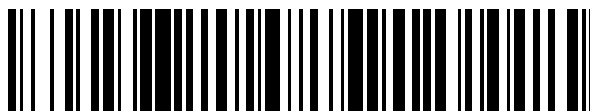


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 683 071**

51 Int. Cl.:

C12N 15/85 (2006.01)

C12N 15/90 (2006.01)

C07K 14/715 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **25.04.2013** **PCT/US2013/038165**

87 Fecha y número de publicación internacional: **31.10.2013** **WO13163394**

96 Fecha de presentación y número de la solicitud europea: **25.04.2013** **E 13720718 (9)**

97 Fecha y número de publicación de la concesión europea: **27.06.2018** **EP 2847335**

54 Título: **Direccionamiento mediado por nucleasas con grandes vectores de direccionamiento**

30 Prioridad:

25.04.2012 US 201261638267 P

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

24.09.2018

73 Titular/es:

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

72 Inventor/es:

FRENDEWEY, DAVID;
AUERBACH, WOJTEK;
VALENZUELA, DAVID M.;
YANCOPOULOS, GEORGE D. y
LAI, KA-MAN VENUS

74 Agente/Representante:

SÁEZ MAESO, Ana

Observaciones :

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

ES 2 683 071 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

Direccionamiento mediado por nucleasas con grandes vectores de direccionamiento

5 Referencia cruzada con solicitudes relacionadas

Esta solicitud reivindica la prioridad de la Solicitud Provisional de Estados Unidos No. 61/638.267 presentada el 25 de abril de 2012.

10 Campo de la invención

Nucleasas y construcciones de ADN, que incluyen vectores de direccionamiento (por ejemplo, vectores de direccionamiento grandes, "LTVEC") para lograr la recombinación homóloga en un locus genómico objetivo. Composiciones y métodos para realizar una modificación genética dirigida a través de recombinación homóloga que se ve facilitada por una ruptura de cadena sencilla o doble en o cerca de un locus genómico objetivo. Composiciones y métodos para promover la eficacia de la recombinación homóloga entre un LTVEC y un locus genómico objetivo en células procariotas o eucariotas que emplean nucleasas modificadas genéticamente.

20 Antecedentes de la invención

La recombinación homóloga usando vectores de direccionamiento que están específicamente diseñados para añadir, eliminar o reemplazar una secuencia particular de ácido nucleico en un locus genómico es un enfoque popular para lograr una modificación genómica deseada en animales no humanos. Una nucleasa que se modifica específicamente por ingeniería genética para introducir un rompimiento de cadena doble o sencilla en o cerca de un locus genómico objetivo se puede usar junto con un vector de direccionamiento para mejorar la eficacia de la recombinación homóloga en el locus genómico objetivo.

Aunque la técnica de la modificación del genoma a través de la recombinación homóloga ha avanzado considerablemente en las últimas dos décadas, todavía existen dificultades para alcanzar una frecuencia de direccionamiento aceptable usando vectores de direccionamiento muy grandes, LTVEC, en muchas circunstancias, por ejemplo, cuando una gran parte de el genoma de un roedor se reemplaza con un fragmento genómico humano grande, o dificultades de direccionamiento a ciertos tipos de células, por ejemplo, fibroblastos u otras células somáticas. Existe una necesidad en la técnica por métodos adicionales y mejorados para modificar grandes loci genómicos de un genoma eucariota usando LTVEC.

35 Sumario

Se proporcionan composiciones y métodos para modificar un locus genómico de interés usando un vector de direccionamiento grande (LTVEC) en combinación con un agente nucleasa, que permite la eliminación, adición (por ejemplo, inserción) y/o el reemplazo eficiente de una gran secuencia de ácido nucleico en el locus genómico de interés.

Se proporcionan composiciones y métodos para modificar un locus genómico objetivo de un mamífero en una célula procariota usando un LTVEC y un agente nucleasa a través de recombinación homóloga bacteriana (BHR), en el que el BHR se facilita mediante una escisión de una cadena sencilla o doble en o cerca del locus genómico objetivo creado por el agente nucleasa. Se proporcionan células procariotas que comprenden un LTVEC y un agente nucleasa que, tras la expresión, es capaz de introducir una escisión de cadena sencilla o doble en o cerca de un sitio objetivo. Se proporcionan composiciones y métodos para reemplazar un locus genómico grande de un animal no humano con una secuencia de ácido nucleico exógeno, por ejemplo, secuencias de ácido nucleico genómico humano homólogas u ortólogas, en una célula procariota recombinogénica empleando varios LTVEC y nucleasas como se describe en la presente memoria.

Se proporcionan composiciones y métodos para modificar un locus genómico de interés usando un LTVEC y un agente nucleasa en diversas células de mamífero pluripotentes. Se proporcionan composiciones y métodos para reemplazar un locus genómico grande de un animal no humano con una secuencia de ácido nucleico exógeno, por ejemplo, secuencias de ácido nucleico genómico humano homólogas u ortólogas, en una célula pluripotente del animal no humano empleando varios LTVEC y nucleasas como se describe en la presente memoria.

Se proporcionan células pluripotentes de un animal no humano que comprenden diversos LTVEC y agentes de nucleasa descritos en este documento.

Se proporcionan composiciones y métodos para generar un animal no humano genéticamente modificado que comprende una o más modificaciones genéticas dirigidas como se describe en este documento.

65 La divulgación se refiere a una célula procariota que comprende un vector de direccionamiento grande (LTVEC) que comprende brazos de homología dirigidos a un locus objetivo, y una secuencia de ácido nucleico que codifica un

agente nucleasa que realiza un rompimiento de cadena sencilla o doble en o cerca del lugar de destino.

En una realización, la célula procariota es capaz de expresar una recombinasa que media la recombinación homóloga bacteriana (BHR). En una realización, la célula procariota es una cepa de *E. coli* competente para la recombinación.

En una realización, el LTVEC varía de aproximadamente 50 kb a aproximadamente 300 kb. En una realización, el LTVEC varía de aproximadamente 50 kb a aproximadamente 75 kb. En una realización, el LTVEC varía de aproximadamente 75 kb a aproximadamente 100 kb. En una realización, el LTVEC varía de aproximadamente 100 kb a 125 kb. En una realización, el LTVEC varía de aproximadamente 125 kb a aproximadamente 150 kb. En una realización, el LTVEC varía de aproximadamente 150 kb a aproximadamente 175 kb. En una realización, el LTVEC varía de aproximadamente 175 kb a aproximadamente 200 kb. En una realización, el LTVEC varía de aproximadamente 200 kb a aproximadamente 225 kb. En una realización, el LTVEC varía de aproximadamente 225 kb a aproximadamente 250 kb. En una realización, el LTVEC varía de aproximadamente 250 kb a aproximadamente 275 kb. En una realización, el LTVEC varía desde aproximadamente 275 kb a aproximadamente de 300 kb.

En una realización, los brazos de homología del vector de direccionamiento se derivan de una biblioteca de BAC, una biblioteca de cósmidos o una biblioteca de fagos P1. En una realización, los brazos de homología se derivan de un locus genómico del animal no humano que no puede direccionarse usando un método convencional. En una realización, los brazos de homología se derivan de un ADN sintético.

En una realización, una suma total del brazo de homología secuencia arriba y el brazo de homología secuencia abajo es de al menos 10 kb. En una realización, el brazo de homología secuencia arriba varía de aproximadamente 5 kb a aproximadamente 100 kb. En una realización, el brazo de homología secuencia abajo varía de aproximadamente 5 kb a aproximadamente 100 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 5 kb a aproximadamente 10 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 10 kb a aproximadamente 20 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 20 kb a aproximadamente 30 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 30 kb a aproximadamente 40 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 40 kb a aproximadamente 50 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 50 kb a aproximadamente 60 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 60 kb a aproximadamente 70 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 70 kb a aproximadamente 80 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 80 kb a aproximadamente 90 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 90 kb a aproximadamente 100 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 100 kb a aproximadamente 110 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 110 kb a aproximadamente 120 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 120 kb a aproximadamente 130 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 130 kb a aproximadamente 140 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 140 kb a aproximadamente 150 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 150 kb a aproximadamente 160 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 160 kb a aproximadamente 170 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 170 kb a aproximadamente 180 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 180 kb a aproximadamente 190 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 190 kb a aproximadamente 200 kb.

En una realización, el LTVEC comprende un casete de selección. En una realización, el casete de selección comprende una secuencia de ácido nucleico que codifica un marcador de selección, en el que la secuencia de ácido nucleico está operativamente unida a un promotor. En una realización, el promotor es activo en una célula procariota. En una realización, el promotor es activo tanto en células procariotas como en células eucariotas. En una realización, el marcador de selección se selecciona de neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa (puro^r), blasticidina S desaminasa (bsr^r), xantina / guanina fosforribosil transferasa (gpt), y timidina quinasa del virus del herpes simple (HSV-k), y una combinación de los mismos.

En una realización, el LTVEC comprende un ácido nucleico insertado que varía de aproximadamente 5 kb a aproximadamente 200 kb. En una realización, el ácido nucleico insertado es de aproximadamente 5 kb a aproximadamente 10 kb. En una realización, el ácido nucleico insertado es de aproximadamente 10 kb a aproximadamente 20 kb. En una realización, el ácido nucleico insertado es de aproximadamente 20 kb a aproximadamente 30 kb. En una realización, el ácido nucleico insertado es de aproximadamente 30 kb a aproximadamente 40 kb. En una realización, el ácido nucleico insertado es de aproximadamente 40 kb a aproximadamente 50 kb. En una realización, el ácido nucleico insertado es de aproximadamente 50 kb a aproximadamente 60 kb a

aproximadamente 70 kb. En una realización, el ácido nucleico insertado es de aproximadamente 80 kb a aproximadamente 90 kb. En una realización, el ácido nucleico insertado es de aproximadamente 90 kb a aproximadamente 100 kb. En una realización, el ácido nucleico insertado es de aproximadamente 100 kb a aproximadamente 110 kb. En una realización, el ácido nucleico insertado es de aproximadamente 120 kb a aproximadamente 130 kb. En una realización, el ácido nucleico insertado es de aproximadamente 130 kb a aproximadamente 140 kb. En una realización, el ácido nucleico insertado es de aproximadamente 140 kb a aproximadamente 150 kb. En una realización, el ácido nucleico insertado es de aproximadamente 150 kb a aproximadamente 160 kb. En una realización, el ácido nucleico insertado es de aproximadamente 160 kb a aproximadamente 170 kb. En una realización, el ácido nucleico insertado es de aproximadamente 170 kb a aproximadamente 180 kb. En una realización, el ácido nucleico insertado es de aproximadamente 180 kb a aproximadamente 190 kb. En una realización, el ácido nucleico insertado es de aproximadamente 190 kb a aproximadamente 200 kb.

En una realización, el LTVEC comprende un ácido nucleico insertado que comprende un ácido nucleico flanqueado con secuencias objetivo de recombinación específicas del sitio. En una realización, el ácido nucleico comprende un ácido nucleico genómico. En una realización, el ácido nucleico genómico se deriva de un ratón, un ser humano o una combinación de los mismos. En una realización, las secuencias objetivo de recombinación específicas del sitio se seleccionan del grupo que consiste en loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attP, att, FRT, rox, y una combinación de las mismas.

En una realización, el LTVEC comprende un ácido nucleico insertado que comprende un alelo condicional. En una realización, el alelo condicional es un alelo multifuncional, como se describe en el documento US 2011/0104799. En una realización, el alelo condicional comprende: (a) una secuencia activadora en orientación sentido con respecto a la transcripción de un gen objetivo, y un casete de selección de fármaco en orientación sentido o antisentido; (b) en orientación antisentido una secuencia de nucleótidos de interés (NSI) y un módulo condicional por inversión (COIN, que utiliza un intrón de división de exones y un módulo de tipo trampa génica invertible; véase, por ejemplo, el documento US 2011/0104799,); y (c) unidades recombinables que se recombinan tras la exposición a una primera recombinasa para formar un alelo condicional que (i) carece de la secuencia activadora y la DSC, y (ii) contiene la NSI en orientación sentido y el COIN en orientación antisentido.

En una realización, el LTVEC comprende un ácido nucleico insertado que comprende un casete de selección. En una realización, el casete de selección comprende una secuencia de ácido nucleico que codifica un marcador de selección, en el que la secuencia de ácido nucleico está operativamente unida a un promotor. En una realización, el promotor es activo en una célula procariota. En una realización, el ácido nucleico es activo en células tanto procariotas como eucariotas. En una realización, el casete de selección está flanqueado con secuencias objetivo de recombinación específicas del sitio. En una realización, el marcador de selección se selecciona de neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa (puro^r), blastidina S desaminasa (bsr^r), xantina / guanina fosforribosil transferasa (gpt), y timidina quinasa del virus del herpes simple (HSV-k), y una combinación de los mismos.

En una realización, el LTVEC comprende un ácido nucleico insertado que comprende un gen informador unido operativamente a un promotor, en el que el gen informador codifica una proteína informadora seleccionada del grupo que consiste en LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Rojo, DsRed, mOrange, mKO, mCitrine, Venus, YPet, proteína fluorescente amarilla mejorada (EYFP), Emerald, proteína fluorescente verde mejorada (EGFP), CyPet, proteína fluorescente cian (CFP), Cerulean, T-Sapphire, luciferasa, fosfatasa alcalina, y una combinación de las mismas. En una realización, el gen informador se expresa bajo el control de un promotor inducible. En una realización, el gen informador se expresa bajo el control de un promotor endógeno. En una realización, el gen informador se expresa bajo el control de un promotor exógeno. En una realización, el gen informador se expresa en un tipo específico de célula. En una realización, el gen informador se expresa de una manera específica del tejido. En una realización, el gen informador se expresa en una forma específica de la etapa de desarrollo.

La divulgación se refiere a una célula eucariota que comprende un vector de direccionamiento grande que comprende brazos de homología dirigidos a un locus objetivo dentro del genoma de la célula eucariota, y una secuencia de ácido nucleico que codifica un agente nucleasa que hace un rompimiento de cadena sencilla o doble en o cerca del lugar de destino.

En una realización, la célula eucariota es una célula pluripotente. En una realización, la célula pluripotente es una célula madre embrionaria (ES). En una realización, la célula pluripotente es una célula ES no humana. En una realización, la célula pluripotente es una célula madre pluripotente inducida (iPS). En una realización, la célula pluripotente inducida (iPS) se deriva de un fibroblasto. En una realización, la célula pluripotente inducida (iPS) se deriva de un fibroblasto humano. En una realización, la célula pluripotente es una célula madre hematopoyética (HSC). En una realización, la célula pluripotente es una célula madre neuronal (NSC). En una realización, la célula pluripotente es una célula madre epiblasto. En una realización, la célula pluripotente es una célula progenitora restringida en el desarrollo. En una realización, la célula pluripotente es una célula pluripotente de roedor. En una realización, la célula pluripotente de roedor es una célula pluripotente de rata. En una realización, la célula pluripotente de rata es una célula ES de rata. En una realización, la célula pluripotente de roedor es una célula

pluripotente de ratón. En una realización, la célula pluripotente es una célula madre embrionaria (ES) de ratón.

