

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 019 688**

51 Int. Cl.:

C12N 15/90

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **05.06.2015** **E 20155939 (0)**

97 Fecha y número de publicación de la concesión europea: **02.04.2025** **EP 3708671**

54 Título: **Métodos y composiciones para modificar un locus objetivo**

30 Prioridad:

06.06.2014 US 201462008832 P
27.06.2014 US 201462017916 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
21.05.2025

73 Titular/es:

REGENERON PHARMACEUTICALS, INC.
(100.00%)
777 Old Saw Mill River Road
Tarrytown, NY 10591, US

72 Inventor/es:

AUERBACH, WOJTEK;
FRENDEWEY, DAVID;
DROQUETT, GUSTAVO;
GAGLIARDI, ANTHONY;
KUNO, JUNKO y
VALENZUELA, DAVID

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 3 019 688 T3

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

DESCRIPCIÓN

Métodos y composiciones para modificar un locus objetivo

Referencia cruzada a solicitudes relacionadas

5 Esta solicitud reivindica el beneficio de la solicitud provisional de patente de Estados Unidos núm. 62/008,832, presentada el 6 de junio de 2014, y de la solicitud provisional de patente de Estados Unidos núm. 62/017,916, presentada el 27 de junio de 2014.

Campo

Los métodos y composiciones se relacionan con el campo de la biología molecular. En particular, se proporcionan métodos para modificar un locus objetivo en una célula.

10 Como un archivo de texto a través de la web de EFS

La copia oficial del listado de secuencias se envía electrónicamente a través de EFS-Web como un listado de secuencias con formato ASCII con un archivo llamado 461003SEQLIST.TXT, creado el 5 de junio de 2015 y con un tamaño de 5 KB, y se presenta simultáneamente con la descripción. El listado de secuencias contenido en esta descripción con formato ASCII es parte de la descripción.

15 Antecedentes

La recombinación homóloga usando vectores de transformación que están específicamente diseñados para añadir, eliminar o reemplazar una secuencia de ácido nucleico particular en un locus genómico es un enfoque popular para lograr una modificación genómica deseada en una célula. Una nucleasa diseñada específicamente para introducir un corte o ruptura de doble cadena en o cerca de un locus objetivo puede usarse en combinación con un vector de transformación para mejorar la eficiencia de la recombinación homóloga en el locus objetivo.

20 Aunque la técnica de la modificación dirigida a través de la recombinación homóloga ha avanzado considerablemente en las últimas dos décadas, aún persisten dificultades para lograr una eficiencia de transformación aceptable utilizando vectores de transformación. Se necesitan métodos que mejoren la eficacia y la eficiencia de las modificaciones específicas producidas. Los documentos

25 Resumen

Se proporcionan métodos para modificar uno o más loci objetivo en una célula.

La presente invención se refiere a las realizaciones como se caracterizan en las reivindicaciones.

30 En un primer aspecto de la invención, se proporciona un método para la modificación en serie de un locus objetivo en una célula, que comprende:

(a) proporcionar la célula que comprende el locus objetivo, en donde el locus objetivo comprende un primer polinucleótido que codifica un primer marcador de selección unido operativamente a un primer promotor y un primer sitio de reconocimiento de nucleasa para un primer agente de nucleasa, en donde el primer sitio de reconocimiento de nucleasa está ubicado en una región codificante del primer marcador de selección o cualquier región no codificante de proteína del primer marcador de selección, y opcionalmente en donde el primer polinucleótido que codifica el primer marcador de selección está flanqueado por el primer sitio objetivo y el segundo sitio objetivo;

35 (b) introducir en la célula:

(i) el primer agente de nucleasa o un polinucleótido que codifica el primer agente de nucleasa, en donde el primer agente de nucleasa induce un corte o ruptura de doble cadena en el primer sitio de reconocimiento de nucleasa; y

40 (ii) un primer vector de transformación que comprende un primer polinucleótido de inserción flanqueado por un primer brazo de homología correspondiente a un primer sitio objetivo ubicado en el locus objetivo y un segundo brazo de homología correspondiente a un segundo sitio objetivo ubicado en el locus objetivo, en donde el primer polinucleótido de inserción comprende un segundo polinucleótido que codifica un segundo marcador de selección unido operativamente a un segundo promotor y un segundo sitio de reconocimiento de nucleasa para un segundo agente de nucleasa, en donde el segundo sitio de reconocimiento de nucleasa está ubicado dentro de una región codificante del segundo marcador de selección o cualquier región no codificante de proteínas del segundo marcador de selección; en donde el primer marcador de selección y el segundo marcador de selección son diferentes,

50

en donde el primer agente de nucleasa es diferente del segundo agente de nucleasa, y

(c) identificar una célula modificada que comprende el primer polinucleótido de inserción en el locus objetivo, en donde la célula modificada tiene la actividad del segundo marcador de selección pero no tiene la actividad del primer marcador de selección, opcionalmente en donde la identificación se lleva a cabo a través de un ensayo de modificación de alelos (MOA);

En una realización, el locus objetivo está en el genoma de la célula. En otra realización, el locus objetivo está ubicado en un vector en la célula. En una realización, el corte o ruptura de doble cadena en el primer sitio de reconocimiento interrumpe la actividad del primer marcador de selección. En otra realización adicional, la etapa de identificación (c) comprende cultivar las células en condiciones que permiten la identificación de células que no tienen una actividad del primer marcador de selección. En una realización, el primer polinucleótido que comprende el primer marcador de selección está flanqueado por un primer sitio objetivo y un segundo sitio objetivo. En una realización, la etapa de identificación (c) comprende identificar al menos una célula que comprende el primer polinucleótido de inserción integrado en el primer y el segundo sitio objetivo. En una realización, el primer polinucleótido de inserción comprende: (a) un primer polinucleótido de interés; y (b) un segundo polinucleótido que codifica un segundo marcador de selección unido operativamente a un segundo promotor activo en la célula, en donde el segundo polinucleótido comprende un segundo sitio de reconocimiento para un segundo agente de nucleasa.

En una realización, el método comprende además (a) introducir en la célula que comprende el primer polinucleótido de inserción integrado en el locus objetivo, (i) un segundo agente de nucleasa, en donde el segundo agente de nucleasa induce un corte o ruptura de doble cadena en el segundo sitio de reconocimiento; y (ii) un segundo vector de transformación que comprende un segundo polinucleótido de inserción flanqueado por un tercer y un cuarto brazos de homología correspondientes a un tercer y un cuarto sitios objetivo ubicados en una proximidad suficiente al segundo sitio de reconocimiento; y (b) identificar al menos una célula que comprende el segundo polinucleótido de inserción integrado en el locus objetivo. En una realización, el corte o ruptura de doble cadena en el segundo sitio de reconocimiento interrumpe la actividad del segundo marcador de selección. En una realización, la etapa de identificación (b) comprende cultivar la célula en condiciones que permiten la identificación de células que no tienen la actividad del segundo marcador de selección. En una realización, el segundo polinucleótido que comprende el segundo marcador seleccionable está flanqueado por el tercer sitio objetivo y el cuarto sitio objetivo. En una realización, la etapa de identificación (b) comprende identificar al menos una célula que comprende el segundo polinucleótido de inserción integrado en el tercer y el cuarto sitios objetivo.

En una realización, el segundo polinucleótido de inserción comprende: (a) un segundo polinucleótido de interés; y (b) un tercer polinucleótido que codifica un tercer marcador de selección unido operativamente a un tercer promotor activo en la célula, en donde el tercer polinucleótido comprende un tercer sitio de reconocimiento para un tercer agente de nucleasa. El primer agente de nucleasa es diferente del segundo agente de nucleasa. El primer marcador de selección es diferente del segundo marcador de selección. En una realización, el primer y el tercer sitios de reconocimiento de nucleasa son idénticos entre sí y son diferentes del segundo sitio de reconocimiento de nucleasa; y, el primer y el tercer agente de nucleasa son idénticos entre sí y son diferentes del segundo agente de nucleasa. En una realización, el primer y tercer marcadores de selección son idénticos. En una realización, uno del primer, el segundo o el tercer marcador de selección proporciona resistencia a un antibiótico. En una realización, el antibiótico comprende G418, higromicina, blastomicina, neomicina o puromicina. En una realización, uno del primer, el segundo o el tercer marcador de selección está unido operativamente a un promotor inducible, y la expresión del marcador seleccionable es tóxica para la célula. En una realización, el primer, el segundo o el tercer marcador de selección comprende hipoxantina-guanina fosforribosiltransferasa (HGPRT) o timidina quinasa del virus del herpes simple (HSV-TK). En una realización, dicha célula es una célula procariota. En una realización la célula es una célula eucariota. En una realización, la célula eucariota es una célula de mamífero. En una realización, la célula de mamífero es una célula de mamífero no humano. En una realización, la célula de mamífero es de un roedor. En una realización, el roedor es una rata o un ratón.

En una realización, la célula es una célula pluripotente. En una realización, la célula de mamífero es una célula madre pluripotente humana inducida (iPS). En una realización, la célula pluripotente es una célula madre embrionaria (ES) no humana. En una realización, la célula pluripotente es una célula madre embrionaria (ES) de ratón o una célula madre embrionaria (ES) de rata. En una realización, la célula pluripotente es una célula madre hematopoyética. En una realización, la célula pluripotente es una célula madre neuronal. En una realización, la célula de mamífero es un fibroblasto humano.

En una realización, el uso combinado del primer vector de transformación con el primer agente de nucleasa da como resultado una mayor eficacia de transformación en comparación con el uso del primer vector de transformación solo. En una realización, la eficacia de transformación del primer vector de transformación aumenta al menos 2 veces en comparación con el uso del primer vector de transformación solo.

En una realización, el primer o el segundo agente de nucleasa comprende una construcción de expresión que comprende una secuencia de ácido nucleico que codifica el agente de nucleasa, y en donde el ácido nucleico está unido operativamente a un cuarto promotor activo en la célula. En una realización, el primer o el segundo agente de nucleasa es un ARNm que codifica una nucleasa. En una realización, el primer o el segundo agente de nucleasa es una nucleasa de dedos de zinc (ZFN). En una realización, el primer o el segundo agente de nucleasa es una nucleasa efectora de tipo activador de la transcripción (TALEN). En una realización, el primer o el segundo agente de nucleasa es una meganucleasa.

En una realización, el primer o el segundo agente de nucleasa comprende una proteína (Cas) asociada a repeticiones palindrómicas cortas intercaladas regularmente y agrupadas (CRISPR) y un ARN guía (ARNg). En una realización, el ARN guía (ARNg) comprende (a) un ARN de Repeticiones palindrómicas cortas intercaladas regularmente y agrupadas (CRISPR) (ARNcr) que se dirige al primero, al segundo o al tercer sitios de reconocimiento; y (b) un ARN CRISPR transactivante (tracrARN). En una realización, el primer o el segundo sitios de reconocimiento están flanqueados inmediatamente por una secuencia de un motivo adyacente a protoespaciador (PAM). En una realización, el locus genómico de interés comprende la secuencia de nucleótidos de SEQ ID NO: 1. En una realización, la proteína Cas es Cas9. En una realización, el ARNg comprende: (a) el ARN quimérico de la secuencia de ácido nucleico de SEQ ID NO: 2; o (b) el ARN quimérico de la secuencia de ácido nucleico de SEQ ID NO: 3. En una realización, el ARNcr comprende la SEQ ID NO: 4; la SEQ ID NO: 5; o la SEQ ID NO: 6. En una realización, el tracrARN comprende la SEQ ID NO: 7 o la SEQ ID NO: 8.

En una realización, el primero, el segundo y/o el tercer sitio de reconocimiento están ubicados en un intrón, un exón, un promotor, una región reguladora del promotor o una región potenciadora del primer, el segundo o el tercer marcador de selección. En una realización, el primer sitio objetivo y el segundo sitio objetivo están inmediatamente adyacentes al primer sitio de reconocimiento. En una realización, el primer sitio objetivo y el segundo sitio objetivo son de aproximadamente 10 nucleótidos a aproximadamente 14 kb a partir del primer sitio de reconocimiento. En una realización, el tercer sitio objetivo y el cuarto sitio objetivo están inmediatamente adyacentes al segundo sitio de reconocimiento. En una realización, el tercer sitio objetivo y el cuarto sitio objetivo son de aproximadamente 10 nucleótidos a aproximadamente 14 kb a partir del segundo sitio de reconocimiento.

En una realización, una suma total del primer brazo de homología y el segundo brazo de homología es de al menos aproximadamente 10 kb. En una realización, una suma total del tercer brazo de homología y el cuarto brazo de homología es de al menos aproximadamente 10 kb. En una realización, el primer polinucleótido de inserción varía de aproximadamente 5 kb a aproximadamente 300 kb de longitud. En una realización, el segundo polinucleótido de inserción varía de aproximadamente 5 kb a aproximadamente 300 kb de longitud.

En una realización, la integración del primer polinucleótido de inserción en el locus 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 estos. En una realización, la integración del segundo polinucleótido de inserción en el locus 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 estos.

En una realización, el primer polinucleótido de inserción comprende un polinucleótido de interés que comprende un polinucleótido humano. En una realización, el segundo polinucleótido de inserción comprende un polinucleótido de interés que comprende un polinucleótido humano. En una realización, el primer polinucleótido de inserción comprende un polinucleótido de interés que comprende una región del locus del receptor alfa de células T.

En una realización, el segundo polinucleótido de inserción comprende un polinucleótido de interés que comprende una región del locus del receptor alfa de células T. En una realización, el primer o el segundo polinucleótido de inserción comprenden un polinucleótido de interés que comprende al menos un segmento génico de región variable y/o un segmento génico de región de unión del locus del receptor alfa de células T. En una realización, la región del locus del receptor alfa de células T es de un ser humano.

En una realización, el primer polinucleótido de inserción comprende un polinucleótido de interés que comprende una secuencia de ácido nucleico de una región variable de cadena pesada de inmunoglobulina humana no reorganizada, unida operativamente a una secuencia de ácido nucleico de una región constante de cadena pesada de inmunoglobulina no humana.

En una realización, la etapa de identificación se lleva a cabo mediante un ensayo de modificación de alelos (MOA). En una realización, el primer polinucleótido de inserción comprende un polinucleótido de interés que comprende una secuencia de ácido nucleico que es homóloga u ortóloga a la secuencia de ácido nucleico en un genoma de la célula. En una realización, el segundo polinucleótido de inserción comprende una secuencia de ácido nucleico que es homóloga u ortóloga a la secuencia de ácido nucleico en un genoma de la célula. En

una realización, el primer polinucleótido de inserción comprende un polinucleótido de interés que comprende una secuencia de ácido nucleico exógena. En una realización, el segundo polinucleótido de inserción comprende un polinucleótido de interés que comprende una secuencia de ácido nucleico exógena.

En algunas realizaciones, los métodos para modificar un locus objetivo en una célula comprenden: (a) proporcionar una célula que comprende un primer locus objetivo que comprende un ácido nucleico que codifica un primer marcador de selección unido operativamente a un primer promotor; (b) introducir en la célula (i) una o más construcciones de expresión que codifican una proteína Cas y un primer ARN guía (ARNg), cada uno de los cuales está unido operativamente a un promotor activo en la célula, en donde la proteína Cas induce un corte o ruptura de doble cadena en un primer sitio objetivo de ARNg en el primer ácido nucleico, interrumpiendo así la expresión o actividad del primer marcador de selección, y (ii) un primer vector de transformación que comprende un primer ácido nucleico de inserción que comprende un segundo ácido nucleico que codifica un segundo marcador de selección unido operativamente a un segundo promotor, en donde el primer ácido nucleico de inserción está flanqueado por un primer y un segundo brazos de homología correspondientes a un primer y un segundo sitios objetivo ubicados en el primer locus objetivo; y (c) identificar una célula modificada que comprende el primer ácido nucleico de inserción en el primer locus objetivo, en donde la célula modificada tiene la actividad del segundo marcador de selección pero no tiene la actividad del primer marcador de selección, y en donde el primer y el segundo marcadores de selección son diferentes. En una realización, el primer ARNg no hibrida con el primer ácido nucleico de inserción. En una realización, el locus objetivo de interés está ubicado en el genoma de la célula. En otra realización, el locus objetivo de interés está ubicado en un vector en la célula. En una realización, la etapa de identificación (c) comprende cultivar la célula en condiciones que permiten la identificación de la célula modificada que tiene actividad del segundo marcador de selección pero no tiene la actividad del primer marcador de selección.

En una realización, el método comprende además (d) introducir en la célula modificada que comprende el primer ácido nucleico de inserción en el primer locus objetivo (i) uno o más ácidos nucleicos que codifican la proteína Cas y un segundo ARNg, cada uno de los cuales está unido operativamente al promotor activo en la célula modificada, en donde la proteína Cas induce el corte o ruptura de doble cadena en un segundo sitio objetivo de ARNg en el primer ácido nucleico de inserción que comprende el segundo ácido nucleico, interrumpiendo así la expresión o actividad del segundo marcador de selección, y (ii) un segundo vector de transformación que comprende un segundo ácido nucleico de inserción que comprende un tercer ácido nucleico que codifica un tercer marcador de selección unido operativamente a un tercer promotor, en donde el segundo ácido nucleico de inserción está flanqueado por un tercer y cuarto brazos de homología correspondientes a un tercer y un cuarto sitios objetivo ubicados en un segundo locus objetivo; y (e) identificar una segunda célula modificada que comprende el segundo ácido nucleico de inserción en el segundo locus objetivo, en donde la segunda célula modificada tiene la actividad del tercer marcador de selección pero no tiene la actividad del segundo marcador de selección, en donde el segundo y el tercer marcadores de selección son diferentes. En una realización, el primer y el segundo loci objetivo están ubicados inmediatamente adyacentes entre sí. En otra realización, el primer o el segundo locus objetivo está ubicado de aproximadamente 10 nucleótidos a aproximadamente 14 kb, aproximadamente 10 nucleótidos a aproximadamente 100 nucleótidos, aproximadamente 100 nucleótidos a aproximadamente 500 nucleótidos, aproximadamente 500 nucleótidos a aproximadamente 1000 nucleótidos, aproximadamente 1 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 10 kb, o aproximadamente 10 kb a aproximadamente 14 kb a partir del primer o el segundo sitio objetivo de ARNg. En una realización, el segundo ARNg no hibrida con el segundo ácido nucleico de inserción. En una realización, la etapa de identificación (e) comprende cultivar la célula modificada en condiciones que permiten la identificación de la segunda célula modificada que tiene actividad del tercer marcador de selección pero no tiene la actividad del segundo marcador de selección.

En una realización, el método comprende además (f) introducir en la segunda célula modificada que comprende el segundo ácido nucleico de inserción en el segundo locus objetivo: (i) la una o más construcciones de expresión que codifican la proteína Cas y un tercer ARNg, cada uno de los cuales está unido operativamente al promotor activo en la segunda célula modificada, en donde la proteína Cas induce el corte o ruptura de doble cadena en un tercer sitio objetivo de ARNg en el segundo ácido nucleico de inserción que comprende el tercer ácido nucleico, interrumpiendo así la expresión o actividad del tercer marcador de selección, y (ii) un tercer vector de transformación que comprende un tercer ácido nucleico de inserción que comprende un cuarto ácido nucleico que codifica un cuarto marcador de selección unido operativamente a un cuarto promotor, en donde el tercer ácido nucleico de inserción está flanqueado por el quinto y sexto brazos de homología correspondientes a los quinto y sexto sitios objetivo ubicados en un tercer locus objetivo; y (g) identificar una tercera célula modificada que comprende el tercer ácido nucleico de inserción en el tercer locus objetivo, en donde la tercera célula modificada tiene la actividad del cuarto marcador de selección pero no tiene la actividad del tercer marcador de selección, en donde el tercer y el cuarto marcadores de selección son diferentes. En una realización, el segundo y el tercer loci objetivo están ubicados inmediatamente adyacentes entre sí. En otra realización, el segundo o el tercer locus objetivo está ubicado de aproximadamente 10 nucleótidos a aproximadamente 14 kb a partir del primer o segundo sitio objetivo de ARNg.

En una realización, el primero, el segundo, el tercero o el cuarto marcador proporcionan resistencia a un antibiótico. En una realización, el antibiótico comprende G418, higromicina, blastomicina, neomicina o

puromicina. En una realización, el primero, segundo, tercero o cuarto marcadores de selección comprenden hipoxantina-guanina fosforribosiltransferasa (HGPRT) o timidina quinasa del virus del herpes simple (HSV-TK). En una realización, el primer, el segundo o el tercer ARNg comprende (i) una secuencia de nucleótidos que hibrida con el primer, el segundo o el tercer sitio objetivo de ARNg y (ii) un ARN CRISPR transactivante (tracrARN). En una realización, el primer, el segundo o el tercer locus objetivo está ubicado muy cerca del primer, el segundo o el tercer sitio objetivo de ARNg de tal manera que el corte o ruptura de doble cadena en el sitio objetivo de ARNg promueve la recombinación homóloga del vector de transformación en el locus objetivo. En una realización, la proteína Cas es Cas9. En una realización, el primer, el segundo o el tercer sitio objetivo de ARNg está flanqueado inmediatamente por una secuencia de un motivo adyacente a protoespaciador (PAM).

En una realización, la célula es una célula procariota. En otra realización, la célula es una célula eucariota. En una realización, la célula eucariota es una célula de mamífero. En una realización, la célula de mamífero es un fibroblasto. En una realización, la célula de mamífero es un fibroblasto humano. En una realización, la célula de mamífero es una célula de mamífero no humano. En una realización, la célula de mamífero es de un roedor. En una realización, el roedor es una rata, un ratón o un hámster.

En una realización, la célula eucariota es una célula pluripotente. En una realización, la célula pluripotente es una célula madre hematopoyética o una célula madre neuronal. En una realización, la célula pluripotente es una célula madre pluripotente humana inducida (iPS). En una realización, la célula pluripotente es una célula madre embrionaria (ES) de ratón o una célula madre embrionaria (ES) de rata.

En una realización, el primer, el segundo o el tercer sitio objetivo de ARNg se encuentra en un intrón, un exón, un promotor o una región reguladora del promotor en el primer, el segundo o el tercer ácido nucleico que codifica el primer, el segundo, o el tercer marcador de selección. En una realización, el primer, el segundo o el tercer vector de transformación es de al menos aproximadamente 10 kb. En una realización, el primer, el segundo o el tercer ácido nucleico de inserción varía de aproximadamente 5 kb a aproximadamente 300 kb.

En una realización, el primero, segundo o tercer ácido nucleico de inserción comprende una región genómica del locus del receptor alfa de células T humanas. En una realización, la región genómica comprende al menos un segmento génico de región variable y/o un segmento génico de la región de unión del locus del receptor alfa de células T humanas.

En una realización, el primer y el tercer marcadores de selección son iguales. En una realización, el primer y tercer marcadores de selección son iguales y el segundo y cuarto marcadores de selección son iguales. En una realización, el primer y el tercer ARNg son iguales.

Además, en esta divulgación se proporcionan métodos para modificar un locus objetivo en una célula como se caracteriza en las reivindicaciones. Dichos métodos comprenden proporcionar una célula que comprende un locus objetivo que comprende un primer polinucleótido que codifica un primer marcador de selección unido operativamente a un primer promotor activo en la célula, en donde el primer polinucleótido comprende además un primer sitio de reconocimiento para un primer agente de nucleasa. Se introduce un primer agente de nucleasa en la célula, en donde el primer agente de nucleasa induce un corte o ruptura de doble cadena en el primer sitio de reconocimiento. Además, se introduce en la célula un primer vector de transformación que comprende un primer polinucleótido de inserción flanqueado por un primer y un segundo brazos de homología que corresponden a un primer y un segundo sitios objetivo ubicados en proximidad suficiente al primer sitio de reconocimiento. Después se identifica al menos una célula que comprende el primer polinucleótido de inserción integrado en el locus objetivo.

