADN de células ES de tipo silvestre de las cepas B6, 129 y de la línea de células ES F1H4. Los conjuntos de cebadores que produjeron bandas de PCR distinguibles específicas para los alelos B6 o 129 y fueron consistentes en la producción de estas dos mismas bandas distinguibles mediante el uso de ADN de F1H4 se seleccionaron para la prueba en clones. Para el cromosoma 19 (la ubicación del gen Lrp5), se seleccionaron seis conjuntos de cebadores-ID 190045, 190061, 190068, 190030, 190033, 190013-para el uso en clones con Lrp5 humanizado genotipificados como "con transformación homocigótica" o "con colapso homocigótico" por ensayos de modificación de alelo (MOA) y PCR convencional. Los ensayos de PCR de SV se espaciaron a lo largo del cromosoma 19 desde el locus Lrp5 hasta el extremo telomérico del cromosoma, en el intervalo de -13,7 a -56,2 Mb del locus Lrp5. Las distancias aproximadas (en Mb) de los ensayos de SV en el cromosoma 19 desde el locus Lrp5 son las siguientes: 13,7 para el ensayo 190045, 19,0 para el ensayo 190061, 35,0 para el ensayo 190068, 37,4 para el ensayo 190030, 48,3 para el ensayo 190033, y 56,2 para el ensayo 190013. Solo el ensayo 190033 se muestra en la Figura 5 (mostrado como SV 48,3), pero los cebadores para los ensayos 190045, 190061, 190068, 190030, 190033, y 190013 se muestran en la Tabla 1.

Las PCR se ejecutaron en el ADN de estos clones así como en ADN control de F1H4, ADN control de 129, y ADN control de B6. Los productos de PCR se fraccionaron por electroforesis en geles de poliacrilamida al 6 %, que posteriormente se tiñeron con GelRed. Los clones que producen dos bandas coincidieron con el control de F1H4, que desde la optimización anterior mostró que la banda superior era específica para el alelo 129 y la banda inferior era específica para el alelo B6. Los clones que produjeron solamente una banda mostraron solo la banda de B6 o la de 129. Los clones AW-A7, AW-F10, BA-D5, BA-F2, BC-H9, y BR-B4 mostraron solamente la banda de B6 para los seis ensayos, mientras que el clon BO-A8 mostró solamente la banda de 129 para los seis ensayos. Como se mencionó previamente, estos clones se genotipificaron como con transformación homocigótica o con colapso homocigótico por MOA y/o PCR, e involucraron varias combinaciones de ARNg (A más F, A más E2, B2, y D). La presencia de solo una única banda alélica sugirió que está teniendo lugar un evento de conversión génica, si no hubiera conversión, ambas bandas aún estarían presentes como en el control de F1H4.

Además, las variantes de nucleótido simple (SNV) entre los alelos 129 y B6 se ensayaron por ensayos de discriminación alélica TaqMan®. Las posiciones aproximadas de los ensayos de SNV en el mapa del cromosoma 19 en la Figura 5 se muestran con puntas de flecha con sus distancias (en Mb) del locus Lrp5 proporcionadas debajo. Las distancias (en Mb) del locus Lrp5 son las siguientes: 0,32 centromérica de Lrp5 (C2), 1,2 telomérica de Lrp5(T3), 11,1 telomérica de Lrp5 (T6), 13,2 telomérica de Lrp5 (T7), 17,5 telomérica de Lrp5 (T8), 25,8 telomérica de Lrp5 (T9), 33,0 telomérica de Lrp5(T10), 38,3 telomérica de Lrp5 (T11), 49,6 telomérica de Lrp5 (T13), y 57,2 telomérica de Lrp5 (T14). Las sondas específicas para 129 y específicas para B6 y los pares de cebadores se muestran en la Tabla 1.

La Tabla 3 muestra siete ejemplos de clones de células ES que exhibieron eventos de conversión génica evidentes en el brazo largo del cromosoma 19 en una dirección telomérica del locus diana Lrp5 por LOH tanto para alelos con SV como con SNV. Los clones de células ES se derivaron de experimentos de transformación independientes que combinaron el LTVEC de humanización de Lrp5 (Figura 1) con uno o dos ARNg, como se indicó. Las posiciones de los sitios de reconocimiento de ARNg se muestran encima de la representación del gen Lrp5 en la Figura 5 (flecha gruesa que apunta hacia la izquierda). Los ensayos de genotipificación indicaron que seis de los siete clones tenían humanizaciones con transformación homocigótica del gen Lrp5, mientras que uno tenía un colapso homocigótico (delección grande entre los sitios de los ARNg). En seis de los siete clones, los alelos 129 se perdieron, dejando solamente los alelos B6. En el otro clon, los alelos B6 se perdieron, dejando solamente los alelos 129. Todos los clones se mantuvieron heterocigóticos para los alelos ensayados en el lado centromérico del locus Lrp5 (es decir, todos los clones eran heterocigóticos para B6/129 con el ensayo de SNV C2). La LOH observada en los siete clones indica que un mecanismo mediante el cual se obtienen alelos homocigóticos genéticamente modificados cuando un LTVEC se combina con uno, o con mayor frecuencia, dos

ARNg es una primera modificación genética dirigida en un alelo seguido de un evento de conversión génica por recombinación dirigida por homología que copia la modificación genética dirigida de un cromosoma a su homólogo.

Tabla 3. Resultados del ensayo de pérdida de heterocigosidad.

Clon	ARNg	Tipo de alelo de Lrp5	Ensayos de pérdida de heterocigosidad (SV y SNV)
AW-A7	A+F	Con transformación homocigótica	Solo se detectaron los alelos B6
AW-F10	A + F	Colapso homocigótico	Solo se detectaron los alelos B6
BO-A8	A + F	Con transformación homocigótica	Solo se detectaron los alelos 129
BA-D5	A + E2	Con transformación homocigótica	Solo se detectaron los alelos B6
BA-F2	A + E2	Con transformación homocigótica	Solo se detectaron los alelos B6
BC-H9	B2	Con transformación homocigótica	Solo se detectaron los alelos B6
BR-B4	D	Con transformación homocigótica	Solo se detectaron los alelos B6

Locus C5 (Hc)

En otro conjunto de experimentos, el LTVEC se diseñó para crear una delección de 76 kb del gen de ratón del componente 5 del complemento (C5 o Hc (complemento hemolítico)) y un reemplazo simultáneo con un fragmento de 97 kb del gen C5 humano homólogo (Figura 6). El locus diana comprendía el exón 2 hasta el codón de terminación del gen C5 (Hc). El LTVEC comprendía el fragmento de 97 kb del gen C5 humano flanqueado por brazos de homología que contienen 35 kb y 31 kb de ADN genómico derivado de partes del locus C5 (Hc) de ratón que flanquean la secuencia de 76 kb del gen C5 (Hc) de ratón prevista para la delección. En experimentos separados, el LTVEC de humanización de C5 (Hc) se combinó con un plásmido que codifica Cas9 y un segundo plásmido que codifica uno de seis ARNg (A, B, C, D, E, y E2; ver la Tabla 1) diseñados para crear rupturas bicatenarias dentro de la región del gen C5 (Hc) de ratón reconocida para la delección. Los ARNg se diseñaron de modo que eviten el reconocimiento de cualquier secuencia en la porción insertada del gen C5 humano. En otros experimentos, se combinó el LTVEC y el plásmido que codifica Cas9 con plásmidos que codifican dos ARNg diferentes que reconocen sitios diferentes dentro de la región del gen C5 (Hc) de ratón reconocida para la delección. En algunos experimentos, se usó un LTVEC de control que transforma el locus Ch25h en lugar del LTVEC de humanización de C5 (Hc). El LTVEC de control, que se diseña para eliminar la secuencia codificante completa de Ch25h (~1 kb) e insertar los casetes de selección de puomicina y neomicina en el locus Ch25h, se usó como medio para seleccionar clones resistentes a fármaco que no se transformaron por recombinación homóloga en el locus C5 (Hc).

Los resultados de la humanización asistida por CRISPR/Cas9 del gen C5(Hc) se muestran en la Tabla 4 y son similares a los resultados obtenidos para la humanización asistida por CRISPR/Cas9 del gen Lrp5. La eficiencia de transformación con el LTVEC solo fue mayor (6,1 %) para la humanización de C5 (Hc) que para Lrp5, pero la adición de Cas9 y ARNg mejoró la eficiencia de transformación para cuatro de los seis ARNg probados. Al igual que para Lrp5, la combinación de ARNg (es decir, el uso de dos ARNg) para la humanización de C5 (Hc) aumentó aún más la eficiencia de transformación total, predominantemente mediante el aumento la frecuencia de eventos de transformación hemocigóticos y homocigóticos. Se encontraron además clones de células ES con delecciones grandes inducidas por CRISPR en ambos alelos (observadas a frecuencias de 1,8 % a 3,6 %). Además, cuando el LTVEC que transforma el locus Ch25h se usó en combinación con dos ARNg de C5(Hc), los clones con alelos homocigóticos que colapsaron entre las dos secuencias de reconocimiento de ARN de CRISPR de los ARNg se observaron a frecuencias de 1,2 % a 6 %, lo que indica que los eventos de colapso ocurren independientemente de los eventos de recombinación homóloga en el locus diana. Al igual que para Lrp5, se usaron ensayos de retención para confirmar los clones transformados correctamente. Los dos ensayos de retención para este tamizaje fueron ensayos TaqMan® que usan los siguientes cebadores y sondas: cebador directo de 7140retU CCCAGCATCTGACGACACC (SEQ ID NO: 125); cebador inverso de 7140retU GACCACTGTGGGCATCTGTAG (SEQ ID NO: 126); sonda TaqMan® de 7140retU CCGAGTCTGCTGTTACTGTTAGCATCA (SEQ ID NO: 127); cebador directo de 7140retD CCCGACACCTTCTGAGCATG (SEQ ID NO: 128); cebador inverso de 7140retD TGCAGGCTGAGTCAGGATTTG (SEQ ID NO: 129); sonda TaqMan® de 7140retD TAGTCAGTTTTGTGACACCCAGA (SEQ ID NO: 130).

Tabla 4. Resultados del tamizaje de la humanización asistida por CRISPR/Cas9 del gen C5 (Hc) mediante el uso de ARNg individuales y ARNg combinados.

			Eficiencia de transformación por tipo de alelo				
ARNg	Distancia del sitio de ARNg de los extremos 5'/3' de la delección dirigida (pb)	LTVEC	Het. Transf. (% de Ef.)	Hemi. Transf. (% de Ef.)	Homo. Transf. (% de Ef.)	Total Transf. (% de Ef.)	Homo. Del. (% de Ef.)
A	100 (5')	C5	16,6			16,6	
B	500 (5')	C5	14,5			14,5	
C	38 200 (5')/ 37 500 (3')	C5	11,4			11,4	
D	43 500 (5')/ 32 200 (3')	C5	7,3			7,3	
E	500 (3')	C5	4,2			4,2	
E2	100 (3')	C5	6,2			6,2	
A+C	A: 100 (5') C: 37 500 (3')	C5	19,6	7,1	0,6	27,3	0,6
A+C	A: 100 (5') C: 37 500 (3')	Ch25h	N/A	N/A	N/A	N/A	6,0
A+E2	A: 100 (5') E2: 100 (3')	C5	19,0	3,6	1,2	23,8	3,0
A+E2	A: 100 (5') E2: 100 (3')	Ch25h	N/A	N/A	N/A	N/A	1,2
Ninguna	N/A	C5	6,1			6,1	

Se usó hibridación fluorescente *in situ* (FISH) para confirmar la humanización dirigida homocigótica del gen C5 (Hc). Los clones de células ES calificados mediante ensayos de PCR cuantitativos y convencionales como con transformación homocigótica de los experimentos de transformación en los que el LTVEC de humanización de C5 (Hc) (Figura 6) se combinó con Cas9 y dos ARNg se enviaron a un servicio comercial de citología para FISH y análisis de cariotipo. Un cromosoma artificial bacteriano (BAC) que porta el gen C5 (Hc) de ratón se etiquetó con un marcador fluorescente rojo y se usó como sonda para identificar loci endógenos, y un BAC que porta el gen C5 humano se etiquetó con un marcador fluorescente verde y se usó como sonda para identificar cromátidas transformadas con el inserto humano. Las sondas BAC etiquetadas se hibridaron con frotis en metafase de los clones transformados y se visualizaron por microscopía de fluorescencia. Los cromosomas en los frotis se visualizaron por tinción con DAPI (4',6-diamidino-2-fenilindol), y los cariotipos separados de cada clon se determinaron por tinción Giemsa. Un resultado típico se muestra en la Figura 7B para el clon O-E. La fotografía compuesta en la Figura 7B muestra que la señal roja de la sonda de BAC de ratón y la señal verde de la sonda de BAC humano se colocan en el locus C5 (Hc) en ambas copias del cromosoma 2 de ratón, la ubicación conocida del gen C5 (Hc). Estos resultados confirman que el fragmento de 97 kb del gen C5 humano en el LTVEC de humanización (Figura 6) se insertó correctamente en el locus C5 (Hc) de ratón previsto en ambos cromosomas 2 homólogos en el clon O-E3. Por el contrario, la fotografía compuesta en la Figura 7A muestra que la señal roja de la sonda de BAC de ratón y la señal verde de la sonda de BAC humano se colocan en una sola copia del cromosoma 2 de ratón (flecha sólida), mientras que solo la señal roja de la sonda de BAC de ratón se localiza en el locus C5 (Hc) en la otra copia del cromosoma 2 de ratón. Estos resultados confirman que el fragmento de 97 kb del gen C5 humano en el LTVEC de humanización (Figura 6) se insertó correctamente en el locus C5 (Hc) de ratón previsto solamente en una copia del cromosoma 2 (transformación heterocigótica) en el clon Q-E9.

50 Locus Ror1

En otro conjunto de experimentos, el LTVEC se diseñó para crear una delección de 110 kb del gen Ror1 (receptor transmembrana de proteína quinasa de tirosina ROR1) de ratón y un reemplazo simultáneo con un fragmento de 134 kb del gen ROR1 humano homólogo (Figura 8). El LTVEC comprendía el fragmento de 134 kb del gen ROR1 humano flanqueado por brazos de homología que contienen 41,8 kb y 96,4 kb de ADN genómico derivado de partes del locus Ror1 de ratón que flanquean la secuencia de 110 kb del gen Ror1 de ratón prevista para la delección. En experimentos separados, el LTVEC de humanización de Ror1 se combinó con un plásmido que codifica Cas9 y un segundo plásmido que codifica uno de seis ARNg (A, B, C, D, E, y F; ver la Tabla 1) diseñados para crear rupturas bicatenarias dentro de la región del gen Ror1 de ratón reconocida para la delección. Los ARNg se diseñaron de modo que eviten el reconocimiento de cualquier secuencia en la porción insertada del gen ROR1 humano. En otros experimentos, se combinó el LTVEC y el plásmido que codifica Cas9 con plásmidos que codifican dos ARNg diferentes que reconocen sitios diferentes dentro del gen Ror1 reconocido para la delección.

Los resultados de la humanización asistida por CRISPR/Cas9 del gen Ror1 se muestran en la Tabla 5 y son similares a los resultados obtenidos para la humanización asistida por CRISPR/Cas9 de los genes Lrp5 y C5 (Hc). La eficiencia de transformación con LTVEC solo fue 0,3 %, y la adición de Cas9 y ARNg aumentó ligeramente la eficiencia de

transformación de dos de los seis ARNg probados. La combinación de los ARNg A y F aumentó la eficiencia de transformación de Ror1 total a 6,3 % mediante el aumento la frecuencia de los eventos de transformación heterocigóticos y hemicigóticos. Se encontraron además clones de células ES con deleciones grandes inducidas por CRISPR en ambos alelos (observadas a una frecuencia de 1,6 %).

Tabla 5. Resultados del tamizaje de la humanización asistida por CRISPR/Cas9 del gen Ror1 mediante el uso de ARNg individuales y ARNg combinados.

ARNg	Distancia del sitio de ARNg de los extremos 5'/3' de la deleción dirigida (pb)	Eficiencia de transformación por tipo de alelo				
		Het. Transf. (% de Ef.)	Hemi. Transf. (% de Ef.)	Homo. Transf. (% de Ef.)	Total Transf. (% de Ef.)	Homo. Del. (% de Ef.)
A	200 (5')	0,7			0,7	
B	1000 (5')	0,0			0,0	
D	54 300 (5')/ 55 500 (3')	0,7			0,7	
C	54 500 (5')/ 55 300 (3')	0,0			0,0	
E	1000 (3')	0,0			0,0	
F	200 (3')	0,3			0,3	
A: + F	A: 200 (5')	4,2	2,1		6,3	1,6
	F: 200 (3')					
Ninguna	N/A	0,3			0,3	

30 Locus Trpa1

En otro conjunto de experimentos, el LTVEC se diseñó para crear una deleción de 45,3 kb del gen Trpa1 (canal catiónico receptor de potencial transitorio, subfamilia A, miembro 1) de ratón y un reemplazo simultáneo con un fragmento de 54,5 kb del gen TRPA1 humano homólogo (Figura 9). El LTVEC comprendía el fragmento de 54,5 kb del gen TRPA1 humano flanqueado por brazos de homología que contienen 41,0 kb y 58,0 kb de ADN genómico derivado de partes del locus Trpa1 de ratón que flanquean la secuencia de 45,3 kb del gen Trpa1 de ratón prevista para la deleción. En experimentos separados, el LTVEC de humanización de Trpa1 se combinó con un plásmido que codifica Cas9 y un segundo plásmido que codifica uno de ocho ARNg (A, A2, B, C, D, E2, E, y F; ver la Tabla 1) diseñados para crear rupturas bicatenarias dentro de la región del gen Trpa1 de ratón reconocida para la deleción. Los ARNg se diseñaron de modo que eviten el reconocimiento de cualquier secuencia en la porción insertada del gen TRPA1 humano. En otros experimentos, se combinó el LTVEC y el plásmido que codifica Cas9 con plásmidos que codifican dos ARNg diferentes que reconocen sitios diferentes dentro del gen Trpa1 reconocido para la deleción.

Los resultados de la humanización asistida por CRISPR/Cas9 del gen Trpa1 se muestran en la Tabla 6 y son similares a los resultados obtenidos para la humanización asistida por CRISPR/Cas9 de los genes Lrp5 y C5 (Hc). La eficiencia de transformación con LTVEC solo fue 0,3 %, y la adición de Cas9 y ARNg aumentó la eficiencia de transformación de seis de los ocho ARNg probados. La combinación de los ARNg B y F aumentó la eficiencia de transformación de Trpa1 total a 3,4 % mediante el aumento la frecuencia de los eventos de transformación heterocigóticos, hemicigóticos, y homocigóticos. Se encontraron además clones de células ES con deleciones grandes inducidas por CRISPR en ambos alelos (observadas a una frecuencia de 0,3 %).

Tabla 6. Resultados del tamizaje de la humanización asistida por CRISPR/Cas9 del gen Trpa1 mediante el uso de ARNg individuales y ARNg combinados.

ARNg	Distancia del sitio de ARNg de los extremos 5'/3' de la deleción dirigida (pb)	Eficiencia de transformación por tipo de alelo				
		Het. Transf. (% de Ef.)	Hemi. Transf. (% de Ef.)	Homo. Transf. (% de Ef.)	Total Transf. (% de Ef.)	Homo. Del. (% de Ef.)
A	100 (5')	1,0			1,0	
A2	500 (5')	2,1			2,1	
B	1000 (5')	1,4			1,4	
C	25 600 (5')/ 19 740 (3')	1,0			1,0	

	D	26 970 (5')/ 18 370 (3')	2,1			2,1	
	E2	1000 (3')	0,0			0,0	
5	E	500 (3')	0,0			0,0	
	F	100 (3')	0,7			0,7	
	B + F	B: 1000 (5')	2,8	0,3	0,3	3,4	0,3
10		F: 100 (3')					
	Ninguna	N/A	0,3			0,3	

15 Tal como ilustran estos ejemplos, el uso de pares de ARN guías en sitios ampliamente separados mejoró el aumento de la humanización heterocigótica en comparación con los ARNg individuales. Además, el uso de pares de ARN guías promovió eventos bialélicos en comparación con los ARNg individuales. Por el contrario a la transformación con un ARNg, la transformación con dos ARNg da como resultado la creación de células con transformación homocigótica (Hum/Hum) en las que ambos alelos tienen una humanización dirigida, células con delección homocigótica (Δ/Δ) en las que ningún alelo se transformó con el LTVEC de humanización pero ambos tienen delecciones grandes, y células con transformación hemicigótica (Hum/ Δ) en las que un alelo tiene una humanización dirigida y el otro tiene una delección grande inducida por el par de ARNg/Cas9. En primer lugar, se encontraron clones transformados correctamente que tenían humanizaciones muy grandes precisas e idénticas en ambos alelos diana (por ejemplo, células que eran homocigóticas para la modificación génica dirigida). Aunque también se observaron clones con transformación homocigótica cuando se usó un ARNg para lograr la humanización de Lrp5, estos aparecieron a una frecuencia mucho menor que cuando se emplearon dos ARNg (ver la Tabla 2). Del mismo modo, no se observó transformación homocigótica cuando se usó un ARNg para lograr la humanización de C5 (Hc) o la humanización de Trpa1, pero se observó transformación homocigótica cuando se usaron dos ARNg con el vector de transformación (ver las Tablas 4 y 6). De manera similar, se encontraron clones transformados correctamente que eran hemicigóticos para la modificación génica (es decir, tenían una humanización dirigida con precisión en un alelo y una delección muy grande, en ocasiones con eliminación del gen, en el otro alelo) para la transformación de Lrp5, la transformación de C5 (Hc), la transformación de Ror1, y la transformación de Trpa1. Tales modificaciones no ocurrieron en absoluto cuando se usó un ARNg para lograr la humanización de Lrp5, C5 (Hc), Ror1, o Trpa1 (ver las Tablas 2, 4, 5, y 6, respectivamente).

