(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 888 250**

(51) Int. CI.:

C12N 15/85 (2006.01)
C12N 9/22 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **16.04.2014 E 18187581 (6)**

(97) Fecha y número de publicación de la concesión europea: **14.07.2021 EP 3456831**

(54) Título: **Modificación direccionada del genoma de rata**

(30) Prioridad:

16.04.2013 US 201361812319 P
11.12.2013 US 201361914768 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
03.01.2022

(73) Titular/es:

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

(72) Inventor/es:

LEE, JEFFREY;
MUJICA, ALEXANDER;
AUERBACH, WOJTEK;
LAI, KA-MAN VENUS;
VALENZUELA, DAVID y
YANCOPOULOS, GEORGE

(74) Agente/Representante:

SÁEZ MAESO, Ana

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

Modificación direccionada del genoma de rata

Referencia a una lista de secuencias presentada como un archivo de texto a través de la web de EFS

La copia oficial de la lista de secuencias se envía electrónicamente a través de la web de EFS como una lista de secuencias en formato ASCII con un archivo llamado 444701SEQLIST.TXT, creado el 16 de abril de 2014 y con un tamaño de 15 kilobytes, y se presenta simultáneamente con la especificación.

Campo de la invención

Células madre totipotentes o pluripotentes no humanas aisladas, en particular células madre embrionarias de rata, que son capaces de mantener la pluripotencia después de una o más modificaciones genéticas en serie *in vitro*, y que son capaces de transmitir las modificaciones genéticas direccionadas a las generaciones siguientes a través de la línea germinal. Composiciones y métodos para modificar un locus de interés genómico de rata a través de recombinación homóloga bacteriana (BHR) en una célula procariota. Composiciones y métodos para modificar genéticamente un locus de interés genómico de rata utilizando un vector de direccionamiento grande (LTVEC) en combinación con endonucleasas. Composiciones y métodos para producir una rata modificada genéticamente que comprende una o más modificaciones genéticas específicas.

Antecedentes de la invención

En tanto que las ratas han sido consideradas como un importante sistema de modelo animal que puede recapitular la patología de varias enfermedades humanas, incluidas, entre otras, cardiovasculares (por ejemplo, hipertensión), metabólicas (por ejemplo, obesidad, diabetes), neurológicas (por ejemplo, patologías del dolor) y una variedad de cánceres, el uso de ratas en el modelado de enfermedades humanas ha sido limitado, en comparación con los ratones, debido en parte a la falta de disponibilidad de células pluripotentes transmisibles en la línea germinal, que puedan mantener su pluripotencia después de una serie de modificaciones genéticas *in vitro*, por ejemplo, una o más electroporaciones en serie, y debido en parte a la falta de tecnologías de direccionamiento eficientes que permitan la introducción o eliminación de grandes secuencias de ADN genómico, o el reemplazo de grandes secuencias de ADN genómico endógeno con secuencias de ácidos nucleicos exógenas en células de rata pluripotentes.

Existe una necesidad en la técnica de composiciones y métodos que permitan cambios direccionados precisos en el genoma de una rata, que puedan abrir o expandir áreas actuales de descubrimiento de dianas y validar agentes terapéuticos de forma más rápida y sencilla.

Resumen

La presente invención se refiere a las realizaciones caracterizadas en las reivindicaciones. En particular, se proporcionan métodos para modificar un locus genómico de interés en una célula de rata pluripotente mediante modificación genética direccionada. Tal método comprende (a) introducir en la célula pluripotente un vector de direccionamiento grande (LTVEC) que comprende un inserto de ácido nucleico flanqueado por un brazo de homología 5' y un brazo de homología 3'; y (b) identificar una célula pluripotente modificada genéticamente que comprende la modificación genética direccionada en el locus genómico de interés, en donde la modificación genética direccionada es capaz de ser transmitida a través de la línea germinal.

De acuerdo con la presente invención, la célula pluripotente se deriva de un animal no humano, es decir, una rata.

En particular, la célula pluripotente es una célula pluripotente de rata.

En una realización, la célula pluripotente es un óvulo fertilizado de rata en la etapa de célula única.

En algunas realizaciones, la suma total de los brazos de homología 5' y 3' de LTVEC es al menos 10 kb. En algunas realizaciones, la suma total de los brazos de homología 5' y 3' de LTVEC es de al menos 10 kb, pero menos de 100 kb o la suma total de los brazos de homología 5' y 3' de LTVEC es de al menos 10 kb pero menos de 150kb. En otras realizaciones, el tamaño de la suma total del total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 10kb a aproximadamente 150kb, de aproximadamente 10 kb a aproximadamente 100kb, de aproximadamente 10kb a aproximadamente 75kb, de 20 kb a aproximadamente 150kb, aproximadamente 20kb a aproximadamente 100kb, aproximadamente 20kb a aproximadamente 75 kb, aproximadamente 30 kb a aproximadamente 150kb, aproximadamente 30kb a aproximadamente 100kb, aproximadamente 30kb a aproximadamente 75 kb, aproximadamente 40kb a aproximadamente 150kb, aproximadamente 40kb a aproximadamente 100 kb, aproximadamente 40kb, aproximadamente 75kb, aproximadamente 50kb a aproximadamente 150kb, aproximadamente 50kb a aproximadamente 100kb, o aproximadamente 50 kb a aproximadamente 75kb, aproximadamente 10kb a aproximadamente 30kb, aproximadamente 20kb a aproximadamente 40kb, aproximadamente 40kb a aproximadamente 60kb, aproximadamente 60kb a aproximadamente 80kb, aproximadamente 80kb, aproximadamente 100kb, aproximadamente 100kb a aproximadamente 120kb, o de

aproximadamente 120kb a aproximadamente 150kb. En una realización, el tamaño de la eliminación es el mismo o similar al tamaño de la suma total del homólogo 5' y 3' del LTVEC

En algunas de tales realizaciones, la modificación genética direccionada es bialélica.

En las realizaciones de la presente invención, la célula pluripotente es una célula de rata pluripotente. En una realización, la célula de rata pluripotente es una célula madre embrionaria de rata. En una realización, la célula de rata pluripotente se deriva de una cepa DA o una cepa ACI. En algunas realizaciones, la célula de rata pluripotente se caracteriza por la expresión de al menos un marcador que consiste en Dnmt3L, Eras, Err-beta, Fbxo15, Fgf4, Gdf3, Klf4, Lef1, LIF, Lin28, Nanog, Oct4, Sox15, Sox2, Utf1, o una combinación de los mismos. En algunos de estos métodos, la célula de rata pluripotente se caracteriza por una o más de las siguientes características: (a) falta de expresión de uno o más marcadores de pluripotencia que comprenden c-Myc, Ecat1 y Rexol; o (b) falta de expresión de marcadores mesodérmicos que comprenden Brachyury y/o Bmpr2; (c) falta de expresión de uno o más marcadores endodérmicos que comprenden Gata6, Sox17 y/o Sox7; o (d) falta de expresión de uno o más marcadores neuronales que comprenden Nestin y/o Pax6. Tales métodos proporcionan que la suma total de los brazos de homología 5' y 3' del LTVEC es de aproximadamente 10kb a aproximadamente 30kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, o de aproximadamente 80kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 120kb, de aproximadamente 120kb a aproximadamente 150kb, o de aproximadamente 10 kb pero menos de aproximadamente 150kb. En algunas realizaciones, la suma total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 16 Kb a aproximadamente 100 Kb. En otras realizaciones, el tamaño de la suma total del total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 10kb a aproximadamente 150kb, de aproximadamente 10kb a aproximadamente 100kb, de aproximadamente 10kb a aproximadamente 75kb, de 20kb a aproximadamente 150kb, aproximadamente 20kb a aproximadamente 100kb, aproximadamente 20kb a aproximadamente 75kb, aproximadamente 30kb a aproximadamente 150kb, aproximadamente 30kb a aproximadamente 100kb, aproximadamente 30kb a aproximadamente 75kb, aproximadamente 40kb a aproximadamente 150kb, aproximadamente 40kb a aproximadamente 100kb, aproximadamente 40kb a aproximadamente 75kb, aproximadamente 50kb a aproximadamente 150kb, aproximadamente 50kb a aproximadamente 100kb, aproximadamente 50kb a aproximadamente 75kb, aproximadamente 10kb a aproximadamente 30kb, aproximadamente 20kb a aproximadamente 40kb, aproximadamente 40kb a aproximadamente 60kb, aproximadamente 60kb, aproximadamente 80kb, aproximadamente 80kb a aproximadamente 100kb, aproximadamente 100kb a aproximadamente 120kb, o desde aproximadamente 120kb a aproximadamente 150kb. 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.

Los métodos además proporcionan que la modificación genética direccionada (a) comprende un reemplazo de una secuencia de ácidos nucleicos de rata endógena con una secuencia de ácidos nucleicos de mamífero homóloga u ortóloga; (b) comprende una eliminación de una secuencia endógena de ácido nucleico de rata; (c) comprende una eliminación de una secuencia endógena de ácido nucleico de rata, en donde la eliminación varía de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150kb, o de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 200kb a aproximadamente 300kb, de aproximadamente 300kb a aproximadamente 400kb, de aproximadamente 400kb a aproximadamente 500kb, de aproximadamente 500kb a aproximadamente 1Mb, desde aproximadamente 1Mb hasta aproximadamente 1.5Mb, desde aproximadamente 1.5Mb hasta aproximadamente 2Mb, desde aproximadamente 2Mb hasta aproximadamente 2.5Mb, o desde aproximadamente 2.5Mb hasta aproximadamente 3Mb; (d) comprende una secuencia de ácidos nucleicos exógena que varía de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 200kb a aproximadamente 250kb, de aproximadamente 250kb a aproximadamente 300kb, de aproximadamente 300kb a aproximadamente 350kb, o de aproximadamente 350kb a aproximadamente 400kb; (e) comprende una secuencia de ácidos nucleicos exógena que comprende una secuencia de ácidos nucleicos homóloga u ortóloga; (f) comprende una secuencia de ácidos nucleicos quimérica que comprende una secuencia de ácidos nucleicos humana y de rata; (g) varía de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 200kb a aproximadamente 250kb, de aproximadamente 250kb a aproximadamente 300kb, de aproximadamente 300kb a aproximadamente 350kb, o de aproximadamente 350kb a aproximadamente 400kb; (h) comprende un alelo condicional flanqueado con secuencias diana de recombinasa específicas del sitio; o, (i) comprende un gen informador unido operativamente a un promotor activo en una célula de rata.

Además, se proporciona un método para modificar un locus genómico de interés en una célula de rata pluripotente mediante modificación genética direccionada, en donde el locus genómico de interés comprende (i) una primera

secuencia de ácidos nucleicos que es complementaria al brazo de homología de rata 5'; y (ii) una segunda secuencia de ácidos nucleicos que es complementaria al brazo de homología de rata 3'. En algunas de tales realizaciones, la primera y la segunda secuencia de ácidos nucleicos están separadas por al menos 5kb. En algunas realizaciones, la primera y la segunda secuencia de ácidos nucleicos están separadas por al menos 5kb, pero menos de 3Mb. En algunos de estos métodos, la primera y la segunda secuencia de ácidos nucleicos están separadas por al menos 5kb pero menos de 10kb, al menos 10kb pero menos de 20kb, al menos 20kb pero menos de 40kb, al menos 40kb pero menos de 60kb, al menos 60kb pero menos de 80kb, al menos aproximadamente 80kb pero menos de 100kb, al menos 100kb pero menos de 150kb, o al menos 150kb pero menos de 200kb, al menos aproximadamente 200kb pero menos de aproximadamente 300kb, al menos aproximadamente 300kb pero menos de aproximadamente 400kb, al menos aproximadamente 400kb pero menos de aproximadamente 500kb, al menos aproximadamente 500kb pero menos de aproximadamente 1Mb, al menos aproximadamente 1Mb pero menos de aproximadamente 1.5Mb, al menos aproximadamente 1.5Mb pero menos de aproximadamente 2Mb, al menos aproximadamente 2Mb pero menos de aproximadamente 2.5Mb, al menos aproximadamente 2.5Mb, pero menos de aproximadamente 3Mb, al menos aproximadamente 1Mb pero menos de aproximadamente 2 Mb, al menos aproximadamente 2Mb pero menos de aproximadamente 3Mb.

En algunas realizaciones, la etapa de introducción comprende además la introducción de un segundo ácido nucleico que codifica un agente de nucleasa que promueve una recombinación homóloga entre el constructo de direccionamiento y el locus genómico de interés en la célula de rata pluripotente. En algunas de tales realizaciones, el agente de nucleasa comprende (a) una proteína quimérica que comprende un dominio de unión a ADN basado en dedos de zinc fusionado a una endonucleasa FokI; o, (b) una proteína quimérica que comprende una nucleasa efectora similar a un activador de transcripción (TALEN) fusionada con una endonucleasa FokI.

En algunos métodos, la etapa de introducción comprende además la introducción en la célula de rata pluripotente: (i) un primer constructo de expresión que comprende un primer promotor unido operativamente a una primera secuencia de ácidos nucleicos que codifica una proteína (Cas) asociada a Repeticiones Palindrómicas Cortas Regularmente Interespaciadas Agrupada (CRISPR), (ii) un segundo constructo de expresión que comprende un segundo promotor unido operativamente a una segunda secuencia de ácidos nucleicos que codifica una secuencia diana genómica unida operativamente a un ARN guía (ARNg), en donde la secuencia diana genómica está inmediatamente flanqueada en el extremo 3' por una secuencia *Protospacer Adyac Motif* (PAM). En una realización, el locus genómico de interés comprende la secuencia de nucleótidos de SEQ ID NO: 1. En una realización, el ARNg comprende una tercera secuencia de ácidos nucleicos que codifica un ARN (ARNcr) de Repeticiones Palindrómicas Cortas Regularmente Interespaciadas Agrupadas (CRISPR) y un ARN CRISPR de transactivación del (ARNtracr). En otra realización, el genoma de la célula de rata pluripotente comprende una región de ADN diana complementaria a la secuencia diana genómica. En algunos de estos métodos, la proteína Cas es Cas9. En algunos de estos métodos, el ARNg comprende (a) el ARN quimérico de la secuencia de ácidos nucleicos de la SEQ ID NO: 2; o, (b) el ARN quimérico de la secuencia de ácidos nucleicos de la SEQ ID NO: 3. En algunos de estos métodos, el ARNcr comprende la secuencia establecida en la SEQ ID NO: 4, la SEQ ID NO: 5 o la SEQ ID NO: 6. En algunos de tales métodos, el ARNtracr comprende la secuencia establecida en la la SEQ ID NO: 7 o la SEQ ID NO: 8.

Se proporciona además un locus genómico de rata que comprende (i) una inserción de una secuencia de ácido nucleico humano homóloga u ortóloga; (ii) un reemplazo de una secuencia de ácido nucleico de rata endógena con la secuencia de ácido nucleico humana homóloga u ortóloga; o (iii) una combinación de los mismos, en donde el locus genómico de la rata es capaz de transmitirse a través de la línea germinal. En algunos de estos locus genómicos de rata, el tamaño de la inserción o reemplazo es de aproximadamente 5kb a aproximadamente 400kb. En algunos de estos locus genómicos de rata, el tamaño de la inserción o reemplazo es de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 200kb a aproximadamente 250kb, de aproximadamente 250kb a aproximadamente 300kb, de aproximadamente 300kb a aproximadamente 350kb, de aproximadamente 350kb a aproximadamente 400kb, de aproximadamente 400 kb a aproximadamente 800 kb, de aproximadamente 800 kb a 1 Mb, de aproximadamente 1 Mb a aproximadamente 1.5 Mb, de aproximadamente 1.5 Mb a aproximadamente 2 Mb, de aproximadamente 2 Mb a aproximadamente 2.5 Mb, de aproximadamente 2.5 Mb a aproximadamente 2.8 Mb, de aproximadamente 2.8 Mb a aproximadamente 3 Mb, 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 1 Mb pero menos de aproximadamente 2 Mb, al menos aproximadamente 2 Mb pero menos de aproximadamente 3 Mb.

Se proporciona además un método para hacer una rata humanizada, que comprende: (a) direccionar a un locus genómico de interés en una célula de rata pluripotente con un constructo de direccionamiento que comprende un ácido nucleico de inserción humana para formar una célula de rata pluripotente modificada genéticamente; (b) introducción de la célula de rata pluripotente modificada genéticamente en un embrión de rata huésped; y (c) la gestación del embrión de rata huésped en una madre sustituta; en donde la madre sustituta produce una progenie de rata que comprende, un locus genómico modificado que comprende: (i) una inserción de una secuencia de ácidos nucleicos humana; (ii) una sustitución de la secuencia de ácidos nucleicos de rata en el locus genómico de interés con una

secuencia de ácidos nucleicos humana homóloga u ortóloga; (iii) una secuencia de ácidos nucleicos quimérica que comprende una secuencia de ácidos nucleicos humana y una de rata; o (iv) una combinación de los mismos, en donde el locus genómico modificado puede transmitirse a través de la línea germinal.

En algunos de estos métodos, el constructo de direccionamiento es un vector de direccionamiento grande (LTVEC), y la suma total de los brazos de homología 5' y 3' del LTVEC es al menos 10 kb pero menor que 100kb o la suma total del 5' los brazos de homología 3' de LTVEC son al menos 10 kb pero menos de 150 kb. En algunos de estos métodos, la suma total de los brazos de homología 5' y 3' del constructo de orientación es de aproximadamente 10kb a aproximadamente 30kb, de aproximadamente 20kb a 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, desde aproximadamente 80kb hasta aproximadamente 100kb, desde aproximadamente 100kb hasta aproximadamente 120kb, o desde aproximadamente 120kb hasta aproximadamente 150kb. En algunos de estos métodos, la secuencia del ácido nucleico humano es de al menos 5kb pero menos de 400kb. En algunos de estos métodos, la secuencia de ácidos nucleicos humana es al menos 5kb pero menos de 10kb, al menos 10kb pero menos de 20kb, al menos 20kb pero menos de 40kb, al menos 40kb pero menos de 60kb, al menos 60kb pero menos de 80kb , al menos aproximadamente 80kb pero menos de 100kb, al menos 100kb pero menos de 150kb, al menos 150kb pero menos de 200kb, al menos 200kb pero menos de 250kb, al menos 250kb pero menos de 300kb, al menos 300kb pero menos de 350kb, o al menos 350kb pero menos de 400kb. En otras realizaciones, el tamaño de la suma total del total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 10kb a aproximadamente 150kb, de aproximadamente 10kb a aproximadamente 100kb, de aproximadamente 10kb a aproximadamente 75kb, de aproximadamente 20kb a aproximadamente 150kb, aproximadamente 20kb a aproximadamente 100kb, aproximadamente 20kb a aproximadamente 75kb, aproximadamente 30kb a aproximadamente 150kb, aproximadamente 30kb a aproximadamente 100kb, aproximadamente 30kb a aproximadamente 75kb, aproximadamente 40kb a aproximadamente 150kb, aproximadamente 40kb a aproximadamente 100kb, aproximadamente 40kb, a aproximadamente 75kb, aproximadamente 50kb a aproximadamente 150 kb, aproximadamente 50 kb a aproximadamente 100 kb, aproximadamente 50 kb a aproximadamente 75 kb, aproximadamente 10 kb a aproximadamente 30 kb, aproximadamente 20 kb a aproximadamente 40 kb, aproximadamente 40 kb a aproximadamente 60 kb, aproximadamente 60 kb a aproximadamente 80 kb, aproximadamente 80 kb, aproximadamente 100 kb, aproximadamente 100 kb a aproximadamente 120kb, o de aproximadamente 120kb a aproximadamente 150kb. 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 algunos métodos para hacer una rata humanizada, la célula pluripotente de rata es una célula madre embrionaria de rata (ES). En algunos de estos métodos, la célula de rata pluripotente se deriva de una cepa DA o una cepa ACI. En algunos de estos métodos, la célula de rata pluripotente se caracteriza por la expresión de al menos un marcador de pluripotencia que comprende Dnmt3L, Eras, Err-beta, Fbxo15, Fgf4, Gdf3, Klf4, Lef1, LIF receptor, Lin28, Nanog, Oct4, Sox15, Sox2, Utf1, y/o una combinación de los mismos. En algunos de estos métodos, la célula de rata pluripotente se caracteriza por una o más de las siguientes características: (a) falta de expresión de uno o más marcadores de pluripotencia que comprenden c-Myc, Ecat1 y/o Rexo1; (b) falta de expresión de uno o más marcadores mesodérmicos que comprenden Brachyury y/o Bmpr2; (c) falta de expresión de uno o más marcadores endodérmicos que comprenden Gata6, Sox17 y/o Sox7; o (d) falta de expresión de uno o más marcadores neuronales que comprenden Nestina y/o Pax6.

Se proporciona además una rata modificada genéticamente que comprende un locus genómico humanizado, en donde la rata modificada genéticamente comprende: (i) una inserción de una secuencia de ácidos nucleicos humana homóloga u ortóloga; (ii) un reemplazo de una secuencia de ácidos nucleicos de rata con una secuencia de ácidos nucleicos humana homóloga u ortóloga en un locus genómico endógeno con una secuencia de ácidos nucleicos humana homóloga u ortóloga; (iii) una secuencia de ácidos nucleicos quimérica que comprende una secuencia de ácidos nucleicos humana y una de rata; o, (iv) una combinación de los mismos, en donde el locus genómico humanizado es capaz de ser transmitido a través de la línea germinal. En algunas de tales ratas modificadas genéticamente, el locus genómico humanizado comprende una secuencia de ácidos nucleicos quimérica que comprende una secuencia de ácidos nucleicos de rata y humana.

También se proporcionan métodos para modificar un locus genómico diana de una rata a través de recombinación homóloga bacteriana (BHR) y comprenden: la introducción en una célula procariota de un vector de direccionamiento grande (LTVEC) que comprende un ácido nucleico de inserción flanqueado con un brazo de homología de rata 5' y el brazo de homología de rata 3', en donde la célula procariota comprende un ácido nucleico de rata y es capaz de expresar una recombinasa que media el BHR en el locus diana, y en donde la suma total de los brazos de homología 5' y 3 'de LTVEC está en menos 10kb pero menos de 100kb o la suma total de los brazos de homología 5' y 3' de LTVEC es al menos 10 kb pero menos de 150kb. En otras realizaciones , el tamaño de la suma total del total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 10kb a aproximadamente 150kb, aproximadamente 10kb a aproximadamente 100kb, aproximadamente 10kb a aproximadamente 75kb, aproximadamente 20kb a aproximadamente 150kb, aproximadamente 20kb a aproximadamente 100kb, aproximadamente 20kb a aproximadamente 75kb, aproximadamente 30kb a aproximadamente 150kb, aproximadamente 30kb a aproximadamente 100kb, aproximadamente 30kb a aproximadamente 75kb, aproximadamente 40kb a aproximadamente 150kb, aproximadamente 40kb a aproximadamente 100kb, aproximadamente 40kb, aproximadamente 75kb, aproximadamente 50kb a aproximadamente 150kb, aproximadamente 50kb a

aproximadamente 100kb, o aproximadamente 50kb a aproximadamente 75kb, aproximadamente 10kb a aproximadamente 30kb, aproximadamente 20kb a aproximadamente 40kb, aproximadamente 40kb a aproximadamente 60kb, aproximadamente 60kb a aproximadamente 80kb, aproximadamente 80kb, aproximadamente 100kb, aproximadamente 100kb a aproximadamente 120kb, o desde aproximadamente 120kb hasta aproximadamente 150kb. 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 algunos de tales métodos, el locus diana del ácido nucleico de rata comprende una primera secuencia de ácidos nucleicos que es complementaria al brazo de homología 5' y una segunda secuencia de ácidos nucleicos que es complementaria al brazo de homología 3'. En algunos de estos métodos, la primera y la segunda secuencia de ácidos nucleicos están separadas por al menos 5kb pero menos de 10kb, al menos 10kb pero menos de 20kb, al menos 20kb pero menos de 40kb, al menos 40kb pero menos de 60kb, al menos 60kb pero menos de 80kb, al menos aproximadamente 80kb pero menos de 100kb, al menos 100kb pero menos de 150kb, o al menos 150kb pero menos de 200kb, al menos aproximadamente 200kb pero menos de aproximadamente 300kb, al menos aproximadamente 300kb pero menos de aproximadamente 400kb, al menos unos 400kb pero menos de unos 500 kb, al menos unos 500kb pero menos de aproximadamente 1Mb, al menos aproximadamente 11Mb pero menos de aproximadamente 2Mb, al menos aproximadamente 2Mb pero menos de aproximadamente 3Mb.

En algunos de estos métodos, la introducción del vector de direccionamiento en la célula procariota conduce a: (i) una eliminación de una secuencia endógena de ácido nucleico de rata del locus genómico diana; (ii) una adición de una secuencia de ácidos nucleicos exógena en el locus genómico diana; (iii) un reemplazo de la secuencia endógena de ácido nucleico de rata con la secuencia de ácidos nucleicos exógena en el locus diana; o (iv) una combinación de los mismos. En algunos de estos métodos, el ácido nucleico insertado comprende (a) un polinucleótido que es homólogo u ortólogo a la secuencia de ácidos nucleicos de rata en el locus genómico diana; o (b) un alelo condicional flanqueado con secuencias de reconocimiento de recombinación específicas del sitio.

Se proporciona además una célula procariota huésped que comprende un vector de direccionamiento que comprende un ácido nucleico de inserción flanqueado con un brazo de homología de rata 5' y un brazo de homología de rata 3', en donde el ácido nucleico de inserción varía de aproximadamente 5k a aproximadamente 400 kb. En algunas células procariotas huésped, el tamaño del ácido nucleico insertado es de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, desde aproximadamente 100kb a aproximadamente 150kb, desde aproximadamente 150kb a aproximadamente 200kb, desde aproximadamente 200kb a aproximadamente 250kb, desde aproximadamente 250kb a aproximadamente 300kb, desde aproximadamente 300kb a aproximadamente 350kb, o desde 350kb a aproximadamente 400kb. En algunas células procariotas huésped, la célula procariota comprende un gen de recombinasa unido operativamente a un promotor constitutivamente activo o un promotor inducible.

Los métodos también proporcionan modificar un locus genómico de interés en una célula mediante modificación genética direccionada que comprende la introducción en la célula.

(a) un vector de direccionamiento grande (LTVEC) que comprende un inserto de ácido nucleico flanqueado con un brazo de homología 5' y un brazo de homología 3', en donde la suma total de los brazos de homología 5' y 3' del LTVEC es al menos 10kb; y

(b) (i) un primer constructo de expresión que comprende un primer promotor unido operativamente a una primera secuencia de ácidos nucleicos que codifica una proteína (Cas) asociada a Repeticiones Palindrómicas Cortas Regularmente Interespaciadas Aglomeradas (CRISPR), (ii) un segundo constructo de expresión que comprende un segundo promotor unido operativamente a una segunda secuencia de ácidos nucleicos que codifica una secuencia diana genómica operativamente unida a un ARN guía (ARNg); e identificar una célula pluripotente genéticamente modificada que comprende la modificación genética direccionada en el locus genómico de interés.

En una realización, el locus genómico de interés comprende la secuencia de nucleótidos expuesta en la SEQ ID NO: 1, en donde el ARNg comprende una tercera secuencia de ácidos nucleicos que codifica un ARN (ARNcr) de Repeticiones Palindrómicas Cortas Regularmente Interespaciadas Aglomeradas (CRISPR) y una transactivación del ARN CRISPR (ARNtracr), y en donde el genoma de la célula comprende una región de ADN diana complementaria a la secuencia diana genómica. En algunos de estos métodos, la proteína Cas es Cas9. En tales métodos, la célula es una célula de rata pluripotente (tal como una célula madre embrionaria).

En otras realizaciones, el tamaño de la suma total del total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 10kb a aproximadamente 150kb, de aproximadamente 10kb a aproximadamente 100kb, de aproximadamente 10kb a aproximadamente 75kb, de aproximadamente 20kb a aproximadamente 150kb, aproximadamente 20kb a aproximadamente 100kb, aproximadamente 20kb a aproximadamente 75kb, aproximadamente 30kb a aproximadamente 150kb, aproximadamente 30kb a aproximadamente 100kb, aproximadamente 30kb a aproximadamente 75kb, aproximadamente 40kb a aproximadamente 150kb, aproximadamente 40kb a aproximadamente 100kb, aproximadamente 40kb a aproximadamente 75kb,

aproximadamente 50kb a aproximadamente 150kb, aproximadamente 50kb a aproximadamente 100kb, o aproximadamente 50kb a aproximadamente 75kb, aproximadamente 10kb a aproximadamente 30kb, aproximadamente 20kb a aproximadamente 40kb, aproximadamente 40kb a aproximadamente 60kb, aproximadamente 60kb a aproximadamente 80kb, aproximadamente 80kb a aproximadamente 100kb, aproximadamente 100kb a aproximadamente 120kb, o desde aproximadamente 120kb a aproximadamente 150kb. 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.

También se proporciona una rata o célula de rata que comprende una modificación genética direccionada en su locus genómico, en donde el locus genómico es un locus gamma del receptor de interleucina-2, un locus ApoE, un locus Rag1, un locus Rag2, o Locus Rag2/Rag1, en donde la modificación genética direccionada comprende: (a) una eliminación de una secuencia endógena de ácido nucleico de rata en el locus genómico; (b) una inserción de un ácido nucleico homólogo, un ácido nucleico ortólogo o un ácido nucleico quimérico que comprende una secuencia de ácidos nucleicos de rata y humano; o (c) una combinación de los mismos. En tal rata o célula de rata, la modificación genética direccionada es transmisible a través de la línea germinal de la rata o una rata propagada desde la célula de rata.

En algunas de tales ratas o células de rata, la eliminación del ácido nucleico de rata endógeno en el locus genómico es de al menos aproximadamente 10kb, o la inserción de la secuencia de ácidos nucleicos exógena en el locus genómico es de al menos aproximadamente 5 kb.

Además, se proporciona una rata o célula de rata, en donde (a) la modificación genética direccionada en el locus gamma del receptor de interleucina-2 da como resultado una disminución o ausencia de la actividad de la proteína gamma del receptor de interleucina-2; (b) la modificación genética direccionada en el locus ApoE produce una disminución o ausencia de la actividad de la proteína ApoE; (c) la modificación genética direccionada en el locus Rag1 produce una disminución o ausencia de la actividad de la proteína Rag1; (d) la modificación genética direccionada en el locus Rag2 da como resultado una disminución o ausencia de la actividad de la proteína Rag2; o, (e) la modificación genética direccionada en el locus Rag2/Rag1 produce una disminución o ausencia de la actividad de la proteína Rag2 y la actividad de Rag1.

En algunas realizaciones, la modificación genética direccionada del locus gamma del receptor de interleucina-2 comprende: (a) una eliminación de la región codificante gamma del receptor de interleucina-2 de rata completa o una porción de la misma; (b) un reemplazo de la región codificante gamma del receptor de interleucina-2 de rata completa o una porción de la misma con una región codificadora gamma del receptor de interleucina-2 humana o una porción de la misma; (c) un reemplazo de un ectodominio de la región codificante gamma del receptor de interleucina-2 de rata con el ectodominio de un receptor gamma de interleucina-2 humano; o, (d) al menos una eliminación de 3 kb del locus gamma del receptor de interleucina-2. En otras ratas o células de rata similares, la modificación genética direccionada del locus ApoE comprende: (a) una eliminación de la región codificante de ApoE completa o una porción de la misma; o, (b) al menos una eliminación de 1.8 kb del locus ApoE que comprende la región de codificación ApoE.

Se proporciona además una rata o célula de rata, en donde la modificación genética direccionada del locus Rag2 comprende: (a) una eliminación de toda la región codificante de Rag2 o una parte de la misma; o (b) al menos una eliminación de 5.7 kb del locus Rag2 que comprende la región codificante de Rag2. En algunas realizaciones, la modificación genética direccionada del locus Rag2/Rag1 comprende: (a) una eliminación de la región codificante de Rag2 completa o una parte de la misma y una eliminación de la región codificante de Rag1 completa o una porción de la misma; o, (b) una eliminación de al menos 16 kb del locus Rag2/Rag1 que comprende la región de codificación Rag2.

Se proporciona además una rata o célula de rata, en donde la modificación genética direccionada comprende una inserción de un casete de expresión que comprende un marcador selectivo en el locus gamma del receptor de interleucina-2, el locus ApoE, el locus Rag1, el locus Rag2 o el locus Rag2/Rag1. En algunas de tales ratas o células de rata, el casete de expresión comprende un gen lacZ unido operativamente al promotor endógeno en el locus genómico y un promotor de ubiquitina humana unido operativamente a un marcador selectivo.

Se proporciona además una rata o célula de rata, en donde la modificación genética direccionada en el locus gamma del receptor de interleucina 2, el locus ApoE, el locus Rag1, el locus Rag2 o el locus Rag2/Rag1 comprende la inserción de un casete de selección autoeliminable. En algunas de tales ratas o células de rata, el casete de selección que se elimina automáticamente comprende un gen marcador selectivo unido operativamente a un promotor activo en la célula de rata y un gen de recombinasa unido operativamente a un promotor específico de células germinales masculinas, en donde el casete autoeliminable está flanqueado por sitios de reconocimiento de recombinación reconocidos por la recombinasa. En algunas de estas ratas o células de rata, el promotor específico de células germinales masculinas es un promotor de protamina-1; el gen de recombinasa codifica Cre, y los sitios de reconocimiento de recombinación son sitios loxP. En una realización, el promotor Protamina-1 es un promotor Protamina-1 de ratón o de rata.

Se proporciona además una rata o célula de rata, en donde la inserción de la secuencia de ácidos nucleicos exógena en el locus genómico comprende un ácido nucleico informador unido operativamente a un promotor gamma receptor de Interleucina-2 endógeno, un promotor ApoE endógeno, un promotor Rag1 endógeno, o un promotor Rag2

endógeno. En algunas de esas ratas o células de rata, el ácido nucleico informador codifica un indicador que comprende β-galactosidasa, 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 o una combinación de los mismos.

Se proporciona además una célula de rata, en donde la célula de rata es una célula de rata pluripotente o una célula madre embrionaria de rata (ES). En algunas de estas células de rata, la célula pluripotente de rata o la célula madre embrionaria de rata (ES) se deriva de una cepa DA o una cepa ACI; (b) se caracteriza por la expresión de al menos un marcador de pluripotencia que comprende Dnmt3L, Eras, Err-beta, Fbxo15, Fgf4, Gdf3, Klf4, Lef1, receptor LIF, Lin28, Nanog, Oct4, Sox15, Sox2, Utf1, o una combinación de los mismos; o (c) se caracteriza por una o más de las siguientes características: (i) falta de expresión de uno o más marcadores de pluripotencia que comprenden c-Myc, Ecat1 y Rexo1; (ii) falta de expresión de marcadores mesodérmicos que comprenden Brachyury y Bmpr2; (iii) falta de expresión de uno o más marcadores endodérmicos que comprenden Gata6, Sox17 y Sox7; o (iv) falta de expresión de uno o más marcadores neuronales que comprenden Nestin y Pax6.

Se proporciona además un método para modificar un locus genómico diana en un locus gamma del receptor de interleucina-2, un locus ApoE, un locus Rag1, un locus Rag2 o un locus Rag2/Rag1 en una célula de rata pluripotente comprendiendo el método: a) introducir en la célula de rata pluripotente un vector de direccionamiento que comprende un ácido nucleico de inserción flanqueado con brazos de homología de rata 5' y 3' homólogos al locus genómico diana; y (b) identificar una célula de rata pluripotente modificada genéticamente que comprende una modificación genética direccionada en el locus genómico diana, en donde la modificación genética direccionada puede ser transmitida a través de la línea germinal de una rata propagada desde la célula de rata pluripotente. En algunos de estos métodos, el vector de direccionamiento es un vector de direccionamiento grande (LTVEC), en donde la suma total de los brazos de homología de rata 5' y 3' es de al menos aproximadamente 10kb. En algunas realizaciones, la suma total de los brazos de homología de rata 5' y 3' es al menos 10kb pero menos de 150kb. En algunas realizaciones, la suma total de los brazos de homología de rata 5' y 3' es al menos aproximadamente 10kb pero menor que aproximadamente 100kb. En algunas realizaciones, la introducción del vector de direccionamiento en la célula de rata pluripotente conduce a: (i) una eliminación de una secuencia de ácidos nucleicos de rata endógena en el locus genómico diana; (ii) una inserción de una secuencia de ácidos nucleicos exógena en el locus genómico diana; o (iii) una combinación de los mismos.

En algunas realizaciones, la eliminación del ácido nucleico de rata endógeno en el locus genómico es de al menos aproximadamente 10 kb; la eliminación de una secuencia endógena de ácido nucleico de rata en el locus genómico varía de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, desde aproximadamente 100kb a aproximadamente 150kb, o desde aproximadamente 150kb a aproximadamente 200kb, desde aproximadamente 200kb a aproximadamente 300kb, desde aproximadamente 300kb a aproximadamente 400kb, desde aproximadamente 400kb a aproximadamente 500kb, desde aproximadamente 500kb a aproximadamente 1Mb, desde aproximadamente 1Mb a aproximadamente 1.5Mb, desde aproximadamente 1.5Mb a aproximadamente 2Mb, desde aproximadamente 2Mb a aproximadamente 2.5Mb, o desde aproximadamente 2.5Mb a aproximadamente 3Mb; la inserción de una secuencia de ácidos nucleicos exógena en el locus genómico es de al menos aproximadamente 5 kb; o la inserción de una secuencia de ácidos nucleicos exógena varía de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 200kb a aproximadamente 250kb, de aproximadamente 250kb a aproximadamente 300kb, de aproximadamente 300kb a aproximadamente 350kb, o de aproximadamente 350kb a aproximadamente 400kb.

En algunas realizaciones, (a) la modificación genética direccionada en el locus gamma del receptor de Interleucina-2 da como resultado una disminución o ausencia de la actividad de la proteína gamma del receptor de Interleucina-2; (b) la modificación genética direccionada en el locus ApoE produce una disminución o ausencia de la actividad de la proteína ApoE; (c) la modificación genética direccionada en el locus Rag1 produce una disminución o ausencia de la actividad de la proteína Rag1; (d) la modificación genética direccionada en el locus Rag2 da como resultado una disminución o ausencia de la actividad de la proteína Rag2; o, (e) la modificación genética direccionada en el locus Rag2/Rag1 produce una disminución o ausencia de la actividad de la proteína Rag2 y la actividad de la proteína Rag1.

En algunas realizaciones, la modificación genética direccionada en el locus gamma del receptor de interleucina-2 comprende (a) una eliminación de la región codificante gamma del receptor de interleucina-2 de rata completa o una porción de la misma; (b) un reemplazo de la región codificante gamma del receptor de interleucina-2 de rata completa o una porción de la misma con una región codificadora gamma del receptor de interleucina-2 humana o una porción de la misma; (c) un reemplazo de un ectodominio de la región codificante gamma del receptor de interleucina-2 de rata con el ectodominio de un gamma del receptor de interleucina-2 humano; o, (d) al menos una eliminación de 3 kb del locus gamma del receptor de Interleucina-2 que comprende la región codificadora gamma del receptor de Interleucina-2.

En algunas realizaciones, la modificación genética direccionada en el locus ApoE comprende: (a) una eliminación de la región codificante de ApoE completa o una parte de la misma; o, (b) al menos una eliminación de 1.8 kb del locus ApoE que comprende la región de codificación ApoE.

En algunas realizaciones, la modificación genética direccionada en el locus Rag2 comprende: (a) una eliminación de la región codificante de Rag2 completa o una parte de la misma; o, (b) al menos una eliminación de 5.7 kb del locus Rag2 que comprende la región de codificación Rag2. En otros métodos, la modificación genética direccionada del locus Rag1/Rag2 comprende: (a) una eliminación de la región codificante de Rag2 completa o una porción de la misma y una eliminación de la región codificante de Rag1 completa o una porción de la misma; o, (b) una eliminación de al menos 16 kb del locus Rag2/Rag1 que comprende las regiones codificantes Rag2 y Rag1.

En algunas de dichas realizaciones para modificar un locus genómico diana, el ácido nucleico de inserción comprende un casete de expresión que comprende un polinucleótido que codifica un marcador selectivo. En algunas de tales realizaciones, el casete de expresión comprende un gen lacZ unido operativamente a un promotor endógeno en el locus genómico y un promotor de ubiquitina humana unido operativamente a un gen marcador selectivo.

En algunas realizaciones, el inserto de ácido nucleico comprende un casete de selección autoeliminable. En algunas de tales realizaciones, el casete de selección autoeliminable comprende un marcador selectivo unido operativamente a un promotor activo en la célula pluripotente de rata y un polinucleótido que codifica una recombinasa unida operativamente a un promotor específico de células germinales masculinas, en donde el casete autoeliminable es flanqueado por sitios de reconocimiento de recombinación reconocidos por la recombinasa. En algunas de tales realizaciones, el promotor específico de células germinales masculinas es un promotor de protamina-1; o, el gen de recombinasa codifica Cre y los sitios de reconocimiento de recombinación son sitios loxP. En algunas realizaciones, el promotor Protamina-1 es un ratón o un promotor Protamina-1 de rata.

En otros métodos, la inserción de la secuencia de ácidos nucleicos exógena en el locus genómico comprende una secuencia de ácidos nucleicos informadora operativamente unida al promotor gamma del receptor de Interleucina-2 endógeno, el promotor ApoE endógeno, el promotor Rag1 endógeno o el promotor Rag2 endógeno. En algunas de tales realizaciones, la secuencia de ácidos nucleicos informadora codifica un informador que comprende β-galactosidasa, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, proteína fluorescente amarilla mejorada (EYFP), Emerald, proteína verde fluorescente mejorada (EGFP), CyPet, proteína fluorescente cian (CFP), Emerald, T-Sapphire, luciferasa, fosfatasa alcalina o una combinación de los mismos.

En algunas realizaciones para modificar un locus genómico diana, la célula pluripotente de rata es una célula madre embrionaria de rata (ES). En algunas de tales realizaciones, la célula de rata pluripotente (a) se deriva de una cepa DA o una cepa ACI; (b) se caracteriza por la expresión de un marcador de pluripotencia que comprende Oct-4, Sox-2, fosfatasa alcalina o una combinación de los mismos; o, (c) se caracteriza por una o más de las siguientes características: (i) falta de expresión de uno o más marcadores de pluripotencia que comprenden c-Myc, Ecat1 y Rexo1; (ii) falta de expresión de marcadores mesodérmicos que comprenden Brachyury y Bmpr2; (iii) falta de expresión de uno o más marcadores endodérmicos que comprenden Gata6, Sox17 y Sox7; o (iv) falta de expresión de uno o más marcadores neuronales que comprenden Nestin y Pax6.

En algunas realizaciones, el método comprende además identificar la modificación genética direccionada en el locus genómico diana, en donde la etapa de identificación emplea un ensayo cuantitativo para evaluar una modificación del alelo (MOA) en el locus genómico diana.

En algunas realizaciones, la etapa de introducción comprende además la introducción de un segundo ácido nucleico que codifica un agente de nucleasa que promueve una recombinación homóloga entre el vector de direccionamiento y el locus genómico diana en la célula de rata pluripotente. En algunas de tales realizaciones, el agente de nucleasa comprende una proteína quimérica que comprende un dominio de unión a ADN basado en dedos de zinc fusionado a una endonucleasa FokI. Algunos de estos métodos resultan en una modificación bialélica del locus genómico diana.

En algunas realizaciones, la etapa de introducción del método comprende además la introducción en la célula de rata pluripotente: un primer constructo de expresión que comprende un primer promotor unido operativamente a una primera secuencia de ácidos nucleicos que codifica una proteína (Cas) asociada a Repeticiones Palindrómicas Cortas Regularmente Interespaciadas Aglomeradas (CRISPR), y una segunda constructo de expresión que comprende un segundo promotor unido operativamente a una segunda secuencia de ácidos nucleicos que codifica una secuencia diana genómica unida operativamente a un ARN guía (ARNg), en donde la secuencia diana genómica está inmediatamente flanqueada en el extremo 3' termina con una secuencia de motivo adyacente Protospacer (PAM). En una realización, la secuencia diana genómica comprende la secuencia de nucleótidos expuesta en la SEQ ID NO: 1. En una realización, el ARNg comprende una tercera secuencia de ácidos nucleicos que codifica un ARN (ARNcr) de Repeticiones Palindrómicas Cortas Regularmente Interespaciadas Aglomeradas (CRISPR) y una transactivación del ARN CRISPR (taARNcr). En algunas de tales realizaciones, la proteína Cas es Cas9. En algunas de tales realizaciones, (a) el ARNg es el ARN quimérico de la secuencia de ácidos nucleicos expuesta en la SEQ ID NO: 2; (b) el ARNg es el ARN quimérico de la secuencia de ácidos nucleicos expuesta en la SEQ ID NO:3; (c) el ARNcr

comprende una secuencia expuesta en la SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6; o, (d) el ARNtracr comprende la secuencia expuesta en la SEQ ID NO: 7 y/o la SEQ ID NO:8.

Breve descripción de los dibujos

El archivo de patente o solicitud contiene al menos un dibujo ejecutado en color. La Oficina proporcionará copias de esta publicación de patente o solicitud de patente con dibujos en color, previa solicitud y pago de la tarifa necesaria.

La Figura 1 muestra los ESC de ratas, que crecen como colonias esféricas compactas que se desprenden y flotan rutinariamente en la placa.

Las Figuras 2A a D representan diversos marcadores de pluripotencia expresados por ESCs de rata: A representa Oct-4 (verde); B representa Sox-2 (rojo); C representa DAPI (azul); D representa una superposición de marcadores de pluripotencia expresados por rESCs.

La Figura 3 muestra que los ESC de rata expresan niveles de luz de fosfatasa alcalina (un marcador de pluripotencia) (izquierda), y el cariotipo para la línea DA.2B es 42X,Y (derecha). Se realizó un cariotipo porque las ESC de rata a menudo se vuelven tetraploides; por lo tanto, las líneas se preseleccionaron contando los cromosomas en metafase, y las líneas con recuentos en su mayoría normales fueron luego formalmente cariotipificadas.

La Figura 4A-B proporciona una fotografía que muestra el análisis del número de cromosomas de la línea celular ACI.G1 de rata ES.

La Figura 5A-B proporciona una fotografía que muestra el análisis del número de cromosoma de la línea celular DA.2B rata ES.

La Figura 6A-B proporciona una fotografía que muestra el análisis del número de cromosoma de la línea celular DA.C2 rata ES.

La Figura 7 muestra una vista más cercana de una rata ESC de la Figura 1.

La Figura 8 representa la producción de quimeras por inyección de blastocistos y la transmisión del genoma ESC de rata a través de la línea germinal; quimeras producidas por inyección de blastocisto utilizando ESC parentales ACI.G1 de rata; las quimeras de alto porcentaje suelen tener hocicos albinos.

La Figura 9 muestra crías F1 de agutí con compañeros de camada albinos, engendrados por la quimera ACI/SD marcada con un asterisco (*) en la figura 8.

La Figura 10 proporciona un esquema del locus ApoE de rata y denota con barras grises el sitio de corte para nucleasas de dedos de zinc (ZFN1 y ZFN2). Las regiones genómicas correspondientes a los brazos de homología 5' y 3' (5 kb y 5.4 kb, respectivamente) se indican mediante los recuadros gris oscuro. El exón 1 del gen ApoE no está codificado y se muestra como una caja abierta más cercana al brazo de homología 5'. Los tres intrones del gen ApoE se indican como líneas. Los Exones 2 y 3 comprenden regiones de codificación y se muestran como recuadros grises punteados. El exón 4 contiene secuencias codificadas y no codificadas como se indica por el sombreado gris punteado y el cuadro abierto.

La Figura 11 proporciona un resumen de la eficacia de la direccionamiento de ApoE cuando se realiza en presencia de nucleasas de dedos de zinc (ZFN1 o ZFN2)

La Figura 12 muestra El direccionamiento del locus Rosa 26 de rata, que se encuentra entre los genes Setd5 y Thumpd3 como en el ratón, con el mismo espacio. El panel A muestra la estructura del ratón Rosa 26 locus. Los transcriptos de ratón Rosa 26 constan de 2 o 3 Exones. El panel B representa la estructura del locus Rosa 26 de rata; el locus de la rata contiene un segundo exón 1 (Exlb) además del exón homólogo al exón 1 de ratón (Ex1a); ningún tercer exón ha sido identificado en ratas. El panel C representa un alelo de Rosa 26 de rata direccionado; se clonaron brazos de homología de 5 kb cada uno por PCR usando ADN genómico de DA rESC; el alelo objetivo contiene un casete Splicing Acceptor (SA)-lacZ-hUB-neo que reemplaza una eliminación de 117 pb en el intrón Rosa 26 de rata.

La Figura 13A representa un cerebro de control de una rata de tipo salvaje de 14 semanas de edad, que se tiñó con X-gal. El cerebro de control mostró un bajo nivel de tinción de fondo para LacZ (vista dorsal).

La Figura 13B representa la expresión de LacZ en el cerebro de una rata heterocigótica rRosa 26 (14 semanas de edad). El informador lacZ se expresó de manera ubicua en todo el cerebro del heterocigoto rRosa 26.

La Figura 13C muestra un corazón y un timo de control (recuadro) de una rata de tipo salvaje de 14 semanas de edad, que se trataron con X-gal. El corazón y el timo de control mostraron un bajo nivel de tinción de fondo para LacZ.

La Figura 13D representa la expresión de LacZ en el corazón y el timo (recuadro) de una rata heterocigótica rRosa 26 de 14 semanas de edad. El informador lacZ se expresó de forma ubicua en todo el corazón y el timo del heterocigoto rROSA 26.

La Figura 13E muestra un pulmón de control de una rata de tipo salvaje de 14 semanas de edad, que se trató con X-gal. El pulmón de control mostró un bajo nivel de tinción de fondo para LacZ.

La Figura 13F representa la expresión de LacZ en el pulmón de una rata heterozigota rRosa 26 de 14 semanas de edad. El indicador lacZ se expresó de forma ubicua en todo el pulmón del heterocigoto rRosa 26.

Las Figuras 13G y H representan la expresión de LacZ en embriones de rata E12.5. En contraste con el embrión de control de tipo salvaje (H), que muestra un bajo nivel de tinción de LacZ de fondo, el embrión heterocigoto rRosa 26 exhibió una expresión ubicua del indicador de LacZ en todo el embrión.

Las Figuras 13I y J representan la expresión de LacZ en embriones de rata E14.5. En contraste con el embrión de control de tipo salvaje (J), que muestra un bajo nivel de tinción de LacZ de fondo, el embrión de rata heterocigoto rRosa 26 exhibió una expresión ubicua del informador de LacZ en todo el embrión.

La Figura 14 ilustra un evento de recombinación homóloga o no homóloga que ocurre dentro de una célula ES de rata después de una electroporación de un vector de direccionamiento que comprende un casete de selección (casete lacZ-neo).

La Figura 15 ilustra el mecanismo por el cual las endonucleasas de edición del genoma (por ejemplo, ZFN y TALEN) introducen una ruptura de doble cadena (DSB) en una secuencia genómica diana y activan la unión de extremos no homóloga (NHEJ) en una célula ES.

La Figura 16 ilustra una técnica de selección de genes que utiliza ZFN/TALEN para mejorar la eficiencia de la recombinación homóloga de un vector de direccionamiento. DSB representa ruptura de doble hebra.

La Figura 17 proporciona un resumen de la producción de quimeras y la transmisión de la línea germinal del locus ApoE de rata modificado. La modificación direccionada fue asistida por nucleasas de dedos de zinc.

La Figura 18 proporciona un esquema del evento de orientación IL2r-γ en combinación con nucleasas de dedos de zinc que apuntan a ZFN U y ZFN D. Los sitios de corte de ZFN se indican en la figura.

La Figura 19 proporciona la eficiencia de direccionamiento cuando se dirige a IL2r-γ en combinación con el sistema CRISPR/Cas9.

La Figura 20 proporciona un esquema del locus ApoE de rata y un plásmido de direccionamiento. El esquema superior muestra la estructura genómica del locus ApoE de rata y las regiones genómicas correspondientes a los brazos de homología 5' y 3' (5 kb y 5,4 kb respectivamente; cajas de color gris oscuro). El exón 1 del gen ApoE no está codificado y se muestra como una caja abierta más cercana al brazo de homología 5'. Los tres intrones del gen ApoE se indican como líneas. Los Exones 2 y 3 comprenden regiones de codificación y se muestran como recuadros grises punteados. El exón 4 contiene secuencias codificadas y no codificadas como se indica por el sombreado gris punteado y el cuadro abierto. El panel inferior muestra el plásmido de direccionamiento. Los brazos de homología 5' y 3' (5 kb y 5.4 kb, respectivamente) se indican mediante los recuadros en gris oscuro. El vector de direccionamiento comprende un gen informador (lacZ) y un casete autoeliminable flanqueado por sitios loxP (flechas abiertas). El casete que se elimina automáticamente comprende un promotor Prm1 de ratón unido operativamente al gen Crei y un casete de selección de fármaco que comprende un promotor de ubiquitina humana unido operativamente a un gen de resistencia a neomicina.

La Figura 21 proporciona un esquema para apuntar al locus ApoE en células ES de rata usando nucleasas de dedo de zinc y un vector de direccionamiento que comprende un gen informador (LacZ) y un casete autoeliminable que comprende un promotor Prm1 de ratón operativamente enlazado al gen Crei y un casete de selección de un medicamento que comprende un promotor de ubiquitina humana operativamente unido a un gen de resistencia a neomicina.

La Figura 22 proporciona un esquema del locus ApoE de rata y un vector de direccionamiento grande (LTVEC). El panel superior muestra la organización genómica del locus ApoE de rata y las regiones genómicas correspondientes a los brazos de homología 5' y 3' (45 kb y 23 kb, respectivamente; las cajas de color gris oscuro). El exón 1 de ApoE no está codificado y se muestra como un armario de caja abierta para el brazo de homología 5'. Los tres intrones del gen ApoE se indican como líneas y los Exones 2 y 3 comprenden regiones codificantes y se muestran como recuadros grises punteados. El exón 4 contiene secuencias codificadas y no codificadas como se indica por el sombreado gris punteado y el cuadro abierto. El panel inferior muestra el LTVEC para modificar el locus ApoE de rata. Los brazos de homología 5' y 3' (45 kb y 23 kb, respectivamente) se indican mediante los recuadros de color gris oscuro. El LTVEC comprende un gen informador (lacZ) y un casete que se elimina automáticamente flanqueado por sitios loxP (flechas abiertas), que comprende un promotor Prm1 de ratón unido operativamente al gen Crei y un casete de selección de fármacos que comprende un promotor de ubiquitina humana unido operativamente a un gen de resistencia a la neomicina.

La Figura 23 proporciona un esquema del locus ApoE de rata y denota con barras grises los sitios de corte para las nucleasas de dedos de zinc (ZFN1 y ZFN2) usadas junto con el vector de direccionamiento grande (LTVEC) para mejorar la recombinación homóloga entre el vector de direccionamiento y la región cromosómica cognada objetivo.

La Figura 24 representa el locus IL2r-γ de rata que se ha interrumpido por una eliminación de 3.2 kb y la inserción de un gen informador (eGFP) y un casete autoeliminable que comprende un casete de selección de fármaco (hUb-neo) y el gen Crei unido de forma funcional a un ratón Prm1 promotor.

La Figura 25 proporciona un resumen de las líneas de células madre embrionarias de rata objetivo transmisoras de línea germinal.

La Figura 26 proporciona otra representación del locus IL2r-γ de rata que ha sido interrumpido por una eliminación de 3.2 kb y la inserción de un gen informador (eGFP) y un casete autoeliminable que comprende el gen Crei unido operativamente a un promotor Prm1 de ratón y un casete de selección de drogas (hUb-Neo).

La Figura 27 proporciona un esquema del locus Rag2 de rata y un vector de direccionamiento grande (LTVEC) para modificar el locus Rag2 de rata. El panel superior muestra la organización genómica del locus Rag2 de rata y las regiones genómicas afines correspondientes a los brazos de homología 5' y 3' (48 kb y 15 kb, respectivamente; cajas de color gris oscuro). Rag2 comprende un solo exón indicado por el sombreado gris punteado. El panel inferior es el LTVEC. Los brazos de homología 5' y 3' (48 kb y 15 kb, respectivamente) se indican mediante los recuadros de color gris oscuro. El LTVEC comprende un gen informador (lacZ) y un casete que se elimina automáticamente flanqueado por sitios loxP (flechas abiertas) que contiene un promotor Prm1 de rata unido operativamente al gen Crei y un casete de selección de fármacos que contiene un promotor de ubiquitina humana unido operativamente a una neomicina gen de resistencia.

La Figura 28 proporciona la estructura genómica del locus Rag1/Rag2 de rata y las regiones genómicas eliminadas por la orientación Rag2 (eliminación Rag2) o la doble orientación Rag2/Rag1 (eliminación Rag2/Rag1).

La Figura 29 proporciona un esquema de los loci Rag2 y Rag1 de rata y un gran vector de direccionamiento (LTVEC) utilizado para modificar los loci. El panel superior muestra la organización genómica de los loci Rag1 y Rag2 y las regiones genómicas afines correspondientes a los brazos de homología 5' y 3' (48 kb y 84 kb, respectivamente; cajas de color gris oscuro). Rag2 y Rag1 comprenden cada uno un solo exón indicado por el sombreado gris punteado. El panel inferior es el LTVEC. Los brazos de homología 5' y 3' (48 kb y 84 kb, respectivamente) se indican mediante los recuadros de color gris oscuro. El LTVEC comprende un gen informador (lacZ) y un casete que se elimina automáticamente flanqueado por sitios loxP (flechas abiertas), que comprende un promotor Prm1 de rata unido operativamente al gen Crei y un casete de selección de fármacos que comprende un promotor de ubiquitina humana unido operativamente a un gen de resistencia a la neomicina.

La Figura 30 muestra que las Il2rg-/y PBMC no expresan marcadores de linfocitos maduros. Se detectaron linfocitos positivos para GFP en sangre periférica en 2 de las 3 quimeras.

La Figura 31 proporciona un esquema del locus de IL-2rg de rata y un plásmido de direccionamiento para la humanización completa del locus de IL-2rg de rata. El panel superior muestra la organización genómica del locus IL-2rg de rata y las regiones genómicas afines correspondientes a los brazos de homología 5' y 3' (4.4 kb y 5.0 kb, respectivamente; recuadros de color gris oscuro). El panel inferior es el plásmido de direccionamiento. Los brazos de homología 5' y 3' (4.4 kb y 5.0 kb, respectivamente) se indican mediante los recuadros en gris oscuro. El plásmido de direccionamiento comprende la región genómica de IL-2rg humana, un gen informador (GFP) y un casete que se elimina automáticamente flanqueado por sitios loxP (flechas abiertas) que contiene un promotor Prm1 de ratón operativamente enlazado al gen Crei y un casete de selección de fármacos que contiene un promotor de ubiquitina humana unido operativamente a un gen de resistencia a neomicina.

La Figura 32 proporciona un esquema del locus de IL-2rg de rata y un plásmido de direccionamiento para la humanización de ectodominios del locus de IL-2rg de rata. El panel superior muestra la organización genómica del locus IL-2rg de rata y las regiones genómicas afines correspondientes a los brazos de homología 5' y 3' (4.4 kb y 5.0 kb, respectivamente; recuadros de color gris oscuro). El panel inferior es el plásmido de direccionamiento. Los brazos de homología 5' y 3' (4.4 kb y 5.0 kb, respectivamente) se indican mediante los recuadros en gris oscuro. El plásmido de direccionamiento comprende el ectodominio humano de la región genómica de IL-2Rg, un gen informador (GFP) y un casete autoeliminable flanqueado por sitios loxP (flechas abiertas) que contiene un promotor Prm1 de ratón operativamente enlazado al gen Crei y un casete de selección de fármacos un promotor de ubiquitina humana operativamente unido a un gen de resistencia a neomicina.

La Figura 33 proporciona una alineación de secuencia de la proteína IL-2rg humana (SEQ ID NO: 20; NP_000197.1); la proteína IL-2rg de rata (SEQ ID NO: 21; NP_543165.1); y la proteína quimérica IL-2rg (SEQ ID NO: 22) que comprende el ectodominio humano de IL-2rg fusionada con el resto de la proteína IL-2rg de rata. La unión entre la IL-2rg humana y la rata se observa mediante la línea vertical.

Descripción detallada de la invención

Glosario

El término "célula madre embrionaria" o "célula ES", como se usa en el presente documento, incluye una célula totipotente o pluripotente derivada de embrión no humano que es capaz de contribuir a cualquier tejido del embrión en desarrollo tras la introducción en un embrión. El término "célula pluripotente" como se usa en el presente documento incluye una célula indiferenciada que posee la capacidad de desarrollarse en más de un tipo de células diferenciadas.

El término "ácido nucleico homólogo" como se usa en el presente documento incluye una secuencia de ácidos nucleicos que es idéntica o sustancialmente similar a una secuencia de referencia conocida. En una realización, el término "ácido nucleico homólogo" se usa para caracterizar una secuencia que tiene una secuencia de aminoácidos que es al menos el 70%, al menos el 75%, al menos el 80%, al menos el 85%, al menos el 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o incluso 100% idéntico a una secuencia de referencia conocida.

El término "ácido nucleico ortólogo", como se usa en el presente documento, incluye una secuencia de ácidos nucleicos de una especie que es funcionalmente equivalente a una secuencia de referencia conocida en otra especie.

El término "vector de direccionamiento grande" o "LTVEC", como se usa en el presente documento, incluye vectores de direccionamiento grandes para células eucariotas que se derivan de fragmentos de ADN genómico clonado más grandes que los utilizados típicamente por otras metodologías destinadas a realizar el abordaje de genes homólogos en células eucarióticas. Ejemplos de LTVEC incluyen, entre otros, cromosoma bacteriano homólogo (BAC) y cromosoma artificial de levadura (YAC).

El término "modificación de alelo" (MOA), tal como se usa en el presente documento, incluye la modificación de la secuencia de ADN exacta de un alelo de un gen o genes o locus (loci) cromosómico en un genoma. Ejemplos de "modificación del alelo (MOA)" como se describe en el presente documento incluyen, pero no se limitan a, eliminaciones, sustituciones o inserciones de tan solo un nucleótido individual o eliminaciones de muchas kilobases que abarcan un gen(es) o locus cromosómico (loci) de interés, así como cualquiera y todas las posibles modificaciones entre estos dos extremos.

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

Las modificaciones genéticas "en serie" incluyen dos o más modificaciones a, por ejemplo, una célula ES de rata, realizadas de forma independiente. Por ejemplo, se realiza una primera modificación a un genoma de células ES de rata empleando un primer constructo de ácido nucleico adecuado. La primera modificación se puede lograr por electroporación, o cualquier otro método conocido en la técnica. Luego se realiza una segunda modificación en el mismo genoma de células ES de rata empleando una segunda constructo de ácido nucleico adecuada. La segunda modificación se puede lograr mediante una segunda electroporación, o cualquier otro método conocido en la técnica. En diversas formas de realización, después de la primera y la segunda modificación genética de la misma célula ES de rata, se pueden lograr modificaciones genéticas en serie de una tercera, una cuarta, una quinta, una sexta, etc. (una detrás de otra), por ejemplo, la electroporación en serie o cualquier otro método adecuado (en serie) conocido en la técnica.

El término "recombinasa específica de 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. Ejemplos de "recombinasa específica de sitio" incluyen, pero no se limitan a, recombinasas de Cre, Flp y Dre.

El término "línea germinal" en referencia a una secuencia de ácidos nucleicos incluye una secuencia de ácidos nucleicos que puede pasarse a la progenie.

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

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

La expresión "unido operativamente" comprende una relación en donde los componentes operan de manera funcional. En un caso, una secuencia de ácidos nucleicos que codifica una proteína puede estar unida operativamente a secuencias reguladoras (por ejemplo, promotor, potenciador, secuencia silenciadora, etc.) para conservar la regulación transcripcional adecuada. En un caso, una secuencia de ácidos nucleicos de una región variable de inmunoglobulina (o segmentos V(D)J) puede unirse operativamente a una secuencia de ácidos nucleicos de una región constante de inmunoglobulina para permitir una recombinación adecuada entre las secuencias en una inmunoglobulina pesada o secuencia de la cadena ligera.

La presente invención se refiere a las realizaciones caracterizadas en las reivindicaciones.