También se proporcionan métodos para modificar un locus objetivo en una célula que comprende: (a) proporcionar una célula que comprende un locus objetivo que comprende un primer polinucleótido que codifica un primer marcador de selección unido operativamente a un primer promotor activo en la célula, en donde el primer polinucleótido comprende además un primer sitio de reconocimiento para un primer agente de nucleasa, (b) introducir en la célula (i) un primer agente de nucleasa, en donde el primer agente de nucleasa induce un corte o ruptura de doble cadena en el primer sitio de reconocimiento; y (ii) un primer vector de transformación que comprende un primer polinucleótido de inserción flanqueado por un primer y un segundo brazos de homología correspondientes a un primer y un segundo sitios objetivo ubicados en proximidad suficiente al primer sitio de reconocimiento; y (c) identificar al menos una célula que comprende el primer polinucleótido de inserción integrado en el locus objetivo. En una realización, el locus objetivo está en el genoma de la célula. En otra realización, el locus objetivo está ubicado en un vector en la célula. En una realización, el corte o ruptura de doble cadena en el primer sitio de reconocimiento interrumpe la actividad del primer marcador de selección. En otra realización adicional, la etapa de identificación (c) comprende cultivar las células en condiciones que permiten la identificación de células que no tienen una actividad del primer marcador de selección. En una realización, el primer polinucleótido que comprende el primer marcador de selección está flanqueado por un primer sitio objetivo y un segundo sitio objetivo. En una realización, la etapa de identificación (c) comprende identificar al menos una célula que comprende el primer polinucleótido de inserción integrado en el primer y el segundo sitio objetivo. En una realización, el primer polinucleótido de inserción comprende: (a) un primer

polinucleótido de interés; y (b) un segundo polinucleótido que codifica un segundo marcador de selección unido operativamente a un segundo promotor activo en la célula, en donde el segundo polinucleótido comprende un segundo sitio de reconocimiento para un segundo agente de nucleasa.

En una realización, el método comprende además (a) introducir en la célula que comprende el primer polinucleótido de inserción integrado en el locus objetivo, (i) un segundo agente de nucleasa, en donde el segundo agente de nucleasa induce un corte o ruptura de doble cadena en el segundo sitio de reconocimiento; y (ii) un segundo vector de transformación que comprende un segundo polinucleótido de inserción flanqueado por un tercer y un cuarto brazos de homología correspondientes a un tercer y un cuarto sitios objetivo ubicados en una proximidad suficiente al segundo sitio de reconocimiento; y (b) identificar al menos una célula que comprende el segundo polinucleótido de inserción integrado en el locus objetivo. En una realización, el corte o ruptura de doble cadena en el segundo sitio de reconocimiento interrumpe la actividad del segundo marcador de selección. En una realización, la etapa de identificación (b) comprende cultivar la célula en condiciones que permiten la identificación de células que no tienen la actividad del segundo marcador de selección. En una realización, el segundo polinucleótido que comprende el segundo marcador de selección está flanqueado por el tercer sitio objetivo y el cuarto sitio objetivo. En una realización, la etapa de identificación (b) comprende identificar al menos una célula que comprende el segundo polinucleótido de inserción integrado en el tercer y el cuarto sitios objetivo. En una realización, el segundo polinucleótido de inserción comprende: (a) un segundo polinucleótido de interés; y (b) un tercer polinucleótido que codifica un tercer marcador de selección unido operativamente a un tercer promotor activo en la célula, en donde el tercer polinucleótido comprende un tercer sitio de reconocimiento para un tercer agente de nucleasa. El primer agente de nucleasa es diferente del segundo agente de nucleasa. El primer marcador de selección es diferente del segundo marcador de selección. En una realización, el primer y el tercer sitios de reconocimiento de nucleasa son idénticos entre sí y son diferentes del segundo sitio de reconocimiento de nucleasa; y, el primer y el tercer agente de nucleasa son idénticos entre sí y son diferentes del segundo agente de nucleasa. En una realización, el primer y tercer marcadores de selección son idénticos. En una realización, uno del primer, el segundo o el tercer marcador de selección proporciona resistencia a un antibiótico. En una realización, el antibiótico comprende G418, higromicina, blastomicina, neomicina o puromicina. En una realización, uno del primer, el segundo o el tercer marcador de selección está unido operativamente a un promotor inducible, y la expresión del marcador de selección es tóxica para la célula. En una realización, el primer, el segundo o el tercer marcador de selección comprende hipoxantina-guanina fosforibosiltransferasa (HGPRT) o timidina quinasa del virus del herpes simple (HSV-TK). En una realización, dicha célula es una célula procariota. En una realización, la célula es una célula eucariota. En una realización, la célula eucariota es una célula de mamífero. En una realización, la célula de mamífero es una célula de mamífero no humano. En una realización, la célula de mamífero es de un roedor. En una realización, el roedor es una rata o un ratón. En una realización, la célula de mamífero es un fibroblasto humano.

En una realización, la célula es una célula pluripotente. En una realización, la célula de mamífero es una célula madre pluripotente humana inducida (iPS). En una realización, la célula pluripotente es una célula madre embrionaria (ES) no humana. En una realización, la célula pluripotente es una célula madre embrionaria (ES) de ratón o una célula madre embrionaria (ES) de rata. En una realización, la célula pluripotente es una célula madre hematopoyética. En una realización, la célula pluripotente es una célula madre neuronal.

En una realización, el uso combinado del primer vector de transformación con el primer agente de nucleasa da como resultado una mayor eficacia de transformación en comparación con el uso del primer vector de transformación solo. En una realización, la eficacia de transformación del primer vector de transformación aumenta al menos 2 veces en comparación con el uso del primer vector de transformación solo.

En una realización, el primer o el segundo agente de nucleasa comprende una construcción de expresión que comprende una secuencia de ácido nucleico que codifica el agente de nucleasa, y el ácido nucleico está unido operativamente a un cuarto promotor activo en la célula. En una realización, el primer o el segundo agente de nucleasa es un ARNm que codifica una nucleasa. En una realización, el primer o el segundo agente de nucleasa es una nucleasa de dedos de zinc (ZFN). En una realización, el primer o el segundo agente de nucleasa es una nucleasa efectora de tipo activador de la transcripción (TALEN). En una realización, el primer o el segundo agente de nucleasa es una meganucleasa.

En una realización, el primer o el segundo agente de nucleasa comprende una proteína (Cas) asociada a repeticiones palindrómicas cortas intercaladas regularmente y agrupadas (CRISPR) y un ARN guía (ARNg). En una realización, el ARN guía (ARNg) comprende (a) un ARN de Repeticiones palindrómicas cortas intercaladas regularmente y agrupadas (CRISPR) (ARNcr) que se dirige al primero, al segundo o al tercer sitios de reconocimiento; y (b) un ARN CRISPR transactivante (tracrARN). En una realización, el primer o el segundo sitios de reconocimiento están flanqueados inmediatamente por una secuencia de un motivo adyacente a protoespaciador (PAM). En una realización, el locus genómico de interés comprende la secuencia de nucleótidos de SEQ ID NO: 1. En una realización, la proteína Cas es Cas9. En una realización, el ARNg comprende: (a) el ARN quimérico de la secuencia de ácido nucleico de SEQ ID NO: 2; o (b) el ARN quimérico de la secuencia de ácido nucleico de SEQ ID NO: 3. En una realización, el ARNcr comprende la SEQ ID NO: 4; la SEQ ID NO: 5; o la SEQ ID NO: 6. En una realización, el tracrARN comprende la SEQ ID NO: 7 o la SEQ

ID NO: 8. En una realización, el primero, el segundo y/o el tercer sitio de reconocimiento están ubicados en un intrón, un exón, un promotor, una región reguladora del promotor o una región potenciadora del primer, el segundo o el tercer marcador de selección. En una realización, el primer sitio objetivo y el segundo sitio objetivo están inmediatamente adyacentes al primer sitio de reconocimiento. En una realización, el primer sitio objetivo y el segundo sitio objetivo son de aproximadamente 10 nucleótidos a aproximadamente 14 kb a partir del primer sitio de reconocimiento. En una realización, el tercer sitio objetivo y el cuarto sitio objetivo están inmediatamente adyacentes al segundo sitio de reconocimiento. En una realización, el tercer sitio objetivo y el cuarto sitio objetivo son de aproximadamente 10 nucleótidos a aproximadamente 14 kb a partir del segundo sitio de reconocimiento. En una realización, una suma total del primer brazo de homología y el segundo brazo de homología es de al menos aproximadamente 10 kb. En una realización, una suma total del tercer brazo de homología y el cuarto brazo de homología es de al menos aproximadamente 10 kb. En una realización, el primer polinucleótido de inserción varía de aproximadamente 5 kb a aproximadamente 300 kb de longitud. En una realización, el segundo polinucleótido de inserción varía de aproximadamente 5 kb a aproximadamente 300 kb de longitud. En una realización, la integración del primer polinucleótido de inserción en el locus 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 estos. En una realización, la integración del segundo polinucleótido de inserción en el locus 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 estos. En una realización, el primer polinucleótido de inserción comprende un polinucleótido de interés que comprende un polinucleótido humano. En una realización, el segundo polinucleótido de inserción comprende un polinucleótido de interés que comprende un polinucleótido humano. En una realización, el primer polinucleótido de inserción comprende un polinucleótido de interés que comprende una región del locus del receptor alfa de células T. En una realización, el segundo polinucleótido de inserción comprende un polinucleótido de interés que comprende una región del locus del receptor alfa de células T. En una realización, el primer o el segundo polinucleótido de inserción comprende un polinucleótido de interés que comprende al menos un segmento génico de región variable y/o un segmento génico de región de unión del locus del receptor alfa de células T. En una realización, la región del locus del receptor alfa de células T es de un ser humano. En una realización, el primer polinucleótido de inserción comprende un polinucleótido de interés que comprende una secuencia de ácido nucleico de una región variable de cadena pesada de inmunoglobulina humana no reorganizada, unida operativamente a una secuencia de ácido nucleico de una región constante de cadena pesada de inmunoglobulina no humana. En una realización, la etapa de identificación se lleva a cabo mediante un ensayo de modificación de alelos (MOA). En una realización, en donde el primer polinucleótido de inserción comprende un polinucleótido de interés que comprende una secuencia de ácido nucleico que es homóloga u ortóloga a la secuencia de ácido nucleico en un genoma de la célula. En una realización, el segundo polinucleótido de inserción comprende una secuencia de ácido nucleico que es homóloga u ortóloga a la secuencia de ácido nucleico en un genoma de la célula. En una realización, el primer polinucleótido de inserción comprende un polinucleótido de interés que comprende una secuencia de ácido nucleico exógena. En una realización, el segundo polinucleótido de inserción comprende un polinucleótido de interés que comprende una secuencia de ácido nucleico exógena.

Breve descripción de las figuras

La figura 1 proporciona un esquema de un evento de transformación genómica en el que una célula que tiene una modificación heterocigota del locus de TCR alfa en el cromosoma 14 de ratón, un alelo del cual es el alelo de TCR alfa humanizado A-neo, que comprende un casete de selección para neomicina ubicado hacia el extremo 5' de ocho segmentos de genes variables (V) humanos y 61 segmentos de genes de unión (J) humanos, es transformada con un vector de transformación del alelo de TCR alfa humanizado B-hyg, que comprende un casete de selección para higromicina y un fragmento de más de 100 kb que comprende 11 segmentos de genes variables humanos adicionales. La electroporación del vector de transformación del alelo B-hyg y los plásmidos que expresan las dos mitades de un par de nucleasas de dedos de zinc (ZFN) que se dirige al casete de neomicina en el alelo de TCR alfa A-neo generó un locus de TCR alfa modificado (alelo B-hyg) que comprende, de 5' a 3', un casete de higromicina, 19 segmentos de genes V humanos y 61 segmentos de genes J humanos ubicados hacia el extremo 5' de la secuencia de nucleótidos de la región constante endógena. El evento de transformación insertó con precisión más de 100 kb de la secuencia del gen de TCR alfa humano en el locus de TCR alfa de ratón.

La figura 2 proporciona un esquema de un evento de transformación genómica en el que una célula que tiene una modificación heterocigota del locus de TCR alfa en el cromosoma 14 de ratón, un alelo del cual es el alelo de TCR alfa humanizado B-hyg, que comprende un casete de selección para higromicina ubicado hacia el extremo 5' de 19 segmentos de genes V humanos y 61 segmentos de genes J humanos es transformada con un vector de transformación del alelo de TCR alfa humanizado C-neo, que comprende un casete de selección para neomicina y un fragmento de más de 100 kb que comprende 11 segmentos de genes variables humanos adicionales. La electroporación del vector de transformación del alelo C-neo y los plásmidos que expresan las dos mitades de un par de nucleasas de dedos de zinc (ZFN) que se dirigen al casete de higromicina en el alelo de TCR alfa B-hyg generó un locus de TCR alfa modificado (alelo C-neo) que comprende, de 5' a 3', un casete

de neomicina, 30 segmentos de genes V humanos y 61 segmentos de genes J humanos ubicados hacia el extremo 5' de la secuencia de nucleótidos de la región constante endógena. El evento de transformación insertó con precisión más de 100 kb de la secuencia del gen de TCR alfa humano en el locus de TCR alfa de ratón.

La figura 3 proporciona una representación esquemática de los casetes de selección por fármacos *neo^r*, que codifica la neomicina fosfotransferasa, e *hyg^r*, que codifica la higromicina B fosfotransferasa. Las posiciones de los sitios de reconocimiento (secuencias dadas a continuación) para las nucleasas de dedos de zinc Neo-ZFN(1,2) y Neo-ZFN(3,4) (ZFN, Fig. 3A) que se dirigen a *neo^r* y las ZFN Hyg-ZFN(1,2) e Hyg-ZFN(3,4) (Fig. 3B) que se dirigen a *hyg^r* se indican mediante cuadros sombreados arriba o abajo de las flechas gruesas que representan las secuencias codificantes de fosfotransferasa respectivas.

10 Descripción detallada

La presente invención se refiere a las realizaciones tal como se caracterizan en las reivindicaciones.

Las presentes invenciones se describirán ahora más completamente a continuación con referencia a los dibujos adjuntos, en los que se muestran algunas, pero no todas las realizaciones de la invención. De hecho, estas invenciones pueden incorporarse en varias formas diferentes y no deben interpretarse como limitadas a las realizaciones que se exponen en la presente descripción; más bien, estas realizaciones se proporcionan para que esta descripción satisfaga los requisitos legales aplicables. Los números iguales se refieren a elementos iguales a todo lo largo.

Varias modificaciones y otras realizaciones de las invenciones que se exponen en la presente descripción serán evidentes para un experto en la técnica a la cual pertenece esta invención, que tienen el beneficio de las enseñanzas presentadas en las descripciones anteriores y los dibujos asociados. Por lo tanto, se debe entender que la invención no se limita a las realizaciones específicas descritas y que las modificaciones y otras realizaciones se incluyen dentro del alcance de las reivindicaciones adjuntas. Aunque términos específicos se emplean en la presente descripción, ellos se usan en un sentido genérico y descriptivo solamente y no para propósitos de limitación.

25 I. Ejemplo general

Se proporcionan métodos para modificar un locus objetivo, por ejemplo, un locus genómico, en una célula. Los métodos emplean agentes de nucleasa y sitios de reconocimiento de agentes de nucleasa para mejorar los eventos de recombinación homóloga de un polinucleótido de inserción en el locus objetivo. Los diversos métodos proporcionados en la presente descripción ubican estratégicamente el sitio de reconocimiento del agente de nucleasa dentro de un polinucleótido que codifica un marcador de selección, un reportero o una proteína exógena (por ejemplo, eGFP o una secuencia humana en una célula de ratón).

Se proporcionan métodos que permiten la modificación en serie (es decir, en mosaico) de polinucleótidos de interés en un locus objetivo (es decir, un locus genómico). Como se explica con más detalle a continuación, se proporcionan métodos para agrupar polinucleótidos de interés en mosaico en un locus objetivo (es decir, un locus genómico) en donde el locus objetivo (es decir, un locus genómico) y los diversos vectores de transformación empleados en el método alternan el uso de un primer marcador de selección que comprende un primer sitio de reconocimiento para un primer agente de nucleasa y un segundo marcador de selección que comprende un segundo sitio de reconocimiento para un segundo agente de nucleasa. Al hacerlo, el método no requiere un suministro constante de nucleasas diseñadas para reconocer nuevos sitios de reconocimiento. En cambio, en realizaciones específicas, la modificación en serie dirigida solo requiere dos agentes de nucleasa y el sitio de reconocimiento correspondiente de los dos agentes de nucleasa. Además, dado que los agentes de nucleasa se dirigen a secuencias exógenas (es decir, el sitio de reconocimiento dentro de un polinucleótido que codifica un marcador de selección) y dado que la eficacia y el efecto por acción fuera del objetivo de cualquier sitio de reconocimiento dado se habrá confirmado previamente, la escisión no específica de una secuencia genómica endógena puede minimizarse a la vez que aumenta el tiempo y la rentabilidad del proceso de modificación en mosaico.

II. Sistema de integración dirigido

Se proporcionan métodos para modificar un locus objetivo en una célula. El sistema emplea agentes de nucleasa, sitios de reconocimiento para el agente de nucleasa, un locus objetivo, marcadores de selección, vectores de transformación y polinucleótidos de inserción. Cada uno de estos componentes se describe con más detalle más abajo.

A. Agentes de nucleasa y sitios de reconocimiento para los agentes de nucleasa

El término "sitio de reconocimiento para un agente de nucleasa" incluye una secuencia de ADN en la que un agente de nucleasa induce un corte o ruptura de doble cadena. El sitio de reconocimiento para un agente de nucleasa puede ser endógeno (o nativo) a la célula o el sitio de reconocimiento puede ser exógeno a la célula. En realizaciones específicas, el sitio de reconocimiento es exógeno a la célula y, por lo tanto, no ocurre

naturalmente en el genoma de la célula. En otras realizaciones adicionales, el sitio de reconocimiento es exógeno a la célula y a los polinucleótidos de interés que se desean colocar en el locus objetivo. En realizaciones adicionales, el sitio de reconocimiento exógeno o endógeno está presente solo una vez en el genoma de la célula huésped. En realizaciones específicas, se identifica un sitio endógeno o nativo que está presente solo una vez dentro del genoma. Dicho sitio se puede usar para diseñar agentes de nucleasa que producirán un corte o ruptura de doble cadena en el sitio de reconocimiento endógeno.

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

Cualquier agente de nucleasa que induzca un corte o ruptura de doble cadena en un sitio de reconocimiento deseado puede usarse en los métodos y composiciones descritos en esta descripción. Se puede emplear un agente de nucleasa natural o nativo siempre que el agente de nucleasa induzca un corte o ruptura de doble cadena en un sitio de reconocimiento deseado. Alternativamente, se puede emplear un agente de nucleasa diseñado genéticamente o modificado. Un "agente de nucleasa diseñado genéticamente" incluye una nucleasa que se diseña (se modifica o deriva) de su forma nativa para reconocer e inducir específicamente un corte o ruptura de doble cadena en el sitio de reconocimiento deseado. Por lo tanto, un agente de nucleasa diseñado genéticamente puede derivarse de un agente de nucleasa natural o puede crearse o sintetizarse artificialmente. La modificación del agente de nucleasa puede ser tan pequeña como un aminoácido en un agente de escisión de proteínas o un nucleótido en un agente de escisión de ácido nucleico. En algunas realizaciones, la nucleasa diseñada genéticamente induce un corte o ruptura de doble cadena en un sitio de reconocimiento, en donde el sitio de reconocimiento no es una secuencia que hubiera sido reconocida por un agente de nucleasa nativo (no diseñado genéticamente o modificado). Producir un corte o ruptura de doble cadena en un sitio de reconocimiento u otro ADN puede denominarse en la presente descripción "cortar" o "escindir" el sitio de reconocimiento u otro ADN.

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

El sitio de reconocimiento se ubica dentro del polinucleótido que codifica el marcador de selección. Dicha posición puede ubicarse dentro de la región codificante del marcador de selección o dentro de las regiones reguladoras, que influyen en la expresión del marcador de selección. Por lo tanto, un sitio de reconocimiento del agente de nucleasa puede ubicarse en un intrón del marcador de selección, un promotor, un potenciador, una región reguladora o cualquier región no codificante de proteínas del polinucleótido que codifica el marcador de selección. En realizaciones específicas, un corte o ruptura de doble cadena en el sitio de reconocimiento interrumpe la actividad del marcador de selección. Se conocen métodos para analizar la presencia o ausencia de un marcador de selección funcional.

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

Se describen ejemplos de nucleasas de tipo TAL adecuadas y métodos para preparar nucleasas de tipo TAL adecuadas, por ejemplo, en las solicitudes de patente de EE. UU. núms. 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 de tipo TAL están diseñadas para cortar en o cerca de una secuencia de ácido nucleico objetivo, *por ejemplo*, en un locus de interés o un locus genómico de interés, en donde la secuencia de ácido nucleico objetivo está en o cerca de una secuencia que será modificada por un vector de transformación. Las nucleasas de tipo TAL adecuadas para su uso con los

diversos métodos y composiciones proporcionados en la presente descripción incluyen aquellas que están diseñadas específicamente para unirse a secuencias de ácido nucleico objetivo o cerca de ellas para ser modificadas por vectores de transformación como se describe en la presente descripción.

En una realización, cada monómero de TALEN comprende 33-35 repeticiones TAL que reconocen un solo par de bases a través de dos residuos hipervariables. En una realización, el agente de nucleasa es una proteína química que comprende un dominio de unión a ADN basado en repeticiones TAL unido operativamente a una nucleasa independiente. En una realización, la nucleasa independiente es una endonucleasa FokI. En una realización, el agente de nucleasa comprende un primer dominio de unión a ADN basado en repeticiones TAL y un segundo dominio de unión a ADN basado en repeticiones TAL, en donde cada uno del primer y segundo dominios de unión a ADN basados en repeticiones TAL está unido operativamente a una subunidad de nucleasa FokI, en donde el primer y el segundo dominios de unión a ADN basados en repeticiones TAL reconocen dos secuencias de ADN objetivo contiguas en cada cadena de la secuencia de ADN objetivo separadas por una secuencia espaciadora de longitud variable (12-20 pb), y en donde las subunidades de nucleasa FokI se dimerizan para crear una nucleasa activa que realiza una ruptura de doble cadena en una secuencia objetivo.

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

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

Cualquier meganucleasa puede usarse en esta descripción, incluyendo, entre otras, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIIVP, I-TII, I-PpII, PI-PsPI, F-SceI, F-SceII, F-SuVI, F-TevI, F-TevII, I-Amal, I-Anil, I-ChuI, I-Cmoel, I-Cpal, I-Cpall, I-Csml, I-Cvul, I-CvuAIP, I-DdII, I-DdIII, I-Dirl, I-Dmol, I-Hmul, I-HmulI, I-HsNIP, I-LlAI, I-Msol, I-Naal, I-NanI, I-NcII, I-NgrIP, I-NitI, I-NjAI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbplIP, I-SpBetaIP, I-SceI, I-SexIP, I-SneiP, I-SpomI, I-SpomCP, I-SpomIIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdelIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbilP, PI-MtuI, PI-MtuHIP, PI-MtuHIIP, PI-Pful, PI-Pfull, PI-Pkol, PI-Pkoll, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-Tful, PI-Tfull, PI-Thyl, PI-TIil, PI-TIilI, o cualquier variante activa o fragmento de estas.

En una realización, la meganucleasa reconoce secuencias de ADN bicatenario 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 dirigida. En una realización, la nucleasa dirigida es una familia LAGLIDADG de nucleasa dirigida. En una realización, la familia LAGLIDADG de nucleasa dirigida se selecciona de I-SceI, I-CreI e I-Dmol.

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

El agente de nucleasa empleado en los diversos métodos y composiciones también puede comprender un sistema CRISPR/Cas. Dichos sistemas pueden emplear una nucleasa Cas9, que en algunos casos, tiene codones optimizados para el tipo de célula deseado en donde se va a expresar. El sistema emplea además una construcción de ARNcr-tracrARN fusionada que funciona con la Cas9 con codones optimizados. Este ARN único a menudo se denomina ARN guía o ARNg. Dentro de un ARNg, la porción de ARNcr se identifica como la "secuencia objetivo" para el sitio de reconocimiento dado y el tracrARN a menudo se denomina "soporte". Se ha demostrado que este sistema funciona en una variedad de células eucariotas y procariotas. Brevemente, un fragmento corto de ADN que contiene la secuencia objetivo se inserta en un plásmido de expresión del ARN guía. El plásmido de expresión del ARNg comprende la secuencia objetivo (en algunas realizaciones alrededor de 20 nucleótidos), una forma de la secuencia de tracrARN (el soporte), así como un promotor adecuado que es activo en la célula y elementos necesarios para el procesamiento adecuado en células eucariotas. Muchos de los sistemas se basan en oligos complementarios personalizados que hibridan para formar un ADN bicatenario y después se clonan en el plásmido de expresión de ARNg. Después el casete de expresión de ARNg y el casete de expresión de Cas9 se introducen en la célula. Ver, por ejemplo, Mali P. y otros (2013) *Science* 15 de febrero de 2013; 339 (6121):823-6; Jinek M y otros *Science*, 17 de agosto de 2012; 337(6096): 816-21; Hwang WY y otros *Nat Biotechnol*, 2013 Mar; 31(3): 227-9; Jiang W y otros *Nat Biotechnol*, 2013 Mar; 31(3): 233-9; y Cong L y otros *Science*, 15 de febrero de 2013;339(6121):819-23.