35 En segundo lugar, se encontraron clones que tenían delecciones muy grandes idénticas (>45 kb) inducidas por eventos de escisión por Cas9 guiados por ambos ARNg en ambos alelos diana (es decir, las células eran homocigóticas para una delección grande, en ocasiones con eliminación del gen, en el locus diana). Estos tipos de mutaciones no requieren que el vector de transformación se dirija contra el mismo gen. Por ejemplo, como se muestra en la Tabla 4, se han obtenido células ES con delecciones homocigóticas inducidas por CRISPR mediante la combinación de Cas9 y dos ARNg con un vector de transformación dirigido contra un gen diferente no relacionado con el reconocido por los ARNg. Por lo tanto, una nucleasa Cas9 guiada por dos ARNg puede inducir una delección grande en las células sin adición de un vector de transformación. En tales casos, la selección por fármaco transitoria o estable proporcionada por un vector que expresa un gen de resistencia a fármaco puede facilitar el aislamiento de clones con delección homocigótica poco frecuentes por enriquecimiento de las células ES que han captado el ADN.

45 Ejemplo 2. Análisis de delecciones grandes inducidas por ARNg combinados.

Estructuras alélicas para delecciones grandes inducidas por ARNg combinados

50 Se realizó un análisis de secuencias adicional en los clones que comprenden delecciones grandes inducidas por eventos de escisión por Cas9 guiados por dos ARNg (ver la Tabla 7). Estas delecciones grandes parecían ser independientes de los eventos de recombinación homóloga dirigidos por LTVEC en el mismo locus en el que se obtuvieron delecciones grandes en el locus Lrp5 aproximadamente a la misma frecuencia cuando se combinaron los ARNg con un LTVEC de Lrp5 o uno que transforma el gen Ch25h a casi 30 Mb de distancia (no se muestran los datos). Para caracterizar las delecciones grandes, se realizaron PCR que abarcan la delección en 37 clones, 15 hemicigóticos y 22 con delecciones bialélicas grandes, de cuatro humanizaciones, y los clones individuales de los productos de PCR se secuenciaron. Las secuencias confirmaron las delecciones grandes, que estaban en el intervalo de 38 kb a 109 kb. Dos de los clones de células ES (clones de Lrp5 AW-A8 y BP-D3) tenían delecciones precisas perfectamente reparadas (68,2 kb) entre los sitios de escisión de Cas9 predichos, mientras que un clon (clon de Hc P-B12) tenía una inserción de un solo par de bases además de la delección de 38,1 kb. Veintisiete de los clones de células ES tenían delecciones que se extendían más allá de los sitios de escisión de Cas9, consistentes con la reparación imprecisa por unión de extremos no homólogos (NHEJ). Los siete clones de células ES restantes tenían mutaciones que combinaban delecciones e inserciones inducidas por NHEJ evidentes (por ejemplo, el clon de Lrp5 BP-F6 y el clon de Hc O-E4), cuatro de los cuales tenían inserciones de más de 200 pb que pudieron ubicarse en el mapa de sus loci genómicos de origen (no se muestran los datos). La inserción de 210 pb en el clon de Lrp5 BO-E9 estaba en una orientación invertida con respecto a una secuencia idéntica que reside aproximadamente 2600 pb fuera del sitio diana del ARNg F en la dirección centromérica (cromosoma 19 +, 3589138-3589347). Esta secuencia estaba presente en el brazo de homología 3' largo del LTVEC de Lrp5. Los clones de Lrp5 BP-

F6 y BP-G7 se derivaron de un experimento en el que se combinaron los ARNg A y F de Lrp5 con Cas9 y un LTVEC que transformó el gen Ch25h a 30 Mb de distancia de Lrp5 en la dirección telomérica. El clon BP-F6 tenía una inserción de 266 pb que parecía derivarse de un extremo del LTVEC de Ch25h dado que estaba formada por un fragmento de 103 pb idéntico a parte de la cadena principal del vector unido a un fragmento de 163 pb que era idéntico a una secuencia cercana a Ch25h y presente además en el brazo largo del LTVEC (cromosoma 19+, 34478136-34478298); este fragmento se insertó en la delección en una orientación invertida con respecto a la secuencia cromosómica endógena. El clon de Hc O-E4 tenía una inserción de 254 pb que estaba invertida con respecto a una secuencia idéntica encontrada dentro de la secuencia eliminada aproximadamente 3,1 kb de distancia del sitio de reconocimiento de ARNg A. La inserción de 1304 pb en el clon Hc S-D5 estaba formada por dos fragmentos: un pedazo de 1238 pb que estaba en la misma orientación que una secuencia idéntica encontrada dentro de la secuencia eliminada aproximadamente 1,4 kb de distancia del sitio predicho de escisión de Cas9 dirigida por ARNg E2 y un segundo pedazo de 66 pb que era una duplicación en una orientación invertida de una secuencia de 25 pb idéntica fuera del sitio de corte de ARNg E2.

Tabla 7. Estructuras alélicas de delecciones grandes inducidas por ARNg combinados.

Gen	Clon de células ES	Genotipo ¹	ARNg	Posiciones dentro de la delección dirigida (pb)	Tamaño de la delección (kb)	Secuencia eliminada adicional (pb)	Inserción (pb)	Clones de PCR
Lrp5	AW-A8	Δ/Δ	A+F	5'-50/50-3'	68,2	-	-	40
	BO-E9	Δ/Δ				12	210	17
	BP-D3	Δ/Δ				-	-	11
	BP-F6	Δ/Δ				30	266	6
	BP-G7	Δ/Δ				77		9
	BA-A7	Δ/Δ	A + E2	5'-5011000-3'	67,3	7		19
	BA-C7	Δ/Δ				84		32
	N-A11	Δ/Δ				14		12
	N-D4	Δ/Δ				10		15
	N-D11	Hum/Δ				20		10
						10		1
	N-E1	Hum/Δ				10		13
	N-E9	Hum/Δ				20		16
	O-C5	Hum/Δ				31		21
	O-D2	Hum/Δ				5		12
	O-E4	Hum/Δ				19	254	18
	O-E5	Hum/Δ	A+C	5'-100/38 200-3'	38,1	35	2	16
	O-E6	Hum/Δ				6		17
	O-F11	Hum/Δ				12	7	18
	O-F12	Hum/Δ				41		6
						35		1
Hc	P-B12	Δ/Δ					1	7
	P-C12	Δ/Δ				20		15
	P-D1	Δ/Δ				33		10
	P-G8	Δ/Δ				5		2
	Q-F5	Hum/Δ	A + E2	5'-1001100-3'	75,6	3	3	15
	Q-F10	Δ/Δ				46		13
	R-A5	Δ/Δ				18		14
	R-A7	Δ/Δ				37		15
	R-A9	Hum/Δ				261		8

	R-C8	Hum/ Δ				180		11
	R-D12	Hum/ Δ				182		10
5	R-F11	Hum/ Δ	A + E2	5'-1001100-3'	75,6	19		11
	S-A11	Δ/Δ				122		11
						46		1
10	S-D5	Δ/Δ				216	1304	8
	Y-B5	Δ/Δ	A+F	5'-2001200-3'	109	18		6
	Y-C7	Δ/Δ				23		7
15	Y-E1	Δ/Δ				12		3
	Trpa1	Δ/Δ	B + F	5'-1000/100-3'	44,6	30		8
20	¹ Hum/+, humanización dirigida de uno de los dos alelos nativos que da como resultado un genotipo heterocigótico; Hum/ Δ , una modificación bialélica en la que un alelo tiene una humanización dirigida y el otro tiene una delección grande inducida por Cas9-ARNg que da como resultado un genotipo hemicigótico; Hum/Hum, una modificación bialélica en la que ambos alelos tienen una humanización dirigida que da como resultado un genotipo homocigótico; Δ/Δ una modificación bialélica en la que ambos alelos tienen una delección grande inducida por Cas9-ARNg.							

Evidencias de conversión génica en alelos homocigóticos

Veintiuno de los 22 clones de células ES con delecciones bialélicas grandes tenían solamente una secuencia individual, única (Tabla 7), lo que indica que eran alelos homocigóticos. Para el clon de Hc S-A11, se encontró la misma secuencia en 11 de 12 clones de PCR. El único clon con una secuencia diferente pudiera sugerir la delección de dos alelos diferentes, pero también se encontró el mismo resultado para dos de los clones de Hc hemicigóticos, N-D11 y O-F12. Los distintos alelos con delección homocigótica en múltiples clones sugieren que pudieran haber surgido de un mecanismo de conversión génica en el que una delección en un cromosoma sirvió como plantilla para la reparación por recombinación homóloga de las escisiones por Cas9 en el cromosoma homólogo. Nos beneficiamos de la composición híbrida F1 129S6SvEvTac (129) y C57BL/6NTac (B6) de la línea de células ES VGF1 (Poueymirou y otros (2007) Nat. Biotechnol. 25:91-99; Valenzuela y otros (2003) Nat. Biotechnol. 21:652-659) para ensayar la conversión génica como pérdida de heterocigosidad (Lefebvre y otros (2001) Nat. Genet. 27:257-258) para variantes estructurales (SV) y de nucleótido simple (SNV) entre las cepas alrededor del locus Lrp5 en el cromosoma 19 (ver la Figura 5 para los cinco ensayos de SV y los diez ensayos de SNV que se usan a continuación) y el locus Hc en el cromosoma 2 (no se muestra). Para confirmar que cualquier pérdida de heterocigosidad no era el resultado de la pérdida de cromosomas completos, se realizaron ensayos de número de copias cromosómicas (CCN) en los sitios que eran idénticos entre las cepas 129 y B6. Para los alelos de Lrp5 humanizados o con delección se ensayaron múltiples SV y SNV posicionadas desde 1,2 Mb de distancia de Lrp5 en la dirección telomérica hasta el extremo del brazo largo del cromosoma 19 (Figura 5). Debido a la ubicación de Lrp5 cercana al centrómero, no se encontraron SV y solamente se encontró una SNV en el lado centromérico del gen. Para Hc, fue posible ensayar múltiples SV y SNV en cualquier lado del gen en el cromosoma 2 (no se muestra). Los resultados de seis de los clones de Lrp5 se muestran en las Figuras 10A-E y 11A-C.

La Figura 10A-E muestra los resultados de cinco ensayos de SV, cuyas posiciones estaban en el intervalo de 13,7 Mb de distancia de Lrp5 a 56,7 Mb de distancia cerca del extremo telomérico del brazo largo. Los cinco ensayos de SV produjeron dos productos de tamaños diferentes para los alelos 129 (mayor) y B6 (menor) en los controles de 129, B6, y VGF1. Las posiciones aproximadas de los ensayos de SV en el mapa del cromosoma 19 se muestran en la Figura 5 (ver el ensayo SV 13,7, ensayo SV 20,0, ensayo SV 36,9, ensayo SV 48,3, y ensayo SV 56,7). El número del ensayo representa el número de Mb teloméricas respecto a Lrp5. Los cebadores de estos ensayos se muestran en la Tabla 1, y los resultados se muestran en la Figura 10A-E. Dos de los clones, BC-H9 (Lrp5^{Hum/Hum}, ARNg B2) y BR-B4 (Lrp5^{Hum/Hum}, ARNg D), mostraron una pérdida de heterocigosidad que retenía la totalidad de los alelos de SV de B6, mientras que un tercer clon, BO-A8 (Lrp5^{Hum/Hum}, ARNg A + F), retuvo la totalidad de los alelos 129. Los otros tres clones, BO-F10 (Lrp5^{Hum/Hum}, ARNg A + F), BO-G11 (Lrp5^{Hum/Hum}, ARNg A + F), y BP-G7 (Lrp5 ^{Δ/Δ} , ARNg A + F), se mantuvieron heterocigóticos.

Además, las variantes de nucleótido simple (SNV) entre los alelos 129 y B6 se ensayaron por ensayos de discriminación alélica TaqMan®. Las posiciones aproximadas de los ensayos de SNV en el mapa del cromosoma 19 en la Figura 5 se muestran con puntas de flecha con los números de ensayo debajo, y sus distancias (en Mb) del locus Lrp5 se proporcionan debajo. Las distancias (en Mb) del locus Lrp5 son las siguientes: 0,32 centromérica de Lrp5 (C2), 1,2 telomérica de Lrp5 (T3), 11,1 telomérica de Lrp5 (T6), 13,2 telomérica de Lrp5 (T7), 17,5 telomérica de Lrp5 (T8), 25,8 telomérica de Lrp5 (T9), 33,0 telomérica de Lrp5 (T10), 38,3 telomérica de Lrp5 (T11), 49,6 telomérica de Lrp5 (T13), y 57,2 telomérica de Lrp5 (T14). Las sondas específicas para 129 y específicas para B6 y los pares de cebadores se muestran en la Tabla 1. Los resultados de tres clones (BC-H9, BO-A8, y BR-B4) que mostraron pérdida de heterocigosidad (LOH) telomérica por ensayos de SV se muestran en la Figura 11A-C. Los ensayos de SNV (Figura 11A-C y datos que no se muestran) confirmaron los eventos de conversión génica en el brazo largo del cromosoma 19 en el lado telomérico de Lrp5 (SNV 1,2

y SNV 57,2; ver la Figura 11B y la Figura 11C, respectivamente), pero el ensayo de SNV 0,32 (ver la Figura 11A) mostró que todos los clones se mantenían heterocigóticos para un alelo a 320 kb de distancia de Lrp5 en el lado centromérico. De los 24 clones de Lrp5^{Hum/Hum} o Lrp5^{Δ/Δ} ensayados, se encontraron seis que tenían evidencias de pérdida de heterocigosidad en el brazo largo completo del cromosoma 19 en el lado telomérico de Lrp5. Cinco de los clones (cuatro Lrp5^{Hum/Hum} y uno Lrp5^{Δ/Δ}) se convirtieron de heterocigóticos a homocigóticos para B6, mientras que un sexto clon (Lrp5^{Hum/Hum}) se convirtió a homocigótico para 129. Los ensayos de CCN demostraron la retención de dos copias del cromosoma 19. Los ensayos de pérdida de heterocigosidad similares de 21 clones homocigóticos para Hc revelaron que dos, R-E2 (Hc^{Hum/Hum}, ARNg A + F) y R-E8 (Hc^{Δ/Δ}, ARNg A + F), mostraron pérdida de heterocigosidad a homocigóticos para 129 para todas las SV y SNV en el lado telomérico del gen Hc mientras retuvieron la heterocigosidad para todos los alelos en el lado centromérico. Los ensayos de CCN indicaron que no hay pérdida del cromosoma 2.

Los resultados demuestran por primera vez que CRISPR/Cas9 puede mejorar la reparación dirigida por homología de humanizaciones grandes en una sola etapa de más de 100 kb, lo que expande las posibilidades de la modificación genética del genoma a gran escala. El beneficio más notable e inesperado de la combinación de LTVEC y ARNg/Cas9 fue su capacidad de promover humanizaciones dirigidas homocigóticas. Aunque se han informado mutaciones bialélicas y eventos de transformación homocigóticos en otros experimentos con CRISPR/Cas9, la mayor parte de estas modificaciones e inserciones génicas han sido órdenes de magnitud menores que nuestros alelos humanizados. Antes del uso de CRISPR/Cas9, nunca se había encontrado transformación homocigótica por un LTVEC, ni se había observado transformación simultánea de más de un gen cuando se combinaban múltiples LTVEC que transforman genes separados. Dada esta experiencia, la transformación homocigótica inducida por ARNg/Cas9 sugirió que en lugar de dos LTVEC que transforman ambos alelos por separado, un evento de transformación inicial en un alelo pudiera servir como plantilla para la conversión homóloga del otro alelo promovida por uno o más cortes de Cas9. La revelación de que las delecciones bialélicas grandes inducidas por pares de ARNg/Cas9 también eran homocigóticas (Tabla 7) proporcionó apoyo adicional para un mecanismo de conversión génica.

Los ensayos de pérdida de heterocigosidad (Figura 5) demostraron que la conversión génica a gran escala de múltiples alelos que cubren un fragmento grande del cromosoma en el lado telomérico del gen diana era responsable de algunas de las humanizaciones y delecciones grandes homocigóticas. Este tipo de conversión génica direccional de largo alcance es consistente con la recombinación mitótica entre las cromátidas replicadas de cromosomas homólogos en la fase G2 del ciclo celular (Lefebvre y otros (2001) Nat. Genet. 27:257-258) (Figura 12). Aunque explicó solamente una minoría de los eventos homocigóticos, este mecanismo podría proporcionar un medio por el cual la escisión por ARNg/Cas9 puede usarse para promover la conversión a gran escala de heterocigótico a homocigótico para múltiples alelos en una porción grande de un cromosoma. La mayor parte de los eventos homocigóticos, sin embargo, parece haber sido el resultado de la conversión génica local cuyo mecanismo merece investigación adicional.

Evidencias adicionales de la conversión génica direccional de largo alcance fueron proporcionadas por el análisis de tres clones obtenidos después de la electroporación de células ES híbridas F1H4 (que se componen de 50 % de la cepa 129SvS6 y 50 % de la cepa C57BL/6N) con plásmidos que codifican los ARNg A y F de Lrp5, un plásmido que codifica Cas9, y un LTVEC que transforma el gen Ch25h a 30 Mb de distancia de Lrp5 en la dirección telomérica. Tres clones se calificaron inicialmente como de tipo silvestre después del tamizaje primario mediante el uso de ensayos TaqMan® dentro de la delección predicha entre los 2 ARNg (500 pb de distancia en el extremo 5' y 2 kb en el extremo 3'), pero ensayos de discriminación alélica TaqMan® posteriores que ensayan variantes de nucleótido simple (SNV) entre los alelos 129 y B6 revelaron sorprendentemente pérdida de heterocigosidad. Los ensayos de SNV usados fueron un ensayo centromérico (SNV 0,32) y dos ensayos teloméricos (SNV 1,2 y SNV 57,2) (ver la Figura 5). Como se muestra en la Tabla 8, el ensayo de SNV centromérico (0,32 Mb) confirmó la retención de la heterocigosidad en los tres clones. Sin embargo, ambos ensayos de SNV teloméricos mostraron que BP-E7 y BP-H4 eran homocigóticos para el alelo 129, y ambos ensayos de SNV teloméricos mostraron que BP-E6 era homocigótico para el alelo B6. Los tres clones mostraron retención de dos copias del cromosoma 19, y los tres clones fueron transgénicos para la transformación con LTVEC (es decir, el locus Ch25h se transformó). Estos resultados abren la posibilidad a la homocigosidad forzada mediante el uso de escisión dirigida por CRISPR/Cas9.

Tabla 8. Resultados del tamizaje de ensayos de discriminación alélica de SNV.

Clon	SNV 0,32	SNV 1,2	SNV 57,2
BP-E7	129 / B6	129 / 129	129 / 129
BP-H4	129 / B6	129 / 129	129 / 129
BP-E6	129 / B6	B6 / B6	B6 / B6

Varios posibles mecanismos pueden explicar los resultados observados en los experimentos de humanización con LTVEC asistida por CRISPR/Cas9 en células ES híbridas F1H4 de ratón (que se componen de 50 % de la cepa 129SvS6 y 50 % de la cepa C57BL/6N) (ver la Figura 16A-F). Tales mecanismos podrían ocurrir a través del intercambio recíproco de cromátidas por entrecruzamiento mitótico (ver la Figura 16A-C), o mediante copia de cromátidas por replicación inducida por ruptura (ver la Figura 16D-E). En cualquier caso, podría ocurrir una modificación heterocigótica en la que el cromosoma 129 o el cromosoma B6 se transforma con el LTVEC antes de la replicación del genoma (ver la Figura 16A y 16D).

Alternativamente, una sola cromátida de 129 o una sola cromátida de B6 podría transformarse con el LTVEC después de la replicación del genoma, seguido de conversión génica entre cromátidas (ver la Figura 16B y 16E). Alternativamente, puede existir falta de transformación con LTVEC en el locus genómico diana, pero puede ocurrir la escisión por Cas9 en el cromosoma 129 o B6 (ver la Figura 16C y 16F). Esta última posibilidad puede explicar los resultados observados con los clones BP-E7, BP-H4, y BP-E6. Los resultados potenciales se muestran en las Figuras 16A-F. Para la Figura 16F, también es posible observar pérdida de heterocigosidad (LOH) que retiene los alelos B6 si la Cas9 escinde una cromátida de 129. En los experimentos descritos anteriormente, se han observado eventos de pérdida de heterocigosidad que dan como resultado la transformación de ambos alelos (Hum/Hum) o que ambos alelos sean alelos de tipo silvestre (+/+).

Ejemplo 3. Efecto de los tamaños de los brazos de homología de LTVEC sobre la eficiencia de transformación

Para determinar el efecto del tamaño de los brazos de homología sobre la eficiencia de transformación, se compararon dos LTVEC diseñados para crear una delección de 76 kb del gen de ratón para el componente 5 del complemento (C5 o Hc (complemento hemolítico)) y un reemplazo simultáneo con un fragmento de 97 kb del gen C5 humano homólogo (Figura 13). El locus diana comprendía el exón 2 hasta el codón de terminación del gen C5 (Hc). El primer LTVEC comprendía el fragmento de 97 kb del gen C5 humano flanqueado por brazos de homología que contienen 35 kb y 31 kb de ADN genómico derivado de partes del locus C5 (Hc) de ratón que flanquean la secuencia de 76 kb del gen C5 (Hc) de ratón prevista para la delección (ver el vector de transformación etiquetado como LTVEC en la Figura 13). El segundo LTVEC comprendía el fragmento de 97 kb del gen C5 humano flanqueado por brazos de homología que contienen 5 kb cada uno de ADN genómico derivado de partes del locus C5 (Hc) de ratón que flanquean la secuencia de 76 kb del gen C5 (Hc) de ratón prevista para la delección (ver el vector de transformación etiquetado como sTVEC en la Figura 13).

En experimentos separados, los LTVEC de humanización de C5 (Hc) se combinaron con un plásmido que codifica Cas9 y un segundo plásmido que codifica uno o dos de seis ARNg (A, B, C, D, E, y E2; ver la Tabla 1) diseñados para crear rupturas bicatenarias dentro de la región del gen C5 (Hc) de ratón reconocida para la delección. Los ARNg se diseñaron de modo que eviten el reconocimiento de cualquier secuencia en la porción insertada del gen C5 humano.

