

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 694 629**

51 Int. Cl.:

C12N 5/0735 (2010.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **26.06.2015 PCT/US2015/038001**

87 Fecha y número de publicación internacional: **30.12.2015 WO15200805**

96 Fecha de presentación y número de la solicitud europea: **26.06.2015 E 15745288 (9)**

97 Fecha y número de publicación de la concesión europea: **26.09.2018 EP 3161128**

54 Título: **Métodos y composiciones para modificaciones genéticas objetivo y métodos de uso**

30 Prioridad:

26.06.2014 US 201462017582 P
26.06.2014 US 201462017627 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
26.12.2018

73 Titular/es:

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

72 Inventor/es:

FRENDEWEY, DAVID;
DROGUETT, GUSTAVO;
GAGLIARDI, ANTHONY;
KUNO, JUNKO;
AUERBACH, WOJTEK y
VALENZUELA, DAVID M.

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 694 629 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 modificaciones genéticas objetivo y métodos de uso

Referencia a un listado de secuencias presentada como un archivo de texto a través de la red del EFS

5 La copia oficial del listado de secuencias se envía electrónicamente a través de la red del EFS como un listado de secuencias en formato ASCII con un archivo llamado 463545SEQLIST.TXT, creado el 25 de junio de 2015 y con un tamaño de 14 kilobytes, y se archiva simultáneamente con la especificación. El listado de secuencias contenido en este documento con formato ASCII es parte de la memoria descriptiva.

Campo de la invención

10 La invención se refiere a los métodos y composiciones para mantener o cultivar células pluripotentes y/o totipotentes y a métodos y composiciones para generar poblaciones de células y animales transgénicos.

Antecedentes

15 Tal vez debido a las características estructurales únicas del cromosoma Y, las estrategias convencionales de reconocimiento génico en células madre embrionarias de ratón para generar mutaciones en los genes ligados a Y han tenido un éxito limitado. Por lo tanto, a menudo, la comprensión de las funciones de los genes murinos ligados a Y se limita a la información obtenida de los estudios de ratones que portan supresiones espontáneas, inserciones aleatorias de trampas génicas o transgenes autosómicos. Se necesitan métodos para mejorar la capacidad de apuntar a un locus genómico en el cromosoma Y.

20 La proteína Sry (región Y determinante del sexo) es el regulador clave de la determinación del sexo masculino en los mamíferos placentarios. El gen *Sry*, también conocido como el factor determinante del testículo (TDF), reside en el cromosoma Y. Se cree que *Sry* es un factor de transcripción que se une al ADN a través de su dominio de grupo de alta movilidad (HMG). La expresión del gen *Sry* de ratón se restringe a la cresta genital en una ventana de tiempo estrecha alrededor del día 11 del desarrollo embrionario; se detectan tanto el ARNm como la proteína Sry. Se debe elaborar suficiente Sry dentro de esta ventana de tiempo para convertir la cresta genital bipotencial hacia el programa de formación de testículos masculinos al tiempo que se inhibe el programa hembra de desarrollo ovárico.
25 En los testículos adultos se detecta un transcrito circular pero no la proteína Sry. Las mutaciones en el gen *Sry* que causan la producción de una proteína Sry inactiva o que alteran el tiempo y la fuerza de la expresión génica pueden provocar la inversión sexual de hombre a mujer, lo que da como resultado animales que tienen un cromosoma X y un Y, pero que son anatómicamente hembras. Las llamadas hembras XY a menudo son estériles o tienen una baja fertilidad. Ser capaz de controlar la determinación del sexo mediante la regulación del gen *Sry* tendría un gran valor
30 en la producción de animales modificados genéticamente.

WANG et al, "TALEN-mediated editing of the mouse Y chromosome," Nat. Biotech. 31 (6): 530-532, (2013) describe un método para modificar un locus genómico objetivo en el cromosoma Y en una célula.

Resumen

La presente invención es como se define en las reivindicaciones.

35 La presente invención se refiere a un método para modificar un locus genómico objetivo en un cromosoma Y en una célula, que comprende:

(a) proporcionar la célula que comprende el locus genómico objetivo en el cromosoma Y, en donde el locus genómico objetivo comprende un sitio de reconocimiento para un agente de nucleasa, y en el que la célula está en un cultivo que comprende un medio base DMEM;

40 (b) la introducción en la célula:

(i) del agente de nucleasa, en el que el agente de nucleasa induce un corte o ruptura de la cadena doble en el sitio de reconocimiento; y

45 (ii) de un vector de direccionamiento grande que comprende un polinucleótido de inserción flanqueado por los brazos de homología primero y segundo correspondientes a los sitios objetivo primero y segundo ubicados dentro del locus genómico objetivo, en donde la suma total del primer brazo de homología y el segundo brazo de homología es al menos de 10 kb, y en el que el vector de direccionamiento experimenta recombinación homóloga con el locus genómico objetivo; e

50 (c) identificar al menos una célula que comprende en su genoma el polinucleótido de inserción integrado en el locus genómico objetivo, en donde la integración del polinucleótido de inserción introduce una modificación genética que comprende la eliminación de una secuencia endógena de ácido nucleico y la sustitución con una secuencia exógena de ácido nucleico en el locus genómico objetivo,

en el que la célula no se produce mediante un proceso que implica modificar la identidad genética de la línea germinal de los seres humanos o que implica el uso de un embrión humano para fines industriales o comerciales, y

en el que el método no es un método para el tratamiento del cuerpo humano o animal mediante terapia.

5 Esta divulgación proporciona un método para elaborar una línea de células madre embrionarias XY (ES) capaz de producir un mamífero hembra XY fértil no humano en una generación F0. El método comprende: (a) modificar una célula madre (ES) embrionaria XY de mamífero no humano para que tenga una modificación que disminuya el nivel y/o la actividad de una proteína Sry; y (b) cultivar la línea de células ES modificada en condiciones que permitan elaborar una línea de células ES capaz de producir un mamífero hembra XY fértil no humano en una generación F0.

10 Esta divulgación también proporciona un método para preparar un mamífero hembra XY fértil no humano en una generación F0. El método comprende: (a) introducir la célula ES XY de mamífero no humano elaborada mediante el método anterior que tiene una modificación que disminuye el nivel y/o la actividad de una proteína Sry en un embrión huésped; (b) gestar el embrión huésped; y, (c) obtener un mamífero hembra no humano XY F0, en donde, al alcanzar la madurez sexual, el mamífero hembra no humano F0 XY es fértil. En una realización de esta divulgación, el mamífero hembra no humano XY F0 es fértil cuando se cruza con un ratón de tipo silvestre. En realizaciones específicas, el ratón de tipo silvestre es C57BL/6. En otra realización, el roedor es una rata o un hámster.

15 En algunas realizaciones, el nivel y/o la actividad disminuidos de la proteína Sry resultan de una modificación genética en el gen Sry. En algunos de estos métodos, la modificación genética en el gen Sry comprende una inserción de uno o más nucleótidos, una eliminación de uno o más nucleótidos, una sustitución de uno o más nucleótidos, una inactivación, una activación, un reemplazo de una secuencia endógena de ácido nucleico con una secuencia homóloga, heteróloga u ortóloga de ácido nucleico, o una combinación de las mismas.

20 En los métodos proporcionados en el presente documento, la modificación genética escogida puede comprender una inserción, una eliminación, una inactivación, un activación, una mutación puntual, o una combinación de los mismos. En otra realización, la modificación genética escogida se encuentra en un autosoma.

25 En algunas realizaciones, la modificación del gen Sry comprende una inserción de un marcador seleccionable y/o un gen informador operativamente unido a un promotor activo en la célula ES de mamífero no humano. En algunas realizaciones, la modificación del gen Sry comprende una inserción de un gen informador operativamente enlazado al promotor endógeno de Sry. En una realización específica, el gen informador codifica la proteína informadora LacZ.

30 En una realización, la etapa de cultivo comprende cultivar la célula ES XY de mamífero no humano en un medio que comprende un medio base y suplementos adecuados para mantener la célula ES de mamífero no humano en cultivo, en donde el medio es un medio de baja osmolalidad. En una realización, el medio de baja osmolalidad exhibe una osmolalidad de aproximadamente 200 mOsm/kg a menos de aproximadamente 329 mOsm/kg. En otras realizaciones, el medio de baja osmolalidad exhibe una o más de las siguientes características: una conductividad de aproximadamente 11 mS/cm a aproximadamente 13 mS/cm; una sal de un metal alcalino y un haluro en una concentración de aproximadamente 50 mM a aproximadamente 110 mM; una concentración de sal de ácido carbónico de aproximadamente 17 mM a aproximadamente 30 mM; una concentración total de sal de haluro de metal alcalino y de sal de ácido carbónico de aproximadamente 85 mM a aproximadamente 130 mM; y/o una combinación de dos o más de los mismos.

35 En algunas realizaciones, tras la introducción de células ES XY de mamífero no humano en un embrión huésped y después de la gestación del embrión huésped, al menos de 80%, al menos de 85%, al menos de 90%, o al menos de 95% de los mamíferos no humanos F0 son hembras XY que, al alcanzar la madurez sexual, el mamífero no humano hembra F0 XY es fértil.

40 En una realización, la célula ES XY de mamífero no humano comprende un locus genómico objetivo en el cromosoma Y que comprende un sitio de reconocimiento para un agente de nucleasa, y en el que el agente de nucleasa induce un corte o rompimiento de la cadena doble en el sitio de reconocimiento. Dicho método puede comprender además exponer la célula ES al agente de nucleasa en presencia de un vector de direccionamiento que comprende un polinucleótido de inserción, en donde después de la exposición al agente de nucleasa y el vector de direccionamiento, la célula ES se modifica para contener el polinucleótido de inserción. En una realización, el agente de nucleasa es un ARNm que codifica una nucleasa. En realizaciones específicas, el agente de nucleasa es (a) una nucleasa de dedo de cinc (ZFN); (b) es una nucleasa efectora de tipo activador de transcripción (TALEN); o (c) una meganucleasa. En otras realizaciones, el agente de nucleasa comprende una proteína (Cas) asociada a repeticiones palindrómicas cortas agrupadas y regularmente espaciadas (CRISPR) y un ARN guía (ARNg). En tales métodos, el ARN guía (ARNg) comprende (a) un ARN de repeticiones palindrómicas cortas agrupadas y regularmente espaciadas (CRISPR) (ARNcr) que se dirige al primer sitio de reconocimiento; y (b) un ARN de CRISPR transactivador (ARNcrtra). En algunos casos, el sitio de reconocimiento está inmediatamente flanqueado por una secuencia del motivo adyacente protoespaciador (PAM). En una realización, la proteína Cas es Cas9.

55 En esta descripción también se proporciona un cultivo *in vitro* que comprende la línea de células ES XY de mamífero no humano de acuerdo con cualquiera de los métodos proporcionados en el presente documento.

Se proporciona un cultivo *in vitro* mediante esta descripción y comprende (a) una célula madre (ES) embrionaria XY de mamífero no humano que tiene una modificación que disminuye el nivel y/o la actividad de una proteína Sry; y, (b) un medio que comprende un medio base y suplementos adecuados para mantener la célula ES de mamífero no humano en cultivo. En una realización, el medio base muestra una osmolalidad de aproximadamente 200 mOsm/kg hasta menos de aproximadamente 329 mOsm/kg. En otras realizaciones, el medio base exhibe una o más de las siguientes características: una conductividad de aproximadamente 11 mS/cm a aproximadamente 13 mS/cm; una sal de un metal alcalino y un haluro en una concentración de aproximadamente 50 mM a aproximadamente 110 mM; una concentración de sal de ácido carbónico de aproximadamente 17 mM a aproximadamente 30 mM; una concentración total de sal de haluro de metal alcalino y de sal de ácido carbónico de aproximadamente 85 mM a aproximadamente 130 mM; y/o una combinación de dos o más de los mismos. En una realización, la célula ES XY de mamífero no humano es de un roedor. En una realización, el roedor es un ratón o una rata. En una realización, la célula ES XY de ratón es una célula ES de ratón VGF1. En una realización, el roedor es una rata o un hámster. En una realización, el nivel y/o la actividad disminuidos de la proteína Sry provienen de una modificación genética en el gen Sry. En una realización, la modificación genética en el gen Sry comprende una inserción de uno o más nucleótidos, una eliminación de uno o más nucleótidos, una sustitución de uno o más nucleótidos, una inactivación, una activación, un reemplazo de una secuencia endógena de ácido nucleico con un secuencia heteróloga de ácido nucleico o una combinación de las mismas. En una realización, la célula ES de mamífero no humano comprende una, dos, tres o más modificaciones genéticas escogidas. En una realización, la modificación genética escogida comprende una inserción, una eliminación, una inactivación, una activación, una mutación puntual, o una combinación de los mismos. En una realización, la modificación genética escogida comprende al menos una inserción de un polinucleótido heterólogo en el genoma de la célula ES XY. En una realización, la modificación genética escogida se encuentra en un autosoma. En una realización, el medio base exhibe 50 ± 5 mM de NaCl, 26 ± 5 mM de carbonato y 218 ± 22 mOsm/kg. En una realización, el medio base exhibe aproximadamente 3 mg/mL de NaCl, 2,2 mg/mL de bicarbonato de sodio y 218 mOsm/kg. En una realización, el medio base exhibe 87 ± 5 mM de NaCl, 18 ± 5 mM de carbonato y 261 ± 26 mOsm/kg. En una realización, el medio base exhibe aproximadamente 5,1 mg/mL de NaCl, 1,5 mg/mL de bicarbonato de sodio y 261 mOsm/kg. En una realización, el medio base exhibe 110 ± 5 mM de NaCl, 18 ± 5 mM de carbonato y 294 ± 29 mOsm/kg. En una realización, el medio base exhibe aproximadamente 6,4 mg/mL de NaCl, 1,5 mg/mL de bicarbonato de sodio y 294 mOsm/kg. En una realización, el medio base exhibe 87 ± 5 mM de NaCl, 26 ± 5 mM de carbonato y 270 ± 27 mOsm/kg. En una realización, el medio base exhibe aproximadamente 5,1 mg/mL de NaCl, 2,2 mg/mL de bicarbonato de sodio y 270 mOsm/kg. En una realización, el medio base exhibe 87 ± 5 mM de NaCl, 26 ± 5 mM de carbonato, 86 ± 5 mM de glucosa y 322 ± 32 mOsm/kg. En una realización, el medio base exhibe aproximadamente 5,1 mg/mL de NaCl, 2,2 mg/mL de bicarbonato de sodio, 15,5 mg/mL de glucosa y 322 mOsm/kg. En una realización, tras la introducción de las células ES XY de mamífero no humano en un embrión huésped y después de la gestación del embrión huésped, al menos de 80% de los mamíferos no humanos F0 son hembras XY que al alcanzar la madurez sexual, el mamífero hembra F0 XY no humano es fértil.

Además, esta descripción proporciona un método para elaborar un mamífero hembra XY fértil no humano en una generación F0, que comprende: (a) cultivar una célula madre (ES) embrionaria XY de mamífero no humano donante que tiene una modificación que disminuye el nivel y/o la actividad de una proteína Sry en un medio que comprende un medio base y suplementos adecuados para mantener la célula ES de mamífero no humano en cultivo, (b) la introducción de la célula ES de mamífero no humano XY donante en un embrión huésped; (c) gestar el embrión huésped; y, (d) obtener un mamífero hembra no humano XY F0, en donde, al alcanzar la madurez sexual, el mamífero hembra no humano XY F0 es fértil. En una realización, el medio exhibe una osmolalidad de aproximadamente 200 mOsm/kg a menos de aproximadamente 329 mOsm/kg. En otras realizaciones, el medio exhibe una característica que comprende uno o más de los siguientes: una conductividad de aproximadamente 11 mS/cm a aproximadamente 13 mS/cm; una sal de un metal alcalino y un haluro en una concentración de aproximadamente 50 mM a aproximadamente 110 mM; una concentración de sal de ácido carbónico de aproximadamente 17 mM a aproximadamente 30 mM; una concentración total de sal de haluro de metal alcalino y de sal de ácido carbónico de aproximadamente 85 mM a aproximadamente 130 mM; y/o una combinación de dos o más de los mismos; en una realización, la célula ES XY de mamífero no humano es de un roedor. En una realización, el roedor es un ratón o una rata. En una realización, la célula ES XY de ratón es una célula ES de ratón VGF1. En una realización, el roedor es una rata o un hámster. En una realización, el nivel y/o la actividad disminuidos de la proteína Sry provienen de una modificación genética en el gen Sry. En una realización, la modificación genética en el gen Sry comprende una inserción de uno o más nucleótidos, una eliminación de uno o más nucleótidos, una sustitución de uno o más nucleótidos, una inactivación, una activación, un reemplazo de una secuencia endógena de ácido nucleico con un secuencia heteróloga de ácido nucleico o una combinación de las mismas. En una realización, la célula ES de mamífero no humano comprende una, dos, tres o más modificaciones genéticas escogidas. En una realización, la modificación genética escogida comprende una inserción, una eliminación, una inactivación, una activación, una mutación puntual, o una combinación de los mismos. En una realización, la modificación genética escogida comprende al menos una inserción de un polinucleótido heterólogo en un genoma de la célula ES XY. En una realización, la modificación genética escogida se encuentra en un autosoma. En una realización, el medio base exhibe 50 ± 5 mM de NaCl, 26 ± 5 mM de carbonato y 218 ± 22 mOsm/kg. En una realización, el medio base exhibe aproximadamente 3 mg/mL de NaCl, 2,2 mg/mL de bicarbonato de sodio y 218 mOsm/kg. En una realización, el medio base exhibe 87 ± 5 mM de NaCl, 18 ± 5 mM de carbonato y 261 ± 26 mOsm/kg. En una realización, el medio base muestra aproximadamente 5,1 mg/mL de NaCl, 1,5 mg/mL de

bicarbonato de sodio y 261 mOsm/kg. En una realización, el medio base exhibe 110 ± 5 mM de NaCl, 18 ± 5 mM de carbonato y 294 ± 29 mOsm/kg. En una realización, el medio base muestra aproximadamente 6,4 mg/mL de NaCl, 1,5 mg/mL de bicarbonato de sodio y 294 mOsm/kg. En una realización, el medio base exhibe 87 ± 5 mM de NaCl, 26 ± 5 mM de carbonato y 270 ± 27 mOsm/kg. En una realización, el medio base exhibe aproximadamente 5,1 mg/mL de NaCl, 2,2 mg/mL de bicarbonato de sodio y 270 mOsm/kg. En una realización, en la que el medio base exhibe 87 ± 5 mM de NaCl, 26 ± 5 mM de carbonato, 86 ± 5 mM de glucosa y 322 ± 32 mOsm/kg. En una realización, en la que el medio base exhibe aproximadamente 5,1 mg/mL de NaCl, 2,2 mg/mL de bicarbonato de sodio, 15,5 mg/mL de glucosa y 322 mOsm/kg.

Se proporcionan además en esta descripción métodos para producir un homocigoto de mamífero no humano transgénico para una mutación genética escogida en la generación F1 que comprende: (a) cruzar una hembra fértil XY F0 que tiene un nivel y/o actividad disminuidos de la proteína Sry con un hermano clonal de cohorte, derivado del mismo clon de células ES, mamífero no humano macho XY F0, en el que el mamífero hembra no humano fértil XY F0 y el mamífero no humano macho XY F0 son heterocigotos para la mutación genética; y, (b) obtener un ratón de progenie F1 que es homocigoto para la modificación genética.

La presente divulgación también proporciona un método para modificar un locus genómico objetivo en el cromosoma Y en una célula y comprende (a) proporcionar una célula que comprende un locus genómico objetivo en el cromosoma Y que comprende un sitio de reconocimiento para un agente de nucleasa, (b) introducir en la célula (i) el agente de nucleasa, en el que el agente de nucleasa induce un corte o rompimiento de la cadena doble en el primer sitio de reconocimiento; y, (ii) un primer vector de direccionamiento que comprende un primer polinucleótido de inserción flanqueado por un primer y un segundo brazo de homología que corresponde a un primer y un segundo sitio objetivo situados en proximidad suficiente al primer sitio de reconocimiento; y, (c) identificar al menos una célula que comprende en su genoma el primer polinucleótido de inserción integrado en el locus genómico objetivo. Esta divulgación establece que la suma total del primer brazo de homología y el segundo brazo de homología es de al menos 4 kb pero menos de 150 kb. Esta descripción establece que la longitud del primer brazo de homología y/o el segundo brazo de homología es de al menos 400 pb pero menor que 1000 pb. Esta divulgación establece que la longitud del primer brazo de homología y/o el segundo brazo de homología es de aproximadamente 700 pb a aproximadamente 800 pb.

Se proporciona además en este descubrimiento un método para modificar un locus genómico objetivo en el cromosoma Y en una célula y comprende: (a) proporcionar una célula que comprende un locus genómico objetivo en el cromosoma Y que comprende un sitio de reconocimiento para un agente de nucleasa, (b) introducir en la célula un primer vector de direccionamiento que comprende un primer polinucleótido de inserción flanqueado por un primer y un segundo brazo de homología que corresponde a un primer y un segundo sitio objetivo; y, (c) identificar al menos una célula que comprende en su genoma el primer polinucleótido de inserción integrado en el locus genómico objetivo. Esta divulgación establece que la longitud del primer brazo de homología y/o el segundo brazo de homología es de al menos 400 pb pero menor que 1.000 pb. Esta divulgación establece que la longitud del primer brazo de homología y/o el segundo brazo de homología es de aproximadamente 700 pb a aproximadamente 800 pb. En una realización, la célula es una célula de mamífero. En una realización, la célula de mamífero es una célula no humana. 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 es una célula pluripotente. En una realización, la célula de mamífero es una célula madre pluripotente inducida (iPS). En una realización, la célula pluripotente es una célula madre (ES) embrionaria no humana. En una realización, la célula pluripotente es una célula madre (ES) embrionaria de roedor, una célula madre (ES) embrionaria de ratón o una célula madre (ES) embrionaria de rata. En una realización, el agente de nucleasa es un ARNm que codifica una nucleasa. En una realización, el agente de nucleasa es una nucleasa de dedo de cinc (ZFN). En una realización, el agente de nucleasa es una nucleasa efectora similar a un activador de transcripción (TALEN). En una realización, el agente de nucleasa es una meganucleasa. En algunas realizaciones, el agente de nucleasa comprende una proteína (Cas) asociada a repeticiones palindrómicas cortas agrupadas y regularmente espaciadas (CRISPR) y un ARN guía (ARNg). En tal método, el ARN guía (ARNg) puede comprender (a) un ARN de repeticiones palindrómicas cortas agrupadas y regularmente espaciadas (CRISPR) (ARNcr) que se dirige al primer sitio de reconocimiento; y (b) un ARN de CRISPR transactivador (ARNcrtra). En una realización, el primero o segundo sitios de reconocimiento están inmediatamente flanqueados por una secuencia del motivo adyacente protoespaciador (PAM). En algunas realizaciones, la proteína Cas es Cas9.

En algunas realizaciones, la modificación comprende una eliminación de una secuencia endógena de ácido nucleico. En algunas realizaciones, la eliminación 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, 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 una realización específica, la eliminación es de al menos 500 kb. En una realización, la célula es una célula de mamífero. En una realización, la célula de mamífero es una

célula no humana. 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 es una célula pluripotente. En una realización, la célula de mamífero es una célula madre pluripotente inducida (iPS). En una realización, la célula pluripotente es una célula madre (ES) embrionaria no humana. En una realización, la célula pluripotente es una célula madre (ES) embrionaria de roedor, una célula madre (ES) embrionaria de ratón o una célula madre (ES) embrionaria de rata. En algunas realizaciones, el agente de nucleasa comprende una proteína (Cas) asociada a repeticiones palindrómicas cortas agrupadas y regularmente espaciadas (CRISPR) y un ARN guía (ARNg). En tal método, el ARN guía (ARNg) puede comprender (a) un ARN de repeticiones palindrómicas cortas agrupadas y regularmente espaciadas (CRISPR) (ARNcr) que se dirige al primer sitio de reconocimiento; y (b) un ARN de CRISPR transactivador (ARNcrtra). En una realización, el primero o segundo sitios de reconocimiento están inmediatamente flanqueados por una secuencia del motivo adyacente protoespaciador (PAM). En algunas realizaciones, la proteína Cas es Cas9. En una realización, el agente de nucleasa es una nucleasa de dedo de cinc (ZFN). En una realización, el agente de nucleasa es una nucleasa efectora similar a un activador de transcripción (TALEN). En una realización, el agente de nucleasa es una meganucleasa.

Métodos para modificar el cromosoma Y que comprende exponer el cromosoma Y a una proteína Cas y un ARN de CRISPR en presencia de un gran vector de direccionamiento (LTVEC) que comprende una secuencia de ácido nucleico de al menos 10 kb y comprende después de la exposición a la proteína Cas, el ARN de CRISPR, y el LTVEC, el cromosoma Y se modifica para contener al menos una secuencia de ácido nucleico de 10 kb. El LTVEC puede comprender una secuencia de ácido nucleico de al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb o al menos 90 kb. En otras realizaciones, el LTVEC comprende una secuencia de ácido nucleico de al menos 100 kb, al menos 150 kb o al menos 200 kb.

También se proporciona en esta divulgación un método para modificar un locus genómico objetivo en el cromosoma Y, que comprende: (a) proporcionar una célula de mamífero que comprende el locus genómico objetivo en el cromosoma Y, en donde el locus genómico objetivo comprende una secuencia objetivo de ARN guía (ARNg); (b) la introducción en la célula de mamífero: (i) un vector de direccionamiento grande (LTVEC) que comprende un primer ácido nucleico flanqueado con brazos de direccionamiento homólogos al locus genómico objetivo, en el que el LTVEC es de al menos 10 kb; (ii) un primer constructo de expresión que comprende un primer promotor unido operativamente a un segundo ácido nucleico que codifica una proteína Cas, y (iii) un segundo constructo de expresión que comprende un segundo promotor unido operativamente a un tercer ácido nucleico que codifica un ARN guía (ARNg) que comprende una secuencia de nucleótidos que se hibrida con la secuencia objetivo de ARNg y un ARN de CRISPR transactivador (ARNcrtra), en donde el primero y el segundo promotores son activos en la célula de mamífero; y (c) identificar una célula de mamífero modificada que comprende una modificación genética escogida en el locus genómico objetivo en el cromosoma Y. En otras realizaciones, el LTVEC es de al menos 15 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb o al menos 90 kb. En otras realizaciones, el LTVEC es al menos 100 kb, al menos 150 kb, o al menos 200 kb. En una realización, la célula de mamífero es una célula de mamífero no humano. En una realización, la célula de mamífero es una célula de fibroblasto. 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 de mamífero es una célula pluripotente. En una realización, la célula pluripotente es una célula madre pluripotente inducida (iPS). En una realización, la célula pluripotente es una célula madre (ES) embrionaria de ratón o una célula madre (ES) embrionaria de rata. En una realización, la célula pluripotente es una célula progenitora humana de desarrollo restringido. En una realización, la proteína Cas es una proteína Cas9. En una realización, la secuencia objetivo de ARNg está inmediatamente flanqueada por una secuencia del motivo adyacente protoespaciador (PAM). En una realización, la suma total de los brazos de homología 5' y 3' del LTVEC es de aproximadamente 10 kb a aproximadamente 150 kb. En una realización, la suma total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 120 kb, o de aproximadamente 120 kb a 150 kb. Esta divulgación establece que la modificación genética escogida comprende: (a) un reemplazo de una secuencia endógena de ácido nucleico con una secuencia de ácido nucleico homóloga u ortóloga; (b) una eliminación de una secuencia endógena de ácido nucleico; (c) una eliminación de una secuencia endógena de ácido nucleico, en donde la eliminación 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, 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; (d) inserción de una secuencia exógena de ácido nucleico; (e) inserción de una secuencia exógena de ácido nucleico 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; (f) inserción de una secuencia exógena de ácido nucleico que comprende una secuencia de ácido nucleico homóloga u ortóloga; (g) inserción de una secuencia quimérica de ácido nucleico que comprende una secuencia de ácido nucleico humana y no humana; (h) inserción de un alelo condicional flanqueado con secuencias objetivo de recombinasa específicas del sitio; (i) inserción de un marcador seleccionable o un gen informador unido operativamente a un tercer promotor activo en la célula de mamífero; o (j) una combinación de los mismos. En una realización, el locus genómico objetivo comprende (i) una secuencia objetivo 5' que es homóloga a un brazo de homología 5'; y (ii) una secuencia objetivo 3' que es homóloga a un brazo de homología 3'. En una realización, la secuencia objetivo 5' y la secuencia objetivo 3' están separadas por al menos 5 kb pero menos de 3 Mb. En una realización, la secuencia objetivo 5' y la secuencia objetivo 3' están separadas por al menos 5 kb pero menos de 10 kb, al menos 10 kb pero menos de 20 kb, al menos 20 kb pero menos de 40 kb, al menos 40 kb pero menos de 60 kb, al menos 60 kb pero menos de 80 kb, al menos aproximadamente 80 kb pero menos de 100 kb, al menos 100 kb pero menos de 150 kb, o al menos 150 kb pero menos de 200 kb, al menos aproximadamente 200 kb pero menos de aproximadamente 300 kb, al menos aproximadamente 300 kb pero menos de aproximadamente 400 kb, al menos aproximadamente 400 kb pero menos de aproximadamente 500 kb, al menos aproximadamente 500 kb pero menos de aproximadamente 1 Mb, al menos aproximadamente 1 Mb pero menos de aproximadamente 1,5 Mb, al menos aproximadamente 1,5 Mb pero menos de aproximadamente 2 Mb, al menos aproximadamente 2 Mb pero menos de aproximadamente 2,5 Mb, o al menos aproximadamente 2,5 Mb pero menos de aproximadamente 3 Mb. En una realización, el primero y segundo constructos de expresión están en una única molécula de ácido nucleico. En una realización, el locus genómico objetivo comprende el locus *Sry*.

Se proporciona además en esta divulgación un método para la modificación genética escogida en el cromosoma Y de un animal no humano, que comprende: (a) modificar un locus genómico de interés en el cromosoma Y de una célula pluripotente no humana de acuerdo con los métodos descritos en el presente documento, produciendo así una célula pluripotente no humana modificada genéticamente que comprende una modificación genética escogida en el cromosoma Y; (b) introducción de la célula pluripotente no humana modificada de (a) en un embrión huésped no humano; y gestar el embrión huésped no humano que comprende la célula pluripotente modificada en una madre sustituta, en donde la madre sustituta produce progenie F0 que comprende la modificación genética escogida, en donde la modificación genética escogida puede transmitirse a través de la línea germinal. En una realización, el locus genómico de interés comprende el locus *Sry*.

Se proporcionan métodos y composiciones para generar modificaciones genéticas escogidas en el cromosoma Y. Las composiciones incluyen un cultivo *in vitro* que comprende una célula animal XY pluripotente y/o totipotente (es decir, células ES XY o células iPS XY) que tiene una modificación que disminuye el nivel y/o la actividad de una proteína *Sry*; y cultivar estas células en un medio que promueva el desarrollo de hembras fértiles F0 XY. Dichas composiciones encuentran uso en varios métodos para elaborar mamíferos no humanos XY hembra fértiles en una generación F0.

Breve descripción de las Figuras.

La FIG. 1 proporciona un esquema de Cas9/ARNg de CRISPR que se dirige al gen *Sry* de ratón. VG-1 (SEQ ID NO: 10); VG-2 (SEQ ID NO: 11); VG-3 (SEQ ID NO: 12). Los cebadores y sondas indicados en la FIG. 1 se proporcionan en las SEQ ID NOS: 13-29.

La FIG. 2 proporciona un esquema de direccionamiento del gen *Sry* con TALEN y CRISPR utilizando un gen informador *lacZ*. El gen *Sry* fue direccionado con un LTVEC y un vector de brazos cortos (TVEC pequeño) con brazos de homología más pequeños que un LTVEC para evitar loci de ataque en el cromosoma Y.

La FIG. 3 ilustra la expresión de *LacZ* en embriones.

La FIG. 4 proporciona un esquema de una gran eliminación de más de 500 kb en el cromosoma Y mediado por ZFN o por los ARN guía de CRISPR en combinación con la ASD endonucleasa de Cas9.

Las FIG. 5A, B y C proporcionan la confirmación de la secuenciación de la gran eliminación del cromosoma Y en varios clones. La FIG. 5A es el resultado de la secuenciación para el clon 1-D5. La secuencia *Kdm5* hacia arriba y *Uspy9* hacia abajo se proporcionan en la SEQ ID NO: 30; 1-D5 1500F (SEQ ID NO: 31); 1-D5 1000R (SEQ ID NO: 32); la FIG. 5B es el resultado de la secuenciación para el clon 5-C4. La secuencia *Kdm5* hacia arriba y *Uspy9* hacia abajo se proporcionan en la SEQ ID NO: 33; 1500F (SEQ ID NO: 34); 1000R (SEQ ID NO: 35); 1000F (SEQ ID NO: 36); y la FIG. 5C es el resultado de la secuenciación para el clon 6-A12. La secuencia *Kdm5* hacia arriba y *Uspy9* hacia abajo se proporcionan en la SEQ ID NO: 37; 1500F (SEQ ID NO: 38); 1000R (SEQ ID NO: 39); 1000F (SEQ ID NO: 40); 1500R (SEQ ID NO: 41). Las regiones encuadradas en la FIG. 5B y la FIG. 5C representan regiones de microhomología.

Descripción detallada

Definiciones

Los términos "proteína", "polipéptido" y "péptido", usados indistintamente en este documento, incluyen formas poliméricas de aminoácidos de cualquier longitud, incluidos los aminoácidos codificados y no codificados y los aminoácidos derivados o modificados química o bioquímicamente. Los términos también incluyen polímeros que se han modificado, tales como polipéptidos que tienen esqueletos peptídicos modificados.

- 5 Los términos "ácido nucleico" y "polinucleótido", usados indistintamente en este documento, incluyen formas poliméricas de nucleótidos de cualquier longitud, incluidos ribonucleótidos, desoxirribonucleótidos, o análogos o versiones modificadas de los mismos. Incluyen ADN o ARN de cadena simple, doble y múltiple, ADN genómico, ADNc, híbridos de ADN-ARN y polímeros que comprenden bases de purina, bases de pirimidina u otras sustancias naturales, modificadas químicamente, modificadas bioquímicamente, no naturales, o bases de nucleótidos derivadas. Por simplicidad, el tamaño del ácido nucleico puede referirse en pb si el ácido nucleico está en forma de doble cadena o de cadena sencilla, en este último caso, siendo los pb los que se forman cuando el ácido nucleico de cadena sencilla se duplica con su cadena exactamente complementaria.

- 15 La "optimización de codones" generalmente incluye un proceso de modificación de una secuencia de ácido nucleico para la expresión mejorada en células huésped particulares reemplazando al menos un codón de la secuencia nativa con un codón que se usa más frecuentemente o lo más frecuentemente en los genes de célula huésped manteniendo la secuencia nativa de aminoácidos. Por ejemplo, un ácido nucleico que codifica una proteína Cas puede modificarse para sustituir codones que tienen una mayor frecuencia de uso en una célula procarionota o eucariota dada, incluida una célula bacteriana, una célula de levadura, una célula humana, una célula no humana, una célula de mamífero, una célula de roedor, una célula de ratón, una célula de rata, una célula de hámster o cualquier otra célula huésped, en comparación con la secuencia de ácido nucleico que se produce naturalmente. Las tablas de uso de codones están disponibles, por ejemplo, en la "Base de datos de uso de codones". Estas tablas se pueden adaptar de varias maneras. Véase Nakamura et al. (2000) *Nucleic Acids Research* 28: 292. Los algoritmos informáticos para la optimización de codones de una secuencia particular para la expresión en un huésped particular también están disponibles (véase, por ejemplo, Gene Forge).