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

Algunos sistemas CRISPR/Cas utilizados en los métodos descritos en la presente descripción no son de origen natural. Un sistema "no natural" incluye cualquier cosa que indique la participación de la acción del hombre, como uno o más componentes del sistema que estén alterados o mutados con relación a su estado natural, al menos sustancialmente libres de al menos otro componente con el que están naturalmente asociados en la naturaleza, o que estén asociados con al menos otro componente con el que no están naturalmente asociados. Por ejemplo, algunos sistemas CRISPR/Cas emplean complejos CRISPR no naturales que comprenden un ARNg y una proteína Cas que no se encuentran juntos de manera natural.

i. Endonucleasas Cas guiadas por ARN

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

Los ejemplos de proteínas Cas incluyen Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e (CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9 (Csn1 o Csx12), Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (CasA), Cse2 (CasB), Cse3 (CasE), Cse4 (CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr2, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, y Cu1966, y homólogos o sus versiones modificadas.

Las proteínas Cas pueden ser de un sistema CRISPR/Cas tipo II. Por ejemplo, la proteína Cas puede ser una proteína Cas9 o derivarse de una proteína Cas9. Las proteínas Cas9 suelen compartir cuatro motivos clave con una arquitectura conservada. Los motivos 1, 2 y 4 son motivos similares a RuvC, y el motivo 3 es un motivo HNH. La proteína Cas9 puede ser de, por ejemplo, *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus sp.*, *Staphylococcus aureus*, *Nocardiopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycoides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microcilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas sp.*, *Crocospaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor beccsii*, *Candidatus Desulforudis*, *Clostridium botulinum*, *Clostridium difficile*, *Finegoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platenis*, *Arthrospira sp.*, *Lyngbya sp.*, *Microcoleus chthonoplastes*, *Oscillatoria sp.*, *Petrotoga mobilis*, *Thermosiphon africanus*, o *Acaryochloris marina*. Otros ejemplos de los miembros de la familia Cas9 se describen en el documento WO 2014/131833. Una enzima preferida es la proteína Cas9 de *S. pyogenes* o derivada de esta. La proteína Cas9 de *S. pyogenes* tiene asignado el número de acceso de SwissProt Q99ZW2.

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

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

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

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

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

Una proteína Cas puede fusionarse con un polipéptido heterólogo que proporciona una localización subcelular. Dichos péptidos heterólogos incluyen, por ejemplo, una señal de localización nuclear (NLS) tal como la NLS de SV40 para dirigirse al núcleo, una señal de localización mitocondrial para dirigirse a las mitocondrias, una señal de retención en ER, y similares. Ver, por ejemplo, Lange y otros (2007) J. Biol. Chem. 282:5101-5105. Dichas señales de localización subcelular pueden ubicarse en el extremo N-terminal, el C-terminal o en cualquier lugar dentro de la proteína Cas. Una NLS puede comprender un tramo de aminoácidos básicos y puede ser una secuencia monopartita o una secuencia bipartita.

Las proteínas Cas también se pueden unir a un dominio de penetración celular. Por ejemplo, el dominio de penetración celular puede derivarse de la proteína TAT del VIH-1, el motivo de penetración celular TLM del virus de la hepatitis B humana, MPG, Pep-1, VP22, un péptido penetrante de células del virus del herpes simple o una secuencia peptídica de poliarginina. Ver, por ejemplo, el documento WO 2014/089290. El dominio de penetración celular puede ubicarse en el extremo N terminal, C terminal o en cualquier lugar dentro de la proteína Cas.

Las proteínas Cas también pueden comprender un polipéptido heterólogo para facilitar el seguimiento o la purificación, como una proteína fluorescente, una etiqueta de purificación o una etiqueta de epítipo. Los ejemplos de proteínas fluorescentes incluyen proteínas fluorescentes verdes (por ejemplo, GFP, GFP-2, tagGFP, turboGFP, eGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), proteínas fluorescentes amarillas (por ejemplo, YFP, eYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1), proteínas fluorescentes azules (por ejemplo, eBFP, eBFP2, Azurite, mKalamal, GFPuv, Sapphire, T-sapphire), proteínas fluorescentes cian (por ejemplo, eCFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), proteínas fluorescentes rojas (mKate, mKate2, mPlum, monómero de DsRed, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred) proteínas fluorescentes naranjas (mOrange, mKO, Kusabira-Orange, Kusabira-Orange monomérico, mTangerine, tdTomato) y cualquier otra proteína fluorescente adecuada. Los ejemplos de etiquetas incluyen glutatión-S-transferasa (GST), proteína de unión a quitina (CBP), proteína de unión a maltosa, tioredoxina (TRX), poli(NANP), etiqueta de purificación de afinidad en tándem (TAP), myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, hemaglutinina (HA), nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, histidina (His), proteína portadora de biotina carboxilo (BCCP) y calmodulina.

Las proteínas Cas se pueden proporcionar en cualquier forma. Por ejemplo, se puede proporcionar una proteína Cas en forma de una proteína, tal como una proteína Cas formando un complejo con un ARNg. Alternativamente, se puede proporcionar una proteína Cas en forma de un ácido nucleico que codifica la proteína Cas, tal como un ARN (por ejemplo, ARN mensajero (ARNm)) o ADN. Opcionalmente, los codones del ácido nucleico que codifica la proteína Cas pueden optimizarse para una traducción eficiente a proteína en una célula u organismo particular.

Los ácidos nucleicos que codifican las proteínas Cas pueden integrarse de manera estable en el genoma de la célula y unirse operativamente a un promotor activo en la célula. Alternativamente, los ácidos nucleicos que codifican las proteínas Cas pueden unirse operativamente a un promotor en una construcción de expresión. Las construcciones de expresión incluyen cualquier construcción de ácido nucleico capaz de dirigir la expresión de un gen u otra secuencia de ácido nucleico de interés (por ejemplo, un gen Cas) y que puede transferir dicha secuencia de ácido nucleico de interés a una célula objetivo. Los promotores que pueden usarse en una construcción de expresión incluyen, por ejemplo, promotores activos en una célula pluripotente de rata, eucariota, mamífero no humano, humano, roedor, ratón o hámster. Ejemplos de otros promotores se describen en otra parte de la presente descripción.

ii. ARN guías (ARNg)

Un "ARN guía" o "ARNg" incluye una molécula de ARN que se une a una proteína Cas y dirige la proteína Cas a una ubicación específica dentro de un ADN objetivo. Los ARN guías pueden comprender dos segmentos: un "segmento dirigido al ADN" y un "segmento de unión a proteínas". "Segmento" incluye un segmento, sección o región de una molécula, como un tramo contiguo de nucleótidos en un ARN. Algunos ARNg comprenden dos moléculas de ARN separadas: un "ARN activador" y un "ARN que dirige al objetivo". Otros ARNg son una molécula de ARN simple (polinucleótido de ARN único), que también se puede llamar un "ARNg de molécula única", un "ARN guía único" o un "ARNsg". Ver, por ejemplo, los documentos WO/2013/176772A1, WO/2014/065596A1, WO/2014/089290A1, WO/2014/093622A2, WO/2014/099750A2, WO/2013142578A1 y WO 2014/131833A1. Los términos "ARN guía" y "ARNg" incluyen tanto ARNg de doble molécula como ARNg de una sola molécula.

Un ejemplo de ARNg de dos moléculas comprende una molécula similar a un ARNcr ("ARN CRISPR" o "ARN que dirige al objetivo" o "ARNcr" o "repetición de ARNcr") y una molécula similar al tracrARN correspondiente ("ARN CRISPR transfuncional" o "ARN activador" o "tracrARN" o "soporte"). Un ARNg comprende tanto el segmento dirigido al ADN (monocatenario) del ARNg como un tramo de nucleótidos que forma la mitad del híbrido de ARNbc del segmento de unión a proteínas del ARNg.

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

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

El segmento dirigido al ADN (ARNcr) de un ARNg dado comprende una secuencia de nucleótidos que es complementaria a una secuencia en un ADN objetivo. El segmento dirigido al ADN de un ARNg interactúa con un ADN objetivo de una manera específica de secuencia mediante hibridación (es decir, apareamiento de bases). Como tal, la secuencia de nucleótidos del segmento dirigido al ADN puede variar y determina la ubicación dentro del ADN objetivo con el que interactuarán el ARNg y el ADN objetivo. El segmento dirigido al ADN de un ARNg sujeto puede modificarse para que hibride con cualquier secuencia deseada dentro de un ADN objetivo. Los ARNcr de origen natural difieren según el sistema Cas9 y el organismo, pero a menudo contienen un segmento de direccionamiento de entre 21 y 72 nucleótidos de longitud, flanqueado por dos repeticiones directas (DR) de una longitud de entre 21 y 46 nucleótidos (ver, por ejemplo, el documento WO2014/131833). En el caso de *S. pyogenes* las DR tienen 36 nucleótidos de largo y el segmento de direccionamiento tiene 30 nucleótidos de largo. La DR localizada en 3' es complementaria e hibrida con el correspondiente tracrARN, que a su vez se une a la proteína Cas9.

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

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

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

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

El segmento de unión a proteínas de un ARNg puede comprender dos tramos de nucleótidos que son complementarios entre sí. Los nucleótidos complementarios del segmento de unión a proteínas se hibridan para formar un híbrido de ARN bicatenario (ARNbc). El segmento de unión a proteínas de un ARNg sujeto interactúa con una proteína Cas, y el ARNg dirige la proteína Cas unida a una secuencia de nucleótidos específica dentro del ADN objetivo a través del segmento dirigido al ADN.

Los ARN guía pueden incluir modificaciones o secuencias que proporcionan características deseables adicionales (por ejemplo, estabilidad modificada o regulada; direccionamiento subcelular; seguimiento con un marcador fluorescente; un sitio de unión para una proteína o complejo proteico; y similares). Los ejemplos de tales modificaciones incluyen, por ejemplo, una caperuza en 5' (por ejemplo, una caperuza de 7-metilguanilato (m7G)); una cola poliadenilada en 3' (es decir, una cola de poli(A) en 3'); una secuencia de ribointerruptor (por ejemplo, para permitir estabilidad regulada y/o accesibilidad regulada por proteínas y/o complejos proteicos); una secuencia de control de estabilidad; una secuencia que forma un híbrido de ARNbc (es decir, una horquilla); una modificación o secuencia que dirige el ARN a una ubicación subcelular (por ejemplo, núcleo, mitocondrias, cloroplastos y similares); una modificación o secuencia que proporciona seguimiento (por ejemplo, conjugación directa a una molécula fluorescente, conjugación a un resto que facilita la detección fluorescente, una secuencia que permite la detección fluorescente, etcétera); una modificación o secuencia que proporciona un sitio de unión para proteínas (por ejemplo, proteínas que actúan sobre el ADN, incluidos activadores transcripcionales, represores transcripcionales, ADN metiltransferasas, ADN desmetilasas, histona acetiltransferasas, histona desacetilasas y similares); y combinaciones de estos.

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

Los ADN que codifican los ARNg pueden integrarse de manera estable en el genoma de la célula y unirse operativamente a un promotor activo en la célula. Alternativamente, los ADN que codifican los ARNg se pueden unir operativamente a un promotor en una construcción de expresión. Dichos promotores pueden ser activos, por ejemplo, en una célula pluripotente de rata, eucariota, mamífero, mamífero no humano, humano, roedor, ratón o hámster. En algunos casos, el promotor es un promotor de ARN polimerasa III, tal como un promotor U6 humano, un promotor U6 polimerasa III de rata o un promotor U6 polimerasa III de ratón. Ejemplos de otros promotores se describen en otra parte de la presente descripción.

Alternativamente, los ARNg se pueden preparar por varios otros métodos. Por ejemplo, los ARNg pueden prepararse mediante transcripción *in vitro* con el uso, por ejemplo, de ARN polimerasa T7 (ver, por ejemplo, los

documentos WO 2014/089290 y WO 2014/065596). Los ARN guía también pueden ser una molécula producida sintéticamente preparada por síntesis química.

iii. Secuencias de reconocimiento de ARN CRISPR

- El término "secuencia de reconocimiento de ARN CRISPR" incluye secuencias de ácido nucleico presentes en un ADN objetivo al que se unirá un segmento dirigido al ADN de un ARNg, siempre que existan condiciones suficientes para la unión. Por ejemplo, las secuencias de reconocimiento de ARN CRISPR incluyen secuencias para las cuales se diseña un ARN guía para que tengan complementariedad, donde la hibridación entre una secuencia de reconocimiento de ARN CRISPR y una secuencia dirigida al ADN promueve la formación de un complejo CRISPR. No se requiere necesariamente una completa complementariedad, siempre que haya suficiente complementariedad para causar hibridación y promover la formación de un complejo CRISPR. Las secuencias de reconocimiento de ARN CRISPR también incluyen sitios de escisión para proteínas Cas, que se describen con más detalle a continuación. Una secuencia de reconocimiento de ARN CRISPR puede comprender cualquier polinucleótido, que puede ubicarse, por ejemplo, en el núcleo o citoplasma de una célula o dentro de un orgánulo de una célula, como una mitocondria o cloroplasto.
- La secuencia de reconocimiento de ARN CRISPR dentro de un ADN objetivo puede ser seleccionada por (es decir, puede unirse a o hibridarse con, o ser complementaria a) una proteína Cas o un ARNg. Las condiciones adecuadas para la unión de ADN/ARN incluyen condiciones fisiológicas normalmente presentes en una célula. En la técnica se conocen otras condiciones adecuadas para la unión de ADN/ARN (por ejemplo, condiciones en un sistema libre de células) (véase, por ejemplo, *Molecular Cloning: A Laboratory Manual*, 3ra Ed. (Sambrook y otros, Harbor Laboratory Press 2001)). La cadena del ADN objetivo que es complementaria e hibrida con la proteína Cas o el ARNg se puede denominar "cadena complementaria" y la cadena del ADN objetivo que es complementaria a la "cadena complementaria" (y por lo tanto no es complementaria a la proteína Cas o ARNg) puede denominarse "cadena no complementaria" o "cadena molde".
- La proteína Cas puede escindir el ácido nucleico en un sitio dentro o fuera de la secuencia de ácido nucleico presente en el ADN objetivo al que se unirá el segmento dirigido al ADN de un ARNg. El "sitio de escisión" incluye la posición de un ácido nucleico en el que una proteína Cas produce una ruptura de cadena sencilla o una ruptura de cadena doble. Por ejemplo, la formación de un complejo CRISPR (que comprende un ARNg hibridado con una secuencia de reconocimiento de ARN CRISPR y en complejo con una proteína Cas) puede dar como resultado la escisión de una o ambas cadenas en o cerca (por ejemplo, dentro de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50 o más pares de bases) de la secuencia de ácido nucleico presente en un ADN objetivo al que se unirá un segmento dirigido al ADN de un ARNg. Si el sitio de escisión está fuera de la secuencia de ácido nucleico a la que se unirá el segmento dirigido al ADN del ARNg, el sitio de escisión todavía se considera dentro de la "secuencia de reconocimiento de ARN CRISPR". El sitio de escisión puede estar en una sola cadena o en ambas cadenas de un ácido nucleico. Los sitios de escisión pueden estar en la misma posición en ambas cadenas del ácido nucleico (produciendo extremos romos) o pueden estar en diferentes sitios en cada cadena (produciendo extremos escalonados). Los extremos escalonados se pueden producir, por ejemplo, mediante el uso de dos proteínas Cas, cada una de las cuales produce una ruptura de cadena sencilla en un sitio de escisión diferente en cada cadena, produciendo así una ruptura de cadena doble. Por ejemplo, una primera nickasa puede crear una ruptura monocatenaria en la primera cadena del ADN bicatenario (ADNbc), y una segunda nickasa puede crear una ruptura monocatenaria en la segunda cadena del ADNbc de manera que se crean secuencias sobresalientes. En algunos casos, la secuencia de reconocimiento de ARN CRISPR de la nickasa en la primera cadena está separada de la secuencia de reconocimiento de ARN CRISPR de la nickasa en la segunda cadena por al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500 o 1000 pares de bases.
- La escisión específica de un sitio del ADN objetivo por Cas9 puede ocurrir en lugares determinados por (i) complementariedad de apareamiento de bases entre el ARNg y el ADN objetivo y (ii) un motivo corto, llamado motivo adyacente a protoespaciador (PAM), en el ADN objetivo. El PAM puede flanquear la secuencia de reconocimiento de ARN CRISPR. Opcionalmente, la secuencia de reconocimiento de ARN CRISPR puede estar flanqueada por el PAM. Por ejemplo, el sitio de escisión de Cas9 puede ser de aproximadamente 1 a aproximadamente 10 o de aproximadamente 2 a aproximadamente 5 pares de bases (por ejemplo, 3 pares de bases) hacia el extremo 5' o hacia el extremo 3' de la secuencia PAM. En algunos casos (por ejemplo, cuando se usa Cas9 de *S. pyogenes* o un Cas9 estrechamente relacionado), la secuencia PAM de la cadena no complementaria puede ser 5'-N₁GG-3', donde N₁ es cualquier nucleótido de ADN y está inmediatamente en 3' de la secuencia de reconocimiento de ARN CRISPR de la cadena no complementaria del ADN objetivo. Como tal, la secuencia PAM de la cadena complementaria sería 5'-CC N₂-3', donde N₂ es cualquier nucleótido de ADN y está inmediatamente en 5' de la secuencia de reconocimiento de ARN CRISPR de la cadena complementaria del ADN objetivo. En algunos de estos casos, N₁ y N₂ pueden ser complementarios y el par de bases N₁- N₂ puede ser cualquier par de bases (por ejemplo, N₁=C y N₂=G; N₁=G y N₂=C; N₁=A y N₂=T, N₁=T y N₂=A).
- Los ejemplos de secuencias de reconocimiento de ARN CRISPR incluyen una secuencia de ADN complementaria al segmento dirigido al ADN de un ARNg, o una secuencia de ADN de este tipo además de

una secuencia PAM. Por ejemplo, el motivo objetivo puede ser una secuencia de ADN de 20 nucleótidos que precede inmediatamente a un motivo NGG reconocido por una proteína Cas (véase, por ejemplo, el documento WO 2014/165825). La guanina en el extremo 5' puede facilitar la transcripción por la ARN polimerasa en las células. Otros ejemplos de secuencias de reconocimiento de ARN CRISPR pueden incluir dos nucleótidos de guanina en el extremo 5' (por ejemplo, GGN₂₀NGG; SEQ ID NO: 10) para facilitar la transcripción eficiente por la polimerasa T7 *in vitro*. Ver, por ejemplo, el documento WO 2014/065596.

La secuencia de reconocimiento de ARN CRISPR puede ser cualquier secuencia de ácido nucleico endógena o exógena a una célula. La secuencia de reconocimiento de ARN CRISPR puede ser una secuencia que codifica un producto génico (por ejemplo, una proteína) o una secuencia no codificante (por ejemplo, una secuencia reguladora) o puede incluir ambas. En una realización, la secuencia objetivo está flanqueada inmediatamente por una secuencia de un motivo adyacente a protoespaciador (PAM). En una realización, el locus de interés comprende la secuencia de nucleótidos de SEQ ID NO: 1. En una realización, el ARNg comprende una tercera secuencia de ácido nucleico que codifica un ARN de Repeticiones Palindrómicas Cortas Intercaladas Regularmente y Agrupadas (CRISPR) (ARNcr) y un ARN CRISPR transactivante (tracrARN). En otra realización, el genoma de la célula pluripotente de rata comprende una región de ADN objetivo complementaria a la secuencia objetivo. En algunos de estos métodos, la proteína Cas es Cas9. En algunas realizaciones, el ARNg comprende (a) el ARN quimérico de la secuencia de ácido nucleico de SEQ ID NO: 2; o (b) el ARN quimérico de la secuencia de ácido nucleico de SEQ ID NO: 3. En algunos de estos métodos, el ARNcr comprende la secuencia establecida en SEQ ID NO: 4, la SEQ ID NO: 5, o la SEQ ID NO: 6. En algunos de estos métodos, el tracrARN comprende la secuencia establecida en SEQ ID NO: 7 o la SEQ ID NO: 8.

También se proporcionan variantes y fragmentos activos de agentes de nucleasa (es decir, un agente de nucleasa diseñado genéticamente). Dichas variantes activas pueden comprender al menos 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más de identidad de secuencia con el agente de nucleasa nativo, en donde las variantes activas retienen la capacidad de cortar en un sitio de reconocimiento deseado y, por lo tanto, retienen la actividad inductora de corte o ruptura de doble cadena. Por ejemplo, cualquiera de los agentes de nucleasa descritos en la presente descripción puede modificarse a partir de una secuencia de endonucleasa nativa y diseñarse para reconocer e inducir un corte o ruptura de doble cadena en un sitio de reconocimiento que no fue reconocido por el agente de nucleasa nativo. Por lo tanto, en algunas realizaciones, la nucleasa diseñada genéticamente tiene una especificidad para inducir un corte o ruptura de doble cadena en un sitio de reconocimiento que es diferente del sitio de reconocimiento del agente de nucleasa nativo correspondiente. Los ensayos de actividad inductora de corte o ruptura de doble cadena son conocidos y generalmente miden la actividad general y la especificidad de la endonucleasa en sustratos de ADN que contienen el sitio de reconocimiento.

Por ejemplo, la Fig. 3 representa las posiciones de los sitios de unión de ZFN y los sitios de corte en los casetes de selección. Los sitios son los siguientes: Neo-ZFN(1,2): SITIO DE UNIÓN A NUCLEASA/sitio de corte GGGCGCCCGGTTCTTTT/gtcaag/ACCGACCTGTCCGGTG (SEQ ID NO: 9); Neo-ZFN(3,4): SITIO DE UNIÓN A NUCLEASA/sitio de corte CCGGTTCTTTTGTG/aagacc/GACCTGTCCGGTGCC (SEQ ID NO: 10); Hyg-ZFN(1,2): SITIO DE UNIÓN A NUCLEASA/sitio de corte TGCGATCGCTGCGGCCGA/tcttag/CCAGACGAGCGGGTTCGG (SEQ ID NO: 11); e Hyg-ZFN(3,4): SITIO DE UNIÓN A NUCLEASA/sitio de corte CGCTGCGGCCGATCT/tagcca/GACGAGCGGGTTCGG (SEQ ID NO: 12).

El agente de nucleasa puede introducirse en la célula mediante cualquier medio conocido en la técnica. El polipéptido que codifica el agente de nucleasa puede introducirse directamente en la célula. Alternativamente, un polinucleótido que codifica el agente de nucleasa puede introducirse en la célula. Cuando se introduce un polinucleótido que codifica el agente de nucleasa en la célula, el agente de nucleasa puede expresarse de forma transitoria, condicional o constitutiva dentro de la célula. Por lo tanto, el polinucleótido que codifica el agente de nucleasa puede estar contenido en un casete de expresión y estar unido operativamente a un promotor condicional, un promotor inducible, un promotor constitutivo o un promotor específico de tejido. Dichos promotores de interés se analizan con más detalle en otra parte de la presente descripción. Alternativamente, el agente de nucleasa se introduce en la célula como un ARNm que codifica un agente de nucleasa.

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

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

levadura, una célula humana, una célula no humana, una célula de mamífero, una célula de roedor, una célula de ratón, una célula de rata o cualquier otra célula huésped de interés, en comparación con la secuencia de polinucleótido de origen natural.

B. Marcadores de selección

- 5 Los diversos métodos y composiciones proporcionados en la presente descripción emplean los agentes de nucleasa y sus sitios de reconocimiento correspondientes en combinación con marcadores de selección. Como se analiza en la presente descripción, la posición del sitio de reconocimiento en el polinucleótido que codifica el marcador de selección permite un método eficiente para identificar eventos de integración en el locus objetivo. Además, en la presente descripción se proporcionan varios métodos en donde se emplean
- 10 marcadores de selección alternos que tienen el sitio de reconocimiento de nucleasa para mejorar la eficiencia y la eficacia a través de los cuales se integran múltiples polinucleótidos de interés dentro de un locus objetivo dado.

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

Los polinucleótidos que codifican los marcadores de selección están unidos operativamente a un promotor activo en la célula. Dichos casetes de expresión y sus diversos componentes reguladores se analizan en más detalle en otra parte de la presente descripción.