Los resultados de la humanización asistida por CRISPR/Cas9 del gen C5(Hc) se muestran en la Tabla 9. La eficiencia de transformación del primer LTVEC solo (brazos de homología de 35 kb y 31 kb) fue mayor que la eficiencia de transformación del segundo LTVEC solo (brazos de homología de 5 kb y 5 kb). Sin embargo, las eficiencias de transformación totales de cada LTVEC cuando se combinó con los ARNg A y E2 fueron casi idénticas (ver la Tabla 9), lo que indica que los tamaños de los brazos de homología de 5 kb (es decir, suma total de 10 kb) son suficientes para facilitar el aumento de la eficiencia de transformación observada cuando se transforma el locus C5 (Hc) mediante el uso de CRISPR/Cas9 en combinación con la transformación con LTVEC.

Tabla 9. Resultados del tamizaje de la humanización asistida por CRISPR/Cas9 del gen C5 (Hc) mediante el uso de LTVEC con diferentes tamaños de los brazos de homología.

		Eficiencias de transformación (%) por tamizaje de LOA			
ARNg	LTVEC	Heterocigótico	Hemicigótico	Homocigótico	Total
A + E2	1	12,8	6,4	3,8	23,0
Ninguna	1	8,3	0	0	8,3
A+E2	2	7,2	13,1	4,3	24,6
Ninguna	2	5,4	0	0	5,4

Ejemplo 4. Efecto de distancias más cortas entre las secuencias de reconocimiento de ARN de CRISPR sobre la eficiencia de transformación

Para determinar el efecto de distancias más cortas entre las secuencias de reconocimiento de ARN de CRISPR y los sitios de escisión sobre la eficiencia de transformación, se diseñó un LTVEC para crear una delección de 18,2 kb del gen de ratón para la hidroxilasa de ácido citidina monofosfato-N-acetilneuramínico (Cmah) y un reemplazo simultáneo con un inserto que comprende un reportero lacZ y un casete de selección de resistencia a higromicina. El LTVEC se usó con dos ARNg que reconocen secuencias poco separadas (Figura 14). El locus diana comprendía los cinco primeros exones del gen Cmah. El LTVEC comprendía el inserto lacZ-hyg^r de 8,8 kb flanqueado por brazos de homología que contienen 120 kb y 57 kb de ADN genómico derivado de partes del locus Cmah de ratón que flanquean la secuencia de 18,2 kb del gen Cmah de ratón prevista para la delección. El LTVEC se combinó con plásmidos que codifican Cas9 y dos ARNg (A y B) diseñados para crear rupturas bicatenarias cerca del extremo 5' de la región del gen Cmah de ratón reconocida para la delección. Los dos ARNg reconocieron secuencias poco separadas cerca del ATG en el extremo 5' de la secuencia prevista para la delección, con los sitios de escisión a 27 pb de separación (ver la Figura 15). La escisión con Cas9 guiada por los dos ARNg produce una secuencia de 27 pb suprimida que tiene extremos romos. Se usó LTVEC solo como control.

Los clones de células ES resistentes a fármaco se tamizaron en cuanto a las humanizaciones dirigidas por ensayos de modificación de alelo (Valenzuela y otros (2003) Nat. Biotechnol. 21:652-659; Frenthewey y otros (2010) Methods Enzymol.

476:295-307) para secuencias dentro de la delección y para secuencias dentro del casete de selección por fármaco y el inserto de gen humano. Los clones se calificaron como transformados correctamente si habían perdido una de las dos secuencias del gen endógeno de ratón y ganado una copia del inserto lacZ-hyg^r. Además, se usaron ensayos de PCR en tiempo real que reconocen secuencias en los brazos de homología del LTVEC, referidos como ensayos de retención, para comprobar la transformación correcta del LTVEC en el genoma de ratón. La determinación del número de copias de estos ensayos de retención proporcionó más claridad para ayudar a distinguir los clones de ES transformados correctamente, que retuvieron un número de copias de dos, de los clones en los que una delección grande inducida por Cas9 del locus diana de ratón coincide con la integración aleatoria del LTVEC en otra parte en el genoma, en cuyo caso los ensayos de retención tuvieron un número de copias de tres (o más). La capacidad de los pares de ARNg de crear delecciones grandes mediadas por Cas9 en el locus diana de ratón significa que podría añadirse ensayos de retención a los ensayos LOA y GOA estándar como se describió previamente para proporcionar más claridad y comprobar la transformación correcta. Por lo tanto, los ensayos de retención se diseñaron y se usaron junto con los ensayos LOA y GOA.

Los resultados del experimento de transformación de Cmah se resumen en la Tabla 10. En el experimento de transformación de control con LTVEC solo, 5,4 % (3/56) de los clones tamizados tenían una mutación heterocigótica de delección y reemplazo (Het); 95 % de los clones se mantuvieron de tipo silvestre (WT) en el locus Cmah. En el experimento de transformación con CRISPR, se observaron cinco tipos de alelos mutantes diferentes además de algunos clones WT. Se observaron tres tipos de alelos con transformación por LTVEC: (1) Het; (2) Hom (delección y reemplazo homocigótico); y (3) Hemi (delección y reemplazo en un alelo y una mutación inducida por ARNg/Cas9 en el otro alelo). Estos tres tipos constituyen el 43,5 % (106/244) de todos los clones tamizados. En comparación con LTVEC solo, se observó un aumento en 8 veces de la transformación del gen Cmah en el que se transformó al menos un alelo. Se observaron además dos tipos de alelos que portaban solamente mutaciones indel por ARNg/Cas9: (1) Het, en el que se detectó un indel en uno de los dos alelos WT; y (2) mutaciones bialélicas de indel, que podían ser homocigóticas (Hom) o hemicigóticas (Hemi). Solo 3,7 % de los clones tamizados se mantuvieron WT sin mutación detectable en el locus Cmah. En general, más de 94 % de los clones tenían mutaciones inducidas por Cas9 cuando se usó la combinación de ARNg A y B.

Tabla 10. Resultados del tamizaje de la transformación de Cmah.

Transformación de control (LTVEC solo)				Transformación con CRISPR: LTVEC + Cas9 + ARNg A + B			
Mutación	Tipo de alelo	Genotipo	% de Clones tamizados	Mutación	Tipo de alelo	Genotipo	% de Clones tamizados
Delección y reemplazo	Het	lacZ/+	5,4	Delección y reemplazo	Het	lacZ/+	2,5
Delección y reemplazo	Hom	lacZ/lacZ	0	Delección y reemplazo	Hom	lacZ/lacZ	15
Delección y reemplazo/ Indel	Hemi	lacZ/indel	0	Delección y reemplazo/ Indel	Hemi	lacZ/indel	26
Indel	Het	indel/+	0	Indel	Het	indel/+	11
Indel	Hom o Hemi	indel/indel	0	Indel	Hom o Hemi	indel/indel	42
Ninguna	WT	+/+	95	Ninguna	WT	+/+	3,7

Ejemplo 5. Colapso grande mediante el uso de pares de ARNg en embriones en etapa de una célula

Para lograr una delección dirigida grande en embriones en etapa de una célula, se diseñó un experimento para crear una delección de 68 kb de la porción del gen Lrp5 (proteína 5 relacionada con el receptor de lipoproteínas de baja densidad) de ratón que codifica el ectodominio y opcionalmente un reemplazo simultáneo con una inserción de 4 nucleótidos a través del uso de una secuencia donante de ADN monocatenario (124 nucleótidos de longitud) con un inserto de 4 nucleótidos flanqueado por dos brazos de homología de 60 nucleótidos. El inserto de 4 nucleótidos creó un sitio de enzima de restricción tras la inserción en el locus diana. En experimentos separados, la proteína Cas9 en forma de proteína se suministró por inyección citoplasmática (CI), o Cas9 en forma de ARNm se suministró por inyección pronuclear (PNI) o electroporación (EP). La Cas9 se combinó con dos ARNg (A + F) diseñados para crear rupturas bicatenarias dentro de la región del gen Lrp5 de ratón reconocida para la delección, y opcionalmente con un donante de recombinación homóloga. Los ARNg se inyectaron en forma de ARN. Después se evaluó la frecuencia de mutaciones monoalélicas y bialélicas resultantes.

Los resultados se resumen en la Tabla 11, que incluye la delección mediada por NHEJ entre los sitios diana de los dos ARN guías o la delección asistida por reparación dirigida por homología con el donante de ADNmc. Las mutaciones bialélicas se observaron cuando los pares de ARN guías y Cas9 se introdujeron junto con el donante de ADNmc por medio de inyección citoplasmática. En cada mutación bialélica observada, un cromosoma se modificó por medio de delección mediada por NHEJ y un cromosoma se modificó por medio de delección asistida por HDR. Estos resultados indican que

las piezo inyecciones citoplasmáticas de ARNm dan como resultado la reparación dirigida por homología consistente con el potencial para la recombinación homocigótica con un donante.

Tabla 11. Comparación de métodos de suministro de Cas9.

Diana	Donante de HR	Suministro	Cas9	Concentración (mg) Cas9/ARNgs/Donante	Mutación monoalélica		Mutación bialélica	
					Colapso por NHEJ	Colapso por HDR	Colapso por NHEJ	Colapso por HDR
Lrp5 - Colapso de 68 kb (ARNgs A+F)	-	PNI	Proteína	50/60/0	7 %	N/A	0 %	N/A
		CI	ARNm	100/100/0	15 %	N/A	N/A	N/A
		EP	ARNm	250/150/0	3 %	N/A	0 %	N/A
Lrp5 - Colapso de 68 kb (ARNgs A+F)	+	PNI	N/A	N/A	N/A	N/A	N/A	N/A
		CI	ARNm	100/50/100	16,6 %	19,4 %	5,5 %	5,5 %
		EP	ARNm	250/200/250	1,4 %	2,1 %	0 %	0 %

Ejemplo 6. Ensayos de retención para distinguir entre inserciones dirigidas e inserciones transgénicas y entre deleciones dirigidas y deleciones que se extienden más allá de la región diana

Las estrategias de tamizaje de modificación de alelo (MOA) estándar (ver, por ejemplo, Figura 17A) determinan el número de copias TaqMan® mediante la comparación de un promedio de valores de Ct de cuatro réplicas biológicas para cada muestra con el Ct medio de todas las muestras. Para la pérdida de alelo, se usan sondas TaqMan® contra las regiones corriente arriba (mTU) y corriente abajo (mTD) de la región del locus genómico diana que se reconoce para la deleción. Para la ganancia de alelo, se usan sondas TaqMan® contra el casete de resistencia a neomicina. Sin embargo, tales sondas podrían diseñarse contra cualquier región del inserto de ácido nucleico. Para un clon diploide, con transformación heterocigótica, el número de copias TaqMan® para cada una de las sondas de mTU, mTD, y Neo debe ser uno. Para los clones diploides, con transformación homocigótica, el número de copias TaqMan® para cada uno de mTU y mTD debe ser cero, y el número de copias TaqMan® para Neo debe ser dos. Del mismo modo, para los clones diploides, no transformados, el número de copias TaqMan® para cada uno de mTU y mTD debe ser dos, y el número de copias TaqMan® para Neo debe ser cero. Para los clones diploides, con colapso heterocigótico, el número de copias TaqMan® para mTU y mTD debe ser uno, y el número de copias para Neo debe ser cero. Para los clones diploides, con colapso homocigótico, el número de copias TaqMan® para cada uno de mTU, mTD, y Neo debe ser cero.

Debido a que los pares de ARNg pueden crear deleciones grandes mediadas por Cas en un locus genómico diana, sin embargo, puede ser útil aumentar los ensayos LOA y GOA estándar para comprobar la transformación correcta por LTVEC. Por ejemplo, los ensayos LOA y GOA solos pueden no distinguir los clones de células transformados correctamente de los clones en los que una deleción grande inducida por Cas del locus genómico diana coincide con la integración aleatoria de un LTVEC en otra parte en el genoma. Debido a que la presión de selección en la célula transformada se basa en el casete de selección, la integración transgénica aleatoria del LTVEC en otra parte en el genoma generalmente incluirá el casete de selección y las regiones adyacentes del LTVEC pero excluirá regiones más distales del LTVEC. Por ejemplo, si los ensayos LOA y GOA se usan para evaluar la integración dirigida del LTVEC, y el ensayo GOA utiliza sondas contra el casete de selección, una deleción heterocigótica en el locus genómico diana combinada con una integración transgénica aleatoria del LTVEC proporcionará la misma lectura que una integración dirigida heterocigótica del LTVEC en el locus genómico diana. Para comprobar la transformación correcta por el LTVEC, pueden usarse ensayos de retención, solos o junto con ensayos LOA y/o GOA.

Cuando se usan ensayos de retención TaqMan®, se usan las sondas corriente arriba y corriente abajo correspondientes a la secuencia diana 5' del brazo de homología 5' (sonda retU) y la secuencia diana 3' del brazo de homología 3' (sonda retD) (ver la Figura 17B, que muestra el uso de ensayos de retención TaqMan® en combinación con ensayos GOA y LOA para tamizar la humanización asistida por CRISPR/Cas9 mediante el uso de selección con neomicina). La Figura 17B muestra además cómo pueden usarse las diferentes sondas dentro del inserto de ácido nucleico para ensayos GOA (ver sonda hTU corriente arriba y sonda hTD corriente abajo). Los resultados de los ensayos GOA, LOA, y de retención para diferentes tipos de modificaciones dirigidas e inserciones transgénicas se muestran en la Tabla 12.

Tabla 12. Lecturas de números de copias predichos de ensayos GOA, LOA, y de retención de los diferentes tipos de modificación.

Tipo de modificación	retU	mTGU	mTM	mTGD	retD	Neo
Con transformación homocigótica	2	0	0	0	2	2

	Con transformación heterocigótica	2	1	1	1	2	1
	Con transformación heterocigótica + Colapso heterocigótico	2	0	0	0	2	1
5	Colapso heterocigótico	2	1	1	1	2	0
	Colapso heterocigótico con inserción transgénica	2	1	1	1	3*	1*
	Colapso homocigótico	2	0	0	0	2	0
10	Colapso homocigótico con inserción transgénica	2	0	0	0	3*	1*

Los ensayos de retención TaqMan pueden usarse además en combinación con ensayos LOA para tamizar las deleciones asistidas por CRISPR/Cas9 mediante el uso de pares de ARNg (ver la Figura 17C). En tales ensayos, los números de copias de retU y retD deben mantenerse en dos en todos los casos. Los números de copias de menos de dos indican deleciones grandes mediadas por Cas9 que se extienden más allá de la región que se reconoce para la deleción. Los resultados de los ensayos LOA y de retención de los diferentes tipos de modificaciones relacionadas con colapso se muestran en la Tabla 13.

Tabla 13. Lecturas de los números de copias predichos de ensayos LOA y de retención para diferentes modificaciones de colapso.

Tipo de modificación	retU	mTGU	mTM	mTGD	retD
Colapso homocigótico	2	0	0	0	2
Colapso heterocigótico	2	1	1	1	2
Colapso heterocigótico con deleción corriente abajo adicional	2	1	1	1	1*

Ejemplo 7. Transformación mediada por CRISPR/Cas9 mediante el uso de cuatro ARN guías.

Para efectuar una deleción precisa, en una sola etapa de una región de aproximadamente 900 kb de un locus de cadena pesada de inmunoglobulina de ratón modificado y su reemplazo simultáneo con un inserto Pgk-Neo (promotor de fosfoglicerato quinasa I unido operativamente al gen de neomicina fosfotransferasa) flanqueado por sitios loxP, las siguientes moléculas de ácidos nucleicos se introdujeron por electroporación en células ES de ratón: (1) un LTVEC (2) un plásmido que codifica una endonucleasa Cas9; y (3) uno o más plásmidos que codifican cuatro ARN guías simples de CRISPR (ARNg). En cada experimento, el LTVEC se linealizó. El locus diana de la modificación era una región de aproximadamente 900 kb de un locus de cadena pesada de inmunoglobulina de ratón con segmentos génicos de región variable (V_H , D_H , J_H) reemplazados con equivalentes humanos (ver la Figura 18). El LTVEC comprendía el inserto Pgk-Neo que tiene una longitud de aproximadamente 2 kb flanqueado por un brazo de homología 5' de 19 kb y un brazo de homología 3' de 13 kb diseñado para dirigir un evento de recombinación homóloga que elimina la región de aproximadamente 900 kb del locus diana e inserta el casete de selección por fármaco que dirige la expresión de neomicina fosfotransferasa para impartir resistencia a G418.

De los cuatro ARNg que se usaron, dos se dirigían a la escisión por Cas9 cerca del extremo 5' del locus diana (5' ARNg_I y 5' ARNg_II en la Figura 18), y dos se dirigían a la escisión por Cas9 cerca del extremo 3' del locus diana (3' ARNg_I y 3' ARNg_II en la Figura 18). Las secuencias diana de 5' ARNg_I y 5' ARNg_II estaban aproximadamente 150 pb separadas entre sí, y las secuencias diana de 3' ARNg_I y 3' ARNg_II se solapaban, con el sitio diana de 3' ARNg_II desplazado 1 pb con relación al sitio diana de 3' ARNg_I.

Las células ES que captaron el LTVEC y lo incorporaron en sus genomas fueron capaces de crecer y formar colonias en una placa de cultivo de tejidos en un medio de crecimiento que contiene el fármaco antibiótico. Se seleccionaron colonias resistentes a fármaco y se tamizaron por el método de modificación de alelo (Valenzuela y otros (2003) Nat. Biotech. 21:652-660; Frenthewey y otros (2010) Methods Enzymol. 476:295-307) para identificar los clones que tenían el alelo humanizado transformado correctamente (ver la Tabla 14 a continuación). Además, los ensayos de PCR en tiempo real que reconocen secuencias en los brazos de homología del LTVEC, referidos como ensayos de retención, se usaron para comprobar la transformación correcta del LTVEC en el genoma de ratón (ver la Tabla 14 a continuación).

Tabla 14. Sondas usadas para confirmar la transformación con LTVEC y 4 ARNg.

Sonda	Ensayo	SEQ ID NO	Distancia a	Kb
5' IgH Brazo 1	Retención	148	5' ARNg	1,5
5' IgH Brazo 2	Retención	173	5' ARNg	0,315
hIgH31	LOA	136	5' ARNg	154

ES 2 731 437 T3

	hlgH1	LOA	142	5' ARNg	747
	hlgH1	LOA	142	3' ARNg	116
5	hlgH9	LOA	139	3' ARNg	3,2
	3' IgH Brazo 1	Retención	176	3' ARNg	0,484
	3' IgH Brazo 2	Retención	157	3' ARNg	1,2
10	mIgM-398	Retención	151	3' ARNg	6,4
	mIgM-1045	Retención	154	3' ARNg	7

15 En las células ES transformadas resultantes, la región de aproximadamente 900 kb se eliminó y se reemplazó con el inserto Pgk-Neo en ambos alelos (ver la Figura 18). Esta delección grande y su reemplazo se lograron con una eficiencia inesperadamente alta (aproximadamente 1,2 % de eficiencia para la delección bialélica).

Listado de secuencias

20 <110> Friendewey, David Lai, Ka-Man Venus Auerbach, Wojtek Lee, Jeffrey D. Mujica, Alexander O. Droguett, Gustavo Trzaska, Sean Hunt, Charleen Gagliardi, Anthony Valenzuela, David M. Voronina, Vera Macdonald, Lynn Murphy, Andrew J. Yancopoulos, George D.

