

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 966 625**

51 Int. Cl.:

A01K 67/027 (2006.01)

C07K 14/745 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.04.2020** **PCT/US2020/026413**

87 Fecha y número de publicación internacional: **08.10.2020** **WO20206139**

96 Fecha de presentación y número de la solicitud europea: **02.04.2020** **E 20722746 (3)**

97 Fecha y número de publicación de la concesión europea: **15.11.2023** **EP 3772929**

54 Título: **Roedores que comprenden un locus del factor de coagulación 12 humanizado**

30 Prioridad:

04.04.2019 US 201962829321 P

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

23.04.2024

73 Titular/es:

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

72 Inventor/es:

TANG, YAJUN;
CHALOTHORN, DAN;
MITNAUL, LYNDON;
MORTON, LORI;
ZAMOLODCHIKOV, DARIA;
ALESSANDRI-HABER, NICOLE y
MACDONALD, LYNN

74 Agente/Representante:

SÁEZ MAESO, Ana

Observaciones:

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

ES 2 966 625 T3

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

DESCRIPCIÓN

Roedores que comprenden un locus del factor de coagulación 12 humanizado

5 Antecedentes

Los factores de coagulación, junto con las plaquetas, son componentes sanguíneos relevantes en el proceso de la sangre durante la lesión de un vaso. Está bien establecido que estos componentes pueden ser impulsores de la trombosis cuando se producen desequilibrios en la regulación (es decir, producción y/o actividad). Se creía que las enfermedades trombóticas surgían principalmente de una activación aberrante en la vía extrínseca (a través del factor tisular), pero más recientemente, estudios preclínicos con ratones deficientes *F12* y moléculas que apuntan al factor XII de coagulación activado (factor XIIa) sugirieron que la vía intrínseca (también conocida como vía de contacto) de la coagulación también está involucrada. El factor XIIa es fundamental para la vía de contacto y puede impulsar la coagulación mediante la escisión del factor XI o puede impulsar la inflamación mediante la escisión de la precalicreína plasmática. Por tanto, la inhibición de la actividad del factor XII puede conducir a una reducción de la coagulación trombótica y a la inflamación asociada con la activación de la vía de contacto.

Sin embargo, sigue existiendo la necesidad de animales no humanos adecuados que proporcionen el verdadero objetivo humano o una aproximación cercana al verdadero objetivo humano de los reactivos orientados al factor XII de coagulación humano, permitiendo así probar la eficacia y el modo de acción de dichos agentes en animales vivos, así como estudios farmacocinéticos y farmacodinámicos.

Resumen

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

Se proporcionan roedores que comprenden un locus (*F12*) del factor XII de coagulación humanizado tal como se caracteriza en las reivindicaciones, así como métodos para fabricar y utilizar dichos roedores. También se proporcionan genomas o células de roedores que comprenden un locus (*F12*) del factor XII de coagulación humanizado, así como métodos para producir y utilizar dichos genomas o células de roedores. También se describen genes *F12* de animales no humanos humanizados, agentes nucleasa y/o vectores orientados para su uso en la humanización de un gen *F12* de animales no humanos, y métodos para producir y usar dichos genes *F12* humanizados.

En un aspecto, se proporcionan genomas de roedores, células de roedores o roedores que comprenden un locus (*F12*) del factor XII de coagulación humanizado como se caracteriza en las reivindicaciones. Dichos genomas de roedor, células de roedor o roedores comprenden en su genoma un locus *F12* endógeno humanizado en donde un segmento del locus *F12* endógeno ha sido eliminado y reemplazado con una secuencia de *F12* humano correspondiente.

En tales genomas de roedores, células de roedores o roedores, una región del locus *F12* endógeno que comprende tanto una secuencia codificante como una secuencia no codificante se ha eliminado y reemplazado con una secuencia de *F12* humano correspondiente que comprende tanto una secuencia codificante como una secuencia no codificante. En algunos de estos genomas de roedores, células de roedores o animales de roedores, el locus *F12* endógeno humanizado comprende el promotor *F12* endógeno, en donde la secuencia de *F12* humana está operativamente ligada al promotor *F12* endógeno. En algunos de estos genomas de roedores, células de roedores o roedores, al menos un intrón y al menos un exón del locus *F12* endógeno se han eliminado y reemplazado con la secuencia de *F12* humano correspondiente.

En tales genomas de roedores, células de roedores o roedores, la región del locus *F12* endógeno desde el codón de inicio hasta el codón de parada se ha eliminado y reemplazado con la secuencia de *F12* humano correspondiente.

En algunos de estos genomas de roedores, células de roedores o roedores, la región no traducida 3' endógena de *F12* no ha sido eliminada ni reemplazada por la secuencia de *F12* humano correspondiente. En algunos de estos genomas de roedores, células de roedores o roedores, la región no traducida 5' endógena de *F12* no ha sido eliminada ni reemplazada por la secuencia de *F12* humano correspondiente.

En algunos de tales genomas de roedores, células de roedores o roedores, la región del locus *F12* endógeno desde el codón de inicio hasta el codón de parada se ha eliminado y reemplazado con la secuencia de *F12* humano correspondiente, y el promotor *F12* endógeno no se ha eliminado y reemplazado con la secuencia de *F12* humano correspondiente.

En algunos de tales genomas de roedores, células de roedores o roedores, la secuencia de *F12* humano en el locus *F12* endógeno humanizado comprende una secuencia de al menos 90 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % idéntica a la secuencia expuesta en la SEQ ID NO: 31 o 32. En algunos de tales genomas de roedores, células de roedores o roedores, el locus *F12* endógeno humanizado codifica una proteína que comprende una secuencia al menos 90%, 95%, 96%, 97%, 98%, 99% o 100% idéntica a la secuencia establecida en la SEQ ID NO: 5 o 46. En algunos de tales genomas de roedores, células de roedores o roedores, el locus *F12* endógeno humanizado

comprende una secuencia codificante que comprende una secuencia al menos 90%, 95%, 96%, 97%, 98%, 99% o 100% idéntica a la secuencia establecida en la SEQ ID NO: 13 o 48. En algunos de tales genomas de roedores, células de roedores o roedores, el locus *F12* endógeno humanizado comprende una secuencia al menos 90 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % idéntica a la secuencia expuesta en la SEQ ID NO: 17 o 18.

- 5 En algunos de estos genomas de roedores, células de roedores o roedores, el locus *F12* endógeno humanizado no comprende un casete de selección ni un gen indicador. En algunos de estos genomas de roedores, células de roedores o roedores, el locus *F12* endógeno humanizado comprende un casete de selección o un gen indicador dentro de un intrón de la secuencia de *F12* humano correspondiente.

- 10 En algunos de tales genomas de roedores, células de roedores o roedores, el roedor es homocigoto para el locus *F12* endógeno humanizado.

En algunos de estos genomas de roedores, células de roedores o roedores, el roedor comprende el locus *F12* endógeno humanizado en su línea germinal.

- 15 Opcionalmente, el roedor es una rata o un ratón. Opcionalmente, el roedor es un ratón.

En tales genomas de roedor, células de roedor o roedores, el roedor expresa una isoforma corta del factor XII de coagulación humano en el cerebro, en donde la isoforma corta del factor XII de coagulación humano está codificada por un ARNm de *F12* detectado mediante sondas contra los exones 11-14 de *F12* humano, pero no mediante sondas contra los exones 1-6 de *F12* humano. Opcionalmente, la isoforma corta del factor XII de coagulación humano es FXII₂₉₇₋₅₉₆. Opcionalmente, la isoforma corta del factor XII de coagulación humano se expresa en neuronas del cerebro. Opcionalmente, la isoforma corta del factor XII de coagulación humano puede activarse mediante calicreína plasmática y convertir pro-HGF en HGF activo.

- 20 En algunos de tales genomas de roedor, células de roedor o roedores, el roedor tiene un tiempo de tromboplastina parcial activado (aPTT) que no es significativamente diferente de un aPTT en un animal de tipo silvestre correspondiente y/o en donde los roedores tienen un tiempo de protrombina (PT) que no es significativamente diferente de un PT en un animal de tipo silvestre correspondiente.

En algunos de estos genomas de roedores, células de roedores o roedores, el roedor tiene niveles plasmáticos del factor XII de coagulación humano de al menos aproximadamente 5 µg/ml.

- 30 En algunos de estos genomas de roedores, células de roedores o roedores, la región del locus *F12* endógeno desde el codón de inicio hasta el codón de parada ha sido eliminada y reemplazada con la secuencia de *F12* humano correspondiente, el promotor *F12* endógeno no se ha eliminado ni reemplazado con la secuencia de *F12* humano correspondiente, y el roedor expresa una isoforma corta del factor XII de coagulación humano en el cerebro, opcionalmente en donde: (i) la isoforma corta del factor XII de coagulación humano está codificada por un ARNm de *F12* detectado por sondas contra los exones 11-14 del *F12* humano pero no por sondas contra los exones 1-6 del *F12* humano; y/o (ii) la isoforma corta del factor XII de coagulación humano es FXII₂₉₇₋₅₉₆; y/o (iii) la isoforma corta del factor XII de coagulación humano se expresa en neuronas del cerebro; y/o (iv) la isoforma corta del factor XII de coagulación humano puede activarse mediante calicreína plasmática y convertir pro-HGF en HGF activo.

- 35 Además, en el presente documento se describen vectores orientados para generar un locus *F12* endógeno humanizado en donde un segmento del locus *F12* endógeno ha sido eliminado y reemplazado con una secuencia de *F12* humano correspondiente, en donde el vector orientado comprende un ácido nucleico de inserto que comprende la secuencia de *F12* humana correspondiente flanqueada por un brazo de homología 5' orientado a una secuencia objetivo 5' en el locus *F12* endógeno y un brazo de homología 3' orientado a una secuencia objetivo 3' en el locus *F12* endógeno.

- 40 Además, en el presente documento se describen genes *F12* de animales no humanos humanizados en donde un segmento del locus *F12* endógeno ha sido eliminado y reemplazado con una secuencia de *F12* humano correspondiente.

- 45 En otro aspecto, se proporcionan métodos para evaluar la actividad de un reactivo orientado al factor XII de coagulación humano *in vivo*. Dichos métodos comprenden: (a) administrar el reactivo orientado al factor XII de coagulación humano al roedor descrito anteriormente; y (b) evaluar la actividad del reactivo orientado al factor XII de coagulación humano en el roedor.

- 50 En algunos de estos métodos, la administración comprende administración mediada por virus adenoasociados (AAV), administración mediada por nanopartículas lipídicas (LNP), administración hidrodinámica (HDD) o inyección.

- 55 En algunos de tales métodos, el paso (b) comprende aislar plasma del roedor y evaluar la actividad del reactivo orientado al factor XII de coagulación humano en el plasma. En algunos de tales métodos, el paso (b) comprende evaluar la actividad del reactivo orientado al factor XII de coagulación humano en el hígado o el cerebro del roedor. En algunos de dichos métodos, el paso (b) comprende evaluar la coagulación o la generación de trombina. En algunos de estos métodos, el paso (b) comprende evaluar la señalización de HGF-Met. En algunos de tales métodos, el paso

(b) comprende medir la expresión de un ARN mensajero *F12* codificado por el locus *F12* endógeno humanizado. En algunos de dichos métodos, el paso (b) comprende medir la expresión de una proteína del factor XII de coagulación codificada por el locus del factor XII de coagulación endógeno humanizado. En algunos de tales métodos, el reactivo orientado al factor XII de coagulación humano es un agente de edición del genoma, y el paso (b) comprende evaluar la modificación del locus *F12* endógeno humanizado. En algunos de dichos métodos, el paso (b) comprende medir la frecuencia de inserciones o deleciones dentro del locus *F12* endógeno humanizado.

En algunos de tales métodos, el reactivo orientado al factor XII de coagulación humano comprende un agente nucleasa diseñado para orientarse a una región de un gen *F12* humano. Opcionalmente, el agente nucleasa comprende una proteína Cas y un ARN guía diseñado para orientar a una secuencia objetivo de ARN guía en el gen *F12* humano. Opcionalmente, la proteína Cas es una proteína Cas9.

En algunos de dichos métodos, el reactivo orientado al factor XII de coagulación humano comprende un ácido nucleico donante exógeno, en donde el ácido nucleico donante exógeno está diseñado para orientarse al gen *F12* humano, y opcionalmente en donde el ácido nucleico del donante exógeno se administra a través de AAV. En algunos de estos métodos, el reactivo orientado al factor XII de coagulación humano es un agente de ARNi o un oligonucleótido antisentido. En algunos de estos métodos, el reactivo orientado al factor XII de coagulación humano es una proteína de unión a antígeno. En algunos de estos métodos, el reactivo orientado al factor XII de coagulación humano es una molécula pequeña.

En otro aspecto, se proporcionan métodos para optimizar la actividad de un reactivo orientado al factor XII de coagulación humano *in vivo*. Dichos métodos comprenden: (I) realizar cualquiera de los métodos anteriores por primera vez en un primer roedor que comprende en su genoma un locus *F12* endógeno humanizado; (II) cambiar una variable y realizar el método del paso (I) una segunda vez con la variable cambiada en un segundo roedor que comprende en su genoma un locus *F12* endógeno humanizado; y (III) comparar la actividad del reactivo orientado al factor XII de coagulación humano en el paso (I) con la actividad del reactivo orientado al factor XII de coagulación humano en el paso (II), y seleccionar el método resultante en la actividad superior.

En algunos de tales métodos, la variable cambiada en el paso (II) es el método de administración para introducir el reactivo orientado al factor XII de coagulación humano en el roedor. En algunos de tales métodos, la variable cambiada en el paso (II) es la vía de administración para introducir el reactivo orientado al factor XII de coagulación humano en el roedor. En algunos de tales métodos, la variable cambiada en el paso (II) es la concentración o cantidad del reactivo orientado al factor XII de coagulación humano introducido en el roedor. En algunos de tales métodos, la variable cambiada en el paso (II) es la forma del reactivo orientado al factor XII de coagulación humano introducido en el roedor. En algunos de tales métodos, la variable cambiada en el paso (II) es el reactivo orientado al factor XII de coagulación humano introducido en el roedor.

En otro aspecto, se proporcionan métodos para fabricar cualquiera de los roedores anteriores. Algunos de estos métodos comprenden: (a) introducir en una célula madre embrionaria (ES) de roedor: (i) un agente nucleasa que se orienta a una secuencia objetivo en el locus *F12* endógeno; y (ii) un vector orientado que comprende un inserto de ácido nucleico que comprende la secuencia de *F12* humano flanqueada por un brazo de homología 5' correspondiente a una secuencia objetivo 5' en el locus *F12* endógeno y un brazo de homología 3' correspondiente a una secuencia objetivo 3' en el locus *F12* endógeno, en donde el vector orientado se recombina con el locus *F12* endógeno para producir una célula ES de roedor modificada genéticamente que comprende en su genoma el locus *F12* endógeno humanizado que comprende la secuencia de *F12* humano; (b) introducir la célula ES de roedor modificada genéticamente en un embrión huésped de roedor; y (c) gestar el embrión huésped de roedor en una madre sustituta, en donde la madre sustituta produce una progenie F0 de roedor modificado genéticamente que comprende en su genoma el locus *F12* endógeno humanizado que comprende la secuencia de *F12* humano. Opcionalmente, el vector orientado es un vector orientado grande de al menos 10 kb de longitud o en donde la suma total de los brazos de homología 5' y 3' tiene al menos 10 kb de longitud.

Algunos de estos métodos comprenden: (a) introducir en un embrión unicelular de roedor: (i) un agente nucleasa que se orienta a una secuencia objetivo en el locus *F12* endógeno; y (ii) un vector orientado que comprende un inserto de ácido nucleico que comprende la secuencia de *F12* humano flanqueada por un brazo de homología 5' correspondiente a una secuencia objetivo 5' en el locus *F12* endógeno y un brazo de homología 3' correspondiente a una secuencia objetivo 3' en el locus *F12* endógeno, en donde el vector orientado se recombina con el locus *F12* endógeno para producir un embrión unicelular de roedor modificado genéticamente que comprende en su genoma el locus *F12* endógeno humanizado que comprende la secuencia de *F12* humano; (b) gestar el embrión unicelular de roedor genéticamente modificado en una madre sustituta para producir un roedor de generación F0 genéticamente modificado que comprende en su genoma el locus *F12* endógeno humanizado que comprende la secuencia de *F12* humano.

En algunos de tales métodos, el agente nucleasa comprende una proteína Cas y un ARN guía. Opcionalmente, la proteína Cas es una proteína Cas9. En algunos de tales métodos, el paso (a) comprende además introducir un segundo ARN guía que se orienta a una segunda secuencia objetivo dentro del locus *F12* endógeno.

Algunos de estos métodos comprenden: (a) introducir en un embrión huésped de roedor una célula madre embrionaria (ES) de roedor modificada genéticamente que comprende en su genoma un locus *F12* endógeno humanizado en

donde un segmento del locus *F12* endógeno ha sido eliminado y reemplazado con una secuencia de *F12* humano correspondiente; y (b) gestar el embrión huésped del roedor en una madre sustituta, en donde la madre sustituta produce una progenie F0 de un roedor genéticamente modificado que comprende el locus *F12* endógeno humanizado.

- Algunos de estos métodos comprenden: (a) introducir en una célula madre embrionaria (ES) de roedor un vector orientado que comprende un inserto de ácido nucleico que comprende la secuencia de *F12* humano flanqueada por un brazo de homología 5' correspondiente a una secuencia objetivo 5' en el locus *F12* endógeno y un brazo de homología 3' correspondiente a una secuencia objetivo 3' en el locus *F12* endógeno, en donde el vector orientado se recombina con el locus *F12* endógeno para producir una célula ES de roedor modificada genéticamente que comprende en su genoma el locus *F12* endógeno humanizado que comprende la secuencia de *F12* humano; (b) introducir la célula ES de roedor modificada genéticamente en un embrión huésped de roedor; y (c) gestar el embrión huésped de roedor en una madre sustituta, en donde la madre sustituta produce una progenie F0 de roedor modificado genéticamente que comprende en su genoma el locus *F12* endógeno humanizado que comprende la secuencia de *F12* humano. Opcionalmente, el vector orientado es un vector orientado grande de al menos 10 kb de longitud o en donde la suma total de los brazos de homología 5' y 3' tiene al menos 10 kb de longitud.
- Algunos de tales métodos comprenden: (a) introducir en un embrión unicelular de roedor un vector orientado que comprende un inserto de ácido nucleico que comprende la secuencia de *F12* humano flanqueada por un brazo de homología 5' correspondiente a una secuencia objetivo 5' en el locus *F12* endógeno y un brazo de homología 3' correspondiente a una secuencia objetivo 3' en el locus *F12* endógeno, en donde el vector orientado se recombina con el locus *F12* endógeno para producir un embrión unicelular de roedor genéticamente modificado que comprende en su genoma el locus *F12* endógeno humanizado que comprende la secuencia de *F12* humano; (b) gestar el embrión unicelular de roedor genéticamente modificado en una madre roedor sustituta para producir un roedor de generación F0 genéticamente modificado que comprende en su genoma el locus *F12* endógeno humanizado que comprende la secuencia de *F12* humano, en donde el roedor de generación F0 genéticamente modificado expresa la isoforma corta del factor XII de coagulación humano en el cerebro.
- En algunos de dichos métodos, el paso (a) comprende además introducir un agente nucleasa o un ácido nucleico que codifica el agente nucleasa, en donde el agente nucleasa se orienta a una secuencia objetivo en el locus *F12* endógeno. Opcionalmente, el agente nucleasa comprende una proteína Cas y un ARN guía. Opcionalmente, la proteína Cas es una proteína Cas9. Opcionalmente, el paso (a) comprende además introducir un segundo ARN guía que se orienta a una segunda secuencia objetivo dentro del locus *F12* endógeno.
- En algunos de estos métodos, el roedor es un ratón o una rata. Opcionalmente, el roedor es un ratón.

En otro aspecto, se proporcionan métodos que evalúan la expresión de una isoforma corta de ARNm *F12* en una muestra biológica. Dichos métodos comprenden: (a) detectar y cuantificar transcritos de ARN usando cebadores y/o sondas contra uno o más de los exones 9-14 de un gen *F12*; y (b) detectar y cuantificar transcritos de ARN usando cebadores y/o sondas contra uno o más de los exones 1-6 del gen *F12* en donde la isoforma corta se detecta mediante cebadores y/o sondas contra los exones 9-14 pero no mediante cebadores y/o sondas contra los exones 1-6.

Breve descripción de las figuras

La Figura 1 (no a escala) muestra un esquema del esquema de orientación para la humanización del locus (*F12*) del factor XII de coagulación de ratón. La parte superior de la figura muestra el locus *F12* endógeno de ratón y el locus *F12* endógeno humano, y la porción inferior de la figura muestra el locus humanizado con o sin el casete de selección de autoeliminación.

La Figura 2 (no a escala) muestra un esquema de los ensayos TAQMAN® para detectar la humanización del locus *F12* de ratón. Los ensayos de ganancia de alelo (GOA) incluyen 7374hU y 7374hD. Los ensayos de pérdida de alelo (LOA) incluyen 7374mTU y 7374mTD.

La Figura 3 muestra una alineación de las proteínas del factor XII de coagulación humano y de ratón (hF12 y mF12, respectivamente). Los péptidos señal están encuadrados y se anotan las regiones de cadena pesada y cadena ligera.

La Figura 4 muestra los niveles del factor XII de coagulación humano en muestras de plasma de ratones *F12* humanizados y ratones de tipo silvestre (VG WT). Se utilizó plasma humano normal combinado como control positivo y plasma humano FXII-ID (FXII-inmunodepletado) como control negativo. Estudiantes *t*-prueba; *** para $p < 0,001$ frente al plasma humano normal para el plasma humano analizado o frente al plasma natural para el plasma de ratón.

La Figura 5 muestra el análisis de transferencia Western de muestras de plasma de ratones de tipo silvestre (FXII^{+/+}), ratones inactivados para *F12* (FXII^{-/-}) y ratones *F12* humanizados (FXII^{hum/hum}). Se midió la expresión del factor XII de coagulación de ratón y del factor XII de coagulación humano. Se utilizó albúmina como control de carga.

La Figura 6 muestra los resultados del tiempo de tromboplastina parcial activada (aPTT) en ratones *F12* de tipo silvestre y humanizados.

La Figura 7A es un esquema que muestra la localización de sondas de qPCR en relación con los dominios de la

proteína FXII. La línea debajo de cada sonda representa la región del ARNm de *F12* amplificada.

La Figura 7B muestra los resultados del análisis de qPCR que demuestra la expresión de todas las regiones de ARNm de *F12* en el hígado, pero solo la expresión de los exones 7-14 en el cerebro total, el hipocampo y el lóbulo temporal. La expresión de *F12* se normaliza con respecto a GAPDH y se representa como porcentaje de los niveles hepáticos.

La Figura 7C muestra que los datos de ARNseq del Portal GTEx muestran una cobertura completa del ARNm de *F12* desde el exón 1 al exón 14 en el hígado humano, pero solo la cobertura del exón 7 al 14 en el hipocampo y la corteza humanos. Los exones *F12* están ordenados de derecha a izquierda ya que el gen está ubicado en la cadena negativa. Los datos se obtuvieron del Portal GTEx en diciembre de 2018.

La Figura 8A es un esquema del análisis del navegador del genoma de UCSC que muestra genes y transcripciones *F12* humanos y de ratón. El *F12* humano tiene una gran isla CpG que consta de 1364 bases (recuento de dinucleótidos CpG = 141; 68.9% de contenido de GC) que abarca los exones 8-12, lo cual falta en el *F12* de ratón, donde la isla CpG tiene un recuento de CpG mucho más bajo de 19 y consta de 231 bases. GenBank muestra 5 transcripciones de *F12* humano que comienzan en el exón 9, que están ausentes en el *F12* de ratón. De estas cinco transcripciones, dos están respaldadas por evidencia experimental, mientras que tres son sintéticas. Los números de acceso para las transcripciones *F12* que comienzan en el exón 9 son: BC012390, KJ901422, KR710723, KR710724, y BT007350. Los datos se obtuvieron del navegador UCSC Genome ([//genome.ucsc.edu](http://genome.ucsc.edu)). El ensamblaje de diciembre de 2013 (GRCh38/hg38) se usó para datos humanos y el ensamblaje de diciembre de 2011 (GRCm38/mm10) se usó para datos de ratones.

La Figura 8B muestra que el ARNm de *F12* encontrado en el cerebro codifica tres marcos de lectura abiertos (ORF), con los ORF 1 y 2 codificando proteínas similares a β FXIIa (N335-S596).

Las Figuras 9A-9D muestran que FXII₂₉₇₋₅₉₆ adquiere actividad catalítica tras la activación por calicreína y puede convertir pro-HGF en HGF.

En la Figura 9A, se incubó FXII₂₉₇₋₅₉₆ en presencia o ausencia de calicreína (15 nM) y se inhibió la actividad de calicreína con aprotinina antes de medir la actividad de FXII₂₉₇₋₅₉₆ con el sustrato cromogénico FXIIa S-2302. La señal de FXII₂₉₇₋₅₉₆ preincubado con calicreína no se debe a la actividad de calicreína residual, ya que no se observa actividad en muestras en donde se omitió FXII₂₉₇₋₅₉₆. Se muestra la media $\pm$ SEM.

En la Figura 9B, se incubó FXII (125 μ M) con cathepsina K (Cath K; 100 nM) o vehículo y se analizaron los productos de reacción mediante SDS-PAGE no reductora. Los productos de escisión de FXII a $\sim$ 40 kDa podrían deberse a la escisión entre L296 y M297, generando FXII₂₉₇₋₅₉₆. La actividad del FXII escindido por cathepsina K (FXII 375 nM, cathepsina K 30 nM) se analizó mediante un ensayo de sustrato cromogénico. Se observó una actividad insignificante para el FXII escindido por cathepsina K en ausencia de activación de calicreína, lo que indica que la escisión de cathepsina K por sí sola no genera FXII activo. El FXII escindido con cathepsina e incubado con calicreína (15 nM) tuvo una actividad mucho mayor que el FXII de longitud completa incubado con calicreína. El FXII, la calicreína y la cathepsina K por sí solos tuvieron una actividad insignificante. Se muestra la media $\pm$ SEM.

En la Figura 9C, HGFA y β FXIIa escindieron pro-HGF 25 μ g/ml, pero no tPA ni trombina (todos a 50 nM).

En la Figura 9D, FXII₂₉₇₋₅₉₆ (225 nM) y FXII (50 nM) se activaron con calicreína (15 nM) y la actividad de calicreína se inhibió con aprotinina antes de la incubación con Pro-HGF (25 μ g/ml). Pro-HGF fue escindido por FXII₂₉₇₋₅₉₆ activado por calicreína y FXII activado por calicreína (flechas), pero no por calicreína inactivada sola. Se incluyeron tPA y HGFA (225 nM) como controles positivos y negativos.

Las Figuras 10A-10B muestran datos de ARNseq de tejidos humanos y de ratón. La Figura 10A muestra la expresión de *F12* en tejidos humanos seleccionados que se obtuvo del portal GTEx en diciembre de 2018. La Figura 10B muestra la expresión de *F12* en tejidos de ratón seleccionados que se obtuvo del análisis de ARNseq de tejido de ratón C57BL/6. Las líneas verticales que cruzan el gráfico indican fragmentos por kilobase de transcripción por millón de lecturas mapeadas (FPKM) = 1.

La Figura 11 muestra que FXII₂₉₇₋₅₉₆ (225 nM) se activó con calicreína (15 nM) y la actividad de calicreína se inhibió con aprotinina antes de la incubación con precalicreína (200 nM). La precalicreína fue escindida por FXII₂₉₇₋₅₉₆ activado por calicreína, pero no solo con calicreína inactivada. La cadena pesada de calicreína no se observó en el carril de control de calicreína debido a los bajos niveles de calicreína utilizados para activar FXII₂₉₇₋₅₉₆.

Definiciones

Los términos "proteína", "polipéptido" y "péptido", usados indistintamente en el presente documento, incluyen formas poliméricas de aminoácidos de cualquier longitud, incluyendo aminoácidos codificados y no codificados y aminoácidos derivatizados o modificados química o bioquímicamente. Los términos también incluyen polímeros que se han modificado, tales como polipéptidos que tienen cadenas principales peptídicas modificadas. El término "dominio" se refiere a cualquier parte de una proteína o polipéptido que tenga una función o estructura particular.

Los términos "ácido nucleico" y "polinucleótido", utilizados indistintamente en el presente documento, incluyen formas poliméricas de nucleótidos de cualquier longitud, incluyendo ribonucleótidos, desoxirribonucleótidos o análogos o versiones modificadas de los mismos. Incluyen ADN o ARN monocatenario, bicatenario y multicatenario, ADN genómico, ADNc, híbridos de ADN-ARN y polímeros que comprenden bases purínicas, bases pirimidínicas u otros naturales, modificados químicamente, modificados bioquímicamente, no naturales o bases de nucleótidos derivatizadas.

El término "genómicamente integrado" se refiere a un ácido nucleico que se ha introducido en una célula de manera que la secuencia de nucleótidos se integra en el genoma de la célula. Puede usarse cualquier protocolo para la incorporación estable de un ácido nucleico en el genoma de una célula.

El término "vector orientado" se refiere a un ácido nucleico recombinante que puede introducirse mediante recombinación homóloga, ligadura mediada por unión de extremos no homólogos o cualquier otro medio de recombinación en una posición objetivo en el genoma de una célula.

El término "vector viral" se refiere a un ácido nucleico recombinante que incluye al menos un elemento de origen viral e incluye elementos suficientes para o que permiten empaquetar en una partícula de vector viral. El vector y/o partícula se puede utilizar con el fin de transferir ADN, ARN u otros ácidos nucleicos al interior de las células *in vitro*, *ex vivo*, o *in vivo*. Se conocen numerosas formas de vectores virales.

El término "aislado" con respecto a células, tejidos, proteínas y ácidos nucleicos incluye células, tejidos, proteínas y ácidos nucleicos que están relativamente purificados con respecto a otros componentes bacterianos, virales, celulares u otros que normalmente pueden estar presentes *in situ*, hasta e incluyendo una preparación sustancialmente pura de células, tejidos, proteínas y ácidos nucleicos. El término "aislado" también incluye células, tejidos, proteínas y ácidos nucleicos que no tienen contrapartes naturales, han sido sintetizados químicamente y, por lo tanto, no están sustancialmente contaminados por otras células, tejidos, proteínas y ácidos nucleicos, o han sido separados o purificados de la mayoría de los otros componentes (por ejemplo, componentes celulares) con los que están acompañados naturalmente (por ejemplo, otras proteínas celulares, polinucleótidos o componentes celulares).

El término "tipo silvestre" incluye entidades que tienen una estructura y/o actividad como la que se encuentra en un estado o contexto normal (en contraste con un estado o contexto mutante, enfermo, alterado, etc.). Los genes y polipéptidos de tipo silvestre a menudo existen en múltiples formas diferentes (por ejemplo, alelos).

El término "secuencia endógena" se refiere a una secuencia de ácido nucleico que se produce de forma natural dentro de una célula o roedor. Por ejemplo, una secuencia de *F12* endógeno de un roedor se refiere a una secuencia de *F12* nativo que se produce naturalmente en el locus *F12* del roedor.

Las moléculas o secuencias "exógenas" incluyen moléculas o secuencias que normalmente no están presentes en una célula en esa forma o ubicación (por ejemplo, locus genómico). La presencia normal incluye la presencia con respecto al paso de desarrollo particular y las condiciones ambientales de la célula. Una molécula o secuencia exógena, por ejemplo, puede incluir una versión mutada de una secuencia endógena correspondiente dentro de la célula, tal como una versión humanizada de la secuencia endógena, o puede incluir una secuencia correspondiente a una secuencia endógena dentro de la célula, pero en una forma diferente (es decir, no dentro de un cromosoma). Por el contrario, las moléculas o secuencias endógenas incluyen moléculas o secuencias que normalmente están presentes en esa forma y ubicación en una célula particular en una etapa de desarrollo particular bajo condiciones ambientales particulares.

El término "heterólogo" cuando se usa en el contexto de un ácido nucleico o una proteína indica que el ácido nucleico o la proteína comprende al menos dos segmentos que no se encuentran juntos de forma natural en la misma molécula. Por ejemplo, el término "heterólogo", cuando se usa con referencia a segmentos de un ácido nucleico o segmentos de una proteína, indica que el ácido nucleico o la proteína comprende dos o más subsecuencias que no se encuentran en la misma relación entre sí (por ejemplo, unidos) en la naturaleza. Como ejemplo, una región "heteróloga" de un vector de ácido nucleico es un segmento de ácido nucleico dentro o unido a otra molécula de ácido nucleico que no se encuentra asociado con la otra molécula en la naturaleza. Por ejemplo, una región heteróloga de un vector de ácido nucleico podría incluir una secuencia codificante flanqueada por secuencias que no se encuentran asociadas con la secuencia codificante en la naturaleza. Asimismo, una región "heteróloga" de una proteína es un segmento de aminoácidos dentro o unido a otra molécula peptídica que no se encuentra asociada con la otra molécula peptídica en la naturaleza (por ejemplo, una proteína de fusión o una proteína con una etiqueta). De manera similar, un ácido nucleico o una proteína pueden comprender un marcador heterólogo o una secuencia de secreción o localización heterólogo.

La "optimización de codones" aprovecha la degeneración de los codones, como se muestra por la multiplicidad de combinaciones de codones de tres pares de bases que especifican un aminoácido, y generalmente incluye un proceso de modificación de una secuencia de ácido nucleico para mejorar la expresión en células huésped particulares reemplazando al menos un codón de la secuencia nativa con un codón que se usa con mayor frecuencia o más frecuentemente en los genes de la célula huésped mientras se mantiene la secuencia de aminoácidos nativa. Por ejemplo, un ácido nucleico que codifica una proteína Cas9 puede modificarse para sustituir codones que tengan una

mayor frecuencia de uso en una célula procariótica o eucariota determinada, incluyendo una célula bacteriana, una célula de levadura, una célula humana, una célula no humana, una célula de mamífero, una célula de roedor, una célula de ratón, una célula de rata, una célula de hámster o cualquier otra célula huésped, en comparación con la secuencia de ácido nucleico de origen natural. Las tablas de uso de codones están disponibles, por ejemplo, en la "Codon Usage Database." Estas tablas se pueden adaptar de varias maneras. Véase Nakamura et al. (2000) Nucleic Acids Research 28:292. También se encuentran disponibles algoritmos informáticos para la optimización de codones de una secuencia particular para su expresión en un huésped particular (véase, por ejemplo, Gene Forge).

El término "locus" se refiere a una ubicación específica de un gen (o secuencia significativa), secuencia de ADN, secuencia que codifica un polipéptido o posición en un cromosoma del genoma de un organismo. Por ejemplo, un "locus del factor XII de coagulación" o "locus *F12*" puede referirse a la ubicación específica de un gen *F12*, secuencia de ADN *F12*, secuencia que codifica el factor de coagulación XII o posición *F12* en un cromosoma del genoma de un organismo que se ha identificado en cuanto a dónde reside dicha secuencia. Un "locus *F12*" puede comprender un elemento regulador de un gen *F12*, incluyendo, por ejemplo, un potenciador, un promotor, una región no traducida (UTR) 5' y/o 3', o una combinación de los mismos.

El término "gen" se refiere a secuencias de ADN en un cromosoma que pueden contener, si están presentes de forma natural, al menos una región codificante y al menos una no codificante. La secuencia de ADN en un cromosoma que codifica un producto (por ejemplo, pero sin limitarse a, un producto de ARN y/o un producto polipeptídico) puede incluir la región codificante interrumpida con intrones no codificantes y la secuencia ubicada adyacente a la región codificante en ambos los extremos 5' y 3' de manera que el gen corresponda al ARNm de longitud completa (incluidas las secuencias no traducidas 5' y 3'). Además, pueden estar presentes otras secuencias no codificantes que incluyen secuencias reguladoras (por ejemplo, entre otras, promotores, potenciadores y sitios de unión de factores de transcripción), señales de poliadenilación, sitios internos de entrada a ribosomas, silenciadores, secuencias aislantes y en un gen pueden estar presentes regiones de unión a la matriz. Estas secuencias pueden estar cerca de la región codificante del gen (por ejemplo, pero sin limitarse a, dentro de 10 kb) o en sitios distantes, e influyen en el nivel o velocidad de transcripción y traducción del gen.

El término "alelo" se refiere a una forma variante de un gen. Algunos genes tienen una variedad de formas diferentes, que se encuentran en la misma posición, o locus genético, en un cromosoma. Un organismo diploide tiene dos alelos en cada locus genético. Cada par de alelos representa el genotipo de un locus genético específico. Los genotipos se describen como homocigotos si hay dos alelos idénticos en un locus particular y como heterocigotos si los dos alelos difieren.

Un "promotor" es una región reguladora del ADN que normalmente comprende una caja TATA capaz de dirigir la ARN polimerasa II para iniciar la síntesis de ARN en el sitio de inicio de la transcripción apropiado para una secuencia de polinucleótidos particular. Un promotor puede comprender adicionalmente otras regiones que influyen en la tasa de inicio de la transcripción. Las secuencias promotoras divulgadas en el presente documento modulan la transcripción de un polinucleótido ligado operativamente. Un promotor puede ser activo en uno o más de los tipos de células divulgados en el presente documento (por ejemplo, una célula eucariota, una célula de mamífero no humano, una célula humana, una célula de roedor, una célula pluripotente, un embrión en etapa unicelular, una célula diferenciada o una combinación de los mismos). Un promotor puede ser, por ejemplo, un promotor constitutivamente activo, un promotor condicional, un promotor inducible, un promotor temporalmente restringido (por ejemplo, un promotor regulado por el desarrollo), o un promotor espacialmente restringido (por ejemplo, un promotor específico de célula o de tejido). Ejemplos de promotores se pueden encontrar, por ejemplo, en el documento WO 2013/176772.

Ejemplos de promotores inducibles incluyen, por ejemplo, promotores regulados químicamente y promotores regulados físicamente. Los promotores regulados químicamente incluyen, por ejemplo, promotores regulados por alcohol (por ejemplo, un promotor del gen de la alcohol deshidrogenasa (*alcA*)), promotores regulados por tetraciclina (por ejemplo, un promotor sensible a tetraciclina, una secuencia operadora de tetraciclina (*tetO*), un promotor *tet-On* promotor o un promotor *tet-Off*), promotores regulados por esteroides (por ejemplo, un receptor de glucocorticoides de rata, un promotor de un receptor de estrógeno o un promotor de un receptor de ecdisona) o promotores regulados por metales (por ejemplo, un promotor de metaloproteína). Los promotores regulados físicamente incluyen, por ejemplo, promotores regulados por temperatura (por ejemplo, un promotor de choque térmico) y promotores regulados por luz (por ejemplo, un promotor inducible por luz o un promotor reprimible por luz).

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

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

"Enlace operable" o estar "ligado operativamente" incluye yuxtaposición de dos o más componentes (por ejemplo, un promotor y otro elemento de secuencia) de manera que ambos componentes funcionen normalmente y permitan la

posibilidad de que al menos uno de los componentes pueda mediar una función que sea ejercida sobre al menos uno de los otros componentes. Por ejemplo, un promotor puede ligarse operativamente a una secuencia codificante si el promotor controla el nivel de transcripción de la secuencia codificante en respuesta a la presencia o ausencia de uno o más factores reguladores de la transcripción. El enlace operable puede incluir que tales secuencias sean contiguas entre sí o actúen *in trans* (por ejemplo, una secuencia reguladora puede actuar a distancia para controlar la transcripción de la secuencia codificante).

"Complementariedad" de los ácidos nucleicos significa que una secuencia de nucleótidos en una cadena de ácido nucleico, debido a la orientación de sus grupos de nucleobases, forma enlaces de hidrógeno con otra secuencia en una cadena de ácido nucleico opuesta. Las bases complementarias en el ADN suelen ser A con T y C con G. En el ARN, suelen ser C con G y U con A. La complementariedad puede ser perfecta o sustancial/suficiente. La complementariedad perfecta entre dos ácidos nucleicos significa que los dos ácidos nucleicos pueden formar un dúplex en donde cada base del dúplex está unida a una base complementaria mediante el emparejamiento Watson-Crick. Complementaria "sustancial" o "suficiente" significa que una secuencia en una cadena no es completa y/o perfectamente complementaria a una secuencia en una cadena opuesta, pero que se produce una unión suficiente entre las bases de las dos cadenas para formar un complejo híbrido estable en conjunto de las condiciones de hibridación (por ejemplo, concentración de sal y temperatura). Estas condiciones se pueden predecir utilizando secuencias y cálculos matemáticos estándar para predecir la T_m (temperatura de fusión) de las cadenas hibridadas, o mediante la determinación empírica de la T_m mediante el uso de métodos de rutina. T_m incluye la temperatura a la que una población de complejos de hibridación formados entre dos cadenas de ácido nucleico se desnaturaliza en un 50% (es decir, una población de moléculas de ácido nucleico bicatenario se disocia a la mitad en cadenas sencillas). A una temperatura por debajo de la T_m , se favorece la formación de un complejo de hibridación, mientras que, a una temperatura por encima de la T_m , se favorece la fusión o separación de las cadenas en el complejo de hibridación. La T_m puede estimarse para un ácido nucleico que tiene un contenido de G+C conocido en una solución acuosa de NaCl 1 M usando, por ejemplo, $T_m = 81.5 \pm 0.41$ (% de G+C), aunque otros cálculos de T_m conocidos tienen en cuenta las características estructurales del ácido nucleico.

La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque es posible que no coincidan las bases. Las condiciones apropiadas para la hibridación entre dos ácidos nucleicos dependen de la longitud de los ácidos nucleicos y del grado de complementación, variables que son bien conocidas. Cuanto mayor sea el grado de complementación entre dos secuencias de nucleótidos, mayor será el valor de la temperatura de fusión (T_m) para híbridos de ácidos nucleicos que tengan esas secuencias. Para las hibridaciones entre ácidos nucleicos con tramos cortos de complementariedad (por ejemplo, complementariedad de más de 35 o menos, 30 o menos, 25 o menos, 22 o menos, 20 o menos, o 18 o menos nucleótidos), la posición de los desajustes se vuelve importante (véase, Sambrook et al., supra, 11.7-11.8). Normalmente, la longitud de un ácido nucleico hibridable es de al menos aproximadamente 10 nucleótidos. Las longitudes mínimas ilustrativas para un ácido nucleico hibridable incluyen al menos aproximadamente 15 nucleótidos, al menos aproximadamente 20 nucleótidos, al menos aproximadamente 22 nucleótidos, al menos aproximadamente 25 nucleótidos y al menos aproximadamente 30 nucleótidos. Además, la temperatura y la concentración de sal de la solución de lavado se pueden ajustar según sea necesario de acuerdo con factores tales como la longitud de la región de complementación y el grado de complementación.

No es necesario que la secuencia del polinucleótido sea 100% complementaria a la de su ácido nucleico objetivo para que sea específicamente hibridable. Además, un polinucleótido puede hibridarse sobre uno o más segmentos de manera que los segmentos intermedios o adyacentes no estén involucrados en el evento de hibridación (por ejemplo, una estructura en bucle o una estructura en horquilla). Un polinucleótido (por ejemplo, ARNg) puede comprender al menos 70 %, al menos 80 %, al menos 90 %, al menos un 95 %, al menos 99 % o 100 % de complementariedad de secuencia con una región objetivo dentro de la secuencia de ácido nucleico objetivo para a los que se orienten. Por ejemplo, un ARNg en donde 18 de 20 nucleótidos son complementarios a una región objetivo y, por lo tanto, se hibridarían específicamente, representaría 90% de complementariedad. En este ejemplo, los nucleótidos no complementarios restantes pueden agruparse o intercalarse con nucleótidos complementarios y no necesitan ser contiguos entre sí ni con nucleótidos complementarios.

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

Los métodos y composiciones proporcionados en el presente documento emplean una variedad de componentes diferentes. Algunos componentes a lo largo de la descripción pueden tener variantes y fragmentos activos. Dichos componentes incluyen, por ejemplo, proteínas Cas, ARN CRISPR, ARNtracr y ARN guía. La actividad biológica para cada uno de estos componentes se describe en otra parte del presente documento. El término "funcional" se refiere a la capacidad innata de una proteína o ácido nucleico (o un fragmento o variante del mismo) para exhibir una actividad o función biológica. Tales actividades o funciones biológicas pueden incluir, por ejemplo, la capacidad de una proteína Cas para unirse a un ARN guía y a una secuencia de ADN objetivo. Las funciones biológicas de los fragmentos o variantes funcionales pueden ser las mismas o, de hecho, pueden cambiarse (por ejemplo, con respecto a su

especificidad, selectividad o eficacia) en comparación con la molécula original, pero conservando la función biológica básica de la molécula.

El término "variante" se refiere a una secuencia de nucleótidos que difiere de la secuencia más prevalente en una población (por ejemplo, en un nucleótido) o una secuencia de proteínas diferente de la secuencia más prevalente en una población (por ejemplo, en un aminoácido).

El término "fragmento", cuando se refiere a una proteína, significa una proteína que es más corta o tiene menos aminoácidos que la proteína de longitud completa. El término "fragmento", cuando se refiere a un ácido nucleico, significa un ácido nucleico que es más corto o tiene menos nucleótidos que el ácido nucleico de longitud completa. Un fragmento puede ser, por ejemplo, cuando se hace referencia a un fragmento de proteína, un fragmento N-terminal (es decir, eliminación de una porción del extremo C-terminal de la proteína), un fragmento C-terminal (es decir, eliminación de una porción del extremo N-terminal de la proteína), o un fragmento interno (es decir, eliminación de una porción de cada uno de los extremos N-terminal y C-terminal de la proteína). Un fragmento puede ser, por ejemplo, cuando se hace referencia a un fragmento de ácido nucleico, un fragmento 5' (es decir, eliminación de una porción del extremo 3' del ácido nucleico), un fragmento 3' (es decir, eliminación de una porción del extremo 5' del ácido nucleico), o un fragmento interno (es decir, eliminación de una porción de cada uno de los extremos 5' y 3' del ácido nucleico).

"Identidad de secuencia" o "identidad" en el contexto de dos polinucleótidos o secuencias de polipéptidos se refiere a los residuos en las dos secuencias que son iguales cuando se alinean para una máxima correspondencia en una ventana de comparación especificada. Cuando se utiliza el porcentaje de identidad de secuencia en referencia a proteínas, las posiciones de los residuos que no son idénticas a menudo difieren mediante sustituciones conservadoras de aminoácidos, donde los residuos de aminoácidos se sustituyen por otros residuos de aminoácidos con propiedades químicas similares (por ejemplo, carga o hidrofobicidad) y por lo tanto no cambian las propiedades funcionales de la molécula. Cuando las secuencias difieren en sustituciones 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 mediante tales sustituciones conservadoras tienen "similitud de secuencia" o "similitud". Los medios para realizar este ajuste son bien conocidos. Normalmente, esto implica calificar una sustitución conservadora como un desajuste parcial en lugar de completa, aumentando así el porcentaje de identidad de secuencia. Así, por ejemplo, cuando a un aminoácido idéntico se le asigna una puntuación de 1 y a una sustitución no conservadora se le asigna una puntuación de cero, a una sustitución conservadora se le asigna una puntuación entre cero y 1. La puntuación de las sustituciones conservadoras se calcula, por ejemplo, tal como se implementa en el programa PC/GENE (Intelligenetics, Mountain View, California).

"Porcentaje de identidad de secuencia" incluye el valor determinado comparando dos secuencias óptimamente alineadas (el mayor número de residuos perfectamente coincidentes) durante una ventana de comparación, en donde la porción de la secuencia de polinucleótidos en la ventana de comparación puede comprender adiciones o deleciones (es decir, espacios) en comparación con la secuencia de referencia (que no comprende adiciones ni deleciones) para una alineación óptima de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en donde aparece la base de ácido nucleico o el residuo de aminoácido idéntico en ambas secuencias para producir el número de posiciones coincidentes, dividiendo el número de posiciones coincidentes por el número total de posiciones en la ventana de comparación, y multiplicar el resultado por 100 para obtener el porcentaje de identidad de secuencia. A menos que se especifique lo contrario (por ejemplo, la secuencia más corta incluye una secuencia heteróloga ligada), la ventana de comparación es la longitud total de la más corta de las dos secuencias que se comparan.

A menos que se indique lo contrario, los valores de identidad/similitud de secuencia incluyen el valor obtenido usando GAP Versión 10 usando los siguientes parámetros: % de identidad y % de similitud para una secuencia de nucleótidos utilizando un peso GAP de 50 y un peso de longitud de 3, y la matriz de puntuación nwsgapdna.cmp; % de identidad y % de similitud para una secuencia de aminoácidos usando un peso GAP de 8 y un peso de longitud de 2, y la matriz de puntuación BLOSUM62; o cualquier programa equivalente del mismo. "Programa equivalente" incluye cualquier programa de comparación de secuencias que, para dos secuencias cualesquiera en cuestión, genera una alineación que tiene coincidencias idénticas de residuos de nucleótidos o aminoácidos y un porcentaje de identidad de secuencia idéntico en comparación con la alineación correspondiente generada por GAP Versión 10.

El término "sustitución conservativa de aminoácidos" se refiere a la sustitución de un aminoácido que normalmente está presente en la secuencia por un aminoácido diferente de tamaño, carga o polaridad similar. Ejemplos de sustituciones conservadoras incluyen la sustitución de un residuo no polar (hidrófobo) tal como isoleucina, valina o leucina por otro residuo no polar. Asimismo, ejemplos de sustituciones conservadoras incluyen la sustitución de un residuo polar (hidrófilo) por otro, como entre arginina y lisina, entre glutamina y asparagina, o entre glicina y serina. Además, la sustitución de un residuo básico tal como lisina, arginina o histidina por otro, o la sustitución de un residuo ácido como ácido aspártico o ácido glutámico por otro residuo ácido son ejemplos adicionales de sustituciones conservadoras. Ejemplos de sustituciones no conservadoras incluyen la sustitución de un residuo de aminoácido no polar (hidrófobo) tal como isoleucina, valina, leucina, alanina o metionina por un residuo polar (hidrófilo) tal como cisteína, glutamina, ácido glutámico o lisina y/o un residuo polar para un residuo no polar. Las categorizaciones típicas de aminoácidos se resumen en Tabla 1 abajo.

Tabla 1. Categorizaciones de aminoácidos.