C. Locus objetivo

- 25 Se proporcionan diversos métodos y composiciones, que permiten la integración de al menos un polinucleótido de inserción en un locus objetivo. El término "locus objetivo" comprende cualquier segmento o región de ADN que se desee integrar en un polinucleótido de inserción. En una realización, el locus objetivo es un locus genómico. El locus objetivo puede ser nativo de la célula, o alternativamente puede comprender un segmento de ADN heterólogo o exógeno. Dichos segmentos de ADN heterólogos o exógenos pueden incluir transgenes, casetes de expresión, polinucleótidos que codifican marcadores de selección, o regiones de ADN heterólogas o exógenas (es decir, regiones heterólogas o exógenas de ADN genómico). El locus objetivo puede comprender cualquiera de los sistemas de integración dirigidos que incluyen, por ejemplo, el sitio de reconocimiento, el
- 30 marcador de selección, polinucleótidos de inserción previamente integrados, polinucleótidos que codifican agentes de nucleasa, promotores, etc. Alternativamente, el locus objetivo puede ubicarse dentro de un cromosoma artificial de levadura (YAC), un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano o cualquier otra región genómica modificada contenida en una célula huésped apropiada. Por lo tanto, en realizaciones específicas, el locus objetivo puede comprender una secuencia de ácido nucleico genómico nativo, heterólogo o exógeno de un procariota, un eucariota, levadura, bacterias, un mamífero no humano, una célula no humana, un roedor, un ser humano, una rata, un ratón, un hámster, un conejo, un cerdo, un bovino,
- 35 un ciervo, una oveja, una cabra, un pollo, un gato, un perro, un hurón, un primate (*por ejemplo*, tití, mono rhesus), mamífero domesticado o un mamífero de uso agrícola o cualquier otro organismo de interés o una combinación de estos.

- Los ejemplos no limitantes del locus objetivo incluyen, un locus genómico que codifica una proteína expresada en una célula B, un locus genómico que expresa un polipéptido en una célula B inmadura, un locus genómico
- 45 que expresa un polipéptido en una célula B madura, loci de inmunoglobulinas (Ig), o loci de receptores de células T, que incluyen por ejemplo un locus del receptor alfa de células T. Tal locus puede ser de un ave (*por ejemplo*, un pollo), un mamífero no humano, un roedor, un ser humano, una rata, un ratón, un hámster, un conejo, un cerdo, un bovino, un ciervo, una oveja, una cabra, un gato, un perro, un hurón, un primate (*por ejemplo*, tití, mono rhesus), mamífero domesticado o un mamífero de uso agrícola o cualquier otro organismo de interés o una combinación de estos.
- 50

En realizaciones adicionales, el locus objetivo no puede transformarse mediante el uso de un método convencional o puede ser transformado solo de manera incorrecta o solo con una eficiencia significativamente baja, en ausencia de un corte o ruptura de doble cadena inducido por un agente de nucleasa.

D. Vectores de transformación y polinucleótidos de inserción

- 55 Como se describió anteriormente, los métodos y composiciones proporcionados en la presente descripción aprovechan los agentes de nucleasa y el posicionamiento estratégico de los sitios de reconocimiento para un agente de nucleasa dentro de un casete de selección en combinación con un evento de recombinación homóloga. Dichos métodos emplean el corte o ruptura de doble cadena en el sitio de reconocimiento en

combinación con la recombinación homóloga para, de ese modo, dirigir la integración de un polinucleótido de inserción en el locus objetivo. La "recombinación homóloga" se usa convencionalmente para incluir el intercambio de fragmentos de ADN entre dos moléculas de ADN en sitios cruzados dentro de las regiones de homología.

5 i. Polinucleótido de inserción

El término "polinucleótido de inserción" comprende un segmento de ADN que se desea integrar en el locus objetivo. En una realización, el polinucleótido de inserción comprende uno o más polinucleótidos de interés. En otras realizaciones, el polinucleótido de inserción puede comprender uno o más casetes de expresión. Un casete de expresión dado puede comprender un polinucleótido de interés, un polinucleótido que codifica un

10 marcador de selección y/o un gen reportero junto con los diversos componentes reguladores que influyen en la expresión. Los ejemplos no limitantes de polinucleótidos de interés, marcadores de selección y genes reporteros (por ejemplo, eGFP) que pueden incluirse dentro del polinucleótido de inserción se analizan en detalle en otra parte de la presente descripción.

En realizaciones específicas, el polinucleótido de inserción puede comprender un ácido nucleico genómico. En una realización, el ácido nucleico genómico se deriva de un ratón, un ser humano, un roedor, un no humano, una rata, un hámster, un conejo, un cerdo, un bovino, un ciervo, una oveja, una cabra, un pollo, un gato, un

15 perro, un hurón, un primate (*por ejemplo*, tití, mono rhesus), mamífero domesticado o un mamífero de uso agrícola o cualquier otro organismo de interés o una combinación de estos.

En realizaciones adicionales, el polinucleótido de inserción comprende un alelo condicional. En una realización, el alelo condicional es un alelo multifuncional, como se describe en el documento US 2011/0104799. En realizaciones específicas, el alelo condicional comprende: (a) una secuencia de actuación en orientación sentido con respecto a la transcripción de un gen objetivo, y un casete para la selección por fármacos 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 divisor de exón y un módulo de tipo genético

20 invertible; ver, 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 de actuación y el DSC, y (ii) contiene la NSI en orientación sentido y el COIN en orientación antisentido.

El polinucleótido de inserción puede ser de aproximadamente 5 kb a aproximadamente 200 kb, de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de

30 aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 60 kb a aproximadamente 70 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 110 kb, de aproximadamente 120 kb a aproximadamente 130 kb, de aproximadamente 130 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente

35 150 kb, de aproximadamente 150 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 170 kb, de aproximadamente 170 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 190 kb, o de aproximadamente 190 kb a aproximadamente 200 kb.

En realizaciones específicas, el polinucleótido de inserción comprende un ácido nucleico flanqueado con secuencias objetivo de recombinación específicas del sitio. Se reconoce que si bien el polinucleótido de inserción completo puede estar flanqueado por dicha secuencia objetivo de recombinación específica del sitio, cualquier región o polinucleótido individual de interés dentro del polinucleótido de inserción también puede estar

40 flanqueado por dichos sitios. El término "sitio de recombinación" incluye una secuencia de nucleótidos que es reconocida por una recombinasa específica de un sitio y que puede servir como sustrato para un evento de recombinación. El término "recombinasa específica de un sitio" incluye un grupo de enzimas que pueden facilitar la recombinación entre sitios de recombinación donde los dos sitios de recombinación están físicamente separados dentro de una molécula de ácido nucleico individual o en moléculas de ácido nucleico separadas. Los ejemplos de recombinasas específicas de un sitio incluyen, pero no se limitan a, recombinasas Cre, Flp y Dre. La recombinasa específica de un sitio se puede introducir en la célula por cualquier medio, incluso mediante la introducción del polipéptido de recombinasa en la célula o mediante la introducción de un

50 polinucleótido que codifica la recombinasa específica del sitio en la célula huésped. El polinucleótido que codifica la recombinasa específica del sitio puede ubicarse dentro del polinucleótido de inserción o dentro de un polinucleótido separado. La recombinasa específica del sitio puede estar unida operativamente a un promotor activo en la célula que incluye, por ejemplo, un promotor inducible, un promotor que es endógeno a la célula, un promotor que es heterólogo a la célula, un promotor específico de la célula, un promotor específico de tejido, o un promotor específico de una etapa del desarrollo. Las secuencias objetivo de recombinación específica de un sitio que pueden flanquear el polinucleótido de inserción o cualquier polinucleótido de interés en el polinucleótido de inserción pueden incluir, entre otras, loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attP, att, FRT, rox y una combinación de estos.

En otras realizaciones, los sitios de recombinación específica de un sitio flanquean un polinucleótido que codifica un marcador de selección y/o un gen reportero contenido dentro del polinucleótido de inserción. En

60

tales casos después de la integración del polinucleótido de inserción en el locus objetivo, las secuencias entre los sitios de recombinación específica de un sitio pueden eliminarse.

En una realización, el polinucleótido de inserción comprende un polinucleótido que codifica un marcador de selección. Dichos marcadores de selección incluyen, entre otros, neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa (puro^r), blasticidina S desaminasa (bsr^r), xantina/guanina fosforribosil transferasa (gpt) o timidina quinasa del virus del herpes simple (HSV-k), o una combinación de estos. En una realización, el polinucleótido que codifica el marcador de selección está unido operativamente a un promotor activo en la célula. Cuando los polinucleótidos de interés se modifican en serie en un locus objetivo (es decir, un locus genómico), el marcador de selección puede comprender un sitio de reconocimiento para un agente de nucleasa, como se describió anteriormente. En una realización, el polinucleótido que codifica el marcador de selección está flanqueado por una secuencia objetivo de recombinación específica de un sitio.

El polinucleótido de inserción puede comprender además un gen reportero unido operativamente a un promotor, en donde el gen reportero codifica una proteína reportera 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 estos. Dichos genes reporteros pueden estar unidos operativamente a un promotor activo en la célula. Dichos promotores pueden ser un promotor inducible, un promotor que es endógeno al gen reportero o a la célula, un promotor que es heterólogo al gen reportero o a la célula, un promotor específico de la célula, un promotor específico de tejido o un promotor específico de una etapa del desarrollo.

ii. Vectores de transformación

Los vectores de transformación se emplean para introducir el polinucleótido de inserción en el locus objetivo. El vector de transformación comprende el polinucleótido de inserción y además comprende un brazo de homología hacia el extremo 5' y hacia el extremo 3', que flanquean el polinucleótido de inserción. Los brazos de homología, que flanquean el polinucleótido de inserción, corresponden a regiones dentro del locus objetivo. Para facilitar la referencia, las regiones correspondientes dentro del locus objetivo se denominan en la presente descripción "sitios objetivo". Por lo tanto, en un ejemplo, un vector de transformación puede comprender un primer polinucleótido de inserción flanqueado por un primer y un segundo brazos de homología correspondientes a un primer y un segundo sitios objetivo ubicados en proximidad suficiente al primer sitio de reconocimiento dentro del polinucleótido que codifica el marcador de selección. Como tal, el vector de transformación de este modo ayuda en la integración del polinucleótido de inserción en el locus objetivo a través de un evento de recombinación homóloga que ocurre entre los brazos de homología y los sitios objetivo correspondientes, por ejemplo, dentro del genoma de la célula.

Un brazo de homología del vector de transformación puede ser de cualquier longitud que sea suficiente para promover un evento de recombinación homóloga con un sitio objetivo correspondiente, que incluye, por ejemplo, 50-100 bases, 100-1000 bases o al menos 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5-50, 5-55, 5-60, 5-65, 5-70, 5-75, 5-80, 5-85, 5-90, 5-95, 5-100, 100-200 o 200-300 kilobases de longitud o más. Como se describe con más detalle a continuación, los vectores de transformación grandes pueden emplear brazos de direccionamiento de mayor longitud.

Los sitios objetivo dentro del locus objetivo que corresponden a los brazos de homología hacia el extremo 5' y hacia el extremo 3' del vector de transformación están ubicados en "proximidad suficiente al sitio de reconocimiento" ubicado en el polinucleótido que codifica el marcador de selección. Los brazos de homología hacia el extremo 5' y hacia el extremo 3' de un vector de transformación están "ubicados en una proximidad suficiente" a un sitio de reconocimiento cuando la distancia es tal que promueve la ocurrencia de un evento de recombinación homóloga entre los sitios objetivo y los brazos de homología tras un corte o ruptura de doble cadena en el sitio de reconocimiento. Por lo tanto, en realizaciones específicas, los sitios objetivo correspondientes al brazo de homología hacia el extremo 5' y/o hacia el extremo 3' del vector de transformación están dentro de al menos 1 nucleótido de un sitio de reconocimiento dado, están dentro de al menos 10 nucleótidos a aproximadamente 14 kb de un sitio de reconocimiento dado o están dentro de aproximadamente 10 nucleótidos a aproximadamente 100 nucleótidos, aproximadamente 100 nucleótidos a aproximadamente 500 nucleótidos, aproximadamente 500 nucleótidos a aproximadamente 1000 nucleótidos, aproximadamente 1 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 10 kb, o aproximadamente 10 kb a aproximadamente 14 kb de un sitio de reconocimiento dado. En realizaciones específicas, el sitio de reconocimiento está inmediatamente adyacente a al menos uno o ambos sitios objetivo.

La relación espacial de los sitios objetivo que corresponden a los brazos de homología del vector de transformación y el sitio de reconocimiento dentro del polinucleótido que codifica el marcador de selección puede variar. Por ejemplo, los sitios objetivo pueden ubicarse en 5' con relación al sitio de reconocimiento, ambos sitios objetivo pueden ubicarse en 3' con relación al sitio de reconocimiento o los sitios objetivo pueden flanquear el sitio de reconocimiento.

Un brazo de homología y un sitio objetivo "se corresponden" o son "correspondientes" entre sí cuando las dos regiones comparten un nivel suficiente de identidad de secuencia entre sí para actuar como sustratos para una reacción de recombinación homóloga. Por "homología" se entiende secuencias de ADN que son idénticas o comparten identidad de secuencia con una secuencia correspondiente. La identidad de secuencia entre un sitio objetivo dado y el brazo de homología correspondiente que se encuentra en el vector de transformación puede ser cualquier grado de identidad de secuencia que permita que ocurra una recombinación homóloga. Por ejemplo, la cantidad de identidad de secuencia compartida por el brazo de homología del vector de transformación (o un fragmento de este) y el sitio objetivo (o un fragmento de este) puede ser al menos 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia, de modo que las secuencias experimenten recombinación homóloga. Además, una región correspondiente de homología entre el brazo de homología y el sitio objetivo correspondiente puede ser de cualquier longitud que sea suficiente para promover la recombinación homóloga en el sitio de reconocimiento escindido. Por ejemplo, un brazo de homología dado y/o un sitio objetivo correspondiente puede comprender regiones de homología correspondientes que son de al menos aproximadamente 50-100 bases, 100-1000 bases, o 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5-50, 5-55, 5-60, 5-65, 5-70, 5-75, 5-80, 5-85, 5-90, 5-95, 5-100, 100-200, o 200-300 kilobases de longitud o más (como se describe en los vectores LTVEC descritos en otra parte de esta descripción) de modo que el brazo de homología tenga suficiente homología para experimentar una recombinación homóloga con los sitios objetivo correspondientes dentro del genoma de la célula.

Para facilitar la referencia, los brazos de homología incluyen un brazo de homología hacia el extremo 5' y hacia el extremo 3'. Esta terminología se refiere a la posición relativa de los brazos de homología con el polinucleótido de inserción dentro del vector de transformación.

Por lo tanto, los brazos de homología del vector de transformación están diseñados para corresponder a un sitio objetivo con el locus objetivo. Por lo tanto, los brazos de homología pueden corresponder a un locus que es nativo de la célula, o alternativamente pueden corresponder a una región de un segmento de ADN heterólogo o exógeno que se integró en el genoma de la célula, que incluye, entre otros, transgenes, casetes de expresión o regiones de ADN heterólogas o exógenas. Alternativamente, los brazos de homología del vector de transformación pueden corresponder a una región de un cromosoma artificial de levadura (YAC), un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano o cualquier otra región modificada genéticamente contenida en una célula huésped apropiada. Aún más, los brazos de homología del vector de transformación pueden corresponder o derivarse de una región de una biblioteca de BAC, una biblioteca de cósmidos o una biblioteca de fagos P1. Por lo tanto, en realizaciones específicas, los brazos de homología del vector de transformación corresponden a un locus que es nativo, heterólogo o exógeno a un procariota, una levadura, un ave (por ejemplo, pollo), un mamífero no humano, un roedor, un ser humano, una rata, un ratón, un hámster, un conejo, un cerdo, un bovino, un ciervo, una oveja, una cabra, un gato, un perro, un hurón, un primate (*por ejemplo*, titi, mono rhesus), mamífero domesticado o un mamífero de uso agrícola o cualquier otro organismo de interés. En realizaciones adicionales, los brazos de homología corresponden a un locus de la célula que no puede transformarse mediante el uso de un método convencional o puede ser transformado solo de manera incorrecta o solo con una eficiencia significativamente baja, en ausencia de un corte o ruptura de doble cadena inducido por un agente de nucleasa. En una realización, los brazos de homología se derivan de un ADN sintético.

En otras realizaciones más, los brazos de homología hacia el extremo 5' y hacia el extremo 3' corresponden al 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 objetivo son de un genoma de ratón de una segunda cepa, en donde la primera cepa y la segunda cepa son diferentes. En otras realizaciones, los brazos de homología son del genoma del mismo animal o son del genoma de la misma cepa, *por ejemplo*, el genoma objetivo es un genoma de ratón de una primera cepa, y los brazos objetivo son de un genoma de ratón del mismo ratón o de la misma cepa.

El vector de transformación (tal como un vector de transformación grande) también puede comprender un casete de selección o un gen reportero como se analiza en otra parte de la presente descripción. El casete de selección puede comprender una secuencia de ácido nucleico que codifica un marcador de selección, en donde la secuencia de ácido nucleico está unida operativamente a un promotor. El promotor puede estar activo en una célula procariota de interés y/o activo en una célula eucariota de interés. Dichos promotores pueden ser un promotor inducible, un promotor que es endógeno al gen reportero o la célula, un promotor que es heterólogo al gen reportero o a la célula, un promotor específico de célula, un promotor específico de tejido o un promotor específico de una etapa del desarrollo. 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 estos. El marcador de selección del vector de transformación puede estar flanqueado por los brazos de homología hacia el extremo 5' y hacia el extremo 3' o puede encontrarse 5' o 3' con respecto a los brazos de homología.

En una realización, el vector de transformación (tal como un vector de transformación grande) comprende un gen reportero unido operativamente a un promotor, en donde el gen reportero codifica una proteína reportera 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 estos. Dichos genes reporteros pueden estar unidos operativamente a un promotor activo en la célula. Dichos promotores pueden ser un promotor inducible, un promotor que es endógeno al gen reportero o la célula, un promotor que es heterólogo al gen reportero o a la célula, un promotor específico de la célula, una forma promotora específica del tejido o un promotor específico de una etapa del desarrollo.

En una realización, el uso combinado del vector de transformación (que incluye, por ejemplo, un vector de transformación grande) con el agente de nucleasa da como resultado una mayor eficacia de transformación en comparación con el uso del vector de transformación solo. En una realización, cuando el vector de transformación se usa junto con el agente de nucleasa, la eficiencia de transformación del vector de transformación aumenta al menos dos veces, al menos tres veces, al menos 4 veces o al menos 10 veces en comparación con el uso del vector de transformación solo.

iii. Vectores de transformación grandes

El término "vector de transformación grande" o "LTVEC" incluye vectores de transformación grandes que comprenden brazos de homología que corresponden a y se derivan de secuencias de ácido nucleico más grandes que las utilizadas típicamente por otros enfoques destinados a realizar recombinación homóloga en las células y/o que comprenden polinucleótidos de inserción que comprenden secuencias de ácido nucleico más grandes que las utilizadas típicamente por otros enfoques destinados a realizar recombinación homóloga en las células. En realizaciones específicas, los brazos de homología y/o el polinucleótido de inserción del LTVEC comprenden la secuencia genómica de una célula eucariota. El tamaño del LTVEC es demasiado grande para permitir la detección de eventos de transformación mediante ensayos convencionales, *por ejemplo*, transferencia Southern y PCR de largo alcance (*por ejemplo*, 1 kb-5 kb). Los ejemplos de LTVEC incluyen, entre otros, vectores derivados de un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano o un cromosoma artificial de levadura (YAC). Se describen ejemplos no limitantes de LTVEC y métodos para hacerlos, *por ejemplo*, en los documentos de patente de Estados Unidos núms. 6,586,251, 6,596,541, 7,105,348 y WO 2002/036789 (PCT/US01/45375).

El LTVEC puede ser de cualquier longitud, que incluye, entre otras, de aproximadamente 20 kb a aproximadamente 300 kb, de aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 75 kb, de aproximadamente 75 kb a aproximadamente 100 kb, de aproximadamente 100 kb a 125 kb, de aproximadamente 125 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 175 kb, aproximadamente 175 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 225 kb, de aproximadamente 225 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 275 kb o de aproximadamente 275 kb a aproximadamente 300 kb.

En una realización, el LTVEC comprende un polinucleótido de inserción que varía de aproximadamente 5 kb a aproximadamente 200 kb, de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 60 kb a aproximadamente 70 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 110 kb, de aproximadamente 120 kb a aproximadamente 130 kb, de aproximadamente 130 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 170 kb, de aproximadamente 170 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 190 kb, o de aproximadamente 190 kb a aproximadamente 200 kb.

En 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 otras realizaciones, los brazos de homología se derivan del locus objetivo (es decir, el locus genómico) de la célula y, en algunos casos, el locus objetivo, para el cual se diseña el LTVEC, no es seleccionable mediante un método convencional. En otras realizaciones más, los brazos de homología se derivan de un ADN sintético. En una realización, una suma total del brazo de homología hacia el extremo 5' y el brazo de homología hacia el extremo 3' en el LTVEC es de al menos 10 kb. En una realización, el brazo de homología hacia el extremo 5' varía de aproximadamente 1 kb a aproximadamente 100 kb. En otras realizaciones, el brazo de homología hacia el extremo 5' varía de aproximadamente 5 kb a aproximadamente 100 kb. En una realización, el brazo de homología hacia el extremo 3' varía de aproximadamente 1 kb a aproximadamente 100 kb. En una realización, el brazo de homología hacia el extremo 3' varía de aproximadamente 5 kb a aproximadamente 100 kb. En otras realizaciones, la suma total de los brazos de homología hacia el extremo 5' y hacia el extremo 3' es de aproximadamente 1 kb a aproximadamente 5 kb, de

aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 70 kb, de aproximadamente 70 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 110 kb, de aproximadamente 110 kb a aproximadamente 120 kb, de aproximadamente 120 kb a aproximadamente 130 kb, de aproximadamente 130 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 170 kb, de aproximadamente 170 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 190 kb, o de aproximadamente 190 kb a aproximadamente 200 kb.

En otras realizaciones, la suma total de los brazos de homología en 5' y 3' del LTVEC es de aproximadamente 10 kb a aproximadamente 30 kb, de aproximadamente 20 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 120 kb, o de aproximadamente 120 kb a 150 kb. En otros casos, la suma total del brazo de homología en 5' y 3' es de aproximadamente 16 Kb a aproximadamente 150 Kb.

En realizaciones adicionales, el LTVEC y el polinucleótido de inserción están diseñados para permitir una delección en el locus objetivo de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 150 kb, o de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 1 Mb, de aproximadamente 1 Mb a aproximadamente 1,5 Mb, de aproximadamente 1,5 Mb a aproximadamente 2 Mb, de aproximadamente 2 Mb a aproximadamente 2,5 Mb, o de aproximadamente 2,5 Mb a aproximadamente 3 Mb.

En otros casos, el LTVEC y el polinucleótido de inserción están diseñados para permitir una inserción en el locus objetivo de una secuencia de ácido nucleico exógena que varía de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, o de aproximadamente 350 kb a aproximadamente 400 kb. En una realización, el polinucleótido de inserción tiene aproximadamente 130 kb o aproximadamente 155 kb.

En una realización, el LTVEC comprende un casete de selección o un gen reportero como se analiza en otra parte de esta descripción.

III. Métodos para integrar un polinucleótido de interés en un locus objetivo

A. Métodos de integración de un polinucleótido de inserción cerca del sitio de reconocimiento por recombinación homóloga

Se proporcionan métodos para modificar un locus objetivo en una célula como se caracteriza en las reivindicaciones. Los métodos comprenden (a) proporcionar una célula que comprende un primer polinucleótido que codifica un primer marcador de selección unido operativamente a un primer promotor activo en la célula, en donde el primer polinucleótido comprende además un primer sitio de reconocimiento para un primer agente de nucleasa; (b) introducir en la célula: (i) un primer agente de nucleasa que induce un corte o ruptura de doble cadena en el primer sitio de reconocimiento, y (ii) un primer vector de transformación que comprende un primer polinucleótido de inserción flanqueado por un primer y un segundo brazos de homología que corresponden a un primer y un segundo sitios objetivo ubicados en una proximidad suficiente al primer sitio de reconocimiento; y (c) identificar al menos una célula que comprende el primer polinucleótido de inserción integrado en el locus objetivo. En realizaciones específicas, el primer polinucleótido que comprende el primer marcador de selección está flanqueado por un primer sitio objetivo y un segundo sitio objetivo, el primer sitio objetivo corresponde al primer brazo de homología en el primer vector de transformación y el segundo sitio objetivo corresponde al segundo brazo de homología en el primer vector de transformación.