25 <120> MÉTODOS Y COMPOSICIONES PARA LA MODIFICACIÓN GENÉTICA DIRIGIDA MEDIANTE EL USO DE PARES DE ARN GUÍAS

<130> 57766-472225

30 <150> US 62/083,005
<151> 2014-11-21

<150> US 62/182,314
<151> 2015-06-19

35 <150> US 62/211,421
<151> 2015-08-28

<160> 176

40 <170> FastSEQ para Windows Versión 4.0

<210> 1

<211> 80

<212> ARN

45 <213> Secuencia artificial

<220>

<223> ARNg

50 <400> 1

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60
ggcaccgagu cggugcuuuu 80

<210> 2

55 <211> 42

<212> ARN

<213> Secuencia artificial

<220>

60 <223> ARNg

<400> 2

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cg 42

65 <210> 3

<211> 30
 <212> ARN
 <213> Secuencia artificial

5 <220>
 <223> ARNcr

<400> 3
 guuuuagagc uagaaauagc aaguuaaaau 30

10 <210> 4
 <211> 33
 <212> ARN
 <213> Secuencia artificial

15 <220>
 <223> ARNcr

<400> 4
 guuuuagagc uagaaauagc aaguuaaaau aag 33

<210> 5
 <211> 26
 <212> ARN
 25 <213> Secuencia artificial

<220>
 <223> ARNcr

30 <400> 5
 gaguccgagc agaagaagaa guuuua 26

<210> 6
 <211> 12
 35 <212> ARN
 <213> Secuencia artificial

<220>
 <223> ARNtracr

40 <400> 6
 aaggcuaguc cg 12

45 <210> 7
 <211> 50
 <212> ARN
 <213> Secuencia artificial

50 <220>
 <223> ARNtracr

<400> 7
 aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 50

55 <210> 8
 <211> 23
 <212> ADN
 <213> Secuencia artificial

60 <220>
 <223> Secuencia de reconocimiento de ARN de CRISPR

<220>
 <221> misc_característica
 65 <222> (2)...(21)
 <223> n = a, t, c o g

<400> 8
 gnnnnnnnnn nnnnnnnnn ngg 23

5 <210> 9
 <211> 23
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Secuencia de reconocimiento de ARN de CRISPR

 <220>
 <221> misc_característica
 15 <222> (1)...(21)
 <223> n = a, t, c o g

 <400> 9
 nnnnnnnnnn nnnnnnnnn ngg 23

20 <210> 10
 <211> 25
 <212> ADN
 <213> Secuencia artificial

25 <220>
 <223> Secuencia de reconocimiento de ARN de CRISPR

 <220>
 <221> misc_característica
 30 <222> (3)...(23)
 <223> n = a, t, c o g

 <400> 10
 ggnnnnnnnn nnnnnnnnn nnngg 25

35 <210> 11
 <211> 20
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> Segmento de orientación al ADN del ARNg A de C5 (Hc)

 <400> 11
 45 atcacaacc agttaaccgg 20

 <210> 12
 <211> 20
 <212> ADN
 50 <213> Secuencia artificial

 <220>
 <223> Segmento de orientación al ADN del ARNg B de C5 (Hc)

 <400> 12
 55 ttccagacga gccgaccgg 20

 <210> 13
 <211> 20
 60 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Segmento de orientación al ADN del ARNg C de C5 (Hc)

65 <400> 13

ES 2 731 437 T3

```

    tgtgtgcat agcgatgctg      20
    <210> 14
    <211> 20
5    <212> ADN
    <213> Secuencia artificial

    <220>
    <223> Segmento de orientación al ADN del ARNg D de C5 (Hc)
10
    <400> 14
    aacaggtacc ctatcctcac      20

    <210> 15
    <211> 20
15    <212> ADN
    <213> Secuencia artificial

    <220>
    <223> Segmento de orientación al ADN del ARNg E de C5 (Hc)
20
    <400> 15
    ggcccgacc tagtctctct      20

25    <210> 16
    <211> 20
    <212> ADN
    <213> Secuencia artificial

30    <220>
    <223> Segmento de orientación al ADN del ARNg E2 de C5 (Hc)

    <400> 16
35    tcgtggtgc atgcgcactg      20

    <210> 17
    <211> 20
    <212> ADN
    <213> Secuencia artificial
40

    <220>
    <223> Segmento de orientación al ADN del ARNg A de Lrp5

    <400> 17
45    ggaaccac agcactacc      20

    <210> 18
    <211> 20
    <212> ADN
50    <213> Secuencia artificial

    <220>
    <223> Segmento de orientación al ADN del ARNg B de Lrp5

55    <400> 18
    gaatcatgca cggctacccc      20

    <210> 19
    <211> 20
60    <212> ADN
    <213> Secuencia artificial

    <220>
    <223> Segmento de orientación al ADN del ARNg B2 de Lrp5
65
    <400> 19

```

tgctcctatg gggaggcgcg 20
 <210> 20
 <211> 20
 5 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Segmento de orientación al ADN del ARNg C de Lrp5
 10 <400> 20
 actgagatca atgaccccga 20
 <210> 21
 <211> 20
 15 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> Segmento de orientación al ADN del ARNg D de Lrp5
 <400> 21
 gggtcgcccg gaacctctac 20
 25 <210> 22
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> Segmento de orientación al ADN del ARNg E2 de Lrp5
 <400> 22
 35 cttggataac attgataccc 20
 <210> 23
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> Segmento de orientación al ADN del ARNg E de Lrp5
 <400> 23
 45 ggggcagagc ccttatatca 20
 <210> 24
 <211> 20
 <212> ADN
 50 <213> Secuencia artificial
 <220>
 <223> Segmento de orientación al ADN del ARNg F de Lrp5
 55 <400> 24
 tcgctcacat taatccctag 20
 <210> 25
 <211> 20
 60 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Segmento de orientación al ADN del ARNg A de Ror1
 65 <400> 25

5 tgtgggcctt tgctgatcac 20
 <210> 26
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Segmento de orientación al ADN del ARNg B de Ror1
 10 <400> 26
 aatctatgat cctatggcct 20
 <210> 27
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> Segmento de orientación al ADN del ARNg D de Ror1
 <400> 27
 tgccaatagc agtgacttga 20
 <210> 28
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 30 <223> Segmento de orientación al ADN del ARNg C de Ror1
 <400> 28
 gggaagaatg ggctattgtc 20
 <210> 29
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 40 <223> Segmento de orientación al ADN del ARNg E de Ror1
 <400> 29
 ggtgtttgt gctgatgacg 20
 <210> 30
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Segmento de orientación al ADN del ARNg F de Ror1
 <400> 30
 ccgtcctagg ccttctacgt 20
 <210> 31
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Segmento de orientación al ADN del ARNg A de Trpa1
 65 <400> 31

gtactgggga atcggtggtc 20

5 <210> 32
 <211> 20
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Segmento de orientación al ADN del ARNg A2 de Trpa1

<400> 32
 cacgcactcc aaatttatcc 20

15 <210> 33
 <211> 20
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Segmento de orientación al ADN del ARNg B de Trpa1

<400> 33
 ctaagtgtgt atcagttacat 20

25 <210> 34
 <211> 20
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> Segmento de orientación al ADN del ARNg C de Trpa1

<400> 34
 tgccctgcac aataagcgca 20

35 <210> 35
 <211> 20
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> Segmento de orientación al ADN del ARNg D de Trpa1

45 <400> 35
 actcattgaa acgttatggc 20

50 <210> 36
 <211> 20
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Segmento de orientación al ADN del ARNg E2 de Trpa1

55 <400> 36
 agtaagggtg gattaaattc 20

60 <210> 37
 <211> 20
 <212> ADN
 <213> Secuencia artificial

65 <220>
 <223> Segmento de orientación al ADN del ARNg E de Trpa1

<400> 37

gccatctaga tcatgtaac 20

5 <210> 38
 <211> 20
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Segmento de orientación al ADN del ARNg F de Trpa1

<400> 38
 gactagaaat gttctgcacc 20

15 <210> 39
 <211> 21
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> cebador directo de 190045

<400> 39
 gagctcatag ccaacagctt g 21

25 <210> 40
 <211> 20
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> cebador directo de 190061

<400> 40
 atgcatcaga tcacgctcag 20

35 <210> 41
 <211> 20
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> cebador directo de 190068

45 <400> 41
 gtccttgg cattccaac 20

50 <210> 42
 <211> 24
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> cebador directo de 190030

55 <400> 42
 ccagtatggt gtcagtaat agcg 24

60 <210> 43
 <211> 19
 <212> ADN
 <213> Secuencia artificial

65 <220>
 <223> cebador directo de 190033

<400> 43

ctgtgcagaa agcagcctc 19
 <210> 44
 <211> 20
 5 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador directo de 190013
 10
 <400> 44
 cctctccctc taggcacctg 20
 <210> 45
 <211> 20
 15 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> cebador inverso de 190045
 <400> 45
 tctttaaggg ctccgtgtc 20
 25 <210> 46
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> cebador inverso de 190061
 <400> 46
 35 aagaccaacc attcaccag 20
 <210> 47
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> cebador inverso de 190068
 <400> 47
 45 ttcccagtc aagtcaaagg 20
 <210> 48
 <211> 20
 <212> ADN
 50 <213> Secuencia artificial
 <220>
 <223> cebador inverso de 190030
 55 <400> 48
 ctgttatctg caaggcacc 20
 <210> 49
 <211> 20
 60 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador inverso de 190033
 65 <400> 49

ES 2 731 437 T3

5 acaactggat cctgattcgc 20
 <210> 50
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador inverso de 190013
 10
 <400> 50
 taagagggca tgggtgagac 20
 <210> 51
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 15
 <220>
 <223> sonda C2 (B6)
 20
 <400> 51
 aattcagaag acctatcgta 20
 25
 <210> 52
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 30
 <220>
 <223> sonda T3 (B6)
 <400> 52
 tatgtgtata ggtgttggga t 21
 35
 <210> 53
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 40
 <220>
 <223> sonda T6 (B6)
 <400> 53
 45 tacattgcta aatgaaacc 19
 <210> 54
 <211> 16
 <212> ADN
 50 <213> Secuencia artificial
 <220>
 <223> sonda T7 (B6)
 55 <400> 54
 cgcagtcatg cacata 16
 <210> 55
 <211> 20
 60 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> sonda T8 (B6)
 65 <400> 55

ttataaagcc cagtatgtac 20

5 <210> 56
 <211> 14
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> sonda T9 (B6)

<400> 56
 tgctgcataa tcag 14

15 <210> 57
 <211> 18
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> sonda T10 (B6)

<400> 57
 tcaggagtga attggata 18

25 <210> 58
 <211> 16
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> sonda T11 (B6)

<400> 58
 ctgctactta cctttg 16

35 <210> 59
 <211> 13
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> sonda T13 (B6)

45 <400> 59
 aggaggaaaa cgc 13

<210> 60
 <211> 17
 <212> ADN
 50 <213> Secuencia artificial

<220>
 <223> sonda T14 (B6)

55 <400> 60
 cctttgttcc tcataag 17

<210> 61
 <211> 20
 60 <212> ADN
 <213> Secuencia artificial

<220>
 <223> sonda C2 (129)

65 <400> 61

ES 2 731 437 T3

aattcagaag acctattgta 20

<210> 62
 <211> 21
 5 <212> ADN
 <213> Secuencia artificial

<220>
 <223> sonda T3 (129)

10 <400> 62
 tatgtgtata ggtgttgca t 21

<210> 63
 <211> 16
 15 <212> ADN
 <213> Secuencia artificial

<220>
 20 <223> sonda T6 (129)

<400> 63
 cattgctaca tgaaac 16

25 <210> 64
 <211> 16
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> sonda T7 (129)

<400> 64
 cgcagtcacg cacgta 16

35 <210> 65
 <211> 23
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> sonda T8 (129)

<400> 65
 45 tgagaattta taaagcccaa tat 23

<210> 66
 <211> 14
 50 <212> ADN
 <213> Secuencia artificial

<220>
 <223> sonda T9 (129)

55 <400> 66
 tgctgcatga tcag 14

<210> 67
 <211> 15
 60 <212> ADN
 <213> Secuencia artificial

<220>
 <223> sonda T10 (129)

65 <400> 67

tcaggagtga atcgg 15

5 <210> 68
 <211> 16
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> sonda T11 (129)

<400> 68
 ctgctagtta cctttg 16

15 <210> 69
 <211> 15
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> sonda T13 (129)

<400> 69
 aggaggaaga cgcag 15

25 <210> 70
 <211> 17
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> sonda T14 (129)

<400> 70
 cttgttctt cataagc 17

35 <210> 71
 <211> 25
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> cebador directo C2

45 <400> 71
 atgagggatt tccttaatca gacaa 25

50 <210> 72
 <211> 29
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> cebador directo T3

55 <400> 72
 tggatggtt attcttactc aaggtttg 29

60 <210> 73
 <211> 22
 <212> ADN
 <213> Secuencia artificial

65 <220>
 <223> cebador directo T6

<400> 73

5 gggcaactga tggaaagaac tc 22
 <210> 74
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador directo T7
 10 <400> 74
 gactgacgca caaactgtc ctt 23
 <210> 75
 <211> 26
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador directo T8
 <400> 75
 20 cccaaagcat ataacaagaa caaatg 26
 <210> 76
 <211> 18
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador directo T9
 <400> 76
 30 gcaggacgca ggcgttta 18
 <210> 77
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador directo T10
 <400> 77
 40 gcatcctcat ggcagtctac atc 23
 <210> 78
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador directo T11
 <400> 78
 50 cctgcccctt gatgagtgtt 20
 <210> 79
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador directo T13
 65 <400> 79

ccctcttga tatgctcgtg tgt 23

5 <210> 80
 <211> 22
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> cebador directo T14

<400> 80
 tcccacaggt ccatgtcttt aa 22

15 <210> 81
 <211> 28
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> cebador inverso C2

<400> 81
 agactacaat gagctacat cataaggt 28

25 <210> 82
 <211> 24
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> cebador inverso T3

<400> 82
 caaccatcta aaactccagt tcca 24

35 <210> 83
 <211> 28
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> cebador inverso T6

45 <400> 83
 tgtgtaacag gacagttgaa tgtagaga 28

<210> 84
 <211> 19
 <212> ADN
 50 <213> Secuencia artificial

<220>
 <223> cebador inverso T7

55 <400> 84
 cttaaaacc gccctgcat 19

60 <210> 85
 <211> 26
 <212> ADN
 <213> Secuencia artificial

65 <220>
 <223> cebador inverso T8

<400> 85

ctacaggaga tgggctgtt ctatgt 26

5 <210> 86
 <211> 22
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> cebador inverso T9

<400> 86
 tcagcgtgat tcgctttagt tc 22

15 <210> 87
 <211> 26
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> cebador inverso T10

<400> 87
 tgcatagctg tttgaataat gacaag 26

25 <210> 88
 <211> 21
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> cebador inverso T11

<400> 88
 tgcatctct ctgtcaagca a 21

35 <210> 89
 <211> 24
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> cebador inverso T13

45 <400> 89
 gcaacaacat aaccacagc ataa 24

50 <210> 90
 <211> 23
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> cebador inverso T14

55 <400> 90
 gctaagcgtt tggaagaaat tcc 23

60 <210> 91
 <211> 20
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> cebador directo de SV 13,7

65 <400> 91

taggctctaa ggatgctggc 20

5 <210> 92
 <211> 20
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> cebador inverso de SV 13,7

<400> 92
 aagcagcttc aaaccctctg 20

15 <210> 93
 <211> 20
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> cebador directo de SV 20,0

<400> 93
 ttacttg gcc ttggaactgc 20

25 <210> 94
 <211> 21
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> cebador inverso de SV 20,0

<400> 94
 tgattcgtaa tcgtcactgc c 21

35 <210> 95
 <211> 20
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> cebador directo de SV 36,9

45 <400> 95
 tcctgtcccg agaaactgtc 20

50 <210> 96
 <211> 20
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> cebador inverso de SV 36,9

55 <400> 96
 agctggcttt cagagagctg 20

60 <210> 97
 <211> 20
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> cebador directo de SV 56,7

65 <400> 97

ttagaaagtg ccaaccaggc 20
 <210> 98
 <211> 20
 5 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador inverso de SV 56,7
 10
 <400> 98
 ctctggctag gaacaatggc 20
 <210> 99
 15 <211> 25
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> cebador m-lr-f
 <400> 99
 gtaggtgca gggctactc agctg 25
 25 <210> 100
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> cebador m-5'-f
 <400> 100
 35 ggaggagagg agaagcagcc 20
 <210> 101
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> cebador m-A
 <400> 101
 45 ggaggagagg agaagcagcc 20
 <210> 102
 <211> 26
 <212> ADN
 50 <213> Secuencia artificial
 <220>
 <223> cebador h-lr-r
 55 <400> 102
 gcaaacagcc ttctccac attogg 26
 <210> 103
 <211> 24
 60 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador m-5'-r
 65 <400> 103

ttgcttcag tagttcaggt gtgc 24
 <210> 104
 <211> 20
 5 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador h-5'-r
 10 <400> 104
 ggcgttgca ggaagtgcc 20
 <210> 105
 15 <211> 22
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> cebador m-F
 <400> 105
 tgaagtgag aggcacatga gg 22
 25 <210> 106
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> cebador m-E2
 <400> 106
 35 tagagtagcc acaggcagca aagc 24
 <210> 107
 <211> 21
 <212> ARN
 <213> Secuencia artificial
 40 <220>
 <223> Segmento de orientación al ADN del ARNg A de Cmah
 <400> 107
 45 gugacaggaa ggcuucucac c 21
 <210> 108
 <211> 21
 <212> ARN
 50 <213> Secuencia artificial
 <220>
 <223> Segmento de orientación al ADN del ARNg B de Cmah
 <400> 108
 55 gcuuacaagc aauuugcuga c 21
 <210> 109
 <211> 108
 60 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Locus Cmah
 65 <400> 109

ES 2 731 437 T3

ttgccagtga ccctgtttgc agttagagtt gacaggaagg ctctcaccg gggacatttt 60 aaatgacca gtcagcaaat tgcttgaag ttttggtt
 ctttcatt 108

5 <210> 110
 <211> 46
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Locus Cmah corriente arriba del corte de ARNg A

<400> 110
 ttgccagtga ccctgtttgc agttagagtt gacaggaagg ctctc 46

15 <210> 111
 <211> 35
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Locus Cmah corriente abajo del corte de ARNg B

<400> 111
 agcaaattgc ttgtaagttt tgggttctt tcatt 35

25 <210> 112
 <211> 27
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> Secuencia del locus Cmah suprimida por los ARNg A y B

<400> 112
 acccgggaca ttttaaatga cccagtc 27

35 <210> 113
 <211> 21
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> cebador m-5'-F de C5

<400> 113
 accccagcat ctgacgacac c 21

45 <210> 114
 <211> 20
 <212> ADN
 <213> Secuencia artificial

50 <220>
 <223> cebador m-5'-R de C5

<400> 114
 agaaagaccg cagtggaacc 20

55 <210> 115
 <211> 21
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> cebador h-5'-R de C5

65 <400> 115

tccccacatg cctagtagga g 21

5 <210> 116
 <211> 22
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> cebador directo TaqMan de Cmah

<400> 116
 gtgaccctgt ttgcagttag ag 22

15 <210> 117
 <211> 23
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> sonda TaqMan de Cmah

<400> 117
 acaggaaggc ttctcacccg gga 23

25 <210> 118
 <211> 24
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> cebador inverso TaqMan de Cmah

<400> 118
 tgaagtcagg aaactgttcc aatg 24

35 <210> 119
 <211> 22
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> cebador directo de 7064retU

45 <400> 119
 cctcctgagc tttcctttgc ag 22

50 <210> 120
 <211> 25
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> cebador inverso de 7064retU

55 <400> 120
 cctagacaac acagacactg tatca 25

60 <210> 121
 <211> 23
 <212> ADN
 <213> Secuencia artificial

65 <220>
 <223> sonda TaqMan de 7064retU

<400> 121

ttctgcctt gaaaaggaga ggc 23
 <210> 122
 <211> 18
 5 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador directo de 7064retD
 10 <400> 122
 cctctgaggc cacctgaa 18
 <210> 123
 15 <211> 23
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> cebador inverso de 7064retD
 <400> 123
 ccctgacaag ttctgccttc tac 23
 25 <210> 124
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> sonda TaqMan de 7064retD
 <400> 124
 35 tgccaagcc tctgcagctt t 21
 <210> 125
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> cebador directo de 7140retU
 <400> 125
 45 cccagcatct gacgacacc 19
 <210> 126
 <211> 21
 <212> ADN
 50 <213> Secuencia artificial
 <220>
 <223> cebador inverso de 7140retU
 55 <400> 126
 gaccactgtg ggcactgtgta g 21
 <210> 127
 <211> 27
 60 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> sonda TaqMan de 7140retU
 65 <400> 127

ccgagtctgc tgttactggt agcatca 27
 <210> 128
 <211> 20
 5 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador directo de 7140retD
 10 <400> 128
 cccgacacct tctgagcatg 20
 <210> 129
 15 <211> 21
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> cebador inverso de 7140retD
 <400> 129
 tgcaggctga gtcaggattt g 21
 25 <210> 130
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> sonda TaqMan de 7140retD
 <400> 130
 35 tagtcacggt ttgtgacacc ccaga 25
 <210> 131
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> cebador directo de LOA para mADAM6-2
 <400> 131
 45 agggctgagg gagaacatat ac 22
 <210> 132
 <211> 20
 <212> ADN
 50 <213> Secuencia artificial
 <220>
 <223> cebador inverso de LOA para mADAM6-2
 55 <400> 132
 aggcctgatg caggagctat 20
 <210> 133
 <211> 26
 60 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> sonda de LOA para mADAM6-2
 65 <400> 133

tcctctcagc tggattaaca gcatca 26

5 <210> 134
 <211> 20
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> cebador directo de LOA para hlgH31

<400> 134
 atcacactca tcccatcccc 20

15 <210> 135
 <211> 20
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> cebador inverso de LOA para hlgH31

<400> 135
 cacaggaag caggaactgc 20

25 <210> 136
 <211> 29
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> sonda de LOA para hlgH31

<400> 136
 cccttccta agtaccacag agtgggctc 29

35 <210> 137
 <211> 19
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> cebador directo de LOA para hlgH9

45 <400> 137
 tcctccaacg acaggtccc 19

<210> 138
 <211> 20
 <212> ADN
 50 <213> Secuencia artificial

<220>
 <223> cebador inverso de LOA para hlgH9

55 <400> 138
 gatgaactga cgggcacagg 20

60 <210> 139
 <211> 24
 <212> ADN
 <213> Secuencia artificial

65 <220>
 <223> sonda de LOA para hlgH9

<400> 139

tcctggaac tctgccccga caca 24
 <210> 140
 <211> 19
 5 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador directo de LOA para hlgH1
 10 <400> 140
 cagtcccgtt gatccagcc 19
 <210> 141
 15 <211> 21
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> cebador inverso de LOA para hlgH1
 <400> 141
 ggatatgcag cactgtgcc a c 21
 25 <210> 142
 <211> 30
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> sonda de LOA para hlgH1
 <400> 142
 35 cccatcaggg atttgtatc tctgtggacg 30
 <210> 143
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> cebador directo de GOA para Neo
 <400> 143
 45 ggtggagagg ctattcggc 19
 <210> 144
 <211> 17
 <212> ADN
 50 <213> Secuencia artificial
 <220>
 <223> cebador inverso de GOA para Neo
 55 <400> 144
 gaacacggcg gcatcag 17
 <210> 145
 <211> 23
 60 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> sonda de GOA para Neo
 65 <400> 145

tgggcacaac agacaatcgg ctg 23
 <210> 146
 <211> 19
 5 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador directo de ensayo de retención de 5' IgH Brazo1
 10 <400> 146
 aggatgctgg gaaacagac 19
 <210> 147
 15 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> cebador inverso de ensayo de retención de 5' IgH Brazo1
 <400> 147
 gaccactctc aggactctca 20
 25 <210> 148
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> sonda de ensayo de retención de 5' IgH Brazo1
 <400> 148
 35 tgggaagtcc caaaggaaac caca 24
 <210> 149
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> cebador directo de ensayo de retención de mIgM398
 <400> 149
 45 gagctcacac cttgacctt ca 22
 <210> 150
 <211> 22
 <212> ADN
 50 <213> Secuencia artificial
 <220>
 <223> cebador inverso de ensayo de retención de mIgM398
 55 <400> 150
 tgggtgggacg aacacattta ca 22
 <210> 151
 <211> 24
 60 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> sonda de ensayo de retención de mIgM398
 65 <400> 151