1. Locus diana que comprende un ácido nucleico de rata

Se proporcionan diversos métodos y composiciones, que permiten la integración de al menos un ácido nucleico de inserción en un locus diana. Como se usa en el presente documento, un "locus genómico de interés" comprende cualquier segmento o región de ADN dentro del genoma que uno desea integrar un ácido nucleico de inserción. Los términos "locus genómico de interés" y "locus genómico diana de interés" se pueden usar de manera intercambiable. El locus genómico de interés puede ser nativo de la célula, o alternativamente puede comprender un segmento de ADN heterólogo o exógeno que se integró en el genoma de la célula. Dichos segmentos de ADN heterólogos o exógenos pueden incluir transgenes, casetes de expresión, fabricantes de selección que codifican polinucleótidos, o regiones heterólogas o exógenas de ADN genómico. El término "locus" se define aquí como un segmento de ADN dentro del ADN genómico. Las modificaciones genéticas como se describen en el presente documento pueden incluir una o más eliminaciones de un locus de interés, adiciones a un locus de interés, reemplazo de un locus de interés y/o cualquier combinación de los mismos. El locus de interés puede comprender regiones codificadoras o regiones reguladoras no codificantes.

El locus genómico de interés puede comprender además cualquier componente de un sistema de integración direccionado que incluye, por ejemplo, un sitio de reconocimiento, un marcador de selección, un ácido nucleico de inserto previamente integrado, polinucleótidos que codifican agentes de nucleasa, promotores, etc. Alternativamente, el locus genómico de interés se puede ubicar dentro de un cromosoma artificial de levadura (YAC), un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano o cualquier otra región genómica diseñada en una célula huésped apropiada. En diversas realizaciones, el locus direccionado puede comprender una secuencia de ácidos nucleicos nativa, heteróloga o exógena de un procariota, un eucariota, una levadura, una bacteria, 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 de origen agropecuario o cualquier otro organismo de interés o una combinación de los mismos.

En realizaciones específicas, el locus genómico de interés comprende un locus diana de un "ácido nucleico de rata". Dicha región comprende un ácido nucleico de una rata que está integrado dentro del genoma de una célula.

Los ejemplos no limitantes del locus diana incluyen un locus genómico que codifica una proteína expresada en una célula B, un locus genómico que expresa un polipéptido en una célula B inmadura, un locus genómico que expresa un polipéptido en una célula B madura, un loci de inmunoglobulina (Ig), o un loci receptor de células T, que incluye, por ejemplo, un locus alfa del receptor de células T. Ejemplos adicionales de locus genómico diana incluyen un locus FcERIa, un locus TLR4, un locus PRLR, un locus Notch4, un locus Accn2, un locus Adamts5, un locus TRPA1, un locus FoIH1, un locus LRP5, un locus receptor IL2, incluido, para, por ejemplo, un locus de receptor de IL2 (IL2Rg), un locus ApoE, un locus Rag1, un locus Rag2, un locus Rag1/Rag2 y un locus ERBB4. Cualquier locus diana puede ser de una rata.

En una realización, el locus diana codifica una secuencia de aminoácidos de región variable de cadena pesada de inmunoglobulina de mamífero. En una realización, el locus diana codifica una secuencia de aminoácidos de región

variable de cadena pesada de inmunoglobulina de rata. En una realización, el locus diana comprende una secuencia de ADN genómico que comprende una secuencia de ácidos nucleicos de región variable de cadena pesada de inmunoglobulina humana de rata, ratón o no reorganizada operativamente unida a una secuencia de ácidos nucleicos de región constante de cadena pesada de inmunoglobulina. En una realización, la secuencia de ácidos nucleicos de la región constante de la cadena pesada de inmunoglobulina es una secuencia de ácidos nucleicos de la región constante de la cadena pesada de inmunoglobulina de rata, ratón o humana seleccionada de un CH1, una bisagra, un CH2, un CH3, y una combinación de los mismos. En una realización, la secuencia de ácidos nucleicos de la región constante de la cadena pesada comprende una CH1-bisagra-CH2-CH3. En una realización, el locus diana comprende una secuencia de ácidos nucleicos de región variable de cadena pesada de rata, ratón o inmunoglobulina humana reorganizada operativamente unida a una secuencia de ácidos nucleicos de región constante de cadena pesada de inmunoglobulina. En una realización, la secuencia de ácidos nucleicos de la región constante de la cadena pesada de inmunoglobulina es una secuencia de ácidos nucleicos de la región constante de la cadena pesada de inmunoglobulina de rata, ratón o humana seleccionada de un CH1, una bisagra, un CH2, un CH3 y una combinación de los mismos. En una realización, la secuencia de ácidos nucleicos de la región constante de la cadena pesada comprende una CH1-bisagra-CH2-CH3.

En una realización, el locus diana comprende una secuencia de ADN genómico que codifica una secuencia de aminoácidos de región variable de cadena ligera de inmunoglobulina de mamífero. En una realización, la secuencia de ADN genómico comprende una secuencia de ácidos nucleicos de región variable de la cadena ligera de λ y/o κ de mamífero no reordenado.

En una realización, la secuencia de ADN genómico comprende una secuencia de ácidos nucleicos de región variable de cadena ligera de λ y/o κ de mamífero reordenada. En una realización, la secuencia de ácidos nucleicos de la región variable de la cadena ligera λ o κ no reorganizada está unida operativamente a una secuencia de ácidos nucleicos de la región constante de la cadena ligera de la inmunoglobulina de mamífero seleccionada de una secuencia de ácidos nucleicos de la región constante de la cadena ligera λ y un ácido nucleico de la región constante de la cadena ligera κ secuencia. En una realización, la secuencia de ácidos nucleicos de la región constante de la cadena ligera de la inmunoglobulina de mamífero es una secuencia de ácidos nucleicos de la región constante de la cadena ligera de la inmunoglobulina de rata. En una realización, la secuencia de ácidos nucleicos de la región constante de la cadena ligera de la inmunoglobulina de mamífero es una secuencia de ácidos nucleicos de la región constante de la cadena ligera de la inmunoglobulina de ratón. En una realización, la secuencia de ácidos nucleicos de la región constante de la cadena ligera de la inmunoglobulina de mamífero es una secuencia de ácidos nucleicos de la región constante de la cadena ligera de la inmunoglobulina humana.

Como se usa en el presente documento, un locus ApoE de rata, un locus gamma (Il-2rg) de rata de receptor de interleucina-2 de rata, un locus Rag2 de rata, un locus Rag1 de rata y/o un locus Rag2/Rag1 de rata comprenden las regiones respectivas de genoma de rata en donde se ubican cada uno de estos genes o combinaciones de genes. Al modificar cualquiera de los locus ApoE de rata, el locus gamma del receptor de interleucina-2 de rata, el locus Rag2 de rata, el locus Rag1 de rata y/o el locus Rag2/Rag1 de rata combinados puede comprender cualquier alteración deseada del locus dado. Ejemplos no limitantes de modificación al locus de rata dado se discuten con más detalle en el presente documento.

Por ejemplo, en realizaciones específicas, uno o más del locus ApoE de rata, el locus gamma del receptor de Interleucina-2 de rata, el locus Rag2 y/o el locus Rag2/Rag1 se modifican de tal manera que la actividad y/o el nivel de la proteína ApoE codificada o la proteína gamma del receptor de interleucina-2 o la proteína Rag1 o la proteína Rag2 o una combinación de las proteínas Rag1 y Rag2 disminuyen. En otras realizaciones, la actividad de la proteína ApoE, la proteína gamma del receptor de la interleucina 2, la proteína Rag1 o la proteína Rag2, o una combinación de las proteínas Rag1 y Rag2 está ausente.

Por "disminuido" se entiende cualquier disminución en el nivel o actividad del gen/proteína codificado en el locus de interés. Por ejemplo, una disminución en la actividad puede comprender (1) una disminución estadísticamente significativa en el nivel general o la actividad de una proteína dada (es decir, ApoE, receptor de interleucina-2 gamma, Rag2, Rag2 o una combinación de Rag1 y Rag2) que incluye, por ejemplo, una disminución de nivel o actividad de 0.5%, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 120% o mayor en comparación con un control apropiado. Los métodos para ensayar una disminución en la concentración y/o la actividad de cualquiera de ApoE, receptor de interleucina-2, Rag1 y Rag2 son conocidos en la técnica.

En otras realizaciones, uno o más del locus ApoE de rata, el locus gamma del receptor de interleucina-2 de rata, el locus Rag2 de rata, el locus Rag1 de rata y/o el locus Rag2/Rag1 de rata comprenden una modificación tal que la actividad y/o el nivel del polipéptido ApoE codificado, el polipéptido gamma del receptor de interleucina-2, el polipéptido Rag2, el polipéptido Rag1, o ambos, el polipéptido Rag1 y Rag2 aumenta. Por "aumento" se entiende cualquier aumento en el nivel o actividad del gen/polipéptido codificado en el locus de interés. Por ejemplo, un aumento en la actividad puede comprender (1) un aumento estadísticamente significativo en el nivel general o la actividad de una proteína dada (es decir, ApoE, receptor de interleucina-2 gamma, Rag1, Rag2 o Rag1 y Rag2) que incluye, por ejemplo, un aumento de nivel o actividad de 0.5%, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 120% o más cuando se compara con un control apropiado. Los métodos para ensayar un aumento en la concentración

y/o la actividad de cualquiera de las proteínas gamma del receptor de ApoE, Rag1, Rag2 e interleucina-2 son conocidos en la técnica.

La modificación genética del locus ApoE de rata, el locus gamma del receptor de Interleucina-2 de rata, el locus Rag2 de rata, el locus Rag1 de rata y/o el locus Rag2/Rag1 de rata puede comprender una eliminación de una secuencia endógena de ácido nucleico de rata en el locus genómico, una inserción de un ácido nucleico exógeno en el locus genómico, o una combinación de los mismos. La eliminación y/o inserción puede ocurrir en cualquier lugar dentro del locus dado como se discute en otra parte en el presente documento.

Otras realizaciones proporcionadas en el presente documento comprenden la modificación de uno o más del locus ApoE de rata, el locus gamma del receptor de Interleucina-2 de rata, el locus Rag2 de rata, el locus Rag1 de rata y/o el locus Rag2/Rag1 de rata mediante la sustitución de una parte del locus ApoE de rata, el locus gamma del receptor de interleucina-2, locus Rag2, locus Rag1 y/o locus Rag2/Rag1 con la porción homóloga u ortóloga correspondiente de un locus ApoE, un locus gamma del receptor de interleucina-2, un locus Rag2, un locus Rag1 y/o un locus Rag2/Rag1 de otro organismo.

En todavía otras realizaciones, la modificación de uno o más del locus ApoE de rata, el locus gamma del receptor de interleucina-2 de rata, el locus Rag2, el locus Rag1 y/o el locus Rag2/Rag1 se lleva a cabo mediante la sustitución de una parte del locus ApoE de rata, el locus gamma del receptor de interleucina-2 de rata y/o el locus Rag2 de rata, y/o el locus Rag1 y/o locus Rag2/Rag1 con un polinucleótido de inserción que comparte en toda su longitud al menos el 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% a una porción de un locus ApoE, un locus gamma del receptor de interleucina-2, un locus Rag2, un locus Rag1 y/o un locus Rag2/Rag1 que está reemplazando.

El polinucleótido de inserción dado y/o la región correspondiente del locus de rata que se está eliminando puede ser una región codificante, un intrón, un exón, una región no traducida, una región reguladora, un promotor, un potenciador o cualquier combinación de los mismos o cualquier porción de los mismos. Además, el polinucleótido de inserto dado y/o la región del locus de rata que se está eliminando puede tener cualquier longitud deseada, incluyendo, por ejemplo, entre 10-100 nucleótidos de longitud, 100-500 nucleótidos de longitud, 500-1kb de nucleótido de longitud, 1Kb a 1.5kb de nucleótido de longitud, 1.5kb a 2kb de nucleótidos de longitud, 2kb a 2.5kb de nucleótidos de longitud, 2.5kb a 3kb de nucleótidos de longitud, 3kb a 5kb de nucleótidos de longitud, 5kb a 8kb de nucleótidos de longitud, 8kb a 10kb de nucleótidos en longitud o más. En otros casos, el tamaño de la inserción o reemplazo es de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb de aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 200kb a aproximadamente 250kb, de aproximadamente 250kb a aproximadamente 300kb, de aproximadamente 300kb a aproximadamente 350kb, de aproximadamente 350kb a aproximadamente 400kb, de aproximadamente 400kb a aproximadamente 800kb, de aproximadamente 800kb a 1Mb, de aproximadamente 300kb a aproximadamente 400kb, de aproximadamente 400kb a aproximadamente 500kb, de aproximadamente 500kb a 1Mb, de aproximadamente 1Mb a aproximadamente 1.5Mb, de aproximadamente 1.5Mb a aproximadamente 2Mb, de aproximadamente 2Mb a aproximadamente 2.5Mb, de aproximadamente 2.5Mb a aproximadamente 2.8Mb, de aproximadamente 2.8Mb a aproximadamente 3Mb. En otras realizaciones, el polinucleótido de inserción dado y/o la región del locus de rata que se está eliminando es al menos 100, 200, 300, 400, 500, 600, 700, 800 o 900 nucleótidos o al menos 1kb, 2kb, 3kb, 4kb, 5kb, 6kb, 7kb, 8kb, 9kb, 10kb, 11kb, 12kb, 13kb, 14 kb, 15kb, 16kb o más.

El polinucleótido insertado dado puede ser de cualquier organismo, incluido, por ejemplo, un roedor, una rata, un ratón, un hámster, un mamífero, un mamífero no humano, un humano, un animal de origen agropecuario o un animal doméstico.

Como se describe con más detalle en el presente documento, se proporcionan diversos métodos para generar modificaciones direccionadas a cualquier locus de interés de rata, que incluyen, por ejemplo, modificaciones direccionadas en el locus ApoE de rata, el locus gamma del receptor de interleucina-2 de rata, el locus Rag2 de rata, el locus Rag1 de rata y/o el locus Rag2/Rag1 de rata. Además, se proporcionan ratas modificadas genéticamente o células de rata pluripotentes modificadas genéticamente (por ejemplo, células ES de rata), que comprenden una eliminación, una inserción, un reemplazo y/o una combinación de los mismos en el locus gamma del receptor de interleucina 2, en el locus ApoE, en el locus Rag2 de rata, en el locus Rag1 de rata, y/o en el locus Rag2/Rag1 de rata. Dichas modificaciones genéticas (incluidas aquellas que resultan en una ausencia, una disminución, un aumento o una modulación en la actividad del locus diana) y también pueden transmitirse a través de la línea germinal. En realizaciones específicas, las modificaciones genéticas dan como resultado una desactivación del locus diana deseado. Tales ratas encuentran uso en una variedad de sistemas experimentales como se discute en otra parte del presente documento.

Por ejemplo, los bloqueos de ApoE (apolipoproteína E) en ratas ofrecen un modelo animal para estudiar la función endotelial, que incluye, entre otros, la formación de placa, los cambios transcripcionales (Pistola de Secuenciación de Transcriptoma Completo (RNA-Seq) y función *ex vivo*. Además, el tamaño más grande de las ratas facilita todos estos ensayos y mejora potencialmente la calidad de los datos de RNA-Seq. ApoE es una molécula de transporte importante y puede transportar lípidos, como el colesterol, a través del torrente sanguíneo. ApoE también puede funcionar en el

sistema nervioso, por ejemplo, para eliminar el β-amiloide del cerebro. Las modificaciones en la ApoE se han implicado en diversas afecciones, que incluyen, por ejemplo, aterosclerosis, hiperlipidemia y enfermedad de Alzheimer. Los animales bloqueados para ApoE muestran un deterioro deficiente de la lipoproteína de la sangre y desarrollan aterosclerosis. Por lo tanto, los animales bloqueados de ApoE proporcionan un modelo para estudiar condiciones y/o procesos como, por ejemplo, la función del endotelio, la formación de placa, los cambios transcripcionales (RNA-Seq), hiperlipidemia, aterosclerosis y enfermedad de Alzheimer. Los ensayos para medir la actividad de ApoE son conocidos en la técnica. Por ejemplo, una disminución en la actividad de ApoE puede medirse analizando una disminución en los niveles de ApoE en una muestra de sangre obtenida de un sujeto mediante inmunoensayos, como por ejemplo mediante ELISA o mediante técnicas de inmunotransferencia. Además, el gran tamaño de las ratas facilita todos estos ensayos y mejora la calidad de los datos.

RAG1 (gen activador de la recombinación 1) y RAG2 (gen activador de la recombinación 2) son enzimas que forman parte de un complejo de múltiples subunidades que tienen actividad de recombinación VDJ y desempeñan un papel importante en el reordenamiento y la recombinación de los genes de inmunoglobulina y receptores de células T en los linfocitos. RAG1 y RAG2 inducen una división del ADN de doble cadena para facilitar la recombinación y la unión de segmentos de los genes del receptor de células T y los receptores de células B (es decir, inmunoglobulina). El bloqueo de RAG1 y/o RAG2 causa una pérdida de células B y células T en el animal que resulta en una inmunodeficiencia grave. Los animales bloqueados para RAG1 y/o RAG2 encuentran uso, por ejemplo, en estudios de xenoinjertos (es decir, xenoinjertos de células humanas en ratas), cáncer, desarrollo de vacunas, enfermedades autoinmunes, enfermedades infecciosas y enfermedades de injerto contra huésped (GVHD). Varios ensayos para medir la actividad de RAG1 y/o RAG2 son conocidos en la técnica e incluyen, por ejemplo, medir la eficacia de recombinación o evaluar la presencia o ausencia de células B y/o células T en un sujeto. Además, el gran tamaño de las ratas facilita todos estos ensayos y potencialmente mejora la calidad de los datos.

El receptor de IL-2 (IL-2R) se expresa en la superficie de ciertas células inmunes y se une a la citoquina interleucina-2 (IL-2). El IL-2R es una proteína de membrana integral que comprende al menos tres cadenas de subunidades separadas, incluida una cadena alfa (IL-2Ra, CD25), una cadena beta (IL-2Rb, CD122) y una cadena gamma (IL2-Rg, CD132). La cadena gamma del receptor IL-2 (también conocida como IL2r-γ o IL2Rg) es una cadena gamma común que es compartida por varios receptores de citoquinas, incluyendo, por ejemplo, los receptores para IL-2, IL-4, IL-7, IL-9, IL-15 e IL-21. IL-2Rg comprende un ectodominio en la superficie extracelular de la célula, que contribuye a la unión del ligando, un dominio transmembrana y un dominio intracelular que puede interactuar con varias moléculas para inducir vías de transducción de señales intracelulares. El gen Il2rg se encuentra en el cromosoma X en mamíferos y ciertas mutaciones en el gen de la cadena gamma en los seres humanos pueden causar una inmunodeficiencia combinada severa ligada al X (XSCID) humana caracterizada por un defecto profundo en las células T. Además, el ectodominio de la cadena gamma puede desprenderse del receptor transmembrana y liberarse como un receptor de cadena gamma soluble. El receptor de cadena gamma soluble puede detectarse en la sangre de un sujeto y puede funcionar para regular la señalización de citoquinas.

En algunas realizaciones, la cadena de IL-2Rg de rata se reemplaza con la cadena de IL2-Rg humana de modo que la rata expresa una cadena de IL-2Rg completamente humana. En otros casos, puede ser útil reemplazar solo el ectodominio de la cadena IL-2Rg de rata con el ectodominio de la cadena IL-2Rg humana. En tales casos, la cadena de IL-2Rg humanizada resultante expresada en una rata comprende un ectodominio humano, y el resto de la molécula proviene de la rata.

La humanización en longitud completa de IL-2Rg es útil porque las ratas que tienen este locus modificado producirán IL-2Rg humana. Esto permitirá la detección de IL-2Rg humano en ratas con anticuerpos específicos para IL-2Rg humano. La ectohumanización (es decir, la sustitución del ectodominio de rata de IL-2Rg con el ectodominio humano de IL-2Rg) dará como resultado un polipéptido de IL-2Rg que se unirá a los ligandos humanos para IL2-Rg, pero debido a que el dominio citoplásmico todavía es de rata, su forma ectohumanizada de IL-2Rg también interactuará con la maquinaria de señalización de rata.

2. Modificación de un locus diana de rata

A. Direccionamiento de vectores e inserción de ácidos nucleicos

i. Inserción de ácidos nucleicos

Como se usa en el presente documento, el "ácido nucleico de inserción" comprende un segmento de ADN que se desea integrar en el locus diana. En una realización, el ácido nucleico de inserción comprende uno o más polinucleótidos de interés. En otras realizaciones, el ácido nucleico 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 indicador junto con los diversos componentes reguladores que influyen en la expresión. Ejemplos no limitantes de polinucleótidos de interés, marcadores de selección y genes informadores que pueden incluirse dentro del inserto de ácido nucleico se discuten en detalle en otra parte de este documento.

En realizaciones específicas, el ácido nucleico insertado puede comprender un ácido nucleico de rata, que puede incluir un segmento de ADN genómico, un ADNc, una región reguladora, o cualquier parte o combinación de los

mismos. En otras realizaciones, el ácido nucleico insertado puede comprender un ácido nucleico de 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, un tití, un mono rhesus), un mamífero domesticado o un mamífero de origen agropecuario o cualquier otro organismo de interés. Como se describe con más detalle en el presente documento, el ácido nucleico insertado empleado en los diversos métodos y composiciones puede dar como resultado la "humanización" de un locus diana que comprende un ácido nucleico de rata.

En una realización, el inserto de ácido nucleico comprende un alelo activado de al menos un exón de un gen endógeno. En una realización, el inserto de ácido nucleico comprende un alelo activado de todo el gen endógeno (es decir, "activación por de intercambio de genes").

En una realización, el inserto de ácido nucleico comprende un elemento regulador, que incluye, por ejemplo, un promotor, un potenciador o un elemento de unión a represor transcripcional.

En realizaciones adicionales, el ácido nucleico de inserción comprende un alelo condicional. En una realización, el alelo condicional es un alelo multifuncional, como se describe en el documento US 2011/0104799. En realizaciones específicas, el alelo condicional comprende: (a) una secuencia de actuación en orientación de sentido con respecto a la transcripción de un gen diana, y un casete de selección de fármaco en orientación de 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 Genetrap; véase, por ejemplo, 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 la DSC, y (ii) contiene el NSI en orientación de sentido y la COIN en orientación antisentido.

El inserto de ácido nucleico varía de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150 kb, de aproximadamente 150 kb a aproximadamente 200 kb, de aproximadamente 200 kb a aproximadamente 250 kb, de aproximadamente 250 kb a aproximadamente 300 kb, de aproximadamente 300 kb a aproximadamente 350 kb, o de aproximadamente 350 kb a aproximadamente 400 kb.

En una realización, el ácido nucleico insertado comprende una eliminación de una secuencia de ADN genómico de rata que varía de aproximadamente 1kb a aproximadamente 200kb, de aproximadamente 2kb a aproximadamente 20kb, o de aproximadamente 0.5kb a aproximadamente 3mb. En una realización, la extensión de la eliminación de la secuencia de ADN genómico es mayor que la longitud total del brazo de homología 5' y el brazo de homología 3'. En una realización, el alcance de la eliminación de la secuencia de ADN genómico varía de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 20kb a aproximadamente 30kb, de aproximadamente 30kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 50kb, de aproximadamente 50kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 70kb, desde aproximadamente 70kb a aproximadamente 80kb, desde aproximadamente 80kb a aproximadamente 90kb, desde aproximadamente 90kb a aproximadamente 100kb, desde aproximadamente 100kb a aproximadamente 110kb, desde aproximadamente 110kb a aproximadamente 120kb, desde aproximadamente 120kb a aproximadamente 130kb, desde aproximadamente 130kb a aproximadamente 140kb, desde aproximadamente 140kb a aproximadamente 150kb, desde aproximadamente 150kb a aproximadamente 160kb, desde aproximadamente 160kb a aproximadamente 170kb, desde aproximadamente 170kb a aproximadamente 180kb, desde aproximadamente 180kb a aproximadamente 190kb, desde aproximadamente 190kb a aproximadamente 200kb, desde aproximadamente 200kb a aproximadamente 250kb, desde aproximadamente 250kb a aproximadamente 300kb, desde aproximadamente 300kb a aproximadamente 350kb, de aproximadamente 350kb a aproximadamente 400kb, de aproximadamente 400kb a aproximadamente 800kb, de aproximadamente 800kb a 1Mb, de aproximadamente 1Mb a aproximadamente 1.5Mb, de aproximadamente 1.5Mb a aproximadamente 2Mb, de aproximadamente 2Mb, a aproximadamente 2.5Mb, de aproximadamente 2.5Mb a aproximadamente 2.8Mb, desde aproximadamente 2.8Mb a aproximadamente 3Mb, desde aproximadamente 200kb a aproximadamente 300kb, desde aproximadamente 300kb a aproximadamente 400kb, desde aproximadamente 400kb a aproximadamente 500kb, desde aproximadamente 500kb a aproximadamente 1Mb, desde aproximadamente 1Mb a aproximadamente 1.5Mb , desde aproximadamente 1.5Mb hasta aproximadamente 2Mb, desde aproximadamente 2Mb hasta aproximadamente 2.5Mb, o desde aproximadamente 2.5Mb hasta aproximadamente 3Mb.

En una realización, el ácido nucleico de inserción comprende una inserción o un reemplazo de una secuencia de ácidos nucleicos de rata con una secuencia de ácidos nucleicos humana homóloga u ortóloga. En una realización, el ácido nucleico insertado comprende una inserción o sustitución de una secuencia de ADN de rata con una secuencia de ácidos nucleicos humana homóloga u ortóloga en un locus endógeno de rata que comprende la secuencia de ADN de rata correspondiente.

En una realización, la modificación genética es una adición de una secuencia de ácido nucleico. En una realización, la secuencia de nucleótidos añadida varía de 5kb a 200kb.

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

En una realización, el inserto de ácido nucleico comprende una inserción o un reemplazo de una secuencia de ácidos nucleicos de rata con una secuencia de ácidos nucleicos humana homóloga u ortóloga. En una realización, el ácido nucleico insertado comprende una inserción o sustitución de una secuencia de ADN de rata con una secuencia de ácidos nucleicos humana homóloga u ortóloga en un locus endógeno de rata que comprende la secuencia de ADN de rata correspondiente.

En una realización, la modificación genética comprende una eliminación de una secuencia no codificadora de proteínas, pero no comprende una eliminación de una secuencia codificadora de proteínas. En una realización, la eliminación de la secuencia no codificadora de proteínas comprende una eliminación de un elemento regulador. En una realización, la modificación genética comprende una eliminación de un promotor. En una realización, la modificación genética comprende una adición de un promotor o un elemento regulador. En una realización, la modificación genética comprende una sustitución de un promotor o un elemento regulador.

En una realización, la secuencia de ácidos nucleicos del vector de direccionamiento puede comprender un polinucleótido que cuando se integra en el genoma producirá una modificación genética de una región del locus ApoE de rata, en donde la modificación genética en el locus ApoE da lugar a una disminución de la actividad de ApoE, aumento de la actividad de ApoE o una modulación de la actividad de ApoE. En una realización, se genera un ApoE bloqueado (“alelo nulo”).

En una realización, la secuencia de ácidos nucleicos del vector de direccionamiento puede comprender un polinucleótido que cuando se integre en el genoma producirá una modificación genética de una región del locus del receptor de interleucina-2 de rata, en donde la modificación genética en el locus del receptor de interleucina-2 da como resiltado una disminución en la actividad del receptor de interleucina-2. En una realización, se genera un bloqueo del receptor de interleucina-2 ("alelo nulo").

En realizaciones adicionales, el inserto de ácido nucleico da como resultado el reemplazo de una porción del locus ApoE de rata, el locus gamma y/o el locus Rag2 del receptor de interleucina-2 y/o el locus Rag1 y/o el locus Rag2/Rag1 con la porción homóloga u ortóloga correspondiente de un locus ApoE, un locus gamma del receptor de interleucina-2, un locus Rag2, un locus Rag1 y/o un locus Rag2/Rag1 de otro organismo.

En otras realizaciones, el inserto de ácido nucleico comprende un polinucleótido que comparte en toda su longitud al menos 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% a una porción de un locus ApoE, un locus gamma del receptor de interleucina-2, un locus Rag2, un locus Rag1 y/o un locus Rag2/Rag1 que está reemplazando.

El polinucleótido de inserto dado y la región correspondiente del locus de rata que se reemplaza pueden ser una región codificante, un intrón, un exón, una región no traducida, una región reguladora, un promotor, un potenciador o cualquier combinación de los mismos. Además, el polinucleótido de inserto dado y/o la región del locus de rata que se está eliminando puede tener cualquier longitud deseada, incluyendo, por ejemplo, entre 10-100 nucleótidos de longitud, 100-500 nucleótidos de longitud, 500-1kb de nucleótido de longitud, 1Kb a 1.5kb de nucleótido de longitud, 1.5kb a 2kb de nucleótidos de longitud, 2kb a 2.5kb de nucleótidos de longitud, 2.5kb a 3kb de nucleótidos de longitud, 3kb a 5kb de nucleótidos de longitud, 5kb a 8kb de nucleótidos de longitud, 8kb a 10kb de nucleótidos en longitud o más. En otros casos, el tamaño de la inserción o reemplazo es de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb de aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 200kb a aproximadamente 250kb, de aproximadamente 250kb a aproximadamente 300kb, de aproximadamente 300kb a aproximadamente 350kb, de aproximadamente 350kb a aproximadamente 400kb, de aproximadamente 400kb a aproximadamente 800kb, de aproximadamente 800kb a 1Mb, de aproximadamente 1Mb a aproximadamente 1.5Mb, de aproximadamente 1.5Mb a aproximadamente 2Mb, de aproximadamente 2Mb, a aproximadamente 2.5Mb, de aproximadamente 2.5Mb a aproximadamente 2.8Mb, de aproximadamente 2.8Mb a aproximadamente 3Mb . En otras realizaciones, el polinucleótido de inserción dado y/o la región del locus de rata que se está eliminando es al menos 100, 200, 300, 400, 500, 600, 700, 800 o 900 nucleótidos o al menos 1 kb, 2kb, 3kb, 4kb, 5kb, 6kb, 7kb, 8kb, 9kb, 10kb, 11kb, 12kb, 13kb, 14 kb, 15kb, 16kb o más.

En una realización, el promotor es promotor constitutivamente activo.

En una realización, el promotor es un promotor inducible. En una realización, el promotor inducible es un promotor regulado químicamente. En una realización, el promotor regulado químicamente es un promotor regulado por alcohol. En una realización, el promotor regulado por alcohol es un promotor del gen de la alcohol deshidrogenasa (alcA). En una realización, el promotor regulado químicamente es un promotor regulado por tetraciclina. En una realización, el

promotor regulado por tetraciclina es un promotor sensible a tetraciclina. En una realización, el promotor regulado por tetraciclina es una secuencia de operador de tetraciclina (tetO). En una realización, el promotor regulado por tetraciclina es un promotor tet-On. En una realización, el promotor regulado por tetraciclina es un promotor tet-Off. En una realización, el promotor regulado químicamente es un promotor regulado por esteroides. En una realización, el promotor regulado por esteroides es un promotor de un receptor de glucocorticoides de rata. En una realización, el promotor regulado por esteroides es un promotor de un receptor de estrógeno. En una realización, el promotor regulado por esteroides es un promotor de un receptor de ecdisona. En una realización, el promotor regulado químicamente es un promotor regulado por metal. En una realización, el promotor regulado por metal es un promotor de metaloproteína. En una realización, el promotor inducible es un promotor regulado físicamente. En una realización, el promotor regulado físicamente es un promotor regulado por temperatura. En una realización, el promotor regulado por temperatura es un promotor de choque térmico. En una realización, el promotor regulado físicamente es un promotor regulado por luz. En una realización, el promotor regulado por luz es un promotor inducible por luz. En una realización, el promotor regulado por la luz es un promotor reprimible por la luz.

En una realización, el promotor es un promotor específico de tejido. En una realización, el promotor es un promotor específico de neuronas. En una realización, el promotor es un promotor específico de la glía. En una realización, el promotor es un promotor específico de células musculares. En una realización, el promotor es un promotor específico de células cardíacas. En una realización, el promotor es un promotor específico de células renales. En una realización, el promotor es un promotor específico de células óseas. En una realización, el promotor es un promotor específico de células endoteliales. En una realización, el promotor es un promotor específico de células inmunes. En una realización, el promotor de células inmunes es un promotor de células B. En una realización, el promotor de células inmunes es un promotor de células T.

En una realización, el promotor es un promotor regulado por el desarrollo. En una realización, el promotor regulado por el desarrollo está activo solo durante una etapa embrionaria de desarrollo. En una realización, el promotor regulado por el desarrollo está activo solo en una célula adulta.

En algunas realizaciones, el ácido nucleico insertado comprende un ácido nucleico flanqueado con secuencias diana de recombinación específicas del sitio. Se reconoce que, mientras que el ácido nucleico del inserto completo puede estar flanqueado por dichas secuencias diana de recombinación específicas del sitio, cualquier región o polinucleótido individual de interés dentro del ácido nucleico del inserto también puede estar flanqueado por dichos sitios. La recombinasa específica de 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 de sitio en la célula huésped. El polinucleótido que codifica la recombinasa específica del sitio puede ubicarse dentro del ácido nucleico del inserto o dentro de un polinucleótido separado. La recombinasa específica de 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 diana de recombinación específicas del sitio, que pueden flanquear el ácido nucleico de inserción o cualquier polinucleótido de interés en el ácido nucleico de inserción, pueden incluir, pero no se limitan a, loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attp, att, FRT, rox, y una combinación de los mismos.

En algunas realizaciones, los sitios de recombinación específicos del sitio flanquean un polinucleótido que codifica un marcador de selección y/o un gen indicador contenido dentro del ácido nucleico del inserto. En tales casos, después de la integración del ácido nucleico de inserción en el locus diana, se pueden eliminar las secuencias entre los sitios de recombinación específicos del sitio.

En una realización, el inserto de ácido nucleico comprende un polinucleótido que codifica un marcador de selección. El marcador de selección puede estar contenido en un casete de selección. Dichos marcadores de selección incluyen, pero no se limitan a, neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa ($puro^f$), blasticidina S desaminasa (bsr^r), xantina/guanina fosforribosil transferasa (gpt) o virus del herpes simple timidina quinasa (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, célula de rata, célula de rata pluripotente o la célula de rata ES. Cuando se apilan en serie los polinucleótidos de interés en un locus diana, el marcador de selección puede comprender un sitio de reconocimiento para un agente de nucleasa, como se describe anteriormente. En una realización, el polinucleótido que codifica el marcador de selección está flanqueado con secuencias diana de recombinación específicas del sitio.

El inserto de ácido nucleico puede comprender además un gen informador unido operativamente a un promotor, en donde el gen informador codifica una proteína reportera seleccionada del grupo que consiste en o que comprende 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/o 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 a la célula, un promotor que es heterólogo al gen informador o a la célula, un promotor específico de la célula, un promotor específico del tejido o un promotor específico de etapa una etapa del desarrollo.

En una realización, el inserto de ácido nucleico puede comprender un ácido nucleico de mamífero que comprende un locus genómico que codifica una proteína expresada en el sistema nervioso, el sistema esquelético, el sistema digestivo, el sistema circulatorio, el sistema muscular, el sistema respiratorio, El sistema cardiovascular, el sistema linfático, el sistema endocrino, el sistema urinario, el sistema reproductivo, o una combinación de los mismos. En una realización, el ácido nucleico de mamífero comprende un locus genómico que codifica una proteína expresada en una médula ósea o una célula derivada de médula ósea. En una realización, el ácido nucleico comprende un locus genómico que codifica una proteína expresada en una célula de bazo.

En una realización, el ácido nucleico de los mamíferos comprende un locus genómico que codifica una proteína expresada en el sistema nervioso, el sistema esquelético, el sistema digestivo, el sistema circulatorio, el sistema muscular, el sistema respiratorio, el sistema cardiovascular, sistema linfático, el sistema endocrino, el sistema urinario, el sistema reproductivo, o una combinación de los mismos. En una realización, el ácido nucleico de mamífero comprende un locus genómico que codifica una proteína expresada en una médula ósea o una célula derivada de médula ósea. En una realización, el ácido nucleico comprende un locus genómico que codifica una proteína expresada en una célula de bazo. En una realización, el locus genómico comprende una secuencia de ADN genómico de ratón, una secuencia de ADN genómico de rata y una secuencia de ADN genómico humano, o una combinación de las mismas. En una realización, el locus genómico comprende, en cualquier orden, secuencias de ADN genómico de rata y humano. En una realización, el locus genómico comprende, en cualquier orden, secuencias de ADN genómico de ratón y humano. En una realización, el locus genómico comprende, en cualquier orden, secuencias de ADN genómico de ratón y rata. En una realización, el locus genómico comprende, en cualquier orden, secuencias de ADN genómico de rata, ratón y humano.

En una realización, el locus genómico comprende una secuencia de ADN genómico de ratón, una secuencia de ADN genómico de rata, una secuencia de ADN genómico humano, o una combinación de las mismas. En una realización, el locus genómico comprende, en cualquier orden, secuencias de ADN genómico de rata y humano. En una realización, el locus genómico comprende, en cualquier orden, secuencias de ADN genómico de ratón y humano. En una realización, el locus genómico comprende, en cualquier orden, secuencias de ADN genómico de ratón y rata. En una realización, el locus genómico comprende, en cualquier orden, secuencias de ADN genómico de rata, ratón y humano.

En una realización, la modificación genética comprende al menos un alelo de enfermedad humana de un gen humano. En una realización, la enfermedad humana es una enfermedad neurológica. En una realización, la enfermedad humana es una enfermedad cardiovascular. En una realización, la enfermedad humana es una enfermedad renal. En una realización, la enfermedad humana es una enfermedad muscular. En una realización, la enfermedad humana es una enfermedad de la sangre. En una realización, la enfermedad humana es un cáncer. En una realización, la enfermedad humana es una enfermedad del sistema inmune.

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

En una realización, la modificación genética produce una forma mutante de una proteína con una característica de unión alterada, localización alterada, expresión alterada y/o patrón de expresión alterado.

En una realización, el inserto de ácido nucleico comprende un casete de selección. En una realización, el casete de selección comprende una secuencia de ácidos nucleicos que codifica un marcador selectivo, en donde la secuencia de ácidos nucleicos está unida operativamente a un promotor activo en células ES de rata. En una realización, el marcador selectivo se selecciona de o comprende un gen de resistencia a la higromicina o un gen de resistencia a la neomicina.

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

En una realización, el inserto de ácido nucleico comprende un elemento regulador. En una realización, el elemento regulador es un promotor. En una realización, el elemento regulador es un potenciador. En una realización, el elemento regulador es un elemento de unión a represor transcripcional.

En una realización, la modificación genética comprende una eliminación de una secuencia no codificadora de proteínas, pero no comprende una eliminación de una secuencia codificadora de proteínas. En una realización, la eliminación de la secuencia no codificadora de proteínas comprende una eliminación de un elemento regulador. En una realización, la modificación genética comprende una eliminación de un elemento regulador. En una realización, la modificación genética comprende una adición de un promotor o un elemento regulador. En una realización, la modificación genética comprende una sustitución de un promotor o un elemento regulador.

ii. Casetes de Expresión

En el presente documento se proporcionan polinucleótidos o moléculas de ácido nucleico que comprenden los diversos componentes empleados en un sistema de integración genómica diana proporcionado en el presente documento (es decir, uno cualquiera o cualquier combinación de agentes de nucleasa, sitios de reconocimiento, ácidos nucleicos de inserción, polinucleótidos de interés, vectores de direccinamiento, marcadores de selección y otros componentes).

Los términos "polinucleótido", "secuencia de polinucleótido", "secuencia de ácido nucleico" y "fragmento de ácido nucleico" se usan indistintamente en el presente 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 sintéticas, 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 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.

Además, se proporcionan polinucleótidos recombinantes que comprenden los diversos componentes del sistema de integración genómica direccionada. Los términos "polinucleótido recombinante" y "constructo de ADN recombinante" se usan indistintamente en el presente 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 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 sola 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. También se proporcionan los elementos genéticos necesarios para transformar, seleccionar y propagar con éxito las células huésped que comprenden cualquiera de los fragmentos de ácido nucleico aislados proporcionados en el presente documento. 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 otros.

En realizaciones específicas, uno o más de los componentes del sistema de integración genómica diana descrito en el presente documento pueden proporcionarse en un casete de expresión para la expresión en una célula procariota, una célula eucariota, una célula bacteriana, una célula de levadura o una célula de mamífero u otro organismo. o tipo de celular de interés. El casete puede incluir secuencias reguladoras 5' y 3' operativamente unidas a un polinucleótido proporcionado en el presente documento. "Conectado operativamente" comprende una relación en donde los componentes operan enlazados funcionan de la manera deseada. Por ejemplo, un 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 unidos operativamente pueden ser contiguos o no contiguos. Cuando se usa para referirse a la unión de dos regiones codificantes de proteínas, unidas operativamente significa que las regiones codificantes están en el mismo marco de lectura. En otro caso, una secuencia de ácidos nucleicos que codifica una proteína puede estar unida operativamente a secuencias reguladoras (por ejemplo, promotor, potenciador, secuencia silenciadora, etc.) para conservar la regulación transcripcional adecuada. En un caso, una secuencia de ácidos nucleicos de una región variable de inmunoglobulina (o segmentos V(D)J) puede unirse operativamente a una secuencia de ácidos nucleicos de una región constante de inmunoglobulina para permitir una recombinación adecuada entre las secuencias en una inmunoglobulina pesada o secuencia de la cadena ligera.

El casete puede contener adicionalmente al menos un polinucleótido adicional de interés para ser cointroducido en el organismo. Alternativamente, el polinucleótido adicional de interés puede proporcionarse en múltiples casetes de expresión. Dicho casete de expresión está provisto de una pluralidad de sitios de restricción y/o sitios de recombinación para 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.

El casete de expresión puede incluir en la dirección de transcripción 5'-3', una región de iniciación de la transcripción y la traducción (es decir, un promotor), un polinucleótido recombinante proporcionado aquí, y una región de terminación de la transcripción y traducción (es decir, la región de terminación) funcional en célula de mamífero o una célula huésped de interés. Las regiones reguladoras (es decir, los promotores, las regiones reguladoras de la transcripción y las regiones de terminación de la traducción) y/o un polinucleótido proporcionado en 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 heterólogos 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 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 unido 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 unido operativamente, puede ser nativa con la célula huésped o puede derivar 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. Hacia este fin, pueden emplearse adaptadores o enlazadores para unir los fragmentos de ADN u otras manipulaciones pueden estar implicadas para proporcionar sitios de restricción convenientes, eliminación de ADN superfluo, eliminación de sitios de restricción, o similares. Para este propósito, pueden estar implicadas mutagénesis *in vitro*, reparación de cebadores, restricción, recocido, resustituciones, 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. Dichas 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 iniciación de la transcripción, un ribosoma sitio de unión, una señal de procesamiento de ARN, un sitio de terminación de la transcripción y/o una señal de poliadenilación.

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

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

Los diversos métodos y composiciones proporcionados en el presente documento pueden emplear marcadores de selección. Se pueden usar diversos marcadores de selección en los métodos y composiciones divulgados en el presente documento. Dichos marcadores de selección pueden, por ejemplo, conferir resistencia a un antibiótico tal como G418, higromicina, blastocidina, neomicina o puromicina. Dichos marcadores de selección incluyen neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa ($puro^f$) y blasticidina S desaminasa (bsr^f). En otras realizaciones más, el marcador de selección está unido operativamente a un promotor inducible y la expresión del marcador de selección es tóxica para la célula. Ejemplos no limitantes de dichos marcadores de selección incluyen xantina/guanina fosforribosil transferasa (gpt), haoxipantantina-guanina fosforribosiltransferasa (HGPRT) o timidina quinasa del virus del herpes simple (HSV-TK). El polinucleótido que codifica los marcadores de selección está unido operativamente a un promotor activo en la célula.

iii. Vectores de direccionamiento

Los vectores de direccionamiento se emplean para introducir el ácido nucleico de inserción en el locus diana del ácido nucleico de rata. El vector de direccionamiento comprende el ácido nucleico de inserto y además comprende un brazo de homología 5' y 3', que flanquean el ácido nucleico de inserto. Los brazos de homología, que flanquean el ácido nucleico de inserción, corresponden a regiones dentro del locus diana del ácido nucleico de rata. Para facilitar la referencia, las regiones genómicas afines correspondientes dentro del locus genómico diana se denominan en el presente documento "sitios objetivo". Por ejemplo, un vector de direccionamiento puede comprender un primer ácido nucleico de inserción flanqueado por un primer y un segundo brazo de homología complementarios a un primer y un segundo sitio de destino. Como tal, el vector de direccionamiento ayuda de ese modo a la integración del ácido nucleico de inserción en el objetivo locus del ácido nucleico de rata a través de un evento de recombinación homóloga que se produce entre los brazos de homología y los sitios diana complementarios dentro del genoma de la célula.

En una realización, el locus diana del ácido nucleico de rata comprende una primera secuencia de ácidos nucleicos que es complementaria al brazo de homología 5' y una segunda secuencia de ácidos nucleicos que es complementaria al brazo de homología 3'. En una realización, la primera y la segunda secuencias de ácidos nucleicos están separadas por al menos 5kb. En otra realización, la primera y la segunda secuencias de ácidos nucleicos están separadas por al menos 5kb pero menos de 200kb. En una realización, la primera y la segunda secuencias de ácidos nucleicos están separadas por al menos 10 kb. En una realización, la primera y la segunda secuencia de ácidos nucleicos están separadas por al menos 20kb, al menos 30kb, al menos 40kb, al menos 50kb, al menos 60kb, al menos 70kb, al menos 80kb, al menos 90kb, al menos 100kb, al menos 110kb, al menos 120kb, al menos 130kb, al menos 140kb, al menos 150kb, al menos 160kb, al menos 170kb, al menos 180kb, al menos 190kb, o al menos 200kb. En otras realizaciones adicionales, la primera y la segunda secuencia de ácidos nucleicos están separadas por al menos 5kb pero menos de 10kb, al menos 5kb pero menos de 3Mb, al menos 10kb pero menos de 20kb, al menos 20kb pero menos de 40kb, al menos 40kb pero menos de 60kb, al menos 60kb pero menos de 80kb, al menos aproximadamente 80kb pero menos de 100kb, al menos 100kb pero menos de 150kb, o al menos 150kb pero menos de 200kb, al menos aproximadamente 200kb pero menos de aproximadamente 300kb, al menos unos 300kb pero menos de unos 400kb, al menos unos

400kb pero menos de unos 500kb, al menos unos 500kb pero menos de aproximadamente 1Mb, al menos aproximadamente 1.5Mb pero menos de unos 2Mb, al menos aproximadamente 1Mb pero menos de unos 1.5Mb, al menos aproximadamente 2Mb, pero menos de 2.5Mb, al menos aproximadamente 2.5Mb pero menos de 3Mb, o al menos aproximadamente 2Mb pero menos de aproximadamente 3Mb.

Un brazo de homología del vector de direccionamiento puede tener cualquier longitud que sea suficiente para promover un evento de recombinación homóloga con un sitio objetivo correspondiente, que incluya, por ejemplo, al menos 5-10kb, 5-15kb, 10-20kb, 20-30kb, 30-40kb, 40-50kb, 50-60kb, 60-70kb, 70-80kb, 80-90kb, 90-100kb, 100-110kb, 110-120kb, 120-130kb, 130-140kb, 140-150kb, 150-160kb, 160-170kb, 170-180kb, 180-190kb, 190-200kb de longitud o más. Como se describe con más detalle a continuación, los vectores de orientación grandes pueden emplear brazos de orientación de mayor longitud. En una realización específica, la suma total del brazo de homología 5' y el brazo de homología 3' es al menos 10kb o la suma total del brazo de homología 5' y el brazo de homología 3' es de al menos aproximadamente 16kb a aproximadamente 100kb o alrededor de 30kb a unos 100kb. En otras realizaciones, el tamaño de la suma total del total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 10kb a aproximadamente 150kb, de aproximadamente 10kb a aproximadamente 100kb, de aproximadamente 10kb a aproximadamente 75kb, de 20kb a aproximadamente 150kb, aproximadamente 20kb a aproximadamente 100kb, aproximadamente 20kb a aproximadamente 75kb, aproximadamente 30kb a aproximadamente 150kb, aproximadamente 30kb a aproximadamente 100kb, aproximadamente 30kb a aproximadamente 75kb, aproximadamente 40kb a aproximadamente 150kb, aproximadamente 40kb a aproximadamente 100kb, aproximadamente 40kb, aproximadamente 75kb, aproximadamente 50kb a aproximadamente 150kb, aproximadamente 50kb a aproximadamente 100kb, o aproximadamente 50kb a aproximadamente 75kb, aproximadamente 10kb a aproximadamente 30kb, aproximadamente 20kb a aproximadamente 40kb, aproximadamente 40kb a aproximadamente 60kb, aproximadamente 60kb a aproximadamente 80kb, aproximadamente 80kb, aproximadamente 100kb, aproximadamente 100kb a aproximadamente 120kb, o desde aproximadamente 120kb hasta aproximadamente 150kb. 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.

Cuando se emplean agentes de nucleasa, las regiones genómicas afines correspondientes a los brazos de homología 5' y 3' de un vector de direccionamiento están "ubicadas en una proximidad suficiente" a los sitios diana de nucleasa para promover la aparición de un evento de recombinación homóloga entre las regiones genómicas afines y los brazos de homología en una ruptura de muesca o doble hebra en el sitio de reconocimiento. Por ejemplo, los sitios objetivo de las nucleasas pueden ubicarse en cualquier lugar entre las regiones genómicas afines correspondientes a los brazos de homología 5' y 3'. En realizaciones específicas, el sitio de reconocimiento es inmediatamente adyacente a al menos una o ambas regiones genómicas afines.

Como se usa en el presente documento, un brazo de homología y un sitio objetivo (es decir, una región genómica relacionada) se "complementan" o son "complementarios" 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 o "complementaria". 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 complementaria de homología entre el brazo de homología y el sitio diana complementario 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 complementario pueden comprender regiones complementarias de homología que son al menos 5-10kb, 5-15kb, 10-20kb, 20-30kb, 30-40kb, 40-50kb, 50-60kb , 60-70kb, 70-80kb, 80-90kb, 90-100kb, 100-110kb, 110-120kb, 120-130kb, 130-140kb, 140-150kb, 150-160kb, 160-170kb, 170-180kb, 180-190kb, 190-200kb de longitud o mayor (como se describe en los vectores LTVEC descritos en otra parte del presente documento) de modo que el brazo de homología tenga suficiente homología para experimentar recombinación homóloga con los sitios diana correspondientes dentro del genoma de la célula. Para facilitar la referencia, los brazos de homología se denominan en el presente documento como un brazo de homología 5' y 3'. Esta terminología se relaciona con la posición relativa de los brazos de homología con respecto al ácido nucleico de inserción dentro del vector de direccionamiento.

Por lo tanto, los brazos de homología del vector de direccionamiento están diseñados para ser complementarios de un sitio objetivo con el locus diana. Por lo tanto, los brazos de homología pueden ser complementarios a un locus que es nativo a la célula, o alternativamente pueden ser complementarios a una región de un segmento heterólogo o exógeno de ADN que se integró en el genoma de la célula, incluidos, entre otros a, transgenes, casetes de expresión, o regiones heterólogas o exógenas del ADN genómico. Alternativamente, los brazos de homología del vector de direccionamiento pueden ser complementarios a una región de un cromosoma artificial humano o cualquier otra región genómica diseñada contenida en una célula huésped apropiada. Aún más, los brazos de homología del vector de direccionamiento pueden ser complementarios o derivados de una región de una biblioteca BAC, una biblioteca de cósmidos o una biblioteca de fagos P1. Por lo tanto, en realizaciones específicas, los brazos de homología del vector

de direccionamiento son complementarios a un locus genómico de rata que es nativo, heterólogo o exógeno a una célula dada. En otras formas de realización, los brazos de homología son complementarios a un locus genómico de rata que no se puede atacar utilizando un método convencional o que pueden ser atacados solo incorrectamente o con una eficiencia significativamente baja, en ausencia de una ruptura de muesca o doble cadena inducida por una nucleasa agente. En una realización, los brazos de homología se derivan de un ADN sintético.

En otras realizaciones más, los brazos de homología 5' y 3' son complementarios al mismo genoma que el genoma direccionado. En una realización, los brazos de homología son de un genoma relacionado, por ejemplo, el genoma direccionado es un genoma de rata de una primera cepa, y los brazos direccionados son de un genoma de rata de una segunda cepa, en donde la primera cepa y la segunda cepa son diferente. En otras realizaciones, los brazos de homología son del genoma del mismo animal o del genoma de la misma cepa, por ejemplo, el genoma objetivo es un genoma de rata de una primera cepa, y los brazos direccionados son de un genoma de rata de la misma rata o de la misma cepa.

El vector de direccionamiento (tal como un vector de direccionamiento grande) también puede comprender un casete de selección o un gen indicador como se describe en otra parte en el presente documento. El casete de selección puede comprender una secuencia de ácidos nucleicos que codifica un marcador de selección, en donde la secuencia de ácidos nucleicos está unida operativamente a un promotor. El promotor puede estar activo en una célula procariota de interés y/o activo en una célula eucariótica de interés. 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, un promotor específico del tejido o un promotor específico de la etapa de desarrollo. En una realización, el marcador de selección se selecciona de o comprende neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa ($puro^r$), blasticidina S desaminasa (bsr^r), xantina/guanina fosforribosil transferasa (gpt), y la timidina quinasa del virus del herpes simple (HSV-k), y/o una combinación de los mismos. El marcador de selección del vector de direccionamiento puede estar flanqueado por los brazos de homología 5' y 3' o puede encontrarse 5' o 3' en los brazos de homología.

En una realización, el vector de direccionamiento (como un vector de direccionamiento grande) comprende un gen informador unido operativamente a un promotor, en donde el gen informador codifica una proteína informadora seleccionada del grupo que consiste en o comprende 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/o 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 del informe o la célula, un promotor que es heterólogo al gen informador o a la célula, un promotor específico de la célula, un promotor específico del tejido o un promotor específico de etapa de desarrollo.

En una realización, el uso combinado del vector de direccionamiento (incluyendo, por ejemplo, un vector de direccionamiento grande) con el agente de nucleasa da como resultado un aumento de la 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, o al menos 4 veces cuando se compara con cuando el objetivo vector se utiliza solo.

Cuando se emplea un vector de direccionamiento, el diseño del vector puede ser tal que permita la inserción de una secuencia dada que es de aproximadamente 5 kb a aproximadamente 200 kb como se describe en el presente documento. En una realización, la inserción es de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 30kb, de aproximadamente 30kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 50kb, de aproximadamente 50kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 70kb, desde aproximadamente 80kb a aproximadamente 90kb, desde aproximadamente 90kb a aproximadamente 100kb, desde aproximadamente 100kb a aproximadamente 110kb, desde aproximadamente 110kb a aproximadamente 120kb, desde aproximadamente 120kb a aproximadamente 130kb, desde aproximadamente 130kb a aproximadamente 140kb, desde aproximadamente 140kb a aproximadamente 150kb, desde aproximadamente 150kb a aproximadamente 160kb, desde aproximadamente 160kb a aproximadamente 170kb, desde aproximadamente 170kb a aproximadamente 180kb, desde aproximadamente 180kb a aproximadamente 190kb, o desde aproximadamente 190kb a aproximadamente 200kb, desde aproximadamente 5kb a aproximadamente 10kb, desde aproximadamente 10kb a aproximadamente 20kb, desde aproximadamente 20kb a aproximadamente 40kb, desde aproximadamente 40kb a aproximadamente 60kb, desde aproximadamente 60kb a aproximadamente 80kb, desde aproximadamente 80kb a aproximadamente 100kb, desde aproximadamente 100kb a aproximadamente 150kb, desde aproximadamente 150kb a aproximadamente 200kb, desde aproximadamente 200kb a aproximadamente 250kb, desde aproximadamente 250kb a aproximadamente 300kb, desde aproximadamente 300kb a aproximadamente 350kb, o desde alrededor de 350kb hasta alrededor de 400kb.

Cuando se emplea un vector de direccionamiento, el diseño del vector puede ser tal que permita el reemplazo de una secuencia dada que es de aproximadamente 5kb a aproximadamente 200kb o de aproximadamente 5kb a aproximadamente 3.0Mb como se describe aquí. En una realización, el reemplazo es de aproximadamente 5kb a

aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 30kb, de aproximadamente 30kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 50kb, de aproximadamente 50kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 70kb, desde aproximadamente 80kb a aproximadamente 90kb, desde aproximadamente 90kb a aproximadamente 100kb, desde aproximadamente 100kb a aproximadamente 110kb, desde aproximadamente 110kb a aproximadamente 120kb, desde aproximadamente 120kb a aproximadamente 130kb, desde aproximadamente 130kb a aproximadamente 140kb, desde aproximadamente 140kb a aproximadamente 150kb, desde aproximadamente 150kb a aproximadamente 160kb, desde aproximadamente 160kb a aproximadamente 170kb, desde aproximadamente 170kb a aproximadamente 180kb, desde aproximadamente 180kb a aproximadamente 190kb, desde aproximadamente 190kb a aproximadamente 200kb, desde aproximadamente 5kb a aproximadamente 10kb, desde aproximadamente 10kb de aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150kb, o de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 200kb a unos 300kb, desde unos 300kb a unos 400kb, desde unos 400kb a unos 500kb, de aproximadamente 500kb a aproximadamente 1Mb, de aproximadamente 1Mb a aproximadamente 1.5Mb, de aproximadamente 1.5Mb a aproximadamente 2Mb, de aproximadamente 2Mb a aproximadamente 2.5Mb, o de aproximadamente 2.5Mb a aproximadamente 3Mb.

En una realización, el vector de direccionamiento comprende un gen de recombinasa específico del sitio. En una realización, el gen de recombinasa específico del sitio codifica una recombinasa Cre. En una realización, el gen de la recombinasa Cre es Crei, en donde dos Exones que codifican la recombinasa Cre están separados por un intrón para evitar su expresión en una célula procariota.

En una realización, el gen de recombinasa Cre comprende además una señal de localización nuclear para facilitar la localización de Cre (o cualquier agente de recombinasa o nucleasa) en el núcleo (por ejemplo, el gen es un gen NL-Cre). En una realización específica, el gen de la recombinasa Cre comprende además una señal de localización nuclear y un intrón (por ejemplo, NL-Crei).

En diversas realizaciones, un promotor adecuado para la expresión del agente de nucleasa (que incluye la recombinasa Cre o Crei discutida anteriormente) se selecciona de o comprende un Prm1, Blimp1, Gata6, Gata4, Igf2, Lhx2, Lhx5 y/o Pax3. En una realización específica, el promotor es el promotor Gata6 o Gata4. Los diversos promotores pueden ser de cualquier organismo, incluido, por ejemplo, un roedor como un ratón o una rata. En otra realización específica, el promotor es un promotor Prm1. En otra realización específica, el promotor es un promotor Prm1 de rata. En otra realización específica, el promotor es un promotor Prm1 de ratón. En otra realización específica, el promotor es un promotor Blimp1 o un fragmento del mismo, por ejemplo, un fragmento de 1kb o 2kb de un promotor Blimp1. Véase, por ejemplo, la Patente de los Estados Unidos 8,697,851 y la Publicación de Solicitud de los Estados Unidos 2013-0312129.

iv. Vectores de direccionamiento grandes

El término "vector de direccionamiento grande" o "LTVEC", como se usa en el presente documento, comprende vectores de direccionamiento grandes que comprenden brazos de homología que corresponden y se derivan de secuencias de ácidos nucleicos más grandes que las utilizadas típicamente por otros enfoques destinados a realizar un direccionamiento homólogo en las células y/o que comprende ácidos nucleicos de inserción que comprenden secuencias de ácidos nucleicos más grandes que las utilizadas típicamente por otras metodologías destinadas a realizar un direccionamiento de recombinación homóloga en células. Por ejemplo, el LTVEC hace posible la modificación de loci grandes que no pueden ser acomodados por los vectores de direccionamiento basados en plásmidos tradicionales debido a sus limitaciones de tamaño. En realizaciones específicas, los brazos de homología y/o el ácido nucleico de inserción de LTVEC comprenden la secuencia genómica de una célula eucariota. El tamaño del LTVEC es demasiado grande para permitir el rastreo de eventos de direccionamiento mediante ensayos convencionales, por ejemplo, inmunotransferencia Southern y PCR de largo alcance (por ejemplo, 1kb-5kb). 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). Ejemplos no limitativos de LTVEC y los métodos para hacerlos se describen, por ejemplo, en la Patente de los Estados Unidos No. 6,586,251, 6,596,541, 7,105,348, y WO 2002/036789 (PCT/US01/45375), y US 2013/0137101.

El LTVEC puede ser de cualquier longitud, incluyendo, pero no limitado a, de aproximadamente 20kb a aproximadamente 400kb, de aproximadamente 20kb a aproximadamente 30kb, de aproximadamente 30kb a 40kb, de aproximadamente 40kb a aproximadamente 50kb, de aproximadamente 50kb a aproximadamente 75kb, de aproximadamente 75kb a aproximadamente 100kb, de aproximadamente 100kb a 125kb, de aproximadamente 125kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 175kb, aproximadamente 175kb a aproximadamente 200kb, de aproximadamente 200kb a aproximadamente 225kb, de aproximadamente 225kb a aproximadamente 250kb, de aproximadamente 250kb a aproximadamente 275kb o de aproximadamente 275kb a aproximadamente 300kb, de aproximadamente 200kb a aproximadamente 300kb, de aproximadamente 300kb a aproximadamente 350kb, de aproximadamente 350kb a aproximadamente 400kb, de aproximadamente 350kb a aproximadamente 550kb. En una realización, el LTVEC es de aproximadamente 100 kb.

En una realización, el LTVEC comprende un inserto de ácido nucleico que varía de aproximadamente 5kb a aproximadamente 200kb, de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 30kb, de aproximadamente 0.5kb a aproximadamente 30kb, de aproximadamente 0.5kb a aproximadamente 40kb, desde aproximadamente 30kb a aproximadamente 150kb, desde aproximadamente 0.5kb a aproximadamente 150kb, desde aproximadamente 30kb a aproximadamente 40kb, desde aproximadamente 40kb a aproximadamente 50kb, desde aproximadamente 60kb a aproximadamente 70kb, desde aproximadamente 80kb a aproximadamente 90kb, desde aproximadamente 90kb a aproximadamente 100kb, desde aproximadamente 100kb a aproximadamente 110kb, desde aproximadamente 120kb a aproximadamente 130kb, desde aproximadamente 130kb a aproximadamente 140kb, desde aproximadamente 140kb a aproximadamente 150kb, desde aproximadamente 150kb a aproximadamente 160kb, desde aproximadamente 160kb a aproximadamente 170kb, desde aproximadamente 170kb a aproximadamente 180kb, desde aproximadamente 180kb a aproximadamente 190kb, o desde aproximadamente 190kb a aproximadamente 200kb, desde aproximadamente 5kb a aproximadamente 10kb, desde aproximadamente 10kb a aproximadamente 20kb, desde aproximadamente 20kb a aproximadamente 40kb, desde aproximadamente 40kb a aproximadamente 60kb, desde aproximadamente 60kb a aproximadamente 80kb, desde aproximadamente 80kb a aproximadamente 100kb, desde aproximadamente 100kb a aproximadamente 150kb, para de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 200kb a aproximadamente 250kb, de aproximadamente 250kb a aproximadamente 300kb, de aproximadamente 300kb a aproximadamente 350kb, o de aproximadamente 350kb a aproximadamente 400kb;

Cuando se emplea un LTVEC, el diseño del vector puede ser tal que permita la sustitución de una secuencia dada que es de aproximadamente 5kb a aproximadamente 200kb o de aproximadamente 5kb a aproximadamente 3Mb como se describe en el presente documento. En una realización, el reemplazo es de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 30kb, de aproximadamente 30kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 50kb, de aproximadamente 50kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 70kb, desde aproximadamente 80kb a aproximadamente 90kb, desde aproximadamente 90kb a aproximadamente 100kb, desde aproximadamente 100kb a aproximadamente 110kb, desde aproximadamente 110kb a aproximadamente 120kb, desde aproximadamente 120kb a aproximadamente 130kb, desde aproximadamente 130kb a aproximadamente 140kb, desde aproximadamente 140kb a aproximadamente 150kb, desde aproximadamente 150kb a aproximadamente 160kb, desde aproximadamente 160kb a aproximadamente 170kb, desde aproximadamente 170kb a aproximadamente 180kb, desde aproximadamente 180kb a aproximadamente 190kb, desde aproximadamente 190kb a aproximadamente 200kb, desde aproximadamente 5kb a aproximadamente 10kb, desde aproximadamente 10kb de aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150kb, o de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 200kb a unos 300kb, desde unos 300kb a unos 400kb, desde unos 400kb a unos 500kb, de aproximadamente 500kb a aproximadamente 1Mb, de aproximadamente 1Mb a aproximadamente 1.5Mb, de aproximadamente 1.5Mb a aproximadamente 2Mb, de aproximadamente 2Mb a aproximadamente 2.5Mb, o de aproximadamente 2.5Mb a aproximadamente 3Mb.