Se pueden usar varios métodos para identificar células que tienen el polinucleótido de inserción integrado en el locus objetivo. En una realización, el corte o ruptura de doble cadena en el primer sitio de reconocimiento interrumpe la actividad del primer marcador de selección. Por lo tanto, en una realización, tales células se identifican cultivando las células en condiciones que identifican células, que no tienen la actividad del marcador

de selección codificado por el polinucleótido que tiene el sitio de reconocimiento, que fue cortado por el agente de nucleasa. Se conocen métodos que emplean dichos marcadores de selección y ensayos para determinar su actividad. Un método adicional para identificar células que tienen el polinucleótido de inserción en el locus objetivo puede comprender identificar al menos una célula que comprende tener el polinucleótido de inserción integrado en el sitio objetivo deseado. Tales métodos pueden incluir identificar al menos una célula que comprende en su genoma el primer polinucleótido de inserción integrado en el primer y el segundo sitios objetivo.

También se pueden emplear métodos adicionales para identificar células que tienen el polinucleótido de inserción integrado en el locus objetivo. La inserción del polinucleótido de inserción en el locus objetivo produce una "modificación del alelo". El término "modificación del alelo" o "MOA" incluye la modificación de la secuencia de ADN exacta de un alelo de un gen(es) o locus (loci) cromosómico en un genoma. Los ejemplos de "modificación del alelo (MOA)" incluyen, entre otras, delecciones, sustituciones o inserciones de tan solo un único nucleótido o delecciones de muchas kilobases que abarcan un(os) gen(es) o locus (loci) cromosómico de interés, así como cualquiera y todas las modificaciones posibles entre estos dos extremos.

En diversas realizaciones, para facilitar la identificación de la modificación dirigida, se emplea un ensayo cuantitativo de alto rendimiento, a saber, el ensayo de modificación de alelos (MOA). El ensayo MOA descrito en esta descripción permite una detección a gran escala de un alelo(s) modificado(s) en un cromosoma parental después de una modificación genética. El ensayo MOA puede llevarse a cabo mediante diversas técnicas analíticas, que incluyen, entre otras, 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 objetivo. Además, el conjunto de cebadores comprende una sonda fluorescente que reconoce la secuencia amplificada. El ensayo cuantitativo también se puede llevar a cabo a través de una variedad de técnicas analíticas, que incluyen, entre otras, hibridación in situ mediada por fluorescencia (FISH), hibridación genómica comparativa, amplificación de ADN isotérmico, hibridación cuantitativa a sonda(s) inmovilizada(s), tecnología de sondas Invader Probes®, MMP assays®, TaqMan® Molecular Beacon y Eclipse™. (Ver, por ejemplo, el documento US2005/0144655).

La presencia de un corte o ruptura de doble cadena en el sitio de reconocimiento dentro del marcador de selección, en diversas realizaciones, aumenta la eficacia y/o frecuencia de recombinación entre un vector de transformación (como un LTVEC) y el locus objetivo. En una realización, la recombinación es recombinación homóloga. En diversas realizaciones, en presencia del corte o ruptura de doble cadena, la eficiencia de transformación de un vector de transformación (tal como un LTVEC) en el locus objetivo es al menos aproximadamente 2 veces mayor, al menos aproximadamente 3 veces mayor, al menos aproximadamente 4 veces, al menos aproximadamente 10 veces mayor que en ausencia del corte o ruptura de doble cadena (usando, por ejemplo, el mismo vector de transformación y los mismos brazos de homología y sitios objetivo correspondientes en el locus de interés pero en ausencia de un agente de nucleasa agregado que realice el corte o la ruptura de la doble cadena).

B. Métodos de integración de múltiples polinucleótidos de interés en el locus objetivo

Los diversos métodos proporcionados en la presente descripción permiten la integración dirigida de múltiples polinucleótidos de interés dentro de un locus objetivo dado. Los métodos emplean el sistema de integración dirigida que se describe en esta descripción, que emplea el posicionamiento estratégico del sitio de reconocimiento del agente de nucleasa dentro de un polinucleótido que codifica un marcador de selección. En realizaciones específicas, el marcador de selección y el sitio de reconocimiento se alternan dentro de cada polinucleótido de inserción. Al hacerlo, la modificación en mosaico de los polinucleótidos de inserción secuenciales dentro de un locus objetivo dado ocurre con una mayor eficiencia y eficacia.

En una realización, el método para modificar un locus objetivo en una célula comprende: (a) proporcionar una célula que comprende un locus que comprende un primer polinucleótido que codifica un primer marcador de selección unido operativamente a un primer promotor activo en la célula, en donde el primer polinucleótido comprende además un primer sitio de reconocimiento para un primer agente de nucleasa; (b) introducir en la célula un primer agente de nucleasa, en donde el primer agente de nucleasa induce un corte o ruptura de doble cadena en el primer sitio de reconocimiento; e, introducir en la célula un primer vector de transformación que comprende un primer polinucleótido de inserción flanqueado por un primer y un segundo brazos de homología, que corresponden a un primer y un segundo sitios objetivo ubicados en proximidad suficiente al primer sitio de reconocimiento; y el primer polinucleótido de inserción comprende además (1) un primer polinucleótido de interés; y (2) un segundo polinucleótido que codifica un segundo marcador de selección unido operativamente a un segundo promotor activo en la célula, en donde el segundo polinucleótido comprende un segundo sitio de reconocimiento para un segundo agente de nucleasa; y (c) identificar al menos una célula que comprende el primer polinucleótido de inserción integrado en el locus objetivo.

En otras realizaciones, se pueden integrar polinucleótidos de interés adicionales en el locus objetivo. Tales métodos para modificar un locus objetivo en una célula comprenden: (a) proporcionar una célula que comprende un locus, que comprende un primer polinucleótido que codifica un primer marcador de selección

unido operativamente a un primer promotor activo en la célula, en donde el primer polinucleótido comprende además un primer sitio de reconocimiento para un primer agente de nucleasa; (b) introducir en la célula un primer agente de nucleasa, en donde el primer agente de nucleasa induce un corte o ruptura de doble cadena en el primer sitio de reconocimiento; e, introducir en la célula un primer vector de transformación que comprende un primer polinucleótido de inserción flanqueado por un primer y un segundo brazos de homología, que corresponden a un primer y un segundo sitios objetivo ubicados en proximidad suficiente al primer sitio de reconocimiento; y el primer polinucleótido de inserción comprende además (1) un primer polinucleótido de interés; y (2) un segundo polinucleótido que codifica un segundo marcador de selección unido operativamente a un segundo promotor activo en la célula, en donde el segundo polinucleótido comprende un segundo sitio de reconocimiento para un segundo agente de nucleasa; (c) identificar al menos una célula que comprende el primer polinucleótido de inserción integrado en el locus objetivo; (d) introducir en la célula que comprende en su genoma el primer polinucleótido de inserción integrado en el locus objetivo, (i) un segundo agente de nucleasa, en donde el segundo agente de nucleasa induce un corte o ruptura de doble cadena en el segundo sitio de reconocimiento; y (ii) un segundo vector de transformación que comprende un segundo polinucleótido de inserción flanqueado por un tercer y un cuarto brazos de homología; y (b) identificar al menos una célula que comprende el segundo polinucleótido de inserción integrado en el locus objetivo. En realizaciones específicas, el corte o ruptura de doble cadena en el segundo marcador de reconocimiento interrumpe la actividad del segundo marcador de selección. En realizaciones adicionales, identificar al menos una célula que comprende el segundo polinucleótido de inserción integrado en el locus objetivo comprende cultivar la célula en condiciones que identifican células que no tienen la actividad del segundo marcador de selección. En otras realizaciones adicionales, el segundo polinucleótido que comprende el segundo marcador de selección está flanqueado por un tercer sitio objetivo y un cuarto sitio objetivo, el tercer sitio objetivo corresponde al tercer brazo de homología en el segundo vector de transformación y el cuarto sitio objetivo corresponde al cuarto brazo de homología en el segundo vector de transformación. En otras realizaciones adicionales, identificar al menos una célula que comprende el segundo polinucleótido de inserción integrado en el locus objetivo comprende identificar al menos una célula que comprende el segundo polinucleótido de inserción integrado en el tercer y cuarto sitios objetivo.

Otros métodos para modificar un locus objetivo en una célula comprenden: (a) proporcionar una célula que comprende un locus objetivo que comprende un primer polinucleótido que codifica un primer marcador de selección unido operativamente a un primer promotor activo en la célula, en donde el primer polinucleótido comprende además un primer sitio de reconocimiento para un primer agente de nucleasa; (b) introducir en la célula (i) un primer agente de nucleasa, en donde el primer agente de nucleasa induce un corte o ruptura de doble cadena en el primer sitio de reconocimiento; y (ii) un primer vector de transformación que comprende un primer polinucleótido de inserción flanqueado por un primer y un segundo brazos de homología correspondientes a un primer y un segundo sitios objetivo ubicados en proximidad suficiente al primer sitio de reconocimiento y el primer polinucleótido de inserción comprende además (1) un primer polinucleótido de interés; y (2) un segundo polinucleótido que codifica un segundo marcador de selección unido operativamente a un segundo promotor activo en la célula, en donde el segundo polinucleótido comprende un segundo sitio de reconocimiento para un segundo agente de nucleasa, y el segundo polinucleótido que comprende el segundo marcador de selección está flanqueado por un tercer sitio objetivo y un cuarto sitio objetivo, el tercer sitio objetivo corresponde al tercer brazo de homología en el segundo vector de transformación y el cuarto sitio objetivo corresponde al cuarto brazo de homología en el segundo vector de transformación; (c) identificar al menos una célula que comprende el primer polinucleótido de inserción integrado en el locus objetivo; (d) introducir en la célula que comprende el primer polinucleótido de inserción integrado en el locus objetivo, (i) un segundo agente de nucleasa, en donde el segundo agente de nucleasa induce un corte o ruptura de doble cadena en el segundo sitio de reconocimiento; y (ii) un segundo vector de transformación que comprende un segundo polinucleótido de inserción flanqueado por un tercer y un cuarto brazos de homología en donde el segundo polinucleótido de inserción comprende (1) un segundo polinucleótido de interés; y (2) un tercer polinucleótido que codifica un tercer marcador de selección unido operativamente a un tercer promotor activo en la célula, en donde el tercer polinucleótido comprende un tercer sitio de reconocimiento para un tercer agente de nucleasa; y (b) identificar al menos una célula que comprende el segundo polinucleótido de inserción integrado en el locus objetivo. En realizaciones específicas, el corte o ruptura de doble cadena en el segundo marcador de reconocimiento interrumpe la actividad del segundo marcador de selección. En realizaciones adicionales, identificar al menos una célula que comprende en su genoma el segundo polinucleótido de inserción integrado en el locus objetivo comprende cultivar la célula en condiciones que identifican células que no tienen la actividad del segundo marcador de selección. En realizaciones adicionales, identificar al menos una célula que comprende en su genoma el segundo polinucleótido de inserción integrado en el locus objetivo comprende identificar al menos una célula que comprende en su genoma el segundo polinucleótido de inserción integrado en el tercer y cuarto sitios objetivo.

Los diversos métodos expuestos anteriormente pueden repetirse secuencialmente para permitir la integración dirigida de cualquier cantidad de polinucleótidos de inserción en un locus objetivo dado. Por lo tanto, los diversos métodos prevén la inserción de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más polinucleótidos de inserción en el locus objetivo. En realizaciones particulares, tales métodos de modificación en mosaico secuencial permiten la reconstrucción de grandes regiones genómicas a partir de una

célula de mamífero (es decir, un ser humano, un no humano, un roedor, un ratón, un mono, una rata, un hámster, un mamífero domesticado o un animal de uso agrícola) en un locus objetivo (es decir, un locus genómico). En tales casos, la transferencia y reconstrucción de regiones genómicas que incluyen regiones codificantes y no codificantes permiten preservar la complejidad de una región dada al retener, al menos en parte, las regiones codificantes, las regiones no codificantes y variaciones del número de copias encontradas dentro de la región genómica nativa. Así, los diversos métodos proporcionan, por ejemplo, métodos para generar regiones genómicas "heterólogas" o "exógenas" dentro de cualquier célula o animal mamífero de interés. En un ejemplo no limitante, se genera una región genómica "humanizada" dentro de un animal no humano.

Cuando se realiza la integración de múltiples polinucleótidos de inserción dentro de un locus objetivo dado, los polinucleótidos que codifican los marcadores de selección y que comprenden el sitio de reconocimiento del agente de nucleasa pueden alternarse entre rondas de integración. Por ejemplo, el primer agente de nucleasa es diferente del segundo agente de nucleasa y el primer marcador de selección es diferente del segundo marcador de selección. En otros ejemplos, al insertar tres polinucleótidos de inserción en un locus objetivo, el primer y el tercer marcador de selección pueden ser idénticos entre sí y, en realizaciones específicas, comprenden además el mismo sitio de reconocimiento, y el segundo marcador de selección es diferente del primer marcador de selección y puede ser diferente del tercer marcador de selección y contienen un sitio de reconocimiento diferente. La selección de los marcadores de selección y los sitios de reconocimiento de esa manera minimiza la cantidad de agentes de nucleasa que deben generarse, y de ese modo mejora la eficiencia y la eficacia de los eventos de integración.

C. Métodos para modificar uno o más loci objetivo mediante el uso de un sistema CRISPR/Cas

Se proporcionan métodos y composiciones para modificar uno o más loci de interés objetivo en una célula mediante el uso de un sistema CRISPR/Cas como se describe en otra parte de esta descripción. Para el sistema CRISPR/Cas, los términos "sitio objetivo" o "secuencia objetivo" pueden usarse indistintamente e incluyen secuencias de ácido nucleico presentes en un ADN objetivo al que se unirá un segmento dirigido al ADN de un ARN guía (ARNg), siempre que existan condiciones suficientes para la unión. Por ejemplo, el sitio objetivo (o secuencia objetivo) dentro de un ADN objetivo es seleccionado por (o se une a, o hibrida con, o es complementario a) la nucleasa Cas o ARNg. Las condiciones adecuadas para la unión de ADN/ARN incluyen condiciones fisiológicas normalmente presentes en una célula. En la técnica se conocen otras condiciones adecuadas para la unión de ADN/ARN (*por ejemplo*, condiciones en un sistema libre de células) (véase, *por ejemplo*, Molecular Cloning: A Laboratory Manual, 3ra Ed. (Sambrook y otros, Harbor Laboratory Press 2001)). La cadena del ADN objetivo que es complementaria e hibrida con la proteína Cas o el ARNg se conoce como la "cadena complementaria" y la cadena del ADN objetivo que es complementaria a la "cadena complementaria" (y por lo tanto no es complementaria a la proteína Cas o ARNg) se conoce como "cadena no complementaria" o "cadena molde".

La proteína Cas puede escindir el ácido nucleico en un sitio dentro de la secuencia objetivo o fuera de la secuencia objetivo. El "sitio de escisión" incluye la posición de un ácido nucleico en donde una proteína Cas produce una ruptura de cadena sencilla o una ruptura de cadena doble. Los extremos cohesivos también se pueden producir mediante el uso de dos proteínas Cas9 que producen una ruptura de cadena sencilla en los sitios de escisión en cada cadena. La escisión específica del sitio del ADN objetivo por Cas9 puede ocurrir en ubicaciones determinadas por (i) complementariedad de apareamiento de bases entre el ARN guía y el ADN objetivo; y (ii) un motivo corto, denominado motivo adyacente a protoespaciador (PAM), en el ADN objetivo. Por ejemplo, el sitio de escisión de Cas9 puede ser de aproximadamente 1 a aproximadamente 10 o de aproximadamente 2 a aproximadamente 5 pares de bases (*por ejemplo*, 3 pares de bases) hacia el extremo 5' de la secuencia PAM. En algunas realizaciones (*por ejemplo*, cuando se usa Cas9 de *S. pyogenes* o un Cas9 estrechamente relacionado), la secuencia PAM de la cadena no complementaria puede ser 5'-XGG-3', donde X es cualquier nucleótido de ADN y X está inmediatamente en 3' de la secuencia objetivo de la cadena no complementaria del ADN objetivo. Como tal, la secuencia PAM de la cadena complementaria sería 5'-CCY-3', donde Y es cualquier nucleótido de ADN y Y está inmediatamente en 5' de la secuencia objetivo de la cadena complementaria del ADN objetivo. En algunas de tales realizaciones, X y Y pueden ser complementarios y el par de bases XY puede ser cualquier par de bases (*por ejemplo*, X=C y Y=G; X=G y Y=C; X=A y Y=T, X=T y Y=A).

Por lo tanto, en algunas realizaciones, los métodos para modificar un locus objetivo de interés en una célula comprenden: (a) proporcionar la célula que comprende un primer locus objetivo que comprende un ácido nucleico que codifica un primer marcador de selección unido operativamente a un primer promotor; (b) introducir en la célula (i) una o más construcciones de expresión que codifican una proteína Cas y un primer ARN guía (ARNg), cada uno de los cuales está unido operativamente a un promotor activo en la célula, en donde la proteína Cas induce un corte o ruptura de doble cadena en un primer sitio objetivo del ARNg en el primer ácido nucleico, interrumpiendo así la expresión o actividad del primer marcador de selección, y (ii) un primer vector de transformación que comprende un primer ácido nucleico de inserción que comprende un segundo ácido nucleico que codifica un segundo marcador de selección unido operativamente a un segundo promotor, en donde el primer ácido nucleico de inserción está flanqueado por un primer y un segundo brazos de homología

correspondientes a un primer y un segundo sitios objetivo ubicados en el primer locus objetivo; y (c) identificar una célula modificada que comprende el primer ácido nucleico de inserción en el primer locus objetivo, en donde la célula modificada tiene la actividad del segundo marcador de selección pero no tiene la actividad del primer marcador de selección, y en donde el primer y el segundo marcadores de selección son diferentes. En una realización, el primer ARNg no hibrida con el primer ácido nucleico de inserción. En una realización, el locus objetivo de interés está ubicado en el genoma de la célula. En otra realización, el locus objetivo de interés está ubicado en un vector en la célula. En una realización, la etapa de identificación (c) comprende cultivar la célula en condiciones que permiten la identificación de la célula modificada que tiene actividad del segundo marcador de selección pero no tiene la actividad del primer marcador de selección.

En una realización, el método comprende además (d) introducir en la célula modificada que comprende el primer ácido nucleico de inserción en el primer locus objetivo (i) uno o más ácidos nucleicos que codifican la proteína Cas y un segundo ARNg, cada uno de los cuales está unido operativamente al promotor activo en la célula modificada, en donde la proteína Cas induce el corte o ruptura de doble cadena en un segundo sitio objetivo de ARNg en el primer ácido nucleico de inserción que comprende el segundo ácido nucleico, interrumpiendo así la expresión o actividad del segundo marcador de selección, y (ii) un segundo vector de transformación que comprende un segundo ácido nucleico de inserción que comprende un tercer ácido nucleico que codifica un tercer marcador de selección unido operativamente a un tercer promotor, en donde el segundo ácido nucleico de inserción está flanqueado por un tercer y cuarto brazos de homología correspondientes a un tercer y un cuarto sitios objetivo ubicados en un segundo locus objetivo; y (e) identificar una segunda célula modificada que comprende el segundo ácido nucleico de inserción en el segundo locus objetivo, en donde la segunda célula modificada tiene la actividad del tercer marcador de selección pero no tiene la actividad del segundo marcador de selección, en donde el segundo y el tercer marcadores de selección son diferentes. En una realización, el primer y el segundo loci objetivo están ubicados inmediatamente adyacentes entre sí. En otra realización, el primer o el segundo locus objetivo está ubicado de aproximadamente 10 nucleótidos a aproximadamente 14 kb a partir del primer o el segundo sitio objetivo de ARNg. En una realización, el segundo ARNg no hibrida con el segundo ácido nucleico de inserción. En una realización, la etapa de identificación (e) comprende cultivar la célula modificada en condiciones que permiten la identificación de la segunda célula modificada que tiene actividad del tercer marcador de selección pero no tiene la actividad del segundo marcador de selección.

En una realización, el método comprende además (f) introducir en la segunda célula modificada que comprende el segundo ácido nucleico de inserción en el segundo locus objetivo: (i) la una o más construcciones de expresión que codifican la proteína Cas y un tercer ARNg, cada uno de los cuales está unido operativamente al promotor activo en la segunda célula modificada, en donde la proteína Cas induce el corte o ruptura de doble cadena en un tercer sitio objetivo de ARNg en el segundo ácido nucleico de inserción que comprende el tercer ácido nucleico, interrumpiendo así la expresión o actividad del tercer marcador de selección, y (ii) un tercer vector de transformación que comprende un tercer ácido nucleico de inserción que comprende un cuarto ácido nucleico que codifica un cuarto marcador de selección unido operativamente a un cuarto promotor, en donde el tercer ácido nucleico de inserción está flanqueado por el quinto y sexto brazos de homología correspondientes a los quinto y sexto sitios objetivo ubicados en un tercer locus objetivo; y (g) identificar una tercera célula modificada que comprende el tercer ácido nucleico de inserción en el tercer locus objetivo, en donde la tercera célula modificada tiene la actividad del cuarto marcador de selección pero no tiene la actividad del tercer marcador de selección, en donde el tercer y el cuarto marcadores de selección son diferentes. En una realización, el segundo y el tercer loci objetivo están ubicados inmediatamente adyacentes entre sí. En otra realización, el segundo o el tercer locus objetivo está ubicado de aproximadamente 10 nucleótidos a aproximadamente 14 kb a partir del primer o segundo sitio objetivo de ARNg.

En una realización, el primero, el segundo, el tercero o el cuarto marcador proporcionan resistencia a un antibiótico. En una realización, el antibiótico comprende G418, higromicina, blastomicina, neomicina o puromicina. En una realización, el primero, segundo, tercero o cuarto marcadores de selección comprenden hipoxantina-guanina fosforribosiltransferasa (HGPRT) o timidina quinasa del virus del herpes simple (HSV-TK). En una realización, el primer, el segundo o el tercer ARNg comprende (i) una secuencia de nucleótidos que hibrida con el primer, el segundo o el tercer sitio objetivo de ARNg y (ii) un ARN CRISPR transactivante (tracrARN). En una realización, el primer, el segundo o el tercer locus objetivo está ubicado muy cerca del primer, el segundo o el tercer sitio objetivo de ARNg de tal manera que el corte o ruptura de doble cadena en el sitio objetivo de ARNg promueve la recombinación homóloga del vector de transformación en el locus objetivo. En una realización, la proteína Cas es Cas9. En una realización, el primer, el segundo o el tercer sitio objetivo de ARNg está flanqueado inmediatamente por una secuencia de un motivo adyacente a protospaciador (PAM).

En realizaciones específicas, el ARNg está diseñado para dirigirse a un primer marcador de selección con antibiótico (*por ejemplo*, Hyg^r) para la inserción de un primer ácido nucleico de inserción que codifica un segundo marcador de selección (*por ejemplo*, Neo^r), por lo que la inserción del primer ácido nucleico de inserción interrumpe la actividad del primer marcador de selección con antibiótico. Se puede diseñar un segundo plásmido de expresión de ARNg para expresar ARNm que se dirige al segundo marcador de selección para la inserción de un segundo ácido nucleico de inserción que codifica el primer marcador de selección, por

lo que la inserción del segundo ácido nucleico de inserción interrumpe la actividad del segundo marcador de selección con antibiótico. De esta manera, solo deben diseñarse ARNg que se dirijan a cada uno de los dos marcadores de selección de antibióticos que pueden usarse en insertos alternos de ácido nucleico. Los ejemplos de ácidos nucleicos que codifican ARNg específicos para marcadores de selección de resistencia a Neo se pueden encontrar en las SEQ ID NO: 17, 18, 19 y 20. Ejemplos de ácidos nucleicos que codifican ARNg específicos para marcadores de selección de resistencia a Hyg se pueden encontrar en las SEQ ID NO: 17, 18, 19 y 20.

En una realización, la célula es una célula procariota. En otra realización, la célula es una célula eucariota. En una realización, la célula eucariota es una célula de mamífero o una célula de mamífero no humano. En una realización, la célula de mamífero es un fibroblasto. En una realización, la célula de mamífero es un fibroblasto humano. En una realización, la célula de mamífero es una célula madre humana adulta. En una realización, la célula de mamífero es una célula progenitora de desarrollo restringido. En una realización, la célula de mamífero es una célula progenitora humana de desarrollo restringido.