Alanina	Ala	A	No polar	Neutral	1.8
Arginina	Arg	R	Polar	Positivo	-4.5
Asparagina	Asn	N	Polar	Neutral	-3.5
Ácido aspártico	Asp	D	Polar	Negativo	-3.5
Cisteína	Cys	C	No polar	Neutral	2.5
Ácido glutámico	Glu	E	Polar	Negativo	-3.5
Glutamina	Gln	Q	Polar	Neutral	-3.5
Glicina	Gly	G	No polar	Neutral	-0.4
Histidina	His	H	Polar	Positivo	-3.2
Isoleucina	Ile	I	No polar	Neutral	4.5
Leucina	Leu	L	No polar	Neutral	3.8
Lisina	Liz	K	Polar	Positivo	-3.9
Metionina	Met	M	No polar	Neutral	1.9
Fenilalanina	Phe	F	No polar	Neutral	2.8
Prolina	Pro	P	No polar	Neutral	-1.6
Serina	Ser	S	Polar	Neutral	-0.8
Treonina	Thr	T	Polar	Neutral	-0.7
Triptófano	Trp	W	No polar	Neutral	-0.9
Tirosina	Tyr	Y	Polar	Neutral	-1.3
Valina	Val	V	No polar	Neutral	4.2

Una secuencia "homóloga" (por ejemplo, secuencia de ácido nucleico) incluye una secuencia que es idéntica o sustancialmente similar a una secuencia de referencia conocida, de modo que sea, por ejemplo, al menos 50 %, al menos 55 %, al menos 60 %, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99% o 100% idéntica a la secuencia de referencia conocida. Las secuencias homólogas pueden incluir, por ejemplo, secuencias ortólogas y secuencias parálogas. Los genes homólogos, por ejemplo, normalmente descienden de una secuencia de ADN ancestral común, ya sea a través de un evento de especiación (genes ortólogos) o un evento de duplicación genética (genes parálogos). Los genes "ortólogos" incluyen genes de diferentes especies que evolucionaron a partir de un gen ancestral común mediante especiación. Los ortólogos suelen conservar la misma función en el curso de la evolución. Los genes "parálogos" incluyen genes relacionados por duplicación dentro de un genoma. Los parálogos pueden desarrollar nuevas funciones en el curso de la evolución.

El término "*in vitro*" incluye ambientes artificiales y procesos o reacciones que ocurren dentro de un ambiente artificial (por ejemplo, un tubo de ensayo o una célula o línea celular aislada). El término "*in vivo*" incluye ambientes naturales (por ejemplo, una célula, organismo o cuerpo) y procesos o reacciones que ocurren dentro de un ambiente natural. El término "*ex-vivo*" incluye células que se han extraído del cuerpo de un individuo y procesos o reacciones que ocurren dentro de dichas células.

El término "gen indicador" se refiere a un ácido nucleico que tiene una secuencia que codifica un producto génico (típicamente una enzima) que se analiza fácil y cuantificablemente cuando se introduce un constructo que comprende la secuencia del gen indicador ligada operativamente a un promotor heterólogo y/o elemento potenciador en células que contienen (o que pueden hacerse que contengan) los factores necesarios para la activación de los elementos promotores y/o potenciadores. Los ejemplos de genes indicadores incluyen, entre otros, genes que codifican la beta-galactosidasa (lacZ), los genes bacterianos de la cloranfenicol acetiltransferasa (cat), los genes de la luciferasa de luciérnaga, los genes que codifican la beta-glucuronidasa (GUS) y los genes que codifican proteínas fluorescentes. Una "proteína informadora" se refiere a una proteína codificada por un gen indicador.

El término "proteína indicadora fluorescente", tal como se utiliza en el presente documento, significa una proteína indicadora que es detectable basándose en la fluorescencia, en donde la fluorescencia puede provenir directamente de la proteína indicadora, de la actividad de la proteína indicadora sobre un sustrato fluorogénico o de una proteína con afinidad para unirse a un compuesto marcado con fluorescencia. Los ejemplos de proteínas fluorescentes incluyen proteínas fluorescentes verdes (por ejemplo, GFP, GFP-2, tagGFP, turboGFP, eGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP y ZsGreen1), proteínas fluorescentes amarillas (por ejemplo, YFP, eYFP, Citrine, Venus, YPet, PhiYFP, y ZsYellow1), proteínas fluorescentes azules (por ejemplo, BFP, eBFP, eBFP2, Azurite, mKalamal, GFPuv, Sapphire, y T-sapphire), proteínas fluorescentes cian (por ejemplo, CFP, eCFP, Cerulean, CyPet, AmCyan1 y Midoriishi-Cyan), proteínas fluorescentes rojas (por ejemplo, RFP, mKate, mKate2, mPlum, monómero DsRed, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry y Jred), proteínas fluorescentes de color naranja (por ejemplo, mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine y tdTomato) y cualquier otra proteína fluorescente adecuada cuya presencia en las células pueda detectarse mediante métodos de citometría de flujo.

La reparación en respuesta a rupturas bicatenarias (DSB) se produce principalmente a través de dos vías de reparación del ADN conservadas: recombinación homóloga (HR) y unión de extremos no homólogos (NHEJ). Véase Kasperek y Humphrey (2011) Semin. Cell Dev. Biol. 22(8):886-897. Asimismo, la reparación de un ácido nucleico objetivo mediada por un ácido nucleico donante exógeno puede incluir cualquier proceso de intercambio de información genética entre los dos polinucleótidos.

El término "recombinación" incluye cualquier proceso de intercambio de información genética entre dos polinucleótidos y puede ocurrir por cualquier mecanismo. La recombinación puede ocurrir mediante reparación orientada por homología (HDR) o recombinación homóloga (HR). HDR o HR incluye una forma de reparación de ácidos nucleicos que puede requerir homología de secuencia de nucleótidos, utiliza una molécula "donante" como plantilla para la reparación de una molécula "objetivo" (es decir, la que experimentó la ruptura bicatenaria) y conduce a la transferencia de información genética del donante al objetivo. Sin desear quedar ligado a ninguna teoría particular, dicha transferencia puede implicar la corrección de desajustes del ADN heterodúplex que se forma entre el objetivo roto y el donante, y/o la fusión de cadenas dependiente de la síntesis, en donde el donante se utiliza para resintetizar información genética que formar parte del objetivo y/o procesos relacionados. En algunos casos, el polinucleótido donante, una porción del polinucleótido donante, una copia del polinucleótido donante o una porción de una copia del polinucleótido donante se integra en el ADN objetivo. Véase Wang et al. (2013) Cell 153:910-918; Mandalos et al. (2012) PLoS ONE 7:e45768:1-9; y Wang et al. (2013) Nat. Biotechnol. 31:530-532.

La unión de extremos no homólogos (NHEJ) incluye la reparación de rupturas bicatenarias en un ácido nucleico mediante la ligadura directa de los extremos rotos entre sí o con una secuencia exógena sin la necesidad de una plantilla homóloga. La ligadura de secuencias no contiguas mediante NHEJ a menudo puede dar lugar a deleciones, inserciones o translocaciones cerca del sitio de la ruptura bicatenaria. Por ejemplo, NHEJ también puede dar como resultado la integración orientada de un ácido nucleico donante exógeno mediante la ligadura directa de los extremos rotos con los extremos del ácido nucleico donante exógeno (es decir, captura basada en NHEJ). Dicha integración orientada mediada por NHEJ puede preferirse para la inserción de un ácido nucleico donante exógeno cuando las vías de reparación orientada por homología (HDR) no son fácilmente utilizables (por ejemplo, en células que no se dividen, células primarias y células que realizan mal la reparación del ADN basada en homología). Además, a diferencia de la reparación orientada por homología, no se necesita conocimiento sobre grandes regiones de identidad de secuencia que flanquean el sitio de escisión, lo que puede ser beneficioso cuando se intenta la inserción orientada en organismos que tienen genomas para los cuales existe un conocimiento limitado de la secuencia genómica. La integración puede proceder mediante la ligadura de extremos romos entre el ácido nucleico donante exógeno y la secuencia genómica escindida, o mediante la ligadura de extremos adhesivos (es decir, que tienen salientes 5' o 3') utilizando un ácido nucleico donante exógeno que está flanqueado por salientes que son compatibles con los generados por un agente nucleasa en la secuencia genómica escindida. Véase, por ejemplo, los documentos US 2011/020722, WO 2014/033644, WO 2014/089290, y Maresca et al. (2013) Genoma Res. 23(3):539-546. Si se ligan extremos romos, puede ser necesaria la resección del objetivo y/o del donante para generar regiones de microhomología necesarias para la unión de fragmentos, lo que puede crear alteraciones no deseadas en la secuencia objetivo.

El término "proteína de unión a antígeno" incluye cualquier proteína que se une a un antígeno. Ejemplos de proteínas de unión a antígeno incluyen un anticuerpo, un fragmento de un anticuerpo de unión a antígeno, un anticuerpo multiespecífico (por ejemplo, un anticuerpo biespecífico), un scFV, un bis-scFV, un diacuerpo, un triacuerpo, un tetracuerpo, un V-NAR, un VHH, un VL, un F(ab), un F(ab)₂, una DVD (proteína de unión a antígeno de dominio

variable dual), una SVD (proteína de unión a antígeno de dominio variable único), un activador biespecífico de células T (BiTE) o un cuerpo de Davis (Patente de EE.UU. No. 8,586,713).

Las composiciones o métodos "que comprenden" o "incluyen" uno o más elementos enumerados pueden incluir otros elementos no mencionados específicamente. Por ejemplo, una composición que "comprende" o "incluye" una proteína puede contener la proteína sola o en combinación con otros ingredientes. La frase transitoria "que consiste esencialmente en" significa que debe interpretarse que el alcance de una reivindicación abarca los elementos especificados enumerados en la reivindicación y aquellos que no afectan materialmente las características básicas y novedosas de la invención reivindicada. Por lo tanto, no se pretende que el término "que consiste esencialmente en" cuando se utiliza en una reivindicación de esta invención sea interpretado como equivalente a "que comprende".

"Opcional" u "opcionalmente" significa que el evento o circunstancia descrito posteriormente puede ocurrir o no y que la descripción incluye casos en donde el evento o circunstancia ocurre y casos en donde no.

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

A menos que se desprenda lo contrario del contexto, el término "aproximadamente" abarca valores ± 5 de un valor indicado.

El término "y/o" se refiere y abarca cualquiera y todas las combinaciones posibles de uno o más de los elementos enumerados asociados, así como la falta de combinaciones cuando se interpreta en la alternativa ("o").

El término "o" se refiere a cualquier miembro de una lista particular y también incluye cualquier combinación de miembros de esa lista.

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

Medias estadísticamente significativas $p \leq 0.05$.

Descripción detallada

I. Descripción general

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

En el presente documento se describen genomas de roedores, células de roedores y roedores que comprenden un locus (*F12*) del factor XII de coagulación humanizado y métodos para usar dichas células animales no humanas y animales no humanos como se caracterizan en las reivindicaciones. También se describen en el presente documento genes *F12* de animales no humanos humanizados que comprenden una modificación genética orientada que humaniza los genes *F12* de animales no humanos y agentes de nucleasa y vectores orientados para su uso en la humanización de un gen *F12* de animales no humanos. Las células de roedor o los roedores que comprenden un locus *F12* humanizado expresan una proteína del factor XII de coagulación humano. Dichas células de roedores y roedores pueden usarse para evaluar la administración o eficacia de agentes orientados al factor XII de coagulación humano (por ejemplo, agentes de edición del genoma CRISPR/Cas9 o proteínas de unión a antígenos) *in vitro*, *ex vivo*, o *in vivo* y puede usarse en métodos para optimizar la administración de eficacia de tales agentes *in vitro*, *ex vivo*, o *in vivo*.

En las células de roedor y los roedores descritos en el presente documento, la mayor parte o la totalidad del ADN genómico *F12* humano se inserta en el locus *F12* de roedor ortólogo correspondiente. En algunas de las células de roedor y de los roedores divulgados en el presente documento, la mayor parte o la totalidad del ADN genómico *F12* de roedor se reemplaza uno por uno con el correspondiente ADN genómico *F12* humano ortólogo. En comparación con los animales no humanos con inserciones de ADNc, los niveles de expresión deberían ser mayores cuando se mantienen la estructura intrón-exón y la maquinaria de empalme porque es más probable que los elementos reguladores conservados queden intactos y las transcripciones empalmadas que se someten al procesamiento del ARN son más estables que los ADNc. Por el contrario, la inserción de ADNc *F12* humano en un locus *F12* de un animal no humano aboliría los elementos reguladores conservados. Es más probable que reemplazar la secuencia genómica animal no humana con la correspondiente secuencia genómica humana ortóloga o insertar la secuencia genómica *F12* humana en el locus *F12* ortólogo no humano correspondiente dé como resultado una expresión fiel del transgén del locus *F12* endógeno. Por ejemplo, la secuencia genómica *F12* humana puede tener diferentes elementos reguladores internos (en comparación con la secuencia genómica *F12* de ratón) que impulsan la expresión de una isoforma más corta en las neuronas, un fenotipo que se recapitula en ratones humanizados *F12*. Los animales transgénicos no humanos con inserción transgénica de secuencias codificantes del factor XII de coagulación humano en un locus genómico aleatorio en lugar del locus *F12* endógeno de animales no humanos tampoco reflejarán con tanta precisión la regulación endógena de la expresión *F12*. Un alelo *F12* humanizado resultante de reemplazar la mayor parte o la totalidad del ADN genómico de animales no humanos uno por uno con el correspondiente ADN

genómico humano ortólogo o la inserción de una secuencia genómica de *F12* humano en el correspondiente locus *F12* ortólogo no humano proporcionará el verdadero objetivo humano o una aproximación cercana del verdadero objetivo humano de los reactivos orientados al factor XII de coagulación humano (por ejemplo, reactivos CRISPR/Cas9 diseñados para orientar al *F12* humano, o anticuerpos o moléculas pequeñas diseñadas para orientar al factor XII de coagulación humano), permitiendo así realizar pruebas de la eficacia y el modo de acción de dichos agentes en animales vivos, así como estudios farmacocinéticos y farmacodinámicos en un entorno donde la proteína humanizada y el gen humanizado son las únicas versiones del factor XII de coagulación y *F12* presentes.

II. Roedores que comprenden un locus (*F12*) del factor XII de coagulación humanizado

Los genomas de roedores, células de roedores y roedores divulgados en el presente documento comprenden un locus (*F12*) del factor XII de coagulación humanizado. Las células o roedores que comprenden un locus *F12* humanizado expresan una proteína del factor XII de coagulación humano.

A. Factor XII (*F12*) de Coagulación

Las células y roedores descritos en el presente documento comprenden un locus (*F12*) del factor XII de coagulación humanizado. El factor XII de coagulación (también conocido como factor XII, FXII, factor de Hageman, factor beta XIIa parte 1, factor beta XIIa parte 2, cadena pesada del factor XIIa de coagulación o cadena ligera del factor XIIa de coagulación) está codificado por gen *F12* (también conocido como *FXII*, *HAE3*, *HAEX*, o *HAF*). El factor XII de coagulación es sintetizado por los hepatocitos en el hígado y secretado a la circulación, donde inicia el sistema de activación por contacto. Aunque normalmente se piensa que está restringida a la circulación, la proteína del factor XII se ha encontrado en el cerebro de pacientes con enfermedad de Alzheimer y esclerosis múltiple, y se ha observado la activación del sistema de contacto en el cerebro y el líquido cefalorraquídeo humanos. El factor XII de coagulación es una serina proteasa circulante que inicia el sistema de activación por contacto. La activación del factor XII en superficies cargadas negativamente o por la calicreína plasmática produce el factor XII activado (factor XIIa). El factor XIIa inicia la cascada de coagulación intrínseca, que conduce a la generación de trombina y la formación de coágulos. El factor XIIa también convierte la precalicreína plasmática en calicreína, lo que activa la vía calicreína-quinina, lo que da como resultado la liberación del péptido vasoactivo y proinflamatorio bradicinina. Además de estas funciones bien estudiadas, también se ha demostrado que el factor XIIa activa el factor de crecimiento de hepatocitos (HGF) y el sistema del complemento *in vitro*. El factor XII circula como un zimógeno de cadena sencilla que, tras la escisión en R353 durante la activación, se convierte en una molécula de dos cadenas enzimáticamente activa (α FXIIa). α FXIIa se procesa adicionalmente a β FXIIa, que consiste en la cadena ligera que contiene el dominio catalítico y solo una pequeña porción de la cadena pesada.

F12 humano se asigna a 5q35.3 en el cromosoma 5 (NCBI RefSeq Gene ID 2161; ensamblaje GRCh38.p12 (GCF_000001405.38); ubicación NC_000005.10 (177402138..177409576, complemento)). Se ha informado que el gen tiene 14 exones. A la proteína del factor XII de coagulación humano de tipo silvestre se le ha asignado el número de acceso UniProt P00748. La secuencia de una isoforma, P00748-1 (idéntica al número de acceso del NCBI NP_000496.2), se establece en la SEQ ID NO: 46. A un ARNm (ADNc) que codifica esta isoforma se le asigna el número de acceso del NCBI NM_000505.3 y se establece en la SEQ ID NO: 34. A una secuencia codificante ejemplar (CDS) para esta isoforma se le asigna CCDS ID CCDS34302.1 y se establece en la SEQ ID NO: 48. Otra CDS ejemplar en donde la posición 619 es C en el exón 6 (rs17876030;c.619G>C; p.Ala207Pro) se establece en la SEQ ID NO: 13 y codifica la proteína expuesta en la SEQ ID NO: 5. La proteína del factor XII de coagulación humano de longitud completa expuesta en la SEQ ID NO: 5 tiene 615 aminoácidos, incluyendo un péptido señal (aminoácidos 1-19), una cadena pesada (aminoácidos 20-372) y una cadena ligera (aminoácidos 373-615). De manera similar, la proteína del factor XII de coagulación humano de longitud completa expuesta en la SEQ ID NO: 46 tiene 615 aminoácidos, incluyendo un péptido señal (aminoácidos 1-19), una cadena pesada (aminoácidos 20-372) y una cadena ligera (aminoácidos 373-615). Las delimitaciones entre estos dominios son las designadas en UniProt. Referencia al factor XII de coagulación humano o *F12* incluye las formas canónicas (tipo silvestre), así como todas las formas alélicas e isoformas. Cualquier otra forma de factor XII de coagulación humano tiene aminoácidos numerados para una alineación máxima con la forma natural (SEQ ID NO: 5 o 46), designándose los aminoácidos alineados con el mismo número.

F12 del ratón asigna a 13; 13 B1 en el cromosoma 13 (NCBI RefSeq Gene ID 58992; ensamblaje GRCm38.p4 (GCF_000001635.24); ubicación NC_000079.6 (55417958..55426804, complemento)). Se ha informado que el gen tiene 14 exones. A la proteína del factor XII de coagulación de ratón de tipo silvestre se le ha asignado el número de acceso UniProt Q80YC5. La secuencia del factor XII de coagulación de ratón (idéntica al número de acceso del NCBI NP_067464.2) se establece en la SEQ ID NO: 1. Un ARNm (ADNc) ejemplar que codifica la isoforma canónica (SEQ ID NO: 1) se le asigna el número de acceso NCBI NM_021489.3 y se establece en la SEQ ID NO: 33. Una secuencia codificante ejemplar (CDS) (CCDS ID CCDS36675.1) se establece en la SEQ ID NO: 9. La proteína canónica del factor XII de coagulación de ratón de longitud completa expuesta en la SEQ ID NO: 1 tiene 597 aminoácidos, incluyendo un péptido señal (aminoácidos 1-19), una cadena pesada (aminoácidos 20-354) y una cadena ligera (aminoácidos 355-597). Las delimitaciones entre estos dominios son las designadas en UniProt. Referencia al factor XII de coagulación de ratón o *F12* incluye las formas canónicas (tipo silvestre), así como todas las formas alélicas e isoformas. Cualquier otra forma de factor XII de coagulación de ratón tiene aminoácidos numerados para un alineamiento máximo con la forma de tipo silvestre, designándose los aminoácidos alineados con el mismo número.

También se conocen secuencias del factor XII de coagulación para muchos otros animales no humanos. Estos incluyen, por ejemplo, rata (número de acceso UniProt D3ZTE0; NCBI RefSeq Gene ID 306761), bovino (número de acceso UniProt P98140; NCBI RefSeq Gene ID 280789), cobaya (número de acceso UniProt Q04962) y cerdo (número de acceso UniProt O97507). ; ID del gen NCBI RefSeq 397474).

5 B. Loci *F12* humanizado

Un locus *F12* humanizado es un locus *F12* en donde un segmento del locus *F12* endógeno ha sido eliminado y reemplazado con una secuencia de *F12* humano ortóloga. Un locus *F12* humanizado puede ser un locus *F12* en donde el gen *F12* completo se reemplaza con la secuencia de *F12* humano ortóloga correspondiente, o puede ser un locus *F12* en donde solo una porción del gen *F12* se reemplaza con la secuencia de *F12* humano ortóloga correspondiente (es decir, humanizado). Alternativamente, un locus *F12* humanizado puede ser un locus *F12* en donde se inserta una porción de un locus *F12* humano ortólogo, o puede ser un locus *F12* en donde se elimina una porción del gen *F12* y se inserta una porción del locus *F12* humano ortólogo. La porción del locus *F12* humano ortólogo que se inserta puede, por ejemplo, comprender más locus *F12* humano que el que se elimina del locus *F12* endógeno. Por ejemplo, toda la secuencia codificante de *F12* en el locus *F12* endógeno puede eliminarse y reemplazarse con la correspondiente secuencia de *F12* humana. Una secuencia de *F12* humano correspondiente a un segmento particular de la secuencia de *F12* endógena se refiere a la región de *F12* humano que se alinea con el segmento particular de la secuencia de *F12* endógena cuando el *F12* humano y el *F12* endógeno están óptimamente alineados (mayor número de residuos perfectamente coincidentes). La secuencia humana ortóloga correspondiente puede comprender, por ejemplo, ADN complementario (ADNc) o ADN genómico. Opcionalmente, la correspondiente secuencia de *F12* humano ortóloga se modifica para optimizar sus codones en función del uso de codones en el roedor. Las regiones reemplazadas (es decir, humanizadas) pueden incluir regiones codificantes tal como un exón, regiones no codificantes tal como un intrón, una región no traducida o una región reguladora (por ejemplo, un promotor, un potenciador o un elemento de unión a represor transcripcional), o cualquier combinación de los mismos. Como ejemplo, los exones correspondientes a 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 o los 14 exones del gen *F12* humano puede humanizarse. Por ejemplo, los exones correspondientes a todos los exones (es decir, los exones 1-14) del gen *F12* humano pueden humanizarse. Asimismo, los intrones correspondientes a 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o los 13 intrones del gen *F12* humano se pueden humanizar o pueden permanecer endógenos. Por ejemplo, los intrones correspondientes a todos los intrones (es decir, intrones 1-13) del gen *F12* humano pueden humanizarse. Las regiones flanqueantes no traducidas que incluyen secuencias reguladoras también pueden humanizarse o permanecer endógenas. Por ejemplo, la región no traducida (UTR) 5', la UTR 3', o tanto la UTR 5' como la UTR 3' pueden humanizarse, o la UTR 5', la UTR 3', o ambas la UTR 5' y la UTR 3' pueden permanecer endógenas. En un ejemplo específico, tanto la UTR 5' como la UTR 3' siguen siendo endógenas. Dependiendo del grado de reemplazo por secuencias ortólogas, las secuencias reguladoras, tal como un promotor, pueden ser endógenas o suministradas por la secuencia ortóloga humana de reemplazo. Por ejemplo, el locus *F12* humanizado puede incluir el promotor *F12* endógeno de roedor (es decir, la secuencia de *F12* humano puede ligarse operativamente al promotor endógeno de roedor).

Una o más o todas las regiones que codifican el péptido señal, la cadena pesada o la cadena ligera pueden humanizarse, o una o más de dichas regiones pueden permanecer endógenas. Secuencias codificantes ejemplares para un péptido señal, una cadena pesada y una cadena ligera del factor XII de coagulación de ratón se exponen en las SEQ ID NOS: 10-12, respectivamente. Secuencias codificantes ejemplares para un péptido señal, una cadena pesada y una cadena ligera del factor XII de coagulación humano se exponen en las SEQ ID NOS: 14-16, respectivamente. Otra secuencia codificante ejemplar para una cadena pesada del factor XII de coagulación humano se establece en la SEQ ID NO: 49.

Por ejemplo, toda o parte de la región del locus *F12* que codifica el péptido señal puede humanizarse, y/o toda o parte de la región del locus *F12* que codifica la cadena pesada puede humanizarse, y/o toda o parte de la región del locus *F12* que codifica la cadena ligera puede humanizarse. Alternativa o adicionalmente, toda o parte de la región del locus *F12* que codifica el péptido señal puede permanecer endógena, y/o toda o parte de la región del locus *F12* que codifica la cadena pesada puede permanecer endógena, y/o toda o parte de la región del locus *F12* que codifica la cadena ligera puede permanecer endógena. De acuerdo con la invención, todas las regiones del locus *F12* que codifican el péptido señal, la cadena pesada y la cadena ligera están humanizadas. Opcionalmente, la CDS de la región humanizada del locus *F12* comprende, consiste esencialmente en, o consiste en una secuencia que es al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 % o aproximadamente 100 % idéntica a la SEQ ID NO: 13 (o sus degenerados). Opcionalmente, la CDS de la región humanizada del locus *F12* comprende, consiste esencialmente en, o consiste en una secuencia que es al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 %, o aproximadamente 100% idéntica a la SEQ ID NO: 31 y/o 32. Opcionalmente, el locus *F12* humanizado codifica una proteína que comprende, consiste esencialmente en, o consiste en una secuencia que es al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos

aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 %, o aproximadamente 100 % idéntica a la SEQ ID NO: 5 o 46. Opcionalmente, el locus *F12* humanizado comprende, consiste esencialmente en, o consiste en una secuencia que es al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 %, o aproximadamente 100 % idéntica a la SEQ ID NO: 17 o 18. En cada caso, la proteína del factor XII de coagulación humanizado puede conservar la actividad de la proteína del factor XII de coagulación nativa y/o la proteína del factor XII de coagulación humano.

La proteína del factor XII de coagulación codificada por el *F12* locus humanizado puede comprender uno o más dominios que son de una proteína del factor XII de coagulación humano. Secuencias de aminoácidos ejemplares para un péptido señal, una cadena pesada y una cadena ligera del factor XII de coagulación de ratón se exponen en las SEQ ID NOS: 2-4, respectivamente. Secuencias de aminoácidos ejemplares para un péptido señal, una cadena pesada y una cadena ligera del factor XII de coagulación humano se exponen en las SEQ ID NOS: 6-8, respectivamente. Otra secuencia de aminoácidos ejemplar para una cadena pesada del factor XII de coagulación humano se establece en la SEQ ID NO: 47.

La proteína del factor XII de coagulación puede comprender uno o más o todos de un péptido señal del factor XII de coagulación humano, una cadena pesada del factor XII de coagulación humano y una cadena ligera del factor XII de coagulación humano como se caracteriza en las reivindicaciones.

Como ejemplo, la proteína del factor XII de coagulación codificada por el locus *F12* humanizado puede comprender un péptido señal del factor XII de coagulación humano. Opcionalmente, el péptido señal del factor XII de coagulación humano comprende, consiste esencialmente en, o consiste en una secuencia que es al menos 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % idéntica a la SEQ ID NO: 6. Opcionalmente, la secuencia codificante para el péptido señal del factor XII de coagulación humano comprende, consiste esencialmente en, o consiste en una secuencia que es al menos 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, o 100% idéntica a la SEQ ID NO: 14. Como otro ejemplo, la proteína del factor XII de coagulación codificada por el locus *F12* humanizado puede comprender una cadena pesada del factor XII de coagulación humano. Opcionalmente, la cadena pesada del factor XII de coagulación humano comprende, consiste esencialmente en, o consiste en una secuencia que es al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 % o aproximadamente 100 % idéntica a la SEQ ID NO: 7 o 47. Opcionalmente, la secuencia codificante para la cadena pesada del factor XII de coagulación humano comprende, consiste esencialmente en, o consiste en una secuencia que es al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 % o aproximadamente 100 % idéntica a la SEQ ID NO: 15 o 49. Como otro ejemplo, la proteína del factor XII de coagulación codificada por el locus *F12* humanizado puede comprender una cadena ligera humana. Opcionalmente, la cadena ligera humana comprende, consiste esencialmente en, o consiste en una secuencia que es al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 %, o aproximadamente 100 % idéntica a la SEQ ID NO: 8. Opcionalmente, la secuencia codificante para la cadena ligera humana comprende, consiste esencialmente en, o consiste en una secuencia que es al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 %, o aproximadamente 100 % idéntica a la SEQ ID NO: 16. Por ejemplo, la proteína del factor XII de coagulación codificada por el locus *F12* humanizado puede comprender, consistir esencialmente en o consistir en una secuencia que es al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 %, o aproximadamente 100 % idéntica a la SEQ ID NO: 5 o 46. Opcionalmente, la CDS *F12* codificada por el locus *F12* humanizado puede comprender, consistir esencialmente en o consistir en una secuencia que es al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 %, o aproximadamente 100 % idéntica a la SEQ ID NO: 13 o 48 (o degenerados de los mismos). En cada caso, la proteína del factor XII de coagulación humanizado puede conservar la actividad de la proteína del factor XII de coagulación nativa y/o la proteína del factor XII de coagulación humano.

Opcionalmente, un locus *F12* humanizado puede comprender otros elementos. Ejemplos de dichos elementos pueden incluir casetes de selección, genes indicadores, sitios de reconocimiento de recombinasas u otros elementos. En un ejemplo, un elemento adicional (por ejemplo, un casete de selección) es un intrón de la secuencia humana insertada en el locus *F12* humanizado. En otro ejemplo, un elemento adicional (por ejemplo, un casete de selección) está situado en 3' de la secuencia humana insertada en el locus *F12* humanizado. Alternativamente, el locus *F12* humanizado puede carecer de otros elementos (por ejemplo, puede carecer de un marcador de selección o un casete de selección). En otros lugares del presente documento se divulgan ejemplos de genes indicadores y proteínas indicadoras adecuados. Ejemplos de marcadores de selección adecuados incluyen neomicina fosfotransferasa (neo_r), higromicina B fosfotransferasa (hyg_r), puromicina-N-acetiltransferasa (puro_r), blasticidina S deaminasa (bsr_r), xantina/guanina fosforribosil transferasa (gpt) y timidina quinasa del virus del herpes simple (HSV-k). Ejemplos de recombinasas

incluyen recombinasas Cre, Flp y Dre. Un ejemplo de 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 procariótica. Dichas recombinasas pueden comprender además una señal de localización nuclear para facilitar la localización en el núcleo (por ejemplo, NLS-Crei). Los sitios de reconocimiento de recombinasa incluyen secuencias de nucleótidos que son reconocidas por una recombinasa específica de sitio y pueden servir como sustrato para un evento de recombinasa. Ejemplos de sitios de reconocimiento de recombinasa incluyen sitios FRT, FRT11, FRT71, attP, att, rox y lox tales como loxP, lox511, lox2272, lox66, lox71, loxM2 y lox5171.

Otros elementos tales como genes indicadores o casetes de selección pueden ser casetes de autodelección flanqueados por sitios de reconocimiento de recombinasa. Véase, por ejemplo, los documento US 8,697,851 y US 2013/0312129. Como ejemplo, el casete de autodelección puede comprender un gen Crei (comprende dos exones que codifican una recombinasa Cre, que están separados por un intrón) ligado operativamente a un promotor *P_{rm1}* de ratón y un gen de resistencia a neomicina ligado operativamente a un promotor de ubiquitina humana. Al emplear el promotor *P_{rm1}*, el casete de autodelección puede eliminarse específicamente en células germinales masculinas de animales F0. El polinucleótido que codifica el marcador de selección puede ligarse operativamente a un promotor activo en una célula a la que se orienta. Ejemplos de promotores se describen en otra parte del presente documento. Como otro ejemplo específico, un casete de selección con autodelección puede comprender una secuencia codificante del gen de resistencia a higromicina ligada operativamente a uno o más promotores (por ejemplo, tanto promotores de ubiquitina humana como EM7) seguido de una señal de poliadenilación, seguido de una secuencia codificante de Crei ligada operativamente a uno o más promotores (por ejemplo, un promotor *mP_{rm1}*), seguido de otra señal de poliadenilación, en donde todo el casete está flanqueado por sitios loxP.

El locus *F12* humanizado también puede ser un alelo condicional. Por ejemplo, el alelo condicional puede ser un alelo multifuncional, como se describe en el documento US 2011/0104799. Por ejemplo, el alelo condicional puede comprender: (a) una secuencia de actuación en orientación sentido con respecto a la transcripción de un gen objetivo; (b) un casete de selección de fármacos (DSC) en orientación sentido o antisentido; (c) una secuencia de nucleótidos de interés (NSI) en orientación antisentido; y (d) un módulo condicional por inversión (COIN, que utiliza un intrón que divide el exón y un módulo invertible similar a una trampa genética) en orientación inversa. Véase, por ejemplo, el documento US 2011/0104799. El alelo condicional puede comprender además unidades recombinables que se recombinan tras la exposición a una primera recombinasa para formar un alelo condicional que (i) carece de la secuencia de activación y de la DSC; y (ii) contiene el NSI en orientación sensorial y el COIN en orientación antisentido. Véase, por ejemplo, el documento US 2011/0104799.

Un locus *F12* humanizado de ejemplo (por ejemplo, un locus *F12* de ratón humanizado) es uno en donde una región desde el codón de inicio hasta el codón de parada se reemplaza con la secuencia humana correspondiente. Véase la Figura 1 y las SEQ ID NOS: 17 y 18. En un ejemplo específico, la secuencia de *F12* humano en el locus *F12* endógeno humanizado comprende una secuencia al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 % o aproximadamente 100 % idéntica a la secuencia expuesta en la SEQ ID NO: 31 o 32. En otro ejemplo específico, el locus *F12* endógeno humanizado codifica una proteína que comprende una secuencia al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 %, o aproximadamente 100% idéntica a la secuencia expuesta en la SEQ ID NO: 5 o 46. En otro ejemplo específico, el locus *F12* endógeno humanizado comprende una secuencia codificante que comprende una secuencia al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 %, o aproximadamente 100% idéntica a la secuencia expuesta en la SEQ ID NO: 13 o 48. En otro ejemplo específico, el locus *F12* endógeno humanizado comprende una secuencia al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 % o aproximadamente 100 % idéntica a la secuencia expuesta en la SEQ ID NO: 17 o 18.

C. Genomas de roedores, células de roedores y roedores que comprenden un locus (*F12*) del factor XII de coagulación humanizado

Se proporcionan genomas de roedores, células de roedores y roedores que comprenden un locus (*F12*) del factor XII de coagulación humanizado como se describe en otra parte del presente documento. Los genomas, células o roedores pueden ser masculinos o femeninos. Los genomas, células o roedores pueden ser heterocigotos u homocigotos para el locus *F12* humanizado. Un organismo diploide tiene dos alelos en cada locus genético. Cada par de alelos representa el genotipo de un locus genético específico. Los genotipos se describen como homocigotos si hay dos alelos idénticos en un locus particular y como heterocigotos si los dos alelos difieren. Un roedor que comprende un locus *F12* humanizado puede comprender el locus *F12* endógeno humanizado en su línea germinal.

Los genomas o células de roedor proporcionados en el presente documento pueden ser, por ejemplo, cualquier genoma o célula de roedor que comprenda un locus *F12* o un locus genómico homólogo u ortólogo al locus *F12* humano. Los genomas provienen o las células son células de roedores. Una célula de roedor puede ser, por ejemplo, una célula de rata, una célula de ratón o una célula de hámster. Otros mamíferos no humanos incluyen, por ejemplo,

primates no humanos, monos, simios, orangutanes, gatos, perros, conejos, caballos, toros, ciervos, bisontes, ganado (por ejemplo, especies bovinas tales como vacas, novillos, etc.; especies ovinas, tales como ovejas, cabras, etc.; y especies porcinas, tales como cerdos y jabalíes). Las aves incluyen, por ejemplo, pollos, pavos, avestruces, gansos, patos, etc. También se incluyen los animales domésticos y los animales agrícolas. El término "no humano" excluye a los humanos.

Las células también pueden estar en cualquier tipo de estado diferenciado o indiferenciado. Por ejemplo, una célula puede ser una célula totipotente de roedor, una célula pluripotente de roedor (por ejemplo, una célula pluripotente de roedor tal como una célula madre embrionaria (ES) de ratón o una célula ES de rata), o una célula no pluripotente de roedor. Las células totipotentes incluyen células indiferenciadas que pueden dar lugar a cualquier tipo de célula, y las células pluripotentes incluyen células indiferenciadas que poseen la capacidad de desarrollarse en más de un tipo de célula diferenciada. Dichas células pluripotentes y/o totipotentes pueden ser, por ejemplo, células ES o células similares a ES, tales como células madre pluripotentes inducidas (iPS). Las células ES incluyen células totipotentes o pluripotentes derivadas de embriones que son capaces de contribuir a cualquier tejido del embrión en desarrollo tras su introducción en un embrión. Las células ES pueden derivar de la masa celular interna de un blastocisto y son capaces de diferenciarse en células de cualquiera de las tres capas germinales de los vertebrados (endodermo, ectodermo y mesodermo).

Las células proporcionadas en el presente documento también pueden ser células germinales (por ejemplo, espermatozoides u ovocitos). Las células pueden ser células mitóticamente competentes o células mitóticamente inactivas, células meióticamente competentes o células meióticamente inactivas. De manera similar, las células también pueden ser células somáticas primarias o células que no son una célula somática primaria. Las células somáticas incluyen cualquier célula que no sea un gameto, una célula germinal, un gametocito o una célula madre indiferenciada. Las células pueden ser, por ejemplo, células hepáticas, tal como por ejemplo hepatoblastos o hepatocitos. Alternativamente, las células pueden ser células neuronales, tales como células neuronales piramidales o células neuronales corticales.

Las células adecuadas proporcionadas en el presente documento también incluyen células primarias. Las células primarias incluyen células o cultivos de células que se han aislado directamente de un organismo, órgano o tejido. Las células primarias incluyen células que no están transformadas ni son inmortales. Incluyen cualquier célula obtenida de un organismo, órgano o tejido que no haya pasado previamente en un cultivo de tejidos o que haya pasado previamente en un cultivo de tejidos pero que sea incapaz de pasar indefinidamente en un cultivo de tejidos. Dichas células pueden aislarse mediante técnicas convencionales e incluyen, por ejemplo, hepatocitos o células neuronales.

Otras células adecuadas proporcionadas en el presente documento incluyen células inmortalizadas. Las células inmortalizadas incluyen células de un organismo multicelular que normalmente no proliferarían indefinidamente pero que, debido a una mutación o alteración, han evadido la senescencia celular normal y, en cambio, pueden seguir dividiéndose. Estas mutaciones o alteraciones pueden ocurrir de forma natural o inducirse intencionalmente. Un ejemplo específico de una línea celular inmortalizada es la línea celular de cáncer de hígado humano HepG2. Son bien conocidos numerosos tipos de células inmortalizadas. Las células inmortalizadas o primarias incluyen células que normalmente se usan para cultivar o para expresar genes o proteínas recombinantes.

Las células proporcionadas en el presente documento también incluyen embriones en etapa unicelular (es decir, ovocitos o cigotos fertilizados). Dichos embriones en etapa unicelular pueden ser de cualquier origen genético (por ejemplo, BALB/c, C57BL/6, 129 o una combinación de los mismos para ratones), pueden ser frescos o congelados y pueden derivarse de reproducción natural o fertilización *in vitro*.

Las células proporcionadas en el presente documento pueden ser células normales y sanas, o pueden ser células enfermas o portadoras de mutantes.

Los roedores que comprenden un locus *F12* humanizado como se describe en el presente documento se pueden producir mediante los métodos descritos en otra parte del presente documento. El término "animal", tal como se utiliza en el presente documento, se refiere a roedores (por ejemplo, ratones, ratas, hámsteres y cobayas). El término "animal no humano" excluye a los humanos. Los roedores preferidos incluyen, por ejemplo, ratones y ratas.

El roedor puede tener cualquier origen genético. Por ejemplo, los ratones adecuados pueden ser de una cepa 129, una cepa C57BL/6, una mezcla de 129 y C57BL/6, una cepa BALB/c o una cepa Swiss Webster. Ejemplos de cepas 129 incluyen 129P1, 129P2, 129P3, 129X1, 129S1 (por ejemplo, 129S1/SV, 129S1/SvIm), 129S2, 129S4, 129S5, 129S9/SvEvH, 129S6 (129/SvEvTac), 129S7, 129S8, 129T1, y 129T2. Véase, por ejemplo, Testing et al. (1999) Mamm. Genome 10(8):836. Los ejemplos de cepas C57BL incluyen C57BL/A, C57BL/An, C57BL/GrFa, C57BL/Ka_wN, C57BL/6, C57BL/6J, C57BL/6ByJ, C57BL/6NJ, C57BL/10, C57BL/10ScSn, C57BL/10Cr, y C57BL/Ola. Los ratones adecuados también pueden ser de una mezcla de una cepa 129 antes mencionada y una cepa C57BL/6 antes mencionada (por ejemplo, 50 % 129 y 50 % C57BL/6). Asimismo, los ratones adecuados pueden ser de una mezcla de las cepas 129 antes mencionadas o una mezcla de las cepas BL/6 antes mencionadas (por ejemplo, la cepa 129S6 (129/SvEvTac)).

De manera similar, las ratas pueden ser de cualquier cepa de rata, incluyendo, por ejemplo, una cepa de rata ACI, una cepa de rata Dark Agouti (DA), una cepa de rata Wistar, una cepa de rata LEA, una cepa de rata Sprague Dawley (SD) o una cepa de rata Fischer tal como Fisher F344 o Fisher F6. Las ratas también se pueden obtener a partir de una cepa derivada de una mezcla de dos o más cepas mencionadas anteriormente. Por ejemplo, una rata adecuada puede ser de una cepa DA o una cepa ACI. La cepa de rata ACI se caracteriza por tener agutí negro, con vientre y patas blancas y un haplotipo *RT1^{av1}*. Estas cepas están disponibles en una variedad de fuentes, incluyendo Harlan Laboratories. La cepa de rata Dark Agouti (DA) se caracteriza por tener un pelaje agutí y un haplotipo *RT1^{av1}*. Estas ratas están disponibles en una variedad de fuentes, incluyendo Charles River y Harlan Laboratories. Algunas ratas adecuadas pueden ser de una cepa de ratas endogámicas. Véase, por ejemplo, el documento US 2014/0235933

Los roedores que comprenden un locus *F12* humanizado como se describe en el presente documento expresan una isoforma corta FXII en el cerebro, tal como en las neuronas del cerebro. Por ejemplo, la isoforma corta FXII puede codificarse mediante un ARNm de *F12* detectado mediante sondas contra los exones 11-14 pero no mediante sondas contra los exones 1-6. En un ejemplo específico, la isoforma corta puede ser FXII₂₉₇₋₅₉₆, una proteína que contiene los dominios catalíticos y ricos en prolina de FXII (M297-S596; SEQ ID NO: 74, codificado por SEQ ID NO: 75). La isoforma corta FXII puede, por ejemplo, activarse mediante calicreína plasmática y puede, por ejemplo, convertir pro-HGF en HGF activo. La isoforma corta FXII también puede activarse, por ejemplo, tras su escisión con calicreína. La isoforma corta FXII activada puede, por ejemplo, activar recíprocamente la precalicreína a calicreína. Por ejemplo, la isoforma corta FXII puede ser capaz de iniciar la vía calicreína-cinina. Por ejemplo, una activación breve de la isoforma FXII puede conducir a la generación de bradicinina.

Los roedores que comprenden un locus *F12* humanizado como se describe en el presente documento pueden tener tiempos de coagulación del tiempo de tromboplastina parcial activado (aPTT) que no son significativamente diferentes de los tiempos de coagulación de aPTT en los correspondientes roedores de tipo silvestre. Asimismo, los roedores que comprenden un locus *F12* humanizado como se describe en el presente documento pueden tener valores de tiempo de protrombina (PT) que no son significativamente diferentes de los valores de PT en los correspondientes roedores de tipo silvestre.

Los roedores que comprenden un locus *F12* humanizado como se describe en el presente documento pueden tener niveles de factor XII de coagulación humano en el plasma de entre aproximadamente 1 a aproximadamente 30, entre aproximadamente 1 a aproximadamente 25, entre aproximadamente 1 a aproximadamente 20, entre aproximadamente 2 a aproximadamente 30, entre aproximadamente 2 a aproximadamente 25, entre aproximadamente 2 a aproximadamente 20, entre aproximadamente 3 a aproximadamente 30, entre aproximadamente 3 a aproximadamente 25, entre aproximadamente 3 a aproximadamente 20, entre aproximadamente 4 a aproximadamente 30, entre aproximadamente 4 a aproximadamente 25, entre aproximadamente 4 a aproximadamente 20, entre aproximadamente 5 a aproximadamente 30, entre aproximadamente 5 a aproximadamente 25, o entre aproximadamente 5 a aproximadamente 20 µg/ml. Por ejemplo, el factor XII de coagulación humano en los roedores puede estar entre aproximadamente 5 a aproximadamente 18 µg/ml. Alternativamente, los roedores que comprenden un locus *F12* humanizado como se describe en el presente documento puede tener niveles de factor XII de coagulación humano en el plasma de al menos aproximadamente 1 µg/ml, al menos aproximadamente 2 µg/ml, al menos aproximadamente 3 µg/ml, al menos aproximadamente 4 µg/ml, al menos aproximadamente 5 µg/ml, al menos aproximadamente 6 µg/ml, al menos aproximadamente 7 µg/ml, al menos aproximadamente 8 µg/ml, al menos aproximadamente 9 µg/ml o al menos aproximadamente 10 µg/ml.

III. Métodos de uso de roedores que comprenden un locus (*F12*) del factor XII de coagulación humanizado para evaluar la eficacia de reactivos orientados a hechos de coagulación humana *in vivo* o *ex vivo*

Se proporcionan diversos métodos para usar los roedores que comprenden un locus *F12* humanizado como se describe en otra parte del presente documento para evaluar u optimizar la administración o eficacia de reactivos orientados al factor XII de coagulación humano (por ejemplo, moléculas o complejos terapéuticos) *in vivo* o *ex vivo*. Debido a que los roedores comprenden un locus *F12* humanizado, los roedores reflejarán con mayor precisión la eficacia de un reactivo orientado al factor XII de coagulación humano. Dichos roedores son particularmente útiles para probar reactivos de edición del genoma diseñados para dirigirse al gen *F12* humano porque los roedores divulgados en el presente documento comprenden loci *F12* endógenos humanizados en lugar de inserciones transgénicas de secuencia de *F12* humano en loci genómicos aleatorios, y los loci *F12* endógenos humanizados comprenden una secuencia de *F12* genómica humana ortóloga de regiones codificantes y no codificantes en lugar de una secuencia de ADNc artificial.

A. Métodos para probar la eficacia de los reactivos orientados al factor XII de coagulación humano *in vivo* o *ex-vivo*

Se proporcionan diversos métodos para evaluar la administración o eficacia de reactivos orientados al factor XII de coagulación humano *in vivo* usando roedores que comprenden un locus *F12* humanizado como se describe en otra parte del presente documento. Dichos métodos pueden comprender: (a) introducir en el roedor un reactivo orientado al factor XII de coagulación humano (es decir, administrar el reactivo orientado al factor XII de coagulación humano al roedor); y (b) evaluar la actividad del reactivo orientado al factor XII de coagulación humano.

El reactivo orientado al factor XII de coagulación humano puede ser cualquier agente biológico o químico que se orienta al locus *F12* humano (el gen *F12* humano), al ARNm de *F12* humano o a la proteína del factor XII de coagulación humano. Ejemplos de reactivos orientados al factor XII de coagulación humano se divulgan en otra parte del presente documento e incluyen, por ejemplo, agentes de edición del genoma o proteínas de unión a antígeno. Por ejemplo, el reactivo orientado al factor XII de coagulación humano puede ser un ácido nucleico orientado a *F12* (por ejemplo, ARN guía CRISPR/Cas, ARN en horquilla corta (ARNsh) o pequeños ARN de interferencia (ARNsi)) o un ácido nucleico que codifica una proteína orientada a *F12* (por ejemplo, una proteína Cas tal como Cas9, una ZFN o una TALEN). Alternativamente, el reactivo orientado a *F12* humano puede ser un anticuerpo orientado al factor XII de coagulación o una proteína de unión a antígeno, o cualquier otra molécula grande o pequeña que se orienta al factor XII de coagulación humano. En un ejemplo, el reactivo orientado al factor XII de coagulación humano es un agente de edición del genoma tal como un agente nucleasa y/o un ácido nucleico donante exógeno (por ejemplo, un vector orientado).

Dichos reactivos orientados al factor XII de coagulación humano se pueden administrar mediante cualquier método de administración (por ejemplo, AAV, LNP, HDD o inyección) como se divulga con más detalle en otra parte del presente documento y mediante cualquier vía de administración. Los medios para administrar complejos y moléculas terapéuticos y las vías de administración se divulgan con más detalle en otra parte del presente documento. En métodos particulares, los reactivos se administran mediante administración mediada por AAV. Por ejemplo, AAV8 se puede utilizar para orientar al hígado. En otros métodos particulares, los reactivos se administran mediante administración mediada por LNP. En otros métodos particulares, los reactivos se administran mediante administración hidrodinámica (HDD). La dosis puede ser cualquier dosis adecuada. Por ejemplo, en algunos métodos en donde los reactivos (por ejemplo, ARNm de Cas9 y ARNg) se administran mediante administración mediada por LNP, la dosis puede estar entre aproximadamente 0.01 and aproximadamente 10 mg/kg, aproximadamente 0.01 y aproximadamente 5 mg/kg, entre aproximadamente 0.01 y aproximadamente 4 mg/kg, entre aproximadamente 0.01 y aproximadamente 3 mg/kg, entre aproximadamente 0.01 y aproximadamente 2 mg/kg, entre aproximadamente 0.01 y aproximadamente 1 mg/kg, entre aproximadamente 0.1 y aproximadamente 10 mg/kg, entre aproximadamente 0.1 y aproximadamente 6 mg/kg; entre aproximadamente 0.1 y aproximadamente 5 mg/kg, entre aproximadamente 0.1 y aproximadamente 4 mg/kg, entre aproximadamente 0.1 y aproximadamente 3 mg/kg, entre aproximadamente 0.1 y aproximadamente 2 mg/kg, entre aproximadamente 0.1 y aproximadamente 1 mg/kg, entre aproximadamente 0.3 y aproximadamente 10 mg/kg, entre aproximadamente 0.3 y aproximadamente 6 mg/kg; entre aproximadamente 0.3 y aproximadamente 5 mg/kg, entre aproximadamente 0.3 y aproximadamente 4 mg/kg, entre aproximadamente 0.3 y aproximadamente 3 mg/kg, entre aproximadamente 0.3 y aproximadamente 2 mg/kg, entre aproximadamente 0.3 y aproximadamente 1 mg/kg, aproximadamente 0.1 mg/kg, aproximadamente 0.3 mg/kg, aproximadamente 1 mg/kg, aproximadamente 2 mg/kg, o aproximadamente 3 mg/kg. En un ejemplo específico, la dosis está entre aproximadamente 0.1 y aproximadamente 6 mg/kg, entre aproximadamente 0.1 y aproximadamente 3 mg/kg, o entre aproximadamente 0.1 y aproximadamente 2 mg/kg.