ES 2 731 437 T3

ccagctgtcg cagagatgaa cccc 24

5 <210> 152
<211> 22
<212> ADN
<213> Secuencia artificial

10 <220>
<223> cebador directo de ensayo de retención de mlgM1045

<400> 152
tcctccaca gacatcctaa cc 22

15 <210> 153
<211> 21
<212> ADN
<213> Secuencia artificial

20 <220>
<223> cebador inverso de ensayo de retención de mlgM1045

<400> 153
gttagcggac ttgctgagga a 21

25 <210> 154
<211> 22
<212> ADN
<213> Secuencia artificial

30 <220>
<223> sonda de ensayo de retención de mlgM1045

<400> 154
tcaccatccc cccctccttt gc 22

35 <210> 155
<211> 22
<212> ADN
<213> Secuencia artificial

40 <220>
<223> cebador directo de ensayo de retención de 3' IgH Brazo2

45 <400> 155
ggatcatgtgg caaggctatt tg 22

50 <210> 156
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> cebador inverso de ensayo de retención de 3' IgH Brazo2

55 <400> 156
agcctggact ttcggttgg t 21

60 <210> 157
<211> 30
<212> ADN
<213> Secuencia artificial

<220>
<223> sonda de ensayo de retención de 3' IgH Brazo2

65 <400> 157

ccactaggta aactgtagc tgtggttga 30

5 <210> 158
 <211> 17
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> cebador directo de mlgHp2 progenitor

<400> 158
 gccatgcaag gccaagc 17

15 <210> 159
 <211> 24
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> cebador inverso de mlgHp2 progenitor

<400> 159
 agttcttgag ccttaggtg ctag 24

25 <210> 160
 <211> 24
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> sonda de mlgHp2 progenitor

<400> 160
 ccaggaaaat gctgccagag cctg 24

35 <210> 161
 <211> 20
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> cebador directo de mlgKd2 progenitor

45 <400> 161
 gcaaacaaaa accactggcc 20

<210> 162
 <211> 19
 <212> ADN
 50 <213> Secuencia artificial

<220>
 <223> cebador inverso de mlgKd2 progenitor

55 <400> 162
 ggccacattc catgggttc 19

60 <210> 163
 <211> 37
 <212> ADN
 <213> Secuencia artificial

65 <220>
 <223> sonda de mlgKd2 progenitor

<400> 163

ctgttctct aaaactggac tccacagtaa atggaaa 37

5 <210> 164
 <211> 20
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> cebador directo de hlgK5 progenitor

<400> 164
 ccccgctctc ctccttttc 20

15 <210> 165
 <211> 19
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> cebador inverso de hlgK5 progenitor

<400> 165
 tgcaagtgtc gccagcaag 19

25 <210> 166
 <211> 32
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> sonda de hlgK5 progenitor

<400> 166
 tcatgtccat taaccattt acctttgccc ca 32

35 <210> 167
 <211> 21
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> Secuencia de orientación al ADN de 3' ARNg_I

45 <400> 167
 gactactacg gtatggacgt c 21

<210> 168
 <211> 21
 <212> ADN
 50 <213> Secuencia artificial

<220>
 <223> Secuencia de orientación al ADN de 3' ARNg_II

55 <400> 168
 gctactacgg tatggacgtc t 21

<210> 169
 <211> 20
 60 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Secuencia de orientación al ADN de 5' ARNg_I

65 <400> 169

ES 2 731 437 T3

gaagctgact agtttacgca 20

<210> 170
<211> 21
5 <212> ADN
<213> Secuencia artificial

<220>
<223> Secuencia de orientación al ADN de 5' ARNg_II

10 <400> 170
gtagcattct tacacctagc a 21

<210> 171
15 <211> 23
<212> ADN
<213> Secuencia artificial

<220>
20 <223> cebador directo de ensayo de retención de 5' IgH Brazo2

<400> 171
ctgagcatac tgctgcctaa cac 23

25 <210> 172
<211> 25
<212> ADN
<213> Secuencia artificial

30 <220>
<223> cebador inverso de ensayo de retención de 5' IgH Brazo2

<400> 172
35 gagcagtgc tttcttagtt aagga 25

<210> 173
<211> 26
<212> ADN
<213> Secuencia artificial

40 <220>
<223> sonda de ensayo de retención de 5' IgH Brazo2

<400> 173
45 tgaaatggca gttcttctcc agctgg 26

<210> 174
<211> 20
<212> ADN
50 <213> Secuencia artificial

<220>
<223> cebador directo de ensayo de retención de 3' IgH Brazo1

55 <400> 174
ggtggagtcc ctggatgatg 20

<210> 175
60 <211> 21
<212> ADN
<213> Secuencia artificial

<220>
65 <223> cebador inverso de ensayo de retención de 3' IgH Brazo1

<400> 175

ES 2 731 437 T3

atccctccag ccataggatt g 21

<210> 176

<211> 28

5 <212> ADN

<213> Secuencia artificial

<220>

10 <223> sonda de ensayo de retención de 3' IgH Brazo1

<400> 176

cttggaggc tcattgagg gagatgct 28

REIVINDICACIONES

1. Un método para hacer una modificación bialélica a un locus genómico diana en un genoma dentro de una célula, que comprende:
- 5 (I) introducir en una población de células:
- (a) una proteína Cas;
- (b) un primer ARN guía que se hibrida con una primera secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana;
- 10 (c) un segundo ARN guía que se hibrida con una segunda secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana; y
- (d) un vector de transformación que comprende un inserto de ácido nucleico flanqueado por un brazo de homología 5' que se hibrida con una secuencia diana 5' dentro del locus genómico diana y un brazo de homología 3' que se hibrida con una secuencia diana 3' dentro del locus genómico diana, en donde si la célula es un embrión en etapa de una célula el vector de transformación tiene una longitud de no más de 5 kb;
- 15 en donde el genoma comprende un par de primer y segundo cromosomas homólogos que comprenden el locus genómico diana, opcionalmente en donde la primera y segunda secuencias de reconocimiento de ARN de CRISPR flanquean la totalidad o parte de una secuencia codificante de un gen; y
- en donde la proteína Cas escinde al menos una de la primera y segunda secuencias de reconocimiento de ARN de CRISPR para generar al menos una ruptura bicatenaria en cada uno del primer y segundo cromosomas homólogos, opcionalmente en donde la introducción del primer y segundo ARN guías da como resultado el aumento de la eficiencia de modificación bialélica en comparación con la introducción del primer ARN guía o el segundo ARN guía solos; y
- 20 (II) identificar una célula que comprende un locus genómico diana modificado que comprende una delección y/o una inserción, en donde la identificación comprende realizar un ensayo cuantitativo de modificación de alelo y un ensayo de retención, en donde el ensayo de modificación de alelo comprende:
- (a) un ensayo de ganancia de alelo para determinar el número de copias de una plantilla de ADN de inserto de ácido nucleico en una muestra de ADN genómico de la célula; y/o
- (b) un ensayo de pérdida de alelo para determinar el número de copias en la muestra de ADN genómico de una plantilla de ADN dentro de una región del locus genómico diana reconocida para la delección, y
- 30 en donde el ensayo de retención determina el número de copias en la muestra de ADN genómico de una plantilla de ADN de la secuencia diana 5' que se pretende retener en el locus genómico diana modificado y/o una plantilla de ADN de la secuencia diana 3' que se pretende retener en el locus genómico diana modificado,
- en donde la célula no se produce mediante el uso de un proceso que involucra la modificación de la identidad genética de la línea germinal de seres humanos o que involucra el uso de un embrión humano para fines industriales o comerciales, y
- 35 en donde el método no es un método para el tratamiento del cuerpo humano o animal mediante terapia o cirugía.
2. El método de la reivindicación 1, en donde el ensayo de retención determina el número de copias de la plantilla de ADN de la secuencia diana 5' en la muestra de ADN genómico y la plantilla de ADN de la secuencia diana 3' en la muestra de ADN genómico, y opcionalmente en donde el ensayo de modificación de alelo comprende el ensayo de pérdida de alelo y el ensayo de ganancia de alelo.
- 40
3. El método de la reivindicación 2, en donde el inserto de ácido nucleico comprende un casete de selección adyacente a un primer brazo de homología que se hibrida con una primera secuencia diana,
- 45 en donde el primer brazo de homología es el brazo de homología 5' y la primera secuencia diana es la secuencia diana 5', o en donde el primer brazo de homología es el brazo de homología 3' y la primera secuencia diana es la secuencia diana 3',
- en donde la identificación comprende:
- (a) obtener ADN de la célula;
- (b) exponer el ADN de la célula a una sonda que se une dentro de la primera secuencia diana, una sonda que se une dentro del inserto de ácido nucleico, y una sonda que se une dentro de un gen de referencia que tiene un número de copias conocido, en donde cada sonda genera una señal detectable tras la unión;
- (c) detectar las señales de la unión de cada una de las sondas; y
- (d) comparar la señal de la sonda del gen de referencia con la señal de la sonda de la primera secuencia diana para determinar un número de copias de la primera secuencia diana, y comparar la señal de la sonda del gen de referencia con la señal de la sonda del inserto de ácido nucleico para determinar un número de copias del inserto de ácido nucleico,
- 50 en donde un número de copias del inserto de ácido nucleico de uno o dos y un número de copias de la primera secuencia diana de dos indica la inserción dirigida del inserto de ácido nucleico en el locus genómico diana, y
- 55 en donde un número de copias del inserto de ácido nucleico de uno o más y un número de copias de la primera secuencia diana de tres o más indica una inserción aleatoria del inserto de ácido nucleico en un locus genómico distinto al locus genómico diana.
- 60
4. El método de cualquier reivindicación precedente, en donde:

- (a) la proteína Cas escinde la primera y segunda secuencias de reconocimiento de ARN de CRISPR en cada uno del primer y segundo cromosomas homólogos para generar al menos dos rupturas bicatenarias en cada uno del primer y segundo cromosomas homólogos; o
- 5 (b) la proteína Cas escinde la primera y segunda secuencias de reconocimiento de ARN de CRISPR en al menos uno del primer y segundo cromosomas homólogos para generar al menos dos rupturas bicatenarias en al menos uno del primer y segundo cromosomas homólogos.
5. El método de cualquier reivindicación precedente, en donde la etapa (I) comprende además introducir en la población de células:
- 10 (e) un tercer ARN guía que se hibrida con una tercera secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana; y
- (f) un cuarto ARN guía que se hibrida con una cuarta secuencia de reconocimiento de ARN de CRISPR dentro del locus genómico diana.
- 15 6. El método de la reivindicación 5, en donde:
- (a) la primera secuencia de reconocimiento de ARN de CRISPR y la tercera secuencia de reconocimiento de ARN de CRISPR se encuentran separadas por aproximadamente 25 pb a aproximadamente 1 kb; y/o
- (b) la segunda secuencia de reconocimiento de ARN de CRISPR y la cuarta secuencia de reconocimiento de ARN de CRISPR se encuentran separadas por aproximadamente 25 pb a aproximadamente 1 kb; y/o
- 20 (c) la primera y tercera secuencias de reconocimiento de ARN de CRISPR son un primer par de secuencias de reconocimiento de ARN de CRISPR, y la segunda y cuarta secuencias de reconocimiento de ARN de CRISPR son un segundo par de secuencias de reconocimiento de ARN de CRISPR, en donde el primer par y el segundo par se encuentran separados por aproximadamente 25 pb a aproximadamente 100 Mb, opcionalmente en donde la proteína Cas escinde al menos dos de la primera, segunda, tercera y cuarta secuencias de reconocimiento de ARN de CRISPR para generar al menos dos rupturas bicatenarias en al menos uno del primer y segundo cromosomas homólogos, opcionalmente en donde la proteína Cas escinde al menos dos de la primera, segunda, tercera y cuarta secuencias de reconocimiento de ARN de CRISPR para generar al menos dos rupturas bicatenarias en el primer y segundo cromosomas homólogos.
- 25 7. El método de cualquier reivindicación precedente, en donde el inserto de ácido nucleico se inserta entre las secuencias diana 5' y 3'.
8. El método de cualquier reivindicación precedente, en donde la célula es diploide, y la modificación bialélica da como resultado homocigosidad, heterocigosidad combinada, o hemicigosidad en el locus genómico diana.
- 35 9. El método de cualquier reivindicación precedente, en donde la modificación bialélica comprende una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo, opcionalmente en donde:
- 40 (a) la modificación bialélica comprende la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer y segundo cromosomas homólogos; opcionalmente en donde la modificación bialélica comprende además la inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer y segundo cromosomas homólogos;
- 45 (b) la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer y segundo cromosomas homólogos, y la inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer cromosoma homólogo pero no en el segundo cromosoma homólogo;
- (c) la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo, y la interrupción del locus genómico diana en el segundo cromosoma homólogo, en donde la interrupción es de reparación mediada por unión de extremos no homólogos (NHEJ) de la al menos una ruptura bicatenaria;
- 50 (d) la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo, una inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer cromosoma homólogo, y la interrupción del locus genómico diana en el segundo cromosoma homólogo, en donde la interrupción es de reparación mediada por unión de extremos no homólogos (NHEJ) de la al menos una ruptura bicatenaria;
- 55 (e) la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo, y una inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer cromosoma homólogo, en donde la secuencia del inserto de ácido nucleico es homóloga u ortóloga a la secuencia eliminada;
- 60 (f) el ácido nucleico eliminado es de aproximadamente 5 kb a aproximadamente 3 Mb; o
- (g) el ácido nucleico eliminado es de al menos 20 kb.
10. El método de cualquier reivindicación precedente, en donde:
- 65 (a) la primera y segunda secuencias de reconocimiento de ARN de CRISPR se ubican cada una al menos 50 pb de ambas secuencias diana 5' y 3';
- (b) la primera y segunda secuencias de reconocimiento de ARN de CRISPR se ubican cada una entre aproximadamente 50 pb a aproximadamente 100 kb de ambas secuencias diana 5' y 3';

- (c) la primera y segunda secuencias de reconocimiento de ARN de CRISPR se encuentran separadas por aproximadamente 1 kb a aproximadamente 3 Mb;
- (d) la primera y segunda secuencias de reconocimiento de ARN de CRISPR se encuentran separadas por al menos 1 kb;
- 5 (e) la primera y segunda secuencias de reconocimiento de ARN de CRISPR se encuentran separadas por aproximadamente 25 pb a aproximadamente 1 kb; o
- (f) la primera y segunda secuencias de reconocimiento de ARN de CRISPR se encuentran separadas por menos de 25 pb.
- 10 11. El método de cualquier reivindicación precedente, en donde el vector de transformación está en forma lineal, y/o en donde el vector de transformación es monocatenario o bicatenario.
12. El método de cualquier reivindicación precedente, en donde la célula es una célula eucariota, opcionalmente en donde la célula eucariota es una célula de mamífero, una célula humana, una célula no humana, una célula de roedor, una célula de ratón, una célula de rata, una célula pluripotente, una célula no pluripotente, una célula pluripotente no humana, una célula pluripotente humana, una célula pluripotente de roedor, una célula pluripotente de ratón, una célula pluripotente de rata, una célula madre embrionaria (ES) de ratón, una célula ES de rata, una célula ES humana, una célula madre adulta humana, una célula progenitora humana restringida por el desarrollo, una célula madre pluripotente inducida (iPS) humana, o un embrión no humano en etapa de una célula, o un embrión de ratón en etapa de una célula.
13. El método de la reivindicación 12, en donde la célula es un embrión no humano en etapa de una célula, y en donde:
- (a) el vector de transformación es de entre aproximadamente 50 nucleótidos a aproximadamente 5 kb de longitud;
- (b) el vector de transformación es ADN monocatenario y es de entre aproximadamente 60 a aproximadamente 200 nucleótidos de longitud;
- (c) la proteína Cas, el primer ARN guía, y el segundo ARN guía se introducen cada uno en el embrión en etapa de una célula por microinyección; o
- (d) la proteína Cas, el primer ARN guía, y el segundo ARN guía se introducen cada uno en el embrión en etapa de una célula en forma de ARN por microinyección en el citoplasma.
14. El método de la reivindicación 12, en donde la célula no es un embrión en etapa de una célula, y en donde:
- (a) el vector de transformación es un vector de transformación grande (LTVEC) que es de al menos 10 kb, opcionalmente en donde el LTVEC es de aproximadamente 50 kb a aproximadamente 300 kb; o
- (b) el vector de transformación es un vector de transformación grande (LTVEC), en donde la suma total de los brazos de homología 5' y 3' del LTVEC es de al menos 10 kb, opcionalmente en donde la suma total de los brazos de homología 5' y 3' del LTVEC es de aproximadamente 10 kb a aproximadamente 200 kb.
15. El método de cualquiera de las reivindicaciones 1-12 y 14, en donde la célula no es un embrión en etapa de una célula,
- en donde el vector de transformación es un vector de transformación grande (LTVEC), en donde los brazos de homología 5' y 3' tienen una suma total de al menos 10 kb;
- en donde la primera y segunda secuencias de reconocimiento de ARN de CRISPR se ubican cada una más de 200 pb de ambas secuencias diana 5' y 3';
- en donde la proteína Cas escinde la primera y segunda secuencias de reconocimiento de ARN de CRISPR en al menos uno del primer y segundo cromosomas homólogos para generar al menos dos rupturas bicatenarias en al menos uno del primer y segundo cromosomas homólogos; y
- en donde la modificación bialélica comprende la delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo y una inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer cromosoma homólogo, en donde la secuencia del inserto de ácido nucleico es homóloga u ortóloga a la secuencia eliminada.
16. El método de cualquier reivindicación precedente, en donde la proteína Cas es Cas9, opcionalmente en donde la proteína Cas tiene actividad nucleasa en ambas cadenas de ADN bicatenario, u opcionalmente en donde la proteína Cas es una nickasa, y en donde el método comprende además introducir en la célula:
- (e) un tercer ARN guía que se hibrida con una tercera secuencia de reconocimiento de ARN de CRISPR; y
- (f) un cuarto ARN guía que se hibrida con una cuarta secuencia de reconocimiento de ARN de CRISPR;
- en donde la proteína Cas forma una mella en una primera cadena de ADN genómico dentro de la primera y segunda secuencias de reconocimiento de ARN de CRISPR, y la proteína Cas forma una mella en una segunda cadena de ADN genómico dentro de la tercera y cuarta secuencias de reconocimiento de ARN de CRISPR, en donde las mellas dentro de la primera y tercera secuencias de reconocimiento de ARN de CRISPR son mellas desplazadas para crear una primera ruptura bicatenaria que tiene extremos escalonados, y las mellas dentro de la segunda y cuarta secuencias de reconocimiento de ARN de CRISPR son mellas desplazadas para crear una segunda ruptura bicatenaria que tiene extremos escalonados.
17. El método de cualquier reivindicación precedente, en donde:

- (a) el primer ARN guía comprende un primer ARN de CRISPR y un primer ARNtracr fusionados entre sí, y/o el segundo ARN guía comprende un segundo ARN de CRISPR y un segundo ARNtracr fusionados entre sí;
- (b) el primer ARN guía comprende un primer ARN de CRISPR y un primer ARNtracr, en donde el primer ARN de CRISPR y el primer ARNtracr son moléculas de ARN separadas, y/o el segundo ARN guía comprende un segundo ARN de CRISPR y un segundo ARNtracr, en donde el segundo ARN de CRISPR y el segundo ARNtracr son moléculas de ARN separadas;
- (c) la proteína Cas se introduce en la célula en forma de una proteína, un ARN mensajero (ARNm) que codifica la proteína Cas, o un ADN que codifica la proteína Cas;
- (d) el primer ARN guía se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el primer ARN guía;
- (e) el segundo ARN guía se introduce en la célula en forma de un ARN o en forma de un ADN que codifica el segundo ARN guía;
- (f) la proteína Cas y el primer ARN guía se introducen en la célula como un primer complejo de ARN y proteína, y/o la proteína Cas y el segundo ARN guía se introducen en la célula como un segundo complejo de ARN y proteína;
- (g) la proteína Cas, el primer ARN guía, y el segundo ARN guía se introducen cada uno en la célula en forma de ADN; o
- (h) la proteína Cas, el primer ARN guía, y el segundo ARN guía se introducen cada uno en la célula en forma de ARN.
18. El método de cualquier reivindicación precedente, en donde la célula se ha modificado para disminuir la unión de extremos no homólogos (NHEJ) y/o para aumentar la conversión génica o la reparación dirigida por homología (HDR), opcionalmente en donde la célula se ha modificado para disminuir la expresión y/o la actividad de uno o más de ADN-PK, PARP1, y ligasa IV, y opcionalmente en donde la disminución de la expresión o la actividad es inducible, reversible, temporalmente específica, y/o espacialmente específica.
19. El método de cualquier reivindicación precedente, en donde el método produce:
- (a) una primera subpoblación de células que comprenden una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer y segundo cromosomas homólogos, y la inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer cromosoma homólogo pero no en el segundo cromosoma homólogo;
- (b) una segunda subpoblación de células que comprenden una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR y la inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer y segundo cromosomas homólogos;
- (c) una tercera subpoblación de células que comprenden una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer y segundo cromosomas homólogos pero no la inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3';
- (d) una cuarta subpoblación de células que comprenden una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo, y la interrupción del locus genómico diana en el segundo cromosoma homólogo, en donde la interrupción es de reparación mediada por unión de extremos no homólogos (NHEJ) de la al menos una ruptura bicatenaria; o
- (e) una quinta subpoblación de células que comprenden una delección entre la primera y segunda secuencias de reconocimiento de ARN de CRISPR en el primer cromosoma homólogo, una inserción del inserto de ácido nucleico entre las secuencias diana 5' y 3' en el primer cromosoma homólogo, y una interrupción del locus genómico diana en el segundo cromosoma homólogo, en donde la interrupción es de reparación mediada por unión de extremos no homólogos (NHEJ) de la al menos una ruptura bicatenaria,
- opcionalmente en donde el método produce la primera subpoblación de células, la segunda subpoblación de células, la tercera subpoblación de células, la cuarta subpoblación de células, y la quinta subpoblación de células.
20. El método de cualquier reivindicación precedente, en donde la célula es una célula madre embrionaria (ES) de ratón o rata, y el método comprende además introducir la célula ES de ratón o rata en un embrión huésped.