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

- 30 "Complementariedad" de ácidos nucleicos significa que una secuencia de nucleótidos en una cadena de ácido nucleico, debido a la orientación de sus grupos nucleobásicos, forma enlaces de hidrógeno con otra secuencia en una cadena opuesta de ácido nucleico. Las bases complementarias en el ADN son típicamente A con T y C con G. En el ARN, son típicamente C con G y U con A. La complementariedad puede ser perfecta o sustancial/suficiente. La perfecta complementariedad entre dos ácidos nucleicos significa que los dos ácidos nucleicos pueden formar un dúplex en el que cada base en el dúplex está unida a una base complementaria por el emparejamiento de Watson-Crick. Complementariedad "sustancial" o "suficiente" significa que una secuencia en una cadena no es completa y/o perfectamente complementaria a una secuencia en una cadena opuesta, pero que se produce una unión suficiente entre las bases en las dos cadenas para formar un complejo híbrido estable junto con las condiciones de hibridación (p. ej., concentración de sal y temperatura). Dichas condiciones se pueden predecir utilizando las secuencias y los cálculos matemáticos estándar para predecir la T_m de las cadenas hibridadas, o mediante la determinación empírica de T_m utilizando métodos de rutina. T_m incluye la temperatura a la cual una población de complejos de hibridación formados entre dos cadenas de ácido nucleico está desnaturalizada al 50%. A una temperatura por debajo de la T_m , se favorece la formación de un complejo de hibridación, mientras que a una temperatura por encima de la T_m , se favorece la fusión o separación de las cadenas en el complejo de hibridación. La T_m puede estimarse para un ácido nucleico que tenga un contenido conocido de G + C en una solución acuosa de NaCl 1 M utilizando, por ejemplo, $T_m = 81,5 + 0,41 (\% \text{ de G} + \text{C})$, aunque otros cálculos de T_m conocidos tienen en cuenta las características estructurales del ácido nucleico.

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

- 60 La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque son posibles los desajustes entre las bases. Las condiciones apropiadas para la hibridación entre dos ácidos nucleicos dependen de la longitud de los ácidos nucleicos y del grado de complementación, variables bien conocidas en la técnica. Cuanto mayor sea el grado de complementación entre dos secuencias de nucleótidos, mayor será el valor de la temperatura de fusión (T_m) para los híbridos de ácidos nucleicos que tienen esas secuencias. Para las hibridaciones

entre ácidos nucleicos con tramos cortos de complementariedad (p. ej., complementariedad sobre 35 o menos, 30 o menos, 25 o menos, 22 o menos, 20 o menos, o 18 o menos nucleótidos) la posición de los desajustes se vuelve importante (véase Sambrook et al., citado más arriba, 11.7-11.8). Normalmente, la longitud de un ácido nucleico hibridable es de al menos aproximadamente 10 nucleótidos. Las longitudes mínimas ilustrativas para un ácido nucleico hibridable incluyen al menos aproximadamente 15 nucleótidos, al menos aproximadamente 20 nucleótidos, al menos aproximadamente 22 nucleótidos, al menos aproximadamente 25 nucleótidos y al menos aproximadamente 30 nucleótidos. Además, la temperatura y la concentración de sal de la solución de lavado se pueden ajustar según sea necesario de acuerdo con factores tales como la longitud de la región de complementación y el grado de complementación.

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

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

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

La "identidad de secuencia" o "identidad" en el contexto de dos polinucleótidos o secuencias de polipéptidos hace referencia a los residuos en las dos secuencias que son iguales cuando se alinean para una correspondencia máxima en una ventana de comparación específica. Cuando se usa el porcentaje de identidad de secuencia en referencia a las proteínas, se reconoce que las posiciones de los residuos que no son idénticas a menudo difieren por las sustituciones de aminoácidos conservadoras, donde los residuos de aminoácidos son sustituidos 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 porcentaje de identidad de secuencia puede ajustarse hacia arriba para corregir la naturaleza conservadora de la sustitución. Se dice que las secuencias que difieren por tales sustituciones conservadoras tienen "similitud de secuencia" o "similitud". Los medios para realizar este ajuste son bien conocidos por los expertos en la técnica. Por lo general, esto implica calificar una sustitución conservadora como una falta de coincidencia parcial en lugar de total, lo que aumenta el porcentaje de identidad de secuencia. Así, por ejemplo, cuando a un aminoácido idéntico se le asigna un puntaje de 1 y a una sustitución no conservadora se le asigna un puntaje de cero, se le asigna un puntaje entre 0 y 1 a la sustitución conservadora. Se calcula el puntaje de las sustituciones conservadoras, por ejemplo, tal como se implementó en el programa PC/GENE (Intelligenetics, Mountain View, California).

El "porcentaje de identidad de secuencia" incluye el valor determinado comparando dos secuencias alineadas de manera óptima en una ventana de comparación, en donde la porción de la secuencia de polinucleótido en la ventana de comparación puede incluir adiciones o eliminaciones (es decir, huecos) en comparación con la secuencia de referencia (que no comprende adiciones o eliminaciones) para una alineación óptima de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las que está presente la base de ácido nucleico o aminoácido idéntico en ambas secuencias para obtener el número de posiciones coincidentes, dividiendo el número de posiciones coincidentes por el número total de posiciones en la ventana de comparación, y multiplicando el resultado por 100 para obtener el porcentaje de identidad de secuencia.

A menos que se indique lo contrario, los valores de identidad/similitud de secuencia incluyen el valor obtenido al usar GAP versión 10 usando los siguientes parámetros: % de identidad y % de similitud para una secuencia de nucleótidos usando ponderación por huecos de 50 y ponderación por longitud de 3, y la matriz de puntuación nwsgapdna.cmp; % de identidad y % de similitud para una secuencia de aminoácidos usando ponderación por huecos de 8 y ponderación por longitud de 2, y la matriz de puntuación BLOSUM62; o cualquier programa equivalente de los mismos. El "programa equivalente" incluye cualquier programa de comparación de secuencias que, para cualquiera de las dos secuencias en cuestión, genere una alineación con idénticas coincidencias de

residuos de nucleótidos o aminoácidos y un porcentaje idéntico de identidad de secuencia en comparación con la alineación correspondiente generada por la versión 10 de GAP.

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

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

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

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

I. Métodos y composiciones para elaborar un animal XY hembra fértil en una generación F0

15 Se conocen métodos para hacer animales no humanos a partir de células ES de donante y embriones huésped. Las células ES de donante se seleccionan por ciertas características que mejoran la capacidad de las células para poblar un embrión huésped y, por lo tanto, contribuyen en parte o en gran parte a un animal formado por las células ES del donante y el embrión huésped. El animal formado puede ser macho o hembra, basado en gran parte en el genotipo de la célula ES (por ejemplo, XY o XX).

20 La mayoría de las líneas de células ES para hacer animales transgénicos tienen un genotipo XY masculino. Debido a la dominancia del cromosoma Y en la determinación del sexo en mamíferos, cuando las células ES XY se introducen en un embrión huésped del blastocisto y se gestan, casi siempre producen en la primera generación (F0) animales fenotípicamente masculinos que son quimeras, es decir, que contienen células derivadas de la célula ES del donante masculino (XY) y células derivadas del embrión huésped, que pueden ser masculinas (XY) o hembra (XX). Las células ES XY, cuando se introducen en un embrión huésped de 8 células mediante el método
25 VelociMouse y se gestan, pueden producir en la primera generación (F0) animales fenotípicamente machos que se derivan completamente de las células ES XY.

El documento WO2011/156723 proporciona métodos y composiciones que emplean un medio de cultivo para mantener células donantes XY en cultivo, de manera que después de la introducción de las células donantes XY en un embrión huésped y la gestación en un huésped adecuado, se producen animales hembra XY fértiles en la
30 población F0. Dichas composiciones encuentran uso en la elaboración de la progenie F1 que son homocigotas para la modificación genética escogida dada.

La presente divulgación proporciona métodos y composiciones que emplean una combinación de células donantes XY que tienen una modificación que disminuye el nivel y/o la actividad de la proteína Sry en combinación con un medio de cultivo que promueve la producción de hembras F0 XY anatómicamente normales, fértiles y fecundas.
35 Dichos métodos y composiciones permiten hacer un animal XY hembra fértil no humano en una generación F0. La combinación de células ES XY que tienen una modificación que disminuye el nivel y/o la actividad de la proteína Sry en combinación con los medios de cultivo descritos en el presente documento aumenta significativamente el porcentaje de progenie XY hembra fértil en la generación F0. Los métodos para la conversión eficiente del sexo de hombre a mujer son valiosos para la industria de animales domésticos. Por ejemplo, las terneras son mucho más
40 valiosas para la industria de ganado lechero que los machos. Lo mismo es cierto para las aves de corral. Para fines de reproducción, ya sea ganado vacuno o porcino u ovino, se prefiere criar muchas hembras con solo unos pocos toros, jabalíes o carneros. Por lo tanto, los diversos métodos proporcionados en este documento encuentran uso en varias industrias de reproducción comercialmente importantes.

También se proporcionan métodos y composiciones para hacer una línea de células madre (ES) embrionarias XY capaz de producir un mamífero hembra XY fértil no humano en una generación F0 sin cultivar en un medio de feminización. En tales métodos, la línea de células ES XY que tiene una modificación que disminuye el nivel y/o la actividad de una proteína Sry puede producir una línea de células ES capaz de producir un mamífero hembra XY fértil no humano en una generación F0 en ausencia de un medio de feminización proporcionado en otro lugar del presente documento (por ejemplo, mediante el cultivo en un medio base, tal como DMEM, descrito en otro lugar del
50 presente documento).

A. Células XY animales que tienen una modificación que disminuye el nivel y/o la actividad de una proteína Sry

Se proporcionan diversas composiciones y métodos en el presente documento que comprenden diversas células pluripotentes y/o totipotentes XY de un animal. El término "célula pluripotente" como se usa en el presente documento incluye una célula indiferenciada que posee la capacidad de desarrollarse en más de un tipo de células
55 diferenciadas. Estas células XY pluripotentes y/o totipotentes pueden ser, por ejemplo, una célula madre (ES) embrionaria o una célula madre pluripotente inducida (iPS). El término "célula madre embrionaria" o "célula ES",

como se usa en el presente documento, incluye una célula totipotente o pluripotente derivada de embrión que es capaz de contribuir a cualquier tejido del embrión en desarrollo tras la introducción en un embrión.

El término "animal", en referencia a células, células pluripotentes y/o totipotentes, células XY, células ES, células iPS, células donantes y/o embriones huésped, incluye mamíferos, peces y aves. Los mamíferos incluyen, por ejemplo, seres humanos, primates no humanos, monos, simios, perros, caballos, toros, venados, bisontes, ovejas, roedores (por ejemplo, ratones, ratas, hámsteres, cobayas), ganado (por ejemplo, especies bovinas, por ejemplo, vacas, novillos, etc., especies ovinas, por ejemplo, ovejas, cabras, etc., y especies porcinas, por ejemplo, cerdos y jabalíes). Las aves incluyen, por ejemplo, pollos, pavos, avestruces, gansos, patos, etc. También se incluyen animales domesticados y animales agrícolas. La frase "animal no humano", en referencia a células, células XY, células ES, células donantes y/o embriones huésped, excluye a los humanos.

En realizaciones específicas, la célula pluripotente es una célula ES XY humana, una célula iPS XY humana, una célula ES XY adulta humana, una célula ES progenitora humana con desarrollo restringido, una célula ES XY no humana, una célula iPS XY no humana, una célula ES XY de roedor, una célula iPS XY de roedor, una célula ES XY de ratón, una célula iPS XY de ratón, una célula ES XY de rata, una célula iPS XY de rata, una célula ES XY de hámster, una célula iPS XY de hámster, una célula ES XY de mono, una célula iPS XY de mono, una célula ES XY de mamífero agrícola, una célula iPS XY agrícola, una célula ES XY de mamífero domesticado o una célula iPS XY domesticada. Además, la célula ES XY o la célula iPS XY pueden ser de una cepa consanguínea, una cepa híbrida o una cepa exógama. Se reconoce además que las células XY pluripotentes y/o totipotentes pueden comprender un cariotipo XYY o un cariotipo XXY.

Las células de ratón pluripotentes y/o totipotentes (es decir, células ES XY o células iPS XY) pueden ser de una cepa 129, una cepa C57BL/6, una mezcla de 129 y C57BL/6, una cepa BALB/c, o una cepa Swiss Webster. En una realización específica, el ratón es 50% 129 y 50% C57BL/6. En una realización, el ratón es una cepa 129 seleccionada del grupo que consiste en una cepa que es 129P1, 129P2, 129P3, 129X1, 129S1 (por ejemplo, 129S1/SV, 129S1/SvIm), 129S2, 129S4, 129S5, 129S9/SvEvH, 129S6 (129/SvEvTac), 129S7, 129S8, 129T1, 129T2. Véase, por ejemplo, Festing et al. (1999) *Mammalian Genome* 10: 836. En una realización, el ratón es una cepa C57BL, y en una realización específica es de C57BL/A, C57BL/An, C57BL/GrFa, C57BL/Kal_wN, C57BL/6, C57BL/6J, C57BL/6ByJ, C57BL/6NJ, C57BL/6NTac, C57BL/10, C57BL/10ScSn, C57BL/10Cr, o C57BL/Ola. En una realización específica, el ratón es una mezcla de una cepa 129 mencionada anteriormente y una cepa C57BL/6 mencionada anteriormente. En otra realización específica, el ratón es una mezcla de las cepas 129 mencionadas anteriormente, o una mezcla de las cepas BL/6 mencionadas anteriormente. En una realización específica, la cepa 129 de la mezcla es una cepa 129S6 (129/SvEvTac). En algunas realizaciones, la célula ES XY de ratón comprende un cromosoma Y derivado de la cepa 129.

En otra realización más, la célula ES de ratón XY es una célula ES de ratón VGF1. Las células ES de ratón VGF1 (también conocidas como F1H4) se derivaron de embriones híbridos producidos al cruzar un ratón hembra C57BL/6NTac con un ratón macho 129S6/SvEvTac. Por lo tanto, las células ES de VGF1 contienen un cromosoma Y del ratón 129S6/SvEvTac. Véase, por ejemplo, Auerbach, W. et al. (2000) *Establishment and chimera analysis of 129/SvEv- and C57BL/6-derived mouse embryonic stem cell lines*. *Biotechniques* 29, 1024-1028, 1030, 1032.

Una célula pluripotente y/o totipotente de rata (es decir, una célula ES XY o una célula iPS XY) puede ser de cualquier cepa de rata, incluida, entre otras, una cepa de rata ACI, una cepa de rata Dark Agouti (DA), una cepa de rata Wistar, una cepa de rata LEA, una cepa de rata Sprague Dawley (SD) o una cepa de rata Fischer tal como Fisher F344 o Fisher F6. Las células pluripotentes y/o totipotentes de rata (es decir, células ES XY o células iPS XY) también pueden obtenerse a partir de una cepa derivada de una mezcla de dos o más cepas citadas anteriormente. En una realización, la célula pluripotente y/o totipotente de rata (es decir, célula ES XY o célula iPS XY) se deriva de una cepa seleccionada de una cepa DA y una cepa ACI. En una realización específica, la célula pluripotente y/o totipotente de rata (es decir, célula ES XY o célula iPS XY) se deriva de una cepa ACI. La cepa de rata ACI se caracteriza por tener agutí negro, con vientre y patas blancas y un haplotipo *RT1^{av1}*. Dichas cepas están disponibles a partir de una variedad de fuentes, incluidos los Laboratorios Harlan. En otras realizaciones, las diversas células pluripotentes y/o totipotentes de rata (es decir, célula ES XY o célula iPS XY) son de una cepa de rata Dark Agouti (DA), que se caracteriza por tener una capa de agutí y un haplotipo *RT1^{av1}*. Tales ratas están disponibles en una variedad de fuentes, incluyendo Charles River y Harlan Laboratories. En una realización adicional, las células pluripotentes y/o totipotentes de rata (es decir, células ES XY o células iPS XY) son de una cepa de rata endogámica. En realizaciones específicas, la línea de células ES de rata es de una rata ACI y comprende la célula ES de rata ACI.G1. En otra realización, la línea de células ES de rata es de una rata DA y comprende la línea de células ES de rata DA.2B o la línea de células ES de rata DA.2C. Véase, por ejemplo, la solicitud de patente de utilidad de EE.UU. No. 14/185.703, presentada el 20 de febrero de 2014.

En diversas realizaciones, la célula pluripotente y/o totipotente (es decir, la célula ES XY o la célula iPS XY), la célula del donante y/o el embrión huésped no son de uno o más de los siguientes: *Akodon spp.*, *Myopus spp.*, *Microtus spp.*, *Talpa spp.* En diversas formas de realización, la célula del donante y/o el embrión huésped no son de ninguna especie cuya característica normal de tipo silvestre sea la fertilidad femenina XY. En varias realizaciones, cuando una modificación genética está presente en la célula pluripotente y/o totipotente (es decir, célula ES XY o célula iPS XY), la célula del donante o el embrión huésped, la modificación genética no es una XYY o XXY, una inversión de

sexo negativa para Tdy, inversión de sexo positiva para Tdy, una modificación XO, una aneuploidía, un genotipo *fgf9^{-/-}*, o una modificación SOX9.

5 Las células XY pluripotentes y/o totipotentes (es decir, una célula ES XY o una célula iPS XY) empleadas en los métodos y composiciones tienen una modificación genética que da como resultado un nivel y/o actividad disminuidos de la proteína Sry. La proteína de la "región Y determinante del sexo" o la proteína "Sry" es un factor de transcripción que es un miembro de la familia de proteínas de unión al ADN del grupo de alta movilidad (HMG). Sry es el factor determinante de los testículos que inicia la determinación del sexo masculino. Se conoce la secuencia de la proteína Sry de una variedad de organismos, incluida la de ratón (No. de acceso Q05738); de rata (GenBank: CAA61882.1); humana (No. de acceso Q05066); de gato (No. de acceso Q67C50), y de caballo (No. de acceso P36389).

10 En general, el nivel y/o la actividad de la proteína Sry disminuye si el nivel de proteína y/o el nivel de actividad de la proteína Sry es estadísticamente más bajo que el nivel de proteína de Sry en una célula de control apropiada que no ha sido genéticamente modificada o sometida a mutagénesis para inhibir la expresión y/o actividad de la proteína Sry. En realizaciones específicas, la concentración y/o actividad de la proteína Sry disminuye en al menos 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, o 90% en relación con una célula de control que no se ha modificado
15 para tener el nivel y/o actividad disminuidos de la proteína Sry.

Una "célula sujeto" es una en la que se ha efectuado una alteración genética, tal como una modificación genética descrita en el presente documento, o es una célula que descende de una célula así alterada y que comprende la alteración. Un "control" o "célula de control" proporciona un punto de referencia para medir los cambios en el fenotipo de la célula sujeto. En una realización, una célula de control es lo más parecida posible a la célula con actividad reducida de Sry, excepto que carece de la modificación genética o mutación que resulta en la actividad reducida (por ejemplo, las células respectivas pueden originarse en la misma línea celular). En otros casos, la célula de control puede comprender, por ejemplo: (a) una célula de tipo silvestre, es decir, del mismo genotipo que el material de partida para la alteración genética que dio lugar a la célula sujeto; (b) una célula del mismo genotipo que el material de partida pero que ha sido modificada genéticamente con una construcción nula (es decir, con una construcción que no tiene ningún efecto conocido sobre el rasgo de interés, tal como una construcción que comprende un gen marcador); (c) una célula que es una progenie no modificada genéticamente de una célula sujeto (es decir, la célula de control y la célula sujeto se originan en la misma línea celular); (d) una célula genéticamente idéntica a la célula sujeto pero que no está expuesta a condiciones o estímulos que induzcan la expresión del gen de interés; o (e) la propia célula sujeto, bajo condiciones en las cuales la modificación genética no produce una alteración en la expresión del polinucleótido de interés.
20
25
30

El nivel de expresión del polipéptido Sry se puede medir directamente, por ejemplo, analizando el nivel del polipéptido Sry en la célula u organismo, o indirectamente, por ejemplo, midiendo la actividad del polipéptido Sry. Se conocen varios métodos para determinar la actividad de la proteína Sry. Véase, Wang et al. (2013) Cell 153: 910-918, Mandalos et al. (2012) PLOS ONE 7: e45768: 1-9, y Wang et al. (2013) Nat Biotechnol. 31: 530-532.

35 En otros casos, las células que tienen la modificación genética escogida que reduce la actividad y/o el nivel del polipéptido Sry se seleccionan usando métodos que incluyen, entre otros, análisis de transferencia Southern, secuenciación de ADN, análisis por PCR o análisis fenotípicos. Dichas células se emplean luego en los diversos métodos, composiciones y kits descritos en el presente documento.

40 Una modificación genética escogida puede comprender una alteración dirigida a un polinucleótido de interés que incluye, por ejemplo, una alteración dirigida a un locus genómico objetivo en el cromosoma Y, una alteración dirigida al gen Sry, o una alteración dirigida a otros polinucleótidos deseados. Dichas modificaciones escogidas incluyen, pero no se limitan a, adiciones de uno o más nucleótidos, eliminaciones de uno o más nucleótidos, sustituciones de uno o más nucleótidos, una inactivación del polinucleótido de interés o una porción del mismo, una activación del polinucleótido de interés o una porción del mismo, un reemplazo de una secuencia endógena de ácido nucleico con una secuencia heteróloga de ácido nucleico, o una combinación de los mismos. En realizaciones específicas, al menos 1, 2, 3, 4, 5, 7, 8, 9, 10 o más nucleótidos se cambian para formar la modificación genómica escogida.
45

Una disminución en el nivel y/o actividad de la proteína Sry puede resultar de una modificación genética en el gen Sry (es decir, una modificación genética en una región reguladora, la región de codificación, y/o intrones, etc.). Dichas modificaciones genéticas incluyen, pero no se limitan a, adiciones, eliminaciones y sustituciones de nucleótidos en el genoma. Dichas modificaciones genéticas pueden incluir una alteración del gen Sry, que incluye, por ejemplo, una inserción de uno o más nucleótidos en el gen Sry, una eliminación de uno o más nucleótidos del gen Sry, una sustitución de uno o más nucleótidos en el gen Sry, una inactivación del gen Sry o una parte del mismo, una activación del gen Sry o una parte del mismo, un reemplazo de una secuencia endógena de ácido nucleico con una secuencia heteróloga de ácido nucleico, o una combinación de las mismas. Por lo tanto, en realizaciones específicas, la actividad de un polipéptido Sry puede reducirse o eliminarse mediante la alteración del gen que codifica el polipéptido Sry. En realizaciones específicas, al menos 1, 2, 3, 4, 5, 7, 8, 9, 10 o más nucleótidos se cambian en el gen Sry. Se pueden utilizar varios métodos para generar la modificación genética adicional escogida. Véase, por ejemplo, Wang et al. (2013) Cell 153: 910-918, Mandalos et al. (2012) PLOS ONE 7: e45768: 1-9, y Wang et al. (2013) Nat Biotechnol. 31: 530-532. Además, los diversos métodos descritos en el presente
50
55

documento para modificar el locus genómico en el cromosoma Y se pueden usar para introducir una modificación genética escogida en el gen *Sry*.

5 En otras realizaciones, la actividad y/o el nivel del polipéptido *Sry* se reducen o eliminan introduciendo en la célula un polinucleótido que inhibe el nivel o la actividad del polipéptido *Sry*. El polinucleótido puede inhibir la expresión del polipéptido *Sry* directamente, impidiendo la traducción del ARN mensajero de *Sry*, o indirectamente, codificando un polipéptido que inhibe la transcripción o traducción del gen que codifica una proteína *Sry*. En otras realizaciones, la actividad del polipéptido *Sry* se reduce o elimina introduciendo en la célula una secuencia que codifica un polipéptido que inhibe la actividad del polipéptido *Sry*.

10 En una realización, las células pluripotentes y/o totipotentes XY (es decir, células ES XY o células iPS XY) comprenden un alelo *Sry* condicional que reduce la actividad y/o el nivel de la proteína *Sry*. Un "alelo *Sry* condicional" incluye un gen *Sry* modificado manipulado para tener el nivel y/o actividad disminuidos de la proteína *Sry* en un tiempo de desarrollo deseado y/o dentro de un tejido de interés deseado. El nivel y/o la actividad reducidos se pueden comparar con una célula de control que carece de la modificación que da lugar al alelo condicional, o en el caso de una actividad reducida en un tiempo de desarrollo deseado con tiempos anteriores y/o
15 posteriores, o en el caso de un tejido deseado, con una actividad media de todos los tejidos. En una realización, el alelo *Sry* condicional comprende un alelo condicional nulo de *Sry* que puede desconectarse en un punto de tiempo de desarrollo deseado y/o en tejidos específicos. Este alelo condicional se puede utilizar para crear hembras XY fértiles derivadas de cualquier clon de reconocimiento génico. Como se describe en otra parte en este documento, un método de este tipo permite la creación de una modificación genética homocigótica deseada en la generación F1.
20 Dichos métodos proporcionan una mirada rápida al fenotipo sin tener que reproducirse hasta la generación F2.

En una realización no limitativa, el alelo *Sry* 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 de selección de fármaco (DSC) 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 que divide el exón y un módulo invertible de tipo trampa de genes; véase, por ejemplo, el documento US 2011/0104799); y (c) unidades recombinables que se recombinan tras la exposición a una primera recombinasa para formar un alelo condicional que (i) carece de la secuencia de activación y el DSC, y (ii) contiene la NSI en orientación sentido y el COIN en orientación antisentido.

30 El alelo condicional del gen *Sry* puede generarse en cualquier tipo de célula, y no está limitado a una célula pluripotente y/o totipotente XY. Dichos tipos de células, junto con métodos no limitantes para atacar un locus genómico en el cromosoma Y, se discuten con más detalle en otra parte de este documento.

Como se discutió en otra parte en este documento, la célula XY pluripotente y/o totipotente (es decir, una célula ES XY o una célula iPS XY) que tiene modificación genética que disminuye el nivel y/o la actividad de la proteína *Sry* puede comprender además al menos una modificación genética adicional dirigida a un polinucleótido de interés. Al menos una modificación genética escogida adicional puede comprender una sustitución de uno o más ácidos nucleicos, un reemplazo de una secuencia endógena de ácido nucleico con una secuencia heteróloga de ácido nucleico, una inactivación y una activación. La modificación genética adicional dirigida puede estar en el cromosoma Y, el cromosoma X o en un autosoma. Se pueden usar varios métodos para generar la modificación genética escogida adicional, incluido el empleo de plásmidos de direccionamiento y vectores de direccionamiento grandes como se describe en otra parte del presente documento. Véanse, también, los documentos US20080092249, WO/1999/005266A2, US20040177390, WO/2008/017234A1, y la patente de EE. UU. No. 7.612.250, para métodos relacionados con la transferencia nuclear. Además, los diversos métodos descritos aquí para modificar el locus genómico en el cromosoma Y (es decir, el gen *Sry*) también se pueden usar para introducir modificaciones genéticas escogidas a polinucleótidos de interés que no están ubicados en el cromosoma Y.

B. Medios para cultivar las células XY pluripotentes y/o totipotentes que tienen una modificación que disminuye el nivel y/o la actividad de una proteína *Sry*

Los medios de cultivo empleados en los diversos métodos y composiciones que promueven a la hembra fértil XY en la generación F0 es tal que mantiene las células pluripotentes y/o totipotentes (es decir, células ES, células iPS, células ES XY, células iPS XY, etc.). Los términos "mantener", "que mantiene" y "mantenimiento" se refieren a la conservación estable de al menos una o más de las características o fenotipos de células pluripotentes y/o totipotentes descritas en el presente documento (incluidas las células ES o las células iPS). Tales fenotipos pueden incluir mantener la pluripotencia y/o la totipotencia, la morfología celular, los perfiles de expresión génica y las otras características funcionales de las células. Los términos "mantener", "que mantiene" y "mantenimiento" también
55 pueden abarcar la propagación de las células, o un aumento en el número de células que se están cultivando. Los términos contemplan además las condiciones de cultivo que permiten que las células permanezcan pluripotentes, mientras que las células pueden o no continuar dividiéndose y aumentando en número.

En algunas realizaciones, las células XY que tienen la modificación genética que reduce el nivel y/o la actividad de la proteína *Sry* se mantienen mediante el cultivo en DMEM que es adecuado para su uso (con suplementos añadidos)

para cultivar o mantener las células pluripotentes y/o totipotentes (es decir, células ES, células iPS, células ES XY, células iPS XY, etc.) en cultivo. En tales casos, las células ES XY cultivadas tienen el potencial de convertirse en animales hembra fértiles, pero aún conservan la pluripotencia y/o totipotencia, de modo que las células pueden implementarse en un embrión receptor y dar lugar a una progenie femenina fértil.

- 5 En otras realizaciones, las células XY que tienen la modificación genética que reduce el nivel y/o la actividad de la proteína Sry se mantienen mediante el cultivo en un medio como se define adicionalmente a continuación durante un tiempo suficiente para que algunas de las células se conviertan en células XY con el potencial para convertirse en animales hembra fértiles, pero aún así retener la pluripotencia y/o totipotencia, de manera que las células pueden implementarse en un embrión receptor y dar lugar a una progenie femenina fértil.
- 10 El medio empleado para mantener las células pluripotentes y/o totipotentes XY (es decir, células ES XY, células iPS XY, etc.) que tienen la modificación genética que reduce el nivel y/o la actividad de la proteína Sry promueve el desarrollo de hembras fértiles F0 XY. Por lo tanto, el cultivo en un medio de este tipo aumenta el número de hembras fértiles F0 XY que se obtienen cuando se compara con el cultivo en un medio de control apropiado (como, por ejemplo, uno basado en DMEM). Por lo tanto, un mayor número de hembras fértiles F0 XY puede comprender al menos 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 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 los animales no humanos F0 (después de la introducción de las células ES XY de animales no humanos en un embrión huésped y la gestación del embrión huésped) son hembras XY y, al alcanzar la madurez sexual, el animal no humano hembra F0 XY es fértil.

La frase "medio base" o "medios base" incluye, por ejemplo, un medio base conocido en la técnica (por ejemplo, DMEM) que es adecuado para usar (con suplementos agregados) para hacer crecer o mantener células pluripotentes y/o totipotentes (es decir, células ES, células iPS, células ES XY, células iPS XY, etc.) en cultivo. Los medios base adecuados para producir una hembra XY fértil (es decir, "DMEM con bajo contenido de sal" o "medio de baja osmolalidad") difieren de los medios base utilizados normalmente para mantener las células ES en cultivo. Para los propósitos de discutir los medios base en general, los medios base que no son adecuados para producir hembras XY fértiles se describen en esta sección como "DMEM" y en la Tabla 1 (por ejemplo, medios DMEM típicos). Con el fin de analizar los medios base adecuados para producir hembras XY fértiles, se utiliza la frase "DMEM con bajo contenido de sal" o "DMEM con bajo contenido de osmolalidad". Las diferencias entre los medios base típicamente utilizados para mantener células pluripotentes y/o totipotentes en cultivo (por ejemplo, DMEM) y los medios base adecuados para producir hembras XY fértiles (por ejemplo, "DMEM con bajo contenido de sal") están articuladas aquí. La frase "DMEM con bajo contenido de sal" se usa por conveniencia; el DMEM adecuado para hacer hembras XY fértiles exhibe características no limitadas a "bajo en sal", pero incluye las descritas en este documento. Por ejemplo, el DMEM que se muestra en la Tabla 1 puede hacerse adecuado para producir hembras XY fértiles mediante la alteración de las concentraciones de cloruro de sodio y/o bicarbonato de sodio como se proporciona en este documento, lo que también resultará en una osmolalidad diferente y una conductividad diferente en comparación con el DMEM que se muestra en la Tabla 1. Un ejemplo de medio base es el medio de Eagle modificado por Dulbeco (DMEM), en varias formas (por ejemplo, Invitrogen DMEM, Cat. No. 11971-025) (Tabla 1). Un DMEM adecuado bajo en sal está disponible comercialmente como KO-DMEM^{MR} (Invitrogen Cat. No. 10829-018). El medio base se complementa típicamente con una serie de suplementos conocidos en la técnica cuando se usa para mantener células en cultivo para uso como células donantes. Dichos suplementos se indican como "suplementos" o "+ suplementos" en esta divulgación.

Tabla 1: Medios base DMEM para el mantenimiento o cultivo de células pluripotentes y/o totipotentes

Componente	Mg/L	mM
Glicina	30	0,4
L-Arginina•HCl	84	0,398
L-Cistina•2HCl	63	0,201
L-Glutamina	584	4
L-Histidina•HCl•H ₂ O	42	0,2
L-Isoleucina	105	0,802
L-Leucina	105	0,802
L-Lisina•HCl	146	0,798
L-Metionina	30	0,201

ES 2 694 629 T3

L-Fenilalanina	66	0,4
L-Serina	42	0,4
L-Treonina	95	0,798
L-Triptófano	16	0,0784
Sal disódica dihidratada de L-Tirosina	104	0,398
L-Valina	94	0,803
Cloruro de colina	4	0,0286
Pantotenato de calcio D	4	$8,39 \times 10^{-3}$
Ácido fólico	4	$9,07 \times 10^{-3}$
Niacinamida	4	0,0328
Piridoxina·HCl	4	0,0196
Riboflavina	0,4	$1,06 \times 10^{-3}$
Tiamina·HCl	4	0,0119
i-Inositol	7,2	0,04
Cloruro de calcio (CaCl ₂) (anhidro)	200	1,8
Nitrato férrico (Fe(NO ₃) ₃ ·9H ₂ O)	0,1	$2,48 \times 10^{-4}$
Sulfato de magnesio (MgSO ₄) (anhidro)	97,67	0,814
Cloruro de potasio (KCl)	400	5,33
D-Glucosa (Dextrosa)	4500	25
Rojo de fenol	15	0,0399
Contenido de NaCl/NaHCO ₃ de DMEM		
Bicarbonato de sodio (NaHCO ₃)	3700	44,05
Cloruro de sodio (NaCl)	6400	110,34
Contenido de NaCl/NaHCO ₃ de DMEM bajo en sales		
Bicarbonato de sodio (NaHCO ₃)	<3700	<44,05
Cloruro de sodio (NaCl)	<6400	<110,34

5 El término "suplementos" o la frase "+ suplementos" incluye elementos agregados al medio base para hacer crecer o mantener células pluripotentes y/o totipotentes (es decir, células ES XY o células iPS XY) en cultivo, por ejemplo, para mantener pluripotencia o totipotencia de células donantes en cultivo. Por ejemplo, los suplementos de medios adecuados para el crecimiento o el mantenimiento de células pluripotentes y/o totipotentes en cultivo incluyen, pero no se limitan a, suero fetal bovino (FBS), glutamina, antibiótico(s), penicilina y estreptomycin (por ejemplo, penstrep), sales piruvato (p. ej., piruvato de sodio), aminoácidos no esenciales (p. ej., MEM NEAA), 2-mercaptoetanol y factor inhibidor de la leucemia (LIF).