En una realización, la célula de mamífero es una célula de mamífero no humano. En una realización, la célula de mamífero es de un roedor. En una realización, el roedor es una rata, un ratón o un hámster. En una realización, la célula eucariota es una célula pluripotente. En una realización, la célula pluripotente es una célula madre hematopoyética o una célula madre neuronal. En una realización, la célula pluripotente es una célula madre pluripotente humana inducida (iPS). En una realización, la célula pluripotente es una célula ES no humana, una célula ES humana, una célula madre embrionaria (ES) de roedor, una célula madre embrionaria (ES) de ratón o una célula madre embrionaria (ES) de rata.

En una realización, el primer, el segundo o el tercer sitio objetivo de ARNg se encuentra en un intrón, un exón, un promotor o una región reguladora del promotor en el primer, el segundo o el tercer ácido nucleico que codifica el primer, el segundo, o el tercer marcador de selección. En una realización, el primer, el segundo o el tercer vector de transformación es de al menos aproximadamente 10 kb. En una realización, el primer, el segundo o el tercer ácido nucleico de inserción varía de aproximadamente 5 kb a aproximadamente 300 kb.

En una realización, el primero, segundo o tercer ácido nucleico de inserción comprende una región genómica del locus del receptor alfa de células T humanas. En una realización, la región genómica comprende al menos un segmento génico de región variable y/o un segmento génico de la región de unión del locus del receptor alfa de células T humanas.

En una realización, el primer y el tercer marcadores de selección son iguales. En una realización, el primer y tercer marcadores de selección son iguales y el segundo y cuarto marcadores de selección son iguales. En una realización, el primer y el tercer ARNg son iguales.

En algunas realizaciones, los métodos para modificar un locus objetivo de interés en una célula comprenden: (a) proporcionar la célula que comprende un primer locus objetivo que comprende un ácido nucleico que codifica un primer marcador de selección unido operativamente a un primer promotor; (b) introducir en la célula (i) una o más construcciones de expresión que codifican una proteína Cas y un primer ARNg, cada uno de los cuales está unido operativamente a un promotor activo en la célula, en donde la proteína Cas induce un corte o ruptura de la doble cadena en un primer sitio objetivo de ARNg en el primer ácido nucleico, interrumpiendo así la expresión o actividad del primer marcador de selección, y (ii) un primer vector de transformación que comprende un primer ácido nucleico de inserción que comprende un segundo ácido nucleico que codifica un segundo marcador de selección unido operativamente a un segundo promotor, en donde el primer ácido nucleico de inserción está flanqueado por un primer y un segundo brazos de homología que corresponden a un primer y un segundo sitios objetivo ubicados en el primer locus objetivo; (c) identificar una célula modificada que comprende el primer ácido nucleico de inserción en el primer locus objetivo, en donde la célula modificada tiene la actividad del segundo marcador de selección pero no tiene la actividad del primer marcador de selección, y en donde el primer y el segundo marcadores de selección son diferentes; (d) introducir en la célula modificada que comprende el primer ácido nucleico de inserción en el primer locus objetivo: (i) uno o más ácidos nucleicos que codifican la proteína Cas y un segundo ARNg, cada uno de los cuales está unido operativamente al promotor activo en la célula modificada, en donde la proteína Cas induce el corte o ruptura de doble cadena en un segundo sitio objetivo de ARNg en el primer ácido nucleico de inserción que comprende el segundo ácido nucleico, interrumpiendo así la expresión o actividad del segundo marcador de selección; y (ii) un segundo vector de transformación que comprende un segundo ácido nucleico de inserción que comprende un tercer ácido nucleico que codifica un tercer marcador de selección unido operativamente a un tercer promotor, en donde el segundo ácido nucleico de inserción está flanqueado por un tercer y cuarto brazos de homología correspondientes a un tercer y un cuarto sitios objetivo ubicados en un segundo locus objetivo; y (e) identificar una segunda célula modificada que comprende el segundo ácido nucleico de inserción en el segundo locus objetivo, en donde la segunda célula modificada tiene la actividad del tercer marcador de selección pero no tiene la actividad del segundo marcador de selección, en donde el primer y el tercer marcadores de selección son iguales, y el segundo y el tercer marcadores de selección son diferentes.

En otras realizaciones, los métodos para modificar un locus objetivo de interés en una célula comprenden: (a) proporcionar la célula que comprende un primer locus objetivo que comprende un ácido nucleico que codifica un primer marcador de selección unido operativamente a un primer promotor; (b) introducir en la célula (i) una o más construcciones de expresión que codifican una proteína Cas y un primer ARNg, cada uno de los cuales está unido operativamente a un promotor activo en la célula, en donde la proteína Cas induce un corte o ruptura de la doble cadena en un primer sitio objetivo de ARNg en el primer ácido nucleico, interrumpiendo así la expresión o actividad del primer marcador de selección, y (ii) un primer vector de transformación que comprende un primer ácido nucleico de inserción que comprende un segundo ácido nucleico que codifica un segundo marcador de selección unido operativamente a un segundo promotor, en donde el primer ácido nucleico de inserción está flanqueado por un primer y un segundo brazos de homología que corresponden a un primer y un segundo sitios objetivo ubicados en el primer locus objetivo; (c) identificar una célula modificada que comprende el primer ácido nucleico de inserción en el primer locus objetivo, en donde la célula modificada tiene la actividad del segundo marcador de selección pero no tiene la actividad del primer marcador de selección, y en donde el primer y el segundo marcadores de selección son diferentes; (d) introducir en la célula modificada que comprende el primer ácido nucleico de inserción en el primer locus objetivo: (i) uno o más ácidos nucleicos que codifican la proteína Cas y un segundo ARNg, cada uno de los cuales está unido operativamente al promotor activo en la célula modificada, en donde la proteína Cas induce el corte o ruptura de doble cadena en un segundo sitio objetivo de ARNg en el primer ácido nucleico de inserción que comprende el segundo ácido nucleico, interrumpiendo así la expresión o actividad del segundo marcador de selección; y (ii) un segundo vector de transformación que comprende un segundo ácido nucleico de inserción que comprende un tercer ácido nucleico que codifica un tercer marcador de selección unido operativamente a un tercer promotor, en donde el segundo ácido nucleico de inserción está flanqueado por un tercer y cuarto brazos de homología correspondientes a un tercer y un cuarto sitios objetivo ubicados en un segundo locus objetivo; (e) identificar una segunda célula modificada que comprende el segundo ácido nucleico de inserción en el segundo locus objetivo, en donde la segunda célula modificada tiene la actividad del tercer marcador de selección pero no tiene la actividad del segundo marcador de selección, en donde el segundo y el tercer marcadores de selección son diferentes; (f) introducir en la segunda célula modificada que comprende el segundo ácido nucleico de inserción en el segundo locus objetivo: (i) la una o más construcciones de expresión que codifican la proteína Cas y un tercer ARNg, cada uno de los cuales está unido operativamente al promotor activo en la segunda célula modificada, en donde la proteína Cas induce el corte o ruptura de doble cadena en un tercer sitio objetivo de ARNg en el segundo ácido nucleico de inserción que comprende el tercer ácido nucleico, interrumpiendo así la expresión o actividad del tercer marcador de selección; y (ii) un tercer vector de transformación que comprende un tercer ácido nucleico de inserción que comprende un cuarto ácido nucleico que codifica un cuarto marcador de selección unido operativamente a un cuarto promotor, en donde el tercer ácido nucleico de inserción está flanqueado por un quinto y un sexto brazos de homología correspondientes a un quinto y un sexto sitios objetivo ubicados en un tercer locus objetivo; y (g) identificar una tercera célula modificada que comprende el tercer ácido nucleico de inserción en el tercer locus objetivo, en donde la tercera célula modificada tiene la actividad del cuarto marcador de selección pero no tiene la actividad del tercer marcador de selección, en donde el tercer y el cuarto marcadores de selección son diferentes. En algunas realizaciones, el primer y tercer marcadores de selección son iguales y el segundo y cuarto marcadores de selección son iguales. En una realización, el primer y el tercer marcadores de selección son iguales, el segundo y el cuarto marcadores de selección son iguales, y el primer y el tercer ARNg son iguales.

IV. Polinucleótidos de Interés

Cualquier polinucleótido de interés puede estar contenido en los diversos polinucleótidos de inserción y, por lo tanto, integrarse en el locus objetivo. Los métodos descritos en la presente descripción proporcionan al menos 1, 2, 3, 4, 5, 6 o más polinucleótidos de interés para integrarse en el locus objetivo.

El polinucleótido de interés dentro del polinucleótido de inserción cuando se integra en el locus objetivo puede introducir una o más modificaciones genéticas en la célula. La modificación genética puede comprender una delección de una secuencia de ácido nucleico endógeno y/o la adición de un polinucleótido exógeno, heterólogo u ortólogo en el locus objetivo. En una realización, la modificación genética comprende un reemplazo de una secuencia de ácido nucleico endógeno con un polinucleótido exógeno de interés en el locus objetivo. Por lo tanto, los métodos proporcionados en esta descripción permiten la generación de una modificación genética que comprende 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 estos. Dichas modificaciones pueden producirse tras la integración del primero, segundo, tercero, cuarto, quinto, sexto, séptimo o cualquier polinucleótido de inserción posterior en el locus objetivo.

El polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo puede comprender una secuencia que es nativa u homóloga a la célula en la que se introduce; el polinucleótido de interés puede ser heterólogo a la célula en la que se introduce; el polinucleótido de interés puede ser exógeno a la célula en la que se introduce; el polinucleótido de interés puede ser ortólogo a la célula en la que se introduce; o el polinucleótido de interés puede ser de una especie diferente a la célula en la que se introduce. El término "homólogo" en referencia a una secuencia incluye una secuencia que es nativa de la célula. El

término "heterólogo" en referencia a una secuencia incluye una secuencia que se origina de una especie extraña, o, si es de la misma especie, se modifica sustancialmente con relación a su forma nativa en cuanto a composición y/o lugar por una intervención humana deliberada. El término "exógeno" en referencia a una secuencia incluye una secuencia que se origina de una especie extraña. El término "ortólogo" incluye un polinucleótido de una especie que es funcionalmente equivalente a una secuencia de referencia conocida en otra especie (es decir, una variante de especie). El polinucleótido de interés puede ser de cualquier organismo de interés, incluidos, entre otros, no humanos, un roedor, un hámster, un ratón, una rata, un ser humano, un mono, un mamífero de uso agrícola o un mamífero de no uso agrícola. El polinucleótido de interés puede comprender además una región codificante, una región no codificante, una región reguladora o un ADN genómico. Por lo tanto, el 1ro, 2do, 3ro, 4to, 5to, 6to, 7mo, y/o cualquiera de los polinucleótidos de inserción posteriores pueden comprender tales secuencias.

En una realización, el polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo es homólogo a una secuencia de ácido nucleico de ratón, un ácido nucleico humano, un ácido nucleico no humano, un ácido nucleico de roedor, un ácido nucleico de rata, un ácido nucleico de hámster, un ácido nucleico de mono, un ácido nucleico de mamífero de uso agrícola o un ácido nucleico de mamífero de no uso agrícola. En otras realizaciones adicionales, el polinucleótido de interés integrado en el locus objetivo es un fragmento de un ácido nucleico genómico. En una realización, el ácido nucleico genómico es un ácido nucleico genómico de ratón, un ácido nucleico genómico humano, un ácido nucleico no humano, un ácido nucleico de roedor, un ácido nucleico de rata, un ácido nucleico de hámster, un ácido nucleico de mono, un ácido nucleico de mamífero de uso agrícola o un ácido nucleico de mamífero de no uso agrícola o una combinación de estos.

En una realización, el polinucleótido de interés puede variar de aproximadamente 500 nucleótidos a aproximadamente 200 kb como se describió anteriormente. El polinucleótido de interés puede ser de aproximadamente 500 nucleótidos a aproximadamente 5 kb, de aproximadamente 5 kb a aproximadamente 200 kb, de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 70 kb, de aproximadamente 70 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 110 kb, de aproximadamente 110 kb a aproximadamente 120 kb, de aproximadamente 120 kb a aproximadamente 130 kb, de aproximadamente 130 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 170 kb, de aproximadamente 170 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 190 kb, o de aproximadamente 190 kb a aproximadamente 200 kb.

El polinucleótido de interés dentro del polinucleótido de inserción y/o insertado en el locus objetivo puede codificar un polipéptido, puede codificar un miARN, o puede comprender cualquier región reguladora o región no codificante de interés que incluya, por ejemplo, una secuencia reguladora, una secuencia promotora, una secuencia potenciadora, una secuencia de unión a un represor transcripcional, o una delección de una secuencia que no codifica proteínas. Además, el polinucleótido de interés dentro del polinucleótido de inserción y/o insertado en el locus objetivo puede codificar una proteína expresada en el sistema nervioso, el sistema esquelético, el sistema digestivo, el sistema circulatorio, el sistema muscular, el sistema respiratorio, el sistema cardiovascular, el sistema linfático, el sistema endocrino, el sistema urinario, el sistema reproductor o una combinación de estos. En una realización, el polinucleótido de interés dentro del polinucleótido de inserción y/o insertado en el locus objetivo codifica una proteína expresada en una médula ósea o una célula derivada de médula ósea. En una realización, el polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo codifica una proteína expresada en una célula del bazo. En otras realizaciones adicionales, el polinucleótido de interés dentro del polinucleótido de inserción y/o insertado en el locus objetivo codifica una proteína expresada en una célula B, codifica una proteína expresada en una célula B inmadura o codifica una proteína expresada en una célula B madura.

En una realización, el polinucleótido de interés dentro del polinucleótido de inserción y/o insertado en el locus objetivo comprende una secuencia de ácido nucleico genómico que codifica una secuencia de aminoácidos de una región variable de cadena pesada de inmunoglobulina. La frase "cadena pesada" o "cadena pesada de inmunoglobulina" incluye una secuencia de una cadena pesada de inmunoglobulina, que incluye una secuencia de una región constante de cadena pesada de inmunoglobulina, de cualquier organismo. Los dominios variables de la 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 estos. Una cadena pesada típica tiene, después del dominio variable (de N-terminal a C-terminal), un dominio CH1, una bisagra, un dominio CH2 y un dominio CH3. Un fragmento funcional de una cadena pesada incluye un fragmento que es capaz de reconocer específicamente un epítipo (por ejemplo, reconocer el epítipo con una KD en el rango micromolar, nanomolar o picomolar), que es capaz de expresarse y secretarse de una célula, y que comprende al menos una CDR. Los dominios variables de la cadena pesada están codificados por la secuencia de nucleótidos de la región variable, que generalmente comprende segmentos VH, DH y JH derivados de un repertorio de segmentos VH, DH y JH presentes en la línea germinal. Las secuencias, ubicaciones y nomenclatura para los 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".

En una realización, el polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo comprende una secuencia de ácido nucleico genómico que codifica una secuencia de aminoácidos de una región variable de cadena pesada de inmunoglobulina humana. En una realización, la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de una región variable de cadena pesada de inmunoglobulina humana no reorganizada unida operativamente a una secuencia de ácido nucleico de una región constante de cadena pesada de inmunoglobulina. En una realización, la secuencia de ácido nucleico de una región constante de cadena pesada de inmunoglobulina es una secuencia de ácido nucleico de una región constante de cadena pesada de inmunoglobulina de ratón o una secuencia de ácido nucleico de una región constante de cadena pesada de inmunoglobulina humana, o una combinación de estas. En una realización, la secuencia de ácido nucleico de una región constante de cadena pesada de inmunoglobulina se selecciona de un CH1, una bisagra, un CH2, un CH3 y una combinación de estos. En una realización, la secuencia de ácido nucleico de una región constante de cadena pesada comprende CH1-bisagra-CH2-CH3. En una realización, la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de una región variable de cadena pesada de inmunoglobulina humana reorganizada unida operativamente a una secuencia de ácido nucleico de una región constante de cadena pesada de inmunoglobulina. En una realización, la secuencia de ácido nucleico de una región constante de cadena pesada de inmunoglobulina es una secuencia de ácido nucleico de una región constante de cadena pesada de inmunoglobulina de ratón o una secuencia de ácido nucleico de una región constante de cadena pesada de inmunoglobulina humana, o una combinación de estas. En una realización, la secuencia de ácido nucleico de una región constante de cadena pesada de inmunoglobulina se selecciona de un CH1, una bisagra, un CH2, un CH3 y una combinación de estos. En una realización, la secuencia de ácido nucleico de una región constante de cadena pesada comprende CH1-bisagra-CH2-CH3.

En una realización, el polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo comprende una secuencia de ácido nucleico genómico que codifica una secuencia de aminoácidos de una región variable de cadena ligera de inmunoglobulina. 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 un VpreB, así como cadenas ligeras sustitutas. Los dominios variables de la cadena ligera generalmente 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 terminal al carboxilo terminal, un dominio variable que incluye FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4, y una secuencia de aminoácidos de la región constante de una cadena ligera. Los dominios variables de la cadena ligera están codificados por la secuencia de nucleótidos de una región variable de cadena ligera, que generalmente comprende segmentos de genes VL de cadena ligera y JL de cadena ligera, derivados de un repertorio de segmentos de genes de cadena ligera V y J presentes en la línea germinal. Las secuencias, ubicaciones y nomenclatura de los segmentos de genes de cadena ligera V 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". Las cadenas ligeras incluyen aquellas, por ejemplo, que no se unen selectivamente ni a un primer ni a un segundo epítipo unidos selectivamente por la proteína de unión al epítipo en la que aparecen. Las cadenas ligeras también incluyen aquellas que se unen y reconocen, o ayudan a la cadena pesada a unirse y reconocer uno o más epítopos unidos selectivamente por la proteína de unión al epítipo en la que aparecen.

En una realización, el polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo comprende una secuencia de ácido nucleico genómico que codifica una secuencia de aminoácidos de una región variable de cadena ligera de inmunoglobulina humana. En una realización, la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de una región variable de cadena ligera λ y/o κ humana no reorganizada. En una realización, la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de una región variable de cadena ligera λ y/o κ humana reordenada. En una realización, la secuencia de ácido nucleico de una región variable de cadena ligera λ y/o κ reorganizada o no está unida operativamente a una secuencia de ácido nucleico de una región constante de cadena ligera de inmunoglobulina humana, de rata o ratón, seleccionada de una secuencia de ácido nucleico de una región constante de cadena ligera λ y una secuencia de ácido nucleico de una región constante de cadena ligera κ .

El polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo puede codificar una proteína extracelular o un ligando para un receptor. En realizaciones específicas, el ligando codificado es una citocina. Las citocinas de interés incluyen una quimiocina seleccionada de CCL, CXCL, CX3CL y XCL. La citocina también puede comprender un factor de necrosis tumoral (TNF). En otras realizaciones, la citocina es una interleucina (IL). En una realización, la interleucina 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 interleucina es IL-2. En realizaciones específicas, dichos polinucleótidos de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo son de un ser humano y, en realizaciones más específicas, pueden comprender una secuencia humana.

El polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo puede codificar una proteína citoplasmática o una proteína de membrana. En una realización, la proteína de membrana es un receptor, tal como un receptor de citocina, un receptor de interleucina, un receptor alfa de interleucina 2, un receptor beta de interleucina 2 o un receptor gamma de interleucina 2.

- 5 El polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo puede comprender un polinucleótido que codifica al menos una región de un receptor de células T, que incluye el receptor alfa de células T. En métodos específicos, cada uno de los polinucleótidos de inserción comprende una región del locus del receptor de células T (es decir, el locus del receptor alfa de células T) de modo que al completarse la integración en serie, una parte o la totalidad del locus del receptor de células T se ha integrado
10 en el locus objetivo. Tales polinucleótidos de inserción pueden comprender al menos uno o más de un segmento variable o un segmento de unión de un locus del receptor de células T (es decir, del locus del receptor alfa de células T). Además, el polinucleótido de interés que codifica la región del receptor de células T puede ser, por ejemplo, de un polinucleótido de mamífero, de un mamífero no humano, de un roedor, de un ratón, de una rata, de un ser humano, de un mono, de un mamífero de uso agrícola o de un mamífero doméstico, que
15 codifica una proteína mutante.

En otras realizaciones, el polinucleótido de interés integrado en el locus objetivo codifica una proteína nuclear. En una realización, la proteína nuclear es un receptor nuclear. En realizaciones específicas, dichos polinucleótidos de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo son de un ser humano y, en realizaciones más específicas, pueden comprender una secuencia genómica humana.

- 20 El polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus genómico objetivo puede comprender una modificación genética en una secuencia codificante. Dichas modificaciones genéticas incluyen, entre otras, una mutación por delección de una secuencia codificante o la fusión de dos secuencias codificantes.

- El polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo puede comprender un polinucleótido que codifica una proteína mutante. En una realización, la proteína mutante se caracteriza por una característica de unión alterada, localización alterada, expresión alterada y/o patrón de expresión alterado. En una realización, el polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo comprende al menos un alelo de una enfermedad, que incluye, por ejemplo, un
25 alelo de una enfermedad neurológica, un alelo de una enfermedad cardiovascular, un alelo de una enfermedad renal, un alelo de una enfermedad muscular, un alelo de una enfermedad de la sangre, un alelo de un gen que provoca cáncer o un alelo de una enfermedad del sistema inmunitario. En tales casos, el alelo de la enfermedad puede ser un alelo dominante o el alelo de la enfermedad es un alelo recesivo. Además, el alelo de la enfermedad puede comprender un alelo de polimorfismo de un solo nucleótido (SNP). El polinucleótido de interés que codifica la proteína mutante puede ser de cualquier organismo, incluido, entre otros, un
30 polinucleótido de un mamífero, de un mamífero no humano, de un roedor, de un ratón, de una rata, de un ser humano, de un mono, de un mamífero de uso agrícola o de un mamífero doméstico, que codifica una proteína mutante.

- El polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo también puede comprender una secuencia reguladora, que incluye, por ejemplo, una secuencia promotora, una
40 secuencia potenciadora o una secuencia de unión a un represor transcripcional. En realizaciones específicas, el polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus objetivo comprende un polinucleótido que tiene una delección de una secuencia que no codifica proteínas, pero no comprende una delección de una secuencia codificante de proteínas. En una realización, la delección de la secuencia que no codifica proteínas comprende una delección de una secuencia reguladora. En otra realización, la delección del
45 elemento regulador comprende una delección de una secuencia promotora. En una realización, la delección del elemento regulador comprende una delección de una secuencia potenciadora. Tal polinucleótido de interés puede ser de cualquier organismo, incluyendo, entre otros, un polinucleótido de un mamífero, de un mamífero no humano, de un roedor, de un ratón, de una rata, de un ser humano, de un mono, de un mamífero de uso agrícola o de un mamífero doméstico, que codifica una proteína mutante.

- 50 V. Métodos de introducción de secuencias y generación de animales transgénicos no humanos

Como se describió anteriormente, en la presente descripción se proporcionan métodos y composiciones para permitir la integración dirigida de uno o más polinucleótidos de interés. Dichos sistemas emplean una variedad de componentes y para facilitar la referencia, en esta descripción el término "sistema de integración dirigido" se refiere genéricamente a todos los componentes requeridos para un evento de integración (es decir, los
55 diversos agentes de nucleasa, sitios de reconocimiento, polinucleótidos de ADN de inserción, vectores de transformación, locus objetivo y polinucleótidos de interés).

Los métodos proporcionados en la presente descripción comprenden introducir en una célula uno o más polinucleótidos o construcciones de polipéptidos que comprenden los diversos componentes del sistema de integración dirigido. El término "introducir" incluye presentar a la célula la secuencia (polipéptido o

polinucleótido) de tal manera que la secuencia tenga acceso al interior de la célula. Los métodos proporcionados en la presente descripción no dependen de un método particular para introducir ningún componente del sistema de integración dirigido en la célula, solo que el polinucleótido obtenga acceso al interior de al menos una célula. Los métodos para introducir polinucleótidos en diversos tipos de células son conocidos en la técnica e incluyen, entre otros, métodos de transfección estable, métodos de transfección transitoria y métodos mediados por virus.

En algunas realizaciones, las células empleadas en los métodos y composiciones tienen una construcción de ADN incorporada establemente en su genoma. "Incorporado establemente" o "introducido establemente" significa la introducción de un polinucleótido en la célula de tal manera que la secuencia de nucleótidos se integra en el genoma de la célula y es capaz de ser heredada por su progenie. Se puede usar cualquier protocolo para la incorporación estable de las construcciones de ADN o los diversos componentes del sistema de integración dirigido.