En una realización, los brazos de homología de LTVEC se derivan de una biblioteca BAC, una biblioteca de cósmidos o una biblioteca de fagos P1. En otras realizaciones, los brazos de homología se derivan del locus genómico diana de la célula y, en algunos casos, el locus genómico diana, que el LTVEC está diseñado para direccionar, no es direccionable utilizando un método convencional. En otras realizaciones más, los brazos de homología se derivan de un ADN sintético.

En una realización, una suma total del brazo de homología 5' y el brazo de homología 3' en el LTVEC es de al menos 10 kb. En otras realizaciones, la suma total de los brazos de homología 5' y 3' del LTVEC es de aproximadamente 10kb a aproximadamente 30kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, de 100kb a aproximadamente 120kb, de aproximadamente 120kb a aproximadamente 140kb, de aproximadamente 140kb a aproximadamente 160kb, de aproximadamente 160kb a aproximadamente 180kb, de aproximadamente 180kb a aproximadamente 200kb. En una realización, la suma total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 30 kb a aproximadamente 100 kb. En otras realizaciones, el tamaño de la suma total del total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 10kb a aproximadamente 150kb, de aproximadamente 10kb a aproximadamente 100kb, de aproximadamente 10kb a aproximadamente 75kb, de 20kb a aproximadamente 150kb, aproximadamente 20kb a aproximadamente 100kb, aproximadamente 20kb a aproximadamente 75kb, aproximadamente 30kb a aproximadamente 150kb, aproximadamente 30kb a aproximadamente 100kb, aproximadamente 30kb a aproximadamente 75kb, aproximadamente 40kb a aproximadamente 150kb, aproximadamente 40kb a aproximadamente 100kb, aproximadamente 40kb a aproximadamente 75kb, aproximadamente 50kb a aproximadamente 150kb, aproximadamente 50kb a aproximadamente 100kb, o aproximadamente 50kb a aproximadamente 75kb, aproximadamente 10kb a aproximadamente 30 kb, aproximadamente 20kb a aproximadamente 40kb, aproximadamente 40kb a aproximadamente 60kb, aproximadamente 60kb a aproximadamente 80kb, aproximadamente 80kb, aproximadamente 100kb, aproximadamente 100kb a aproximadamente 120kb, o desde

aproximadamente 120kb hasta aproximadamente 150kb. 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 otras realizaciones, el brazo de homología 5' varía de aproximadamente 5 kb a aproximadamente 100 kb. En una realización, el brazo de homología 3' varía de aproximadamente 5 kb a aproximadamente 100 kb. En otras realizaciones, la suma total de los brazos de homología 5' y 3' es de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 30kb, de aproximadamente 30kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 50kb, de aproximadamente 50kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 70kb, de aproximadamente 70kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 90kb, de aproximadamente 90kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 110kb, de aproximadamente 110kb a aproximadamente 120kb de aproximadamente 120kb a aproximadamente 130kb, de aproximadamente 130kb a aproximadamente 140kb, de aproximadamente 140kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 160kb, de aproximadamente 160kb a aproximadamente 170kb, de aproximadamente 170kb a aproximadamente 180kb, de aproximadamente 180kb a aproximadamente 190kb, de aproximadamente 190kb a aproximadamente 200kb, o de aproximadamente 30kb a aproximadamente 100kb, aproximadamente 10kb a aproximadamente 30kb, aproximadamente 20kb a aproximadamente 40kb, aproximadamente 40kb a aproximadamente 60kb, aproximadamente 60kb a aproximadamente 80kb, aproximadamente 80kb a aproximadamente 100kb, aproximadamente 100kb a aproximadamente 120kb , o desde aproximadamente 120kb hasta aproximadamente 150kb.

En una realización, el LTVEC comprende un ácido nucleico de inserción que es homólogo u ortólogo a una secuencia de ácidos nucleicos de rata flanqueada por los brazos de homología de LTVEC. En una realización, la secuencia de ácidos nucleicos de inserción es de una especie distinta de una rata. En una realización, el ácido nucleico de inserción que es homólogo u ortólogo a la secuencia de ácidos nucleicos de rata es un ácido nucleico de mamífero. En una realización, el ácido nucleico de mamífero es un ácido nucleico de ratón. En una realización, el ácido nucleico de mamífero es un ácido nucleico humano. En una realización, el ácido nucleico de inserción es un ADN genómico. En una realización, el inserto es de 5kb a 200kb como se describe anteriormente.

En una realización, el LTVEC comprende un casete de selección o un gen indicador. Varias formas del gen de casete de selección y informador que pueden emplearse se discuten en otra parte en el presente documento.

Como se describe en otra parte en el presente documento, el LTVEC también puede usarse en los métodos proporcionados en el presente documento en combinación con un agente de nucleasa que promueve una recombinación homóloga entre el vector de direccionamiento y el locus diana de un ácido nucleico de rata en una célula de rata pluripotente.

En una realización, el vector de direccionamiento grande (LTVEC) comprende un gen de recombinasa específico del sitio. En una realización, el gen de recombinasa específico del sitio codifica una recombinasa Cre. En una realización, el gen de la recombinasa Cre es Crei, en donde dos Exones que codifican la recombinasa Cre están separados por un intrón para evitar su expresión en una célula procariota. En una realización, el gen de recombinasa Cre comprende además una señal de localización nuclear para facilitar la localización de Cre (o cualquier agente de recombinasa o nucleasa) en el núcleo (por ejemplo, el gen es un gen NL-Cre). En una realización específica, el gen de la recombinasa Cre comprende además una señal de localización nuclear y un intrón (por ejemplo, NL-Crei).

En diversas realizaciones, un promotor adecuado para la expresión del agente de nucleasa (que incluye la recombinasa Cre o Crei discutida anteriormente) se selecciona entre Prm1, Blimp1, Gata6, Gata4, Igf2, Lhx2, Lhx5 y Pax3. En una realización específica, el promotor es el promotor Gata6 o Gata4. Los diversos promotores pueden ser de cualquier organismo, incluido, por ejemplo, un roedor como un ratón o una rata. En otra realización específica, el promotor es un promotor Prm1. En otra realización específica, el promotor es un promotor Prm1 de rata. En otra realización específica, el promotor es un promotor Prm1 de ratón. En otra realización específica, el promotor es un promotor Blimp1 o un fragmento del mismo, por ejemplo, un fragmento de 1 kb o 2 kb de un promotor Blimp1. Consulte, por ejemplo, la Patente de los Estados Unidos 8,697,851 y la Publicación de Solicitud de los Estados Unidos 2013-0312129.

En una realización, el LTVEC comprende un ácido nucleico de inserción que puede producir una eliminación, adición, reemplazo o una combinación de los mismos de una región del locus ApoE de rata, el locus IL-2Rg, el locus Rag2, el locus Rag1 y/o el locus Rag2/Rag1 como se explica en detalle en otra parte del presente documento. En realizaciones específicas, la modificación genética en el locus ApoE da como resultado una disminución, un aumento o una modulación en la actividad ApoE, la actividad IL-2Rg, la actividad Rag2, la actividad Rag1 y/o la actividad Rag2 y Rag1. En una realización, se genera un ApoE bloqueado, y un IL-2Rg bloqueado, un Rag2 bloqueado, un Rag1 bloqueado, se genera un Rag2/Rag1 bloqueado. Como se explica más adelante, los agentes de nucleasa pueden emplearse con cualquiera de los sistemas de orientación LTVEC para dirigirse a cualquier locus genómico de interés.

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

Como se describe en detalle anteriormente, los agentes de nucleasa pueden utilizarse en los métodos y composiciones descritos aquí para ayudar en la modificación del locus diana tanto en una célula procariota como en una célula de rata pluripotente. Dicho agente de nucleasa puede promover la recombinación homóloga entre el vector de direccionamiento y el locus diana. En una realización, el agente de nucleasa comprende un agente de endonucleasa.

Como se usa en el presente documento, el término "sitio de reconocimiento para un agente de nucleasa" comprende una secuencia de ADN en donde se induce una ruptura de muesca o 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 para la célula y para los polinucleótidos de interés que se desea posicionar en el locus genómico diana. 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 una ruptura de muesca o de doble cadena en el sitio de reconocimiento endógeno.

La longitud del sitio de reconocimiento puede variar, e incluye, por ejemplo, sitios de reconocimiento que tienen al menos 4, 6, 8, 10, 12, 14, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70 o más nucleótidos de longitud. 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 subrangos (por ejemplo, 9-18 nucleótidos). El sitio de reconocimiento podría ser palindrómico, es decir, la secuencia en una cadena se lee igual en la dirección opuesta en la cadena complementaria. 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 escisión/muesca independientemente de si el sitio de escisión/muesca 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 sobresalientes de una sola hebra, también llamados "extremos pegajosos", que puede ser 5' sobresalientes, o 3' sobresalientes.

Cualquier agente de nucleasa que induzca una ruptura de muesca o de doble cadena en un sitio de reconocimiento deseado puede usarse en los métodos y composiciones descritos aquí. Puede emplearse un agente de nucleasa natural o nativo siempre que el agente de nucleasa induzca una ruptura de muesca o de doble cadena en un sitio de reconocimiento deseado. Alternativamente, puede emplearse un agente de nucleasa modificado o manipulado. Un "agente de nucleasa diseñado" comprende una nucleasa que está diseñada (modificada o derivada) de su forma nativa para reconocer específicamente e inducir una ruptura de muesca o de doble cadena en el sitio de reconocimiento deseado. Por lo tanto, un agente de nucleasa diseñado 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 una ruptura de muesca o de doble cadena en un sitio de reconocimiento, en donde el sitio de reconocimiento no era una secuencia que hubiera sido reconocida por un agente de nucleasa nativo (no manipulado o no modificado). Producir una ruptura de muesca o de doble hebra 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 secuencia. Los ensayos para medir la ruptura de doble cadena de un sitio de reconocimiento por un agente de nucleasa son conocidos en la técnica y generalmente miden la capacidad de una nucleasa para cortar el sitio de reconocimiento.

El sitio de reconocimiento del agente de nucleasa se puede colocar en cualquier lugar dentro o cerca del locus diana. El sitio de reconocimiento puede estar ubicado dentro de una región de codificación de un gen, o dentro de regiones reguladoras, que influyen en la expresión del gen. Por lo tanto, un sitio de reconocimiento del agente de nucleasa se puede ubicar en un intrón, un exón, un promotor, un potenciador, una región reguladora o cualquier región codificante no proteica.

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 rupturas de doble cadena en secuencias diana específicas en el genoma de un organismo procariótico o eucariota. Las nucleasas efectoras TAL se crean al fusionar un efector similar a un activador de transcripción (TAL) nativo o

diseñado, o una parte funcional de la misma, con el dominio catalítico de una endonucleasa, como, por ejemplo, FokI. El dominio de unión al ADN del efector TAL, único y modular, 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 diseñarse para reconocer sitios específicos del ADN y, por lo tanto, usarse para hacer rupturas 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.

Ejemplos de nucleasas TAL adecuadas y métodos para preparar nucleasas TAL adecuadas se describen, por ejemplo, en la Solicitud de Patente de los Estados Unidos No. 2011/0239315 A1, 2011/0269234 A1, 2011/0145940 A1, 2003/0232410 A1, 2005/0208489 A1, 2005/0026157 A1, 2005/0064474 A1, 2006/0188987 A1 y 2006/0063231 A1. En diversas realizaciones, las nucleasas efectoras TAL están diseñadas para cortar en o cerca de una secuencia de ácidos nucleicos diana en, por ejemplo, un locus genómico de interés, en donde la secuencia de ácidos nucleicos diana 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 a o cerca de secuencias de ácidos nucleicos diana para ser modificadas por vectores de direccionamiento como se describe aquí.

En una realización, cada monómero de TALEN comprende 12-25 repeticiones de TAL, en donde cada repetición de TAL se une a un subsitio de 1 pb. En una realización, el agente 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 TAL y un segundo dominio de unión a ADN basado en repetición TAL, en donde cada uno del primer y el segundo dominio de unión de ADN basado en repetición TAL está unido operativamente a una nucleasa FokI, en donde el primer y el segundo dominio de unión al ADN basado en la repetición de TAL reconocen dos secuencias de ADN diana contiguas en cada hebra de la secuencia de ADN diana separadas por aproximadamente 6 pb a aproximadamente 40 pb del sitio deescisión, y en donde las nucleasas FokI se dimerizan y forman una ruptura de doble cadena en una secuencia diana.

En una realización, el agente de nucleasa comprende un primer dominio de unión a ADN basado en repetición TAL y un segundo dominio de unión a ADN basado en repetición TAL, en donde cada uno del primer y el segundo dominio de unión de ADN basado en repetición TAL es ligado operativamente a una nucleasa FokI, en donde el primer y el segundo dominio de unión a ADN basado en la repetición de TAL reconocen dos secuencias de ADN diana contiguas en cada cadena de la secuencia de ADN diana separadas por un sitio de escisión de 5 pb o 6 pb, y en donde las nucleasas de FokI dimerizan y hacer una ruptura de doble hebra.

El agente de nucleasa empleado en los diversos métodos y composiciones divulgados en el presente documento puede comprender además una nucleasa de dedo de zinc (ZFN). En una realización, cada monómero de la ZFN comprende 3 o más dominios de unión a ADN basados en dedos de zinc, en donde cada dominio de unión a ADN basado en dedos de zinc se une a un subsitio de 3pb. En otras realizaciones, el ZFN es una proteína quimérica que comprende un dominio de unión a ADN basado en dedos de zinc operativamente unido a una nucleasa independiente. En una realización, la endonucleasa independiente es una endonucleasa FokI. En una realización, el agente de nucleasa comprende un primer ZFN y un segundo ZFN, en donde cada uno del primer ZFN y el segundo ZFN está unido operativamente a una nucleasa FokI, en donde el primer y el segundo ZFN reconocen dos secuencias de ADN diana contiguas en cada cadena de la secuencia de ADN diana separada por aproximadamente 6pb a aproximadamente 40pb del sitio de escisión o de aproximadamente 5pb a aproximadamente 6pb del sitio de escisión, y en donde las nucleasas FokI se dimerizan y producen una ruptura de doble hebra. Véase, por ejemplo, US20060246567; US20080182332; US20020081614; US20030021776; WO/2002/057308A2; US20130123484; US20100291048; y, WO/2011/017293A2.

En una realización de los métodos proporcionados en el presente documento, el agente de nucleasa comprende (a) una proteína quimérica que comprende un dominio de unión a ADN basado en dedos de zinc fusionado a una endonucleasa FokI; o, (b) una proteína quimérica que comprende una Nucleasa Efectora de Efecto de Activador de Transcripción (TALEN) fusionada a una endonucleasa FokI.

En otra realización más, el agente de nucleasa es una meganucleasa. Las meganucleasas se han clasificado en cuatro familias según motivos de secuencia conservada, las familias son las familias de caja LAGLIDADG (SEQ ID NO: 16), GIY-YIG, H-N-H y His-Cys. Estos motivos participan en la coordinación de iones metálicos e hidrólisis de enlaces fosfodiéster. Las HEasas 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 and 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 por manipulación. 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, véase 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 and Zhao, (2005) Nucleic Acids Res 33: e154; WO2005105989; WO2003078619; WO2006097854; WO2006097853; WO2006097784; y WO2004031346.

Cualquier meganucleasa se puede usar aquí, incluyendo, pero no limitándose a, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, PI-PspI, F-SceI, F-SceII, F-SuvI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoeI, I-CpaI, I-CpaII, I-CsmI, I-CvuI, I-CvuAIP, I-DdiI, I-DdiII, I-DirI, I-DmoI, I-HmuI, I-HmuII, I-HsNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-ScaI, I-SexIP, I-SneIP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIP PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, PI-TliII, 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 diana perfectamente coincidente en el genoma. En una realización, la meganucleasa es una nucleasa de alojamiento. En una realización, la nucleasa de alojamiento es una familia LAGLIDADG (SEQ ID NO: 16) de nucleasa de alojamiento. En una realización, la familia LAGLIDADG (SEQ ID NO: 16) de nucleasa de alojamiento se selecciona entre I-SceI, I-CreI y I-DmoI.

Los agentes de nucleasa pueden comprender además endonucleasas de restricción, que incluyen endonucleasas de Tipo I, Tipo II, Tipo III y Tipo IV. Las endonucleasas de restricción 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 escinden en un lado y en 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, and Belfort et al., (2002) en Mobile DNA II, pp. 761-783, Eds. Craigie et al, (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, por ejemplo, una nucleasa Cas9, que en algunos casos está optimizada con codones para el tipo de célula deseado en donde se va a expresar. El sistema emplea además un constructo ARNcr-ARNtracr 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 diana" para el sitio de reconocimiento dado y el ARNtracr se conoce a menudo como el "andamio". En resumen, un fragmento de ADN corto que contiene la secuencia diana se inserta en un plásmido de expresión de ARN guía. El plásmido de expresión de ARNg comprende la secuencia diana (en algunas realizaciones alrededor de 20 nucleótidos), una forma de la secuencia de ARNtracr (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 reconocen 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 Cas9 se introducen en la célula. Véase, por ejemplo, Mali P et al. (2013) Science 2013 Feb 15; 339 (6121): 823-6; Jinek M et al. Science 2012 17 de agosto; 337 (6096): 816-21; Hwang WY et al. Nat Biotechnol 2013 Mar; 31 (3): 227-9; Jiang W et al. Nat Biotechnol 2013 Mar; 31(3): 233-9; y, Cong L et al. Ciencia 2013 15 de febrero; 339 (6121): 819-23.

En una realización, el método para modificar un locus genómico de interés en una célula de rata pluripotente comprende además la introducción en la célula de rata pluripotente: (a) un primer constructo de expresión que comprende un primer promotor unido operativamente a una primera secuencia de ácidos nucleicos que codifica una proteína (Cas) asociada con Repeticiones Palindrómicas Cortas Regularmente Interespaciadas Aglomeradas (CRISPR); (b) un segundo constructo de expresión que comprende un segundo promotor unido operativamente a una secuencia diana genómica unida a un ARN guía (ARNg), en donde la secuencia diana genómica está flanqueada en el extremo 3' por una secuencia de Motivo Protoespaciador Adyacente (PAM). En una realización, la secuencia diana genómica comprende la secuencia de nucleótidos de GNNNNNNNNNNNNNNNNNGG (GN_{1-20} GG; SEQ ID NO: 1). En una realización, la secuencia diana genómica comprende la SEQ ID NO: 23, en donde N tiene una longitud de entre 1 y 20 nucleótidos. En otra realización, la secuencia diana genómica comprende entre 14 y 20 nucleótidos de longitud de SEQ ID NO: 1.

En una realización, el ARNg comprende una tercera secuencia de ácidos nucleicos que codifica un ARN de CRISPR (ARNCR) de Repeticiones Palindrómicas Cortas Regularmente Interespaciadas Aglomeradas y un ARN de CRISPR (ARNtracr) transactivante. En realizaciones específicas, la proteína Cas es Cas9.

En algunas realizaciones, el ARNg comprende (a) el ARN quimérico de la secuencia de ácidos nucleicos 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUU-3' (SEQ ID NO: 2); o, (b) el ARN quimérico de la secuencia de ácidos nucleicos 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCG-3' (SEQ ID NO: 3).

En otra realización, el ARNcr comprende 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAA-3' (SEQ ID NO: 4); 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAAUAAG (SEQ ID NO: 5); o 5'-GAGUCCGAGCAGAAGAAGAAGUUUUA-3' (SEQ ID NO: 6).

En aún otras realizaciones, el ARNtracr comprende, 5'-AAGGCUAGUCCG-3' (SEQ ID NO: 7) o 5'-AAGGCUAGUCCGU UAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' (SEQ ID NO: 8).

En una realización, la proteína Cas es una proteína Cas de tipo I. En una realización, la proteína Cas es una proteína Cas de tipo II. En una realización, la proteína Cas de tipo II es Cas9. En una realización, la primera secuencia de ácidos nucleicos codifica una proteína Cas optimizada en codón humano.

En una realización, el primer ácido nucleico comprende una mutación que altera al menos un residuo de aminoácido de los sitios activos de nucleasa en la proteína Cas, en donde la proteína Cas mutante genera una ruptura en una sola cadena de la región del ADN objetivo, y en donde la mutación disminuye la recombinación no homóloga en la región del ADN objetivo.

En una realización, el primer ácido nucleico que codifica la proteína Cas comprende además una señal de localización nuclear (NLS). En una realización, la señal de localización nuclear es una señal de localización nuclear SV40.

En una realización, el segundo promotor que impulsa la expresión de la secuencia diana genómica y el ARN guía (ARNg) es un promotor de la ARN polimerasa III. En una realización, el promotor de la ARN polimerasa III es un promotor U6 humano. En una realización, el promotor de ARN polimerasa III es un promotor de U6 polimerasa III de rata. En una realización, el promotor de ARN polimerasa III es un promotor de U6 polimerasa III de ratón.

En una realización, las secuencias de ácidos nucleicos que codifican ARNcr y el ARNtracr se unen a través de un bucle sintético, en donde, tras la expresión, el ARNcr y el ARNtracr forman un dúplex de ARNcr:ARNtracr.

En una realización, el primer constructo de expresión y el segundo constructo de expresión se expresan a partir de un mismo plásmido.

En una realización, los constructos de expresión primero y segundo se introducen junto con el LTVEC. En una realización, los constructos de expresión primero y segundo se introducen por separado del LTVEC durante un período de tiempo.

En una realización, el método comprende introducir una pluralidad de el segundo constructo y una pluralidad de LTVEC para la edición multiplex de loci diana distintos como se describe en el presente documento.

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

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 se puede expresar 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 o que comprende un agente de nucleasa.

En una realización, el ARNc y el ARNtracr se expresan como transcripciones de ARN separadas.

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 operativamente a un promotor activo en la célula. En otras realizaciones, el polinucleótido que codifica el agente de nucleasa está en el mismo vector de direccionamiento que comprende el ácido nucleico 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 ácido nucleico de inserción.

Cuando el agente de nucleasa se proporciona a la célula a través de la introducción de un polinucleótido que codifica el agente de nucleasa, 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.

En una realización, el agente de endonucleasa se introduce junto con el LTVEC. En una realización, el agente de endonucleasa se introduce por separado del LTVEC durante un período de tiempo. En una realización, el agente de endonucleasa se introduce antes de la introducción de la LTVEC. En una realización, el agente de endonucleasa se introduce en la célula ES de rata después de la introducción de la LTVEC.

En una realización, el agente de endonucleasa es un constructo de expresión que comprende una secuencia de ácidos nucleicos que codifica una endonucleasa, en donde la secuencia de ácidos nucleicos está unida operativamente a un promotor. En una realización, el promotor es un promotor constitutivamente activo. En una realización, el promotor es un promotor inducible. En una realización, el promotor es activo en la célula de rata pluripotente. En una realización, el agente de endonucleasa es un ARNm que codifica una endonucleasa.

B. Métodos para integrar un polinucleótido de interés en un locus diana

Se proporcionan métodos para modificar un locus diana de interés. En una realización, un locus diana en una célula de rata pluripotente se dirige a la modificación genética. Dicho método comprende: (a) introducir en la célula de rata pluripotente un vector direccionado que comprende un inserto de ácido nucleico flanqueado con un brazo de homología de rata 5' y un brazo de homología de rata 3'; y (b) identificar una célula de rata pluripotente modificada genéticamente que comprende la modificación genética direccionada en el locus diana, en donde la modificación genética direccionada puede transmitirse a través de la línea germinal. En realizaciones específicas, la suma total del brazo de homología 5' y el brazo de homología 3' es al menos 10kb y se emplea un vector de direccionamiento grande.

En otras realizaciones, el tamaño de la suma total del total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 10kb a aproximadamente 150kb, aproximadamente 10kb a aproximadamente 100kb, aproximadamente 10kb a aproximadamente 75kb, aproximadamente 20kb a aproximadamente 150kb, aproximadamente 20kb a aproximadamente 100kb, aproximadamente 20kb a aproximadamente 75kb, aproximadamente 30kb a aproximadamente 150kb, aproximadamente 30kb a aproximadamente 100kb, aproximadamente 30kb a aproximadamente 75kb, aproximadamente 40kb a aproximadamente 150kb, aproximadamente 40kb, aproximadamente 100kb, aproximadamente 40kb a aproximadamente 75kb, aproximadamente 50kb a aproximadamente 150kb, aproximadamente 50kb a aproximadamente 100kb, o aproximadamente 50kb a aproximadamente 75kb, aproximadamente 10kb a aproximadamente 30kb, aproximadamente 20kb a aproximadamente 40kb, aproximadamente 40kb a aproximadamente 60kb, aproximadamente 60kb a aproximadamente 80kb, aproximadamente 80kb a aproximadamente 100kb, aproximadamente 100kb a aproximadamente 120kb, o desde aproximadamente 120kb a aproximadamente 150kb. 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.

La célula pluripotente de rata puede ser una célula madre embrionaria de rata. En una realización específica, (a) la célula ES de rata se deriva de una cepa DA o una cepa ACI; o, (b) la célula ES de rata se caracteriza por la expresión de un marcador de pluripotencia que comprende Oct-4, Sox-2, fosfatasa alcalina o una combinación de los mismos. En otros casos, la célula madre embrionaria de rata empleada comprende una célula ES de rata como se describe en la Solicitud de Patente de los Estados Unidos No. 14/185,703, presentada el 20 de febrero de 2014.

Como se describe en otra parte del presente documento, el ácido nucleico insertado puede ser cualquier secuencia de ácido nucleico. En realizaciones no limitantes, (a) el inserto de ácido nucleico comprende un reemplazo de una secuencia endógena de ácido nucleico de rata con una secuencia de ácidos nucleicos de mamífero homóloga u ortóloga; (b) el inserto de ácido nucleico comprende una eliminación de una secuencia endógena de ácido nucleico de rata; (c) el inserto de ácido nucleico comprende una eliminación de una secuencia endógena de ácido nucleico de rata, en donde la eliminación varía de 5kb a 200kb o de 5kb a 3Mb (como se describe en detalle en el presente

documento); (d) el inserto de ácido nucleico comprende una adición de una secuencia de ácidos nucleicos exógena (que incluye, por ejemplo, una secuencia de ácidos nucleicos exógena que varía de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb de aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 200kb a aproximadamente 250kb, de aproximadamente 250kb a aproximadamente 300kb, de aproximadamente 300kb a aproximadamente 350kb, o de aproximadamente 350kb a aproximadamente 400kb); (e) el inserto de ácido nucleico comprende una secuencia de ácidos nucleicos exógena que comprende una secuencia de ácidos nucleicos homóloga u ortóloga; (f) la secuencia de ácidos nucleicos homóloga u ortóloga de (a) en donde la secuencia de ácidos nucleicos es una secuencia de ácidos nucleicos humana; (g) el inserto de ácido nucleico comprende la secuencia de ácidos nucleicos homóloga u ortóloga de (a) en donde la secuencia de ácidos nucleicos es una secuencia de ácidos nucleicos quimérica que comprende una secuencia de ácidos nucleicos de rata y humana; (h) el ácido nucleico del inserto comprende la secuencia de ácidos nucleicos exógena de (e), en donde el ácido nucleico del inserto varía de aproximadamente 5 kb a aproximadamente 200 kb; (i) el ácido nucleico de inserción comprende un alelo condicional flanqueado con secuencias diana de recombinasa específicas del sitio; (j) el inserto de ácido nucleico comprende un gen indicador unido operativamente a un promotor; (k) el inserto de ácido nucleico comprende uno o más segmentos del gen V_H de la cadena pesada de la inmunoglobulina humana no reordenado, uno o más segmentos del gen D de la cadena pesada de la inmunoglobulina humana no reordenado, y uno o más J_H de la cadena pesada de la inmunoglobulina humana no reordenado, segmentos del gen, que están unidos de manera funcional a una secuencia de ácidos nucleicos de región constante de cadena pesada de roedor; (1) el ácido nucleico inserto comprende una secuencia de ácidos nucleicos de región variable de cadena pesada de inmunoglobulina humana reorganizada operativamente unida a una secuencia de ácidos nucleicos de región constante de cadena pesada de roedor; (m) el inserto de ácido nucleico comprende uno o más segmentos del gen Y_κ o V_λ de inmunoglobulina humana no reorganizados y uno o más segmentos del gen J_κ o J_λ de inmunoglobulina humana no reorganizados, que están operativamente enlazados a una inmunoglobulina λ de mamíferos λ o κ cadena ligera cadena ligera región constante región nucleica secuencia ácida (n) el inserto de ácido nucleico comprende una secuencia de ácidos nucleicos de región variable de la cadena ligera λ o κ reorganizada humana unida operativamente a una secuencia de ácidos nucleicos de región constante de la cadena ligera λ o κ de cadena ligera κ; (o) la secuencia de ácidos nucleicos de la región constante de la cadena pesada de mamífero de (k) y/o (1) comprende una secuencia de ácidos nucleicos de la región constante de rata, una secuencia de ácidos nucleicos de la región constante humana, o una combinación de las mismas; o, (p) el ácido nucleico de la región constante de la cadena ligera de λ o κ de los mamíferos de (m) y/o (n) comprende una secuencia de ácidos nucleicos de la región constante de rata, una secuencia de ácidos nucleicos de la región constante humana, o una combinación de los mismos.

En una realización, el ácido nucleico inserto comprende uno o más segmentos funcionales del gen V_H humano que comprenden V_H1-2, V_H1-3, V_H1-8, V_H1-18, V_H1-24, V_H1-45, V_H1-46, V_H1-58, V_H1-69, V_H2-5, V_H2-26, V_H2-70, V_H3-7, V_H3-9, V_H3-11, V_H3-13, V_H3-15, V_H3-16, V_H3-20, V_H3-21, V_H3-23, V_H3-30, V_H3-30-3, V_H 3-30-5, V_H3-33, V_H3-35, V_H3-38, V_H3-43, V_H3-48, V_H3-49, V_H3-53, V_H3-64, V_H3-66, V_H3-72, V_H3-73, V_H3-74, V_H4-4, V_H4-28, V_H4-30-1, V_H4-30-2, V_H4-30-4, V_H4-31, V_H4-34, V_H4-39, V_H4-59, V_H4-61, V_H5-51, V_H6-1, V_H7-4-1, V_H7-81, o una combinación de los mismos.

En una realización, el ácido nucleico insertado comprende uno o más segmentos funcionales del gen D humano que comprenden D1-1, D1-7, D1-14, D1-20, D1-26, D2-2, D2-8, D2-15, D2-21, D3-3, D3-9, D3-10, D3-16, D3-22, D4-4, D4-11, D4-17, D4-23, D5-12, D5-5, D5-18, D5-24, D6-6, D6-13, D6-19, D6-25, D7-27, o una combinación de los mismos.

En una realización, el ácido nucleico insertado comprende uno o más segmentos génicos J_H funcionales que comprenden J_H1, J_H2, J_H3, J_H4, J_H5, J_H6, o una combinación de los mismos. En una realización, el ácido nucleico de inserción comprende uno o más segmentos génicos V_K humanos que comprenden Vκ4-1, Vκ5-2, Vκ 7-3, Vκ 2-4, Vκ1-5, Vκ1-6, Vκ3-7, Vκ1-8, Vκ1-9, Vκ2-10, Vκ3-11, Vκ1-12, Vκ1-13, Vκ2-14, Vκ3-15, Vκ1-16, Vκ1-17, Vκ2-18, Vκ2-19, Vκ3-20, Vκ6-21, Vκ1-22, Vκ1-23, Vκ2-24, Vκ3-25, Vκ2-26, Vκ1-27, Vκ2-28, Vκ2-29, Vκ2-30, Vκ3-31, Vκ1-32, Vκ1-33, Vκ3-34, Vκ1-35, Vκ2-36, Vκ1-37, Vκ2-38, Vκ1-39, Vκ2-40, o una combinación de los mismos.

En una realización, el ácido nucleico insertado comprende uno o más segmentos del gen Vλ humano que comprenden Vλ3-1, Vλ4-3, Vλ2-8, Vλ3-9, Vλ3-10, Vλ2-11, Vλ3-12, Vλ2-14, Vλ3-16, Vλ2-18, Vλ3-19, Vλ3-21, Vλ3-22, Vλ2-23, Vλ3-25, Vλ3-27, o una combinación de los mismos.

En una realización, el ácido nucleico insertado comprende uno o más segmentos del gen Jκ humano que comprenden Jκ1, Jκ2, Jκ3, Jκ4, Jκ5, o una combinación de los mismos.

En realizaciones específicas, tras la modificación del locus diana en una célula de rata pluripotente, la modificación genética se transmite a través de la línea germinal.

En una realización, la secuencia de ácidos nucleicos insertada comprende un polinucleótido que cuando se integra en el genoma producirá una modificación genética de una región del locus ApoE de rata, en donde la modificación

genética en el locus ApoE produce una disminución en la actividad ApoE, un aumento en la actividad de ApoE o una modulación de la actividad de ApoE. En una realización, se genera un ApoE bloqueado.

En una realización, la secuencia de ácidos nucleicos insertada comprende un polinucleótido que cuando se integra en el genoma producirá una modificación genética de una región del locus gamma del receptor de Interleucina-2 de rata, en donde la modificación genética en el locus del receptor gamma de la interleucina-2 da como resultado una disminución en la actividad del receptor de interleucina-2, un aumento en la actividad gamma del receptor de interleucina-2, o una modulación de la actividad del receptor de interleucina-2. En una realización, se genera un bloqueo del receptor de interleucina-2.

En otra realización más, la secuencia de ácidos nucleicos insertada comprende un polinucleótido que cuando se integra en el genoma producirá una modificación genética de una región del locus Rag1 de rata, el locus Rag2 de rata y/o el locus de rata Rag2/Rag1, en donde la modificación genética en la rata Rag1, Rag2 y/o Rag2/Rag1 produce una disminución en la actividad de la proteína Rag1, Rag2 o Rag1 y Rag2, un aumento en la actividad de la proteína Rag1, Rag2 o Rag1 y Rag2, o una modulación en actividad de la proteína Rag1, Rag2 o Rag 1 y Rag2. En una realización, se genera un bloqueo de Rag1, Rag2 o Rag2/Rag1.

En realizaciones adicionales, el ácido nucleico insertado da como resultado el reemplazo de una parte del locus ApoE de rata, el locus gamma del receptor de interleucina-2 y/o el locus Rag2, y/o el locus Rag1 y/o el locus Rag2/Rag1 con la porción ortóloga correspondiente de un locus ApoE, un locus gamma del receptor de interleucina-2, un locus Rag2, un locus Rag1 y/o un locus Rag2/Rag1 de otro organismo.

En todavía otras realizaciones, el inserto de ácido nucleico comprende un polinucleótido que comparte en toda su longitud al menos 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%. 98%, 99% a una porción de un locus ApoE, un locus gamma del receptor de interleucina-2, un locus Rag2, un locus Rag1 y/o un locus Rag2/Rag1 que está reemplazando.

El polinucleótido de inserción dado y la región correspondiente del locus de rata que se reemplaza pueden ser una región codificante, un intrón, un exón, una región no traducida, una región reguladora, un promotor o un potenciador o cualquier combinación de los mismos. Además, el polinucleótido de inserto dado y/o la región del locus de rata que se reemplaza puede tener cualquier longitud deseada, incluyendo por ejemplo, entre 10-100 nucleótidos de longitud, 100-500 nucleótidos de longitud, 500-1 kb de nucleótido de longitud, 1Kb a 1.5kb de nucleótido de longitud, 1.5kb a 2kb de nucleótidos de longitud, 2kb a 2.5kb de nucleótidos de longitud, 2.5kb a 3kb de nucleótidos de longitud, 3kb a 5kb de nucleótidos de longitud, 5kb a 8kb de nucleótidos de longitud, 8kb a 10kb de nucleótidos en longitud o más. En otros casos, el tamaño de la inserción o reemplazo es de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb de aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 200kb, de aproximadamente 200kb a aproximadamente 250kb, de aproximadamente 250kb a aproximadamente 300kb, de aproximadamente 300kb a aproximadamente 350kb, de aproximadamente 350kb a aproximadamente 400kb, de aproximadamente 400kb a aproximadamente 800kb, de aproximadamente 800kb a 1Mb, de aproximadamente 1Mb a aproximadamente 1.5Mb, de aproximadamente 1.5Mb a aproximadamente 2Mb, de aproximadamente 2Mb, a aproximadamente 2.5Mb, de aproximadamente 2.5Mb a aproximadamente 2.8Mb, de aproximadamente 2.8Mb a aproximadamente 3Mb. En otras realizaciones, el polinucleótido de inserción dado y/o la región del locus de rata que se reemplaza es al menos 100, 200, 300, 400, 500, 600, 700, 800 o 900 nucleótidos o al menos 1 kb, 2kb, 3kb, 4kb, 5kb, 6kb, 7kb, 8kb, 9kb, 10kb, 11kb, 12kb, 13kb, 14 kb, 15kb, 16kb o más.

i. Métodos para modificar un locus diana de un ácido nucleico de rata a través de recombinación homóloga bacteriana (BHR)

Se proporcionan métodos y composiciones para modificar un locus diana de un ácido nucleico de rata mediante recombinación homóloga bacteriana (BHR) en una célula procariota. Tales métodos encuentran uso en la utilización de recombinación homóloga bacteriana en una célula procariota para modificar genéticamente un locus diana de un ácido nucleico de rata para crear un vector de direccionamiento. Dicho vector de direccionamiento que comprende el locus diana modificado genéticamente puede introducirse en una célula eucariota, por ejemplo, una célula de rata pluripotente. La "recombinación homóloga" incluye el intercambio de fragmentos de ADN entre dos moléculas de ADN en los sitios de cruce dentro de las regiones de homología. Por lo tanto, "recombinación homóloga bacteriana" o "BHR" incluye la recombinación homóloga que se produce en las bacterias.

Se proporcionan métodos para modificar un locus diana de un ácido nucleico de rata mediante recombinación homóloga bacteriana (BHR) que comprenden la introducción en una célula procariota de un vector de direccionamiento que comprende un ácido nucleico de inserción flanqueado con un brazo de homología de rata 5' y un brazo de homología de rata 3', en donde la célula procariota comprende un ácido nucleico de rata y es capaz de expresar una recombinasa que media en el BHR en el locus diana. Dichos vectores de direccionamiento pueden incluir cualquiera de los grandes vectores de direccionamiento divulgados en el presente documento.

En una realización, el método comprende la introducción en una célula procariota: (i) un primer constructo que comprende un ácido nucleico de rata que tiene una secuencia de ADN de interés; (ii) un segundo constructo de direccionamiento que comprende un inserto de ácido nucleico flanqueado con un brazo de homología 5' de rata y un brazo de homología 3' de rata, y (iii) un tercer constructo que codifica una recombinasa que media la recombinación homóloga bacteriana. En una realización, el primero, el segundo y el tercer constructo se introducen en la célula procariota por separado durante un período de tiempo. En una realización, la célula procariota comprende un ácido nucleico que codifica la recombinasa, y el método no requiere la introducción del tercer constructo. En una realización, la recombinasa se expresa bajo el control de un promotor inducible.

En una realización, el primer constructo que comprende el ácido nucleico de rata se deriva de un cromosoma artificial bacteriano (BAC) o un cromosoma artificial de levadura (YAC).

Se puede seleccionar una célula procariota que comprende el ácido nucleico de inserción en el locus genómico diana. Este método puede repetirse en serie como se describe en el presente documento para permitir la introducción de múltiples ácidos nucleicos de inserción en el locus de rata objetivo en la célula procariota. Una vez que el locus de ácido nucleico de rata diana se "construye" dentro de la célula procariota, un vector de direccionamiento que comprende el locus diana de rata modificado puede aislarse de la célula procariota e introducirse en un locus genómico diana dentro de una célula de mamífero (es decir, una célula de rata pluripotente, o una célula madre embrionaria de rata).

Las células de rata preferidas para recibir vectores de direccionamiento se describen en la Solicitud de los Estados Unidos 14/185.703, presentada el 20 de febrero de 2014, cuyos contenidos se resumen en el presente documento. Estas células de rata son células de rata pluripotentes capaces de mantener su pluripotencia después de una o más modificaciones genéticas específicas *in vitro*, y son capaces de transmitir las modificaciones genéticas específicas a través de la línea germinal.

Las células de rata pluripotentes sometidas a electroporación se colocan en placas a una alta densidad para la selección de células resistentes al fármaco que comprenden el vector de direccionamiento. El proceso de selección de medicamentos elimina la mayoría de las células en placa (~99%), dejando atrás colonias individuales, cada una de las cuales es un clon derivado de una sola célula. De las células restantes, la mayoría de las células (~ 80-100%) contienen el vector de direccionamiento (que comprende un casete de selección de medicamentos) integrado en una ubicación aleatoria en el genoma. Por lo tanto, las colonias se seleccionan individualmente y se genotipifican para identificar células ES de rata que albergan el vector de direccionamiento en la ubicación genómica correcta (por ejemplo, utilizando la modificación del ensayo de alelos que se describe a continuación).

Se puede usar un ensayo cuantitativo de alto rendimiento, a saber, la modificación del alelo (MOA), para la genotipificación. Un ensayo de este tipo permite una exploración a gran escala de uno o varios alelos modificados en un cromosoma parental después de una modificación genética. El ensayo de MOA se puede llevar a cabo mediante diversas técnicas analíticas, que incluyen, entre otras, una PCR cuantitativa, por ejemplo, una PCR en tiempo real (qPCR). Por ejemplo, la PCR en tiempo real comprende un primer conjunto de cebadores que reconoce el locus diana y un segundo conjunto de cebadores que reconoce un locus de referencia no direccionado. Además, el conjunto de cebadores comprende una sonda fluorescente que reconoce la secuencia amplificada. En una realización, el ensayo cuantitativo se lleva a cabo a través de Invader Probes®. En una realización, el ensayo cuantitativo se lleva a cabo mediante ensayos MMP®. En una realización, el ensayo cuantitativo se lleva a cabo a través de TaqMan® Molecular Beacon. En una realización, el ensayo cuantitativo se lleva a cabo mediante la tecnología de sonda Eclipse™. (Véase, por ejemplo, US2005/0144655).

La célula de rata pluripotente seleccionada o las células ES de rata que comprenden la modificación genética direccionada pueden introducirse en un embrión de rata huésped, por ejemplo, un embrión de rata en estadio premórula o blastocisto, e implantarse en el útero de una madre sustituta para generar una rata fundadora (rata F0). Posteriormente, la rata fundadora se puede criar en una rata de tipo salvaje para crear heterocigotos de progenie F1 para la modificación genética. El apareamiento de la rata F1 heterozigota puede producir una progenie homozigota para la modificación genética. El apareamiento de la rata F1 heterozigota puede producir una progenie homozigota para la modificación genética. En algunas realizaciones, pueden llevarse a cabo diversas modificaciones genéticas de los loci diana divulgados en el presente documento utilizando un vector de direccionamiento grande (LTVEC) como se describe en detalle en otra parte del presente documento. Por ejemplo, un LTVEC puede derivarse del ADN del Cromosoma Artificial Bacteriano (BAC) utilizando la tecnología de ingeniería genética VELOCIGENE® (Véase, por ejemplo, la Patente de los Estados Unidos No. 6,586,251 y Valenzuela, D.M. et al. (2003), High-throughput engineering of the mouse genome coupled with high-resolution expression analysis, Nature Biotechnology 21(6): 652-659).

El uso de recombinación homóloga bacteriana (BHR) para generar un vector de direccionamiento grande (LTVEC) evita las limitaciones de los plásmidos para alojar un fragmento grande de ADN genómico y la consiguiente baja eficiencia de introducir una modificación direccionada en un locus endógeno en células de rata pluripotentes. Se pueden realizar una o más modificaciones genéticas específicas en la generación de un LTVEC. Un LTVEC de ejemplo producido en la célula procariota puede comprender un ácido nucleico de inserción que porta una secuencia genómica de rata con una o más modificaciones genéticas o un ácido nucleico exógeno (por ejemplo, un homólogo u ortólogo

de un ácido nucleico de rata), que está flanqueado por brazos homólogos de rata complementarios a regiones genómicas específicas.

También se proporcionan células procariotas huésped que comprenden los diversos vectores de direccionamiento divulgados en el presente documento. Dichas células procariotas incluyen, pero no se limitan a, bacterias tales como E. coli. En una realización, una célula procariota huésped comprende un vector de direccionamiento que comprende un ácido nucleico de inserción flanqueado con un brazo de homología de rata 5' y un brazo de homología de rata 3', en donde el ácido nucleico de inserción varía de aproximadamente 5 kb a aproximadamente 200 kb.

La célula procariota huésped puede comprender además un ácido nucleico que codifica un polipéptido de recombinasa o el ácido nucleico que codifica el polipéptido de recombinasa está unido operativamente a un promotor inducible.

Además, se proporcionan diversos métodos y composiciones, que emplean el LTVEC como se describe en el presente documento en combinación con una célula procariota para producir modificaciones genéticas direccionadas. Tales composiciones y métodos se discuten en otra parte en el presente documento.

Se proporcionan métodos para modificar un locus diana de un ácido nucleico mediante recombinación homóloga bacteriana (BHR) que comprenden la introducción en una célula procariota de un vector de direccionamiento que comprende un ácido nucleico de inserción flanqueado con un brazo de homología 5' y un brazo de homología 3', en donde la célula procariota comprende ácidos nucleicos correspondientes a los brazos de homología 5' y 3' y la célula procariota es capaz de expresar una recombinasa que media el BHR en el locus diana. Dichos vectores de direccionamiento pueden incluir cualquiera de los grandes vectores de direccionamiento divulgados en el presente documento. Tales métodos pueden emplear un LTVEC como se describe en detalle en el presente documento y además emplean el sistema CRISPR/Cas como se describe en otro lugar en el presente documento.

ii. Métodos para modificar un locus de interés diana en una célula de rata pluripotente

Se proporciona además un método para modificar un locus diana de interés en una célula de rata pluripotente mediante modificación genética direccionada, que comprende (a) introducir en la célula de rata pluripotente un vector de direccionamiento que comprende un inserto de ácido nucleico flanqueado con un brazo de homología de rata 5' y un brazo de homología de rata 3', en donde la suma total del brazo de homología 5' y el brazo de homología 3' es al menos 10 kb; y (b) identificar una célula de rata pluripotente modificada genéticamente que comprende la modificación genética direccionada en el locus de interés diana. En una realización, la suma total del brazo de homología 5' y el brazo de homología 3' es de al menos aproximadamente 16 kb a aproximadamente 30 kb. En realizaciones específicas, la modificación genética direccionada puede transmitirse a través de la línea germinal. Dichos vectores de direccionamiento pueden incluir cualquiera de los grandes vectores de direccionamiento divulgados en el presente documento.

En un aspecto, se proporciona un método para modificar un locus genómico de interés en una célula de rata pluripotente mediante modificación genética direccionada, que comprende: (a) proporcionar una célula de rata pluripotente que es capaz de mantener su pluripotencia después de al menos una modificación genética direccionada su genoma y es capaz de transmitir la modificación direccionada a una línea germinal de una generación F1; (b) introducir un vector de direccionamiento grande (LTVEC) en la célula de rata pluripotente, en donde el LTVEC comprende un inserto de ácido nucleico flanqueado con un brazo de homología 5' y un brazo de homología 3', en donde el brazo de homología 5' y el brazo de homología 3' comprende un fragmento de ADN genómico de rata; y (c) identificar una célula de rata pluripotente modificada genéticamente que comprende la modificación genética direccionada.

Se pueden usar diversos métodos para identificar células que tienen el ácido nucleico de inserción integrado en el locus diana de interés. La inserción del ácido nucleico insertado en el locus diana de interés da como resultado una "modificación del alelo". El término "modificación de alelo" y los métodos para la detección del alelo modificado se discuten con más detalle en otra parte de este documento.

En un aspecto, se proporciona un método para modificar un locus genómico de interés en una célula de rata pluripotente a través de la orientación de genes mediada por endonucleasa, comprendiendo el método: (a) proporcionar una célula de rata pluripotente aislada que es capaz de transmitir el genoma genéticamente modificado a una línea germinal de una generación F1; (b) introducir en la célula de rata pluripotente un agente endonucleasa; en donde el agente de endonucleasa hace una ruptura de muesca o de doble cadena en una secuencia de ADN diana ubicada en el locus genómico de interés, y en donde la ruptura de muesca o de doble cadena en la secuencia de DNA diana en la célula ES de rata induce: (i) reparación de ADN mediada por la unión por el extremo no homólogo (NHEJ) de la muesca o la ruptura de doble cadena, en donde la reparación de ADN mediada por NHEJ genera un alelo mutante que comprende una inserción o una eliminación de una secuencia de ácidos nucleicos en la secuencia de ADN diana; o (ii) reparación de ADN mediada por recombinación homóloga que resulta en la restauración de una secuencia de ácidos nucleicos de tipo salvaje; y (c) identificar el locus genómico de interés modificado.

En un aspecto, se proporciona método para modificar un locus genómico de interés en una célula madre embrionaria de rata aislada (ES) a través de un agente de nucleasa, que comprende: (a) proporcionar una célula ES de rata aislada que es capaz de transmitir la modificación genética direccionada a una línea germinal de una generación F1; (b)

introducción en la célula ES de rata: (i) un vector de direccionamiento grande (LTVEC) que comprende un ácido nucleico de inserto flanqueado con un brazo de homología 5' de rata y un brazo de homología 3' de rata, en donde el inserto es una secuencia de ácidos nucleicos que es de al menos 5 kb; y (ii) un agente de endonucleasa, en donde el agente de endonucleasa realiza una ruptura de muesca o de doble cadena en una secuencia de ADN diana localizada en el locus genómico de interés, y en donde la secuencia diana no está presente en el ácido nucleico de inserción; y (c) identificar la modificación genética direccionada en la célula madre embrionaria (ES) de rata.

En un aspecto, se proporciona un método para modificar un locus genómico de interés en una célula de rata pluripotente mediante ingeniería de genoma guiada por ARNcomprendiendo el método: (a) proporcionar una célula de rata pluripotente que puede transmitir el gen genéticamente modificado genoma a una línea germinal de una generación F1; (b) introducción en la célula de rata pluripotente: (i) un primer constructo de expresión que comprende un primer promotor unido operativamente a una primera secuencia de ácidos nucleicos que codifica una proteína (CAS) asociada a Repeticiones Palindrómicas Cortas Regularmente Interespaciadas Aglomeradas (CRISPR), (ii) un segundo constructo de expresión que comprende un segundo promotor unido operativamente a una secuencia diana genómica unida a un ARN guía (ARNg), en donde la secuencia diana genómica está flanqueada en el extremo 3 por una secuencia de Motivo Protoespaciador Adyacente (PAM). En una realización, la secuencia diana genómica comprende la secuencia de nucleótidos de GNNNNNNNNNNNNNNNNNNGG ($GN_{1-20}GG$; SEQ ID NO: 1). En una realización, la secuencia diana genómica comprende la SEQ ID NO: 1, en donde N tiene una longitud de entre 14 y 20 nucleótidos. En una realización, el ARNg comprende una tercera secuencia de ácidos nucleicos que codifica un ARN de Repeticiones Palindrómicas Cortas Regularmente Interespaciadas Aglomeradas (CRISPR) (ARNcr) y una cuarta secuencia de ácidos nucleicos que codifica un ARN CRISPR (ARNtracr) transactivador. En una realización, tras la expresión, la proteína Cas forma un complejo CRISPR-Cas que comprende el ARNcr y el ARNtracr, y el complejo CRISPR-Cas produce una ruptura de muesca o doble cadena en una secuencia de ADN objetivo ubicada en el locus genómico de interés, y en donde la ruptura de la muesca o la doble cadena en la secuencia de ADN diana en la célula de rata pluripotente induce: (i) la reparación de ADN mediada por la unión de extremo no homóloga (NHEJ) de la muesca o la ruptura de doble cadena creada por el complejo CRISPR-Cas, en donde el NHEJ genera un alelo mutante que comprende una inserción o una eliminación de una secuencia de ácidos nucleicos en la secuencia de ADN diana; o (ii) reparación de ADN mediada por recombinación homóloga que resulta en la restauración de una secuencia de ácidos nucleicos de tipo salvaje; y (c) identificar el locus genómico de interés modificado.

En una realización, la célula de rata pluripotente es una célula madre pluripotente de rata inducida (iPS). En una realización, la célula de rata pluripotente es una célula progenitora de desarrollo restringido.

La presencia de una muesca o una ruptura de doble cadena en el sitio de reconocimiento dentro del marcador de selección, en diversas realizaciones, aumenta la eficiencia y/o la frecuencia de recombinación entre un vector de direccionamiento (es decir, un LTVEC) y el locus diana de interés. En una realización, la recombinación es recombinación homóloga. En otra realización, la recombinación es una inserción por unión de extremo no homóloga. En diversas realizaciones, en presencia de la ruptura de la muesca o de la doble cadena, la eficiencia de la orientación de un vector de direccionamiento (es decir, un LTVEC) en el locus genómico diana es al menos aproximadamente 2 veces mayor, al menos aproximadamente 3 veces mayor, al menos aproximadamente 4 veces más que en ausencia de la ruptura de la muesca o de la doble cadena (utilizando, por ejemplo, el mismo vector de direccionamiento y los mismos brazos de homología y los correspondientes sitios de destino en el locus genómico de interés pero en ausencia de un agente de nucleasa agregado) eso hace que el muesca o la doble hebra se rompan).

En una realización, la modificación genética direccionada en el locus diana es bialélica. Por "bialélico" se entiende que ambos alelos de un gen comprenden la modificación genética direccionada. En ciertas realizaciones, el uso combinado de un vector de direccionamiento (incluyendo, por ejemplo, un LTVEC) con un agente de nucleasa da como resultado una modificación genética direccionada bialélica del locus genómico de interés en una célula en comparación con el uso del vector de direccionamiento solo. Cuando el vector de direccionamiento se usa junto con un agente de nucleasa, la eficiencia de direccionamiento bialélico aumenta al menos dos veces, al menos tres veces, al menos 4 veces o más en comparación con cuando el vector de direccionamiento se usa solo. En otras formas de realización, la eficacia de la orientación biallelica es al menos 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 2%, 3%, 4% o 5% o mayor.

Se proporcionan composiciones que comprenden una rata modificada genéticamente que tiene una modificación genética direccionada en el locus gamma del receptor de interleucina-2 o en el locus ApoE. Los diversos métodos y composiciones aquí proporcionados permiten que estos loci modificados se transmitan a través de la línea germinal.

En realizaciones específicas, una rata modificada genéticamente o una célula de rata pluripotente modificada genéticamente comprende un locus genómico que tiene una modificación genética direccionada en el locus del receptor gamma interleucina-2 o que tiene una modificación genética direccionada en el locus ApoE, en donde el receptor gamma interleucina-2 el locus genómico o el locus ApoE comprenden: (i) una eliminación de al menos una porción del locus del receptor gamma interleucina-2 o al menos una porción del locus ApoE; (ii) una inserción de una secuencia de ácidos nucleicos heteróloga en el locus ApoE o en el locus del receptor gamma interleucina-2; o (iii) una combinación de los mismos, en donde el locus genómico modificado genéticamente es capaz de ser transmitido a través de la línea germinal.

Se proporcionan adicionalmente métodos que permiten que se produzcan tales ratas modificadas genéticamente y que se produzcan tales células de rata pluripotentes modificadas genéticamente. Tales métodos incluyen un método para modificar un locus genómico de ApoE o un locus de receptor de interleucina-2 gamma en una célula de rata pluripotente mediante modificación genética direccionada. El método comprende (a) introducir en la célula de rata pluripotente un vector de direccionamiento que comprende un ácido nucleico insertado flanqueado con un brazo de homología de rata 5' al locus ApoE y un brazo de homología de rata 3' al locus ApoE, (b) que identifica una célula de rata pluripotente modificada genéticamente que comprende la modificación genética direccionada en el locus genómico de interés de ApoE, en donde la modificación genética direccionada es capaz de transmitirse a través de la línea germinal.

Los métodos adicionales incluyen (a) introducir en la célula de rata pluripotente un vector de direccionamiento que comprende un inserto de ácido nucleico flanqueado con un brazo de homología de rata 5' al locus gamma del receptor de la interleucina 2 y un brazo de homología de rata 3' al locus gamma del receptor de la interleucina 2, (b) identificación de una célula de rata pluripotente modificada genéticamente que comprende la modificación genética direccionada en el locus gamma del receptor de interleucina-2, donde la modificación genética direccionada es capaz de transmitirse a través de la línea germinal.

iii. Métodos de integración de polinucleótidos múltiples de interés en el locus direccionado

Los diversos métodos y composiciones proporcionados en el presente documento permiten la integración direccionada de múltiples polinucleótidos de interés con un locus diana dado. Los diversos métodos expuestos anteriormente se pueden repetir secuencialmente para permitir la integración direccionada de cualquier número de ácidos nucleicos de inserción en un locus específico dado. 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 insertos de ácidos nucleicos en el locus diana. En realizaciones particulares, tales métodos de apilamiento secuencial permiten la reconstrucción de grandes regiones genómicas 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 de origen agropecuario) en un locus diana. En tales casos, la transferencia y reconstrucción de 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 regiones codificantes. Variaciones en el número de copias encontradas dentro de la región genómica nativa. 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, particularmente dentro de una célula huésped procariota o y de acuerdo con la presente invención dentro de una célula de rata pluripotente o una célula ES de rata. En un ejemplo no limitante, se genera una región genómica "humanizada" dentro de una rata.

3. Un locus genómico humanizado

En el presente documento se proporcionan diversos métodos y composiciones que comprenden un locus de rata humanizado. Como se usa en el presente documento, por locus genómico "humanizado" se entiende una región de un genoma no humano que comprende al menos una secuencia de ácidos nucleicos humana. Un "locus de rata humanizado" comprende una región de ADN de rata que tiene insertada una secuencia de ADN humano. La secuencia de ADN humano puede ser una secuencia de ADN humano natural o puede modificarse desde su forma nativa. En realizaciones específicas, el ADN humano comparte al menos 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99% de identidad de secuencia con una secuencia humana nativa. Si una secuencia humana no es una secuencia humana nativa, al menos tiene mayor identidad de secuencia con una secuencia humana nativa que con una secuencia de rata ortóloga. Además, la secuencia de ADN humano puede comprender un ADNc, una región de ADN genómico humano, una región reguladora no codificante, o cualquier parte de una región de codificación, genómica o reguladora del ADN humano. La secuencia de ADN humano insertada en el locus de rata puede comprender cualquiera de los polinucleótidos de inserción como se describe en otra parte en el presente documento. En realizaciones específicas, la secuencia de ADN humano es ortóloga al locus diana de rata, mientras que, en otros casos, la secuencia de ADN humano es homóloga al locus diana de rata.

En una realización, la modificación genética direccionada es una inserción o un reemplazo de una secuencia de ácidos nucleicos de rata endógena con una secuencia de ácidos nucleicos humana homóloga u ortóloga. En una realización, la modificación genética direccionada comprende una inserción o reemplazo de una secuencia endógena de ácido nucleico de rata con una secuencia de ácidos nucleicos humana homóloga u ortóloga en un locus de rata endógeno que comprende la correspondiente secuencia de ácidos nucleicos de rata.

Los métodos para hacer un locus de rata humanizado (o una rata o célula de rata que comprende el locus de rata humanizado) comprenden la introducción en el locus diana que comprende un ácido nucleico de rata una secuencia de ácidos nucleicos humana. En una realización, se proporciona un método para hacer una rata humanizada. Dicho método comprende (a) modificar un genoma de una célula de rata pluripotente con un vector de direccionamiento que comprende un ácido nucleico de inserción que comprende una secuencia de ácidos nucleicos humana para formar una célula donante; (b) introducir la célula donante en un embrión de rata huésped; y (c) gestar el embrión de rata huésped en una madre sustituta; en donde la madre sustituta produce una progenie de rata que comprende la secuencia del ácido nucleico humano. En realizaciones específicas, el locus de rata humanizado es capaz de

transmitirse a través de la línea germinal. En una realización adicional, el vector de direccionamiento comprende un vector de direccionamiento grande (LTVEC) y el ácido nucleico de inserción que comprende una secuencia de ácidos nucleicos humana es de al menos 5 kb.

En otros métodos, el locus de rata humanizado se fabrica modificando un locus diana de un ácido nucleico de rata mediante recombinación homóloga bacteriana (BHR). El método comprende introducir en una célula procariota un vector de direccionamiento que comprende un ácido nucleico de inserción flanqueado con un brazo de homología de rata 5' y un brazo de homología de rata 3', en donde el ácido nucleico de inserción comprende una secuencia de ácidos nucleicos humana, y en donde la célula procariota comprende un ácido nucleico de rata y es capaz de expresar una recombinasa que media el BHR en el locus diana.

El locus genómico de rata humanizado puede comprender (a) una inserción de una secuencia de ácidos nucleicos humana homóloga u ortóloga; (b) un reemplazo de una secuencia endógena de ácido nucleico de rata con una secuencia de ácidos nucleicos humana homóloga u ortóloga; o (c) una combinación de los mismos. En realizaciones específicas, el locus genómico de rata humanizado es capaz de transmitirse a través de la línea germinal. En otras realizaciones más, la secuencia ortóloga humana reemplaza la secuencia correspondiente encontrada en la rata.

Se puede usar cualquier secuencia de ácidos nucleicos humana en los métodos y composiciones proporcionados en el presente documento. Ejemplos no limitantes de secuencias de ácidos nucleicos humanos que pueden usarse en los métodos y composiciones se discuten en detalle en otra parte de este documento.

La secuencia de ácidos nucleicos humana para la inserción en el locus de interés de rata puede ser de cualquier tamaño. En una realización, la secuencia de ácidos nucleicos humana puede ser de aproximadamente 500 nucleótidos a aproximadamente 200kb, de aproximadamente 500 nucleótidos a aproximadamente 5kb, de aproximadamente 5kb a aproximadamente 200kb, de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 30kb, desde aproximadamente 30kb a aproximadamente 40kb, desde aproximadamente 40kb a aproximadamente 50kb, desde aproximadamente 60kb a aproximadamente 70kb, desde aproximadamente 80kb a aproximadamente 90kb, desde aproximadamente 90kb a aproximadamente 100kb, desde aproximadamente 100kb a aproximadamente 110kb, desde aproximadamente 120kb a aproximadamente 130kb, desde aproximadamente 130kb a aproximadamente 140kb, desde aproximadamente 140kb a aproximadamente 150kb, desde aproximadamente 150kb a aproximadamente 160kb, desde aproximadamente 160kb a aproximadamente 170kb, desde aproximadamente 170kb a aproximadamente 180kb, desde aproximadamente 180kb a aproximadamente 190kb, o desde aproximadamente 190kb a unos 200kb. En una realización específica, la secuencia de ácidos nucleicos humana es de al menos 5 kb.

En una realización, se proporcionan un locus genómico de rata en donde la secuencia de ácidos nucleicos humana homóloga u ortóloga comprende (a) uno o más segmentos del gen V_H de la cadena pesada de la inmunoglobulina humana no reorganizados, uno o más segmentos del gen de la cadena pesada D de la inmunoglobulina humana no reorganizados y uno o más segmentos de inmunoglobulina humana no reorganizados del gen J_H de cadena pesada, que están unidos operativamente a una secuencia de ácidos nucleicos de región constante de cadena pesada de mamífero; (b) una secuencia de ácidos nucleicos de región variable de cadena pesada de inmunoglobulina humana reorganizada operativamente unida a una secuencia de ácidos nucleicos de región constante de cadena pesada de inmunoglobulina de mamífero; (c) uno o más segmentos del gen V_κ o V_λ de la inmunoglobulina humana no reorganizados y uno o más segmentos del gen J_κ o J_λ de la inmunoglobulina humana no reorganizados, que están operativamente enlazados a una secuencia de ácidos nucleicos de la región ligera de la cadena ligera de un mamífero, inmunoglobulina λ o κ; o, (d) una secuencia de ácidos nucleicos de la región variable de la cadena ligera λ o κ de la inmunoglobulina humana reorganizada operativamente unida a una secuencia de ácidos nucleicos de la región constante de la cadena ligera λ o κ de la cadena ligera de la inmunoglobulina de mamífero.

En otra realización, se proporciona un locus genómico de rata en donde (a) la secuencia de ácidos nucleicos de región constante de cadena pesada de inmunoglobulina de mamífero es una secuencia de ácidos nucleicos de región constante de rata, una secuencia de ácidos nucleicos de región constante humana, o una combinación de los mismos; o, (b) la secuencia de ácidos nucleicos de región constante de cadena ligera de cadena ligera de λ o κ de mamífero es una secuencia de ácidos nucleicos de región constante de rata, una secuencia de ácidos nucleicos de región constante humana, o una combinación de los mismos.

En una realización específica, se proporciona un locus genómico de rata en donde la secuencia de ácidos nucleicos de la región constante de la cadena pesada de inmunoglobulina se selecciona de o comprende un CH1, una bisagra, un CH2, un CH3 y/o una combinación de los mismos.