10 En una realización, el medio base comprende uno o más suplementos adecuados para mantener células pluripotentes en cultivo, incluyendo, por ejemplo, células ES XY o células iPS XY que tienen una capacidad reducida para contribuir al programa de desarrollo de la determinación del sexo masculino después de la inyección en una transferencia embrionaria e intrauterina a un ratón madre sustituto.

En una realización específica, los uno o más suplementos adecuados para mantener la célula pluripotente en cultivo son FBS (90 mL de FBS/0,5 L de medio base), glutamina (2,4 mmoles/0,5 L de medio base), piruvato de sodio (0,6 mmoles/0,5 L de medio base), aminoácidos no esenciales (<0,1 mmol/0,5 L de medio base), 2-mercaptoetanol, LIF y uno o más antibióticos.

- 5 En otras realizaciones, los medios para mantener células pluripotentes en cultivo, incluyendo, por ejemplo, células ES XY o células iPS XY que tienen una capacidad reducida para contribuir al programa de desarrollo de la determinación del sexo masculino después de la inyección en un embrión y transferencia intrauterina a una ratón madre sustituto, comprende aproximadamente 500 mL de medio base en el que se agregan los siguientes suplementos: aproximadamente 90 mL de FBS (por ejemplo, Hylcone FBS Cat. No. SH30070.03), aproximadamente 10 2,4 milimoles de glutamina (por ejemplo, aproximadamente 12 mL de una solución de glutamina 200 mM, por ejemplo, Invitrogen Cat. No. 25030-081, penicilina:estreptomina (por ejemplo, 60.000 unidades de penicilina G sódica y 60 mg de sulfato de estreptomina, con aproximadamente 51 mg de NaCl; por ejemplo, aproximadamente 6 mL de pennstrep de Invitrogen, Cat. No. 15140-122), aproximadamente 0,6 milimoles de piruvato de sodio (por ejemplo, 6 mL de piruvato de sodio 100 mM, Invitrogen Cat. No. 1 1360-070), aproximadamente 0,06 milimoles de 15 aminoácidos no esenciales (por ejemplo, aproximadamente 6 mL de MEM NEAA, por ejemplo, MEM NEAA de Invitrogen Cat. No. 1 140-050), aproximadamente 1,2 mL. 2-mercaptoetanol y aproximadamente 1,2 microgramos de LIF (por ejemplo, aproximadamente 120 microlitros de una preparación de LIF de 106 unidades/mL; por ejemplo, aproximadamente 120 microlitros de Millipore ESGRO^{MR}-LIF, Cat. No. ESG1 107). Al componer medios base para 20 mantener las células ES XY o iPS XY para producir hembras XY fértiles, típicamente se emplean los mismos suplementos en aproximadamente las mismas cantidades, pero la composición del medio base diferirá (de DMEM, por ejemplo, del medio descrito en la tabla anterior) y la diferencia o diferencias corresponden a la diferencia o diferencias enseñadas aquí.

En algunas realizaciones, los suplementos incluyen medios acondicionados con Wnt, por ejemplo, medios acondicionados con Wnt-3a.

- 25 En una realización, la célula pluripotente, que incluye, por ejemplo, una célula ES XY o una célula iPS XY que tiene una capacidad reducida para contribuir al programa de desarrollo de la determinación del sexo masculino después de la inyección en un embrión y la transferencia intrauterina a un ratón madre sustituto, se mantiene en un cultivo *in vitro* en un medio que comprende medio base y suplementos, en donde el medio base exhibe una o más de las siguientes características: (a) una osmolalidad de aproximadamente 200 mOsm/kg a menos de aproximadamente 30 329 mOsm/kg; (b) una conductividad de aproximadamente 11 mS/cm a aproximadamente 13 mS/cm; (c) una sal de un metal alcalino y un haluro en una concentración de aproximadamente 50 mM a aproximadamente 110 mM; (d) una concentración de sal de ácido carbónico de aproximadamente 17 mM a aproximadamente 30 mM; (e) una concentración total de sal de haluro de metal alcalino y sal de ácido carbónico de aproximadamente 85 mM a aproximadamente 130 mM; y/o (f) una combinación de dos o más de los mismos. En otras realizaciones, las células pluripotentes y/o totipotentes XY (es decir, células ES XY o células iPS XY) se mantienen en un cultivo *in vitro* en un 35 medio como se describe en el documento WO2011/156723.

- En una realización, el medio base es un DMEM con bajo contenido de sal. En una realización específica, el DMEM bajo en sal tiene una concentración de NaCl de 85-130 mM. En una realización, el medio base es un DMEM de baja osmolalidad. En una realización específica, el DMEM de baja osmolalidad tiene una osmolalidad de 250-310 40 mOsm/kg. En una realización, el medio base es un DMEM de baja conductividad. En una realización específica, el DMEM de baja conductividad tiene una conductividad de 11-13 mS/cm.

- En otras realizaciones, el medio base exhibe una osmolalidad de no más de aproximadamente 320, 310, 300, 290, 280, 275, 270, 260, 250 o 240 mOsm/kg. En una realización, el medio base o el medio que comprende el medio base y los suplementos exhiben una osmolalidad de no más de aproximadamente 240-320, 250-310, 275-295, o 45 260-300 mOsm/kg. En una realización específica, el medio base o el medio que comprende el medio base y los suplementos exhiben una osmolalidad de aproximadamente 270 mOsm/kg.

- En otras realizaciones, el medio base exhibe una conductividad de no más de aproximadamente 10,0; 10,5; 11,0; 11,5; 12,0; 12,5; 13,0; 13,5 o 14,0 mS/cm. En una realización, el medio base exhibe una conductividad de no más de aproximadamente 10-14 mS/cm o 11-13 mS/cm. En una realización específica, el medio base exhibe una 50 conductividad de aproximadamente 12-13 mS/cm.

- En una realización específica, el medio base exhibe una conductividad de aproximadamente 12-13 mS/cm y una osmolalidad de aproximadamente 260-300 mOsm/kg. En una realización específica adicional, el medio base comprende cloruro de sodio a una concentración de aproximadamente 90 mM de NaCl. En una realización específica adicional, la concentración de cloruro de sodio es aproximadamente 70-95 mM. En una realización específica adicional, el medio base comprende bicarbonato de sodio a una concentración de menos de 55 aproximadamente 35 mM. En una realización específica adicional, la concentración de bicarbonato de sodio es de aproximadamente 20-30 mM.

En una realización, el medio base exhibe una concentración de una sal de un metal alcalino y un haluro de no más de aproximadamente 100 mM. En una realización, la sal del metal alcalino y el haluro es NaCl. En una realización, la

concentración de la sal del metal alcalino y el haluro no es superior a 90, 80, 70, 60 o 50 mM. En una realización, la concentración en el medio base de la sal del metal alcalino y el haluro es aproximadamente 60-105, 70-95, u 80-90 mM. En una realización específica, la concentración es de aproximadamente 85 mM.

5 En una realización, el medio base exhibe una concentración de una sal de ácido carbónico. En una realización, la sal de ácido carbónico es una sal de sodio. En una realización, la sal de sodio es bicarbonato de sodio. En una realización, la concentración de sal de ácido carbónico en el medio base no es superior a 40, 35, 30, 25 o 20 mM. En una realización, la concentración de sal de ácido carbónico en el medio base es aproximadamente 10-40, en otra realización aproximadamente 20-30 mM. En una realización específica, la concentración es de aproximadamente 25 o 26 mM. En otras realizaciones más, la concentración de bicarbonato de sodio es aproximadamente 26 mM, aproximadamente 18 mM, aproximadamente 18 mM a aproximadamente 26 mM o aproximadamente 18 mM a aproximadamente 44 mM.

15 En una realización, la suma de la concentración de la sal del metal alcalino y el haluro y la sal de ácido carbónico en el medio base no es mayor que 140, 130, 120, 110, 100, 90 o 80 mM. En una realización, la suma de la concentración de la sal del metal alcalino y el haluro y la sal del ácido carbónico en el medio base es aproximadamente 80-140, 85-130, 90-120, 95-120 o 100-120 mM. En una realización específica, la suma de la concentración de la sal del metal alcalino y el haluro y la sal del ácido carbónico en el medio base es de aproximadamente 115 mM.

20 En una realización, la relación molar de la sal del metal alcalino y el haluro y la sal del ácido carbónico es superior a 2,5. En una realización, la relación es aproximadamente 2,6-4,0, 2,8-3,8, 3-3,6 o 3,2-3,4. En una realización, la relación es 3,3-3,5. En una realización específica, la relación es 3,4.

25 En una realización, el medio base muestra una osmolalidad de aproximadamente 250-310 mOsm/kg, y una concentración de una sal de un metal alcalino y un haluro de aproximadamente 60-105 mM. En una realización adicional, el medio base tiene una concentración de una sal de ácido carbónico de aproximadamente 20-30 mM. En una realización adicional, la suma de las concentraciones de la sal de un metal alcalino y haluro y la sal de ácido carbónico es aproximadamente 80-140 mM. En una realización adicional, la conductividad del medio base es de aproximadamente 12-13 mS/cm.

30 En una realización, el medio base comprende aproximadamente 50 ± 5 mM de NaCl y aproximadamente 26 ± 5 mM de carbonato, con una osmolalidad de aproximadamente 218 ± 22 mOsm/kg. En una realización específica, el medio básico comprende aproximadamente 3 mg/mL de NaCl y 2,2 mg/mL de bicarbonato de sodio, con una osmolalidad de aproximadamente 218 mOsm/kg.

En otra realización, el medio base comprende aproximadamente 87 ± 5 mM de NaCl y aproximadamente 18 ± 5 mM, con una osmolalidad de aproximadamente 261 ± 26 mOsm/kg. En una realización específica, el medio básico comprende aproximadamente 5,1 mg/mL de NaCl y aproximadamente 1,5 mg/mL de bicarbonato de sodio, con una osmolalidad de aproximadamente 261 mOsm/kg.

35 En otra realización, el medio base comprende aproximadamente 110 ± 5 mM de NaCl y aproximadamente 18 ± 5 mM de carbonato, con una osmolalidad de aproximadamente 294 ± 29 mOsm/kg. En una realización específica, el medio base comprende aproximadamente 6,4 mg/mL de NaCl y aproximadamente 1,5 mg/mL de bicarbonato de sodio, con una osmolalidad de aproximadamente 294 mOsm/kg.

40 En otra realización, el medio base exhibe aproximadamente 87 ± 5 mM de NaCl y aproximadamente 26 ± 5 mM de carbonato, con una osmolalidad de aproximadamente 270 ± 27 mOsm/kg. En una realización específica, el medio base exhibe aproximadamente 5,1 mg/mL de NaCl y aproximadamente 2,2 mg/mL de bicarbonato de sodio, con una osmolalidad de aproximadamente 270 mOsm/kg.

45 En otra realización, el medio base comprende aproximadamente 87 ± 5 mM de NaCl, aproximadamente 26 ± 5 mM de carbonato y aproximadamente 86 ± 5 mM de glucosa, con una osmolalidad de aproximadamente 322 ± 32 mOsm/kg. En una realización específica, el medio básico comprende aproximadamente 5,1 mg/mL de NaCl, aproximadamente 2,2 mg/mL de bicarbonato de sodio y aproximadamente 15,5 mg/mL de glucosa, con una osmolalidad de aproximadamente 322 mOsm/kg.

50 Los medios base adicionales que pueden emplearse en los diversos métodos y composiciones descritos aquí incluyen un medio base que comprende 50 ± 5 mM de NaCl y 26 ± 5 mM de carbonato, con una osmolalidad de 218 ± 22 mOsm/kg. En una realización particular, el medio básico comprende aproximadamente 3 mg/mL de NaCl y 2,2 mg/mL de bicarbonato de sodio, con una osmolalidad de aproximadamente 218 mOsm/kg.

En otras realizaciones, el medio base comprende 50 ± 5 mM de NaCl y 26 ± 5 mM de carbonato, con una osmolalidad de 218 ± 22 mOsm/kg. En una realización específica, el medio básico comprende aproximadamente 3 mg/mL de NaCl y 2,2 mg/mL de bicarbonato de sodio, con una osmolalidad de aproximadamente 218 mOsm/kg.

55 En otras realizaciones, los medios DMEM con alto contenido de glucosa (LifeTech) con concentraciones de NaHCO_3 como se describe en el presente documento, incluidos, aproximadamente 44 mM, 26 mM o 18 mM, se

complementaron con aminoácidos no esenciales 0,1 mM, piruvato de sodio 1 mM, 2-mercaptoetanol 0,1 mM, L-glutamina 2 mM, 50 µg/mL de cada uno de penicilina y estreptomina (LifeTech), FBS al 15% (Hyclone) y 2.000 U/mL de LIF (Millipore).

C. Método para realizar modificaciones genéticas escogidas

5 Se pueden usar varios métodos para realizar modificaciones genéticas escogidas que disminuyen el nivel y/o la actividad de la proteína Sry. Por ejemplo, en un caso, la modificación genética escogida emplea un sistema que generará una modificación genética específica mediante un evento de recombinación homóloga. En otros casos, la célula animal puede modificarse utilizando agentes de nucleasa que generan una ruptura de cadena sencilla o doble en una ubicación genómica escogida. La ruptura de la cadena sencilla o doble se repara luego por la vía de unión de extremos no homólogos (NHEJ). Tales sistemas encuentran uso, por ejemplo, en la generación de una pérdida escogida de modificaciones genéticas funcionales. Los métodos no limitantes para generar dicha modificación genética escogida se discuten en detalle en otra parte del presente documento, que incluye, por ejemplo, el uso de plásmidos de direccionamiento, vectores de direccionamiento pequeños (TVEC pequeños) o vectores de direccionamiento grandes. Véase, también, Wang et al. (2013) Cell 153: 910-918, Mandalos et al. (2012) PLOS ONE 7: e45768: 1-9, y Wang et al. (2013) Nat Biotechnol. 31: 530-532.

Se reconoce que en realizaciones específicas, la modificación genética escogida del gen *Sry* y/o la modificación genética escogida de cualquier otro polinucleótido de interés puede ocurrir mientras la célula pluripotente (es decir, la célula ES) se mantiene en los medios de cultivo descritos aquí (por ejemplo, un medio que promueve el desarrollo de las hembras fértiles F0 XY). Alternativamente, la modificación genética escogida del gen *Sry* y/o cualquier otro polinucleótido de interés puede ocurrir mientras la célula pluripotente (es decir, la célula ES) se mantiene en diferentes medios de cultivo, y luego se transfiere a los medios de cultivo descritos aquí (por ejemplo, un medio que promueve el desarrollo de las hembras fértiles F0 XY).

D. Método de cultivo y mantenimiento de una célula pluripotente y/o totipotente en cultivo

Se proporciona un método para mantener o cultivar una célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) en un cultivo *in vitro* en esta divulgación, en el que la célula comprende una modificación que disminuye el nivel y/o la actividad de una proteína Sry y la célula se mantiene en un cultivo *in vitro* en las condiciones descritas en este documento. Dichos métodos para mantener o cultivar una célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) en un cultivo *in vitro* promueven un aumento en el número de animales hembra fértiles F0 XY tras la introducción de las células ES XY de animales no humanos en un embrión huésped y después de la gestación de los embriones huésped.

Aunque cualquier medio divulgado en el presente documento puede emplearse para tales métodos de mantenimiento o cultivo, un ejemplo no limitante incluye el cultivo en un medio que comprende un medio base y suplementos adecuados para mantener o cultivar la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) en cultivo, en donde el medio base o el medio que comprende el medio base y los suplementos exhibe una osmolalidad de aproximadamente 200 mOsm/kg hasta menos de aproximadamente 329 mOsm/kg.

La presente divulgación establece que el medio base o el medio que comprende el medio base y los suplementos exhiben una o más de las siguientes características: una conductividad de aproximadamente 11 mS/cm a aproximadamente 13 mS/cm; una sal de un metal alcalino y un haluro en una concentración de aproximadamente 50 mM a aproximadamente 110 mM; una concentración de sal de ácido carbónico de aproximadamente 17 mM a aproximadamente 30 mM; una concentración total de sal de haluro de metal alcalino y de sal de ácido carbónico de aproximadamente 85 mM a aproximadamente 130 mM; y/o una combinación de dos o más de los mismos.

Esta divulgación establece que el método comprende mantener o cultivar la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) en un medio de cultivo adecuado que comprende un medio base y suplementos, en donde la base medio o el medio que comprende el medio base y los suplementos comprenden una osmolalidad de aproximadamente 240-320 mOsm/kg, una conductividad de aproximadamente 10-14 mS/cm, una concentración de sal de haluro de metal alcalino de aproximadamente 50-105 mM, una sal de concentración de ácido carbónico de 10-40 mM, y/o una concentración combinada de sal de metal alcalino y sal de ácido carbónico de aproximadamente 80-140 mM. Esta divulgación establece que la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) se mantiene en el medio (con suplementos para mantener células ES) durante un período de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 días, o 2 semanas, 3 semanas o 4 semanas antes de la introducción en un embrión huésped. Esta divulgación establece que la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) se mantiene en el medio (medio base bajo en sal con suplementos para mantener las células ES) durante aproximadamente 2-4 semanas antes a la introducción en el embrión huésped.

Esta divulgación establece que la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) se mantiene en un medio con un medio base bajo en sal durante al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 o 12 días, 2 semanas, 3 semanas o 4 semanas antes de introducir la célula del donante en un embrión huésped. En una

realización específica, la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) se mantiene en un medio con un medio base bajo en sal por lo menos 2-4 semanas antes de la introducción de la célula en el embrión huésped.

5 Esta divulgación establece que la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) se mantiene (por ejemplo, congelada) en un medio que promueve hembras F0 fértiles XY y la célula del donante se descongela. en y se mantiene en el medio que promueve las hembras F0 fértiles XY durante al menos 1, 2, 3 o 4 días o más antes de introducir la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) en el embrión huésped. En una realización específica, la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) se pasa al menos una vez en un medio que promueve hembras F0 fértiles XY, la
10 célula se congela en el medio que promueve hembras F0 fértiles XY, y la célula se descongela en un medio que promueve hembras F0 fértiles XY y se cultiva durante 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 días, 2 semanas, 3 semanas, 4 semanas o más antes de la introducción en el embrión huésped.

15 Esta divulgación establece que la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) se mantiene en el medio que promueve hembras F0 fértiles XY durante un período de uno, dos, tres o cuatro días antes de la introducción en un embrión huésped. En una realización, la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) se mantiene en el medio que promueve hembras F0 fértiles XY durante un período de 3 días.

20 Esta divulgación establece que la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) mantiene el medio que promueve a las hembras F0 fértiles XY antes de la introducción en el embrión huésped durante aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 o 12 días, 2 semanas, 3 semanas, o 4 semanas o más. En una realización específica, la célula del donante se mantiene en el medio que promueve hembras F0 fértiles XY durante al menos una semana antes de la introducción en el embrión huésped. En una realización específica, la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) se mantiene en el medio que promueve hembras F0 fértiles XY durante 2 a 4 semanas antes de la introducción en el embrión huésped.

25 Por lo tanto, en esta divulgación se proporciona un método para mantener o cultivar una célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY), en donde la célula se mantiene en condiciones que promueven o favorecen el desarrollo de un animal XY hembra después de la introducción de la célula XY en un embrión huésped y después de la gestación en un huésped hembra adecuado.

30 En un aspecto, se proporciona en esta divulgación un método para mantener o cultivar una célula pluripotente y/o totipotente XY donante (es decir, una célula ES XY o una célula iPS XY) en cultivo, en las condiciones descritas en este documento, en las que luego de la introducción de la célula ES XY donante en un embrión huésped para formar un embrión F0 y la gestación del embrión F0 en un animal adecuado, el embrión F0 se convierte en un animal F0 que es al menos 70%, 75%, 80%, 85 %, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más XY y es una hembra que, al alcanzar la madurez sexual, es fértil.

35 E. Generación de embriones F0 y progenie F1 con una modificación genética específica

Los diversos métodos y composiciones que emplean la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula iPS XY) que tienen un nivel y/o actividad disminuidos de la proteína Sry proporcionados en el presente documento pueden usarse para generar un animal genéticamente modificado. Varios métodos para introducir modificaciones genéticas se describen en detalle en otra parte de este documento.

40 i. Método para elaborar un animal no humano XY hembra fértil en una generación F0

45 En esta descripción se proporciona un método para elaborar un animal no humano XY hembra fértil en una generación F0. Tales métodos comprenden: (a) mantener o cultivar una célula pluripotente y/o totipotente XY de un animal no humana donante (es decir, una célula ES XY o una célula iPS XY) que tiene una modificación que disminuye el nivel y/o la actividad de una proteína Sry en un medio que promueve el desarrollo de células ES hembra fértiles XY; (b) la introducción de la célula pluripotente y/o totipotente XY XY de animal no humano XY (es decir, una célula ES XY o una célula iPS XY) en un embrión huésped; (c) gestar el embrión huésped; y, (d) obtener un animal no humano hembra F0 XY, en donde, al alcanzar la madurez sexual, el animal no humano hembra F0 XY es fértil. En realizaciones específicas, la célula del donante XY animal no humana donante puede comprender al menos una modificación genética escogida adicional en un polinucleótido de interés. Tales modificaciones se
50 discuten en detalle en otra parte de este documento.

Las células ES XY que tienen una modificación que disminuye el nivel y/o la actividad de una proteína Sry se pueden mantener sin un medio con bajo contenido de sal y pueden convertirse en una hembra fértil XY.

55 Esta divulgación establece que el medio que promueve el desarrollo de animales hembra F0 fértiles XY puede comprender un medio con bajo contenido en sal que comprende un medio base y suplementos adecuados para mantener o cultivar células ES de mamíferos no humanos en cultivo, en donde el medio base bajo en sal exhibe una característica que comprende uno o más de los siguientes: una osmolalidad de aproximadamente 200 mOsm/kg a menos de aproximadamente 329 mOsm/kg; una conductividad de aproximadamente 11 mS/cm a aproximadamente

13 mS/cm; una sal de un metal alcalino y un haluro en una concentración de aproximadamente 50 mM a aproximadamente 110 mM; una concentración de sal de ácido carbónico de aproximadamente 17 mM a aproximadamente 30 mM; una concentración total de sal de haluro de metal alcalino y de sal de ácido carbónico de aproximadamente 85 mM a aproximadamente 130 mM; y/o una combinación de dos o más de los mismos.

- 5 Esta divulgación establece que tales métodos para elaborar un animal no humano XY hembra fértil en una generación F0 pueden realizarse utilizando los medios descritos en el presente documento que incluyen, entre otros, (a) un medio base que comprende 50 ± 5 mM de NaCl, 26 ± 5 mM de carbonato y 218 ± 22 mOsm/kg; (b) un medio base que comprende aproximadamente 3 mg/mL de NaCl, 2,2 mg/mL de bicarbonato de sodio y 218 mOsm/kg; (c) un medio base que comprende 87 ± 5 mM de NaCl, 18 ± 5 mM de carbonato y 261 ± 26 mOsm/kg; (d) un medio base que comprende aproximadamente 5,1 mg/mL de NaCl, 1,5 mg/mL de bicarbonato de sodio y 261 mOsm/kg; (e) un medio base que comprende 110 ± 5 mM de NaCl, 18 ± 5 mM de carbonato y 294 ± 29 mOsm/kg; (f) un medio base que comprende aproximadamente 6,4 mg/mL de NaCl, 1,5 mg/mL de bicarbonato de sodio y 294 mOsm/kg; (g) un medio base que comprende 87 ± 5 mM de NaCl, 26 ± 5 mM de carbonato y 270 ± 27 mOsm/kg; (h) un medio base que comprende aproximadamente 5,1 mg/mL de NaCl, 2,2 mg/mL de bicarbonato de sodio y 270 mOsm/kg; (i) un medio base que comprende 87 ± 5 mM de NaCl, 26 ± 5 mM de carbonato, 86 ± 5 mM glucosa y 322 ± 32 mOsm/kg; y/o (j) un medio base que comprende aproximadamente 5,1 mg/mL de NaCl, 2,2 mg/mL de bicarbonato de sodio, 15,5 mg/mL de glucosa y 322 mOsm/kg.

La célula pluripotente y/o totipotente XY genéticamente modificada (es decir, una célula ES XY o una célula iPS XY) que tiene una modificación que disminuye el nivel y/o la actividad de una proteína Sry y se ha cultivado en el medio que promueve el desarrollo de hembras fértiles F0 XY puede implantarse en un embrión huésped. Las células que se han implantado en un embrión huésped se denominan en el presente documento "células donantes". Esta divulgación establece que la célula pluripotente y/o totipotente XY genéticamente modificada (es decir, una célula ES XY o una célula iPS XY) proviene de la misma cepa que el embrión huésped o de una cepa diferente que el embrión huésped. Del mismo modo, la madre sustituta puede ser de la misma cepa que la célula pluripotente y/o totipotente XY genéticamente modificada (es decir, una célula ES XY o una célula iPS XY) y/o el embrión huésped, o la madre sustituta puede ser de una cepa diferente que la célula pluripotente y/o totipotente XY modificada genéticamente (es decir, una célula ES XY o una célula iPS XY) y/o el embrión huésped. Esta divulgación establece que la célula del donante XY se implanta en un embrión huésped XX.

Se puede emplear una variedad de embriones huésped en los métodos y composiciones divulgados en el presente documento. Esta divulgación establece que las células pluripotentes y/o totipotentes XY (es decir, la célula ES XY o la célula iPS XY) que tienen la modificación genética escogida que resulta en un nivel y/o actividad disminuidos de la proteína Sry se introducen en un embrión en etapa de premórula de un organismo correspondiente, por ejemplo, un embrión en etapa de 8 células. Véase, por ejemplo, los documentos US 7.576.259, US 7.659.442, US 7.294.754 y US 2008-0078000 A1. Esta divulgación establece que las células ES del donante pueden implantarse en un embrión huésped en la fase de 2 células, en la fase de 4 células, en la fase de 8 células, en la fase de 16 células, en la fase de 32 células o en la fase de 64 células. En otra realización, el embrión huésped es un blastocisto. Esta divulgación establece que el embrión huésped se encuentra en una etapa seleccionada de un embrión preblastocisto, una etapa de premórula, una etapa de mórula, una etapa de mórula sin compactar y una etapa de mórula compactada. En una realización, cuando se emplea un embrión de ratón, la etapa de embrión huésped se selecciona de una Etapa 1 de Theiler (TS1), una TS2, una TS3, una TS4, una TS5 y una TS6, con referencia a las etapas de Theiler descritas en Theiler (1989) "The House Mouse: Atlas of Mouse Development", Springer-Verlag, Nueva York. Esta divulgación establece que la etapa de Theiler se selecciona de TS1, TS2, TS3 y TS4. Esta divulgación establece que el embrión huésped comprende una zona pelúcida, y la célula del donante es una célula ES XY que se introduce en el embrión huésped a través de un agujero en la zona pelúcida, mientras que en otras realizaciones, el embrión huésped es un embrión sin zona. En otras realizaciones específicas, se agrega el embrión huésped de la etapa de mórula.

Las técnicas de transferencia nuclear también se pueden usar para generar los animales genéticamente modificados de esta descripción. En resumen, los métodos para la transferencia nuclear incluyen las etapas de: (1) enuclear un ovocito; (2) aislar una célula o núcleo donante que se combina con el ovocito enucleado; (3) insertar la célula o el núcleo en el ovocito enucleado para formar una célula reconstituida; (4) implantar la célula reconstituida en el útero de un animal para formar un embrión; y (5) permitir que el embrión se desarrolle. En tales métodos, los ovocitos generalmente se recuperan de animales muertos, aunque también pueden aislarse de oviductos y/o ovarios de animales vivos. Los ovocitos se pueden madurar 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 el núcleo del donante en el ovocito enucleado para formar una célula reconstituida generalmente se realiza mediante microinyección de una célula del donante en la zona pelúcida antes de la fusión. La fusión se puede inducir mediante la aplicación de un pulso eléctrico de corriente continua a través del plano de contacto/fusión (electrofusión), mediante la exposición de las células a sustancias químicas que promueven la fusión, como el polietilenglicol, o mediante 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 por el esperma, niveles crecientes de cationes divalentes en el ovocito y reducción de la fosforilación de proteínas celulares (como los inhibidores de la 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. Véase, por ejemplo, los documentos US20080092249, WO/1999/005266A2, US20040177390, WO/2008/017234A1, y la patente de los Estados Unidos No. 7.612.250.

5 El embrión huésped que comprende las células pluripotentes y/o totipotentes XY genéticamente modificadas (es decir, una célula ES XY o una célula iPS XY) que tiene el nivel y/o actividad disminuidos de la proteína Sry se incuba hasta la etapa de blastocisto y luego se implanta en una madre sustituta para producir un animal F0. Los animales que tienen el locus genómico modificado genéticamente pueden identificarse mediante un ensayo de modificación del alelo (MOA) como se describe en este documento.

10 Esta divulgación establece que el embrión huésped que comprende las células pluripotentes y/o totipotentes XY genéticamente modificadas (es decir, una célula ES XY o una célula iPS XY) que tiene el nivel y/o actividad disminuidos de la proteína Sry se mantiene en un medio que promueve el desarrollo de células ES hembra fértiles XY (es decir, un medio base bajo en sal) durante uno, dos, tres o cuatro días o más antes de la implantación en un huésped adecuado. Tales métodos permiten favorecer la generación de un animal hembra fértil F0.

Esta divulgación establece que el embrión huésped cultivado se implanta en una madre sustituta, y el embrión huésped hospedado se gesta en la madre sustituta.

15 Esta divulgación establece que tras la introducción de las células pluripotentes y/o totipotentes XY de animales no humanos (es decir, una célula ES XY o una célula iPS XY) en un embrión huésped y después de la gestación del embrión huésped, al menos 15%, 20%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97 %, 98%, 99% o 100% de los animales no humanos F0 son hembras XY que, al alcanzar la madurez sexual, el mamífero hembra no humano XY F0 es fértil.

20 En esta descripción se proporciona además un embrión de F0 que comprende una masa celular interna que tiene al menos una célula madre heteróloga que comprende una célula ES XY o una célula iPS XY que tiene una modificación genética escogida que disminuye el nivel y/o la actividad de la proteína Sry.

25 Los diversos métodos descritos en este documento para generar un animal no humano XY hembra fértil en una generación F0 pueden emplear células pluripotentes y/o totipotentes XY (es decir, una célula ES XY o una célula iPS XY) que tienen (1) la modificación genética para reducir el nivel y/o la actividad del polipéptido Sry; y, esta divulgación establece que, (2) una o más modificaciones genéticas escogidas adicionales en un polinucleótido de interés. Como se describe en otra parte del presente documento, al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más modificaciones genéticas escogidas adicionales se pueden hacer en la célula pluripotente y/o totipotente XY (es decir, una célula ES XY o una célula XY iPS). En tales casos, el animal no humano XY hembra fértil F0 puede comprender una o más de estas modificaciones genéticas específicas adicionales.

30 Esta divulgación establece que el animal no humano XY hembra fértil F0 produce 1, 2, 3, 4, 5, 6, 7, 8 o 9 camadas durante su vida útil. Esta divulgación establece que el animal no humano XY hembra fértil F0 produce al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 crías por camada. En una realización, el animal no humano XY hembra fértil F0 produce aproximadamente 4-6 crías por camada. Esta divulgación establece que el animal no humano XY hembra fértil F0 produce 2-6 camadas, en donde cada camada tiene al menos 2, 3, 4, 5 o 6 crías. En una realización, al menos aproximadamente 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 70%, 80%, 85%, 90%, 95% o 100% de las crías son crías hembra fértiles XY.

35 También se proporciona un método para generar una camada de roedores (es decir, un ratón o una camada de rata) en esta divulgación y comprende la introducción de una célula donante pluripotente y/o totipotente XY (es decir, una célula ES del donante XY o una célula iPS donante XY) que tienen el nivel y/o la actividad disminuidos de la proteína Sry preparados de acuerdo con los métodos establecidos en el presente documento en embriones huésped, gestando los embriones en una madre segregada adecuada y obteniendo una progenie F0 que comprende al menos un roedor hembra XY que al alcanzar la madurez sexual es un roedor hembra XY fértil. En una realización, el porcentaje de roedores hembra XY F0 nacidos que al alcanzar la madurez sexual son fértiles es de alrededor del 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60% 70%, 75%, 80%, 85%, 95% o 100%.

Esta descripción establece que la progenie F0 producida a partir de tales métodos son aproximadamente 3%, aproximadamente 10% o más, o aproximadamente 63% o más derivadas de la célula XY donante modificada genéticamente.

40 Los métodos y composiciones proporcionados en esta divulgación permiten al menos 1%, 3%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75% o más de los animales F0 tengan la modificación genética escogida (es decir, la disminución en el nivel de proteína Sry y/o la actividad y/o la modificación genética escogida para un polinucleótido de interés) para transmitir la modificación genética a la progenie F1.

45 Esta divulgación establece que el animal no humano XY hembra de generación F0 y/o el animal no humano XY macho es al menos de 90%, 92%, 94%, 96%, 98%, 99% o 99,8% derivado de la célula del donante. Esta divulgación establece que el animal no humano XY hembra F0 y/o el animal no humano XY macho F0 tiene un color de pelaje que es 100% derivado de la célula del donante.

La presente divulgación establece que el animal XY hembra no humano en la generación F0 es un roedor (es decir, un ratón o una rata) y tiene un color de pelaje 100% derivado de la célula del donante. Esta descripción establece que el animal no humano XY hembra formado en la generación F0 es al menos de 90%, 92%, 94%, 96%, 98% o 99,8% derivado de la célula del donante XY. Esta divulgación establece que el animal XY hembra no humano en la generación F0 es aproximadamente el 100% derivado de la célula del donante. Esta divulgación establece que la contribución de una célula embrionaria huésped al animal XY hembra no humano en la generación F0 se determina mediante un ensayo cuantitativo que es capaz de detectar 1 célula en 2.000 (0,05%), y ningún tejido del animal XY hembra es positivo para la contribución de la célula embrionaria huésped.

ii. Diversos métodos de reproducción de la generación F0 XY fértil hembra

Esta divulgación establece que la generación F0 XY fértil hembra resultante derivada de las células pluripotentes y/o totipotentes XY (es decir, la célula ES XY o la célula iPS XY) que tiene la modificación genética que disminuye el nivel y/o la actividad de la proteína Sry se cruza con un animal para obtener descendientes de la generación F1. Esta divulgación establece que la hembra fértil XY F0 se cruza con un animal de tipo silvestre. Esta divulgación establece que el mamífero hembra no humano XY F0 es fértil cuando se cruza con un ratón de tipo silvestre. Esta divulgación establece que el ratón de tipo silvestre es C57BL/6. La progenie F1 puede genotiparse utilizando cebadores y/o sondas específicos para determinar si la modificación genética escogida que comprende el nivel y/o la actividad disminuidos de la proteína Sry está presente. Además, si las modificaciones genéticas escogidas adicionales estaban presentes en la generación F0, la progenie F1 puede genotiparse utilizando cebadores y/o sondas específicos que determinan si tales modificaciones están presentes. Entonces se puede identificar una progenie F1 apropiada para un uso deseado. Esta divulgación establece que se seleccionan las progenes F1 que carecen de la modificación genética que redujo el nivel y/o la actividad de la proteína Sry. Esta descripción establece que se selecciona la progenie F1 que carece de la modificación genética que redujo el nivel y/o la actividad de la proteína Sry y que comprende al menos una modificación genética escogida adicional.