En una realización, la célula eucariota es una célula inmortalizada de ratón o rata. En una realización, la célula eucariota es una célula humana inmortalizada. En una realización, la célula eucariota es un fibroblasto humano. En una realización, la célula eucariota es una célula cancerosa. En una realización, la célula eucariota es una célula cancerosa humana.

En una realización, el LTVEC varía de aproximadamente 50 kb a aproximadamente 300 kb. En una realización, el LTVEC varía de aproximadamente 50 kb a aproximadamente 75 kb. En una realización, el LTVEC varía de aproximadamente 75 kb a aproximadamente 100 kb. En una realización, el LTVEC varía de aproximadamente 100 kb a 125 kb. En una realización, el LTVEC varía de aproximadamente 150 kb a aproximadamente 175 kb. En una realización, el LTVEC varía de aproximadamente 175 kb a aproximadamente 200 kb. En una realización, el LTVEC varía de aproximadamente 200 kb a aproximadamente 225 kb. En una realización, el LTVEC varía de aproximadamente 225 kb a aproximadamente 250 kb. En una realización, el LTVEC varía de aproximadamente 250 kb a aproximadamente 275 kb. En una realización, el LTVEC varía de aproximadamente 275 kb a aproximadamente 300 kb.

En una realización, los brazos de homología del vector de direccionamiento se derivan de una biblioteca de BAC, una biblioteca de cósmidos o una biblioteca de fagos P1. En una realización, los brazos de homología se derivan de un locus genómico del animal no humano que no puede direccionarse usando un método convencional. En una realización, los brazos de homología se derivan de un ADN sintético.

En una realización, una suma total del brazo de homología secuencia arriba y el brazo de homología secuencia abajo es de al menos 10 kb. En una realización, el brazo de homología secuencia arriba varía de aproximadamente 5 kb a aproximadamente 100 kb. En una realización, el brazo de homología secuencia abajo varía de aproximadamente 5 kb a aproximadamente 100 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 5 kb a aproximadamente 10 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 10 kb a aproximadamente 20 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 20 kb a aproximadamente 30 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 30 kb a aproximadamente 40 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 40 kb a aproximadamente 50 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 50 kb a aproximadamente 60 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 60 kb a aproximadamente 70 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 70 kb a aproximadamente 80 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 80 kb a aproximadamente 90 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 90 kb a aproximadamente 100 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 100 kb a aproximadamente 110 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 110 kb a aproximadamente 120 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 120 kb a aproximadamente 130 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 130 kb a aproximadamente 140 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 140 kb a aproximadamente 150 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 150 kb a aproximadamente 160 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 160 kb a aproximadamente 170 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 170 kb a aproximadamente 180 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 180 kb a aproximadamente 190 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 190 kb a aproximadamente 200 kb.

En una realización, el LTVEC comprende un casete de selección. En una realización, el casete de selección comprende una secuencia de ácido nucleico que codifica un marcador de selección, en el que la secuencia de ácido nucleico está operativamente unida a un promotor. En una realización, el promotor es activo en una célula procariota. En una realización, el promotor es activo tanto en células procariotas como en células eucariotas. En una realización, el marcador de selección se selecciona de neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa (puro^r), blastidina S desaminasa (bsr^r), xantina / guanina fosforribosil transferasa (gpt), y timidina quinasa del virus del herpes simple (HSV-k), y una combinación de los mismos.

En una realización, el LTVEC comprende un ácido nucleico insertado que varía de aproximadamente 5 kb a aproximadamente 200 kb. En una realización, el ácido nucleico insertado es de aproximadamente 5 kb a aproximadamente 10 kb. En una realización, el ácido nucleico insertado es de aproximadamente 10 kb a aproximadamente 20 kb. En una realización, el ácido nucleico insertado es de aproximadamente 20 kb a aproximadamente 30 kb. En una realización, el ácido nucleico insertado es de aproximadamente 30 kb a

aproximadamente 40 kb. En una realización, el ácido nucleico insertado es de aproximadamente 40 kb a aproximadamente 50 kb. En una realización, el ácido nucleico insertado es de aproximadamente 60 kb a aproximadamente 70 kb. En una realización, el ácido nucleico insertado es de aproximadamente 80 kb a aproximadamente 90 kb. En una realización, el ácido nucleico insertado es de aproximadamente 90 kb a aproximadamente 100 kb. En una realización, el ácido nucleico insertado es de aproximadamente 100 kb a aproximadamente 110 kb. En una realización, el ácido nucleico insertado es de aproximadamente 120 kb a aproximadamente 130 kb. En una realización, el ácido nucleico insertado es de aproximadamente 130 kb a aproximadamente 140 kb. En una realización, el ácido nucleico insertado es de aproximadamente 140 kb a aproximadamente 150 kb. En una realización, el ácido nucleico insertado es de aproximadamente 150 kb a aproximadamente 160 kb. En una realización, el ácido nucleico insertado es de aproximadamente 160 kb a aproximadamente 170 kb. En una realización, el ácido nucleico insertado es de aproximadamente 170 kb a aproximadamente 180 kb. En una realización, el ácido nucleico insertado es de aproximadamente 180 kb a aproximadamente 190 kb. En una realización, el ácido nucleico insertado es de aproximadamente 190 kb a aproximadamente 200 kb.

En una realización, el LTVEC comprende un ácido nucleico insertado que comprende un ácido nucleico flanqueado con secuencias objetivo de recombinación específicas del sitio. En una realización, el ácido nucleico comprende un ácido nucleico genómico. En una realización, el ácido nucleico genómico se deriva de un ratón, un ser humano o una combinación de los mismos. En una realización, las secuencias objetivo de recombinación específicas del sitio se seleccionan del grupo que consiste en loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attP, att, FRT, rox, y una combinación de los mismos.

En una realización, el LTVEC comprende un ácido nucleico insertado que comprende un alelo condicional. En una realización, el alelo condicional es un alelo multifuncional, como se describe en el documento US 2011/0104799. En una realización, el alelo condicional comprende: (a) una secuencia activadora en orientación sentido con respecto a la transcripción de un gen objetivo, y un casete de selección de fármaco en orientación sentido o antisentido; (b) en orientación antisentido una secuencia de nucleótidos de interés (NSI) y un módulo condicional por inversión (COIN, que utiliza un intrón de división de exones y un módulo de tipo trampa génica invertible; véase, por ejemplo, el documento US 2011/0104799,); y (c) unidades recombinables que se recombinan tras la exposición a una primera recombinasa para formar un alelo condicional que (i) carece de la secuencia activadora y la DSC, y (ii) contiene el NSI en orientación sentido y el COIN en orientación antisentido.

En una realización, el LTVEC comprende un ácido nucleico insertado que comprende un casete de selección. En una realización, el casete de selección comprende una secuencia de ácido nucleico que codifica un marcador de selección, en el que la secuencia de ácido nucleico está operativamente unida a un promotor. En una realización, el promotor es activo en una célula procariota. En una realización, el ácido nucleico es activo en células tanto procariotas como eucariotas. En una realización, el casete de selección está flanqueado con secuencias objetivo de recombinación específicas del sitio. En una realización, el marcador de selección se selecciona de neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa (puro^r), blastidina S desaminasa (bsr^r), xantina / guanina fosforribosil transferasa (gpt), y timidina quinasa del virus del herpes simple (HSV-k), y una combinación de los mismos.

En una realización, el LTVEC comprende un ácido nucleico insertado que comprende un gen informador unido operativamente a un promotor, en el que el gen informador codifica una proteína informadora seleccionada del grupo que consiste en LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Rojo, DsRed, mOrange, mKO, mCitrine, Venus, YPet, proteína fluorescente amarilla mejorada (EYFP), Emerald, proteína fluorescente verde mejorada (EGFP), CyPet, proteína fluorescente cian (CFP), Cerulean, T-Sapphire, luciferasa, fosfatasa alcalina, y una combinación de las mismas. En una realización, el gen informador se expresa bajo el control de un promotor inducible. En una realización, el gen informador se expresa bajo el control de un promotor endógeno. En una realización, el gen informador se expresa en un tipo de célula específico. En una realización, el gen informador se expresa de una forma específica del tejido. En una realización, el gen informador se expresa en una forma específica de la etapa de desarrollo.

La divulgación se refiere a un método para modificar un locus genómico objetivo de una célula de mamífero mediante recombinación homóloga bacteriana (BHR) en una célula procariota que comprende:

(a) introducción en una célula procariota que comprende un locus genómico objetivo de un mamífero:

(i) un vector de direccionamiento que comprende un ácido nucleico insertado flanqueado con un primer brazo de homología secuencia arriba y un primer brazo de homología secuencia abajo, y
(ii) un agente nucleasa que produce una ruptura de cadena sencilla o doble en o cerca del locus genómico objetivo, y

(i) seleccionar una célula procariota dirigida que comprende el ácido nucleico insertado,

en el que la célula procariota es capaz de expresar una recombinasa que media la BHR.

En una realización, el locus genómico objetivo se selecciona de un locus FcER1a, un locus TLR4, un locus PRLR, un locus Notch4, un locus Accn2, un locus Adamts5, un locus TRPA1, un locus FoiH1, un locus LRP5, y un locus ERBB4.

- 5 En una realización, el locus genómico objetivo está presente en un vector de direccionamiento grande (LTVEC) que comprende un segundo brazo de homología secuencia arriba y un segundo brazo de homología secuencia abajo. En una realización, una suma total del segundo brazo de homología secuencia arriba y el segundo brazo de homología secuencia abajo es de al menos 10 kb. En una realización, el segundo brazo de homología secuencia arriba varía de aproximadamente 5 kb a aproximadamente 100 kb. En una realización, el segundo brazo de homología secuencia abajo varía de aproximadamente 5 kb a aproximadamente 100 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 5 kb a aproximadamente 10 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 10 kb a aproximadamente 20 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 20 kb a aproximadamente 30 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 30 kb a aproximadamente 40 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 40 kb a aproximadamente 50 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 50 kb a aproximadamente 60 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 60 kb a aproximadamente 70 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 70 kb a aproximadamente 80 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 80 kb a aproximadamente 90 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 90 kb a aproximadamente 100 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 100 kb a aproximadamente 110 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 110 kb a aproximadamente 120 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 120 kb a aproximadamente 130 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 130 kb a aproximadamente 140 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 140 kb a aproximadamente 150 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 150 kb a aproximadamente 160 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 160 kb a aproximadamente 170 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 170 kb a aproximadamente 180 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 180 kb a aproximadamente 190 kb. En una realización, los segundos brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 190 kb a aproximadamente 200 kb.
- 40 En una realización, el mamífero es un humano y el direccionamiento es de una célula humana *ex vivo*. En una realización, el mamífero es un roedor. En una realización, el roedor se selecciona de un ratón, una rata y un hámster.
- 45 En una realización, el agente nucleasa se introduce junto con el vector de direccionamiento. En una realización, el agente nucleasa se introduce por separado del vector de direccionamiento durante un período de tiempo. En una realización, el agente nucleasa se introduce antes de la introducción del vector de direccionamiento. En una realización, el agente nucleasa se introduce después de la introducción del vector de direccionamiento.
- 50 En una realización, el uso combinado del vector de direccionamiento con el agente nucleasa da como resultado una mayor eficacia de direccionamiento en comparación con el uso del vector de direccionamiento solo. En una realización, cuando el vector de direccionamiento se usa junto con el agente nucleasa, la eficacia de direccionamiento del vector de direccionamiento se incrementa al menos dos veces en comparación con cuando el vector de direccionamiento se usa solo. En una realización, cuando el vector de direccionamiento se usa junto con el agente nucleasa, la eficacia de dirección del vector de direccionamiento se incrementa al menos tres veces en comparación con cuando el vector de direccionamiento se usa solo. En una realización, cuando el vector de direccionamiento se usa junto con el agente nucleasa, la eficacia de dirección del vector de direccionamiento se incrementa al menos cuatro veces en comparación con cuando el vector de direccionamiento se usa solo.
- 55 En una realización, la célula procariota es una cepa de *E. coli* competente para recombinación. En una realización, la célula procariota comprende un ácido nucleico que codifica la recombinasa. En una realización, la célula procariota no comprende el ácido nucleico que codifica la recombinasa, y el ácido nucleico que codifica la recombinasa se introduce en la célula procariota. En una realización, el ácido nucleico comprende un ADN o un ARNm que codifica la recombinasa. En una realización, el ácido nucleico que codifica la recombinasa es pABG. En una realización, la recombinasa se expresa bajo el control de un promotor inducible. En una realización, la expresión de la recombinasa está controlada por arabinosa.
- 60
- 65

En una realización, el agente nucleasa es una construcción de expresión que comprende una secuencia de ácido nucleico que codifica una nucleasa, en la que la secuencia de ácido nucleico está operativamente unida a un promotor. En una realización, el promotor es un promotor constitutivamente activo. En una realización, el promotor es un promotor inducible. En una realización, el promotor es activo en la célula procariota. En una realización, el agente nucleasa es un ARNm que codifica una endonucleasa.

En una realización, el agente nucleasa es una nucleasa de dedo de cinc (ZFN). En una realización, cada monómero de la ZFN comprende 3 o más dominios de unión a ADN basados en dedos de cinc, en donde cada dominio de unión a ADN basado en dedos de cinc se une a un subsitio de 3 pb. En una realización, la ZFN es una proteína química que comprende un dominio de unión a ADN basado en dedos de cinc operativamente unido a una nucleasa independiente. En una realización, la endonucleasa independiente es una endonucleasa FokI. En una realización, el agente nucleasa comprende una primera ZFN y una segunda ZFN, en donde cada una de la primera ZFN y la segunda ZFN está unida operativamente a una nucleasa FokI, en donde la primera y la segunda ZFN reconocen dos secuencias de ADN objetivo contiguas en cada cadena de la secuencia de ADN objetivo separada por un sitio de escisión de aproximadamente 6 pb a aproximadamente 40 pb, y en el que las nucleasas FokI se dimerizan y hacen un rompimiento de doble cadena.

En una realización, el agente nucleasa es una nucleasa efectora similar al activador de la transcripción (TALEN). En una realización, cada monómero de TALEN comprende 12-25 repeticiones de TAL, en donde cada repetición de TAL se une a un subsitio de 1 pb. En una realización, el agente nucleasa es una proteína química que comprende un dominio de unión a ADN basado en la repetición de TAL operativamente unido a una nucleasa independiente. En una realización, la nucleasa independiente es una endonucleasa FokI. En una realización, el agente nucleasa comprende un primer dominio de unión a ADN basado en la repetición de TAL y un segundo dominio de unión a ADN basado en la repetición de TAL, en el que cada uno del primer y segundo dominio de unión a ADN basado en la repetición de TAL está operativamente unido a una nucleasa FokI, en la que el primer y el segundo dominio de unión de ADN basado en la repetición de TAL reconocen dos secuencias de ADN objetivo contiguas en cada cadena de la secuencia de ADN objetivo separadas por aproximadamente 6 pb a aproximadamente 40 pb del sitio de escisión, y en donde las nucleasas FokI se dimerizan y producen una ruptura de doble cadena en una secuencia objetivo.