Los métodos para evaluar la actividad del reactivo orientado al factor XII de coagulación humano son bien conocidos y se proporcionan en otra parte del presente documento. La evaluación de la actividad puede realizarse en cualquier tipo de célula, cualquier tipo de tejido o cualquier tipo de órgano como se divulga en otra parte del presente documento. En algunos métodos, la evaluación de la actividad se realiza en plasma. En algunos métodos, la evaluación de la actividad se realiza en las células del hígado. Si el reactivo orientado al factor XII de coagulación es un reactivo de edición del genoma (por ejemplo, un agente nucleasa), dichos métodos pueden comprender evaluar la modificación del locus *F12* humanizado. Como ejemplo, la evaluación puede comprender medir la actividad de unión de extremos no homólogos (NHEJ) en el locus *F12* humanizado. Esto puede comprender, por ejemplo, medir la frecuencia de inserciones o deleciones dentro del locus *F12* humanizado. Por ejemplo, la evaluación puede comprender secuenciar el locus *F12* humanizado en una o más células aisladas del roedor (por ejemplo, secuenciación de próxima generación). La evaluación puede comprender aislar un órgano objetivo (por ejemplo, hígado) o tejido de roedor y evaluar la modificación del locus *F12* humanizado en el órgano o tejido objetivo. La evaluación también puede comprender evaluar la modificación del locus *F12* humanizado en dos o más tipos de células diferentes dentro del órgano o tejido objetivo. De manera similar, la evaluación puede comprender aislar un órgano o tejido no objetivo (por ejemplo, dos o más órganos o tejidos no objetivo) del roedor y evaluar la modificación del locus *F12* humanizado en el órgano o tejido no objetivo.

Dichos métodos también pueden comprender medir los niveles de expresión del ARNm producido por el locus *F12* humanizado, o medir los niveles de expresión de la proteína codificada por el locus *F12* humanizado. Por ejemplo, los niveles de proteína se pueden medir en una célula, tejido u tipo de órgano particular (por ejemplo, hígado), o los niveles secretados se pueden medir en el plasma o suero. Métodos para evaluar la expresión de ARNm de *F12* o proteína expresada del locus *F12* humanizado se proporcionan en otra parte del presente documento y son bien conocidos.

Como ejemplo específico, si el reactivo orientado al factor XII de coagulación humano es un reactivo de edición del genoma (por ejemplo, un agente nucleasa), se puede evaluar el porcentaje de edición (por ejemplo, el número total de inserciones o deleciones observadas sobre el número total de secuencias leídas en la reacción de PCR de un conjunto de células lisadas) en el locus *F12* humanizado (por ejemplo, en células hepáticas).

Los diversos métodos proporcionados anteriormente para evaluar la actividad *in vivo* también se pueden usar para evaluar la actividad de los reactivos orientados al factor XII de coagulación humano *ex vivo* como se describe en otra parte del presente documento.

En algunos métodos, el reactivo orientado al factor XII de coagulación humano es un agente nucleasa, tal como un agente nucleasa CRISPR/Cas, que se orienta al gen *F12* humano. Dichos métodos pueden comprender, por ejemplo: (a) introducir en el agente nucleasa de roedor diseñado para escindir el gen *F12* humano (por ejemplo, proteína Cas tal como Cas9 y un ARN guía diseñado para orientar a una secuencia objetivo de ARN guía en el gen *F12* humano); y (b) evaluar la modificación del locus *F12* humanizado.

En el caso de una nucleasa CRISPR/Cas, por ejemplo, la modificación del locus *F12* humanizado se inducirá cuando el ARN guía forme un complejo con la proteína Cas y dirige la proteína Cas al locus *F12* humanizado, y el complejo Cas/ARN guía escinde la secuencia objetivo del ARN guía, lo que desencadena la reparación por parte de la célula (por ejemplo, mediante unión de extremos no homólogos (NHEJ) si no está presente ninguna secuencia donante).

Opcionalmente, se pueden introducir dos o más ARN guía, cada uno diseñado para orientar a una secuencia objetivo de ARN guía diferente dentro del gen *F12* humano. Por ejemplo, se pueden diseñar dos ARN guía para escindir una secuencia genómica entre las dos secuencias objetivo de ARN guía. La modificación del locus *F12* humanizado se inducirá cuando el primer ARN guía forma un complejo con la proteína Cas y dirige la proteína Cas al locus *F12* humanizado, el segundo ARN guía forma un complejo con la proteína Cas y dirige la proteína Cas al locus *F12* humanizado, el primer complejo Cas/ARN guía escinde la primera secuencia objetivo de ARN guía, y el segundo complejo Cas/ARN guía escinde la segunda secuencia objetivo de ARN guía, dando como resultado la escisión de la secuencia intermedia.

Adicional o alternativamente, también se introduce en el roedor un ácido nucleico donante exógeno (por ejemplo, vector orientado) capaz de recombinarse con un gen *F12* humano y modificarlo. Opcionalmente, el agente nucleasa o la proteína Cas se pueden unir al ácido nucleico donante exógeno como se describe en otra parte del presente documento. La modificación del locus *F12* humanizado se inducirá, por ejemplo, cuando el ARN guía forma un complejo con la proteína Cas y orienta la proteína Cas al locus *F12* humanizado, el complejo Cas/ARN guía escinde la secuencia objetivo del ARN guía, y el locus *F12* humanizado se recombina con el ácido nucleico del donante exógeno para modificar el locus *F12* humanizado. El ácido nucleico donante exógeno puede recombinarse con el locus *F12* humanizado, por ejemplo, mediante reparación dirigida por homología (HDR) o mediante inserción mediada por NHEJ. Se puede utilizar cualquier tipo de ácido nucleico donante exógeno, cuyos ejemplos se proporcionan en otra parte del presente documento.

B. Métodos para optimizar la administración o eficacia del reactivo de orientación al factor XII de coagulación humano *in vivo* o *ex-vivo*

Se proporcionan diversos métodos para optimizar la administración de reactivos orientados al factor XII de coagulación humano a una célula o roedor u optimizar la actividad o eficacia de reactivos orientados al factor XII de coagulación humano *in vivo*. Dichos métodos pueden comprender, por ejemplo: (a) realizar el método para probar la eficacia de reactivos orientados al factor XII de coagulación humano como se describió anteriormente por primera vez en un primer roedor o una primera célula que comprende un locus *F12* humanizado; (b) cambiar una variable y realizar el método una segunda vez en un segundo roedor (es decir, de la misma especie) o una segunda célula que comprende un locus *F12* humanizado con la variable cambiada; y (c) comparar la actividad de los reactivos orientados al factor XII de coagulación humano en el paso (a) con la actividad de los reactivos orientados al factor XII de coagulación humano en el paso (b), y seleccionar el método resultante en la actividad superior.

Los métodos para medir la administración, eficacia o actividad de reactivos orientados al factor XII de coagulación humano se describen en otra parte del presente documento. Por ejemplo, tales métodos pueden comprender medir la modificación del locus *F12* humanizado. Una modificación más eficaz del locus *F12* humanizado puede significar diferentes cosas dependiendo del efecto deseado dentro del roedor o la célula. Por ejemplo, una modificación más eficaz del locus *F12* humanizado puede significar uno o más o todos los niveles más altos de modificación, mayor precisión, mayor consistencia o especificidad. Niveles más altos de modificación (es decir, mayor eficacia) del locus *F12* humanizado se refiere a un mayor porcentaje de células orientadas dentro de un tipo de célula objetivo particular, dentro de un tejido objetivo particular o dentro de un órgano objetivo particular (por ejemplo, hígado). Una mayor precisión se refiere a una modificación más precisa del locus *F12* humanizado (por ejemplo, un mayor porcentaje de células objetivo que tienen la misma modificación o que tienen la modificación deseada sin inserciones y deleciones adicionales no deseadas (por ejemplo, indeles NHEJ)). Una mayor consistencia se refiere a una modificación más consistente del locus *F12* humanizado entre diferentes tipos de células, tejidos u órganos objetivo si se orienta a más de un tipo de célula, tejido u órgano (por ejemplo, modificación de un mayor número de tipos de células dentro del hígado). Si el objetivo es un órgano en particular, una mayor consistencia también puede referirse a una modificación más consistente en todas las ubicaciones dentro del órgano (por ejemplo, el hígado). Una mayor especificidad puede referirse a una mayor especificidad con respecto al locus o loci genómicos objetivo, una mayor especificidad con respecto al tipo de célula objetivo, una mayor especificidad con respecto al tipo de tejido objetivo o una mayor especificidad con respecto al órgano objetivo. Por ejemplo, una mayor especificidad del locus genómico se refiere a una menor modificación de los loci genómicos fuera del objetivo (por ejemplo, un porcentaje menor de células objetivo

que tienen modificaciones en loci genómicos no deseados y fuera del objetivo en lugar de o además de la modificación del locus genómico objetivo). Del mismo modo, una mayor especificidad del tipo de célula, tejido o tipo de órgano se refiere a una menor modificación de tipos de células, tipos de tejido o tipos de órganos fuera del objetivo si se está orientando a un tipo de célula, tipo de tejido o tipo de órgano en particular (por ejemplo, cuando un tipo de célula, tipo de tejido o tipo de órgano en particular está siendo el objetivo) (por ejemplo, el hígado), hay menos modificación de las células en órganos o tejidos que no son los objetivos previstos).

Alternativamente, dichos métodos pueden comprender medir la expresión del ARNm de *F12* o de la proteína del factor XII de coagulación. En un ejemplo, un agente orientado al factor XII de coagulación humano más eficaz da como resultado una mayor disminución en la expresión de la proteína del ARNm de *F12* o del factor XII de coagulación. Alternativamente, tales métodos pueden comprender medir la actividad del factor XII de coagulación. En un ejemplo, un agente orientado al factor XII de coagulación humano más eficaz da como resultado una mayor disminución en la actividad del factor XII de coagulación.

La variable que se cambia puede ser cualquier parámetro. Como ejemplo, la variable cambiada puede ser el embalaje o el método de administración (por ejemplo, vehículos de administración) mediante el cual el reactivo o reactivos orientados al factor XII de coagulación humana se introducen en la célula o roedor. En otras partes del presente documento se divulgan ejemplos de métodos de administración vehículos de entrega, tales como LNP, HDD y AAV. Por ejemplo, la variable cambiada puede ser el serotipo de AAV. De manera similar, la administración puede comprender administración mediada por LNP, y la variable cambiada puede ser la formulación de LNP. Como otro ejemplo, la variable cambiada puede ser la vía de administración para la introducción del reactivo o reactivos orientados al factor XII de coagulación humano en la célula o roedor. En otra parte del presente documento se divulgan ejemplos de vías de administración, tales como instilación intravenosa, intravítrea, intraparenquimatosa y nasal.

Como otro ejemplo, la variable cambiada puede ser la concentración o cantidad del reactivo o reactivos orientados al factor XII de coagulación humano introducidos. Como otro ejemplo, la variable cambiada puede ser la concentración o la cantidad de un reactivo orientado al factor XII de coagulación humano introducido (por ejemplo, ARN guía, proteína Cas, ácido nucleico donante exógeno, agente de ARNi o ASO) en relación con la concentración o la cantidad de otro reactivo orientado al factor XII de coagulación humano introducido (por ejemplo, ARN guía, proteína Cas, ácido nucleico donante exógeno, agente de ARNi o ASO).

Como otro ejemplo, la variable cambiada puede ser el momento de introducir el reactivo o reactivos orientados al factor XII de coagulación humano en relación con el momento de evaluar la actividad o eficacia de los reactivos. Como otro ejemplo, la variable cambiada puede ser el número de veces o la frecuencia con la que se introducen el reactivo o reactivos orientados al factor XII de coagulación humano. Como otro ejemplo, la variable cambiada puede ser el momento de la introducción de un reactivo orientado al factor XII de coagulación humano introducido (por ejemplo, ARN guía, proteína Cas, ácido nucleico donante exógeno, agente de ARNi o ASO) en relación con el momento de introducción de otro reactivo orientado al factor XII de coagulación humano introducido (por ejemplo, ARN guía, proteína Cas, ácido nucleico donante exógeno, agente de ARNi o ASO).

Como otro ejemplo, la variable cambiada puede ser la forma en donde se introducen el reactivo o reactivos orientados al factor XII de coagulación humano. Por ejemplo, un ARN guía puede introducirse en forma de ADN o en forma de ARN. Una proteína Cas (por ejemplo, Cas9) se puede introducir en forma de ADN, en forma de ARN o en forma de una proteína (por ejemplo, formando un complejo con un ARN guía). Un ácido nucleico donante exógeno puede ser ADN, ARN, monocatenario, bicatenario, lineal, circular, etc. De manera similar, cada uno de los componentes puede comprender diversas combinaciones de modificaciones para la estabilidad, para reducir los efectos fuera del objetivo, para facilitar la administración, etc. Asimismo, los agentes de ARNi y los ASO, por ejemplo, pueden comprender diversas combinaciones de modificaciones para la estabilidad, para reducir los efectos fuera del objetivo, para facilitar la administración, etc.

Como otro ejemplo, la variable cambiada puede ser el reactivo o reactivos orientados al factor XII de coagulación humano que se introducen. Por ejemplo, si el reactivo orientado al factor XII de coagulación humano comprende un ARN guía, la variable cambiada puede ser introducir un ARN guía diferente con una secuencia diferente (por ejemplo, orientarse a una secuencia objetivo de ARN guía diferente). De manera similar, si el reactivo orientado al factor XII de coagulación humano comprende un agente de ARNi o un ASO, la variable cambiada puede ser la introducción de un agente de ARNi o ASO diferente con una secuencia diferente. Asimismo, si el reactivo orientado al factor XII de coagulación humano comprende una proteína Cas, la variable cambiada puede ser introducir una proteína Cas diferente (por ejemplo, introducir una proteína Cas diferente con una secuencia diferente, o un ácido nucleico con una secuencia diferente (por ejemplo, con codones optimizados) pero que codifica la misma secuencia de aminoácidos de la proteína Cas. Asimismo, si el reactivo orientado al factor XII de coagulación humano comprende un ácido nucleico donante exógeno, la variable cambiada puede introducir un ácido nucleico donante exógeno diferente con una secuencia diferente (por ejemplo, un ácido nucleico de inserto diferente o brazos de homología diferentes (por ejemplo, brazos de homología más largos o cortos o brazos de homología orientados a una región diferente del gen *F12* humano)).

En un ejemplo específico, el reactivo orientado al factor XII de coagulación humano comprende una proteína Cas y un ARN guía diseñado para orientar a una secuencia objetivo de ARN guía en un gen *F12* humano. En tales métodos, la

variable cambiada puede ser la secuencia de ARN guía y/o la secuencia objetivo de ARN guía. En algunos de tales métodos, la proteína Cas y el ARN guía se pueden administrar cada uno en forma de ARN, y la variable cambiada puede ser la relación de ARNm de Cas a ARN guía (por ejemplo, en una formulación de LNP). En algunos de tales métodos, la variable cambiada puede ser modificaciones del ARN guía (por ejemplo, un ARN guía con una modificación se compara con un ARN guía sin la modificación).

C. Reactivos orientados al factor XII de coagulación humano

Un reactivo orientado al factor XII de coagulación humano puede ser cualquier reactivo que se orienta a un gen *F12* humano, un ARNm de *F12* humano o una proteína del factor XII de coagulación humano. Por ejemplo, puede ser un reactivo de edición del genoma tal como un agente nucleasa que escinde una secuencia objetivo dentro del gen *F12* humano y/o una secuencia donante exógena que se recombina con un gen *F12* humano, puede ser un oligonucleótido antisentido orientado a un ARNm de *F12* humano, puede ser una proteína de unión a antígeno orientada a un epítipo de una proteína del factor XII de coagulación humano, o puede ser una molécula pequeña orientada al factor XII de coagulación humano. Los reactivos orientados al factor XII de coagulación humano en los métodos divulgados en el presente documento pueden ser reactivos orientados al factor XII de coagulación humano conocidos, pueden ser reactivos orientados al factor XII de coagulación humano putativos (por ejemplo, reactivos candidatos diseñados para orientar al factor XII de coagulación humano), o pueden ser reactivos que se están cribando para determinar la actividad orientada al factor XII de coagulación humano.

(1) Agentes nucleasa orientados al gen *F12* humano

Un reactivo orientado al factor XII de coagulación humano puede ser un reactivo de edición del genoma tal como un agente nucleasa que escinde una secuencia objeto dentro del gen *F12* humano. Una secuencia objetivo de nucleasa incluye una secuencia de ADN en donde un agente de nucleasa induce una mella o ruptura bicatenaria. La secuencia objetivo de un agente nucleasa puede ser endógena (o nativa) de la célula o la secuencia objetivo puede ser exógena a la célula. Una secuencia objetivo que es exógena a la célula no se encuentra de forma natural en el genoma de la célula. La secuencia objetivo también puede ser exógena a los polinucleótidos de interés que se desea posicionar en el locus objetivo. En algunos casos, la secuencia objetivo está presente sólo una vez en el genoma de la célula huésped.

La longitud de la secuencia objetivo puede variar, e incluye, por ejemplo, secuencias objetivo que tienen aproximadamente 30-36 pb para un par de nucleasas con dedos de zinc (ZFN) (es decir, aproximadamente 15-18 pb para cada ZFN), aproximadamente 36 pb para una nucleasa efectora similar a un activador de la transcripción (TALEN), o aproximadamente 20 pb para un ARN guía CRISPR/Cas9.

Cualquier agente nucleasa que induzca una mella o ruptura bicatenaria en una secuencia objetivo deseada se puede utilizar en los métodos y composiciones divulgados en el presente documento. Se puede emplear un agente de nucleasa natural o nativo siempre que el agente de nucleasa induzca una mella o ruptura bicatenaria en una secuencia objetivo deseada. Alternativamente, se puede emplear un agente nucleasa modificado o diseñado. Un "agente de nucleasa diseñado" incluye una nucleasa que está diseñada (modificada o derivada) de su forma nativa para reconocer e inducir específicamente una mella o ruptura bicatenaria en la secuencia objetivo deseada. Por tanto, un agente de nucleasa modificado genéticamente puede derivarse de un agente de nucleasa nativo de origen natural o puede crearse o sintetizarse artificialmente. La nucleasa diseñada puede inducir una mella o ruptura bicatenaria en una secuencia objetivo, por ejemplo, en donde la secuencia objetivo no es una secuencia que habría sido reconocida por un agente de nucleasa nativo (no modificado o no diseñado). La modificación del agente nucleasa puede ser tan pequeña como un aminoácido en un agente de escisión de proteínas o un nucleótido en un agente de escisión de ácidos nucleicos. A la producción de una mella o ruptura bicatenaria en una secuencia objetivo u otro ADN se le puede denominar en el presente documento "cortar" o "escindir" la secuencia objetivo u otro ADN.

También se proporcionan variantes activas y fragmentos de las secuencias objetivo ejemplificadas. 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 respecto a la secuencia objetivo dada, en donde las variantes activas conservan la actividad biológica y, por lo tanto, son capaces de ser reconocidas y escindidas por un agente nucleasa de una manera específica de secuencia. Son bien conocidos los ensayos para medir la ruptura bicatenaria de una secuencia objetivo mediante un agente nucleasa. Véase, *por ejemplo*, Frendewey et al. (2010) *Methods in Enzymology* 476:295-307.

La secuencia objetivo del agente nucleasa se puede colocar en cualquier lugar dentro o cerca del locus *F12*. La secuencia objetivo puede ubicarse dentro de una región codificante de la gen *F12*, o dentro de regiones reguladoras que influyen en la expresión del gen. Una secuencia objetivo del agente nucleasa puede localizarse en un intrón, un exón, un promotor, un potenciador, una región reguladora o cualquier región codificante no proteica.

Un tipo de agente nucleasa es una nucleasa efectora similar a un activador de la transcripción (TALEN). Las nucleasas efectoras TAL son una clase de nucleasas de secuencia específica que se pueden usar para realizar rupturas bicatenarias en secuencias objetivo específicas en el genoma de un organismo procariótico o eucariota. Las nucleasas efectoras TAL se crean fusionando un efector similar a un activador de la transcripción (TAL) nativo o diseñado, o una

parte funcional del mismo, al dominio catalítico de una endonucleasa, tal como, por ejemplo, FokI. El exclusivo dominio modular de unión al ADN efector TAL permite el diseño de proteínas con potencialmente cualquier especificidad de reconocimiento de ADN determinada. Por lo tanto, los dominios de unión al ADN de las nucleasas efectoras TAL pueden modificarse para reconocer sitios objetivo de ADN específicos y, por lo tanto, usarse para realizar rupturas bicatenarias en secuencias objetivo deseadas. Véase el documento WO 2010/079430; Morbitzer et al. (2010) Proc. Natl. Acad. Sci. EE. UU. 107(50):21617-21622; Scholze & Boch (2010) Virulence 1:428-432; Christian et al. Genetics (2010) 186:757-761; Li et al. (2010) Nuc. Acids Res. (2010) 39(1):359-372; and Miller et al. (2011) Nature Biotechnology 29:143-148.

Se describen ejemplos de nucleasas TAL adecuadas y métodos para preparar nucleasas TAL adecuadas, por ejemplo, en los documentos US 2011/0239315 A1, US 2011/0269234 A1, US 2011/0145940 A1, US 2003/0232410 A1, US 2005/0208489 A1, US 2005/0026157 A1, US 2005/0064474 A1, US 2006/0188987 A1, y US 2006/0063231 A1. En diversas realizaciones, las nucleasas efectoras TAL se modifican para cortar en o cerca de una secuencia de ácido nucleico objetivo en, por ejemplo, un locus de interés o un locus genómico de interés, en donde la secuencia de ácido nucleico objetivo está en o cerca de una secuencia que se va a modificar mediante un vector orientado. Las nucleasas TAL adecuadas para su 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 ácido nucleico objetivo que se modificarán mediante vectores orientados como se describe en el presente documento.

En algunos TALEN, cada monómero de TALEN comprende 33-35 repeticiones TAL que reconocen un solo par de bases a través de dos residuos hipervariables. En algunos TALEN, el agente nucleasa es una proteína quimérica que comprende un dominio de unión al ADN basado en repeticiones TAL ligado operativamente a una nucleasa independiente tal como una endonucleasa FokI. Por ejemplo, el agente nucleasa puede comprender un primer dominio de unión a ADN basado en repeticiones TAL y un segundo dominio de unión a ADN basado en repeticiones TAL, en donde cada uno de los dominios de unión a ADN basados en repeticiones TAL primero y segundo está operativamente ligado a una nucleasa FokI, en donde el primer y el segundo dominio de unión a ADN basado en repeticiones TAL reconocen dos secuencias de ADN objetivo contiguas en cada cadena de la secuencia de ADN objetivo separadas por una secuencia espaciadora de longitud variable (12-20 pb), y en donde las subunidades de nucleasa FokI se dimerizan para crear una nucleasa activa que rompe una doble cadena en una secuencia objetivo.

El agente nucleasa empleado en los diversos métodos y composiciones divulgados en el presente documento puede comprender además una nucleasa con dedos de zinc (ZFN). En algunas ZFN, cada monómero de la ZFN comprende 3 o más dominios de unión a ADN basados en dedos de zinc, en donde cada dominio de unión a ADN basado en dedos de zinc se une a un subsitio de 3 pb. En otras ZFN, la ZFN es una proteína quimérica que comprende un dominio de unión a ADN basado en dedos de zinc ligado operativamente a una nucleasa independiente tal como una endonucleasa FokI. Por ejemplo, el agente nucleasa puede comprender una primera ZFN y una segunda ZFN, en donde cada una de la primera ZFN y la segunda ZFN está operativamente ligada a una subunidad de nucleasa FokI, en donde la primera y la segunda ZFN reconocen dos secuencias de ADN objetivo contiguas en cada cadena de la secuencia de ADN objetivo separada por aproximadamente 5-7 pb espaciador, y en donde las subunidades de nucleasa FokI se dimerizan para crear una nucleasa activa que rompe la doble cadena. Véase, *por ejemplo*, los documentos US20060246567; US20080182332; US20020081614; US20030021776; WO/2002/057308A2; US20130123484; US20100291048; WO/2011/017293A2; y Gaj et al. (2013) Trends in Biotechnology, 31(7):397-405.

Otro tipo de agente nucleasa es una meganucleasa modificada. Las meganucleasas se han clasificado en cuatro familias según motivos de secuencia conservados, las familias son las familias LAGLIDADG, GIY-YIG, H-N-H e His-Cys. Estos motivos participan en la coordinación de iones metálicos y en la hidrólisis de enlaces fosfodiéster. Las meganucleasas se destacan por sus largas secuencias objetivo y por tolerar algunos polimorfismos de secuencia en sus sustratos de ADN. Los dominios, la estructura y la función de la meganucleasa son conocidos, véase, *por ejemplo*, Guhan y Muniyappa (2003) Crit Rev Biochem Mol Biol 38:199-248; Lucas et al., (2001) Nucleic Acids Res 29:960-9; Jurica y Stoddard, (1999) Cell Mol Life Sci 55:1304-26; Stoddard, (2006) Q Rev Biophys 38:49-95; y Moure et al., (2002) Nat Struct Biol 9:764. En algunos ejemplos, se utiliza una variante natural y/o una meganucleasa derivada genéticamente modificada. Se conocen métodos para modificar la cinética, las interacciones de cofactores, la expresión, las condiciones óptimas y/o la especificidad de la secuencia objetivo y la detección de 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; los documentos WO2005105989; WO2003078619; WO2006097854; WO2006097853; WO2006097784; y WO2004031346.

Puede usarse cualquier meganucleasa, incluyendo, por ejemplo, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAII, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIIP, I-TII, I-PpoI, PI-PsPI, F-SceI, F-SceII, F-SuvI, F-TevI, F-TevII, I-Amal, I-Aml, I-Chul, I-Cmoel, I-Cpal, I-Cpall, I-Csml, I-Cvul, I-CvuAIP, I-DdII, I-DdIII, I-Dirl, I-Dmol, I-Hmul, I-Hmull, I-HsNIP, I-LlaI, I-Msol, I-Naal, 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-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdelIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIP, PI-MtuHIIP, PI-Pful, PI-Pfull, PI-

PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetalP, PI-Scel, PI-Tful, PI-Tfull, PI-Thyl, PI-Tiil, PI-Tiill, o cualquier variante o fragmento activo de las mismas.

Las meganucleasas pueden reconocer, por ejemplo, secuencias de ADN bicatenario de 12 a 40 pares de bases. En algunos casos, la meganucleasa reconoce una secuencia objetivo perfectamente coincidente en el genoma.

- 5 Algunas meganucleasas son nucleasas dirigidas. Un tipo de nucleasa dirigida es una familia LAGLIDADG de nucleasas dirigidas que incluye, por ejemplo, I-Scel, I-Crel e I-Dmol.

Los agentes nucleasas pueden comprender además sistemas CRISPR/Cas como se describe con más detalle a continuación.

- 10 También se proporcionan variantes activas y fragmentos de agentes nucleasa (es decir, un agente nucleasa modificado). 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 nucleasa nativo, en donde las variantes activas conservan la capacidad de cortar en una secuencia objetivo deseada y, por tanto, conservan la actividad inductora de mella o ruptura bicatenaria. Por ejemplo, cualquiera de los agentes nucleasa descritos en el presente documento puede modificarse a partir de una secuencia de endonucleasa nativa y diseñarse para reconocer e inducir una mella o ruptura bicatenaria en una secuencia objetivo que no fue reconocida por el agente nucleasa nativo. Por lo tanto, algunas nucleasas modificadas genéticamente tienen una especificidad para inducir una mella o ruptura bicatenaria en una secuencia objetivo que es diferente de la correspondiente secuencia objetivo del agente de nucleasa nativo. Se conocen ensayos para determinar la actividad inductora de mella o ruptura bicatenaria y generalmente miden la actividad general y la especificidad de la endonucleasa sobre sustratos de ADN que contienen la secuencia objetivo.

- 15 El agente nucleasa puede introducirse en una célula o roedor mediante cualquier medio conocido. Un polipéptido que codifica el agente nucleasa puede introducirse directamente en la célula o en el roedor. Alternativamente, se puede introducir en la célula o roedor un polinucleótido que codifica el agente nucleasa. Cuando se introduce un polinucleótido que codifica el agente nucleasa, el agente nucleasa puede expresarse de forma transitoria, condicional o constitutiva dentro de la célula. El polinucleótido que codifica el agente nucleasa puede estar contenido en un casete de expresión y estar ligado operativamente a un promotor condicional, un promotor inducible, un promotor constitutivo o un promotor específico de tejido. Ejemplos de promotores se analizan con más detalle en otra parte del presente documento. Alternativamente, el agente nucleasa se puede introducir en la célula como un ARNm que codifica el agente nucleasa.

- 25 Un polinucleótido que codifica un agente nucleasa puede integrarse de manera estable en el genoma de una célula y ligarse operativamente a un promotor activo en la célula. Alternativamente, un polinucleótido que codifica un agente nucleasa puede estar en un vector orientado.

- 30 Cuando el agente nucleasa se proporciona a la célula mediante la introducción de un polinucleótido que codifica el agente nucleasa, dicho polinucleótido que codifica un agente nucleasa se puede modificar para sustituir codones que tengan una mayor frecuencia de uso en la célula de interés, en comparación con la secuencia de polinucleótidos de origen natural que codifica el agente nucleasa. Por ejemplo, el polinucleótido que codifica el agente nucleasa se puede modificar para sustituir codones que tengan una mayor frecuencia de uso en una célula eucariótica de interés dada, es decir, 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 de origen natural.

(2) Sistemas CRISPR/Cas orientados al gen *F12* humano

- 40 Un tipo particular de reactivo orientado al factor XII de coagulación humano puede ser un sistema de repeticiones palindrómicas cortas agrupadas regularmente intercaladas (CRISPR)/asociado a CRISPR (Cas) que se orientan al gen humano *F12*. Los sistemas CRISPR/Cas incluyen transcripciones y otros elementos implicados en la expresión o dirección de la actividad de los genes Cas. Un sistema CRISPR/Cas puede ser, por ejemplo, un sistema de tipo I, de tipo II, de tipo III o de tipo V (por ejemplo, subtipo V-A o subtipo V-B). Los sistemas CRISPR/Cas utilizados en las composiciones y métodos descritos en el presente documento pueden ser de origen no natural. Un sistema "que no se produce naturalmente" incluye cualquier cosa que indique la participación de la mano del hombre, tal como uno o más componentes del sistema que sean alterados o mutados de su estado natural, que esté al menos sustancialmente libre de al menos otro componente con el que están asociados de forma natural en la naturaleza, o estando asociados con al menos otro componente con el que no están asociados de forma natural. Por ejemplo, algunos sistemas CRISPR/Cas emplean complejos CRISPR que no se producen de forma natural que comprenden un ARNg y una proteína Cas que no se producen juntos de forma natural, emplean una proteína Cas que no se produce de forma natural o emplean un ARNg que no se produce de forma natural.

- 55 Proteínas Cas y polinucleótidos que codifican proteínas Cas. Las proteínas Cas generalmente comprenden al menos un dominio de unión o reconocimiento de ARN que puede interactuar con los ARN guía (ARNg). Las proteínas cas también pueden comprender dominios de nucleasa (por ejemplo, dominios de ADNasa o dominios de ARNasa), dominios de unión a ADN, dominios de helicasa, dominios de interacción proteína-proteína, dominios de dimerización y otros dominios. Algunos de estos dominios (por ejemplo, dominios DNasa) pueden provenir de una proteína Cas nativa. Se pueden agregar otros dominios similares para producir una proteína Cas modificada. Un dominio de

nucleasa posee actividad catalítica para la escisión de ácidos nucleicos, que incluye la ruptura de los enlaces covalentes de una molécula de ácido nucleico. La escisión puede producir extremos romos o escalonados, y puede ser monocatenario o bicatenario. Por ejemplo, una proteína Cas9 de tipo silvestre normalmente creará un producto de escisión romo. Alternativamente, una proteína Cpf1 de tipo silvestre (por ejemplo, FnCpf1) puede dar como resultado un producto de escisión con un saliente 5' de 5 nucleótidos, ocurriendo la escisión después del par de bases 18 de la secuencia PAM en la cadena no objetivo y después de la base 23 en la cadena objetivo. Una proteína Cas puede tener actividad de escisión completa para crear una ruptura bicatenaria en un locus genómico objetivo (por ejemplo, una ruptura de doble cadena con extremos romos), o puede ser una nickasa que crea una ruptura monocatenaria en un locus genómico objetivo.

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

Una proteína Cas ejemplar es una proteína Cas9 o una proteína derivada de una proteína Cas9. Las proteínas Cas9 provienen de un sistema CRISPR/Cas de tipo II y normalmente comparten cuatro motivos clave con una arquitectura conservada. Los motivos 1, 2 y 4 son motivos similares a RuvC y el motivo 3 es un motivo HNH. Las proteínas Cas9 ejemplares son de *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus sp.*, *Staphylococcus aureus*, *Nocardia* *dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycoides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microscilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas sp.*, *Crocospaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor beccii*, *Candidatus Desulforudis*, *Clostridium botulinum*, *Clostridium difficile*, *Fingoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochrochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira sp.*, *Lyngbya sp.*, *Microcoleus chthonoplastes*, *Oscillatoria sp.*, *Petrotoga mobilis*, *Thermosiphon africanus*, *Acaryochloris marina*, *Neisseria meningitidis*, o *Campylobacter jejuni*. Ejemplos adicionales de los miembros de la familia Cas9 se describen en el documento WO 2014/131833. Cas9 de *S. pyogenes* (SpCas9) (al que se le ha asignado el número de acceso de SwissProt Q99ZW2) es una proteína Cas9 ejemplar. Cas9 de *S. aureus* (SaCas9) (al que se le ha asignado el número de acceso de UniProt J7RUA5) es otra proteína Cas9 ejemplar. Cas9 de *Campylobacter jejuni* (CjCas9) (al que se le ha asignado el número de acceso de UniProt Q0P897) es otra proteína Cas9 ejemplar. Véase, por ejemplo, Kim et al. (2017) Nat. Com. 8:14500. SaCas9 es más pequeño que SpCas9 y CjCas9 es más pequeño que SaCas9 y SpCas9. Cas9 de *Neisseria meningitidis* (Nme2Cas9) es otra proteína Cas9 ejemplar. Véase, por ejemplo, Edraki et al. (2019) Mol. Cell 73(4):714-726. Proteínas Cas9 de *Streptococcus thermophilus* (por ejemplo, *Streptococcus thermophilus* LMD-9 Cas9 codificado por el locus CRISPR1 (St1Cas9) o *Streptococcus thermophilus* Cas9 del locus CRISPR3 (St3Cas9)) son otras proteínas Cas9 ejemplares. Cas9 de *Francisella novicida* (FnCas9) o la variante Cas9 de RHA *Francisella novicida* que reconoce una PAM alternativa (sustituciones E1369R/E1449H/R1556A) son otras proteínas Cas9 ejemplares. Estas y otras proteínas Cas9 ejemplares se revisan, por ejemplo, en Cebrián-Serrano y Davies (2017) Mamm. Genome 28(7):247-261. Una secuencia de proteína Cas9 ejemplar puede comprender, consistir esencialmente en o consistir en la SEQ ID NO: 35. Un ADN ejemplar que codifica la proteína Cas9 puede comprender, consistir esencialmente en o consistir en la SEQ ID NO: 36.

Otro ejemplo de proteína Cas es una proteína Cpf1 (CRISPR de *Prevotella* y *Francisella* 1). Cpf1 es una proteína grande (aproximadamente 1300 aminoácidos) que contiene un dominio de nucleasa similar a RuvC homólogo al dominio correspondiente de Cas9 junto con una contraparte del característico grupo rico en arginina de Cas9. Sin embargo, Cpf1 carece del dominio de nucleasa HNH que está presente en las proteínas Cas9, y el dominio similar a RuvC es contiguo en la secuencia de Cpf1, en contraste con Cas9, donde contiene inserciones largas que incluyen el dominio HNH. Véase, por ejemplo, Zetsche et al. (2015) Cell 163(3):759-771. Las proteínas Cpf1 ejemplares son de *Francisella tularensis* 1, *Francisella tularensis* subsp. *novicida*, *Prevotella albensis*, *Lachnospiraceae bacterium MC2017 1*, *Butyrivibrio proteoclasticus*, *Peregrinibacteria bacterium GW2011_GWA2_33_10*, *Parcubacteria bacterium GW2011_GWC2_44_17*, *Smithella sp.* SCADC, *Acidaminococcus sp.* BV3L6, *Lachnospiraceae bacterium MA2020*, *Candidatus Methanoplasma termitum*, *Eubacterium eligens*, *Moraxella bovoculi* 237, *Leptospira inadai*, *Lachnospiraceae bacterium ND2006*, *Porphyromonas crevioricanis* 3, *Prevotella disiens*, y *Porphyromonas macacae*. Cpf1 de *Francisella novicida* U112 (FnCpf1; número de acceso UniProt asignado A0Q7Q2) es una proteína Cpf1 ejemplar.

Las proteínas Cas pueden ser proteínas de tipo silvestre (es decir, aquellas que se encuentran en la naturaleza), proteínas Cas modificadas (es decir, variantes de la proteína Cas) o fragmentos de proteínas Cas de tipo silvestre o modificadas. Las proteínas Cas también pueden ser variantes o fragmentos activos con respecto a la actividad catalítica de proteínas Cas de tipo silvestre o modificadas. Las variantes o fragmentos activos con respecto a la actividad catalítica pueden comprender al menos 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más identidad de secuencia con la proteína Cas de tipo silvestre o modificada o una porción de la misma, en donde

las variantes activas conservan la capacidad de cortar en un sitio de escisión deseado y, por lo tanto, conservan la actividad inductora de mellas o de inducción de ruptura bicatenaria. Se conocen ensayos para determinar la actividad inductora de mellas o de ruptura bicatenaria y generalmente miden la actividad global y la especificidad de la proteína Cas sobre sustratos de ADN que contienen el sitio de escisión.

- 5 Las proteínas Cas pueden modificarse para aumentar o disminuir una o más de la afinidad de unión a ácidos nucleicos, la especificidad de unión a ácidos nucleicos y la actividad enzimática. Las proteínas Cas también se pueden modificar para cambiar cualquier otra actividad o propiedad de la proteína, tal como la estabilidad. Por ejemplo, se pueden modificar, eliminar o inactivar uno o más dominios de nucleasa de la proteína Cas, o se puede truncar una proteína Cas para eliminar dominios que no son esenciales para la función de la proteína o para optimizar (por ejemplo, mejorar o reducir) la actividad o una propiedad de la proteína Cas.

10 Un ejemplo de proteína Cas modificada es la proteína SpCas9-HF1 modificada, que es una variante de alta fidelidad de Cas9 de *Streptococcus pyogenes* que alberga alteraciones (N497A/R661A/Q695A/Q926A) diseñadas para reducir los contactos de ADN no específicos. Véase, por ejemplo, Kleinstiver et al. (2016) Nature 529(7587):490-495. Otro ejemplo de proteína Cas modificada es la variante eSpCas9 modificada (K848A/K1003A/R1060A) diseñada para reducir los efectos fuera del objetivo. Véase, por ejemplo, Slaymaker et al. (2016) Science 351(6268):84-88. Otras variantes de SpCas9 incluyen K855A y K810A/K1003A/R1060A. Estas y otras proteínas Cas modificadas se revisan, por ejemplo, en Cebrián-Serrano y Davies (2017) Mamm. Genome 28(7):247-261. Otro ejemplo de una proteína Cas9 modificada es xCas9, que es una variante de SpCas9 que puede reconocer una gama ampliada de secuencias PAM. Véase, por ejemplo, Hu et al. (2018) Nature 556:57-63.

- 20 Las proteínas cas pueden comprender al menos un dominio de nucleasa, tal como un dominio de ADNasa. Por ejemplo, una proteína CpfI de tipo silvestre generalmente comprende un dominio similar a RuvC que escinde ambas cadenas de ADN objetivo, quizás en una configuración dimerica. Las proteínas Cas también pueden comprender al menos dos dominios de nucleasa, tales como dominios de ADNasa. Por ejemplo, una proteína Cas9 de tipo silvestre comprende generalmente un dominio de nucleasa similar a RuvC y un dominio de nucleasa similar a HNH. Cada uno de los dominios RuvC y HNH puede cortar una cadena diferente de ADN bicatenario para realizar una ruptura bicatenaria en el ADN. Véase, por ejemplo, Jinek et al. (2012) Science 337(6096):816-.

30 Uno o más o todos los dominios de nucleasa pueden eliminarse o mutarse de modo que ya no sean funcionales o tengan una actividad de nucleasa reducida. Por ejemplo, si uno de los dominios de nucleasa se elimina o muta en una proteína Cas9, la proteína Cas9 resultante puede denominarse nickasa y puede generar una ruptura monocatenaria dentro de un ADN objetivo bicatenario, pero no una ruptura bicatenaria (es decir, puede escindir la cadena complementaria o la cadena no complementaria, pero no ambas). Si ambos dominios de nucleasa se eliminan o mutan, la proteína Cas resultante (por ejemplo, Cas9) tendrá una capacidad reducida para escindir ambas cadenas de un ADN bicatenario (por ejemplo, una proteína Cas nula o inactiva para nucleasa, o una proteína Cas catalíticamente muerta (dCas)). Un ejemplo de una mutación que convierte Cas9 en una nickasa es una mutación D10A (aspartato a alanina en la posición 10 de Cas9) en el dominio RuvC de Cas9 de *S. pyogenes*. Asimismo, H939A (histidina a alanina en la posición del aminoácido 839), H840A (histidina a alanina en la posición del aminoácido 840), o N863A (asparagina a alanina en la posición del aminoácido N863) en el dominio HNH de Cas9 de *S. pyogenes* puede convertir el Cas9 en un nickasa. Otros ejemplos de mutaciones que convierten Cas9 en una nickasa incluyen las mutaciones correspondientes a Cas9 de *S. thermophilus*. Véase, por ejemplo, Sapranaukas et al. (2011) Nucleic Acids Res. 39(21):9275-9282 y el documento WO 2013/141680. Estas mutaciones se pueden generar utilizando métodos tales como la mutagénesis orientada al sitio, la mutagénesis mediada por PCR o la síntesis genética total. Se pueden encontrar ejemplos de otras mutaciones que crean nickasas, por ejemplo, en los documentos WO 2013/176772 y WO 2013/142578. Si todos los dominios de nucleasa se eliminan o mutan en una proteína Cas (por ejemplo, ambos dominios de nucleasa se eliminan o mutan en una proteína Cas9), la proteína Cas resultante (por ejemplo, Cas9) tendrá una capacidad reducida para escindir ambas cadenas de un ADN bicatenario (por ejemplo, una proteína Cas nula o inactiva para nucleasas). Un ejemplo específico es un doble mutante Cas9 de *S. pyogenes* D10A/H840A o un doble mutante correspondiente en un Cas9 de otra especie cuando se alinea de manera óptima con Cas9 de *S. pyogenes*. Otro ejemplo específico es un doble mutante Cas9 de D10A/N863A *S. pyogenes* o un doble mutante correspondiente en un Cas9 de otra especie cuando se alinea de manera óptima con Cas9 de *S. pyogenes*.

- 50 Los ejemplos de mutaciones inactivadoras en los dominios catalíticos de xCas9 son los mismos que los descritos anteriormente para SpCas9. También se conocen ejemplos de mutaciones inactivadoras en los dominios catalíticos de las proteínas Cas9 de *Staphylococcus aureus*. Por ejemplo, la enzima Cas9 de *Staphylococcus aureus* (SaCas9) puede comprender una sustitución en la posición N580 (por ejemplo, sustitución N580A) y una sustitución en la posición D10 (por ejemplo, sustitución D10A) para generar una proteína Cas inactiva a nucleasa. Véase, por ejemplo, el documento WO 2016/106236. También se conocen ejemplos de mutaciones inactivadoras en los dominios catalíticos de Nme2Cas9 (por ejemplo, combinación de D16A y H588A). También se conocen ejemplos de mutaciones inactivadoras en los dominios catalíticos de St1Cas9 (por ejemplo, combinación de D9A, D598A, H599A y N622A). También se conocen ejemplos de mutaciones inactivadoras en los dominios catalíticos de St3Cas9 (por ejemplo, combinación de D10A y N870A). También se conocen ejemplos de mutaciones inactivadoras en los dominios catalíticos de CjCas9 (por ejemplo, combinación de D8A y H559A). También se conocen ejemplos de mutaciones inactivadoras en los dominios catalíticos de FnCas9 y RHA FnCas9 (por ejemplo, N995A).

También se conocen ejemplos de mutaciones inactivadoras en los dominios catalíticos de las proteínas Cpfl. Con referencia a las proteínas Cpfl de *Francisella novicida* U112 (FnCpf1), *Acidaminococcus* sp. BV3L6 (AsCpf1), *Lachnospiraceae bacterium* ND2006 (LbCpf1), y *Moraxella bovoculi* 237 (MbCpf1 Cpfl), tales mutaciones pueden incluir mutaciones en las posiciones 908, 993 o 1263 de AsCpf1 o posiciones correspondientes en ortólogos de Cpfl, o posiciones 832, 925, 947 o 1180 de LbCpf1 o posiciones correspondientes en ortólogos de Cpfl. Dichas mutaciones pueden incluir, por ejemplo, una o más de las mutaciones D908A, E993A y D1263A de AsCpf1 o mutaciones correspondientes en ortólogos de Cpfl, o D832A, E925A, D947A y D1180A de LbCpf1 o mutaciones correspondientes en ortólogos de Cpfl. Véase, *por ejemplo*, el documento US 2016/0208243

Las proteínas Cas (por ejemplo, proteínas Cas activas para nucleasas o proteínas Cas inactivas para nucleasas) también pueden ligarse operativamente a polipéptidos heterólogos como proteínas de fusión. Por ejemplo, una proteína Cas puede fusionarse con un dominio de escisión o un dominio de modificación epigenética. Véase el documento WO 2014/089290. Las proteínas Cas también pueden fusionarse con un polipéptido heterólogo proporcionando una estabilidad aumentada o disminuida. El dominio fusionado o polipéptido heterólogo puede localizarse en el extremo N, el extremo C o internamente dentro de la proteína Cas.

Como ejemplo, una proteína Cas puede fusionarse con uno o más polipéptidos heterólogos que proporcionan localización subcelular. Dichos polipéptidos heterólogos pueden incluir, por ejemplo, una o más señales de localización nuclear (NLS) tales como la NLS SV40 monopartita y/o una NLS alfa-importina bipartita para orientarse al núcleo, una señal de localización mitocondrial para orientarse a las mitocondrias, una señal de retención ER y similares. Véase, *por ejemplo*, Lang et al. (2007) J. Biol. Chem. 282(8):5101-5105. Dichas señales de localización subcelular pueden ubicarse en el extremo N, el extremo C o en cualquier lugar dentro de la proteína Cas. Una NLS puede comprender un tramo de aminoácidos básicos y puede ser una secuencia monopartita o una secuencia bipartita. Opcionalmente, una proteína Cas puede comprender dos o más NLS, incluyendo una NLS (por ejemplo, una NLS de alfa-importina o una NLS monopartita) en el extremo N y una NLS (por ejemplo, una NLS SV40 o una NLS bipartita) en el extremo C. Una proteína Cas también puede comprender dos o más NLS en el extremo N y/o dos o más NLS en el extremo C.

Las proteínas Cas también pueden estar ligadas operativamente a un dominio de penetración celular o a un dominio de transducción de proteínas. Por ejemplo, el dominio de penetración celular puede derivarse de la proteína TAT del VIH-1, el motivo de penetración celular TLM del virus de la hepatitis B humana, MPG, Pep-1, VP22, un péptido de penetración celular del virus del herpes simple o una secuencia peptídica de poliarginina. Véase, *por ejemplo*, los documentos WO 2014/089290 y WO 2013/176772. El dominio de penetración celular puede ubicarse en el extremo N, el extremo C o en cualquier lugar dentro de la proteína Cas.

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

Las proteínas cas también pueden estar unidas a ácidos nucleicos donantes exógenos o ácidos nucleicos marcados. Tal unión (es decir, ligador físico) se puede lograr mediante interacciones covalentes o interacciones no covalentes, y la unión puede ser directa (por ejemplo, mediante fusión directa o conjugación química, que se puede lograr mediante la modificación de residuos de cisteína o lisina en la proteína o modificación de la inteína), o puede lograrse a través de uno o más conectores o moléculas adaptadoras interpuestas tales como estreptavidina o aptámeros. Véase, *por ejemplo*, Pierce et al. (2005) Mini Rev. Med. Chem. 5(1):41-55; Duckworth et al. (2007) Angew. Chem. Int. Ed. Engl. 46(46):8819-8822; Schaeffer y Dixon (2009) Australian J. Chem. 62(10): 1328-1332; Goodman et al. (2009) Chembiochem. 10(9):1551-1557; y Khatwani et al. (2012) Bioorg. Med. Chem. 20(14):4532-4539. Las estrategias no covalentes para sintetizar conjugados de proteína y ácido nucleico incluyen los métodos de biotina-estreptavidina y níquel-histidina. Los conjugados covalentes de proteína y ácido nucleico se pueden sintetizar conectando proteínas y ácidos nucleicos funcionalizados apropiadamente utilizando una amplia variedad de químicas. Algunas de estas químicas implican la unión directa del oligonucleótido a un residuo de aminoácido en la superficie de la proteína (por ejemplo, una lisina amina o una cisteína tiol), mientras que otros esquemas más complejos requieren una modificación postraduccional de la proteína o la participación de un dominio proteico catalítico o reactivo. Los métodos para la unión covalente de proteínas a ácidos nucleicos pueden incluir, por ejemplo, reticulación química de oligonucleótidos con residuos de lisina o cisteína de proteína, ligadura de proteínas expresadas, métodos quimioenzimáticos y el uso de fotoaptámeros. El ácido nucleico donante exógeno o el ácido nucleico marcado puede estar unido al extremo C, al extremo N o a una región interna dentro de la proteína Cas. En un ejemplo, el ácido nucleico donante exógeno o el

ácido nucleico marcado está unido al extremo C o al extremo N de la proteína Cas. Asimismo, la proteína Cas puede estar unida al extremo 5', al extremo 3' o a una región interna dentro del ácido nucleico donante exógeno o del ácido nucleico marcado. Es decir, el ácido nucleico donante exógeno o el ácido nucleico marcado puede unirse en cualquier orientación y polaridad. Por ejemplo, la proteína Cas puede unirse al extremo 5' o al extremo 3' del ácido nucleico donante exógeno o al ácido nucleico marcado.