En una realización, el locus genómico de rata comprende uno o más segmentos funcionales del gen V_H humano que comprenden V_H1-2, V_H1-3, V_H1-8, V_H1-18, V_H1-24, V_H1-45, V_H1-46, V_H1-58, V_H1-69, V_H2-5, V_H2-26, V_H2-70, V_H3-7, V_H3-9, V_H3-11, V_H3-13, V_H3-15, V_H3-16, V_H3-20, V_H3-21, V_H3-23, V_H3-30, V_H3-30-3, V_H3-30-5, V_H3-33, V_H3-35, V_H3-38, V_H3-43, V_H3-48, V_H3-49, V_H3-53, V_H3-64, V_H3-66, V_H3-72, V_H3-73, V_H3-74, V_H4-4, V_H4-28, V_H4-30-1, V_H4-30-2, V_H4-30-4, V_H4-31, V_H4-34, V_H4-39, V_H4-59, V_H4-61, V_H5-51, V_H6-1, V_H7-4-1, V_H7-81, o una combinación de los mismos.

En una realización, el locus genómico de rata comprende uno o más segmentos funcionales del gen D humano que comprenden D1-1, D1-7, D1-14, D1-20, D1-26, D2-2, D2-8, D2-15, D2-21, D3-3, D3-9, D3-10, D3-16, D3-22, D4-4, D4-11, D4-17, D4-23, D5-12, D5-5, D5-18, D5-24, D6-6, D6-13, D6-19, D6-25, D7-27, o una combinación de los mismos.

En una realización, el locus genómico de rata comprende uno o más segmentos funcionales del gen J_H que comprenden J_H1, J_H2, J_H3, J_H4, J_H5, J_H6, y/o una combinación de los mismos. En una realización, el ácido nucleico insertado comprende uno o más segmentos génicos V_K humanos que comprenden Vκ4-1, Vκ5-2, Vκ7-3, Vκ2-4, Vκ1-5, Vκ1-6, Vκ3-7, Vκ1-8, Vκ1-9, Vκ2-10, Vκ3-11, Vκ1-12, Vκ1-13, Vκ2-14, Vκ3-15, Vκ1-16, Vκ1-17, Vκ2-18, Vκ2-19, Vκ3-20, Vκ6-21, Vκ1-22, Vκ1-23, Vκ2-24, Vκ3-25, Vκ2-26, Vκ1-27, Vκ2-28, Vκ2-29, Vκ2-30, Vκ3-31, Vκ1-32, Vκ1-33, Vκ3-34, Vκ1-35, Vκ2-36, Vκ1-37, Vκ2-38, Vκ1-39, Vκ2-40, o una combinación de los mismos.

En una realización, el locus genómico de rata comprende uno o más segmentos del gen Vλ humano que comprenden Vλ3-1, Vλ4-3, Vλ2-8, Vλ3-9, Vλ3-10, Vλ2-11, Vλ3-12, Vλ2-14, Vλ3 -16, Vλ2-18, Vλ3-19, Vλ3-21, Vλ3-22, Vλ2-23, Vλ3-25, Vλ3-27, o una combinación de los mismos.

En una realización, el locus genómico de rata comprende uno o más segmentos de genes J_K humanos que comprenden Jκ1, Jκ2, Jκ3, Jκ4, Jκ5, o una combinación de los mismos.

En aún otra realización, el locus genómico de rata comprende un locus genómico humanizado que comprende una secuencia de ácidos nucleicos del receptor de interleucina-2 humana (IL2R) o una variante o un fragmento de la misma. En realizaciones específicas, la secuencia de ácidos nucleicos de IL2R comprende un receptor alfa de interleucina-2, un receptor beta de interleucina-2, o una secuencia de ácidos nucleicos gamma del receptor de interleucina-2 o variantes o fragmentos de la misma.

En realizaciones adicionales, un locus genómico de rata comprende un locus genómico humanizado que comprende una porción del locus ApoE humano, el locus gamma del receptor de Interleucina-2 humano, el locus Rag2 humano, el locus Rag1 humano y/o el Rag2 humano El locus/Rag1 reemplaza la porción homóloga u ortóloga correspondiente del locus ApoE de rata, el locus gamma del receptor de Interleucina-2 de rata, el locus Rag2 de rata, el locus Rag1 de rata y/o el locus Rag2/Rag1 de rata. En una realización, el ectodominio de rata de IL-2Rg se reemplaza con el ectodominio de IL-2Rg humano, con el resto de la molécula que es de la rata.

En otra realización, se proporciona una rata modificada genéticamente que comprende un locus genómico humanizado. Dichas ratas modificadas genéticamente comprenden (a) una inserción de una secuencia de ácidos nucleicos humana homóloga u ortóloga; (b) un reemplazo de la secuencia de ácidos nucleicos de rata con una secuencia de ácidos nucleicos humana homóloga u ortóloga en un locus genómico endógeno; o (c) una combinación de los mismos, en donde el locus genómico humanizado es capaz de ser transmitido a través de la línea germinal.

También se proporcionan ratas modificadas genéticamente que comprenden cualquiera de los diversos loci genómicos humanizados divulgados en el presente documento y descritos anteriormente.

4. Polinucleótidos de interés.

Cualquier polinucleótido de interés puede estar contenido en los diversos ácidos nucleicos del inserto y, por lo tanto, integrarse en el locus diana. Los métodos divulgados en el presente documento proporcionan al menos 1, 2, 3, 4, 5, 6 o más polinucleótidos de interés para integrarse en el locus genómico diana.

El polinucleótido de interés dentro del ácido nucleico de inserción cuando se integra en el locus genómico diana puede introducir una o más modificaciones genéticas en la célula. La modificación genética puede comprender una eliminación de una secuencia de ácidos nucleicos endógena y/o la adición de un polinucleótido exógeno o heterólogo u ortólogo en el locus genómico diana. En una realización, la modificación genética comprende un reemplazo de una secuencia de ácidos nucleicos endógena con un polinucleótido exógeno de interés en el locus genómico diana. Por lo tanto, los métodos proporcionados en el presente documento permiten la generación de una modificación genética que comprende un bloqueo, 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. Dichas modificaciones pueden ocurrir tras la integración del primero, segundo, tercero, cuarto, quinto, sexto, séptimo o cualquier ácido nucleico de inserción posterior en el locus genómico diana.

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus diana puede comprender una secuencia que es nativa de la célula en donde 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 donde se introduce; el polinucleótido de interés puede ser ortólogo a la célula en donde se introduce; o el polinucleótido de interés puede ser de una especie diferente a la célula en donde se introduce. Como se usa en el presente documento, "nativo" en referencia a una secuencia insertada en el locus diana es una secuencia que es nativa de la célula que tiene el locus diana o nativa de la célula de la que se derivó el locus diana (es decir, de una rata). Como se usa en el presente documento, "heterólogo" en referencia a una secuencia incluye una secuencia que se origina a partir de una especie extraña, o, si es de la misma especie, es sustancialmente diferente o está modificada de su forma nativa en

composición y/o locus genómico por intervención humana deliberada. Como se usa en el presente documento, "exógeno" en referencia a una secuencia es una secuencia que se origina a partir de una especie extraña. 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 mamífero de origen agropecuario 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 1º, 2º, 3º, 4º, 5º, 6º, 7º y/o cualquiera de los ácidos nucleicos del inserto posterior pueden comprender tales secuencias.

En una realización, el polinucleótido de interés dentro del ácido nucleico del inserto y/o integrado en el locus diana es nativo de una secuencia de ácidos nucleicos de ratón, un ácido nucleico humano, un ácido nucleico no humano, un ácido nucleico de roedor, un ácido nucleico de rata, un ácido nucleico de hámster, un ácido nucleico de mono, un ácido nucleico de mamífero de origen agropecuario o un ácido nucleico de mamífero no agrícola. En otras realizaciones adicionales, el polinucleótido de interés integrado en el locus diana es un fragmento de un ácido nucleico genómico. En una realización, el ácido nucleico genómico es un ácido nucleico genómico de ratón, un ácido nucleico genómico humano, un ácido nucleico no humano, un ácido nucleico de roedor, un ácido nucleico de rata, un ácido nucleico de hámster, un ácido nucleico de mono, un ácido nucleico de mono. ácido nucleico de mamífero o un ácido nucleico de mamífero no agrícola o una combinación de los mismos.

En una realización, el polinucleótido de interés puede variar desde aproximadamente 500 nucleótidos hasta aproximadamente 200kb como se describe anteriormente. El polinucleótido de interés puede ser de aproximadamente 500 nucleótidos a aproximadamente 5kb, de aproximadamente 5kb a aproximadamente 200kb, de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 30kb, de aproximadamente 30kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 50kb, desde aproximadamente 60kb a aproximadamente 70kb, desde aproximadamente 80kb a aproximadamente 90kb, desde aproximadamente 90kb a aproximadamente 100kb, desde aproximadamente 100kb a aproximadamente 110kb, desde aproximadamente 120kb a aproximadamente 130kb, desde aproximadamente 130kb a aproximadamente 140kb, desde aproximadamente 140kb a aproximadamente 150kb, desde aproximadamente 150kb a aproximadamente 160kb, desde aproximadamente 160kb a aproximadamente 170kb, desde aproximadamente 170kb a aproximadamente 180kb, desde aproximadamente 180kb a aproximadamente 190kb, o desde aproximadamente 190kb a aproximadamente 200kb, desde aproximadamente 5kb a aproximadamente 10kb, desde aproximadamente 10kb a aproximadamente 20kb, desde aproximadamente 20kb a aproximadamente 40kb, desde aproximadamente 40kb a aproximadamente 60kb, desde aproximadamente 60kb a aproximadamente 80kb, desde aproximadamente 80kb a aproximadamente 100kb, desde aproximadamente 100kb a aproximadamente 150kb, desde aproximadamente 150kb a aproximadamente 200kb, desde aproximadamente 200kb a aproximadamente 250kb, desde aproximadamente 250kb a aproximadamente 300kb, desde aproximadamente 300kb a aproximadamente 350kb, o de aproximadamente 350kb a aproximadamente 400kb.

El polinucleótido de interés dentro del inserto de ácido nucleico y/o insertado en el locus genómico diana puede codificar un polipéptido, puede codificar un ARNmi, o puede comprender cualquier región reguladora o no codificante de interés que incluya, por ejemplo, una secuencia reguladora, una secuencia promotora, una secuencia potenciadora, una secuencia de unión a represor transcripcional, o una eliminación de una secuencia no codificadora de proteínas, pero no comprende una eliminación de una secuencia codificadora de proteínas. Además, el polinucleótido de interés dentro del inserto de ácido nucleico y/o insertado en el locus genómico diana 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 los mismos. En una realización, el polinucleótido de interés dentro del inserto de ácido nucleico y/o insertado en el locus genómico diana codifica una proteína expresada en una médula ósea o una célula derivada de médula ósea. En una realización, el polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus diana codifica una proteína expresada en una célula de bazo. En otras realizaciones adicionales, el polinucleótido de interés dentro del inserto de ácido nucleico y/o insertado en el locus diana codifica una proteína expresada en una célula B, codifica una proteína expresada en una célula B inmadura o codifica una proteína expresada en una célula B madura.

El polinucleótido de interés dentro del polinucleótido de inserción puede comprender una porción de un locus ApoE, un locus IL-2-Rg, un locus Rag1, un locus Rag2 y/o un locus Rag2/Rag1. Tales porciones de estos loci dados se discuten en otra parte en el presente documento, al igual que las diversas regiones homólogas y ortólogas de cualquier organismo de interés que pueda emplearse.

En una realización, el polinucleótido de interés dentro del inserto de ácido nucleico y/o insertado en el locus diana comprende una secuencia de ácidos nucleicos genómica que codifica una secuencia de aminoácidos de región variable de cadena pesada de inmunoglobulina. La expresión "cadena pesada" o "cadena pesada de inmunoglobulina" se describe en otra parte de este documento.

En una realización, el polinucleótido de interés dentro del ácido nucleico de inserción y/o integrado en el locus diana comprende una secuencia de ácidos nucleicos genómica que codifica una secuencia de aminoácidos de región variable de cadena pesada de inmunoglobulina humana.

En una realización, la secuencia de ácidos nucleicos genómica comprende uno o más segmentos génicos V_H de cadena pesada de inmunoglobulina humana no reorganizados, uno o más segmentos génicos D de cadena pesada D de inmunoglobulina humana no reorganizados, y uno o más segmentos génicos J_H de cadena pesada de inmunoglobulina humana no reorganizados, que están unidos operativamente a una secuencia de ácidos nucleicos de región constante de cadena pesada de mamífero. En una realización, la secuencia de ácidos nucleicos genómica comprende una secuencia de ácidos nucleicos de región variable de cadena pesada de inmunoglobulina humana reorganizada operativamente unida a una secuencia de ácidos nucleicos de región constante de cadena pesada de mamífero. En una realización, la secuencia de ácidos nucleicos genómica comprende una o más inmunoglobulinas humanas Y_κ o V_λ no reorganizadas, segmentos génicos y una o más inmunoglobulina humana J_κ o J_λ no reorganizadas, segmentos génicos, que están operativamente enlazados a una inmunoglobulina de mamífero λ o κ cadena ligera cadena de la región constante secuencia de ácido nucleico. En una realización, la secuencia de ácidos nucleicos genómica comprende una secuencia de ácidos nucleicos de región variable de la cadena ligera λ o κ reorganizada operativamente unida a una secuencia de ácidos nucleicos de región constante de la cadena ligera de la λ o κ de la cadena ligera de κ. En una realización, la secuencia de ácidos nucleicos de la región constante de la cadena pesada comprende una secuencia de ácidos nucleicos de la región constante de rata, una secuencia de ácidos nucleicos de la región constante humana, o una combinación de las mismas. En una realización, el ácido nucleico de la región constante de la cadena ligera de λ o κ comprende una secuencia de ácidos nucleicos de la región constante de rata, una secuencia de ácidos nucleicos de la región constante humana, o una combinación de los mismos.

En una realización, la secuencia de ácidos nucleicos de la región constante de la cadena pesada de inmunoglobulina se selecciona de o comprende un CH1, una bisagra, un CH2, un CH3 y/o una combinación de los mismos. En una realización, la secuencia de ácidos nucleicos de la región constante de la cadena pesada comprende una CH1-bisagra-CH2-CH3.

En una realización, el polinucleótido de interés dentro del ácido nucleico de inserción y/o integrado en el locus diana comprende una secuencia de ácidos nucleicos genómica que codifica una secuencia de aminoácidos de región variable de cadena ligera de inmunoglobulina. La expresión "cadena ligera" incluye una secuencia de cadena ligera de inmunoglobulina de cualquier organismo, y se describe en otra parte en el presente documento.

En una realización, el polinucleótido de interés dentro del ácido nucleico de inserción y/o integrado en el locus genómico diana comprende una secuencia de ácidos nucleicos genómica que codifica una secuencia de aminoácidos de región variable de cadena ligera de inmunoglobulina humana.

En una realización, la secuencia de ácidos nucleicos genómica comprende una o más segmentos génicos inmunoglobulina humana V_κ o V_λ no reordenados y uno o más segmentos génicos de inmunoglobulina humana no reorganizados J_κ o J_λ, , que están operativamente enlazados a una secuencia de ácidos nucleicos de la región constante de la cadena ligera de la cadena ligera y la cadena ligera λ o κ de inmunoglobulina de roedor. En una realización, la secuencia de ácidos nucleicos genómica comprende una secuencia de ácidos nucleicos de región variable de cadena ligera λ o κ reorganizada operativamente unida operativamente a una secuencia de ácidos nucleicos de región constante de cadena ligera λ o κ de cadena ligera de roedor. En una realización, la secuencia de ácidos nucleicos de región constante de cadena ligera comprende una secuencia de ácidos nucleicos de región constante de rata, una secuencia de ácidos nucleicos de región constante humana, o una combinación de los mismos. En una realización, el ácido nucleico de la región constante de la cadena ligera de λ o κ comprende una secuencia de ácidos nucleicos de la región constante de rata, una secuencia de ácidos nucleicos de la región constante humana, o una combinación de los mismos.

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus diana puede codificar una proteína extracelular o un ligando para un receptor. En realizaciones específicas, el ligando codificado es una citoquina. Las citoquinas de interés incluyen una quimioquina seleccionada de o que comprende CCL, CXCL, CX3CL y/o XCL. La citoquina también puede comprender un factor de necrosis tumoral (TNF). En otras realizaciones más, la citoquina es una interleucina (IL). En una realización, la interleucina se selecciona de o comprende IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10 , IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19, IL-20, IL-21, IL-22, IL -23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32, IL-33, IL-34, IL-35, y/o IL-36. En una realización, la interleucina es IL-2. En realizaciones específicas, tales polinucleótidos de interés dentro del inserto de ácido nucleico y/o integrados en el locus genómico diana son de un humano y, en realizaciones más específicas, pueden comprender la secuencia genómica humana.

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico diana puede codificar la apolipoproteína E (ApoE).

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus diana puede codificar una proteína citoplásmica o una proteína de membrana. En una realización, la proteína de membrana es un receptor, tal como un receptor de citoquinas, un receptor de interleucina, un receptor alfa de interleucina 2, un receptor beta de interleucina 2, un receptor gamma de interleucina 2 o un receptor tirosina quinasa. En otros casos, el polinucleótido de interés dentro del ácido nucleico insertado y/o integrado en el locus diana puede comprender una región ortóloga u homóloga del locus diana.

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus diana puede comprender un polinucleótido que codifica al menos una región de un receptor de células T, incluido el receptor de células T alfa. En métodos específicos, cada uno de los ácidos nucleicos del inserto comprende una región genómica del locus del receptor de las células T (es decir, el locus alfa del receptor de las células T) de manera que, al completarse la integración en serie, una porción o la totalidad del locus del receptor de las células T genómicas tiene sido integrado en el locus diana. Dichos ácidos nucleicos de inserción pueden comprender al menos uno o más de un segmento variable o un segmento de unión de un locus receptor de células T (es decir, del locus alfa del receptor de células T). En otras realizaciones adicionales, el polinucleótido de interés que codifica la región del receptor de células T puede ser, por ejemplo, de un mamífero, un mamífero no humano, un roedor, un ratón, una rata, un ser humano, un mono, un mamífero de origen agropecuario o un polinucleótido de mamífero doméstico que codifica una proteína mutante.

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

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico diana puede comprender una modificación genética en una secuencia codificadora. Dichas modificaciones genéticas incluyen, pero no se limitan a, una mutación por eliminación de una secuencia codificadora o la fusión de dos secuencias de codificación.

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus diana puede comprender un polinucleótido que codifica una proteína mutante, que incluye, por ejemplo, una proteína mutante humana. 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 inserto de ácido nucleico y/o integrado en el locus diana comprende al menos un alelo de enfermedad, que incluye, por ejemplo, un alelo de una enfermedad neurológica, un alelo de una enfermedad cardiovascular, un alelo de una enfermedad renal, un alelo de una enfermedad muscular, un alelo de una enfermedad de la sangre, un alelo de un gen causante de cáncer o un alelo de una enfermedad del sistema inmunológico. 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 de origen agropecuario o un polinucleótido de mamíferos domésticos que codifica una proteína mutante.

En una realización, la modificación genética produce una forma mutante de una proteína con una característica de unión alterada, localización alterada, expresión alterada y/o patrón de expresión alterado.

En una realización, la modificación genética produce una eliminación, adición, reemplazo o una combinación de las mismas de una región del locus ApoE de rata, en donde la modificación genética en el locus ApoE da como resultado una disminución en la actividad de ApoE, en una realización, se genera un ApoE bloqueado.

En una realización, la modificación genética produce una eliminación, adición, reemplazo o una combinación de los mismos de una región del locus Rag1 de rata, en donde la modificación genética en el locus Rag1 da como resultado una disminución en la actividad de Rag1. En una realización, se genera un Rag1 bloqueado. En una realización, la modificación genética produce una eliminación, adición, reemplazo o una combinación de los mismos de una región del locus Rag2 de rata, en donde la modificación genética en el locus Rag2 da como resultado una disminución en la actividad de Rag2. En una realización, se genera un Rag2 bloqueado. En una realización, la modificación genética produce una eliminación, adición, reemplazo o una combinación de las mismas de una región del locus Rag1/Rag2 de rata, en donde la modificación genética en el locus Rag1/Rag2 da como resultado una disminución en la actividad Rag1 y una disminución en Actividad de Rag2. En una realización, se genera un Rag1/Rag2 bloqueado.

En una realización, la modificación genética produce una eliminación, adición, reemplazo o una combinación de las mismas de una región del locus gamma del receptor de interleucina-2 de rata, en donde la modificación genética en el locus gamma del receptor de interleucina-2 da como resultado una disminución en el receptor de interleucina-2 gamma. En una realización, se genera una desactivación gamma del receptor de interleucina-2.

Como se discutió en otra parte en el presente documento, las realizaciones adicionales proporcionadas en el presente documento comprenden uno o más del locus ApoE de rata, el locus gamma del receptor de Interleucina-2 de rata, el locus Rag2, el locus Rag1 y/o el locus Rag2/Rag1 se modifican a través de el reemplazo de una porción del locus ApoE de rata, el locus gamma del receptor de interleucina-2, el locus Rag2, el locus Rag1 y/o el locus Rag2/Rag1 con la porción ortóloga correspondiente de un locus ApoE, un locus gamma del receptor de interleucina-2, un locus Rag2, un locus Rag1 y/o un locus Rag2/Rag1 de otro organismo.

En una realización, se generan múltiples modificaciones genéticas. En una realización, una modificación genética produce una eliminación, adición, reemplazo o una combinación de los mismos de una región del locus gamma del receptor de Interleucina-2 de rata, en donde la modificación genética en el locus gamma del receptor de Interleucina-

2 da como resultado una disminución en el receptor gamma de la interleucina-2 y una segunda modificación genética producen una eliminación, adición, reemplazo o una combinación de los mismos de una región del locus Rag2 de rata, en donde la modificación genética en el locus Rag2 da como resultado una disminución en la actividad Rag2. En una realización, se genera un bloqueo de receptor gamma de interleucina-2/Rag2. Tal rata tiene un fenotipo SCID.

En una realización, el ácido nucleico de los mamíferos comprende un locus genómico que codifica una proteína expresada en el sistema nervioso, el sistema esquelético, el sistema digestivo, el sistema circulatorio, el sistema muscular, el sistema respiratorio, el sistema cardiovascular, sistema linfático, el sistema endocrino, el sistema urinario, el sistema reproductivo, o una combinación de los mismos. En una realización, el ácido nucleico de mamífero comprende un locus genómico que codifica una proteína expresada en una médula ósea o una célula derivada de médula ósea. En una realización, el ácido nucleico comprende un locus genómico que codifica una proteína expresada en una célula de bazo. En una realización, el locus genómico comprende una secuencia de ADN genómico de ratón, una secuencia de ADN genómico de rata y una secuencia de ADN genómico humano, o una combinación de las mismas. En una realización, el locus genómico comprende, en cualquier orden, secuencias de ADN genómico de rata y humano. En una realización, el locus genómico comprende, en cualquier orden, secuencias de ADN genómico de ratón y humano. En una realización, el locus genómico comprende, en cualquier orden, secuencias de ADN genómico de ratón y rata. En una realización, el locus genómico comprende, en cualquier orden, secuencias de ADN genómico de rata, ratón y humano.

En una realización, el ácido nucleico inserto comprende una modificación genética en una secuencia codificadora de un gen. En una realización, la modificación genética comprende una mutación por eliminación en la secuencia codificadora. En una realización, la modificación genética comprende una fusión de dos secuencias codificantes endógenas.

En una realización, la modificación genética comprende una eliminación de una secuencia no codificadora de proteínas, pero no comprende una eliminación de una secuencia codificadora de proteínas. En una realización, la eliminación de la secuencia no codificadora de proteínas comprende una eliminación de un elemento regulador. En una realización, la modificación genética comprende una adición de un promotor. En una realización, la modificación genética comprende una sustitución de un promotor o elemento regulador. En una realización, el elemento regulador es un potenciador. En una realización, el elemento regulador es un elemento de unión a represor transcripcional.

En una realización, la modificación genética comprende la colocación de una secuencia de ácidos nucleicos humana que codifica una proteína humana mutante. En una realización, la modificación genética comprende al menos un alelo de enfermedad humana de un gen humano. En una realización, la enfermedad humana es una enfermedad neurológica. En una realización, la enfermedad humana es una enfermedad cardiovascular. En una realización, la enfermedad humana es una enfermedad renal. En una realización, la enfermedad humana es una enfermedad muscular. En una realización, la enfermedad humana es una enfermedad de la sangre. En una realización, la enfermedad humana es un cáncer. En una realización, la enfermedad humana es una enfermedad del sistema inmune. En una realización, el alelo de la enfermedad humana es un alelo dominante. En una realización, el alelo de la enfermedad humana es un alelo recesivo. En una realización, el alelo de la enfermedad humana comprende un alelo de polimorfismo de nucleótido único (SNP).

El polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus diana también puede comprender una secuencia reguladora, que incluye, por ejemplo, una secuencia promotora, una secuencia potenciadora o una secuencia de unión a represor transcripcional. En realizaciones específicas, el polinucleótido de interés dentro del inserto de ácido nucleico y/o integrado en el locus genómico diana comprende un polinucleótido que tiene una eliminación de una secuencia no codificadora de proteínas, pero no comprende una eliminación de una secuencia codificadora de proteínas. En una realización, la eliminación de la secuencia no codificadora de proteínas 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 de origen agropecuario o un polinucleótido de mamífero doméstico que codifique una proteína mutante.

5. Métodos de introducción de secuencias y generación de ratas transgénicas

Como se describió anteriormente, se proporcionan métodos y composiciones en el presente documento para permitir la integración direccionada de uno o más polinucleótidos de interés en un locus diana. Dichos sistemas emplean una variedad de componentes y, para facilitar la referencia, en el presente documento el término "sistema de integración direccionado" comprende genéticamente todos los componentes necesarios para un evento de integración (es decir, en ejemplos no limitantes, los diversos agentes de nucleasas, sitios de reconocimiento, polinucleótidos de ADN de inserción, vectores de direccionamiento, locus genómico diana y/o polinucleótidos de interés).

Los métodos proporcionados en el presente 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 direccionada. "Introducción" significa presentar a la célula la secuencia (polipéptido o polinucleótido) de tal

manera que la secuencia obtenga acceso al interior de la célula. Los métodos proporcionados en el presente documento no dependen de un método particular para introducir cualquier componente del sistema de integración genómica diana 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 de forma estable en su genoma. "Incorporado de forma estable" o "introducido de 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 diana.

Los protocolos de transfección y los protocolos para introducir polipéptidos o secuencias de polinucleótidos en las células pueden variar. Métodos de transfección no limitantes incluyen métodos de transfección basados en 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, and Kriegler, M (1991). Transfer and Expression: A Laboratory Manual. New York: W.H. Freeman and Company. pp. 96-97); dendrímeros; o polímeros catiónicos tales como DEAE-dextrano o polietilenimina. Los métodos no químicos incluyen la electroporación; sonoporación; y transfección óptica. La transfección basada 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 la transfección.

En una realización, la introducción de uno o más de los polinucleótidos en una célula está mediada por electroporación, por inyección intracitoplásmica, por una infección viral, por adenovirus, por lentivirus, por retrovirus, por transfección, por transfección mediada por lípidos o es mediada a través de Nucleofection™.

En una realización, la introducción de uno o más de los polinucleótidos en una célula comprende, además: la introducción de un constructo de expresión que comprende una secuencia de ácidos nucleicos de interés operativamente unida a un promotor. En una realización, el promotor es un promotor constitutivamente activo. En una realización, el promotor es un promotor inducible. En una realización, el promotor es activo en la célula madre embrionaria de rata.

En una realización, el constructo de expresión se introduce junto con el LTVEC. En una realización, el constructo de expresión se introduce por separado del LTVEC durante un período de tiempo.

En una realización, la introducción de uno o más polinucleótidos en la célula puede realizarse múltiples veces durante un período de tiempo. En una realización, la introducción de uno o más polinucleótidos en la célula se realiza al menos dos veces durante un período de tiempo, al menos tres veces durante un período de tiempo, al menos cuatro veces durante un período de tiempo, al menos cinco veces durante un período de tiempo, al menos seis veces durante un período de tiempo, al menos siete veces durante un período de tiempo, al menos ocho veces durante un período de tiempo, al menos nueve veces durante un período de tiempo, al menos diez veces durante un período de tiempo, al menos once veces, al menos doce veces durante un período de tiempo, al menos trece veces durante un período de tiempo, al menos catorce veces durante un período de tiempo, al menos quince veces durante un período de tiempo, al menos dieciséis veces durante un período de tiempo, al menos diecisiete veces durante un período de tiempo, al menos dieciocho veces durante un período de tiempo, al menos diecinueve veces durante un período de tiempo, o al menos veinte veces durante un período de tiempo.

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

En una realización, la etapa de selección comprende un ensayo cuantitativo para evaluar la modificación del alelo (MOA) de un cromosoma parental. En una realización, el ensayo cuantitativo se lleva a cabo a través de una PCR cuantitativa. En una realización, la PCR cuantitativa es una PCR en tiempo real (qPCR). En una realización, la PCR en tiempo real comprende un primer conjunto de cebadores que reconoce el locus diana y un segundo conjunto de cebadores que reconoce un locus de referencia no direccionado. En una realización, el conjunto de cebadores comprende una sonda fluorescente que reconoce la secuencia amplificada. En una realización, el ensayo cuantitativo se lleva a cabo mediante hibridación in situ mediada por fluorescencia (FISH). En una realización, el ensayo cuantitativo se lleva a cabo mediante hibridación genómica comparativa. En una realización, el ensayo cuantitativo se lleva a cabo a través de amplificación de ADN isotérmica. En una realización, el ensayo cuantitativo se lleva a cabo a través de amplificación de ADN isotérmica. En una realización, el ensayo cuantitativo se lleva a cabo mediante hibridación cuantitativa a una(s) sonda(s) inmovilizada(s). En una realización, el ensayo cuantitativo se lleva a cabo a través de Invader Probes®. En una realización, el ensayo cuantitativo se lleva a cabo mediante ensayos MMP®. En una realización, el ensayo cuantitativo se lleva a cabo a través de TaqMan® Molecular Beacon. En una realización, el

ensayo cuantitativo se lleva a cabo mediante la tecnología de sonda Eclipse™. (Véase, por ejemplo, US2005/0144655).

También se proporciona un método para hacer una rata humanizada, que comprende: (a) modificar un genoma de una célula de rata pluripotente con un vector de direccionamiento que comprende un ácido nucleico de inserción que comprende una secuencia de ácidos nucleicos humana para formar una célula donante; (b) introducir la célula donante en un embrión de rata huésped; y (c) gestar el embrión de rata huésped en una madre sustituta; en donde la madre sustituta produce una progenie de rata que comprende la secuencia del ácido nucleico humano. En una realización, la célula donante se introduce en un embrión de rata huésped que está en la etapa de blastocisto o en una etapa de premórula (es decir, una etapa de 4 células o una etapa de 8 células). Además, la etapa (a) también se puede realizar con un vector de direccionamiento grande (LTVEC) y/o una secuencia de ácidos nucleicos humana de al menos 5Kb de longitud. En otras realizaciones adicionales, la modificación genética es capaz de transmitirse a través de la línea germinal.

Se pueden generar ratas modificadas genéticamente empleando los diversos métodos divulgados en el presente documento. Tales métodos comprenden (1) integrar uno o más polinucleótidos de interés en el locus diana de una célula de rata pluripotente para generar una célula de rata pluripotente modificada genéticamente que comprende el ácido nucleico de inserción en el locus genómico diana que emplea los métodos divulgados en el presente documento; (2) seleccionar la célula de rata pluripotente modificada genéticamente que tiene uno o más polinucleótidos de interés en el locus genómico diana; (3) introducción de la célula de rata pluripotente modificada genéticamente en un embrión huésped de rata; y (4) implantar el embrión de rata huésped que comprende la célula de rata pluripotente modificada genéticamente en una madre sustituta. Se genera una progenie de la célula de rata pluripotente genéticamente modificada. En una realización, la célula donante se introduce en un embrión huésped de rata en la etapa de blastocisto o en la etapa de premórula (es decir, la etapa de 4 células o la etapa de 8 células). La progenie que es capaz de transmitir la modificación genética a través de la línea germinal se genera. La célula de rata pluripotente puede ser una célula ES de rata como se describe en otra parte en el presente documento.

Las técnicas de transferencia nuclear también se pueden usar para generar ratas modificadas genéticamente. En resumen, los métodos para la transferencia nuclear incluyen los pasos de: (1) enuclear 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 CC 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.

Se proporciona un método para hacer una rata modificada genéticamente, que comprende la modificación de un locus genómico de interés en una célula de rata pluripotente empleando la orientación de genes mediada por endonucleasa para introducir una modificación en un locus genómico de rata de interés para formar una célula de rata pluripotente modificada, manteniendo el célula de rata pluripotente modificada en condiciones suficientes para mantener la pluripotencia, empleando la célula de rata pluripotente modificada como célula donante en un embrión huésped de rata, y gestando el embrión huésped que comprende la célula de rata pluripotente modificada en una madre sustituta, en donde el embrión huésped es gestado por la madre sustituta y nace una progenie de rata modificada genéticamente.

En un ejemplo, la secuencia diana se ubica en un intrón. En un ejemplo, la secuencia diana se encuentra en un exón. En un ejemplo, la secuencia diana se encuentra en un promotor. En un ejemplo, la secuencia diana está ubicada en una región reguladora promotora. En un ejemplo, la secuencia diana se encuentra en una región potenciadora.

En un ejemplo, el paso de introducción se realiza varias veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana. En un ejemplo, la etapa se realiza al menos dos veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos tres veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos cuatro veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos cinco veces durante un período de tiempo utilizando una

pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos seis veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos siete veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos ocho veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos nueve veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos diez veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos once veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos doce veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos trece veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos catorce veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos quince veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos dieciséis veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos diecisiete veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos dieciocho veces durante un período o tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, al menos diecinueve veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana, o al menos veinte veces durante un período de tiempo utilizando una pluralidad de endonucleasas que reconocen distintas secuencias diana.

En un ejemplo, el paso de introducción está mediado por electroporación, por inyección intracitoplásmica, por un adenovirus, por lentivirus, por retrovirus, por transfección, por transfección mediada por lípidos o está mediado por Nucleofection™.

En una realización, el método comprende además la introducción de un ácido nucleico exógeno en la célula de rata pluripotente modificada genéticamente. En un ejemplo, el ácido nucleico exógeno es un transgén. En un ejemplo, el ácido nucleico exógeno se introduce en un locus endógeno. En un ejemplo, el ácido nucleico exógeno se introduce ectópicamente (por ejemplo, en un locus diferente de su locus endógeno).

Se proporciona un método para hacer una rata modificada genéticamente, que comprende modificar un locus genómico de interés en una célula de rata pluripotente empleando ingeniería de genoma guiada por ARN para introducir una modificación en un locus genómico de interés de rata para formar una célula de rata pluripotente modificada, manteniendo la célula de rata pluripotente modificada en condiciones suficientes para mantener la pluripotencia, empleando la célula de rata pluripotente modificada como célula donante en un embrión huésped de rata, y gestando el embrión huésped que comprende la célula de rata pluripotente modificada en una madre sustituta, en donde el embrión huésped es gestado por la madre sustituta y nace una progenie de rata modificada genéticamente.

En un ejemplo, el método tiene una tasa de direccionamiento que oscila entre aproximadamente el 2% y aproximadamente el 80%.

En un ejemplo, el método comprende cointroducir una pluralidad de el segundo constructo de expresión que comprende secuencias diana genómicas distintas para la edición múltiple de loci genómicos distintos. En un ejemplo, el método comprende la introducción de una pluralidad de el segundo constructo de expresión que comprende secuencias diana genómicas distintas para la edición múltiple de loci genómicos distintos durante un período de tiempo.

En un ejemplo, la introducción de pasos se realiza varias veces durante un período de tiempo. En un ejemplo, la introducción de la etapa (b) se realiza al menos dos veces durante un período de tiempo, al menos tres veces durante un período de tiempo, al menos cuatro veces durante un período de tiempo, al menos cinco veces durante un período de tiempo , al menos seis veces durante un período de tiempo, al menos siete veces durante un período de tiempo, al menos ocho veces durante un período de tiempo, al menos nueve veces durante un período de tiempo, al menos diez veces durante un período de tiempo, al menos once veces durante un período de tiempo, al menos doce veces durante un período de tiempo, al menos trece veces durante un período de tiempo, al menos catorce veces durante un período de tiempo, al menos quince veces durante un período de tiempo , al menos dieciséis veces durante un período de tiempo, al menos diecisiete veces durante un período de tiempo, al menos dieciocho veces durante un período de tiempo, al menos diecinueve veces durante un período de tiempo, al menos veinte veces durante un período de tiempo.

En un ejemplo, el primer constructo de expresión y el segundo constructo de expresión se expresan a partir de un mismo plásmido.

En un ejemplo, la etapa de introducción está mediada por electroporación, por inyección intracitoplasmática, por un adenovirus, por lentivirus, por retrovirus, por transfección, por transfección mediada por lípidos o está mediada por Nucleofection™.

En un ejemplo, el método comprende además la introducción de un ácido nucleico exógeno en la célula de rata pluripotente que comprende el alelo mutante.

En un ejemplo, el ácido nucleico exógeno es un transgén. En una realización, el ácido nucleico exógeno se introduce en un locus endógeno. En un ejemplo, el ácido nucleico exógeno se coloca ectópicamente (por ejemplo, en un lugar diferente de su locus endógeno).

En un ejemplo, el método comprende además la introducción de un ácido nucleico exógeno en la célula de rata pluripotente modificada genéticamente. En un ejemplo, el ácido nucleico exógeno es un transgén. En un ejemplo, el ácido nucleico exógeno se introduce en un locus endógeno. En un ejemplo, el ácido nucleico exógeno se introduce ectópicamente (por ejemplo, en un locus diferente de su locus endógeno).

En un aspecto, se proporciona un método para hacer una rata humanizada, que comprende el método de acuerdo con las reivindicaciones adjuntas, y que comprende la modificación de un genoma de una célula de rata pluripotente con un LTVEC que comprende un inserto que comprende una secuencia humana de al menos 5 kb, y empleando la célula de rata pluripotente como una célula donante, introduciendo la célula donante en un embrión huésped, y gestando el embrión huésped en una madre sustituta, en donde la madre sustituta da a luz una progenie de rata que comprende la humanización.

Se proporcionan otros métodos para hacer una rata modificada genéticamente que comprende en su línea germinal una o más modificaciones genéticas como se describe en el presente documento, que comprende: (a) modificar un locus de rata objetivo contenido en una célula procariota que emplea los diversos métodos divulgados en el presente documento; (b) seleccionar una célula procariota modificada que comprende la modificación genética en el locus de rata seleccionado; (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 de rata pluripotente para generar una célula pluripotente modificada genéticamente que comprende el ácido nucleico de inserción en el locus genómico direccionado; (e) seleccionar la célula pluripotente de rata modificada genéticamente; (f) introducción de la célula de rata pluripotente modificada genéticamente en un embrión de rata huésped en una etapa de premórula; y (g) implantar el embrión de rata huésped que comprende la célula de rata pluripotente modificada genéticamente en una madre sustituta para generar una generación F0 derivada de la célula de rata pluripotente modificada genéticamente. En tales métodos, el vector de direccionamiento puede comprender un vector de direccionamiento grande. La célula de rata pluripotente puede ser una célula ES de rata. En métodos adicionales, la etapa de aislamiento (c) comprende además (c1) la linealización del vector de direccionamiento modificado genéticamente (es decir, el LTVEC modificado genéticamente). En todavía realizaciones adicionales, la etapa de introducción (d) comprende además (d1) la introducción de un agente de nucleasa como se describe en el presente documento en la célula de rata pluripotente. En una realización, los pasos 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 de rata pluripotente. En una realización, 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.

Métodos adicionales para modificar un locus genómico diana de una célula de mamífero mediante recombinación homóloga bacteriana (BHR) en una célula procariota se proporcionan y comprenden: (a) proporcionar una célula procariota que comprende un locus diana que comprende un ácido nucleico de rata, (b) introducir en la célula procariota un vector direccionado que comprende un ácido nucleico de inserción flanqueado con un brazo de homología de rata 5' y un brazo de homología de rata 3', en donde el ácido nucleico de inserción comprende una región de mamífero (que incluye, por ejemplo, un inserto de ADN de un ser humano), y (c) seleccionar una célula procariota direccionada que comprende el ácido nucleico insertado en el locus de rata objetivo, en donde la célula procariota es capaz de expresar una recombinasa que media en la BHR. La etapa (a1) puede comprender proporcionar una célula procariota que comprende un locus diana que comprende un ácido nucleico de rata que comprende un primer polinucleótido que comprende un primer sitio de reconocimiento para un primer agente de nucleasa, y la etapa (b1) puede comprender además expresar en la célula procariota un agente de nucleasa eso hace una ruptura de muesca o de doble cadena en o cerca del primer sitio de reconocimiento. Las etapas (a)-(c) se pueden repetir en serie como se describe en el presente documento para permitir la introducción de múltiples ácidos nucleicos de inserción en el locus de rata objetivo en la célula procariota. Una vez que el locus genómico diana se "construye" con la célula procariota, un vector de direccionamiento que comprende el locus de rata objetivo modificado puede aislarse de la célula procariota e introducirse en un locus genómico diana dentro de una célula de rata pluripotente. Las células de rata pluripotentes (es decir, células ES de rata) que comprenden el locus genómico modificado pueden transformarse en ratas modificadas genéticamente.

En algunas realizaciones, pueden llevarse a cabo diversas modificaciones genéticas de los loci genómicos diana divulgados en el presente documento mediante una serie de reacciones de recombinación homóloga (BHR) en células bacterianas utilizando un LTVEC derivado de ADN del Cromosoma Artificial Bacteriano (BAC) utilizando VELOCIGENE® tecnología de ingeniería genética (véase, por ejemplo, la Patente de Estados Unidos No. 6,586,251 y Valenzuela, DM et al. (2003), High-throughput engineering of the mouse genome coupled with high-resolution expression analysis, Nature Biotechnology 21(6): 652-659).

En algunas realizaciones, las células ES de rata direccionadas que comprenden diversas modificaciones genéticas como se describen en el presente documento se utilizan como células ES de inserción y se introducen en un embrión en etapa premórula de un organismo correspondiente, por ejemplo, un embrión de ratón en etapa de 8 células, a través del método VELOCIMOUSE® (Véase, por ejemplo, US 7,576,259, US 7,659,442, US 7,294,754, y US 2008-

0078000 A1). El embrión de rata que comprende las células ES de rata modificadas genéticamente se incuba hasta la etapa de blastocisto y luego se implanta en una madre sustituta para producir un F0. Las ratas que tienen el locus genómico modificado genéticamente pueden identificarse mediante un ensayo de modificación del alelo (MOA) como se describe en el presente documento. La rata de generación F0 resultante derivada de las células de rata ES modificadas genéticamente se cruza con una rata de tipo salvaje para obtener descendientes de generación F1. Tras la genotipificación con cebadores y/o sondas específicas, las ratas F1 que son heterozigotas para el locus genómico genéticamente modificado se cruzan entre sí para producir ratas que son homozigotas para el locus genómico genéticamente modificado. Alternativamente, una rata hembra F0 y una rata macho F0 que tienen la modificación genética pueden cruzarse para obtener una rata F1 homozigota para la modificación genética.

Se proporciona un genoma de rata modificado genéticamente, que comprende una modificación direccionada de una secuencia de ácidos nucleicos de rata endógena con una secuencia de ácidos nucleicos de rata homóloga u ortóloga.

En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga tiene una longitud de aproximadamente 5kb a aproximadamente 200kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 5kb a aproximadamente 10kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 10kb a aproximadamente 20kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga oscila entre aproximadamente 20kb y aproximadamente 30kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga oscila entre aproximadamente 30kb y aproximadamente 40kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 40kb a aproximadamente 50 kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 50kb a aproximadamente 60kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 60kb a aproximadamente 70kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga oscila entre aproximadamente 70kb y aproximadamente 80kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 80kb a aproximadamente 90kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 90kb a aproximadamente 100kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga oscila entre aproximadamente 100kb y aproximadamente 110kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga oscila entre aproximadamente 110kb y aproximadamente 120kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 120kb a aproximadamente 130kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 140kb a aproximadamente 150kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 150kb a aproximadamente 160kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 160kb a aproximadamente 170kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 170kb a aproximadamente 180kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 180kb a aproximadamente 190kb. En una realización, la secuencia de ácidos nucleicos no rata homóloga u ortóloga varía de aproximadamente 190kb a aproximadamente 200kb. Varios polinucleótidos de interés que pueden emplearse en el ácido nucleico de inserción se describen en otra parte en el presente documento.

6. Células

Los diversos métodos y composiciones descritos aquí emplean un sistema de direccionamiento de locus genómico en una célula, en donde, la célula es una célula pluripotente.

En kas realizaciones dela presente invención, la célula pluripotente es una célula de rata pluripotente. En una realización, la célula de rata pluripotente es una célula madre embrionaria de rata (ES). En una realización, la célula de rata pluripotente es una célula madre pluripotente inducida (iPS) o es una célula progenitora de desarrollo restringido. En otras realizaciones, la célula de rata pluripotente es capaz de mantener su pluripotencia después de al menos una modificación genética direccionada de su genoma y es capaz de transmitir la modificación direccionada a una línea germinal de una generación F1.

En una rfealización, la célula pluripotente es un óvulo fertilizado de rata en la etapa de célula única.

Las diversas células empleadas en el método y las composiciones divulgadas en el presente documento también pueden comprender células procariotas, tales como una célula bacteriana, que incluye *E. coli*. En realizaciones específicos, la célula procariota es una cepa de *E. coli* competente para la recombinación. En un ejemplo, la célula procariota comprende un ácido nucleico que codifica la recombinasa, mientras que, en otros casos, la célula procariota no comprende el ácido nucleico que codifica la recombinasa, y el ácido nucleico que codifica la recombinasa se introduce en la célula procariota. En una realización, el ácido nucleico que codifica la recombinasa comprende un ADN o un ARNm. En algunas realizaciones, el ácido nucleico que codifica la recombinasa es pABG. En una realización, la recombinasa se expresa bajo el control de un promotor inducible. En una realización, la expresión de la recombinasa está controlada por arabinosa.

A. Células madre embrionarias (ES) de rata

Como se describe en detalle anteriormente, las diversas composiciones y métodos proporcionados en el presente documento pueden emplear células madre embrionarias (ES) de rata. En una realización, la célula de rata pluripotente es una célula ES de rata. En una realización, la célula ES de rata se deriva de una cepa de rata es una rata Wistar, una cepa LEA, una cepa Sprague Dawley, una cepa Fischer, F344, F6 y Dark Agouti o ACI. En una realización, la cepa de rata es una mezcla de dos o más de una cepa seleccionada del grupo que consiste en Wistar, LEA, Sprague Dawley, Fischer, F344, F6 y Dark Agouti. En una realización, la célula ES de rata se deriva de una cepa consanguínea. En una realización, la célula ES de rata se deriva de una cepa seleccionada de una cepa DA y una cepa ACI. En una realización específica, la célula ES de rata se deriva de una cepa ACI. En una realización, la célula ES de rata se deriva de un blastocisto de rata.

En otras realizaciones, la célula ES de rata se caracteriza por la expresión de al menos un marcador de pluripotencia. En realizaciones específicas, la célula ES de rata se caracteriza por la expresión de un marcador de pluripotencia que comprende Oct-4, Sox-2, fosfatasa alcalina o una combinación de los mismos. En una realización, la célula ES de rata es una célula ES de rata macho (XY) o una célula ES de rata hembra (XX).

En una realización, después de una a 15 modificaciones genéticas en serie, las células ES de rata modificadas genéticamente tras la exposición al medio de diferenciación son capaces de diferenciarse en una pluralidad de tipos de células.

En una realización, después de una a 15 modificaciones genéticas en serie, las células ES de rata modificadas genéticamente son capaces de mantenerse en un estado indiferenciado en cultivo. En una realización, las células ES de rata modificadas genéticamente y cultivadas en estado indiferenciado, cuando se emplean como células donantes en un embrión huésped de rata, pueblan el embrión y forman un blastocisto que comprende de una a quince modificaciones genéticas. En una realización, el blastocisto, cuando se implanta en una madre sustituta en condiciones adecuadas para la gestación, se desarrolla en una progenie de rata F0 que comprende de una a 15 modificaciones genéticas.

Se proporciona una célula ES de rata aislada que es capaz de mantener la pluripotencia después de una o más modificaciones genéticas *in vitro*, y que es capaz de transmitir un genoma genéticamente modificado a una línea germinal de una generación F1.

En una realización, la célula ES de rata mantiene su pluripotencia para desarrollarse en una pluralidad de tipos celulares después de una o más modificaciones genéticas en serie *in vitro* (por ejemplo, dos, tres, cuatro, cinco o seis o más modificaciones genéticas en serie). En una realización, la modificación genética está mediada por una electroporación, por inyección intracitoplasmática, por una infección viral, por un adenovirus, por lentivirus, por retrovirus, por transfección, por transfección mediada por lípidos o por Nucleofaction™.

En una realización, la célula ES de rata mantiene su pluripotencia para desarrollarse en una pluralidad de tipos celulares después de una única ronda de electroporación con un ácido nucleico exógeno. En una realización, la célula ES de rata mantiene su pluripotencia para convertirse en una pluralidad de tipos de células después de una ronda 2ª, 3ª, 4ª, 5ª, 6ª, 7ª, 8ª, 9ª, 10ª, 11ª, 12ª, 13ª, 14ª o 15ª de electroporación con un ácido nucleico exógeno.

En otras realizaciones, las células ES de rata empleadas son las descritas en la Solicitud de los Estados Unidos No. 14/185,703, presentada el 20 de febrero de 2014.

La célula de rata pluripotente empleada en los diversos métodos y composiciones divulgadas en el presente documento puede caracterizarse por la expresión de al menos un marcador de pluripotencia que comprende Dnmt3L, Eras, Err-beta, Fbxo15, Fgf4, Gdf3, Klf4, Lef1, LIF receptor, Lin28, Nanog, Oct4, Sox15, Sox2, Utf1, y/o una combinación de los mismos. En otros casos, la célula de rata pluripotente empleada en los diversos métodos y composiciones descritas aquí se caracteriza por una o más de las siguientes características: (a) falta de expresión de uno o más marcadores de pluripotencia que comprenden c-Myc, Ecat1 y/o Rexo1; (b) falta de expresión de uno o más marcadores mesodérmicos que comprenden Brachyury y/o Bmpr2; (c) falta de expresión de uno o más marcadores endodérmicos que comprenden Gata6, Sox17 y/o Sox7; o (d) falta de expresión de uno o más marcadores neuronales que comprenden Nestina y/o Pax6. Como se usa en el presente documento, "falta de expresión" en lo que se refiere a la expresión de un marcador de pluripotencia significa que la expresión del marcador de pluripotencia es igual o inferior al fondo experimental según se determina para cada experimento individual.

En una realización no limitativa, las células ES de rata proporcionadas en el presente documento tienen una o más de cualquiera de las siguientes propiedades:

(a) tener competencia en la línea germinal, es decir, cuando la célula ES de rata se implanta en un embrión huésped de rata, el genoma de la línea celular ES de rata se transmite a una descendencia;

(b) tener competencia en la línea germinal después de al menos una modificación genética direccionada, lo que significa que cuando la célula ES de rata que tiene la modificación genética direccionada se implanta en un embrión huésped de rata, la modificación genética direccionada dentro del genoma de la línea celular ES de rata se transmite en una descendencia;

(c) tener pluripotencia *in vitro*;

(d) tener totipotencia *in vitro*;

(e) cuando se cultiva in *vitro*, se adhieren a una capa de células alimentadoras;

(f) cuando se cultivan *in vitro*, forman colonias de tipo esfera cuando se colocan en placas *in vitro* sobre una capa de células alimentadoras;

(g) mantener la pluripotencia cuando se cultiva *in vitro* en condiciones que comprenden una capa de células alimentadoras que no están modificadas genéticamente para expresar el factor inhibidor de la leucemia (LIF), donde los medios de cultivo comprenden una concentración suficiente de LIF;

(h) mantener la pluripotencia cuando se cultiva *in vitro* en condiciones que comprenden una capa de células alimentadoras, en donde los medios de cultivo comprenden LIF de ratón o una variante activa o fragmento de la misma;

(i) comprenden una firma molecular que se caracteriza por

i) la expresión de uno o más de los genes específicos de células ES de rata que comprenden Adheres Junctions Associate Protein (Ajap1), Claudina 5 (Cldn5), factor 9 de intercambio de nucleótido guanina Cdc42 (Arhgef9), proteína quinasa dependiente de calcio/calmodulina (Camk4), efrina-Al (Efna1), receptor de EPH A4 (Epha4), proteína de unión gap beta 5 (Gjb5), factor 1 de unión al factor de crecimiento insulínico similar (Igfbpl1), interleulina 36 beta(Il1f8), receptor interleucina 28, alfa (Il28ra), factor 1 de determinación izquierda-derecha (Lefty1), receptor del factor inhibidor de la leucemia alfa (Lifr), receptor 2 del ácido lisofosfatídico (Lpar2), receptor de pentraxina neuronal (Ntm), proteína tirosina fosfatasa no receptor tipo 18 (Ptpn18), homeobox tipo 2 Caudal (Cdx2), Fibronectina tipo III y dominios de repetición de anquirina 1 (Fank1), Caja E1 Forkhead (factor 2 de transcripción tiroidea) (Foxe1), Peludo/potenciador de la división relacionado con el motivo 2 YRPW (Hey2), Caja E1 Forkhead (factor 2 de transcripción tiroidea) (Foxe1), Peludo/potenciador de la división relacionado con motivo 2 YRPW (Hey2), factor 1 de unión al potenciador linfoide (Lef1), tipo sal 3 (Drosophila) (Sall3), Homeobox SATB 1 (Satb1), miR-632, o una combinación de los mismos;

ii) la expresión de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 , 23, 24, 25 o más de los genes específicos de células ES de rata que comprenden Adheres Junctions Associate Protein (Ajap1), Claudina 5 (Cldn5), factor 9 de intercambio de nucleótido guanina Cdc42 (Arhgef9), proteína quinasa IV dependiente de calcio/calmodulina (Camk4), efrina-Al (Efna1), receptor de EPH A4 (Epha4), proteína de la unión beta 5 (Gjb5), factor 1 de unión al factor de crecimiento insulínico similar (Igfbpl1), Interleulin 36 beta(Il1f8), receptor de interleucina 28, alfa (Il28ra), factor 1 de determinación izquierda-derecha (Lefty1), receptor de factor inhibidor de la leucemia alfa (Lifr), receptor 2 de ácido lisofosfatídico (Lpar2), receptor de pentraxina neuronal (Ntm), proteína tirosina fosfatasa no receptor tipo 18 (Ptpn18)), Homeobox tipo Caudal 2 (Cdx2), fibronectina tipo III y dominios de repetición de anquirina 1 (Fank1), Caja E1 Forkhead (factor 2 de transcripción tiroidea) (Foxe1), velludo/potenciador de la división relacionado con el motivo 2 YRPW (Hey2) Caja E1 Forkhead (factor 2 de transcripción de la tiroides) (Foxe1), velloso/potenciador de la división relacionado con el motivo 2 YRPW (Hey2), factor 1 de unión al reforzador linfoide (Lef1), tipo sal 3 (Drosophila) (Sall3), SATB Homeobox 1 (Satb1), miR-632, o una combinación de los mismos;

iii) al menos un aumento de 20 veces en la expresión de uno o más de los genes específicos de células ES de rata como se establece en la Tabla 9 cuando se compara con una célula ES de ratón F1H4;

iv) al menos un aumento de 20 veces en la expresión de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 o más de los genes específicos de células ES de rata como se establece en la Tabla 9 cuando se comparan con una célula ES de ratón F1H4;

v) la expresión de uno o más de los genes específicos de células ES de rata como se establece en la Tabla 10;

vi) la expresión de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50 o más de los genes específicos de células ES de rata como se establece en la Tabla 10;

vii) al menos un aumento de 20 veces en la expresión de uno o más de los genes específicos de células ES de rata como se establece en la Tabla 10 cuando se compara con una célula ES de ratón F1H4;

viii) al menos un aumento de 20 veces en la expresión de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50 o más de los genes específicos de células ES de rata como se establece en la Tabla 10 cuando se comparan con una célula ES de ratón F1H4;

ix) al menos una disminución de 20 veces en la expresión de uno o más de los genes específicos de células ES de rata como se establece en la Tabla 8 cuando se compara con una célula ES de ratón F1H4;

x) al menos una disminución de 20 veces en la expresión de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50 o más de los genes específicos de células ES de rata como se establece en la Tabla 8 cuando se comparan con una célula ES de ratón F1H4;

xi) cualquier combinación de expresión de los genes específicos de células ES de rata de las partes (i)-(x);

xii) un nivel de expresión relativo de marcadores de pluripotencia como se muestra en la Tabla 11 para al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 o 18 de los marcadores de pluripotencia listados. Véase, columna de clasificación de pluripotencia de la Tabla 11 para los niveles de expresión relativos;

xiii) un nivel de expresión relativo de los marcadores mesodérmicos como se muestra en la Tabla 11 para al menos 2, 3 o 4 de los marcadores mesodérmicos listados. Véase, columna de clasificación mesodérmica en la Tabla 11 para los niveles de expresión relativos;

xiv) un nivel de expresión relativo de marcadores endodérmicos como se muestra en la Tabla 11 para al menos 2, 3, 4, 5 o 6 de los marcadores endodérmicos listados. Consulte la columna de clasificación endodérmica en la Tabla 11 para los niveles de expresión relativos;

xv) un nivel de expresión relativo de marcadores neurales como se muestra en la Tabla 11 para al menos 2 y 3 de los marcadores neuronales enumerados. Véase, columna de clasificación neuronal en la Tabla 11 para los niveles de expresión relativos;

xvi) un nivel de expresión relativo de marcadores de trofectodermo como se muestra en la Tabla 11 para los marcadores de trofectodermo listados. Consulte la columna de clasificación del trofectodermo en la Tabla 11 para los niveles de expresión relativos;

xvii) cualquier nivel de expresión relativa de uno o más (2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30) de los marcadores de pluripotencia, marcadores mesodérmicos, marcadores endodérmicos, marcadores neurales y/o marcadores de trofectodermo expuestos en la Tabla 11;

xviii) el nivel de expresión relativa de cada uno de los marcadores establecidos en la Tabla 11;

xix) cualquier combinación de las firmas establecidas en xii-xiix; y/o

xx) cualquier combinación de la firma establecida en i-xiix;

(j) tener la capacidad de producir una rata F0;

(k) son susceptibles de subcultivarse y mantener el estado indiferenciado;

(l) tienen el mismo número de cromosomas que una célula de rata normal;

(m) mantener la pluripotencia *in vitro* sin requerir señalización LIF paracrina;

(n) auto renovación, lo que significa que se dividen indefinidamente mientras mantienen la pluripotencia;

(o) las células ES de rata expresan al menos un marcador de pluripotencia que comprende Dnmt3L, Eras, Err-beta, Fbxo15, Fgf4, Gdf3, Klf4, Lef1, receptor LIF, Lin28, Nanog, Oct4, Sox15, Sox2, Utf1 y/o una combinación de los mismos;

(p) las células ES de rata no expresan uno o más marcadores de diferenciación que comprenden c-Myc, Ecat1 y/o Rexo1;

(q) las células ES de rata no expresan uno o más marcadores mesodérmicos que comprenden Brachyury, Bmpr2 y/o una combinación de los mismos;

(r) las células ES de rata no expresan uno o más marcadores endodérmicos que comprenden Gata6, Sox 17, Sox7 y/o una combinación de los mismos; y/o

(s) las células ES de rata no expresan uno o más marcadores neurales que comprenden Nestina, Pax6 y/o una combinación de los mismos.

Una o más de las características reseñadas en (a)-(s) pueden estar presentes en una célula ES de rata, una población de células ES de rata o una línea celular ES de rata empleadas en los métodos y composiciones proporcionadas en el presente documento, en donde las células ES de rata no han sufrido una modificación genética direccionada. Además, después de una o más modificaciones genéticas en el locus diana de la rata, como se describe en detalle más arriba, la una o más de las características reseñadas en (a)-(s) pueden conservarse en la célula ES de rata después de la modificación genética del locus diana.

En una realización, la célula ES de rata exhibe una eficacia de recombinación homóloga de al menos 2%, al menos 3%, al menos 4%, al menos 5%, al menos 6%, al menos 7%, al menos 8%, al menos 9%, al menos 10%, al menos 11%, al menos 12%, al menos 13%, al menos 14%, al menos 15%, al menos 20%, al menos 25%, al menos 30 %, al menos 35%, al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, o al menos 80%.

En una realización, la eficiencia de recombinación homóloga que emplea la célula ES de rata es mayor que 4%.

En una realización, la célula ES de rata tiene un tiempo de duplicación que varía de 24 horas a 36 horas. En una realización, la célula ES de rata tiene un tiempo de duplicación de 25 horas.

En una realización, la célula ES de rata se puede pasar hasta al menos 15 veces en medio 2i (Millipore Cat. SF016-200). En una realización, la célula ES de rata se puede pasar al menos 14 veces en medio 2i (Millipore Cat. No. SF016-200). En una realización, la célula ES de rata se puede pasar al menos 13, 12, 11, 10, 9, 8, 7, 6 o 5 veces en medio 2i.

En una realización, cuando se trasplantan a un embrión de rata en etapa premórula, la célula ES de rata puede contribuir a al menos el 90% de las células en una generación F0. En una realización, cuando se trasplantan a un embrión de rata en etapa premórula, la célula ES de rata puede contribuir a al menos el 95%, 96%, 97%, 98% o 99% de las células en una generación F0.

En realizaciones específicas, las diversas células ES de rata y líneas celulares empleadas en los diversos métodos y composiciones proporcionados en el presente documento se usan para generar una modificación direccionada en un locus diana. La célula ES de rata que tiene estas modificaciones genéticas específicas puede ser competente en la línea germinal, lo que significa que cuando la célula ES de rata que tiene la modificación genética específica se implanta en un embrión huésped de rata, la modificación genética específica de la célula ES de rata se transmite a la descendencia (es decir, la población F1). Así, en diversos aspectos, las células ES de rata en los diversos métodos y composiciones se emplean para obtener una alta frecuencia, o alta eficacia, de la transmisión de la línea germinal de un genoma de célula de rata a partir de células ES de rata que han sufrido una modificación genética direccionada. En diversas realizaciones, la frecuencia de transmisión de la línea germinal es mayor que 1:600, mayor que 1:500, mayor que 1:400, mayor que 1:300, mayor que 1:200 y mayor que 1:100. En diversas realizaciones, la frecuencia de transmisión de la línea germinal es mayor que 1%, mayor que 2%, mayor que 3%, mayor que 4%, mayor que 5%, mayor que 6%, mayor que 7%, mayor que 8%, más del 9%, más del 10%, hasta aproximadamente el 16%, más del 25%, más del 50%, más del 60%, más del 65%, más del 70%, más del 75% o más. En diversas realizaciones, la frecuencia de transmisión de la línea germinal varía de 9% a 16%. En diversos aspectos, el porcentaje de la progenie derivada de rESC donante en la generación F1 es 1% o más, 2% o más, 3% o más, 10% o más, 20% o más, 30% o más, 40% o más, 50% o más, 60% o más, de 3% a alrededor de 10% o más; de 3% o más a aproximadamente 63%, de aproximadamente 10% a aproximadamente 30%, de aproximadamente 10% a aproximadamente 50%, de aproximadamente 30% a aproximadamente 70%, de aproximadamente 30% a aproximadamente 60%, de aproximadamente 20% hasta aproximadamente el 40%, desde aproximadamente el 20% hasta el 65%, o desde aproximadamente el 40% hasta el 70%. Por lo tanto, una célula ES de rata que tiene una modificación genética direccionada tiene la capacidad de transmitir su genoma a la población F1.

Una célula ES de rata que tiene una modificación genética direccionada puede ser pluripotente y/o totipotente. Se pueden usar diversos métodos para determinar si una célula ES de rata es pluripotente. Por ejemplo, la célula ES puede analizarse para la expresión de diversos marcadores pluripotentes que incluyen, entre otros, Oct-4, Sox2, fosfatasa alcalina o una combinación de los mismos. Véase, por ejemplo, Okamoto, K. et al., Cell, 60: 461-472 (1990), Scholer, HR et al., EMBO J. 9: 2185-2195 (1990)) y Nanog (Mitsui, K. et al., Cell, 113: 631-642 (2003), Chambers, I. et al., Cell, 113: 643-655 (2003) para diversos métodos de análisis de la presencia o el nivel de tales marcadores. Véanse también las Figuras 2 y 3 proporcionadas aquí. Otros marcadores de pluripotencia incluyen, por ejemplo, la presencia de al menos 1, 2, 3, 4 o 5 marcadores de pluripotencia que comprenden Nanog, Klf4, Dppa2, Fgf4, Rex1, Eras, Err-beta y/o Sall 3. Otros marcadores de pluripotencia incluyen, por ejemplo, la ausencia de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 marcador de pluripotencia que comprende T/Brachyury, Flk1, Nodal, Bmp4, Bmp2, Gata6, Sox17, Hhex1, Sox7 y/o Pax6.