En una realización, cada monómero de la nucleasa reconoce una secuencia objetivo de al menos 9 nucleótidos. En una realización, la secuencia objetivo es de aproximadamente 9 a aproximadamente 12 nucleótidos de longitud. En una realización, la secuencia objetivo es de aproximadamente 12 a aproximadamente 15 nucleótidos de longitud. En una realización, la secuencia objetivo es de aproximadamente 15 a aproximadamente 18 nucleótidos de longitud. En una realización, la secuencia objetivo es de aproximadamente 18 a aproximadamente 21 nucleótidos de longitud.

En una realización, una secuencia objetivo del agente nucleasa está localizada en un intrón. En una realización, la secuencia objetivo está ubicada en un exón. En una realización, la secuencia objetivo está ubicada en un promotor. En una realización, la secuencia objetivo está en una región que no codifica proteínas. En una realización, la región que no codifica proteínas es una región reguladora. En una realización, la secuencia objetivo está ubicada en una región reguladora del promotor. En una realización, la secuencia objetivo está localizada en una región potenciadora.

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

En una realización, el vector de direccionamiento es un vector de direccionamiento grande (LTVEC).

En una realización, el LTVEC varía de aproximadamente 50 kb a aproximadamente 300 kb. En una realización, el LTVEC varía de aproximadamente 50 kb a aproximadamente 75 kb. En una realización, el LTVEC varía de aproximadamente 75 kb a aproximadamente 100 kb. En una realización, el LTVEC varía de aproximadamente 100 kb a 125 kb. En una realización, el LTVEC varía de aproximadamente 125 kb a aproximadamente 150 kb. En una realización, el LTVEC varía de aproximadamente 150 kb a aproximadamente 175 kb. En una realización, el LTVEC varía de aproximadamente 175 kb a aproximadamente 200 kb. En una realización, el LTVEC varía de aproximadamente 200 kb a aproximadamente 225 kb. En una realización, el LTVEC varía de aproximadamente 225 kb a aproximadamente 250 kb. En una realización, el LTVEC varía de aproximadamente 250 kb a aproximadamente 275 kb. En una realización, el LTVEC varía de aproximadamente 275 kb a aproximadamente 300 kb.

En una realización, los brazos de homología del vector de direccionamiento se derivan de una biblioteca de BAC, una biblioteca de cósmidos o una biblioteca de fagos P1. En una realización, los brazos de homología se derivan de un locus genómico del animal no humano que no puede direccionarse usando un método convencional. En una realización, los brazos de homología se derivan de un ADN sintético.

En una realización, una suma total del brazo de homología secuencia arriba y el brazo de homología secuencia abajo es de al menos 10 kb. En una realización, el brazo de homología secuencia arriba varía de aproximadamente 5 kb a aproximadamente 100 kb. En una realización, el brazo de homología secuencia abajo varía de aproximadamente 5 kb a aproximadamente 100 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 5 kb a aproximadamente 10 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 10 kb a aproximadamente 20 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 20 kb a aproximadamente 30 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 30 kb a aproximadamente 40 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 40 kb a aproximadamente 50 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 50 kb a aproximadamente 60 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 60 kb a aproximadamente 70 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 70 kb a aproximadamente 80 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 80 kb a aproximadamente 90 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 90 kb a aproximadamente 100 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 100 kb a aproximadamente 110 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 110 kb a aproximadamente 120 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 120 kb a aproximadamente 130 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 130 kb a aproximadamente 140 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 140 kb a aproximadamente 150 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 150 kb a aproximadamente 160 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 160 kb a aproximadamente 170 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 170 kb a aproximadamente 180 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 180 kb a aproximadamente 190 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 190 kb a aproximadamente 200 kb.

En una realización, el vector de direccionamiento comprende un casete de selección. En una realización, el casete de selección comprende una secuencia de ácido nucleico que codifica un marcador de selección, en el que la secuencia de ácido nucleico está operativamente unida a un promotor. En una realización, el promotor es activo en una célula procariota. En una realización, el promotor es activo tanto en células procariotas como en células eucariotas. En una realización, el marcador de selección se selecciona de neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa (puro^r), blasticidina S desaminasa (bsr^r), xantina / guanina fosforribosil transferasa (gpt), y timidina quinasa del virus del herpes simple (HSV-k), y una combinación de los mismos.

En una realización, el ácido nucleico insertado es de aproximadamente 5 kb a aproximadamente 200 kb. En una realización, el ácido nucleico insertado es de aproximadamente 5 kb a aproximadamente 10 kb. En una realización, el ácido nucleico insertado es de aproximadamente 10 kb a aproximadamente 20 kb. En una realización, el ácido nucleico insertado es de aproximadamente 20 kb a aproximadamente 30 kb. En una realización, el ácido nucleico insertado es de aproximadamente 30 kb a aproximadamente 40 kb. En una realización, el ácido nucleico insertado es de aproximadamente 40 kb a aproximadamente 50 kb. En una realización, el ácido nucleico insertado es de aproximadamente 60 kb a aproximadamente 70 kb. En una realización, el ácido nucleico insertado es de aproximadamente 80 kb a aproximadamente 90 kb. En una realización, el ácido nucleico insertado es de aproximadamente 90 kb a aproximadamente 100 kb. En una realización, el ácido nucleico insertado es de aproximadamente 100 kb a aproximadamente 110 kb. En una realización, el ácido nucleico insertado es de aproximadamente 120 kb a aproximadamente 130 kb. En una realización, el ácido nucleico insertado es de aproximadamente 130 kb a aproximadamente 140 kb. En una realización, el ácido nucleico insertado es de aproximadamente 140 kb a aproximadamente 150 kb. En una realización, el ácido nucleico insertado es de aproximadamente 150 kb a aproximadamente 160 kb. En una realización, el ácido nucleico insertado es de aproximadamente 160 kb a aproximadamente 170 kb. En una realización, el ácido nucleico insertado es de aproximadamente 170 kb a aproximadamente 180 kb. En una realización, el ácido nucleico insertado es de aproximadamente 180 kb a aproximadamente 190 kb. En una realización, el ácido nucleico insertado es de aproximadamente 190 kb a aproximadamente 200 kb.

En una realización, el ácido nucleico insertado comprende un ácido nucleico flanqueado con secuencias objetivo de recombinación específicas del sitio. En una realización, el ácido nucleico comprende un ácido nucleico genómico. En una realización, el ácido nucleico genómico se deriva de un ratón, un ser humano o una combinación de los mismos. En una realización, las secuencias objetivo de recombinación específicas del sitio se seleccionan del grupo que consiste en loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attp, att, FRT, rox y una combinación de los mismos.

En una realización, el ácido nucleico insertado comprende un alelo condicional. En una realización, el alelo condicional es un alelo multifuncional, como se describe en el documento US 2011/0104799. En una realización, el alelo condicional comprende: (a) una secuencia activadora en orientación sentido con respecto a la transcripción de un gen objetivo, y un casete de selección de fármaco en orientación sentido o antisentido; (b) en orientación antisentido una secuencia de nucleótidos de interés (NSI) y un módulo condicional por inversión (COIN, que utiliza un intrón de división de exones y un módulo de tipo trampa génica invertible; véase, por ejemplo, el documento US 2011/0104799,); y (c) unidades recombinables que se recombinan tras la exposición a una primera recombinasa para formar un alelo condicional que (i) carece de la secuencia activadora y la DSC, y (ii) contiene el NSI en orientación sentido y el COIN en orientación antisentido.

En una realización, el ácido nucleico insertado comprende un casete de selección. En una realización, el casete de selección comprende una secuencia de ácido nucleico que codifica un marcador de selección, en el que la secuencia de ácido nucleico está operativamente unida a un promotor. En una realización, el promotor es activo en una célula procariota. En una realización, el ácido nucleico es activo en células tanto procariotas como eucariotas. En una realización, el casete de selección está flanqueado con secuencias objetivo de recombinación específicas del sitio. En una realización, el marcador de selección se selecciona de neomicina fosfotransferasa (*neo^r*), higromicina B fosfotransferasa (*hyg^r*), puromicina-N-acetiltransferasa (*puro^r*), blasticidina S desaminasa (*bsr^r*), xantina / guanina fosforribosil transferasa (*gpt*), y timidina quinasa del virus del herpes simple (HSV-k), y una combinación de los mismos.

En una realización, el ácido nucleico insertado comprende un gen informador unido operativamente a un promotor, en el que el gen informador codifica una proteína informadora seleccionada del grupo que consiste en LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, proteína fluorescente amarilla mejorada (EYFP), Emerald, proteína fluorescente verde mejorada (EGFP), CyPet, proteína fluorescente cian (CFP), Cerulean, T-Sapphire, luciferasa, fosfatasa alcalina y una combinación de las mismas. En una realización, el gen informador se expresa bajo el control de un promotor inducible. En una realización, el gen informador se expresa bajo el control de un promotor endógeno. En una realización, el gen informador se expresa bajo el control de un promotor exógeno. En una realización, el gen informador se expresa en un tipo de célula específico. En una realización, el gen informador se expresa de una manera específica del tejido. En una realización, el gen informador se expresa en una etapa específica de la etapa de desarrollo.

En una realización, la integración del ácido nucleico insertado en el locus genómico objetivo introduce una o más modificaciones genéticas como se describe en este documento. En una realización, la modificación genética es una eliminación de una secuencia de ácido nucleico endógeno. En una realización, la modificación genética es una adición de una secuencia de ácido nucleico exógeno en el locus genómico objetivo. En una realización, la modificación genética es un reemplazo de una secuencia de ácido nucleico endógeno con una secuencia de ácido nucleico exógeno en el locus genómico objetivo. En una realización, la secuencia de ácido nucleico exógeno no es una secuencia de ácido nucleico de ratón. En una realización, la secuencia de ácido nucleico exógeno es una secuencia de ácido nucleico humano. En una realización, la modificación genética es una desactivación, una eliminación, una inserción, un reemplazo ("activación"), una mutación puntual, un intercambio de dominios, un intercambio de exones, un intercambio de intrones, un intercambio de secuencias reguladoras, un intercambio de genes, o una combinación de los mismos.

En una realización, el ácido nucleico insertado es homólogo a una secuencia de ácido nucleico de ratón. En una realización, el ácido nucleico insertado es un ácido nucleico humano. En una realización, el ácido nucleico insertado es un fragmento de un ácido nucleico genómico. En una realización, el ácido nucleico genómico es un ácido nucleico genómico de ratón, un ácido nucleico genómico humano o una combinación de los mismos. En una realización, el ácido nucleico insertado varía de aproximadamente 5 kb a aproximadamente 200 kb como se describió anteriormente.

En una realización, el ácido nucleico insertado es ortólogo a una secuencia de ácido nucleico de ratón. En una realización, el ácido nucleico insertado es un ácido nucleico humano. En una realización, el ácido nucleico insertado es un fragmento de un ácido nucleico genómico. En una realización, el ácido nucleico genómico es un ácido nucleico genómico de ratón, un ácido nucleico genómico humano o una combinación de los mismos. En una realización, el ácido nucleico insertado varía de aproximadamente 5 kb a aproximadamente 200 kb como se describió anteriormente.

En una realización, el ácido nucleico insertado comprende un locus genómico que codifica una proteína expresada en el sistema nervioso, el sistema esquelético, el sistema digestivo, el sistema circulatorio, el sistema muscular, el sistema respiratorio, el sistema cardiovascular, el sistema linfático, el sistema endocrino, el sistema urinario, el sistema reproductivo o una combinación de los mismos. En una realización, el ácido nucleico insertado comprende un locus genómico que codifica una proteína expresada en una médula ósea o una célula derivada de médula ósea. En una realización, el ácido nucleico insertado comprende un locus genómico que codifica una proteína expresada en una célula de bazo. En una realización, el locus genómico comprende una secuencia de ácido nucleico genómico de ratón, una secuencia de ácido nucleico genómico humano, o una combinación de los mismos. En una realización, el ácido nucleico comprende un locus genómico que codifica una proteína expresada en una célula B. En una realización, el ácido nucleico comprende un locus genómico que codifica una proteína expresada en una célula B.

inmadura. En una realización, el ácido nucleico comprende un locus genómico que codifica una proteína expresada en una célula B madura.

5 En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico genómico que codifica una secuencia de aminoácidos de la región variable de cadena pesada de inmunoglobulina humana.

10 En una realización, la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de la región variable de cadena pesada de inmunoglobulina humana no reordenada ligada operativamente a una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina es una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina de ratón o secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina humana, o una combinación de las mismas. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina se selecciona de un C_H1, una bisagra, un C_H2, un C_H3 y una combinación de los mismos. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada comprende C_H1-bisagra-C_H2-C_H3. En una realización, la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de región variable de cadena pesada de inmunoglobulina humana reordenada operativamente unida a una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina es una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina de ratón o una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina humana o una combinación de las mismas. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina se selecciona de una C_H1, una bisagra, una C_H2, una C_H3 y una combinación de los mismos. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada comprende una C_H1-bisagra-C_H2-C_H3.

25 En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico genómico que codifica una secuencia de aminoácidos de la región variable de cadena ligera de inmunoglobulina humana. En una realización, la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de región variable de cadena ligera λ y/o κ humana no reordenada. En una realización, la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de la región variable de cadena ligera λ y/o κ humana reordenada. En una realización, la secuencia de ácido nucleico de la región variable de cadena ligera λ y/o κ reordenada o no reordenada se une operativamente a una secuencia de ácido nucleico de la región constante de cadena ligera de inmunoglobulina de ratón, de rata o humana seleccionada de una secuencia de ácido nucleico de la región constante de cadena ligera λ y una secuencia de ácido nucleico de la región constante de cadena ligera κ.

35 En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico humano. En una realización, la secuencia de ácido nucleico humano codifica una proteína extracelular. En una realización, la secuencia de ácido nucleico humano codifica un ligando para un receptor. En una realización, el ligando es una citoquina. En una realización, la citoquina es una quimioquina seleccionada de CCL, CXCL, CX3CL y XCL. En una realización, la citoquina es un factor de necrosis tumoral (TNF). En una realización, la citoquina es una interleuquina (IL). En una realización, la interleuquina se selecciona de IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19, IL-20, IL-21, IL-22, IL-23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32, IL-33, IL-34, IL-35, e IL-36. En una realización, la interleuquina es IL-2. En una realización, la secuencia de ácido nucleico genómico humano codifica una proteína citoplasmática. En una realización, la secuencia de ácido nucleico genómico humano codifica una proteína de membrana. En una realización, la proteína de membrana es un receptor. En una realización, el receptor es un receptor de citoquina. En una realización, el receptor de citoquina es un receptor de interleuquina. En una realización, el receptor de interleuquina es un receptor alfa de interleuquina 2. En una realización, el receptor de interleuquina es un receptor beta de interleuquina 2. En una realización, el receptor de interleuquina es un receptor gamma de interleuquina 2. En una realización, la secuencia de ácido nucleico genómico humano codifica una proteína nuclear. En una realización, la proteína nuclear es un receptor nuclear.