Las proteínas Cas se pueden proporcionar en cualquier forma. Por ejemplo, una proteína Cas se puede proporcionar en forma de una proteína, tal como una proteína Cas formando un complejo con un ARNg. Alternativamente, se puede proporcionar una proteína Cas en forma de un ácido nucleico que codifica la proteína Cas, tal como un ARN (por ejemplo, ARN mensajero (ARNm)) o ADN. Opcionalmente, el ácido nucleico que codifica la proteína Cas se puede optimizar con codones para una traducción eficaz en proteína en una célula u organismo particular. Por ejemplo, el ácido nucleico que codifica la proteína Cas puede modificarse para sustituir codones que tengan una mayor frecuencia de uso en una célula bacteriana, una célula de levadura, una célula humana, una célula no humana, una célula de mamífero, una célula de roedor, una célula de ratón, una célula de rata o cualquier otra célula huésped de interés, en comparación con la secuencia de polinucleótidos de origen natural. Cuando se introduce en la célula un ácido nucleico que codifica la proteína Cas, la proteína Cas puede expresarse de forma transitoria, condicional o constitutiva en la célula.

Las proteínas Cas proporcionadas como ARNm pueden modificarse para mejorar la estabilidad y/o las propiedades de inmunogenicidad. Las modificaciones pueden realizarse en uno o más nucleósidos dentro del ARNm. Ejemplos de modificaciones químicas de las nucleobases de ARNm incluyen pseudouridina, 1-metil-pseudouridina y 5-metil-citidina. Por ejemplo, se puede utilizar ARNm Cas poliadenilado y protegido que contiene N1-metilpseudouridina. Asimismo, los ARNm de Cas pueden modificarse mediante el agotamiento de uridina utilizando codones sinónimos.

Los ácidos nucleicos que codifican proteínas Cas pueden integrarse de forma estable en el genoma de una célula y ligarse operativamente a un promotor activo en la célula. Alternativamente, los ácidos nucleicos que codifican proteínas Cas pueden ligarse operativamente a un promotor en un constructo de expresión. Los constructos de expresión incluyen cualquier constructo de ácido nucleico capaz de dirigir la expresión de un gen u otra secuencia de ácido nucleico de interés (por ejemplo, un gen Cas) y que puede transferir dicha secuencia de ácido nucleico de interés a una célula objetivo. Por ejemplo, el ácido nucleico que codifica la proteína Cas puede estar en un vector orientado que comprende un inserto de ácido nucleico y/o un vector que comprende un ADN que codifica un ARNg. Alternativamente, puede estar en un vector o plásmido que esté separado del vector orientado que comprende el inserto de ácido nucleico y/o separado del vector que comprende el ADN que codifica el ARNg. Los promotores que pueden usarse en un constructo de expresión incluyen promotores activos, por ejemplo, en una o más de una célula eucariota, una célula humana, una célula no humana, una célula de mamífero, una célula de mamífero no humana, una célula de roedor, una célula de ratón, una célula de rata, una célula de hámster, una célula de conejo, una célula pluripotente, una célula madre embrionaria (ES) o un cigoto. Dichos promotores pueden ser, por ejemplo, promotores condicionales, promotores inducibles, promotores constitutivos o promotores específicos de tejido. Opcionalmente, el promotor puede ser un promotor bidireccional que dirige la expresión tanto de una proteína Cas en una dirección como de un ARN guía en la otra dirección. Dichos promotores bidireccionales pueden consistir en (1) un promotor Pol III unidireccional, convencional y completo que contiene 3 elementos de control externos: un elemento de secuencia distal (DSE), un elemento de secuencia proximal (PSE) y una caja TATA; y (2) un segundo promotor Pol III básico que incluye un PSE y una caja TATA fusionados al extremo 5' del DSE en orientación inversa. Por ejemplo, en el promotor H1, el DSE es adyacente al PSE y a la caja TATA, y el promotor puede volverse bidireccional creando un promotor híbrido en donde la transcripción en la dirección inversa se controla añadiendo una caja PSE y TATA derivada del promotor de U6. Véase, *por ejemplo*, el documento US 2016/0074535. El uso de un promotor bidireccional para expresar genes que codifican una proteína Cas y un ARN guía permite simultáneamente la generación de casetes de expresión compactos para facilitar la administración.

ARN guía. Un "ARN guía" o "ARNg" es una molécula de ARN que se une a una proteína Cas (por ejemplo, proteína Cas9) y orienta la proteína Cas a una ubicación específica dentro de un ADN objetivo. Los ARN guía pueden comprender dos segmentos: un "segmento orientado al ADN" y un "segmento de unión a proteínas". "Segmento" incluye una sección o región de una molécula, tal como un tramo contiguo de nucleótidos en un ARN. Algunos ARNg, tal como los de Cas9, pueden comprender dos moléculas de ARN separadas: un "ARN activador" (por ejemplo, ARNtracr) y un "ARN objetivo" (por ejemplo, ARN CRISPR o ARNcr). Otros ARNg son una sola molécula de ARN (polinucleótido de ARN único), que también puede denominarse "ARNg de una sola molécula", "ARN de guía única" o "ARNg". Véase, *por ejemplo*, los documentos WO 2013/176772, WO 2014/065596, WO 2014/089290, WO 2014/093622, WO 2014/099750, WO 2013/142578, y WO 2014/131833. Para Cas9, por ejemplo, un ARN guía único puede comprender un ARNcr fusionado a un ARNtracr (por ejemplo, mediante un enlazador). Para Cpf1, por ejemplo, sólo se necesita un ARNcr para lograr la unión y/o la escisión de una secuencia objetivo. Los términos "ARN guía" y "ARNg" incluyen tanto ARNg de doble molécula (es decir, modulares) como ARNg de una sola molécula.

Un ARNg de dos moléculas ejemplar comprende una molécula similar a ARNcr ("ARN CRISPR" o "ARN objetivo" o "ARNcr" o "repeticón de ARNcr") y una molécula correspondiente similar a ARNtracr ("ARN CRISPR de acción trans" o "ARN activador" o "ARNtracr"). Un ARNcr comprende tanto el segmento orientado al ADN (monocatenario) del ARNg como un tramo de nucleótidos (es decir, la cola del ARNcr) que forma la mitad del dúplex de ARNbc del segmento de unión a proteínas del ARNg. Un ejemplo de una cola de ARNcr, ubicada corriente abajo (3') del segmento orientado al

ADN, comprende, consiste esencialmente en, o consiste en GUUUUAGAGCUAUGCU (SEQ ID NO: 37). Cualquiera de los segmentos orientados al ADN divulgados en el presente documento se puede unir al extremo 5' de SEQ ID NO: 37 para formar un ARNcr.

Un ARNtracr correspondiente (ARN activador) comprende un tramo de nucleótidos que forma la otra mitad del dúplex de ARNbc del segmento de unión a proteínas del ARNg. Un tramo de nucleótidos de un ARNcr es complementario y se hibridan con un tramo de nucleótidos de un ARNtracr para formar el dúplex de ARNbc del dominio de unión a proteínas del ARNg. Como tal, se puede decir que cada ARNcr tiene un ARNtracr correspondiente. Un ejemplo de una secuencia de ARNtracr comprende, consiste esencialmente en, o consiste en AGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACC GAGUCGGUGCUUU (SEQ ID NO: 38). Otros ejemplos de secuencias de ARNtracr comprenden, consisten esencialmente en o consisten en

AAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGCUUUU (SEQ ID NO: 80),

o

GUUGGAACCAUUCAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA
ACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO: 81).

En sistemas en donde se necesitan tanto un ARNcr como un ARNtracr, el ARNcr y el ARNtracr correspondiente se hibridan para formar un ARNg. En sistemas en donde sólo se necesita un ARNcr, el ARNcr puede ser el ARNg. El ARNcr proporciona además el segmento orientado al ADN monocatenario que se hibrida con la cadena complementaria de un ADN objetivo. Si se utiliza para la modificación dentro de una célula, la secuencia exacta de una determinada molécula de ARNcr o ARNtracr puede diseñarse para que sea específica de la especie en donde se utilizarán las moléculas de ARN. Véase, *por ejemplo*, Mali et al. (2013) Science 339(6121):823-826; Jinek et al. (2012) Science 337(6096):816-821; Hwang et al. (2013) Nat. Biotechnol. 31(3):227-229; Jiang et al. (2013) Nat. Biotechnol. 31(3):233-239; y Cong et al. (2013) Science 339(6121):819-823.

El segmento orientado al ADN (ARNcr) de un ARNg determinado comprende una secuencia de nucleótidos que es complementaria a una secuencia de la cadena complementaria del ADN objetivo, como se describe con más detalle a continuación. El segmento de un ARNg orientado al ADN interactúa con el ADN objetivo de una manera específica de secuencia mediante hibridación (es decir, emparejamiento de bases). Como tal, la secuencia de nucleótidos del segmento orientado al ADN puede variar y determina la ubicación dentro del ADN objetivo con la que interactuarán el ARNg y el ADN objetivo. El segmento orientado al ADN de un ARNg en cuestión se puede modificar para hibridar con cualquier secuencia deseada dentro de un ADN objetivo. Los ARNcr naturales difieren según el sistema CRISPR/Cas y el organismo, pero a menudo contienen un segmento orientado de entre 21 y 72 nucleótidos de longitud, flanqueado por dos repeticiones directas (DR) de una longitud de entre 21 y 46 nucleótidos (véase, *por ejemplo*, el documento WO 2014/131833). En el caso de *S. pyogenes*, las DR tienen 36 nucleótidos de largo y el segmento orientado tiene 30 nucleótidos de largo. La DR ubicada en 3' es complementaria y se hibrida con el ARNtracr correspondiente, que a su vez se une a la proteína Cas.

El segmento orientado al ADN puede tener, por ejemplo, una longitud de al menos aproximadamente 12, 15, 17, 18, 19, 20, 25, 30, 35 o 40 nucleótidos. Dichos segmentos orientados a ADN pueden tener, por ejemplo, una longitud de aproximadamente 12 a aproximadamente 100, de aproximadamente 12 a aproximadamente 80, de aproximadamente 12 a aproximadamente 50, de aproximadamente 12 a aproximadamente 40, de aproximadamente 12 a aproximadamente 30, de aproximadamente de 12 a aproximadamente 25, o de aproximadamente 12 a aproximadamente 20 nucleótidos. Por ejemplo, el segmento orientado al ADN puede tener de aproximadamente 15 a aproximadamente 25 nucleótidos (por ejemplo, de aproximadamente 17 a aproximadamente 20 nucleótidos, o aproximadamente 17, 18, 19 o 20 nucleótidos). Véase, *por ejemplo*, el documento US 2016/0024523. Para Cas9 de *S. pyogenes*, un segmento típico orientado al ADN tiene entre 16 y 20 nucleótidos de longitud o entre 17 y 20 nucleótidos de longitud. Para Cas9 de *S. aureus*, un segmento típico orientado al ADN tiene entre 21 y 23 nucleótidos de longitud. Para Cpf1, un segmento típico orientado al ADN tiene al menos 16 nucleótidos de longitud o al menos 18 nucleótidos de longitud.

Los ARNtracr pueden tener cualquier forma (por ejemplo, ARNtracr de longitud completa o ARNtracr parciales activos) y de longitudes variables. Pueden incluir transcripciones primarias o formularios procesados. Por ejemplo, los ARNtracr (como parte de un ARN guía único o como una molécula separada como parte de un ARNg de dos moléculas) pueden comprender, consistir esencialmente en o consistir en toda o una porción de una secuencia de ARNtracr de tipo silvestre (por ejemplo, aproximadamente o más de aproximadamente 20, 26, 32, 45, 48, 54, 63, 67, 85 o más nucleótidos de una secuencia de ARNtracr de tipo silvestre). Ejemplos de secuencias de ARNtracr de tipo silvestre de *S. pyogenes* incluyen versiones de 171 nucleótidos, 89 nucleótidos, 75 nucleótidos y 65 nucleótidos. Véase, *por*

ejemplo, Deltcheva et al. (2011) Nature 471(7340):602-607; el documento WO 2014/093661. Los ejemplos de ARNtracr dentro de ARN de guía única (ARNsg) incluyen los segmentos de ARNtracr que se encuentran en las versiones +48, +54, +67, y +85 de ARNsg, donde "+n" indica que hasta el +n nucleótido de ARNtracr de tipo silvestre está incluido en el ARNsg. Véase el documento US 8,697,359.

- 5 El porcentaje de complementariedad entre el segmento orientado al ADN del ARN guía y la cadena complementaria del ADN objetivo puede ser al menos 60% (por ejemplo, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 97%, al menos 98%, al menos 99% o 100%). El porcentaje de complementariedad entre el segmento orientado al ADN y la cadena complementaria del ADN objetivo puede ser al menos del 60 % en aproximadamente 20 nucleótidos contiguos. Como ejemplo, el porcentaje de complementariedad entre el segmento orientado al ADN y la cadena complementaria del ADN objetivo puede ser 100% sobre los 14 nucleótidos contiguos en el extremo 5' de la cadena complementaria del ADN objetivo y tan bajo como 0% sobre el resto. En tal caso, se puede considerar que el segmento orientado al ADN tiene una longitud de 14 nucleótidos. Como otro ejemplo, el porcentaje de complementariedad entre el segmento orientado al ADN y la cadena complementaria del ADN objetivo puede ser 100 % sobre los siete nucleótidos contiguos en el extremo 5' de la cadena complementaria del ADN objetivo y tan bajo como 0 % sobre el resto. En tal caso, se puede considerar que el segmento orientado al ADN tiene una longitud de 7 nucleótidos. En algunos ARN guía, al menos 17 nucleótidos dentro del segmento orientado al ADN son complementarios a la cadena complementaria del ADN objetivo. Por ejemplo, el segmento orientado al ADN puede tener una longitud de 20 nucleótidos y puede comprender 1, 2 o 3 desajustes con la cadena complementaria del ADN objetivo. En un ejemplo, los desajustes no son adyacentes a la región de la cadena complementaria correspondiente a la secuencia del motivo adyacente al protoespaciador (PAM) (es decir, el complemento inverso de la secuencia PAM) (por ejemplo, los desajustes están en el extremo 5' del segmento del ARN guía orientado al ADN, o los desajustes son al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, o 19 pares de bases de la región de la cadena complementaria correspondiente a la secuencia PAM).

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

- 30 Los ARN de guía única pueden comprender un segmento orientado al ADN unido a una secuencia de armazón (es decir, la secuencia de unión a proteínas o de unión a Cas del ARN guía). Por ejemplo, dichos ARN guía pueden tener un segmento orientado al ADN en 5' y una secuencia de armazón en 3'. Las secuencias de armazón ejemplares comprenden, consisten esencialmente en o consisten en:

GUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGA
AAAAGUGGCACCGAGUCGGUGCU (versión 1; SEQ ID NO: 39);

GUUGGAACCAUUCAAAAACAGCAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCA
ACUUGAAAAAGUGGCACCGAGUCGGUGC (versión 2; SEQ ID NO: 40);

- 35 GUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGA
AAAAGUGGCACCGAGUCGGUGC (versión 3; SEQ ID NO: 41);

y GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUU AAAUAAGGCUAGUCCGUU
AUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (versión 4; SEQ ID NO: 42). Otras secuencias de armazón ejemplares comprenden, consisten esencialmente en o consisten en:

GUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGA
AAAAGUGGCACCGAGUCGGUGCUUUUUUU (versión 5; SEQ ID NO: 82);

- 40 GUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUUGA
AAAAGUGGCACCGAGUCGGUGCUUUUU (versión 6; SEQ ID NO: 83);

o

GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUU AAAUAAGGCUAGUCCGUU
AUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (versión 7; SEQ ID NO: 84).

Los ARN guía que se orientan a cualquier secuencia objetivo de ARN guía pueden incluir, por ejemplo, un segmento orientado al ADN en el extremo 5' del ARN guía fusionado a cualquiera de las secuencias de armazón de ARN guía

ejemplares en el extremo 3' del ARN guía. Es decir, cualquiera de los segmentos orientados al ADN divulgados en el presente documento puede unirse al extremo 5' de cualquiera de las SEQ ID NOS: 39-42 para formar un único ARN guía (ARN guía quimérico). Las versiones 1, 2, 3 y 4 del ARN guía, tal como se divulgan en otra parte del presente documento, se refieren a segmentos orientados al ADN (es decir, secuencias guía o guías) unidos con las versiones 1, 2, 3 y 4 del almacén, respectivamente. Asimismo, cualquiera de los segmentos orientados al ADN divulgados en el presente documento se puede unir al extremo 5' de una cualquiera de las SEQ ID NOS: 82-84 para formar un único ARN guía (ARN guía quimérico). Las versiones 5, 6 y 7 del ARN guía, como se divulgan en otra parte del presente documento, se refieren a segmentos orientados al ADN (es decir, secuencias guía o guías) unidos con las versiones de almacén 5, 6 y 7, respectivamente.

Los ARN guía pueden incluir modificaciones o secuencias que proporcionan características deseables adicionales (por ejemplo, estabilidad modificada o regulada; orientación subcelular; seguimiento con un marcador fluorescente; un sitio de unión para una proteína o complejo proteico; y similares). Ejemplos de tales modificaciones incluyen, por ejemplo, una tapa 5' (por ejemplo, una tapa 7-metilguanilato (m7G)); una cola poliadenilada 3' (es decir, una cola poli(A) 3'); una secuencia de riboswitch (por ejemplo, para permitir la estabilidad regulada y/o la accesibilidad regulada por proteínas y/o complejos proteicos); una secuencia de control de estabilidad; una secuencia que forma un dúplex de ARNbc (es decir, una horquilla); una modificación o secuencia que se orienta al ARN a una ubicación subcelular (por ejemplo, núcleo, mitocondrias, cloroplastos y similares); una modificación o secuencia que proporciona seguimiento (por ejemplo, conjugación directa con una molécula fluorescente, conjugación con una fracción que facilita la detección fluorescente, una secuencia que permite la detección fluorescente, etc.); una modificación o secuencia que proporciona un sitio de unión para proteínas (por ejemplo, proteínas que actúan sobre el ADN, incluyendo ADN metiltransferasas, ADN desmetilasas, histonas acetiltransferasas, histonas desacetilasas y similares); y combinaciones de los mismos. Otros ejemplos de modificaciones incluyen estructuras dúplex de bucle madre modificadas, regiones abultadas diseñadas, horquillas 3' diseñadas de la estructura dúplex de bucle madre, o cualquier combinación de las mismas. Véase, *por ejemplo*, el documento US 2015/0376586 Un abultamiento puede ser una región desapareada de nucleótidos dentro del dúplex formado por la región similar a ARNcr y la región mínima similar a ARNtracr. Un abultamiento puede comprender, en un lado del dúplex, un 5'-XXX-3' desapareado donde X es cualquier purina e Y puede ser un nucleótido que puede formar un par oscilante con un nucleótido en la cadena opuesta y una región de nucleótidos desapareada en el otro lado del dúplex.

Los ácidos nucleicos no modificados pueden ser propensos a la degradación. Los ácidos nucleicos exógenos también pueden inducir una respuesta inmune innata. Las modificaciones pueden ayudar a introducir estabilidad y reducir la inmunogenicidad. Los ARN guía pueden comprender nucleósidos modificados y nucleótidos modificados que incluyen, por ejemplo, uno o más de los siguientes: (1) alteración o sustitución de uno o ambos oxígenos de fosfato no ligadores y/o de uno o más de los oxígenos de fosfato ligadores en el enlace principal de fosfodiéster; (2) alteración o sustitución de un constituyente del azúcar ribosa tal como alteración o sustitución del 2' hidroxilo del azúcar ribosa; (3) sustitución de la fracción fosfato con conectores desfosfo; (4) modificación o sustitución de una nucleobase natural; (5) sustitución o modificación del esqueleto de ribosa-fosfato; (6) modificación del extremo 3' o del extremo 5' del oligonucleótido (por ejemplo, eliminación, modificación o sustitución de un grupo fosfato terminal o conjugación de una fracción); y (7) modificación del azúcar. Otras posibles modificaciones del ARN guía incluyen modificaciones o sustitución de uracilos o tratos de poliuracilo. Véase, *por ejemplo*, los documentos WO 2015/048577 y US 2016/0237455. Se pueden realizar modificaciones similares a los ácidos nucleicos que codifican Cas, tal como los ARNm de Cas. Por ejemplo, los ARNm de Cas pueden modificarse mediante el agotamiento de uridina utilizando codones sinónimos.

Como ejemplo, los nucleótidos en el extremo 5' o 3' de un ARN guía pueden incluir enlaces fosforotioato (por ejemplo, las bases pueden tener un grupo fosfato modificado que es un grupo fosforotioato). Por ejemplo, un ARN guía puede incluir enlaces fosforotioato entre los 2, 3 o 4 nucleótidos terminales en el extremo 5' o 3' del ARN guía. Como otro ejemplo, los nucleótidos en el extremo 5' y/o 3' de un ARN guía pueden tener modificaciones 2'-O-metilo. Por ejemplo, un ARN guía puede incluir modificaciones 2'-O-metilo en los nucleótidos 2, 3 o 4 terminales en el extremo 5' y/o 3' del ARN guía (por ejemplo, el extremo 5'). Véase, *por ejemplo*, el documento WO 2017/173054 A1 y Finn et al. (2018) Cell Rep. 22(9):2227-2235. En un ejemplo específico, el ARN guía comprende análogos de 2'-O-metilo y enlaces internucleotídicos de fosforotioato 3' en los primeros tres residuos de ARN terminales 5' y 3'. En otro ejemplo específico, el ARN guía se modifica de manera que todos los grupos 2'OH que no interactúan con la proteína Cas9 se reemplazan con análogos 2'-O-metilo, y la región de la cola del ARN guía, que tiene una interacción mínima con Cas9, está modificado con enlaces internucleotídicos fosforotioato 5' y 3'. Véase, *por ejemplo*, Yin et al. (2017) Nat. Biotech. 35(12):1179-1187. Otros ejemplos de ARN guía modificados se proporcionan, por ejemplo, en el documento WO 2018/107028 A1. Estas modificaciones químicas pueden, por ejemplo, proporcionar mayor estabilidad y protección contra las exonucleasas para guiar los ARN, permitiéndoles persistir dentro de las células durante más tiempo que los ARN guía no modificados. Estas modificaciones químicas también pueden, por ejemplo, proteger contra respuestas inmunes intracelulares innatas que pueden degradar activamente el ARN o desencadenar cascadas inmunes que conducen a la muerte celular.

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

proporcionar como una molécula de ADN o como moléculas de ADN separadas que codifican el ARNcr y el ARNtracr, respectivamente.

Cuando un ARNg se proporciona en forma de ADN, el ARNg puede expresarse de forma transitoria, condicional o constitutiva en la célula. Los ADN que codifican los ARNg pueden integrarse de manera estable en el genoma de la célula y ligarse operativamente a un promotor activo en la célula. Alternativamente, los ADN que codifican los ARNg pueden ligarse operativamente a un promotor en un constructo de expresión. Por ejemplo, el ADN que codifica el ARNg puede estar en un vector que comprende un ácido nucleico heterólogo, tal como un ácido nucleico que codifica una proteína Cas. Alternativamente, puede estar en un vector o un plásmido que esté separado del vector que comprende el ácido nucleico que codifica la proteína Cas. Los promotores que pueden usarse en dichos constructos de expresión incluyen promotores activos, por ejemplo, en una o más de una célula eucariota, una célula humana, una célula no humana, una célula de mamífero, una célula de mamífero no humana, una célula de roedor, una célula de ratón, una célula de rata, una célula de hámster, una célula de conejo, una célula pluripotente, una célula madre embrionaria (ES), una célula madre adulta, una célula progenitora de desarrollo restringido, una célula madre pluripotente inducida (iPS), o una embrión en etapa unicelular. Dichos promotores pueden ser, por ejemplo, promotores condicionales, promotores inducibles, promotores constitutivos o promotores específicos de tejido. Tales promotores pueden ser también, por ejemplo, promotores bidireccionales. Los ejemplos específicos de promotores adecuados incluyen un promotor de la ARN polimerasa III, tal como un promotor U6 humano, un promotor de la polimerasa III U6 de rata o un promotor de la polimerasa III U6 de ratón.

Alternativamente, los ARNg se pueden preparar mediante otros diversos métodos. Por ejemplo, los ARNg se pueden preparar mediante transcripción *in vitro* utilizando, por ejemplo, la ARN polimerasa T7 (véase, *por ejemplo*, los documentos WO 2014/089290 y WO 2014/065596). Los ARN guía también pueden ser una molécula producida sintéticamente y preparada mediante síntesis química. Por ejemplo, se puede sintetizar químicamente un ARN guía para incluir análogos de 2'-O-metilo y enlaces internucleotídicos de fosforotioato 3' en los primeros tres residuos de ARN terminales 5' y 3'.

Los ARN guía (o ácidos nucleicos que codifican los ARN guía) pueden estar en composiciones que comprenden uno o más ARN guía (por ejemplo, 1, 2, 3, 4 o más ARN guía) y un portador que aumenta la estabilidad del ARN guía (por ejemplo, prolongando el período bajo condiciones dadas de almacenamiento (por ejemplo, -20°C, 4°C o temperatura ambiente) durante el cual los productos de degradación permanecen por debajo de un umbral, tal como por debajo del 0,5% en peso del ácido nucleico o proteína de partida; o el aumento de la estabilidad *in vivo*). Los ejemplos no limitantes de dichos vehículos incluyen microesferas de poli(ácido láctico) (PLA), microesferas de poli(ácido D,L-láctico-coglicólico) (PLGA), liposomas, micelas, micelas inversas, coqueatos lipídicos y microtúbulos lipídicos. Tales composiciones pueden comprender además una proteína Cas, tal como una proteína Cas9, o un ácido nucleico que codifica una proteína Cas.

Guía de secuencias objetivo de ARN. Los ADN objetivo para los ARN guía incluyen secuencias de ácido nucleico presentes en un ADN a las que se unirá un segmento de un ARNg orientado al ADN, siempre que existan condiciones suficientes para la unión. Las condiciones de unión de ADN/ARN adecuadas incluyen condiciones fisiológicas normalmente presentes en una célula. Otras condiciones de unión de ADN/ARN adecuadas (por ejemplo, condiciones en un sistema libre de células) se conocen en la técnica (véase, *por ejemplo*, Molecular Cloning: A Laboratory Manual, 3rd Ed. (Sambrook et al., Harbor Laboratory Press 2001)). La cadena del ADN objetivo que es complementaria y se hibrida con el ARNg puede denominarse "cadena complementaria", y la cadena del ADN objetivo que es complementaria a la "cadena complementaria" (y por lo tanto no es complementaria a la proteína Cas o ARNg) puede denominarse "cadena no complementaria" o "cadena plantilla".

El ADN objetivo incluye tanto la secuencia de la cadena complementaria con la que se hibrida el ARN guía como la secuencia correspondiente de la cadena no complementaria (por ejemplo, adyacente al motivo adyacente al protoespaciador (PAM)). El término "secuencia objetivo de ARN guía" como se usa en el presente documento se refiere específicamente a la secuencia en la cadena no complementaria correspondiente (es decir, el complemento inverso de) la secuencia con la que se hibrida el ARN guía en la cadena complementaria. Es decir, la secuencia objetivo de ARN guía se refiere a la secuencia de la cadena no complementaria adyacente al PAM (por ejemplo, corriente arriba o 5' del PAM en el caso de Cas9). Una secuencia objetivo de ARN guía es equivalente al segmento orientado al ADN de un ARN guía, pero con timinas en lugar de uracilos. Como ejemplo, una secuencia objetivo de ARN guía para una enzima SpCas9 puede hacer referencia a la secuencia corriente arriba de 5'-NGG-3' PAM en la cadena no complementaria. Un ARN guía está diseñado para tener complementariedad con la cadena complementaria de un ADN objetivo, donde la hibridación entre el segmento del ARN guía orientado al ADN y la cadena complementaria del ADN objetivo promueve la formación de un complejo CRISPR. No se requiere necesariamente una complementariedad total, siempre que haya suficiente complementariedad para provocar la hibridación y promover la formación de un complejo CRISPR. Si en el presente documento se hace referencia a un ARN guía como orientado a una secuencia objetivo de ARN guía, lo que se quiere decir es que el ARN guía se hibrida con la secuencia de cadena complementaria del ADN objetivo que es el complemento inverso de la secuencia objetivo de ARN guía en la cadena no complementaria.

Una secuencia objetivo de ADN objetivo o ARN guía puede comprender cualquier polinucleótido y puede ubicarse, por ejemplo, en el núcleo o citoplasma de una célula o dentro de un orgánulo de una célula, tal como una mitocondria o

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

- 5 La unión específica de sitio y la escisión de un ADN objetivo por una proteína Cas pueden ocurrir en ubicaciones determinadas tanto por (i) la complementariedad del emparejamiento de bases entre el ARN guía y la cadena complementaria del ADN objetivo como (ii) un motivo corto, llamado motivo adyacente protoespaciador (PAM), en la cadena no complementaria del ADN objetivo. El PAM puede flanquear la secuencia objetivo del ARN guía. Opcionalmente, la secuencia objetivo del ARN guía puede estar flanqueada en el extremo 3' por el PAM (por ejemplo, para Cas9). Alternativamente, la secuencia objetivo del ARN guía puede estar flanqueada en el extremo 5' por el PAM (por ejemplo, para Cpf1). Por ejemplo, el sitio de escisión de las proteínas Cas puede ser de aproximadamente 1 a aproximadamente 10 o de aproximadamente 2 a aproximadamente 5 pares de bases (por ejemplo, 3 pares de bases) corriente arriba o corriente abajo de la secuencia PAM (por ejemplo, dentro de la secuencia objetivo del ARN guía). En el caso de SpCas9, la secuencia PAM (es decir, en la cadena no complementaria) puede ser 5'-N₁GG-3', donde N₁ es cualquier nucleótido de ADN, y donde el PAM está inmediatamente en 3' de la secuencia objetivo del ARN guía en la cadena no complementaria del ADN objetivo. Como tal, la secuencia correspondiente al PAM en la cadena complementaria (es decir, el complemento inverso) sería 5'-CCN₂-3', donde N₂ es cualquier nucleótido de ADN y está inmediatamente en 5' de la secuencia con la que el segmento del ARN guía que se orienta al ADN se hibrida en la cadena complementaria del ADN objetivo. En algunos de estos casos, N₁ y N₂ puede ser complementario y el par base Ni-N₂ puede ser cualquier par de bases (por ejemplo, Ni=C y N₂=G; Ni=G y N₂=C; Ni=A y N₂=T; o Ni=T, y N₂=A). En el caso de Cas9 de *S. aureus*, el PAM puede ser NNGRRT o NNGRR, donde N puede ser A, G, C o T, y R puede ser G o A. En el caso de Cas9 de *C. jejuni*, el PAM puede ser, por ejemplo, NNNNACAC o NNNNRYAC, donde N puede ser A, G, C o T, y R puede ser G o A. En algunos casos (por ejemplo, para FnCpf1), la secuencia PAM puede estar corriente arriba del extremo 5' y tienen la secuencia 5'-TTN-3'.
- 25 Un ejemplo de una secuencia objetivo de ARN guía es una secuencia de ADN de 20 nucleótidos que precede inmediatamente a un motivo NGG reconocido por una proteína SpCas9. Por ejemplo, dos ejemplos de secuencias objetivo de ARN guía más PAM son GN₁₉NGG (SEQ ID NO: 43) o N₂₀NGG (SEQ ID NO: 44). Véase, *por ejemplo*, el documento WO 2014/165825. La guanina en el extremo 5' puede facilitar la transcripción mediante la ARN polimerasa en las células. Otros ejemplos de secuencias objetivo de ARN guía más PAM pueden incluir dos nucleótidos de guanina en el extremo 5' (por ejemplo, GGN₂₀NGG; SEQ ID NO: 45) para facilitar la transcripción eficiente por la polimerasa T7 *in vitro*. Véase, *por ejemplo*, el documento WO 2014/065596. Otras secuencias objetivo de ARN guía más PAM pueden tener entre 4-22 nucleótidos de longitud de las SEQ ID NOS: 43-45, incluyendo el 5' G o GG y el 3' GG o NGG. Aún otras secuencias objetivo de ARN guía, PAM, pueden tener entre 14 y 20 nucleótidos de longitud de las SEQ ID NOS: 43-45.
- 35 La formación de un complejo CRISPR hibridado con un ADN objetivo puede dar como resultado la escisión de una o ambas cadenas del ADN objetivo dentro o cerca de la región correspondiente a la secuencia objetivo del ARN guía (es decir, la secuencia objetivo del ARN guía en la cadena no complementaria del ADN objetivo y el complemento inverso en la cadena complementaria con la que se hibrida el ARN guía). Por ejemplo, el sitio de escisión puede estar dentro de la secuencia objetivo del ARN guía (por ejemplo, en una ubicación definida con respecto a la secuencia PAM). El "sitio de escisión" incluye la posición de un ADN objetivo en donde una proteína Cas produce una ruptura monocatenaria o una ruptura bicatenaria. El sitio de escisión puede estar en una sola cadena (por ejemplo, cuando se usa una nickasa) o en ambas cadenas de un ADN bicatenario. Los sitios de escisión pueden estar en la misma posición en ambas cadenas (produciendo extremos romos; por ejemplo, Cas9)) o pueden estar en diferentes sitios en cada cadena (produciendo extremos escalonados (es decir, salientes); por ejemplo, Cpf1). Se pueden producir extremos escalonados, por ejemplo, usando dos proteínas Cas, cada una de las cuales produce una ruptura monocatenaria en un sitio de escisión diferente en una cadena diferente, produciendo así una ruptura bicatenaria. Por ejemplo, una primera nickasa puede crear una ruptura monocatenaria en la primera cadena de ADN bicatenario (ADNds), y una segunda nickasa puede crear una ruptura monocatenaria en la segunda cadena de ADNds de manera que se crean secuencias salientes. En algunos casos, la secuencia objetivo del ARN guía o el sitio de escisión de la nickasa en la primera cadena está separada de la secuencia objetivo del ARN guía o el sitio de escisión de la nickasa en la segunda cadena por al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500 o 1,000 pares de bases.

(3) Ácidos nucleicos donantes exógenos orientados al gen *F12* humano

- Los métodos y composiciones divulgados en el presente documento pueden utilizar ácidos nucleicos donantes exógenos para modificar el locus *F12* humanizado después de la escisión del locus *F12* humanizado con un agente nucleasa o independientemente de la escisión del locus *F12* humanizado con un agente nucleasa. En tales métodos que utilizan un agente nucleasa, la proteína del agente nucleasa escinde el locus *F12* humanizado para crear una ruptura monocatenaria (mella) o una rotura bicatenaria, y el ácido nucleico donante exógeno recombina el locus *F12* humanizado mediante ligadura mediada por unión de extremos no homólogos (NHEJ) o mediante un evento de reparación dirigido por homología. Opcionalmente, la reparación con el ácido nucleico del donante exógeno elimina o altera la secuencia objetivo de la nucleasa de modo que el agente de nucleasa no pueda volver a orientar los alelos que se han orientado.

El ácido nucleico del donante exógeno puede orientar a cualquier secuencia en el gen *F12* humano. Algunos ácidos nucleicos donantes exógenos comprenden brazos de homología. Otros ácidos nucleicos donantes exógenos no comprenden brazos de homología. Los ácidos nucleicos donantes exógenos pueden ser capaces de insertarse en un locus *F12* humanizado mediante reparación dirigida por homología, y/o pueden ser capaces de insertarse en un locus *F12* humanizado mediante unión de extremos no homólogos.

Los ácidos nucleicos donantes exógenos pueden comprender ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN), pueden ser monocatenarios o bicatenarios y pueden estar en forma lineal o circular. Por ejemplo, un ácido nucleico donante exógeno puede ser un oligodesoxinucleótido monocatenario (ssODN). Véase, *por ejemplo*, Yoshimi et al. (2016) Nat. Commun. 7:10431. Los ácidos nucleicos donantes exógenos pueden ser ácidos nucleicos desnudos o pueden ser administrados por virus, tal como AAV. En un ejemplo específico, el ácido nucleico del donante exógeno puede administrarse mediante AAV y puede ser capaz de insertarse en un locus *F12* humanizado mediante unión de extremos no homólogos (por ejemplo, el ácido nucleico donante exógeno puede ser uno que no comprenda brazos de homología).

Un ácido nucleico donante exógeno ejemplar tiene una longitud de entre aproximadamente 50 nucleótidos y aproximadamente 5 kb, una longitud de entre aproximadamente 50 nucleótidos y aproximadamente 3 kb, o una longitud de entre aproximadamente 50 y aproximadamente 1,000 nucleótidos. Otros ácidos nucleicos donantes exógenos ejemplares tienen una longitud de entre aproximadamente 40 y aproximadamente 200 nucleótidos. Por ejemplo, un ácido nucleico donante exógeno puede estar entre aproximadamente 50-60, 60-70, 70-80, 80-90, 90-100, 100-110, 110-120, 120-130, 130-140, 140-150, 150-160, 160-170, 170-180, 180-190 o 190-200 nucleótidos de longitud. Alternativamente, un ácido nucleico donante exógeno puede estar entre aproximadamente 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900 o 900-1000 nucleótidos de longitud. Alternativamente, un ácido nucleico donante exógeno puede tener una longitud de entre aproximadamente 1-1.5, 1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, o 4.5-5 kb. Alternativamente, un ácido nucleico donante exógeno puede tener, por ejemplo, no más de 5 kb, 4.5 kb, 4 kb, 3.5 kb, 3 kb, 2.5 kb, 2 kb, 1.5 kb, 1 kb, 900 nucleótidos, 800 nucleótidos, 700 nucleótidos, 600 nucleótidos, 500 nucleótidos, 400 nucleótidos, 300 nucleótidos, 200 nucleótidos, 100 nucleótidos o 50 nucleótidos de longitud. Los ácidos nucleicos donantes exógenos (por ejemplo, vectores orientados) también pueden ser más largos.

En un ejemplo, un ácido nucleico donante exógeno es un ssODN que tiene entre aproximadamente 80 nucleótidos y aproximadamente 200 nucleótidos de longitud. En otro ejemplo, un donante de ácidos nucleicos exógeno es un ssODN que tiene entre aproximadamente 80 nucleótidos y aproximadamente 3 kb de longitud. Un ssODN de este tipo puede tener brazos de homología, por ejemplo, que tienen cada uno una longitud de entre aproximadamente 40 nucleótidos y aproximadamente 60 nucleótidos. Un ssODN de este tipo también puede tener brazos de homología, por ejemplo, que tienen cada uno una longitud de entre aproximadamente 30 nucleótidos y 100 nucleótidos. Los brazos de homología pueden ser simétricos (por ejemplo, cada uno de 40 nucleótidos o cada 60 nucleótidos de longitud) o pueden ser asimétricos (por ejemplo, un brazo de homología de 36 nucleótidos de longitud y un brazo de homología de 91 nucleótidos de longitud).

Los ácidos nucleicos donantes exógenos pueden incluir modificaciones o secuencias que proporcionan características deseables adicionales (por ejemplo, estabilidad modificada o regulada; seguimiento o detección con un marcador fluorescente; un sitio de unión para una proteína o complejo de proteínas; y así sucesivamente). Los ácidos nucleicos donantes exógenos pueden comprender uno o más marcadores fluorescentes, marcadores de purificación, marcadores de epítomos o una combinación de los mismos. Por ejemplo, un ácido nucleico donante exógeno puede comprender uno o más marcadores fluorescentes (por ejemplo, proteínas fluorescentes u otros fluoróforos o colorantes), tales como al menos 1, al menos 2, al menos 3, al menos 4 o al menos 5 marcadores fluorescentes. Los marcadores fluorescentes ejemplares incluyen fluoróforos tales como fluoresceína (por ejemplo, 6-carboxifluoresceína (6-FAM)), Texas Red, HEX, Cy3, Cy5, Cy5.5, Pacific Blue, 5-(y-6)-carboxitetrametilrodamina (TAMRA), y Cy7. Hay disponible comercialmente una amplia gama de colorantes fluorescentes para marcar oligonucleótidos (por ejemplo, de Integrated DNA Technologies). Dichos marcadores fluorescentes (por ejemplo, marcadores fluorescentes internos) se pueden usar, por ejemplo, para detectar un ácido nucleico donante exógeno que se ha integrado directamente en un ácido nucleico objetivo escindido que tiene extremos sobresalientes compatibles con los extremos del ácido nucleico donante exógeno. El marcador o etiqueta puede estar en el extremo 5', en el extremo 3' o internamente dentro del ácido nucleico donante exógeno. Por ejemplo, se puede conjugar un ácido nucleico donante exógeno en el extremo 5' con el fluoróforo IR700 de Integrated DNA Technologies (5'IRDYE®700).

Los ácidos nucleicos donantes exógenos también pueden comprender insertos de ácido nucleico que incluyen segmentos de ADN para integrarse en el locus *F12* humanizado. La integración de un inserto de ácido nucleico en un locus *F12* humanizado puede dar como resultado la adición de una secuencia de ácido nucleico de interés para el locus *F12* humanizado, la delección de una secuencia de ácido nucleico de interés en el locus *F12* humanizado o la sustitución de una secuencia de ácido nucleico de interés en el locus *F12* humanizado (es decir, delección e inserción). Algunos ácidos nucleicos donantes exógenos están diseñados para la inserción de un inserto de ácido nucleico en el locus *F12* humanizado sin ninguna delección correspondiente en el locus *F12* humanizado. Otros ácidos nucleicos donantes exógenos están diseñados para eliminar una secuencia de ácido nucleico de interés en el locus *F12* humanizado sin ninguna inserción correspondiente de un inserto de ácido nucleico. Aún se diseñan otros ácidos

nucleicos donantes exógenos para eliminar una secuencia de ácido nucleico de interés en el locus *F12* humanizado y reemplazarla con un inserto de ácido nucleico.

El inserto de ácido nucleico o el ácido nucleico correspondiente en el locus *F12* humanizado que se elimina y/o reemplaza puede tener diversas longitudes. Un inserto de ácido nucleico ejemplar o un ácido nucleico correspondiente en el locus *F12* humanizado que se elimina y/o reemplaza tiene una longitud de entre aproximadamente 1 nucleótido y aproximadamente 5 kb o tiene una longitud de entre aproximadamente 1 nucleótido y aproximadamente 1,000 nucleótidos. Por ejemplo, un inserto de ácido nucleico o un ácido nucleico correspondiente en el locus *F12* humanizado que se elimina y/o reemplaza puede estar entre aproximadamente 1-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-110, 110-120, 120-130, 130-140, 140-150, 150-160, 160-170, 170-180, 180-190 o 190-120 nucleótidos de longitud. Asimismo, un inserto de ácido nucleico o un ácido nucleico correspondiente en el locus *F12* humanizado que se elimina y/o reemplaza puede estar entre 1-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900 o 900-1000 nucleótidos de longitud. Asimismo, un inserto de ácido nucleico o un ácido nucleico correspondiente en el locus *F12* humanizado que se elimina y/o reemplaza puede tener entre aproximadamente 1-1.5, 1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, o 4.5-5 kb de longitud o más.

El inserto de ácido nucleico puede comprender una secuencia que sea homóloga u ortóloga a toda o parte de la secuencia objetivo para el reemplazo. Por ejemplo, el inserto de ácido nucleico puede comprender una secuencia que comprende una o más mutaciones puntuales (por ejemplo, 1, 2, 3, 4, 5 o más) en comparación con una secuencia orientada para reemplazo en el locus *F12* humanizado. Opcionalmente, dichas mutaciones puntuales pueden dar como resultado una sustitución conservadora de aminoácidos (por ejemplo, sustitución de ácido aspártico [Asp, D] por ácido glutámico [Glu, E]) en el polipéptido codificado.

Algunos ácidos nucleicos donantes exógenos pueden codificar una proteína exógena no codificada ni expresada por un locus *F12* endógeno de tipo silvestre (por ejemplo, puede comprender un ácido nucleico de inserto que codifica una proteína exógena).

Ácidos nucleicos donantes para inserción mediada por unión de extremos no homólogos. Algunos ácidos nucleicos donantes exógenos son capaces de insertarse en un locus *F12* humanizado mediante unión de extremos no homólogos. En algunos casos, dichos ácidos nucleicos donantes exógenos no comprenden brazos de homología. Por ejemplo, dichos ácidos nucleicos donantes exógenos se pueden insertar en una ruptura bicatenaria de extremo romo después de la escisión con un agente nucleasa. En un ejemplo específico, el ácido nucleico del donante exógeno puede administrarse mediante AAV y puede ser capaz de insertarse en un locus *F12* humanizado mediante unión de extremos no homólogos (por ejemplo, el ácido nucleico donante exógeno puede ser uno que no comprenda brazos de homología).

Otros ácidos nucleicos donantes exógenos tienen regiones monocatenarias cortas en el extremo 5' y/o en el extremo 3' que son complementarias a uno o más salientes creados por escisión mediada por nucleasa en el locus *F12* humanizado. Estos salientes también pueden denominarse brazos de homología 5' y 3'. Por ejemplo, algunos ácidos nucleicos donantes exógenos tienen regiones monocatenarias cortas en el extremo 5' y/o el extremo 3' que son complementarias a uno o más salientes creados por escisión mediada por nucleasa en las secuencias objetivo 5' y/o 3' en el locus *F12* humanizado. Algunos de tales ácidos nucleicos donantes exógenos tienen una región complementaria sólo en el extremo 5' o sólo en el extremo 3'. Por ejemplo, algunos de estos ácidos nucleicos donantes exógenos tienen una región complementaria sólo en el extremo 5' complementario a un saliente creado en una secuencia objetivo 5' en el locus *F12* humanizado o sólo en el extremo 3' complementario a un saliente creado en una secuencia objetivo 3' en el locus *F12* humanizado. Otros ácidos nucleicos donantes exógenos tienen regiones complementarias en los extremos 5' y 3'. Por ejemplo, otros ácidos nucleicos donantes exógenos tienen regiones complementarias en los extremos 5' y 3', por ejemplo, complementarias a los salientes primero y segundo, respectivamente, generados por escisión mediada por nucleasa en el locus *F12* humanizado. Por ejemplo, si el ácido nucleico donante exógeno es bicatenario, las regiones complementarias monocatenarias pueden extenderse desde el extremo 5' de la cadena superior del ácido nucleico donante y el extremo 5' de la cadena inferior del ácido nucleico donante, creando salientes 5' en cada extremo. Alternativamente, la región complementaria monocatenaria puede extenderse desde el extremo 3' de la cadena superior del ácido nucleico donante y desde el extremo 3' de la cadena inferior de la plantilla, creando salientes 3'.

Las regiones complementarias pueden tener cualquier longitud suficiente para promover la ligadura entre el ácido nucleico donante exógeno y el ácido nucleico objetivo. Las regiones complementarias ejemplares tienen entre aproximadamente 1 y aproximadamente 5 nucleótidos de longitud, entre aproximadamente 1 y aproximadamente 25 nucleótidos de longitud, o entre aproximadamente 5 y aproximadamente 150 nucleótidos de longitud. Por ejemplo, una región complementaria puede ser al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25 nucleótidos de longitud. Alternativamente, la región complementaria puede ser aproximadamente 5-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-110, 110-120, 120-130, 130-140 o 140-150 nucleótidos de longitud, o más.

Estas regiones complementarias pueden ser complementarias de los salientes creados por dos pares de nickasas. Se pueden crear dos rupturas bicatenarias con extremos escalonados utilizando nickasas primera y segunda que escinden cadenas opuestas de ADN para crear una primera ruptura bicatenaria, y nickasas tercera y cuarta que

escinden cadenas opuestas de ADN para crear una segunda ruptura de bicatenaria. Por ejemplo, se puede usar una proteína Cas para mellar la primera, segunda, tercera y cuarta secuencias objetivo de ARN guía correspondientes con el ARN guía primero, segundo, tercero y cuarto. Las secuencias objetivo de ARN guía primera y segunda se pueden posicionar para crear un primer sitio de escisión de modo que las mellas creadas por las nickasas primera y segunda en las cadenas de ADN primera y segunda creen una ruptura bicatenaria (es decir, el primer sitio de escisión comprende las mellas dentro de las secuencias objetivo de ARN guía primera y segunda). Asimismo, las secuencias objetivo de ARN guía tercera y cuarta se pueden posicionar para crear un segundo sitio de escisión de manera que las mellas creadas por las nickasas tercera y cuarta en las cadenas de ADN primera y segunda creen una ruptura bicatenaria (es decir, el segundo sitio de escisión comprende las mellas dentro de las secuencias objetivo de ARN guía tercera y cuarta). Preferiblemente, las mellas dentro de las secuencias objetivo de ARN guía primera y segunda y/o las secuencias objetivo de ARN guía tercera y cuarta pueden ser mellas desplazadas que crean salientes. La ventana de compensación puede ser, por ejemplo, al menos aproximadamente 5 pb, 10 pb, 20 pb, 30 pb, 40 pb, 50 pb, 60 pb, 70 pb, 80 pb, 90 pb, 100 pb o más. Véase Ran et al. (2013) Cell 154:1380-1389; Mali et al. (2013) Nat. Biotech. 31:833-838; y Shen et al. (2014) Nat. Methods 11:399-404. En tales casos, se puede diseñar un ácido nucleico donante exógeno bicatenario con regiones complementarias monocatenarias que son complementarias a los salientes creados por las mellas dentro de las secuencias objetivo de ARN guía primera y segunda y por las mellas dentro de las secuencias objetivo de ARN guía tercera y cuarta. Un ácido nucleico donante exógeno de este tipo puede insertarse entonces mediante ligadura mediada por unión de extremos no homólogos.