En realizaciones específicas, la expresión y/o el nivel de expresión de estos marcadores pueden determinarse usando RT-PCR. Hay varios kits disponibles para determinar el nivel y/o la presencia de fosfatasa alcalina, incluyendo, por ejemplo, un kit de tinción de tejidos ALP (Sigma) y el Kit de sustrato de fosfatasa alcalina roja vector I (Funakoshi) y similares. Ensayos adicionales incluyen hibridación in situ, inmunohistoquímica, inmunofluorescencia. En realizaciones específicas, la célula ES de rata se caracteriza por la expresión de al menos un marcador de pluripotencia, que incluye, por ejemplo, la expresión de Oct-4, Sox2, fosfatasa alcalina o una combinación de los mismos, y preferiblemente los tres marcadores.

Las diversas células ES de rata empleadas en el método y las composiciones proporcionadas en el presente documento son capaces de mantener la pluripotencia y/o totipotencia mientras se mantienen en condiciones de cultivo *in vitro*. Por lo tanto, las diversas células ES de rata proporcionadas en el presente documento pueden, en algunas realizaciones, subcultivarse mientras se mantiene el estado indiferenciado. Diversos métodos para cultivar células ES de rata se discuten con más detalle en otra parte de este documento y en la Solicitud de Patente de los Estados Unidos No. 14/185,103, presentada el 20 de febrero de 2014.

En algunas realizaciones, las células madre embrionarias de rata empleadas aquí se han aislado del embrión de rata empleando varias técnicas de aislamiento, purificación y expansión de cultivo que se discuten en detalle en la Solicitud

de los Estados Unidos. En la Solicitud de Patente de los Estados Unidos No. 14/185,103, presentada el 20 de febrero, 2014.

Un embrión de rata o célula ES de rata "aislada" se ha eliminado de su entorno natural. El término "aislado" puede significar libre de 70%, 80%, 90%, 95%, 96%, 97%, 98% o 99% de los constituyentes con los que se encuentra un componente en su estado natural. Como se usa en el presente documento, una "línea celular" de rata ES comprende una población de células de rata aisladas que se desarrollaron a partir de una única célula ES de rata y, por lo tanto, la población de células dentro de una línea celular dada tiene una composición genética uniforme distinta a la de las mutaciones o los cambios cariotípicos ocurre durante la propagación o durante modificaciones genéticas específicas. Por ejemplo, las células ES de rata pueden caracterizarse por un alto nivel de euploidía. Sin embargo, en algunas líneas celulares el nivel de euploidía es inferior al 100% debido a cambios cariotípicos en la propagación de la línea desde una sola célula. Además, una población dada de células ES de rata puede comprender al menos $1x10^3$, $1x10^4$, $1x10^5$, $1x10^6$, $1x10^7$, $1x10^8$, $1x10^9$ o $1x10^{10}$ células o más. Algunas poblaciones celulares tienen células suficientes para permitir la selección de una célula modificada deseada pero no un número excesivamente mayor para reducir la posibilidad de que se desarrollen mutaciones o cambios cariotípicos en la línea celular. Por ejemplo, algunas poblaciones de células tienen de $1x10^3$ a $1x10^6$ células.

Como se analiza en otra parte del presente documento, se proporcionan diversos métodos para la modificación genética direccionada de una línea celular ES de rata. Cuando se llevan a cabo tales métodos, al menos una célula dentro de una línea de células ES de rata contiene la modificación genética direccionada. A través de diversas técnicas de cultivo y/o selección, se producen líneas celulares de rata ES que tienen una o más modificaciones genéticas direccionadas deseadas.

En realizaciones específicas, una célula ES de rata, una población de células ES de rata o una línea celular ES de rata (que no han sufrido una modificación genética direccionada y/o tienen una modificación genética específica) son euploides y así tener un número de cromosoma que es un múltiplo exacto del número haploide. En una realización adicional, una célula ES de rata, una población de células ES de rata o una línea celular ES de rata (que no han sufrido una modificación genética direccionada y/o tienen una modificación genética específica) son diploides, y así tener dos conjuntos haploides de cromosomas homólogos. Cuando se hace referencia a una población de células ES de rata o una población de células de una población determinada de células ES de rata o una línea de células ES de rata (que no han sufrido una modificación genética específica y/o tienen una modificación genética específica), al menos aproximadamente 50%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de las células con la población dada son euploides y/o diploides. En otros casos, al referirse a una población de células ES de rata o una población de células de una línea celular de ES de rata determinada (que no han sufrido una modificación genética específica y/o tienen una modificación genética específica), al menos aproximadamente el 50% a 95%, aproximadamente 60% a 90%, aproximadamente 60% a 95%, aproximadamente 60% a 85%, aproximadamente 60% a 80%, aproximadamente 70% a 80%, aproximadamente 70% a 85%, aproximadamente 70% a aproximadamente 90%, aproximadamente 70% a aproximadamente 95%, aproximadamente 70% a aproximadamente 100%, aproximadamente 80% a aproximadamente 100%, aproximadamente 80% a aproximadamente 95%, aproximadamente 80% a aproximadamente 90%, aproximadamente 90% a aproximadamente 100%, aproximadamente 90% a aproximadamente 99%, aproximadamente 90% a aproximadamente 98%, aproximadamente 90% a aproximadamente 97%, aproximadamente 90% a aproximadamente 95% de las células dentro de la población dada son euploides y/o diploides.

En otras realizaciones adicionales, una célula ES de rata, una población de células ES de rata o una línea celular ES de rata (que no han sufrido una modificación genética direccionada y/o tienen una modificación genética específica) tienen 42 cromosomas. Cuando se hace referencia a una población de células ES de rata o una población de células de una línea celular ES de rata dada (que no han sufrido una modificación genética específica y/o tienen una modificación genética específica) al menos aproximadamente el 50%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de las células con la población dada tienen 42 cromosomas. En otros casos, al referirse a una población de células ES de rata o una población de células de una línea celular de ES de rata dada (que no han sufrido una modificación genética específica y/o tienen una modificación genética específica) al menos aproximadamente del 50% al 95%, aproximadamente del 60% al 90%, aproximadamente del 60% al 95%, aproximadamente del 60% al 85%, aproximadamente del 60% al 80%, aproximadamente del 70% al 80%, aproximadamente del 70% al 85%, aproximadamente del 70% a aproximadamente 90% aproximadamente del 70% a aproximadamente 95%, aproximadamente del 70% a aproximadamente 100%, aproximadamente del 80% a aproximadamente 100%, aproximadamente del 80% a aproximadamente 95%, aproximadamente del 80% a aproximadamente 90%, aproximadamente del 90% a aproximadamente 100% aproximadamente del 90% a aproximadamente 99%, aproximadamente 90% a aproximadamente 98%, aproximadamente 90% a aproximadamente 97%, aproximadamente 90% a aproximadamente 95% de las células dentro de la población dada tienen 42 cromosomas.

En otras realizaciones, una célula ES de rata, una población de células ES de rata o una línea celular ES de rata (que no se han sometido a una modificación genética direccionada y/o tienen una modificación genética específica) aquí proporcionadas forman colonias similares a esferas cuando se siembra una capa de células alimentadoras *in vitro*. La morfología "similar a una esfera" se refiere a la forma de las colonias de células ES de rata en cultivo, en lugar de la forma de células ES individuales. Las colonias de células ES de rata son de tipo esférico. Las colonias, que están

ligeramente adheridas a las células alimentadoras, parecen circulares (tienen una morfología circular). Las colonias que flotan libremente son de tipo esférico. Las colonias de células ES de rata son de tipo esférico y muy compactas, lo que significa que los límites entre las células son muy difíciles de ver. El borde de la colonia aparece brillante y nítido. Los núcleos individuales son difíciles de distinguir porque las células son muy pequeñas (de modo que el núcleo ocupa la mayor parte del volumen de la célula). Las células ES de ratón forman colonias alargadas y se unen fuertemente a las células alimentadoras. La morfología de mESC puede variar con la tensión; por ejemplo, las colonias B6 son más redondas y más abovedadas que las colonias F1H4, pero aún son más alargadas que las rESC. Las colonias de células ES humanas son más planas y más dispersas que las colonias mESC. Las colonias ES de rata instantánea no son planas y no se parecen a las colonias de células ES humanas.

En otras realizaciones adicionales, una célula ES de rata, una población de células ES de rata o una línea celular ES de rata (que no han sufrido una modificación genética direccionada y/o tienen una modificación genética direccionada) tienen una morfología circular. A continuación, se proporciona una escala de morfología para un círculo, donde una puntuación de 10 representa un círculo perfecto y una puntuación de 1 representa una elipse.

Escala de morfología de un círculo:

10 = Un círculo con una estructura que tiene un eje longitudinal y un eje vertical que corre a través del centro de la estructura y tiene la misma longitud.

9 = Una estructura que tiene un eje longitudinal y un eje vertical que corre a través del centro de la estructura, en donde uno de los ejes está entre 0.9999 y 0.9357 de la longitud del otro eje.

8 = Una estructura que tiene un eje longitudinal y un eje vertical que recorre el centro de la estructura, en donde uno de los ejes está entre 0.9357 y 0.875 de la longitud del otro eje.

7 = Una estructura que tiene un eje longitudinal y un eje vertical que atraviesa el centro de la estructura, en donde uno de los ejes está entre 0.875 y aproximadamente 0.8125 de la longitud del otro eje.

6 = Una estructura que tiene un eje longitudinal y un eje vertical que atraviesa el centro de la estructura, en donde uno de los ejes está entre 0.8125 y 0.750 de la longitud del otro eje.

5 = Una estructura que tiene un eje longitudinal y un eje vertical que atraviesa el centro de la estructura, donde uno de los ejes está entre 0.750 y 0.6875 de la longitud del otro eje.

4 = Una estructura que tiene un eje longitudinal y un eje vertical que recorre el centro de la estructura, en donde uno de los ejes está entre 0.6875 y 0.625 de la longitud del otro eje.

3 = Una estructura que tiene un eje longitudinal y un eje vertical que atraviesa el centro de la estructura, en donde uno de los ejes está entre 0.625 y 0.5625 de la longitud del otro eje.

2 = Una estructura que tiene un eje longitudinal y un eje vertical que atraviesa el centro del círculo, en donde uno de los ejes está entre 0.5625 y 0.523 de la longitud del otro eje.

1 = Una elipse se define como que tiene un eje longitudinal y un eje vertical que recorre el centro de la estructura, en donde uno de los ejes está entre 0.523 y 0.500 de la longitud del otro eje.

En una realización no limitativa, la población de células ES de rata o una población de células de una línea celular ES de una rata dada (que no han sufrido una modificación genética direccionada y/o tienen una modificación genética específica) tienen al menos aproximadamente el 50%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de las células con la población dada tienen una puntuación de morfología circular de 10, 9 u 8. En otras realizaciones, la población de células ES de rata o una población de células de una rata dada la línea celular ES (que no ha sufrido una modificación genética específica y/o tiene una modificación genética específica) tiene al menos aproximadamente 50% a 95%, aproximadamente 60% a 90%, aproximadamente 60% a 95%, aproximadamente 60% a 85%, aproximadamente 60% a 80%, aproximadamente 70% a 80%, aproximadamente 70% a 85%, aproximadamente 70% a aproximadamente 90%, aproximadamente 70% a aproximadamente 95%, aproximadamente 70% a aproximadamente 100%, aproximadamente 80% a aproximadamente 100%, aproximadamente 80% a aproximadamente 95%, aproximadamente 80% a aproximadamente 90%, aproximadamente 90% a aproximadamente 100%, aproximadamente 90% a aproximadamente 99%, aproximadamente 90% a aproximadamente 98%, aproximadamente 90% a aproximadamente 97%, aproximadamente 90% a aproximadamente 95% de las células dentro de la población dada tienen una puntuación de morfología circular de 10, 9 u 8.

En otra realización no limitante, la población de células ES de rata o una población de células de una línea celular de ES de rata dada (que no han sufrido una modificación genética direccionada y/o tienen una modificación genética específica) tienen al menos aproximadamente 50%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de las células con la población dada tienen una puntuación de morfología circular de 7, 6, 5, 4 o 3. En otras realizaciones no limitantes, la población de células ES de rata o una

población de células de una línea celular ES de rata dada (que no han sufrido una modificación genética específica y/o tienen una modificación genética específica) tienen al menos aproximadamente 50% a 95%, aproximadamente 60% a 90%, aproximadamente 60% a 95%, aproximadamente 60% a 85%, aproximadamente 60% a 80%, aproximadamente 70% a 80%, aproximadamente 70% a 85%, aproximadamente 70% a aproximadamente 90%, aproximadamente 70% a aproximadamente 95%, aproximadamente 70% a aproximadamente 100%, aproximadamente 80% a aproximadamente 100%, aproximadamente 80% a aproximadamente 95%, aproximadamente 80% a aproximadamente 90%, aproximadamente 90% a aproximadamente 100%, aproximadamente 90% a aproximadamente 99%, aproximadamente 90% a aproximadamente 98%, aproximadamente 90% a aproximadamente 97%, aproximadamente 90% a aproximadamente 95% de las células dentro de la población dada tienen una puntuación de morfología circular de 7, 6, 5, 4 o 3.

En otras realizaciones adicionales, se forman colonias de tipo esfera cuando las células ES de rata (que no han sufrido una modificación genética direccionada y/o tienen una modificación genética específica) se colocan sobre una capa de células alimentadoras *in vitro*. A continuación, se proporciona una escala de morfología para una esfera, donde una puntuación de 10 representa una esfera perfecta y una puntuación de 1 representa una estructura elíptica tridimensional.

Escala de morfología de una estructura similar a una esfera:

10 = Una esfera es una estructura que tiene un eje X y un eje Y y un eje Z, cada uno de los cuales atraviesa el centro de la estructura y tiene la misma longitud.

9 = Una estructura que tiene un eje X y un eje Y y un eje Z que se extiende a través del centro de la estructura, en donde uno de los ejes está entre 0.9999 y 0.9357 la longitud de al menos uno de los otros ejes.

8 = Una estructura que tiene un eje X y un eje Y y un eje Z que se extiende a través del centro de la estructura, en donde uno de los ejes está entre 0.9357 y 0.875, la longitud de al menos uno o ambos ejes.

7 = Una estructura que tiene un eje X y un eje Y y un eje Z que se extiende a través del centro de la estructura, en donde uno de los ejes está entre 0,875 y 0,8125 de la longitud de al menos uno o ambos de los otros ejes.

6 = Una estructura que tiene un eje X y un eje Y y un eje Z que corre a través del centro de la estructura, en donde uno de los ejes está entre 0.8125 y 0.750 la longitud de al menos uno o ambos de los otros ejes.

5 = Una estructura que tiene un eje X y un eje Y y un eje Z que se extiende a través del centro de la estructura, en donde uno de los ejes es 0.750 a 0.6875, la longitud de al menos uno o ambos ejes.

4 = Una estructura que tiene un eje X y un eje Y y un eje Z que se ejecuta a través del centro de la estructura, en donde uno de los ejes es 0.6875 a 0.625 la longitud de al menos uno o ambos de los otros ejes.

3 = Una estructura que tiene un eje X y un eje Y y un eje Z que se extiende a través del centro de la estructura, en donde uno de los ejes está entre 0.625 y 0.5625 la longitud de al menos uno o ambos ejes.

2 = Una estructura que tiene un eje X y un eje Y y un eje Z que se extiende a través del centro de la estructura, en donde uno de los ejes está entre 0.5625 y 0.523, la longitud de al menos uno o ambos ejes.

1 = Una estructura que tiene un eje X y un eje Y y un eje Z que se extiende a través del centro de la estructura, en donde uno de los ejes está entre 0.523 y 0.500 de la longitud de al menos uno o ambos ejes.

En una realización no limitativa, la población de células ES de rata o una población de células de una línea celular ES de rata dada (que no han sufrido una modificación genética direccionada y/o tienen una modificación genética específica) tienen al menos aproximadamente el 50%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de las colonias que se forman cuando las células se colocan sobre una capa de células alimentadoras *in vitro* tienen una morfología similar a una esfera de 10, 9 u 8. En otras realizaciones, la población de células ES de rata o una población de células de una línea celular ES de rata dada (que no han sufrido una modificación genética específica y/o tienen una modificación genética específica) tienen al menos aproximadamente 50% a 95%, aproximadamente 60% a 90%, aproximadamente 60% a 95%, aproximadamente 60% a 85%, aproximadamente 60% a 80%, aproximadamente 70% a 80%, aproximadamente 70% a 85%, aproximadamente 70% a aproximadamente 90%, aproximadamente 70% a aproximadamente 95%, aproximadamente 70% a aproximadamente 100%, aproximadamente 80% a aproximadamente 100%, aproximadamente 80% a aproximadamente 95%, aproximadamente 80% a aproximadamente 90%, aproximadamente 90% a aproximadamente 100%, aproximadamente el 90% a aproximadamente el 99%, aproximadamente el 90% a aproximadamente el 98%, aproximadamente el 90% a aproximadamente el 97%, aproximadamente el 90% a aproximadamente el 95% de las colonias que se forman cuando las células se colocan sobre una capa de células alimentadoras *in vitro* tienen una morfología en forma de esfera de un 10, 9 u 8.

En otra realización no limitante, la población de células ES de rata o una población de células de una línea celular ES de rata dada (que no han sufrido una modificación genética direccionada y/o tienen una modificación genética

específica) tienen al menos aproximadamente el 50%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de las colonias que se forman cuando las células se colocan sobre una capa de células alimentadoras *in vitro* tienen una morfología similar a una esfera de 7, 6, 5, 4 o 3. En otras realizaciones, la población de células ES de rata o una población de células de una línea celular de ES de rata dada (que no han sufrido una modificación genética específica y/o tienen una modificación genética específica) tienen al menos aproximadamente 50% a 95%, aproximadamente 60% a 90%, aproximadamente 60% a 95%, aproximadamente 60% a 85%, aproximadamente 60% a 80%, aproximadamente 70% a 80%, aproximadamente 70% a 85%, aproximadamente 70% a aproximadamente 90%, aproximadamente 70% a aproximadamente 95%, aproximadamente 70% a aproximadamente 100%, aproximadamente 80% a aproximadamente 100%, aproximadamente 80% a aproximadamente 95%, aproximadamente 80% a aproximadamente 90%, aproximadamente 90% a aproximadamente 100%, aproximadamente 90% a aproximadamente 99%, aproximadamente 90% a aproximadamente 98%, aproximadamente 90% a aproximadamente 97%, aproximadamente 90% a aproximadamente 95% de las colonias que se forman cuando las células se colocan en una capa de células alimentadoras *in vitro* tienen una morfología similar a una esfera de 7, 6, 5, 4 o 3.

Una célula ES de rata dada, empleada en los diversos métodos y composiciones proporcionadas en el presente documento puede ser una célula ES de rata macho (XY), una población masculina (XY) de células ES de rata, o una línea celular ES de rata macho (XY). En otras realizaciones, una población de células ES de rata o una línea celular ES de rata empleada en el presente documento puede ser una célula ES de rata hembra (XX), una población femenina (XX) de células ES de rata, o una línea celular ES de rata hembra (XX). Cualquier célula ES de rata, población de células ES de rata o línea celular ES de rata puede comprender la euploidía y/o diploidía como se describió anteriormente.

Las diversas células ES de rata empleadas en los métodos y composiciones pueden ser de cualquier cepa de rata, incluyendo, 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 como Fisher F344 o Fisher F6. Las diversas células ES de rata también pueden obtenerse a partir de una cepa derivada de una mezcla de dos o más cepas citadas anteriormente. En una realización, la célula ES de rata se deriva de una cepa seleccionada de una cepa DA y una cepa ACI. En una realización específica, la célula ES de rata se deriva de una cepa ACI. La cepa de rata ACI se caracteriza por ser agutí negro, con vientre y pies blancos y un haplotipo RT1av1. Dichas cepas están disponibles en una variedad de fuentes, incluidos los Laboratorios Harlan. En otras realizaciones, las diversas células ES de rata son de una cepa de rata Dark Agouti (DA), que se caracteriza por tener una capa de agutí y un haplotipo RT1av1. Dichas ratas están disponibles en una variedad de fuentes, incluidos Charles River y Harlan Laboratories. En una realización adicional, las diversas células ES de rata empleadas en el presente documento son de una cepa de rata consanguínea.

En realizaciones específicas, la línea celular ES de rata es de una rata ACI y comprende la célula ES rata ACI.G1 como se describe en detalle en la Solicitud de Patente de los Estados Unidos No. 14/185,103, presentada el 20 de febrero de 2014. En otra realización, la línea celular ES de rata es de una rata DA y comprende la línea celular ES de rata DA.2B o la línea celular ES de rata DA.2C como se describe en detalle en la Solicitud de Patente de los Estados Unidos No. 14/185,103, presentada el 20 de febrero de 2014.

Una célula ES de rata dada proporcionada en el presente documento puede obtenerse a partir de un embrión de rata en diversas etapas del desarrollo del embrión de rata. Los embriones de rata empleados para derivar las células ES de rata pueden ser un embrión en etapa de mórula, un embrión en etapa de blastocisto o un embrión de rata en una etapa de desarrollo entre un embrión en etapa de mórula y un embrión en etapa de blastocisto. Por lo tanto, en realizaciones específicas, el embrión de rata empleado está en o entre las etapas de Witschi de 5 y 7. En otras realizaciones, el embrión de rata empleado está en la etapa de Witschi 5, 6 o 7.

En una realización, la célula ES de rata se obtiene de un blastocisto de rata. En otras realizaciones, la célula ES de rata se obtiene de un blastocisto de una rata superovulada. En otras realizaciones, las células ES de rata se obtienen de un embrión en estadio de 8 células, que luego se cultivan *in vitro* hasta que se convierte en un estadio de blastocistema de mórula, un embrión entre los estadios 5 y 7 de Witschi, o en un embrión en la etapa 5, 6 o 7 de Witschi. En ese momento, los embriones se colocan en placas. Los embriones en etapa de mórula comprenden una bola compacta de células sin cavidad interna. Los embriones en etapa de blastocisto tienen una cavidad interna visible (el blastocoel) y contienen una masa celular interna (ICM). Las células ICM forman células ES.

B. Derivación y propagación de células madre embrionarias de rata (ES)

Los métodos de derivación y propagación de células madre embrionarias de rata son conocidos en la técnica y se describen, por ejemplo, en la Solicitud de Patente de los Estados Unidos No. 14/185,703, presentada el 20 de febrero de 2014. En realizaciones específicas, tales métodos comprenden (a) proporcionar un cultivo *in vitro* que comprende una capa de células alimentadoras y una población de células madre embrionarias (ES) de rata aisladas; (b) cultivo *in vitro* en condiciones que son suficientes para mantener la pluripotencia y/o totipotencia de la célula ES de rata aislada. Tales métodos permiten de este modo la propagación de una población de células ES de rata y/o una línea celular de ES de rata.

Se proporcionan métodos para cultivar una línea de células madre embrionarias de rata. Dichos métodos comprenden cultivar *in vitro* una capa de células alimentadoras y una línea celular ES de rata, en donde las condiciones de cultivo mantienen la pluripotencia de las células ES de rata y comprenden un medio que tiene un factor inhibidor de la leucemia de ratón (LIF) o una variante activa o fragmento de la misma. Los métodos pueden comprender además pasar y cultivar *in vitro* las células de la línea celular ES de rata, en donde cada cultivo posterior *in vitro* comprende cultivar las células ES de rata en la capa de células alimentadoras en condiciones que mantienen la pluripotencia de las células ES de rata y comprende un medio con ratón LIF o una variante activa o fragmento de la misma.

Los medios de cultivo empleados en los diversos métodos y composiciones pueden mantener las células ES de rata. Los términos "mantener" y "mantenimiento" se refieren a la conservación estable de al menos una o más de las características o fenotipos de las células ES de rata que se describen aquí. Dichos 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 madre de rata descritas en el presente documento. El término "mantener" también puede abarcar la propagación de células madre, o un aumento en el número de células madre que se cultivan. El término contempla además las condiciones de cultivo que permiten que las células madre permanezcan pluripotentes, mientras que las células madre pueden o no continuar dividiéndose y aumentando en número.

El término "célula de alimentación" o "capa de células de alimentación" comprende un cultivo de células que crecen *in vitro* y secretan al menos un factor en el medio de cultivo que se utiliza para apoyar el crecimiento de otra célula de interés en el cultivo. Las células alimentadoras empleadas aquí ayudan a mantener la pluripotencia de las células ES de rata, y en realizaciones específicas, una o más de las otras características o fenotipos descritos aquí. Se pueden usar diversas células alimentadoras, incluidos, por ejemplo, fibroblastos embrionarios de ratón, incluidos fibroblastos embrionarios de ratón obtenidos entre el día 12 y el día 16 del embarazo. En realizaciones específicas, la capa de células alimentadoras comprende una monocapa de fibroblastos embrionarios de ratón inactivados mitóticamente (MEF).

Los cultivos *in vitro* de las células ES de rata comprenden además una cantidad eficaz de factor inhibidor de leucemia (LIF) o una variante activa o fragmento de la misma. El factor inhibidor de la leucemia (LIF) pertenece a la familia de receptores IL-6. LIF se une a un receptor de membrana heterodimérico formado por una subunidad específica de LIF, gp190 o LIFR, y la subunidad gp130, que se comparte con los otros miembros de la familia IL-6. LIF inhibe la diferenciación de células madre embrionarias en ratones y contribuye a la auto renovación de las células madre. El LIF humano y el ratón comparten un 79% de homología de secuencia y exhiben actividad de especies cruzadas. La LIF de rata (rtLIF) es una proteína de 22.1 kDa que contiene 202 residuos de aminoácidos que exhibe una identidad de secuencia de aminoácidos del 91% con la LIF murina (Takahama et al. 1998). Hay seis posibles sitios de glicosilación ligada a asparagina (N-glicosilación) que se conservan entre el polipéptido LIF de las diversas especies y un sitio adicional de Asn150 que es específico para LIF de rata. La estructura terciaria del LIF del ratón y su función se describen con más detalle en Aikawa et al. (1998) Biosci. Biotechnol. Biochem. 62 1318-1325 and Senturk et al. (2005) Immunology of Pregnancy, editor Gil Mor., Patente de los Estados Unidos No. 5,750,654 and D P Gearing (1987) EMBO Journal 1987-12-20. Una secuencia LIF de ratón parcial se informa en el sitio web SwissProt con el número de acceso P09056.

La actividad LIF del ratón se evalúa por su capacidad para inducir la diferenciación de las células de leucemia mieloide M1. La actividad específica es 1 x 10^6 unidades/ml (Cat. No. 03-0011 de Stemgent) y 1 x 10^7 unidades/ml (Cat. No. 03-0011-100 de Stemgent), donde 50 unidades se definen como la cantidad del LIF de ratón requerido para inducir la diferenciación en el 50% de las colonias M1 en 1 ml de medio. Véase, también, Williams, R.L. et al. (1988) Nature 336: 684-687.; Metcalf, D. et al. (1988) Leukemia 2: 216-221; Niwa, H. et al. (2009) Nature 460: 118-122; Xu, J. et al. (2010) Cell Biol Int. 34: 791-797; Fukunaga, N. et al. (2010) Cell Reprogram. 12: 369-376; y Metcalf D. (2003) Stem Cells 21: 5-14. Una "cantidad efectiva de LIF" comprende una concentración de LIF que permite que las células ES de rata de un cultivo *in vitro* permanezcan en un estado pluripotente indiferenciado. Diversos marcadores que se pueden usar para analizar las células que permanecen en un estado pluripotente se discuten en otra parte de este documento.

El polipéptido LIF empleado en los diversos métodos y composiciones proporcionados en el presente documento puede ser de cualquier organismo, incluido un mamífero, un roedor, un ser humano, una rata o un ratón. En una realización, el polipéptido LIF es de un ratón. En otras realizaciones adicionales, el polipéptido LIF de ratón comprende la secuencia de aminoácidos expuesta en el número de acceso SwissProt: P09056 y también se expone en la SEQ ID NO: 9.

En otras realizaciones, se puede usar una variante o fragmento activo del polipéptido LIF de ratón como se establece en la SEQ ID NO: 9 o en el número de acceso SwissProt: P09056. Dichas variantes y fragmentos activos (incluidas las variantes activas que tienen al menos 75%, 80%, 85% 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99% de identidad de secuencia con la SEQ ID NO: 9 se analiza con más detalle en otra parte del presente documento.

El polipéptido LIF o la variante activa o fragmento de la misma puede proporcionarse al cultivo *in vitro* de diversas maneras. En una realización, la cantidad eficaz del polipéptido LIF o la variante activa o fragmento de la misma se añade a los medios de cultivo. En otras realizaciones, las células alimentadoras se han modificado genéticamente para sobreexpresar el polipéptido LIF o la variante activa o fragmento de la misma. Dichas células alimentadoras incluyen células alimentadoras preparadas a partir de fibroblastos de ratón DIA-M tratados con mitomicina o irradiados

con rayos gamma que expresan LIF asociado a la matriz. El método para generar y utilizar dichas células alimentadoras modificadas genéticamente se puede encontrar, por ejemplo, en See, Buehr et al. (2003) Biol Reprod 68: 222-229, Rathjen et al. (1990) Cell 62 1105-1115, y Buehr et al. (2008) Cell 135: 1287-1298. El LIF heterólogo expresado en las células alimentadoras puede ser del mismo organismo que las células alimentadoras o de un organismo diferente al de la célula alimentadora. Además, el LIF heterólogo expresado en las células alimentadoras puede ser del mismo organismo o de un organismo diferente al de las células ES que la capa de alimentación soporta.

En aún otras realizaciones, las células alimentadoras empleadas en los diversos métodos divulgados en el presente documento no están modificadas genéticamente para expresar un polipéptido LIF heterólogo o una variante activa o fragmento de la misma. Por lo tanto, en realizaciones particulares, la monocapa de fibroblastos embrionarios de ratón inactivados mitóticamente empleado en los métodos no se ha modificado genéticamente para expresar un polipéptido LIF heterólogo.

En otras realizaciones, el polipéptido LIF o la variante activa o fragmento de la misma se añade a los medios de cultivo. Cuando se agrega LIF a los medios de cultivo, el LIF puede ser de cualquier organismo, incluido un mamífero, un roedor, un humano, una rata o un ratón. En una realización, el LIF presente en los medios de cultivo es de un ratón. En otras realizaciones adicionales, el polipéptido LIF de ratón comprende la secuencia de aminoácidos expuesta en la SEQ ID NO: 9. En otras realizaciones, puede usarse una variante o fragmento activo del polipéptido LIF de ratón como se expone en la SEQ ID NO: 9. Dichas variantes y fragmentos activos (incluidas las variantes activas que tienen al menos 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99% la identidad de secuencia a la SEQ ID NO: 9) se discute con más detalle en otra parte del presente documento.

En realizaciones específicas, las células ES de rata y las líneas celulares ES de rata proporcionadas en el presente documento mantienen la pluripotencia *in vitro* sin requerir señal LIF paracrina.

En realizaciones específicas, LIF o una variante activa o fragmento de la misma está presente en los medios de cultivo en cualquier concentración que mantenga las células ES de rata. El polipéptido LIF o una variante activa o fragmento de la misma está presente en el medio de cultivo a aproximadamente 25U/ml a aproximadamente 50U/ml, a aproximadamente 50U/ml a aproximadamente 100U/ml, a aproximadamente 100U/ml a aproximadamente 125U/ml, a aproximadamente 125U/ml a aproximadamente 150U/ml, a aproximadamente 150U/ml a aproximadamente 175U/ml, a aproximadamente 175U/ml a aproximadamente 200U/ml, a aproximadamente 200U/ml a aproximadamente 225U/ml, a aproximadamente 225Um a aproximadamente 250U/ml, de aproximadamente 250U/ml a aproximadamente 300U/ml, de aproximadamente 300U/ml a aproximadamente 325U/ml, de aproximadamente 325U/ml a aproximadamente 350U/ml, de aproximadamente 350U/ml a aproximadamente 400U/ml, de aproximadamente 400U/ml a aproximadamente 425U/ml, a aproximadamente 425U/ml a aproximadamente 450U/ml, a aproximadamente 450U/ml a aproximadamente 475U/ml, a aproximadamente 475U/ml a aproximadamente 500U/ml, a aproximadamente 75U/ml a aproximadamente 500U/ml o mayor. En otras realizaciones, el polipéptido LIF o variante activa o fragmento de la misma está presente en el medio de cultivo a aproximadamente 25U/ml a aproximadamente 50U/ml, a aproximadamente 25U/ml a aproximadamente 100U/ml, a aproximadamente 75U/ml a aproximadamente 125U/ml , de aproximadamente 50U/ml a aproximadamente 150U/ml, de aproximadamente 90U/ml a aproximadamente 125U/ml, de aproximadamente 90U/ml a aproximadamente 110U/ml, de aproximadamente 80U/ml a aproximadamente 150U/ml, o de aproximadamente 80U/ml hasta aproximadamente 125U/ml. En una realización específica, el polipéptido LIF o variante activa o fragmento de la misma está presente en el medio de cultivo a aproximadamente 100 U/ml.

Cuando se emplea LIF de ratón, el polipéptido LIF de ratón o variante activa o fragmento de la misma está presente en el medio de cultivo en cualquier concentración que mantenga las células ES de rata. El polipéptido LIF de ratón o variante activa o fragmento de la misma está presente a aproximadamente 25U/ml a aproximadamente 50U/ml, a aproximadamente 50U/ml a aproximadamente 100U/ml, a aproximadamente 100U/ml a aproximadamente 125U/ml, a aproximadamente 125U/ml a aproximadamente 150U/ml, aproximadamente 150U/ml a aproximadamente 175U/ml, aproximadamente 175U/ml a aproximadamente 200U/ml, aproximadamente 200U/ml a aproximadamente 225U/ml, aproximadamente 225U/ml a aproximadamente 250U/ml, aproximadamente 250U/ml a aproximadamente 300U/ml, a aproximadamente 300U/ml a aproximadamente 325U/ml, a aproximadamente 325U/ml a aproximadamente 350U/ml, a aproximadamente 350U/ml a aproximadamente 400U/ml, a aproximadamente 400U/ml a aproximadamente 425U/ml, a aproximadamente 425U/ml a aproximadamente 450U/ml, a aproximadamente 450U/ml a aproximadamente 475U/ml, a aproximadamente 475U/ml a aproximadamente 500U/ml, a aproximadamente 75U/ml a aproximadamente 500U/ml o más. En otras realizaciones, el polipéptido LIF de ratón o variante activa o fragmento de la misma está presente a aproximadamente 25U/ml a aproximadamente 50U/ml, a aproximadamente 25U/ml a aproximadamente 100U/ml, a aproximadamente 75U/ml a aproximadamente 125U/ml, a aproximadamente 50U/ml a aproximadamente 150U/ml, a aproximadamente 90U/ml a aproximadamente 125U/ml, a aproximadamente 90U/ml a aproximadamente 110U/ml, a aproximadamente 80U/ml a aproximadamente 150U/ml, o aproximadamente a 80U/ml a aproximadamente 125U/ml. En una realización específica, el polipéptido LIF de ratón o variante activa o fragmento de la misma está presente en el medio de cultivo a aproximadamente 100U/ml.

Los medios de cultivo empleados mantienen células ES de rata. Como tal, en realizaciones específicas, los medios de cultivo empleados en los diversos métodos y composiciones mantendrán la pluripotencia de todas o la mayoría (es decir, más del 50%) de las células ES de rata en una línea celular durante un período de al menos 5, 10 o 15 pases.

En una realización, los medios de cultivo comprenden uno o más compuestos que ayudan a mantener la pluripotencia. En una realización, los medios de cultivo comprenden un inhibidor de la ruta MEK y un inhibidor del glucógeno sintasa quinasa-3 (GSK-3). También se describen ejemplos en los que los medios pueden comprender además componentes adicionales que ayudan a mantener las células ES, incluyendo, por ejemplo, inhibidores del receptor de FGF, inhibidores de ROCK y/o inhibidores de ALK (receptor de TGFb). Un ejemplo no limitante de un inhibidor del receptor de FGF incluye PD184352. Un ejemplo no limitante de un inhibidor de ROCK incluye Y-27632, y un ejemplo no limitativo de un inhibidor de ALK (receptor de TGFb) incluye A-83-01. En realizaciones, el medio 2i se usa con 10 uM ROCKi cuando se descongela el rESC crioconservado o cuando se vuelve a colocar el rESC después de la disociación con tripsina.

En otras realizaciones, los medios comprenden una combinación de inhibidores que consisten en un inhibidor de la ruta de MEK y un inhibidor del glucógeno sintasa quinasa-3 (GSK-3).

En una realización no limitativa, los medios de cultivo comprenden un inhibidor de GSK-3 que comprende CHIR99021 y/o un inhibidor de MEK que comprende PD0325901. En otras realizaciones, los medios comprenden una combinación de inhibidores que consisten en CHIR99021 y PD0325901. Cualquiera de estos compuestos se puede obtener, por ejemplo, de Stemgent. En realizaciones específicas, CHIR99021 está presente en los medios de cultivo a una concentración de aproximadamente 0.5 µ a aproximadamente 3 µM, aproximadamente 0.5 µ a aproximadamente 3.5 µM, aproximadamente 0.5 µM a aproximadamente 4 µM, aproximadamente 0.5 µM a aproximadamente 1 µM, aproximadamente 1 µM a aproximadamente 1.5 µM, aproximadamente 1.5 µM a aproximadamente 2 µM, aproximadamente 2 µM a aproximadamente 2.5 µM, aproximadamente 2.5 a aproximadamente 3 µM, 3 µM a aproximadamente 3.5 µM. En realizaciones adicionales, CHIR99021 está presente en los medios de cultivo a una concentración de aproximadamente 3 µM. En otras realizaciones, PD0325901 está presente en los medios de cultivo a una concentración de aproximadamente 0.4 µM a aproximadamente 1 uM, aproximadamente 0.4 µM a aproximadamente 1.5 uM, aproximadamente 0.4 µM a aproximadamente 2 µM, aproximadamente 0.4 µM a aproximadamente 0.8 µM, 0.8 µM a aproximadamente 1.2 µM, aproximadamente 1.2 a aproximadamente 1.5 µM. En realizaciones adicionales, PD0325901 está presente en los medios de cultivo a una concentración de aproximadamente 1 µM. En realizaciones específicas, CHIR99021 está presente en los medios de cultivo a una concentración de aproximadamente 3 µM y PD0325901 está presente en una concentración de aproximadamente 1 µM.

En una realización no limitativa, los medios de cultivo empleados en los diversos métodos y composiciones divulgados en el presente documento son medios 2i que comprenden: medio basal DMEM/F12 (a una concentración de 1x (50%)); medios neurobasales (a una concentración de 1x (50%)); penicilina/estreptomicina (a una concentración del 1%); L-glutamina (a una concentración de 4 mM); 2-mercaptoetanol (a una concentración de 0.1 mM); suplemento de N2 (a una concentración de 1x); Suplemento B27 (a una concentración 1x); LIF (a una concentración de 100U/ml); PD0325901 (inhibidor de MEK) (a una concentración de 1 µM) y CHIR99021 (inhibidor de GSK) (a una concentración de 3 µM).

Los medios adicionales que pueden emplearse incluyen los descritos en Li et al. (2008) Cell 135: 1299-1310, Yamamoto et al. (2012) Transgenic Rats 21: 743-755, Ueda et al. (2008) PLoS ONE 3(6): e2800, Meek et al. (2010) PLoS ONE 4 (12): e14225; Tong et al. (2010) Nature 467: 211-213; Publicación de Patente de los Estados Unidos 2012/0142092, Buehr et al. (2008) Cell 135: 1287-1298, Li et al. (135) Cell 1299-1310, cada una de las cuales se incorpora aquí como referencia en su totalidad. Cuando se emplean dichos medios, la concentración y la fuente de LIF se pueden modificar como se describe en el presente documento. En realizaciones específicas, los diversos medios de cultivo se utilizan en combinación con LIF de ratón o una variante activa o fragmento de la misma, e incluso en realizaciones adicionales, los diversos medios de cultivo comprenden una LIF de ratón o una variante activa o fragmento de la misma a una concentración de aproximadamente 50U/ml a aproximadamente 100U/ml, aproximadamente 50U/ml a aproximadamente 150U/ml, o aproximadamente 100U/ml.

La temperatura de los cultivos de células ES de rata, tanto para la producción de la línea celular ES como para el cultivo y mantenimiento de la línea ES, se realiza típicamente a aproximadamente 35ºC hasta aproximadamente 37.5ºC. En una realización específica, la temperatura es de 37.0°C. El cultivo se realiza típicamente al 7.5% de CO_2.

7. Identidad de secuencia

Los métodos y composiciones proporcionados en el presente documento emplean una variedad de diferentes componentes del sistema de integración genómica direccionada (es decir, agentes de nucleasa, sitios de reconocimiento, ácidos nucleicos de inserción, polinucleótidos de interés, vectores de direccionamiento, marcadores de selección y otros componentes). A lo largo de la descripción, se reconoce que algunos componentes del sistema de integración genómica diana pueden tener variantes y fragmentos activos. Tales componentes incluyen, por ejemplo, agentes de nucleasa (es decir, agentes de nucleasa modificados por ingeniería genética), sitios de reconocimiento de agentes de nucleasa, polinucleótidos de interés, sitios diana y brazos de homología correspondientes del vector de direccionamiento. La actividad biológica para cada uno de estos componentes se describe en otra parte de este documento.

Como se usa en el presente documento, "identidad de secuencia" o "identidad" en el contexto de dos polinucleótidos o secuencias polipeptídicas 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 cero 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).

Como se usa en el presente documento, "porcentaje de identidad de secuencia" significa 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 comprender adiciones o eliminaciones (es decir, brechas) como en comparación con la secuencia de referencia (que no incluye adiciones o eliminaciones) para una alineación óptima de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las que se produce 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 proporcionados en el presente documento se refieren al valor obtenido al usar GAP Versión 10 utilizando los siguientes parámetros: % de identidad y % de similitud para una secuencia de nucleótidos usando GAP Peso de 50 y Longitud Peso de 3, y el nwsgapdna.cmp matriz de puntuación % de identidad y % de similitud para una secuencia de aminoácidos usando GAP Peso de 8 y Longitud Peso de 2, y la matriz de puntuación BLOSUM62; o cualquier programa equivalente de los mismos. "Programa equivalente" significa cualquier programa de comparación de secuencias que, para cualquiera de las dos secuencias en cuestión, genere una alineación que tenga coincidencias idénticas de residuos de nucleótidos o aminoácidos y un porcentaje idéntico de identidad de secuencia en comparación con la alineación correspondiente generada por GAP Versión 10.

A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que entiende comúnmente un experto en la técnica a la que pertenece esta invención. Aunque cualquier método y materiales similares o equivalentes a los divulgados en el presente documento también se pueden usar en la práctica o prueba de la invención descrita, ahora se describen los métodos y materiales preferidos.

Se debe tener en cuenta que tal como se usa en el presente documento y en las reivindicaciones adjuntas, las formas singulares "un/una", "y", y "el/la" incluyen referencias plurales a menos que el contexto indique claramente lo contrario. Todos los términos técnicos y científicos utilizados aquí tienen el mismo significado.

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

Ejemplos

Los siguientes ejemplos se presentan para proporcionar a los expertos en la técnica una divulgación y descripción completas de cómo hacer y usar la presente invención. Se han realizado esfuerzos para garantizar la precisión con respecto a las cifras utilizadas (por ejemplo, cantidades, temperatura, etc.), pero se deben tener en cuenta algunos errores y desviaciones experimentales. A menos que se indique lo contrario, las partes son partes en peso, el peso molecular es el peso molecular promedio en peso, la temperatura está en grados centígrados y la presión está en o cerca de la atmosférica.

Ejemplo 1. Derivación y caracterización de células ES de rata

1.1. Caracterización de células ES de rata

Como se muestra en la figura 1, las ESC de rata crecen como colonias esféricas compactas, que de manera rutinaria se desprenden y flotan en la placa (primer plano, figura 7). Las ESC de rata expresan marcadores de pluripotencia que incluyen Oct-4 (figura 2A) y Sox2 (figura 2B), y expresan altos niveles de fosfatasa alcalina (figura 3, panel izquierdo). El cariotipo para la línea DA.2B es 42X,Y (figura 3, panel derecho). Las ESC de rata a menudo se vuelven

tetraploides; por lo tanto, las líneas fueron preescrutadas contando los cromosomas en metafase; las líneas con recuentos en su mayoría normales fueron luego formalmente cariotipo.

Los blastocistos ACI se obtuvieron de hembras superovuladas obtenidas comercialmente. Los blastocistos DA se cultivaron a partir de embriones de 8 células congelados obtenidos comercialmente. Las zonas pelucidae se eliminaron con Acid Tyrodes; y los blastocistos se colocaron en placas sobre MEF inactivados mitóticamente. Los crecimientos se seleccionaron y expandieron utilizando métodos estándar. Todos los blastocistos se cultivaron en placas, se cultivaron y se expandieron utilizando medios 2i (Li et al. (2008) Germline competent embryonic stem cells derived from rat blastocysts, Cell 135: 1299-1310).

	Tabla 1. Derivación de células ES de rata	
	ACI	DA
Fuente de embriones	Blastocistos (superovulación)	Embriones congelados de 8 células cultivados a blastocistos
Blastocistos sembrados:	107	22
Crecimientos:	32 (30% de blastos)	10 (45% de blastos)
Líneas:	16 (50% de crecimientos)	9 (90% de crecimientos)
Cariotipificado :	3; todos los 42X,Y	6: 3 42X,X
		3 42X,Y
GLT validado	1 (ACI.G1)	1 42X,X (DA.2C)
		1 42X,Y (DA.2B)

1.2: Producción de ratas

Las ratas quiméricas se produjeron mediante inyección de blastocistos y transmisión del genoma ESC de rata. En la figura 8 se muestran las quimeras producidas por microinyección de blastocistos utilizando ESC parentales ACI.G1 de rata. Las crías F1 agutí con compañeros de camada albinos, engendrados por la quimera ACI/SD marcadas con un asterisco (*) en la figura 8 se muestran en la figura 9.

Transmisión de la línea germinal de la ESC de rata parental.

Se evaluaron tres líneas de ESC de euploides de rata para la pluripotencia mediante microinyección en blastocistos SD albinos. Las quimeras se identificaron por el color del pelaje agutí, lo que indica la contribución de la rata ESC. Para cada línea, la mayoría de las quimeras transmitieron el genoma rESC a la descendencia F1 (Tabla 2).

Tabla 2. Transmisión de la línea germinal de rESC parental					
Línea	Quimeras criadas	Transmisores de línea germinal	Cachorros totales de quimeras GLT	Cachorros derivados de rESC	Eficiencia GLT (%)
ACI.G1	5	3(60%)	103	11	11
DA.2B	5	4(80%)	129	11	9
DA.2C (XX)	3	2(66%)	45	7	16

1.3.: Derivación de células madre embrionarias de rata.

Protocolo de superovulación, ratas.

Día 0: inyectado con suero de yegua preñada: IP, 20 U (0.4 ml).

Día 1: no hay acción.

Día 2: (46 horas más tarde): inyectado con hCG, IP, 50 U (1 ml).

- establecimiento de apareamientos solteros femeninos.

Día 3: Conexiones comprobadas. Las hembras fueron conectadas. Este es el día 0.5.

Día 6 (e3.5): Hembras sometidas a eutanasia y embriones enjuagados.

Protocolo de derivación celular ES (superovulación).

Día 0:

1) Rata hembra sometida a eutanasia con CO_2.

2) Abdomen ventral frotado con etanol al 70%; utilizando unas tijeras, se abre la pared del cuerpo ventral para exponer las vísceras.

3) Se diseccionaron los oviductos y los cuernos uterinos y se colocaron en una placa de cultivo de tejidos que contenía medio caliente N2B27. Se lavó la mayor cantidad de sangre posible y se transfirió a una nueva placa con N2B27.

4) Utilizando una jeringa de 1 ml y una aguja roma de 27 g, se lavó con el medio a través de los cuernos uterinos y oviductos para expulsar los blastocistos al medio.

5) Se recogieron los blastocitos con una pipeta bucal y se transfirieron a una placa de cultivo de embriones que contenía KSOM + 2i (1 µMP00325901, 3 µM CHIR99021). KSOM es un medio de cultivo producido por Millipore. El número de catálogo es MR-106-D.

6) Cultivado durante la noche a 37°; 7.5% de CO_2.

Protocolo de derivación celular ES (embriones congelados).

Día 0:

1) Se descongelaron los embriones congelados de 8 células (obtenidos comercialmente) en medio M2. Se cultiva 10 minutos a temperatura ambiente.

2) Transferido a KSOM + 2i y cultivo durante la noche.

Protocolo de derivación celular ES (igual para ambos)

Día 1:

1) Se transfieren los embriones soemtidos a cavitación a medio 2i y cultivo durante la noche.

2) Cultivo continuo de embriones sin cavitación en KSOM + 2i

Día 2:

1) Se transfieren todos los embriones restantes al medio 2i (ya sea que se hayan sometido a cavitación o no).

2) Cultivado durante la noche; Se continuó cultivando los embriones anteriores en medio 2i.

Día 3:

1) Transferir los embriones durante 30-60 segundos con Acid Tyrodes para eliminar la zona pelúcida.

2) Embriones lavados 3x en medio 2i para eliminar los Acid Tyrodes.

3) Se depositó cada embrión en un pozo separado de una placa de alimentación de 96 pozos (el pozo contiene una monocapa de fibroblastos embrionarios de ratón inactivados mitóticamente (MEF).

4) Cultivado durante la noche en medio 2i.

Días 4-5:

1) Embriones en placa monitorizados para detectar la presencia de un crecimiento (una masa de células amorfas no diferenciadas). Las excrecencias están listas para la transferencia cuando tienen aproximadamente el doble del tamaño del embrión sembrado.

2) Cada día: eliminar los medios gastados con una micropipeta y reemplazarlos con medios 2i nuevos.

3) Crecimiento transferido a nuevos pozos de alimentación:

a. Se eliminaron los medios gastados y se lavó suavemente con PBS.

b. Se eliminó el PBS y se añadieron 30 µl de tripsina al 0.05%; incubar durante 10 minutos.

c. Se detuvo la reacción de tripsina al agregar 30 µl de 2i + 10% de FBS.

d. Separar suavemente las células con una micropipeta y transferir todo el contenido del pozo a un nuevo pozo en una placa de alimentación de 24 pozos. Este fue el pase 1 (P1).

e. Se cultiva durante la noche en medio 2i.

Día 5-8: (el tiempo depende de cuán rápido se expande cada línea)

1) Medios modificados cada día (medios 2i) y monitorizados para detectar la presencia de colonias con una morfología ESC.

2) Cuando aparecen colonias, continuar el cultivo hasta que las colonias se expandan a ~50% de confluencia.

3) Colonias tripsinizadas y pasadas como antes; sembrado en alimentadores, 1 pozo por línea, en un plato de 6 pozos. Este fue el pase 2 (P2).

En marcha:

1) Alimentación continua y monitorizo de cada línea hasta aproximadamente 50% de confluencia.

2) Células tripsinizadas como es usual.

3) se inactivó la tripsina con 2i + 10% de FBS; formación de pellas con las células por centrifugación (5', 1200 rpm en la centrífuga de mesa Beckman-Coulter).

4) Aspirar el sobrenadante y resuspender suavemente las células en 400 µl de medio de congelación (70% 2i, 20% FBS, 10% DMSO).

5) Distribuir las células en 2 viales y congelar a -80°. Este fue el pase 3 (P3).

6) Para el almacenamiento a largo plazo, transferir los viales al almacenamiento de N_2 líquido.

Los medios 2i se prepararon como sigue en la Tabla 3.

Reactivo	Vendedor	Concentración
Medio basal DMEM/F12	Invitrogen/Life Technologies	1x
Medios neurobasales		1x
Penicilina/Estreptomicina	Invitrogen/Life Technologies	1%
L-Glutamina	Invitrogen/Life Technologies	4 mM
2-Mercaptoetanol	Invitrogen/Life Technologies	0.1 mM
Suplemento N2	Invitrogen/Life Technologies	1x
Suplemento B27	Invitrogen/Life Technologies	1x
LIF	Millipore	100 U/ml
PD0325901 (Inhibidor de MEK).	Stemgent	1 uM
CHIR99021 (inhibidor de GSK).	Stemgent	3 uM

Materiales: Gonadotropina sérica de yegua embarazada (PMSG)

Gonadotropina coriónica de orina en embarazo humana (HCG)

Ratas hembras (5-12 semanas de edad)

Ratas machos (12 semanas a 8 meses de edad), una por jaula.

Jeringas/agujas

Sala de animales con luces encendidas 6:00-18:00.

Procedimiento:

Día 1: 8:00-10:00 AM

Inyectar hembras con 20 UI de PMSG (0.4 ml), IP

Desechar el PMSG no utilizado.

Día 3: 8:00-10:00 AM (48 horas después de la inyección de PMSG)

Inyectar hembras con 50 UI de HCG (1 ml), IP

Colocar una hembra por macho en jaula de apareamiento.

Desechar la HCG no utilizada.

Día 4: 8:00-10:00 AM (24 horas después de la inyección de HCG)

Compruebe las hembras para los enchufes.

Proveedores de hormonas

PMSG: Sigma #G-4877 (1000 UI). Resuspender en PBS a un final [ ] de 50 UI/ml. Almacenar a -20° en alícuotas de 1 ml.

HCG: Sigma #CG-5 (5000 UI). Resuspender en PBS a un final [ ] de 50 UI/ml. Almacenar a -20° en alícuotas de 1 ml.

1.4.: Cariotipificar líneas de células madre embrionarias de rata

Las líneas celulares ES de rata generadas en el presente documento fueron cariotipificadas, y los resultados se resumen en las Tablas 4-7.

Tabla 4

Resultados de cariotipificación ACI.G1	Número de células
Número de células cariotipificadas	7
Número de células analizadas	20
Cantidad de 42 células XY	18
Número de células anormales	2
40, XY, -5, -9	1
41, XY, -14	1
42, XY	18
Otras notas: A dos células analizadas les faltaban autosomas diferentes, lo que puede ser una ocurrencia esporádica debido a un artefacto técnico. El 90% de las células analizadas tenían un cariotipo normal 42, XY.	
La Figura 4 proporciona una fotografía que muestra el análisis del número de cromosoma de la línea celular ES de rata ACI.G1.	

Tabla 5

Resultados de cariotipificación DA.2B	Número de células
Número de células cariotipificadas	6
Número de células analizadas	20
Cantidad de 42 células XY	20
Número de células anormales	0

Resultados de cariotipificación DA.2B	Número de células
42, XY	20
. Otras notas: Todas las células analizadas tenían un cariotipo diploide 42, XY normal.	
La Figura 5 proporciona una fotografía que muestra el análisis del número de cromosoma de la línea celular ES.2 de rata DA.2B.	

Tabla 6

Resultados de cariotipificación DA.C2	Número de células
Número de células cariotipificadas	5
Número de células analizadas	20
Cantidad de 42 células XY	20
Número de células anormales	0
42, XX	
Otras notas: el 100% de las células analizadas tenían cariotipo de rata XX hembra normal.	
La Figura 6 proporciona una fotografía que muestra el análisis del número de cromosoma de la línea celular ES de rata DA.C2.	

Tabla 7.

Cepa	Blastocistos sembrados	Líneas establecidas	Líneas cariotipificadas	Cariotipos
BN x SD F1	41	8(20%)	5	Todas las líneas fueron de alto % de poliploides complejos.
ACI	27	16 (60%)	3	G1: 90% 42 XY; otros fueron 70-85% euploides
DA	20	9 (45%)	6	2B: 100% 42 XY; 2C: 100% 42 XX; otros eran 95-100% euploides
F344	4	1 (25%)	0	
Totales	92	34 (37%)		

1.5.: Electroporación del vector en células madre embrionarias de rata

1. Pase de células de rata ES 24-48 horas antes de la electroporación.

2. Se cambiaron los medios a RVG2i + ROCKi (10 µM Y-27632) 24 horas antes de la electroporación

3. Medios modificados 30' antes de la tripsinización.

4. Alícuotas de ADN para electroporación.

5. Permitir que el ADN se caliente a temperatura ambiente durante >10 min.

6. ADN calentado durante 5' a 62°C. Colocar el ADN en hielo.

7. Células tripsinizadas:

a. Recolectar colonias flotantes. Placa lavada para recoger tantos flotadores como sea posible.

b. Colonias en pellas: 3' a 750 rpm.

c. Pellas lavados 1x con 5-10 ml de PBS y centrifugado/pella

d. Sobrenadante aspirado; agregar 500λ de tripsina, 0.05% + 1% de suero de pollo.

i. No se agruparon más de 1 placa de 10 cm de colonias por tubo. Si hay demasiadas colonias empaquetadas en el fondo del tubo durante la tripsinización, se agruparán y la mayoría de las células se perderán.

e. 4' a 37°. Colonias pipeteadas varias veces para minimizar la aglomeración.

f. Pasos repetidos 1-2 X: 4' a 37°.

g. Se inactivó la tripsina con 500λ RVG2i + 10% FBS.

8. Células en pellas: 5' a 1200 rpm.

9. Resuspender las células en 10 ml de PBS. Contar dos alícuotas de 20λ para determinar el número total de células.

10. Células en pellas (5'/1200 rpm); calcular el número total de células y el volumen total de resuspensión para lograr la concentración celular correcta (objetivo #/75 µl de regulador EP).

11. Resuspender en un volumen mínimo de regulador EP; medir el volumen total y ajustar al volumen deseado con el regulador EP. El regulador de electroporación es vendido por Millipore. El catálogo # es ES-003-D. Véase, Valenzuela et al. (2003) Nature Biotechnology 21: 652-659.

12. Añadir células 75λ a ADN 50λ; transferir la solución de células/ADN 125λ a un pozo de una cubeta BTX de 48 pozos.

a. Se llenaron los pozos vacíos en la misma columna con regulador EP 125λ.

13. Pulsó la cubeta una vez en el electroporador BTX:

a. Ajustes: 400V; Ω; 100 µF (la configuración puede variar)

14. Colocar la cubeta en hielo durante 15' para recuperarla.

15. Se eliminaron las células en 5 ml de RVG2i + 10 µM ROCKi.

16. Se añadieron a una placa de 15 cm con 20 ml de RVG2i + 10µM ROCKi. La placa tiene 2x MEF neoR (u otros MEF según el proyecto). El marcador seleccionable neoR es el gen de la neomicina fosfotransferasa (neo) de Beck et al. (1982) Gene, 19: 327-36 o en la Patente de los Estados Unidos No, 7,205,148 o 6,596,541.

17. Incubar a 37°. Comenzar la selección 48 horas más tarde.

El inhibidor de ROCK utilizado fue Y-27632.

1.6: Selección de una modificación genética direccionada en una célula madre embrionaria de rata.

1. Células pasadas durante 24-48 horas antes de la electroporación.

2. Medios modificados a RVG2i + ROCKi (10 µM Y-27632) 24 h. antes de la electroporación

3. Medios modificados 30' antes de la tripsinización.

4. Alícuotas de ADN para electroporación.

5. Permitir que el ADN se caliente a temperatura ambiente durante >10 min.

6. ADN calentado durante 5' a 62°C. Colocar el ADN en hielo.

7. Células tripsinizadas:

a. Recolectar colonias flotantes. Placa lavada para recoger tantos flotadores como sea posible.

b. Colonias en pellas: 3' a 750 rpm.

c. Pellas lavadas 1x con 5-10 ml de PBS y centrifugado/pella

d. Sobrenadante aspirado; agregar 500λ de tripsina, 0.05% + 1% de suero de pollo.

i. No se agruparon más de 1 placa de 10 cm de colonias por tubo. Si hay demasiadas colonias empacadas en el fondo del tubo durante la tripsinización, se agruparán y la mayoría de las células se perderán.

e. 4' a 37°. Colonias pipeteadas varias veces para minimizar la aglomeración

f. Repetido 1-2 X: 4' a 37°.

g. Se inactivó la tripsina con 500λ RVG2i + 10% FBS.

8. Células en pellas: 5' a 1200 rpm.

9. Células resuspendidas en 10 ml de PBS. Contar dos alícuotas de 20λ para determinar el número total de células.

10. Células en pellas (5'/1200 rpm); calcular el número total de células y el volumen total de resuspensión para lograr la concentración celular correcta (objetivo #/75 µl de regulador EP).

11. Resuspender en un volumen mínimo de regulador EP; volumen total medido y ajustado al volumen objetivo con regulador EP.

12. Se agregaron células de 75λ a 50λ de ADN; transferir la solución de células/ADN 125λ a un pozo de una cubeta BTX de 48 pozos.

a. Se llenaron los pozos vacíos en la misma columna con regulador EP 125λ.

13. Pulsó la cubeta una vez en el electroporador BTX:

a. Ajustes: 400V; 100 µF (la configuración puede variar)

14. Colocar la cubeta en hielo durante 15' para recuperarla.

15. Se retiraron las células en 5 ml de RVG2i + ROCKi 10 µM.

16. Añadido a una placa de 15 cm con 20 ml de RVG2i + 10µM ROCKi. La placa tenía 2x MEF neoR (u otros MEF según el proyecto).

17. Incubado a 37°. Comenzó la selección 48 horas después.

18. El protocolo de selección G418 fue el siguiente:

a. Día 2 (2º día después del EP): células incubadas en medios 2i + G418, 75 µg/ml.

b. Día 3: células incubadas en medio 2i sin G418

c. Día 4: células incubadas en medios 2i + G418, 75 µg/ml.

d. Día 5: células incubadas en medio 2i sin G418

e. Día 6: células incubadas en medios 2i + G418, 75 µg/ml.

f. Día 7: células incubadas en medio 2i sin G418

g. Día 8: células incubadas en medios 2i + G418, 75 µg/ml.

h. Día 9: células incubadas en medio 2i sin G418

i. Día 10: células incubadas en medios 2i + G418, 75 µg/ml.

j. Día 11: células incubadas en medio 2i sin G418

k. Día 12: se recogieron colonias para expandirlas para la selección. Cada colonia se disoció en tripsina al 0.05% + suero de pollo al 1% durante 10 minutos y luego se colocó en 1 pozo de una placa de alimentación de 96 pozos.

19. Colonias expandidas por 3 días en medios 2i.

20. Pase de clones 1:1 a nuevas placas de alimentación de 96 pozos.

21. Clones expandidos por 3 días en medios 2i.

22. Para cada clon, colonias disociadas en tripsina. Se congelaron 2/3 de cada clon y se almacenaron a -80°; se sembró el 1/3 restante en placas de laminina (placas de 96 pozos recubiertas con 10 µg/ml de laminina).

23. Cuando las placas de laminina fueron confluentes, pasaron al laboratorio de detección para la genotipificación de los clones.

1.7. Firma molecular de las células madre embrionarias de rata

Los genes listados en la Tabla 8 se expresaron 20 veces más bajo en células ES de rata que los genes correspondientes en células ES de ratón. Los genes listados en la Tabla 9 se expresaron a niveles 20 veces superiores en células ES de rata que en los genes correspondientes en células ES de ratón.

Los datos de micromatrices en las Tablas 8 y 9 se generaron de la siguiente manera. Se cultivaron células ES de rata (ACI.G2 y DA.2B) y células ES de ratón (F1H4) en medio 2i durante 3 pases hasta que confluyeron. Las células F1H4 se cultivaron en placas recubiertas de gelatina en ausencia de alimentadores. Se derivaron células ES de ratón F1H4 de embriones heterocigotos 129S6/SvEvTac y C57BL/6NTac (véase, por ejemplo, la Patente de los Estados Unidos No. 7,294,754 y Poueymirou, W.T., Auerbach, W., Frendewey, D., Hickey, J.F., Escaravage, J.M., Esau, L., Dore, A.T., Stevens, S., Adams, N.C., Dominguez, M.G., Gale, N.W., Yancopoulos, G.D., DeChiara, T.M., Valenzuela, D.M. (2007)).

Se usó el siguiente protocolo para la preparación de la muestra: se marcaron tubos Eppendorf de 1.5 mL con el ID de la muestra. Las células cultivadas en una placa se enjuagaron en solución salina tamponada con fosfato (PBS) a 37ºC. Se eliminó el PBS y se agregaron 300 ul de Trizol®. Se usó un raspador para romper las células en Trizol® (Life Technology). Las células lisadas se recogieron en Trizol® en un tubo Epperdorf de 1.5 mL. Para las células cultivadas en suspensión, las células se enjuagaron en PBS a 37ºC y se recogieron en un tubo de 1.5 mL. Las células se hicieron girar hacia abajo; el PBS fue eliminado; y se añadieron 300 ul de Trizol® a las células. Las membranas celulares se rompieron por pipeteo. Las muestras se clasificaron para FACS con 10 a 10^5 células, el volumen se concentró a menos de 100 uL. Se añadieron 4 volúmenes de regulador de lisis de ARN y se mezclaron mediante pipeteo. Para la muestra, se añadió 320 uL de regulador de lisis de ARN a la muestra de 80 uL. Las muestras se almacenaron a -20°C.