50 En una realización, el ácido nucleico insertado comprende una modificación genética en una secuencia codificante. En una realización, la modificación genética comprende una mutación por eliminación de una secuencia codificante. En una realización, la modificación genética comprende una fusión de dos secuencias codificantes endógenas.

55 En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico humano que codifica una proteína humana mutante. En una realización, la proteína humana mutante se caracteriza por una característica de unión alterada, localización alterada, expresión alterada y/o patrón de expresión alterado. En una realización, la secuencia de ácido nucleico humano comprende al menos un alelo de enfermedad humana. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad neurológica. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad cardiovascular. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad renal. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad muscular. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad de la sangre. En una realización, el alelo de la enfermedad humana es un alelo de un gen causante de cáncer. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad del sistema inmune. En una realización, el alelo de la enfermedad humana es un alelo dominante. En una realización, el alelo de la enfermedad

humana es un alelo recesivo. En una realización, el alelo de la enfermedad humana comprende un alelo de polimorfismo de un solo nucleótido (SNP).

En una realización, el ácido nucleico insertado comprende una secuencia reguladora. En una realización, la secuencia reguladora es una secuencia promotora. En una realización, la secuencia reguladora es una secuencia potenciadora. En una realización, la secuencia reguladora es una secuencia de unión a represor transcripcional. En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico humano, en el que la secuencia de ácido nucleico humano comprende una eliminación de una secuencia que no codifica una proteína, pero no comprende una eliminación de una secuencia que codifica una proteína. En una realización, la eliminación de la secuencia que no codifica proteína comprende una eliminación de una secuencia reguladora. En una realización, la eliminación del elemento regulador comprende una eliminación de una secuencia promotora. En una realización, la eliminación del elemento regulador comprende una eliminación de una secuencia potenciadora.

En un aspecto, se proporciona un método para modificar un locus genómico objetivo en una célula madre de embrión de ratón (ES), que comprende: (a) introducir en la célula ES de ratón: (i) una nucleasa de dedo de cinc (ZFN) que realiza un rompimiento de doble cadena en o cerca de un locus genómico objetivo; y (ii) un vector de direccionamiento grande (LTVEC) que comprende un ácido nucleico insertado flanqueado por un brazo de homología secuencia arriba y un brazo de homología secuencia abajo, donde el ácido nucleico insertado varía de 5 kb a 30 kb de longitud, en el que la suma total de los brazos de homología secuencia arriba y secuencia abajo tienen una longitud de al menos 10 kb, en los que los brazos de homología secuencia arriba y secuencia abajo tienen una longitud de entre 5 kb y 200 kb, y en los que el LTVEC varía de 50 kb a 300 kb de longitud; y (b) seleccionar una célula ES de ratón específica que comprende el ácido nucleico insertado en el locus genómico objetivo, donde la integración del ácido nucleico insertado en el locus genómico objetivo da como resultado el reemplazo de una secuencia de ácido nucleico endógeno en el locus genómico objetivo con el ácido nucleico insertado.

En una realización, el ácido nucleico insertado comprende un gen informador unido operativamente a un informador exógeno.

En una realización, la célula ES de ratón es una célula madre pluripotente inducida (iPS). En una realización, la célula pluripotente inducida (iPS) se deriva de un fibroblasto. En una realización, la célula pluripotente es una célula madre hematopoyética (HSC). En una realización, la célula pluripotente es una célula madre neuronal (NSC). En una realización, la célula pluripotente es una célula madre epiblasto. En una realización, la célula pluripotente es una célula progenitora restringida en el desarrollo.

En una realización, la ES de ratón es una célula cancerosa.

En una realización, el locus genómico objetivo se selecciona de un locus FcER1a, un locus TLR4, un locus PRLR, un locus Notch4, un locus Accn2, un locus Adamts5, un locus TRPA1, un locus FcIH1, un locus LRP5, y un locus ERBB4.

En una realización, el locus genómico objetivo comprende una secuencia de ácido nucleico genómico de un ratón.

En una realización, la ZFN se introduce junto con el vector de direccionamiento grande (LTVEC). En una realización, la ZFN se introduce por separado del LTVEC durante un período de tiempo. En una realización, la ZFN se introduce antes de la introducción del LTVEC. En una realización, la ZFN se introduce después de la introducción del LTVEC.

En una realización, el uso combinado del LTVEC con la ZFN da como resultado una eficacia de direccionamiento aumentada en comparación con el uso del LTVEC solo, opcionalmente en donde la eficacia de direccionamiento del LTVEC se incrementa al menos dos veces en comparación con el uso del LTVEC solo. En una realización, cuando el LTVEC se usa junto con la ZFN, la eficacia de direccionamiento del LTVEC se incrementa al menos tres veces en comparación con cuando el LTVEC se usa solo. En una realización, cuando el LTVEC se usa junto con la ZFN, la eficacia de direccionamiento del LTVEC se incrementa al menos cuatro veces en comparación con cuando el LTVEC se usa solo.

En una realización, (I) la ZFN es una construcción de expresión que comprende una secuencia de ácido nucleico que codifica una proteína ZFN, y en la que la secuencia de ácido nucleico está operativamente unida a un promotor activo en la célula; o (II) la ZFN es un ARNm que codifica una proteína ZFN. En una realización, el promotor es un promotor constitutivamente activo. En una realización, el promotor es un promotor inducible. En una realización, el promotor es activo en la célula de mamífero. En una realización, el agente nucleasa es un ARNm que codifica una endonucleasa.

En una realización, cada monómero de ZFN comprende 3 o más dominios de unión a ADN basados en dedos de cinc, en donde cada dominio de unión a ADN basado en dedos de cinc se une a un subsitio de 3 pb. En una realización, la ZFN es una proteína química que comprende un dominio de unión a ADN basado en dedos de cinc operativamente unido a una nucleasa independiente. En una realización, la endonucleasa independiente es una endonucleasa Fok1. En una realización, el agente nucleasa comprende una primera ZFN y una segunda ZFN, en

donde cada una de la primera ZFN y la segunda ZFN está unida operativamente a una nucleasa FokI, en donde la primera y la segunda ZFN reconocen dos secuencias de ADN objetivo contiguas en cada cadena de la secuencia de ADN objetivo separada por un sitio de escisión de aproximadamente 6 pb a aproximadamente 40 pb, y en el que las nucleasas FokI se dimerizan y forman un rompimiento de doble cadena.

En una realización, cada monómero de ZFN reconoce una secuencia objetivo de al menos 9 nucleótidos. En una realización, la secuencia objetivo es de aproximadamente 9 a aproximadamente 12 nucleótidos de longitud. En una realización, la secuencia objetivo es de aproximadamente 12 a aproximadamente 15 nucleótidos de longitud. En una realización, la secuencia objetivo es de aproximadamente 15 a aproximadamente 18 nucleótidos de longitud. En una realización, la secuencia objetivo es de aproximadamente 18 a aproximadamente 21 nucleótidos de longitud.

En una realización, una secuencia de ácido nucleico objetivo de la ZFN está localizada en un intrón, un exón, un promotor, una región que no codifica proteínas, una región reguladora, una región reguladora del promotor o una región potenciadora en el locus genómico objetivo.

La familia LAGLIDADG de nucleasa de asentamiento se selecciona de I-SceI, I-CreI e I-DmI.

En una realización, el LTVEC varía de aproximadamente 50 kb a aproximadamente 300 kb. En una realización, el LTVEC varía de aproximadamente 50 kb a aproximadamente 75 kb. En una realización, el LTVEC varía de aproximadamente 75 kb a aproximadamente 100 kb. En una realización, el LTVEC varía de aproximadamente 100 kb a 125 kb. En una realización, el LTVEC varía de aproximadamente 125 kb a aproximadamente 150 kb. En una realización, el LTVEC varía de aproximadamente 150 kb a aproximadamente 175 kb. En una realización, el LTVEC varía de aproximadamente 175 kb a aproximadamente 200 kb. En una realización, el LTVEC varía de aproximadamente 200 kb a aproximadamente 225 kb. En una realización, el LTVEC varía de aproximadamente 225 kb a aproximadamente 250 kb. En una realización, el LTVEC varía de aproximadamente 250 kb a aproximadamente 275 kb. En una realización, el LTVEC varía de aproximadamente 275 kb a aproximadamente 300 kb.

En una realización, los brazos de homología del LTVEC se derivan de una biblioteca de BAC, una biblioteca de cósmidos o una biblioteca de fagos P1. En una realización, los brazos de homología se derivan de un locus genómico del animal no humano que no puede direccionarse usando un método convencional. En una realización, los brazos de homología se derivan de un ADN sintético.

En una realización, una suma total del brazo de homología secuencia arriba y el brazo de homología secuencia abajo es de al menos 10 kb. En una realización, el brazo de homología secuencia arriba varía de aproximadamente 5 kb a aproximadamente 100 kb. En una realización, el brazo de homología secuencia abajo varía de aproximadamente 5 kb a aproximadamente 100 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 5 kb a aproximadamente 10 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 10 kb a aproximadamente 20 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 20 kb a aproximadamente 30 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 30 kb a aproximadamente 40 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 40 kb a aproximadamente 50 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 50 kb a aproximadamente 60 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 60 kb a aproximadamente 70 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 70 kb a aproximadamente 80 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 80 kb a aproximadamente 90 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 90 kb a aproximadamente 100 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 100 kb a aproximadamente 110 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 110 kb a aproximadamente 120 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 120 kb a aproximadamente 130 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 130 kb a aproximadamente 140 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 140 kb a aproximadamente 150 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 150 kb a aproximadamente 160 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 160 kb a aproximadamente 170 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 170 kb a aproximadamente 180 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 180 kb a aproximadamente 190 kb. En una realización, los brazos de homología secuencia arriba y secuencia abajo varían de aproximadamente 190 kb a aproximadamente 200 kb.

En una realización, el vector de direccionamiento comprende un casete de selección. En una realización, el casete de selección comprende una secuencia de ácido nucleico que codifica un marcador de selección, en el que la secuencia de ácido nucleico está operativamente unida a un promotor. En una realización, el promotor es activo en una célula de mamífero. En una realización, el promotor es activo tanto en células procariotas como en células

eucariotas. En una realización, el marcador de selección se selecciona de neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa (puro^r), blasticidina S desaminasa (bsr^r), xantina / guanina fosforribosil transferasa (gpt), y timidina quinasa del virus del herpes simple (HSV-k), y una combinación de los mismos.

5 En una realización, el ácido nucleico insertado es de aproximadamente 5 kb a aproximadamente 10 kb. En una realización, el ácido nucleico insertado es de aproximadamente 10 kb a aproximadamente 20 kb. En una realización, el ácido nucleico insertado es de aproximadamente 20 kb a aproximadamente 30 kb.

10 En una realización, el ácido nucleico insertado comprende un ácido nucleico flanqueado con secuencias objetivo de recombinación específicas del sitio. En una realización, el ácido nucleico comprende un ácido nucleico genómico. En una realización, el ácido nucleico genómico se deriva de un ratón, un ser humano o una combinación de los mismos. En una realización, las secuencias objetivo de recombinación específicas del sitio se seleccionan del grupo que consiste en loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attp, att, FRT, rox y una
15 combinación de los mismas.

En una realización, el ácido nucleico insertado comprende un alelo condicional. En una realización, el alelo condicional es un alelo multifuncional, como se describe en el documento US 2011/0104799. En una realización, el alelo condicional comprende: (a) una secuencia activadora en orientación sentido con respecto a la transcripción de un gen objetivo, y un casete de selección de fármaco en orientación sentido o antisentido; (b) en orientación antisentido una secuencia de nucleótidos de interés (NSI) y un módulo condicional por inversión (COIN, que utiliza un intrón de división de exones y un módulo de tipo trampa génica invertible; véase, por ejemplo, el documento US 2011/0104799); y (c) unidades recombinables que se recombinan tras la exposición a una primera recombinasa para formar un alelo condicional que (i) carece de la secuencia activadora y la DSC, y (ii) contiene el NSI en orientación
20 sentido y el COIN en orientación antisentido.

En una realización, el ácido nucleico insertado comprende un casete de selección. En una realización, el casete de selección comprende una secuencia de ácido nucleico que codifica un marcador de selección, en el que la secuencia de ácido nucleico está operativamente unida a un promotor. En una realización, el promotor es activo en una célula de mamífero. En una realización, el ácido nucleico es activo en una célula eucariota. En una realización, el casete de selección está flanqueado con secuencias objetivo de recombinación específicas del sitio. En una realización, el marcador de selección se selecciona de neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa (puro^r), blasticidina S desaminasa (bsr^r), xantina/guanina fosforribosil transferasa (gpt), y timidina quinasa del virus del herpes simple (HSV-k), y una combinación de los
30 mismos.

En una realización, el ácido nucleico insertado comprende un gen informador unido operativamente a un promotor, en el que el gen informador codifica una proteína informadora seleccionada del grupo que consiste en LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, proteína fluorescente amarilla mejorada (EYFP), Emerald, proteína fluorescente verde mejorada (EGFP), CyPet, proteína fluorescente cian (CFP), Cerulean, T-Sapphire, luciferasa, fosfatasa alcalina y una combinación de las mismas. En una realización, el gen informador se expresa bajo el control de un promotor inducible. En una realización, el gen informador se expresa bajo el control de un promotor endógeno. En una realización, el gen informador se expresa bajo el control de un promotor exógeno. En una realización, el gen informador se expresa en un tipo de célula específico. En una
40 realización, el gen informador se expresa de una manera específica de tejido. En una realización, el gen informador se expresa en una forma específica de la etapa de desarrollo.

En una realización, la integración del ácido nucleico insertado en el locus genómico objetivo introduce una o más modificaciones genéticas como se describe en este documento. En una realización, la modificación genética es una eliminación de una secuencia de ácido nucleico endógena. En una realización, la modificación genética es una adición de una secuencia de ácido nucleico exógeno en el locus genómico objetivo. En una realización, la modificación genética es un reemplazo de una secuencia de ácido nucleico endógeno con una secuencia de ácido nucleico exógeno en el locus genómico objetivo. En una realización, la secuencia de ácido nucleico exógeno no es una secuencia de ácido nucleico de ratón. En una realización, la secuencia de ácido nucleico exógeno es una
50 secuencia de ácido nucleico humano. En una realización, la integración del ácido nucleico insertado en el locus genómico objetivo da como resultado una desactivación, una eliminación, una inserción, un reemplazo ("activación"), una mutación puntual, un intercambio de dominios, un intercambio de exones, un intercambio de intrones, un intercambio de secuencias reguladoras, un intercambio de genes, o una combinación de los mismos.