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

En una realización, el agente de nucleasa se introduce en la célula simultáneamente con el vector de transformación o el vector de transformación grande (LTVEC). Alternativamente, el agente de nucleasa se introduce por separado del vector de transformación o el LTVEC durante un período de tiempo. En una realización, el agente de nucleasa se introduce antes de la introducción del vector de transformación o el LTVEC, mientras que en otras realizaciones, el agente de nucleasa se introduce después de la introducción del vector de transformación o el LTVEC.

Se pueden generar animales mamíferos no humanos empleando los diversos métodos descritos en esta descripción. Tales métodos comprenden (1) integrar uno o más polinucleótidos de interés en el locus objetivo de una célula pluripotente del animal no humano para generar una célula pluripotente genéticamente modificada que comprende el polinucleótido de inserción en el locus objetivo empleando los métodos descritos en esta descripción; (2) seleccionar la célula pluripotente genéticamente modificada que tiene uno o más polinucleótidos de interés en el locus objetivo; (3) introducir la célula pluripotente modificada genéticamente en un embrión huésped del animal no humano en una etapa pre-mórula; y (4) implantar el embrión huésped que comprende la célula pluripotente modificada genéticamente en una madre sustituta para generar una generación F0 derivada de la célula pluripotente modificada genéticamente. El animal no humano puede ser un mamífero no humano, un roedor (por ejemplo, un ratón, una rata, un hámster), un mono, un mamífero de uso agrícola o un mamífero doméstico. La célula pluripotente puede ser una célula ES no humana, una célula ES de roedor (por ejemplo, una célula ES de ratón, una célula ES de rata o una célula ES de hámster), una célula ES de mono, una célula ES de mamífero de uso agrícola o una célula ES de un mamífero domesticado. Ver, por ejemplo, la publicación de EE. UU. núm. 2014/0235933; publicación de Estados Unidos núm. 2014/0310828; y Tong y otros (2010) *Nature*, 467(7312):211-213.

Las técnicas de transferencia nuclear también se pueden utilizar para generar animales mamíferos no humanos. Brevemente, los métodos para la transferencia nuclear incluyen las etapas de: (1) enuclear un ovocito; (2) aislar una célula o núcleo de donante para combinar con el ovocito enucleado; (3) insertar la célula o núcleo en el ovocito enucleado para formar una célula reconstituida; (4) implantar la célula reconstituida en el útero de un animal para formar un embrión; y (5) permitir que se desarrolle el embrión. En tales métodos, los ovocitos generalmente se obtienen de animales fallecidos, aunque también pueden aislarse a partir de oviductos y/u ovarios de animales vivos. Los ovocitos pueden madurarse en una variedad de medios conocidos por los expertos en la técnica antes de la enucleación. La enucleación del ovocito se puede realizar de varias maneras bien conocidas por los expertos en la técnica. La inserción de la célula o núcleo donante en el ovocito enucleado para formar una célula reconstituida suele realizarse mediante microinyección de una célula donante debajo de la zona pelúcida antes de la fusión. La fusión puede ser inducida por la aplicación de un pulso eléctrico de CD a través del plano de contacto/fusión (electrofusión), por la exposición de las células a sustancias químicas que promueven la fusión, como el polietilenglicol, o por medio de un virus inactivado, como el virus Sendai. Una célula reconstituida se activa típicamente por medios eléctricos y/o no eléctricos antes, durante y/o después de la fusión del donante nuclear y el ovocito receptor. Los métodos de activación incluyen pulsos eléctricos, choque inducido químicamente, penetración de espermatozoides, aumento de los niveles de cationes divalentes en el ovocito y reducción de la fosforilación de proteínas celulares (como inhibidores de quinasa) en el ovocito. Las células reconstituidas activadas, o embriones, se cultivan típicamente en un medio bien conocido por los expertos en la técnica y luego se transfieren al útero de un animal. Ver, por ejemplo, los documentos

US20080092249, WO/1999/005266A2, US20040177390, WO/2008/017234A1y la patente de los Estados Unidos núm. 7,612,250.

Se proporcionan otros métodos para obtener un animal no humano que comprenda en su línea germinal una o más modificaciones genéticas como se describe en esta descripción, que comprenden: (a) modificar un locus objetivo de un animal no humano en una célula procariota empleando los diversos métodos descritos en esta descripción; (b) seleccionar una célula procariota modificada que comprende la modificación genética en el locus objetivo; (c) aislar el vector de transformación genéticamente modificado de la célula procariota modificada; (d) introducir el vector de transformación modificado genéticamente en una célula pluripotente del animal no humano para generar una célula pluripotente modificada genéticamente que comprende el ácido nucleico de inserción en el locus objetivo; (e) seleccionar la célula pluripotente modificada genéticamente; (f) introducir la célula pluripotente modificada genéticamente en un embrión huésped del animal no humano en una etapa pre-mórula; y (g) implantar el embrión huésped que comprende la célula pluripotente modificada genéticamente en una madre sustituta para generar una generación F0 derivada de la célula pluripotente modificada genéticamente. En tales métodos, el vector de transformación puede comprender un vector de transformación grande. El animal no humano puede ser un mamífero no humano, un roedor, un ratón, una rata, un hámster, un mono, un mamífero de uso agrícola o un mamífero doméstico. La célula pluripotente puede ser una célula ES no humana, una célula ES de roedor (por ejemplo, una célula ES de ratón, una célula ES de rata o una célula ES de hámster), una célula ES de mono, una célula ES de mamífero de uso agrícola o una célula ES de mamífero doméstico.

En otros métodos, la etapa de aislamiento (c) comprende además (c1) linealizar el vector de transformación modificado genéticamente (es decir, el LTVEC modificado genéticamente). En otras realizaciones adicionales, la etapa de introducción (d) comprende además (d1) introducir un agente de nucleasa como se describe en la presente descripción en la célula pluripotente. En otras realizaciones, la etapa de introducción (d) comprende además (d2) en donde la célula pluripotente del mamífero no humano comprende un locus objetivo que comprende un primer polinucleótido que codifica un primer marcador de selección unido operativamente a un primer promotor activo en la célula, en donde el primer polinucleótido comprende además un primer sitio de reconocimiento para un primer agente de nucleasa, e introducir un agente de nucleasa en la célula pluripotente, en donde el primer agente de nucleasa induce un corte o ruptura de doble cadena en el primer sitio de reconocimiento. Además, se introduce en la célula pluripotente un primer vector de transformación que comprende el vector de transformación modificado genéticamente del genoma de la célula procariota modificada. El vector de transformación modificado comprende un primer y un segundo brazos de homología que corresponden a un primer y un segundo sitios objetivo en suficiente proximidad al primer sitio de reconocimiento dentro del genoma de la célula pluripotente del mamífero no humano. En una realización, las etapas de selección (b) y/o (e) se llevan a cabo aplicando un agente de selección como se describe en la presente descripción a la célula procariota o la célula pluripotente. En una realización, las etapas de selección (b) y/o (e) se llevan a cabo mediante un ensayo de modificación de alelos (MOA) como se describe en la presente descripción.

Se proporcionan métodos adicionales para modificar un locus objetivo de una célula de mamífero mediante recombinación homóloga bacteriana (BHR) en una célula procariota y comprenden: (a) proporcionar una célula procariota que comprende un locus objetivo que comprende un primer polinucleótido que codifica un primer marcador de selección unido operativamente a un primer promotor activo en la célula procariota, en donde el primer polinucleótido comprende además un primer sitio de reconocimiento para un primer agente de nucleasa, (b) introducir en la célula procariota un vector de transformación que comprende un polinucleótido de inserción flanqueado con un primer brazo de homología hacia el extremo 5' y un primer brazo de homología hacia el extremo 3', en donde el polinucleótido de inserción comprende una región de mamífero, e introducir en la célula procariota un agente de nucleasa que produce un corte o ruptura de doble cadena en o cerca del primer sitio de reconocimiento, y (c) seleccionar una célula procariota específica que comprende el polinucleótido de inserción en el locus objetivo, en donde la célula procariota es capaz de expresar proteínas recombinogénicas y enzimas que median la BHR. Las etapas (a) - (c) se pueden repetir en serie como se describe en esta descripción para permitir la introducción de múltiples polinucleótidos de inserción en el locus objetivo en la célula procariota. Una vez que el locus objetivo es "construido" con la célula procariota, un vector de transformación que comprende el locus objetivo modificado puede aislarse de la célula procariota e introducirse en un locus objetivo dentro de una célula de mamífero no humano. Las células de mamíferos que comprenden el locus modificado pueden convertirse en animales transgénicos no humanos.

En algunas realizaciones, varias modificaciones genéticas de los loci objetivo descritos en esta descripción pueden llevarse a cabo mediante una serie de reacciones de recombinación homóloga (BHR) en células bacterianas con el uso de ADN del cromosoma artificial bacteriano (BAC) que emplea tecnología de ingeniería genética VELOCIGENE® (ver, por ejemplo, la patente núm. 6,586,251 y Valenzuela, DM y otros (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 mamífero no humano específicas (es decir, de mamíferos no humanos, roedores (por ejemplo, ratones, ratas o hámsters), mamíferos de uso agrícola, mamíferos

domésticos, monos, etc.) que comprenden diversas modificaciones genéticas como se describe en esta descripción, se introducen en un embrión en etapa pre-mórula de un organismo correspondiente, por ejemplo, un embrión de ratón en etapa de 8 células, a través del método VELOCIMOUSE® (ver, por ejemplo, los documentos US 7,576,259, US 7,659,442, US 7,294,754 y US 2008-0078000 A1). El embrión de mamífero no humano que comprende las células ES no humanas genéticamente modificadas se incuba hasta la etapa de blastocisto y luego se implanta en una madre sustituta para producir un F0. En algunas otras realizaciones, las células ES de mamífero humano específicas que comprenden diversas modificaciones genéticas, como se describe en la presente descripción, se introducen en un embrión en etapa de blastocisto. Los mamíferos no humanos que portan el locus modificado genéticamente pueden identificarse mediante el ensayo de modificación de alelos (MOA) como se describe en la presente descripción. El mamífero no humano resultante de la generación F0 derivado de las células ES no humanas genéticamente modificadas se cruza con un mamífero no humano de tipo silvestre para obtener la descendencia de la generación F1. Después de la genotipificación con cebadores y/o sondas específicos, los mamíferos no humanos F1 que son heterocigotos para el locus modificado genéticamente se cruzan entre sí para producir mamíferos no humanos que son homocigotos para el locus modificado genéticamente.

VI. Células

Los diversos métodos descritos en esta descripción emplean un sistema de direccionamiento al locus en una célula. Dichas células incluyen células procariotas tales como células bacterianas que incluyen *E. coli*, o células eucariotas como de levaduras, insectos, anfibios, aves (por ejemplo, células de pollo), células vegetales o de mamíferos, incluidas, entre otras, una célula de ratón, una célula de rata, una célula de conejo, una célula de cerdo, una célula de bovino, una célula de venado, una célula de oveja, una célula de cabra, una célula de gato, una célula de perro, una célula de hurón, una célula de primate (por ejemplo, mono tití, mono rhesus) y similares y células de mamíferos domesticados o células de mamíferos de uso agrícola. Algunas células son no humanas, particularmente células de mamíferos no humanos. En algunas realizaciones, para los mamíferos para los cuales las células pluripotentes genéticamente modificables adecuadas no están fácilmente disponibles, se emplean otros métodos para reprogramar las células somáticas a células pluripotentes, por ejemplo, mediante la introducción en las células somáticas de una combinación de factores inductores de pluripotencia, que incluyen, entre otros, Oct3/4, Sox2, KLF4, Myc, Nanog, LIN28 y Glis1.

En una realización, la célula eucariota es una célula pluripotente. En una realización, la célula pluripotente es una célula madre embrionaria (ES) no humana. El término "célula madre embrionaria" o "célula ES" incluye una célula totipotente o pluripotente derivada del embrión que es capaz de proliferación indiferenciada in vitro, y es capaz de contribuir a cualquier tejido del embrión en desarrollo tras su introducción en un embrión. El término "célula pluripotente" incluye una célula indiferenciada que posee la capacidad de convertirse en más de un tipo de célula diferenciada. El término "línea germinal" en referencia a una secuencia de polinucleótido incluye una secuencia de ácido nucleico que se puede pasar a la progenie.

La célula pluripotente puede ser una célula ES no humana o una célula madre pluripotente inducida (iPS). En una realización, la célula pluripotente inducida (iPS) se deriva de un fibroblasto. En realizaciones específicas, la célula pluripotente inducida (iPS) se deriva de un fibroblasto humano. En algunas realizaciones, la célula pluripotente es una célula madre hematopoyética (HSC), una célula madre neuronal (NSC) o una célula madre epiblastica. La célula pluripotente también puede ser una célula progenitora de desarrollo restringido. En realizaciones adicionales, la célula pluripotente es una célula pluripotente de roedor. En una realización, la célula pluripotente de roedor es una célula pluripotente de rata o una célula ES de rata. En otras realizaciones, la célula pluripotente de roedor es una célula pluripotente de ratón o una célula ES de ratón.

En otras realizaciones, la célula de mamífero puede ser células de ratón, células de ratas o células humanas inmortalizadas. En una realización, la célula de mamífero es un fibroblasto humano, mientras que en otras realizaciones, la célula de mamífero es una célula cancerosa, que incluye una célula cancerosa humana.

En otras realizaciones adicionales, el mamífero es un ser humano y la transformación se lleva a cabo utilizando una célula humana ex vivo.

En una realización, la célula de mamífero es una célula humana aislada de un paciente que tiene una enfermedad y/o comprende un polinucleótido humano que codifica una proteína 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 una enfermedad humana. En una realización, la secuencia de ácido nucleico humano comprende al menos un alelo de una 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 que provoca cáncer. En una realización, el alelo de la enfermedad humana es un alelo de una enfermedad del sistema inmunitario.

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 nucleótido simple (SNP).

Cuando la célula comprende una célula procariota, en realizaciones específicas, la célula procariota es una cepa de *E. coli competente para la recombinación*. En una realización, la célula procariota comprende un ácido nucleico que codifica proteínas y enzimas recombinogénicas. En una realización, la célula procariota no comprende el ácido nucleico que codifica las proteínas y enzimas recombinogénicas, y el ácido nucleico que codifica las proteínas y enzimas recombinogénicas se introduce en la célula procariota. En una realización, el ácido nucleico comprende un ADN o un ARNm que codifica las proteínas y enzimas recombinogénicas. En una realización, el ácido nucleico que codifica las proteínas y enzimas recombinogénicas es pABG. En una realización, las proteínas y enzimas recombinogénicas se expresan bajo el control de un promotor inducible. En una realización, la expresión de las proteínas y enzimas recombinogénicas está controlada por arabinosa.

VII. Casetes de expresión

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

Los términos "polinucleótido", "secuencia de polinucleótido", "secuencia de ácido nucleico" y "fragmento de ácido nucleico" se usan indistintamente en la presente descripción. Estos términos abarcan secuencias de nucleótidos y similares. Un polinucleótido puede ser un polímero de ARN o ADN que es de cadena simple o doble, que contiene opcionalmente bases nucleotídicas sintéticas, no naturales o alteradas. Un polinucleótido en la forma de un polímero de ADN puede estar compuesto por uno o más segmentos de ADNc, ADN genómico, ADN sintético o mezclas de estos. Los polinucleótidos pueden comprender desoxirribonucleótidos y ribonucleótidos e incluyen tanto moléculas naturales como análogos sintéticos, y cualquier combinación de estos. Los polinucleótidos proporcionados en la presente descripción también abarcan todas las formas de secuencias que incluyen, entre otras, formas monocatenarias, formas bicatenarias, horquillas, estructuras de tallo y bucle, y similares.

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

En realizaciones específicas, uno o más de los componentes del sistema de integración dirigido descrito en esta descripción pueden proporcionarse en un casete de expresión para su expresión en una célula procariota, una célula eucariota, una bacteria, una célula de levadura, o una célula de mamífero u otro organismo o tipo de célula de interés. El casete puede incluir secuencias reguladoras en 5' y 3' unidas operativamente a un polinucleótido proporcionado en la presente descripción. "Unido operativamente" incluye una unión funcional entre dos o más elementos. Por ejemplo, una unión operativa entre un polinucleótido de interés y una secuencia reguladora (es decir, un promotor) es una unión funcional que permite la expresión del polinucleótido de interés. Los elementos unidos operativamente pueden ser contiguos o no. Cuando se usa para referirse a la unión de dos regiones codificantes de proteínas, unidas operativamente significa que las regiones codificantes están en el mismo marco de lectura. En otro caso, una secuencia de ácido nucleico que codifica una proteína puede estar unida operativamente a secuencias reguladoras (por ejemplo, promotor, potenciador, secuencia silenciadora, etc.) para mantener una regulación transcripcional adecuada. En un caso, una secuencia de ácido nucleico de una región variable de inmunoglobulina (o segmentos V(D)J) puede estar unida operativamente a una secuencia de ácido nucleico de una región constante de inmunoglobulina para permitir la recombinación adecuada entre las secuencias en una secuencia de cadena pesada o ligera de inmunoglobulina.

El casete puede contener adicionalmente al menos un polinucleótido de interés adicional para ser introducido conjuntamente en el organismo. Alternativamente, el polinucleótido de interés adicional puede proporcionarse en múltiples casetes de expresión. Dicho casete de expresión está provisto de una pluralidad de sitios de

restricción y/o sitios de recombinación para que la inserción de un polinucleótido recombinante esté bajo la regulación transcripcional de las regiones reguladoras. El casete de expresión puede contener adicionalmente genes marcadores de selección.

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

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

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

Se pueden usar varios promotores en los casetes de expresión proporcionados en la presente descripción. Los promotores pueden seleccionarse en función del resultado deseado. Se reconoce que se pueden mejorar diferentes aplicaciones mediante el uso de diferentes promotores en los casetes de expresión para modular el momento, la ubicación y/o el nivel de expresión del polinucleótido de interés. Tales construcciones de expresión también pueden contener, si se desea, una región reguladora del promotor (por ejemplo, una que confiera expresión inducible, constitutiva, regulada ambientalmente o por el desarrollo o específica/selectiva de células o tejidos), un sitio de inicio de la transcripción, un sitio de unión al ribosoma, una señal de procesamiento de ARN, un sitio de terminación de la transcripción y/o una señal de poliadenilación.

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

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

VIII. Identidad de secuencia

Los métodos proporcionados en la presente descripción emplean una variedad de componentes diferentes del sistema de integración dirigido (es decir, agentes de nucleasa, sitios de reconocimiento, polinucleótidos de inserción, polinucleótidos de interés, vectores de transformación, marcadores de selección y otros componentes). Se reconoce a lo largo de la descripción que algunos componentes del sistema de integración dirigido pueden tener variantes y fragmentos activos. Dichos componentes incluyen, por ejemplo, agentes de nucleasa (es decir, agentes de nucleasa modificados por ingeniería genética), sitios de reconocimiento de agentes de nucleasa, polinucleótidos de interés, sitios objetivo y brazos de homología correspondientes del vector de transformación. La actividad biológica para cada uno de estos componentes se describe en otra parte de la presente descripción.

Como se usa en la presente descripción, "identidad de secuencia" o "identidad" en el contexto de dos polinucleótidos o secuencias de polipéptidos hace referencia a los residuos en las dos secuencias que son iguales cuando se alinean por correspondencia máxima en una ventana de comparación especificada. Cuando se usa el porcentaje de identidad de secuencia como referencia para las proteínas se reconoce que las posiciones de los residuos que no son idénticos difieren frecuentemente en sustituciones de aminoácidos conservadores, donde los residuos de aminoácidos se sustituyen por otros residuos de aminoácidos con

propiedades químicas similares (por ejemplo, carga o hidrofobicidad) y por lo tanto no cambian las propiedades funcionales de la molécula. Cuando las secuencias difieren en las sustituciones conservadoras, el por ciento de identidad de secuencia se puede ajustar hacia arriba para corregir la naturaleza conservadora de la sustitución. Las secuencias que difieren en tales sustituciones conservadoras se dice que tienen "similitud de secuencia" o "similitud". Los medios para hacer este ajuste son bien conocidos por los expertos en la técnica. Típicamente esto implica calificar una sustitución conservadora como una incompatibilidad parcial en lugar de completa, de ese modo aumenta el porcentaje de identidad de secuencia. Así, por ejemplo, cuando un aminoácido idéntico recibe una puntuación de 1 y una sustitución no conservadora recibe una puntuación de cero, una sustitución conservadora recibe una puntuación entre cero y 1. La puntuación de las sustituciones conservadoras se calcula, por ejemplo, como se implementa en el programa PC/GENE (Intelligenetics, Mountain View, California).

Como se usa en la presente descripción, el "porcentaje de identidad de secuencia" significa el valor determinado por la comparación de dos secuencias óptimamente alineadas sobre una ventana de comparación, en donde la porción de la secuencia de polinucleótidos en la ventana de comparación puede comprender adiciones o deleciones (es decir, interrupciones) en comparación con la secuencia de referencia (que no comprende adiciones o deleciones) para la alineación óptima de las dos secuencias. El porcentaje se calcula al determinar la cantidad de posiciones en las que se presenta el residuo de aminoácido o la base del ácido nucleico idéntico en ambas secuencias para obtener la cantidad de posiciones coincidentes, dividir la cantidad de posiciones coincidentes por la cantidad total de posiciones en la ventana de comparación y multiplicar el resultado por 100 para obtener el porcentaje de identidad de secuencia.

A menos que se indique lo contrario, los valores de identidad/similitud de secuencia proporcionados en esta descripción se refieren al valor obtenido mediante el uso de GAP Versión 10 con los siguientes parámetros: % de identidad y % de similitud para una secuencia de nucleótidos usando un peso por interrupción de 50 y un peso por longitud de 3, y la matriz de puntuación nwsgapdna.cmp; % de identidad y % de similitud para una secuencia de aminoácidos usando un peso por interrupción de 8 y un peso por longitud de 2, y la matriz de puntuación BLOSUM62; o cualquier programa equivalente. "Programa equivalente" significa cualquier programa de comparación de secuencias que, para cualquiera de las dos secuencias en cuestión, genera una alineación que tiene coincidencias idénticas de residuos de nucleótidos o aminoácidos y un por ciento idéntico de identidad de secuencia cuando se compara con la alineación correspondiente generada por GAP Versión 10.

Los siguientes ejemplos se ofrecen en forma de ilustración y no en forma de limitación.

Ejemplos

Los experimentos de transformación génica secuencial representados en las Figs. 1 y 2 demostraron el valor de combinar un gran vector de transformación basado en BAC (LTVEC) con una nucleasa de dedos de zinc (ZFN) diseñada para reconocer y escindir una secuencia objetivo en un casete para la selección por fármacos.

Para la primera etapa en la transformación secuencial (Fig. 1), se construyó un LTVEC para crear una modificación (alelo de TCR α B-hyg) que inserta 136 kb de ADN que codifica 11 dominios variables (V) del receptor alfa de células T humanas (TCR α) en el locus de TCR α de ratón correspondiente. Se electroporaron 0,02 mg del LTVEC construido en 10 millones de células madre embrionarias de ratón (ES) que portaban una modificación creada previamente (alelo de TCR α A-neo) en el locus de TCR α , que reemplazó los segmentos de genes variables (V) y de unión (J) de ratón con V y J humanos. Después de la recuperación de la célula ES electroporada en un medio de cultivo, se añadió higromicina para seleccionar colonias derivadas de células que habían incorporado el LTVEC en sus genomas. El ensayo de modificación de alelos (MOA) de las colonias aisladas dio como resultado la identificación de cuatro clones transformados correctamente entre 136 colonias resistentes a la higromicina analizadas, para una eficacia de transformación del 2,9 % (Tabla 1, Experimento 1). Además de la inserción de los 11 V adicionales, los clones transformados correctamente tenían un casete de resistencia a la higromicina (hyg^r) que había reemplazado el casete de resistencia a la neomicina (G418) (neo^r).

El Experimento 2 fue idéntico al Experimento 1, excepto por las adiciones de 0,02 mg de cada uno de los dos plásmidos que expresaron cada mitad de Neo-ZFN(1,2), que se une a las secuencias de reconocimiento en el gen neor y cataliza una ruptura de doble cadena en el ADN. La inclusión de Neo-ZFN(1,2) dio como resultado 55 clones transformados correctamente de 568 clones resistentes a la higromicina seleccionados, para una eficiencia de transformación del 9,7 %, lo que representa una eficiencia de transformación 3,3 veces mayor en comparación con una electroporación con el LTVEC solo (Tabla 1, compare los Experimentos 1 y 2).