Ácidos nucleicos donantes para inserción mediante reparación orientada por homología. Algunos ácidos nucleicos donantes exógenos comprenden brazos de homología. Si el ácido nucleico donante exógeno también comprende un inserto de ácido nucleico, los brazos de homología pueden flanquear el inserto de ácido nucleico. Para facilitar la referencia, los brazos de homología se denominan en el presente documento brazos de homología 5' y 3' (es decir, corriente arriba y corriente abajo). Esta terminología se refiere a la posición relativa de los brazos de homología con respecto al inserto de ácido nucleico dentro del ácido nucleico donante exógeno. Los brazos de homología 5' y 3' corresponden a regiones dentro del locus *F12* humanizado, que se denominan en el presente documento "secuencia objetivo 5'" y "secuencia objetivo 3'", respectivamente.

Un brazo de homología y una secuencia objetivo "corresponden" o "correspondientes" entre sí cuando las dos regiones comparten un nivel suficiente de identidad de secuencia entre sí para actuar como sustratos para una reacción de recombinación homóloga. El término "homología" incluye secuencias de ADN que son idénticas o comparten identidad de secuencia con una secuencia correspondiente. La identidad de secuencia entre una secuencia objetivo determinada y el brazo de homología correspondiente encontrado en el ácido nucleico donante exógeno puede ser cualquier grado de identidad de secuencia que permita que se produzca la recombinación homóloga. Por ejemplo, la cantidad de identidad de secuencia compartida por el brazo de homología del ácido nucleico donante exógeno (o un fragmento del mismo) y la secuencia objetivo (o un fragmento de la misma) puede ser al menos 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia, de modo que las secuencias experimenten recombinación homóloga. Además, una región correspondiente de homología entre el brazo de homología y la secuencia objetivo correspondiente puede tener cualquier longitud que sea suficiente para promover la recombinación homóloga. Los brazos de homología ejemplares tienen entre aproximadamente 25 nucleótidos y aproximadamente 2.5 kb de longitud, tienen entre aproximadamente 25 nucleótidos y aproximadamente 1.5 kb de longitud, o tienen entre aproximadamente 25 y aproximadamente 500 nucleótidos de longitud. Por ejemplo, un brazo de homología dado (o cada uno de los brazos de homología) y/o la secuencia objetivo correspondiente pueden comprender regiones de homología correspondientes que están entre aproximadamente 25-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450 o 450-500 nucleótidos de longitud, tales que los brazos de homología tengan homología suficiente para sufrir recombinación homóloga con las secuencias objetivo correspondientes dentro del ácido nucleico objetivo. Alternativamente, un brazo de homología determinado (o cada brazo de homología) y/o la secuencia objetivo correspondiente pueden comprender regiones de homología correspondientes que están entre aproximadamente 0.5 kb a aproximadamente 1 kb, aproximadamente 1 kb a aproximadamente 1.5 kb, aproximadamente 1.5 kb a aproximadamente 2 kb, o aproximadamente 2 kb a aproximadamente 2.5 kb de longitud. Por ejemplo, los brazos de homología pueden tener cada uno de ellos aproximadamente 750 nucleótidos de longitud. Los brazos de homología pueden ser simétricos (cada uno aproximadamente del mismo tamaño en longitud) o pueden ser asimétricos (uno más largo que el otro).

Cuando se usa un agente de nucleasa en combinación con un ácido nucleico donante exógeno, las secuencias objetivo 5' y 3' se ubican preferiblemente en suficiente proximidad al sitio de escisión de la nucleasa (por ejemplo, dentro de suficiente proximidad a la secuencia objetivo de la nucleasa) como para promover la aparición de un evento de recombinación homóloga entre las secuencias objetivo y los brazos de homología tras una ruptura monocatenaria (mella) o una ruptura bicatenaria en el sitio de escisión de la nucleasa. El término "sitio de escisión de nucleasa" incluye una secuencia de ADN en donde un agente de nucleasa (por ejemplo, una proteína Cas9 forma un complejo con un ARN guía) crea una mella o ruptura bicatenaria. Las secuencias objetivo dentro del locus objetivo que corresponden a los brazos de homología 5' y 3' del ácido nucleico donante exógeno están "ubicadas en suficiente proximidad" a un sitio de escisión de nucleasa si la distancia es tal que promueva la aparición de un evento de recombinación homóloga entre las secuencias objetivo 5' y 3' y los brazos de homología tras una ruptura monocatenaria o una ruptura bicatenaria en el sitio de escisión de la nucleasa. Así, las secuencias objetivo correspondientes a los brazos de homología 5' y/o

3' del ácido nucleico donante exógeno pueden estar, por ejemplo, dentro de al menos 1 nucleótido de un sitio de escisión de nucleasa dado o dentro de al menos 10 nucleótidos a aproximadamente 1.000 nucleótidos de un sitio de escisión de nucleasa determinado. Como ejemplo, el sitio de escisión de la nucleasa puede estar inmediatamente adyacente a al menos una o ambas secuencias objetivo.

- 5 La relación espacial de las secuencias objetivo que corresponden a los brazos de homología del ácido nucleico donante exógeno y el sitio de escisión de la nucleasa puede variar. Por ejemplo, las secuencias objetivo pueden ubicarse en 5' con respecto al sitio de escisión de la nucleasa, las secuencias objetivo pueden ubicarse en 3' con respecto al sitio de escisión de la nucleasa, o las secuencias objetivo pueden flanquear el sitio de escisión de la nucleasa.

10 (4) Otros reactivos orientados al factor XII de coagulación humano

La actividad de cualquier otro reactivo orientado al factor XII de coagulación humano conocido o putativo también se puede evaluar utilizando el roedor divulgado en el presente documento. De manera similar, se puede detectar cualquier otra molécula para determinar la actividad orientada al factor XII de coagulación humano utilizando los roedores divulgados en el presente documento.

- 15 Como ejemplo, un reactivo orientado al factor XII de coagulación humano puede ser una proteína de unión a antígeno orientada a un epítipo de una proteína del factor XII de coagulación humano. El término "proteína de unión a antígeno" incluye cualquier proteína que se une a un antígeno. Ejemplos de proteínas de unión a antígeno incluyen un anticuerpo, un fragmento de un anticuerpo de unión a antígeno, un anticuerpo multiespecífico (por ejemplo, un anticuerpo biespecífico), un scFV, un bis-scFV, un diacuerpo, un triacuerpo, un tetracuerpo, un V-NAR, un VHH, un VL, un F(ab), un F(ab)₂, una DVD (proteína de unión a antígeno de dominio variable dual), una SVD (proteína de unión a antígeno de dominio variable único), un activador biespecífico de células T (BiTE) o un cuerpo de Davis (Patente de EE.UU. No. 8,586,713). Otros reactivos orientados al factor XII de coagulación humano incluyen moléculas pequeñas que se orientan a una proteína del factor XII de coagulación humano.

- 25 Otros reactivos orientados al factor XII de coagulación humano pueden incluir agentes de ARNi. Un "agente de ARNi" es una composición que comprende una pequeña molécula oligonucleotídica de ARN bicatenario o similar a ARN (por ejemplo, ARN químicamente modificado) capaz de facilitar la degradación o inhibición de la traducción de un ARN objetivo, tal como ARN mensajero (ARNm) de una manera específica de secuencia. El oligonucleótido en el agente de ARNi es un polímero de nucleósidos ligados, cada uno de los cuales puede modificarse o no modificarse independientemente. Los agentes de ARNi operan a través del mecanismo de interferencia del ARN (es decir, induciendo la interferencia del ARN mediante la interacción con la maquinaria de la vía de interferencia del ARN (complejo silenciador inducido por ARN o RISC) de células de mamíferos). Si bien se cree que los agentes de ARNi, tal como se usa ese término en el presente documento, operan principalmente a través del mecanismo de interferencia de ARN, los agentes de ARNi divulgados no están vinculados ni limitados a ninguna vía o mecanismo de acción en particular. Los agentes de ARNi divulgados en el presente documento comprenden una cadena sentido y una cadena antisentido, e incluyen, entre otros, ARN de interferencia cortos (ARNip), ARN bicatenario (ARNds), microARN (ARNmi), ARN en horquilla corto (ARNsh) y sustratos del cortador de cubitos. La cadena antisentido de los agentes de ARNi divulgados en el presente documento es al menos parcialmente complementaria a una secuencia (es decir, una sucesión u orden de nucleobases o nucleótidos, descrita con una sucesión de letras usando nomenclatura estándar) en el ARN objetivo.

- 40 Otros reactivos orientados al factor XII de coagulación humano pueden incluir oligonucleótidos antisentido (ASO). Los ASO monocatenarios y la interferencia de ARN (ARNi) comparten un principio fundamental en que un oligonucleótido se une a un ARN objetivo mediante el emparejamiento de bases Watson-Crick. Sin querer ceñirnos a ninguna teoría, durante la ARNi, un pequeño dúplex de ARN (agente de ARNi) se asocia con el complejo silenciador inducido por ARN (RISC), una cadena (la cadena pasajera) se pierde y la cadena restante (la cadena guía) coopera con RISC para unirse al ARN complementario. Argonaute 2 (Ago2), el componente catalítico del RISC, luego escinde el ARN objetivo. La cadena guía siempre está asociada con la cadena sentido complementaria o con una proteína (RISC). Por el contrario, un ASO debe sobrevivir y funcionar como una sola cadena. Los ASO se unen al ARN objetivo y bloquean los ribosomas u otros factores, tal como los factores de empalme, para que no se unan al ARN o recluten proteínas tal como las nucleasas. Se eligen diferentes modificaciones y regiones objetivo para los ASO con base en el mecanismo de acción deseado. Un gapmer es un oligonucleótido ASO que contiene de 2 a 5 nucleótidos modificados químicamente (por ejemplo, LNA o 2'-MOE) en cada extremo que flanquea un espacio central de ADN de 8 a 10 bases. Después de unirse al ARN objetivo, el híbrido ADN-ARN actúa como sustrato para la RNasa H.

D. Administración de reactivos orientados al factor XII de coagulación humano a roedores o células

- 55 Los métodos divulgados en el presente documento pueden comprender la introducción en un roedor o célula de diversas moléculas (por ejemplo, reactivos orientados al factor XII de coagulación humano, tales como moléculas o complejos terapéuticos), incluyendo ácidos nucleicos, proteínas, complejos de ácido nucleico-proteína, complejos de proteínas, o moléculas pequeñas. "Introducir" incluye presentar a la célula o al roedor la molécula (por ejemplo, ácido nucleico o proteína) de tal manera que obtenga acceso al interior de la célula o al interior de las células dentro del roedor. La introducción se puede lograr por cualquier medio, y dos o más de los componentes (por ejemplo, dos de

los componentes, o todos los componentes) se pueden introducir en la célula o animal no humano de manera simultánea o secuencial en cualquier combinación. Por ejemplo, se puede introducir una proteína Cas en una célula o roedor antes de la introducción de un ARN guía, o se puede introducir después de la introducción del ARN guía. Como otro ejemplo, se puede introducir un ácido nucleico donante exógeno antes de la introducción de una proteína Cas y un ARN guía, o se puede introducir después de la introducción de la proteína Cas y el ARN guía (por ejemplo, el ácido nucleico donante exógeno se puede introducir administrado aproximadamente 1, 2, 3, 4, 8, 12, 24, 36, 48 o 72 horas antes o después de la introducción de la proteína Cas y el ARN guía). Véase, *por ejemplo*, los documentos US 2015/0240263 y US 2015/0110762. Además, se pueden introducir dos o más de los componentes en la célula o en el animal no humano mediante el mismo método de administración o métodos de administración diferentes. De manera similar, dos o más de los componentes pueden introducirse en un roedor por la misma vía de administración o por diferentes vías de administración.

En algunos métodos, los componentes de un sistema CRISPR/Cas se introducen en un roedor o una célula. Se puede introducir un ARN guía en un roedor o célula en forma de ARN (por ejemplo, ARN transcrito *in vitro*) o en forma de ADN que codifica el ARN guía. Cuando se introduce en forma de ADN, el ADN que codifica un ARN guía puede ligarse operativamente a un promotor activo en una célula del roedor. Por ejemplo, se puede administrar un ARN guía a través de AAV y expresarlo *in vivo* bajo un promotor U6. Dichos ADN pueden estar en uno o más constructos de expresión. Por ejemplo, dichos constructos de expresión pueden ser componentes de una única molécula de ácido nucleico. Alternativamente, se pueden separar en cualquier combinación entre dos o más moléculas de ácido nucleico (es decir, los ADN que codifican uno o más ARN CRISPR y los ADN que codifican uno o más ARNtracr pueden ser componentes de moléculas de ácido nucleico separadas).

Asimismo, las proteínas Cas se pueden aportar en cualquier forma. Por ejemplo, una proteína Cas se puede proporcionar en forma de una proteína, tal como una proteína Cas formando un complejo con un ARNg. Alternativamente, se puede proporcionar una proteína Cas en forma de un ácido nucleico que codifica la proteína Cas, tal como un ARN (por ejemplo, ARN mensajero (ARNm)) o ADN. Opcionalmente, el ácido nucleico que codifica la proteína Cas se puede optimizar con codones para una traducción eficaz en proteína en una célula u organismo particular. Por ejemplo, el ácido nucleico que codifica la proteína Cas se puede modificar para sustituir codones que tengan una mayor frecuencia de uso en 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 de origen natural. Cuando un ácido nucleico que codifica la proteína Cas se introduce en un roedor, la proteína Cas puede expresarse de forma transitoria, condicional o constitutiva en una célula del roedor.

Los ácidos nucleicos que codifican proteínas Cas o ARN guía pueden ligarse operativamente a un promotor en un constructo de expresión. Los constructos de expresión incluyen cualquier constructo de ácido nucleico capaz de dirigir la expresión de un gen u otra secuencia de ácido nucleico de interés (por ejemplo, un gen Cas) y que puede transferir dicha secuencia de ácido nucleico de interés a una célula objetivo. Por ejemplo, el ácido nucleico que codifica la proteína Cas puede estar en un vector que comprende un ADN que codifica uno o más ARNg. Alternativamente, puede estar en un vector o plásmido que esté separado del vector que comprende el ADN que codifica uno o más ARNg. Los promotores adecuados que pueden usarse en un constructo de expresión incluyen promotores activos, por ejemplo, en una o más de una célula de roedor, una célula de ratón, una célula de rata, una célula de hámster, una célula pluripotente, una célula madre embrionaria (ES), una célula madre adulta, una célula progenitora de desarrollo restringido, una célula madre pluripotente inducida (iPS) o un embrión en etapa de una sola célula. Dichos promotores pueden ser, por ejemplo, promotores condicionales, promotores inducibles, promotores constitutivos o promotores específicos de tejido. Opcionalmente, el promotor puede ser un promotor bidireccional que dirige la expresión tanto de una proteína Cas en una dirección como de un ARN guía en la otra dirección. Dichos promotores bidireccionales pueden consistir en (1) un promotor Pol III unidireccional, convencional y completo que contiene 3 elementos de control externos: un elemento de secuencia distal (DSE), un elemento de secuencia proximal (PSE) y una caja TATA; y (2) un segundo promotor Pol III básico que incluye un PSE y una caja TATA fusionados al extremo 5' del DSE en orientación inversa. Por ejemplo, en el promotor H1, el DSE es adyacente al PSE y a la caja TATA, y el promotor puede volverse bidireccional creando un promotor híbrido en donde la transcripción en la dirección inversa se controla añadiendo una caja PSE y TATA derivada del promotor de U6. Véase, *por ejemplo*, el documento US 2016/0074535 El uso de un promotor bidireccional para expresar genes que codifican una proteína Cas y un ARN guía permite simultáneamente la generación de casetes de expresión compactos para facilitar la administración.

Las moléculas (por ejemplo, proteínas Cas o ARN guía o agentes de ARNi o ASO) introducidas en el roedor o célula se pueden proporcionar en composiciones que comprenden un vehículo que aumenta la estabilidad de las moléculas introducidas (por ejemplo, prolongando el período en condiciones dadas de almacenamiento (por ejemplo, -20°C, 4°C o temperatura ambiente) para los cuales los productos de degradación permanecen por debajo de un umbral, como por debajo del 0.5% en peso del ácido nucleico o proteína de partida; o aumentando la estabilidad *in vivo*). Los ejemplos no limitantes de dichos vehículos incluyen microesferas de poli(ácido láctico) (PLA), microesferas de poli(ácido D,L-láctico-coglicólico) (PLGA), liposomas, micelas, micelas inversas, coqueatos lipídicos y microtúbulos lipídicos.

En el presente documento se describen diversos métodos y composiciones para permitir la introducción de una molécula (por ejemplo, un ácido nucleico o una proteína) en una célula o roedor. Se conocen métodos para introducir

moléculas en diversos tipos de células e incluyen, por ejemplo, métodos de transfección estable, métodos de transfección transitoria y métodos mediados por virus.

Los protocolos de transfección, así como los protocolos para introducir moléculas en las células pueden variar. Los métodos de transfección no limitantes incluyen métodos de transfección basados en productos químicos que utilizan liposomas; nanopartículas; fosfato de calcio (Graham et al. (1973) *Virology* 52 (2): 456-67, Bacchetti et al. (1977) *Proc. Natl. Acad. Sci. EE. UU.* 74 (4): 1590-4, y Kriegler, M (1991). *Transfer and Expression: A Laboratory Manual*. Nueva York: WH Freeman y compañía. págs. 96-97); dendrímeros; o polímeros catiónicos tales como DEAE-dextrano o polietilenimina. Los métodos no químicos incluyen electroporación, sonoporación y transfección óptica. La transfección basada en partículas incluye el uso de una pistola genética o transfección asistida por imán (2006) *Current Pharmaceutical Biotechnology* 7, 277-28). También se pueden utilizar métodos virales para la transfección.

La introducción de ácidos nucleicos o proteínas en una célula también puede estar mediada por electroporación, por inyección intracitoplasmática, por infección viral, por adenovirus, por virus adenoasociados, por lentivirus, por retrovirus, por transfección, por transfección mediada por lípidos o por nucleofección. La nucleofección es una tecnología de electroporación mejorada que permite que los sustratos de ácidos nucleicos se entreguen no solo al citoplasma sino también a través de la membrana nuclear y al núcleo. Además, el uso de nucleofección en los métodos divulgados en el presente documento normalmente requiere muchas menos células que la electroporación regular (por ejemplo, sólo aproximadamente 2 millones en comparación con 7 millones mediante electroporación regular). En un ejemplo, la nucleofección se realiza utilizando sistema LONZA® NUCLEOFECTOR™.

La introducción de moléculas (por ejemplo, ácidos nucleicos o proteínas) en una célula (por ejemplo, un cigoto) también se puede lograr mediante microinyección. En cigotos (es decir, embriones en etapa unicelular), la microinyección puede realizarse en el pronúcleo materno y/o paterno o en el citoplasma. Si la microinyección es en un solo pronúcleo, es preferible el pronúcleo paterno debido a su mayor tamaño. La microinyección de un ARNm se realiza preferiblemente en el citoplasma (por ejemplo, para administrar ARNm directamente a la maquinaria de traducción), mientras que es preferible la microinyección de una proteína Cas o un polinucleótido que codifica una proteína Cas o que codifica un ARN en el núcleo/pronúcleo. Alternativamente, la microinyección se puede llevar a cabo mediante inyección tanto en el núcleo/pronúcleo como en el citoplasma: primero se puede introducir una aguja en el núcleo/pronúcleo y se puede inyectar una primera cantidad, y mientras se retira la aguja del embrión en etapa unicelular se puede inyectar una segunda cantidad en el citoplasma. Si se inyecta una proteína Cas en el citoplasma, la proteína Cas comprende preferiblemente una señal de localización nuclear para asegurar la administración al núcleo/pronúcleo. Los métodos para llevar a cabo la microinyección son bien conocidos. Véase, por ejemplo, Nagy et al. (Nagy A, Gertsenstein M, Vintersten K, Behringer R., 2003, *Manipulating the Mouse Embryo*. Cold Spring Harbor, Nueva York: Cold Spring Harbor Laboratory Press); véase también Meyer et al. (2010) *Proc. Natl. Acad. Sci. EE. UU.* 107: 15022-15026 y Meyer et al. (2012) *Proc. Natl. Acad. Sci. EE. UU.* 109:9354-9359.

Otros métodos para introducir moléculas (por ejemplo, ácido nucleico o proteínas) en una célula o animal no humano pueden incluir, por ejemplo, administración mediante vectores, administración mediada por partículas, administración mediada por exosomas, administración mediada por nanopartículas lipídicas, administración mediada por péptidos penetrantes de células o administración mediada por dispositivo implantable. Como ejemplos específicos, se puede introducir un ácido nucleico o una proteína en una célula o en un animal no humano en un vehículo tal como una microesfera de poli(ácido láctico) (PLA), un poli(ácido D,L-láctico-coglicólico) (PLGA), un liposoma, una micela, una micela inversa, un cocleato lipídico o un microtúbulo lipídico. Algunos ejemplos específicos de administración a un animal no humano incluyen administración hidrodinámica, administración mediada por virus (por ejemplo, administración mediada por virus adenoasociados (AAV)) y administración mediada por nanopartículas lipídicas.

La introducción de ácidos nucleicos y proteínas en células o animales no humanos se puede lograr mediante administración hidrodinámica (HDD). La administración hidrodinámica ha surgido como un método para la administración de ADN intracelular *in vivo*. Para la administración de genes a las células parenquimatosas, sólo es necesario inyectar secuencias de ADN esenciales a través de un vaso sanguíneo seleccionado, lo que elimina los problemas de seguridad asociados con los vectores virales y sintéticos actuales. Cuando se inyecta en el torrente sanguíneo, el ADN es capaz de llegar a las células de los diferentes tejidos accesibles a la sangre. La administración hidrodinámica emplea la fuerza generada por la inyección rápida de un gran volumen de solución en la sangre incompresible en la circulación para superar las barreras físicas del endotelio y las membranas celulares que impiden que compuestos grandes e impermeables a la membrana ingresen a las células parenquimatosas. Además de la administración de ADN, este método es útil para la administración intracelular eficiente de ARN, proteínas y otros compuestos pequeños *in vivo*. Véase, por ejemplo, Bonamassa et al. (2011) *Pharm. Res.* 28(4):694-701.

La introducción de ácidos nucleicos también se puede lograr mediante administración mediada por virus, tal como administración mediada por AAV o administración mediada por lentivirus. Otros virus/vectores virales ejemplares incluyen retrovirus, adenovirus, virus vaccinia, poxvirus y virus del herpes simple. Los virus pueden infectar células en división, células que no se dividen o células tanto en división como en no división. Los virus pueden integrarse en el genoma del huésped o, alternativamente, no integrarse en el genoma del huésped. Estos virus también pueden modificarse para que tengan una inmunidad reducida. Los virus pueden ser competentes en replicación o pueden tener replicación defectuosa (por ejemplo, defectuosos en uno o más genes necesarios para rondas adicionales de replicación y/o empaquetamiento del virión). Los virus pueden causar expresión transitoria, expresión duradera (por

ejemplo, al menos 1 semana, 2 semanas, 1 mes, 2 meses o 3 meses) o expresión permanente (por ejemplo, de Cas9 y/o ARNg). Los títulos virales ejemplares (por ejemplo, títulos de AAV) incluyen 10^{12} , 10^{13} , 10^{14} , 10^{15} y 10^{16} genomas vectoriales/ml.

El genoma de ADNss AAV consta de dos marcos de lectura abiertos, Rep y Cap, flanqueados por dos repeticiones terminales invertidas que permiten la síntesis de la cadena de ADN complementaria. Al construir un plásmido de transferencia de AAV, el transgén se coloca entre las dos ITR y se pueden suministrar Rep y Cap *in trans*. Además de Rep y Cap, AAV puede requerir un plásmido auxiliar que contenga genes de adenovirus. Estos genes (E4, E2a y VA) median la replicación de AAV. Por ejemplo, el plásmido de transferencia, Rep/Cap, y el plásmido auxiliar pueden transfectarse en células HEK293 que contienen el gen de adenovirus E1+ para producir partículas infecciosas de AAV. Alternativamente, los genes auxiliares Rep, Cap y adenovirus se pueden combinar en un único plásmido. Se pueden utilizar células y métodos de empaquetamiento similares para otros virus, tales como los retrovirus.

Se han identificado múltiples serotipos de AAV. Estos serotipos difieren en los tipos de células que infectan (es decir, su tropismo), lo que permite la transducción preferencial de tipos de células específicas. Los serotipos para el tejido del CNS incluyen AAV1, AAV2, AAV4, AAV5, AAV8 y AAV9. Los serotipos del tejido cardíaco incluyen AAV1, AAV8 y AAV9. Los serotipos del tejido renal incluyen AAV2. Los serotipos del tejido pulmonar incluyen AAV4, AAV5, AAV6 y AAV9. Los serotipos del tejido del páncreas incluyen AAV8. Los serotipos de células fotorreceptoras incluyen AAV2, AAV5 y AAV8. Los serotipos del tejido del epitelio pigmentario de la retina incluyen AAV1, AAV2, AAV4, AAV5 y AAV8. Los serotipos del tejido del músculo esquelético incluyen AAV1, AAV6, AAV7, AAV8 y AAV9. Los serotipos para el tejido hepático incluyen AAV7, AAV8 y AAV9, y particularmente AAV8.

El tropismo se puede refinar aún más mediante el pseudotipado, que es la mezcla de una cápside y un genoma de diferentes serotipos virales. Por ejemplo, AAV2/5 indica un virus que contiene el genoma del serotipo 2 empaquetado en la cápside del serotipo 5. El uso de virus pseudotipados puede mejorar la eficiencia de la transducción, así como alterar el tropismo. También se pueden utilizar cápsides híbridas derivadas de diferentes serotipos para alterar el tropismo viral. Por ejemplo, AAV-DJ contiene una cápside híbrida de ocho serotipos y muestra una alta infectividad en una amplia gama de tipos de células *in vivo*. AAV-DJ8 es otro ejemplo que muestra las propiedades de AAV-DJ, pero con una mayor absorción cerebral. Los serotipos de AAV también pueden modificarse mediante mutaciones. Ejemplos de modificaciones mutacionales de AAV2 incluyen Y444F, Y500F, Y730F y S662V. Ejemplos de modificaciones mutacionales de AAV3 incluyen Y705F, Y731F y T492V. Ejemplos de modificaciones mutacionales de AAV6 incluyen S663V y T492V. Otras variantes de AAV pseudotipadas/modificadas incluyen AAV2/1, AAV2/6, AAV2/7, AAV2/8, AAV2/9, AAV2.5, AAV8.2 y AAV/SASTG.

Para acelerar la expresión transgénica, se pueden utilizar variantes de AAV autocomplementarias (scAAV). Debido a que el AAV depende de la maquinaria de replicación del ADN de la célula para sintetizar la cadena complementaria del genoma del ADN monocatenario del AAV, la expresión del transgén puede retrasarse. Para abordar este retraso, se puede utilizar scAAV que contiene secuencias complementarias que son capaces de hibridarse espontáneamente tras la infección, eliminando la necesidad de síntesis de ADN de la célula huésped. Sin embargo, también se pueden utilizar vectores de AAV monocatenarios (ssAAV).

Para aumentar la capacidad de empaquetamiento, se pueden dividir transgenes más largos entre dos plásmidos de transferencia de AAV, el primero con un donante de empalme en 3' y el segundo con un aceptor de empalme en 5'. Tras la coinfección de una célula, estos virus forman concatémeros, se unen y se puede expresar el transgén de longitud completa. Aunque esto permite una expresión transgénica más prolongada, la expresión es menos eficiente. Métodos similares para aumentar la capacidad utilizan recombinación homóloga. Por ejemplo, un transgén se puede dividir entre dos plásmidos de transferencia, pero con una superposición sustancial de secuencias de modo que la coexpresión induce recombinación homóloga y expresión del transgén de longitud completa.

La introducción de ácidos nucleicos y proteínas también se puede lograr mediante la administración mediada por nanopartículas lipídicas (LNP). Por ejemplo, se puede utilizar la administración mediada por LNP para administrar una combinación de ARNm de Cas y ARN guía o una combinación de proteína Cas y ARN guía. La administración a través de tales métodos da como resultado una expresión transitoria de Cas y los lípidos biodegradables mejoran la eliminación, mejoran la tolerabilidad y disminuyen la inmunogenicidad. Las formulaciones de lípidos pueden proteger las moléculas biológicas de la degradación y al mismo tiempo mejorar su absorción celular. Las nanopartículas lipídicas son partículas que comprenden una pluralidad de moléculas lipídicas asociadas físicamente entre sí mediante fuerzas intermoleculares. Estos incluyen microesferas (incluyendo vesículas unilaminares y multilaminares, por ejemplo, liposomas), una fase dispersa en una emulsión, micelas o una fase interna en una suspensión. Estas nanopartículas lipídicas se pueden utilizar para encapsular uno o más ácidos nucleicos o proteínas para su administración. Las formulaciones que contienen lípidos catiónicos son útiles para administrar polianiones tales como ácidos nucleicos. Otros lípidos que se pueden incluir son lípidos neutros (es decir, lípidos zwitteriónicos o sin carga), lípidos aniónicos, lípidos auxiliares que mejoran la transfección y lípidos furtivos que aumentan el tiempo durante el cual pueden existir las nanopartículas *in vivo*. Se pueden encontrar ejemplos de lípidos catiónicos, lípidos neutros, lípidos aniónicos, lípidos auxiliares y lípidos furtivos adecuados en el documento WO 2016/010840 A1. Una nanopartícula lipídica ejemplar puede comprender un lípido catiónico y uno o más componentes diferentes. En un ejemplo, el otro componente puede comprender un lípido auxiliar tal como colesterol. En otro ejemplo, los otros componentes pueden comprender un lípido auxiliar tal como colesterol y un lípido neutro tal como DSPC. En otro ejemplo, los otros

componentes pueden comprender un lípido auxiliar tal como colesterol, un lípido neutro opcional tal como DSPC y un lípido furtivo tal como S010, S024, S027, S031 o S033.

La LNP puede contener uno, más o todos los siguientes: (i) un lípido para encapsulación y escape endosomal; (ii) un lípido neutro para la estabilización; (iii) un lípido auxiliar para la estabilización; y (iv) un lípido furtivo. Véase, *por ejemplo*, Finn et al. (2018) Cell Reports 22:1-9 y el documento WO 2017/173054 A1. En ciertas LNP, la carga puede incluir un ARN guía o un ácido nucleico que codifica un ARN guía. En ciertas LNP, la carga puede incluir un ARNm que codifica una nucleasa Cas, tal como Cas9, y un ARN guía o un ácido nucleico que codifica un ARN guía.

El lípido para la encapsulación y el escape endosómico puede ser un lípido catiónico. El lípido también puede ser un lípido biodegradable, tal como un lípido ionizable biodegradable. Un ejemplo de un lípido adecuado es el lípido A o LP01, que es (9Z,12Z)-3-((4,4-bis(octiloxi)butanoil)oxi)-2-(((3-(dietilamino)propoxi)carbonil)oxi)metil)propilo octadeca-9,12-dienoato, también llamado 3-((4,4-bis(octiloxi)butanoil)oxi)-2-(((3-(dietilamino)propoxi)carbonil)oxi)metil)propil (9Z,12Z)-octadeca-9,12-dienoato. Véase, *por ejemplo*, Finn et al. (2018) Cell Reports 22:1-9 y el documento WO 2017/173054 A1. Otro ejemplo de un lípido adecuado es el lípido B, que es ((5-((dimetilamino)metil)-1,3-fenilen)bis(oxi))bis(octano-8,1-diil)bis(decanoato), también llamado ((5-((dimetilamino)metil)-1,3-fenilen)bis(oxi))bis(octano-8,1-diil)bis(decanoato). Otro ejemplo de un lípido adecuado es el lípido C, que es 2-((4-(((3-(dimetilamino)propoxi)carbonil)oxi)hexadecanoil)oxi)propano-1,3-diil(9Z,9'Z, 12Z,12'Z)-bis(octadeca-9,12-dienoato). Otro ejemplo de un lípido adecuado es el lípido D, que es 3-octilundecanoato de 3-(((3-(dimetilamino)propoxi)carbonil)oxi)-13-(octanoiloxi)tridecilo. Otros lípidos adecuados incluyen 4-(dimetilamino)butanoato de heptatriaconta-6,9,28,31-tetraen-19-ilo (también conocido como Dlin-MC3-DMA (MC3)).

Algunos de estos lípidos adecuados para su uso en las LNP descritas en el presente documento son biodegradables *in vivo*. Por ejemplo, las LNP que comprenden dicho lípido incluyen aquellas en donde al menos 75 % del lípido se elimina del plasma en 8, 10, 12, 24 o 48 horas, o 3, 4, 5, 6, 7 o 10 días. Como otro ejemplo, al menos 50 % de la LNP se elimina del plasma en 8, 10, 12, 24 o 48 horas, o 3, 4, 5, 6, 7 o 10 días.

Dichos lípidos pueden ser ionizables dependiendo del pH del medio en donde se encuentran. Por ejemplo, en un medio ligeramente ácido, los lípidos pueden estar protonados y, por tanto, tener una carga positiva. Por el contrario, en un medio ligeramente básico, tal como, por ejemplo, la sangre, donde el pH es aproximadamente 7.35, los lípidos pueden no estar protonados y, por tanto, no tener carga. En algunas realizaciones, los lípidos pueden protonarse a un pH de al menos aproximadamente 9, 9.5 o 10. La capacidad de dicho lípido para soportar una carga está relacionada con su pKa intrínseco. Por ejemplo, el lípido puede tener, independientemente, un pKa en el intervalo de aproximadamente 5.8 a aproximadamente 6.2.

Los lípidos neutros funcionan para estabilizar y mejorar el procesamiento de las LNP. Ejemplos de lípidos neutros adecuados incluyen una variedad de lípidos neutros, sin carga o zwitteriónicos. Los ejemplos de fosfolípidos neutros adecuados para su uso en la presente divulgación incluyen, entre otros, 5-heptadecilbenceno-1,3-diol (resorcinol), dipalmitoilfosfatidilcolina (DPPC), distearoilfosfatidilcolina (DSPC), fosfocolina (DOPC), dimiristoilfosfatidilcolina (DMPC), fosfatidilcolina (PLPC), 1,2-distearoil-sn-glicero-3-fosfocolina (DAPC), fosfatidiletanolamina (PE), fosfatidilcolina de huevo (EPC), dilaurilfosfatidilcolina (DLPC), dimiristoilfosfatidilcolina (DMPC), 1-miristoil-2 - palmitoil fosfatidilcolina (MPPC), 1-palmitoil-2-miristoil fosfatidilcolina (PMPC), 1-palmitoil-2-estearoil fosfatidilcolina (PSPC), 1,2-diaraquidilol-sn-glicero-3-fosfocolina (DBPC), 1-estearoil-2-palmitoil fosfatidilcolina (SPPC), 1,2-dieicosenilol-sn-glicero-3-fosfocolina (DEPC), palmitoiloleoil fosfatidilcolina (POPC), lisofosfatidilcolina, dioleoil fosfatidiletanolamina (DOPE), dilinoleoilfosfatidilcolina distearoilfosfatidiletanolamina (DSPE), dimiristoil fosfatidiletanolamina (DMPE), dipalmitoil fosfatidiletanolamina (DPPE), palmitoiloleoil fosfatidiletanolamina (POPE), lisofosfatidiletanolamina y combinaciones de los mismos. Por ejemplo, el fosfolípido neutro se puede seleccionar del grupo que consiste en distearoilfosfatidilcolina (DSPC) y dimiristoil fosfatidiletanolamina (DMPE).

Los lípidos auxiliares incluyen lípidos que mejoran la transfección. El mecanismo por el cual el lípido auxiliar mejora la transfección puede incluir mejorar la estabilidad de las partículas. En ciertos casos, el lípido auxiliar puede mejorar la fusogenicidad de la membrana. Los lípidos auxiliares incluyen esteroides, esteroles y alquil resorcinol. Ejemplos de lípidos auxiliares adecuados incluyen colesterol, 5-heptadecilresorcinol y hemisuccinato de colesterol. En un ejemplo, el lípido auxiliar puede ser colesterol o hemisuccinato de colesterol.

Los lípidos furtivos incluyen lípidos que alteran el tiempo que las nanopartículas pueden existir *in vivo*. Los lípidos furtivos pueden ayudar en el proceso de formulación, por ejemplo, reduciendo la agregación de partículas y controlando el tamaño de las partículas. Los lípidos furtivos pueden modular las propiedades farmacocinéticas de la LNP. Los lípidos furtivos adecuados incluyen lípidos que tienen un grupo principal hidrófilo ligador a una fracción lipídica.

El grupo principal hidrófilo de lípido furtivo puede comprender, por ejemplo, una fracción polimérica seleccionada de polímeros con base en PEG (a veces denominado poli(óxido de etileno)), poli(oxazolina), poli(alcohol vinílico), poli(glicerol), poli(N-vinilpirrolidona), poliaminoácidos y poli N-(2-hidroxipropil)metacrilamida. El término PEG significa cualquier polietilenglicol u otro polímero de polialquilenéter. En determinadas formulaciones de LNP, el PEG es un PEG-2K, también denominado PEG 2000, que tiene un peso molecular medio de aproximadamente 2,000 daltons. Véase, *por ejemplo*, el documento WO 2017/173054 A1.

La fracción lipídica del lípido furtivo puede derivarse, por ejemplo, de diacilglicerol o diacilglucamida, incluyendo aquellos que comprenden un grupo dialquilglicerol o dialquilglicerol que tiene una longitud de cadena alquílica que comprende independientemente de aproximadamente C4 a aproximadamente C40 átomos de carbono saturados o insaturados, en donde la cadena puede comprender uno o más grupos funcionales tales como, por ejemplo, una amida o un éster. El grupo dialquilglicerol o dialquilglicerol puede comprender además uno o más grupos alquilo sustituidos.

Como ejemplo, el lípido furtivo puede seleccionarse de PEG-dilauroilglicerol, PEG-dimiristoilglicerol (PEG-DMG), PEG-dipalmitoilglicerol, PEG-diestearoilglicerol (PEG-DSPE), PEG-dilaurilglicerol, PEG-dimiristilglicerol, PEG-dipalmitoilglicerol y PEG-diestearoilglicerol, PEG-colesterol (1-[8'-(Colest-5-en-3[beta]-oxi)carboxamido-3',6'-dioxaoctanil]carbamoil-[omega]-metil-poli(etilenglicol), PEG-DMB (3,4-ditetradecoxibencil-[omega]-metil-poli(etilenglicol)éter), 1,2-dimiristoil-sn-glicero-3-fosfoetanolamina-N-[metoxi(polietilenglicol)-2000] (PEG2k-DMG), 1,2-diestearoil-sn-glicero-3-fosfoetanolamina-N-[metoxi(polietilenglicol)-2000] (PEG2k-DSPE), 1,2-diestearoil-sn-glicerol, metoxipolietilenglicol (PEG2k-DSG), poli(etilenglicol)-2000-dimetacrilato (PEG2k-DMA) y 1,2-diesteariloxipropil-3-amina-N-[metoxi(polietilenglicol)-2000] (PEG2k-DSA). En un ejemplo particular, el lípido furtivo puede ser PEG2k-DMG.

Las LNP pueden comprender diferentes relaciones molares respectivas de los lípidos componentes en la formulación. El % en moles del lípido CCD puede ser, por ejemplo, de aproximadamente 30 % en moles a aproximadamente 60 % en moles, de aproximadamente 35 % en moles a aproximadamente 55 % en moles, de aproximadamente 40 % en moles a aproximadamente 50 % en moles, de aproximadamente 42 % en moles a aproximadamente 47 % en moles, o aproximadamente 45 %. El % en moles del lípido auxiliar puede ser, por ejemplo, de aproximadamente 30 % en moles a aproximadamente 60 % en moles, de aproximadamente 35 % en moles a aproximadamente 55 % en moles, de aproximadamente 40 % en moles a aproximadamente 50 % en moles, de aproximadamente 41 % en moles a aproximadamente 46 % en moles, o aproximadamente 44 % en moles. El % en moles del lípido neutro puede ser, por ejemplo, de aproximadamente 1 % en moles a aproximadamente 20 % en moles, de aproximadamente 5 % en moles a aproximadamente 15 % en moles, de aproximadamente 7 % en moles a aproximadamente 12 % en moles, o aproximadamente 9 % en moles. El % en moles del lípido furtivo puede ser, por ejemplo, de aproximadamente 1 % en moles a aproximadamente 10 % en moles, de aproximadamente 1 % en moles a aproximadamente 5 % en moles, de aproximadamente 1 % en moles a aproximadamente 3 % en moles, aproximadamente 2 % en moles, o aproximadamente 1 % en moles.

Las LNP pueden tener diferentes proporciones entre los grupos amino cargados positivamente del lípido biodegradable (N) y los grupos fosfato cargados negativamente (P) del ácido nucleico que se va a encapsular. Esto puede representarse matemáticamente mediante la ecuación N/P. Por ejemplo, la relación N/P puede ser de aproximadamente 0.5 a aproximadamente 100, de aproximadamente 1 a aproximadamente 50, de aproximadamente 1 a aproximadamente 25, de aproximadamente 1 a aproximadamente 10, de aproximadamente 1 a aproximadamente 7, de aproximadamente 3 a aproximadamente 5, de aproximadamente 4 a aproximadamente 5, aproximadamente 4, aproximadamente 4.5 o aproximadamente 5. La relación N/P también puede ser de aproximadamente 4 a aproximadamente 7 o de aproximadamente 4.5 a aproximadamente 6. En ejemplos específicos, la relación N/P puede ser 4.5 o puede ser 6.

En algunas LNP, la carga puede comprender Cas ARNm y ARNg. El ARNm de Cas y los ARNg pueden estar en diferentes proporciones. Por ejemplo, la formulación de LNP puede incluir una relación de Cas ARNm a ácido nucleico de ARNg que varía de aproximadamente 25:1 a aproximadamente 1:25, que varía de aproximadamente 10:1 a aproximadamente 1:10, que varía de aproximadamente 5:1 a aproximadamente 1:5, o aproximadamente 1:1. Alternativamente, la formulación de LNP puede incluir una relación de Cas ARNm a ácido nucleico de ARNg de aproximadamente 1:1 a aproximadamente 1:5, o aproximadamente 10:1. Alternativamente, la formulación de LNP puede incluir una relación de Cas ARNm a ácido nucleico de ARNg de aproximadamente 1:10, 25:1, 10:1, 5:1, 3:1, 1:1, 1:3, 1:5, 1:10 o 1:25. Alternativamente, la formulación de LNP puede incluir una relación de Cas ARNm a ácido nucleico de ARNg de aproximadamente 1:1 a aproximadamente 1:2. En ejemplos específicos, la relación de ARNm de Cas a ARNg puede ser aproximadamente 1:1 o aproximadamente 1:2.

En algunas LNP, la carga puede comprender ácido nucleico y ARNg de donantes exógenos. El ácido nucleico del donante exógeno y los ARNg pueden estar en diferentes relaciones. Por ejemplo, la formulación de LNP puede incluir una relación de ácido nucleico donante exógeno a ácido nucleico de ARNg que varía de aproximadamente 25:1 a aproximadamente 1:25, que varía de aproximadamente 10:1 a aproximadamente 1:10, que varía de aproximadamente 5:1 a aproximadamente 1:5, o aproximadamente 1:1. Alternativamente, la formulación de LNP puede incluir una relación de ácido nucleico donante exógeno a ácido nucleico de ARNg de aproximadamente 1:1 a aproximadamente 1:5, aproximadamente 5:1 a aproximadamente 1:1, aproximadamente 10:1 o aproximadamente 1:10. Alternativamente, la formulación de LNP puede incluir una relación de ácido nucleico donante exógeno a ácido nucleico de ARNg de aproximadamente 1:10, 25:1, 10:1, 5:1, 3:1, 1:1, 1:3, 1:5, 1:10 o 1:25.

Un ejemplo específico de una LNP adecuada tiene una relación de nitrógeno a fosfato (N/P) de 4.5 y contiene lípidos catiónicos biodegradables, colesterol, DSPC y PEG2k-DMG en una relación molar de 45:44:9:2. El lípido catiónico biodegradable puede ser (9Z,12Z)-3-((4,4-bis(octiloxi)butanoil)oxi)-2-(((3-(dietilamino)propoxi)carbonil)oxi)metil)propil octadeca-9,12-dienoato, también llamado 3-((4,4-bis(octiloxi)butanoil)oxi)-2-(((3-(dietilamino)propoxi)carbonil)oxi)metil)propil (9Z,12Z)-octadeca-9,12-dienoato. Véase, por ejemplo, Finn et al.

(2018) Cell Reports 22:1-9. El ARNm de Cas9 puede estar en una relación de 1:1 en peso con respecto al ARN guía. Otro ejemplo específico de una LNP adecuada contiene Dlin-MC3-DMA (MC3), colesterol, DSPC y PEG-DMG en una relación molar de 50:38.5:10:1.5.

Otro ejemplo específico de una LNP adecuada tiene una relación de nitrógeno a fosfato (N/P) de 6 y contiene lípidos catiónicos biodegradables, colesterol, DSPC y PEG2k-DMG en una relación molar de 50:38.9:3. El lípido catiónico biodegradable puede ser (9Z,12Z)-3-((4,4-bis(octiloxi)butanoil)oxi)-2-(((3-(dietilamino)propoxi)carbonil)oxi)metil)propil octadeca-9,12-dienoato, también llamado 3-((4,4-bis(octiloxi)butanoil)oxi)-2-(((3-(dietilamino)propoxi)carbonil)oxi)metil)propil (9Z,12Z)-octadeca-9,12-dienoato. El ARNm de Cas9 puede estar en una relación de 1:2 en peso con respecto al ARN guía.

El modo de administración se puede seleccionar para disminuir la inmunogenicidad. Por ejemplo, una proteína Cas y un ARNg pueden administrarse mediante diferentes modos (por ejemplo, administración bimodal). Estos diferentes modos pueden conferir diferentes propiedades farmacodinámicas o farmacocinéticas a la molécula administrada en cuestión (por ejemplo, Cas o codificación de ácido nucleico, ARNg o codificación de ácido nucleico, o plantilla de reparación/ácido nucleico donante exógeno). Por ejemplo, los diferentes modos pueden dar como resultado una distribución tisular diferente, una vida media diferente o una distribución temporal diferente. Algunos modos de administración (por ejemplo, la administración de un vector de ácido nucleico que persiste en una célula mediante replicación autónoma o integración genómica) dan como resultado una expresión y presencia más persistente de la molécula, mientras que otros modos de administración son transitorios y menos persistentes (por ejemplo, la administración de un ARN o de una proteína). La administración de proteínas Cas de una manera más transitoria, por ejemplo, como ARNm o proteína, puede garantizar que el complejo Cas/ARNg solo esté presente y activo durante un corto período de tiempo y puede reducir la inmunogenicidad causada por los péptidos de la enzima Cas de origen bacteriano siendo exhibida en la superficie de la célula por moléculas MHC. Esta administración transitoria también puede reducir la posibilidad de modificaciones fuera del objetivo.

Administración *in vivo* puede ser por cualquier vía adecuada incluyendo, por ejemplo, parenteral, intravenosa, oral, subcutánea, intraarterial, intracraneal, intratecal, intraperitoneal, tópica, intranasal o intramuscular. Los modos sistémicos de administración incluyen, por ejemplo, vías orales y parenterales. Ejemplos de vías parenterales incluyen las vías intravenosa, intraarterial, intraósea, intramuscular, intradérmica, subcutánea, intranasal e intraperitoneal. Un ejemplo específico es la infusión intravenosa. La instilación nasal y la inyección intravítrea son otros ejemplos específicos. Los modos de administración locales incluyen, por ejemplo, intratecal, intracerebroventricular, intraparenquimatosa (por ejemplo, administración intraparenquimatosa localizada al cuerpo estriado (por ejemplo, en el caudado o en el putamen), corteza cerebral, circunvolución precentral, hipocampo (por ejemplo, en la circunvolución dentada o región CA3), corteza temporal, amígdala, corteza frontal, tálamo, cerebelo, médula, hipotálamo, tectum, tegmento o sustancia negra), vías intraocular, intraorbitaria, subconjuntival, intravítrea, subretiniana y transescleral. Cantidades significativamente más pequeñas de los componentes (en comparación con enfoques sistémicos) pueden ejercer un efecto cuando se administran localmente (por ejemplo, intraparenquimatosa o intravítrea) en comparación con cuando se administran sistémicamente (por ejemplo, por vía intravenosa). Los modos de administración locales también pueden reducir o eliminar la incidencia de efectos secundarios potencialmente tóxicos que pueden ocurrir cuando se administran sistémicamente cantidades terapéuticamente efectivas de un componente.

Administración *in vivo* puede ser por cualquier vía adecuada incluyendo, por ejemplo, parenteral, intravenosa, oral, subcutánea, intraarterial, intracraneal, intratecal, intraperitoneal, tópica, intranasal o intramuscular. Un ejemplo específico es la infusión intravenosa. Las composiciones que comprenden los ARN guía y/o las proteínas Cas (o ácidos nucleicos que codifican los ARN guía y/o las proteínas Cas) se pueden formular usando uno o más vehículos, diluyentes, excipientes o auxiliares fisiológica y farmacéuticamente aceptables. La formulación puede depender de la vía de administración elegida. El término "farmacéuticamente aceptable" significa que el vehículo, diluyente, excipiente o auxiliar es compatible con los demás ingredientes de la formulación y no es sustancialmente perjudicial para el receptor de la misma.

La frecuencia de administración y el número de dosificaciones pueden depender de la vida media de los ácidos nucleicos del donante exógeno, los ARN guía o las proteínas Cas (o los ácidos nucleicos que codifican los ARN guía o las proteínas Cas) y la vía de administración, entre otros factores. La introducción de ácidos nucleicos o proteínas en la célula o en un animal no humano se puede realizar una o varias veces durante un período de tiempo. Por ejemplo, la introducción se puede realizar 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.