60 En una realización, el ácido nucleico insertado es homólogo a una secuencia de ácido nucleico de ratón. En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico homóloga a la secuencia de ácido nucleico reemplazada en el locus genómico objetivo de la célula ES de ratón. En una realización, el ácido nucleico insertado es un ácido nucleico humano. En una realización, el ácido nucleico insertado es un fragmento de un ácido nucleico genómico. En una realización, el ácido nucleico genómico es un ácido nucleico genómico de ratón, un ácido nucleico genómico humano o una combinación de los mismos. En una realización, el ácido nucleico insertado varía de aproximadamente 5 kb a aproximadamente 200 kb como se describió anteriormente.

En una realización, el ácido nucleico insertado es ortólogo a una secuencia de ácido nucleico de ratón. En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico ortóloga a la secuencia de ácido nucleico reemplazada en el locus genómico objetivo de la célula ES de ratón. En una realización, el ácido nucleico insertado es un ácido nucleico humano. En una realización, el ácido nucleico insertado es un fragmento de un ácido nucleico genómico. En una realización, el ácido nucleico genómico es un ácido nucleico genómico de ratón, un ácido nucleico genómico humano o una combinación de los mismos. En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico de una especie diferente.

En una realización, el ácido nucleico insertado comprende un locus genómico que codifica una proteína expresada en el sistema nervioso, el sistema esquelético, el sistema digestivo, el sistema circulatorio, el sistema muscular, el sistema respiratorio, el sistema cardiovascular, el sistema linfático, el sistema endocrino, el sistema urinario, el sistema reproductivo o una combinación de los mismos. En una realización, el ácido nucleico insertado comprende un locus genómico que codifica una proteína expresada en una médula ósea o una célula derivada de médula ósea. En una realización, el ácido nucleico insertado comprende un locus genómico que codifica una proteína expresada en una célula de bazo. En una realización, el locus genómico comprende una secuencia de ácido nucleico genómico de ratón, una secuencia de ácido nucleico genómico humano, o una combinación de los mismos. En una realización, el ácido nucleico comprende un locus genómico que codifica una proteína expresada en una célula B. En una realización, el ácido nucleico comprende un locus genómico que codifica una proteína expresada en una célula B inmadura. En una realización, el ácido nucleico comprende un locus genómico que codifica una proteína expresada en una célula B madura.

En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico genómico que codifica una secuencia de aminoácidos de la región variable de cadena pesada de inmunoglobulina humana.

En una realización, la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de la región variable de cadena pesada de inmunoglobulina humana no reordenada funcionalmente unida a una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina es una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina de ratón o secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina humana, o una combinación de las mismas. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina se selecciona de un C_H1, una bisagra, un C_H2, un C_H3 y una combinación de las mismas. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada comprende un C_H1-bisagra-C_H2-C_H3. En una realización, la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de región variable de cadena pesada de inmunoglobulina humana reordenada operativamente unida a una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina es una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina de ratón o una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina humana o una combinación de las mismas. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina se selecciona de un C_H1, una bisagra, un C_H2, un C_H3 y una combinación de las mismas. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada comprende C_H1-bisagra-C_H2-C_H3.

En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico genómico que codifica una secuencia de aminoácidos de la región variable de cadena ligera de inmunoglobulina humana. En una realización, la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de región variable de cadena ligera λ y/o κ humana no reordenada. En una realización, la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de región variable de cadena ligera λ o κ humana reordenada operablemente unida a una secuencia de ácido nucleico de región constante de cadena ligera de inmunoglobulina humana o de ratón seleccionada de una secuencia de ácido nucleico de región constante de cadena ligera λ y una secuencia de ácido nucleico de la región constante de cadena ligera κ. En una realización, la secuencia de ácido nucleico de región variable de cadena ligera λ o κ no reordenada o reordenada se une operativamente a una secuencia de ácido nucleico de región constante de cadena ligera de inmunoglobulina humana o de ratón seleccionada de una secuencia de ácido nucleico de región constante de cadena ligera λ y una secuencia de ácido nucleico de región constante de cadena ligera κ.

En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico humano. En una realización, la secuencia de ácido nucleico humano codifica una proteína extracelular. En una realización, la secuencia de ácido nucleico humano codifica un ligando para un receptor. En una realización, el ligando es una citoquina. En una realización, la citoquina es una quimioquina seleccionada de CCL, CXCL, CX3CL y XCL. En una realización, la citoquina es un factor de necrosis tumoral (TNF). En una realización, la citoquina es una interleuquina (IL). En una realización, la interleuquina se selecciona de IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19, IL-20, IL-21, IL-22, IL-23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32, IL-33, IL-34, IL-35, y IL-36. En una realización, la interleuquina es IL-2. En una realización, la secuencia de ácido nucleico genómico humano codifica una proteína citoplasmática. En una realización, la

secuencia de ácido nucleico genómico humano codifica una proteína de membrana. En una realización, la proteína de membrana es un receptor. En una realización, el receptor es un receptor de citoquina. En una realización, el receptor de citoquina es un receptor de interleuquina. En una realización, el receptor de interleuquina es un receptor alfa de interleuquina 2. En una realización, el receptor de interleuquina es un receptor beta de interleuquina 2. En una realización, el receptor de interleuquina es un receptor gamma de interleuquina 2. En una realización, la secuencia de ácido nucleico genómico humano codifica una proteína nuclear. En una realización, la proteína nuclear es un receptor nuclear.

En una realización, el ácido nucleico insertado comprende una modificación genética en una secuencia codificante. En una realización, la modificación genética comprende una mutación por eliminación de una secuencia codificante. En una realización, la modificación genética comprende una fusión de dos secuencias codificantes endógenas.

En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico humano que codifica una proteína humana mutante. En una realización, la proteína humana mutante se caracteriza por una característica de unión alterada, localización alterada, expresión alterada y/o patrón de expresión alterado. En una realización, la secuencia de ácido nucleico humano comprende al menos un alelo de enfermedad humana. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad neurológica. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad cardiovascular. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad renal. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad muscular. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad de la sangre. En una realización, el alelo de la enfermedad humana es un alelo de un gen causante de cáncer. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad del sistema inmune. En una realización, el alelo de la enfermedad humana es un alelo dominante. En una realización, el alelo de la enfermedad humana es un alelo recesivo. En una realización, el alelo de la enfermedad humana comprende un alelo de polimorfismo de un solo nucleótido (SNP).

En una realización, el ácido nucleico insertado comprende una secuencia reguladora. En una realización, la secuencia reguladora es una secuencia promotora. En una realización, la secuencia reguladora es una secuencia potenciadora. En una realización, la secuencia reguladora es una secuencia de unión a represor transcripcional. En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico humano, en el que la secuencia de ácido nucleico humano comprende una eliminación de una secuencia que codifica una proteína, pero no comprende una eliminación de una secuencia que codifica una proteína. En una realización, la eliminación de la secuencia de codificación no proteica comprende una eliminación de una secuencia reguladora. En una realización, la eliminación del elemento regulador comprende una eliminación de una secuencia promotora. En una realización, la eliminación del elemento regulador comprende una eliminación de una secuencia potenciadora. En una realización, la etapa de selección (b) se lleva a cabo mediante una modificación del ensayo de alelos (MOA). En una realización, el método de acuerdo con este aspecto comprende adicionalmente (c) introducir la célula ES de ratón modificada en un embrión en estadio de premórula; y (d) incubar el embrión hasta la etapa de blastocisto, opcionalmente en el que el método comprende producir un ratón F0 a partir del embrión implantado en una madre sustituta; y (e) identificar un ratón que porta el locus genómico genéticamente modificado.

La divulgación se refiere a una célula de mamífero hecha con un método como se describe en la presente memoria. En una realización, la célula de mamífero comprende un ácido nucleico insertado que comprende una o más modificaciones genéticas como se describe en este documento en un locus genómico objetivo.

En una realización, la célula de mamífero es una célula pluripotente. En una realización, la célula pluripotente es una célula madre embrionaria (ES). En una realización, la célula pluripotente es una célula madre pluripotente inducida (iPS). En una realización, la célula pluripotente inducida (iPS) se deriva de un fibroblasto. En una realización, la célula pluripotente inducida (iPS) se deriva de un fibroblasto humano. En una realización, la célula pluripotente es una célula madre hematopoyética (HSC). En una realización, la célula pluripotente es una célula madre neuronal (NSC). En una realización, la célula pluripotente es una célula madre de epiblasto. En una realización, la célula pluripotente es una célula progenitora restringida en el desarrollo.

En una realización, la célula pluripotente es una célula pluripotente de ratón. En una realización, la célula pluripotente es una célula madre embrionaria (ES) de ratón.

En una realización, la célula de mamífero es una célula inmortalizada de ratón o de rata. En una realización, la célula de mamífero es una célula inmortalizada humana. En una realización, la célula de mamífero es un fibroblasto humano. En una realización, la célula de mamífero es una célula cancerosa. En una realización, la célula de mamífero es una célula cancerosa humana.

En una realización, el locus genómico objetivo se selecciona a partir de un locus FcER1a, un locus TLR4, un locus PRLR, un locus Notch4, un locus Accn2, un locus ADAMTS5, un locus TRPA1, un locus FolH1, un locus de LRP5, y un locus ERBB4.

En una realización, el locus genómico objetivo comprende una o más modificaciones genéticas como se describe en

este documento. En una realización, la modificación genética es una eliminación de una secuencia de ácido nucleico endógena. En una realización, la modificación genética es una adición de una secuencia de ácido nucleico exógeno en el locus genómico objetivo. En una realización, la modificación genética es un reemplazo de una secuencia de ácido nucleico endógeno con una secuencia de ácido nucleico exógeno en el locus genómico objetivo. En una
 5 realización, la secuencia de ácido nucleico exógeno no es una secuencia de ácido nucleico de ratón. En una realización, la secuencia de ácido nucleico exógeno es una secuencia de ácido nucleico humano. En una realización, el locus genómico objetivo comprende una modificación seleccionada de una desactivación, una eliminación, una inserción, un reemplazo ("activación"), una mutación puntual, un intercambio de dominios, un intercambio de exones, un intercambio de intrones, un intercambio de secuencias reguladoras, un intercambio de
 10 genes, o una combinación de los mismos.

En una realización, el locus genómico objetivo comprende un ácido nucleico insertado que es homólogo a una secuencia de ácido nucleico de ratón. En una realización, el ácido nucleico insertado es un ácido nucleico humano. En una realización, el ácido nucleico insertado es un fragmento de un ácido nucleico genómico. En una realización,
 15 el ácido nucleico genómico es un ácido nucleico genómico de ratón, un ácido nucleico genómico humano o una combinación de los mismos. En una realización, el ácido nucleico insertado varía de aproximadamente 5 kb a aproximadamente 200 kb como se describió anteriormente.

En una realización, el locus genómico objetivo comprende un ácido nucleico insertado que es ortólogo a una secuencia de ácido nucleico de ratón. En una realización, el ácido nucleico insertado es un ácido nucleico humano. En una realización, el ácido nucleico insertado es un fragmento de un ácido nucleico genómico. En una realización,
 20 el ácido nucleico genómico es un ácido nucleico genómico de ratón, un ácido nucleico genómico humano o una combinación de los mismos. En una realización, el ácido nucleico insertado varía de aproximadamente 5 kb a aproximadamente 200 kb como se describió anteriormente.

En una realización, el locus genómico objetivo comprende un alelo condicional. En una realización, el alelo condicional es un alelo multifuncional, como se describe en el documento US 2011/0104799. En una realización, el alelo condicional comprende: (a) una secuencia activadora en orientación sentido con respecto a la transcripción de un gen objetivo, y un casete de selección de fármaco en orientación sentido o antisentido; (b) en orientación
 25 antisentido una secuencia de nucleótidos de interés (NSI) y un módulo condicional por inversión (COIN, que utiliza un intrón de división de exones y un módulo de tipo trampa génica invertible; véase, por ejemplo, el documento US 2011/0104799;); y (c) unidades recombinables que se recombinan tras la exposición a una primera recombinasa para formar un alelo condicional que (i) carece de la secuencia activadora y la DSC, y (ii) contiene el NSI en orientación sentido y el COIN en orientación antisentido.

En una realización, el ácido nucleico insertado comprende un gen informador unido operativamente a un promotor, en el que el gen informador codifica una proteína informadora seleccionada del grupo que consiste de LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, proteína fluorescente amarilla (EYFP), Emerald, proteína fluorescente verde potenciada (EGFP), CyPet, proteína fluorescente cian (CFP),
 35 Cerulean, T-Sapphire, luciferasa, fosfatasa alcalina, y una combinación de las mismas. En una realización, el gen informador se expresa bajo el control de un promotor inducible. En una realización, el gen informador se expresa bajo el control de un promotor endógeno. En una realización, el gen informador se expresa bajo el control de un promotor exógeno. En una realización, el gen informador se expresa en un tipo de célula específico. En una realización, el gen informador se expresa de una manera específica del tejido. En una realización, el gen informador se expresa en una forma específica de la etapa de desarrollo.

En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico humano que codifica una proteína expresada en el sistema nervioso, el sistema esquelético, el sistema digestivo, el sistema circulatorio, el sistema muscular, el sistema respiratorio, el sistema cardiovascular, el sistema linfático, el sistema endocrino, el
 40 sistema urinario, el sistema reproductivo o una combinación de los mismos. En una realización, la secuencia de ácido nucleico humano codifica una proteína expresada en una médula ósea o una célula derivada de médula ósea. En una realización, el genoma de la célula ES de ratón comprende un locus genómico humano que codifica una proteína expresada en una célula de bazo. En una realización, el ácido nucleico humano codifica una proteína expresada en una célula B. En una realización, el ácido nucleico humano codifica una proteína expresada en una
 45 célula B inmadura. En una realización, el ácido nucleico humano codifica una proteína expresada en una célula B madura.

En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico humano que codifica una secuencia de aminoácidos de la región variable de cadena pesada de inmunoglobulina humana. En una
 50 realización, la secuencia de ácido nucleico humano comprende una secuencia de ácido nucleico de región variable de cadena pesada de inmunoglobulina humana no reordenada. En una realización, la secuencia de ácido nucleico de la región variable de cadena pesada de inmunoglobulina no reordenada está operativamente unida a una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina de ratón, a una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina humana. En una realización, la
 55 secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina se selecciona de un C_H1, una bisagra, un C_H2, un C_H3 y una combinación de las mismas. En una realización, la secuencia de ácido nucleico

de la región constante de cadena pesada comprende una C_H1-bisagra-C_H2-C_H3. En una realización, la secuencia de ácido nucleico humano comprende una secuencia de ácido nucleico de región variable de cadena pesada de inmunoglobulina humana reordenada. En una realización, la secuencia de ácido nucleico de la región variable de cadena pesada de inmunoglobulina humana está unida operativamente a una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina de ratón, a una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina humana. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina se selecciona de un C_H1, una bisagra, un C_H2, un C_H3 y una combinación de las mismas. En una realización, la secuencia de ácido nucleico de la región constante de cadena pesada comprende una C_H1-bisagra-C_H2-C_H3.