En un ejemplo no limitativo, después del genotipado con cebadores y/o sondas específicos, los animales F1 son heterocigotos para la modificación genética escogida con el polinucleótido de interés y carecen de la modificación escogida que reduce el nivel y/o la actividad de las proteínas Sry se cruzan entre sí. Tal cruce produce una progenie F2 que es homocigótica para el locus genómico de interés genéticamente modificado y no comprende la modificación genética para reducir los niveles y/o actividad de la proteína Sry.

Además, en esta descripción se proporciona un método para producir un homocigoto de un animal no humano transgénico para una modificación genética escogida en la generación F1. El método comprende (a) cruzar un animal no humano hembra fértil XY F0 con una modificación genética específica que disminuye el nivel y/o la actividad de la proteína Sry con un animal no humano macho XY F0, en donde el animal no humano hembra fértil XY F0 y el animal no humano macho XY F0 son heterocigotos para la misma modificación genética de un polinucleótido de interés, y (b) obtener una progenie F1 que sea homocigótica para la modificación genética escogida en el polinucleótido de interés. Esta divulgación establece que la progenie F1 seleccionada es homocigótica para la modificación genética escogida en el polinucleótido de interés y carece de la modificación genética escogida que disminuye la actividad y/o el nivel de la proteína Sry. Dichos métodos pueden emplearse para desarrollar parejas reproductoras de animales no humanos, cada uno completamente derivado de una célula ES del donante o célula iPS, en la misma generación F0.

Se pueden emplear diversos métodos para obtener los animales F0 descritos anteriormente. Esta divulgación establece que se aísla un clon de células XY con una modificación escogida en un polinucleótido de interés en cualquier cromosoma. Se reconoce que se pueden usar varios métodos para generar la modificación escogida en el polinucleótido de interés. En una segunda etapa, se introduce una modificación escogida en el gen Sry de manera que la modificación disminuye el nivel y/o la actividad de la proteína Sry. Tales métodos emplearán además el cultivo de la célula ES XY en un medio que promueva el desarrollo de hembras fértiles F0 XY, como se describe detalladamente en otra parte del presente documento. Los métodos de modificación escogida del gen Sry se divulgan en detalle en otro lugar del presente documento y pueden comprender, por ejemplo, el uso de un vector de direccionamiento (incluido un LTVEC) ya sea solo o en combinación con una nucleasa como se describe en otro lugar en el presente documento (es decir, un sistema Talen o CRISPR o ZFN). Se aísla un subclón que comprende la primera modificación escogida en el polinucleótido de interés y la segunda modificación escogida del gen Sry que disminuye el nivel y/o la actividad de la proteína Sry. Tanto el clon XY original con la modificación escogida en el polinucleótido de interés como el subclón XY que comprende tanto la modificación escogida en el gen Sry como el polinucleótido de interés se introducen en embriones huésped no humanos separados, como se explica en otra parte en este documento. La divulgación establece que, los embriones huésped no humanos comprenden un embrión premórmula (es decir, un embrión en estado de 8 células). Cada uno de los embriones huésped no humanos que comprenden las células pluripotentes modificadas se introduce en una madre sustituta para la gestación. Cada una de las madres sustitutas produce una progenie F0 que comprende la modificación del genoma objetivo (es decir, un macho XY F0 que tiene la modificación escogida en el polinucleótido de interés y una hembra fértil XY F0 que tiene la modificación escogida en el polinucleótido de interés y la modificación genética que disminuye el nivel y/o la actividad de la proteína Sry). La descripción establece que cada una de las modificaciones genómicas escogidas es capaz de ser transmitida a través de la línea germinal. Cada uno de estos animales F0 se reproducen entre sí, para generar un animal F1 que comprende una modificación escogida homocigótica en el polinucleótido de interés. Se

espera que una cuarta parte de la generación F1 sea homocigótica para la modificación escogida en el polinucleótido de interés. La progenie F1 puede seleccionarse para retener la modificación escogida para el gen *Sry* o la progenie F1 puede seleccionarse para no retener la modificación escogida para el gen *Sry*.

5 En otra realización, la introducción de la modificación escogida del gen *Sry* empleando un vector de direccionamiento (y, en realizaciones específicas, nucleasas como Talen, Crispr o Zfn) puede ocurrir simultáneamente con el direccionamiento del vector para la modificación genética del polinucleótido de interés. Dichos métodos permiten la generación de una célula ES XY que tiene una modificación genética que disminuye el nivel y/o la actividad de la proteína *Sry* y además comprende la modificación escogida al polinucleótido de interés.

10 La divulgación describe que la progenie de la generación F1 comprende un genoma completamente derivado de la célula ES del donante. En otras realizaciones, la frecuencia de cruzamientos de ratones macho de generación F0 y hembra de generación F0 que dan lugar a ratones derivados completamente de células ES es del 100%.

II. Métodos y composiciones para modificar un locus genómico objetivo desafiante o un locus genómico objetivo en el cromosoma Y

15 Se proporcionan métodos y composiciones que permiten modificar un locus genómico objetivo en el cromosoma Y en una célula. Además, en esta divulgación se proporcionan métodos que permiten modificar un locus genómico "desafiante". El término "locus desafiante" incluye una región cromosómica que es difícil de identificar mediante las estrategias convencionales de selección de genes. Dichos loci pueden ubicarse en el cromosoma Y, el cromosoma X o un autosoma. Esta divulgación establece que los loci de ataque están ubicados dentro o cerca de regiones cromosómicas pobres en genes, ricas en repeticiones y/o en gran parte heterocromáticas. Véase, por ejemplo, Bernardini et al., Proc. Natl Acad Sci. USA 111: 7600-7605 (2014). Esta divulgación establece que, un locus desafiante se encuentra dentro o en la proximidad a regiones cromosómicas en las que la accesibilidad del ADN cromosómico está limitada por la estructura de la cromatina. Esta divulgación establece que un locus desafiante se encuentra dentro o cerca de las regiones cromosómicas caracterizadas por un alto porcentaje de heterocromatina, tal como al menos aproximadamente -20%, al menos aproximadamente -30%, al menos aproximadamente 40%, al menos aproximadamente 50%, al menos aproximadamente el 60%, o al menos aproximadamente el 70% de heterocromatina. Esta divulgación establece que un locus desafiante se ubica dentro o cerca de regiones cromosómicas que han sufrido duplicaciones y reordenamientos o que se caracterizan por la presencia de repeticiones o repeticiones invertidas. Véase, por ejemplo, Gubbay et al., Proc. Natl Acad Sci. USA 89: 7953-7957 (1992).

30 El término "cromatina" incluye complejos de nucleoproteínas que compactan y organizan el material genético celular para contenerlo dentro de las células. El término "heterocromatina" incluye regiones en el genoma que se encuentran en un estado altamente condensado y generalmente son silenciosas transcripcionalmente. La heterocromatina generalmente se enrolla más estrechamente y generalmente tiene más secuencias de ADN repetitivas que la eucromatina. El término "eucromatina" incluye regiones en el genoma caracterizadas por dominios de cromatina más extensos y menos condensados que a menudo son transcripcionalmente activos y accesibles.

35 El término "exposición" incluye el uso de cualquier método por el cual los componentes deseados se ponen en proximidad inmediata o contacto directo.

40 Se proporcionan métodos y composiciones en esta divulgación que permiten modificar un locus genómico objetivo desafiante o un locus genómico objetivo en el cromosoma Y en una célula. Quizás debido a las características estructurales únicas del cromosoma Y, las estrategias convencionales de selección de genes en células madre embrionarias de ratón para generar mutaciones en los genes ligados a Y han tenido un éxito limitado. Por lo tanto, a menudo, la comprensión de las funciones de los genes murinos ligados a Y se limita a la información obtenida de los estudios de ratones que portan eliminaciones espontáneas, inserciones aleatorias de trampas de genes o transgenes autosómicos. Los métodos proporcionados en esta divulgación en el presente documento que permiten el direccionamiento de un locus genómico en el cromosoma Y empleando un vector de direccionamiento en ausencia o en combinación con un agente de nucleasa.

45 Algunos de estos métodos utilizan un vector de direccionamiento pequeño o TVEC pequeño. Un "TVEC pequeño" incluye un vector de direccionamiento que comprende brazos de homología corta. La longitud de un brazo de homología en un TVEC pequeño puede ser de aproximadamente 400-1.000 pb. Un brazo de homología del TVEC pequeño puede ser de cualquier longitud que sea suficiente para promover un evento de recombinación homóloga con un sitio objetivo correspondiente, que incluye, por ejemplo, de aproximadamente 400 pb a aproximadamente 500 pb, de aproximadamente 500 pb a aproximadamente 600 pb, de aproximadamente 600 pb a aproximadamente 700 pb, de alrededor de 700 pb a aproximadamente 800 pb, de aproximadamente 800 pb a aproximadamente 900 pb, o de aproximadamente 900 pb a aproximadamente 1.000 pb. Una longitud preferida de un brazo de homología en un TVEC pequeño es de aproximadamente 700 pb a aproximadamente 800 pb. En otra realización, la suma total de los brazos de homología 5' y 3' del TVEC pequeño es de aproximadamente 0,5 kb, 1 kb, 1,5 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, aproximadamente 0,5 kb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 1,5 kb, aproximadamente 1,5 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 4 kb, aproximadamente 4 kb a

aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 6 kb, aproximadamente 6 kb a aproximadamente 7 kb, aproximadamente 8 kb a aproximadamente 9 kb, o es al menos 10 kb. En tales métodos, la corta longitud de los brazos de homología aumenta la eficacia de focalización en comparación con un vector de direccionamiento con brazos de homología más largos. Debido a la naturaleza del cromosoma Y, que tiene secuencias altamente repetitivas, los brazos cortos de los TVEC pequeños permiten una focalización altamente específica en el cromosoma Y.

Se proporcionan métodos para modificar un locus genómico objetivo en el cromosoma Y en una célula que comprende: (a) proporcionar una célula que comprende un locus genómico objetivo en el cromosoma Y que comprende un sitio de reconocimiento para un agente de nucleasa, (b) introducir en la célula un primer vector de direccionamiento que comprende un primer polinucleótido de inserción flanqueado por un primer y un segundo brazo de homología que corresponde a un primer y un segundo sitio objetivo; y (c) identificar al menos una célula que comprende en su genoma el primer polinucleótido de inserción integrado en el locus genómico objetivo en el cromosoma Y. Esta divulgación establece que la suma total del primer brazo de homología y el segundo brazo de homología del vector de direccionamiento es de aproximadamente 0,5 kb, 1 kb, 1,5 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, aproximadamente 0,5 kb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 1,5 kb, aproximadamente 1,5 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 6 kb, aproximadamente 6 kb a aproximadamente 7 kb, aproximadamente 8 kb a aproximadamente 9 kb, o es al menos 10 kb o al menos 10 kb y menos de 150 kb. En algunas realizaciones, se emplea un TVEC pequeño. En realizaciones específicas, se emplea un LTVEC. Se pueden realizar métodos similares cuando se direcciona a un locus genómico objetivo desafiante. En una realización no limitativa, tales métodos se realizan empleando los medios de cultivo que promueven el desarrollo de hembras fértiles F0 XY divulgadas en el presente documento y generando así animales hembra fértiles F0 XY. En otro caso, los métodos descritos en el presente documento se emplean para producir una modificación genética escogida en el gen Sry, como se discute en otro lugar del presente documento.

Se proporcionan además métodos para modificar un locus genómico objetivo en el cromosoma Y en una célula que comprende: (a) proporcionar una célula que comprende un locus genómico objetivo en el cromosoma Y que comprende un sitio de reconocimiento para un agente de nucleasa, (b) introducción en la célula (i) del agente de nucleasa, en el que el agente de nucleasa induce un corte o rompimiento de la cadena doble en el primer sitio de reconocimiento; y, (ii) un primer vector de direccionamiento que comprende un primer polinucleótido de inserción flanqueado por un primer y un segundo brazo de homología que corresponde a un primer y un segundo sitio objetivo situados en proximidad suficiente del primer sitio de reconocimiento; y (c) identificar al menos una célula que comprende en su genoma el primer polinucleótido de inserción integrado en el locus genómico objetivo en el cromosoma Y. Esta divulgación describe que la suma total del primer brazo de homología y el segundo brazo de homología del vector de direccionamiento es de aproximadamente 0,5 kb, 1 kb, 1,5 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, aproximadamente 0,5 kb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 1,5 kb, aproximadamente 1,5 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 6 kb, aproximadamente 6 kb a aproximadamente 7 kb, aproximadamente 8 kb a aproximadamente 9 kb, o es al menos 10 kb o al menos 10 kb y menos de 150 kb. En algunas realizaciones, se emplea un TVEC pequeño. En realizaciones específicas, se emplea un LTVEC. Se pueden realizar métodos similares cuando se direcciona a un locus genómico objetivo desafiante. En una realización no limitativa, tales métodos se realizan empleando los medios de cultivo que promueven el desarrollo de hembras fértiles F0 XY divulgadas en el presente documento y generando así animales hembra fértiles F0 XY. En otro caso, los métodos descritos en el presente documento se emplean para producir una modificación genética escogida en el gen Sry, como se describe en otro lugar del presente documento.

Se reconoce que los diversos métodos descritos en este documento para generar una modificación escogida en un locus genómico del cromosoma Y (o cualquier locus genómico desafiante) que emplea un vector de direccionamiento, un TVEC pequeño o un LTVEC se pueden realizar en cualquier tipo de célula, y no se limita a una célula pluripotente y/o totipotente XY. Tales tipos de células incluyen, pero no se limitan a, una célula humana, una célula no humana, una célula de mamífero, una célula de mamífero no humano, una célula de roedor, una célula de ratón, una célula de rata, una célula de hámster, una célula de fibroblasto o cualquier otra célula huésped. Dichas células incluyen células pluripotentes, que incluyen, por ejemplo, células madre pluripotentes inducidas (iPS), células madre (ES) embrionarias de ratón, células madre (ES) embrionarias de rata, células (ES) embrionarias humanas o células progenitoras humanas con desarrollo restringido.

Los métodos se divulgan adicionalmente para generar una gran eliminación en el cromosoma Y empleando cualquiera de los diversos agentes de nucleasas proporcionados en el presente documento (por ejemplo, ARNg CRISPR en combinación con Cas9; ZFN o TALEN). Dicha eliminación en el cromosoma Y puede ser una eliminación de una secuencia endógena de ácido nucleico. La eliminación puede variar de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a

aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 600 kb, de aproximadamente 600 kb a aproximadamente 700 kb, de aproximadamente 700 kb a aproximadamente 800 kb, de aproximadamente 800 kb a aproximadamente 900 kb, de aproximadamente 900 kb a aproximadamente 1 Mb, 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 una realización, la eliminación es mayor que 500 kb. En otra realización, la eliminación es de aproximadamente 500 kb a aproximadamente 600 kb. En una realización específica, la eliminación es de aproximadamente 500 kb. Dicha eliminación en el cromosoma Y puede ser una eliminación de cualquier secuencia de ácido nucleico. En una realización, la eliminación comprende un gen que está asociado con la fertilidad/infertilidad. La eliminación en el cromosoma Y puede comprender una eliminación de múltiples genes. En tales métodos, se pueden eliminar 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más genes. En realizaciones específicas, el gen Kdm5d (demetilasa 5d específica de Lisina (K); por ejemplo, Entrez Gene ID 20592 (mus musculus)) y/o el gen Usp9y (peptidasa 9 específica de ubiquitina, ligada a y; por ejemplo, Entrez Gene ID 107868 (mus musculus)) está destinado para su eliminación. En otras realizaciones, el gen Sry está destinado a la eliminación.

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

El término "sitio de reconocimiento para un agente de nucleasa" incluye una secuencia de ADN en la que se induce un corte o rompimiento de doble cadena por un agente de nucleasa. El sitio de reconocimiento para un agente de nucleasa puede ser endógeno (o nativo) a la célula o el sitio de reconocimiento puede ser exógeno a la célula. En realizaciones específicas, el sitio de reconocimiento es exógeno a la célula y, por lo tanto, no está presente de forma natural 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 desea posicionar 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 nativo o endógeno que ocurre solo una vez dentro del genoma. Dicho sitio puede usarse luego para diseñar agentes de nucleasa que producirán un corte o ruptura de cadena doble en el sitio de reconocimiento endógeno.

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

En una realización, cada monómero del agente de nucleasa reconoce un sitio de reconocimiento de al menos 9 nucleótidos. En otras realizaciones, el sitio de reconocimiento tiene una longitud de aproximadamente 9 a aproximadamente 12 nucleótidos, de aproximadamente 12 a aproximadamente 15 nucleótidos, de aproximadamente 15 a aproximadamente 18 nucleótidos, o de aproximadamente 18 a aproximadamente 21 nucleótidos de longitud, y cualquier combinación de dichos subintervalos (por ejemplo, 9-18 nucleótidos). Se reconoce que un agente de nucleasa dado puede unirse al sitio de reconocimiento y escindir ese sitio de unión o, alternativamente, el agente de nucleasa puede unirse a una secuencia que es diferente del sitio de reconocimiento. Además, el término sitio de reconocimiento comprende tanto el sitio de unión del agente de nucleasa como el sitio de corte/escisión independientemente de si el sitio de corte/escisión está dentro o fuera del sitio de unión del agente de nucleasa. En otra variación, la escisión por el agente de nucleasa puede ocurrir en posiciones de nucleótidos inmediatamente opuestas entre sí para producir un corte de extremo romo o, en otros casos, las incisiones pueden escalonarse para producir salientes de una sola cadena, también llamados "extremos adhesivos", que puede ser salientes 5', o salientes 3'.

Cualquier agente de nucleasa que induce un corte o ruptura de cadena doble en un sitio de reconocimiento deseado puede usarse en los métodos y composiciones descritos en este documento. Puede emplearse un agente de nucleasa natural o nativo siempre que el agente de nucleasa induzca un corte o ruptura de cadena doble en un sitio de reconocimiento deseado. Alternativamente, puede emplearse un agente de nucleasa modificado o manipulado. Un "agente de nucleasa manipulado" incluye una nucleasa que está manipulada (modificada o derivada) de su forma nativa para reconocer específicamente e inducir un corte o ruptura de cadena doble en el sitio de reconocimiento deseado. Por lo tanto, un agente de nucleasa manipulado puede derivarse de un agente de nucleasa nativo, natural o puede ser creado o sintetizado artificialmente. La modificación del agente de nucleasa puede ser tan poco como un aminoácido en un agente de escisión de proteínas o un nucleótido en un agente de escisión de ácido nucleico. En algunas realizaciones, la nucleasa modificada genéticamente induce un corte o ruptura de cadena doble en un sitio de reconocimiento, en donde el sitio de reconocimiento no era una secuencia que hubiera sido reconocida por un agente de nucleasa nativo (no manipulado o no modificado). Producir un corte o rompimiento de doble cadena en un sitio de reconocimiento u otro ADN se puede denominar aquí como "cortar" o "escindir" el sitio de reconocimiento u otro ADN.

También se proporcionan variantes activas y fragmentos de los sitios de reconocimiento ejemplificados. Dichas variantes activas pueden comprender al menos 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 % o más de identidad de secuencia con el sitio de reconocimiento dado, en donde las variantes activas retienen la actividad biológica y, por lo tanto, pueden ser reconocidas y escindidas por un agente de

nucleasa de una manera específica de la secuencia. Los ensayos para medir el rompimiento 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®, Frendewey D. et al., *Methods in Enzymology*, 2010, 476: 295-307).

5 En realizaciones específicas, el sitio de reconocimiento se coloca dentro del polinucleótido que codifica el marcador de selección. Dicha posición puede ubicarse dentro de la región de codificación del marcador de selección o dentro de las regiones reguladoras, lo que influye en la expresión del marcador de selección. Por lo tanto, un sitio de reconocimiento del agente de nucleasa se puede ubicar en un intrón del marcador de selección, un promotor, un potenciador, una región reguladora o cualquier región 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.

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

25 Ejemplos de nucleasas TAL adecuadas, y métodos para preparar nucleasas TAL adecuadas, se describen, por ejemplo, en las solicitudes de patente de Estados Unidos Nos. 2011/0239315 A1, 2011/0269234 A1, 2011/0145940 A1, 2003/0232410 A1, 2005/0208489 A1, 2005/0026157 A1, 2005/0064474 A1, 2006/0188987 A1, y 2006/0063231 A1. En varias realizaciones, las nucleasas efectoras TAL están diseñadas para cortar en o cerca de una secuencia de ácido nucleico objetivo en, por ejemplo, un locus de interés o un locus genómico de interés, en donde la secuencia de ácido nucleico objetivo está en o cerca de una secuencia a modificar por un vector de direccionamiento. Las nucleasas TAL adecuadas para uso con los diversos métodos y composiciones proporcionadas en el presente documento incluyen aquellas que están diseñadas específicamente para unirse en o cerca de secuencias de ácido nucleico objetivo para ser modificadas por vectores de direccionamiento como se describe en el presente documento.

30 En una realización, cada monómero de TALEN comprende 33-35 repeticiones de TAL que reconocen un solo par de bases a través de dos residuos hipervariables. En una realización, el agente de nucleasa es una proteína quimérica que comprende un dominio de unión a ADN basado en la repetición de 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 repetición de TAL y un segundo dominio de unión a ADN basado en repetición de TAL, en el que cada uno del primer y el segundo dominio de unión a ADN basado en repetición de TAL está unido operativamente a una nucleasa *FokI*, en la que el primer y el segundo dominio de unión a ADN basado en la repetición de 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 hace una ruptura de doble cadena en una secuencia objetivo.

45 El agente de nucleasa empleado en los diversos métodos y composiciones divulgados en el presente documento puede comprender además una nucleasa de dedo de cinc (ZFN). En una realización, cada monómero de la ZFN comprende 3 o más dominios de unión a ADN basados en dedos de cinc, en donde cada dominio de unión a ADN basado en dedos de cinc se une a un subsitio de 3 pb. En otras realizaciones, el ZFN es una proteína quimérica que comprende un dominio de unión a ADN basado en dedos de cinc operativamente unido a una nucleasa independiente. En una realización, la endonucleasa independiente es una endonucleasa *FokI*. En una realización, el agente 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 la que la primera y la segunda ZFN reconocen dos secuencias de ADN objetivo contiguas en cada una 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 para hacer un rompimiento de doble cadena. Véase, por ejemplo, los documentos US20060246567; US20080182332; US20020081614; US20030021776; WO/2002/057308A2; US20130123484; US20100291048; WO/2011/017293A2; y Gaj et al. (2013) *Trends in Biotechnology*, 31 (7): 397-405.

60 En otra realización más, el agente de nucleasa es una meganucleasa. Las meganucleasas se han clasificado en cuatro familias según los motivos de secuencia conservados, las familias son las familias de cajas LAGLIDADG, GIY-YIG, H-N-H y His-Cys. Estos motivos participan en la coordinación de iones metálicos e hidrólisis de enlaces

fosfodiéster. Las meganucleasas son notables por sus largos sitios de reconocimiento y por tolerar algunos polimorfismos de secuencia en sus sustratos de ADN. Los dominios de meganucleasa, la estructura y la función son conocidos, véase, por ejemplo, Guhan y Muniyappa (2003) *Crit Rev Biochem Mol Biol* 38: 199-248; Lucas et al., (2001) *Nucleic Acids Res* 29: 960-9; Jurica y Stoddard, (1999) *Cell Mol Life Sci* 55: 1304-26; Stoddard, (2006) *Q Rev Biophys* 38: 49-95; y Moure et al., (2002) *Nat Struct Biol* 9: 764. En algunos ejemplos, se utiliza una variante natural y/o una meganucleasa derivada modificada. Se conocen métodos para modificar la cinética, las interacciones del cofactor, la expresión, las condiciones óptimas y/o la especificidad del sitio de reconocimiento, y la detección de la actividad, ver por ejemplo, Epinat et al., (2003) *Nucleic Acids Res* 31: 2952-62; Chevalier et al., (2002) *Mol Cell* 10: 895-905; Gimble et al., (2003) *Mol Biol* 334: 993-1008; Seligman et al., (2002) *Nucleic Acids Res* 30: 3870-9; Sussman et al., (2004) *J Mol Biol* 342: 31-41; Rosen et al., (2006) *Nucleic Acids Res* 34: 4791-800; Chames et al., (2005) *Nucleic Acids Res* 33: e178; Smith et al., (2006) *Nucleic Acids Res* 34: e149; Gruen et al., (2002) *Nucleic Acids Res* 30: e29; Chen y Zhao, (2005) *Nucleic Acids Res* 33: e154; WO2005105989; WO2003078619; WO2006097854; WO2006097853; WO2006097784; y WO2004031346.

Se puede usar cualquier meganucleasa en el presente documento, que incluye, entre otras, I-Scel, I-Scell, I-Scelll, I-ScelIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIP, I-Crel, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TilI, I-PpoI, PI-PspI, F-Scel, F-Scell, F-SuVI, F-TevI, F-TevII, I-Amal, I-Anil, I-Chul, I-Cmoel, I-Cpal, I-Cpall, I-CsmI, I-Cvul, I-CvuAIP, I-DdII, I-DdIII, I-Dirl, I-Dmol, I-Hmul, I-Hmull, I-HsNIP, I-Llal, I-Msol, I-Naal, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-Njal, I-Nsp236IP, I-PakI, I-PbolP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-Scal, I-SexIP, I-SnelP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdelP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbilP, PI-MtuI, PI-MtuHIP, PI-MtuHIIP, PI-Pful, PI-Pfull, PI-Pkol, PI-Pkoll, PI-Rma43812IP, PI-SpBetaIP, PI-Scel, PI-Tful, PI-Tfull, PI-Thyl, PI-TilI, PI-TIilI, o cualquier variante activa o fragmento de las mismas.

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

Los agentes de la nucleasa pueden comprender además endonucleasas de restricción, que incluyen endonucleasas de tipo I, tipo II, tipo III y tipo IV. Las endonucleasas de restricción tipo I y tipo III reconocen sitios de reconocimiento específicos, pero típicamente se escinden en una posición variable del sitio de unión a la nucleasa, que puede estar a cientos de pares de bases del sitio de escisión (sitio de reconocimiento). En los sistemas de tipo II, la actividad de restricción es independiente de cualquier actividad de metilasa, y la escisión 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 los sitios de reconocimiento no palindrómicos y se escinden fuera del sitio de reconocimiento, las enzimas tipo IIb cortan secuencias dos veces con ambos sitios fuera del sitio de reconocimiento, y las enzimas tipo II reconocen un sitio de reconocimiento asimétrico y se escinden en un lado y a una distancia definida de aproximadamente 1-20 nucleótidos desde el sitio de reconocimiento. Las enzimas de restricción de tipo IV se dirigen al ADN metilado. Las enzimas de restricción se describen y clasifican con más detalle, por ejemplo, en la base de datos REBASE (página web en rebase.neb.com; Roberts et al., (2003) *Nucleic Acids Res* 31: 418-20), Roberts et al., (2003) *Nucleic Acids Res* 31: 1805-12, y Belfort et al., (2002) en *Mobile DNA II*, páginas 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 está optimizada por codones para el tipo de célula deseado en el que se va a expresar. El sistema emplea además un constructo ARNcr-ARNcrtra fusionado que funciona con el Cas9 optimizado por codones. Este ARN único a menudo se conoce como un 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 ARNcrtra se conoce a menudo como el "andamio". Se ha demostrado que este sistema funciona en una variedad de células eucariotas y procariotas. En resumen, un fragmento de ADN corto que contiene la secuencia objetivo se inserta en un plásmido de expresión de ARN guía. El plásmido de expresión de ARNg comprende la secuencia objetivo (en algunas realizaciones alrededor de 20 nucleótidos), una forma de la secuencia de ARNcrtra (el andamio), así como un promotor adecuado que es activo en la célula y elementos necesarios para el procesamiento adecuado en células eucarióticas. Muchos de los sistemas se basan en oligos complementarios personalizados que se hibridan para formar un ADN de doble cadena y luego se clonan en el plásmido de expresión de ARNg. El casete de expresión de ARNg y el casete de expresión de Cas9 se introducen luego en la célula. Véase, por ejemplo, Mali P et al., (2013) *Science*. Febrero 15 de 2013; 339 (6121): 823-6; Jinek M et al., *Science* 17 de agosto de 2012; 337 (6096): 816-21; Hwang WY et al., *Nat Biotechnol* Marzo de 2013 31 (3): 227-9; Jiang W et al., *Nat Biotechnol* Marzo de 2013; 31 (3): 233-9; y, Cong L et al., *Science* 15 de febrero de v; 339 (6121): 819-23.

Los métodos y composiciones descritos en este documento pueden utilizar sistemas o componentes (Cas) asociados a CRISPR/ repeticiones palindrómicas cortas agrupadas y regularmente espaciadas (CRISPR) de tales sistemas para modificar un genoma dentro de una célula. Los sistemas CRISPR/Cas incluyen transcritos y otros elementos involucrados en la expresión o la 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 este documento emplean sistemas CRISPR/Cas utilizando complejos CRISPR (que comprenden un ARN guía (ARNg) complejado con una proteína Cas) para la escisión de ácidos nucleicos dirigida al sitio.

5 Algunos sistemas CRISPR/Cas utilizados en los métodos descritos en la presente memoria no ocurren de forma natural. Un sistema "no natural" incluye cualquier cosa que indique la participación de la mano del hombre, tal como uno o más componentes del sistema se alteren o muten de su estado natural, quedando al menos sustancialmente libres de al menos otro componente con que están naturalmente asociados en la naturaleza, o están asociados con al menos otro componente con el que no están asociados naturalmente. Por ejemplo, algunos sistemas CRISPR/Cas emplean complejos CRISPR de origen no natural que comprenden un ARNg y una proteína Cas que
10 no ocurren juntos de forma natural.

(i) A. 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, descritos con más detalle a continuación). Las proteínas Cas también pueden comprender dominios de nucleasa (por ejemplo, dominios de DNasa o RNasa), dominios de unión a
15 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 del ácido nucleico. La escisión incluye la ruptura de los enlaces covalentes de una molécula de ácido nucleico. La escisión puede producir extremos romos o extremos escalonados, y puede ser de cadena sencilla o de cadena doble.

Los ejemplos de proteínas Cas incluyen Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e (CasD), Cas6, Cas6e, Cas6f, Cas7a1, Cas8a2, Cas8b, Cas8c, Cas9 (Csn1 o Csx12), Cas10, Cas1Od, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (CasA), Cse2 (CasB), Cse3 (CasE), Cse4 (CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm4, Csm2, Csm2 Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 y Cu1966, y homólogos o versiones modificadas de los mismos.
20

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 pseudomycooides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microscilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas sp.*, *Crocospaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor beccsii*, *Candidatus Desulfuridis*, *Clostridium botulinum*, *Clostridium difficile*, *Fingoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira sp.*, *Lyngbya sp.*, *Microcoleus chthonoplastes*, *Oscillatoria sp.*, *Petrotoga mobilis*, *Thermosiphon africanus*, o *Acaryochloris marina*. Los ejemplos adicionales de los miembros de la familia Cas9 se describen en el documento WO 2014/131833. La proteína cas9 de *S. pyogenes* o derivado del mismo es una enzima preferida. A la proteína cas9 de *S. pyogenes* se le asignó el número de acceso de SwissProt Q99ZW2.
25
30
35
40

Las proteínas Cas pueden ser proteínas de tipo silvestre (es decir, aquellas que se producen en la naturaleza), proteínas Cas modificadas (es decir, variantes de la proteína Cas) o fragmentos de proteínas Cas de tipo silvestre o proteínas Cas modificadas. Las proteínas Cas también pueden ser variantes activas o fragmentos de proteínas Cas de tipo silvestre o modificadas. Las variantes o fragmentos activos pueden comprender al menos 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más de identidad de secuencia con el tipo silvestre o proteína Cas modificada o una porción de la misma, en la que las variantes activas conservan la capacidad de cortar en un sitio de escisión deseado y, por lo tanto, retienen la actividad inductora de corte o de rompimiento de doble cadena. Los ensayos para la actividad de inducción del corte o de inducción del rompimiento de doble cadena son conocidos y generalmente miden la actividad general y la especificidad de la proteína Cas en sustratos de ADN que contienen el sitio de escisión.
45
50

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 pueden modificarse para cambiar cualquier otra actividad o propiedad de la proteína, tal como la estabilidad. Por ejemplo, uno o más dominios de nucleasa de la proteína Cas pueden modificarse, eliminarse o inactivarse, o una proteína Cas puede truncarse para eliminar los dominios que no son esenciales para la función de la proteína o para optimizar (por ejemplo, mejorar o reducir) la actividad de la proteína Cas.
55

Algunas proteínas Cas comprenden al menos dos dominios de nucleasa, tales como dominios de DNasa. Por ejemplo, una proteína Cas9 puede comprender un dominio de nucleasa de tipo RuvC y un dominio de nucleasa de

tipo HNH. Los dominios RuvC y HNH pueden cortar cada uno una cadena diferente de ADN de doble cadena para hacer una ruptura de doble cadena en el ADN. Véase, por ejemplo, Jinek et al.. (2012) Science 337: 816-821.

Uno o ambos de los dominios de nucleasa pueden eliminarse o mutarse de modo que ya no sean funcionales o tengan una actividad de nucleasa reducida. Si uno de los dominios de la nucleasa se elimina o se muta, la proteína Cas resultante (por ejemplo, Cas9) puede denominarse como una nickasa y puede generar un rompimiento de una sola cadena en una secuencia de reconocimiento de ARN de CRISPR dentro de un ADN de doble cadena pero no un rompimiento de doble cadena (es decir, puede romper la cadena complementaria o la cadena no complementaria, pero no ambas). Si los dos dominios de nucleasas se eliminan o mutan, la proteína Cas resultante (por ejemplo, Cas9) tendrá una capacidad reducida para dividir ambas cadenas de un ADN de doble cadena. Un ejemplo de una mutación que convierte Cas9 en una nickasa es una mutación D10A (aspartato por alanina en la posición 10 de Cas9) en el dominio de RuvC de Cas9 de *S. pyogenes*. Asimismo, H939A (histidina por alanina en la posición del aminoácido 839) o H840A (histidina por alanina en la posición del aminoácido 840) en el dominio HNH de Cas9 de *S. pyogenes* puede convertir la Cas9 en una nickasa. Otros ejemplos de mutaciones que convierten Cas9 en una nickasa incluyen las mutaciones correspondientes a Cas9 de *S. thermophilus*. Véase, por ejemplo, Sapranaukas et al., (2011) Nucleic Acids Research 39: 9275-9282 y el documento WO 2013/141680. Dichas mutaciones pueden generarse utilizando métodos como la mutagénesis dirigida al sitio, mutagénesis mediada por PCR o síntesis total de genes. Ejemplos de otras mutaciones que crean nickasas se pueden encontrar, 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. Véase el documento WO 2014/089290. Las proteínas Cas también se pueden fusionar con un polipéptido heterólogo que proporciona una mayor o menor estabilidad. El dominio fusionado o polipéptido heterólogo puede estar ubicado en el extremo N, el extremo C o internamente dentro de la proteína Cas.