E. Medición de la administración, actividad o eficacia de los reactivos orientados al factor XII de coagulación humano *in vivo* o *ex-vivo*

Los métodos divulgados en el presente documento pueden comprender además detectar o medir la actividad de reactivos orientados al factor XII de coagulación humano. Medir la actividad de dichos reactivos puede comprender, por ejemplo, medir la expresión o actividad del factor XII de coagulación o la actividad prevista. Por ejemplo, la evaluación puede comprender medir la coagulación o la generación de trombina. Alternativamente, la evaluación puede comprender medir el potencial de la actividad de señalización del factor de crecimiento de hepatocitos (HGF)-tirosina-proteína quinasa Met (Met) (por ejemplo, en neuronas), tal como evaluar la conversión de pro-HGF en HGF activo.

Como ejemplo, si el reactivo orientado al factor XII de coagulación humano es un reactivo de edición del genoma (por ejemplo, CRISPR/Cas diseñado para orientar al locus *F12* humano), la medición puede comprender evaluar el locus *F12* humanizado para detectar modificaciones.

Se pueden usar diversos métodos para identificar células que tienen una modificación genética orientada. El cribado puede comprender un ensayo cuantitativo para evaluar la modificación del alelo (MOA) de un cromosoma parental. Véase, *por ejemplo*, los documentos US 2004/0018626; US 2014/0178879; US 2016/0145646; WO 2016/081923; y Frendewey et al. (2010) *Methods Enzymol.* 476:295-307. Por ejemplo, el ensayo cuantitativo se puede llevar a cabo mediante una PCR cuantitativa como una PCR en tiempo real (qPCR). La PCR en tiempo real puede utilizar un primer conjunto de cebadores que reconoce el locus objetivo y un segundo conjunto de cebadores que reconoce un locus de referencia no objetivo. El conjunto de cebadores puede comprender una sonda fluorescente que reconoce la secuencia amplificada. Otros ejemplos de ensayos cuantitativos adecuados incluyen hibridación *in situ* mediada por fluorescencia (FISH), hibridación genómica comparativa, amplificación isotérmica de ADN, hibridación cuantitativa con sonda(s) inmovilizada(s), sondas INVADER®, sondas de baliza molecular TAQMAN® o tecnología de sonda ECLIPSE™ (véase, *por ejemplo*, el documento US 2005/0144655).

La secuenciación de próxima generación (NGS) también se puede utilizar para la detección. La secuenciación de próxima generación también puede denominarse "NGS" o "secuenciación masiva en paralelo" o "secuenciación de alto rendimiento". NGS se puede utilizar como herramienta de detección además de los ensayos MOA para definir la naturaleza exacta de la modificación genética objetivo y si es consistente entre tipos de células, tipos de tejidos o tipos de órganos.

La evaluación de la modificación del locus *F12* humanizado en un roedor puede realizarse en cualquier tipo de célula de cualquier tejido u órgano. Por ejemplo, la evaluación puede realizarse en múltiples tipos de células del mismo tejido u órgano o en células de múltiples ubicaciones dentro del tejido u órgano. Esto puede proporcionar información sobre qué tipos de células dentro de un tejido u órgano objetivo están siendo objetivo o qué secciones de un tejido u órgano están siendo alcanzadas por el reactivo orientado al factor XII de coagulación humano. Como otro ejemplo, la evaluación puede realizarse en múltiples tipos de tejido o en múltiples órganos. En los métodos en donde se orienta a un tejido, órgano o tipo de célula en particular, esto puede proporcionar información sobre la eficacia con la que se orienta a ese tejido u órgano y si existen efectos fuera del objetivo en otros tejidos u órganos.

Si el reactivo está diseñado para inactivar el locus *F12* humanizado, afectar la expresión del locus *F12* humanizado, prevenir la traducción del ARNm de *F12* humanizado o eliminar la proteína del factor XII de coagulación humanizado, la medición puede comprender evaluar la expresión del ARNm o de la proteína de *F12* humanizado. Esta medición puede realizarse dentro de tipos de células u órganos particulares (por ejemplo, el hígado o tipos de células o regiones particulares dentro del hígado, o el cerebro o tipos de células o regiones particulares con el cerebro, tal como la capa piramidal del hipocampo (neuronas piramidales) y neuronas corticales), o puede implicar la medición de los niveles séricos o plasmáticos de la proteína del factor XII de coagulación humanizado.

Un ejemplo de un ensayo que se puede utilizar son los ensayos de hibridación *in situ* (ISH) de ARN RNASCOPE™ y BASESCOPE™, que son métodos que pueden cuantificar transcripciones editadas específicas de células, incluyendo cambios de nucleótidos individuales, en el contexto de tejido fijo intacto. El ensayo ISH ARN BASESCOPE™ puede complementar NGS y qPCR en la caracterización de la edición de genes. Mientras que NGS/qPCR puede proporcionar valores promedio cuantitativos de secuencias editadas y de tipo silvestre, no proporcionan información sobre la heterogeneidad o el porcentaje de células editadas dentro de un tejido. El ensayo ISH BASESCOPE™ puede proporcionar una vista panorámica de un tejido completo y la cuantificación de transcripciones de tipo silvestre versus editadas con resolución unicelular, donde se puede cuantificar el número real de células dentro del tejido objetivo que contiene la transcripción de ARNm editada. El ensayo BASESCOPE™ logra la detección de ARN de una sola molécula utilizando sondas oligo emparejadas ("ZZ") para amplificar la señal sin antecedentes no específicos. Sin embargo, el sistema de amplificación de señal y diseño de sonda BASESCOPE™ permite la detección de ARN de una sola molécula con una sonda ZZ y puede detectar diferencialmente ediciones y mutaciones de un solo nucleótido en tejido fijo intacto.

La producción y secreción de la proteína del factor XII de coagulación humanizado se puede evaluar por cualquier medio conocido. Por ejemplo, la expresión puede evaluarse midiendo los niveles del ARNm codificado o los niveles de la proteína codificada en el roedor (por ejemplo, midiendo los niveles en plasma o suero) usando ensayos conocidos.

La actividad de la proteína del factor XII de coagulación humanizado se puede evaluar por cualquier medio conocido. Dichos ensayos podrían incluir, por ejemplo, evaluar la activación de precalicreína a calicreína, evaluar la activación de FXI, evaluar la activación de pro-HGF a HGF o evaluar la activación de la vía clásica del complemento.

IV. Métodos para crear roedores que comprendan un locus *F12* humanizado

- 5 **[0268]** Se proporcionan diversos métodos para elaborar un genoma de roedor, una célula de roedor o un roedor que comprende un locus (*F12*) del factor XII de coagulación humanizado como se divulga en otra parte del presente documento. Asimismo, se proporcionan diversos métodos para producir un gen o locus (*F12*) del factor XII de coagulación humanizado para producir un genoma de roedor o una célula de roedor que comprende un locus (*F12*) del factor XII de coagulación humanizado como se divulga en otra parte del presente documento. Cualquier método o protocolo conveniente para producir un organismo modificado genéticamente es adecuado para producir dicho roedor modificado genéticamente. Véase, por ejemplo, Poueymirou et al. (2007) Nat. Biotechnol. 25(1):91-99; documentos US 7,294,754; US 7,576,259; US 7,659,442; US 8,816,150; US 9,414,575; US 9,730,434; and US 10,039,269 (que describe las células ES de ratón y el método VELOCIMOUSE® para fabricar un ratón genéticamente modificado). Véanse también los documentos US 2014/0235933 A1, US 2014/0310828 A1, US 10,385,359, y US 10,329,582 (que describe células ES de rata y métodos para producir una rata modificada genéticamente). Véase también Cho et al. (2009) Curr. Protoc. Cell. Biol. 42:19.11.1-19.11.22 (doi: 10.1002/0471143030.cb1911s42) y Gama Sosa et al. (2010) Brain Struct. Funct. 214(2-3):91-109. Estos animales no humanos genéticamente modificados pueden generarse, por ejemplo, mediante activación genética en un locus *F12* objetivo.

- 20 Por ejemplo, el método para producir un roedor que comprende un locus *F12* humanizado puede comprender: (1) modificar el genoma de una célula pluripotente para que comprenda el locus *F12* humanizado; (2) identificar o seleccionar la célula pluripotente modificada genéticamente que comprende el locus *F12* humanizado; (3) introducir la célula pluripotente modificada genéticamente en un embrión huésped de roedor; y (4) implantar y gestar el embrión huésped en una madre sustituta. Por ejemplo, el método para producir un roedor que comprende un locus *F12* humanizado puede comprender: (1) proporcionar una célula pluripotente (es decir, una célula madre embrionaria (ES) tal como una célula ES de ratón o una célula ES de rata) que comprende el locus *F12* humanizado; (2) introducir la célula pluripotente modificada genéticamente en un embrión huésped de roedor; y (3) gestar el embrión huésped en una madre sustituta.

- 30 Como otro ejemplo, el método para producir un roedor que comprende un locus *F12* humanizado puede comprender: (1) modificar el genoma de una célula pluripotente (es decir, una célula madre embrionaria (ES) tal como una célula ES de ratón o una célula ES de rata) para que comprenda el locus *F12* humanizado; (2) identificar o seleccionar la célula pluripotente modificada genéticamente que comprende el locus *F12* humanizado; (3) introducir la célula pluripotente modificada genéticamente en un embrión huésped de roedor; y (4) gestar el embrión huésped en una madre sustituta. La célula donante se puede introducir en un embrión huésped en cualquier etapa, tal como la etapa de blastocisto o la etapa de premórula (es decir, la etapa de 4 células o la etapa de 8 células). Opcionalmente, el embrión huésped que comprende una célula pluripotente modificada (es decir, una célula ES de roedor) puede incubarse hasta la etapa de blastocisto antes de implantarse y gestarse en la madre sustituta para producir un animal no humano F0. La madre sustituta puede entonces producir un roedor de generación F0 que comprenda el locus *F12* humanizado (y capaz de transmitir la modificación genética a través de la línea germinal).

- 40 Los métodos pueden comprender además identificar una célula o animal que tenga un locus genómico objetivo modificado. Se pueden usar diversos métodos para identificar células y animales que tienen una modificación genética orientada.

- 45 Alternativamente, el método para producir los roedores descritos en otra parte del presente documento puede comprender: (1) modificar el genoma de un embrión en etapa unicelular para que comprenda el locus *F12* humanizado usando los métodos descritos anteriormente para modificar células pluripotentes; (2) seleccionar el embrión genéticamente modificado; y (3) gestar el embrión genéticamente modificado en una madre sustituta. Se genera descendencia que es capaz de transmitir la modificación genética a través de la línea germinal.

- 50 También se pueden utilizar técnicas de transferencia nuclear para generar roedores. Brevemente, los métodos para la transferencia nuclear pueden incluir los pasos de: (1) enuclear un ovocito o proporcionar un ovocito enucleado; (2) aislar o proporcionar una célula o núcleo donante para combinarlo 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 fallecidos, aunque también pueden aislarse de oviductos y/o ovarios de animales vivos. Los ovocitos pueden madurar en una variedad de medios bien conocidos antes de la enucleación. La enucleación del ovocito se puede realizar de varias maneras bien conocidas. La inserción de la célula o núcleo del donante en el ovocito enucleado para formar una célula reconstituida puede realizarse mediante microinyección de una célula del donante debajo de la zona pelúcida antes de la fusión. La fusión se puede inducir mediante la aplicación de un pulso eléctrico de DC a través del plano de contacto/fusión (electrofusión), mediante la exposición de las células a sustancias químicas que promueven la fusión, tal como el polietilenglicol, o mediante un virus inactivado, tal como el virus Sendai. Una célula reconstituida puede activarse por medios eléctricos y/o no eléctricos antes, durante y/o después de la fusión del ovocito nuclear donante y receptor. Los métodos de activación incluyen pulsos eléctricos,

choques inducidos químicamente, penetración de espermatozoides, niveles crecientes de cationes divalentes en el ovocito y reducción de la fosforilación de proteínas celulares (por ejemplo, mediante inhibidores de quinasas) en el ovocito. Las células reconstituidas activadas, o embriones, pueden cultivarse en medios bien conocidos y luego transferirse al útero de un animal. Véanse, por ejemplo, los documentos US 2008/0092249, WO 1999/005266, US 2004/0177390, WO 2008/017234, y US 7,612,250.

La célula modificada o el embrión en etapa unicelular se puede generar, por ejemplo, mediante recombinación mediante (a) introducción en la célula uno o más ácidos nucleicos donantes exógenos (por ejemplo, vectores orientados) que comprenden un ácido nucleico de inserto flanqueado, por ejemplo, por brazos de homología 5' y 3' correspondientes a sitios objetivo 5' y 3' (por ejemplo, sitios objetivo que flanquean las secuencias endógenas destinadas a la delección y sustitución con el ácido nucleico inserto), en donde el ácido nucleico de inserto comprende una secuencia de *F12* humano para generar un locus *F12* humanizado; e (b) identificar al menos una célula que comprende en su genoma el ácido nucleico de inserto integrado en el locus *F12* endógeno (es decir, identificar al menos una célula que comprende el locus *F12* humanizado). Asimismo, se puede generar un genoma de roedor modificado o un gen *F12* de roedor humanizado, por ejemplo, mediante recombinación (a) poniendo en contacto el genoma o gen con uno o más ácidos nucleicos donantes exógenos (por ejemplo, vectores orientados) que comprenden brazos de homología 5' y 3' correspondientes a los sitios objetivo 5' y 3' (por ejemplo, sitios objetivo que flanquean las secuencias endógenas destinadas a la delección y sustitución con un ácido nucleico de inserto (por ejemplo, que comprende una secuencia de *F12* humano para generar un locus *F12* humanizado) flanqueado por los brazos de homología 5' y 3'), en donde los ácidos nucleicos donantes exógenos están diseñados para la humanización del locus *F12* endógeno de roedor.

Alternativamente, la célula pluripotente modificada o el embrión en etapa unicelular se pueden generar (a) introduciendo en la célula: (i) un agente nucleasa, en donde el agente nucleasa induce una mella o ruptura bicatenaria en un sitio objetivo dentro del locus *F12* endógeno; y (ii) uno o más ácidos nucleicos donantes exógenos (por ejemplo, vectores orientados) que comprenden un ácido nucleico de inserto flanqueado por, por ejemplo, brazos de homología 5' y 3' correspondientes a sitios objetivo 5' y 3' (por ejemplo, sitios objetivo que flanquean las secuencias endógenas destinadas a la delección y sustitución con el ácido nucleico de inserto), en donde el ácido nucleico de inserto comprende una secuencia de *F12* humano para generar un locus *F12* humanizado; y (c) identificar al menos una célula que comprende en su genoma el ácido nucleico de inserto integrado en el locus *F12* endógeno (es decir, identificar al menos una célula que comprende el locus *F12* humanizado). Asimismo, se puede generar un genoma de roedor modificado o un gen *F12* de roedor humanizado poniendo en contacto el genoma o gen con: (i) un agente nucleasa, en donde el agente nucleasa induce una mella o ruptura bicatenaria en un sitio objetivo dentro del locus o gen *F12* endógeno; y (ii) uno o más ácidos nucleicos donantes exógenos (por ejemplo, vectores orientados) que comprenden un ácido nucleico de inserto (por ejemplo, que comprende una secuencia de *F12* humano para generar un locus *F12* humanizado) flanqueado por, por ejemplo, brazos de homología 5' y 3' que corresponden a sitios objetivo 5' y 3' (por ejemplo, sitios objetivo que flanquean las secuencias endógenas destinadas a la delección y sustitución con el ácido nucleico de inserto), en donde los ácidos nucleicos donantes exógenos están diseñados para la humanización del locus *F12* endógeno. Puede usarse cualquier agente nucleasa que induzca una mella o ruptura bicatenaria en un sitio de reconocimiento deseado. Ejemplos de nucleasas adecuadas incluyen una nucleasa efectora similar a un activador de la transcripción (TALEN), una nucleasa con dedos de zinc (ZFN), una meganucleasa y sistemas de repeticiones palindrómicas cortas agrupadas regularmente intercaladas (CRISPR)/asociadas a CRISPR (Cas) (por ejemplo, sistemas CRISPR/Cas9) o componentes de dichos sistemas (por ejemplo, CRISPR/Cas9). Véase, por ejemplo, los documentos US 2013/0309670 y US 2015/0159175. En un ejemplo, la nucleasa comprende una proteína Cas9 y un ARN guía. En otro ejemplo, la nucleasa comprende una proteína Cas9 y dos o más, tres o más, o cuatro o más ARN guía.

El paso de modificar el genoma puede, por ejemplo, utilizar plantillas de reparación exógenas o ácidos nucleicos donantes (por ejemplo, vectores orientados) para modificar un locus *F12* para que comprenda un locus *F12* humanizado divulgado en el presente documento. Como ejemplo, el vector orientado puede ser para generar un gen *F12* humanizado en un locus *F12* endógeno (por ejemplo, locus *F12* de roedor endógeno), en donde el vector orientado comprende un inserto de ácido nucleico que comprende la secuencia de *F12* humano para integrarse en el locus *F12* flanqueado por un brazo de homología 5' orientado a una secuencia objetivo 5' en el locus *F12* endógeno y un brazo de homología 3' orientado a una secuencia objetivo 3' en el locus *F12* endógeno. La integración de un inserto de ácido nucleico en el locus *F12* puede resultar en la adición de una secuencia de ácido nucleico de interés en el locus *F12*, la delección de una secuencia de ácido nucleico de interés en el locus *F12* o la sustitución de una secuencia de ácido nucleico de interés en el locus *F12* (es decir, delección e inserción). Los brazos de homología pueden flanquear un ácido nucleico de inserto que comprende la secuencia de *F12* humano para generar el locus *F12* humanizado (por ejemplo, para eliminar un segmento del locus *F12* endógeno y reemplazarlo con una secuencia de *F12* humano ortólogo).

Las plantillas de reparación exógenas o los ácidos nucleicos donantes pueden ser para inserción mediada por unión de extremos no homólogos o para recombinación homóloga. Las plantillas de reparación exógenas o ácidos nucleicos donantes pueden comprender ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN), pueden ser monocatenarios o bicatenarios y pueden estar en forma lineal o circular. Por ejemplo, una plantilla de reparación puede ser un oligodesoxinucleótido monocatenario (ssODN).

Los ácidos nucleicos donantes exógenos también pueden comprender una secuencia heteróloga que no está presente en un locus *F12* endógeno no orientado. Por ejemplo, un donante de ácidos nucleicos exógeno puede comprender un casete de selección, tal como un casete de selección flanqueado por sitios de reconocimiento de recombinasa.

Algunos ácidos nucleicos donantes exógenos comprenden brazos de homología. Si el ácido nucleico donante exógeno también comprende un inserto de ácido nucleico, los brazos de homología pueden flanquear el inserto de ácido nucleico. Para facilitar la referencia, los brazos de homología se denominan en el presente documento brazos de homología 5' y 3' (es decir, corriente arriba y corriente abajo). Esta terminología se refiere a la posición relativa de los brazos de homología con respecto al inserto de ácido nucleico dentro del ácido nucleico donante exógeno. Los brazos de homología 5' y 3' corresponden a regiones dentro del locus *F12*, que se denominan en el presente documento "secuencia objetivo 5'" y "secuencia objetivo 3'", respectivamente.

Un brazo de homología y una secuencia objetivo "corresponden" o "correspondientes" entre sí cuando las dos regiones comparten un nivel suficiente de identidad de secuencia entre sí para actuar como sustratos para una reacción de recombinación homóloga. El término "homología" incluye secuencias de ADN que son idénticas o comparten identidad de secuencia con una secuencia correspondiente. La identidad de secuencia entre una secuencia objetivo determinada y el brazo de homología correspondiente encontrado en el ácido nucleico donante exógeno puede ser cualquier grado de identidad de secuencia que permita que se produzca la recombinación homóloga. Por ejemplo, la cantidad de identidad de secuencia compartida por el brazo de homología del ácido nucleico donante exógeno (o un fragmento del mismo) y la secuencia objetivo (o un fragmento de la misma) puede ser al menos 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia, de modo que las secuencias experimenten recombinación homóloga. Además, una región correspondiente de homología entre el brazo de homología y la secuencia objetivo correspondiente puede tener cualquier longitud que sea suficiente para promover la recombinación homóloga. En algunos vectores orientados, la mutación prevista en el locus *F12* endógeno se incluye en un ácido nucleico de inserto flanqueado por los brazos de homología.

En células distintas de los embriones en etapa unicelular, el ácido nucleico del donante exógeno puede ser un "vector orientado grande" o "LTVEC", que incluye vectores orientados que comprenden brazos de homología que corresponden y se derivan de secuencias de ácido nucleico más grandes que las típicamente utilizadas por otros enfoques destinados a realizar recombinación homóloga en células. Véase, *por ejemplo*, el documento US 2004/0018626; WO 2013/163394; US 9,834,786; US 10,301,646; WO 2015/088643; US 9,228,208; US 9,546,384; US 10,208,317; y US 2019-0112619. Los LTVEC también incluyen vectores orientados que comprenden insertos de ácido nucleico que tienen secuencias de ácido nucleico más grandes que las utilizadas normalmente por otros enfoques destinados a realizar recombinación homóloga en células. Por ejemplo, los LTVEC hacen posible la modificación de loci grandes que no pueden adaptarse a los vectores orientados tradicionales basados en plásmidos debido a sus limitaciones de tamaño. Por ejemplo, el locus objetivo puede ser (es decir, los brazos de homología 5' y 3' pueden corresponder a) un locus de la célula que no puede ser objetivo usando un método convencional o que puede ser objetivo sólo incorrectamente o sólo con una eficiencia significativamente baja en la ausencia de una mella o ruptura bicatenaria inducida por un agente nucleasa (por ejemplo, una proteína Cas). Los LTVEC pueden tener cualquier longitud y normalmente tienen al menos 10 kb de longitud. La suma total del brazo de homología 5' y el brazo de homología 3' en un LTVEC suele ser de al menos 10 kb. Se describen la generación y uso de grandes vectores orientados (LTVEC) derivados de ADN de cromosoma artificial bacteriano (BAC) mediante reacciones de recombinación homóloga bacteriana (BHR) utilizando la tecnología de ingeniería genética VELOCIGENE®, por ejemplo, en el documento US 6,586,251 y Valenzuela et al. (2003) Nat. Biotechnol. 21(6):652-659. Se describe la generación de LTVEC mediante métodos de ensamblaje *in vitro*, por ejemplo, en los documentos US 2015/0376628 y WO 2015/200334.

Los métodos pueden comprender además identificar una célula o animal que tenga un locus genómico objetivo modificado. Se pueden usar diversos métodos para identificar células y animales que tienen una modificación genética orientada. El paso de selección puede comprender, por ejemplo, un ensayo cuantitativo para evaluar la modificación del alelo (MOA) de un cromosoma parental. Véanse, *por ejemplo*, los documentos US 2004/0018626; US 2014/0178879; US 2016/0145646; WO 2016/081923; y Frendewey et al. (2010) Methods Enzymol. 476:295-307. Por ejemplo, el ensayo cuantitativo se puede llevar a cabo mediante una PCR cuantitativa como una PCR en tiempo real (qPCR). La PCR en tiempo real puede utilizar un primer conjunto de cebadores que reconoce el locus objetivo y un segundo conjunto de cebadores que reconoce un locus de referencia no objetivo. El conjunto de cebadores puede comprender una sonda fluorescente que reconoce la secuencia amplificada.

Otros ejemplos de ensayos cuantitativos adecuados incluyen hibridación *in situ* mediada por fluorescencia (FISH), hibridación genómica comparativa, amplificación isotérmica de ADN, hibridación cuantitativa con sonda(s) inmovilizada(s), sondas INVADER®, sondas de baliza molecular TAQMAN® o tecnología de sonda ECLIPSE™ (véase, por ejemplo, el documento US 2005/0144655).

Una célula pluripotente adecuada es una célula madre embrionaria (ES) de roedor (por ejemplo, una célula ES de ratón o una célula ES de rata). La célula pluripotente modificada puede generarse, por ejemplo, mediante recombinación (a) introduciendo en la célula uno o más ácidos nucleicos donantes exógenos (por ejemplo, vectores orientados) que comprenden un ácido nucleico de inserto flanqueado, por ejemplo, por brazos de homología 5' y 3'

que corresponden a sitios objetivo 5' y 3', en donde el ácido nucleico de inserto comprende una secuencia de *F12* humano para generar un locus *F12* humanizado; y (b) identificar al menos una célula que comprende en su genoma el ácido nucleico de inserto integrado en el locus *F12* endógeno (es decir, identificar al menos una célula que comprende el locus *F12* humanizado). La célula pluripotente modificada puede generarse, por ejemplo, mediante recombinação (a) introduciendo en la célula uno o más vectores orientados que comprenden un ácido nucleico de inserto flanqueado por brazos de homología 5' y 3' que corresponden a sitios objetivo 5' y 3', en donde el ácido nucleico de inserto comprende un locus *F12* humanizado; y (b) identificar al menos una célula que comprende en su genoma el ácido nucleico de inserto integrado en el locus genómico objetivo.

Alternativamente, la célula pluripotente modificada puede generarse (a) introduciendo en la célula: (i) un agente nucleasa, en donde el agente nucleasa induce una mella o ruptura bicatenaria en un sitio objetivo dentro del locus *F12* endógeno; y (ii) uno o más ácidos nucleicos donantes exógenos (por ejemplo, vectores orientados) que comprenden opcionalmente un ácido nucleico de inserto flanqueado por, por ejemplo, brazos de homología 5' y 3' correspondientes a sitios objetivo 5' y 3' ubicados en suficiente proximidad al sitio objetivo de nucleasa, en donde el ácido nucleico de inserto comprende una secuencia de *F12* humano para generar un locus *F12* humanizado; y (c) identificar al menos una célula que comprende en su genoma el ácido nucleico de inserto integrado en el locus *F12* endógeno (es decir, identificar al menos una célula que comprende el locus *F12* humanizado). Alternativamente, la célula pluripotente modificada puede generarse (a) introduciendo en la célula: (i) un agente nucleasa, en donde el agente nucleasa induce una mella o ruptura bicatenaria en un sitio de reconocimiento dentro del locus genómico objetivo; y (ii) uno o más vectores orientados que comprenden un ácido nucleico de inserto flanqueado por brazos de homología 5' y 3' que corresponden a sitios objetivo 5' y 3' ubicados en suficiente proximidad al sitio de reconocimiento, en donde el ácido nucleico de inserto comprende el locus *F12* humanizado; y (c) identificar al menos una célula que comprende una modificación (por ejemplo, integración del ácido nucleico de inserto) en el locus genómico objetivo. Puede usarse cualquier agente nucleasa que induzca una mella o ruptura bicatenaria en un sitio de reconocimiento deseado. Ejemplos de nucleasas adecuadas incluyen una nucleasa efectora similar a un activador de la transcripción (TALEN), una nucleasa con dedos de zinc (ZFN), una meganucleasa y sistemas de repeticiones palindrómicas cortas agrupadas regularmente intercaladas (CRISPR)/asociadas a CRISPR (Cas)(por ejemplo, sistemas CRISPR/Cas9) o componentes de dichos sistemas (por ejemplo, CRISPR/Cas9). Véase, *por ejemplo*, el documento US 2013/0309670 y US 2015/0159175.

La célula donante puede introducirse en un embrión huésped en cualquier etapa, tal como la etapa de blastocisto o la etapa de premórula (es decir, la etapa de 4 células o la etapa de 8 células). Se genera descendencia que es capaz de transmitir la modificación genética a través de la línea germinal. Véase, *por ejemplo*, Patente de EE. UU. No. 7,294,754.

Alternativamente, el método para producir los roedores descritos en otra parte del presente documento puede comprender: (1) modificar el genoma de un embrión de roedor en etapa unicelular para que comprenda el locus *F12* humanizado usando los métodos descritos anteriormente para modificar células pluripotentes; (2) seleccionar el embrión de roedor modificado genéticamente; y (3) implantar y gestar el roedor genéticamente modificado en una madre roedor sustituta. Alternativamente, el método para producir los roedores descritos en otra parte del presente documento puede comprender: (1) modificar el genoma de un embrión de roedor en etapa unicelular para que comprenda el locus *F12* humanizado usando los métodos descritos anteriormente para modificar células pluripotentes; (2) seleccionar el embrión de roedor modificado genéticamente; y (3) gestar el embrión de roedor genéticamente modificado en una madre roedor sustituta. Se genera descendencia que es capaz de transmitir la modificación genética a través de la línea germinal.

También se pueden utilizar técnicas de transferencia nuclear para generar animales mamíferos no humanos. Brevemente, los métodos para la transferencia nuclear pueden incluir los pasos de: (1) enucleación de un ovocito o proporcionar un ovocito enucleado; (2) aislar o proporcionar una célula o núcleo donante para combinarlo 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 fallecidos, aunque también pueden aislarse de oviductos y/o ovarios de animales vivos. Los ovocitos pueden madurar en una variedad de medios bien conocidos antes de la enucleación. La enucleación del ovocito se puede realizar de varias maneras bien conocidas. La inserción de la célula o núcleo del donante en el ovocito enucleado para formar una célula reconstituida puede realizarse mediante microinyección de una célula del donante debajo de la zona pelúcida antes de la fusión. La fusión se puede inducir mediante la aplicación de un pulso eléctrico de DC a través del plano de contacto/fusión (electrofusión), mediante la exposición de las células a sustancias químicas que promueven la fusión, tal como el polietilenglicol, o mediante un virus inactivado, tal como el virus Sendai. Una célula reconstituida puede activarse por medios eléctricos y/o no eléctricos antes, durante y/o después de la fusión del ovocito nuclear donante y receptor. Los métodos de activación incluyen pulsos eléctricos, choques inducidos químicamente, penetración de espermatozoides, niveles crecientes de cationes divalentes en el ovocito y reducción de la fosforilación de proteínas celulares (por ejemplo, mediante inhibidores de quinasas) en el ovocito. Las células reconstituidas activadas, o embriones, pueden cultivarse en medios bien conocidos y luego transferirse al útero de un animal. Véanse, *por ejemplo*, los documentos US 2008/0092249, WO 1999/005266, US 2004/0177390, WO 2008/017234, y la Patente de EE. UU. No. 7,612,250.

Los diversos métodos proporcionados en el presente documento permiten la generación de un roedor F0 genéticamente modificado en donde las células del roedor F0 genéticamente modificado comprenden el locus *F12*

humanizado. Se reconoce que dependiendo del método utilizado para generar el roedor F0, variará el número de células dentro del roedor F0 que tienen el locus *F12* humanizado. Con ratones, por ejemplo, la introducción de las células ES del donante en un embrión en etapa premórula de un organismo correspondiente (por ejemplo, un embrión de ratón en etapa de 8 células) a través de, por ejemplo, el método VELOCIMOUSE® permite que un mayor porcentaje de la población celular del ratón F0 comprenda células que tengan la secuencia de nucleótidos de interés que comprende la modificación genética orientada. Por ejemplo, al menos 50 %, 60 %, 65 %, 70 %, 75 %, 85 %, 86 %, 87 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de la contribución celular del animal F0 no humano puede comprender una población celular que tiene la modificación objetivo.

Las células del roedor F0 genéticamente modificado pueden ser heterocigotas para el locus *F12* humanizado o pueden ser homocigotas para el locus *F12* humanizado.

V. Métodos para evaluar o utilizar la isoforma corta *F12*

Se proporcionan diversos métodos para evaluar o utilizar la isoforma corta del factor XII de coagulación expresada en células neuronales humanas. Como se describe en el Ejemplo 3, la isoforma corta FXII puede codificarse mediante un ARNm de *F12* detectado mediante sondas contra los exones 11-14 pero no mediante sondas contra los exones 1-6. En un ejemplo específico, la isoforma corta puede ser FXII₂₉₇₋₅₉₆, (SEQ ID NO: 74, codificado por SEQ ID NO: 75) una proteína que contiene los dominios catalíticos y ricos en prolina de FXII (M297-S596).

La evaluación puede comprender, por ejemplo, medir la expresión de la isoforma corta o medir la actividad de la isoforma corta en una muestra biológica. La expresión de la isoforma corta se puede medir a nivel de ARNm usando cebadores y/o sondas (por ejemplo, para PCR cuantitativa) contra el exón 9, 10, 11, 12, 13 o 14, tales como sondas contra los exones 11-14. Esto se puede hacer en comparación con las sondas contra los exones 1, 2, 3, 4, 5 o 6, que no detectarán la isoforma corta. Alternativamente, la expresión de la isoforma corta se puede medir a nivel de ARNm usando cebadores y/o sondas (por ejemplo, para PCR cuantitativa) contra el exón 7, 8, 9, 10, 11, 12, 13 o 14, tales como sondas contra los exones 11-14. Esto se puede hacer en comparación con las sondas contra los exones 1, 2, 3, 4, 5 o 6, que no detectarán la isoforma corta. Alternativamente, la expresión de la isoforma corta se puede medir a nivel de ARNm usando cebadores y/o sondas (por ejemplo, para PCR cuantitativa) contra el exón 9, 10, 11, 12, 13 o 14, tales como sondas contra los exones 11-14. Esto se puede hacer en comparación con sondas contra los exones 1, 2, 3, 4, 5, 6, 7 u 8 que pueden no detectar la isoforma corta. Por ejemplo, tales métodos pueden comprender detectar y cuantificar transcritos de ARN usando cebadores y/o sondas contra los exones de un gen *F12* como se describió anteriormente. Esto puede comprender, por ejemplo, PCR con transcripción inversa (RT-PCR) o PCR cuantitativa (por ejemplo, RT-qPCR). En RT-qPCR, las transcripciones de ARN se cuantifican mediante transcripción inversa en ADNc primero y luego se lleva a cabo la qPCR. Como en la PCR estándar, el ADN se amplifica mediante 3 pasos repetidos: desnaturalización, hibridación y elongación. Sin embargo, en qPCR, el marcado fluorescente permite la recopilación de datos a medida que avanza la PCR.

La evaluación de la actividad de la isoforma corta puede comprender, por ejemplo, medir la actividad de señalización del factor de crecimiento de hepatocitos (HGF)-tirosina-proteína quinasa Met (Met) (por ejemplo, en neuronas), tal como evaluar la conversión de pro-HGF en HGF activo. Dichos ensayos también podrían incluir, por ejemplo, evaluar la activación de precalicreína a calicreína, evaluar la activación de FXI, evaluar la activación de pro-HGF a HGF o evaluar la activación de la vía clásica del complemento.

La evaluación (por ejemplo, medir la expresión o la actividad) se puede realizar en cualquier muestra biológica, tal como cualquier órgano, tipo de tejido o tipo de célula. Por ejemplo, la medición de la expresión se puede realizar en el cerebro, como en las neuronas (por ejemplo, neuronas piramidales o neuronas corticales). El órgano, tejido o célula puede ser de un ser humano o puede ser de un animal no humano (por ejemplo, un roedor con un locus *F12* humanizado como se describe en detalle en otra parte del presente documento).

Si diferentes versiones de una secuencia están asociadas con un número de acceso en diferentes momentos, se entenderá por la versión asociada con el número de acceso en la fecha de presentación efectiva de esta solicitud. La fecha de presentación efectiva significa la anterior entre la fecha de presentación real o la fecha de presentación de una solicitud de prioridad con referencia al número de adhesión, si corresponde. Asimismo, si se publican diferentes versiones de una publicación, sitio web o similar en momentos diferentes, se entenderá por la versión publicada más recientemente en la fecha de presentación efectiva de la solicitud, a menos que se indique lo contrario. Cualquier característica, paso, elemento, realización o aspecto de la invención se puede utilizar en combinación con cualquier otro a menos que se indique específicamente lo contrario. Aunque la presente invención se ha descrito con cierto detalle a modo de ilustración y ejemplo con fines de claridad y comprensión, será evidente que se pueden practicar ciertos cambios y modificaciones dentro del alcance de las reivindicaciones adjuntas.

Breve descripción de las secuencias

Las secuencias de nucleótidos y aminoácidos enumeradas en el listado de secuencias adjunto se muestran utilizando abreviaturas de letras estándar para bases de nucleótidos y códigos de tres letras para aminoácidos. Las secuencias de nucleótidos siguen la convención estándar de comenzar en el extremo 5' de la secuencia y avanzar (es decir, de izquierda a derecha en cada línea) hasta el extremo 3'. Sólo se muestra una cadena de cada secuencia de nucleótidos,

pero se entiende que la cadena complementaria está incluida por cualquier referencia a la cadena presentada. Cuando se proporciona una secuencia de nucleótidos que codifica una secuencia de aminoácidos, se entiende que también se proporcionan variantes degeneradas en codones de la misma que codifican la misma secuencia de aminoácidos. Las secuencias de aminoácidos siguen la convención estándar de comenzar en el extremo amino de la secuencia y avanzar (es decir, de izquierda a derecha en cada línea) hasta el extremo carboxi.

5

Tabla 2. Descripción de Secuencias.

SEQ ID NO	Tipo	Descripción
1	Proteína	Proteína del factor XII de coagulación de ratón (Q80YC5-1; NP 067464.2)
2	Proteína	Péptido de señal de proteína del factor XII de coagulación de ratón
3	Proteína	Cadena pesada de proteína del factor XII de coagulación de ratón
4	Proteína	Cadena ligera de proteína del factor XII de coagulación de ratón
5	Proteína	Proteína del factor XII de coagulación humana
6	Proteína	Péptido de señal de proteína del factor XII de coagulación humano
7	Proteína	Cadena pesada de proteína del factor XII de coagulación humana
8	Proteína	Cadena ligera de proteína del factor XII de coagulación humana
9	ADN	CDS <i>F12</i> de ratón (CCDS36675.1)
10	ADN	Péptido de señal CDS <i>F12</i> de ratón
11	ADN	Cadena pesada CDS <i>F12</i> de ratón
12	ADN	Cadena ligera CDS <i>F12</i> de ratón
13	ADN	CDS <i>F12</i> humano
14	ADN	Péptido de señal CDS <i>F12</i> humano
15	ADN	Cadena pesada CDS <i>F12</i> humano
16	ADN	Cadena ligera CDS <i>F12</i> humano
17	ADN	MAID7374 (Región humanizada de <i>F12</i> con casete de eliminación automática Neo)
18	ADN	MAID7375 (Región humanizada de <i>F12</i> sin casete de eliminación automática Neo)
19	ADN	7374hTU Fwd
20	ADN	7374hTU Sonda
21	ADN	7374hTU Rev.
22	ADN	7374hTD Fwd
23	ADN	7374hTD Sonda

ES 2 966 625 T3

SEQ ID NO	Tipo	Descripción
24	ADN	7374hTD Rev.
25	ADN	7374mTU Fwd
26	ADN	7374mTU Sonda
27	ADN	7374mTU Rev.
28	ADN	7374mTD Fwd
29	ADN	7374mTD Sonda
30	ADN	7374mTD Rev.
31	ADN	Secuencia de <i>F12</i> humano en MAID7374 (Parte I)
32	ADN	Secuencia de <i>F12</i> humano en MAID7374 (Parte II)
33	ADN	ARNm de <i>F12</i> de ratón (NM 021489.3)
34	ADN	ARNm de <i>F12</i> humano (NM 000505.3)
35	Proteína	Cas9
36	ADN	Cas9
37	ARN	Cola de ARNcr
38	ARN	ARNtracr
39	ARN	Guía de armazón de ARN v1
40	ARN	Guía de armazón de ARN v2
41	ARN	Guía de armazón de ARN v3
42	ARN	Guía de armazón de ARN v4
43	ADN	Guía de secuencia objetivo de ARN Plus PAM v1
44	ADN	Guía de secuencia objetivo de ARN Plus PAM v2
45	ADN	Guía de secuencia objetivo de ARN Plus PAM v3
46	Proteína	Proteína del factor XII de coagulación humano (P00748-1; NP 000496.2)
47	Proteína	Cadena pesada de proteína del factor XII de coagulación humano v2
48	ADN	CDS <i>F12</i> humano (CCDS34302.1)
49	ADN	Cadena pesada humana CDS <i>F12</i> v2
50	ADN	<i>F12</i> Exones 1-2 Fwd

ES 2 966 625 T3

SEQ ID NO	Tipo	Descripción
51	ADN	F12 Exones 3-4 Fwd
52	ADN	F12 Exones 5-6 Fwd
53	ADN	F12 Exón 7 Fwd
54	ADN	F12 Exón 9 Fwd
55	ADN	F12 Exón 11 Fwd
56	ADN	F12 Exón 14 Fwd
57	ADN	GAPDH Fwd
58	ADN	F12 Exones 1-2 Rev.
59	ADN	F12 Exones 3-4 Rev.
60	ADN	F12 Exones 5-6 Rev.
61	ADN	F12 Exón 7 Rev
62	ADN	F12 Exón 9 Rev
63	ADN	F12 Exón 11 Rev
64	ADN	F12 Exón 14 Rev
65	ADN	GAPDH Rev
66	ADN	F12 Exones 1-2 Sonda
67	ADN	F12 Exones 3-4 Sonda
68	ADN	F12 Exones 5-6 Sonda
69	ADN	F12 Exón 7 Sonda
70	ADN	F12 Exón 9 Sonda
71	ADN	F12 Exón 11 Sonda
72	ADN	F12 Exón 14 Sonda
73	ADN	GAPDH Sonda
74	Proteína	Proteína de FXII ₂₉₇₋₅₉₆ (ORF1 del Ejemplo 3)
75	ADN	ADN de FXII ₂₉₇₋₅₉₆ (ORF1 del Ejemplo 3)
76	Proteína	Proteína de FXII ₃₅₁₋₅₉₆ (ORF2 del Ejemplo 3)
77	ADN	ADN de FXII ₃₅₁₋₅₉₆ (ORF2 del Ejemplo 3)

SEQ ID NO	Tipo	Descripción
78	Proteína	Proteína de FXII ₅₂₇₋₅₉₆ (ORF3 del Ejemplo 3)
79	ADN	ADN de FXII ₅₂₇₋₅₉₆ (ORF3 del Ejemplo 3)
80	ARN	ARNtracr v2
81	ARN	ARNtracr v3
82	ARN	Guía de armazón de ARN v5
83	ARN	Guía de armazón de ARN v6
84	ARN	Guía de armazón de ARN v7

Ejemplos

Ejemplo 1. Generación de ratones que comprenden un locus (*F12*) del factor XII de coagulación humanizado

5 Un vector orientado grande (LTVEC) que comprende un brazo de homología 5' que comprende 62.4 kb del locus *F12* de ratón (de la construcción del genoma GRCm38/mm10) y un brazo de homología 3' que comprende 107.2 kb del locus *F12* de ratón (de la construcción del genoma GRCm38/mm10) se generó para reemplazar una región de 8.6 kb (8,670 pb) del gen *F12* de ratón con 7.2 kb (7,236 pb) de la secuencia humana correspondiente de *F12* (de la construcción del genoma GRCh38/hg38). La información sobre *F12* de ratón y humano se proporciona en la Tabla 3.

10 La información sobre el LTVEC se proporciona en la Tabla 4. Se describen la generación y el uso de vectores objetivo grandes (LTVEC) derivados del ADN del cromosoma artificial bacteriano (BAC) mediante reacciones de recombinación homóloga bacteriana (BHR) utilizando la tecnología de ingeniería genética VELOCIGENE®, por ejemplo, en el documento US 6,586,251 y Valenzuela et al. (2003) Nat. Biotechnol. 21(6):652-659. La generación de LTVEC mediante métodos de ensamblaje *in vitro* se describe, por ejemplo, en los documentos US 2015/0376628 y WO 2015/200334.

Tabla 3. Factor XII (*F12*) de coagulación humano y de ratón.

	Símbolo oficial	ID del gen NCBI	Fuente principal	ID de ARNm de RefSeq	ID UniProt	Ensamblaje genómico	Ubicación
Ratón	F12	58992	MGI:1891012	NM 021489	Q80YC5	GRCm38/mm10	Cr 13
Humano	F12	2161	HGNC:3530	NM 000505	P00748	GRCh38/hg38	Cr 5

15

Tabla 4. Vector orientado de humanización *F12* del ratón.

	Construcción del genoma	Inicio	Fin	Longitud (pb)
Brazo de ratón de 5'	GRCm38/mm10	Chr13: 55,426,769	Chr13: 55,489,207	62,439
Inserto humano	GRCh38/hg38	Chr5: 177,402,292	Chr5: 177,409,527	7,236
Brazo de ratón de 3'	GRCm38/mm10	Chr13: 55,310,888	Chr13: 55,418,098	107,211

Específicamente, se eliminó una región desde el codón de inicio ATG hasta el codón de parada del locus *F12* de ratón. Se insertó una región correspondiente del *F12* humano desde el codón de inicio ATG hasta el codón de parada en

lugar de la región de ratón eliminada. Se incluyó un casete loxP-mPrr1-Crei-pA-hUb1-em7-Neo-pA-loxP (4,766 pb) dentro del intrón 4 humano (81 pb después del exón 4 y 220 pb antes del exón 5). La región de *F12* humano corriente arriba del casete de autoeliminación incluía 3,874 pb (ATG humano al exón 4 y 81 pb del intrón 4), y la región de *F12* humano corriente abajo del casete de autoeliminación incluía 3,362 pb (220 pb del intrón humano 4 y el exón 5 al codón de parada). Este es el alelo MAID 7374. Véase la parte inferior de Figura 1. Después de la delección del casete, loxP y los sitios de clonación permanecieron en el intrón 4 del *F12* humano. Este es el alelo MAID 7375.

Las secuencias para el péptido señal, la cadena pesada y la cadena ligera del factor XII de coagulación de ratón se exponen en las SEQ ID NOS: 2-4, respectivamente, con las secuencias codificantes correspondientes establecidas en las SEQ ID NOS: 10-12, respectivamente. Las secuencias para el péptido señal, la cadena pesada y la cadena ligera del factor XII de coagulación humano se exponen en las SEQ ID NOS: 6-8, respectivamente, con las secuencias codificantes correspondientes establecidas en las SEQ ID NOS: 14-16, respectivamente. La proteína del factor XII de coagulación humanizado codificada esperada es idéntica a la proteína del factor XII de coagulación humano. Véase Figura 1. Se proporciona una alineación de las proteínas del factor XII de coagulación humana y de ratón en la Figura 3. Las secuencias codificantes de *F12* de ratón y humano se exponen en las SEQ ID NOS: 9 y 13, respectivamente. Las secuencias de proteínas del factor XII de coagulación humana y de ratón se exponen en las SEQ ID NOS: 1 y 5, respectivamente. Las secuencias para la secuencia codificante de *ALB* humanizado esperada y la proteína del factor XII de coagulación humanizado esperada se exponen en las SEQ ID NOS: 13 y 5, respectivamente.

Para generar el alelo mutante, el vector orientado grande descrito anteriormente se introdujo en células madre embrionarias de ratón F1H4. Las células ES de ratón F1H4 se derivaron de embriones híbridos producidos cruzando un ratón hembra C57BL/6NTac con un ratón macho 12956/SvEvTac. Véanse, por ejemplo, los documentos US 2015-0376651 y WO 2015/200805. Después de la selección de antibióticos, TAQMAN® recogió, expandió y analizó las colonias. Véase la Figura 2. Se realizaron ensayos de pérdida de alelo para detectar la pérdida del alelo endógeno de ratón y se realizaron ensayos de ganancia de alelo para detectar la ganancia del alelo humanizado usando los cebadores y sondas establecidos en Tabla 5.

Tabla 5. Ensayos de cribado.

Ensayo	Descripción	Cebador/Sonda	Secuencia
7374hTU	Inserción humana corriente arriba	Fwd	TCGGTGGCAGGCTATGACTTATAG (SEQ ID NO: 19)
		Sonda (FAM)	CAGTTCCTGCCTTCTTCTCTCCC (SEQ ID NO: 20)
		Rev	GGCTTCCCAAGGTGGAATCTAC (SEQ ID NO: 21)
7374hTD	Inserción humana corriente abajo	Fwd	AAGGGCATGAGTGGGTTTACAAG (SEQ ID NO: 22)
		Sonda (Cal)	CGCCTGGAGCAGCTTTGTCCATC (SEQ ID NO: 23)
		Rev	ACACAGAGCGCCTTCTTCA (SEQ ID NO: 24)
7374mTU	LOA del ratón corriente arriba	Fwd	GTTCTGCCTTCTCTCTCCTA (SEQ ID NO: 25)
		Sonda (FAM)	TAGGCTCCACCATGGAAAGACTCCA (SEQ ID NO: 26)
		Rev	CCCATCAGGTGCGTCCTTA (SEQ ID NO: 27)
7374mTD	LOA de ratón corriente abajo	Fwd	TCGCTGCTCCAACCTCTAACG (SEQ ID NO: 28)

Ensayo	Descripción	Cebador/Sonda	Secuencia
		Sonda (Cal)	ACGCCATTCTCCCTGGGATGCTT (SEQ ID NO: 29)
		Rev	ATCGGTGCCTCCCTCCAAGAAG (SEQ ID NO: 30)

Los ensayos de modificación de alelo (MOA) que incluyen ensayos de pérdida de alelo (LOA) y ganancia de alelo (GOA) se describen, por ejemplo, en los documentos US 2014/0178879; US 2016/0145646; WO 2016/081923; y Frendewey et al. (2010) *Methods Enzymol.* 476:295-307. El ensayo de pérdida de alelo (LOA) invierte la lógica de detección convencional y cuantifica el número de copias en una muestra de ADN genómico del locus nativo al que se dirigió la mutación. En un clon celular heterocigoto orientado correctamente, el ensayo LOA detecta uno de los dos alelos nativos (para genes que no están en el cromosoma X o Y), siendo el otro alelo alterado por la modificación orientada. El mismo principio se puede aplicar a la inversa como un ensayo de ganancia de alelo (GOA) para cuantificar el número de copias del vector orientado insertado en una muestra de ADN genómico.

Se generaron ratones F0 a partir de células ES modificadas utilizando el método VELOCIMOUSE®. Específicamente, se inyectaron clones de células ES de ratón seleccionados mediante el ensayo MOA descrito anteriormente en embriones en etapa de 8 células utilizando el método VELOCIMOUSE®. Véanse, por ejemplo, los documentos US 7,576,259; US 7,659,442; US 7,294,754; US 2008/0078000; y Poueymirou et al. (2007) *Nat. Biotechnol.* 25(1):91-99. En el método VELOCIMOUSE®, se inyectan células madre embrionarias (ES) de ratón específicas mediante inyección asistida por láser en embriones en etapa previa a la mórula, por ejemplo, embriones en etapa de ocho células, lo que produce de manera eficiente ratones de generación F0 que son completamente derivados de células ES.