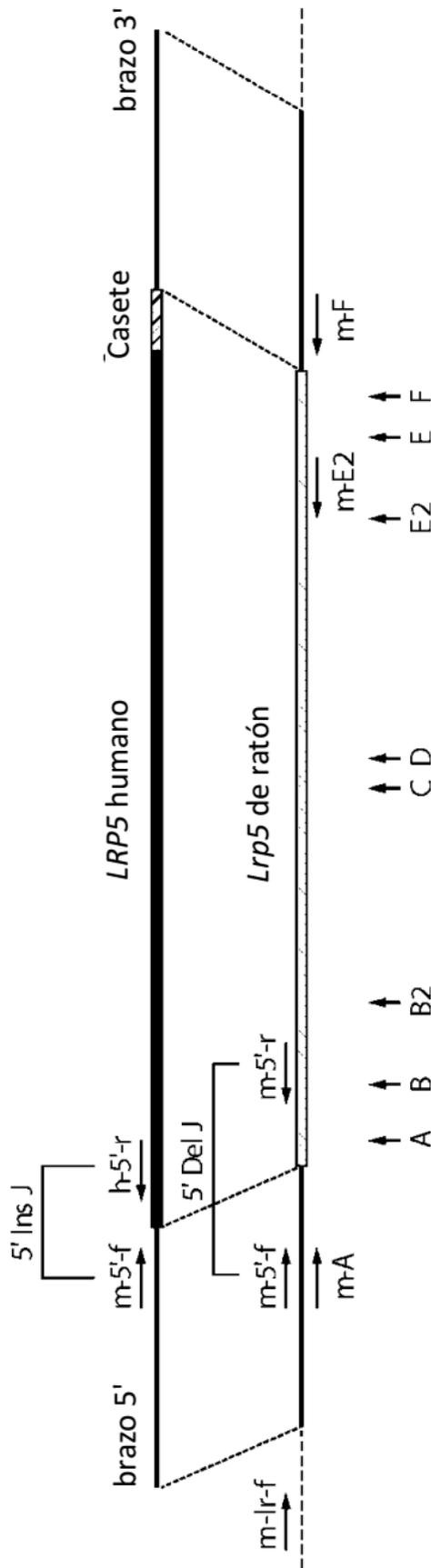


FIGURA 1

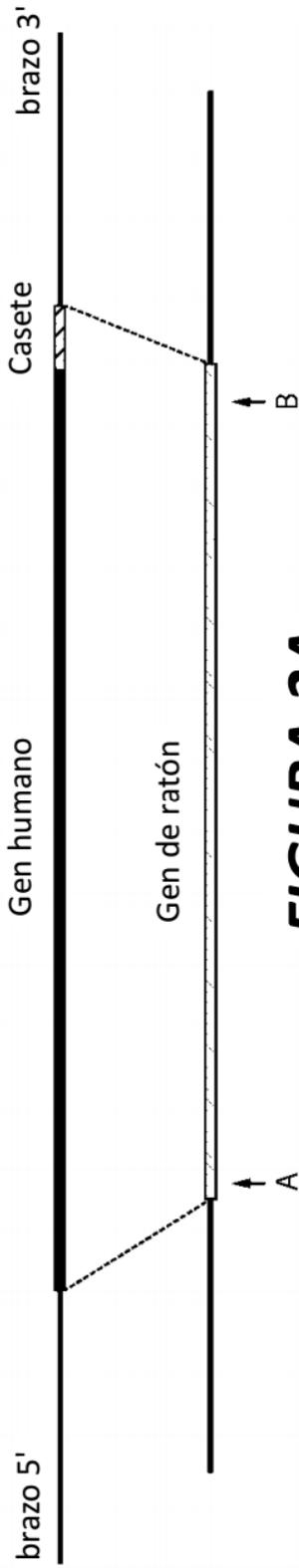


FIGURA 2A



FIGURA 2B



FIGURA 2C



FIGURA 2D



FIGURA 2E

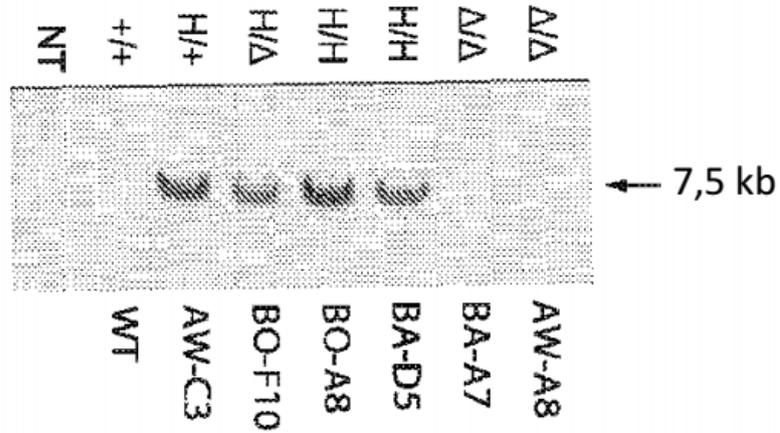


FIGURA 3A

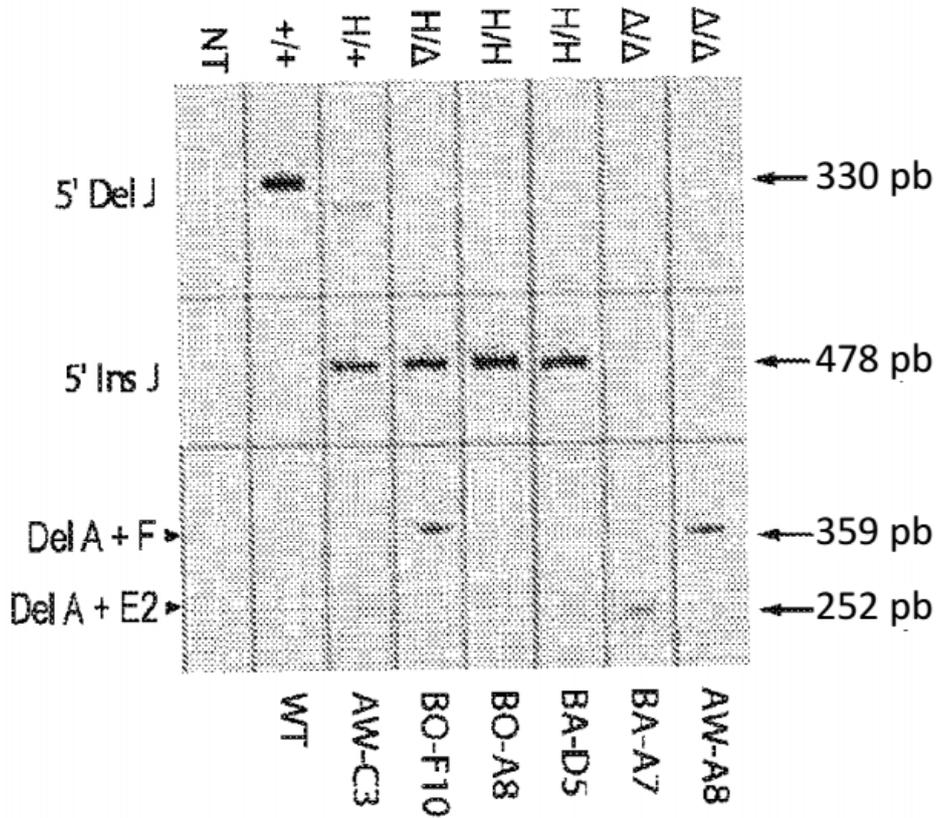


FIGURA 3B

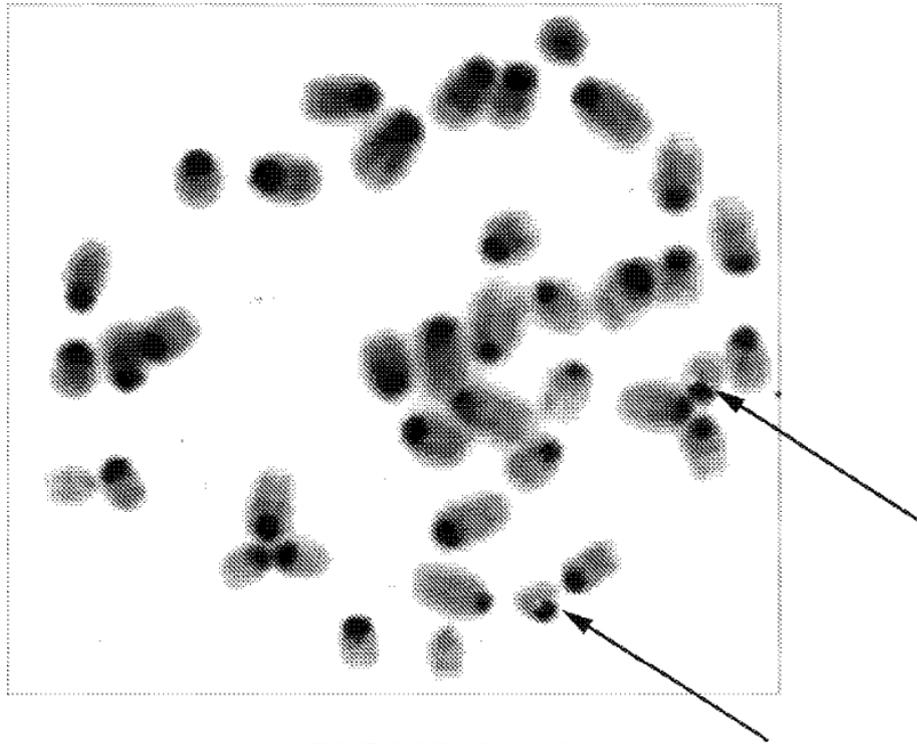


FIGURA 4A

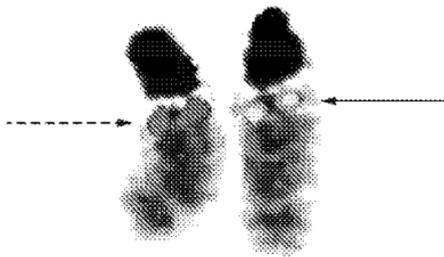


FIGURA 4B



FIGURA 4C

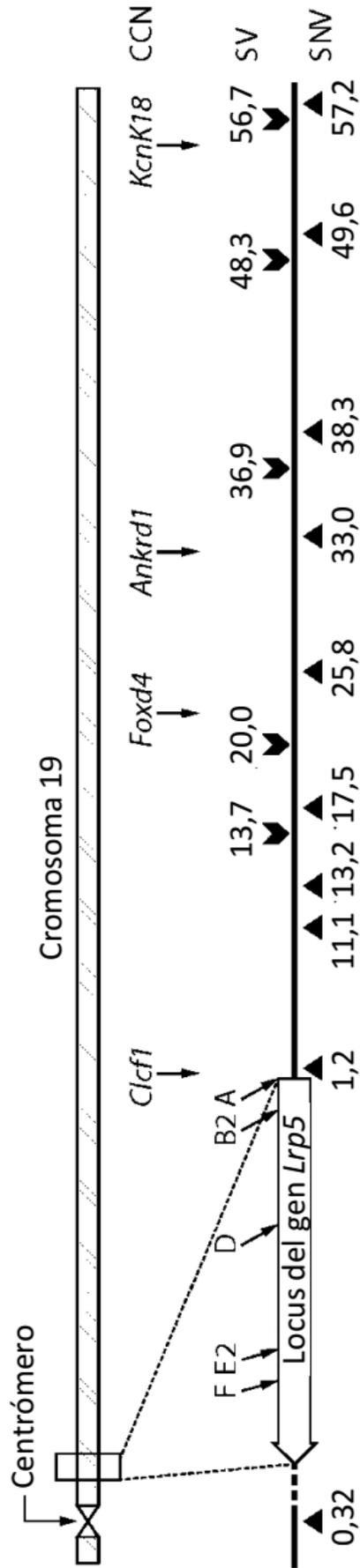


FIGURA 5

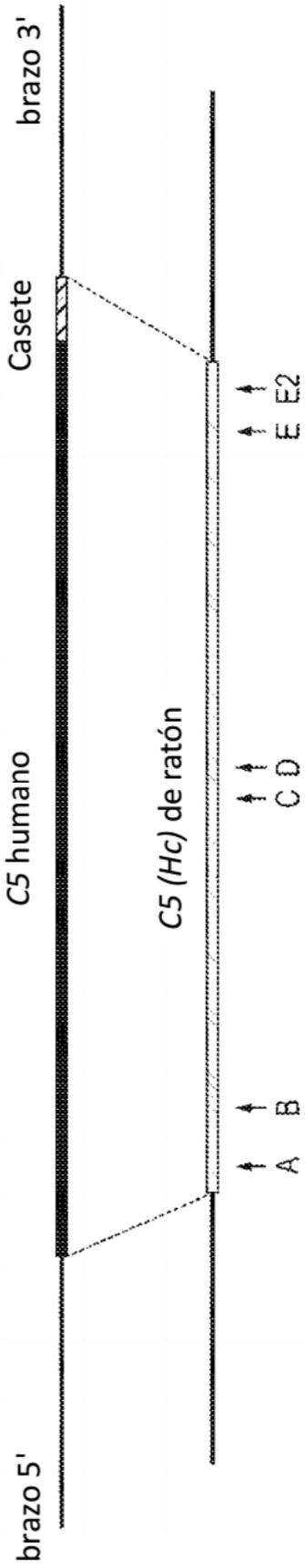


FIGURA 6

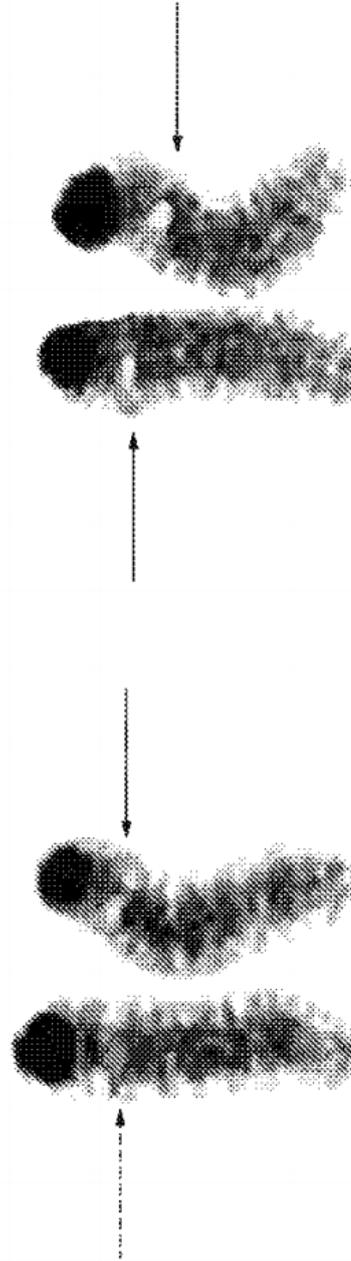


FIGURA 7A

FIGURA 7B

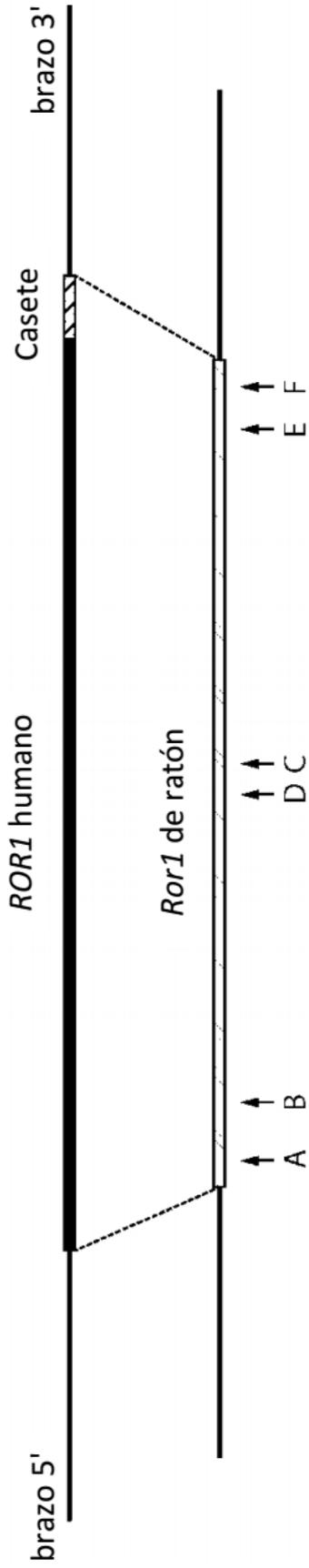


FIGURA 8

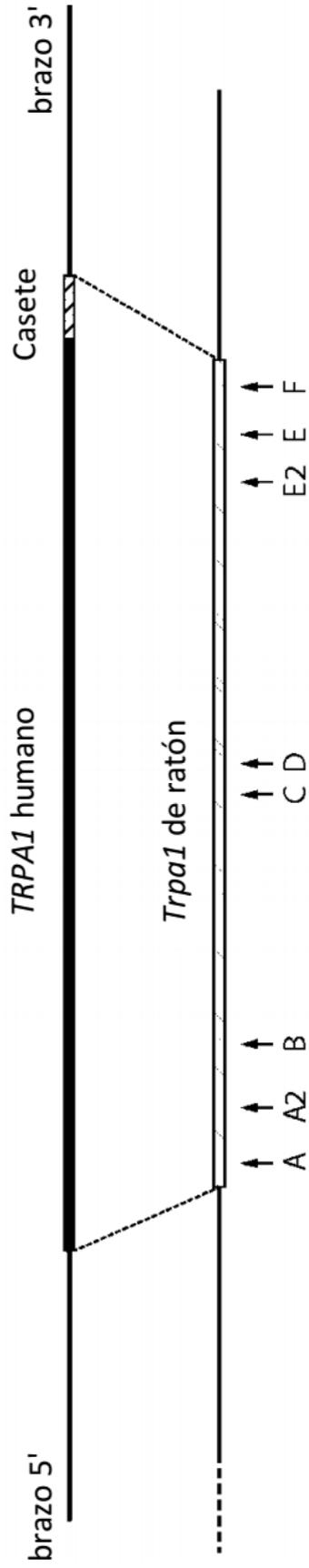
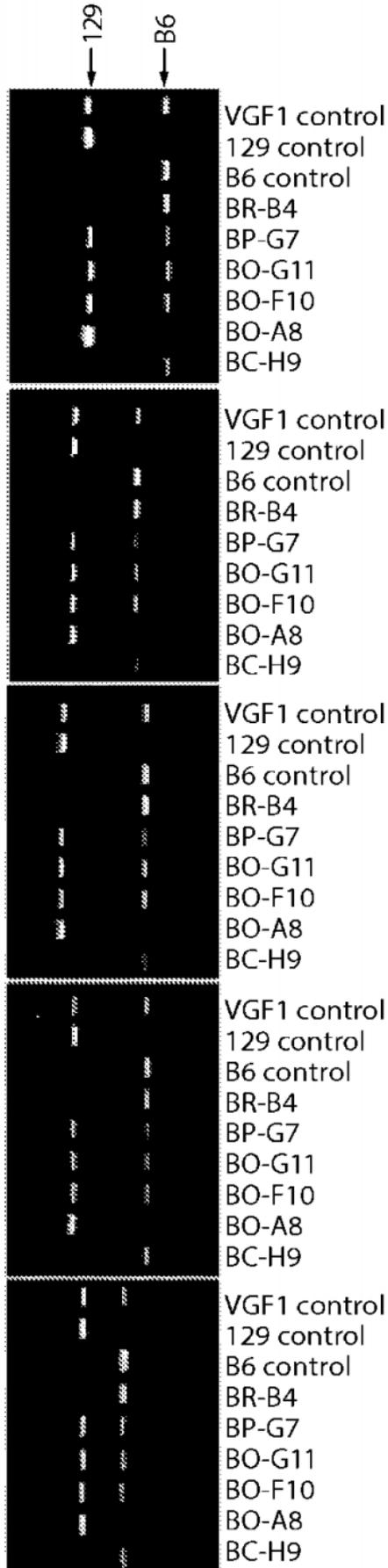