Se utilizó RNA-Seq para medir el nivel de expresión de los genes de ratón y rata. Las lecturas de secuenciación fueron mapeadas al genoma de referencia de ratón y rata por Tophat, y los RPKM (fragmentos por kilobase de exón por millón de fragmentos mapeados) fueron calculados para genes de ratón y rata. Se seleccionaron genes de homología basados en el símbolo del gen y luego se usó la prueba t para comparar el nivel de expresión de cada gen entre el ratón y la rata. miR-632 estaba en el top 10 más alto expresado en ESC de ratas, pero no se expresó en células ES de ratón. Aunque no existen datos comparativos de miR-632, con base en el nivel de su expresión en comparación con otros genes expresados en ESC de ratas y su función conocida en el desarrollo embrionario, miR-632 se seleccionó como un marcador para las células ES de ratas.

Tabla 8. Los genes listados se expresaron en niveles 20 veces más bajos en células ES de rata que en los genes correspondientes en células ES de ratón.

ID	Símbolo	Nombre de gen Entrez	Ubicación	Tipo(s)
Abcb1b	Abcb1b	Casete de unión a ATP, subfamilia B (MDR/TAP), miembro 1B	Membrana de plasma	Transportador
Acta2	ACTA2	actina, alfa 2, músculo liso, aorta	Citoplasma	Otros
Actg2	ACTG2	actina, gamma 2, músculo liso, entérico	Citoplasma	Otros
Aebp1	AEBP1	Proteína 1 de unión a AE	Núcleo	Peptidasa
Angptl2	ANGPTL2	Angiopoyetina tipo 2	Espacio extracelular	Otros
Ankrd1	ANKRD1	dominio de repetición 1 de ankirina (músculo cardíaco)	Citoplasma	regulador de transcripción
Anxa1	ANXA1	anexina A1	Membrana de plasma	Otros
Anxa6	ANXA6	anexina A6	Membrana de plasma	Otros
Anxa8	ANXA8L2	anexina A8 tipo 2	Membrana de plasma	Otros
Arhgef25	ARHGEF2 5	Factor de intercambio de nucleótidos de guanina Rho (FMAM) 25	Citoplasma	Otros

ID	Símbolo	Nombre de gen Entrez	Ubicación	Tipo(s)
Axl	AXL	Receptor AXL tirosina quinasa	Membrana de plasma	Quinasa
Basp1	BASP1	proteína 1 de señalización unida a la membrana, abundante en el cerebro	Núcleo	regulador de transcripción
Bgn	BGN	Biglicano	Espacio extracelular	Otros
Bst2	BST2	médula ósea antígeno de células estromales 2	Membrana de plasma	Otros
Btf3	BTF3	factor de transcripción básico 3	Núcleo	regulador de transcripción
Btg2	BTG2	Familia BTG, miembro 2	Núcleo	regulador de transcripción
Caps1	CAPSL	similar a la calcifosina	Otros	Otros
Cav1	CAV1	caveolina 1, proteína caveolae, 22 kDa	Membrana de plasma	receptor transmembrana
Ccdc80	CCDC80	dominio de espiral que contiene 80	Núcleo	Otros
Ccnd2	CCND2	ciclina D2	Núcleo	Otros
Cd248	CD248	Molécula CD248, endosialina	Membrana de plasma	Otros
Cd44	CD44	Molécula CD44 (grupo sanguíneo indio)	Membrana de plasma	enzima
Cd97	CD97	Molécula CD97	Membrana de plasma	Receptor acoplado a proteína G
Cdc42ep5	CDC42EP5	Proteína efectora CDC42 (unión a Rho GTPasa) 5	Citoplasma	Otros
Cdh11	CDH11	cadherina 11, tipo 2, OB-cadherina (osteoblasto)	Membrana de plasma	Otros
Cdkn2a	CDKN2A	inhibidor de quinasa dependiente de ciclina 2A	Núcleo	regulador de transcripción
Cdo1	CDO1	cisteína dioxigenasa tipo 1	Citoplasma	enzima
Clip3	CLIP3	Dominio CAP-GLY que contiene la proteína enlazadora 3	Citoplasma	Otros
Cln5	CLN5	Lipofuscinosis ceroide, neuronal 5	Citoplasma	Otros
Cnn1	CNN1	Calponina 1, básica, musculo liso	Citoplasma	Otros
Col1a1	COL1A1	colágeno, tipo I, alfa 1	Espaciador extracelular	Otros
Col1 a2	COL1A2	colágeno, tipo I, alfa 2	Espaciador extracelular	Otros
Col3a1	COL3A1	colágeno, tipo III, alfa 1	Espaciador extracelular	Otros
Col5a2	COL5A2	colágeno, tipo V, alfa 2	Espaciador extracelular	Otros
Col6a2	COL6A2	colágeno, tipo VI, alfa 2	Espaciador extracelular	Otros
Cryab	CRYAB	cristalina, alfa B	Núcleo	Otros

ID	Símbolo	Nombre de gen Entrez	Ubicación	Tipo(s)
Csf1	CSF1	factor estimulante de colonias 1 (macrófagos)	Espaciador extracelular	citoquina
Cth	CTH	cistationasa (cistationina gamma-liasa)	Citoplasma	enzima
Cthrc1	CTHRC1	repetición de triple hélice de colágeno que contiene 1	Espaciador extracelular	Otros
Ctsc	CTSC	catepsina C	Citoplasma	peptidasa
Cyr61	CYR61	Inductores angiogénicos ricos en cisteína, 61	Espaciador extracelular	Otros
Ddx58	DDX58	Polipéptido de la caja 58 DEAD (Asp-Glu-Ala-Asp)	Citoplasma	enzima
Dkk3	DKK3	inhibidor dickkopf WNT de la vía de señalización 3	Espaciador extracelular	citoquina
Dmc1	DMC1	Recombinasa 1 de ADN meiótico	Núcleo	enzima
Dpysl3	DPYSL3	Dihidropirimidinasa tipo 3	Citoplasma	enzima
Dse	DSE	dermatan sulfato epimerasa	Citoplasma	enzima
Dusp1	DUSP1	Fosfatasa de especificidad dual 1	Núcleo	fosfatasa
Dusp27	DUSP27	Fosfatasa 27 de especificidad dual (putativo)	Otros	fosfatasa
Dusp9	DUSP9	Fosfatasa de especificidad dual 9	Núcleo	fosfatasa
Ece2	ECE2	enzima 2 convertidora de endotelina	Membrana de plasma	peptidasa
Ecm1	ECM1	proteína 1 de matriz extracelular	Espaciador extracelular	Transportador
Egr1	EGR1	respuesta de crecimiento temprano 1	Núcleo	regulador de transcripción
Emp1	EMP1	Proteína 1 de membrana epitelial	Membrana de plasma	Otros
Emp3	EMP3	Proteína 3 de membrana epitelial	Membrana de plasma	Otros
Ephx2	EPHX2	epóxido hidrolasa 2, citoplasmática	Citoplasma	enzima
F3	F3	factor de coagulación III (tromboplastina, factor tisular)	Membrana de plasma	receptor transmembrana
Fau	FAU	Virus del sarcoma murino de Finkel-Biskis-Reilly (FBR-MuSV) expresado de forma ubicua	Citoplasma	Otros
Fbn1	FBN1	fibrilina 1	Espacio extracelular	Otros
Fbxo15	FBXO15	proteína 15 de caja F	Otros	regulador de transcripción
Fhl2	FHL2	cuatro y medio dominios LIM 2	Núcleo	regulador de transcripción
Flnc	FLNC	filamina C, gamma	Citoplasma	Otros
Fos	FOS	homólogo de oncogén viral de osteosarcoma murino FBJ	Núcleo	regulador de transcripción

ID	Símbolo	Nombre de gen Entrez	Ubicación	Tipo(s)
Fundc2	FUNDC2	Dominio FUN 14 que contiene 2	Citoplasma	Otros
Gjb3	GJB3	proteína de unión brecha, beta 3, 31kDa	Membrana de plasma	Transportador
Gpa33	GPA33	glicoproteína A33 (transmembrana)	Membrana de plasma	Otros
Gpbp1l1	GPBP1L1	Proteína 1 tipo 1 de unión al promotor rico en GC	Otros	Otros
Gpc3	GPC3	glipicano 3	Membrana de plasma	Otros
Grb10	GRB10	proteína 10 unida al receptor del factor de crecimiento	Citoplasma	Otros
Gstm1	GSTM5	glutatión S-transferasa mu 5	Citoplasma	enzima
Hap1	HAP1	proteína 1 asociada a la huntingtina	Citoplasma	Otros
Hist1h2bc	HIST2H2B E (incluye otros)	Aglomerado de histonas 2, H2be	Núcleo	Otros
Hmga2	HMGA2	Grupo de alta movilidad AT-gancho 2	Núcleo	enzima
Hmgn3	Hmgn3	dominio 3 de unión a grupo nucleosómico de alta movilidad	Núcleo	Otros
Hormad1	HORMAD 1	Dominio HORMA que contiene 1	Núcleo	Otros
Hsd17b14	HSD17B14	hidroxiesteroide (17-beta) deshidrogenasa 14	Citoplasma	enzima
Hspb1	HSPB1	proteína 1 de choque térmico 27kDa	Citoplasma	Otros
Hspb8	HSPB8	proteína 8 de choque térmico 22kDa	Citoplasma	Quinasa
Htra1	HTRA1	HtrA serina peptidasa 1	Espacio extracelular	peptidasa
Ifi204	Ifi204 (incluye otros)	gen 204 activado por interferón	Núcleo	regulador de transcripción
Ifi44	IFI44	proteína 44 inducida por interferón	Citoplasma	Otros
Ifit1	IFIT1B	Proteína 1B inducida por interferón con repeticiones de tetratricopéptido	Citoplasma	Otros
Ifitm3	IFITM2	proteína 2 transmembrana inducida por interferón	Citoplasma	Otros
Igf2	IGF2	factor 2 de crecimiento similar a la insulina (somatomedina A)	Espacio extracelular	factor de crecimiento
Igfbp7	IGFBP7	proteína 7 de unión al factor de crecimiento similar a la insulina	Espacio extracelular	Transportador
Il1rl1	IL1RL1	Interleucina 1 similar al receptor 1	Membrana de plasma	receptor transmembrana
Inhba	INHBA	inhibina, beta A	Espacio extracelular	factor de crecimiento
Inhbb	INHBB	inhibina, beta B	Espacio extracelular	factor de crecimiento
Irf7	IRF7	factor 7 regulador de interferón	Núcleo	regulador de transcripción

ID	Símbolo	Nombre de gen Entrez	Ubicación	Tipo(s)
Isg15	ISG15	modificador ISG15 de tipo ubiquitina	Espacio extracelular	Otros
Itga5	ITGA5	integrina, alfa 5 (receptor de fibronectina, polipéptido alfa)	Membrana de plasma	receptor transmembrana
Jun	JUN	protooncogéno jun	Núcleo	regulador de transcripción
Junb	JUNB	protooncogéno jun B	Núcleo	regulador de transcripción
Lgals3bp	LGALS3B P	Lectina, proteína 3 de unión a galactósido, soluble,	Membrana de plasma	receptor transmembrana
Lgals9	LGALS9	lectina, galactósido de unión, soluble, 9	Espacio extracelular	Otros
Lmna	LMNA	lamina A/C	Núcleo	Otros
Lox	LOX	lisil oxidasa	Espacio extracelular	enzima
Loxl2	LOXL2	lisil oxidasa tipo 2	Espacio extracelular	enzima
Loxl3	LOXL3	lisil oxidasa tipo 3	Espacio extracelular	enzima
Lrp1	LRP1	Proteína 1 relacionada con el receptor de lipoproteínas de baja densidad	Membrana de plasma	receptor transmembrana
Mageb16	MAGEB16	antígeno de melanoma familia B, 16	Otros	Otros
Mcam	MCAM	molécula de adhesión celular de melanoma	Membrana de plasma	Otros
Mgp	MGP	proteína matriz Gla	Espacio extracelular	Otros
Mmp2	MMP2	matriz de metalopeptidasa 2 (gelatinasa A, gelatinasa 72kDa, colagenasa tipo IV 72kDa)	Espacio extracelular	peptidasa
Mxra8	MXRA8	8 asociada con remodelación de matriz.	Otros	Otros
Myl9	MYL9	Miosina, cadena ligera 9, reguladora.	Citoplasma	Otros
Mylpf	MYLPF	Cadena ligera de miosina, músculo esquelético rápido fosforilable	Citoplasma	Otros
Nab2	NAB2	proteína de unión 2 NGFI-A (proteína de unión 2 EGR1)	Núcleo	regulador de transcripción
Ndufb4	NDUFB4	NADH deshidrogenasa (ubiquinona) 1 subcomplejo beta, 4, 15 kDa	Citoplasma	Transportador
Npm1	NPM1	nucleofosmina (fosfoproteína nucleolar B23, numatrina)	Núcleo	regulador de transcripción
Nr0b1	NROB1	Subfamilia de receptores nucleares 0, grupo B, miembro 1	Núcleo	receptor nuclear dependiente de ligando
Nr4a1	NR4A1	Subfamilia de receptores nucleares 4, grupo A, miembro 1	Núcleo	receptor nuclear dependiente de ligando
Nrp2	NRP2	neuropilina 2	Membrana de plasma	Quinasa

ID	Símbolo	Nombre de gen Entrez	Ubicación	Tipo(s)
Oas1a	OAS1	2'-5'-oligoadenilato sintetasa 1, 40/46kDa	Citoplasma	enzima
Oasl2	Oasl2	2'-5'oligoadenilato sintetasa similar a 2	Otros	enzima
P4ha2	P4HA2	prolil 4-hidroxilasa, polipéptido alfa II	Citoplasma	enzima
Parp3	PARP3	familia de poli(ADP-ribosa)polimerasa, miembro 3	Núcleo	enzima
Pcolce	PCOLCE	potenciador de procolágeno C-endopeptidasa	Espacio extracelular Space	Otros
Pcyt1b	PCYT1B	Fosfato de citidililtransferasa 1, colina, beta	Citoplasma	enzima
Pdgfc	PDGFC	factor de crecimiento derivado de plaquetas C	Espacio extracelular Space	factor de crecimiento
Phlda1	PHLDA1	Dominio de homología similar a Pleckstrina, familia A, miembro 1	Citoplasma	Otros
Phlda2	PHLDA2	Dominio de homología similar a Pleckstrina, familia A, miembro 2	Citoplasma	Otros
Pla2g1b	PLA2G1B	fosfolipasa A2, grupo IB (páncreas)	Espacio extracelular	enzima
Pla2g4a	PLA2G4A	fosfolipasa A2, grupo IVA (citosólico, dependiente de calcio)	Citoplasma	enzima
Porcn	PORCN	homólogo de puercoespín (Drosophila)	Citoplasma	Otros
Postn	POSTN	periostina, factor especifico osteoblasto	Espacio extracelular	Otros
Prrx1	PRRX1	Homeobox relacionados emparejados 1	Núcleo	regulador de transcripción
Prss23	PRSS23	proteasa, serina, 23	Espacio extracelular	peptidasa
Psmb8	PSMB8	subunidad proteasoma (prosoma, macropain), tipo beta, 8	Citoplasma	peptidasa
Ptgs2	PTGS2	prostaglandina-endoperóxido sintasa 2 (prostaglandina G/H sintasa y ciclooxigenasa)	Citoplasma	enzima
Ptn	PTN	pleiotrofina	Espacio extracelular	factor de crecimiento
Ptrf	PTRF	polimerasa I y factor de liberación de transcripción	Núcleo	regulador de transcripción
Rarg	RARG	receptor de ácido retinoico, gamma	Núcleo	receptor nuclear dependiente de ligando
Rgs16	RGS16	Regulador de la señalización G de la proteína 16.	Citoplasma	Otros
Rn45s	Rn45s	ARN prerribosómico 45S	Otros	Otros
Rpl10a	RPL10A	proteína ribosómica L10a	Otros	Otros
Rpl31	RPL31	proteína ribosómica L31	Otros	Otros
Rpl37a	RPL37A	proteína ribosómica L37a	Citoplasma	Otros

ID	Símbolo	Nombre de gen Entrez	Ubicación	Tipo(s)
Rps10	RPS10-NUDT3	Lectura RPS10-NUDT3	Citoplasma	Otros
Rps14	RPS14	proteína ribosómica S14	Citoplasma	regulador de traducción
Rps20	Rps20	proteína ribosómica S20	Citoplasma	Otros
Rps26	RPS26	proteína ribosómica S26	Citoplasma	Otros
Rps9	RPS9	proteína ribosómica S9	Citoplasma	regulador de traducción
S100a4	S100A4	proteína A4 de unión al calcio S100	Citoplasma	Otros
S100a6	S100A6	proteína A6 de unión al calcio S100	Citoplasma	Transportador
Schip1	SCHIP1	proteína 1 de interacción con schwannomina	Citoplasma	Otros
Sdc2	SDC2	sindecan 2	Membrana de plasma	Otros
Serpine1	SERPINE1	inhibidor de la serpina peptidasa, clado E (nexina, inhibidor del activador del plasminógeno tipo 1), miembro 1	Espacio extracelular	Otros
Serpine2	SERPINE2	inhibidor de la serpinapeptidasa, clado E (nexina, inhibidor del activador del plasminógeno tipo 1), miembro 2	Espacio extracelular	Otros
Serpinf1	SERPINF1	inhibidor de la serpinapeptidasa, clado F (alfa-2 antiplasmina, factor derivado del epitelio pigmentario), miembro 1	Espacio extracelular	Otros
Sh3gl2	SH3GL2	SH3-dominio GRB2 tipo 2	Membrana del plasma	enzima
Slc19a2	SLC19A2	portador de soluto familia 19 (transportador de tiamina), miembro 2	Membrana del plasma	Transportador
Slc25a5	SLC25A5	miembro de la familia de portadores de solutos 25 (portador mitocondrial; translocador de nucleótidos de adenina), miembro 5	Citoplasma	Transportador
Slc29a1	SLC29A1	familia de portadores de solutos 29 (transportador de nucleósidos en equilibrio), miembro 1	Membrana del plasma	Transportador
Slc35f2	SLC35F2	portador de soluto familia 35, miembro F2	Otros	Otros
Snrpn	SNRPN	pequeño polipéptido ribonucleoproteína nuclear N	Núcleo	Otros
Snx22	SNX22	selección de nexina 22	Otros	Transportador
Sparc	SPARC	Proteína secretada, ácida, rica en cisteína (osteonectina)	Espacio extracelular	Otros
Spp1	SPP1	fosfoproteína secretada 1	Espacio extracelular	citoquina
Sult4a1	SULT4A1	familia de sulfotransferasa 4A, miembro 1	Citoplasma	enzima
Tagln	TAGLN	Transgelina	Citoplasma	Otros

ID	Símbolo	Nombre de gen Entrez	Ubicación	Tipo(s)
Tcea3	TCEA3	factor de alargamiento de la transcripción A (SII), 3	Núcleo	regulador de transcripción
Tgfb3	TGFB3	factor de crecimiento transformante, beta 3	Espacio extracelular	Factor de crecimiento
Thbs 1	THBS1	trombospondina 1	Espacio extracelular	Otros
Thbs2	THBS2	trombospondina 2	Espacio extracelular	Otros
Tm4sf1	TM4SF1	miembro 1 de la familia de seis transmembrana 4 L seis	Membrana del plasma	Otros
Tmbim1	TMBIM1	Motivo inhibidor de BAX transmembrana que contiene 1	Citoplasma	Otros
Tmem176b	TMEM176 B	proteína transmembrana 176B	Otros	Otros
Tnc	TNC	Tenascina C	Espacio extracelular	Otros
Tpd5211	TPD52L1	proteína tumoral tipo D52 1	Citoplasma	Otros
Tpm2	TPM2	Tropomiosina 2 (beta)	Citoplasma	Otros
Usp18	USP18	peptidasa 18 específica de ubiquitina	Citoplasma	Peptidasa
Vim	VIM	Vimentina	Citoplasma	Otros
Wfdc2	WFDC2	Dominio central de cuatro disulfuros WAP 2	Espacio extracelular	Otros
Wisp2	WISP2	Proteína 2 de la vía de señalización inducible WNT1	Espacio extracelular	Factor de crecimiento
Ybx1	YBX1	Proteína 1 de unión de caja Y	Núcleo	regulador de transcripción

Tabla 9. Los genes listados se expresaron a niveles 20 veces superiores en células ES de rata que en los genes correspondientes en células ES de ratón.

ID	Símbolo	Nombre de Entrez Gene	Ubicación	Tipo(s)
Ajap1	Ajap1	proteína 1 asociada a la unión adherente	Otros	Otros
Amd1	AMD1	adenosilmetionina descarboxilasa 1	Citoplasma	Enzima
Ankrd2	ANKRD 2	dominio de repetición de anquirina 2 (estiramiento muscular sensible)	Núcleo	Regulador de transcripción
Arhgef9	ARHGE F9	Cdc42 guanina factor de intercambio de nucleótidos (GEF) 9	Citoplasma	Otros
Atp5h	Atp5h	ATP sintasa, transporte de H +, complejo F0 mitocondrial, subunidad d	Citoplasma	Enzima
Btg3	BTG3	Familia BTG, miembro 3	Núcleo	Otros
Car6	CA6	anhidrasa carbónica VI	Espacio extracelular	Enzima
Camk4	CAMK4	proteína quinasa IV dependiente de calcio/calmodulina	Núcleo	Quinasa
Capn12	CAPN12	Calpaina 12	Otros	Peptidasa

ID	Símbolo	Nombre de Entrez Gene	Ubicación	Tipo(s)
Cct6b	CCT6B	chaperonina que contiene TCP1, subunidad 6B (zeta 2)	Citoplasma	Transportador
Cdx2	CDX2	homeobox 2 tipo caudal	Núcleo	Regulador de transcripción
Cldn5	CLDN5	claudina 5	Membrana del plasma	Otros
Clec3a	CLEC3A	Familia 3 de dominios de lectina tipo C, miembro A	Otros	Otros
Clic6	CLIC6	canal intracelular de cloruro 6	Membrana del plasma	canal de iones
Dhrsx	DHRSX	deshidrogenasa/reductasa (familia SDR) con enlace X	Otros	Enzima
Dpysl2	DPYSL2	Dihidropirimidinasa tipo 2	Citoplasma	Enzima
Dusp26	DUSP26	Fosfatasa 26 de especificidad dual (putativo)	Citoplasma	Enzima
Eci3	Eci3	enoil-coenzima A delta isomerasa 3	Otros	Enzima
Eef2k	EEF2K	factor de alargamiento eucariótico-2 quinasa	Citoplasma	Quinasa
Efna1	EFNA1	Efrina-A1	Membrana del plasma	Otros
Epha4	EPHA4	Receptor EPH A4	Membrana del plasma	Quinasa
Fank1	FANK1	Fibronectina tipo III y dominios 1 de repetición de anquirina	Núcleo	Regulador de transcripción
Fhit	FHIT	tríada de histidina frágil	Citoplasma	Enzima
Filip1	FILIP1	proteína 1 Filamina A interactuante	Citoplasma	Otros
Fmod	FMOD	Fibromodulina	Espacio extracelular	Otros
Foxe1	FOXE1	Caja E1 Fokhead (factor de transcripción tiroidea 2)	Núcleo	regulador de transcripción
Fry	FRY	homólogo *furry* (Drosophila)	Espacio extracelular	Otros
Gjb5	GJB5	proteína de unión en muesca, beta 5, 31.1kDa	Membrana del plasma	Transportador
Gpx2	GPX2	glutatión peroxidasa 2 (gastrointestinal)	Citoplasma	Enzima
Grxcr2	GRXCR 2	glutareedoxina, rica en cisteína 2	Otros	Otros
Hecw2	HECW2	Dominio HECT, C2 y WW que contiene la proteína ligasa 2 de ubiquitina E3	Espacio extracelular	Enzima
Hey2	HEY2	velloso/potenciador de división relacionado con el motivo 2 de YRPW.	Núcleo	regulador de transcripción
Icos	Icos	coestimulador inducible de células T	Membrana del plasma	Otros
Ifitm1	IFITM1	proteína 1 transmembrana inducida por interferón	Membrana del plasma	receptor transmembrana
Il1f8	IL1F8 (IL36B)	miembro de la familia Interleucina-1 (Interleucina 36 beta)	Espacio extracelular	Citoquina

ID	Símbolo	Nombre de Entrez Gene	Ubicación	Tipo(s)
Il28ra	IL-28RA	receptor de interleucina 28, alfa	Membrana del plasma	Receptor de citoquina
Igfbpl1	IGFBPL 1	proteína tipo 1 que se une al factor de crecimiento similar a la insulina	Otros	Otros
Ipcef1	IPCEF1	Proteína 1 de interacción para los factores de intercambio de citohesina	Citoplasma	Enzima
Lctl	Lctl	similar a la lactasa	Citoplasma	Otros
Ldhd	LDHD	lactato deshidrogenasa D	Citoplasma	Enzima
Lef1	LEF1	factor 1 de unión al potenciador linfoide	Núcleo	regulador de transcripción
Lefty1	LEFTY1	factor 1 de determinación izquierda-derecha	Espacio extracelular	Factor de crecimiento
Lifr	LIFR	receptor alfa de factor inhibidor de la leucemia	Membrana del plasma	Receptor transmembrana
Lpar2	LPAR2	Receptor 2 de ácido lisofosfatídico	Membrana del plasma	Receptor acoplado a proteína G
Mog	MOG	glicoproteína de oligodendrocito de mielina	Espacio extracelular	Otros
Mom5	MORN5	Repetición de MORN que contiene 5	Otros	Otros
Pigz	NCBP2	proteína subunidad 2 de unión a la cápsula nuclear, 20kDa	Núcleo	Otros
Nptxr	NPTXR	receptor de pentraxina neuronal	Membrana del plasma	Receptor transmembrana
Ntm	NTM	neurotrimina	Membrana del plasma	Otros
Nutf2	NUTF2	factor 2 de transporte nuclear	Núcleo	Transportador
Ocln	OCLN	Ocludina	Membrana del plasma	Enzima
Olr1	OLR1	receptor 1 de lipoproteína oxidada de baja densidad (similar a la lectina)	Membrana del plasma	receptor transmembrana
Pabpc4	PABPC4	proteína de unión a poli(A), citoplásmica 4 (forma inducible)	Citoplasma	regulador de traducción
Pde11a	PDE11A	Fosfodiesterasa 11A	Citoplasma	Enzima
Pdyn	PDYN	prodinorfina	Espacio extracellular e	Transportador
Per3	PER3	reloj 3 del periodo circadiano	Núcleo	Otros
Pllp	PLLP	Plasmolipina	Membrana del plasma	Transportador
Ppp1r14c	PPP1R14 C	proteína fosfatasa 1, subunidad 14C reguladora (inhibidora)	Citoplasma	Otros
Pramel6	Pramel6	Antígeno expresado preferentemente en melanoma tipo 6	Otros	Otros
Ptpn18	PTPN18	proteína tirosina fosfatasa, no receptora tipo 18 (derivado del cerebro)	Núcleo	Fosfatasa
Pycr1	PYCR1	pirrolina-5-carboxilato reductasa 1	Citoplasma	Enzima

ID	Símbolo	Nombre de Entrez Gene	Ubicación	Tipo(s)
Rab26	RAB26	RAB26, miembro de la familia de oncogenes RAS	Membrana del plasma	Enzima
Ramp2	RAMP2	Proteína 2 que modifica la actividad del receptor (acoplado a proteína G)	Membrana del plasma	Transportador
Rbm24	RBM24	proteína 24 del motivo de unión de ARN	Otros	Otros
Rhag	RHAG	Glicoproteína asociada al Rh	Membrana del plasma	Peptidasa
Rpl3	RPL3	proteína ribosómica L3	Citoplasma	Otros
Sall3	SALL3	3 tipo sal (Drosophila)	Núcleo	Otros
Satbl	SATB1	SATB Homeobox 1	Núcleo	regulador de transcripción
Scg2	SCG2	Secretogranina II	Espacio extracelular	Citoquina
Slc15a1	SLC15A 1	portador de soluto de la familia 15 (transportador de oligopéptidos), miembro 1	Membrana del plasma	Transportador
Slc1a1	SLC1A1	portador de soluto de la familia 1 (transportador de glutamato de alta afinidad neuronal/epitelial, sistema Xag), miembro 1	Membrana del plasma	Transportador
Slc24a5	Slc24a5	portador de soluto de la familia 24 (intercambiador de sodio/potasio/calcio), miembro 5	Otros	Otros
Slc37a2	SLC37A 2	portador de soluto de la familia 37 (transportador de glucosa-6-fosfato), miembro 2	Otros	Transportador
40424	SNTB1	sintrofina, beta 1 (proteína A1 asociada a distrofina, 59 kDa, componente básico 1)	Membrana del plasma	Otros
St6galnac3	ST6GAL NAC3	ST6 (alfa-N-acetil-neuraminil-2,3-beta-galactosil-1,3)-N-acetilgalactosaminida alfa-2,6-sialiltransferasa 3	Citoplasma	Enzima
Tex12	TEX12	expresado en testículo 12	Núcleo	Otros
Tex15	TEX15	expresado en testículo 15	Espacio extracelular	Otros
Tfap2a	TFAP2A	factor de transcripción AP-2 alfa (activador de la proteína 2 alfa de unión al potenciador)	Núcleo	regulador de transcripción
Tmc1	TMC1	canal transmembrana tipo 1	Membrana del plasma	Otros
Tmem130	TMEM1 30	proteína transmembrana 130	Otros	Otros
Tmem30b	TMEM3 0B	proteína transmembrana 30B	Otros	Otros
Tomm20	TOMM2 0	Translocasa del homólogo 20 de la membrana mitocondrial externa (levadura)	Citoplasma	Transportador
Tox3	TOX3	Miembro 3 de la familia de la caja de grupo de alta movilidad TOX	Otros	Otros
Ttc25	TTC25	dominio 25 de repetición tetratricopéptido	Citoplasma	Otros
Tymp	TYMP	timidina fosforilasa	Espacio extracelular	Factor de crecimiento
Ubb	Ubb	Ubiquitina B	Citoplasma	Otros
Vamp7	VAMP7	proteína 7 de membrana asociada a la vesícula	Citoplasma	Transportador

ID	Símbolo	Nombre de Entrez Gene	Ubicación	Tipo(s)
Wfdc12	Wfdc12	Dominio 12 central de cuatro disulfuros WAP	Espacio extracelular	Otros
Wfdc15a	Wfdc15a	Dominio 15a central de cuatro disulfuros WAP	Otros	Otros
Wfdc6a	Wfdc6a	Dominio 6a central de cuatro disulfuros WAP	Otros	Otros

Tabla 10. Un subconjunto de genes de la Tabla 9, que se expresan a niveles 20 veces más altos en células ES de rata que los genes correspondientes en células ES de ratón.

ID	Nombre de Gen Entrez
Ajap1	Proteína asociada con uniones de adherencia
Cldn5	Claudina 5
Arhgef9	Cdc42 factor 9 de intercambio de nucleótido guanina
Camk4	Proteína quinasa IV dependiente de calcio/calmodulina
Efna1	efrina-A1
Epha4	Receptor EPH A4
Gjb5	proteína beta 5 de brecha de unión
Igfbpl1	proteína tipo 1 de unión al factor de crecimiento similar a la insulina
Il1f8	Interleucina 36 beta
Il28ra	receptor Interleucina 28, alfa
Lefty1	factor 1 de determinación izquierda-derecha
Lifr	Receptor alfa del factor inhibidor de leucemia
Lpar2	Receptor 2 de ácido lisofosfatídico
Ntm	Receptor de pentraxina neuronal
Ptpn18	Proteína tipo 18 no receptora de tirosina fosfatasa
Cdx2	Homeobox 2 tipo caudal
Fank1	Fibronectina tipo III y dominios 1 de repetición de anquirina
Foxe1	Caja E1 Forkhead (factor 2 de transcripción tiroidea)
Hey2	velloso/potenciador de la división relacionado con el motivo 2 de YRPW
Lef1	Factor 1 de unión al potenciador linfoide
Sall3	3 tipo Sal (Drosophila)
Satb1	homeobox 1 SATB

También se ha desarrollado una firma molecular adicional que emplea los marcadores/genes de pluripotencia para las células ES de rata. La Tabla 11 proporciona una lista de genes y sus rangos de expresión de los datos de perfiles de ARN. Se aisló ARNm a partir de células ES de rata y se comparó el nivel de expresión de diversos marcadores entre sí. El término "rango" significa los niveles de expresión comparativos de genes individuales: cuanto más alto es el rango (1 es el más alto), más alta es la expresión. Por ejemplo, el rango de 13 de Oct4 significa que, de todos los genes analizados, se expresó más alto que todos menos 12 genes. Fue fundamento este experimento cualquier valor de expresión por debajo de 30; 6107 genes tenían valores de expresión de 30 o más.

Tabla 11. Firmas moleculares de células ES de rata que emplean diversos marcadores/genes de pluripotencia, mesodérmicos, endodérmicos, neurales y trofectodérmicos.

Pluripotencia	Rango de pluripotencia	Mesodérmico	Rango Mesodérmico	Endodérmico	Rango endodérmico	Neural	Rango neural	Trofectodérmico	Rango Trofectodérmico
c-Myc	8248	Braquiuria	7542	Gata6	11195	Nestina	7761	Cdx2	739
Dnmt3L	127	Flk1	No probado	Sox17	11418	Pax6	13570		
Dppa2	No probado	Nodal	3050	Hhex1	4571	Sox2	681		
Dppa5	No probado	Bmp4	3072	Nodal	3050				
Ecat1	9714	Bmpr2	6382	Ext1	6091				
Eras	2541			Sox7	10284				
Err-beta	1368								
Fbxo15	1369								
Fgf4	3440								
Fthl17	No probado								
Gdf3	2771		Rango >6107 = expresión bkg						
Klf4	836								
Lef1	1313								
receptor LIF	724								
Lin28	828								
Nanog	774								
Oct4	13								
Rexo1	6119								
Sox15	4524								
Sox2	681								
SSEA1	No probado								
SSEA4	No probado								
Stella	No probado								

Pluripotencia	Rango de pluripotencia	Mesodérmico	Rango Mesodérmico	Endodérmico	Rango endodérmico	Neural	Rango neural	Trofectodérmico	Rango Trofectodérmico
Tcl1	No probado								
Utf1	1501								

Ejemplo 2: Inactivación de loci genómicos en ratas

2.1: Inactivación de loci genómicos endógenos usando un agente de endonucleasa

Para introducir un alelo mutante en un locus genómico de rata endógeno, las células ES de rata descritas en el presente documento se someten a electroporación con vectores de expresión (o ARNm) que expresan ZFNs 1 y 2 (o TALENs 1 y 2). Estas proteínas se unen a sus secuencias diana en cadenas opuestas, separadas por aproximadamente 6 pb a aproximadamente 40 pb. Se forma una ruptura de doble cadena dentro del locus diana, que la célula intenta reparar mediante la unión final no homóloga (NHEJ). En muchos casos, el NHEJ da como resultado la creación de una eliminación, que a menudo interrumpe la función del gen (la mayoría de las veces produce una mutación de cambio de marco). Para identificar un clon positivo que comprenda un alelo mutante, las células sometidas a electroporación se colocan en placas a baja densidad, porque no se realiza la selección del fármaco. Las colonias se seleccionan y ensayan en el sitio objetivo para ver si se produjo una mutación (por ejemplo, utilizando una modificación del ensayo de alelo (MOA) descrito anteriormente). Las células ES seleccionadas que comprenden el alelo mutante se introducen luego en un embrión de rata huésped, por ejemplo, un embrión de rata en estadio premórula o blastocisto, y se implantan en el útero de una madre sustituta para generar una rata fundadora (rata F0). Posteriormente, la rata fundadora se cría en una rata de tipo salvaje para crear heterocigotos de progenie F1 para el alelo mutante. El apareamiento de la rata F1 heterozigota puede producir progenie homozigota para el alelo mutante.

2.2.: Rata ESC direccionada a la inactivación del gen de la apolipoproteína E (ApoE) de rata utilizando nucleasas de dedos de zinc

Las nucleasas de dedo de zinc usan dominios de unión a ADN modulares específicos de secuencia para dirigir la actividad de la endonucleasa a una secuencia diana única en el genoma. Los ZFN están diseñados como un par de monómeros. Cada monómero contiene un dominio de escisión no específico de la endonucleasa FokI fusionada a 3 o más dominios de unión al ADN con dedos de zinc. Cada dedo de zinc se une a un subsitio de 3 pb y la especificidad se logra mediante los sitios objetivo combinados de ambos monómeros. Los ZFN producen rupturas de doble cadena (DSB) en el ADN, y las mutaciones (inserciones o supresiones) ocurren con frecuencia durante la unión de extremos no homólogos (NHEJ). Los DSB también estimulan la reparación direccionada por homología (HDR) mediante recombinación homóloga si se proporciona una secuencia donante con ZFN.

Dichos ZFN se emplearon en combinación con los diversos métodos y composiciones divulgados en el presente documento para mejorar la eficacia de direccionamiento. El locus de apolipoproteína E (ApoE) de rata se seleccionó como se describe en el Ejemplo 3.2(a)(i), excepto que los vectores de expresión que expresan ZFNs 1 y 2 también se introdujeron en las células ES de rata. Consúltese la Figura 10 que proporciona un esquema del evento de orientación de ApoE en combinación con rTZFNIP y rTZFN2P. La eficiencia de la direccionamiento se determinó como se explica a continuación en el Ejemplo 6 y los resultados se muestran en la Figura 11. Sorprendentemente, la eficiencia del direccionamiento aumentó 8-10 veces.

Se construyó un vector de direccionamiento de plásmidos con un casete de casete de selección de fármacos que se elimina automáticamente y un gen lacZ como gen informador. Se logró una buena eficiencia de direccionamiento y se produjo un alto % de quimeras. Las nucleasas de dedo de zinc (ZFN) también se probaron en combinación con vectores de direccionamiento para examinar su efecto en la mejora de la eficiencia de direccionamiento. El vector de direccionamiento se coexpresó con los vectores de expresión para 2 pares de ZFN que cortaron el locus ApoE. Los clones ESC de rata sometidos a electroporación con el vector de direccionamiento y un conjunto de ZFN mostraron una eficacia de direccionamiento de 8 a 10 veces mayor que la de los clones ESC de rata sometidos a electroporación con un vector de direccionamiento solo. Además, se detectó una orientación homozigota bialélica en aproximadamente el 2% de nuestros clones. Se obtuvo un alto porcentaje de quimeras de dos de estos clones direccionados.

Los clones ESC de rata direccionados a ApoE (con asistencia de ZFN) se microinyectaron en blastocistos SD, que luego se transfirieron a hembras receptoras pseudopreñadas SD, utilizando técnicas estándar. Las quimeras se identificaron por el color del pelaje; las quimeras F0 macho fueron criadas a hembras SD. Las crías de la línea germinal F1 fueron genotipificadas para detectar la presencia del alelo ApoE objetivo (Figura 17). Hubo un alto porcentaje de quimeras en dos de estos clones direccionados.

Una rata ApoE bloqueada proporciona un medio para estudiar diversos tipos de trastornos y enfermedades. En los seres humanos, la apolipoproteína se encuentra en quilomicrones, HDL, LDL y VLDL. ApoE es esencial para el catabolismo normal de los constituyentes lipoproteínicos ricos en triglicéridos. Los defectos en APOE dan como resultado numerosos estados de enfermedad que incluyen, por ejemplo, hipercolesterolemia familiar, hiperlipemia, betalipoproteinemia, disbetalipoproteinemia familiar, hiperlipoproteinemia de tipo III (HLP III), riesgo de enfermedad coronaria. Una isoforma (ApoE4) está asociada con la enfermedad de Alzheimer familiar y esporádica de inicio tardío, posiblemente también con EM.

En ratones, ApoE se encuentra principalmente en HDL; transporta el colesterol, como en los humanos. Los ratones deficientes en ApoE (2 KO independientes) tienen 5 veces más colesterol que el plasma normal; desarrollaron depósitos ricos en células espumosas en sus aortas proximales a los 3 meses de edad (comparable al síndrome humano).

Los bloqueos de ApoE en ratas ofrecen un modelo animal para estudiar la función endotelial, que incluye, entre otros, la formación de placa, los cambios transcripcionales (RNA-Seq), la función *ex vivo*. Además, un tamaño mayor de ratas facilitaría todos estos ensayos y potencialmente mejoraría la calidad de los datos de RNA-Seq.

2.3. Inactivación del locus gamma (IL2r-γ) del receptor de interleucina-2 de rata utilizando nucleasas de dedos de zinc

El locus gamma (IL2r-y) del receptor de interleucina-rata se seleccionó como se describe en el Ejemplo 3.3(a), excepto que los vectores de expresión que expresan ZFN U (ZFN en sentido ascendente) y ZFN D (ZFN en sentido descendente) también se introdujeron en as células ES de rata. La Figura 18 proporciona un esquema del evento de orientación IL2r-γ en combinación con ZFN U y ZFN D. La secuencia del locus IL2r-γ que estos dedos de zinc se unen se indica en la Figura 18. La eficiencia de la orientación se determinó como se explica a continuación en el Ejemplo 3.3(a) y los resultados se muestran en la Figura 18. En resumen, los clones direccionados como homozigotas se confirmaron mediante PCR. Para el par ZFN1: 173 clones mutantes de 192 seleccionados (90%) y para el par ZFN2: 162 clones de 192 (84%) seleccionados.

Los clones ESC de rata direccionados a IL2r-γ (con asistencia ZFN) se microinyectaron en blastocistos SD, que luego se transfirieron a hembras receptoras pseudopreñadas SD, utilizando técnicas estándar. Las quimeras se identificaron por el color del pelaje; Las quimeras F0 macho fueron criadas a hembras SD. Las crías de la línea germinal F1 fueron genotipificadas para detectar la presencia del alelo IL2r-γ objetivo.

2.4.: Inactivación del receptor de interleucina-2 de rata (IL2r-γ) utilizando CRISPR/Cas9

El locus IL2r-γ de rata se seleccionó como se describe en el Ejemplo 3.3(a), excepto que el sistema CRISPR/Cas9 también se introdujo en las células ES de rata para ayudar en la eficiencia de la direccionamiento. SBI: System Biosciences Cas9 "SmartNuclease" se emplearon vectores todo en uno y la expresión de Cas9 fue direccionada por el promotor CAG, EF1a, PGK o CMV. El ARNg personalizado se ligó en un vector y se expresó mediante el promotor H1. Se diseñaron 4 ARNgs contra Il2g. La eficiencia de direccionamiento cuando se emplean los diversos ARN de guía se muestra en la Figura 19.

Ejemplo 3: Modificación direccionada de loci genómicos de rata

3.1: Direccionamiento en Rata ESC: El locus de rata Rosa 26..

El locus Rosa 26 de rata se encuentra entre los genes Setd5 y Thumpd3 como en el ratón, con el mismo espaciado. El locus Rosa 26 de rata (Figura 12, Panel B) difiere del locus Rosa 26 de ratón (Figura 12, Panel A). Las transcripciones de Rosa 26 de ratón consisten en 2 o 3 Exones. El locus de rata contiene un segundo exón 1 (Ex1b) además del exón homólogo al exón 1 de ratón (Ex1a). No ha sido identificado ningún tercer exón en ratas. La orientación de un alelo Rosa 26 de rata se representa en la Figura 12 (parte inferior), donde se clonaron brazos de homología de 5 kb cada uno por PCR utilizando ADN genómico de rata ESC DA. El alelo específico contiene un casete SA (aceptor de empalme)-lacZ-hUb-neo que reemplaza una eliminación de 117 pb en el intrón Rosa 26 de rata.

Se determinó la eficiencia de la direccionamiento en el locus Rosa 26 de rata (Tabla 12). El vector linealizado se sometió a electroporación en ESC de rata con DA o ACI, y las colonias transfectadas se cultivaron en medios 2i + G418, utilizando técnicas estándar. Las colonias individuales se seleccionaron y se cribaron utilizando un ensayo de pérdida de Alelo (LOA) (Valenzuela, D. et. Al. (2003). High-throughput engineering of the mouse genome coupled with high-resolution expression analysis, Nature Biotech. 21: 652-660).

Tabla 12. Eficiencia de direccionamiento en rata Rosa 26

Línea celular	Colonias escogidas	Positivos reconfirmados	Eficiencia de direccionamiento (%)
DA.2B	192	4	2.1
ACI.G1	96	4	4.2

Producción de quimeras y transmisión de la línea germinal utilizando clones ESC direccionados a Rosa 26. Se volvieron a inyectar clones ESC de rata direccionados a Rosa 26 reconfirmados en blastocistos SD, que luego se transfirieron a hembras receptoras pseudopreñadas SD, utilizando técnicas estándar. Las quimeras se identificaron por el color del pelaje; las quimeras F0 macho fueron criadas a hembras SD. Las crías de la línea germinal (agutí) F1 fueron genotipificadas por la presencia del alelo Rosa 26 diana; nueve de 22 cachorros de agutí fueron genotipificados como heterocigotos en el locus Rosa 26 (Tabla 13).

Tabla 13. Transmisión de la línea germinal utilizando Rosa 26 rESC direccionado						
Línea celular	clones R26 inyectados	Clones productores de quimeras	Clones de transmisión de la línea germinal	Cachorros totales	Cachorros derivados de rESC	Cachorros derivados de ESC (%)
DA.2B	4	3	2	AH7: 64	AH7:41	AH7: 63
				AE3: 112	AE3:6	AE3:3
ACI.G1	4	4	1	DE9: 39	DE9: 4	10

3.2.(a)(i): Direccionamiento del locus de apolipoproteína E (ApoE) de rata.

El locus de apolipoproteína E (ApoE) de rata se direccionó para alterar la función de ApoE. El direccionamiento del locus ApoE se realizó utilizando un vector de direccionamiento que comprende un casete lacZ-hUb-neo flanqueado con un homólogo de brazos de homología 5' y 3' al locus ApoE. La figura 20 muestra un locus ApoE de rata modificado genéticamente que se ha interrumpido por una eliminación de 1.8 kb y la inserción de un casete lacZ-hUb-neo, que incluye además un casete Cre que se puede borrar y que contiene un gen Crei controlado por un promotor de protamina. Las condiciones de electroporación fueron las siguientes: 6 µg de ADN; 2.05 x 10^6 células; 400V; 200 uF: 342 V, 593 usec; placa en 15 cm 2x neoR MEF densos en 2i + 10 uM ROCKi.

Se determinó la eficacia de direccionamiento en el locus ApoE y se muestra en la Tabla 14. El vector linealizado se sometió a electroporación en ratas ESC DA.2B derivadas de la cepa DA, y las colonias transfectadas se cultivaron usando técnicas estándar. Las colonias individuales se seleccionaron y se cribaron utilizando un ensayo de pérdida de alelo (LOA).

	Tabla 14. Eficiencia de direccionamiento en la rata ApoE			
Línea celular	Vector	Colonias escogidas	Direccionadas	Eficiencia de direccionamiento (%)
DA.2B	ApoE-mSDC	192	7	3.7
DA.2B	ApoE-mSDC	192	15	7.8

Se realizó la producción de quimeras y la transmisión de la línea germinal utilizando clones ESC direccionados contra ApoE. Los clones ESC de rata direccionados a ApoE se microinyectaron en blastocistos SD, que luego se transfirieron a hembras receptoras pseudopreñadas SD, utilizando técnicas estándar. Las quimeras se identificaron por el color del pelaje; las quimeras F0 macho fueron criadas a hembras SD. Los cachorros de F1 fueron genotipificados para detectar la presencia del alelo ApoE objetivo (Tabla 15).

Tabla 15 Resultados de la microinyección

Exp	Clon	Cachorros	Quimeras
1	ApoE-AF5	4	3 (90, 90, 90)
2	ApoE- BC4	5	0

En la figura 21 también se proporcionan datos de direccionamiento adicionales para ApoE.

3.2.(a)(ii). Direccionamiento de ApoE en ratas con un vector de direccionamiento

La figura 20 proporciona un esquema del locus ApoE de rata y un plásmido de direccionamiento. El esquema superior de la figura 20 muestra la estructura genómica del locus ApoE de rata y las regiones genómicas correspondientes a los brazos de homología 5' y 3' (5 kb y 5.4 kb, respectivamente; cuadros de color gris oscuro). El exón 1 de ApoE no está codificado y se muestra como una caja abierta más cercana al brazo de homología 5'. Los 3 intrones de ApoE se indican como líneas y los Exones 2 y 3 comprenden regiones de codificación y se muestran como recuadros grises punteados. El exón 4 contiene secuencias codificadas y no codificadas como se indica por el sombreado gris punteado y el cuadro abierto.

El esquema inferior en la figura 20 es el vector de direccionamiento. Los brazos de homología 5' y 3' (5 kb y 5,4 kb respectivamente) están indicados por los recuadros en gris oscuro. El vector de direccionamiento comprende un gen informador (lacZ) y un casete autoeliminable flanqueado por sitios loxP (flechas abiertas). El casete que se elimina

automáticamente comprende el gen Crei unido operativamente a un promotor Prm1 de ratón y un casete de selección que comprende un gen de resistencia a neomicina unido operativamente a un promotor de ubiquitina humana.

El gen Crei comprende dos Exones que codifican una recombinasa de Cre, que están separados por un intrón (Crei) para evitar su expresión en una célula procariota. Véase, por ejemplo, la Patente de los Estados Unidos 8,697,851 y la Publicación de Solicitud de Patente de los Estados Unidos 2013-0312129, que describen el casete de autoeliminación en detalle. Al emplear el promotor Prm1, el casete que se elimina automáticamente se puede eliminar específicamente en células germinales masculinas de rata F0. El vector de direccionamiento se sometió a electroporación en las células ES de rata obtenidas en el Ejemplo 1 y las células se colocaron en placas sobre MEF resistentes a neomicina densa de 15 cm 2x en 2i + 10 uM de ROCKi. Las células ES de rata transformadas se cultivaron, seleccionaron y mantuvieron como se describe en el Ejemplo 1.

Como se muestra en la Tabla 23, se seleccionaron 384 colonias y se obtuvieron 23 clones direccionados. La eficiencia de direccionamiento fue de 5.99%. Se inyectaron 3 clones en blastocistos como se describe aquí en el Ejemplo 1. Se obtuvieron 3 clones que producen quimeras y 2 de los clones transmitieron la modificación direccionada a través de la línea germinal.

3.2.(a)(iii). Direccionamiento de ApoE en ratas con un vector de direccionamiento en combinación con nucleasas con dedos de zinc

El vector de direccionamiento empleado en el Ejemplo 3.2(a)(ii) se usó en combinación con nucleasas de dedos de zinc para apuntar al locus ApoE de rata. La Tabla 16 proporciona un resumen de la organización genómica del locus ApoE de rata. Las posiciones mostradas en la Tabla 16 se tomaron de la compilación 5.0 de la Secuencia de referencia del genoma de rata (ENSMBL). ApoE está en el cromosoma 1 en la cadena (-).

Tabla 16. Resumen del locus ApoE de rata y las posiciones de los sitios de unión a nucleasa de dedo de Zinc y los sitios de corte.

Característica	Inicio	Fin	longitud	Notas
Exón 1	81881110	81881182	73	5' no codificado
Exón2	81880269	81880332	64	contiene ATG
ATG	81880309	81880311	3	codón de inicio
Exón3	81879607	81879775	169	
Sitio de unión ZFNla	81879707	81879693	15	CAGGCCCTGAACCGC (SEQ ID NO: 10)
sitio de corte ZFN1	81879692	81879687	6	TTCTGG (SEQ ID NO: 11)
Sitio de unión ZFN1b	81879686	81879671	16	GATTACCTGCGCTGGG (SEQ ID NO: 12)
Intrón 3-4	81879776	81879207	400	
sitio de unión ZF21a	81879591	81879577	15	TTCACCCTCCGCACC (SEQ ID NO: 13)
sitio de corte ZFN2	81879576	81879570	7	TGCTGAG (SEQ ID NO: 14)
Sitio de unión ZF21b	81879569	81879552	18	TATCCAGATCCAGGGGTT(SEQ ID NO: 15)
Exón 4	81878371	81879208	838	contiene TGA
TGA	81878482	81878484	3	
Eliminación de apoE	81878482	81880311	1830	

La figura 10 proporciona un esquema del locus ApoE de rata y denota con barras grises el sitio de corte para ZFN1 y ZFN2. El sitio de corte para ZFN1 está en el exón 3 y el sitio de corte para ZNF2 está en el intrón 3. La posición exacta de ambos sitios ZFN se muestra en la Tabla 16. Las regiones genómicas correspondientes a los brazos de homología 5' y 3' (5 kb y 5.4 kb, respectivamente) se indican mediante los recuadros en gris oscuro. El exón 1 de ApoE no está codificado y se muestra como una caja abierta más cercana al brazo de homología 5'. Los tres intrones del gen ApoE se indican como líneas y los Exones 2 y 3 comprenden regiones codificantes y se muestran como recuadros grises punteados. El exón 4 contiene secuencias codificadas y no codificadas como se indica por el sombreado gris punteado y el cuadro abierto.

El vector de direccionamiento empleado fue el mismo que en el Ejemplo 3.2(a)(ii) y se muestra en la figura 20. Los ZFN se introdujeron como dos plásmidos de expresión, uno para cada mitad del par de ZFN. Se utilizaron 20 µg del plásmido para ZFN1 y 20 µg del plásmido para ZFN2. Los ZFN fueron comprados de Sigma. La expresión de cada ZFN fue direccionada por el promotor CMV.

El vector de direccionamiento se sometió a electroporación en las células ES de rata obtenidas en el Ejemplo 1 y las células se sembraron en placas con 15 cm 2x de MOS neoR densos en 2i + 10 uM ROCKi. Las células ES de rata transformadas se cultivaron, seleccionaron y mantuvieron como se describe en el Ejemplo 1.

Como se muestra en la Tabla 23, se seleccionaron 384 colonias y se obtuvieron 290 clones direccionados. La eficiencia de direccionamiento fue del 75.52%. Se inyectaron 2 clones en blastocistos como se describe aquí en el Ejemplo 1. Se obtuvieron dos clones que producen quimeras y uno de los clones transmitió la modificación direccionada a través de la línea germinal.

Además, al emplear ZFN1 y ZFN2 se produjeron 8 clones direccionados bialélicos con una eficiencia del 2.08%.

3.2.(b)(i): Modificación direccionada del locus de apolipoproteína E (ApoE) de rata utilizando un vector de direccionamiento grande (LTC).

El direccionamiento del locus ApoE se realiza utilizando un vector de direccionamiento grande (LTVEC) que comprende un casete Prm1-Crei lacZ-de ratón flanqueado con un brazo de homología 5' al locus ApoE de aproximadamente 45 kb y un brazo de homología 3' al locus ApoE alrededor de 23 Kb. La figura 22 muestra el locus ApoE de rata en donde el locus ApoE se ha interrumpido por una eliminación de 1.83 kb y la inserción del gen lacZ y un casete autoeliminable que comprende el casete mPrm1-Crei y un casete de selección hUb-neo. Los métodos empleados en el ejemplo 3.2(a)(i) se pueden usar para introducir este vector en células ES de rata.

Ejemplo 3.2.(b)(ii). Direccionamiento del locus ApoE de rata con un vector de direccionamiento grande (LTVEC)

La figura 22 proporciona un esquema del locus ApoE de rata y un vector de direccionamiento grande (LTEVC). El esquema superior de la figura 22 muestra la organización genómica del locus ApoE de rata y las regiones genómicas correspondientes a los brazos de homología 5' y 3' (45 kb y 23 kb, respectivamente; cuadros de color gris oscuro). El exón 1 de ApoE no está codificado y se muestra como una caja abierta más cercana al brazo de homología 5'. Los 3 intrones de ApoE se indican como líneas y los Exones 2 y 3 comprenden regiones de codificación y se muestran como recuadros grises punteados. El exón 4 contiene secuencias codificadas y no codificadas como se indica por el sombreado gris punteado y el cuadro abierto.

El esquema inferior en la figura 22 es el LTVEC. Los brazos de homología 5' y 3' (45 kb y 23 kb, respectivamente) se indican mediante los recuadros de color gris oscuro. El vector de direccionamiento comprende un gen informador (lacZ) y un casete que se elimina automáticamente flanqueado por sitios loxP (flechas abiertas), que comprende el gen Crei unido operativamente a un promotor Prm1 de ratón y un casete de selección de fármacos que comprende un gen de resistencia a neomicina unido operativamente a un promotor de ubiquitina humana. Crei comprende dos Exones que codifican la recombinasa Cre que están separados por un intrón (Crei) para evitar su expresión en una célula procariota. Consulte, por ejemplo, la Patente de los Estados Unidos 8,697,851 y la Publicación de Solicitud de los Estados Unidos 2013-0312129, que describe en detalle el casete que se elimina automáticamente. Al emplear un promotor Prm1 de ratón, el casete que se elimina automáticamente se puede eliminar específicamente en células germinales masculinas de rata F0.

El LTVEC se sometió a electroporación en las células ES de rata obtenidas en el Ejemplo 1 y las células se sembraron en placas con 15 cm 2x de MOS neoR densos en 2i + 10 uM ROCKi. Las células ES de rata transformadas se cultivaron, seleccionaron y mantuvieron como se describe en el Ejemplo 1.

Como se muestra en la Tabla 23, se seleccionaron 288 colonias y se obtuvieron 8 clones direccionados. La eficiencia de direccionamiento fue del 2.78%. Se inyectaron 3 clones en un embrión huésped en una etapa de blastocisto como se describe aquí en el Ejemplo 2 para producir ratas quiméricas (F0). Además, se produjo un clon direccionado bialélico que proporciona una eficiencia bialélica del 0.35%.

3.2.(b)(iii). Direccionamiento de ApoE en ratas con un vector de direccionamiento grande (LTVEC) en combinación con nucleasas de dedos de zinc

El LTVEC empleado en el Ejemplo 3.2.(b)(ii) se usó en combinación con nucleasas de dedos de zinc para apuntar al locus ApoE de rata. La Tabla 16 proporciona un resumen de la organización genómica del locus ApoE de rata y las posiciones mostradas se tomaron de la compilación 5.0 de la secuencia de referencia del genoma de rata (ENSMBL).

La figura 23 proporciona un esquema del locus ApoE de rata y denota con barras grises el sitio de corte para ZFN1 y ZFN2. El sitio de corte para ZFN1 está en t exón 3 y el sitio de corte para ZNF2 está en el intrón 3. La posición exacta de ambos sitios ZFN se muestra en la Tabla 16. Los brazos de homología 5' y 3' (45 kb y 23 kb, respectivamente) se denotan por las casillas gris oscuro. El exón 1 del gen ApoE no está codificado y se muestra como una caja abierta más cercana al brazo de homología 5'. Los tres intrones del gen ApoE se indican como líneas. Los Exones 2 y 3

comprenden regiones de codificación y se muestran como recuadros grises punteados. El exón 4 contiene secuencias codificadas y no codificadas como se indica por el sombreado gris punteado y el cuadro abierto.

El LTVEC empleado fue el mismo que en el Ejemplo 3.2(b)(ii) y se muestra en la figura 22. Los ZFN se introdujeron como dos plásmidos de expresión, uno para cada mitad del par de ZFN. Se usaron 20 µg del plásmido para ZFN1 y 20 µg del plásmido para ZFN2. Los ZFNs fueron comprados de Sigma. La expresión de cada ZFN fue direccionada por el promotor CMV.

El vector de direccionamiento se sometió a electroporación en las células ES de rata obtenidas en el Ejemplo 1 y las células se sembraron en placas con 15 cm 2x de MOS neoR densos en 2i + 10 uM ROCKi. Las células ES de rata transformadas se cultivaron, seleccionaron y mantuvieron como se describe en el Ejemplo 1.

Como se muestra en la Tabla 23, se seleccionaron 288 colonias y se obtuvieron 16 clones direccionados. La eficiencia de direccionamiento fue del 5.56%. Se inyectó un clon en blastocistos como se describe en el presente documento en el Ejemplo 2.

Además, el empleo de ZFN1 y ZFN2 produjo un clon bialélico direccionado, con una eficiencia del 0.35%.

3.3(a): Direccionamiento del locus gamma del receptor de Interleucina-2 de rata (IL2r-γ)

El locus del receptor de interleucina-2 gamma (IL2r-γ) de rata fue direccionado para interrumpir la función IL2r-γ. IL2r-γ desempeña un papel importante para la señalización de IL-2, IL-4, IL-7, IL-9, IL-15, IL-21 y las mutaciones en IL2r-γ están asociadas con defectos graves en el desarrollo de células T, B y NK.

El direccionamiento del locus IL2r-γ se realizó utilizando un vector de direccionamiento que comprende un casete eGFP-hUb-neo flanqueado con una homología de 5' y 3' brazos homólogos al locus IL2r-γ. La figura 26 representa la estructura genómica del locus IL2r-γ de rata en donde el locus IL2r-γ se ha interrumpido por una eliminación de 3.2 kb. El locus IL2r-γ objetivo también comprendía un gen eGFP y un casete autoeliminable que contiene Crei unido operativamente a un promotor Protamina 1 de ratón y un casete de selección de fármaco que comprende un promotor hUb unido operativamente a un gen de resistencia a neomicina.

La eficacia de direccionamiento en el locus IL2r-γ se determinó y se muestra en la Tabla 17. El vector linealizado se sometió a electroporación en DA.2B rata ESC y las colonias transfectadas se cultivaron utilizando técnicas estándar. Las colonias individuales se seleccionaron y se cribaron utilizando un ensayo de pérdida de alelo (LOA).

	Tabla 17. Eficiencia de direccionamiento de IL2r-γ de rata			
Línea celular	Vector	Colonias escogidas	Direccionado	Eficiencia de direccionamiento (%)
DA.2B	Il2rg-floxed neo	136	1	0.7
DA.2B	Il2rg-mSDC	96	4	4.2

Se realizó la producción de quimeras y la transmisión de la línea germinal utilizando clones de ESC direccionados a IL2r-γ. Los clones de ESC de rata direccionados a IL2r-γ se microinyectaron en blastocistos SD, que luego se transfirieron a hembras receptoras pseudopreñadas SD, utilizando técnicas estándar. Las quimeras se identificaron por el color del pelaje; las quimeras F0 macho fueron criadas a hembras SD. Las crías de la línea germinal F1 se genotipificaron para determinar la presencia del alelo IL2r-γ direccionado (Tabla 18).

Tabla 18 Resultados de la microinyección

Exp	Clon	Cachorros	Quimeras
1	Il2rg-AA1	5	2 (90, 70)
2	Il2rg-AA1	10	3 (90, 90, 80)

El fenotipo de la quimera $Il2rg^{-/Y}$ #3 se estudió más a fondo. Las células mononucleares de sangre periférica (PBMC) se tiñeron con anticuerpos que reconocen antígenos en varios linajes linfoides. Se detectaron PBMC positivas para GFP en 2 de las quimeras. Además, las células GFP+ fueron negativas para el marcador de células T CD3, y en su mayoría fueron negativas para el marcador de células B B220 y el marcador de células NK CD161a. Vea la figura 30. Las pequeñas poblaciones de doble positivo son consistentes con el fenotipo bloqueado Il2rg publicado en ratones. Estos datos se obtuvieron de una rata quimérica, que contiene células gamma-positivas para el receptor de IL2, y esto puede complicar el análisis del fenotipo.

3.3(b): Modificación direccionada del locus gamma del receptor de Interleucina-2 de rata (IL2r-γ)

El locus del receptor de interleucina-2 gamma (IL2r-γ) de rata fue direccionado para interrumpir la función IL2r-γ en ratas. La figura 26 muestra la estructura genómica del locus Il2rg de rata y el vector de direccionamiento introducido en el locus. eGFP se eligió como informador para que el inmunofenotipo de las ratas genéticamente modificadas pudiera examinarse utilizando FACS. El casete autoeliminable (hUb-Neo; Prm1-Cre) se usó para eliminar el casete de la sección del fármaco y el gen Cre específicamente en células germinales masculinas de la rata F0. Además, el vector de direccionamiento se diseñó para eliminar toda la región de codificación (aproximadamente 3.2 kb) del gen Il2rg de rata.

El tamaño de la eliminación en ESC de ratas se confirmó mediante PCR usando cebadores específicos para el locus Il2rg de rata. Tras la microinyección de los clones direccionados en embriones huésped en una etapa de blastocisto, se obtuvo un alto porcentaje de quimeras. Esas quimeras se han establecido para la cría. Para determinar si la orientación funcionó como se esperaba, la sangre periférica de las quimeras se recolectó antes de la reproducción, y el fenotipo de las células inmunitarias en la sangre periférica se analizó mediante FACS. Como se muestra en la figura 30, se detectaron células positivas para GFP en la sangre periférica en 2 de las 3 quimeras examinadas (panel superior derecho), y las ratas quiméricas contenían menos del 1% de células T, menos del 1% de células B, y menos del 1% de células NK, que son positivas para GFP (es decir, células Il2rg KO).