El experimento 3 fue idéntico al experimento 2, excepto que los plásmidos que codifican Neo-ZFN(3,4) reemplazaron a los de Neo-ZFN(1,2). La inclusión de Neo-ZFN(3,4) dio como resultado 42 clones transformados correctamente de 360 clones resistentes a la higromicina seleccionados, para una eficiencia de transformación del 11,7 %, lo que representa una eficiencia de transformación 4 veces mayor en comparación con una electroporación con LTVEC solo (Tabla 1, comparar los Experimentos 1 y 3).

Para la segunda etapa en la transformación secuencial (Fig. 2), la electroporación de 0,002 mg de un LTVEC diseñado para crear una modificación (alelo de TCR α C-neo) que inserta 157 kb de ADN que codifica 11 dominios variables (V) adicionales del TCR α humano, diferentes de aquellos en los alelos de TCR α A-neo o B-hyg en 10 millones de células madre embrionarias (ES) de ratón que portaban el alelo de TCR α B-hyg que se obtuvo en la primera etapa de transformación secuencial (Fig. 1), introdujo el LTVEC en las células ES. Después de la recuperación de la célula ES electroporada en un medio de cultivo, se añadió G418 para seleccionar colonias que se derivaban de células que habían incorporado el LTVEC en sus genomas. La selección por MOA de las colonias aisladas dio como resultado la identificación de dos clones transformados correctamente entre 192 colonias resistentes a G418 seleccionadas, para una eficacia de transformación del 1,0 % (Tabla 1, Experimento 4). Además de la inserción de los 11 V adicionales, los clones transformados correctamente tenían un casete Neor que había reemplazado al casete Hygr.

El experimento 5 fue idéntico al experimento 4, excepto por las adiciones de 0,02 mg de cada uno de los dos plásmidos que expresaban cada mitad del Hyg-ZFN(1,2), que se une a las secuencias de reconocimiento en hyg^r y cataliza una ruptura de doble cadena en el ADN. La inclusión de Hyg-ZFN(1,2) dio como resultado 40 clones transformados correctamente de 192 clones resistentes G418 seleccionados, para una eficiencia de transformación del 21 %, lo que representa una eficiencia de transformación 21 veces mayor en comparación con una electroporación con el LTVEC solo (Tabla 1, compare los Experimentos 4 y 5).

El experimento 6 fue idéntico al experimento 5, excepto que los plásmidos que codifican Hyg-ZFN(3,4) reemplazaron a los de Hyg-ZFN(1,2). La inclusión de Hyg-ZFN(3,4) dio como resultado 42 clones transformados correctamente de 192 clones resistentes a la higromicina seleccionados, para una eficiencia de transformación del 22 %, lo que representa una eficiencia de transformación 22 veces mayor en comparación con una electroporación con el LTVEC solo (Tabla 1, compare los Experimentos 4 y 6).

Los experimentos resumidos en la Tabla 1 establecieron que la inclusión de ZFN que se dirigen a los casetes de selección neo^r o hyg^r con LTVEC en experimentos de transformación secuencial puede mejorar la eficiencia de transformación en un factor de 3 a 20 veces en comparación con los experimentos de transformación que incluyen solo el LTVEC. Las mayores eficiencias de transformación que se obtienen por la inclusión de las ZFN en los experimentos de transformación secuencial promovieron la inserción correcta prevista de piezas muy grandes (136 kb y 157 kb) de ADN humano precisamente en la posición cromosómica deseada de un alelo previamente modificado. La transformación mejorada con ZFN aumenta en gran medida la probabilidad de éxito en un proyecto de transformación y ofrece un ahorro importante de tiempo, mano de obra y costes de materiales para la selección de células ES.

Tabla 1. Mejora de la transformación génica secuencial por nucleasas de dedos de zinc que reconocen las secuencias neo^r e hyg^r

Experimento	ADN electroporado	Célula ES receptora	Eficiencia de transformación ¹ (%)	Mejora de la transformación ²
1	i. Alelo de TCR α humanizado B-hyg LTVEC ³	Alelo de TCR α humanizado A-neo ³	2,9 (4/136)	n.a.
2	i. Alelo de TCR α humanizado B-hyg LTVEC ii. Plásmido de Neo-ZFN1 ⁴ iii. Plásmido de Neo-ZFN2 ⁴	Alelo de TCR α humanizado A-neo	9,7 (55/568)	3,3X
3	i. Alelo de TCR α humanizado B-hyg LTVEC ii. Plásmido de Neo-ZFN3 ⁵ iii. Plásmido de Neo-ZFN4 ⁵	Alelo de TCR α humanizado A-neo	11,7 (42/360)	4,0X
4	i. Alelo de TCR α humanizado C-neo LTVEC ⁶	Alelo de TCR α humanizado B-hyg ⁶	1,0 (2/192)	n.a.

5	i. Alelo de TCR α humanizado C-neo LTVEC	Alelo de TCR α humanizado B-hyg	21 (40/192)	21X
	ii. Plásmido de Hyg-ZFN1 ⁷			
	iii. Plásmido de Hyg-ZFN2 ⁷			
6	i. Alelo de TCR α humanizado B-hyg LTVEC	Alelo de TCR α humanizado B-hyg	22 (42/192)	22X
	ii. Plásmido de Hyg-ZFN3 ⁸			
	iii. Plásmido de Hyg-ZFN4 ⁸			

¹Proporción de clones transformados correctamente a clones totales seleccionados (mostrados entre paréntesis), expresada como porcentaje

²Proporción de la eficiencia de transformación de un experimento que combinó un vector de transformación con plásmidos de ZFN con relación a un experimento que utilizó el vector de transformación solo

³Ver la Fig. 1

⁴Neo-ZFN(1,2) (ver la Fig. 3) SITIO DE UNIÓN A NUCLEASA/sitio de corte:
GGGCGCCCGGTTCTTTT/gtcaag/ACCGACCTGTCCGGTG (SEQ ID NO: 9)

⁵Neo-ZFN(3,4) (ver la Fig. 3) SITIO DE UNIÓN A NUCLEASA/sitio de corte:
CCGTTCTTTTGTG/aagacc/GACCTGTCCGGTGCC (SEQ ID NO: 10)

⁶Ver la Fig. 2

⁷Hyg-ZFN(1,2) (ver la Fig. 3) SITIO DE UNIÓN A NUCLEASA/sitio de corte:
TGCGATCGCTGCGGCCGA/tcttag/CCAGACGAGCGGGTTTCGG (SEQ ID NO: 11)

⁸Hyg-ZFN(3,4) (ver la Fig. 3) SITIO DE UNIÓN A NUCLEASA/sitio de corte:
CGCTGCGGCCGATCT/tagcca/GACGAGCGGGTTTCGG (SEQ ID NO: 12)

Tabla 2. ARNg de muestra para usar en la transformación de marcadores de resistencia a la neomicina y la higromicina

SEQ ID NO		ARNg
13	Neo Crispr # 1	UGCGCAAGGAACGCCCCGUCG
14	Neo Crispr # 2	GGCAGCGCGGCUAUCGUGGC
15	Neo Crispr # 3	ACGAGGCAGCGCGGCUAUCG
16	Neo Crispr # 4	GCUCUGAUGCCGCCGUGUUC
17	Hyg Crispr # 1	ACGAGCGGGUUCGGCCCAUU
18	Hyg Crispr # 6	CUUAGCCAGACGAGCGGGUU
19	Hyg Crispr # 10	GCCGAUCUUAGCCAGACGAG
20	Hyg Crispr # 16	CGACCUGAUGCAGCUCUCGG

Todas las solicitudes de patente y publicaciones mencionadas en esta descripción son indicativas del nivel de los expertos en la técnica a la cual pertenece la invención. A menos que sea evidente lo contrario del contexto de cualquier realización, aspecto, etapa o característica de la invención, se puede utilizar en combinación con cualquier otro. La referencia a un intervalo incluye cualquier número entero dentro del intervalo, cualquier subintervalo dentro del intervalo. La referencia a múltiples intervalos incluye compuestos de tales intervalos.

5

REIVINDICACIONES

1. Un método para la modificación en serie de un locus objetivo en una célula, que comprende:

(a) proporcionar la célula que comprende el locus objetivo, en donde el locus objetivo comprende un primer polinucleótido que codifica un primer marcador de selección unido operativamente a un primer promotor y un primer sitio de reconocimiento de nucleasa para un primer agente de nucleasa, en donde el primer sitio de reconocimiento de nucleasa se encuentra en una región codificante del primer marcador de selección o en cualquier región no codificante de proteína del primer marcador de selección, y opcionalmente en donde el primer polinucleótido que codifica el primer marcador de selección está flanqueado por el primer sitio objetivo y el segundo sitio objetivo;

(b) introducir en la célula:

(i) el primer agente de nucleasa o un polinucleótido que codifica el primer agente de nucleasa, en donde el primer agente de nucleasa induce un corte o una ruptura de doble cadena en el primer sitio de reconocimiento de nucleasa; y

(ii) un primer vector de transformación que comprende un primer polinucleótido de inserción flanqueado por un primer brazo de homología que corresponde a un primer sitio objetivo ubicado en el locus objetivo y un segundo brazo de homología que corresponde a un segundo sitio objetivo ubicado en el locus objetivo, en donde el primer polinucleótido de inserción comprende un segundo polinucleótido que codifica un segundo marcador de selección unido operativamente a un segundo promotor y un segundo sitio de reconocimiento de nucleasa para un segundo agente de nucleasa, en donde el segundo sitio de reconocimiento de nucleasa se encuentra dentro de una región codificante del segundo marcador de selección o en cualquier región no codificante de proteínas del segundo marcador de selección,

en donde el primer marcador de selección y el segundo marcador de selección son diferentes,

en donde el primer agente de nucleasa es diferente del segundo agente de nucleasa, y

(c) identificar una célula modificada que comprende el primer polinucleótido de inserción en el locus objetivo, en donde la célula modificada tiene la actividad del segundo marcador de selección, pero no tiene la actividad del primer marcador de selección, opcionalmente en donde la identificación se lleva a cabo mediante un ensayo de modificación de alelos (MOA).

2. El método de la reivindicación 1, que comprende además una o más rondas adicionales de integración de polinucleótidos de inserción, en donde los vectores de transformación empleados en el método alternan el uso del primer marcador de selección que comprende el primer sitio de reconocimiento de nucleasa para el primer agente de nucleasa y el segundo marcador de selección que comprende el segundo sitio de reconocimiento de nucleasa para el segundo agente de nucleasa.

3. El método de la reivindicación 1 o 2, en donde la etapa de identificación (c) comprende:

(i) cultivar la célula bajo condiciones que permitan la identificación de células que no tienen la actividad del primer marcador de selección; o

(ii) identificar al menos una célula que comprende el primer polinucleótido de inserción integrado en el primer y segundo sitios objetivo.

4. El método de cualquier reivindicación anterior, en donde:

(I) el primer marcador de selección o el segundo marcador de selección confiere resistencia a un antibiótico,

opcionalmente en donde el antibiótico comprende G418, higromicina, blasticidina, neomicina o puromicina, y

(II) el primer marcador de selección o el segundo marcador de selección está operativamente unido a un promotor inducible, y la expresión del marcador de selección es tóxica para la célula,

opcionalmente en donde el primer marcador de selección o el segundo marcador de selección comprende hipoxantina-guanina fosforribosiltransferasa (HGPRT) o timidina quinasa del virus del herpes simple (HSV-TK).

5. El método de cualquier reivindicación anterior, en donde el uso combinado del primer vector de transformación con el primer agente de nucleasa resulta en una mayor eficiencia de transformación en comparación con el uso del primer vector de transformación solo,

opcionalmente, en donde la eficiencia de transformación del primer vector de transformación se incrementa al menos dos veces en comparación con el uso del primer vector de transformación solo.

- 5 6. El método de cualquier reivindicación anterior, en donde el primer sitio de reconocimiento de nucleasa se encuentra en un intrón, un exón, un promotor, una región reguladora del promotor o una región potenciadora del primer marcador de selección, y el segundo sitio de reconocimiento de nucleasa se encuentra en un intrón, un exón, un promotor, una región reguladora del promotor o una región potenciadora del segundo marcador de selección.
7. El método de cualquiera de las reivindicaciones anteriores, en donde la etapa (b) comprende la introducción en la célula del polinucleótido que codifica el primer agente de nucleasa, y en donde:
- 10 (I) el polinucleótido que codifica el primer agente de nucleasa comprende una construcción de expresión que comprende una secuencia de ácido nucleico que codifica primero el agente de nucleasa, en donde la secuencia de ácido nucleico que codifica el agente de nucleasa está operativamente unida a un promotor activo en la célula; o
- (II) el polinucleótido que codifica la primera nucleasa comprende un ARNm que codifica el agente de nucleasa.
8. El método de cualquiera de las reivindicaciones anteriores, en donde el primer agente de nucleasa o el segundo agente de nucleasa es:
- 15 (I) una nucleasa de dedo de zinc (ZFN);
- (II) una nucleasa efectora del tipo activador de la transcripción (TALEN);
- (III) una meganucleasa; o
- (IV) una proteína asociada a repeticiones palindrómicas cortas intercaladas regularmente y agrupadas (CRISPR) (Cas) y un ARN guía (ARNg).
- 20 9. El método de la reivindicación 8, en donde el primer agente de nucleasa o el segundo agente de nucleasa es la nucleasa de dedo de zinc (ZFN), y el primer sitio de reconocimiento de nucleasa o el segundo sitio de reconocimiento de nucleasa comprende una cualquiera de las SEQ ID NOS: 9-12.
10. El método de la reivindicación 8, en donde el primer agente de nucleasa o el segundo agente de nucleasa es la proteína Cas y el ARNg, en donde la proteína Cas es Cas9, y en donde el ARNg comprende:
- 25 (a) un ARN de repeticiones palindrómicas cortas intercaladas regularmente y agrupadas (CRISPR) (ARNcr) que se dirige al primer sitio de reconocimiento de nucleasa o al segundo sitio de reconocimiento de nucleasa, en donde el primer sitio de reconocimiento de nucleasa o el segundo sitio de reconocimiento de nucleasa está inmediatamente flanqueado por una secuencia de un motivo adyacente a protoespaciador (PAM); y
- (b) un ARN CRISPR transactivador (tracrRNA),
- 30 opcionalmente en donde el locus objetivo comprende la secuencia de nucleótidos de SEQ ID NO: 1, el ARNg comprende SEQ ID NO: 2, 3, 4, 5, 6, 7 u 8, o el ARNg comprende una cualquiera de las SEQ ID NOS: 13-20, y
- opcionalmente en donde el primer agente de nucleasa comprende la proteína Cas y un primer ARNg, el segundo agente de nucleasa comprende la proteína Cas y un segundo ARNg, y el primer ARNg y el segundo ARNg son diferentes.
- 35 11. El método de la reivindicación 10, en donde el primer marcador de selección confiere resistencia a la neomicina, y el primer agente de nucleasa es una proteína Cas9 y un ARN guía que comprende una cualquiera de las SEQ ID NOS: 13-16, y en donde el segundo marcador de selección confiere resistencia a la higromicina, y el segundo agente de nucleasa es una proteína Cas9 y un ARN guía que comprende una cualquiera de SEQ ID NOS: 17-20,
- 40 opcionalmente, en donde el primer marcador de selección confiere resistencia a la neomicina, y el primer agente de nucleasa es la proteína Cas9 y el ARN guía que comprende la SEQ ID NO: 13, y en donde el segundo marcador de selección confiere resistencia a la higromicina, y el segundo agente de nucleasa es la proteína Cas9 y el ARN guía que comprende la SEQ ID NO: 17.
- 45 12. El método de cualquier reivindicación anterior, en donde:
- (I) el primer sitio objetivo y el segundo sitio objetivo son inmediatamente adyacentes al primer sitio de reconocimiento de la nucleasa; o
- (II) el primer sitio objetivo y el segundo sitio objetivo son aproximadamente de 10 nucleótidos a aproximadamente 14 kb del primer sitio de reconocimiento de la nucleasa.
- 50 13. El método de cualquier reivindicación anterior, en donde:

- (I) la suma total del primer brazo de homología y el segundo brazo de homología es de al menos aproximadamente 10 kb; y/o
- (II) cada uno del primer brazo de homología y el segundo brazo de homología varía de aproximadamente 5 kb a aproximadamente 100 kb; y/o
- 5 (III) el primer vector de transformación es de al menos aproximadamente 10 kb o de aproximadamente 20 kb a aproximadamente 300 kb; y/o
- (IV) el primer polinucleótido de inserción varía de aproximadamente 5 kb a aproximadamente 300 kb en longitud.
- 10 14. El método de cualquier reivindicación anterior, en donde el primer polinucleótido de inserción comprende además un primer polinucleótido de interés, en donde:
- (I) el primer polinucleótido de interés comprende un polinucleótido humano;
- (II) el primer polinucleótido de interés comprende un polinucleótido que codifica una región de un receptor de células T,
- opcionalmente en donde el receptor de células T es un receptor alfa de células T, y
- 15 opcionalmente en donde el primer polinucleótido de interés comprende al menos un segmento génico de la región variable y/o un segmento génico de la región de unión del locus del receptor alfa de células T;
- (III) el primer polinucleótido de interés comprende una secuencia de ácido nucleico genómico que codifica una secuencia de aminoácidos de la región variable de la cadena pesada de inmunoglobulina humana, opcionalmente en donde el primer polinucleótido de interés comprende una secuencia de ácido nucleico de la
- 20 región variable de la cadena pesada de inmunoglobulina humana no reordenada, operativamente unida a una secuencia de ácido nucleico de la región constante de la cadena pesada de inmunoglobulina no humana;
- (IV) el primer polinucleótido de interés comprende una secuencia de ácido nucleico genómico que codifica una secuencia de aminoácidos de la región variable de la cadena ligera de inmunoglobulina humana,
- opcionalmente, en donde la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de la región variable de la cadena ligera λ y/o κ humana no reordenada, o en donde la secuencia de ácido nucleico genómico comprende una secuencia de ácido nucleico de la región variable de la cadena ligera λ y/o κ humana reordenada;
- 25 (V) el primer polinucleótido de interés comprende al menos un alelo de una enfermedad; o
- (VI) el primer polinucleótido de interés comprende una secuencia de ácido nucleico exógena o una secuencia de ácido nucleico que es homóloga u ortóloga a una secuencia de ácido nucleico en el genoma de la célula.
- 30 15. El método de cualquier reivindicación anterior, en donde el primer vector de transformación está diseñado para eliminar una secuencia de aproximadamente 5 kb a aproximadamente 3 Mb del locus objetivo.
16. El método de cualquier reivindicación anterior, en donde la integración del primer polinucleótido de inserción en el locus 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 estos.
- 35 17. El método de cualquier reivindicación anterior, que comprende la identificación de una célula modificada que comprende una modificación genética que consiste en la sustitución de una secuencia de ácido nucleico endógena con un polinucleótido exógeno de interés en el locus objetivo.
- 40 18. El método de cualquier reivindicación anterior, en donde el locus objetivo está en el genoma de la célula o se encuentra en un cromosoma artificial de levadura (YAC), un cromosoma artificial bacteriano (BAC) o un cromosoma artificial humano en la célula, opcionalmente en donde el locus objetivo comprende un locus de inmunoglobulina o un locus de receptor de células T, opcionalmente en donde el locus del receptor de células T es un locus de receptor alfa de células T.
- 45 19. El método de cualquier reivindicación anterior, en donde el corte o la ruptura de doble cadena inducida por el primer agente de nucleasa interrumpe la actividad del primer marcador de selección.
20. El método de cualquier reivindicación anterior,
- en donde la inserción del primer polinucleótido de inserción en el locus objetivo interrumpe la actividad del primer marcador de selección.

21. El método de cualquier reivindicación anterior, en donde la célula es una célula eucariota, opcionalmente en donde la célula es una célula de mamífero, opcionalmente en donde la célula es:

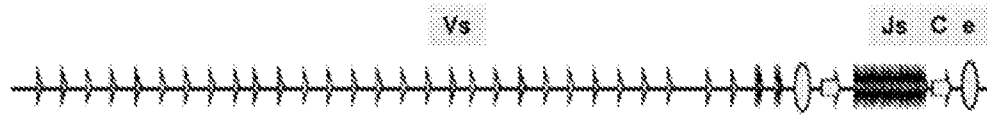
- (I) una célula de mamífero no humano;
- (II) una célula pluripotente;
- 5 (III) una célula madre pluripotente inducida humana;
- (IV) un fibroblasto humano;
- (V) una célula madre embrionaria (ES) no humana;
- (VI) una célula madre embrionaria (ES) de ratón o una célula madre embrionaria (ES) de rata;
- (VII) una célula madre hematopoyética;
- 10 (VIII) una célula madre neuronal; o
- (IX) una célula de roedor.

22. El método de la reivindicación 21,

en donde la célula es la célula ES de ratón o la célula ES de rata.

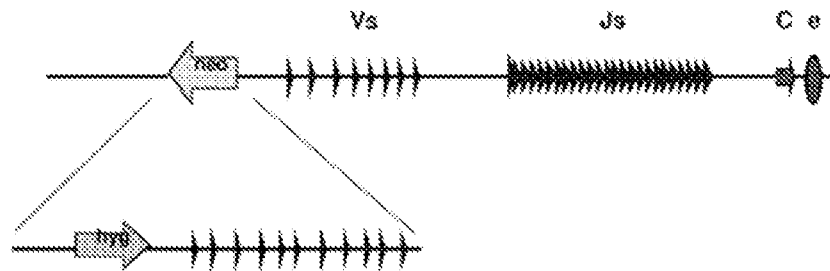
Figura 1

Locus del receptor alfa de células T de ratón (cr. 14a)



Alelo de TCR α humanizado A-neo (cr. 14b)

8 Vs y 61 Js

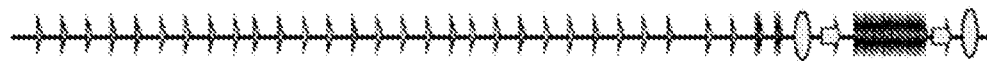


Vector de transformación del alelo de TCR α humanizado B-hyg

11 Vs adicionales

↓ Electroporación + neo ZFN

Locus del receptor alfa de células T de ratón (cr. 14a)



Alelo de TCR α humanizado B-hyg (cr. 14b)

19 Vs y 61 Js

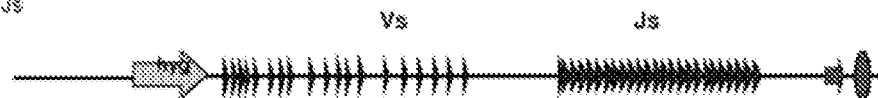


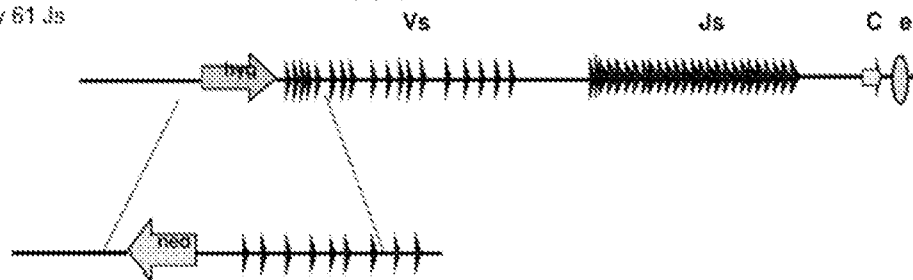
Figura 2

Locus del receptor alfa de células T de ratón (cr. 14a)



Alelo de TCRα humanizado B-hyg (cr. 14b)

19 Vs y 61 Js

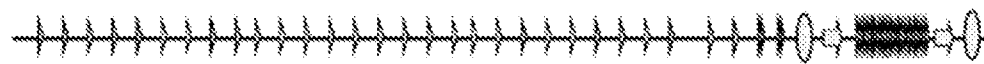


Vector de transformación del alelo de TCRα humanizado C-neo

11 Vs adicionales

↓ Electroporación + hyg ZFN

Locus del receptor alfa de células T de ratón (cr. 14a)



Alelo de TCRα humanizado C-neo (cr. 14b)

30 Vs y 61 Js

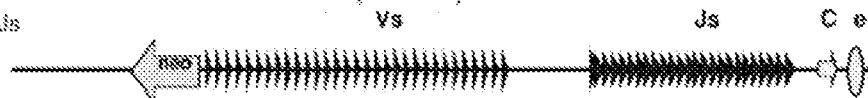


Figura 3