Una proteína Cas puede fusionarse con un polipéptido heterólogo que proporciona una localización subcelular. Tales péptidos heterólogos incluyen, por ejemplo, una señal de localización nuclear (NLS) tal como el NLS de SV40 para dirigirse al núcleo, una señal de localización mitocondrial para dirigirse a la mitocondria, una señal de retención de ER y similares. Véase, por ejemplo, Lange et al., (2007) J. Biol. Chem. 282: 5101-5105. Dichas señales de localización subcelular pueden ubicarse en el extremo N, el extremo C 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 pueden estar unidas 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 que penetra en las células del virus del Herpes simple, o secuencia peptídica de poliarginina. Véase, por ejemplo, el documento WO 2014/089290. El dominio de penetración celular puede estar ubicado en el extremo N, el extremo C o en cualquier lugar dentro de la proteína Cas.

Las proteínas Cas también pueden comprender un polipéptido heterólogo para facilitar el seguimiento o la purificación, 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, DsRed-Monómero, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monómero, HcRed-Tándem, HcRed1 eqFP611, mRaspberry, mStrawberry, Jred), proteínas fluorescentes anaranjadas (m-Orange, 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, tiorredoxina (TRX), poli (NANP), etiqueta de purificación por afinidad en tándem (TAP), myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, hemaglutinina (HA), nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, histidina (His) proteína portadora de biotina carboxilo (BCCP) y calmodulina.

Las proteínas Cas pueden proporcionarse en cualquier forma. Por ejemplo, una proteína Cas puede proporcionarse en forma de una proteína, tal como una proteína Cas complejada con un ARNg. Alternativamente, puede proporcionarse una proteína Cas en forma de un ácido nucleico que codifica la proteína Cas, tal como un ARN (por ejemplo, ARN mensajero (ARNm)) o ADN. Opcionalmente, el ácido nucleico que codifica la proteína Cas puede ser de codón optimizado para una traducción eficiente en 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 de manera operativa a un promotor activo en la célula. Alternativamente, los ácidos nucleicos que codifican proteínas Cas pueden unirse operativamente a un promotor en un constructo de expresión. Los constructos de expresión incluyen cualquier de los constructos de ácido nucleico capaces de dirigir la expresión de un gen u otra secuencia de ácido nucleico de interés (por ejemplo, un gen Cas) y que puede transferir dicha secuencia de ácido nucleico de interés a una célula objetivo. Los promotores que pueden usarse en un constructo

de expresión incluyen, por ejemplo, promotores activos en una célula de rata pluripotente, eucariótica, mamífero, mamífero no humano, humano, roedor, ratón o hámster. Ejemplos de otros promotores se describen en otra parte de este documento.

(ii) B. ARN guía (ARNg)

- 5 Un "ARN guía" o "ARNg" incluye una molécula de ARN que se une a una proteína Cas y dirige a la proteína Cas en una ubicación específica dentro de un ADN objetivo. Los ARN guía 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, tal como un tramo contiguo de nucleótidos en un ARN. Algunos ARNg comprenden dos moléculas de ARN separadas: un "ARN activador" y un "ARN receptor". Otros ARNg son una molécula sencilla de ARN (polinucleótido sencillo de ARN), que también se puede llamar un "ARNg de molécula sencilla", un "ARN guía sencillo" o un "ARNgs". Véase, 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 los ARNg de molécula doble como ARNg de molécula sencilla.

- 15 Un ejemplo de ARNg de dos moléculas comprende una molécula similar a ARNcr ("ARN de CRISPR" o "ARN receptor" o "ARNcr" o "repetición ARNcr") y una molécula correspondiente similar a ARNcrtra ("ARN de CRISPR de activación trans" o "ARN activador" o "ARNcrtra" o "andamio"). Un ARNcr comprende tanto el segmento de direccionamiento de ADN (de cadena sencilla) del ARNg y un tramo de nucleótidos que forma una mitad del dúplex de ARNbc del segmento de unión a la proteína del ARNg.

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

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

- 30 El segmento de direccionamiento de ADN (ARNcr) de un ARNg dado comprende una secuencia de nucleótidos que es complementaria a una secuencia en un ADN objetivo. El segmento de direccionamiento de ADN de un ARNg interactúa con un ADN objetivo de una manera específica de secuencia mediante hibridación (es decir, emparejamiento de bases). Como tal, la secuencia de nucleótidos del segmento de direccionamiento de ADN puede variar y determina la ubicación dentro del ADN objetivo con la que el ARNg y el ADN objetivo interactuarán. El segmento de direccionamiento de ADN de un ARNg sujeto puede modificarse para hibridar 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, flanqueados por dos repeticiones directas (DR) de una longitud de entre 21 y 46 nucleótidos (véase, por ejemplo, el documento WO2014/131833). En el caso de *S.pyogenes*, las DR tienen una longitud de 36 nucleótidos y el segmento de direccionamiento tiene una longitud de 30 nucleótidos. La DR localizada en 3' es complementaria e hibrida con el correspondiente ARNcrtra, que a su vez se une a la proteína Cas9.

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

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

Los ARNcrtra pueden estar en cualquier forma (por ejemplo, ARNcrtra de longitud completa o ARNcrtra parcialmente activos) y de longitudes variables. Pueden incluir transcritos primarios o formas procesados. Por ejemplo, los ARNcrtra (como parte de un ARN guía sencillo o como una molécula separada como parte de un ARNg de dos moléculas) puede comprender o consistir de todo o una porción de una secuencia de ARNcrtra 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 ARNcrtra de tipo silvestre). Ejemplos de secuencias de ARNcrtra de tipo silvestre de *S. pyogenes* incluyen versiones de 171 nucleótidos, 89 nucleótidos, 75 nucleótidos y 65 nucleótidos. Véase, por ejemplo, Deltcheva et al., (2011) Nature 471: 602-607; el documento WO 2014/093661. Los ejemplos de ARNcrtra dentro de un ARN guía sencillo (ARNgs) incluyen los segmentos de ARNcrtra encontrados dentro de las versiones +48, +54, +67 y +85 de ARNgs, donde "+n" indica que hasta el nucleótido +n de ARNcrtra de tipo silvestre está incluido en el ARNgs. Véase la patente de los Estados Unidos No. 8.697.359.

El porcentaje de complementariedad entre la secuencia de direccionamiento de ADN y la secuencia de reconocimiento de ARN de CRISPR dentro del ADN objetivo puede ser al menos del 60% (por ejemplo, al menos del 65%, al menos de 70%, al menos de 75%, al menos de 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 de direccionamiento de ADN y la secuencia de reconocimiento de ARN de CRISPR dentro del ADN objetivo puede ser de al menos de 60% sobre aproximadamente 20 nucleótidos contiguos. Como ejemplo, el porcentaje de complementariedad entre la secuencia de direccionamiento de ADN y la secuencia de reconocimiento del ARN de CRISPR dentro del ADN objetivo es del 100% sobre los 14 nucleótidos contiguos en el extremo 5' de la secuencia de reconocimiento del ARN de 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 de direccionamiento de ADN tiene una longitud de 14 nucleótidos. Como otro ejemplo, el porcentaje de complementariedad entre la secuencia de direccionamiento de ADN y la secuencia de reconocimiento del ARN de CRISPR dentro del ADN objetivo es del 100% sobre los siete nucleótidos contiguos en el extremo 5' de la secuencia de reconocimiento del ARN de 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 de direccionamiento de ADN puede considerarse que es de 7 nucleótidos de longitud.

El segmento de unión a proteínas de un ARNg puede comprender dos tramos de nucleótidos que son complementarios entre sí. Los nucleótidos complementarios del segmento de unión a proteínas se hibridan para formar un dúplex de ARN de doble cadena (ARNbc). El segmento de unión a proteínas de un ARNg sujeto interactúa con una proteína Cas, y el ARNg dirige la proteína Cas unida a una secuencia de nucleótidos específica dentro del ADN objetivo a través del segmento de direccionamiento de 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 de proteínas; y similares). Los ejemplos de dichas modificaciones incluyen, por ejemplo, una tapa 5' (por ejemplo, una tapa de 7-metilguanilato (m7G)); una cola poliadenilada 3' (es decir, una cola poli(A) 3'); una secuencia de riboconmutador (por ejemplo, para permitir una estabilidad regulada y/o accesibilidad regulada por proteínas y/o complejos de proteínas); una secuencia de control de estabilidad; una secuencia que forma un dúplex de ARNbc (es decir, una horquilla); una modificación o secuencia que dirige el ARN a una ubicación subcelular (por ejemplo, núcleo, mitocondrias, cloroplastos y similares); una modificación o secuencia que proporciona un seguimiento (por ejemplo, conjugación directa con una molécula fluorescente, conjugación con un resto que facilita la detección fluorescente, una secuencia que permite la detección

fluorescente, etc.); una modificación o secuencia que proporciona un sitio de unión para proteínas (por ejemplo, proteínas que actúan sobre el ADN, incluidos activadores transcripcionales, represores transcripcionales, ADN metiltransferasas, ADN desmetilasas, histonas acetiltransferasas, histonas desacetilasas y similares); y combinaciones de los mismos.

5 Los ARN guía pueden proporcionarse en cualquier forma. Por ejemplo, el ARNg puede proporcionarse en forma de ARN, ya sea como dos moléculas (ARNcr y ARNcrtra separados) o como una molécula (ARNgs), 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 única molécula de ARN sencilla (ARNgs) o moléculas de ARN separadas (por ejemplo, ARNcr y ARNcrtra separados). En este último caso, el ADN que codifica el ARNg puede proporcionarse como moléculas de ADN separadas que codifican el ARNcr y el ARNcrtra, respectivamente.

10 Los ADN que codifican los ARNg pueden integrarse de forma estable en el genoma de la célula y unirse operativamente a un promotor activo en la célula. Alternativamente, los ADN que codifican los ARNg pueden unirse operativamente a un promotor en un constructo de expresión. Dichos promotores pueden ser activos, por ejemplo, en una célula pluripotente de rata, eucariótica, 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 este documento.

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

(iii) C. Secuencias de reconocimiento de ARN de CRISPR

25 El término "secuencia de reconocimiento de ARN de CRISPR" incluye secuencias de ácido nucleico presentes en un ADN objetivo al que se unirá un segmento de direccionamiento de ADN de un ARNg, siempre que existan condiciones suficientes para la unión. Por ejemplo, las secuencias de reconocimiento de ARN de CRISPR incluyen secuencias en las que un ARN guía está diseñado para tener complementariedad, donde la hibridación entre una secuencia de reconocimiento de ARN de CRISPR y una secuencia de direccionamiento de ADN promueve la formación de un complejo CRISPR. No se requiere necesariamente una completa complementariedad, siempre que haya suficiente complementariedad para causar la hibridación y promover la formación de un complejo de CRISPR. Las secuencias de reconocimiento de ARN de CRISPR también incluyen sitios de escisión para proteínas Cas, que se describen con más detalle a continuación. Una secuencia de reconocimiento de ARN de CRISPR puede comprender cualquier polinucleótido, que puede ubicarse, por ejemplo, en el núcleo o citoplasma de una célula o dentro de un orgánulo de una célula, tal como una mitocondria o un cloroplasto.

35 La secuencia de reconocimiento de ARN de CRISPR dentro de un ADN objetivo puede ser dirigida por (es decir, estar unida por, o hibridar con, o ser complementaria de) una proteína Cas o un ARNg. Las condiciones de unión adecuadas a ADN/ARN incluyen condiciones fisiológicas normalmente presentes en una célula. Otras condiciones de unión adecuadas a ADN/ARN (por ejemplo, condiciones en un sistema libre de células) son conocidas en la técnica (véase, por ejemplo, *Molecular Cloning: A Laboratory Manual*, tercera edición (Sambrook et al., Harbor Laboratory Press 2001)). La cadena del ADN objetivo que es complementaria e hibrida con la proteína Cas o el ARNg puede denominarse "cadena complementaria" y la cadena del ADN objetivo que es complementaria a la "cadena complementaria" (y por lo tanto no es complementaria a proteína Cas o ARNg) puede denominarse "cadena no complementaria" o "cadena de plantilla".

45 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 de direccionamiento de 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 una sola cadena o una ruptura de una doble cadena. Por ejemplo, la formación de un complejo de CRISPR (que comprende un ARNg hibridado a una secuencia de reconocimiento de ARN de CRISPR y complejado con una proteína Cas) puede dar como resultado la escisión de una o ambas cadenas en o cerca de (por ejemplo, dentro de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50 o más pares de bases de) la secuencia de ácido nucleico presente en un ADN objetivo al que se unirá el segmento de direccionamiento de ADN de un ARNg. Si el sitio de escisión está fuera de la secuencia de ácido nucleico a la que se unirá el segmento de direccionamiento del ADN del ARNg, el sitio de escisión todavía se considera que está dentro de la "secuencia de reconocimiento de ARN de 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 (que producen extremos romos) o pueden estar en sitios diferentes en cada cadena (que producen extremos escalonados). Se pueden producir extremos escalonados, por ejemplo, utilizando dos proteínas Cas, cada una de las cuales produce una ruptura de una sola cadena en un sitio de escisión diferente en cada cadena, lo que produce un rompimiento de doble cadena. Por ejemplo, una primera nickasa puede crear una ruptura de una sola cadena en la primera cadena de ADN de doble cadena (ADNbc), y una segunda nickasa puede crear una ruptura de una sola cadena en la segunda cadena de ADNbc de manera que se crean

secuencias sobresalientes. En algunos casos, la secuencia de reconocimiento de ARN de CRISPR de la nickasa en la primera cadena se separa de la secuencia de reconocimiento de ARN de CRISPR de la nickasa en la segunda cadena por al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500 o 1.000 pares de bases.

5 La escisión específica del sitio del ADN objetivo por Cas9 puede ocurrir en ubicaciones determinadas por (i) complementariedad de emparejamiento de bases entre el ARNg y el ADN objetivo y (ii) un motivo corto, llamado el motivo adyacente protoespaciador (PAM), en el ADN objetivo. El PAM puede flanquear la secuencia de reconocimiento de ARN de CRISPR. Opcionalmente, la secuencia de reconocimiento de ARN de 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) secuencia arriba o secuencia abajo de la secuencia de PAM. En algunos casos (por ejemplo, cuando se usa Cas9 de *S. pyogenes* o un Cas9 estrechamente relacionado), la secuencia de PAM de la cadena no complementaria puede ser 5'-N₁GG-3', donde N₁ es cualquier nucleótido de ADN y es inmediatamente 3' de la secuencia de reconocimiento de ARN de CRISPR de la cadena no complementaria del ADN objetivo. Como tal, la secuencia de PAM de la cadena complementaria sería 5'-CC N₂-3', donde N₂ es cualquier nucleótido de ADN y es inmediatamente 5' de la secuencia de reconocimiento de ARN de 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).

20 Los ejemplos de secuencias de reconocimiento de ARN de CRISPR incluyen una secuencia de ADN complementaria al segmento de direccionamiento de ADN de un ARNg, o tal secuencia de ADN además de una secuencia de 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 de CRISPR pueden incluir dos nucleótidos de guanina en el extremo 5' (por ejemplo, GGN₂₀NGG; SEQ ID NO: 9) para facilitar la transcripción eficiente por la polimerasa T7 *in vitro*. Véase, por ejemplo, el documento WO 2014/065596.

La secuencia de reconocimiento de ARN de CRISPR puede ser cualquier secuencia endógena o exógena de ácido nucleico o a una célula. La secuencia de reconocimiento de ARN de CRISPR puede ser una secuencia que codifica un producto génico (por ejemplo, una proteína) o una secuencia no codificante (por ejemplo, una secuencia reguladora) o puede incluir ambas.

30 En una realización, la secuencia objetivo está inmediatamente flanqueada por una secuencia de motivo adyacente protoespaciador (PAM). En una realización, el locus de interés comprende la secuencia de nucleótidos de la SEQ ID NO: 1. En una realización, el ARNg comprende una tercera secuencia de ácido nucleico que codifica un ARN de Repeticiones Palindrómicas Cortas agrupadas y regularmente espaciadas (CRISPR) (ARNcr) y un ARN de CRISPR transactivador (ARNcrtra). 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 la SEQ ID NO: 2; o (b) el ARN quimérico de la secuencia de ácido nucleico de la SEQ ID NO: 3. En algunos de estos métodos, el ARNcr comprende la secuencia expuesta en las SEQ ID NO: 4, SEQ ID NO: 5 o SEQ ID NO: 6. En algunos de estos métodos, el ARNcrtra comprende la secuencia expuesta en la SEQ ID NO: 7 o la SEQ ID NO: 8.

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

55 El agente de nucleasa se puede introducir en la célula por cualquier medio conocido en la técnica. El polipéptido que codifica el agente de nucleasa se puede introducir directamente en la célula. Alternativamente, un polinucleótido que codifica el agente de nucleasa se puede introducir en la célula. Cuando un polinucleótido que codifica el agente de nucleasa se introduce 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 discuten con más detalle en otra parte del presente documento. 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 está integrado de manera estable en el genoma de la célula y está unido de forma operativa a un promotor activo en la célula. En otras realizaciones, el polinucleótido que codifica el agente de nucleasa está en el mismo vector de direccionamiento que comprende el 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 direccionamiento que comprende el polinucleótido de inserción.

Cuando se proporciona el agente de nucleasa a la célula a través de la introducción de un polinucleótido que codifica el agente de nucleasa, tal 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 a la secuencia de polinucleótidos 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 procariota o eucariota de interés, incluida una célula bacteriana, una célula de levadura, una célula humana, una célula no humana, una célula de mamífero, una célula de roedor, una célula de ratón, una célula de rata o cualquier otra célula huésped de interés, en comparación con la secuencia de polinucleótidos que se produce de forma natural.

B. Empleo del sistema CRISPR/Cas en combinación con un vector de direccionamiento grande (LTVEC) o un vector de direccionamiento pequeño (TVEC pequeño) para modificar un loci genómico desafiante o un locus de cromosoma Y

Los métodos no limitantes para modificar un locus genómico desafiante o un locus del cromosoma Y comprenden la exposición del cromosoma (es decir, el cromosoma Y) a una proteína Cas y un ARN de CRISPR en presencia de un gran vector de direccionamiento (LTVEC) que comprende una secuencia de ácido nucleico de al menos 10 kb, en donde después de la exposición a la proteína Cas, el ARN de CRISPR y el LTVEC, el cromosoma (es decir, el cromosoma Y) se modifica para contener al menos 10 kb de secuencia de ácido nucleico.

El método de esta divulgación puede emplear cualquiera de los LTVEC o TVEC pequeños descritos en este documento. Esta divulgación describe que el LTVEC o TVEC pequeño comprende una secuencia de ácido nucleico de al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 150 kb o al menos 200 kb. En otras realizaciones, la suma total de los brazos de homología 5' y 3' del LTVEC es de aproximadamente 10 kb a aproximadamente 150 kb, aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 120 kb, o de aproximadamente 120 kb a 150 kb. Esta divulgación describe que la suma total de los brazos de homología 5' y 3' del TVEC pequeño es aproximadamente 0,5 kb, 1 kb, 1,5 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, aproximadamente 0,5 kb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 1,5 kb, aproximadamente 1,5 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 6 kb, aproximadamente 6 kb a aproximadamente 7 kb, aproximadamente 8 kb a aproximadamente 9 kb, o es al menos 10 kb.

También se proporciona un método para modificar un locus objetivo desafiante o un locus genómico objetivo en el cromosoma Y, que comprende: (a) proporcionar una célula de mamífero que comprende el locus objetivo desafiante o un locus genómico objetivo en el cromosoma Y, en donde el locus genómico objetivo comprende una secuencia objetivo de ARN guía (ARNg); (b) introducción en la célula de mamífero: (i) un vector de direccionamiento grande (LTVEC) que comprende un primer ácido nucleico flanqueado con brazos de direccionamiento homólogos al locus genómico objetivo, en el que el LTVEC es de al menos 10 kb; (ii) un primer constructo de expresión que comprende un primer promotor unido operativamente a un segundo ácido nucleico que codifica una proteína Cas, y (iii) un segundo constructo de expresión que comprende un segundo promotor unido operativamente a un tercer ácido nucleico que codifica un ARN guía (ARNg) que comprende una secuencia de nucleótidos que se hibrida con la secuencia objetivo de ARNg y un ARN de CRISPR transactivador (ARNcrtra), en donde el primero y el segundo promotores son activos en la célula de mamífero; y, (c) identificar una célula de mamífero modificada que comprende una modificación genética escogida en el locus genómico objetivo desafiante o en el locus genómico objetivo en el cromosoma Y. En realizaciones específicas, el primer y segundo constructo de expresión están en una única molécula de ácido nucleico. En otras realizaciones, el locus genómico objetivo de los cromosomas Y es el locus *Sry*.

Como se describió anteriormente, en una realización, la proteína Cas puede comprender una proteína Cas9. En otra realización, la secuencia objetivo de ARNg está inmediatamente flanqueada por una secuencia de un motivo adyacente protoespaciador (PAM).

El método de esta descripción puede emplear cualquiera de los LTVEC o TVEC pequeños descritos en este documento. Esta descripción establece que el LTVEC o TVEC pequeño es al menos 0,5 kb, al menos 1 kb, al menos 5 kb, al menos 10 kb, al menos 15 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, al menos 60 kb, al menos 70 kb, al menos 80 kb, al menos 90 kb, al menos 100 kb, al menos 150 kb o al menos 200 kb. En otras realizaciones, la suma total de los brazos de homología 5' y 3' del LTVEC es de aproximadamente 10 kb a

aproximadamente 150 kb, aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 120 kb, o de aproximadamente 120 kb a 150 kb.

- 5 Los diversos métodos que emplean el sistema CRISPR/Cas (o cualquier método descrito aquí) se pueden realizar en, por ejemplo, células de mamífero, células de mamífero no humano, células de fibroblasto, células de roedor, células de rata, células de ratón o hámster. La célula puede ser una célula pluripotente, una célula madre pluripotente inducida (iPS), una célula madre (ES) embrionaria de ratón, una célula madre (ES) embrionaria de rata, una célula madre (ES) embrionaria humana o una célula progenitora humana con desarrollo restringido.
- 10 Como se explica en detalle a continuación, después de la modificación de un locus genómico desafiante o un locus genómico de interés en el cromosoma Y (es decir, el locus *Sry*) de una célula pluripotente no humana que emplea, por ejemplo, el esquema CRISPR/CAS expuesto anteriormente, la célula pluripotente no humana genéticamente modificada que se produce puede introducirse en un embrión huésped no humano; y se gesta el embrión huésped no humano que comprende la célula pluripotente modificada en una madre sustituta. La madre sustituta produce una
- 15 progenie F0 que comprende la modificación genética escogida. En realizaciones específicas, la modificación genética escogida puede transmitirse a través de la línea germinal.

C. Marcadores de selección

- Se pueden usar diversos marcadores de selección en los métodos y composiciones descritos en el presente documento que proporcionan la modificación de un locus genómico objetivo en el cromosoma Y o un locus genómico objetivo desafiante. Dichos marcadores se describen en otra parte del presente documento e incluyen, pero no se limitan a, marcadores de selección que imparten resistencia a un antibiótico como G418, higromicina, blasticidina, neomicina o puromicina. El polinucleótido que codifica los marcadores de selección está unido operativamente a un promotor activo en la célula. Tales casetes de expresión y sus diversos componentes reguladores se analizan con más detalle en otra parte en este documento.
- 20

25 D. Locus genómico objetivo

Se proporcionan varios métodos y composiciones de esta divulgación que permiten la integración de al menos un polinucleótido de inserción en un locus genómico objetivo en el cromosoma Y o un locus genómico objetivo desafiante. Como se usa en este documento, un "locus genómico objetivo en el cromosoma Y" comprende cualquier segmento o región de ADN en el cromosoma Y que se desea integrar a un polinucleótido de inserción.

- 30 El locus genómico en el cromosoma Y o un locus genómico objetivo desafiante al que se está dirigiendo puede ser nativo a la célula, o alternativamente puede comprender un segmento heterólogo o exógeno de ADN que se integró en el cromosoma de la célula. Dichos segmentos de ADN heterólogos o exógenos pueden incluir transgenes, casetes de expresión, un polinucleótido que codifica marcadores de selección, o regiones heterólogas o exógenas de ADN genómico. El locus genómico objetivo en el cromosoma Y o el locus genómico objetivo desafiante puede
- 35 comprender cualquiera de los sistemas de integración genómica objetivo que incluyen, por ejemplo, el sitio de reconocimiento, el marcador de selección, polinucleótidos de inserto previamente integrados, polinucleótidos que codifican agentes de nucleasa, promotores, etc. Alternativamente, el locus genómico objetivo en el cromosoma Y o el locus genómico objetivo desafiante puede ubicarse dentro de un cromosoma artificial de levadura (YAC), un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano o cualquier otra región genómica manipulada contenida en una célula huésped apropiada. Por lo tanto, esta descripción establece que el locus genómico objetivo
- 40 en el cromosoma Y o el locus genómico objetivo desafiante puede comprender una secuencia de ácido nucleico genómico nativo, heterólogo o exógeno de un mamífero no humano, una célula no humana, un roedor, un humano, una rata, un ratón, un hámster, un conejo, un cerdo, un bovino, 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 agrícola o cualquier otro organismo de interés o una combinación de los mismos.
- 45

Los ejemplos no limitativos del locus genómico objetivo en el cromosoma Y incluyen, el gen *Sry*, el gen *Uty*, el gen *Eif2s3y*, el gen *Ddx3y*, el gen, el gen *Ubely*, el gen *Tspy*, el gen *Usp9y*, el gen *Zfy1* y el gen *Zfy2* y la región del cromosoma Y que abarca los genes *Kdm5d*, *Eif2s3y*, *Tspy*, *Uty*, *Ddx3y* y *Usp9y*. Dicho locus en el cromosoma Y puede ser de un mamífero no humano, un mamífero, un roedor, un humano, una rata, un ratón, un hámster, un

50 conejo, un cerdo, un bovino, un ciervo, una oveja, una cabra, un pollo, un gato, un perro, un hurón, un primate (por ejemplo, un tití, un mono rhesus), un mamífero domesticado o un mamífero agrícola o cualquier otro organismo de interés o una combinación de ellos. Dichas células incluyen células pluripotentes, que incluyen, por ejemplo, células madre pluripotentes inducidas (iPS), células madre (ES) embrionarias de ratón, células madre (ES) embrionarias de rata, células madre (ES) embrionarias humanas o células progenitoras humanas con desarrollo restringido.

- 55 Como se describe en otra parte en este documento, se proporcionan diversos métodos y composiciones que comprenden células pluripotentes y/o totipotentes XY (tales como células ES XY o células iPS) que tienen una actividad o nivel disminuidos de la proteína *Sry*. Los diversos métodos descritos en este documento para modificar el

locus genómico en el cromosoma Y también se pueden usar para introducir modificaciones genéticas escogidas a polinucleótidos de interés que no están localizados en el cromosoma Y.

E. Vectores de direccionamiento y polinucleótidos de inserción

5 Como se describió anteriormente, los métodos y composiciones proporcionados en este documento emplean vectores de direccionamiento solos o en combinación con un agente de nucleasa. La "recombinación homóloga" se usa convencionalmente para referirse al intercambio de fragmentos de ADN entre dos moléculas de ADN en los sitios de cruce dentro de las regiones de homología.

I. Polinucleótido de inserción

10 Como se usa en este documento, el "polinucleótido de inserción" comprende un segmento de ADN que se desea integrar en el locus genómico objetivo. En realizaciones específicas, el locus genómico objetivo está en el cromosoma Y. En otras realizaciones, el locus genómico objetivo es un locus genómico desafiante. 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 marcador de selección y/o un gen informador junto con los diversos componentes reguladores que influyen en la expresión. Los ejemplos no limitantes de polinucleótidos de interés, marcadores de selección y genes informadores que pueden incluirse dentro del polinucleótido de inserción se discuten en detalle en otra parte del presente documento.

20 En realizaciones específicas, el polinucleótido de inserción puede comprender un ácido nucleico genómico. En una realización, el ácido nucleico genómico se deriva de un animal, un ratón, un humano, un no 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 perro, un hurón, un primate (por ejemplo, tití, mono rhesus), un mamífero domesticado o un mamífero agrícola, un ave, o cualquier otro organismo de interés o una combinación de ellos.

25 En otras realizaciones, 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 accionamiento en orientación sentido con respecto a la transcripción de un gen objetivo, y un casete de selección de fármaco en orientación sentido o antisentido; (b) en orientación antisentido una secuencia de nucleótidos de interés (NSI) y un módulo condicional por inversión (COIN, que utiliza un intrón que divide el exón y un módulo invertible de tipo trampa de genes; véase, por ejemplo, el documento US 2011/0104799); y (c) unidades recombinables que se recombinan tras la exposición a una primera recombinasa para formar un alelo condicional que (i) carece de la secuencia de activación y el DSC, y (ii) contiene la NSI en orientación sentido y la COIN en orientación antisentido.

35 El polinucleótido de inserción puede ser de aproximadamente 5 kb a aproximadamente 200 kb, de aproximadamente 5kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 60 kb a aproximadamente 70 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 110 kb, de aproximadamente 120 kb a aproximadamente 130 kb, de aproximadamente 130 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 170 kb, de aproximadamente 170 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 190 kb, de aproximadamente 190 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 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.

45 En realizaciones específicas, el polinucleótido de inserción comprende un ácido nucleico flanqueado con secuencias objetivo de recombinación específica del sitio. Se reconoce que mientras que 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 flanqueado por dichos sitios. El término "sitio de recombinación" como se usa en este documento incluye una secuencia de nucleótidos que es reconocida por una recombinasa específica del sitio y que puede servir como un sustrato para un evento de recombinación. El término "recombinasa específica del sitio" como se usa en el presente documento 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 única molécula de ácido nucleico o en moléculas de ácido nucleico separadas. Los ejemplos de recombinasas específicas del sitio incluyen, pero no se limitan a, recombinasas Cre, Flp y Dre. La recombinasa específica del sitio puede introducirse en la célula por cualquier medio, incluso introduciendo el polipéptido de recombinasa en la célula o introduciendo un polinucleótido que codifica la recombinasa específica del sitio en la célula huésped. El polinucleótido que codifica la recombinasa específica del sitio puede estar ubicado dentro del polinucleótido de inserción o dentro de un polinucleótido separado. La recombinasa específica del sitio puede estar operativamente unida 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 etapa de desarrollo. Las secuencias objetivo de recombinación específicas del sitio que pueden flanquear el polinucleótido de inserción o cualquier polinucleótido de interés en el polinucleótido de inserción pueden incluir, pero no están limitados a, loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attp, att, FRT, rox, y una combinación de los mismos.

En otras realizaciones, los sitios de recombinación específicos del sitio flanquean un polinucleótido que codifica un marcador de selección y/o un gen informador contenido dentro del polinucleótido de inserción. En tales casos, después de la integración del polinucleótido de inserción en el locus genómico objetivo, se pueden eliminar las secuencias entre los sitios de recombinación específicas del sitio.

En una realización, el polinucleótido de inserción comprende un polinucleótido que codifica un marcador de selección. Dichos marcadores de selección incluyen, pero no se limitan, a neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hig^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 los mismos. En una realización, el polinucleótido que codifica el marcador de selección está unido operativamente a un promotor activo en la célula. Cuando se colocan en serie los polinucleótidos de interés en un locus genómico objetivo, 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 con secuencias objetivo de recombinación específicas del sitio.

El polinucleótido de inserción puede comprender además un gen informador unido operativamente a un promotor, en el que el gen informador codifica una proteína indicadora seleccionada del grupo que consiste en LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, proteína fluorescente amarilla mejorada (EYFP), Emerald, proteína fluorescente verde mejorada (EGFP), CyPet, proteína fluorescente cian (CFP), Cerulean, T-Sapphire, luciferasa, fosfatasa alcalina y una combinación de las mismas. Dichos genes informadores pueden unirse operativamente a un promotor activo en la célula. Dichos promotores pueden ser un promotor inducible, un promotor que es endógeno al gen informador o a la célula, un promotor que es heterólogo al gen informador o a la célula, un promotor específico de la célula, una forma del promotor específico del tejido o una promotor específico de la etapa de desarrollo.

ii. Vectores de direccionamiento

Los vectores de direccionamiento se emplean para introducir el polinucleótido de inserción en el locus genómico objetivo en el cromosoma Y o en un locus objetivo desafiante o en otro cromosoma de interés. El vector de direccionamiento comprende el polinucleótido de inserción y además comprende un brazo de homología secuencia arriba y secuencia abajo que flanquea al polinucleótido de inserción. Los brazos de homología que flanquean al polinucleótido de inserción corresponden a regiones genómicas dentro del locus genómico objetivo. Para facilitar la referencia, las regiones genómicas correspondientes dentro del locus genómico objetivo se denominan en este documento "sitios objetivo". Por lo tanto, en un ejemplo, un vector de direccionamiento puede comprender un primer polinucleótido de inserción flanqueado por un primer y un segundo brazo de homología que corresponde a un primer y un segundo sitio objetivo localizados lo suficientemente cerca del primer sitio de reconocimiento dentro del polinucleótido que codifica el marcador de selección. Como tal, el vector de direccionamiento ayuda de este modo a la integración del polinucleótido de inserción en el locus genómico objetivo a través de un evento de recombinación homóloga que se produce entre los brazos de homología y los sitios objetivo correspondientes dentro del genoma de la célula.

Un brazo de homología del vector de direccionamiento puede ser de cualquier longitud que sea suficiente para promover un evento de recombinación homóloga con un sitio objetivo correspondiente, que incluye, por ejemplo, de aproximadamente 400 pb a aproximadamente 500 pb, de aproximadamente 500 pb a aproximadamente 600 pb, de aproximadamente 600 pb a aproximadamente 700 pb, de aproximadamente 700 pb a aproximadamente 800 pb, de aproximadamente 800 pb a aproximadamente 900 pb, o de aproximadamente 900 pb a aproximadamente 1000 pb; 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. Esta divulgación describe que, la suma total de los brazos de direccionamiento es de al menos 0,5 kb, 1 kb, 1,5 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb o al menos 10 kb. Esta divulgación describe que la suma total de los brazos de homología está entre aproximadamente 0,5 kb y aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 1,5 kb, aproximadamente 1,5 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 6 kb, aproximadamente 6 kb a aproximadamente 7 kb, aproximadamente 7 kb a aproximadamente 8 kb, aproximadamente 8 kb a aproximadamente 9 kb, o aproximadamente 10 kb a aproximadamente 150 kb. Como se describe con más detalle a continuación, los vectores de direccionamiento grandes pueden emplear brazos de direccionamiento de mayor longitud.

Los sitios objetivo dentro del locus genómico objetivo que corresponden a los brazos de homología secuencia arriba y secuencia abajo del vector objetivo se ubican en "proximidad suficiente al sitio de reconocimiento" ubicado en el polinucleótido que codifica el marcador de selección. Tal como se usa en el presente documento, los brazos de

homología secuencia arriba y secuencia abajo de un vector de direccionamiento se encuentran "lo suficientemente cerca" de un sitio de reconocimiento cuando la distancia es tal que promueve la aparición de un evento de recombinación homóloga entre los sitios objetivo y los brazos de homología tras un corte o rompimiento de cadena doble en el sitio de reconocimiento. Por lo tanto, en realizaciones específicas, los sitios objetivo correspondientes al brazo de homología secuencia arriba y/o secuencia abajo del vector de direccionamiento están dentro de al menos 10 nucleótidos hasta aproximadamente 14 kb de un sitio de reconocimiento dado. En realizaciones específicas, el sitio de reconocimiento es 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 direccionamiento al sitio de reconocimiento dentro del polinucleótido que codifica el marcador de selección puede variar. Por ejemplo, ambos sitios de destino pueden ubicarse 5' del sitio de reconocimiento, ambos sitios de destino pueden ubicarse 3' del sitio de reconocimiento, o los sitios objetivo pueden flanquear el sitio de reconocimiento.