3.4(a). Direccionamiento del locus Rag2 en ratas con un vector de direccionamiento grande (LTVEC)

La Tabla 19 proporciona un resumen de la organización genómica del locus Rag2 de rata y las posiciones mostradas se tomaron de la compilación 5.0 de la secuencia de referencia del genoma de rata (ENSMBL). Rag2 está en el cromosoma 3 en la cadena (+).

Tabla 19. Resumen de la organización genómica del locus Rag2 de rata.

Características	Inicio	Fin	Longitud	Notas
Exón 1	97,851,317	97,851,448	132	
Exón 2	97,854,635	97,854,693	59	
Exón 3	97,858,260	97,859,615	1,356	contiene la secuencia codificadora completa
ATG	97,856,286	97,856,288	3	codón de inicio
TGA	97,857,867	97,857,869	3	codón de detención
Eliminación de Rag2	97,856,289	97,859,784	3,496	

La figura 27 proporciona un esquema del locus Rag2 de rata y un vector de direccionamiento grande (LTVEC). El esquema superior de la figura 27 muestra la organización genómica del locus ApoE de rata y las regiones genómicas correspondientes a los brazos de homología 5' y 3' (48 Kb y 15 Kb, respectivamente; recuadro gris oscuro). Rag2 comprende un solo exón indicado por el sombreado gris punteado.

El esquema inferior en la Figura 27 es el LTVEC. Los brazos de homología 5' y 3' (48 kb y 15 kb, respectivamente) se indican mediante los recuadros de color gris oscuro. El LTVEC comprende un gen informador (lacZ) y un casete autoeliminable flanqueado por sitios loxP (flechas abiertas). El casete que se elimina automáticamente comprende un promotor Prm1 de ratón unido operativamente al gen Crei y un casete de selección de fármaco que comprende un promotor de ubiquitina humana unido operativamente a un gen de resistencia a neomicina. Crei comprende dos Exones que codifican la recombinasa Cre y están separados por un intrón (Crei) para evitar su expresión en una célula procariota. Consulte, por ejemplo, la Patente de los Estados Unidos 8,697,851 y la Publicación de Solicitud de los Estados Unidos 2013-0312129, que describen en detalle el casete que se elimina automáticamente. Al emplear un promotor Prm1 de ratón, el casete que se elimina automáticamente se puede eliminar específicamente en células germinales masculinas de rata F0.

El LTVEC se sometió a electroporación en las células ES de rata obtenidas en el Ejemplo 1 y las células se sembraron en placas con 15 cm 2x de MEF neoR densos en 2i + 10 uM ROCKi. Las células ES de rata transformadas se cultivaron y se mantuvieron como se describe en el Ejemplo 1. Las colonias se seleccionan como se describe en otra parte en el presente documento y se obtienen clones direccionados. Los clones direccionados luego se inyectan en un embrión huésped como se describe en otra parte del presente documento para producir una rata F0.

3.4.(b). Direccionamiento del locus Rag1 y Rag 2 en ratas

La figura 28 proporciona la estructura genómica del locus Rag1/Rag2 de rata. CDS denota la secuencia codificadora y las cajas grises representan Exones. Rag2 está en la línea "más" con la transcripción a la derecha. Rag1 está en la cadena "menos" con la transcripción a la izquierda. Mbp = millones de pares de bases.

La Tabla 20 proporciona un resumen de la organización genómica del locus Rag2 y Rag1 de rata y las posiciones mostradas se tomaron de la compilación 5.0 de la secuencia de referencia del genoma de rata (ENSMBL). Rag1 está en el cromosoma 3 en la cadena (-).

Tabla 20. Resumen de la organización genómica del locus Rag1 de rata.

Características	Inicio	Fin	Longitud	Notas
Exón 1	97,877,145	97,877,066	80	
Exón 2	97,872,503	97,866,047	6,457	contiene la secuencia codificadora completa
ATG	97,872,489	97,872,487	3	Codón de inicio
TAA	97,869,369	97,869,367	3	Codón de detención
Eliminación de Rag1-2	97,856,289	97,872,486	16,198	

La figura 29 proporciona un esquema del locus Rag2 y Rag1 de rata y un vector de direccionamiento grande (LTVEC). El esquema superior de la figura 29 muestra la organización genómica de los loci Rag1 y Rag2 y las regiones genómicas correspondientes a los brazos de homología 5' y 3' (48 kb y 84 kb, respectivamente; cajas de color gris oscuro). Rag2 y Rag 1 comprenden cada uno un solo exón indicado por el sombreado gris punteado. El esquema inferior en la figura 29 es el LTVEC. Los brazos de homología 5' y 3' (48 kb y 84 kb, respectivamente) se indican mediante los recuadros de color gris oscuro. El LTVEC comprende un gen informador (lacZ) y un casete autoeliminable flanqueado por sitios loxP (flechas abiertas). El casete que se elimina automáticamente comprende un promotor Prm1 de rata unido operativamente al gen Crei y un casete de selección de fármaco que comprende un promotor de ubiquitina humana unido operativamente a un gen de resistencia a neomicina. Crei comprende dos Exones que codifican la recombinasa Cre y están separados por un intrón (Crei) para evitar su expresión en una célula procariota. Consulte, por ejemplo, la Patente de los Estados Unidos 8,697,851 y la Publicación de Solicitud de los Estados Unidos 2013-0312129, que describe en detalle el casete que se elimina automáticamente. Al emplear un promotor Prm1 de rata que impulsa la expresión de Crei específicamente en células germinales masculinas, el casete que se puede eliminar de las células germinales masculinas de rata F0.

El LTVEC se sometió a electroporación en las células ES de rata obtenidas en el Ejemplo 1 y las células se sembraron en placas con 15 cm 2x de MOS neoR densos en 2i + 10 uM ROCKi. Las células ES de rata transformadas se cultivaron y se mantuvieron como se describe en el Ejemplo 1.

Las colonias se criban como se describe en otra parte en el presente documento y se obtienen clones direccionados. Los clones direccionados luego se inyectan en un embrión huésped como se describe en otra parte del presente documento para producir una rata F0.

Ejemplo 4. Humanización.

4.1. Humanización de Loci Genómicos de rata

La humanización de loci genómicos de rata se lleva a cabo empleando las células ES de rata descritas en el presente documento, que son capaces de mantener su pluripotencia después de una o más electroporaciones *in vitro*, y son capaces de transmitir las modificaciones genéticas direccionadas a las generaciones posteriores. Además, para evitar las limitaciones de los plásmidos para acomodar un fragmento grande de ADN genómico, y para superar la baja eficiencia de la introducción de una modificación genética direccionada en un locus endógeno en células ES de rata, se llevan a cabo una o más modificaciones genéticas específicas en bacterias, por ejemplo, *E. coli*, utilizando recombinación homóloga bacteriana (BHR) y empleando un vector de direccionamiento grande (LTVEC). El LTVEC descrito aquí, por ejemplo, incluye un fragmento grande de una secuencia genómica de rata endógena con una o más modificaciones o comprende un ácido nucleico exógeno (por ejemplo, un ácido nucleico humano homólogo u ortólogo) flanqueado con brazos de homología de rata complementarios a regiones genómicas específicas.

4.2. Humanización de locus de inmunoglobulina de rata

La humanización de un locus de cadena pesada de inmunoglobulina endógena de rata se lleva a cabo mediante la eliminación de una o más secuencias de ácidos nucleicos de cadena pesada de inmunoglobulina endógena de rata (por ejemplo, uno o más segmentos de genes V_H endógenos, uno o más segmentos de genes D humanos y uno o más más segmentos de genes J_H humanos); e introduciendo en el locus de inmunoglobulina modificado un vector de

direccionamiento, por ejemplo, un vector de direccionamiento grande (LTVEC) que comprende: (i) una o más secuencias de ácidos nucleicos de región variable humana no reorganizadas (por ejemplo, uno o más segmentos de genes V_H humanos, uno o más segmentos del gen D, y uno o más segmentos del gen J_H humano), o una o más secuencias de ácidos nucleicos de la región variable humana reorganizadas (por ejemplo, uno o más segmentos del gen V-D-J reorganizado humano); (ii) un casete de selección (por ejemplo, un gen de resistencia a la neomicina flanqueado con sitios loxP); y (iii) brazos de homología de ratas 5' y 3'.

En resumen, uno o más segmentos génicos de la región variable de cadena pesada de inmunoglobulina endógena de rata (es decir, uno o más segmentos génicos V_H, uno o más segmentos génicos D humanos y uno o más segmentos génicos J_H humanos) en un clon BAC de rata son eliminados o inactivados direccionando el locus de la cadena pesada de inmunoglobulina endógena de rata con un casete de selección flanqueado por brazos de homología de rata. Más específicamente, se construye un vector de direccionamiento para contener un casete de selección (por ejemplo, un gen de resistencia a neomicina flanqueado con sitios loxP) flanqueado con brazos de homología de rata 5' y 3' que son complementarios a las secuencias genómicas de rata objetivo (por ejemplo, rata corriente arriba y corriente abajo) secuencias de ADN genómico que abarcan uno o más segmentos del gen V_H de rata, uno o más segmentos del gen D humano y uno o más segmentos del gen J_H humano).

A continuación, las células bacterianas que contienen un fragmento grande de ADN genómico de rata que abarca un locus de cadena pesada de inmunoglobulina de rata se seleccionan y se introducen con un plásmido (por ejemplo, pABG) que codifica una recombinasa unida operativamente a un promotor inducible de forma transitoria. El vector de direccionamiento construido anteriormente se introduce luego en las células bacterianas competentes para la recombinación. Después de la electroporación, las células bacterianas se tratan con un inductor (por ejemplo, arabinósido) para iniciar la recombinación homóloga entre el vector de direccionamiento y la secuencia genómica diana de rata en el clon BAC. Las células transformadas se colocan en placas de alta densidad y se someten a selección de medicamentos para encontrar colonias que sean resistentes a los medicamentos. Se seleccionan las colonias resistentes a los fármacos y se criban para la modificación direccionada.

Con el fin de facilitar la identificación de la modificación genética direccionada, se emplea un ensayo cuantitativo de alto rendimiento, a saber, la modificación del ensayo del alelo (MOA), que permite un rastreo a gran escala de un alelo(s) modificado en un cromosoma parental tras una modificación genética. El ensayo de MOA se puede llevar a cabo mediante diversas técnicas analíticas, que incluyen, entre otras, una PCR cuantitativa, por ejemplo, una PCR en tiempo real (qPCR). Por ejemplo, la PCR en tiempo real comprende un primer conjunto de cebadores que reconoce el locus diana y un segundo conjunto de cebadores que reconoce un locus de referencia no direccionado. Además, el conjunto de cebadores puede comprender una sonda fluorescente que reconoce la secuencia amplificada. Alternativamente, el ensayo cuantitativo se puede llevar a cabo mediante una variedad de técnicas analíticas, que incluyen, entre otras, hibridación in situ mediada por fluorescencia (FISH), hibridación genómica comparativa, amplificación de ADN isotérmica, hibridación cuantitativa con una(s) sonda(s) inmovilizada(s), Invader Probes®, MMP assays®, TaqMan® Molecular Beacon y la tecnología de sonda Eclipse™. (Véase, por ejemplo, US2005/0144655).

Las células bacterianas que comprenden el clon BAC de rata modificado, es decir, un clon BAC que contiene una secuencia de ADN genómico de rata en donde se han eliminado uno o más segmentos génicos de la región variable de la cadena pesada endógenos (segmentos génicos V_H, D y/o J_H) o se inactivan, luego se someten a electroporación con un vector de direccionamiento grande (LTVEC) que comprende: (i) una o más secuencias de ácidos nucleicos de la región variable humana no reorganizadas (por ejemplo, uno o más segmentos génicos V_H humanos no reorganizados, uno o más segmentos génicos D humanos, y uno o más segmentos del gen J_H humano), o una o más secuencias de ácidos nucleicos de la región variable humana reorganizadas (por ejemplo, uno o más segmentos del gen V-D-J humano reorganizados).

El inicio de la recombinación homóloga en las células bacterianas y la selección de clones positivos se realizan como se describió anteriormente. Las secuencias de ácidos nucleicos de región variable de cadena pesada de inmunoglobulina humana no reorganizadas o reordenadas, cuando se direccionan al locus de cadena pesada de inmunoglobulina endógena, se unen operativamente a una secuencia de ácidos nucleicos de región constante de cadena pesada de inmunoglobulina endógena de rata. Alternativamente, el locus endógeno de la región constante de la cadena pesada de la rata se puede inactivar, por ejemplo, eliminando uno o más segmentos génicos de la región constante de la cadena pesada de la rata (CH) del locus de la región constante de la cadena pesada endógena, y se puede reemplazar con una constante de la cadena pesada humana secuencia de ácidos nucleicos de la región.

Del mismo modo, la humanización de una inmunoglobulina endógena de rata κ o λ locus de cadena ligera se lleva a cabo mediante la eliminación de una o más inmunoglobulina endógena de rata κ y/o λ secuencias de ácidos nucleicos de región variable de la cadena variable (por ejemplo, uno o más gen de $V_κ$ de rata endógeno segmentos y uno o más segmentos endógenos del gen $J_κ$ de rata); y direccionamiento del locus de la cadena ligera de inmunoglobulina modificada con un vector de direccionamiento, por ejemplo, un vector de direccionamiento grande (LTVEC), que comprende: (i) una o más secuencias de ácidos nucleicos de región variable de cadena ligera de inmunoglobulina humana no reorganizadas (por ejemplo, uno o más segmentos degenes $V_κ$ humanos y uno o más segmentos del gen $J_κ$ humano), o una o más secuencias de ácidos nucleicos de la región variable humana reorganizadas (por ejemplo, uno o más segmentos del gen $V_κ$-$J_κ$ reorganizados humanos); (ii) un casete de selección (por ejemplo, un gen de resistencia a la neomicina flanqueado con sitios loxP); y (iii) brazos de homología de ratas 5' y 3'.

Las secuencias de ácidos nucleicos de región variable de cadena ligera de inmunoglobulina humana no reordenadas o reordenadas, cuando se direccionan al locus de cadena ligera de inmunoglobulina endógena, se unen operativamente a la secuencia de ácidos nucleicos de región constante de cadena ligera de inmunoglobulina endógena de rata.

El LTVEC así producido en las células bacterianas comprende, por ejemplo, un inserto de ácido nucleico que contiene un locus de cadena pesada o cadena ligera de inmunoglobulina de rata humanizada en donde uno o más segmentos génicos de región variable de cadena ligera o cadena de rata endógena de rata se han reemplazado con uno o segmentos de genes de región variable de cadena más pesada o ligera más humana; y brazos homólogos de rata (por ejemplo, que van desde 5kb a 150kb) complementarios a secuencias diana genómicas específicas. El LTVEC que comprende la modificación genética descrita anteriormente luego se linealiza y se somete a electroporación en las células ES de rata. Las células ES de rata sometidas a electroporación se colocan en placas a una alta densidad para seleccionar células ES resistentes al fármaco que comprenden el vector de direccionamiento. El proceso de selección de medicamentos elimina la mayoría de las células en placa (~99%), dejando atrás colonias individuales, cada una de las cuales es un clon derivado de una sola célula. De las células restantes, la mayoría de las células (~80-100%) contienen el vector de direccionamiento integrado en una ubicación aleatoria en el genoma. Por lo tanto, las colonias se seleccionan y se genotipifican individualmente para identificar células ES de rata que comprenden el vector de direccionamiento en la ubicación genómica correcta (por ejemplo, utilizando el ensayo de modificación del alelo (MOA) descrito anteriormente).

Para aumentar la eficacia de la modificación genética direccionada, las células ES de rata se someten a electroporación con vectores de expresión (o ARNm) que expresan ZFNs 1 y 2 (o TALENs 1 y 2) junto con el LTVEC. Los brazos de homología del vector de direccionamiento se encuentran fuera del sitio de destino de ZFN, por lo tanto, el ZFN no separa el vector de direccionamiento. La ruptura de doble hebra producida por los ZFN estimula la reparación direccionada por homología (HDR), que de otro modo representa un porcentaje muy pequeño de reparaciones que se producen normalmente en células de mamíferos (en comparación con la unión terminal no homóloga; NHEJ).

Alternativamente, se pueden introducir vectores de expresión que contienen una nucleasa asociada a CRISPR de tipo II (por ejemplo, Cas9), un ARN guía (que incluye CRISPR-RNA (ARNcr) y transactivación de CRISPR ARN (ARNtracr)), como se describe aquí, en las células bacterianas junto con el LTVEC para aumentar la eficacia de la recombinación homóloga en el locus genómico diana. Las células sometidas a electroporación se colocan en placas de alta densidad y se someten a selección de fármacos para encontrar colonias que sean resistentes a los fármacos. Las colonias resistentes al fármaco se seleccionan y criban para la modificación direccionada utilizando el ensayo de modificación del alelo (MOA) como se describe en el presente documento. Siguiendo estos procedimientos, se puede lograr una mejora en la eficiencia de la direccionamiento. Por ejemplo, la cantidad de mejora puede ser pequeña (por ejemplo, mejora de 10% a 15%) o grande (por ejemplo, mejora de 10% a 80%).

Las células ES de rata seleccionadas que comprenden la modificación genética direccionada se introducen luego en un embrión de rata huésped, por ejemplo, un embrión de rata en estadio premórula o blastocisto, y se implantan en el útero de una madre sustituta para generar una rata fundadora (rata F0). Posteriormente, la rata fundadora se cría en una rata de tipo salvaje para crear heterocigotos de progenie F1 para la modificación genética. El apareamiento de la rata F1 heterozigota puede producir una progenie homozigota para la modificación genética.

4.3(a). Reemplazo de IL2rg de rata con gamma de receptor de IL2 humano

La tabla 21 proporciona un resumen de la organización genómica del locus gamma del receptor de interleucina 2 de rata y las posiciones mostradas se tomaron de la compilación 5.0 de la secuencia de referencia del genoma de rata (ENSMBL). IL2rg está en el cromosoma X en la cadena (-).

Tabla 21. Resumen de la organización genómica del locus Il2rg de rata

Características	Inicio	Fin	Longitud	Notas
Exón 1	72,021,388	72,021,516	129	contiene ATG
ATG	72,017,500	72,017,502	3	Codón de inicio
Exón2	72,021,007	72,021,160	154	
sitio de unión ZFN1a	72,021,014	72,021,028	15	CAGGCCCTGAACCGC (SEQ ID NO: 17)
sitio de corte ZFN1	72,021,008	72,021,013	6	TTCTGG (SEQ ID NO: 18)
sitio de unión ZEN1b	72,020,993	72,021,007	15	GATTACCTGCGCTGGG (SEQ ID NO: 20)
Exón 3	72,020,606	72,020,790	185	
Exón 4	72,020,274	72,020,413	140	

Características	Inicio	Fin	Longitud	Notas
Exón 5	72,019,662	72,019,824	163	
Exón 6	72,019,101	72,019,197	97	
Exón 7	72,018,844	72,018,910	67	
Exón 8	72,017,856	72,018,506	651	contiene TGA
TGA	72,018,321	72,018,323	3	Codón de detención
Eliminación de il2rg	72,018,323	72,021,502	3,180	

El esquema inferior en la figura 26 es el vector de direccionamiento para la eliminación de IL2rg 3.2kb. El vector de direccionamiento comprende un gen informador (eGFP) unido operativamente al promotor endógeno y un casete autoeliminable flanqueado por sitios loxP (flechas abiertas). El casete que se elimina automáticamente comprende el gen Crei unido operativamente a un promotor Prm1 de ratón y un casete de selección que comprende un gen de resistencia a neomicina unido operativamente a un promotor de ubiquitina humana.

El gen Crei comprende dos Exones que codifican una recombinasa Cre, que están separados por un intrón (Crei) para evitar su expresión en una célula procariota. Véase, por ejemplo, la Patente de los Estados Unidos 8,697,851 y la Publicación de Solicitud de Patente de los Estados Unidos 2013-0312129, que describen el casete de autoeliminación en detalle. Al emplear el promotor Prm1 de ratón, el casete de expresión de Cre y el casete de selección de fármaco se pueden eliminar específicamente en células germinales masculinas de rata F0. El vector de direccionamiento se sometió a electroporación en las células ES de rata obtenidas en el Ejemplo 1 y las células se colocaron en placas sobre MEF resistentes a neomicina densa de 15 cm 2x en 2i + 10 uM de ROCKi. Las células ES de rata transformadas se cultivaron, seleccionaron y mantuvieron como se describe en el Ejemplo 1.

Como se muestra en la Tabla 23, se examinaron 168 colonias y se obtuvieron 6 clones direccionados. La eficiencia de direccionamiento fue de 3.57%.

Los clones se inyectan en blastocistos como se describe aquí en el Ejemplo 1. Se obtienen clones que producen ratas F0 y se obtienen ratas F0 que transmiten la modificación direccionada a través de la línea germinal.

Ejemplo 4.3(b). Reemplazo del ectodominio IL2rg de rata con el ectodominio IL2rg humano

La humanización de longitud completa del receptor de IL 2 gamma es útil porque las ratas que tienen este locus modificado producirán Il2rg humano; y esto permitiría la detección de Il2rg humano en ratas con anticuerpos específicos para Il2rg humano.

La ectohumanización (es decir, reemplazando el ectodominio de rata de Il2rg con el ectodominio humano de Il2rg) dará como resultado un polipéptido Il2rg que unirá los ligandos humanos para Il2rg, pero debido a que el dominio citoplasmático todavía es rata, su forma ectohumanizada de Il2rg también interactuará con la maquinaria de señalización de rata. La figura 33 proporciona una secuencia de alineación de la proteína IL-2rg humana (SEQ ID NO: 20; NP_000197.1); la proteína IL-2rg de rata (SEQ ID NO: 21; NP_543165.1); y la proteína quimérica IL-2rg (SEQ ID NO: 22) que comprende el ectodominio humano de IL-2rg fusionada con el resto de la proteína IL-2rg de rata. La unión entre la IL-2rg humana y la rata se observa mediante la línea vertical.

La Tabla 22 proporciona un resumen de la organización genómica del locus gamma del receptor de interleucina 2 de rata y las posiciones mostradas se tomaron de la compilación 5.0 de la secuencia de referencia del genoma de rata (ENSMBL). IL2rg está en el cromosoma X en la cadena (-). También se señala la posición del ectodominio de IL2rg.

Tabla 22. Resumen de la organización genómica del locus Il2rg de rata

Características	Inicio	Fin	Longitud	Notas
Exón 1	71,111,444	71,111,543	100	contiene ATG
ATG	71,111,537	71,111,539	3	codón de inicio
Exón 2	71,110,897	71,111,050	154	
Exón 3	71,110,504	71,110,688	185	
Exón 4	71,110,156	71,110,295	140	

Características	Inicio	Fin	Longitud	Notas
Exón 5	71,109,228	71,109,390	163	
Exón 6	71,108,599	71,108,645	47	contiene el dominio transmembrana
Exón 7	71,108,277	71,108,346	70	
Exón 8	71,107,404	71,107,921	518	contiene TGA
TGA	71,108,736	71,108,738	3	Codón de detención
humanización integral de longitud completa	71,107,404	71,111,539	4,136	(ATG a TGA más 3' poli-A)
ectohumanización	71,108,679	71,111,539	2,861	(ATG al inicio del dominio transmembrana)

Se construyeron vectores de direccionamiento de plásmidos para reemplazar el ectodominio de rata de la región codificante gamma del receptor de interleucina 2 con el ectodominio humano como se muestra en la Figura 31. El vector de direccionamiento se sometió a electroporación en las células ES de rata obtenidas en el Ejemplo 1 y las células fueron sembradas en 15 cm 2x de MEF densos resistentes a neomicina en 2i + 10 uM ROCKi. Las células ES de rata transformadas se cultivaron, seleccionaron y mantuvieron como se describe en el Ejemplo 1.

Como se muestra en la Tabla 23, se examinaron 192 colonias y se obtuvieron 13 clones direccionados. La eficiencia de direccionamiento fue de 6.77%.

Los clones se inyectan en blastocistos como se describe aquí en el Ejemplo 1. Se obtienen clones que producen ratas F0 y se obtienen ratas F0 que transmiten la modificación direccionada a través de la línea germinal.

Ejemplo 5. Resumen

Tabla 23. Resumen de direccionamiento de rata con diversos tipos de vectores y agentes de nucleasa discutidos en los Ejemplos 3 y 4.

Tabla 23. Resumen de direccionamiento

Ejemplo #	Locus	Vector	Colonias cribadas	Clones direccionados	Eficiencia de direccionamiento	Direccionamiento bialélico	Eficiencia bialélica	Clones inyectados	Clones productores de quimeras.	Clones transmitidos a través de la línea germinal.	Notas
3.2(a)(ii)	ApoE	Plásmido	384	23	5.99%			3	3	2	
3.2(a)(iii)	ApoE + ZFN	Plásmido	384	290	75.52%	8	2.08%	2	2	1	Estos 2 clones son bialélicos direccionados.
3.3(a)	Il2rg	Plásmido	232	5	2.16%			6	5		
3.2(b)(ii)	ApoE LTVEC	LTVEC	288	8	2.78%	1	0.35%	3	1		
3.2(b)(iii)	ApoE LTVEC + ZFN	LTVEC	288	16	5.56%	1	0.35%	1	N/A		Este clon es bialélico direccionado.
4.3(a)	Il2rg Humanización 1	Plásmido	168	6	3.57%						Reemplaza Il2rg completa de rata entera con Il2rg humana
4.3(b)	Il2rg Humanización 2	Plásmido	192	13	6.77%			2	N/A		reemplaza ectodominio Il2rg de rata con ectodominio Il2rg humano
3.4(a)	Rag2	LTVEC	270	N/A							Eliminación predicha de 5.7 KB
3.4(b)	Rag1-2	LTVEC	256	N/A							Eliminación predicha de 16.2 kb

Todas las publicaciones y solicitudes de patente mencionadas en la especificación son indicativas del nivel de los expertos en la técnica a los que pertenece esta invención. A menos que sea evidente en el contexto de cualquier realización, aspecto, paso o característica de la invención puede usarse en combinación con cualquier otro. La referencia a un rango incluye cualquier número entero dentro del rango, cualquier subrango dentro del rango. La referencia a múltiples rangos incluye composiciones de tales rangos.

Listado de secuencias

<110> Lee, Jeffrey D. Mujica, Alexander O. Auerbach, Wojtek Lai, Venus Ka-Man Valenzuela, David M. Vancopoulos, George D.

<120> Modificación direccionada de genoma de rata

<130> 57766-444071

<160> 23

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 23

<212> ADN

<213> Secuencia artificial

<220>

<223> una secuencia diana genómica que está enlazada a un ARN de guía (ARNg)

<220>

<221> características_misceláneas

<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

<212> ADN

<213> Secuencia artificial

<220>

<223> un ARN guía (ARNg)

<400> 2

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60
ggcaccgagu cggugcuuuu 80

<210> 3

<211> 42

<212> ADN

<213> Secuencia artificial

<220>

<223> un ARN guía (ARNg)

<400> 3

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cg 42

<210> 4

<211> 30

<212> ADN

<213> Secuencia artificial

<220>

<223> a ARNcr

<400> 4

guuuuagagc uagaaauagc aaguuaaaau 30

<210> 5

<211> 33

<212> ADN

<213> Secuencia artificial

<220>

<223> a ARNcr

<400> 5

guuuuagagc uagaaauagc aaguuaaaau aag 33

<210> 6

<211> 26

<212> ADN

<213> Secuencia artificial

<220>

<223> a ARNcr

<400> 6

gaguccgagc agaagaagaa guuuua 26

<210> 7

<211> 12

<212> ADN

<213> Secuencia artificial

<220>

<223> a ARNtracr

<400> 7

aaggcuaguc cg 12

<210> 8

<211> 50

<212> ADN

<213> Secuencia artificial

<220>

<223> a ARNtracr

<400> 8

aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 50

<210> 9

<211> 203

<212> PRT

<213> Mus Musculus

<400> 9

```
Met Lys Val Leu Ala Ala Gly Ile Val Pro Leu Leu Leu Leu Val Leu
 1               5                  10                  15
His Trp Lys His Gly Ala Gly Ser Pro Leu Pro Ile Thr Pro Val Asn
            20                  25                  30
Ala Thr Cys Ala Ile Arg His Pro Cys His Gly Asn Leu Met Asn Gln
        35                  40                  45
Ile Lys Asn Gln Leu Ala Gln Leu Asn Gly Ser Ala Asn Ala Leu Phe
    50                  55                  60
Ile Ser Tyr Tyr Thr Ala Gln Gly Glu Pro Phe Pro Asn Asn Val Glu
65                  70                  75                  80
Lys Leu Cys Ala Pro Asn Met Thr Asp Phe Pro Ser Phe His Gly Asn
                85                  90                  95
Gly Thr Glu Lys Thr Lys Leu Val Glu Leu Tyr Arg Met Val Ala Tyr
            100                 105                 110
Leu Ser Ala Ser Leu Thr Asn Ile Thr Arg Asp Gln Lys Val Leu Asn
        115                 120                 125
Pro Thr Ala Val Ser Leu Gln Val Lys Leu Asn Ala Thr Ile Asp Val
    130                 135                 140
Met Arg Gly Leu Leu Ser Asn Val Leu Cys Arg Leu Cys Asn Lys Tyr
145                 150                 155                 160
Arg Val Gly His Val Asp Val Pro Pro Val Pro Asp His Ser Asp Lys
                165                 170                 175
Glu Ala Phe Gln Arg Lys Lys Leu Gly Cys Gln Leu Leu Gly Thr Tyr
            180                 185                 190
Lys Gln Val Ile Ser Val Val Val Gln Ala Phe
        195                 200
```

<210> 10

<211> 15

<212> ADN

<213> Secuencia artificial

<220>

<223> Sitio de unión ZEN1a

<400> 10

caggccctga accgc 15

<210> 11

<211> 6

<212> ADN

<213> Secuencia artificial

<220>

<223> Sitio de corte ZFN1

<400> 11

ttctgg 6

<210> 12

<211> 16

<212> ADN

<213> Secuencia artificial

<220>

<223> Sitio de unión ZFN1b

<400> 12

gattacctgc gctggg 16

<210> 13

<211> 15

<212> ADN

<213> Secuencia artificial

<220>

<223> Sitio de unión ZF21A

<400> 13

ttcaccctcc gcacc 15

<210> 14

<211> 7

<212> ADN

<213> Secuencia artificial

<220>

<223> Sitio de corte ZFN2

<400> 14

tgctgag 7

<210> 15

<211> 18

<212> ADN

<213> Secuencia artificial

<220>

<223> Sitio de unión ZF21b

<400> 15

tatccagatc caggggtt 18

<210> 16

<211> 9

<212> PRT

<213> Secuencia artificial

<220>

<223> dominio conservado de una familia de nucleasas de alojamiento

<400> 16

Leu Ala Gly Leu Ile Asp Ala Asp Gly
1 5

<210> 17

<211> 15

<212> ADN

<213> Secuencia artificial

<220>

<223> Sitio de unión ZEN1a

<400> 17

caggccctga accgc 15

<210> 18

<211> 6

<212> ADN

<213> Secuencia artificial

<220>

<223> Sitio de corte ZFN1

<400> 18

ttctgg 6

<210> 19

<211> 16

<212> ADN

<213> Secuencia artificial

<220>

<223> Sitio de unión ZFN1b

<400> 19

gattacctgc gctggg 16

<210> 20

<211> 369

<212> PRT

<213> Homo sapiens

<400> 20

```
Met Leu Lys Pro Ser Leu Pro Phe Thr Ser Leu Leu Phe Leu Gln Leu
 1               5                  10                  15
Pro Leu Leu Gly Val Gly Leu Asn Thr Thr Ile Leu Thr Pro Asn Gly
            20                  25                  30
Asn Glu Asp Thr Thr Ala Asp Phe Phe Leu Thr Thr Met Pro Thr Asp
        35                  40                  45
Ser Leu Ser Val Ser Thr Leu Pro Leu Pro Glu Val Gln Cys Phe Val
    50                  55                  60
Phe Asn Val Glu Tyr Met Asn Cys Thr Trp Asn Ser Ser Ser Glu Pro
65                  70                  75                  80
Gln Pro Thr Asn Leu Thr Leu His Tyr Trp Tyr Lys Asn Ser Asp Asn
                85                  90                  95
Asp Lys Val Gln Lys Cys Ser His Tyr Leu Phe Ser Glu Glu Ile Thr
            100                 105                 110
Ser Gly Cys Gln Leu Gln Lys Lys Glu Ile His Leu Tyr Gln Thr Phe
        115                 120                 125
Val Val Gln Leu Gln Asp Pro Arg Glu Pro Arg Arg Gln Ala Thr Gln
    130                 135                 140
Met Leu Lys Leu Gln Asn Leu Val Ile Pro Trp Ala Pro Glu Asn Leu
145                 150                 155                 160
Thr Leu His Lys Leu Ser Glu Ser Gln Leu Glu Leu Asn Trp Asn Asn
                165                 170                 175
Arg Phe Leu Asn His Cys Leu Glu His Leu Val Gln Tyr Arg Thr Asp
            180                 185                 190
Trp Asp His Ser Trp Thr Glu Gln Ser Val Asp Tyr Arg His Lys Phe
        195                 200                 205
Ser Leu Pro Ser Val Asp Gly Gln Lys Arg Tyr Thr Phe Arg Val Arg
    210                 215                 220
Ser Arg Phe Asn Pro Leu Cys Gly Ser Ala Gln His Trp Ser Glu Trp
225                 230                 235                 240
Ser His Pro Ile His Trp Gly Ser Asn Thr Ser Lys Glu Asn Pro Phe
                245                 250                 255
Leu Phe Ala Leu Glu Ala Val Val Ile Ser Val Gly Ser Met Gly Leu
            260                 265                 270
Ile Ile Ser Leu Leu Cys Val Tyr Phe Trp Leu Glu Arg Thr Met Pro
        275                 280                 285
Arg Ile Pro Thr Leu Lys Asn Leu Glu Asp Leu Val Thr Glu Tyr His
    290                 295                 300
Gly Asn Phe Ser Ala Trp Ser Gly Val Ser Lys Gly Leu Ala Glu Ser
305                 310                 315                 320
Leu Gln Pro Asp Tyr Ser Glu Arg Leu Cys Leu Val Ser Glu Ile Pro
                325                 330                 335
Pro Lys Gly Gly Ala Leu Gly Glu Gly Pro Gly Ala Ser Pro Cys Asn
            340                 345                 350
Gln His Ser Pro Tyr Trp Ala Pro Pro Cys Tyr Thr Leu Lys Pro Glu
        355                 360                 365
Thr
```

<210> 21

<211> 368

<212> PRT

<213> Rattus norvegicus

<400> 21

```
Met Leu Lys Pro Leu Leu Pro Ser Arg Ser Phe Leu Leu Leu Gln Leu
 1               5                  10                  15
Leu Leu Leu Arg Val Gly Trp Ser Ser Lys Val Leu Met Ser Ser Gly
            20                  25                  30
Asn Glu Asp Thr Lys Ser Asp Leu Leu Leu Thr Ser Met Asp Leu Lys
        35                  40                  45
His Leu Ser Val Pro Thr Leu Pro Leu Pro Glu Val Gln Cys Phe Val
    50                  55                  60
Phe Asn Val Glu Tyr Met Asn Cys Thr Trp Asn Ser Ser Ser Glu Pro
65                  70                  75                  80
Gln Pro Thr Asn Leu Thr Met His Tyr Arg Tyr Lys Gly Ser Asp Asn
                85                  90                  95
Asn Thr Phe Gln Glu Cys Ser His Tyr Leu Phe Ser Lys Glu Ile Thr
            100                 105                 110
Ser Gly Cys Gln Ile Gln Lys Glu Asp Ile Gln Leu Tyr Gln Thr Phe
        115                 120                 125
Val Val Gln Leu Gln Asp Pro Gln Lys Pro Gln Arg Arg Ala Glu Gln
    130                 135                 140
Lys Leu Asn Leu Gln Asn Leu Val Ile Pro Trp Ala Pro Glu Asn Leu
145                 150                 155                 160
Thr Leu Tyr Asn Leu Ser Glu Ser Gln Val Glu Leu Arg Trp Lys Ser
                165                 170                 175
Arg Tyr Ile Glu Arg Cys Leu Gln Tyr Leu Val Gln Tyr Arg Ser Asn
            180                 185                 190
Arg Asp Arg Ser Trp Thr Glu Gln Ile Val Asp His Glu Pro Arg Phe
        195                 200                 205
Ser Leu Pro Ser Val Asp Glu Gln Lys Leu Tyr Thr Phe Arg Val Arg
    210                 215                 220
Ser Arg Phe Asn Pro Ile Cys Gly Ser Thr Gln Gln Trp Ser Lys Trp
225                 230                 235                 240
Ser Gln Pro Ile His Trp Gly Ser His Thr Ala Glu Glu Asn Pro Ser
                245                 250                 255
Leu Phe Ala Leu Glu Ala Val Leu Ile Pro Val Gly Thr Met Gly Leu
            260                 265                 270
Ile Ile Thr Leu Ile Phe Val Tyr Cys Trp Leu Glu Arg Met Pro Arg
        275                 280                 285
Ile Pro Ala Ile Lys Asn Leu Glu Asp Leu Val Thr Glu Tyr His Gly
    290                 295                 300
Asn Phe Ser Ala Trp Ser Gly Val Ser Lys Gly Leu Thr Glu Ser Leu
305                 310                 315                 320
Gln Pro Asp Tyr Ser Glu Arg Phe Cys His Val Ser Glu Ile Pro Pro
                325                 330                 335
Lys Gly Gly Ala Leu Gly Glu Gly Pro Gly Gly Ser Pro Cys Ser Leu
            340                 345                 350
His Ser Pro Tyr Trp Pro Pro Pro Cys Tyr Ser Leu Lys Pro Glu Ala
        355                 360                 365
```

<210> 22

<211> 368

<212> PRT

<213> Secuencia artificial

<220>

<223> receptor gamma de IL-2 quimérico que comprende la proteína gamma del receptor de IL-2 de rata que tiene el ectodominio del receptor gamma de IL-2 de humano

<400> 22

```
Met Leu Lys Pro Ser Leu Pro Phe Thr Ser Leu Leu Phe Leu Gln Leu
 1               5                  10                  15
Pro Leu Leu Gly Val Gly Leu Asn Thr Thr Ile Leu Thr Pro Asn Gly
            20                  25                  30
Asn Glu Asp Thr Thr Ala Asp Phe Phe Leu Thr Thr Met Pro Thr Asp
        35                  40                  45
Ser Leu Ser Val Ser Thr Leu Pro Leu Pro Glu Val Gln Cys Phe Val
    50                  55                  60
Phe Asn Val Glu Tyr Met Asn Cys Thr Trp Asn Ser Ser Ser Glu Pro
65                  70                  75                  80
Gln Pro Thr Asn Leu Thr Leu His Tyr Trp Tyr Lys Asn Ser Asp Asn
                85                  90                  95
Asp Lys Val Gln Lys Cys Ser His Tyr Leu Phe Ser Glu Glu Ile Thr
            100                 105                 110
Ser Gly Cys Gln Leu Gln Lys Lys Glu Ile His Leu Tyr Gln Thr Phe
        115                 120                 125
Val Val Gln Leu Gln Asp Pro Arg Glu Pro Arg Arg Gln Ala Thr Gln
    130                 135                 140
Met Leu Lys Leu Gln Asn Leu Val Ile Pro Trp Ala Pro Glu Asn Leu
145                 150                 155                 160
Thr Leu His Lys Leu Ser Glu Ser Gln Leu Glu Leu Asn Trp Asn Asn
                165                 170                 175
Arg Phe Leu Asn His Cys Leu Glu His Leu Val Gln Tyr Arg Thr Asp
            180                 185                 190
Trp Asp His Ser Trp Thr Glu Gln Ser Val Asp Tyr Arg His Lys Phe
        195                 200                 205
Ser Leu Pro Ser Val Asp Gly Gln Lys Arg Tyr Thr Phe Arg Val Arg
    210                 215                 220
Ser Arg Phe Asn Pro Leu Cys Gly Ser Ala Gln His Trp Ser Glu Trp
225                 230                 235                 240
Ser His Pro Ile His Trp Gly Ser Asn Thr Ser Lys Glu Asn Pro Phe
                245                 250                 255
Leu Phe Ala Leu Glu Ala Val Leu Ile Pro Val Gly Thr Met Gly Leu
            260                 265                 270
Ile Ile Thr Leu Ile Phe Val Tyr Cys Trp Leu Glu Arg Met Pro Arg
        275                 280                 285
Ile Pro Ala Ile Lys Asn Leu Glu Asp Leu Val Thr Glu Tyr His Gly
    290                 295                 300
Asn Phe Ser Ala Trp Ser Gly Val Ser Lys Gly Leu Thr Glu Ser Leu
305                 310                 315                 320
Gln Pro Asp Tyr Ser Glu Arg Phe Cys His Val Ser Glu Ile Pro Pro
                325                 330                 335
Lys Gly Gly Ala Leu Gly Glu Gly Pro Gly Gly Ser Pro Cys Ser Leu
            340                 345                 350
His Ser Pro Tyr Trp Pro Pro Pro Cys Tyr Ser Leu Lys Pro Glu Ala
        355                 360                 365
```

<210> 23

<211> 23

<212> ADN

<213> Secuencia artificial

<220>

<223> una secuencia diana genómica que está enlazada a un ARN de guía (ARNg)

<220>

<221> características_misceláneas

<222> (2)...(21)

<223> n= A, T, C o G

<220>

<221> características_misceláneas

<222> (2)...(21)

<223> n puede ser de 1 a 20 nucleótidos

<400> 23

gnnnnnnnnn nnnnnnnnnn ngg 23

REIVINDICACIONES

1. Un método para la modificación direccionada de un locus genómico de interés en una célula de rata pluripotente, que comprende:

(a) introducir en la célula de rata pluripotente un vector objetivo grande (LTVEC) que comprende un ácido nucleico inserto flanqueado por un brazo de homología 5' complementario a una primera secuencia de ácidos nucleicos en el locus genómico de interés y un brazo de homología 3' complementario a una segunda secuencia de ácidos nucleicos en el locus genómico de interés, en donde el brazo de homología 5' es al menos 5 kb y el brazo de homología 3' es al menos 5 kb; y

(b) identificar una célula de rata pluripotente modificada genéticamente que comprende una modificación genética direccionada en el locus genómico de interés,

en donde la modificación genética direccionada puede transmitirse a través de la línea germinal, y

en donde la célula de rata pluripotente que se usará para la modificación direccionada del locus genómico de interés se puede obtener mediante el cultivo de células madre embrionarias de rata aisladas en una capa de células alimentadoras que no está modificada para expresar el factor inhibidor de la leucemia (LIF) con un medio que comprende aproximadamente 50 U/mL hasta aproximadamente 150 U/mL LIF y una combinación de inhibidores que consiste en un inhibidor de la vía MEK y un inhibidor de GSK3;

2. El método de la reivindicación 1, en donde la célula de rata pluripotente es una célula madre embrionaria de rata (ES), en donde opcionalmente la célula ES de rata se deriva de una cepa DA o una cepa ACI.

3. El método de la reivindicación 1 o 2, en donde la célula de rata pluripotente:

(i) se caracteriza por la expresión de al menos un marcador de pluripotencia seleccionado del grupo que consiste en Dnmt3L, Eras, Err-beta, Fbxo15, Fgf4, Gdf3, Klf4, Lef1, receptor LIF, Lin28, Nanog, Oct4, Sox15, Sox2 y Utf1.

(ii) falta de expresión de uno o más de los marcadores de pluripotencia c-Myc, Ecat1 y Rexo1;

(iii) falta de expresión de uno o más de los marcadores mesodérmicos Brachyury y Bmpr2;

(iv) falta de expresión de uno o más de los marcadores endodérmicos Gata6, Sox17 y Sox7; o

(v) falta de expresión de uno o más de los marcadores neuronales Nestin y Pax6.

4. El método de una cualquiera de las reivindicaciones 1-3, en donde la concentración de LIF en el medio está entre aproximadamente 75 U/mL y aproximadamente 125 U/mL, preferiblemente entre aproximadamente 90 U/mL y aproximadamente 110 U/mL, y más preferiblemente alrededor de 100 U/mL.

5. El método de una cualquiera de las reivindicaciones 1-4, en donde el inhibidor de la ruta de MEK es el inhibidor de MEK PD0325901 y/o el inhibidor de GSK3 es CHIR99021.

6. El método de la reivindicación 5, en donde el inhibidor de la ruta de MEK es el inhibidor de MEK PD0325901 a una concentración de 0.8 µM a 1.2 µM, y el inhibidor de GSK3 es CHIR99021 a una concentración de 2.5 µM a 3.5 µM.

7. El método de la reivindicación 6, en donde la concentración de LIF en el medio es de aproximadamente 100 U/mL, el inhibidor de la vía MEK es el inhibidor de MEK PD0325901 a una concentración de aproximadamente 1 µM, y el inhibidor de GSK3 es CHIR99021 a una concentración de aproximadamente 3 µM.

8. El método de la reivindicación 7, en donde el medio es un medio 2i que comprende: medio basal DMEM/F12 a una concentración de lx; medio neurobasal a una concentración de lx; penicilina/estreptomicina a una concentración del 1%; L-glutamina a una concentración de 4 mM; 2-mercaptoetanol a una concentración de 0.1 mM; suplemento de N2 a una concentración de 1x; suplemento B27 a una concentración 1x; LIF a una concentración de 100U/mL; PD0325901 a una concentración de 1 µM, y CHIR99021 a una concentración de 3 µM.

9. El método de una cualquiera de las reivindicaciones 1-8, en donde el locus genómico diana:

(i) es un locus gamma del receptor de interleucina-2, un locus ApoE, un locus Rag1, un locus Rag2 o un locus Rag2/Rag1;

(ii) es un locus de inmunoglobulina,

opcionalmente en donde el locus genómico diana codifica una secuencia de aminoácidos de la región variable de la cadena pesada de inmunoglobulina de mamífero, u

opcionalmente en donde el locus genómico diana comprende una secuencia de ADN genómico que codifica una secuencia de aminoácidos de la región variable de la cadena ligera de inmunoglobulina de mamífero, opcionalmente

en donde la secuencia de ADN genómico comprende una secuencia de ácido nucleico de la región variable de la cadena ligera λ y/o κ de mamífero no reordenada; o

(iii) es un locus de receptor de células T, en donde opcionalmente el locus de receptor de células T es un locus alfa del receptor de células T.

10. El método de una cualquiera de las reivindicaciones 1-9, en donde:

(i) el LTVEC es de aproximadamente 20 kb a aproximadamente 400 kb;

(ii) el brazo de homología 5' varía de aproximadamente 5 kb a aproximadamente 100 kb o el brazo de homología 3' varía de aproximadamente 5 kb a aproximadamente 100 kb;

(iii) la suma total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 10 kb a aproximadamente 150 kb; o

(iv) la suma total de los brazos de homología 5' y 3' es de al menos 10 kb.

11. El método de una cualquiera de las reivindicaciones 1-10, en donde el ácido nucleico de inserción:

(i) comprende un polinucleótido de interés que comprende una secuencia de ácido nucleico genómico que codifica una secuencia de aminoácidos de la región variable de cadena pesada de inmunoglobulina humana;

(ii) comprende un polinucleótido de interés que comprende una secuencia de ácido nucleico genómico que codifica una secuencia de aminoácidos de la región variable de cadena ligera de inmunoglobulina humana;

(iii) comprende un polinucleótido de interés que comprende un polinucleótido que codifica al menos una región de un receptor de células T, en donde opcionalmente el receptor de células T es un receptor alfa de células T; o

(iv) comprende un polinucleótido de interés que comprende al menos un alelo de la enfermedad.

12. El método de una cualquiera de las reivindicaciones 1-11, en donde el ácido nucleico de inserción:

(i) es de un humano;

(ii) comprende un alelo activado de al menos un exón de un gen endógeno;

(iii) comprende un elemento regulador;

(iv) comprende un alelo condicional;

(v) comprende un ácido nucleico flanqueado por secuencias diana de recombinación específicas del sitio, en donde opcionalmente las secuencias diana de recombinación específicas del sitio flanquean un polinucleótido que codifica un marcador de selección y/o un gen indicador;

(vi) comprende un polinucleótido que codifica un marcador de selección, en donde opcionalmente el ácido nucleico de inserción comprende un casete de selección autoeliminable; o

(vii) comprende un gen indicador unido operativamente a un promotor.

13. El método de una cualquiera de las reivindicaciones 1-12, en donde la modificación genética direccionada es bialélica.

14. El método de una cualquiera de las reivindicaciones 1-13, en donde la modificación genética direccionada comprende:

(i) reemplazo de una secuencia de ácido nucleico de rata endógena por una secuencia de ácido nucleico homóloga u ortóloga;

(ii) eliminación de una secuencia de ácido nucleico de rata endógena, en la que opcionalmente la eliminación varía de aproximadamente 5 kb a aproximadamente 3 Mb;

(iii) una secuencia de ácido nucleico exógena que varía de aproximadamente 5 kb a aproximadamente 400 kb;

(iv) una secuencia de ácidos nucleicos exógena que comprende una secuencia de ácidos nucleicos que es homóloga u ortóloga a una secuencia de ácidos nucleicos de rata endógena;

(v) una secuencia de ácidos nucleicos quimérica que comprende una secuencia de ácidos nucleicos humana y una secuencia de ácidos nucleicos de rata;

(vi) un alelo condicional flanqueado por secuencias diana de recombinasa específicas del sitio; o

(vii) un gen informador unido operativamente a un promotor activo en una célula de rata.

15. El método de una cualquiera de las reivindicaciones 1-14, en donde la modificación genética direccionada comprende:

(i) inserción de una secuencia de ácidos nucleicos humana que es homóloga u ortóloga a una secuencia de ácidos nucleicos de rata en un locus genómico endógeno, opcionalmente en donde el tamaño de la inserción es de aproximadamente 5 kb a aproximadamente 500 kb;

(ii) reemplazo de la secuencia de ácidos nucleicos de rata en el locus genómico endógeno con la secuencia de ácidos nucleicos humana homóloga u ortóloga, opcionalmente en donde el tamaño del reemplazo es de aproximadamente 5 kb a aproximadamente 500 kb;

(iii) una secuencia de ácidos nucleicos quimérica que comprende una secuencia de ácidos nucleicos humana y una secuencia de ácidos nucleicos de rata; o

(iv) una combinación de los mismos.

16. El método de una cualquiera de las reivindicaciones 1-15, en donde introducir la etapa (a) comprende además introducir un ácido nucleico que codifica un agente de nucleasa que promueve una recombinación homóloga entre el LTVEC y el locus genómico de interés en la célula de rata pluripotente.

17. El método de la reivindicación 16, en donde el agente de nucleasa comprende:

(i) una nucleasa de dedo de zinc; o

(ii) una nucleasa efectora de tipo activador de transcripción (TALEN).

18. El método de la reivindicación 16, en donde el agente de nucleasa comprende un sistema CRISPR/Cas que comprende una nucleasa Cas9 y una ARN guía (ARNg) que comprende un ARNr-ARNtracr fusionado,

opcionalmente, en donde:

(i) El ARNg comprende:

(I) un ARN quimérico de la secuencia de ácidos nucleicos de la SEQ ID NO: 2; o

(II) un ARN quimérico de la secuencia de ácidos nucleicos de la SEQ ID NO: 3;

(ii) el ARNcr comprende la SEQ ID NO: 4; SEQ ID NO: 5; o SEQ ID NO: 6; o

(iii) el ARNtracr comprende la SEQ ID NO: 7 o la SEQ ID NO: 8.

19. El método de una cualquiera de las reivindicaciones 1-18, en donde la etapa de introducción está mediada por electroporación.

20. El método de una cualquiera de las reivindicaciones 1-19, en donde la etapa de identificación emplea un ensayo cuantitativo para evaluar una modificación del alelo (MOA) en el locus genómico de interés.

21. El método de una cualquiera de las reivindicaciones 1-20, en donde las etapas (a) y (b) se repiten secuencialmente para permitir la integración direccionada de al menos dos ácidos nucleicos de inserción en el locus genómico de interés.

22. Un método para la modificación direccionada de un locus genómico de interés en una célula de rata pluripotente, que comprende:

(a) direccionar un locus genómico de interés en una célula de rata pluripotente según el método de una cualquiera de las reivindicaciones 1-21 para formar una célula de rata pluripotente modificada genéticamente, en donde el ácido nucleico de inserción comprende un ácido nucleico humano; y

(b) introducir la célula de rata pluripotente modificada genéticamente en un embrión de rata huésped.

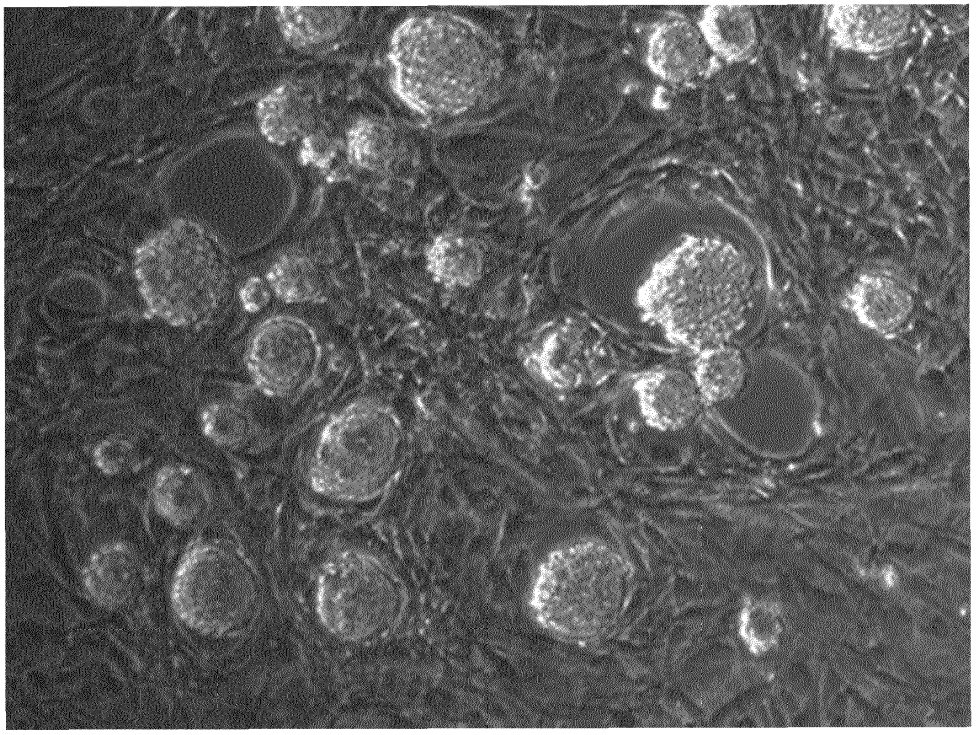


FIG. 1

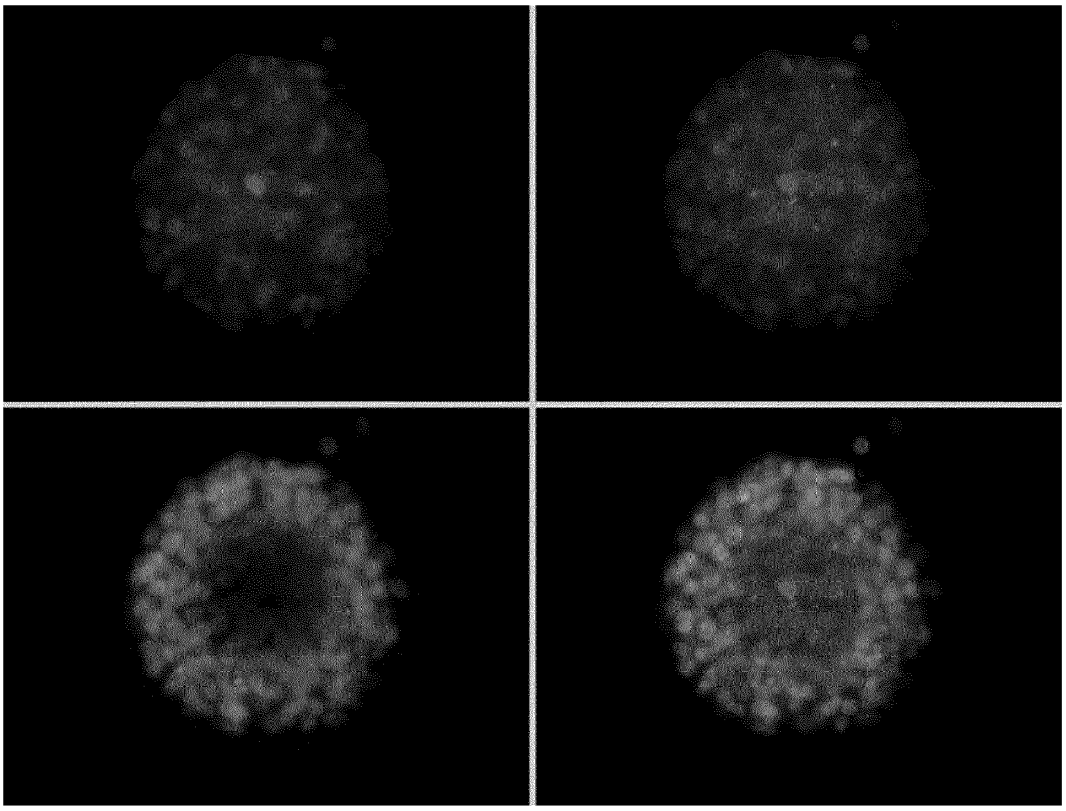


FIG. 2

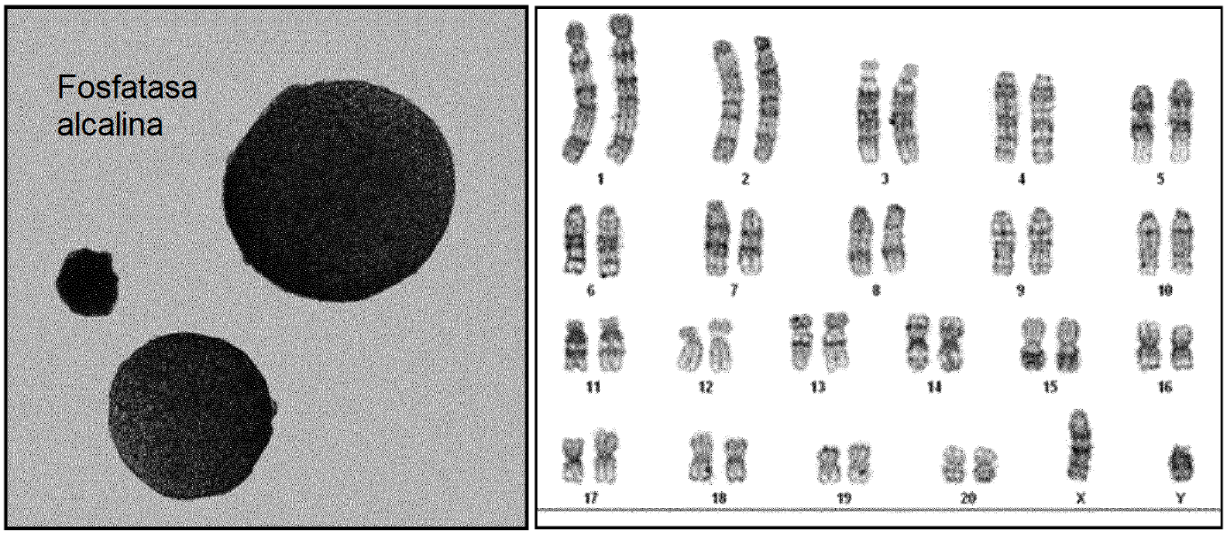

Fosfatasa
alcalina
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
X
Y

FIG. 3

Resultados de célula: *42,XX*

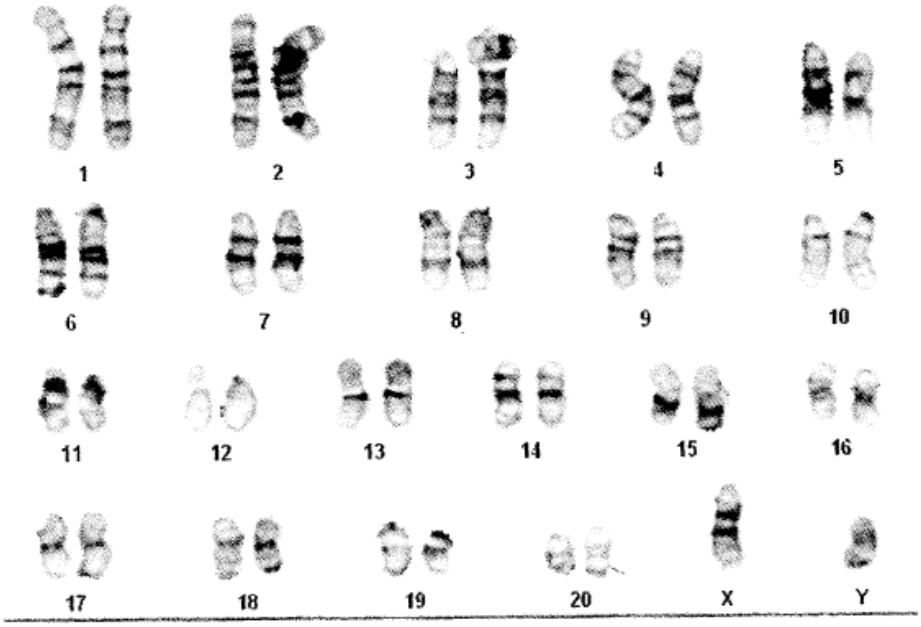


Etiqueta - Lámina/Célula *met-4 - S01-01*

X, Y: ,

FIG. 4a

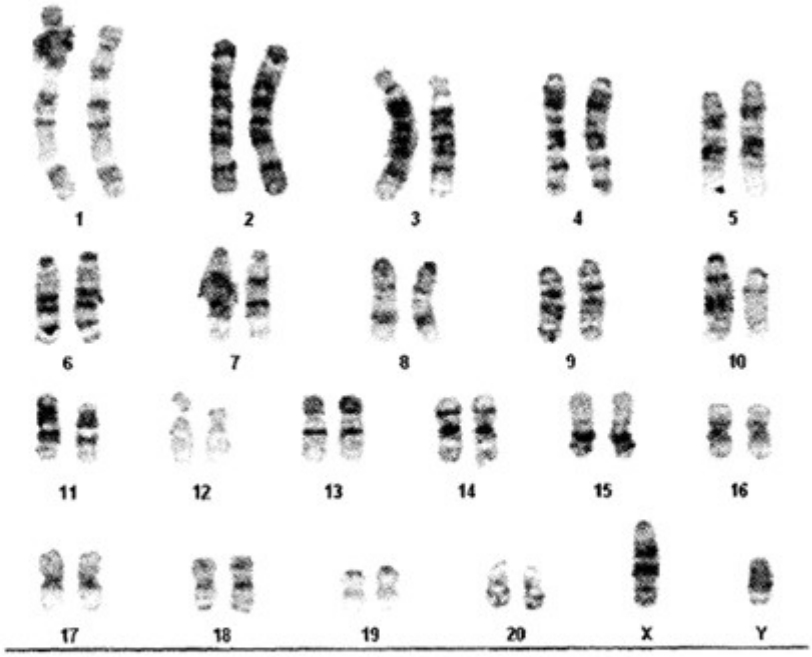
Resultados de célula: *42,XY*



Etiqueta - Lámina/Célula *met-1 - S01-01* ***X, Y:*** .

FIG. 4b

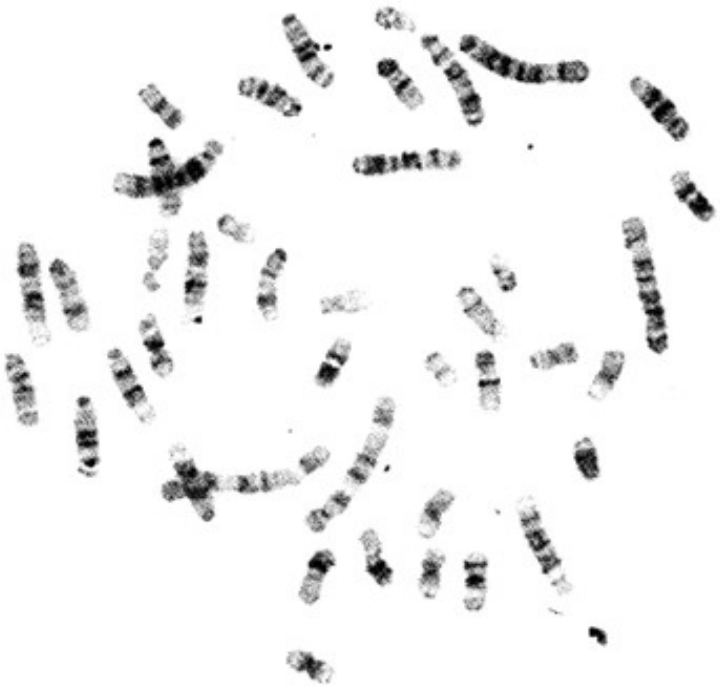
Resultados de célula: *42,XY*



Etiqueta - Lámina/Célula *met-1 - S01-01* ***X, Y:*** .