En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico humano que codifica una secuencia de aminoácidos de la región variable de cadena ligera de inmunoglobulina humana. En una realización, la secuencia de ácido nucleico humano comprende una secuencia de ácido nucleico de región variable de cadena ligera λ y/o κ humana no reordenada. En una realización, la secuencia de ácido nucleico humano comprende una secuencia de ácido nucleico de la región variable de cadena ligera λ y/o κ humana reordenada. En una realización, la secuencia de ácido nucleico de la región variable de cadena ligera λ y/o κ no reordenada está unida operativamente a una secuencia de ácido nucleico de región constante de cadena ligera de inmunoglobulina humana o de ratón seleccionada de una secuencia de ácido nucleico de la región constante de cadena ligera λ y una secuencia de ácido nucleico de la región constante de cadena ligera κ.

En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico humano. En una realización, la secuencia de ácido nucleico humano codifica una proteína extracelular. En una realización, la secuencia de ácido nucleico humano codifica un ligando para un receptor. En una realización, el ligando es una citoquina. En una realización, la citoquina es una quimioquina seleccionada de CCL, CXCL, CX3CL y XCL. En una realización, la citoquina es un factor de necrosis tumoral (TNF). En una realización, la citoquina es una interleuquina (IL). En una realización, la interleuquina se selecciona de IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19, IL-20, IL-21, IL-22, IL-23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32, IL-33, IL-34, IL-35, y IL-36. En una realización, la interleuquina es IL-2. En una realización, la secuencia de ácido nucleico genómico humano codifica una proteína citoplasmática. En una realización, la secuencia de ácido nucleico genómico humano codifica una proteína de membrana. En una realización, la proteína de membrana es un receptor. En una realización, el receptor es un receptor de citoquina. En una realización, el receptor de citoquina es un receptor de interleuquina. En una realización, el receptor de interleuquina es un receptor alfa de interleuquina 2. En una realización, el receptor de interleuquina es un receptor beta de interleuquina 2. En una realización, el receptor de interleuquina es un receptor gamma de interleuquina 2. En una realización, la secuencia de ácido nucleico genómico humano codifica una proteína nuclear. En una realización, la proteína nuclear es un receptor nuclear.

En una realización, el ácido nucleico insertado comprende una modificación genética en una secuencia codificante. En una realización, la modificación genética comprende una mutación por eliminación de una secuencia codificante. En una realización, la modificación genética comprende una fusión de dos secuencias codificantes endógenas.

En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico humano que codifica una proteína humana mutante. En una realización, la proteína humana mutante se caracteriza por una característica de unión alterada, localización alterada, expresión alterada y/o patrón de expresión alterado. En una realización, la secuencia de ácido nucleico humano comprende al menos un alelo de enfermedad humana. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad neurológica. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad cardiovascular. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad renal. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad muscular. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad de la sangre. En una realización, el alelo de la enfermedad humana es un alelo de un gen causante de cáncer. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad del sistema inmune. En una realización, el alelo de la enfermedad humana es un alelo dominante. En una realización, el alelo de la enfermedad humana es un alelo recesivo. En una realización, el alelo de la enfermedad humana comprende un alelo de polimorfismo de un solo nucleótido (SNP).

En una realización, el ácido nucleico insertado comprende una secuencia reguladora. En una realización, la secuencia reguladora es una secuencia promotora. En una realización, la secuencia reguladora es una secuencia potenciadora. En una realización, la secuencia reguladora es una secuencia de unión a represor transcripcional. En una realización, el ácido nucleico insertado comprende una secuencia de ácido nucleico humano, en el que la secuencia de ácido nucleico humano comprende una eliminación de una secuencia que no codifica una proteína, pero no comprende una eliminación de una secuencia que codifica una proteína. En una realización, la eliminación de la secuencia que no codifica proteína comprende una eliminación de una secuencia reguladora. En una realización, la eliminación del elemento regulador comprende una eliminación de una secuencia promotora. En una realización, la eliminación del elemento regulador comprende una eliminación de una secuencia potenciadora.

La divulgación se refiere a un método para fabricar un animal no humano que comprende en su línea germinal una o

más modificaciones genéticas como se describe en la presente invención, que comprende:

(a) modificar un locus genómico de interés de un animal no humano en una célula procariota que emplea un vector de direccionamiento grande (LTVEC) y un agente nucleasa que genera una ruptura de cadena simple o doble en o cerca del locus genómico de interés, en donde el LTVEC comprende un ácido nucleico insertado flanqueado por brazos de homología secuencia arriba y secuencia abajo, y la célula procariota es capaz de expresar una recombinasa;

(b) seleccionar una célula procariota modificada que comprende una LTVEC genéticamente modificada;

(c) aislar el LTVEC genéticamente modificado;

(d) introducir el LTVEC genéticamente modificado en una célula pluripotente del animal no humano para generar una célula pluripotente genéticamente modificada que comprende el ácido nucleico insertado en el locus genómico de interés;

(e) seleccionar la célula pluripotente genéticamente modificada;

(f) introducir la célula pluripotente genéticamente modificada en un embrión huésped del animal no humano en una etapa de premórula; y

(g) implantar el embrión huésped que comprende la célula pluripotente genéticamente modificada en una madre sustituta para generar una generación F0 derivada de la célula pluripotente genéticamente modificada.

En una realización, el animal no humano es un mamífero. En una realización, el mamífero es un roedor. En una realización, el roedor se selecciona de un ratón, una rata y un hámster. En una realización, el animal no humano es un ratón, y la célula pluripotente es una célula ES de ratón. En una realización, el animal no humano es una rata, y la célula pluripotente es una célula ES de rata.

En una realización, el locus genómico objetivo comprende una o más modificaciones genéticas como se describe en este documento.

En una realización, la etapa de aislamiento (c) comprende además (c)' la linealización del LTVEC genéticamente modificado.

En una realización, la introducción de la etapa (d) comprende además (d)' introducir un agente nucleasa como se describe en la presente memoria en la célula pluripotente. En una realización, el agente nucleasa es una nucleasa de dedo de cinc (ZFN). En una realización, el agente nucleasa es una nucleasa efectora similar al activador de la transcripción (TALEN).

En una realización, las etapas de selección (b) y/o (e) se llevan a cabo aplicando un agente seleccionable como se describe en la presente memoria a la célula procariota o a la célula pluripotente.

En una realización, las etapas de selección (b) y/o (e) se llevan a cabo a través de una modificación del ensayo de alelos (MOA) como se describe en este documento.
Breve descripción de los dibujos

La FIG. 1 ilustra el sitio de escisión de nucleasa de dedo de cinc (ZFN) en el gen *Il2rg* de ratón. La figura no fue escalada proporcionalmente. Las secuencias encuadradas en el panel inferior representan secuencias objetivo de ZFN.

Descripción detallada

Esta invención no se limita a métodos particulares, y a condiciones experimentales descritas, ya que tales métodos y condiciones pueden variar. También debe entenderse que la terminología utilizada en este documento tiene el propósito de describir realizaciones particulares solamente, y no pretende ser limitativa, ya que el alcance de la presente invención se define por las reivindicaciones.

A menos que se defina lo contrario, todos los términos y frases usados en la presente invención incluyen los significados que los términos y frases han alcanzado en la técnica, a menos que lo contrario esté claramente indicado o claramente evidente a partir del contexto en el que se usa el término o frase. Aunque puede usarse cualquier método y material similar o equivalente a los descritos en la presente memoria en la práctica o prueba de la presente invención, ahora se describen métodos y materiales particulares.

Definiciones

El término "célula madre embrionaria" o "célula ES" como se usa en este documento incluye una célula totipotente o pluripotente derivada de embrión que es capaz de proliferación indiferenciada *in vitro*, y es capaz de contribuir a cualquier tejido del embrión en desarrollo tras la introducción en un embrión. El término "célula pluripotente" como se usa en el presente documento incluye una célula indiferenciada que posee la capacidad de desarrollarse en más de un tipo de célula diferenciada.

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

La frase "cadena pesada" o "cadena pesada de inmunoglobulina" incluye una secuencia de cadena pesada de inmunoglobulina, que incluye una secuencia de región constante de cadena pesada de inmunoglobulina, de cualquier organismo. Los dominios variables de cadena pesada incluyen tres CDR de cadena pesada y cuatro regiones marco (FR), a menos que se especifique lo contrario. Los fragmentos de cadenas pesadas incluyen CDR, CDR y FR, y combinaciones de los mismos. Una cadena pesada típica tiene, después del dominio variable (desde el N-terminal hasta el C-terminal), un dominio C_H1, una bisagra, un dominio C_H2 y un dominio C_H3. Un fragmento funcional de una cadena pesada incluye un fragmento que es capaz de reconocer específicamente un epítipo (por ejemplo, reconocer el epítipo con un K_D en el rango micromolar, nanomolar o picomolar), que es capaz de expresarse y secretarse desde una célula, y que comprende al menos una CDR. Los dominios variables de cadena pesada están codificados por la secuencia de nucleótidos de la región variable, que generalmente comprende los segmentos V_H, D_H y J_H derivados de un repertorio de segmentos V_H, D_H y J_H presentes en la línea germinal. Las secuencias, ubicaciones y nomenclatura para segmentos de cadena pesada V, D y J para diversos organismos se pueden encontrar en la base de datos IMGT, a la que se puede acceder a través de internet en la red mundial (www) en la URL "imgt.org".

La frase "cadena ligera" incluye una secuencia de cadena ligera de inmunoglobulina de cualquier organismo, y a menos que se especifique lo contrario incluye cadenas ligeras kappa (κ) y lambda (λ) humanas y una VpreB, así como cadenas ligeras sustitutivas. Los dominios variables de cadena ligera típicamente incluyen tres CDR de cadena ligera y cuatro FR, a menos que se especifique lo contrario. Generalmente, una cadena ligera de longitud completa incluye, desde el extremo amino hasta el extremo carboxilo, un dominio variable que incluye FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4, y una secuencia de aminoácidos de la región constante de cadena ligera. Los dominios variables de cadena ligera están codificados por la secuencia de nucleótidos de la región variable de cadena ligera, que generalmente comprende segmentos del gen V_L de cadena ligera y J_L de cadena ligera, derivados de un repertorio de segmentos de genes V y J de cadena ligera presentes en la línea germinal. Secuencias, ubicaciones y nomenclatura para los segmentos del gen V y J de la cadena ligera para varios organismos se pueden encontrar en la base de datos IMGT, a la que se puede acceder a través de Internet en la red mundial (www) en la URL "imgt.org". Las cadenas ligeras incluyen aquellas, por ejemplo, que no se unen selectivamente a un primer o a un segundo epítopos selectivamente unidos por la proteína de unión a epítopos en la que aparecen. Las cadenas ligeras también incluyen aquellas que se unen y reconocen, o ayudan a la cadena pesada con unión y reconocimiento, uno o más epítopos unidos selectivamente por la proteína de unión al epítipo en la que aparecen.

El término "ácido nucleico homólogo" como se usa en la presente memoria incluye una secuencia de ácido nucleico que es idéntica o sustancialmente similar a una secuencia de referencia conocida. En una realización, el término "ácido nucleico homólogo" se usa para caracterizar una secuencia de ADN o ARN que tiene 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 el 98%, al menos el 99%, o incluso el 100% de identidad con una secuencia de referencia conocida.

El término "ácido nucleico ortólogo" como se usa en la presente memoria incluye una secuencia de ácido nucleico de una especie que es funcionalmente equivalente a una secuencia de referencia conocida en otra especie.

El término "vector de direccionamiento grande" o "LTVEC" como se usa en el presente documento incluye vectores de direccionamiento grandes para células eucariotas que se derivan de fragmentos de ácido nucleico genómico clonado mayores que los utilizados típicamente por otros enfoques destinados a realizar direccionamiento homólogo en células eucariotas. El tamaño del LTVEC es demasiado grande para permitir el cribado de eventos de direccionamiento mediante ensayos convencionales, por ejemplo, transferencia Southern y PCR de largo alcance (por ejemplo, 1 kb-5 kb). Los ejemplos del LTVEC incluyen, pero no se limitan a, vectores derivados de un cromosoma artificial bacteriano (BAC) y un cromosoma artificial de levadura (YAC).

El término "modificación de alelo" o "MOA" incluye la modificación de la secuencia de ADN exacta de un alelo de un gen o genes o locus (loci) cromosómico en un genoma. Los ejemplos de "modificación del alelo (MOA)" incluyen, pero no se limitan a eliminaciones, sustituciones o inserciones de tan solo un nucleótido o eliminaciones de muchas kilobases que abarcan un gen o genes o locus (loci) cromosómico de interés, así como todas y cada una de las modificaciones posibles entre estos dos extremos.

El término "nucleasa" como se usa en la presente memoria incluye un agente que induce una ruptura en una secuencia de ácido nucleico, por ejemplo, una ruptura de cadena sencilla o doble en una secuencia de ADN de doble cadena. Las nucleasas incluyen aquellas que se unen a una secuencia preseleccionada o específica y se

cortan en o cerca de la secuencia preseleccionada o específica, por ejemplo, nucleasas de dedos de zinc modificadas por ingeniería genética y nucleasas efectoras TAL modificadas por ingeniería genética. Las nucleasas no están limitadas a ZFN y a la nucleasa efectora TAL, sino que pueden ser cualquier nucleasa adecuada para usar con un LTVEC para lograr una eficacia de direccionamiento mejorada. Los ejemplos no limitantes incluyen otras nucleasas basadas en dedos de zinc y meganucleasas modificadas por ingeniería genética que cortan en secuencias preseleccionadas o específicas.

Las nucleasas efectoras TAL adecuadas para su uso con los procedimientos descritos en la presente memoria incluyen cualquier nucleasa TAL conocida en la técnica. Ejemplos de nucleasas TAL adecuadas, y métodos para preparar nucleasas TAL adecuadas, se divulgan, por ejemplo, en las solicitudes de patente estadounidenses Nos 2011/0239315 A1, 2011/0269234 A1, 2011/0145940 A1, 2003/0232410 A1, 2005/0208489 A1, 2005/0026157 A1, 2005/0064474 A1, 2006/0188987 A1 y 2006/0063231 A1. En diversas realizaciones, las nucleasas efectoras TAL se modifican por ingeniería genética para que corten en o cerca de una secuencia de ácido nucleico objetivo, por ejemplo, en un genoma de interés, en el que la secuencia de ácido nucleico objetivo está en o cerca de una secuencia para ser modificada por un LTVEC. Las nucleasas efectoras TAL son proteínas que comprenden un dominio de endonucleasa y uno o más dominios de unión al ADN del efector TAL, en donde uno o más dominios de unión al ADN del efector TAL comprenden una secuencia que reconoce una secuencia de ácido nucleico preseleccionada o específica. Las nucleasas TAL adecuadas para su uso con los métodos descritos en la presente memoria incluyen aquellas que están diseñadas específicamente para unirse en o cerca de secuencias de ácido nucleico objetivo para ser modificadas por LTVEC como se describe en este documento.