FIGURA 9



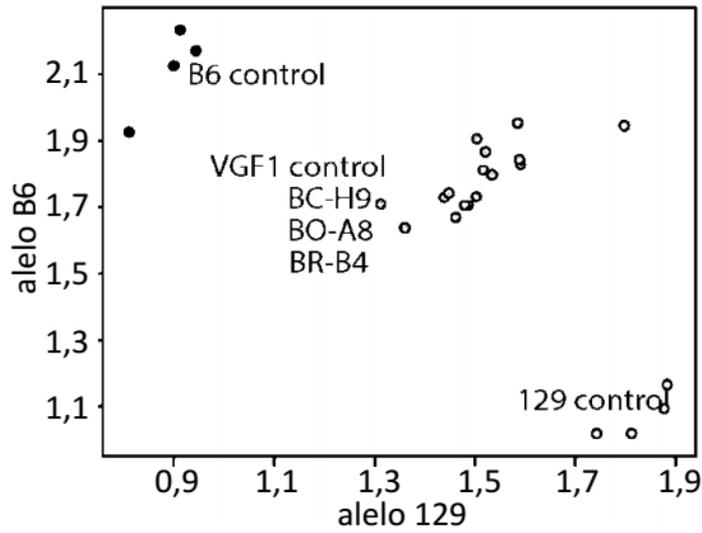


FIGURA 11A

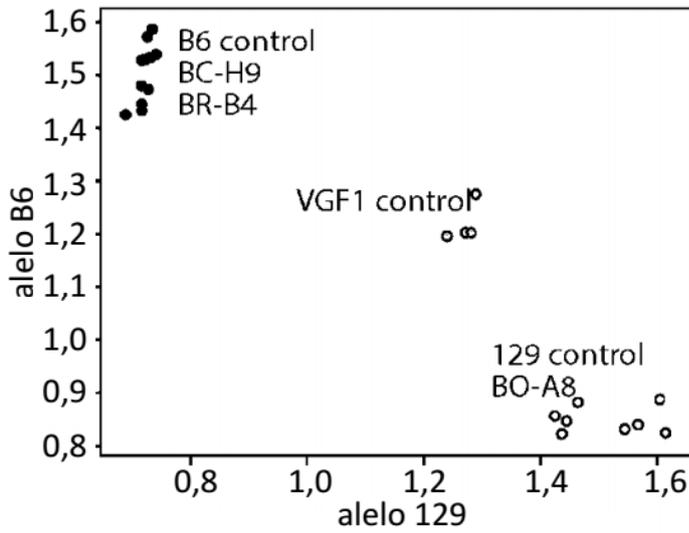


FIGURA 11B

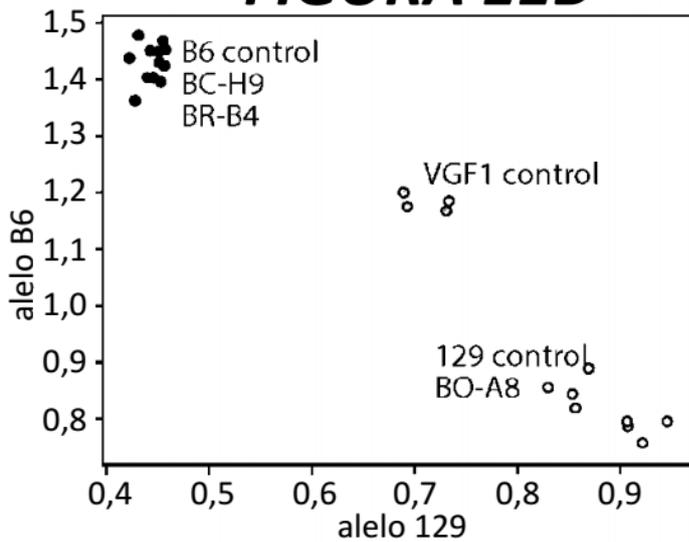


FIGURA 11C

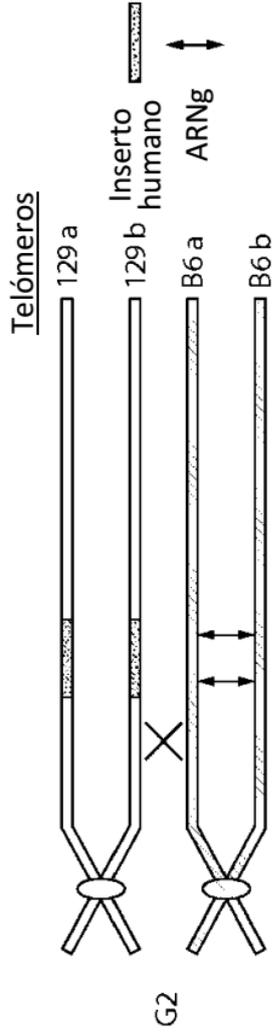


FIGURA 12A

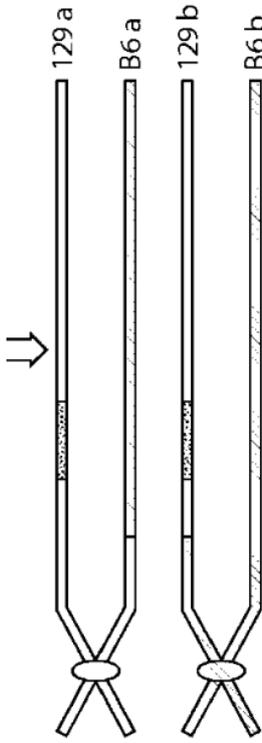


FIGURA 12B

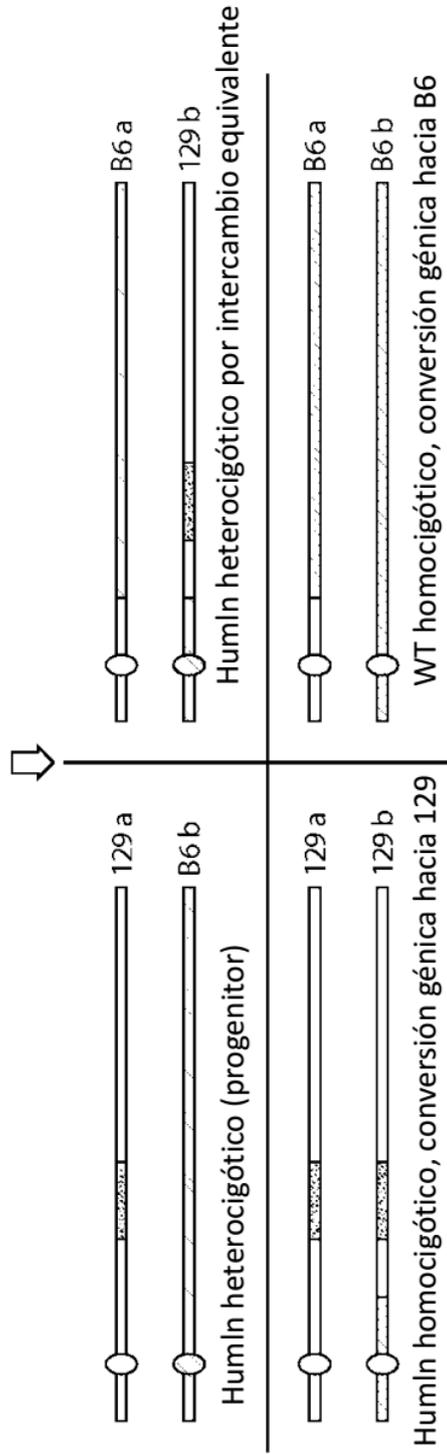


FIGURA 12C

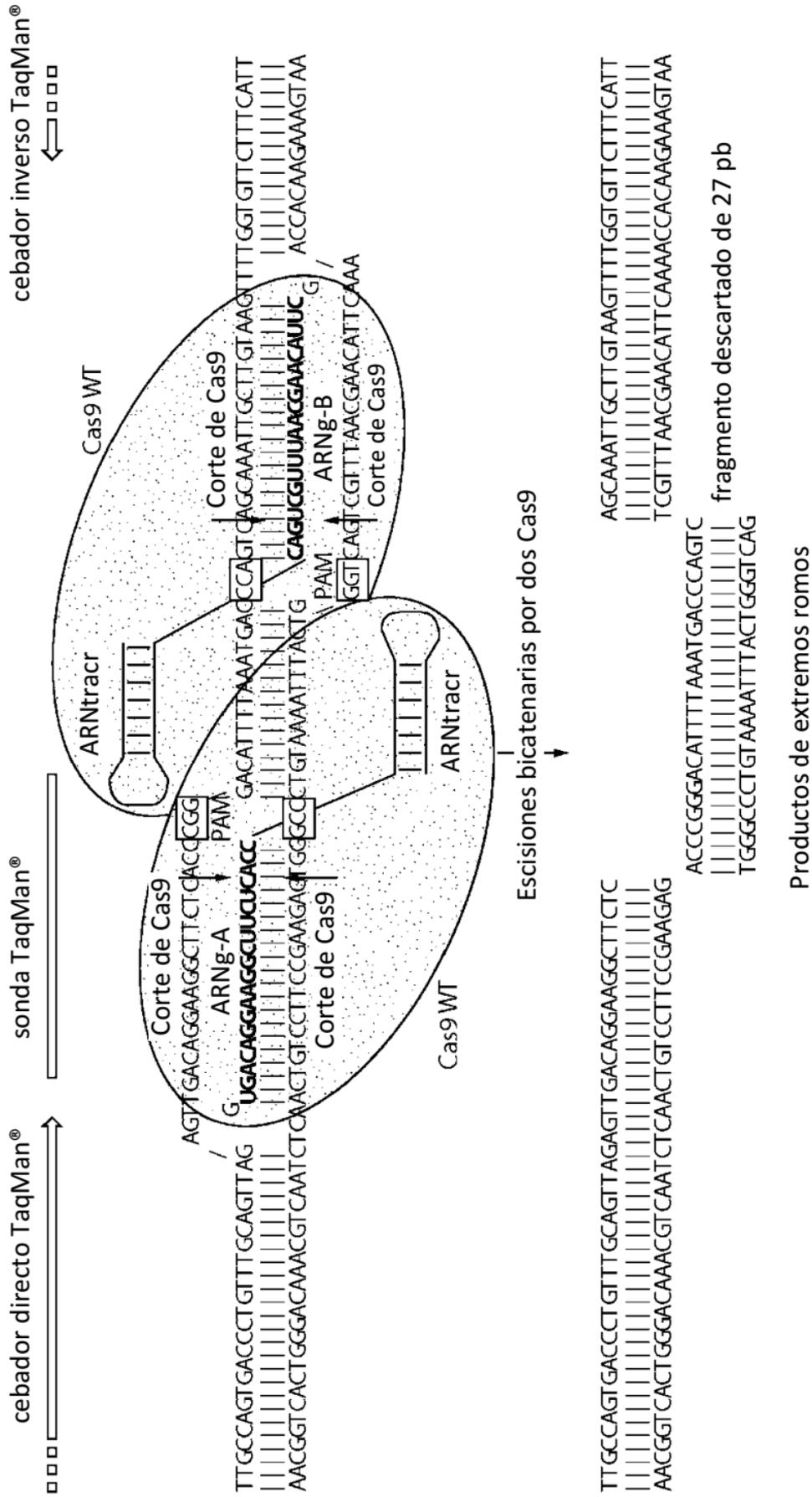


FIGURA 15

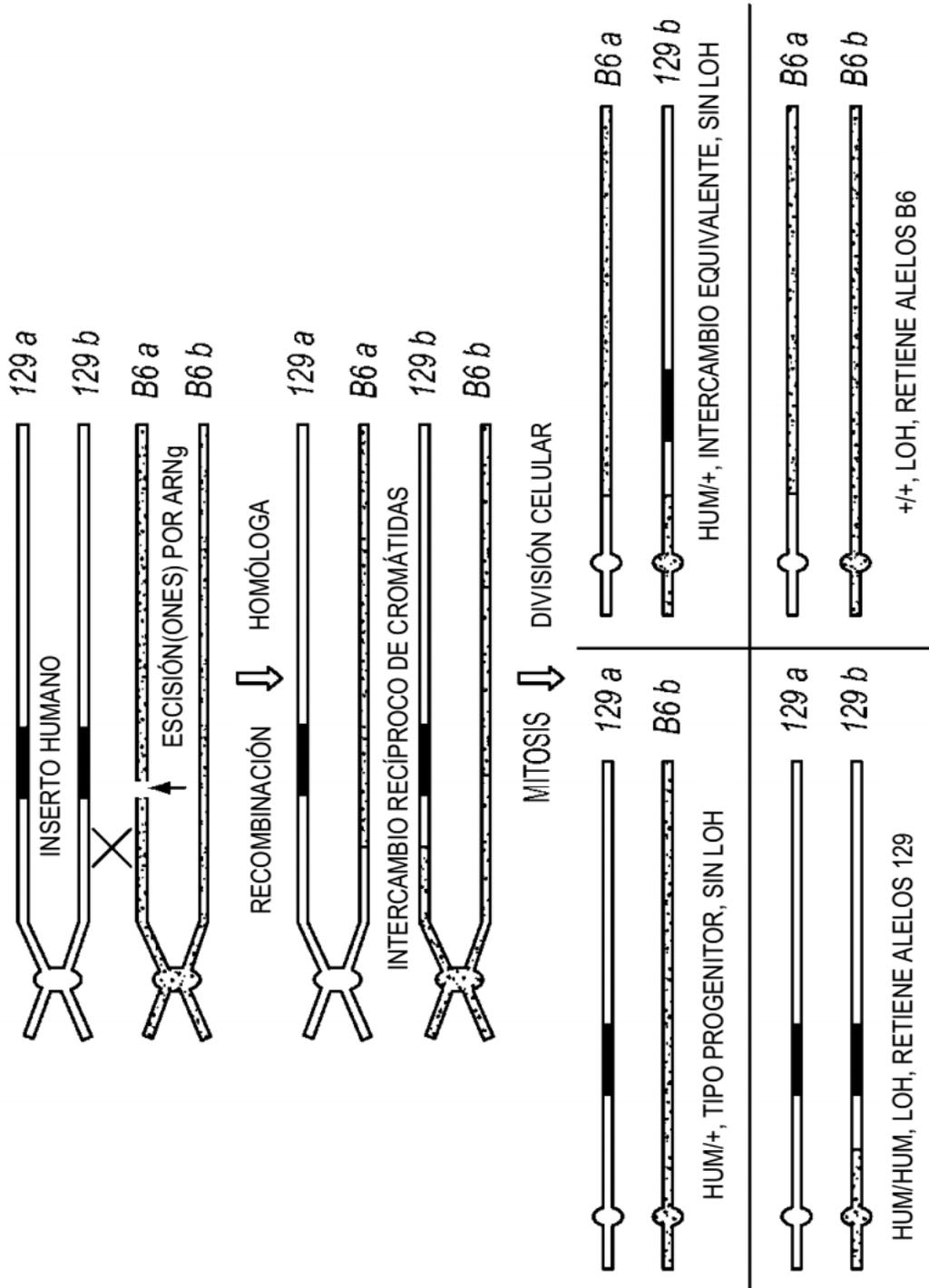


FIGURA 16A