FIG. 5a

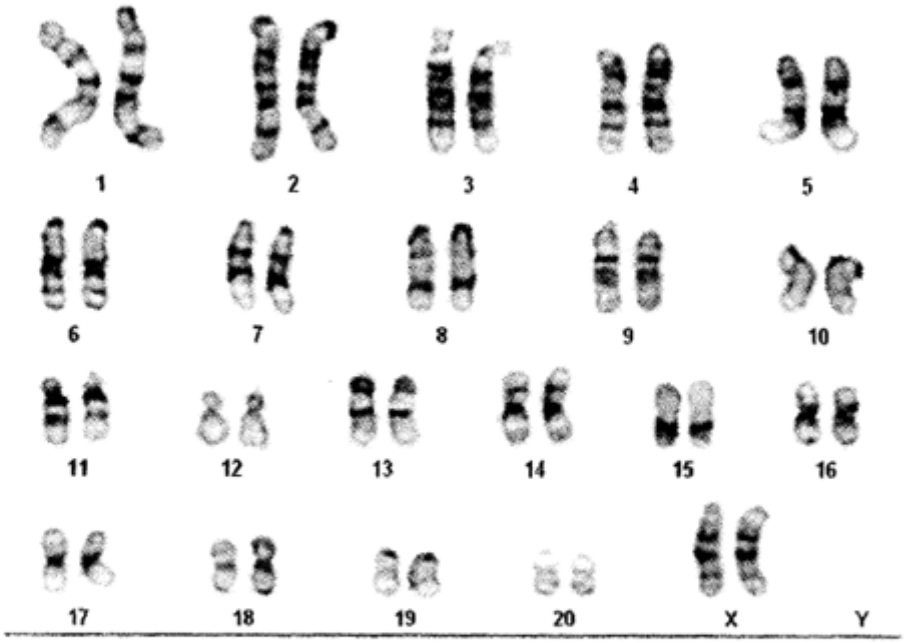
Resultados de célula: *42,XY*



Etiqueta - Lámina/Célula *met-1 - S01-01* ***X, Y:***

FIG. 5b

Resultados de célula: *42,XX*



Etiqueta - Lámina/Célula *met-4 - S01-01* ***X, Y:***

FIG. 6a

Resultados de célula: *42,XX*

Figura 6b



Etiqueta - Lámina/Célula *met-4 - S01-01*

X, Y: ,

FIG. 6b

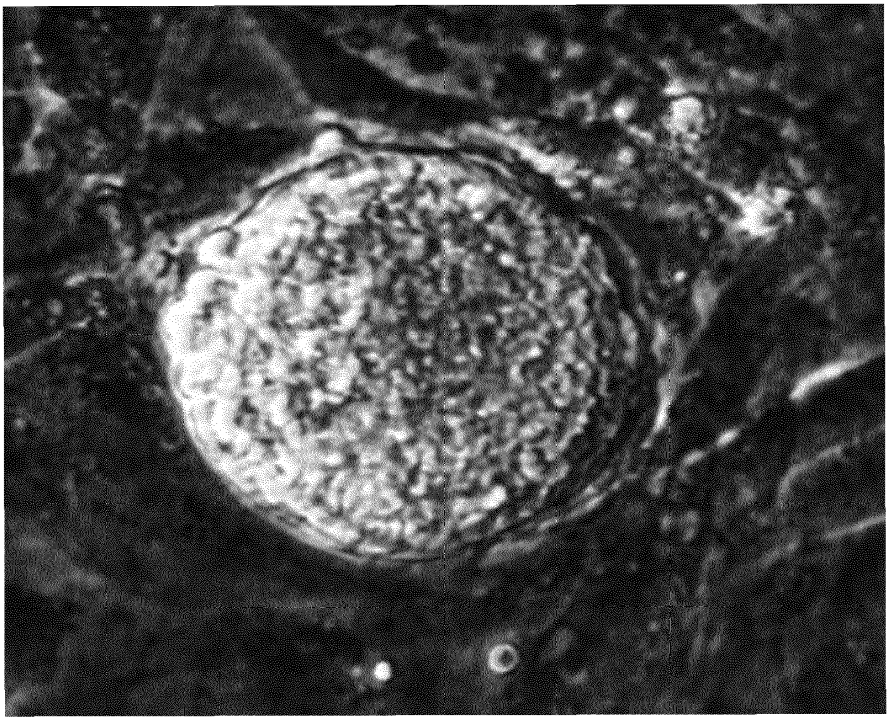


FIG. 7



FIG. 8



FIG. 9

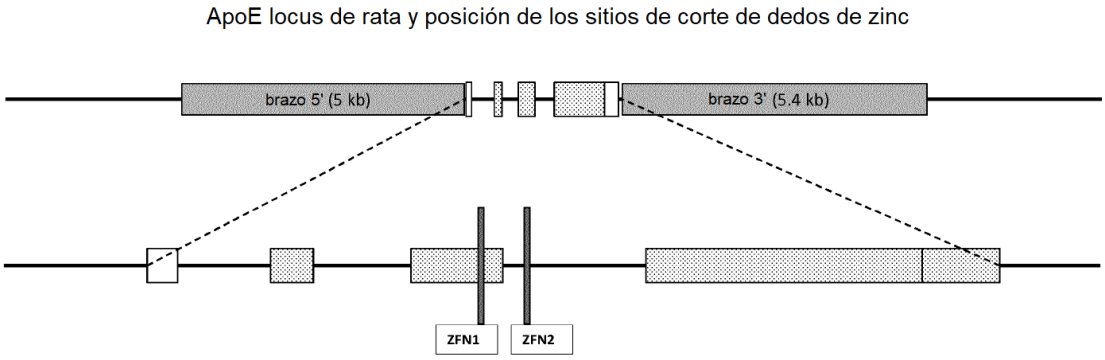

ApoE locus de rata y posición de los sitios de corte de dedos de zinc
brazo 5' (5 kb)
brazo 3' (5.4 kb)
ZFN1
ZFN2

FIG. 10

ApoE ZFNs de rata: eficiencia mejorada de definición de diana

ADN:	Colonias	Proyectado	Heterocigoto como diana	Homocigoto como diana	Dobles "mezclados"	Corte/ sin diana
vector	330	184	15 (8.2%)	0	0	N/A
vector + ZFN 1	560	192	**132** (68.8%)	**6** (3.1%)	**18** (9.4%)	**17** (8.9%)
vector + ZFN 2	410	192	**136** (70.8%)	**2** (1.0%)	**6** (3.1%)	**18** (9.4%)

Heterocigoto como diana:	1 copia de LOA U & D perdida;	1 copia de la sonda ZFN perdida
Dobles mezclados:	1 copia de LOA U & D perdida;	2 copias de la sonda ZFN perdida
Homocigoto como diana:	2 copias de LOA U & D perdida;	2 copias de la sonda ZFN perdida

FIG. 11

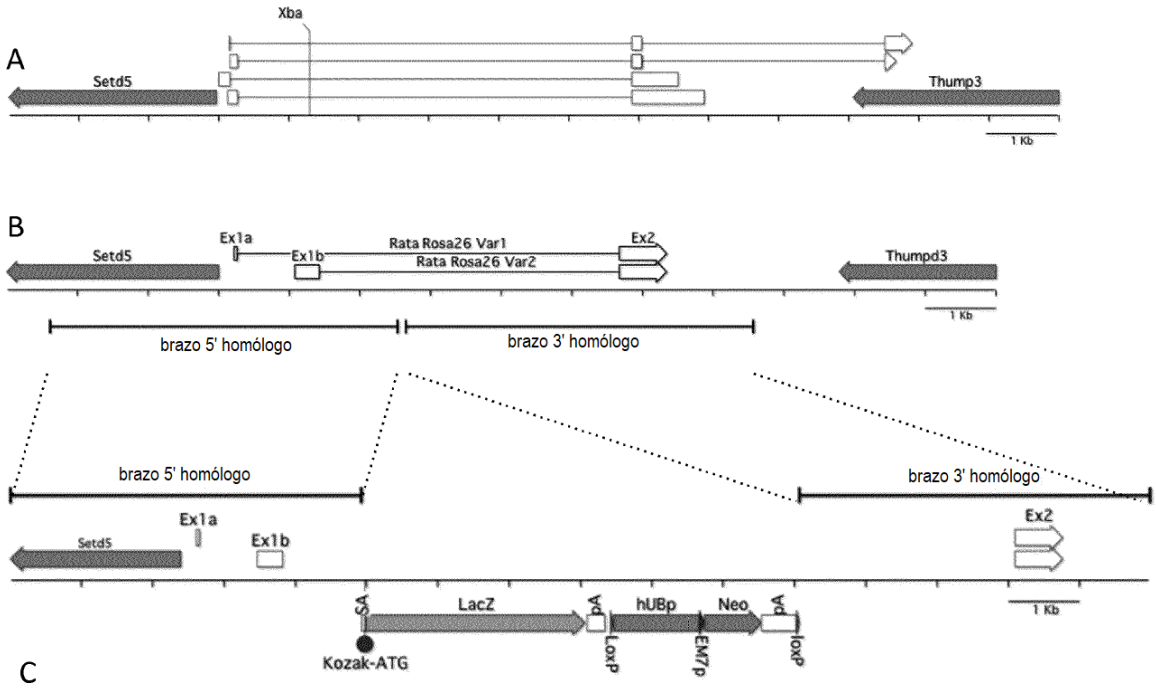

A
Xba
Setd5
Thump3
1 Kb
B
Setd5
Ex1a
Ex1b
Rata Rosa26 Var1
Rata Rosa26 Var2
Ex2
Thumpd3
1 Kb
brazo 5' homólogo
brazo 3' homólogo
brazo 5' homólogo
brazo 3' homólogo
Ex1a
Setd5
Ex1b
Ex2
1 Kb
SA
LacZ
pA
hUBp
Neo
pA
Kozak-ATG
LoxP
EM7p
loxP
C

FIG. 12

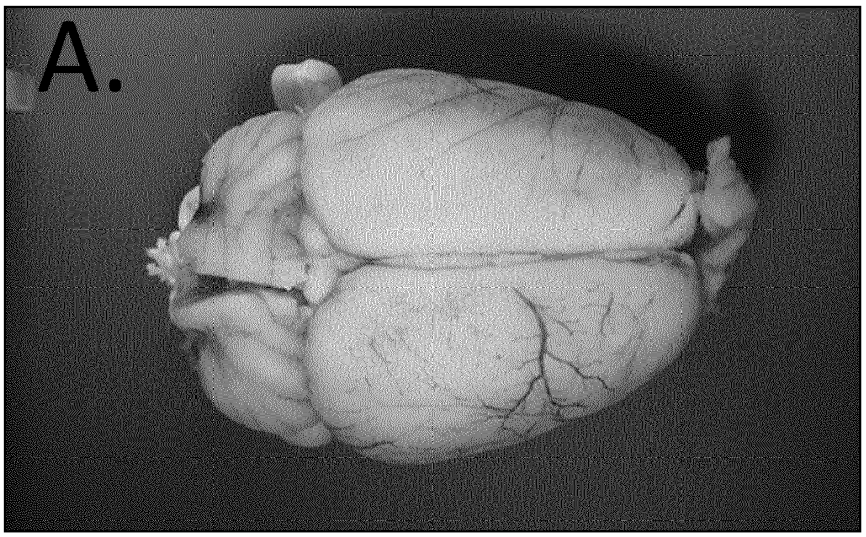


FIG. 13a

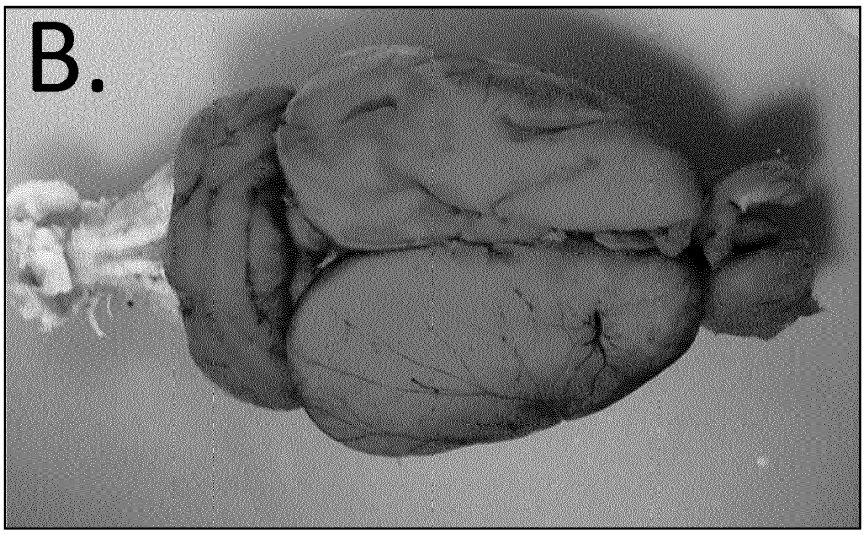


FIG. 13b

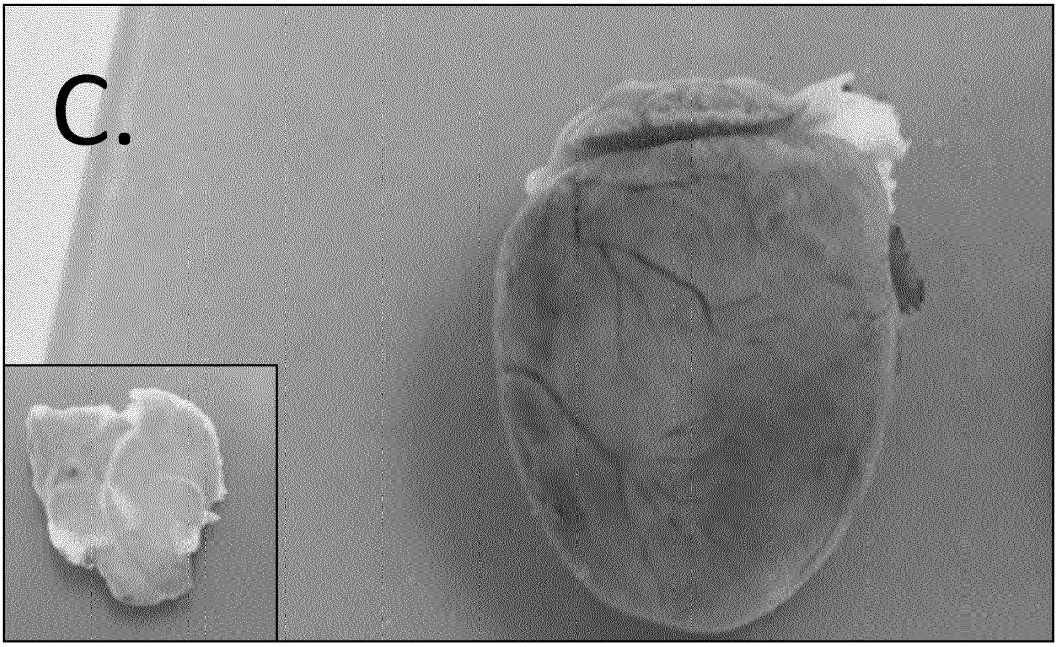


FIG. 13c

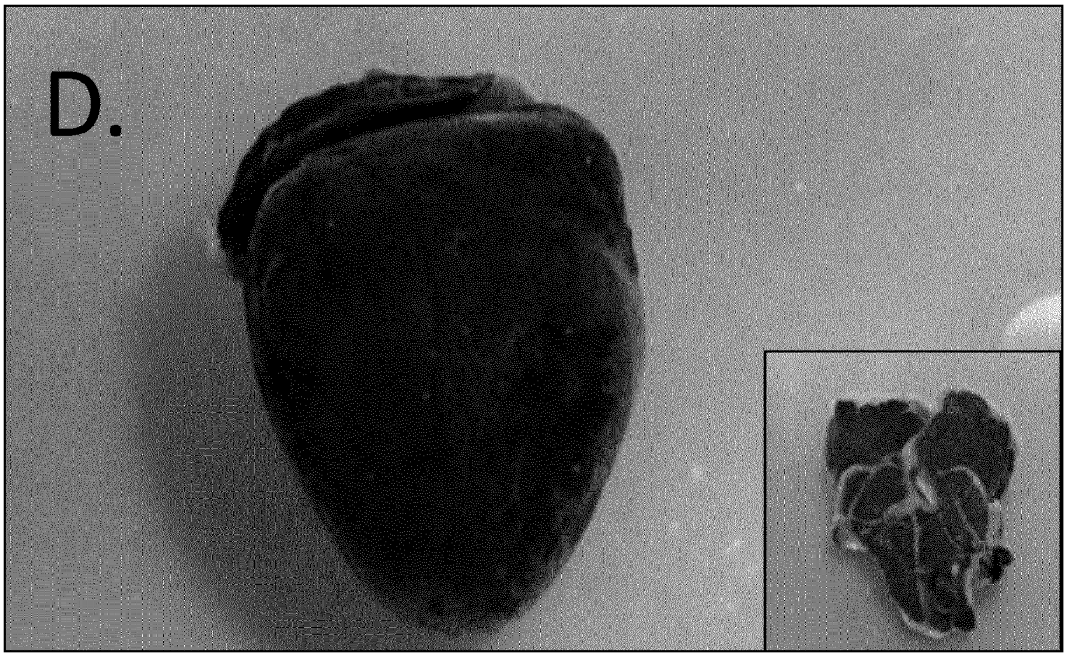


FIG. 13d

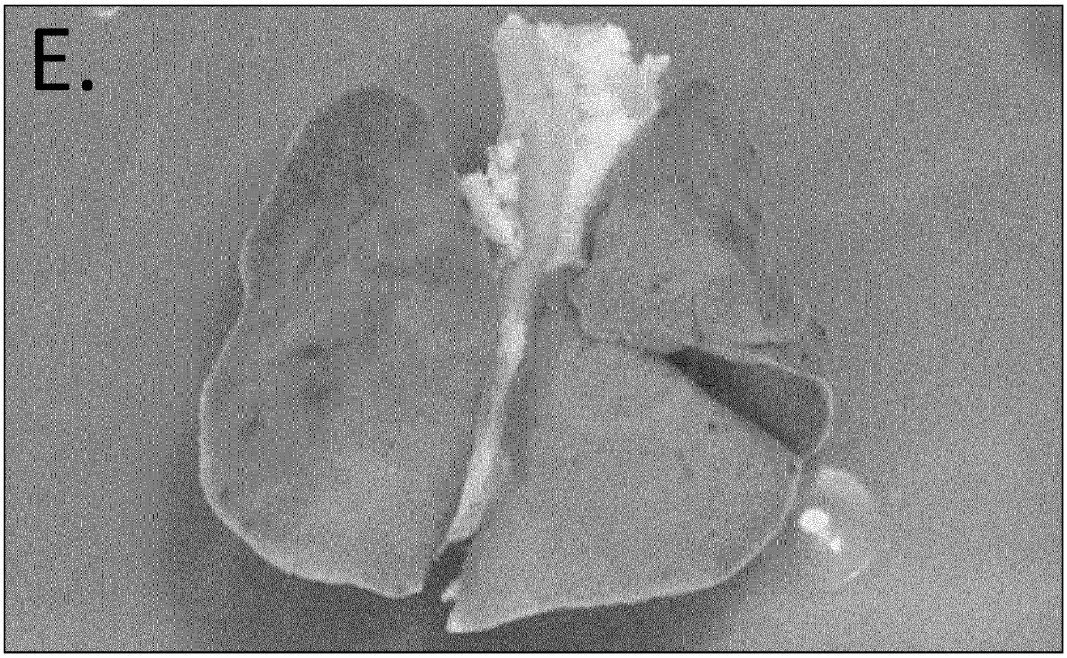


FIG. 13e

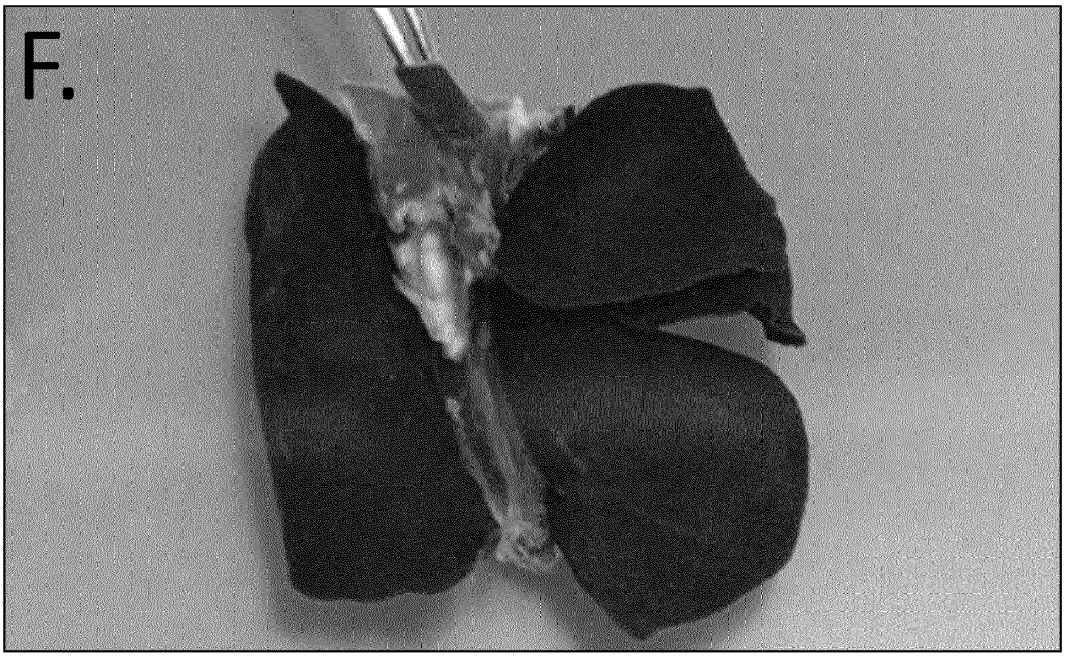


FIG. 13f

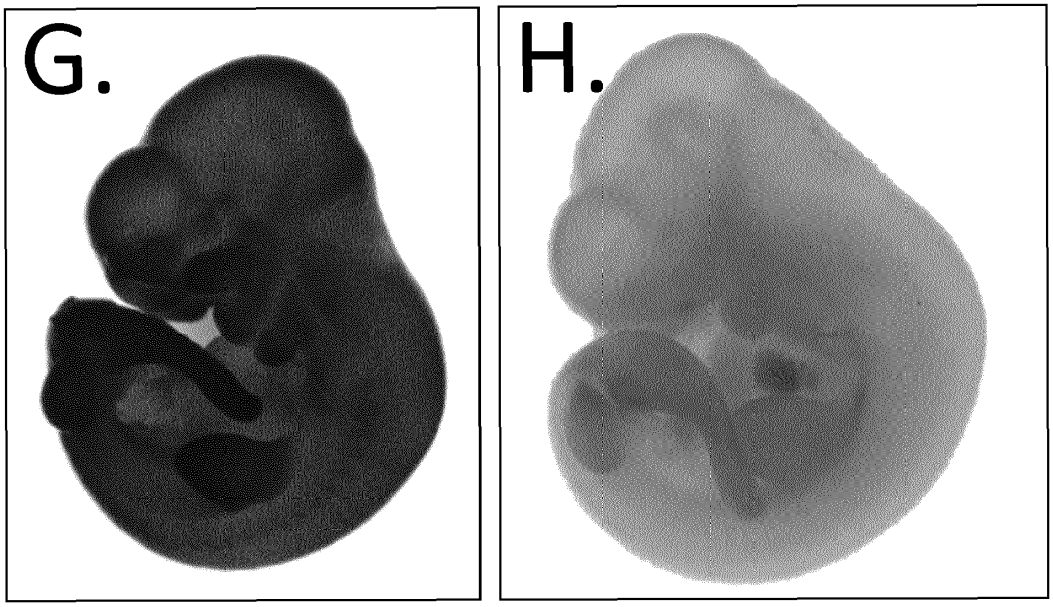


FIG. 13g, h

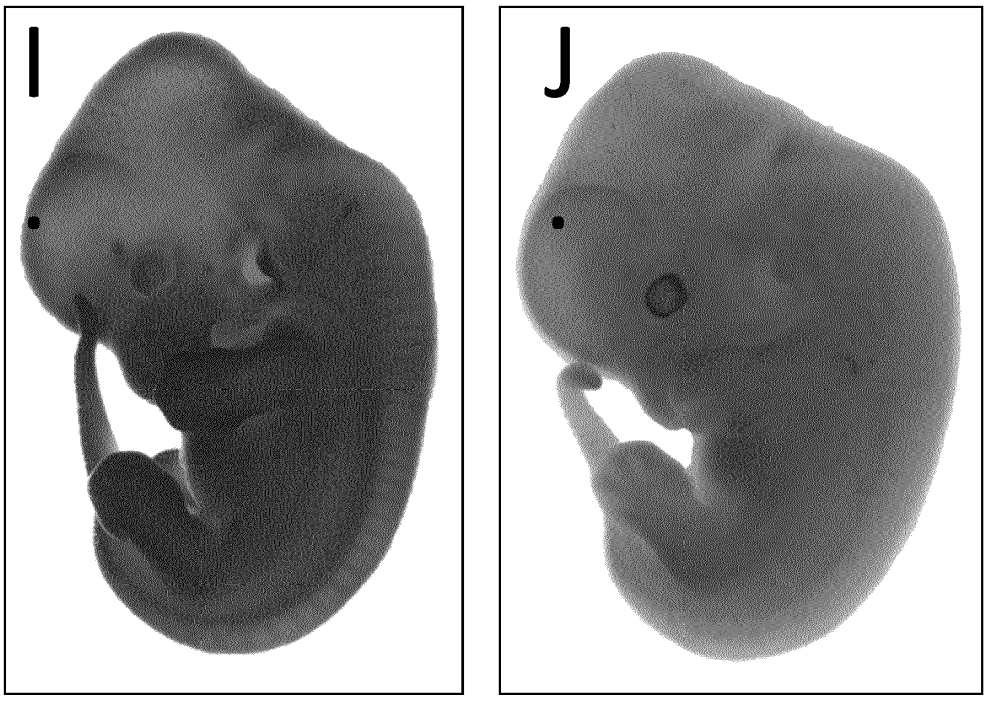


FIG. 13i,j

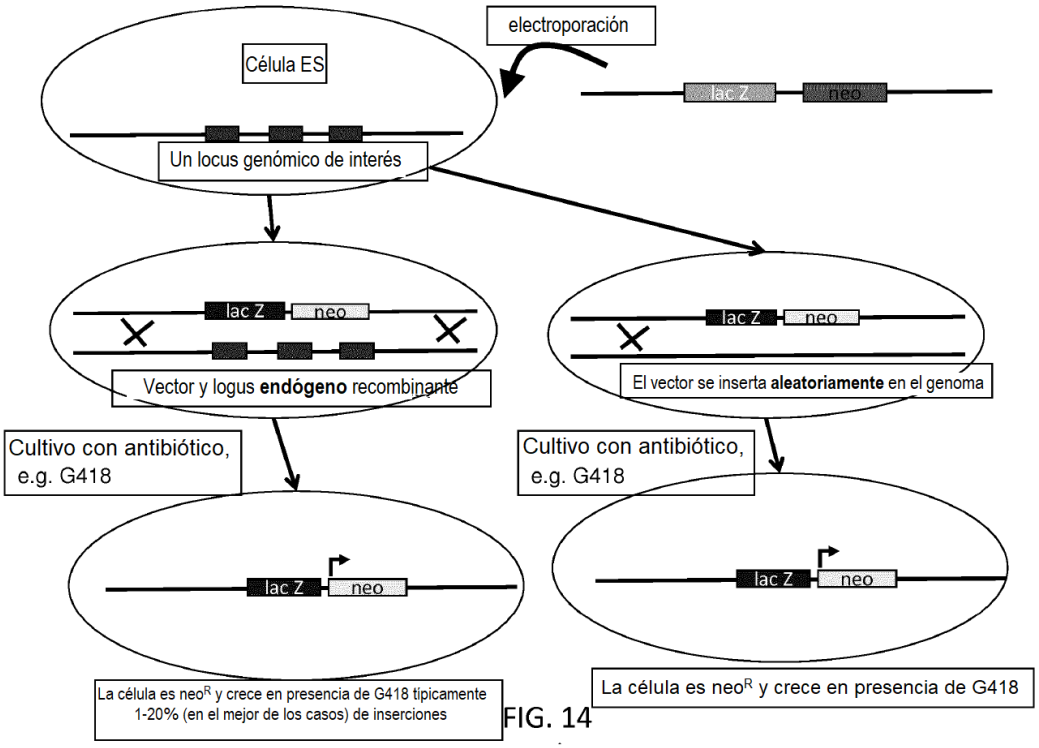

electroporación
Célula ES
lac Z
neo
Un locus genómico de interés
lac Z
neo
Vector y logus **endógeno** recombinante
lac Z
neo
El vector se inserta **aleatoriamente** en el genoma
Cultivo con antibiótico, e.g. G418
Cultivo con antibiótico, e.g. G418
lac Z
neo
lac Z
neo
La célula es neo^R y crece en presencia de G418 típicamente 1-20% (en el mejor de los casos) de inserciones
La célula es neo^R y crece en presencia de G418

FIG. 14

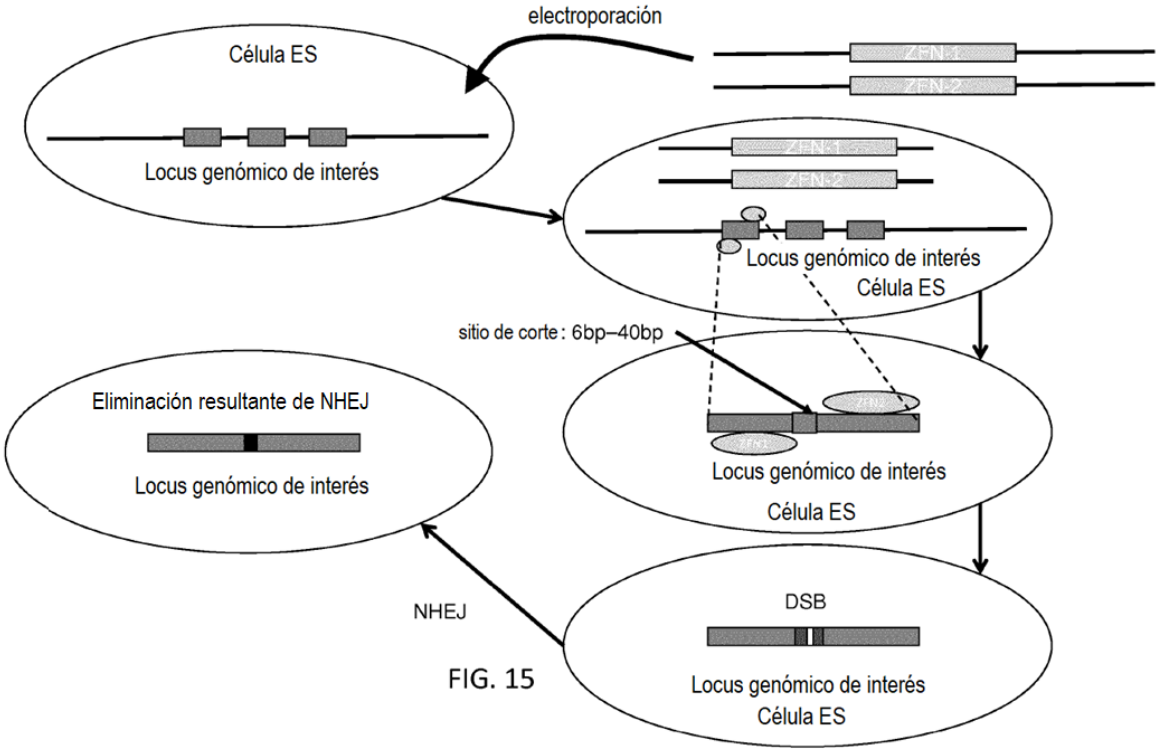
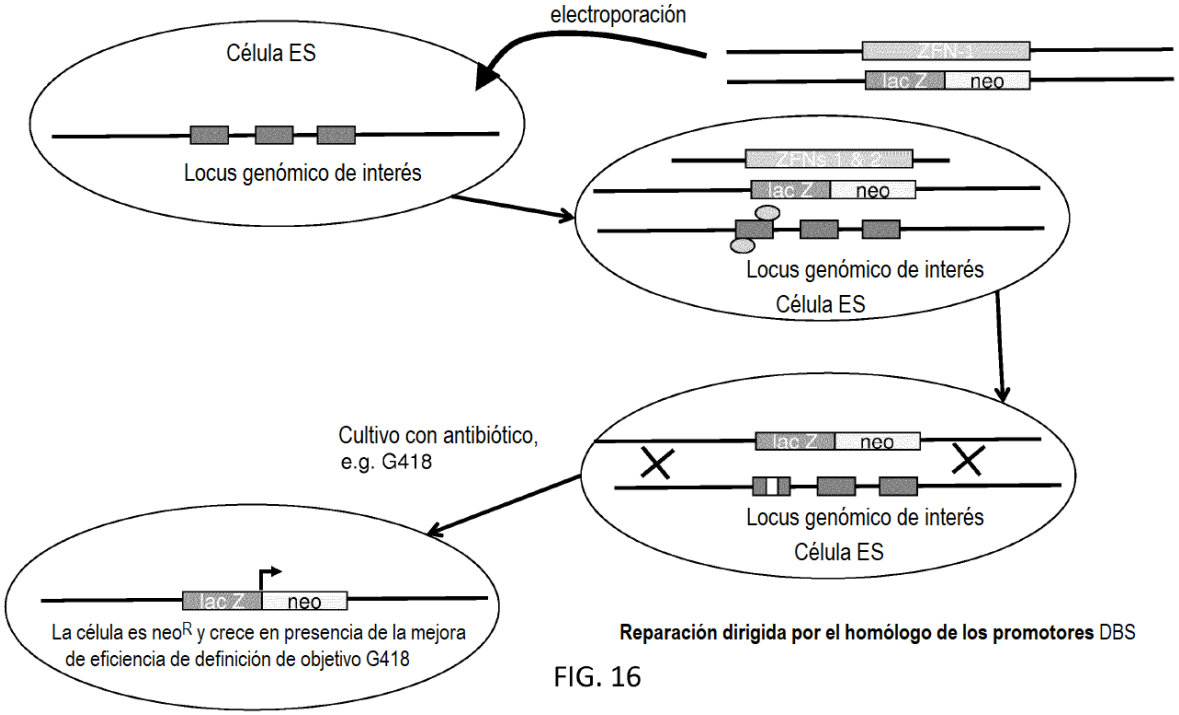

electroporación
Célula ES
Locus genómico de interés
ZFN-1
ZFN-2
Locus genómico de interés
Célula ES
sitio de corte: 6bp–40bp
Eliminación resultante de NHEJ
Locus genómico de interés
Locus genómico de interés
Célula ES
NHEJ
DSB
Locus genómico de interés
Célula ES

FIG. 15



Reparación dirigida por el homólogo de los promotores DBS

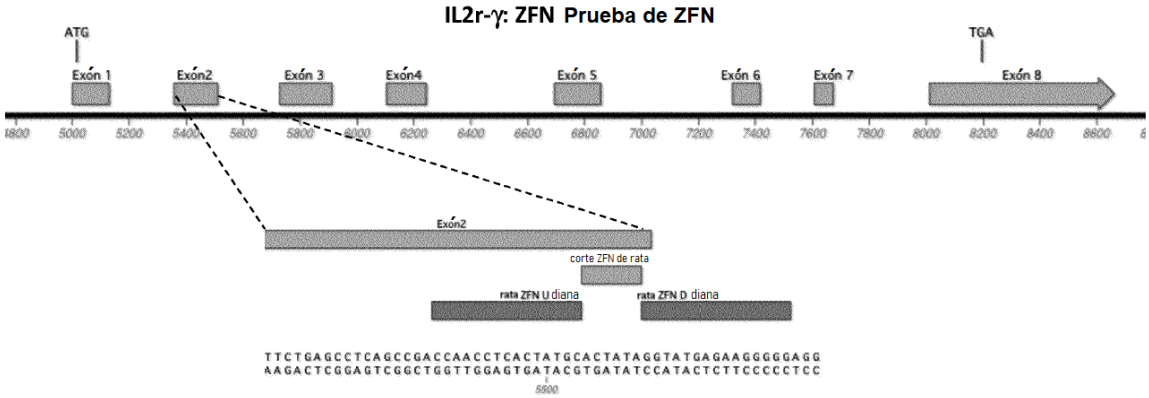
FIG. 16

Clon			Cahorros	Quimeras	
	ApoE-ZFN1-AB5 Homocigoto como diana		12	4 (90, 90, 80, 80)	
	ApoE-ZFN1-AE5 Homocigoto como diana		6	3 (90, 80, 70)	



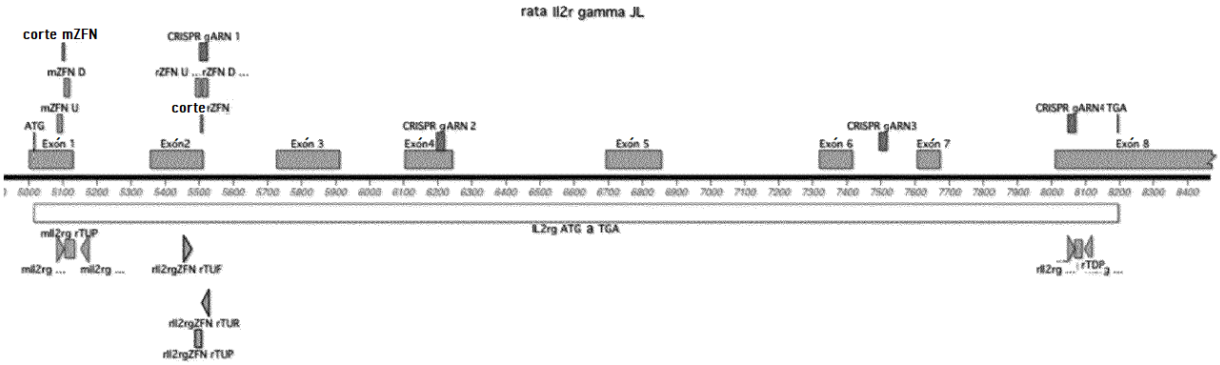
Quimeras ApoE-ZFN1-AB5

FIG. 17



Placa	Colonias cribadas	Como diana	Eficiencia
7/18: Vector solo	96	4	4.2%
7/18: Vector + ZFN	96	3	3.1%

FIG. 18



Constructo(s)	ADN (µg)	Colonias	Candidatos
Il2rg vector de plásmido	6 µg	30	3 (todos "débil")
plásmido + SBI g ARN1	6 µg/40 µg	22	1 fuerte, 1 débil
plásmido + SBI g ARN2	6 µg/40 µg	45	2 fuerte, 2 débil
plásmido + SBI g ARN3	6 µg/40 µg	66	1 fuerte, 2 débil
plásmido + SBI g ARN4	6 µg/40 µg	59	0

FIG. 19

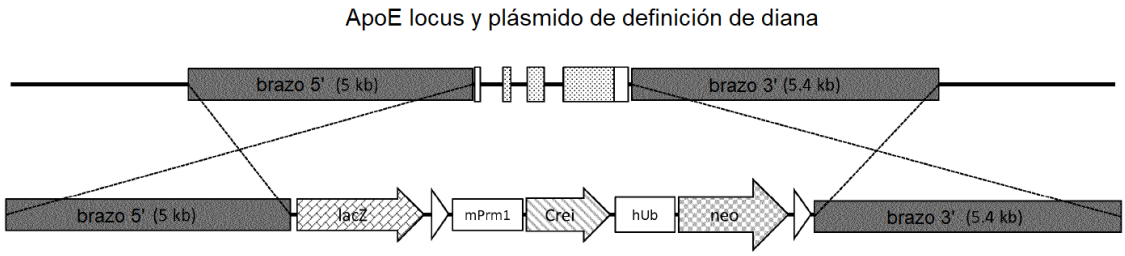
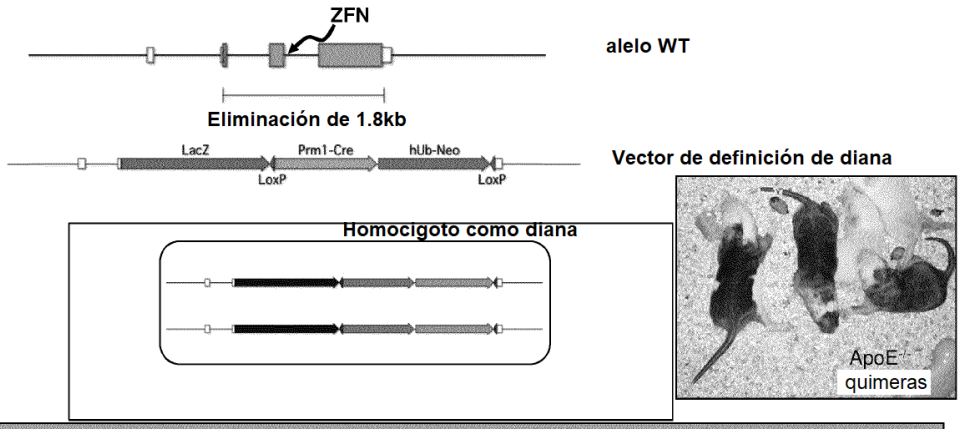

ApoE locus y plásmido de definición de diana
brazo 5' (5 kb)
brazo 3' (5.4 kb)
brazo 5' (5 kb)
lacZ
mPrm1
Crei
hUb
neo
brazo 3' (5.4 kb)

FIG. 20



ADN	Heterocigoto como diana	Homocigoto como diana	Microinyectada	Quimeras (% de quimerismo)	Cría
vector solo	15/192 (8%)	0			
vector + ZFN par 1	156/192 **(81%)**	**6/192 (3%)**	2 clones	7 (70-90%)	
vector + ZFN par 2	134/192 **(70%)**	**2/192 (1%)**			

FIG. 21

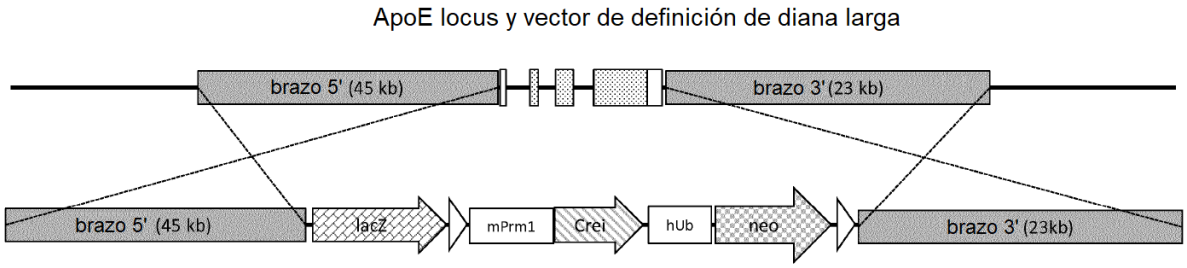

ApoE locus y vector de definición de diana larga
brazo 5' (45 kb)
brazo 3' (23 kb)
brazo 5' (45 kb)
lacZ
mPrm1
Crei
hUb
neo
brazo 3' (23kb)

FIG. 22

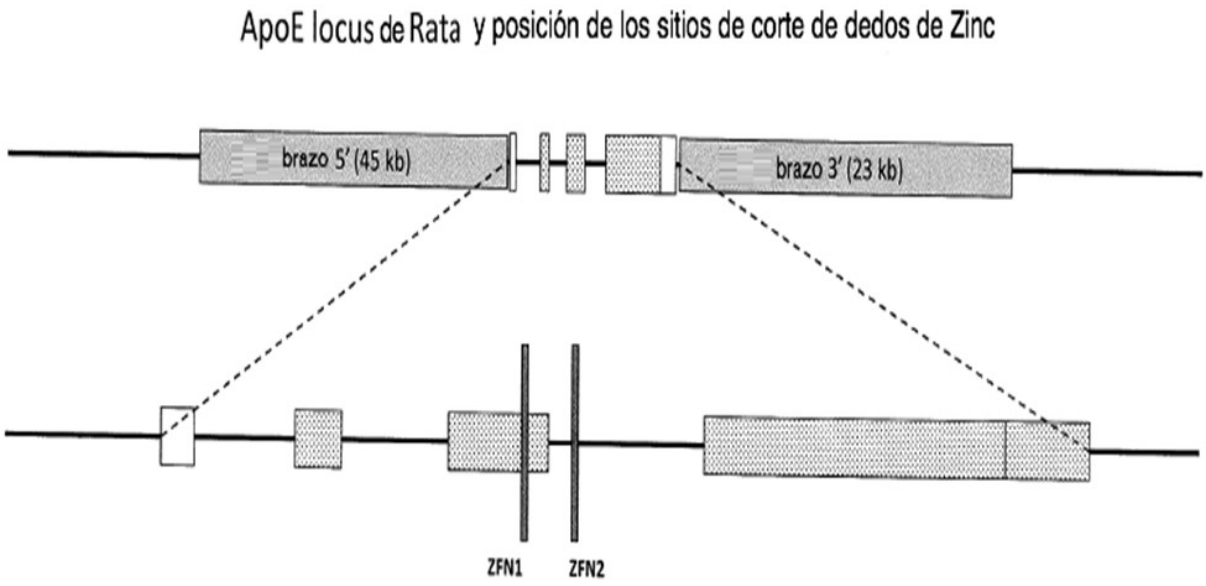

ApoE locus de Rata y posición de los sitios de corte de dedos de Zinc
brazo 5' (45 kb)
brazo 3' (23 kb)
ZFN1
ZFN2

FIG. 23

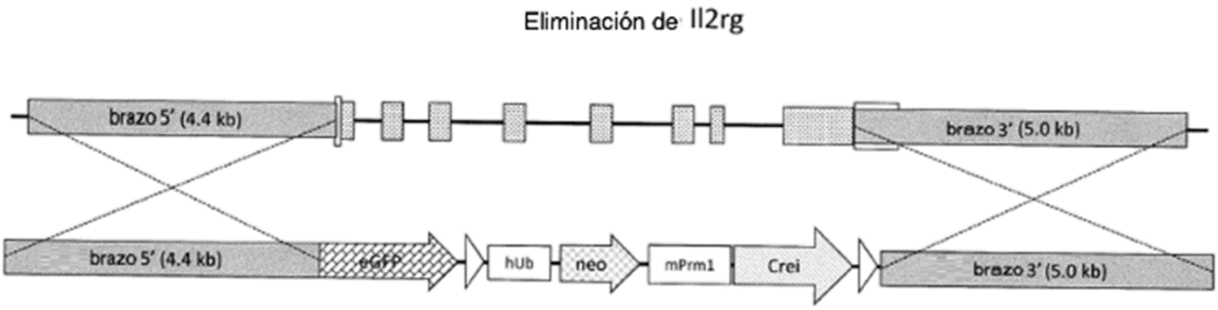

Eliminación de Il2rg
brazo 5' (4.4 kb)
brazo 3' (5.0 kb)
brazo 5' (4.4 kb)
hUb
neo
mPrm1
Crei
brazo 3' (5.0 kb)

FIG. 24

Transmisión de línea germinal confirmada

Línea	Cariotipo	Quimeras criadas	Transmisores de línea germinal	Cachorros derivados rESC	GLT (%)
ACI.G1	XY	5	3 (60%)	11/103	11%
Oscuro Agouti 2B	XY	5	4 (80%)	11/129	9%
Oscuro Agouti 2C	XX	3	2 (66%)	7/45	16%

Línea celular	Positivas	Eficiencia de definición de diana
DA.2B	8/192	4.2%
ACI.G1	4/192	2.1%



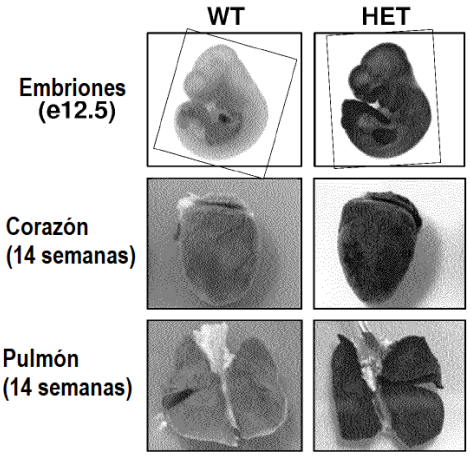
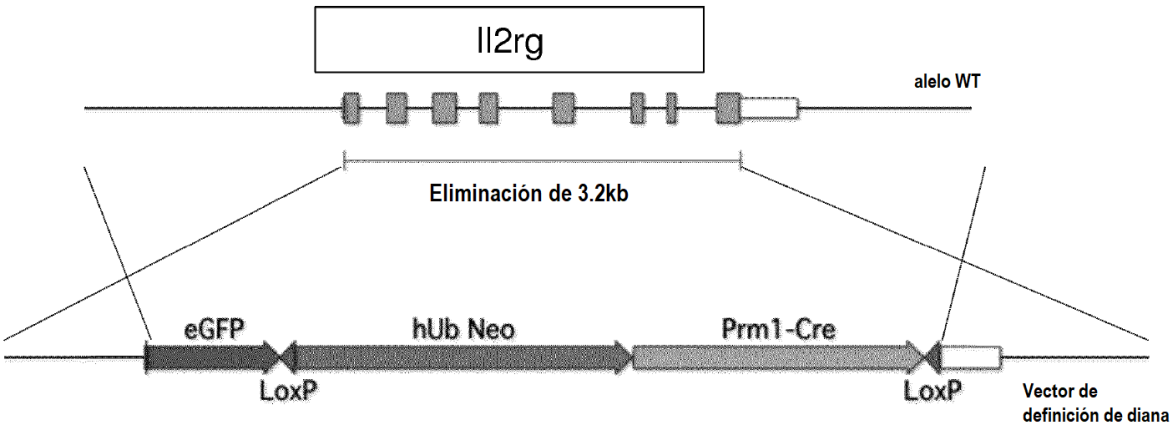


FIG. 25



Gen	Clones como diana		Quimeras (% de quimerismo)	
Il2rg	4/232 (2%)		5 (70 – 90%)	

FIG. 26

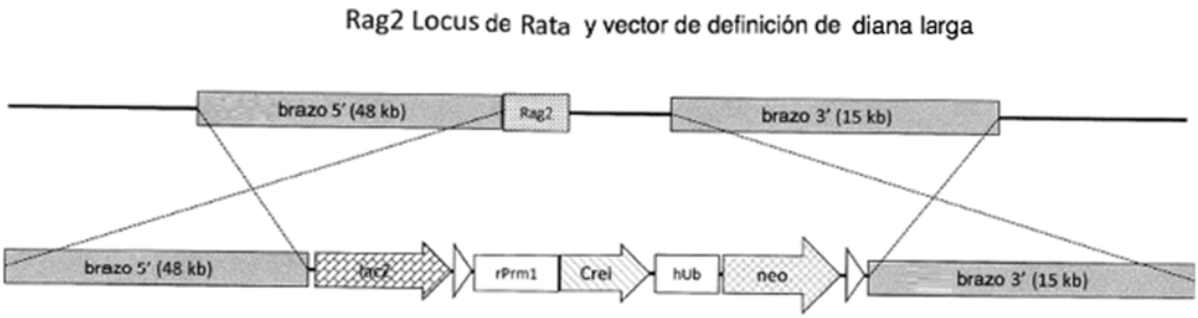

Rag2 Locus de Rata y vector de definición de diana larga
brazo 5' (48 kb)
Rag2
brazo 3' (15 kb)
brazo 5' (48 kb)
rPrm1
Crei
hUb
neo
brazo 3' (15 kb)

FIG. 27

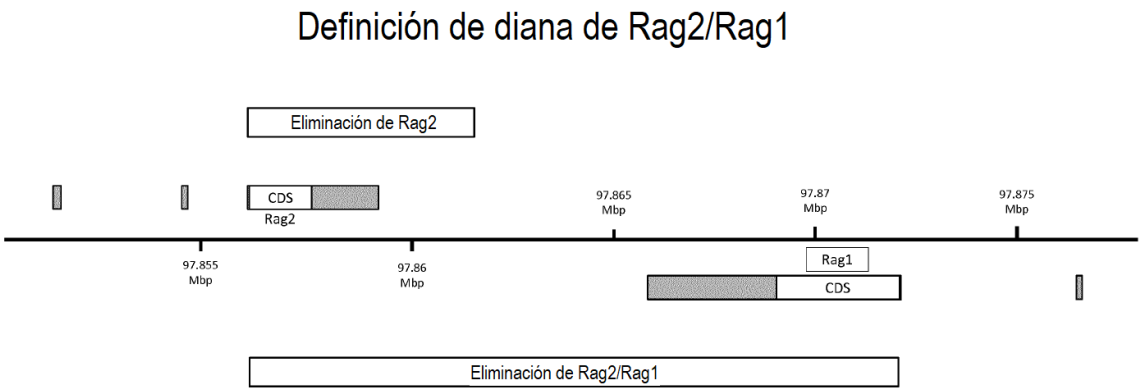

Definición de diana de Rag2/Rag1
Eliminación de Rag2
CDS
Rag2
97.865
Mbp
97.87
Mbp
97.875
Mbp
97.855
Mbp
97.86
Mbp
Rag1
CDS
Eliminación de Rag2/Rag1

FIG. 28

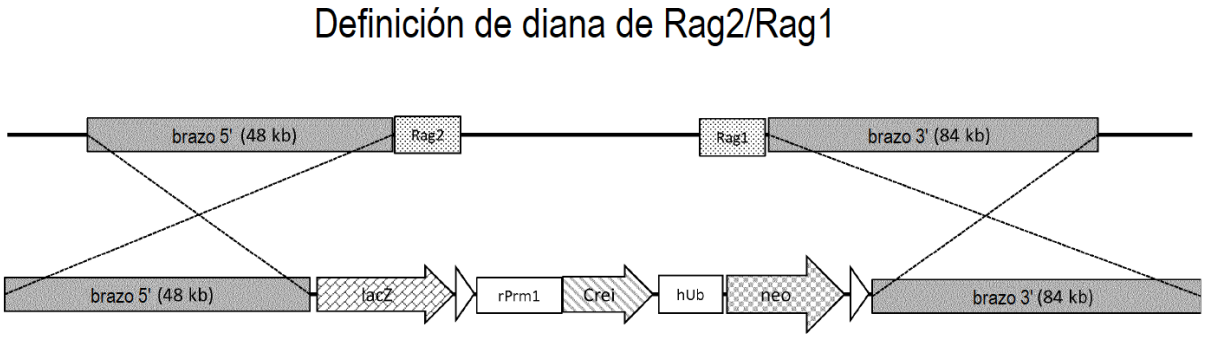

Definición de diana de Rag2/Rag1
brazo 5' (48 kb)
Rag2
Rag1
brazo 3' (84 kb)
brazo 5' (48 kb)
lacZ
rPrm1
Crei
hUb
neo
brazo 3' (84 kb)

FIG. 29

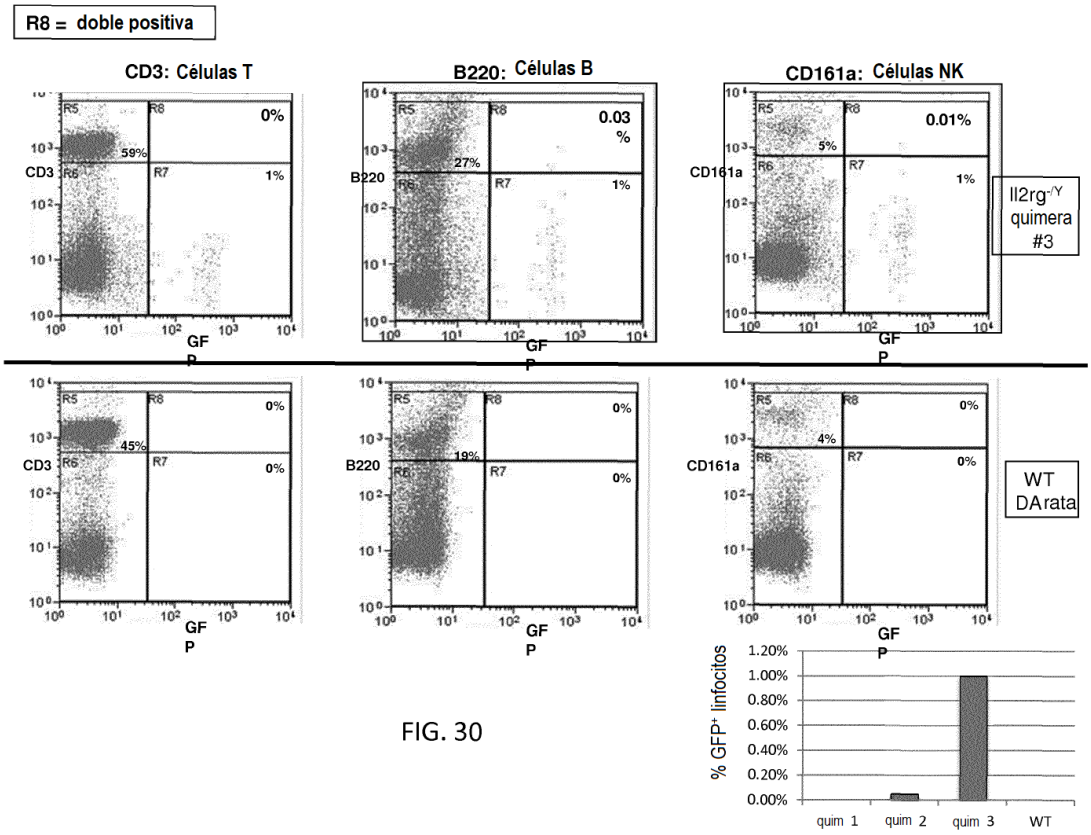

R8 = doble positiva
CD3: Células T
B220: Células B
CD161a: Células NK
0%
59%
1%
0.03 %
27%
1%
0.01%
5%
1%
Il2rg-/Y quimera #3
0%
45%
0%
0%
19%
0%
0%
4%
0%
WT DArata
GFP
CD3
B220
CD161a
% GFP+ linfocitos
1.20%
1.00%
0.80%
0.60%
0.40%
0.20%
0.00%
quim 1
quim 2
quim 3
WT

FIG. 30

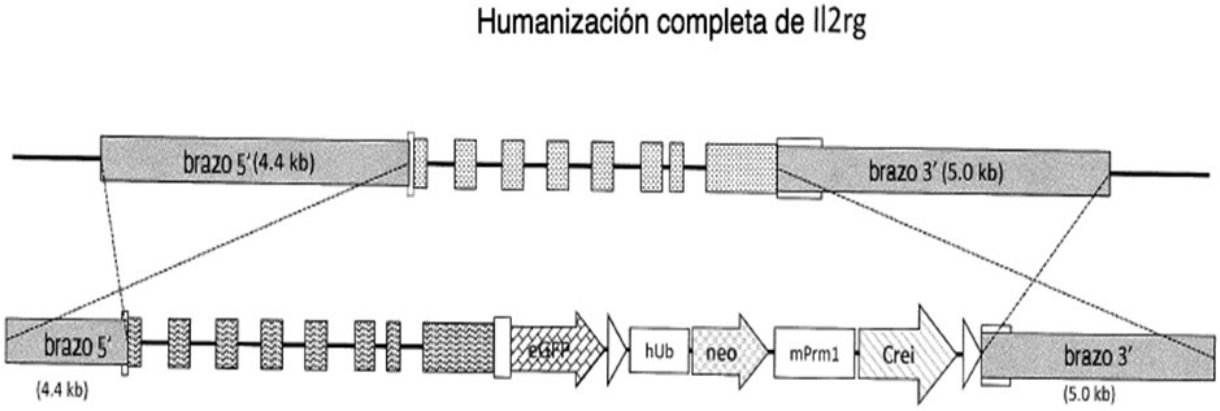

Humanización completa de Il2rg
brazo 5' (4.4 kb)
brazo 3' (5.0 kb)
brazo 5'
(4.4 kb)
eGFP
hUb
neo
mPrm1
Crei
brazo 3'
(5.0 kb)

FIG. 31

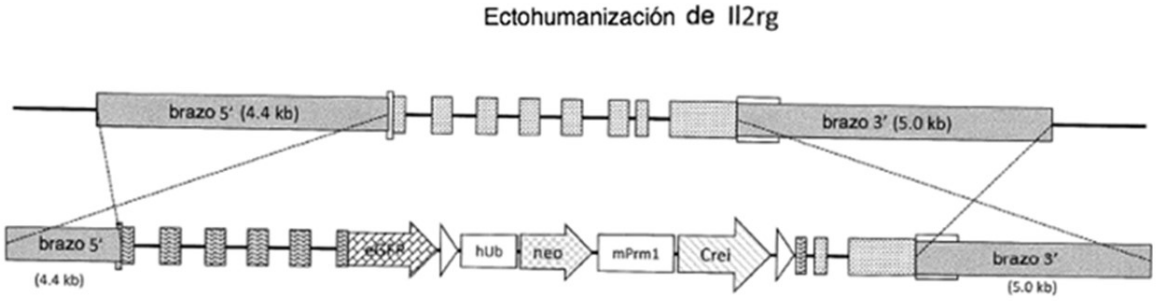

Ectohumanización de Il2rg
brazo 5' (4.4 kb)
brazo 3' (5.0 kb)
brazo 5'
(4.4 kb)
hUb
neo
mPrm1
Crei
brazo 3'
(5.0 kb)

FIG. 32

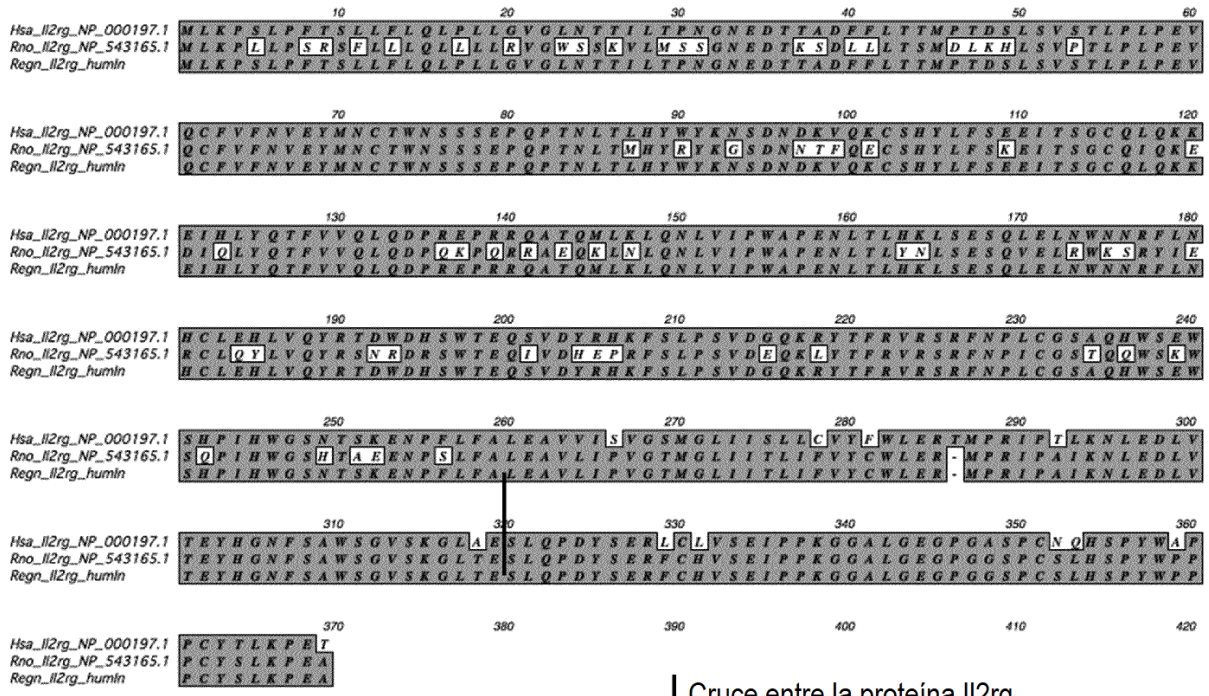

Hsa_Il2rg_NP_000197.1
Rno_Il2rg_NP_543165.1
Regn_Il2rg_humIn
Cruce entre la proteína Il2rg humana/rata

FIG. 33