Ejemplo 2. Validación de ratones que comprenden un locus (*F12*) del factor XII de coagulación humanizado

Para validar los ratones *F12* humanizados, se midieron los niveles del factor XII de coagulación humano en muestras de plasma de ratones *F12* humanizados con casete eliminado utilizando un kit ELISA de factor XII de coagulación humano (kit ELISA de antígeno total hFXII Molecular Innovations Cat# HFXIIKT-TOT). Se detectó factor XII de coagulación humano en niveles de 23.6 µg/ml en plasma humano normal y de 5-18 µg/ml en plasma de ratones *F12* humanizados. Se detectó una pequeña cantidad en plasma de ratón de tipo silvestre, quizás debido a una reactividad cruzada mínima con el factor de coagulación XII de ratón. Véase Figura 4.

También se realizó un análisis de transferencia Western (protocolo proporcionado en el Ejemplo 3) de muestras de plasma para medir la expresión del factor XII de coagulación de ratón y del factor XII de coagulación humano en ratones de tipo silvestre y ratones *F12* humanizados. El análisis de transferencia Western de plasma muestra la presencia de factor XII de ratón y la falta de factor XII humano en plasma de ratón de tipo silvestre (FXII^{+/+}) y la presencia de factor XII humano y la falta de factor XII de ratón en plasma de ratón *F12* humanizado (FXII^{hum/hum}). No se detectó factor XII humano ni de ratón en plasma de ratón inactivado de *F12*. Véase Figura 5.

Para validar aún más los ratones *F12* humanizados, se realizaron dos ensayos de plasma funcional: (1) tiempo de tromboplastina parcial activada (aPTT) y (2) tiempo de protrombina (PT). La prueba aPTT evalúa todos los factores de coagulación de las vías intrínsecas y comunes de la cascada de coagulación midiendo el tiempo que tarda en formarse un coágulo después de la adición de calcio y ácido elálgico, mientras que la prueba PT evalúa todos los factores de coagulación de las vías extrínsecas y comunes de la cascada de coagulación midiendo el tiempo que tarda en formarse un coágulo después de la adición de calcio y factor tisular.

No se observaron diferencias significativas en los tiempos de coagulación del aPTT entre ratones de tipo silvestre y ratones *F12* humanizados. Véase Figura 6. De manera similar, los valores de PT no fueron significativamente diferentes entre los ratones de tipo silvestre y los ratones *F12* humanizados, y cada uno tuvo valores de 9-11 segundos (datos no mostrados).

Ensayo de coagulación: PT - Tiempo de protrombina

Este ensayo se utiliza como examen prequirúrgico para detectar posibles problemas de sangrado. Es sensible a deficiencias en el sistema de coagulación extrínseco (factores II, V, VII y X; el PT del plasma humano normal es de ~10-13 segundos). El instrumento utilizado es el analizador de Hemostasia Start4 (Diagnostic Stago). El reactivo utilizado es TriniCLOT PT Excel 6 ml (Stago T1106). Un procedimiento ejemplar es el siguiente:

1. Encender el analizador. La máquina realizará automáticamente una "autocomprobación" y se calentará. El ensayo se realizará únicamente a 37°C.

2. Presionar la tecla "1" en el teclado para el "Modo de prueba", confirmar con la tecla Enter, presionar la tecla "1" para la prueba PT y luego confirme con la tecla Enter.

3. Ingresar el ID# del paciente o el #de muestra para mostrar en la pantalla de trabajo.
4. Colocar las tiras de cubetas en el área de incubación para precalentarlas a 37 °C.
5. Dispensar una bola en cada cubeta de las tiras de cubetas.
6. Dispensar 50 µL de muestra de plasma (control, muestra del paciente o de tratamiento farmacológico) en cada cubeta (no generar burbujas).
7. Opcionalmente, agregar 5 µL de inhibidores diluidos en serie 2X en plasma (solo cuando se pruebe la respuesta a la dosis del fármaco) e incube la mezcla a 37 °C durante 5 min.
8. Iniciar el cronómetro correspondiente a la columna de incubación (presionar la tecla del cronómetro) para una incubación de 180 seg. Cuando el instrumento emita un pitido, transferir rápidamente la tira de cubetas a la columna de prueba.
9. Cebear la Finnpiptette con los reactivos de inicio (reactivo PT precalentado a 37 °C).
10. En la columna de prueba, activar la Finnpiptette presionando la tecla de control de la pipeta (la bola comienza a moverse hacia adelante y hacia atrás en la cubeta) y luego dispense los 100 µL precalentados de reactivo PT en cada cubeta.
11. Cuando se forme el coágulo, la bola dejará de moverse y se registrará el tiempo de coagulación (el resultado se imprimirá después de que se hayan realizado los cuatro tiempos de coagulación en la misma cubeta-tira).

TTPA modificado: Tiempo de tromboplastina parcial activado

Este ensayo se utiliza para detectar anomalías en el sistema de coagulación intrínseco y es sensible a deficiencias de los factores VIII, IX, X, XI y XII (aPTT a los 28-34 segundos en plasma humano normal). El instrumento utilizado es el analizador de Hemostasia Start4 (Diagnostica Stago, Manual Ref 0931079A). Los reactivos utilizados son los siguientes: APTT-XL Ácido elágico ACTCV 4 ml (Thermo Scientific, Cat#95059-804), CaCl₂ 0.02 M 10 ml (Thermo Scientific, Cat#95059-808) o TriniCLOT aPTT S 10 ml (Stago T1201). Un procedimiento ejemplar es el siguiente:

1. Encender el analizador. La máquina realizará automáticamente una "autocomprobación" y se calentará. El ensayo se realizará únicamente a 37°C.
2. Presionar la tecla "1" en el teclado para el "Modo de prueba", confirmar con la tecla Enter, presionar la tecla "2" para la prueba APTT y luego confirme con la tecla Enter.
3. Ingresar el ID# del paciente o el #de muestra para mostrar en la pantalla de trabajo.
4. Colocar las tiras de cubetas en el área de incubación para precalentarlas a 37 °C.
5. Dispensar una bola en cada cubeta de la tira de cubetas. Hay 4 cubetas conectadas en cada tira de cubetas.
6. Dispensar 50 µL de muestra de plasma (control, muestra del paciente o de tratamiento farmacológico) en cada cubeta (pipetear el plasma lentamente, no generar burbujas) y calentar a 37°C durante 1 min.
7. Opcionalmente, agregar 5 µL de inhibidores diluidos en serie 2X en plasma (solo cuando se pruebe la respuesta a la dosis del fármaco) e incube la mezcla a 37 °C durante 5 min.
8. Agregar 50 µL de desencadenante (es decir, ácido elágico o caolín) a cada cubeta.
9. Iniciar el cronómetro correspondiente a la columna de incubación (presionar la tecla del cronómetro) para una incubación de 300 segundos y luego transferir rápidamente la tira de cubetas a la columna de prueba.
10. Cebear la Finnpiptette con los reactivos de inicio (0.02 M de CaCl₂ precalentado a 37°C).
11. En la columna de prueba, activar la Finnpiptette presionando la tecla de control de la pipeta (la bola comienza a moverse hacia adelante y hacia atrás en la cubeta) y luego dispensar los 50 µL de CaCl₂ precalentados a cada cubeta.
12. Cuando se forme el coágulo, la bola dejará de moverse y se registrará el tiempo de coagulación. El resultado se cebará automáticamente después de que se hayan registrado los tiempos de coagulación de las cuatro cubetas.

Ejemplo 3. La diferencia entre la secuencia genómica de *F12* humano y ratón puede impulsar la expresión de la isoforma corta humana, un fenotipo que se recapitula en ratones *F12* humanizados

El factor XII de coagulación (FXII) es sintetizado por los hepatocitos en el hígado y secretado a la circulación, donde inicia el sistema de activación por contacto. Aunque normalmente se piensa que está restringida a la circulación, la proteína FXII se ha encontrado en el cerebro de pacientes con enfermedad de Alzheimer y esclerosis múltiple, y se ha

observado la activación del sistema de contacto en el cerebro y el líquido cefalorraquídeo humanos. Sin embargo, no se ha dilucidado la fuente de la proteína FXII en el cerebro y se desconoce su papel potencial en el cerebro. Usando hibridación *in situ* y ARNseq, se muestran que las neuronas en el cerebro humano y de ratón humanizado FXII expresan una isoforma FXII más corta, observándose la expresión más alta en las neuronas piramidales. Esta transcripción de ARNm de *F12* más corta contiene un marco de lectura abierto que codifica la porción de FXII que abarca sus dominios catalíticos y ricos en prolina. Una versión recombinante de esta proteína FXII más corta es activada por la calicreína plasmática y convierte el factor de crecimiento prohepatocito (HGF) en HGF activo. La señalización de HGF-Met desempeña un papel en el desarrollo y la supervivencia neuronal, y su desregulación se ha implicado en trastornos del desarrollo neurológico y la neurodegeneración. Aquí se muestra por primera vez que una isoforma corta de ARNm de *F12* se produce localmente en el cerebro y se planteó la posibilidad de que el FXII derivado del cerebro pueda estar involucrado en la señalización de HGF-Met en las neuronas.

El factor XII de coagulación (FXII) es una serina proteasa circulante que inicia el sistema de activación por contacto. La activación del FXII en superficies cargadas negativamente o por la calicreína plasmática produce FXII activado (FXIIa). FXIIa inicia la cascada de coagulación intrínseca, que conduce a la generación de trombina y la formación de coágulos. FXIIa también convierte la precalicreína plasmática en calicreína, lo que activa la vía calicreína-quinina, lo que da como resultado la liberación del péptido vasoactivo y proinflamatorio bradixinina. Además de estas funciones bien estudiadas, también se ha demostrado que FXIIa activa el factor de crecimiento de hepatocitos (HGF) y el sistema del complemento *in vitro*. Véase, *por ejemplo*, Shimomura et al. (1995) Eur. J. Biochem. 229:257-261, Peek et al. (2002) J. Biol. Chem. 277:47804-47809, y Ghebrehiwet et al. (1981) J. Exp. Med. 153:665-676. El FXII circula como un zimógeno de cadena sencilla que, tras la escisión en R353 durante la activación, se convierte en una molécula de dos cadenas enzimáticamente activa (α FXIIa). α FXIIa se procesa adicionalmente a β FXIIa, que consta de la cadena ligera que contiene el dominio catalítico y solo una pequeña porción de la cadena pesada.

El FXII es sintetizado por los hepatocitos en el hígado y secretado a la circulación. Aunque normalmente se piensa que está restringida a la circulación, se ha encontrado que la proteína FXII está colocalizada con placas de A β en el cerebro de pacientes con enfermedad de Alzheimer (AD) y en lesiones cerebrales de pacientes con esclerosis múltiple (MS). Véase, *por ejemplo*, Yasuhara et al. (1994) Brain Res. 654:234-240 y Gobel et al. (2016) Nat. Commun. 7:11626. Dado que la barrera hematoencefálica (BBB) normalmente impide el movimiento de las proteínas circulantes hacia el parénquima cerebral, el FXII no debería ingresar al cerebro humano sano. Por lo tanto, el FXII observado en el cerebro de pacientes con AD y MS podría provenir de la circulación a través de una BBB disfuncional o podría sintetizarse localmente. No se ha caracterizado bien si la fuente de proteína FXII encontrada previamente en el cerebro es periférica o local. Véase, *por ejemplo*, Yasuhara et al. (1994) Brain Res. 654:234-240 y Gobel et al. (2016) Nat. Commun. 7:11626.

El FXII en el cerebro puede desencadenar la activación local del sistema de contacto, que se ha observado en el cerebro y el líquido cefalorraquídeo humanos. Véase, *por ejemplo*, Ashby et al. (2012) Neurobiol. Aging 33:1345-1355 y Bergamaschini et al. (2001) Mech. Ageing Dev. 122:1971-1983. FXII de cerebro también podría activar HGF, que desempeña un papel en el desarrollo y la supervivencia neuronal a través de su receptor Met, o el sistema del complemento, que media la integridad y plasticidad sináptica. Véase, *por ejemplo*, Kato (2017) Biomed. Reps. 7:495-503, Wright y Harding (2015) J. Alzheimers Dis. 45:985-1000, y Hammond et al. (2018) Annu. Rev. Cell Dev. Biol. 34:523-544. La fuga de FXII desde la periferia probablemente se produciría de forma patológica, mientras que la síntesis local de ARNm de *F12* en el cerebro podría ser un mecanismo para la activación fisiológica de estos sistemas en el cerebro. Aquí, mostramos que las neuronas del cerebro humano expresan una nueva transcripción de ARNm de *F12* que potencialmente codifica una proteína similar a β FXIIa, y que esta supuesta proteína FXII cerebral puede activarse mediante calicreína plasmática y convertir pro-HGF en HGF activo.

Los estudios publicados muestran resultados contradictorios con respecto a la expresión de *F12* en el cerebro. Un estudio que utilizó cebadores de PCR que abarcaban los exones 7-9 detectó ARNm de *F12* cerebral (Yasuhara et al. (1994) Brain Res. 654:234-240), mientras que otro que utilizó cebadores dentro de los exones 3-5 prácticamente no mostró señal de *F12* (Neth et al. (2001) Thromb. Haemost. 85:1043-1047). Para investigar si la discrepancia entre estos dos estudios podría deberse al exón elegido para la detección, se diseñaron sondas qPCR contra siete regiones dentro de *F12* (exones 1-2, exones 3-4, exones 5-6, exón 7, exón 9, exón 11 y exón 14; Figura 7A y Tabla 6). Se descubrió que, si bien todas las sondas detectaron *F12* en el hígado humano, solo las sondas contra los exones 7, 9, 11 y 14 detectaron *F12* en el cerebro humano (Figura 7B).

Tabla 6. Cebadores y sondas de qPCR.

Objetivo de sonda	Cebador delantero	Cebador inverso	Secuencia de sonda
F12 Exones 1-2	TGGACCAACGGACGGAT G (SEQ ID NO: 50)	CCCAAGGTGGAATCGAA AGTG (SEQ ID NO: 58)	CATGAGGGCTCTGCTGCTC CTGG (SEQ ID NO: 66)
F12 Exones 3-4	TGCCACTTCCCTTCCAG (SEQ ID NO: 51)	GTAGCACACACAGGCTG AG (SEQ ID NO: 59)	ACCACGGCAGCTGTACC ACAAAT (SEQ ID NO: 67)
F12 Exones 5-6	ACTGTCTCTGTCCACAAC ACC (SEQ ID NO: 52)	CCGACAAGCTGAGGCTC AAAG (SEQ ID NO: 60)	TCACTGGAAACCACTGCC AGAAAGAGA (SEQ ID NO: 68)
F12 Exón 7	ACCAACCCGTGCCTCCAT (SEQ ID NO: 53)	GCAGAAAGCTCCGGTGT AG (SEQ ID NO: 61)	CTGCCTAGAGGTGGAGGG CCACC (SEQ ID NO: 69)
F12 Exón 9	GTGGTCTTGTGCTGAA C (SEQ ID NO: 54)	GTTGGGGTCTGGCCTGT G (SEQ ID NO: 62)	CGACCGGCTGAGCTGGGA GTACT (SEQ ID NO: 70)
F12 Exón 11	AGGATCTGACGGTGGTGC TC (SEQ ID NO: 55)	GTAGCTGACGGGGGAGA AG (SEQ ID NO: 63)	CCAGGAACGCCGTAACCA CAGCT (SEQ ID NO: 71)
F12 Exón 14	ACACCGATGTGGCTACT AC (SEQ ID NO: 56)	ACTCGGAATCACCAAG GA (SEQ ID NO: 64)	CTGGATCCGGGAGCACAC CGTTT (SEQ ID NO: 72)
GAPDH	CCAGGTGGTCTCTCTGA CT (SEQ ID NO: 57)	GCTTGACAAAAGTGTCTGT TGA (SEQ ID NO: 65)	TCAACACGACACCCACT CCTC (SEQ ID NO: 73)

Los resultados están respaldados por datos de ARNseq disponibles a través del Proyecto Genotype-Tissue Expression (GTEx) (GT Consortium, The Genotype-Tissue Expression (GTEx) Project, Nat. Genet., 45 (2013) 580-585). Los datos muestran una alta expresión de ARNm de *F12* en el hígado, con una expresión mucho menor en otros tejidos. Fuera del hígado, algunos de los niveles más altos de expresión de *F12* se encuentran en regiones del cerebro (Figura 10A). Como se esperaba, el análisis exón por exón de las lecturas de ARNm de *F12* en hígado humano muestra cobertura para los 14 exones. Sin embargo, las lecturas de ARNm de *F12* en la corteza humana y el hipocampo solo tienen cobertura entre el exón 7 y el exón 14 (Figura 7C).

A continuación, se utilizó la hibridación *in situ* de ARN para investigar la distribución del ARNm de *F12* en el hipocampo humano. Para ello se diseñaron dos sondas RNAscope contra *F12*: una contra los exones 1-6 y otra contra los exones 11-14. De acuerdo con nuestros resultados de qPCR y los datos públicos de GTEx, la sonda solo detectó el ARNm de *F12* contra los exones 11-14 en el hipocampo humano (datos no mostrados). La región con la señal de ARNm de *F12* más abundante fue la capa piramidal del hipocampo, con alguna señal en las células granulares de la circunvolución dentada (datos no mostrados). La señal del RNAscope fuera de los núcleos celulares teñidos con hematoxilina sugiere que la sonda no detecta pre-ARNm sino más bien ARNm maduro. Debido a las dificultades para obtener tejido cerebral humano con una calidad de ARN consistentemente alta y adecuada para el análisis de RNAscope, se recurrió al tejido de ratón para estudiar más a fondo la expresión de *F12* en el cerebro.

Para investigar la distribución de ARNm de *F12* en cerebro de ratón, se realizó análisis de RNAscope en ratones de tipo silvestre (FXII^{+/+}), usando ratones inactivados *F12* (FXII^{-/-}) que no producen proteína del factor XII como control. Como era de esperar, tanto las sondas del exón 1-6 como del exón 11-14 detectaron *F12* en hígado de ratón de tipo silvestre (datos no mostrados). De manera similar al cerebro humano, solo la sonda contra los exones 11-14 detectó ARNm de *F12* en cerebro de ratón de tipo silvestre (datos no mostrados), mientras que no se detectó señal en el cerebro o el hígado de ratones FXII^{-/-}, lo que confirma la especificidad de la sonda para *F12*. La cantidad de señal de la sonda del exón 11-14 de *F12* en el cerebro de ratón parecía menos abundante que la observada en el cerebro humano, lo que nos llevó a comparar los niveles de expresión de ARNm de *F12* extrahepático en tejidos de ratón versus humanos. En comparación con la expresión extrahepática significativa de *F12* en humanos, donde varios tejidos, incluyendo el cerebro, tenían niveles detectables de ARNm de *F12* (Figura 10A), la expresión de *F12* en ratones no se detectó fuera del hígado (Figura 10B). Estos datos respaldan los niveles de ARNm de *F12* detectados por RNAscope en cerebro de ratón de tipo silvestre en comparación con el cerebro humano y sugieren que los niveles de ARNm de *F12* encontrados en cerebro de ratón por RNAscope no son suficientes para ser detectados en tejido a granel por ARNseq.

Para investigar si la expresión diferencial del ARNm de *F12* corto en el cerebro humano versus el de ratón se debe a elementos específicos dentro del gen *F12* humano o a diferencias más generales entre especies, a continuación, se analizaron ratones humanizados *F12* (FXII^{hum/hum}), que solo producen factor XII humano. Ambas sondas detectaron ARNm de *F12* humano de longitud completa en hígado de ratón FXII^{hum/hum} (datos no mostrados), mientras que solo la sonda contra los exones 11-14 de *F12* humano produjo señal en el cerebro de FXII^{hum/hum}. En comparación con los ratones de tipo silvestre, el cerebro del ratón FXII^{hum/hum} tenía una señal mucho mayor, lo que sugiere que el *F12* humano puede tener características reguladoras distintas en comparación con el *F12* del ratón que favorecen su expresión en el cerebro. De manera similar al patrón observado en el cerebro humano, la expresión de *F12* en el cerebro de ratones FXII^{hum/hum} parecía ser principalmente neuronal, con una señal intensa en la capa piramidal del hipocampo y las neuronas corticales (datos no mostrados). Se observó poca o ninguna señal en el cuerpo calloso, que está compuesto principalmente por glía, y en las células endoteliales que recubren el ventrículo lateral (datos no mostrados).

Para investigar más a fondo la posible expresión neuronal de *F12*, se combinó la tecnología RNAscope con inmunofluorescencia utilizando la sonda contra los exones 11-14 de *F12* y el marcador neuronal NeuN en secciones consecutivas de cerebro de ratón FXII^{hum/hum}. Se observó una superposición en la distribución de la inmunofluorescencia de NeuN y la señal del RNAscope de *F12* en secciones cerebrales consecutivas (datos no mostrados), lo que sugiere que *F12* se expresa principalmente en neuronas. La expresión más fuerte de *F12* se detectó en las neuronas piramidales de la región CA2/CA3 del hipocampo (datos no mostrados).

Dado que la expresión de *F12* parece ser principalmente neuronal, a continuación, se investigó si ARNseq podría detectar el ARNm de *F12* en neuronas primarias y líneas celulares neuronales. De hecho, el ARNm de *F12* se expresó en neuronas humanas derivadas de iPS (5.82 [1.65-7.38] FPKM) pero no fue detectable en neuronas corticales primarias de ratón (datos no mostrados). Al igual que en el tejido cerebral humano, el análisis de lectura de ARNm exón por exón mostró que las neuronas iPS humanas expresaban una isoforma más corta de ARNm de *F12* (datos no mostrados). Esta intrigante diferencia en la expresión del ARNm de *F12* en el cerebro entre humanos y ratones llevó a observar las diferencias en el gen *F12* entre las especies.

Los resultados plantean la posibilidad de que la secuencia genómica de *F12* humano pueda tener diferentes elementos reguladores internos que impulsen la expresión de la isoforma más corta en las neuronas. De hecho, los datos disponibles a través del navegador UCSC (Kent y cols. (2002) Genoma Res. 12:996-1006) muestran que si bien se encuentran múltiples elementos reguladores en la región que abarca los exones 7-10 en los genes *F12* humanos y de ratón, una característica única del *F12* humano es una gran isla CpG que abarca los exones 8-12. Las islas CpG son áreas del genoma a menudo asociadas con sitios de inicio de la transcripción y se caracterizan por una composición

elevada de bases G y C y una mayor frecuencia de dinucleótidos CpG. En muchos casos, la presencia de islas CpG en sitios no asociados con transcripciones conocidas condujo al descubrimiento de transcripciones novedosas (Deaton et al. (2011) Genes Dev. 25:1010-1022). Por lo tanto, la presencia de una gran isla CpG en la secuencia *F12* humana pero no en la de ratón puede ser la base de su mayor transcripción en el cerebro humano.

A continuación, se examinó la secuencia de ARNm de *F12* detectada en el cerebro humano en busca de la presencia de marcos de lectura abiertos (ORF) para predecir las supuestas proteínas que se traducirían. Se encontraron tres ORF, definidos por la presencia de codones de inicio y parada en marco con *F12* de longitud completa, en la secuencia de *F12* neuronal humana (Figura 8B). De estos tres ORF, sólo los ORF 2 y 3 se conservan en ratones, con una mala alineación de la secuencia de aminoácidos en la región ORF 1. Curiosamente, GENBANK® (Benson et al. (2013) Nucleic Acids Res. 41:D36-42) enumera cinco ARNm de *F12* humanos cortos que comienzan en el exón 9, que corresponden al ORF 1, mientras que no se han identificado ARNm de *F12* humanos que comienzan en el exón 10 (correspondiente al ORF 2). No se han descrito ARNm de *F12* de ratón que comiencen en el exón 9 o 10 (Figura 8A). El ATG encontrado en el exón 9 humano no se conserva en ratones. En conjunto, estos datos respaldan una mayor transcripción de una isoforma corta de *F12* que comienza en el exón 9 y codifica ORF 1 en humanos, e identifican diferencias en la secuencia y los elementos reguladores en los genes *F12* humanos y de ratón que puede ayudar a explicar la menor expresión del ARNm corto de *F12* en el cerebro del ratón.

Aquí, se mostró que el ARNm de *F12* se expresa en neuronas en el cerebro humano y de ratón FXII^{hum/hum} usando qPCR, RNAscope y ARNseq. Utilizando RNAscope, se identificaron las neuronas como el tipo de célula principal que expresa el ARNm de *F12* corto en el cerebro y se confirmó su expresión en neuronas glutamatérgicas derivadas de iPS mediante ARNseq. Curiosamente, observamos niveles mucho más altos de ARNm de *F12* en ratones FXII^{hum/hum} y cerebro humano en comparación con el cerebro de ratón de tipo silvestre mediante RNAscope. Esta diferencia entre especies fue respaldada aún más por los datos de ARNseq que muestran una expresión insignificante de *F12* en el cerebro de ratón en comparación con el cerebro humano. El análisis de los elementos reguladores en los genes *F12* humanos y de ratón sugirió que la expresión de una transcripción de *F12* humano que comienza en el exón 9 puede deberse a una gran isla CpG que abarca los exones 8-12 que se encuentra en el *F12* humano, pero no en el ratón.

Para determinar el posible papel funcional de la isoforma corta FXII del cerebro, se evaluaron los tres ORF predichos en la transcripción de *F12* del cerebro (Figura 8B). La traducción de ORF 1 daría como resultado una proteína que contiene los dominios catalíticos y ricos en prolina de FXII (M297-S596; SEQ ID NOS: 74 (proteína) y 75 (ADN)). La traducción de ORF 2 daría como resultado una proteína que comprende el dominio catalítico y tres residuos adicionales de la cadena pesada (M351-S596; SEQ ID NOS: 76 (proteína) y 77 (ADN)). ORF 3 generaría una proteína que incluye una porción del dominio catalítico que no contiene la tríada catalítica (M527-S596; SEQ ID NOS: 78 (proteína) y 79 (ADN)). Debido a que la proteína resultante de ORF 2 tendría una cisteína desapareada (C467), lo que probablemente provocaría inestabilidad, y debido a la falta de un dominio catalítico intacto en la proteína resultante de ORF 3, se decidió centrarse en la proteína que se generaría desde ORF 1 (FXII₂₉₇₋₅₉₆). Esta elección se vio respaldada aún más por la presencia de transcripciones de *F12* en GENBANK® comenzando en ORF 1 pero no en ORF 2 y 3 (Figura 8A).

Se evaluó si la supuesta isoforma neuronal FXII (FXII₂₉₇₋₅₉₆) podría activar las vías desencadenadas por FXII de longitud completa. La escisión en R353 por la calicreína plasmática es necesaria para la activación del FXII de longitud completa y da como resultado la separación de las cadenas pesada y ligera. Se probó si el FXII₂₉₇₋₅₉₆ humano recombinante, que carece de una gran porción de la cadena pesada, requiere escisión por la calicreína plasmática para realizar actividad enzimática. Como se muestra en Figura 8B, no se observó actividad enzimática cuando FXII₂₉₇₋₅₉₆ se incubó con el sustrato FXIIa S-2302 en ausencia de calicreína, mientras que FXII₂₉₇₋₅₉₆ adquirió actividad enzimática después de la preincubación con calicreína (Figura 9A). Todas las reacciones se trataron con el inhibidor de calicreína aprotinina antes de la adición del sustrato cromogénico para garantizar que la actividad enzimática no fuera el resultado de la escisión del sustrato mediada por calicreína. Estos resultados sugirieron que, de manera similar al FXII de longitud completa, la supuesta isoforma cerebral FXII se activa después de la escisión por la calicreína. A continuación, se examinó si el FXII₂₉₇₋₅₉₆ activado puede activar recíprocamente la precalicreína a calicreína. Si bien no se observó escisión de la precalicreína con FXII₂₉₇₋₅₉₆ no activado, el FXII₂₉₇₋₅₉₆ activado convirtió la precalicreína en calicreína como se ve a través de la generación de la cadena pesada de calicreína. (Figura 11). Así, FXII₂₉₇₋₅₉₆ es capaz de iniciar la vía calicreína-cinina, lo que sugiere que la activación de FXII₂₉₇₋₅₉₆ puede conducir a la generación de bradicinina.

Recientemente se demostró que la escisión de FXII dentro de su dominio rico en prolina, que libera una porción de la cadena pesada, aumenta la susceptibilidad de FXII a la activación por calicreína en R353. Véase, por ejemplo, de Maat et al. (2019) J. Thromb. Haemost. 17:183-194. La hipótesis es que FXII₂₉₇₋₅₉₆, en donde falta la cadena pesada N-terminal del dominio rico en prolina, esencialmente se escindiría previamente y, por lo tanto, podría ser susceptible a la activación por niveles de calicreína que no activarían el FXII de longitud completa. Porque el FXII₂₉₇₋₅₉₆ recombinante utilizado en este estudio se produjo en *E. coli* y carecía de glicosilación, no se pudo comparar directamente su susceptibilidad a la activación con la del FXII de longitud completa purificado a partir de plasma humano. Por lo tanto, se buscó generar enzimáticamente una proteína similar al FXII₂₉₇₋₅₉₆ del FXII completo. Análisis de supuestos sitios de escisión en la secuencia de aminoácidos de FXII utilizando el software Prosper (Song et al. (2012) PLoS One 7:e50300) reveló un sitio de escisión de catepsina K entre L296 y M297, que generaría una proteína FXII truncada idéntica a la isoforma potencialmente expresada en el cerebro (FXII₂₉₇₋₅₉₆).

Como se muestra en Figura 9B, la incubación de FXII de longitud completa con catepsina K generó fragmentos de FXII estables con tamaños que apoyan la escisión entre L296 y M297, lo que produciría dos fragmentos de ~35 kDa que se esperaría que funcionaran a ~40 kDa debido a la glicosilación. A continuación, se comparó la actividad enzimática del FXII de longitud completa y el FXII truncado generado por la escisión de catepsina K. Como se esperaba, el FXII de longitud completa y el FXII escindido por catepsina K tuvieron una actividad insignificante por sí solos, pero el FXII escindido por catepsina K tuvo una actividad dramáticamente mayor en comparación con el FXII de longitud completa cuando se incubaron con calicreína plasmática (Figura 9B). Los resultados concuerdan con un estudio reciente que muestra que la catepsina K actúa sinérgicamente con la calicreína para activar FXII (de Maat et al. (2019) J. Thromb. Haemost. 17:183-194) y sugieren que la supuesta isoforma FXII₂₉₇₋₅₉₆ de cerebro, que es similar al FXII escindido por catepsina-K, podría tener una mayor susceptibilidad a la activación por la calicreína plasmática y posiblemente otras proteasas.

Además de activar la precalicreína y la vía de coagulación intrínseca, se ha demostrado que FXIIa convierte pro-HGF en HGF activo. Véase, *por ejemplo*, Shimomura et al. (1995) Eur. J. Biochem. 229:257-261 y Peek et al. (2002) J. Biol. Chem. 277:47804-47809. Se confirmó la capacidad del activador del factor de crecimiento de hepatocitos (HGFA) y β FXIIa para escindir pro-HGF (Figura 9C). Además, se encontró que el FXII₂₉₇₋₅₉₆ activado por calicreína escindió el pro-HGF a la forma activa de dos cadenas (Figura 9D), lo que sugiere que el FXII de cerebro podría contribuir a la activación local del HGF. Esta capacidad de activar pro-HGF es específica, ya que otras serina proteasas encontradas en el cerebro, como la trombina y el activador del plasminógeno tisular no convirtieron pro-HGF en HGF (Figuras 9C y 9D).

La evidencia acumulada sugiere que la expresión de las proteínas del sistema de coagulación y de contacto ocurre fuera del hígado, especialmente en condiciones patológicas. Véanse, *por ejemplo*, Ashby y cols. (2012) Neurobiol. Aging 33:1345-1355, Yin et al. (2010) Am. J. Pathol. 176:1600-1606, y Takano et al. (2003) Brain Res. 978:72-82. También está surgiendo un papel para la expresión extrahepática del FXII: estudios recientes han demostrado que el FXII se expresa en los fibroblastos pulmonares (Jablonska et al. (2010) J. Biol. Chem. 285:11638-11651) y que el FXII expresado por neutrófilos contribuye a la activación y migración de neutrófilos a sitios de inflamación (Stavrou et al. (2018) J. Clin. Invest. 128:944-959). Aquí, se muestra por primera vez que el ARNm de *F12* se expresa en neuronas en el cerebro humano y de ratón FXII^{hum/hum} utilizando qPCR, RNAscope y ARNseq. Estudios anteriores han producido resultados contradictorios con respecto a la expresión de *F12* en el cerebro humano. Utilizando cebadores que abarcaban los exones 7 a 9, un grupo detectó ARNm de *F12* (Yasuhara et al. (1994) Brain Res. 654:234-240), mientras que qPCR realizado utilizando cebadores y una sonda dentro de los exones 3-5 prácticamente no mostró señal de *F12* (Neth et al. (2001) Thromb. Haemost. 85:1043-1047). Estos estudios utilizaron métodos diferentes y, por lo tanto, no son directamente comparables; sin embargo, Yasuhara et al. concluyó que el ARNm de *F12* se encuentra en el cerebro, mientras que Neth et al. llegó a la conclusión de que el ARNm de *F12* se encuentra exclusivamente en el hígado. Los resultados ayudan a resolver esta discrepancia, ya que se encontró que el ARNm de *F12* correspondiente a los exones 7-14 pero no a los exones 1-6, se encuentra en el cerebro (Figura 7C), lo que respalda ambos informes.

Utilizando RNAscope, se identificaron las neuronas como el tipo de célula principal que expresa el ARNm de *F12* corto en el cerebro y se confirmó su expresión en neuronas glutamatérgicas derivadas de iPS mediante ARNseq. Curiosamente, se observaron niveles mucho más altos de ARNm de *F12* en el cerebro humano y del ratón FXII^{hum/hum} en comparación con el cerebro del ratón WT mediante RNAscope. Esta diferencia entre especies fue respaldada aún más por los datos de ARNseq que muestran una expresión insignificante de *F12* en el cerebro de ratón en comparación con el cerebro humano. El análisis de los elementos reguladores en los genes *F12* humanos y de ratón sugirió que la expresión de una transcripción de *F12* humano que comienza en el exón 9 puede deberse a una gran isla CpG que abarca los exones 8-12 que se encuentra en el *F12* humano, pero no en el ratón. Se desconoce la importancia funcional de esta divergencia entre las neuronas humanas y de ratón.

Un trabajo reciente ha sugerido implicaciones funcionales para una posible proteína de FXII de cerebro que abarca FXII₂₉₇₋₅₉₆. La activación del FXII se puede lograr mediante niveles subumbrales de calicreína plasmática cuando el FXII es procesado por primera vez por otras enzimas como la elastasa o la catepsina K. Véase, *por ejemplo*, de Maat et al. (2019) J. Thromb. Haemost. 17:183-194. Se cree que la escisión de la cadena pesada de FXII descubre el sitio de escisión de activación de FXII (R353-V354), lo que permite su activación más eficiente en solución. Véase, *por ejemplo*, de Maat et al. (2019) J. Thromb. Haemost. 17:183-194. Con base en análisis *in silico*, el supuesto FXII₂₉₇₋₅₉₆ neuronal puede generarse mediante la escisión del FXII de longitud completa con catepsina K. Si bien aún no se ha demostrado el sitio exacto de escisión de la catepsina K en el FXII, los resultados con el FXII escindido por catepsina K sugieren que la supuesta isoforma neuronal del FXII podría activarse mediante niveles bajos de calicreína. En este sentido, se han encontrado proteínas y ARNm de precalicreína plasmática en diversas regiones del cerebro (Ashby et al. (2012) Neurobiol. Aging 33:1345-1355 y Neth et al. (2001) Thromb. Haemost. 85:1043-1047), y en niveles comparables (Cerf y Raidoo (2000) Metab. Brain Dis. 15:315-323) a los que se utilizaron en este estudio. Además, la expresión de precalicreína plasmática se localizó en las neuronas del hipocampo (Cerf y Raidoo (2000) Metab. Brain Dis. 15:315-323), las células en donde se encontraron niveles más altos de ARNm de *F12* corto, lo que sugiere que puede ocurrir la activación del FXII neuronal puede ocurrir *in vivo*. Además, los resultados plantean la posibilidad de que la isoforma neuronal del FXII pueda tener una mayor susceptibilidad a la activación por otras proteasas, tal vez producidas por neuronas u otras células cerebrales, que no activarían el FXII de longitud completa.

La activación de pro-HGF a HGF es necesaria para su señalización a través del receptor Met. Si bien se ha establecido un papel para la señalización de HGF-Met en las neuronas (Kato (2017) Biomed. Rep. 7:495-503 y Tyndall y Walikonis (2006) Cell Cycle 5:1560-1568), no se ha caracterizado bien cómo se activa el pro-HGF en el cerebro. HGFA es el principal activador de HGF en la circulación. Véase, *por ejemplo*, Itoh et al. (2004) Gastroenterology 127:1423-1435. Dado que los ratones inactivados para HGFA no tienen anomalías en el desarrollo (Itoh et al. (2004) Gastroenterology 127:1423-1435) mientras que los ratones inactivados para HGF son embrionariamente letales (Uehara et al. (1995) Nature 373:702-705), es probable que otros activadores de HGF participen en la activación de HGF en los tejidos. Curiosamente, FXII es un homólogo de HGFA (Ponczek et al. (2008) J. Thromb. Haemost. 6:1876-1883) y se ha demostrado que convierte pro-HGF en HGF activo *in vitro*. Véase, *por ejemplo*, Shimomura et al. (1995) Eur. J. Biochem. 229:257-261 y Peek et al. (2002) J. Biol. Chem. 277:47804-47809. La hipótesis es que el supuesto FXII neuronal puede participar en la conversión de pro-HGF a HGF activo en el cerebro, contribuyendo a la señalización neuronal de HGF-Met.

La señalización neuronal de Met se ha implicado en una variedad de procesos fisiológicos, incluyendo el crecimiento de neuritas, la arborización dendrítica, la potenciación a largo plazo y la memoria. Véase, *por ejemplo*, Kato (2017) Biomed. Rep. 7:495-503, Wright y Harding (2015) J. Alzheimers Dis. 45:985-1000, y Tyndall y Walikonis (2006) Cell Cycle 5:1560-1568. Se observó una expresión enriquecida del ARNm de *F12* en las neuronas piramidales neocorticales y del hipocampo, lo que plantea la posibilidad de que la activación de HGF por el FXII neuronal putativo pueda ocurrir en estas células. Curiosamente, la delección de Met en las neuronas que surgen del palio dorsal (que incluyen las neuronas del hipocampo y neocorticales) produce déficits en la alternancia espontánea y en la hipoactividad (Thompson and Levitt (2015) J. Neurodev. Disord. 7:35), respaldando la importancia de la señalización de HGF-Met en células que expresan ARNm de *F12*. Una mayor investigación sobre la función de la isoforma corta FXII puede arrojar luz sobre su papel en estas poblaciones de células.

Además de los conocimientos sobre el papel de la señalización neuronal de HGF-Met derivada de *in vitro* y estudios en animales, la evidencia acumulada sugiere que la disregulación en la señalización de HGF-Met contribuye a los trastornos del desarrollo neurológico y la neurodegeneración en humanos. Véase, *por ejemplo*, Eagleson et al. (2017) Biol. Psychiatry 81:424-433, Peng et al. (2013) Int. Rev. Neurobiol. 113:135-165, y Wright y Harding (2015) J. Alzheimers Dis. 45:985-1000. En particular, la expresión de la proteína Met está disminuida en las neuronas del hipocampo de los pacientes con EA (Hamasaki et al. (2014) Neuropathology 34:284-290), mientras que los niveles de proteína HGF aumentan en el cerebro con EA (Fenton et al. (1998) Brain Res. 779:262-270). Si los niveles neuronales de FXII están alterados en la AD u otros estados patológicos y si el FXII neuronal desempeña un papel en los procesos patológicos son cuestiones abiertas. Los estudios futuros arrojarán luz sobre los detalles de la expresión neuronal del FXII y sobre el papel de la isoforma neuronal del FXII en la activación fisiológica y patológica del HGF y las vías de contacto en el cerebro.

Materiales y métodos

qPCR. El ARN total del hígado humano, el cerebro humano completo, el hipocampo, el lóbulo temporal y la corteza fueron de Clontech (Takara Bio USA; Mountain View, CA). El ARNm se transcribió de forma inversa a ADNc utilizando SuperScript® VILO™ Mezcla maestra (Invitrogen de Life Technologies; Waltham, MA). El ADNc se diluyó a 0.5-5 ng/μL, y la entrada de ADNc de 2.5-25 ng se amplificó con SensiFAST Hi-ROX MasterMix (Bioline; Taunton, MA) utilizando el sistema de detección de secuencias ABI 7900HT (Applied Biosystems; Foster City, CA). Los cebadores y sondas contra *F12* humano y GAPDH utilizados se muestran en la Tabla 5. Los datos se presentan como % de expresión del gen objetivo observado en el tejido de control ($2^{-(\Delta\Delta Ct)} \times 100$).

Tejido cerebral humano. Para el análisis de RNAscope, se utilizaron dos fuentes de tejido cerebral humano. Se cortó tejido cerebral de hipocampo recién congelado del Harvard Brain Tissue Resource Center (Belmont, MA) a 10 μm, se montó en portaobjetos VWR superfrost y se almacenó a -80 °C. Las secciones de hipocampo humano incluidas en parafina y fijadas con formalina de 5 μm premontadas eran de Abcam (Cambridge, MA).

Colección de plasma. Para la recolección de plasma, se extrajo sangre de la vena cava utilizando agujas preparadas con citrato de sodio. La sangre se centrifugó a 1500 xg a temperatura ambiente durante 10 minutos y se recogió y almacenó plasma pobre en plaquetas a -80 °C. Para la recolección de tejido, los ratones se perfundieron transcardialmente con solución salina al 0.9%. Se extrajeron los cerebros y los hígados, se congelaron en hielo seco en un compuesto de temperatura de corte óptima y se almacenaron a -80 °C. Se montaron secciones de diez μm en portaobjetos Superfrost plus (VWR; Radnor, PA) y se almacenaron a -80 °C.

Hibridación de ARN *in situ*. La hibridación *in situ* de ARN se realizó utilizando el sistema RNAscope 2.5 HD RED (Advanced Cell Diagnostics; Newark, CA) siguiendo el protocolo del fabricante para tejido incluido en parafina, recién congelado o fijado con formalina. Las sondas (números de catálogo ACDBio 313901 (PPIB (humano)), 313911 (PPIB (ratón)), 310043 (DAPB), 493191 (exón *F12* 1-6 (humano)), 473781 (exón *F12* 11-14 (humano)), 493201 (exón *F12* 1-6 (ratón)), y 483971 (exón *F12* 11-14 (ratón))) se usaron en un ensayo de RNAscope manual de plex único, y la señal colorimétrica de las sondas de RNAscope se capturó con un objetivo de 40X usando un escáner de portaobjetos Aperio (Leica Microsystems; Buffalo Grove, IL). Para confirmar que el tejido era adecuado para el análisis de ARN, se usó una sonda contra un gen expresado ubicuamente (ciclofilina B; PPIB) como control positivo, y se usó una sonda contra el gen bacteriano (4-hidroxi-tetrahidrodipicolinato reductasa; DapB) como un control negativo. Debido a la alta

especificidad de las sondas RNAscope, se diseñaron sondas RNAscope separadas específicas para *F12* de ratón y humano.

Inmunofluorescencia/RNAscope. Se realizaron inmunofluorescencia y RNAscope en secciones consecutivas de cerebro de ratón FXII^{hum/hum}. RNAscope se ejecutó como se describió anteriormente usando la sonda contra exones *F12* 11-14, excepto que el tejido se tiñó con DAPI y las imágenes se adquirieron utilizando un objetivo de 20X con Ex/Em de 590/617 para RNAscope RED y 358/461 para DAPI en un escáner de portaobjetos Zeiss AxioScan Z.1 Carl Zeiss Microscopy; Thornwood, NY). La señal de RNAscope se pseudocoloreó de blanco usando el software Halo (Indica Labs; Corrales, NM). Para la inmunofluorescencia, el tejido se fijó posteriormente con paraformaldehído al 4 % (Electron Microscopy Sciences; Hatfield, PA) en tampón fosfato y se bloqueó con suero de burro al 3 % (Jackson Laboratory; Bar Harbor, ME) que contenía Triton-X al 0,25 % (Sigma Aldrich; St. Louis, MO) en PBS. Luego, los portaobjetos se incubaron con anticuerpo NeuN (Abcam; Cat# Ab177487) diluido 1:500 en tampón de bloqueo durante la noche a 4 °C, luego con anticuerpo secundario conjugado Alexa-488 (ThermoFisher) y DAPI (ThermoFisher) diluidos en suero de burro al 3% durante 1 hora a temperatura ambiente. Después del lavado en PBS, los portaobjetos se cubrieron con medio de montaje Aqua-Mount (ThermoFisher). Las imágenes se adquirieron utilizando un objetivo de 20X con Ex/Em de 490/525 para Alexa-488 y 358/461 para DAPI en un escáner de diapositivas AxioScan Z.1.

Ensayo de actividad FXIIa. Para medir activación de FXII₂₉₇₋₅₉₆ por calicreína plasmática, FXII₂₉₇₋₅₉₆ humano recombinante (Proteintech; Rosemont, IL) se incubó con calicreína plasmática 15 nM (Molecular Innovations; Novi, MI) durante 18 horas a 37 °C en HEPES 20 mM que contenía NaCl 140 mM, pH 7.4 (solución salina tamponada con HEPES; HBS). Luego se inhibió la actividad de la calicreína utilizando 100 KUI/ml de aprotinina (Sigma-Aldrich; St. Louis, MO) y se midió la actividad de FXIIa utilizando el sustrato cromogénico S-2302 (H-D-Pro-Phe-Arg-pNA-2HCl; Diapharma; West Chester, OH) a una concentración de 0.8 mM. Los ensayos se realizaron a 37 °C en placas MaxiSorp de poliestireno de 96 pocillos de fondo plano (Thermo Fisher). La actividad enzimática se detectó como un cambio en la absorbancia a 405 nm utilizando un lector de microplacas Spectramax i3X (Molecular Devices; Downingtown, PA).

El FXII escindido por catepsina K se generó mediante incubación de FXII (1.25 uM; Molecular Innovations) con catepsina K (100 nM; Enzo Life Sciences) en HBS. Las reacciones se detuvieron añadiendo tampón de muestra no reductor y calentando a 100 °C durante 5 min. Los fragmentos de FXII se separaron mediante SDS-PAGE y se visualizaron mediante tinción con azul de Coomassie. La actividad proteolítica del FXII escindido por catepsina K se analizó incubando FXII (375 nM) con catepsina K (30 nM) o vehículo con calicreína plasmática (15 nM) durante 15 minutos. Luego se inhibió la actividad de calicreína con aprotinina y se midió la actividad de FXIIa con el sustrato cromogénico S-2302.

Para analizar la escisión de la precalicreína plasmática, FXII₂₉₇₋₅₉₆ (300 nM) se incubó primero con calicreína plasmática (15 nM) o vehículo a 37°C durante 18 horas. Luego se inactivó la actividad de calicreína usando aprotinina (100 KUI/ml; Sigma). La precalicreína purificada a partir de plasma humano (200 nM; Molecular Innovations) se incubó con FXII₂₉₇₋₅₉₆ activado con calicreína o vehículo (225 nM) o FXII (225 nM) a 37°C durante 2 horas en HBS. Las reacciones se detuvieron añadiendo tampón de muestra reductor. Se realizó SDS-PAGE como se describió anteriormente y las transferencias se analizaron con un anticuerpo contra la precalicreína humana (Abcam; Cat# ab1006).

Transferencia Western Plasma de ratones FXII^{-/-}, FXII^{+/+} y FXII^{hum/hum} se analizó en condiciones reductoras mediante transferencia Western. La SDS-PAGE se realizó utilizando geles de Tris-glicina al 4-20 % (Invitrogen), los geles se transfirieron a membranas de PVDF de 0.2 µm (BioRad; Hercules, CA) y se analizaron con anticuerpos contra el factor XII de ratón (CloudClone Corp; Katy, TX; Cat# PAA677Mu01), factor XII humano (Abeam; Cat# Ab196670) y albúmina (Thermo Scientific Pierce; Waltham, MA; Cat# PA1-29335).

Para analizar la activación de HGF, FXII₂₉₇₋₅₉₆ (300 nM) se incubó primero con calicreína plasmática (15 nM) o vehículo a 37°C durante 18 horas. Luego se inactivó la actividad de calicreína usando aprotinina (100 KUI/ml; Sigma). Se incubó pro-HGF humano recombinante (25 µg/ml; R&D Systems; Minneapolis, MN) con FXII₂₉₇₋₅₉₆ activado con calicreína o vehículo (225 nM), activador del factor de crecimiento de hepatocitos humano recombinante (225 nM; Sino Biological; Wayne, PA), activador del plasminógeno tisular de dos cadenas (225 nM; Oxford Biomedical Research; Rochester Hills, MI), o βFXIIa (50 nM; Molecular Innovations) a 37°C durante 5 horas en HBS. Las reacciones se detuvieron añadiendo tampón de muestra reductor. Se realizó SDS-PAGE como se describió anteriormente y las transferencias se analizaron con un anticuerpo contra HGF humano (R&D Systems; Cat# AF-294-SP).

Cultivo de células. Neuronas glutamatérgicas iPS humanas (iCell® Glutaneuronas; Dinámica Celular; Madison, WI) se cultivaron de acuerdo con las instrucciones del fabricante y la diferenciación se validó mediante ARNseq. Las células mostraron una alta expresión de dos marcadores neuronales bien conocidos, la proteína asociada a microtúbulos 2 (MAP2) y el transportador vesicular de glutamato VGLUT2 (SCL17A6) y una baja expresión de GFAP (marcador de astrocitos) y NES (marcador de células madre neurales y células endoteliales en proliferación) (atos no mostrados). Se cultivaron células de neuroblastoma humano SH-SY5Y y células de neuroblastoma de ratón Neuro-2a de American Type Tissue Collection (ATCC; Manassas, VA) en EMEM (Irvine Scientific; Santa Ana, CA) que contenían L-glutamato, penicilina, estreptomycin y 10% de suero fetal bovino (ThermoFisher). Se cultivaron neuronas corticales primarias de ratones C57BL/6 (Lonza) de acuerdo con las instrucciones del fabricante.