En realizaciones específicas, el locus genómico objetivo comprende (i) una secuencia objetivo 5' que es homóloga a un brazo de homología 5'; y (ii) una secuencia objetivo 3' que es homóloga a un brazo de homología 3'. En realizaciones específicas, la secuencia objetivo 5' y la secuencia objetivo 3' están separadas por al menos 5 kb pero menos de 3 Mb, al menos 5 kb pero menos de 10 kb, al menos 10 kb pero menos de 20 kb, al menos 20 kb pero menos de 40 kb, al menos 40 kb pero menos de 60 kb, al menos 60 kb pero menos de 80 kb, al menos aproximadamente 80 kb pero menos de 100 kb, al menos 100 kb pero menos de 150 kb, o al menos 150 kb pero menos de 200 kb, al menos aproximadamente 200 kb pero menos de aproximadamente 300 kb, al menos aproximadamente 300 kb pero menos de aproximadamente 400 kb, al menos aproximadamente 400 kb pero menos de aproximadamente 500 kb, al menos aproximadamente 500 kb pero menos de aproximadamente 1 Mb, al menos aproximadamente 1 Mb pero menos de aproximadamente 1,5 Mb, al menos aproximadamente 1,5 Mb pero menos de aproximadamente 2 Mb, al menos aproximadamente 2 Mb pero menos de aproximadamente 2,5 Mb, o al menos aproximadamente 2,5 Mb pero menos de aproximadamente 3 Mb.

Como se usa en este documento, un brazo de homología y un sitio objetivo "corresponden" o se "corresponden" entre sí cuando las dos regiones comparten un nivel suficiente de identidad de secuencia entre sí para actuar como sustratos para una reacción de recombinación homóloga. Por "homología" se entiende secuencias de ADN que son idénticas o comparten la identidad de secuencia con una secuencia correspondiente. La identidad de secuencia entre un sitio objetivo dado y el brazo de homología correspondiente encontrado en el vector de direccionamiento puede ser cualquier grado de identidad de secuencia que permita que se produzca una recombinación homóloga. Por ejemplo, la cantidad de identidad de secuencia compartida por el brazo de homología del vector de direccionamiento (o un fragmento del mismo) y el sitio objetivo (o un fragmento del mismo) 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 experimentan recombinación homóloga. Además, una región de homología correspondiente 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 pueden comprender regiones de homología correspondientes que van de aproximadamente 400 pb a aproximadamente 500 pb, de aproximadamente 500 pb a aproximadamente 600 pb, de aproximadamente 600 pb a aproximadamente 700 pb, de aproximadamente 700 pb a aproximadamente 800 pb, de aproximadamente 800 pb a aproximadamente 900 pb, o de aproximadamente 900 pb a aproximadamente 1.000 pb (tal como se describe para los vectores TVEC pequeño descritos en otra parte en el presente documento); o al menos aproximadamente 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 en el presente documento) 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 se denominan en este documento un brazo de homología secuencia arriba y secuencia abajo. Esta terminología se refiere a la posición relativa de los brazos de homología para el polinucleótido de inserción dentro del vector de direccionamiento.

Los brazos de homología del vector de direccionamiento están, por lo tanto, diseñados para corresponder a un sitio objetivo con el locus genómico objetivo en el cromosoma Y o dentro de un locus objetivo desafiante. Por lo tanto, los brazos de homología pueden corresponder a un locus genómico que es nativo a la célula, o alternativamente, pueden corresponder a una región de un segmento heterólogo o exógeno de ADN que se integró en el cromosoma Y, incluidos, entre otros, transgenes, casetes de expresión, o regiones heterólogas o exógenas del ADN genómico. Alternativamente, los brazos de homología del vector de direccionamiento pueden corresponder a una región de un cromosoma artificial de levadura (YAC), un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano o cualquier otra región genómica modificada por ingeniería genética contenida en una célula huésped apropiada. Aún más, los brazos de homología del vector de direccionamiento pueden corresponder o derivarse de una región de una biblioteca de BAC, una biblioteca de cósmidos, o una biblioteca de fagos Pl. Así, en realizaciones específicas, los brazos de homología del vector de direccionamiento corresponden a un locus genómico en el cromosoma Y o a un locus objetivo desafiante que es nativo, heterólogo o exógeno a un mamífero no humano, un roedor, un humano, una rata, un ratón, un hámster, un conejo, un cerdo, un bovino, 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 agrícola, un ave,

o cualquier otro organismo de interés. En otras formas de realización, los brazos de homología corresponden a un locus genómico de la célula que no se puede alcanzar con un método convencional o se puede alcanzar de forma incorrecta o solo con una eficiencia significativamente baja, en ausencia de un corte o rompimiento de cadena doble inducida por un agente de nucleasa. En una realización, los brazos de homología se derivan de un ADN sintético.

- 5 En otras realizaciones adicionales, los brazos de homología secuencia arriba y secuencia abajo 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 dirigidos 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 del genoma de la misma cepa, por ejemplo, el genoma seleccionado es un genoma de ratón de una primera cepa, y los brazos dirigidos son de un genoma de ratón del mismo ratón o de la misma cepa.

10 El vector de direccionamiento (tal como un vector de direccionamiento grande) también puede comprender un casete de selección o un gen informador como se describe en otra parte en este documento. El casete de selección puede comprender una secuencia de ácido nucleico que codifica un marcador de selección, en el que la secuencia de ácido nucleico está unida operativamente a un promotor. Dichos promotores pueden ser un promotor inducible, un promotor que es endógeno al gen informador o la célula, un promotor que es heterólogo al gen informador o a la célula, un promotor específico de la célula, una forma del promotor específico del tejido o un promotor específico de la etapa de desarrollo. En una realización, el marcador de selección se selecciona entre neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hig^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 los mismos. El marcador de selección del vector de direccionamiento puede estar flanqueado por los brazos de homología secuencia arriba y secuencia abajo o puede encontrarse 5' o 3' a los brazos de homología.

15 En una realización, el vector de direccionamiento (tal como un vector de direccionamiento grande) comprende un gen informador unido operativamente a un promotor, en el que el gen informador codifica una proteína informadora seleccionada del grupo que consiste en LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, proteína fluorescente amarilla mejorada (EYFP), Emerald, proteína fluorescente verde mejorada (EGFP), CyPet, proteína fluorescente cian (CFP), Cerulean, T-Sapphire, luciferasa, fosfatasa alcalina y una combinación de los mismos. Dichos genes informadores pueden unirse operativamente a un promotor activo en la célula. Dichos promotores pueden ser un promotor inducible, un promotor que es endógeno al gen informador o la célula, un promotor que es heterólogo al gen informador o a la célula, una forma del promotor específico del tejido o un promotor específico de la etapa de desarrollo.

20 En una realización no limitante, el uso combinado del vector de direccionamiento (que incluye, por ejemplo, un vector de direccionamiento grande) con el agente de nucleasa da como resultado una mayor eficacia de direccionamiento en comparación con el uso del vector de direccionamiento solo. En una realización, cuando el vector de direccionamiento se usa junto con el agente de nucleasa, la eficiencia de direccionamiento del vector de direccionamiento se incrementa al menos dos veces, al menos tres veces, al menos 4 veces, o al menos 10 veces cuando se compara con cuando se usa sólo el vector de direccionamiento.

iii. Vectores de direccionamiento grandes

25 El término "vector de direccionamiento grande" o "LTVEC" como se usa en este documento incluye vectores de direccionamiento grandes que comprenden brazos de homología que corresponden y se derivan de secuencias de ácido nucleico más grandes que las utilizadas típicamente por otros enfoques destinados a realizar un direccionamiento homólogo en células y/o que comprende polinucleótidos de inserción que comprenden secuencias de ácido nucleico más grandes que las utilizadas típicamente por otros enfoques destinados a realizar un direccionamiento de recombinación homóloga en células. En realizaciones específicas, los brazos de homología y/o el polinucleótido de inserción del LTVEC comprenden una secuencia genómica de una célula eucariota. El tamaño del LTVEC es demasiado grande para permitir el cribado de eventos de direccionamiento mediante ensayos convencionales, por ejemplo, transferencia Southern y PCR de gran intervalo (por ejemplo, 1 kb-5 kb). Los ejemplos de LTVEC incluyen, pero no están limitados a, vectores derivados de un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano o un cromosoma artificial de levadura (YAC). Los ejemplos no limitativos de LTVEC y los métodos para hacerlos se describen, por ejemplo, en las patentes de Estados Unidos Nos. 6.586.251, 6.596.541, 7.105.348 y el documento WO 2002/036789 (PCT/US01/45375).

30 El LTVEC puede ser de cualquier longitud, incluyendo, pero no limitado a, al menos aproximadamente 10 kb, aproximadamente 15 kb, aproximadamente 20 kb, aproximadamente 30 kb, aproximadamente 40 kb, aproximadamente 50 kb, aproximadamente 70 kb, aproximadamente 80 kb, aproximadamente 80 kb, aproximadamente 90 kb, aproximadamente 100 kb, aproximadamente 150 kb, aproximadamente 200 kb, de aproximadamente 10 kb a aproximadamente 15 kb, aproximadamente 15 kb a aproximadamente 20 kb, aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 300 kb, de aproximadamente 50 kb a aproximadamente 75 kb, de aproximadamente 75 kb a aproximadamente 100 kb, de aproximadamente 100 kb a 125 kb, de aproximadamente 125 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 175 kb, de aproximadamente

175 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 225 kb, de aproximadamente 225 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 275 kb o de aproximadamente 275 kb a aproximadamente 300 kb.

5 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, 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.

20 En otros casos, el diseño de LTVEC puede ser tal que permita la sustitución de una secuencia dada que es de aproximadamente 5 kb a aproximadamente 200 kb o de aproximadamente 5 kb a aproximadamente 3,0 Mb como se describe en este documento. En una realización, el reemplazo es de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 70 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 110 kb, de aproximadamente 110 kb a aproximadamente 120 kb, de aproximadamente 120 kb a aproximadamente 130 kb, de aproximadamente 130 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 170 kb, de aproximadamente 170 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 190 kb, de aproximadamente 190 kb a aproximadamente 200 kb, de aproximadamente 5kb a aproximadamente 10 kb, de aproximadamente 10 kb de 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 1Mb 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 3Mb.

40 En una realización, los brazos de homología de LTVEC se derivan de una biblioteca de BAC, una biblioteca de cósmidos o una biblioteca de fagos PI. En otras realizaciones, los brazos de homología se derivan del locus genómico objetivo de la célula y, en algunos casos, el locus genómico objetivo para el que se diseña LTVEC para direccionamiento no puede ser direccionado utilizando un método convencional. En aún otras realizaciones, los brazos de homología se derivan de un ADN sintético.

45 En una realización, la suma total del brazo de homología secuencia arriba y el brazo de homología secuencia abajo en el LTVEC es al menos de 10 kb. La divulgación establece que el brazo de homología secuencia arriba varía desde aproximadamente 5 kb a aproximadamente 100 kb. En una realización, el brazo de homología secuencia abajo varía de aproximadamente 5 kb a aproximadamente 100 kb. La divulgación establece que la suma total de los brazos de homología secuencia arriba y secuencia abajo es de aproximadamente 5 kb a aproximadamente 10 kb, de aproximadamente 10 kb a aproximadamente 20 kb, de aproximadamente 20 kb a aproximadamente 30 kb, de aproximadamente 30 kb a aproximadamente 40 kb, de aproximadamente 40 kb a aproximadamente 50 kb, de aproximadamente 50 kb a aproximadamente 60 kb, de aproximadamente 60 kb a aproximadamente 70 kb, de aproximadamente 70 kb a aproximadamente 80 kb, de aproximadamente 80 kb a aproximadamente 90 kb, de aproximadamente 90 kb a aproximadamente 100 kb, de aproximadamente 100 kb a aproximadamente 110 kb, de aproximadamente 110 kb a aproximadamente 120 kb, de aproximadamente 120 kb a aproximadamente 130 kb, de aproximadamente 130 kb a aproximadamente 140 kb, de aproximadamente 140 kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 160 kb, de aproximadamente 160 kb a aproximadamente 170 kb, de aproximadamente 170 kb a aproximadamente 180 kb, de aproximadamente 180 kb a aproximadamente 190 kb, o de aproximadamente 190 kb a aproximadamente 200 kb. En una realización, el tamaño de la eliminación es el mismo o similar al tamaño de la suma total de los brazos de homología 5' y 3' del LTVEC.

En una realización, el LTVEC comprende un casete de selección o un gen informador como se divulgó en otra parte en este documento.

60 iv. Métodos para integrar un polinucleótido de inserción cerca del sitio de reconocimiento en el cromosoma Y mediante recombinación homóloga

Se proporcionan métodos para modificar un locus genómico objetivo en el cromosoma Y en una célula que comprende: (a) proporcionar una célula que comprende un locus genómico objetivo en el cromosoma Y, (b) introducir en la célula un primer vector de direccionamiento que comprende un primer polinucleótido de inserción flanqueado por un primer y un segundo brazo de homología que corresponde a un primer y un segundo sitio objetivo; y (c) identificar al menos una célula que comprende en su genoma el primer polinucleótido de inserción integrado en el locus genómico objetivo en el cromosoma Y. Se pueden realizar métodos similares para dirigir un locus cromosómico desafiante. Como se explica en detalle en otra parte del presente documento, la suma total del primer brazo de homología y el segundo brazo de homología del vector de direccionamiento es de aproximadamente 0,5 kb, 1 kb, 1,5 kb, 2 kb, 3 kb, 4 kb, 5kb, 6kb, 7kb, 8 kb, 9 kb, aproximadamente 0,5 kb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 1,5 kb, aproximadamente 1,5 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 6 kb, aproximadamente 6 kb a aproximadamente 7 kb, aproximadamente 8 kb a aproximadamente 9 kb, o es al menos 10 kb o al menos 10 kb y menos de 150 kb. En realizaciones específicas, se emplea un LTVEC. Esta divulgación describe que se emplea un TVEC pequeño. En una realización no limitante, tales métodos se realizan empleando los medios de cultivo que promueven el desarrollo de hembras fértiles F0 XY descritas en el presente documento y generando de este modo animales hembra fértiles F0 XY. En otro caso, los métodos descritos en el presente documento se emplean para producir una modificación genética escogida en el gen *Sry*, como se describe en otro lugar del presente documento.

También se proporcionan métodos para modificar un locus genómico objetivo en el cromosoma Y en una célula que comprende: (a) proporcionar una célula que comprende un locus genómico objetivo en el cromosoma Y que comprende un sitio de reconocimiento para un agente de nucleasa, (b) introducir en la célula de (i) el agente de nucleasa, en el que el agente de nucleasa induce un corte o rompimiento de la cadena doble en el primer sitio de reconocimiento; y, (ii) un primer vector de direccionamiento que comprende un primer polinucleótido de inserción flanqueado por un primer y un segundo brazo de homología que corresponde a un primer y un segundo sitio objetivo situados en una proximidad suficiente al primer sitio de reconocimiento; y (c) identificar al menos una célula que comprende en su genoma el primer polinucleótido de inserción integrado en el locus genómico objetivo en el cromosoma Y. Se pueden realizar métodos similares para dirigir a un locus objetivo desafiante. Como se explica en detalle en otra parte de este documento, la suma total del primer brazo de homología y el segundo brazo de homología del vector de direccionamiento es aproximadamente 0,5 kb, 1 kb, 1,5 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, aproximadamente 0,5 kb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 1,5 kb, aproximadamente 1,5 kb a aproximadamente 2 kb, aproximadamente 2 kb a aproximadamente 3 kb, aproximadamente 3 kb a aproximadamente 4 kb, aproximadamente 4 kb a aproximadamente 5 kb, aproximadamente 5 kb a aproximadamente 6 kb, aproximadamente 6 kb a aproximadamente 7 kb, aproximadamente 8 kb a aproximadamente 9 kb, o es al menos 10 kb o al menos 10 kb y menos de 150 kb. En realizaciones específicas, se emplea un LTVEC. Esta divulgación describe que se emplea un TVEC pequeño. En una realización no limitante, tales métodos se realizan empleando los medios de cultivo que promueven el desarrollo de hembras fértiles F0 XY descritas en el presente documento y generando de este modo animales hembra fértiles F0 XY. En otro caso, los métodos descritos en el presente documento se emplean para producir una modificación genética escogida en el gen *Sry*, como se describe en otro lugar del presente documento.

También se pueden emplear varios métodos para identificar células que tienen el polinucleótido de inserción integrado en el locus objetivo genómico. La inserción del polinucleótido de inserción en el locus objetivo genómico da como resultado una "modificación del alelo". El término "modificación del alelo" o "MOA" incluye la modificación de la secuencia exacta de ADN de un alelo de un gen o genes o locus (loci) cromosómico en un genoma. Los ejemplos de "modificación del alelo (MOA)" incluyen, pero no se limitan a, eliminaciones, sustituciones o inserciones tan pocas como un solo nucleótido o eliminaciones de muchos kilobases que abarcan un gen o genes o locus (loci) cromosómicos de interés, así como todas y cada una de las posibles modificaciones entre estos dos extremos.

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

En diversas realizaciones, en presencia de un corte o rompimiento de doble cadena, la eficiencia de direccionamiento de un vector de direccionamiento (tal como un LTVEC o un TVEC pequeño) en el locus genómico objetivo es al menos aproximadamente 2 veces mayor, al menos aproximadamente 3 veces superior, al menos aproximadamente 4 veces superior que en ausencia del corte o rompimiento de doble cadena (utilizando, por

ejemplo, el mismo vector de direccionamiento y los mismos brazos de homología y los correspondientes sitios objetivo en el locus genómico de interés pero en la ausencia de un agente de nucleasa objetivo que hace el corte o el rompimiento de doble cadena).

5 Los diversos métodos expuestos anteriormente se pueden secuenciar repetida para permitir la integración objetivo de cualquier número de polinucleótidos de inserción en un locus genómico objetivo dado en el cromosoma Y o en un locus objetivo desafiante. Por lo tanto, los diversos métodos proporcionan 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 genómico objetivo en el cromosoma Y o en un locus objetivo desafiante. En realizaciones particulares, tales métodos de mosaico secuencial permiten la reconstrucción de grandes regiones genómicas de una célula animal o de una célula de mamífero (es decir, un humano, un no humano, un roedor, un ratón, un mono, una rata, un hámster, un mamífero domesticado o un animal agrícola) en un locus genómico objetivo en un cromosoma Y. En tales casos, la transferencia y reconstrucción de las regiones genómicas que incluyen tanto regiones codificantes como no codificantes permiten preservar la complejidad de una región determinada al retener, al menos en parte, las regiones codificantes, las regiones no codificantes y las variaciones del número de copias encontradas dentro de la región genómica nativa. 10 Por lo tanto, los diversos métodos proporcionan, por ejemplo, métodos para generar regiones genómicas "heterólogas" o "exógenas" dentro de cualquier célula de mamífero o animal de interés. En un ejemplo no limitativo, se genera una región genómica "humanizada" dentro de un animal no humano. 15

Se reconoce además que junto con la modificación del locus genómico objetivo en el cromosoma Y, los diversos métodos y composiciones divulgados en el presente documento pueden emplearse también para generar una modificación genética objetivo en otro cromosoma. 20

v. Polinucleótidos de interés

Cualquier polinucleótido de interés puede estar contenido en los diversos polinucleótidos de inserción y, por lo tanto, integrado en el locus genómico objetivo en el cromosoma Y o en un locus objetivo desafiante. Los métodos descritos en el presente documento proporcionan al menos 1, 2, 3, 4, 5, 6 o más polinucleótidos de interés que se integran en el locus genómico objetivo. 25

El polinucleótido de interés dentro del polinucleótido de inserción cuando se integra en el locus genómico objetivo en el cromosoma Y o en un locus objetivo desafiante puede introducir una o más modificaciones genéticas en la célula. La modificación genética puede comprender una eliminación de una secuencia endógena de ácido nucleico y/o la adición de un polinucleótido exógeno o heterólogo u ortólogo en el locus genómico objetivo. En una realización, la modificación genética comprende un reemplazo de una secuencia de ácido nucleico endógena con un polinucleótido exógeno de interés en el locus genómico objetivo. Por lo tanto, los métodos proporcionados en la presente memoria permiten la generación de una modificación genética que comprende una inactivación, una eliminación, una inserción, un reemplazo ("activación"), una mutación puntual, un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes o una combinación de los mismos en un locus genómico objetivo en el cromosoma Y. Dichas modificaciones pueden ocurrir tras la integración de los polinucleótidos de inserción primero, segundo, tercero, cuarto, quinto, sexto, séptimo o posterior en el locus genómico objetivo. 30 35

El polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus genómico objetivo puede comprender una secuencia que es nativa u homóloga a la célula en la que se introduce; el polinucleótido de interés puede ser heterólogo a la célula a 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. Como se usa en este documento, "homólogo" en referencia a una secuencia es una secuencia que es nativa de la célula. Como se usa en este documento, "heterólogo" en referencia a una secuencia es una secuencia que se origina a partir de una especie foránea, o, si es de la misma especie, se modifica sustancialmente a partir de su forma nativa en composición y/o locus genómico por intervención humana deliberada. Como se usa en este documento, "exógeno" en referencia a una secuencia es una secuencia que se origina a partir de una especie foránea. Tal como se usa en el presente documento, "ortólogo" es un polinucleótido de una especie que es funcionalmente equivalente a una secuencia de referencia conocida en otra especie (es decir, una variante de la especie). El polinucleótido de interés puede ser de cualquier organismo de interés que incluye, entre otros, no humano, un roedor, un hámster, un ratón, una rata, un humano, un mono, un ave, un mamífero agrícola o un mamífero no agrícola. El polinucleótido de interés puede comprender además una región codificante, una región no codificante, una región reguladora o un ADN genómico. Por lo tanto, el primero, segundo, tercero, cuarto, quinto, sexto, séptimo y/o cualquiera de los polinucleótidos de inserción posteriores pueden comprender dichas secuencias. 40 45 50

En una realización, el polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus genómico objetivo en el cromosoma Y 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 agrícola o un ácido nucleico de mamífero no agrícola. En aún 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 55 60

genómico de ratón, un ácido nucleico genómico humano, un ácido nucleico no humano, un ácido nucleico de roedor, un ácido nucleico de rata, un ácido nucleico de hámster, un ácido nucleico de mono, un ácido nucleico de mamífero agrícola o un ácido nucleico de mamífero no agrícola o una combinación de los mismos.

5 En una realización, el polinucleótido de interés puede oscilar de aproximadamente 500 nucleótidos hasta 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 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.

20 El polinucleótido de interés dentro del polinucleótido de inserción y/o insertado en el locus genómico objetivo en el cromosoma Y o en un locus objetivo desafiante puede codificar un polipéptido, puede codificar un ARNm, puede codificar un ARN largo no codificante, o puede comprender cualquier región reguladora o regiones no codificantes de interés que incluyen, por ejemplo, una secuencia reguladora, una secuencia promotora, una secuencia potenciadora, una secuencia de unión a represor transcripcional, o una eliminación de una secuencia codificante no proteica, pero no comprende una eliminación de una secuencia codificante de proteínas. Además, el polinucleótido de interés dentro del polinucleótido de inserción y/o insertado en el locus genómico objetivo en el cromosoma Y o en un locus objetivo desafiante puede codificar una proteína expresada en el sistema nervioso, el sistema esquelético, el sistema digestivo, el sistema circulatorio, el sistema muscular, el sistema respiratorio, el sistema cardiovascular, el sistema linfático, el sistema endocrino, el sistema urinario, el sistema reproductivo, o una combinación de ellos.

25 El polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus genómico objetivo en el cromosoma Y o en un locus objetivo desafiante puede comprender una modificación genética en una secuencia de codificación. Dichas modificaciones genéticas incluyen, pero no se limitan a, una mutación por eliminación de una secuencia de codificación o la fusión de dos secuencias de codificación.

30 El polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus genómico objetivo en el cromosoma Y o en un locus objetivo desafiante 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 genómico en el cromosoma Y o en un locus objetivo desafiante comprende al menos un alelo de enfermedad. En tales casos, el alelo de la enfermedad puede ser un alelo dominante o el alelo de la enfermedad es un alelo recesivo. Además, el alelo de la enfermedad puede comprender un alelo de polimorfismo de nucleótido único (SNP). El polinucleótido de interés que codifica la proteína mutante puede ser de cualquier organismo, incluido, entre otros, un mamífero, un mamífero no humano, un roedor, un ratón, una rata, un humano, un mono, un mamífero agrícola o un polinucleótido de mamífero doméstico que codifica una proteína mutante.

45 El polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus genómico objetivo en el cromosoma Y o en un locus objetivo desafiante también puede comprender una secuencia reguladora, que incluye, por ejemplo, una secuencia promotora, una secuencia potenciadora, una secuencia de unión al represor transcripcional, o una secuencia de terminación transcripcional. En realizaciones específicas, el polinucleótido de interés dentro del polinucleótido de inserción y/o integrado en el locus genómico objetivo en el cromosoma Y o en un locus objetivo desafiante comprende un polinucleótido que tiene una eliminación de una secuencia no codificante de proteína, pero no comprende una eliminación de una secuencia codificante de proteína. En una realización, la eliminación de la secuencia no codificadora de proteína comprende una eliminación de una secuencia reguladora. En otra realización, la eliminación del elemento regulador comprende una eliminación de una secuencia promotora. En una realización, la eliminación del elemento regulador comprende una eliminación de una secuencia potenciadora. Dicho polinucleótido de interés puede ser de cualquier organismo, incluido, entre otros, un mamífero, un mamífero no humano, un roedor, un ratón, una rata, un humano, un mono, un mamífero agrícola o un polinucleótido de mamífero doméstico que codifica una proteína mutante.

55 Los diversos métodos divulgados en el presente documento pueden emplearse para generar una variedad de modificaciones en un locus genómico desafiante o en el locus del cromosoma Y (como en el caso de *Sry*). Tales modificaciones incluyen, por ejemplo, un reemplazo de una secuencia endógena de ácido nucleico con una secuencia de ácido nucleico homóloga u ortóloga; una eliminación de una secuencia endógena de ácido nucleico; una eliminación de una secuencia endógena de ácido nucleico, en la que la eliminación 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 300 kb, de aproximadamente 300 kb a aproximadamente 400 kb, de aproximadamente 400 kb a aproximadamente 500 kb, de aproximadamente 500 kb a aproximadamente 600 kb, de aproximadamente 600 kb a aproximadamente 700 kb, de aproximadamente 700 kb a aproximadamente 800 kb, de aproximadamente 800 kb a aproximadamente 900 kb, de aproximadamente 900 kb a aproximadamente 1 Mb, 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; una inserción de una secuencia exógena de ácido nucleico; una inserción de una secuencia exógena de ácido nucleico 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; una inserción de una secuencia exógena de ácido nucleico que comprende una secuencia de ácido nucleico homóloga u ortóloga; una inserción de una secuencia quimérica de ácido nucleico que comprende una secuencia de ácido nucleico humana y no humana; una inserción de un alelo condicional flanqueado con secuencias objetivo de recombinasa específicas del sitio; una inserción de un marcador seleccionable o un gen informador operativamente enlazado a un tercer promotor activo en la célula de mamífero; o una combinación de los mismos.

III. Métodos para la introducción de secuencias y generación de animales transgénicos

Como se describió anteriormente, los métodos y composiciones se proporcionan en este documento para permitir la modificación genética escogida de uno o más polinucleótidos de interés ubicados en el cromosoma Y, en un locus objetivo desafiante, o una disminución en el nivel y/o actividad de la proteína Sry. Se reconoce además que, adicionalmente a la modificación genética objetivo a una secuencia en el cromosoma Y o en un locus cromosómico objetivo desafiante, se puede realizar una modificación genética objetivo adicional en otros cromosomas. Tales sistemas que permiten estas modificaciones genéticas objetivo pueden emplear una variedad de componentes y por facilidad de referencia, en este documento el término "sistema de integración genómica objetivo" incluye genéricamente todos los componentes necesarios para un evento de integración (es decir, los diversos agentes de nucleasa, sitios de reconocimiento, polinucleótidos de inserción de ADN, vectores de direccionamiento, locus genómico objetivo y polinucleótidos de interés).

Los métodos proporcionados en este documento comprenden la introducción en una célula de uno o más polinucleótidos o constructos polipeptídicos que comprenden los diversos componentes del sistema de integración genómica objetivo. "Introducción" significa presentar a la célula la secuencia (de polipéptido o de polinucleótido) de tal manera que la secuencia obtenga acceso al interior de la célula. Los métodos proporcionados en este documento no dependen de un método particular para introducir ningún componente del sistema de integración genómica objetivo en la célula, solo que el polinucleótido obtiene 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 estables, métodos de transfección transitoria y métodos mediados por virus.

En algunas realizaciones, las células empleadas en los métodos y composiciones tienen un constructo de ADN incorporado en forma estable en su genoma. "Incorporado en forma estable" o "introducido en forma estable" significa la introducción de un polinucleótido en la célula de modo que la secuencia de nucleótidos se integre en el genoma de la célula y pueda ser heredada por su progenie. Se puede usar cualquier protocolo para la incorporación estable de los constructos de ADN o los diversos componentes del sistema de integración genómica objetivo.

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 con base en compuestos químicos que incluyen el uso de liposomas; nanopartículas; fosfato de calcio (Graham et al., (1973). *Virology* 52 (2): 456-67, Bacchetti et al., (1977) *Proc Natl Acad Sci USA* 74 (4): 1590-4 y, Kriegler, M (1991). *Transfer and Expression: A Laboratory Manual*. Nueva York: WH Freeman and Company. Páginas 96-97); dendrímeros; o polímeros catiónicos tales como DEAE-dextrano o polietilimina. Los métodos no químicos incluyen electroporación; sonoporación; y transfección óptica. La transfección con base en partículas incluye el uso de una pistola de genes, transfección asistida por imán (Bertram, J. (2006) *Current Pharmaceutical Biotechnology* 7, 277-28). Los métodos virales también se pueden utilizar para transfección.

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

Los animales no humanos pueden generarse empleando los diversos métodos descritos en el presente documento. Tales métodos de esta divulgación comprenden (1) integrar uno o más polinucleótidos de interés en el locus genómico objetivo del cromosoma Y de una célula pluripotente del animal no humano para generar una célula pluripotente modificada genéticamente que comprende el polinucleótido de inserción en el locus genómico objetivo del cromosoma Y que emplea los métodos divulgados en este documento; (2) seleccionar la célula pluripotente modificada genéticamente que tiene uno o más polinucleótidos de interés en el locus genómico objetivo del cromosoma Y; (3) introducir la célula pluripotente modificada genéticamente en un embrión huésped del animal no humano en una etapa premórula; 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. Se pueden emplear métodos similares para dirigirse a un locus cromosómico objetivo desafiante. El animal no humano puede ser un mamífero no humano, un roedor, un ratón, una rata, un hámster, un mono, un mamífero agrícola o un mamífero doméstico, o un pez o un ave.

La célula pluripotente puede ser una célula ES humana, una célula ES no humana, una célula ES de roedor, una célula ES de ratón, una célula ES de rata, una célula ES de hámster, una célula ES de mono, una célula ES de mamífero agrícola o una célula ES de mamífero domesticada. En otras realizaciones, la célula pluripotente es una célula de mamífero, una célula humana, una célula de mamífero no humano, una célula pluripotente humana, una célula ES humana, una célula madre adulta humana, una célula de progenitor humana de desarrollo restringido, una célula iPS humana, una célula humana, una célula de roedor, una célula de rata, una célula de ratón, una célula de hámster. En una realización, la modificación genética objetivo disminuye el nivel y/o la actividad de la proteína Sry. En tales casos, la célula pluripotente puede comprender una célula ES XY o una célula iPS XY. Los métodos para cultivar tales células para promover el desarrollo de animales no humanos hembra XY fértiles F0 se describen en detalle en otra parte de este documento.

Las técnicas de transferencia nuclear también pueden usarse para generar los animales mamíferos no humanos de esta divulgación. En resumen, los métodos para la transferencia nuclear incluyen los pasos de: (1) enucleo un ovocito; (2) aislar una célula o núcleo donante para combinar con el ovocito enucleado; (3) insertar la célula o el núcleo en el ovocito enucleado para formar una célula reconstituida; (4) implantar la célula reconstituida en el útero de un animal para formar un embrión; y (5) permitir que el embrión se desarrolle. En tales métodos, los ovocitos generalmente se recuperan de animales muertos, aunque también pueden aislarse de oviductos y/o ovarios de animales vivos. Los ovocitos se pueden madurar 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 el núcleo del donante en el ovocito enucleado para formar una célula reconstituida generalmente se realiza mediante microinyección de una célula del donante en la zona pelúcida antes de la fusión. La fusión se puede inducir mediante la aplicación de un pulso eléctrico de CD a través del plano de contacto/fusión (electrofusión), mediante la exposición de las células a sustancias químicas que promueven la fusión, tales como el polietilenglicol, o mediante un virus inactivado, tal 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 por el esperma, niveles crecientes de cationes divalentes en el ovocito y reducción de la fosforilación de proteínas celulares (tal como por medio de inhibidores de quinasa) en el ovocito. Las células reconstituidas activadas, o los embriones, se cultivan típicamente en un medio bien conocido por los expertos en la técnica y luego se transfieren al útero de un animal. Véase, por ejemplo, los documentos US20080092249, WO/1999/005266A2, US20040177390, WO/2008/017234A1, y la patente de Estados Unidos No. 7,612,250.

En esta descripción se proporcionan otros métodos para elaborar un animal no humano que comprende en su línea germinal una o más modificaciones genéticas como se describe en el presente documento, que comprende: (a) modificar un locus genómico objetivo en el cromosoma Y de un animal no humano en una célula procariota que emplea los diversos métodos descritos en el presente documento; (b) seleccionar una célula procariota modificada que comprende la modificación genética en el locus genómico objetivo; (c) aislar el vector de direccionamiento modificado genéticamente del genoma de la célula procariota modificada; (d) introducir el vector de direccionamiento modificado genéticamente en una célula pluripotente del animal no humano para generar una célula pluripotente genéticamente modificada que comprende el ácido nucleico de inserción en el locus genómico objetivo del cromosoma Y; (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 de premórula; y (g) implantar el embrión huésped que comprende la célula pluripotente genéticamente modificada en una madre sustituta para generar una generación F0 derivada de la célula pluripotente genéticamente modificada. En tales métodos, el vector de direccionamiento puede comprender un vector de direccionamiento grande o un TVEC pequeño. Se pueden emplear métodos similares para dirigirse a un locus objetivo desafiante. El animal no humano puede ser un mamífero no humano, un roedor, un ratón, una rata, un hámster, un mono, un mamífero agrícola o un mamífero doméstico. La célula pluripotente puede ser una célula ES humana, una célula ES no humana, una célula ES de roedor, una célula ES de ratón, una célula ES de rata, una célula ES de hámster, una célula ES de mono, una célula ES de mamífero agrícola o una célula ES de mamífero doméstico. En otras realizaciones, la célula pluripotente es una célula de mamífero, una célula humana, una célula de mamífero no humano, una célula pluripotente humana, una célula ES humana, una célula madre humana adulta, una célula progenitora humana de desarrollo restringido, una célula iPS humana, una célula humana, una célula de roedor, una célula de rata, una célula de ratón, una célula

de hámster. En una realización, la modificación genética objetivo disminuye el nivel y/o la actividad de la proteína Sry. En tales casos, la célula pluripotente puede comprender una célula ES XY o una célula iPS XY. Los métodos para cultivar tales células para promover el desarrollo de animales no humanos hembra XY fértiles F0 se describen en detalle en otra parte de este documento.