La frase "operativamente unida" incluye una relación en la que los componentes operativamente enlazados funcionan de la manera prevista. En un caso, una secuencia de ácido nucleico que codifica una proteína se puede unir operativamente a secuencias reguladoras (por ejemplo, secuencia promotora, potenciadora, silenciadora, etc.) para retener la regulación transcripcional apropiada. En un caso, una secuencia de ácido nucleico de una región variable de inmunoglobulina (o segmentos V(D)J) puede unirse operativamente a una secuencia de ácido nucleico de una región constante de inmunoglobulina para permitir la recombinación apropiada entre las secuencias en una secuencia de cadena pesada o ligera de inmunoglobulina.

El término "promotor" y "elemento regulador del promotor", y similares, como se usan en este documento incluyen un elemento de secuencia de nucleótidos dentro de un fragmento de ácido nucleico o gen que controla la expresión de ese gen.

El término "sitio de recombinación" como se usa en el presente documento incluye una secuencia de nucleótidos que es reconocida por una recombinasa específica del sitio y que puede servir como un sustrato para un evento de recombinación.

El término "recombinasa específica del sitio" como se usa en este documento incluye un grupo de enzimas que pueden facilitar la recombinación entre "sitios de recombinación" donde los dos sitios de recombinación se separan físicamente dentro de una molécula de ácido nucleico individual o en moléculas de ácido nucleico separadas. Los ejemplos de "recombinasa específica del sitio" incluyen, pero sin limitación, las recombinasas Cre, Flp y Dre.

Modificación de loci genómicos usando un LTVEC y un agente nucleasa

Aunque se han realizado progresos en el direccionamiento de varios loci genómicos de animales no humanos, todavía quedan muchos loci genómicos o tipos de células que no pueden ser dirigidos con estrategias de direccionamiento convencionales. Las razones de esta falla pueden variar, pero, tal como se usan en el presente documento, incluyen loci o células que o bien no se seleccionan con éxito en absoluto, o bien están dirigidas de forma inapropiada o con una eficacia significativamente baja mediante métodos de direccionamiento convencionales. Los métodos de direccionamiento convencionales incluyen direccionamiento usando recombinación homóloga empleando vectores de direccionamiento convencionales. Los loci que son difíciles de dirigir incluyen loci que no pueden ser dirigidos incluso con LTVEC solos, es decir, en ausencia de ayuda en la forma de un rompimiento recombinogénico de cadena sencilla o doble, o que están dirigidos con LTVEC de manera incorrecta o con una baja eficiencia en la ausencia del rompimiento recombinogénico de cadena sencilla o doble.

Se proporcionan composiciones y métodos para direccionamiento de secuencias de ácido nucleico que emplean un agente nucleasa capaz de formar un rompimiento recombinogénico de cadena sencilla o doble, junto con un vector de direccionamiento grande, o LTVEC, en el que la secuencia de ácido nucleico direccionada (o una secuencia cerca de la secuencia de ácido nucleico direccionada) se modifica por el LTVEC. Las composiciones y métodos son útiles para modificar secuencias de ácido nucleico genómicas que son difíciles o imposibles de modificar usando estrategias de direccionamiento convencionales, incluso cuando se usan solamente LTVEC.

Se proporcionan composiciones y métodos para emplear un LTVEC para realizar una modificación en un ácido nucleico objetivo, por ejemplo, un locus en un genoma, en el que el ácido nucleico objetivo comprende una secuencia objetivo que se va a modificar mediante una secuencia del LTVEC (por recombinación homóloga del ácido nucleico objetivo con el LTVEC), en el que se realiza una ruptura de cadena sencilla o doble en el ácido

nucleico objetivo en o cerca de la secuencia objetivo.

La presencia de una ruptura de cadena sencilla o doble en o cerca del ácido nucleico objetivo, en diversas realizaciones, aumenta la eficacia y/o frecuencia de recombinación entre un LTVEC y un ácido nucleico objetivo. En una realización, la recombinación es recombinación homóloga. En otra realización, la recombinación es una inserción por unión final no homóloga. En diversas realizaciones, en presencia del rompimiento de cadena sencilla o doble, la eficacia de direccionamiento de una secuencia de LTVEC en el locus genómico objetivo es al menos aproximadamente 2 veces mayor, al menos aproximadamente 3 veces mayor, al menos aproximadamente 4 veces mayor que en la ausencia de la ruptura de cadena sencilla o doble (usando, por ejemplo, el mismo LTVEC y el mismo ácido nucleico objetivo que comprende la misma secuencia objetivo pero en ausencia de una nucleasa añadida que hace que la cadena sencilla o doble se rompa).

Los LTVEC adecuados para su uso con la invención y los métodos para prepararlos se describen, por ejemplo, en las patentes de Estados Unidos Nos. 6.586.251, 6.596.541, 7.105.348 y el documento WO 2002/036789 (PCT/US01/45375).

Aunque las realizaciones dirigidas a introducir un LTVEC en una célula pluripotente de ratón, por ejemplo, una célula ES de ratón, se discuten ampliamente, también se proporcionan en la presente memoria otros métodos que introducen un LTVEC en una variedad de tipos de células de mamífero. Tales células de mamífero incluyen cualquier célula de mamífero que puede modificarse genéticamente de acuerdo con el método descrito en la presente memoria, que incluye, por ejemplo, una célula de ratón, una célula de rata, una célula de conejo, una célula de cerdo, una célula bovina, una célula de venado, una célula de oveja, una célula de cabra, una célula de pollo, una célula de gato, una célula de perro, una célula de hurón, una célula de primate (por ejemplo, mono tití, mono rhesus) y similares. En algunas realizaciones, para aquellos mamíferos para los cuales no están disponibles fácilmente células pluripotentes genéticamente modificables adecuadas, se emplean otros métodos para reprogramar células somáticas en células pluripotentes, por ejemplo, mediante la introducción en células somáticas de una combinación de factores inductores de pluripotencia, que incluyen, pero sin limitarse a, Oct3/4, Sox2, KLF4, Myc, Nanog, LIN28 y Glis1.

En una realización, los brazos de homología secuencia arriba y secuencia abajo son del mismo genoma que el genoma objetivo. En una realización, los brazos de homología son de un genoma relacionado, por ejemplo, el genoma objetivo es un genoma de ratón de una primera cepa, y los brazos de direccionamiento son de un genoma de ratón de una segunda cepa, en donde la primera cepa y la segunda cepa son diferentes. En una realización, los brazos de direccionamiento se derivan de una biblioteca de BAC, una biblioteca de cósmidos o una biblioteca de fagos P1. En una realización, los brazos de homología se derivan de un ADN sintético. En una realización, los brazos de homología se derivan de un gen que no puede direccionarse usando métodos convencionales. En una realización específica, los brazos de homología son de un gen que no puede ser dirigido usando tecnología de direccionamiento convencional, o pueden ser dirigidos solo incorrectamente o solo con una eficacia significativamente baja, en ausencia de una ruptura de cadena sencilla o doble inducida por un agente nucleasa.

En diversas realizaciones, para facilitar la identificación de la modificación dirigida, se emplea un ensayo cuantitativo de alto rendimiento, concretamente, la modificación del ensayo de alelos (MOA). El ensayo MOA descrito en la presente memoria permite un cribado a gran escala de un alelo o alelos modificados en un cromosoma parental después de una modificación genética. El ensayo MOA se puede llevar a cabo mediante diversas técnicas analíticas, que incluyen, pero no se limitan a, una PCR cuantitativa, por ejemplo, una PCR en tiempo real (qPCR). Por ejemplo, la PCR en tiempo real comprende un primer conjunto de cebadores que reconoce el locus objetivo y un segundo conjunto de cebadores que reconoce un locus de referencia no dirigido. Además, el conjunto de cebadores comprende una sonda fluorescente que reconoce la secuencia amplificada. El ensayo cuantitativo también puede llevarse a cabo a través de una variedad de técnicas analíticas, que incluyen, pero no se limitan a, hibridación *in situ* mediada por fluorescencia (FISH), hibridación genómica comparativa, amplificación isotérmica de ADN, hibridación cuantitativa con una sonda o sondas inmovilizadas, tecnología de sonda Invader Probes®, MMP assays®, TaqMan® Molecular Beacon y Eclipse™. (Véase, por ejemplo, el documento US2005/0144655).

En algunas realizaciones, diversas modificaciones genéticas de los loci genómicos objetivo descritos en la presente memoria se pueden llevar a cabo mediante una serie de reacciones de recombinación homóloga (BHR) en células bacterianas usando un LTVEC derivado de ADN de cromosoma artificial bacteriano (BAC) usando tecnología de modificación genética VELOCIGENE® (véase, por ejemplo, la patente de Estados Unidos No. 6.586.251 y Valenzuela, DM et al., (2003), High-throughput engineering of the mouse genome coupled with high-resolution expression analysis, Nature Biotechnology 21(6): 652-659).

En algunas realizaciones, las células ES de ratón dirigidas que comprenden diversas modificaciones genéticas como se describe en este documento se usan como células ES donadoras y se introducen en un embrión de ratón en estadio premórula, por ejemplo, un embrión de ratón en estadio de 8 células, mediante el método VELOCIMOUSE® (véanse, por ejemplo, los documentos US 7.576.259, US 7.659.442, US 7.294.754 y US 2008-0078000 A1). El embrión de ratón que comprende las células ES genéticamente modificadas, se incuba hasta la etapa de blastocisto y luego se implanta en una madre sustituta para producir un ratón F0. Los ratones que portan el locus genómico

genéticamente modificado se pueden identificar mediante la modificación del ensayo de alelos (MOA) como se describe en este documento. El ratón de la generación F0 resultante derivado de las células ES genéticamente modificadas se cruza con un ratón de tipo silvestre para obtener descendientes de la generación F1. Después de la determinación del genotipo con cebadores y/o sondas específicas, las crías F1 que son heterocigotos para el locus genómico genéticamente modificado se cruzan entre sí para producir ratones que son homocigóticos para el locus genómico genéticamente modificado.

Debe observarse que tal como se usa en el presente documento y en las reivindicaciones adjuntas, las formas singulares "un, uno, una", "y", y "el, la" incluyen referencias al plural a menos que el contexto indique claramente lo contrario. Todos los términos técnicos y científicos utilizados en este documento tienen el mismo significado.

Las publicaciones discutidas en este documento se proporcionan únicamente para su divulgación antes de la fecha de presentación de la presente solicitud. Nada en este documento debe interpretarse como una admisión de que la invención descrita no tiene derecho a anteceder a dicha publicación en virtud de la invención anterior. Además, las fechas de publicación proporcionadas pueden ser diferentes de las fechas de publicación reales, que pueden necesitar confirmación independiente.

La invención descrita puede incorporarse en otras formas específicas sin apartarse del espíritu o atributos esenciales de la misma y, en consecuencia, debe hacerse referencia a las reivindicaciones adjuntas, en lugar de a la especificación anterior, para indicar el alcance de la invención.

Ejemplos

Los siguientes ejemplos se presentan para proporcionar a los expertos en la materia una divulgación y descripción completas de cómo hacer y usar la presente invención, y no están destinados a limitar el alcance de lo que los inventores consideran su invención ni pretende representar que los experimentos a continuación son todos o los únicos experimentos realizados. Se han realizado esfuerzos para garantizar la precisión con respecto a los números utilizados (por ejemplo, cantidades, temperatura, etc.) pero se deben tener en cuenta algunos errores y desviaciones experimentales. A menos que se indique lo contrario, las partes son partes en peso, el peso molecular es el peso molecular medio ponderado, la temperatura es en grados centígrados y la presión está en o cerca de la atmosférica.

Ejemplo 1. Direccionamiento mejorado de LTVEC por una nucleasa de dedo de cinc (ZFN)

Para probar si inducir un rompimiento de doble cadena en un gen mediante una nucleasa de dedo de cinc (ZFN) podría potenciar el direccionamiento del mismo gen mediante un LTVEC (un vector de direccionamiento basado en BAC grande), se realizaron tres electroporaciones en células ES F1H4 con un LTVEC que direcciona al gen *Il2rg* y plásmidos que codifican cada mitad de un par de ZFN en las siguientes combinaciones: (1) 1,5 µg de *Il2rg* LTVEC solo; (2) 20 µg de *Il2rg*ZFN-1 + 20 µg de *Il2rg* LTVEC ZFN-2 + 1,5 µg de *Il2rg* LTVEC; y (3) 20 µg de *Il2rg* ZFN-1 + 20 µg de *Il2rg* ZFN-2 sin LTVEC.

ZFN-1 y ZFN-2 usados en la presente invención se diseñaron para contener (1) dominios de unión a ADN con dedos de zinc que reconocen dos secuencias de ADN objetivo contiguas en cada cadena de la secuencia objetivo separadas por un sitio de escisión de 6 pb; y (2) nucleasas FokI que se dimerizan y rompen una cadena doble en el sitio objetivo. Más específicamente, el dominio de dedo de zinc de ZFN-1 se diseñó para reconocer 5'-AGCTCCAAGGTCCTC-3' (SEQ ID NO: 1) en la cadena sentido del exón 1; y el dominio de dedo de cinc de ZFN-2 se diseñó para reconocer 5'-GTCTTCATTCGCACT-3' (SEQ ID NO: 2) en la cadena antisentido del exón 1 en el gen *Il2rg*.

El LTVEC (VelociGene MAID 5057L1) se diseñó para eliminar las aproximadamente 3.000 pares de bases (3 kb) del gen *Il2rg*, incluyendo toda la secuencia codificante para la cadena del receptor gamma de IL-2 y para reemplazar la secuencia endógena con un casete de selección que expresa higromicina fosfotransferasa, que imparte resistencia a higromicina B. LTVEC 5057L1 se deriva del clon BAC parental 290o15 aislado de una biblioteca BAC comercialmente disponible (BAC ES Release 2; Incyte Genomics) y contiene dos grandes brazos de homología con aproximadamente 90 kb (brazo de homología 1) y aproximadamente 8 kb (brazo de homología 2).

Para las electroporaciones (1) y (2), se aislaron colonias resistentes a higromicina B. Para la electroporación (3), se permitió la formación de colonias sin selección de fármacos. Las colonias se recogieron de cada electroporación y se cultivaron en placas de 96 pozos, seguido de cribado de eventos de direccionamiento mediante ensayos de pérdida de alelo (LOA) diseñados para detectar la eliminación de 3 kb creadas mediante direccionamiento correcto del gen *Il2rg* por el LTVEC. Los ensayos LOA también se usaron para detectar eventos de escisión inducidos por la ZFN en el exón 1.