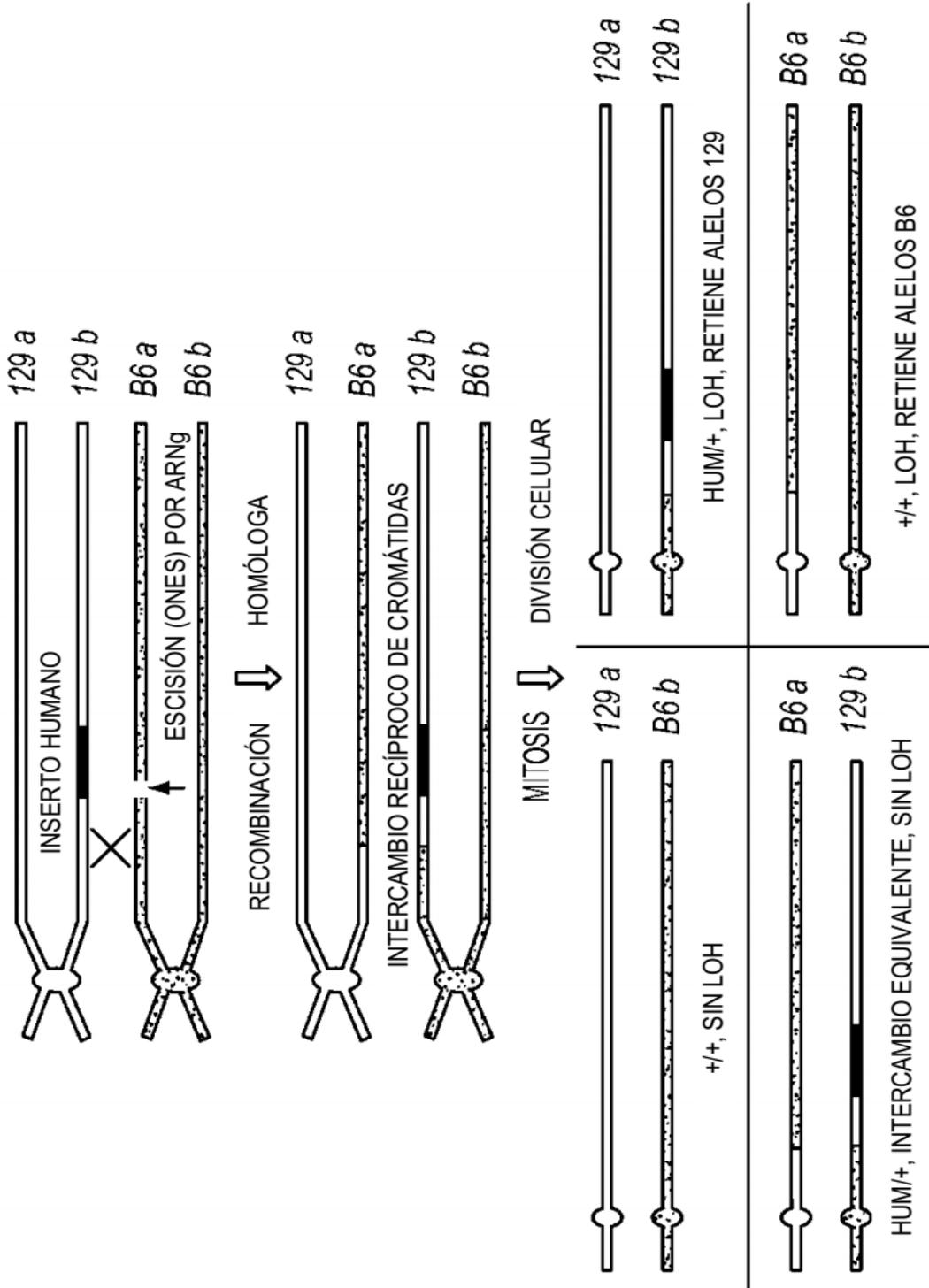


FIGURA 16B

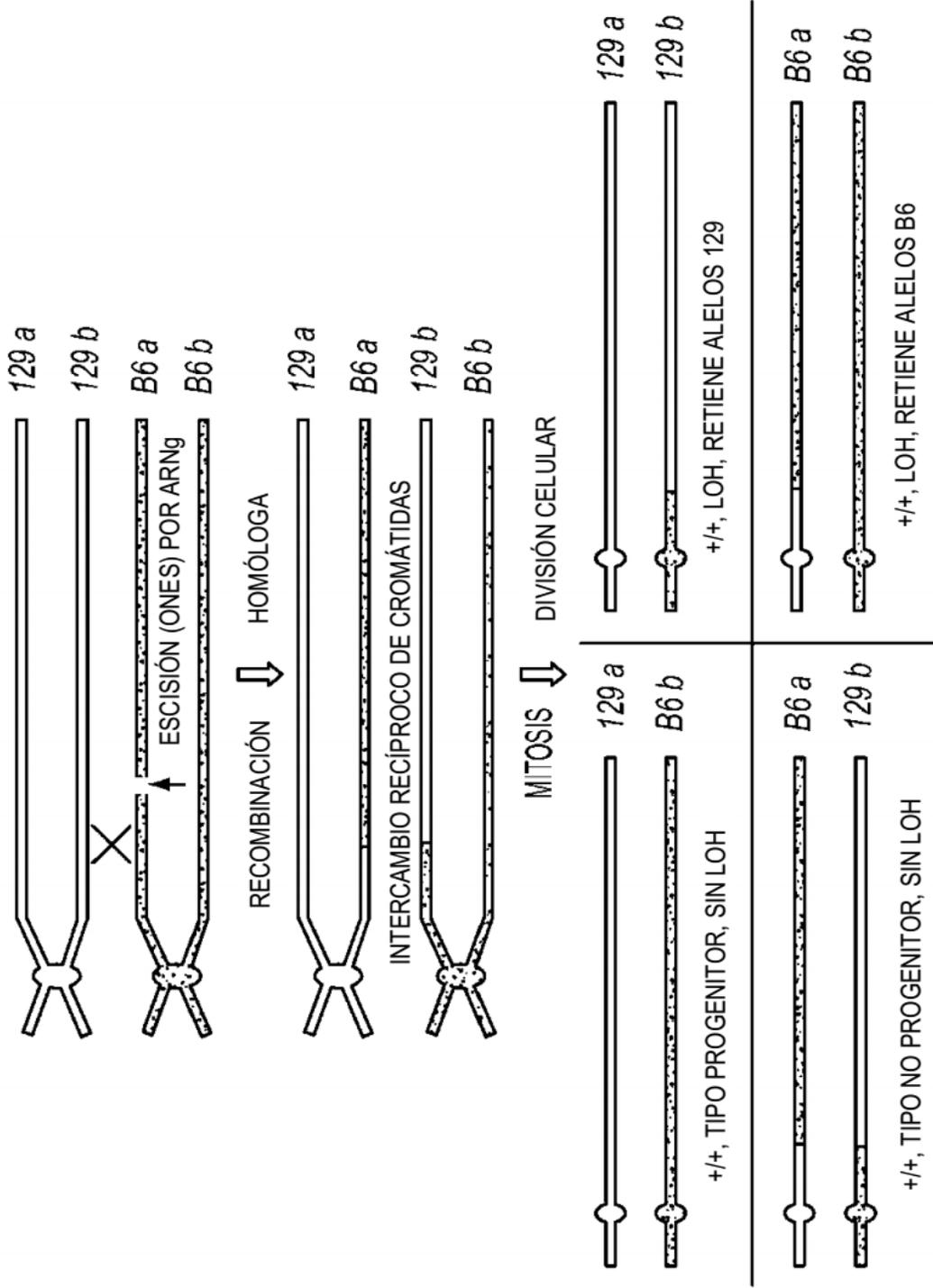


FIGURA 16C

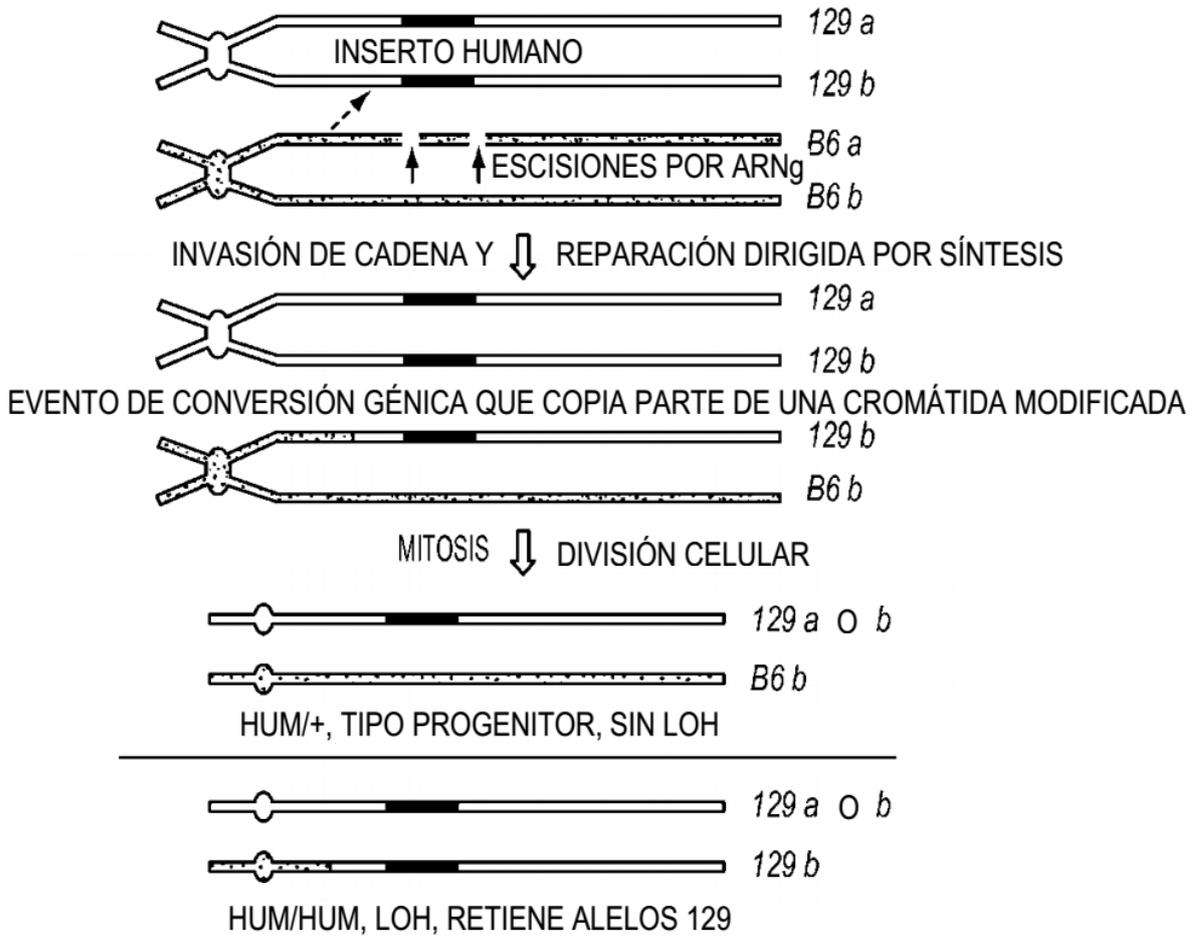


FIGURA 16D

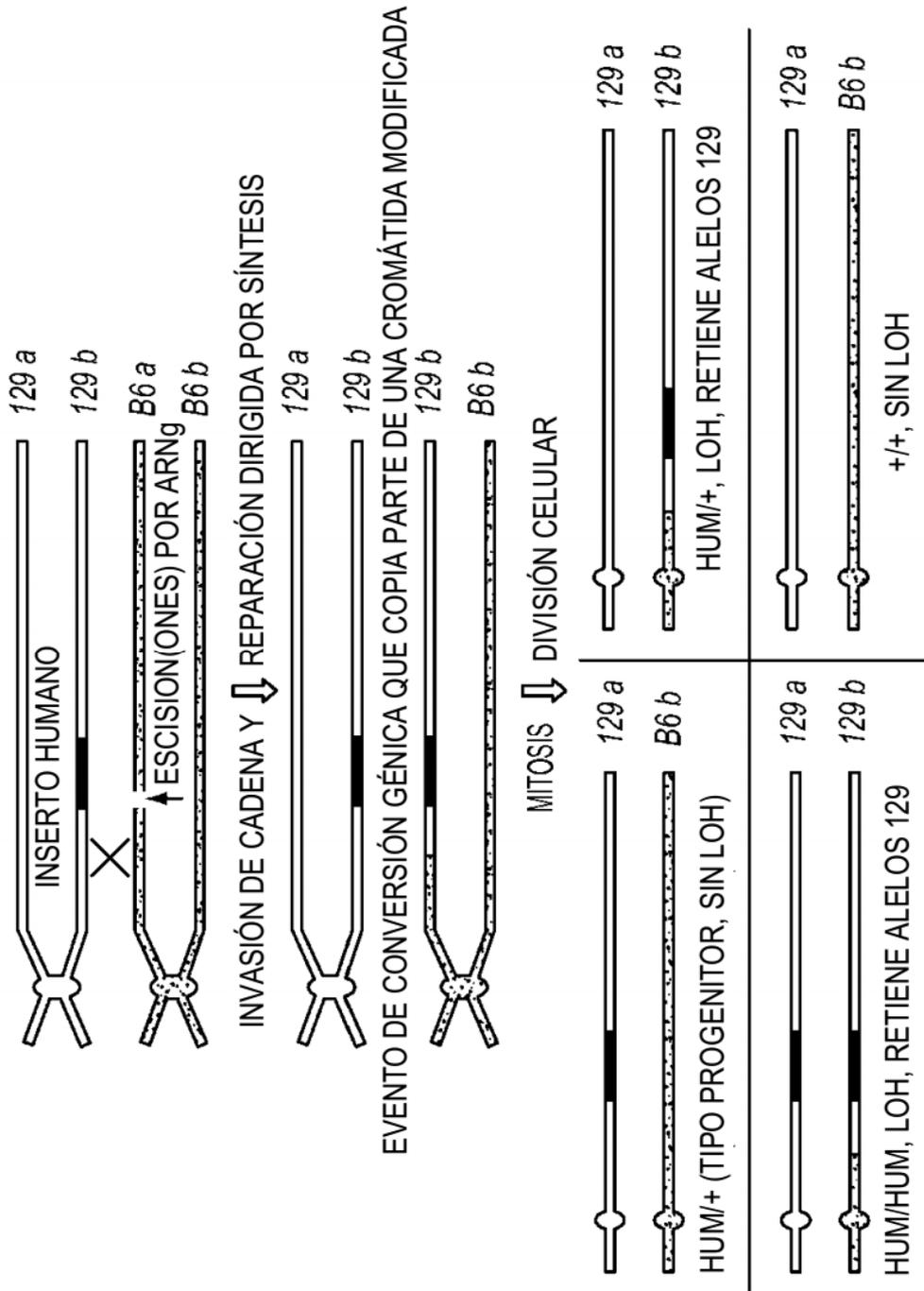


FIGURA 16E

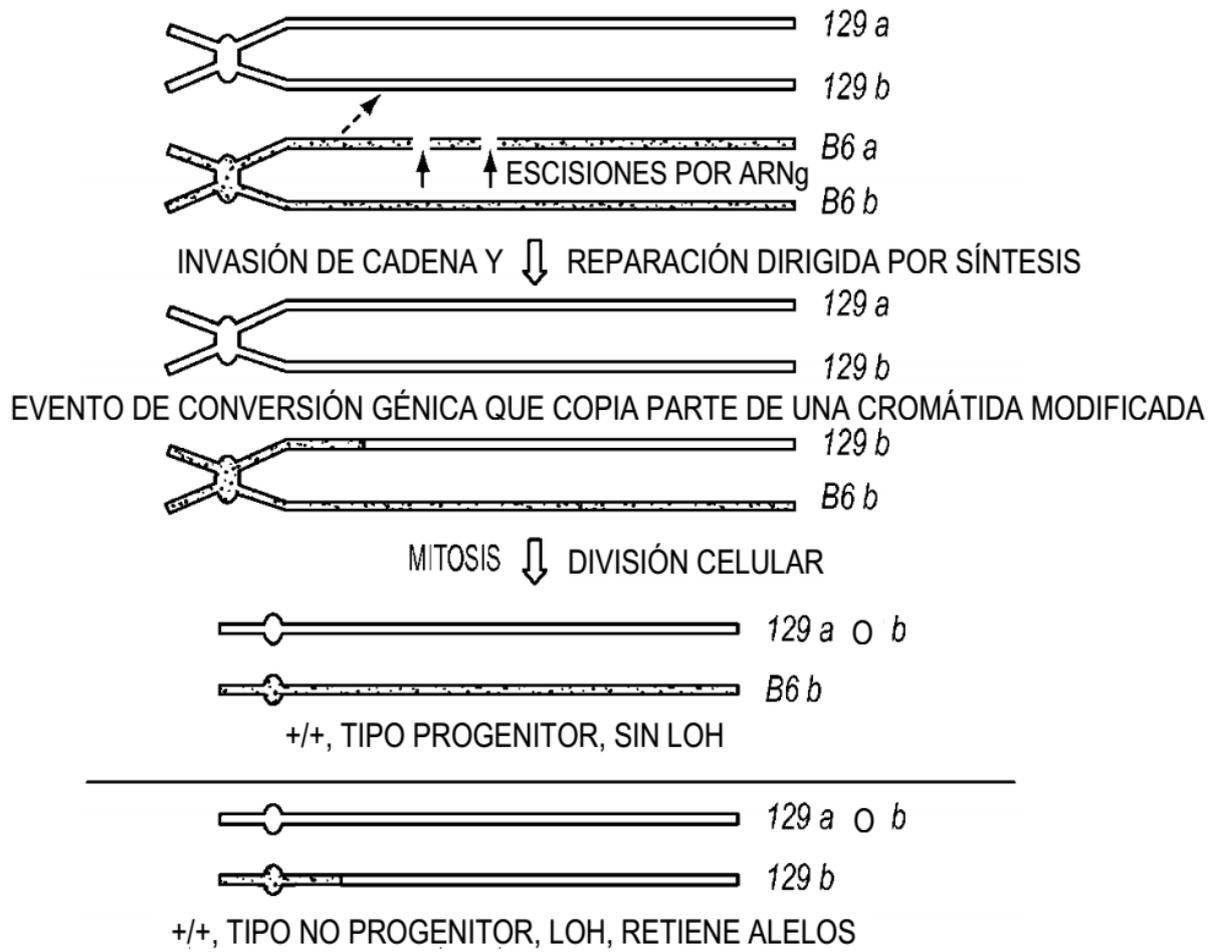


FIGURA 16F

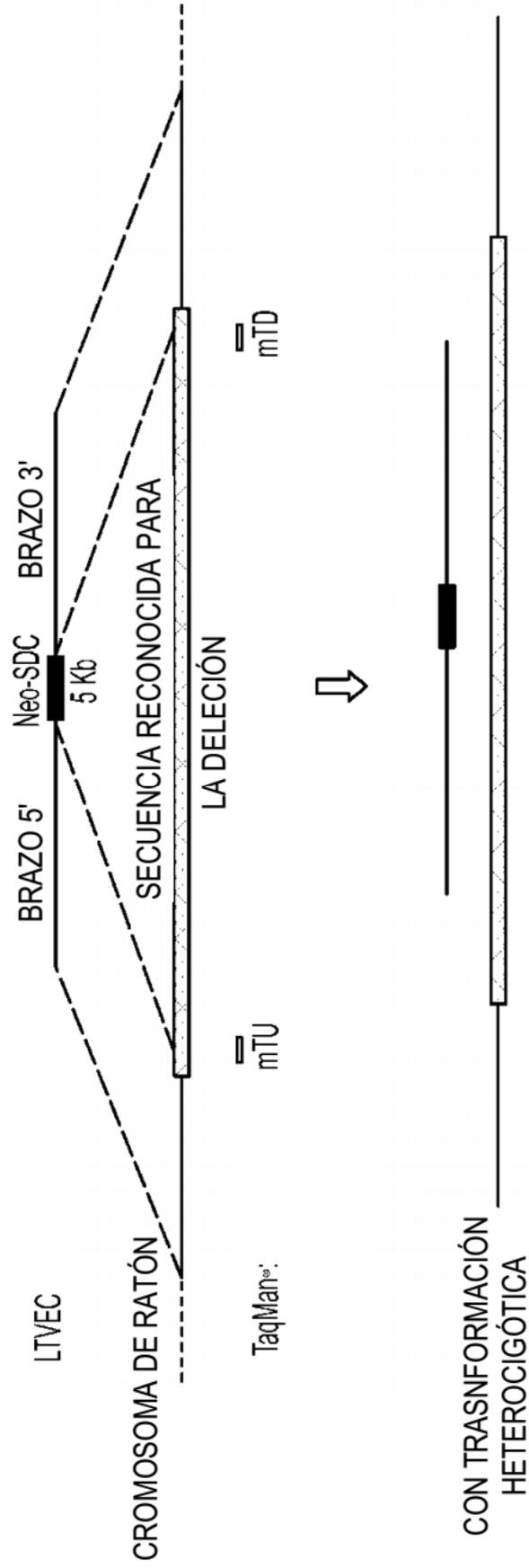


FIGURA 17A

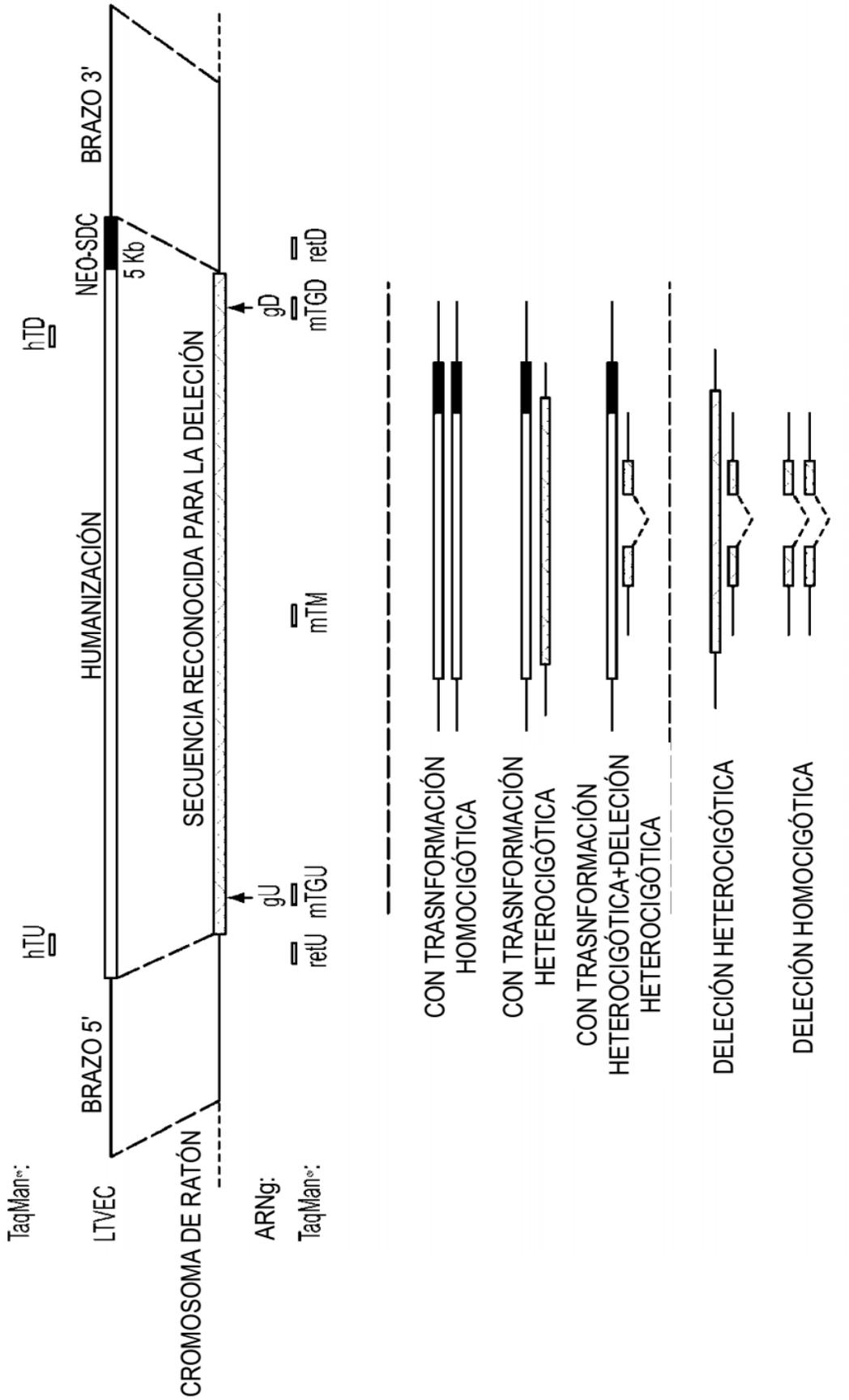


FIGURA 17B

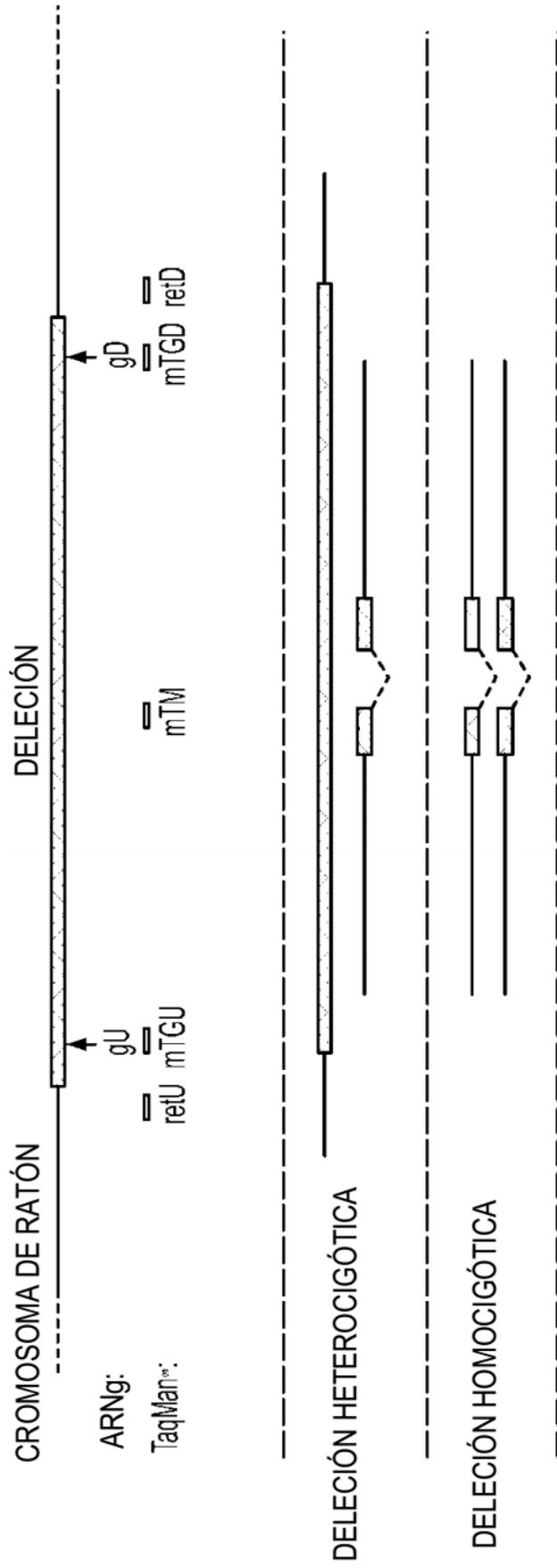


FIGURA 17C

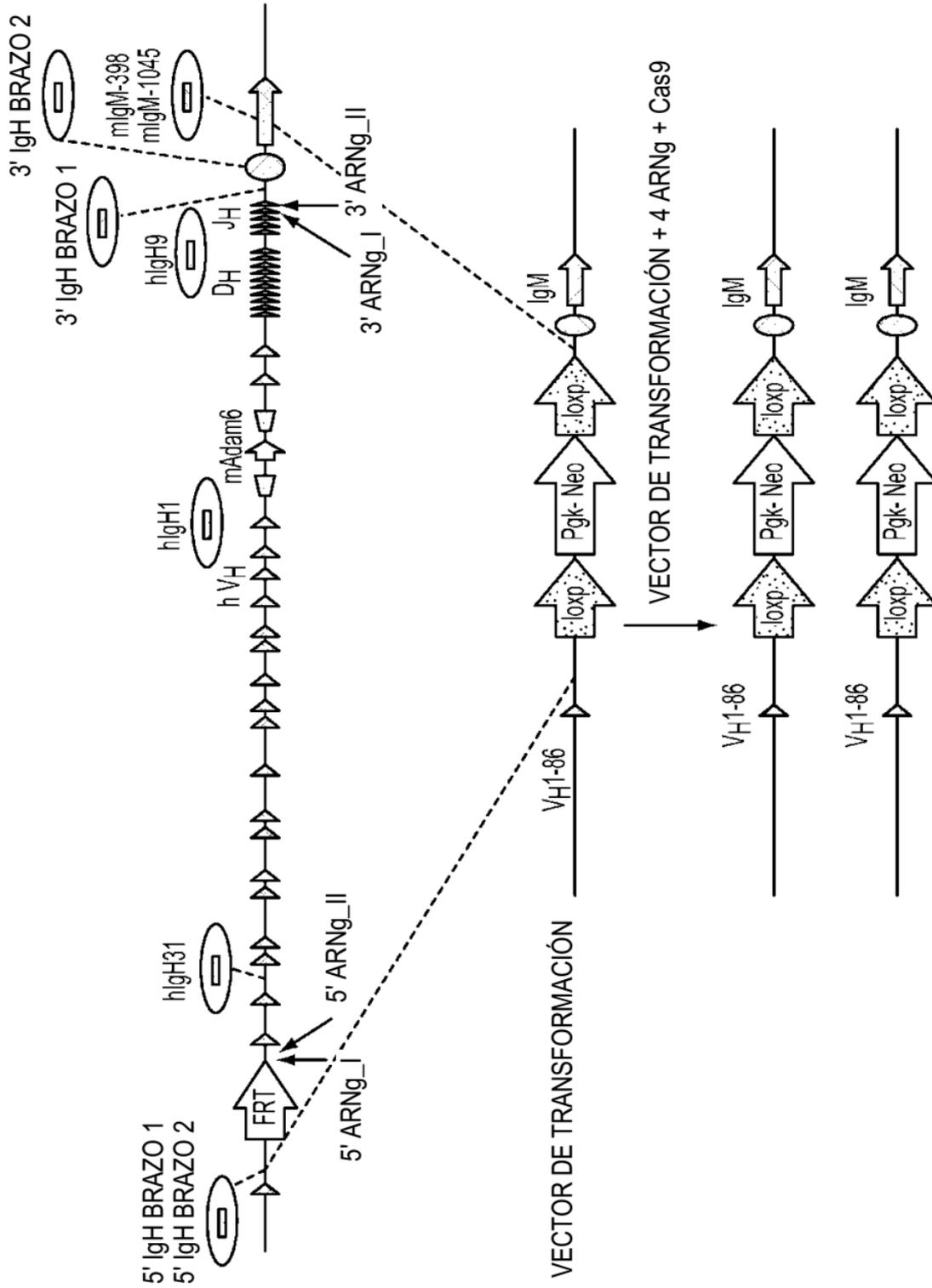


FIGURA 18