5 En otros métodos de esta divulgación, la etapa de aislamiento (c) comprende además (c1) linealizar del vector de direccionamiento genéticamente modificado (es decir, el LTVEC genéticamente modificado). En otros métodos adicionales de esta divulgación, la etapa de introducción (d) comprende además (d1) introducir un agente de nucleasa como se describe en el presente documento en la célula pluripotente. En una realización, las etapas de selección (b) y/o (e) se llevan a cabo aplicando un agente seleccionable como se describe en el presente documento a la célula procariota o la célula pluripotente. En esta divulgación, los pasos de selección (b) y/o (e) se llevan a cabo mediante un ensayo de modificación del alelo (MOA) como se describe en el presente documento.

10 Se proporcionan métodos adicionales para modificar un locus genómico objetivo de una célula animal mediante recombinación homóloga bacteriana (BHR) en una célula procariota y comprenden: (a) proporcionar una célula procariota que comprende un locus genómico objetivo del cromosoma Y; (b) introducir en la célula procariota un vector de direccionamiento (como se describió anteriormente) que comprende un polinucleótido de inserción flanqueado con un primer brazo de homología secuencia arriba y un primer brazo de homología secuencia abajo, en el que el polinucleótido de inserción comprende una región genómica de mamífero e introducir en la célula procariota un agente de nucleasa que hace un corte o un rompimiento de cadena doble en o cerca del primer sitio de reconocimiento, y (c) seleccionar una célula procariota dirigida que comprende el polinucleótido de inserción en el locus genómico objetivo del cromosoma, en donde la célula procariota es capaz de expresar una recombinasa que media la BHR. Se pueden emplear métodos similares para dirigir un locus objetivo desafiante. Las etapas (a)-(c) se pueden repetir en serie como se describe en este documento para permitir la introducción de múltiples polinucleótidos de inserción en el locus genómico objetivo en la célula procariota. Una vez que el locus genómico objetivo se "construye" con la célula procariota, un vector de direccionamiento que comprende el locus genómico objetivo modificado del cromosoma Y puede aislarse de la célula procariota e introducirse en un locus genómico objetivo del cromosoma Y dentro de una célula de mamífero. Las células de mamífero que comprenden el locus genómico modificado del cromosoma Y pueden transformarse en animales transgénicos no humanos.

15 Se proporcionan métodos adicionales para modificar un locus genómico objetivo de una célula animal a través de recombinación homóloga bacteriana (BHR) en una célula procariota y comprenden: (a) proporcionar una célula procariota que comprende un locus genómico objetivo del cromosoma Y; (b) introducir en la célula procariota un vector de direccionamiento (como se describió anteriormente) que comprende un polinucleótido de inserción flanqueado con un primer brazo de homología secuencia arriba y un primer brazo de homología secuencia abajo, en donde el polinucleótido de inserción comprende una región genómica de mamífero, y (c) seleccionar una célula procariota objetivo que comprende el polinucleótido de inserción en el locus genómico objetivo del cromosoma, en donde la célula procariota es capaz de expresar una recombinasa que media la BHR. Se pueden emplear métodos similares para apuntar a un locus objetivo desafiante. Las etapas (a)-(c) se pueden repetir en serie como se describe en este documento para permitir la introducción de múltiples polinucleótidos de inserción en el locus genómico objetivo en la célula procariota. Una vez que el locus genómico objetivo se "construye" con la célula procariota, un vector de direccionamiento que comprende el locus genómico objetivo modificado del cromosoma Y puede aislarse de la célula procariota e introducirse en un locus genómico objetivo del cromosoma Y dentro de una célula de mamífero. Las células de mamífero que comprenden el locus genómico modificado del cromosoma Y pueden transformarse en animales transgénicos no humanos.

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

25 En esta divulgación, las células pluripotentes y/o totipotentes XY escogidas (es decir, las células ES XY o las células iPS XY) que comprenden diversas modificaciones genéticas como se describen en el presente documento se utilizan como células donantes de inserción y se introducen en un embrión en etapa de premórula a partir de una organismo correspondiente, por ejemplo, un embrión de ratón en estadio de 8 células, a través del método VELOCIMOUSE® (véase, por ejemplo, los documentos US 7.576.259, US 7.659.442, US 7.294.754 y US 2008-0078000 A1). El embrión animal no humano que comprende las células pluripotentes y/o totipotentes XY genéticamente modificadas (es decir, las células ES XY o células iPS XY) se incuba hasta la etapa de blastocisto y luego se implanta en una madre sustituta para producir una generación F0. En esta divulgación, se introducen células ES escogidas de mamíferos que comprenden diversas modificaciones genéticas como se describe en el presente documento en un embrión en etapa de blastocisto. Los animales no humanos que portan el locus genómico genéticamente modificado del cromosoma Y pueden identificarse mediante un ensayo de modificación del alelo (MOA) como se describe en esta divulgación. El animal no humano resultante de la generación F0 derivado de las células pluripotentes y/o totipotentes XY genéticamente modificadas (es decir, células ES XY o células iPS XY) se cruza con un animal no humano de tipo silvestre para obtener una descendencia de generación F1. Después del

genotipado con cebadores y/o sondas específicas, los animales no humanos F1 que son heterocigotos para el locus genómico genéticamente modificado se cruzan entre sí para producir descendientes del animal no humano de generación F2 que son homocigotos para el locus genómico genéticamente modificado del cromosoma Y o para el locus objetivo desafiante genéticamente modificado.

5 IV. Células y casetes de expresión

Los diversos métodos descritos en este documento emplean un sistema de direccionamiento de locus genómico para el cromosoma Y o para un locus objetivo desafiante en una célula. Dichas células incluyen células procariotas tales como células bacterianas que incluyen *E. coli*, o células eucariotas como células de levadura, insecto, anfibio, planta o mamífero, incluidas, entre otras, una célula de ratón, una célula de rata, una célula de conejo, una célula de cerdo, una célula bovina, una célula de venado, una célula de oveja, célula de cabra, célula de pollo, célula de gato, 10 célula de perro, célula de hurón, célula de primate (por ejemplo, tití, mono rhesus) y similares, y células de mamíferos domesticados o células de mamíferos agrícolas. Algunas células son células no humanas, particularmente células de mamífero no humanos. En algunas realizaciones, para aquellos mamíferos para los cuales las células pluripotentes genéticamente modificadas no están fácilmente disponibles, se emplean otros 15 métodos para reprogramar células somáticas en células pluripotentes, por ejemplo, a través de la introducción en células somáticas de una combinación de factores inductores de pluripotencia, incluyendo, pero no limitado a, Oct3/4, Sox2, KLF4, Myc, Nanog, LIN28 y Glis1. En tales métodos, la célula también puede ser una célula de mamífero, una célula humana, una célula de mamífero no humano, una célula no humana, una célula de un roedor, una rata, un ratón, un hámster, una célula de fibroblastos o cualquier otra célula huésped. En otras realizaciones, la 20 célula es una célula pluripotente, una célula madre pluripotente inducida (iPS), una célula madre (ES) embrionaria no humana. Dichas células incluyen células pluripotentes, que incluyen, por ejemplo, células madre pluripotentes inducidas (iPS), células madre (ES) embrionarias de ratón, células madre (ES) embrionarias de rata, células (ES) embrionarias humanas o células progenitoras humanas con desarrollo restringido, células madre (ES) embrionaria de roedor, una célula madre (ES) embrionaria de ratón o una célula madre (ES) embrionaria de rata.

Los términos "polinucleótido", "secuencia de polinucleótido", "secuencia de ácido nucleico" y "fragmento de ácido nucleico" se usan de manera intercambiable en este documento. Estos términos abarcan secuencias de nucleótidos y similares. Un polinucleótido puede ser un polímero de ARN o ADN de cadena simple o doble, que opcionalmente contiene bases de nucleótidos sintéticos, no naturales o alteradas. Un polinucleótido en forma de un polímero de ADN puede estar compuesto por uno o más segmentos de ADNc, ADN genómico, ADN sintético o mezclas de los 30 mismos. Los polinucleótidos pueden comprender desoxirribonucleótidos y los ribonucleótidos incluyen tanto moléculas naturales como análogos sintéticos, y cualquier combinación de estos. Los polinucleótidos proporcionados en el presente documento también abarcan todas las formas de secuencias que incluyen, pero no se limitan a, formas monocatenarias, formas bicatenarias, horquillas, estructuras de vástago y bucle, y similares.

También se proporcionan polinucleótidos recombinantes. Los términos "polinucleótido recombinante" y "constructo de ADN recombinante" se usan indistintamente en este documento. Un constructo 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, un constructo recombinante puede comprender secuencias reguladoras y secuencias de codificación que se derivan de diferentes fuentes, o 40 secuencias reguladoras y secuencias de codificación derivadas de la misma fuente, pero dispuestas de una manera diferente a la encontrada en la naturaleza. Tal constructo se puede usar solo o se puede usar en conjunto con un vector. Si se usa un vector, entonces la elección del vector depende del método que se usa para transformar las células huésped como es bien conocido por los expertos en la técnica. Por ejemplo, se puede usar un vector plasmídico. La selección puede realizarse mediante el análisis Southern del ADN, el análisis Northern de la expresión del ARNm, el análisis de inmunotransferencia de la expresión de proteínas o el análisis fenotípico, entre 45 otros.

En realizaciones específicas, uno o más de los componentes descritos en el presente documento pueden proporcionarse en un casete de expresión para la expresión en la célula pluripotente y/o totipotente. El casete puede incluir secuencias reguladoras 5' y 3' operativamente enlazado a un polinucleótido proporcionado en el presente documento. "operativamente enlazado" significa un enlace funcional entre dos o más elementos. Por ejemplo, un 50 enlace operable entre un polinucleótido de interés y una secuencia reguladora (es decir, un promotor) es un enlace funcional que permite la expresión del polinucleótido de interés. Los elementos enlazados operativamente pueden ser contiguos o no contiguos. Cuando se usa para referirse a la unión de dos regiones codificantes de proteínas, operativamente enlazadas 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 enlazarse operativamente a secuencias reguladoras (por ejemplo, promotor, potenciador, secuencia silenciadora, etc.) para conservar la regulación transcripcional adecuada. El casete puede contener adicionalmente al menos un polinucleótido adicional de interés para ser introducido conjuntamente en la célula ES. Alternativamente, el polinucleótido adicional de interés puede proporcionarse en múltiples casetes de expresión. Dicho casete de expresión está provisto de una pluralidad de 55 sitios de restricción y/o sitios de recombinación para la inserción de un polinucleótido recombinante para estar bajo la regulación transcripcional de las regiones reguladoras. El casete de expresión puede contener adicionalmente genes marcadores de selección. 60

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

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

En la preparación del casete de expresión, los diversos fragmentos de ADN pueden manipularse, para proporcionar las secuencias de ADN en la orientación apropiada. Para 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, se pueden involucrar mutagénesis *in vitro*, reparación del cebador, restricción, hibridación, sustituciones nuevas, por ejemplo, transiciones y transversiones.

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

Los siguientes ejemplos se ofrecen a modo de ilustración y no a modo de limitación.

Ejemplos

Ejemplo 1. Selección del gen del cromosoma Y asistido por TALEN o CRISPR

Una eliminación dirigida que comprende un alelo de reemplazo lacZ para *Sry* se creó con un vector de direccionamiento que comprende, en orden, un brazo de homología secuencia arriba de aproximadamente 700 pb, una secuencia codificante de beta-galactosidasa (*lacZ*) seguida de una señal de poliadenilación, un casete de resistencia a neomicina flanqueado por sitios loxP que comprenden un promotor de ubiquitina C humana, que incluye el primer exón, el primer intrón y parte del segundo exón, una secuencia codificante de neomicina fosfotransferasa y una señal de poliadenilación y un brazo de homología secuencia abajo de aproximadamente 650 pb. El alelo creado mediante la orientación correcta del gen *Sry* con el vector de direccionamiento comprende una eliminación del marco de lectura abierto de *Sry* de aproximadamente 1 kb y la sustitución por el casete lacZ-neo, de manera que la secuencia codificante de la beta-galactosidasa se fusiona en marco con el codón de inicio de *Sry*. El vector de direccionamiento se usó para dirigir el gen *Sry* tanto en las líneas de células ES híbridas F1 C57BL/6 VGB6 (también conocida como B6A6) y la C57BL6/129 VGF1 (también conocida como F1H4). Las células ES de ratón VGF1 (F1H4) se derivaron de embriones híbridos producidos al cruzar un ratón hembra C57BL/6NTac con un ratón macho 129S6/SvEvTac. Por lo tanto, las células ES de VGF1 contienen un cromosoma Y del ratón 129S6/SvEvTac. Los ratones XY hembra producidos a partir de la línea celular VGF1 contienen un cromosoma Y derivado de un ratón 129S6/SvEvTac.

Ejemplo 2. Mutaciones inducidas por TALEN o CRISPR en el gen *Sry* del cromosoma Y

Las mutaciones por eliminación, presumiblemente el resultado de la reparación de uniones de extremos no homólogos (NHEJ) de rompimientos de ADN de doble cadena, que van desde 3 pb hasta 1,2 kb y más fueron creadas en el gen *Sry* por la acción de ARN guía de TALEN o de CRISPR, en combinación con la endonucleasa de ADN Cas9 (véase, la Figura 1). Las células ES que comprenden las mutaciones inducidas por TALEN y CRISPR en el gen *Sry* también portaban inserciones transgénicas al azar del LTVEC VG12778 proyecto NIH KOMP (disponible a través de Internet en la web ([www](http://www.velocigene.com/komp/detail/12778)) en la URL "www.velocigene.com/komp/detail/12778"), que comprende la eliminación de la secuencia de codificación de *Sry* y el reemplazo con un casete de inserción que comprende lacZ fusionado en el marco con el codón de inicio de *Sry* y un gen de resistencia a la neomicina flanqueado por brazos de homología

de 38 y 37 kb y con base en un BAC de la biblioteca de bMQ (129S7/SvEv Brd-Hprt b-m2). El LTVEC comprende en su brazo de homología todos los elementos de control conocidos para la expresión de *Sry*. Su beta-galactosidasa codificada por *lacZ* sirve como informador para la expresión específica del tejido y específica de la etapa de desarrollo del gen *Sry*. Las mutaciones inducidas por TALEN y CRISPR acompañadas por inserciones de LTVEC se crearon tanto en las líneas de células ES de VGB6 (también conocida como B6A6) como en VGF1 (también conocida como F1H4).

Se obtuvo un TALEN (TALEN-1) diseñado para dirigir parte de la secuencia de codificación del motivo de unión al ADN de la caja HMG (secuencia de reconocimiento en sentido ascendente: 5'-TCCCGTGGTGAGAGGCAC-3' (SEQ ID NO: 72); secuencia de reconocimiento en sentido descendente: 5'-TATTTTGCATGCTGGGAT-3' (SEQ ID NO: 73) en el gen *Sry*. TALEN-1 participó activamente en la creación de mutaciones NHEJ en el locus de *Sry* en múltiples experimentos.

Se crearon células ES de ratón tanto VGB6 como VGF1 con mutaciones inducidas por TALEN. La Tabla 1 contiene una lista de todos los clones y los tamaños de las mutaciones de eliminación que ellas portan. (ND en la Tabla 1 indica que se detectó una mutación mediante un ensayo de qPCR, pero no se determinó la naturaleza molecular exacta de la mutación). Todos los clones también portan al menos una copia de LTVEC VG12778 del proyecto NIH KOMP.

Tabla 1

Mutaciones inducidas por TALEN y CRISPR en el gen <i>Sry</i>			
Células ES	Clon	Agente que induce mutación	Eliminación (pb)
VGB6	DE7	TALEN-1	9
	DE11	TALEN-1	303*
	DG5	TALEN-1	627
	DH1	TALEN-1	ND
	EA2	TALEN-1	ND
	ED4	TALEN-1	>1.200
	EF4	TALEN-1	16
	EG7	TALEN-1	>1.200
VGB6	RD3	TALEN-1	>1.200
	RE9	TALEN-1	ND
	RF3	TALEN-1	9
	RG7	TALEN-1	15
	SF7	TALEN-1	3
	SG11	TALEN-1	6
	SH2	TALEN-1	>200
	SH11	TALEN-1	2
VGF1	TB1	TALEN-1	11
	TC2	TALEN-1	5
	UA5	TALEN-1	15
	UB5	TALEN-1	1201
	UE12	TALEN-1	9
	WE11	TALEN-1	>1.200

ES 2 694 629 T3

VGB6	OG6	CRISPR-2	9
	QE8	CRISPR-3	5
VGF1	AU-B6	CRISPR-4	5
	AU-C12	CRISPR-4	8
	AW-H5	CRISPR-5	22
*También contenía una inserción de 50 pb			

Los resultados de las microinyecciones de los clones de mutantes se exponen en la Tabla 2 y los resultados de reproducción de hembras con inversión de sexo se exponen en la Tabla 3.

Tabla 2

VelociMice de generación F0 producida por microinyección de clones de células ES mutantes Sry en embriones de 8 células				
Células ES	Clon	Mutación Sry	VM hembra	VM macho
VGB6	ED4	Eliminación >1 kb	2	0
	EG7	Eliminación >1 kb	19	0
	GB4	Ninguna	0	3
	GG1	Ninguna	0	5
	DE11	Eliminación de 303 pb; inserción de 50 pb	1	0
	DG5	Eliminación de 627 pb	11	0
VGF1	TA3	Ninguna	0	5
	TA4	Ninguna	0	11
	TB1	Eliminación de 11 pb	2	0
	TC2	Eliminación de 5 pb	8	0
	TH4	Ninguna	2	6
	UB5	Eliminación de 1.201 pb	6	0
	WE11	Eliminación >1.2 kb	7	0
	UA5	Eliminación de 15 pb	4	0
	UE12	Eliminación de 9 pb	8	0

5

Tabla 3

Resultados de reproducción de VelociMice hembra XY con mutaciones en el gen Sry					
Células ES	Clon	Eliminación de Sry (pb)	ID # de la hembra XY	Camadas producidas	Cachorros nacidos
VGB6	EG7	>1.200	1460403	0	0*
			1460404	0	
			1460405	0	
			1460406	1	

ES 2 694 629 T3

			1460408	0	
			1460409	0	
			1460410	0	
VGB6	DG5	627	1460428	0	
			1460429	0	
			1460430	0	
			1460431	0	
			1460432	0	
			1460436	0	
			1460437	0	
			1460438	0	
			1460410	0	
VGf1	UB5	1.201	1525585	5	33
			1525586	5	25
			1525587	5	32
			1525588	3	35
			1525589	3	19
VGf1	WE11	>1.200	1525573	3	4
			1525574	5	21
			1525575	4	11
			1525576	4	14
			1525577	4	16
			1525578	2	4
			1525579	2	6
VGf1	TB1	11	1525700	5	30
			1525701	4	28
VGf1	TC2	5	1525706	1	2
			1525707	5	10
			1525708	1	6
			1525709	4	17
			1525710	1	3
			1525711	3	9
			1525712	4	12
			1525713	2	7
VGf1	UA5	15	1594102	2	5
			1594103	2	7

			1594104	1	3
			1594105	2	15
VGF1	UE12	9	1594117	1	6
			1594118	2	12
			1594119	1	11
			1594120	2	8
			1594121	2	21
			1594122	2	15
			1594123	2	10
* ID # 1460406 hembra XY tuvo que ser sacrificada antes de nacer porque tuvo una crisis a corto plazo y no pudo ser administrada. Sus cachorros muertos (4 machos, 5 hembras) fueron recuperados por disección y ninguno portaba la mutación Sry.					

5 Todas las VelociMice con mutaciones de *Sry* derivadas de células ES de VGB6 eran hembras, como se esperaba por la inactivación de *Sry* (Tabla 2). Aquellos sin mutaciones de *Sry*, pero que portan al menos una copia del LTVEC VG12778 NIH KOMP produjeron solo VelociMice macho (Tabla 2, clones GB4 y GG1). Cuando se reprodujeron para prueba 17 VelociMice B6 hembra mutantes de *Sry*, solo una quedó embarazada después de aproximadamente cuatro meses de preparación del engendramiento (Tabla 3), y esa hembra tuvo que ser sacrificada antes del nacimiento porque tuvo una crisis cerca al término y no pudo dar a luz. Sus cachorros muertos (4 machos, 5 hembras) recuperados por disección eran todos WT; ninguno portaba la mutación de *Sry*. Se llegó a la conclusión de que casi todos los ratones mutantes de *Sry* elaborados a partir de células ES de VGB6 son estériles, lo que está de acuerdo con la literatura sobre las mutaciones de *Sry*. Sin embargo, nuestros datos demostraron resultados muy diferentes con los clones de VGF1.

10 Primero, las células ES de VGF1 se mantuvieron, como es habitual, en nuestro medio de crecimiento de fuerza osmótica de tipo KO-DMEM que es feminizante: algunos de los clones XY microinyectados cultivados en este medio producirán hembras XY fértiles, es decir, un fenómeno de hembra XY, aunque no portan mutaciones. Un ejemplo es el clon TH4, que no tiene ninguna mutación de *Sry*, pero que porta al menos una copia del LTVEC VG12778 NIH KOMP. Este clon produjo 2 VelociMice hembra y 6 macho (Tabla 2). Otros dos clones de VGF1 sin mutaciones de *Sry* (TA3 y TA4, Tabla 2) produjeron solo VelociMice machos. Se quería determinar si las células ES XY de VGF1 con mutaciones en *Sry* pueden también ser feminizadas por el medio. En otras palabras, ¿producirían, a diferencia de las células ES mutantes de *Sry* VGB6, algunas hembras mutantes de *Sry* XY fértiles? (Téngase en cuenta que las células ES de VGB6 no se pueden mantener en medios de baja fuerza osmótica tipo KO-DMEM y retienen la capacidad de producir ratones). La respuesta es sí, como se muestra en la Tabla 3.

15 Se microinyectaron seis clones de células ES de VGF1 con pequeñas eliminaciones inducidas por TALEN que van desde 5 pb hasta más de 1 kb. Todos produjeron VelociMice hembra, 32 de los cuales fueron criados. Sorprendentemente, todos los VelociMice hembra XY mutantes de *Sry* eran fértiles; cada uno produjo al menos una camada (Tabla 3). Muchas de las hembras XY mutantes de *Sry* produjeron varias camadas con tamaños de camada normales, mientras que algunas de las hembras XY produjeron solo una o dos camadas pequeñas. De los 299 ratones F1 de estas crías que habían sido genotipadas, aproximadamente la mitad (146, 49%) son machos XY normales o hembras XX normales. 174 (58%) de los ratones F1 eran hembras fenotípicas, mientras que 125 (42%) eran machos fenotípicos. 26 de las hembras (el 15% de las hembras, el 8,7% de la generación F1 total) eran hembras XY que heredaron un alelo *Sry* mutante. Debido a los eventos meióticos de no disyunción asociados con los ovocitos XY, se observaron varios genotipos aberrantes - XXY, XYY, XO, XXYY - algunos de los cuales incluían alelos *Sry* mutantes en la progenie F1 de VelociMice hembra XY mutantes de *Sry*.

20 Se ha descubierto un método para la creación eficiente de VelociMice hembra XY fértiles a partir de células ES XY. Si se crean mutaciones inactivadoras en el gen *Sry* en células ES que se han mantenido en el medio de cultivo feminizante, se obtiene una alta proporción de ratones hembras XY fértiles que cuando se crían con machos que producen principalmente ratones machos y hembras con cromosomas X y Y normales.

Ejemplo 3. Recuperación de embriones en KO-DMEM o DMEM después de mutaciones inducidas por TALEN en el Gen *Sry* del cromosoma Y

40 El direccionamiento correcto del *Sry* de ratón por LTVEC se confirmó o se negó mediante el genotipado de la descendencia F1 derivada de hembras F0, que eran XY y portaban la mutación de *Sry*. La segregación conjunta en ratones F1 del casete LacZ/Neo con la mutación de *Sry* (según lo evaluado por los ensayos LOA de *Sry*) sugiere

fuertemente el direccionamiento correcto. La falla de LacZ/Neo para segregarse conjuntamente con la mutación indica que el clon original contenía una mutación de eliminación de *Sry* (inducida por TALEN) acoplada con una inserción transgénica LacZ/Neo en otras partes del genoma.

5 Los descendientes de hembras XY con mutaciones de *Sry* exhibieron una variedad de cariotipos anormales con una frecuencia alta (incluidos XXY, XYY y XO). El recuento de cromosomas sexuales se evaluó mediante el uso de pérdida no relacionada de ensayos de alelos (LOA) para los genes en los cromosomas X e Y. El número de copias de *Sry* se determinó luego utilizando los ensayos de LOA. La presencia del alelo *Sry* mutante se infirió en ratones en los que el número de copias del cromosoma Y superó el número de copias de *Sry* (por ejemplo, 1 copia de Y y 0 copias de *Sry*, o 2 copias de Y y 1 copia de *Sry*). Finalmente, la presencia de LacZ y Neo se determinó mediante ensayos TaqMan.

En el conjunto original de clones, que fueron creados por LTVEC de *Sry* junto con la nucleasa TALEN y cultivados en KO-DMEM, fue evidente que el casete LacZ/Neo no estaba segregándose conjuntamente con la mutación de *Sry*. Una camada de muestra de estos clones se muestra en la Tabla 4.

Tabla 4: Selección de clones generados por LTVEC de *Sry* junto con la nucleasa TALEN

Ratón	Sexo	Cromos. X Copia #	Cromos. Y Copia #	<i>Sry</i> Copia #	LacZ	Neo	Genotipo	Comentarios
1656721	M	1	1	1	0	0	X+Y+	
1656722	M	1	1	1	1	1	X+Y+	LacZ/Neo presente pero mutación de <i>Sry</i> ausente
1656723	M	1	1	1	0	0	X+Y+	
1656724	M	1	2	1	0	0	X+Y+YΔ	Mutación de <i>Sry</i> presente pero LacZ/Neo ausente
1656725	F	2	0	0	1	1	X+X+	LacZ/Neo presente pero mutación de <i>Sry</i> ausente
1656726	F	2	1	0	0	0	X+X+ YΔ	Mutación de <i>Sry</i> presente pero LacZ/Neo ausente
1656727	F	2	1	0	1	1	X+X+ YΔ	
1656728	F	1	0	0	1	1	X+	LacZ/Neo presente pero mutación de <i>Sry</i> ausente
1656729	F	1	0	0	0	0	X+	

15 En el siguiente conjunto de clones, que fueron creados por LTVEC de *Sry* junto con nucleasa TALEN y se cultivaron en DMEM, el casete LacZ/Neo se segregó conjuntamente completamente con la mutación de *Sry*, lo que indica el direccionamiento correcto. Una camada típica de estos clones se muestra en la Tabla 5.

Tabla 5: Resultados de cribado para clones creados por LTVEC de *Sry* junto con nucleasa TALEN

Ratón	Sexo	Cromos. X Copia #	Cromos. Y Copia #	<i>Sry</i> Copia #	LacZ	Neo	Genotipo
1848360	M	1	1	1	0	0	X+Y+
1848361	M	1	1	1	0	0	X+Y+
1848362	M	1	1	1	0	0	X+Y+
1848363	M	1	1	1	0	0	X+Y+

1848364	M	1	2	1	1	1	X+Y+ YΔ
1848365	F	2	1	0	1	1	X+X+ YΔ
1848366	F	1	1	0	1	1	X+YΔ
1848367	F	1	1	0	1	1	X+X+ YΔ

Ejemplo 4: Direccionamiento asistido por TALEN y CRISPR de *Sry* por TVEC pequeños o LTVEC

5 Como se representa en la FIG. 2, se creó una eliminación objetivo que comprende un alelo de reemplazo de *lacZ* para *Sry*, ya sea con un LTVEC o un vector de direccionamiento pequeño (TVEC pequeño) junto con la nucleasa TALEN o los ARN guía de CRISPR, en combinación con la ADN endonucleasa de Cas9. El TVEC pequeño
 10 comprendía, en orden, un brazo de homología secuencia arriba de aproximadamente 700-800 pb, una secuencia codificante de beta-galactosidasa (*lacZ*) seguida de una señal de poliadenilación, un casete de resistencia a neomicina flanqueado por sitios loxP que comprende un promotor de ubiquitina C humana, incluido el primer exón, primer intrón y parte del segundo exón, una secuencia codificante de neomicina fosfotransferasa, y una señal de poliadenilación y un brazo de homología secuencia abajo de aproximadamente 700-800 pb. El alelo creado mediante el direccionamiento correcto del gen *Sry* con el vector de direccionamiento comprende una eliminación del marco de lectura abierto de *Sry* de aproximadamente 1 kb y el reemplazo por el casete *lacZ*-neo, de manera que la secuencia codificante de la beta-galactosidasa se fusiona en el marco en el codón de inicio de *Sry*. El vector de direccionamiento se usó para dirigir al gen *Sry* en la línea de células ES híbrida F1 C57BL6/129 VGF1 (también conocida como F1H4) y en la línea de células ES de VGB6 (también conocida como B6A6).
 15

Como se ilustra en la Tabla 6, los clones producidos utilizando cuatro ARNg diferentes y un par TALEN se produjeron y se cribaron para la escisión y pérdida del alelo mediante los ensayos TaqMan.

Tabla 6: Resultados de cribado para escisión y pérdida del alelo

				Direccionamiento de TVEC pequeño	
Localización	ARNg o TALEN	Clones cribados	KO	Eficiencia total de direccionamiento (%)	
Caja HMG	ARNg 2	192	4	2,1	
Caja HMG	ARNg 3	192	5	2,6	
Extremo 3'	ARNg 4	192	3	1,6	
Extremo 3'	ARNg 5	192	5	2,6	
Caja HMG	TALEN par 1	384	1	0,3	
n.a.	ninguno	384	1	0,3	

20 Los clones transgénicos del LTVEC produjeron embriones con el mismo patrón de patrón de *lacZ*. La FIG. 3 ilustra la expresión de *LacZ* en los embriones.

La Tabla 7 reporta sobre los resultados de fertilidad de las hembras XY derivadas de células ES cultivadas en un medio convencional basado en DMEM que tenía mutaciones de *Sry* escogidas de eliminación-reemplazo de LTVEC asistidas por TALEN. Se compararon inesperadamente con los resultados de un experimento similar con células ES cultivadas en medio basado en KO-DEMEM (Tabla 3), el direccionamiento de LTVEC en el medio basado en DMEM produjo clones con eliminaciones de *Sry* correctamente dirigidas e inserciones de *lacZ*-neo. Cuarenta de cada 41 hembras XY^{*Sry(lacZ)*} derivadas de cuatro clones objetivo produjeron crías nacidas vivas tras el apareamiento - una tasa de fertilidad del 98%. Por lo tanto, se han ideado dos nuevas formas de producir hembras XY altamente fértiles a partir de células ES mutantes: (1) mutaciones en *Sry* inactivadoras inducidas por TALEN en células ES cultivadas en un medio basado en KO-DEMEM; y (2) mutaciones de eliminación-reemplazo preciso dirigido por LTVEC asistida por TALEN en células ES cultivadas en medio basado en DMEM.
 30

Tabla 7: Producción de hembras XY mutantes de *Sry* por TALEN

Clon	Línea de células	Descripción	VelociMice hembra	Crías	Hembra XY	Tasas de
------	------------------	-------------	-------------------	-------	-----------	----------

ES 2 694 629 T3

	ES	del alelo	XY	hembra XY	fértiles	fertilidad (%)	
DMEM	X-C4	VGf1	lacZ-neo dirigido	5	2	2	100
	X-E10	VGf1	lacZ-neo dirigido	1	1	1	100
	X0F3	VGf1	lacZ-neo dirigido	5	3	3	100
	X-G3	VGf1	lacZ-neo dirigido	9	3	2	67
		VGf1 Total		53	41	40	98

Ejemplo 5: Gran eliminación en el cromosoma Y mediada por las ZFN

Como se ilustra en la Figura 4, grandes eliminaciones, de 500 kb o más, se hicieron en el cromosoma Y utilizando las ZFN dirigidas a los genes Kdm5d y Usp9y. La Tabla 8 proporciona ejemplos de secuencias de dedo de cinc en el cromosoma Y.

5

Tabla 8: Secuencia de dedo de cinc en el cromosoma Y

Nombre objetivo	Placa cromosoma Y	Secuencia de dedo de cinc	ZFN #	SEQ ID NO:
KDM5D	NM011419-r43102a1	ttAGGTAGGTAGACAGGGATgttttctg	ZFN1	42
	NM011419-43108a1	atCCAGTCtCTGAAGGAAGCTctgacta	ZFN2	43
	NM011419-r19880a1	caAAAGCTTCAGGGGGActcttactc	ZFN3	44
	NM011419-19887a1	ttTGAGCAgGCTACACAGGAGtatactt	ZFN4	45
	NM011419-r17347a1	aaGCGGTGgCAATAGGCAaaagatgtgg	ZFN5	46
	NM011419-17353a1	ctGAAGTCCCCAAGGGAGTAtggagatg	ZFN6	47
	NM011419-r17350a1	agAAAGCGGTGGCAaTAGGCAaaagatg	ZFN7	48
	NM011419-17356a1	aaGTCCCCAAGGGAGTAtggagatgccc	ZFN8	49
DDX3Y	NM012008-r8130a1	acTCCAACGACTATGACcactccgttca	ZFN1	50
	NM012008-8136a1	acAGATCAGATGAAGATgactggtcaaa	ZFN2	51
	NM012008-r7172a1	ctTTCAAGGAAAAAAGaacaaaacca	ZFN3	52
	NM012008-7178a1	ggTCTGTGATAAGGACAGTTcaggatgg	ZFN4	53
	NM012008-r20472a1	taAATCTGACTGAGAATGGGtagtagaa	ZFN5	54
	NM012008-20479a1	caGATGGTCCAGGAGAGGCTtgaaggc	ZFN6	55
	NM012008-r7267a1	at TGGGCTTCCcTCTGGAat cacgagat	ZFN7	56
	NM012008-7274a1	ttTCAGTGATCGTGGAAAGTGgatccagg	ZFN8	57
USP9Y	NM148943-r92561a1	ctGGTTTGGAAATCGTActgtaaaagac	ZFN1	58
	NM148943-92567a1	gcAAAGAGGTTGAGGATtggacatatt	ZFN2	59

	NM148943-r11830a1	gaGGAGTTGTTGGAGAAGTCcattgga	ZFN3	60
	NM148943-11836a1	atATGAACAAGGCCAAGgtgatgctcca	ZFN4	61
	NM148943-r108581a1	acTCAGAAGAAGGATTAGGAatgctttg	ZFN5	62
	NM148943-108588a1	atGCTTAGaAATGTATCAGTTcatcttg	ZFN6	63
	NM148943-r16244a1	tcCATAAGGATTTTGGAAAagacacag	ZFN7	64
	NM148943-16251a1	agGCTGTGAGTGATGGAAGttgaaat	ZFN8	65

5 En un experimento, se electroporaron 3,3 millones de células ES de los clones D-G5 y E-G7 de VGB6 (Tabla 3) con los pares de ARNm de ZFN Kdm5d-ZFN5 (NM011419-r17347a1)/ZFN6 (NM011419-17353a1) y Usp9y-ZFN3 (NM148943-r11830a1)/ZFN4 (NM148943-11836a1) (10 µg cada uno) y con un LTVEC dirigido al gen Ch25 (0,67 µg), para proporcionar una selección de resistencia a la puromicina. Las colonias resistentes a puromicina se seleccionaron y se cribaron para la eliminación. Los resultados se muestran en la Tabla 9.