Tabla 1. Ensayos LOA para la eliminación dirigida de 3 kb

| Ensayo 5057U2 (intrón 2) | | |
|--------------------------|-----------------------------|--------------|
| Cebador directo | 5'-GGAGGGTAGCACGGGAAGAAG-3' | SEQ ID NO: 3 |

| | | |
|--------------------------|------------------------------------|--------------|
| Cebador inverso | 5'-GCTGGCTACCCACTTGATTGG-3' | SEQ ID NO: 4 |
| Sonda TaqMan® | 5'-TCAAGCAGTCTCTCCAGCTAACCTCCCT-3' | SEQ ID NO: 5 |
| Ensayo 5057D2 (intrón 7) | | |
| Cebador directo | 5'-CAGGATGTGGCTGACCAAATG-3' | SEQ ID NO: 6 |
| Cebador inverso | 5'-GGCTCCTAATGCCCTGTAGTTTC-3' | SEQ ID NO: 7 |
| Sonda TaqMan® | 5'-CCGTCTCTCTGCCTAGCCCACCCT-3' | SEQ ID NO: 8 |

Tabla 2. Ensayo LOA para la escisión con ZFN

| | | |
|-----------------|----------------------------------|---------------|
| Cebador directo | 5'-CAGCTGCTCCTGCTGAGG-3' | SEQ ID NO: 9 |
| Cebador inverso | 5'-CCTACCAGCTTTGATGTCTTCATTTC-3' | SEQ ID NO: 10 |
| Sonda TaqMan® | 5'-AGCTCCAAGGTCCTCATGTCCAGT-3' | SEQ ID NO: 11 |

Tabla 3. Comparación de la eficiencia de direccionamiento de LTVEC

| EP | LTVEC (µg) | ZFN-1 (µg) | ZFN-2 (µg) | Número de clones ensayados | Clones dirigidos correctamente por LTVEC | Eficacia de direccionamiento (%) | Clones con escisión con ZFN | Eficacia de escisión con ZFN |
|----|------------|------------|------------|----------------------------|--|----------------------------------|-----------------------------|------------------------------|
| 1 | 1,5 | 0 | 0 | 26 | 5 | 19,2 | ND ¹ | ND ¹ |
| 2 | 1,5 | 20 | 20 | 102 | 76 | 74,5 | 7 ² | 6,9 |
| 3 | 0 | 20 | 20 | 192 | 0 | 0 | 6 | 3,1 |

¹ no determinado² entre los clones que no fueron correctamente seleccionados por el LTVEC

Como se muestra en la Tabla 3, cuando los plásmidos que codifican una ZFN diseñada para escindir un sitio en un gen objetivo se combinaron con un LTVEC que dirige al mismo gen (EP # 2), se logró una mejora significativa (aproximadamente un aumento de alrededor de 4 veces en el experimento descrito) de eficacia de direccionamiento comparado con LTVEC solo (EP # 1).

Listado de secuencias

<110> Regeneron Pharmaceuticals, Inc.
 Frendewey, David
 Auerbach, Wojtek
 Valenzuela, David M.
 Yancopoulos, George D.

<120> Direccionamiento mediada por nucleasa con vectores de direccionamiento grandes

<130> 7950A-WO

<150> 61 / 638.267
 <151> 2012-04-25

<160> 11

<170> PatentIn versión 3.5

<210> 1
 <211> 15
 <212> ADN
 <213> Artificial

<220>
 <223> Sintético

<400> 1
 agctccaagg tcctc 15

<210> 2
 <211> 15
 <212> ADN
 <213> Artificial

<220>

<223> Sintético
 <400> 2
 gtcttcattc gcact 15
 5
 <210> 3
 <211> 21
 <212> ADN
 <213> Artificial
 10
 <220>
 <223> Sintético
 <400> 3
 15 ggagggtagc acggaagaa g 21
 <210> 4
 <211> 21
 <212> ADN
 20 <213> Artificial
 <220>
 <223> Sintético
 25 <400> 4
 gctggctacc cacttgattg g 21
 <210> 5
 <211> 29
 30 <212> ADN
 <213> Artificial
 <220>
 <223> Sintético
 35 <400> 5
 tcaagcagtc tctccagct aacctccct 29
 <210> 6
 40 <211> 21
 <212> ADN
 <213> Artificial
 <220>
 45 <223> Sintético
 <400> 6
 caggatgtgg ctgaccaa g 21
 50 <210> 7
 <211> 23
 <212> ADN
 <213> Artificial
 55 <220>
 <223> Sintético
 <400> 7
 60 ggctcctaat gccctgtagt ttc 23
 <210> 8
 <211> 24
 <212> ADN
 <213> Artificial
 65 <220>

| | | |
|----|-----------------------------|----|
| | <223> Sintético | |
| | <400> 8 | |
| 5 | ccgtctctct gcctagccca ccct | 24 |
| | <210> 9 | |
| | <211> 18 | |
| | <212> ADN | |
| 10 | <213> Artificial | |
| | <220> | |
| | <223> Sintético | |
| | <400> 9 | |
| 15 | cagctgctcc tgctgagg | 18 |
| | <210> 10 | |
| | <211> 25 | |
| | <212> ADN | |
| 20 | <213> Artificial | |
| | <220> | |
| | <223> Sintético | |
| | <400> 10 | |
| 25 | cctaccagct ttgatgtctt cattc | 25 |
| | <210> 11 | |
| | <211> 24 | |
| 30 | <212> ADN | |
| | <213> Artificial | |
| | <220> | |
| | <223> Sintético | |
| 35 | <400> 11 | |
| | agctccaagg tcctcatgtc cagt | 24 |

REIVINDICACIONES

1. Un método para modificar un locus genómico objetivo en una célula madre embrionaria (ES) de ratón, que comprende:

(a) introducción en la célula ES de ratón:

(i) una nucleasa de dedo de zinc (ZFN) que provoca un rompimiento de doble cadena en o cerca de el locus genómico objetivo; y

(ii) un vector de direccionamiento grande (LTVEC) que comprende un ácido nucleico insertado flanqueado por un brazo de homología secuencia arriba y un brazo de homología secuencia abajo,

en el que el ácido nucleico insertado varía de 5 kb a 30 kb de longitud,

en el que la suma total de los brazos de homología secuencia arriba y secuencia abajo tiene una longitud de al menos 10 kb,

en el que los brazos de homología secuencia arriba y secuencia abajo están entre 5 kb y 200 kb de longitud, y

en el que el LTVEC varía de 50 kb a 300 kb de longitud; y

(b) seleccionar una célula ES de ratón dirigida que comprende el ácido nucleico insertado en el locus genómico objetivo, en el que la integración del ácido nucleico insertado en el locus genómico objetivo da como resultado la sustitución de una secuencia de ácido nucleico endógeno en el locus genómico objetivo con el ácido nucleico insertado.

2. El método de la reivindicación 1, en el que el ácido nucleico insertado comprende un gen informador operativamente enlazado a un informador exógeno.

3. El método de la reivindicación 1, en el que el uso combinado de LTVEC con la ZFN da como resultado mayor eficacia de direccionamiento en comparación con el uso del LTVEC solo, opcionalmente en el que la eficacia de direccionamiento es mayor al menos dos veces en comparación con el uso de LTVEC solo.

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

(I) la ZFN es una construcción de expresión que comprende una secuencia de ácido nucleico que codifica una proteína ZFN, y en la que el ácido nucleico está operativamente enlazado a un promotor activo en la célula; o

(II) la ZFN es un ARNm que codifica una proteína ZFN.

5. El método de la reivindicación 1, en el que la secuencia objetivo de la ZFN está localizada en un intrón, un exón, un promotor, una región reguladora del promotor, o una región potenciadora en el locus genómico objetivo.

6. El método de la reivindicación 1, en el que:

(I) el ácido nucleico insertado comprende un casete de selección;

(II) el ácido nucleico insertado comprende una secuencia de ácido nucleico humano; o

(III) el ácido nucleico insertado comprende un alelo condicional.

7. El método de la reivindicación 1, en el que:

(I) el ácido nucleico insertado comprende una secuencia de ácido nucleico homóloga a la secuencia de ácido nucleico reemplazada en el locus genómico objetivo de la célula ES de ratón;

(II) el ácido nucleico insertado comprende una secuencia de ácido nucleico ortóloga a la secuencia de ácido nucleico reemplazada en el locus genómico objetivo de la célula ES de ratón; o

(III) el ácido nucleico insertado comprende una secuencia de ácido nucleico de una especie diferente.

8. El método de la reivindicación 1, en el que el ácido nucleico insertado comprende una secuencia de ácido nucleico de la región variable de cadena pesada de inmunoglobulina humana operativamente enlazada a una secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina,

opcionalmente, en el que la secuencia de la región constante de cadena pesada de inmunoglobulina es una secuencia de la región constante de cadena pesada de inmunoglobulina de ratón, una secuencia de la región constante de cadena pesada de inmunoglobulina humana, o una combinación de las mismas, y

opcionalmente, en el que la secuencia de ácido nucleico de la región constante de cadena pesada de inmunoglobulina se selecciona de un C_H1, una bisagra, un C_H2, un C_H3 y una combinación de las mismas.

9. El método de la reivindicación 1, en el que el ácido nucleico insertado comprende una secuencia de ácido nucleico de la región variable de cadena ligera λ o κ humana no reordenada operablemente enlazada a una secuencia de ácido nucleico de la región constante de cadena ligera de inmunoglobulina humana o de ratón seleccionada de una secuencia de ácido nucleico de la región constante de cadena ligera λ y una secuencia de ácido nucleico de la región constante de cadena ligera κ.

- 5 10. El método de la reivindicación 1, en el que el ácido nucleico insertado comprende una secuencia de ácido nucleico de región variable de cadena ligera λ o κ humana reordenada operablemente enlazada a una secuencia de ácido nucleico de la región constante de cadena ligera de inmunoglobulina humana o de ratón seleccionada de una secuencia de ácido nucleico de la región constante de cadena ligera λ y una secuencia de ácido nucleico de la región constante de cadena ligera κ .
- 10 11. El método de la reivindicación 1, en el que la integración del ácido nucleico insertado en el locus genómico objetivo da como resultado una desactivación, una activación, una mutación puntual, un intercambio de dominios, un intercambio de exones, un intercambio de intrones, un intercambio de secuencias reguladoras, un intercambio de genes, o una combinación de los mismos.
- 15 12. El método de la reivindicación 1, en el que la selección de la etapa (b) se lleva a cabo mediante una modificación del ensayo de alelos (MOA).
13. El método de la reivindicación 1, que comprende además:
- 20 (c) introducir la célula ES de ratón modificada en un embrión en estadio de premórula; y
(d) incubar el embrión hasta la etapa de blastocisto, opcionalmente en el que el método comprende producir un ratón F0 a partir del embrión implantado en una madre sustituta; y
(e) identificar un ratón portador del locus genómico genéticamente modificado.

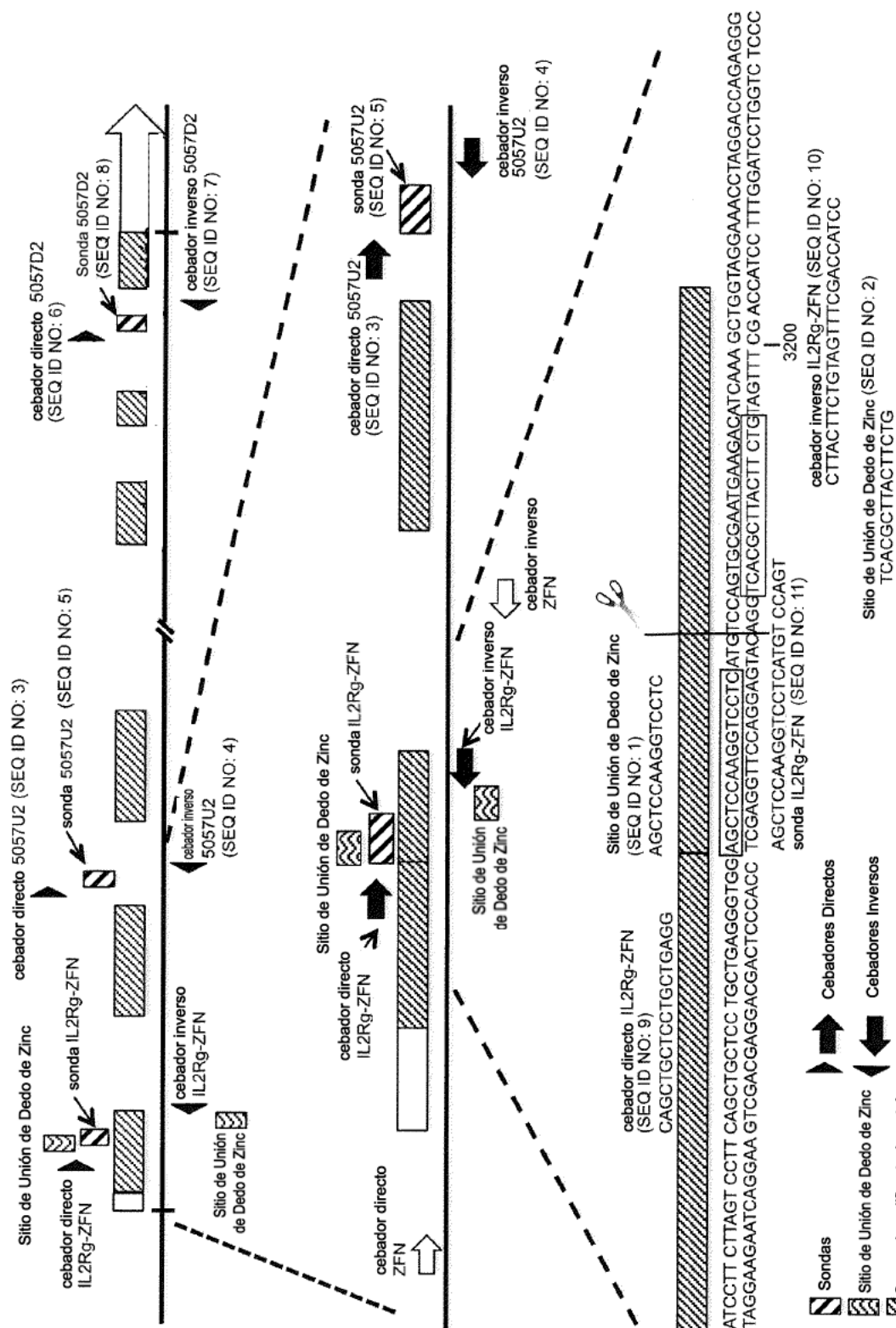


FIG. 1