Tabla 9: Resultados del cribado para la eliminación del cromosoma Y grande en 12778D-G5 y 12778E-G7

Clon parental	# de colonias resistentes a puromicina	# de colonias cribadas	# de clones eliminados confirmados
12778D-G5	244	192	4
12778E-G7	638	384	8

10 La Tabla 10 muestra los tamaños exactos de las eliminaciones superiores a 500 kb que se determinaron con precisión para un clon de eliminación (4306A-D5) derivado del clon parental E-G7 (Tabla 3) y dos clones de eliminación (4306E-C4 y 4306F-A12) derivados del clon parental D-G5 (Tabla 3).

Tabla 10: Eliminaciones mediadas por ZFN de Kdm5d y Usp9y

Clon	Coordenadas de eliminación en el cromosoma Y	Tamaño (pb)
4306A-D5	250569-785404	534835
4306E-C4	520363-785402	535039
4306F-A12	250373-785404	535031

15 La eliminación de los genes Kdm5d, Eif2s3y, Uty, Ddx3y y Usp9y (FIG. 4) se confirmó en los ensayos de eliminación de pérdida de alelos y la secuenciación del ADN como se muestra en la FIG. 5. El clon 4306A-D5 produjo nueve VelociMice derivadas de células ES totalmente hembras XY tras la microinyección en embriones en estadio de 8 células y se transfirió a madres sustitutas. Ninguna de las hembras XY del clon 4306A-D5 era fértil.

Ejemplo 6: Gran eliminación en el cromosoma Y mediado por CRISPR/Cas

20 Se realizó una gran eliminación en el cromosoma Y dirigida a la región entre los genes *Kdm5d* y *Usp9y* utilizando ARN guía de CRISPR en combinación con la ADN endonucleasa de Cas9. Los ARNg se diseñaron para dirigir al gen *Kdm5d* y al gen *Usp9y*. Los siguientes ARNg se diseñaron para dirigir *Kdm5d*: Kdm5dgA (Guía # 1) UUUGCCGAUAUGCUCUCGU (SEQ ID NO: 66); Kdm5dgB (Guía # 2) UUGCCGAUAUGCUCUCGUG (SEQ ID NO: 67); y Kdm5dgC (Guía # 5) CGGGCAUCUCCACAACACCCU (SEQ ID NO: 68). Los siguientes ARNGs se diseñaron para dirigir *Usp9y*: Usp9ygA (Guía # 1) UAGCUCGUUGUGUAGCACCU (SEQ ID NO: 69); Usp9ygB (Guía # 1) UAUAGUUUCUUCGGGGUAAC (SEQ ID NO: 70); y Usp9ygC (Guía # 2) GGAUACCCUUCUAUAGGCC (SEQ ID NO: 71).

Las células ES de ratón VGF1 se sometieron a electroporación con 5 µg de un plásmido que expresaba Cas9 y 10 µg de cada uno de los plásmidos que expresaban el ARNg B de *Kdm5d* y el ARNg C de *Usp9y* y con un LTVEC que dirige al gen *Ch25h* (0,67 µg), para proporcionar una selección de resistencia a la puromicina.

30 Como se ilustra en la FIG. 4, se usaron Kdm5dgB (ARNg B) y Usp9ygC (ARNg C) para dirigir la eliminación de los genes *Kdm5d* y *Usp9y*. Los clones resultantes se cribaron para su eliminación mediante ensayos de pérdida de alelo

para las secuencias en los genes *Kdm5d* y *Usp9y* y los genes intermedios (*Eif2s3y*, *Uty* y *Ddx3y*) y para los genes fuera de la eliminación dirigida (*Zfy2* y *Sry*). Como se muestra en la Tabla 11, se obtuvieron cuatro clones que comprenden la gran eliminación. El clon R-A8 produjo VelociMice derivadas de siete células ES totalmente macho XY y tres hembra XY tras la microinyección en embriones en estadio de 8 células y se transfirió a madres sustitutas.

5 Tabla 11: Ensayo TaqMan que confirma la eliminación grande mediada por los ARN guía de CRISPR y Cas9

Clon	Determinación del número de copia de la pérdida de alelo					Nota
	19178TD (<i>Eif2s3y</i>)	16697TD (<i>Uty</i>)	Ddx3yZF12 (<i>Ddx3y</i>)	<i>Zfy2</i>	<i>Sry</i>	
Q-F1	0	0	0	1	1	Gran eliminación
R-A8	0	0	0	1	1	Gran eliminación
R-C2	0	0	0	~0,5	1	Gran eliminación, pérdida parcial de Y
R-E11	1	1	1	1	1	Clon añadido como control de tipo silvestre

10 Todas las publicaciones y solicitudes de patente mencionadas en la memoria descriptiva son indicativas del nivel de los expertos en la técnica a los que pertenece esta invención. Si la información asociada con una cita, tal como un número de depósito cambia con el tiempo, se pretende la versión de la información vigente en la fecha de presentación efectiva de la solicitud, la fecha de presentación efectiva significa la fecha de presentación real o la fecha de una solicitud de prioridad proporcionando primero la cita. A menos que sea evidente en el contexto de cualquier realización, aspecto, etapa o característica de la invención puede usarse 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.

15 Listado de secuencias

<110> Friendewey, David

Droguett, Gustavo

Gagliardi, Anthony

20 Kuno, Junko

Auerbach, Wojtek

Valenzuela, David m.

<120> MÉTODOS Y COMPOSICIONES PARA LAS MODIFICACIONES GENÉTICAS OBJETIVO Y MÉTODOS DE USO.

25 <130> 057766-463545

<150> US 62/017.582

<151> 2014-06-26

<150> US 62/017.627

<151> 2014-06-26

30 <160> 71

<170> FastSEQ para Windows versión 4.0

<210> 1

<211> 23

<212> ADN

35 <213> un locus objetivo que está unido a un ARN guía artificial

<220>
 <223> un locus objetivo que está unido a un ARN guía (ARNg)
 <220>
 <221> característica nueva

5 <222> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
 <223> n = A, T, C o G
 <400> 1 gnnnnnnnnn nnnnnnnnnn ngg 23
 <210> 2
 <211> 80

10 <212> ARN
 <213> Secuencia artificial
 <220>
 <223> un ARN guía (ARNg)
 <400> 2

15 **guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60**
ggcaccgagu cggugcuuuu 80

<210> 3
 <211> 42
 <212> ARN
 <213> Secuencia artificial

20 <220>
 <223> un ARN guía (ARNg)
 <400> 3

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cg 42
 <210> 4

25 <211> 30
 <212> ARN
 <213> Secuencia artificial
 <220>
 <223> un ARNcr

30 <400> 4

guuuuagagc uagaaauagc aaguuaaaau 30
 <210> 5
 <211> 33
 <212> ARN

35 <213> Secuencia artificial
 <220>

- <223> un ARNcr
 <400> 5
 guuuuagagc uagaaauagc aaguuaaaau aag 33
 <210> 6
- 5 <211> 26
 <212> ARN
 <213> Secuencia artificial
 <220>
 <223> un ARNcr
- 10 <400> 6
 gaguccgagc agaagaagaa guuuua 26
 <210> 7
 <211> 12
 <212> ARN
- 15 <213> Secuencia artificial
 <220>
 <223> un ARNcrtra
 <400> 7
 aaggcuaguc cg 12
- 20 <210> 8
 <211> 50
 <212> ARN
 <213> Secuencia artificial
 <220>
- 25 <223> un ARNcrtra
 <400> 8
 aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 50
 <210> 9
 <211> 23
- 30 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> un locus objetivo que está unido a un ARN guía (ARNg)
 <220>
- 35 <221> característica nueva
 <222> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
 <223> n = A, T, C o G

<400> 9
gnnnnnnnnn nnnnnnnnnn ngg 23
<210> 10
<211> 23
5 <212> ADN
<213> Secuencia artificial
<220>
<223> Secuencia objetivo sintética de ARNg de VG-1
<400> 10
10 ccatgaatgc atttatggtg tgg 23
<210> 11
<211> 23
<212> ADN
<213> Secuencia artificial
15 <220>
<223> Secuencia objetivo sintética de ARNg de VG-2
<400> 11
ccgtggtgag aggcacaagt tgg 23
<210> 12
20 <211> 23
<212> ADN
<213> Secuencia artificial
<220>
<223> Secuencia objetivo sintética de ARNg de VG-3
25 <400> 12
gcaagcagct gggatgcagg tgg 23
<210> 13
<211> 20
<212> ADN
30 <213> Secuencia artificial
<220>
<223> Cebador sryTF1 sintético
<400> 13
cggtgtgaga ggcacaagtt 20
35 <210> 14
<211> 19
<212> ADN

<213> Secuencia artificial
 <220>
 <223> Cebador sryTR1 sintético
 <400> 14
 5 gagatcagca agcagctgg 19
 <210> 15
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> Sonda sryTP1 sintética
 <400> 15
 cccagcagaa tcccagcatg ca 22
 <210> 16
 15 <211> 18
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador sryTF2 sintético
 20 <400> 16
 tggagggccca tgtcaagc 18
 <210> 17
 <211> 18
 <212> ADN
 25 <213> Secuencia artificial
 <220>
 <223> Cebador sryTR2 sintético
 <400> 17
 acaagttggc ccagcaga 18
 30 <210> 18
 <211> 26
 <212> ADN
 <213> Secuencia artificial
 <220>
 35 <223> Sonda sryTP2 sintética
 <400> 18
 tgaatgcatt tatggtgtgg tcct 26

<210> 19
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 5 <220>
 <223> Cebador sryTF3 sintético
 <400> 19
 atgaatgcat ttatggtg 20
 <210> 20
 10 <211> 18
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador sryTR3 sintético
 15 <400> 20
 aggtggaaaa gcctaca 18
 <210> 21
 <211> 23
 <212> ADN
 20 <213> Secuencia artificial
 <220>
 <223> Sonda sryTP3 sintética
 <400> 21
 ccgtggtgag aggcacaagt tgg 23
 25 <210> 22
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 <220>
 30 <223> Cebador sryTF4 sintético
 <400> 22
 gtgtgtccc gtgtgaga 19
 <210> 23
 <211> 20
 35 <212> ADN
 <213> Secuencia artificial
 <220>

<223> Cebador sryTR4 sintético
 <400> 23
 agatcagcaa gcagctggga 20
 <210> 24
 5 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sonda sryTP4 sintética
 10 <400> 24
 aagttggccc agcagaatcc cagc 24
 <210> 25
 <211> 24
 <212> ADN
 15 <213> Secuencia artificial
 <220>
 <223> Cebador sryTF5 sintético
 <400> 25
 catgcaaaat acagagatca gcaa 24
 20 <210> 26
 <211> 14
 <212> ADN
 <213> Secuencia artificial
 <220>
 25 <223> Cebador sryTR5 sintético
 <400> 26
 tinción ggaaaagcct 14
 <210> 27
 <211> 15
 30 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sonda sryTP5 sintético
 <400> 27
 35 cagctgggat gcagg 15
 <210> 28
 <211> 20

ES 2 694 629 T3

<212> ADN
<213> Secuencia artificial
<220>
<223> Cebador 648TUF sintético

5 <400> 28
gatcagcaag cagctgggat 20
<210> 29
<211> 19
<212> ADN

10 <213> Secuencia artificial
<220>
<223> Sonda 648TUP sintética
<400> 29
caggtggaaa agccttaca 19

15 <210> 30
<211> 73
<212> ADN
<213> Secuencia artificial
<220>

20 <223> Sintética
<400> 30
aattaccaca tcttttgcct attgccaccg cttgaggagt tgttgagaa gtctcattgg 60
agaatcaga tga 73

<210> 31
<211> 65

25 <212> ADN
<213> Secuencia artificial
<220>
<223> Sintético
<400> 31
aattaccaca tctttttgcc tattgtccaa atactggaga agtctcattg caacgaatca 60
gatga 65

30 <210> 32
<211> 63
<212> ADN
<213> Secuencia artificial

35 <220>

ES 2 694 629 T3

<223> Sintético
<400> 32
aattaccaca tcttttgcct attgacaaaa tgttgagaa gtctcatggg aagaatcaaa 60
ttg 63
<210> 33
5 <211> 90
<212> ADN
<213> Secuencia artificial
<220>
<223> Sintético
10 <400> 33
ctatcagata ggatatttta gagttttcat attgtatgga ggagttggtg gagaagtctc 60
attggaagaa tcagatgatg agatcttaat 90
<210> 34
<211> 70
<212> ADN
15 <213> Secuencia artificial
<220>
<223> Sintético
<400> 34
ctatcagata ggatatttta gagttggttg agaagtctca ttggaagaa tcagatgatg 60
agatcttaat 70
20 <210> 35
<211> 69
<212> ADN
<213> Secuencia artificial
<220>
25 <223> Sintético
<400> 35
ctatcagata ggatatttta gagttggttg agaagtctca ttggaagaat cagatgatga 60
gatcttaat 69
<210> 36
30 <211> 69
<212> ADN
<213> Secuencia artificial
<220>
<223> Sintético
35 <400> 36

ES 2 694 629 T3

ctatcacata ggatatttta gagttggttg cgaagtctca ttggaagaat cagatgatga 60
gatcttaat 69

<210> 37

<211> 69

<212> ADN

5 <213> Secuencia artificial

<220>

<223> Sintético

<400> 37

tcagatagga tatttttagag ttttcatatt gtatggttat aggaggagtt gttggagaag 60
tctcattgg 69

10 <210> 38

<211> 49

<212> ADN

<213> Secuencia artificial

<220>

15 <223> Sintético

<400> 38

tcagatagga tatttttagag ttttcatatt gttggagaag tctcattgg 49

<210> 39

<211> 49

20 <212> ADN

<213> Secuencia artificial

<220>

<223> Sintético

<400> 39

25 tcagataggg tatttttagag ttttcatatt gttggagaag tctcatggg 49

<210> 40

<211> 49

<212> ADN

<213> Secuencia artificial

30 <220>

<223> Sintético

<400> 40

tccatagga tatttttagag ttttcatatt gctgccgaag tctcattgc 49

<210> 41

35 <211> 38

- <212> ADN
- <213> Secuencia artificial
- <220>
- <223> Sintético
- 5 <400> 41
 ttttagagtt tttcatattg ttgggagaag tctattgg 38
- <210> 42
- <211> 28
- <212> ADN
- 10 <213> Secuencia artificial
- <220>
- <223> Secuencia objetivo sintética de ZFN1 de KDM5D
- <400> 42
 ttaggtaggt agacagggat gtttctg 28
- 15 <210> 43
- <211> 28
- <212> ADN
- <213> Secuencia artificial
- <220>
- 20 <223> Secuencia objetivo sintética de ZFN2 de KDM5D
- <400> 43
 atccagtctc tgaaggaagc tctgacta 28
- <210> 44
- <211> 28
- 25 <212> ADN
- <213> Secuencia artificial
- <220>
- <223> Secuencia objetivo sintética de ZFN3 de KDM5D
- <400> 44
- 30 caaaagcttc agggggactc ttacactc 28
- <210> 45
- <211> 28
- <212> ADN
- <213> Secuencia artificial
- 35 <220>
- <223> Secuencia objetivo sintética de ZFN4 de KDM5D
- <400> 45

tttgagcagg ctacacagga gtatactt 28
 <210> 46
 <211> 28
 <212> ADN
 5 <213> Secuencia artificial
 <220>
 <223> Secuencia objetivo sintética de ZFN5 de KDM5D
 <400> 46
 aagcggtagc aataggcaaa agatgtgg 28
 10 <210> 47
 <211> 28
 <212> ADN
 <213> Secuencia artificial
 <220>
 15 <223> Secuencia objetivo sintética de ZFN6 de KDM5D
 <400> 47
 ctgaagtccc caaggagta tggagatg 28
 <210> 48
 <211> 28
 20 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Secuencia objetivo sintética de ZFN7 de KDM5D
 <400> 48
 25 agaaagcggg ggcaatagc aaaagatg 28
 <210> 49
 <211> 28
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> Secuencia objetivo sintética de ZFN8 de KDM5D
 <400> 49
 aagtcccaa gggagtatg agatgcc 28
 <210> 50
 35 <211> 28
 <212> ADN
 <213> Secuencia artificial

- <220>
<223> Secuencia objetivo sintética de ZFN1 de DDX3Y
<400> 50
actccaacga ctatgaccac tccgttca 28
- 5 <210> 51
<211> 28
<212> ADN
<213> Secuencia artificial
<220>
- 10 <223> Secuencia objetivo sintética de ZFN2 de DDX3Y
<400> 51
acagatcaga tgaagatgac tggcaaaa 28
<210> 52
<211> 28
- 15 <212> ADN
<213> Secuencia artificial
<220>
<223> Secuencia objetivo sintética de ZFN3 de DDX3Y
<400> 52
- 20 ctttcaagga aaaaaagaac aaaaccca 28
<210> 53
<211> 28
<212> ADN
<213> Secuencia artificial
- 25 <220>
<223> Secuencia objetivo sintética de ZFN4 de DDX3Y
<400> 53
ggtctgtgat aaggacagtt caggatgg 28
<210> 54
- 30 <211> 28
<212> ADN
<213> Secuencia artificial
<220>
<223> Secuencia objetivo sintética de ZFN5 de DDX3Y
- 35 <400> 54
taaactgac tgagaatggg tagtagaa 28
<210> 55

- <211> 28
<212> ADN
<213> Secuencia artificial
<220>
- 5 <223> Secuencia objetivo sintética de ZFN6 de DDX3Y
<400> 55
cagatggtcc aggagaggct ttgaaggc 28
<210> 56
<211> 28
- 10 <212> ADN
<213> Secuencia artificial
<220>
<223> Secuencia objetivo sintética de ZFN7 de DDX3Y
<400> 56
- 15 attgggcttc cctctggaat cacgagat 28
<210> 57
<211> 28
<212> ADN
<213> Secuencia artificial
- 20 <220>
<223> Secuencia objetivo sintética de ZFN8 de DDX3Y
<400> 57
tttcagtgat cgtggaagtg gatccagg 28
<210> 58
- 25 <211> 28
<212> ADN
<213> Secuencia artificial
<220>
<223> Secuencia objetivo sintética de ZFN1 de USP9Y
- 30 <400> 58
ctggtttga aatcgtagt taaagac 28
<210> 59
<211> 28
<212> ADN
- 35 <213> Secuencia artificial
<220>
<223> Secuencia objetivo sintética de ZFN2 de USP9Y

<400> 59
gcaaagaggt tgaggattg gacatatt 28
<210> 60
<211> 28
5 <212> ADN
<213> Secuencia artificial
<220>
<223> Secuencia objetivo sintética de ZFN3 de USP9Y
<400> 60
10 gaggagttgt tggagaagtc tcattgga 28
<210> 61
<211> 28
<212> ADN
<213> Secuencia artificial
15 <220>
<223> Secuencia objetivo sintética de ZFN4 de USP9Y
<400> 61
atatgaacaa ggccaaggtg atgctcca 28
<210> 62
20 <211> 28
<212> ADN
<213> Secuencia artificial
<220>
<223> Secuencia objetivo sintética de ZFN5 de USP9Y
25 <400> 62
actcagaaga aggattagga atgctttg 28
<210> 63
<211> 28
<212> ADN
30 <213> Secuencia artificial
<220>
<223> Secuencia objetivo sintética de ZFN6 de USP9Y
<400> 63
atgcttagaa atgtatcagt tcatcttg 28
35 <210> 64
<211> 28
<212> ADN

<213> Secuencia artificial
 <220>
 <223> Secuencia objetivo sintética de ZFN7 de USP9Y
 <400> 64
 5 tccataagga ttttgaaaa agacacag 28
 <210> 65
 <211> 28
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> Secuencia objetivo sintética de ZFN8 de USP9Y
 <400> 65
 aggctgtgag tggatggaag tttgaaat 28
 <210> 66
 15 <211> 20
 <212> ARN
 <213> Secuencia artificial
 <220>
 <223> ARNg 1 sintético de Kdm5dgA
 20 <400> 66
 uuugccgaau augcucucgu 20
 <210> 67
 <211> 20
 <212> ARN
 25 <213> Secuencia artificial
 <220>
 <223> ARNg 2 sintético de Kdm5dgB
 <400> 67
 uugccgaaua ugcucucgug 20
 30 <210> 68
 <211> 20
 <212> ARN
 <213> Secuencia artificial
 <220>
 35 <223> ARNg 5 sintético de Kdm5dgC
 <400> 68
 cgggcaucuc cauacuccu 20

<210> 69
<211> 20
<212> ARN
<213> Secuencia artificial
5 <220>
<223> ARNg 1 sintético de Usp9ygA
<400> 69
uagcucguug uguagcaccu 20
<210> 70
10 <211> 20
<212> ARN
<213> Secuencia artificial
<220>
<223> ARNg 1 sintético de Usp9ygB
15 <400> 70
uauaguuuucu ucgggguaac 20
<210> 71
<211> 20
<212> ARN
20 <213> Secuencia artificial
<220>
<223> ARNg 2 sintético de Usp9ygC
<400> 71
ggauaccuu cuauaggccc 20
25

REIVINDICACIONES

1. Un método para modificar un locus genómico objetivo en un cromosoma Y en una célula, que comprende:
- (a) proporcionar la célula que comprende el locus genómico objetivo en el cromosoma Y, en donde el locus genómico objetivo comprende un sitio de reconocimiento para un agente de nucleasa, y en el que la célula está en un cultivo que comprende un medio base DMEM;
- (b) la introducción en la célula:
- (i) del agente de nucleasa, en el que el agente de nucleasa induce un corte o ruptura de la cadena doble en el sitio de reconocimiento; y
- (ii) de un vector de direccionamiento grande que comprende un polinucleótido de inserción flanqueado por los brazos de homología primero y segundo correspondientes a los sitios objetivo primero y segundo ubicados dentro del locus genómico objetivo, en donde la suma total del primer brazo de homología y el segundo brazo de homología es al menos de 10 kb, y en el que el vector de direccionamiento experimenta recombinación homóloga con el locus genómico objetivo; e
- (c) identificar al menos una célula que comprende en su genoma el polinucleótido de inserción integrado en el locus genómico objetivo, en donde la integración del polinucleótido de inserción introduce una modificación genética que comprende la eliminación de una secuencia endógena de ácido nucleico y la sustitución con una secuencia exógena de ácido nucleico en el locus genómico objetivo,
- en el que la célula no se produce mediante un proceso que implica modificar la identidad genética de la línea germinal de los seres humanos o que implica el uso de un embrión humano para fines industriales o comerciales, y
- en el que el método no es un método para el tratamiento del cuerpo humano o animal mediante terapia.
2. El método de la reivindicación 1, en el que la suma total del primer brazo de homología y el segundo brazo de homología es inferior a 150 kb.
3. El método de la reivindicación 1 o 2, en el que la célula es una célula de mamífero, opcionalmente en la que la célula de mamífero es de un roedor, y opcionalmente en la que el roedor es una rata o un ratón.
4. El método de una cualquiera de las reivindicaciones 1-3, en el que la célula es una célula pluripotente, opcionalmente en la que la célula pluripotente es una célula madre pluripotente inducida (iPS) o una célula madre (ES) embrionaria no humana, opcionalmente en el que célula ES no humana es una célula ES de roedor, una célula ES de rata o una célula ES de ratón.
5. El método de la reivindicación 4, en el que la célula es una célula ES de ratón.
6. El método de una cualquiera de las reivindicaciones 1-5, en el que el agente de nucleasa es:
- (a) una nucleasa de dedo de cinc (ZFN);
- (b) una nucleasa efectora de tipo activador de transcripción (TALEN);
- (c) una meganucleasa;
- (d) una proteína (Cas) asociada a las repeticiones palindrómicas cortas agrupadas y regularmente espaciadas (CRISPR) y un ARN guía (ARNg), o
- (e) un ARNm que codifica una nucleasa.
7. El método de la reivindicación 6, en el que el agente de nucleasa es la proteína Cas y el ARNg, en el que la proteína Cas es Cas9, y en el que el ARNg comprende:
- (a) un ARN de CRISPR (ARNcr) que se dirige al sitio de reconocimiento, en el que el sitio de reconocimiento está inmediatamente flanqueado por una secuencia del motivo adyacente protoespaciador (PAM); y
- (b) un ARN de CRISPR transactivador (ARNcrtra).
8. El método de una cualquiera de las reivindicaciones 1-7, en el que el polinucleótido de inserción es de aproximadamente 5 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 de longitud.
9. El método de una cualquiera de las reivindicaciones 1-8, en el que el polinucleótido de inserción comprende un alelo condicional, un polinucleótido que codifica un marcador de selección, un ácido nucleico flanqueado por

- 5 secuencias objetivo de recombinación específicas del sitio, o un gen informador unido operativamente a un promotor, en el que el gen informador codifica una proteína indicadora seleccionada del grupo que consiste en LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, proteína fluorescente amarilla mejorada (EYFP), Emerald, proteína fluorescente verde mejorada (EGFP), CyPet, proteína fluorescente cian (CFP), Cerulean, T-Sapphire, luciferasa, fosfatasa alcalina y una combinación de las mismas.
10. El método de una cualquiera de las reivindicaciones 1-9, en el que la modificación comprende un intercambio de dominio, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio de genes, o una combinación de los mismos.
- 10 11. El método de una cualquiera de las reivindicaciones 1-10, en el que la secuencia endógena de ácido nucleico eliminada 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, o de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 300 kb, de aproximadamente 300 kb a 15 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.
- 20 12. El método de una cualquiera de las reivindicaciones 1-11, en el que la secuencia endógena de ácido nucleico eliminada es de al menos 500 kb.
13. El método de una cualquiera de las reivindicaciones 1-12, en el que la modificación comprende un reemplazo de una secuencia endógena de ácido nucleico con una secuencia de ácido nucleico homóloga u ortóloga.
- 25 14. El método de una cualquiera de las reivindicaciones 1-13, en el que el locus genómico objetivo en el cromosoma Y es el gen de *Sry*, el gen *Uty*, el gen *Eif2s3y*, el gen *Ddx3y*, el gen *Ube1y*, el gen *Tspy*, el gen *Usp9y*, el gen *Zfy1*, el gen *Zfy2*, o la región que abarca los genes *Kdm5d*, *Eif2s3y*, *Tspy*, *Uty*, *Ddx3y* y *Usp9y*.
15. El método de una cualquiera de las reivindicaciones 1-13, en el que el locus genómico objetivo en el cromosoma Y comprende el gen *Sry*.

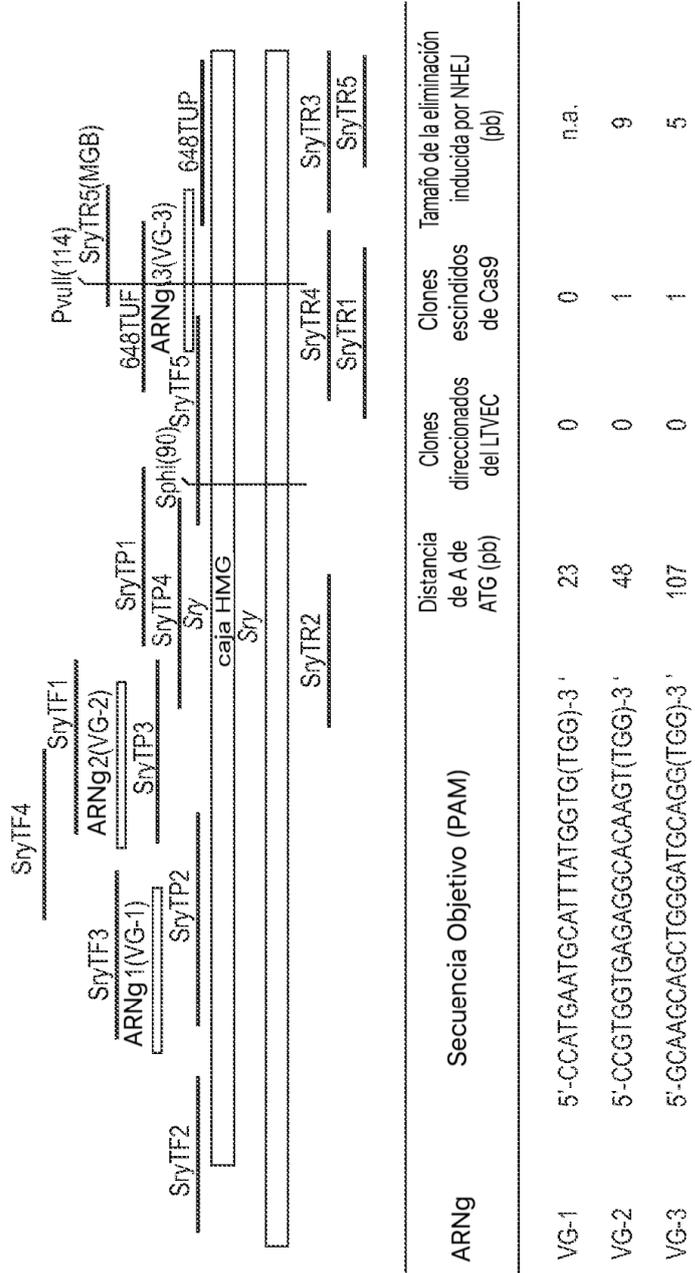


FIG. 1

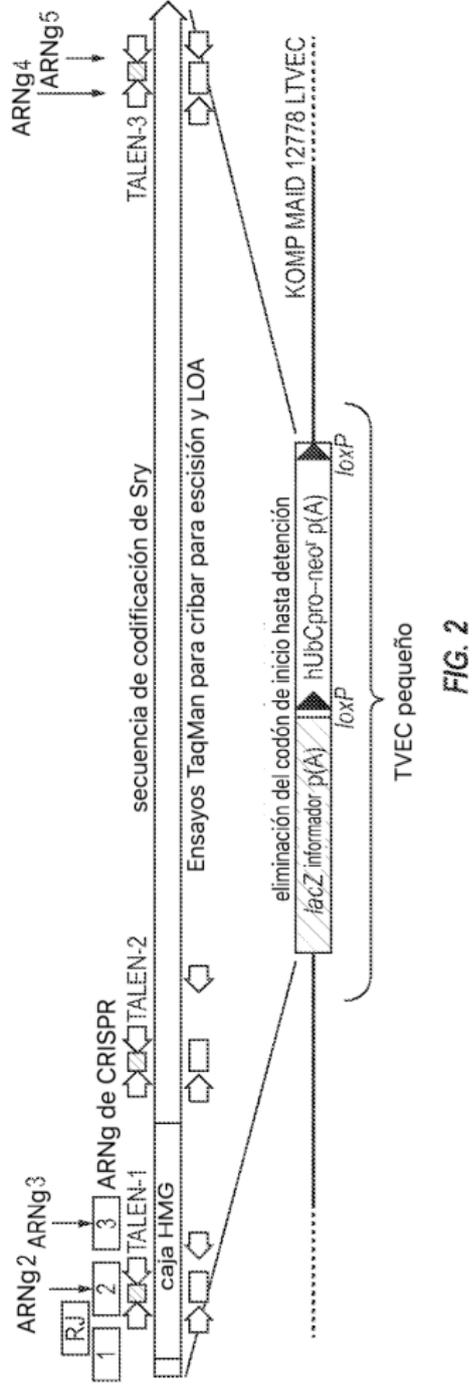
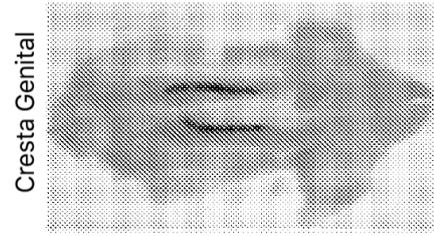


FIG. 2



Montaje Completo

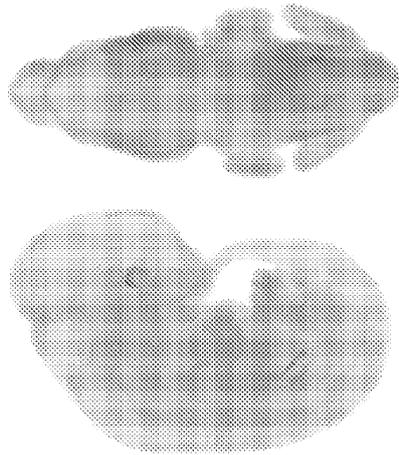


FIG. 3

Tinción de lacZ
≈ E.11.5
21 somitas de cola

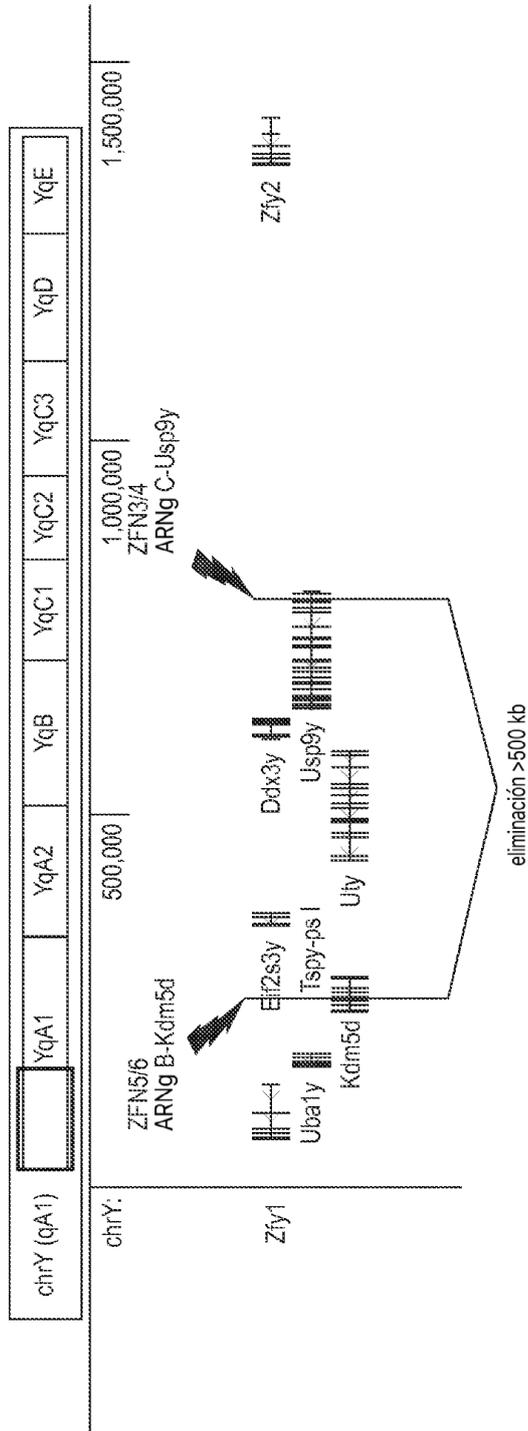


FIG. 4