ARNseq. Los protocolos para la extracción de ARN y la preparación de la biblioteca de secuenciación fueron similares a los descritos anteriormente (Atanasio et al. (2016) Sci. Rep. 6:23204,). Se prepararon bibliotecas de ARN-seq específicas de cadena a partir de 500 ng de ARN utilizando el kit KAPA en cadena ARNm-Seq (KAPA Biosystems). Las bibliotecas se amplificaron mediante PCR de doce ciclos. La secuenciación se realizó en Illumina HiSeq®2000 (Illumina) mediante ejecución de lectura única multiplexada (para neuronas humanas iPS, neuronas primarias de ratón y tejido de ratón) o ejecución de lectura emparejada (para líneas celulares SH-SY5Y y Neuro-2A). Los archivos FASTQ resultantes se analizaron mediante FastQC para garantizar una calidad de datos suficiente. Las lecturas se asignaron al genoma humano (hg19) o al genoma del ratón (mm10) utilizando el software comercial ArrayStudio (OmicSoft) y se permitieron dos desajustes. En las muestras, 83%-90% de las lecturas se asignaron de forma única. Las expresiones genéticas se derivaron de lecturas de secuenciación sin procesar utilizando el canal de análisis OmicSoft ArrayStudio ARNseq (Qiagen Inc.). Para cada gen, se identificaron y contaron las lecturas asignadas de forma única a los exones del gen. Los recuentos de lectura resultantes se resumieron a nivel genético como expresión bruta.

REIVINDICACIONES

1. Un roedor que comprende en su genoma un locus *F12* endógeno humanizado en el que un segmento del locus *F12* endógeno ha sido eliminado y reemplazado con una secuencia de *F12* humano correspondiente,
- 5 en donde una región del locus *F12* endógeno que comprende tanto la secuencia codificante como la secuencia no codificante ha sido eliminada y reemplazada con una secuencia de *F12* humano correspondiente que comprende tanto la secuencia codificante como la secuencia no codificante,
- 10 en donde la región del locus *F12* endógeno desde el codón de inicio hasta el codón de parada se ha eliminado y reemplazado con la secuencia de *F12* humano correspondiente, y
- en donde el roedor expresa una isoforma corta del factor XII de coagulación humano en el cerebro, en donde:
 - 15 (i) la isoforma corta del factor XII de coagulación humano está codificada por un ARNm de *F12* detectado por sondas contra los exones 11-14 del *F12* humano, pero no por sondas contra los exones 1-6 del *F12* humano; y/o
 - (ii) la isoforma corta del factor XII de coagulación humano es FXII₂₉₇₋₅₉₆.
- 20 2. El roedor de la reivindicación 1, en donde el locus *F12* endógeno humanizado comprende el promotor *F12* endógeno, en donde la secuencia de *F12* humano está operativamente unida al promotor *F12* endógeno.
3. El roedor de cualquier reivindicación anterior, en donde:
 - (I) la región endógena 3' no traducida de *F12* no se ha eliminado ni reemplazada con la correspondiente secuencia de *F12* humano; y/o
 - 25 (II) la región endógena 5' no traducida de *F12* no se ha eliminado ni reemplazada con la correspondiente secuencia de *F12* humano.
4. El roedor de cualquier reivindicación anterior, en donde:
 - (I) la secuencia de *F12* humano en el locus *F12* endógeno humanizado comprende una secuencia al menos 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, o 100 % idéntica a la secuencia expuesta en la SEQ ID NO: 31 o 32; y/o
 - 30 (II) el locus *F12* endógeno humanizado codifica una proteína que comprende una secuencia al menos 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, o 100 % idéntica a la secuencia expuesta en la SEQ ID NO: 5 o 46; y/o
 - (III) el locus *F12* endógeno humanizado comprende una secuencia codificante que comprende una secuencia al menos 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, o 100 % idéntica a la secuencia expuesta en la SEQ ID NO: 13 o 48; y/o
 - 35 (IV) el locus *F12* endógeno humanizado comprende una secuencia al menos 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, o 100 % idéntica a la secuencia expuesta en la SEQ ID NO: 17 o 18.
5. El roedor de cualquier reivindicación anterior, en donde el locus *F12* endógeno humanizado comprende un casete de selección o un gen indicador dentro de un intrón de la secuencia de *F12* humano correspondiente, o en donde el locus *F12* endógeno humanizado no comprende un casete de selección o un gen indicador.
- 40 6. El roedor de cualquier reivindicación anterior, en donde el roedor es homocigoto para el locus *F12* endógeno humanizado.
7. El roedor de cualquier reivindicación anterior, en donde el roedor comprende el locus *F12* endógeno humanizado en su línea germinal.
8. El roedor de cualquier reivindicación anterior, en donde el roedor es una rata o un ratón.
- 45 9. El roedor de la reivindicación 8, en donde el roedor es el ratón.
10. El roedor de cualquier reivindicación anterior, en donde el roedor tiene un tiempo de tromboplastina parcial activado (aPTT) que no es significativamente diferente de un aPTT en un roedor de tipo silvestre correspondiente y/o en donde el roedor tiene un tiempo de protrombina (PT) que no es significativamente diferente de un PT en el correspondiente roedor de tipo silvestre.
- 50 11. El roedor de cualquier reivindicación anterior, en donde el roedor tiene niveles plasmáticos del factor XII de coagulación humano de al menos aproximadamente 5 µg/ml.
12. Un método para evaluar la actividad de un reactivo orientado al factor XII de coagulación humano *in vivo*, que comprende:

- (a) administrar el reactivo orientado al factor XII de coagulación humano al roedor de una cualquiera de las reivindicaciones 1-11; y
- (b) evaluar la actividad del reactivo orientado al factor XII de coagulación humano en el roedor.
- 5 13. El método de la reivindicación 12, en donde la administración comprende administración mediada por virus adenoasociados (AAV), administración mediada por nanopartículas lipídicas (LNP), administración hidrodinámica (HDD) o inyección.
14. El método de la reivindicación 12 o 13, en donde:
- (I) el paso (b) comprende aislar plasma del roedor y evaluar la actividad del reactivo orientado al factor XII de coagulación humano en el plasma; y/o
- 10 (II) el paso (b) comprende evaluar la actividad del reactivo orientado al factor XII de coagulación humano en el hígado o el cerebro del roedor; y/o
- (III) el paso (b) comprende evaluar la coagulación o la generación de trombina; y/o
- (IV) el paso (b) comprende evaluar la señalización de HGF-Met; y/o
- 15 (V) el paso (b) comprende medir la expresión de un ARN mensajero de *F12* codificado por el locus *F12* endógeno humanizado; y/o
- (VI) el paso (b) comprende medir la expresión de una proteína del factor XII de coagulación codificada por el locus del factor XII de coagulación endógeno humanizado; y/o
- (VII) el reactivo orientado al factor XII de coagulación humano es un agente de edición del genoma, y el paso (b) comprende evaluar la modificación del locus *F12* endógeno humanizado, opcionalmente en donde el paso (b) comprende medir la frecuencia de inserciones o deleciones dentro del locus *F12* endógeno humanizado.
- 20 15. El método de una cualquiera de las reivindicaciones 12-14, en donde:
- (I) el reactivo orientado al factor XII de coagulación humano comprende un agente nucleasa diseñado para orientarse a una región de un gen *F12* humano, opcionalmente en donde el agente nucleasa comprende una proteína Cas y un ARN guía diseñado para orientarse a una secuencia objetivo de ARN guía en el gen *F12* humano, y opcionalmente en donde la proteína Cas es una proteína Cas9;
- 25 (II) el reactivo orientado al factor XII de coagulación humano comprende un ácido nucleico donante exógeno, en donde el ácido nucleico donante exógeno está diseñado para orientarse al gen *F12* humano y, opcionalmente, en donde el ácido nucleico donante exógeno se administra mediante AAV;
- 30 (III) el reactivo orientado al factor XII de coagulación humano es un agente de ARNi o un oligonucleótido antisentido;
- (IV) el reactivo orientado al factor XII de coagulación humano es una proteína de unión a antígeno; o
- (V) el reactivo orientado al factor XII de coagulación humano es una molécula pequeña.
16. Un método para optimizar la actividad de un reactivo orientado al factor XII de coagulación humano *in vivo*, que comprende:
- 35 (I) realizar el método de una cualquiera de las reivindicaciones 12-15 por primera vez en un primer roedor que comprende en su genoma un locus *F12* endógeno humanizado;
- (II) cambiar una variable y realizar el método del paso (I) una segunda vez con la variable cambiada en un segundo roedor que comprende en su genoma un locus *F12* endógeno humanizado; y
- 40 (III) comparar la actividad del reactivo orientado al factor XII de coagulación humano en el paso (I) con la actividad del reactivo orientado al factor XII de coagulación humano en el paso (II), y seleccionar el método que da como resultado la mayor actividad.
17. El método de la reivindicación 16, en donde:
- 45 (I) la variable cambiada en el paso (II) es el método de administración para introducir el reactivo orientado al factor XII de coagulación humano en el roedor;
- (II) la variable cambiada en el paso (II) es la vía de administración para introducir el reactivo orientado al factor XII de coagulación humano en el roedor;

(III) la variable cambiada en el paso (II) es la concentración o cantidad del reactivo orientado al factor XII de coagulación humano introducido en el roedor;

(IV) la variable cambiada en el paso (II) es la forma del reactivo orientado al factor XII de coagulación humano introducido en el roedor; o

5 (V) la variable cambiada en el paso (II) es el reactivo orientado al factor XII de coagulación humano introducido en el roedor.

18. Un método para fabricar el roedor de una cualquiera de las reivindicaciones 1-11, que comprende:

(a) introducir en un embrión huésped de roedor una célula madre embrionaria (ES) de roedor modificada genéticamente que comprende en su genoma el locus *F12* endógeno humanizado; y

10 (b) gestar el embrión huésped de roedor en una madre roedor sustituta, en donde la madre roedor sustituta produce una progenie F0 de roedor modificado genéticamente que comprende el locus *F12* endógeno humanizado, en donde la progenie F0 de roedor modificado genéticamente expresa la isoforma corta del factor XII de coagulación humano en el cerebro.

15 19. Un método para fabricar el roedor de una cualquiera de las reivindicaciones 1-11, que comprende:

(a) introducir en una célula madre embrionaria (ES) de roedor un vector orientado que comprende un inserto de ácido nucleico que comprende la secuencia de *F12* humano flanqueada por un brazo de homología 5' correspondiente a una secuencia objetivo 5' en el locus *F12* endógeno y un brazo de homología 3' correspondiente a una secuencia objetivo 3' en el locus *F12* endógeno, en donde el vector orientado se recombina con el locus *F12* endógeno para producir una célula ES de roedor modificada genéticamente que comprende en su genoma el locus *F12* endógeno humanizado;

20 (b) introducir la célula ES de roedor modificada genéticamente en un embrión huésped de roedor; y

(c) gestar el embrión huésped de roedor en una madre roedor sustituta, en donde la madre roedor sustituta produce una progenie F0 de roedor modificado genéticamente que comprende en su genoma el locus *F12* endógeno humanizado, en donde la progenie F0 de roedor modificado genéticamente expresa la isoforma corta del factor XII de coagulación humano en el cerebro.

25 20. Un método para fabricar el roedor de una cualquiera de las reivindicaciones 1-11, que comprende:

(a) introducir en un embrión de etapa unicelular de roedor un vector orientado que comprende un inserto de ácido nucleico que comprende la secuencia de *F12* humano flanqueada por un brazo de homología 5' correspondiente a una secuencia objetivo 5' en el locus *F12* endógeno y un brazo de homología 3' correspondiente a una secuencia objetivo 3' en el locus *F12* endógeno,

30 en donde el vector orientado se recombina con el locus *F12* endógeno para producir un embrión de etapa unicelular de roedor genéticamente modificado que comprende en su genoma el locus *F12* endógeno humanizado; y

(b) gestar el embrión de etapa unicelular de roedor genéticamente modificado en una madre roedor sustituta para producir un roedor de generación F0 genéticamente modificado que comprende en su genoma el locus *F12* endógeno humanizado que comprende la secuencia de *F12* humano, en donde el roedor de generación F0 genéticamente modificado expresa la isoforma corta del factor XII de coagulación humana en el cerebro.

35 21. Un método para evaluar la expresión de una isoforma corta de ARNm de *F12* en una muestra biológica, que comprende:

(a) detectar y cuantificar transcripciones de ARN usando cebadores y/o sondas contra uno o más de los exones 9-14 de un gen *F12*; y

40 (b) detectar y cuantificar transcripciones de ARN usando cebadores y/o sondas contra uno o más de los exones 1-6 del gen *F12*

en donde la isoforma corta se detecta mediante cebadores y/o sondas contra los exones 9-14 pero no mediante cebadores y/o sondas contra los exones 1-6.

50

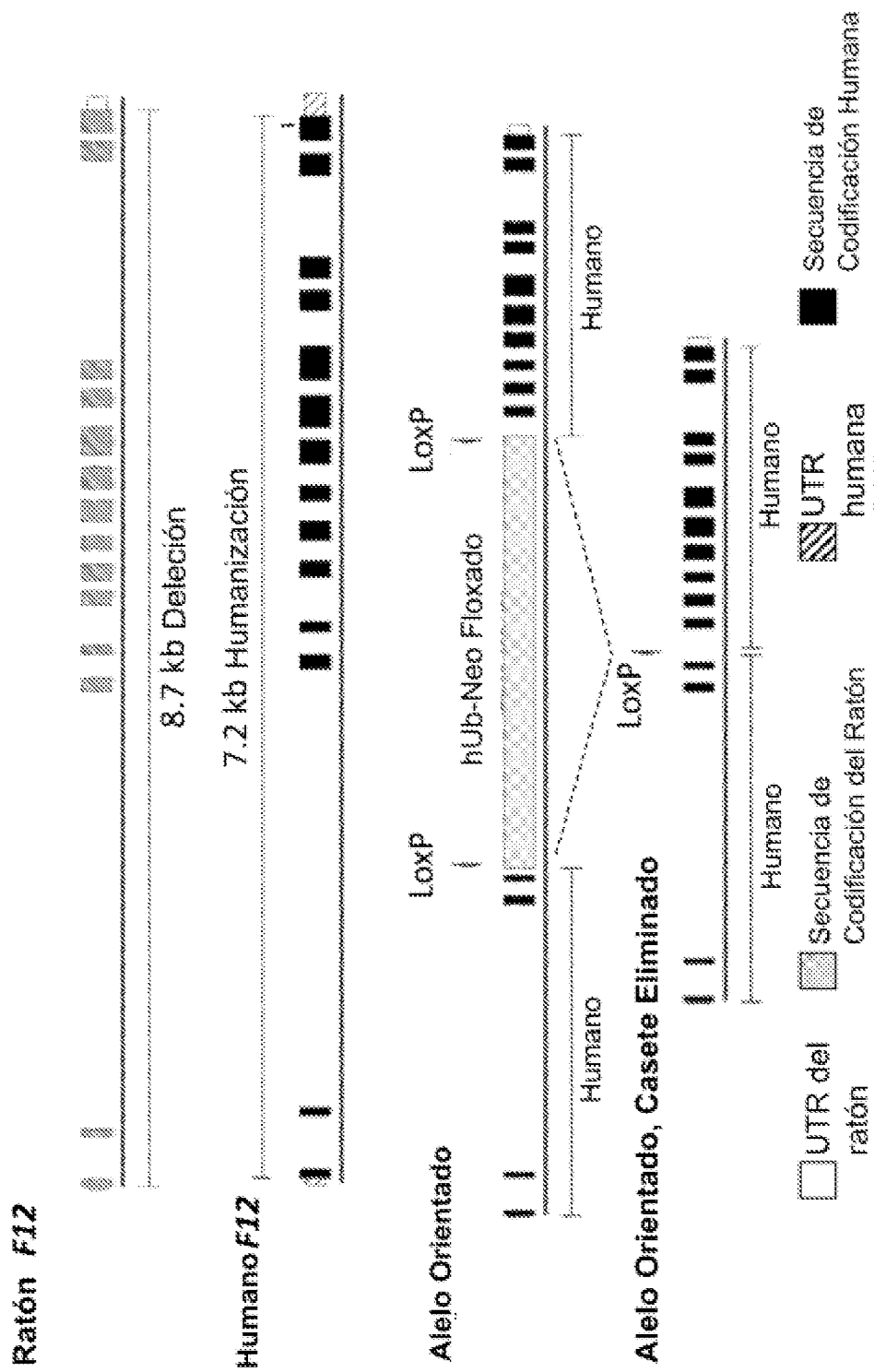


FIG. 1

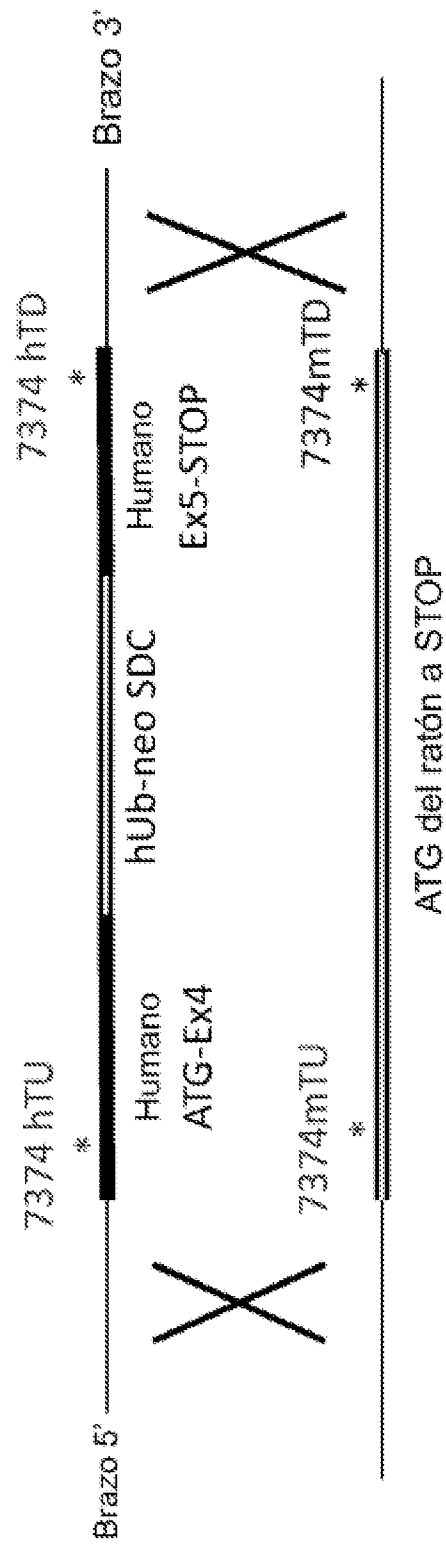


FIG. 2

hF12 MRALLLLGFLLVLSLESTLSIPPWEAPKEHKYKAEHHTVVLTVTGEPCHFPFQYHRQL
mF12 MTALLFLGSLMSLDLTLSAPEWKSCKKFKDAPDGTVVLTVDGRLCHEFPFQYHRQL
hF12 YHKCTHKGRPGPQPMWCATTNFDQDQRWGYCLEPKVKVDHCSKHSKSPCKGGTCVNMMP
mF12 HHKCIHKRRPGSRPWCATTNFEDEDDQGWGYCLEPKVKVDHCSKHNPCHKGTCINTP
hF12 SGPHCLCFQHLTGNHCQKEKCFEPQLLRFFHKNEIWYHTEQAAVARCOCKGPDARCO
mF12 NGPHCLCPHELTGKRCQKEKCFEPQLLKEFFHENEINWRTGPGGVARCECKGSEARCK
hF12 RLASQACRTNECLHGGRCLEVEGHRILCHCPVGYTGFECDVDTKASCYDGRGLSYRGL
mF12 PVASQACSNPCNLNGSCLLVEDHPLCRCPCTGYTGFECDLLDNATCYEGRGLSYRGO
hF12 ARTTISGAPCQFWASEATYRNVTAEQARNWGLCGHAFCRNPNDIRPNCFTVLNRDRL
mF12 AGTTQSGAPCQFWTVEATYRNMTEKQALSWGHLGHAFCRNPNDIRPWCFTVWSGDRRL
hF12 SWEYCDLAQCQTPTQAAPTFPVSPRLHVPILMPAQAPAPKPPQPTTRTPQSQTFCALP
mF12 SWDYCGLEQCQTPTFAPLVPES QEESPSQAPSLSHAPNDST
hF12 AKREQPPSLTRNGPLSCGQRLRKSLSSTHVVGGLVALRGHAHPYIAALYMGHSFCAG
mF12 _____DHQTSLSKTNMTMCGGQRFERKGLSSEFMVVGGLVALPGSHPYIAALYMGNNFCAG
hF12 SLIAPCWVLTAAHCLQDRPAPEDLTVVLQGERNHSCEPCQTLAVRSYRLHEAFSPV
mF12 SLIAPCWVLTAAHCLQNRPAPEELTVVLGQDRHNQSCENCQTLAVRSYRLHEGFSSI
hF12 SYQHDLALLRLQEDADGSCALLSPYVQVCLPSCAARPESETTLCQVAGWGHQFEGAE
mF12 TYQHDLALLRLQESKTNSCALLSPHVQVCLPSCAARPESETVLCVAGWGHQFEGAE
hF12 EYASFLEAQVPFLSLERCAPDVHGSSILPGMLCAGFLEGGTDACQDSSGGPLVCE
mF12 EYSTFLQEAQVPFIALDRCSNSNVHGDAIILPGMLCAGFLEGGTDACQDSSGGPLVCE
hF12 DQAAERRLTLQGIISWGS CGGDRNKPGVYTDVAYYLAWIREHTVS (SEQ ID NO: 5)
mF12 EGTAEHQLTLRGVISWGS CGGDRNKPGVYTDVANYLANTQKHIA (SEQ ID NO: 1)

FIG. 3

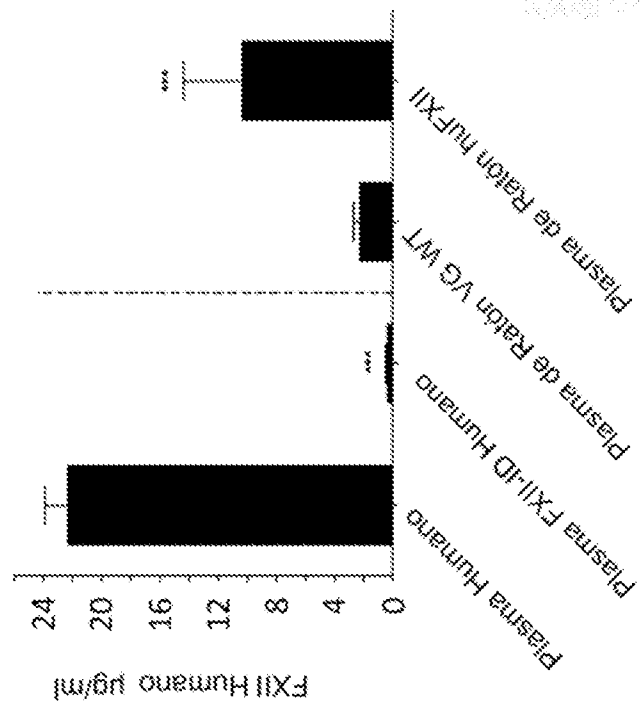


FIG. 4

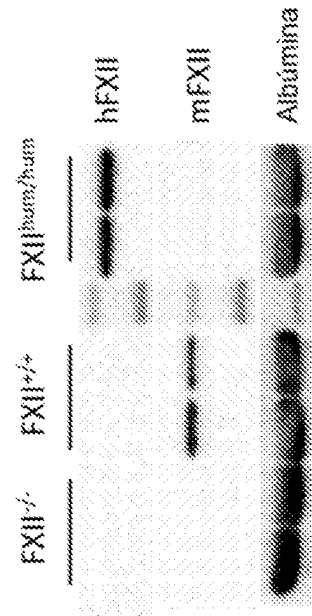


FIG. 5

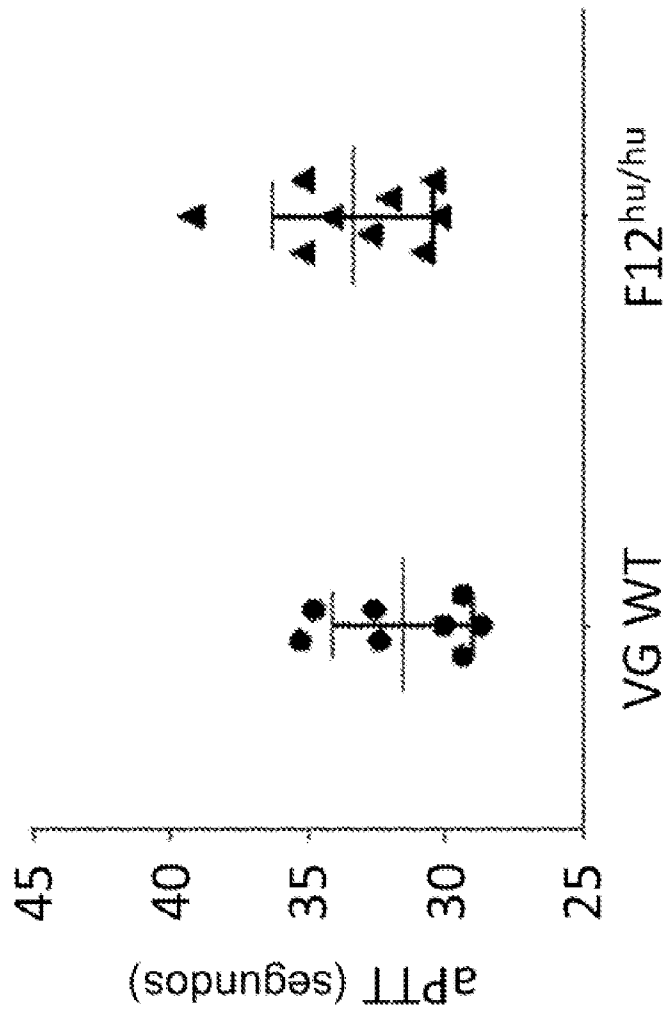


FIG. 6

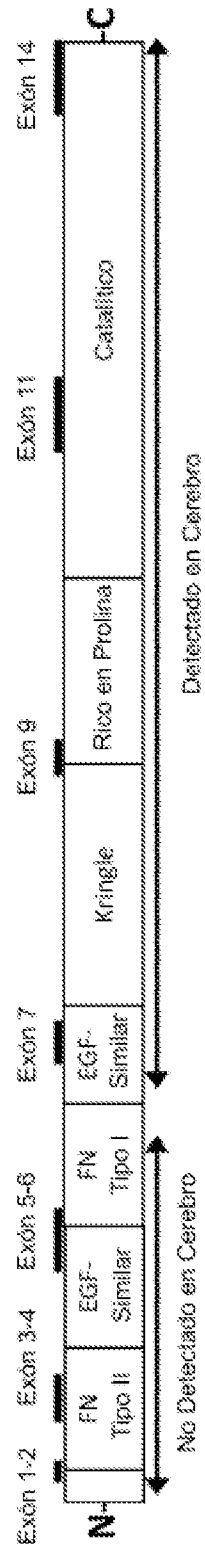


FIG. 7A

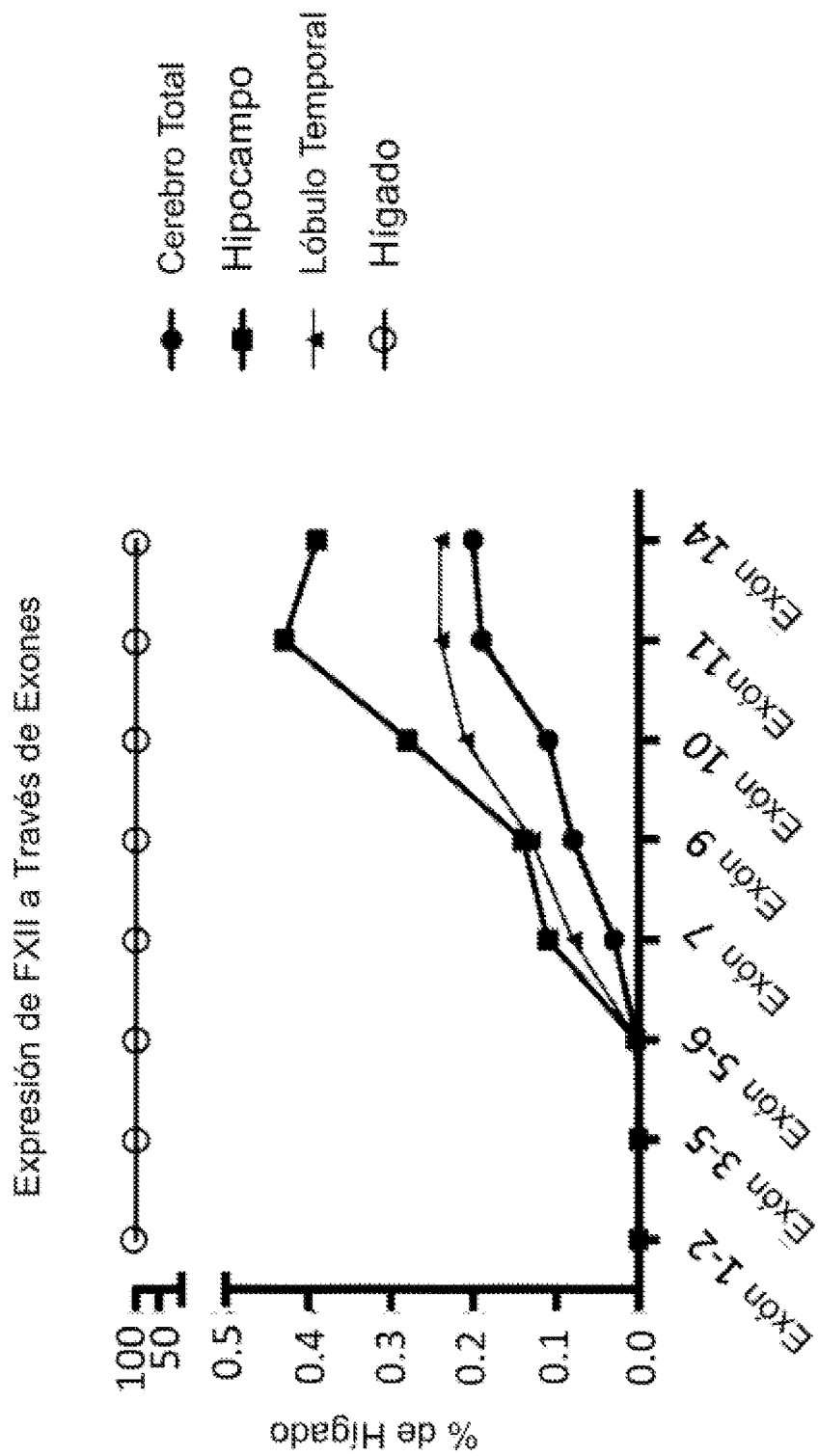


FIG. 7B

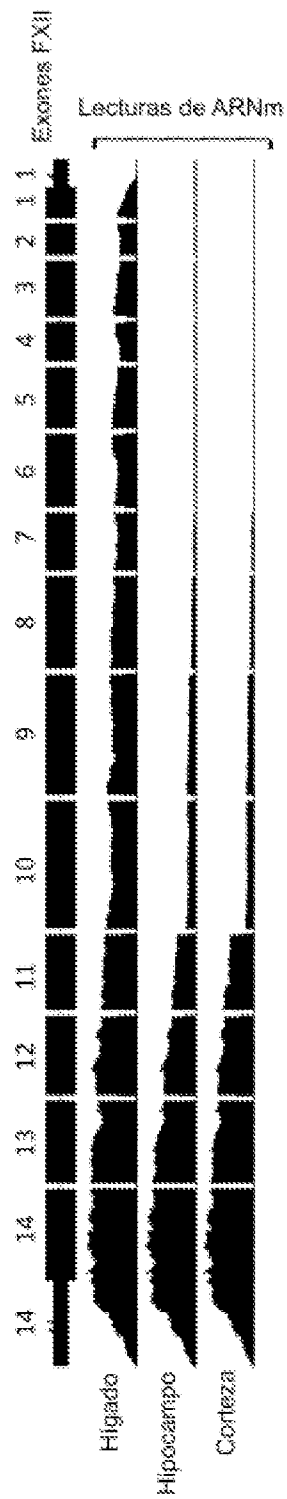


FIG. 7C

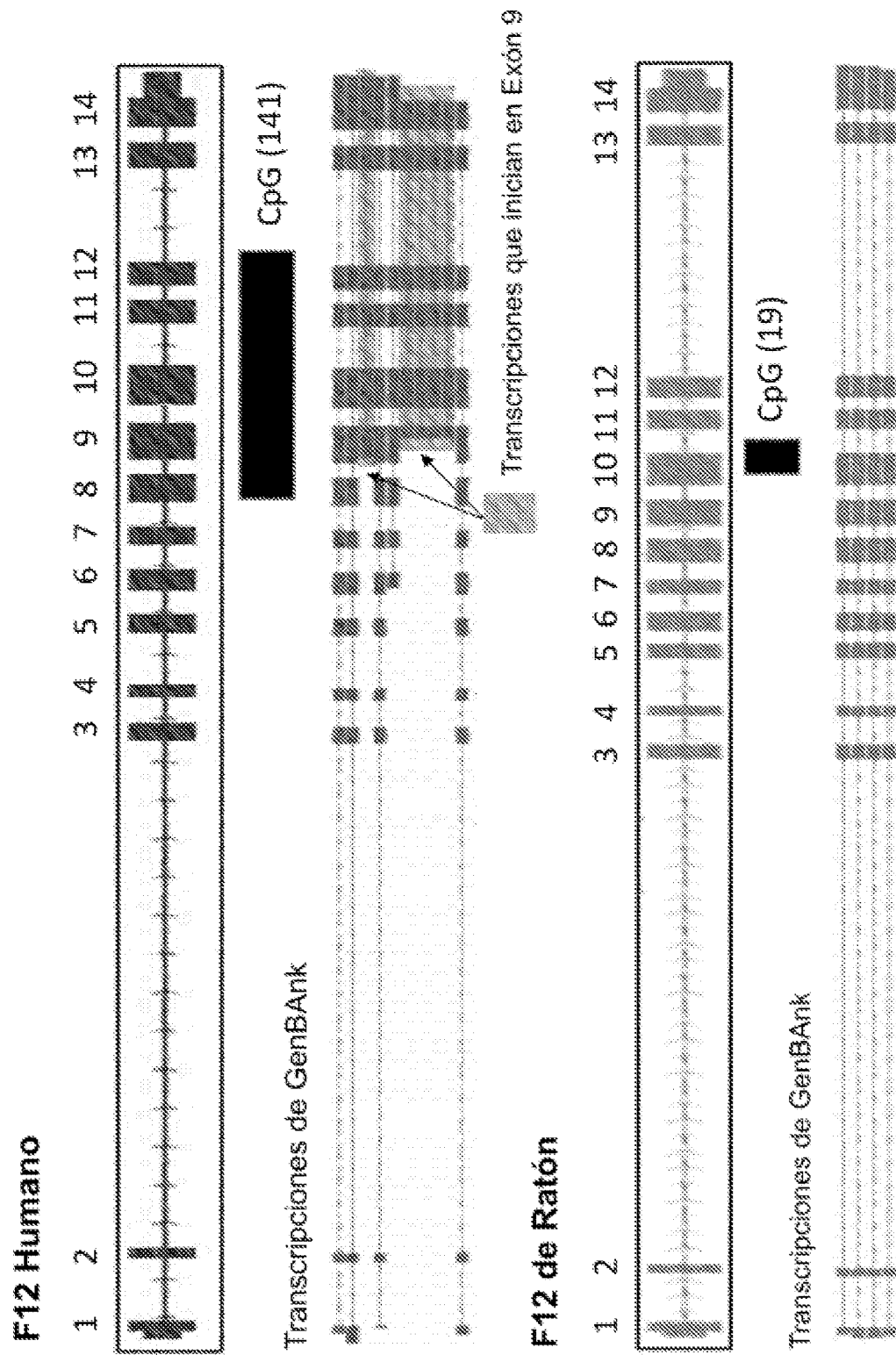


FIG. 8A

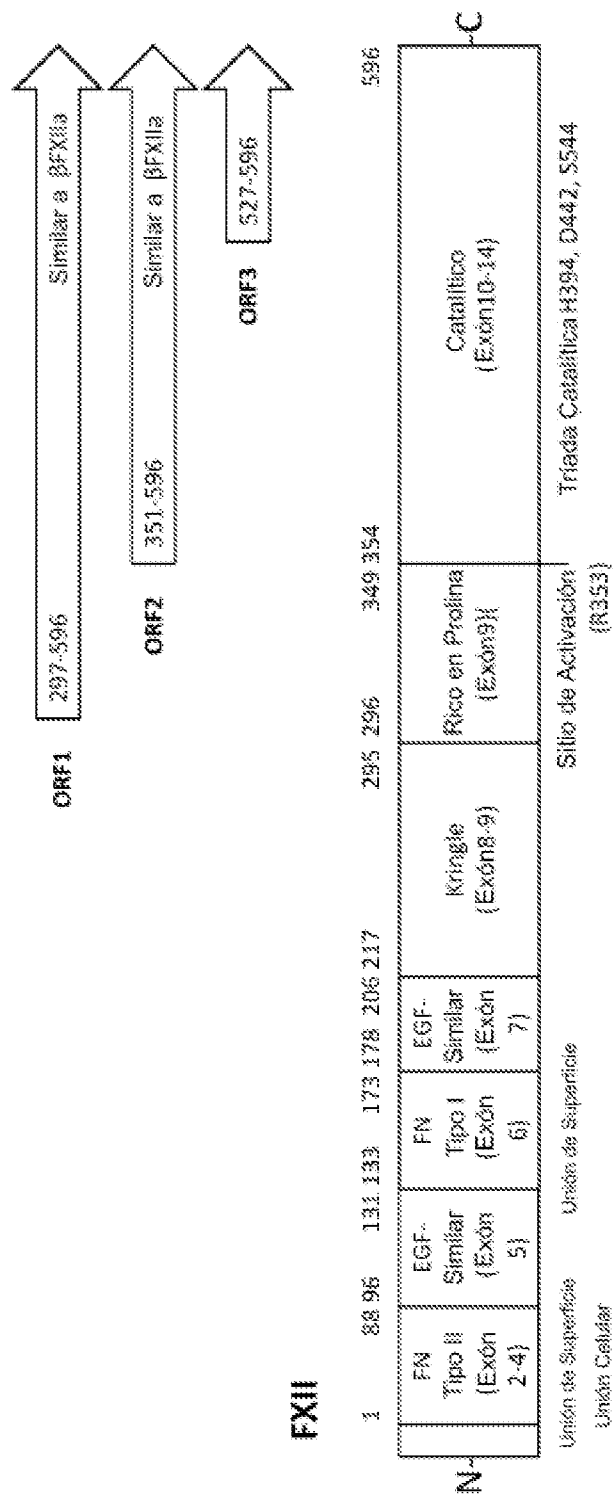


FIG. 8B

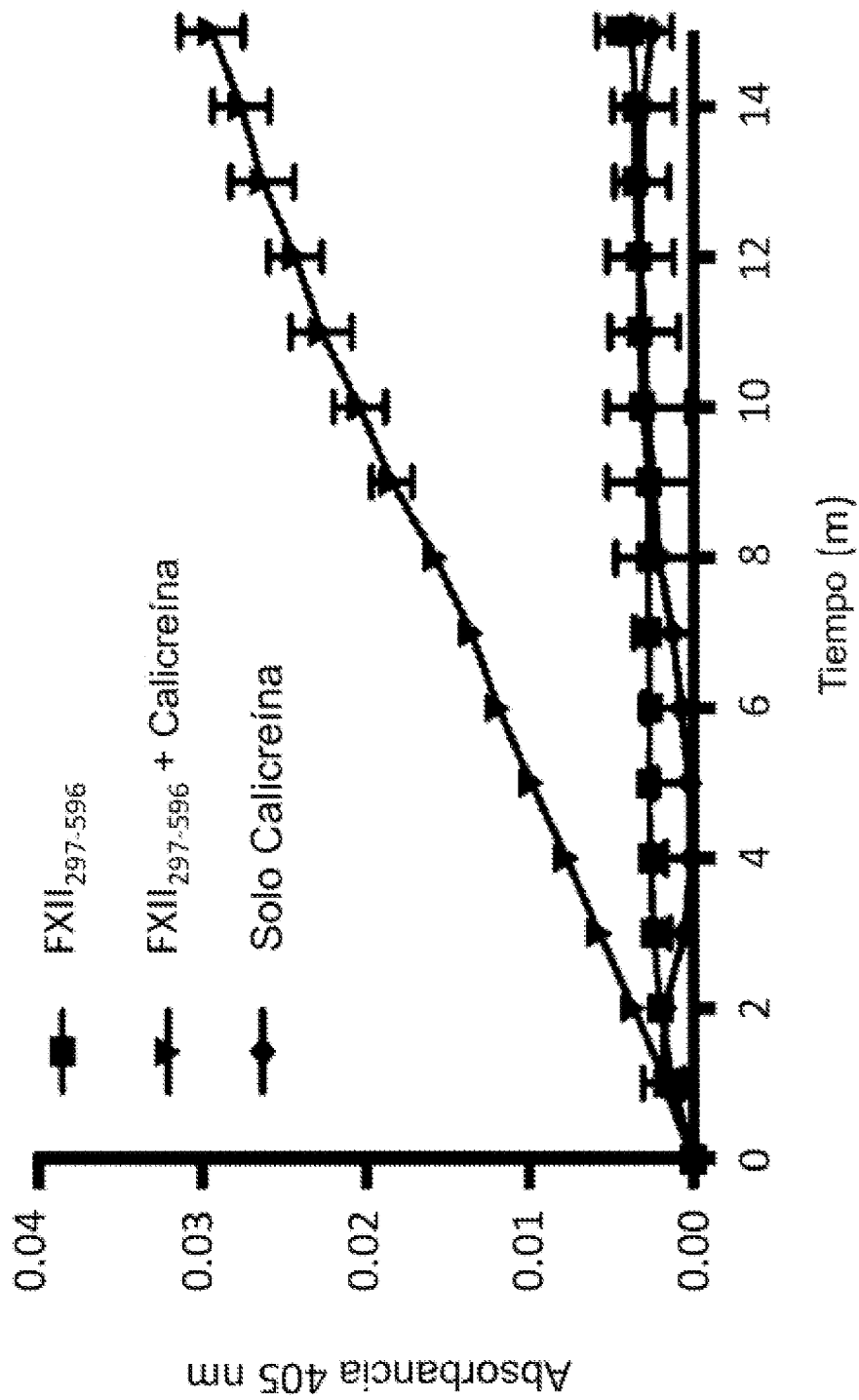


FIG. 9A

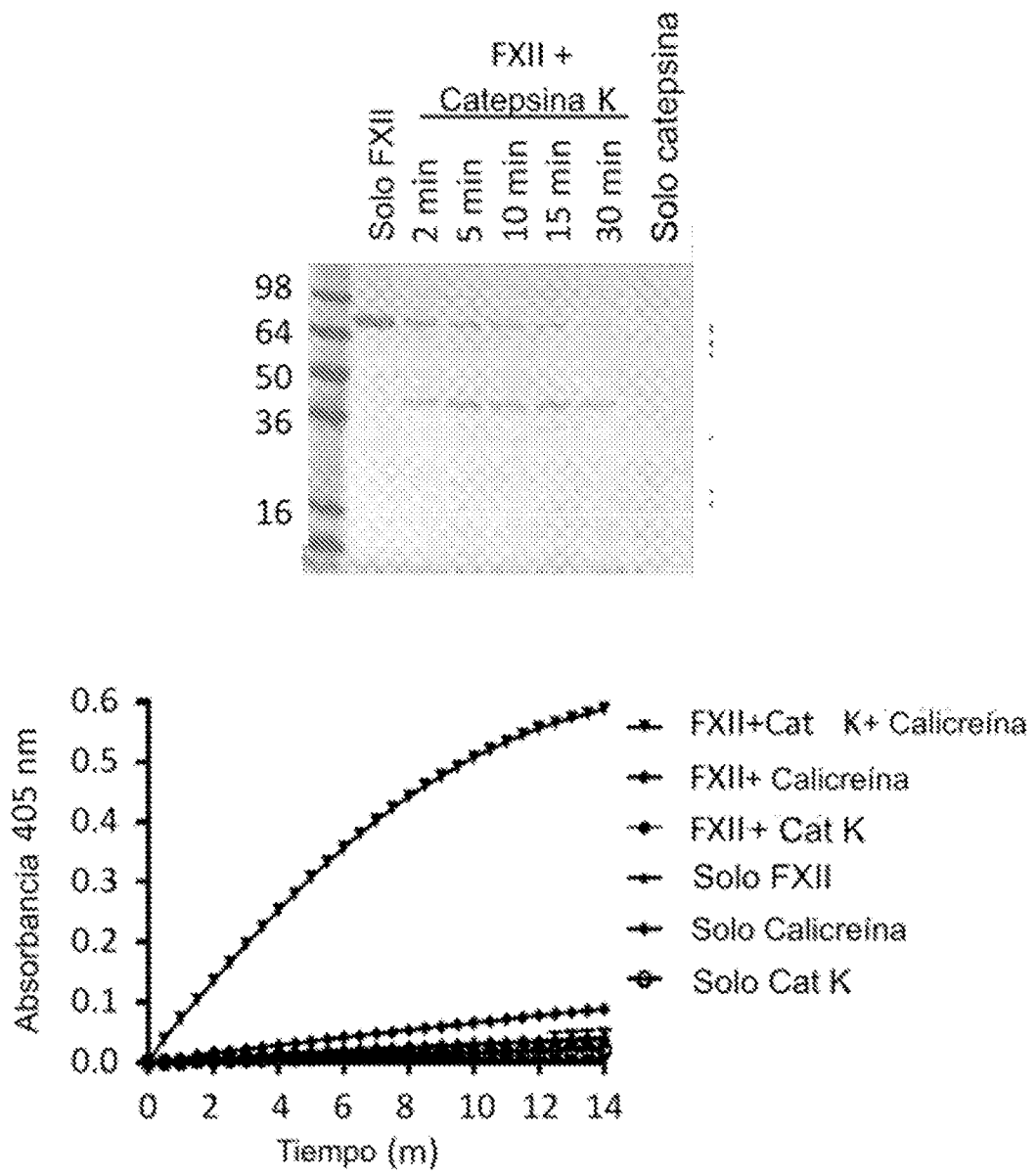


FIG. 9B

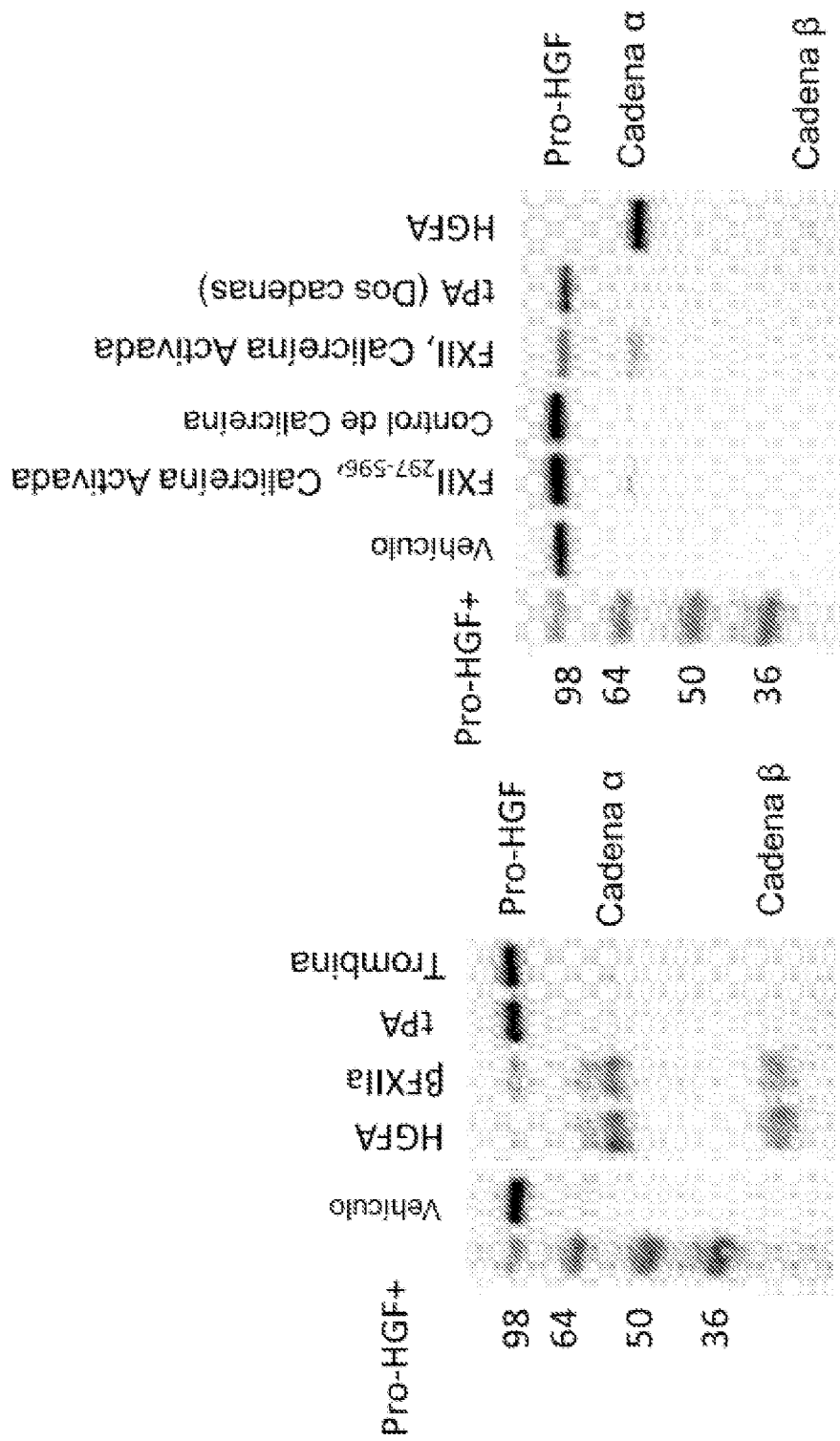
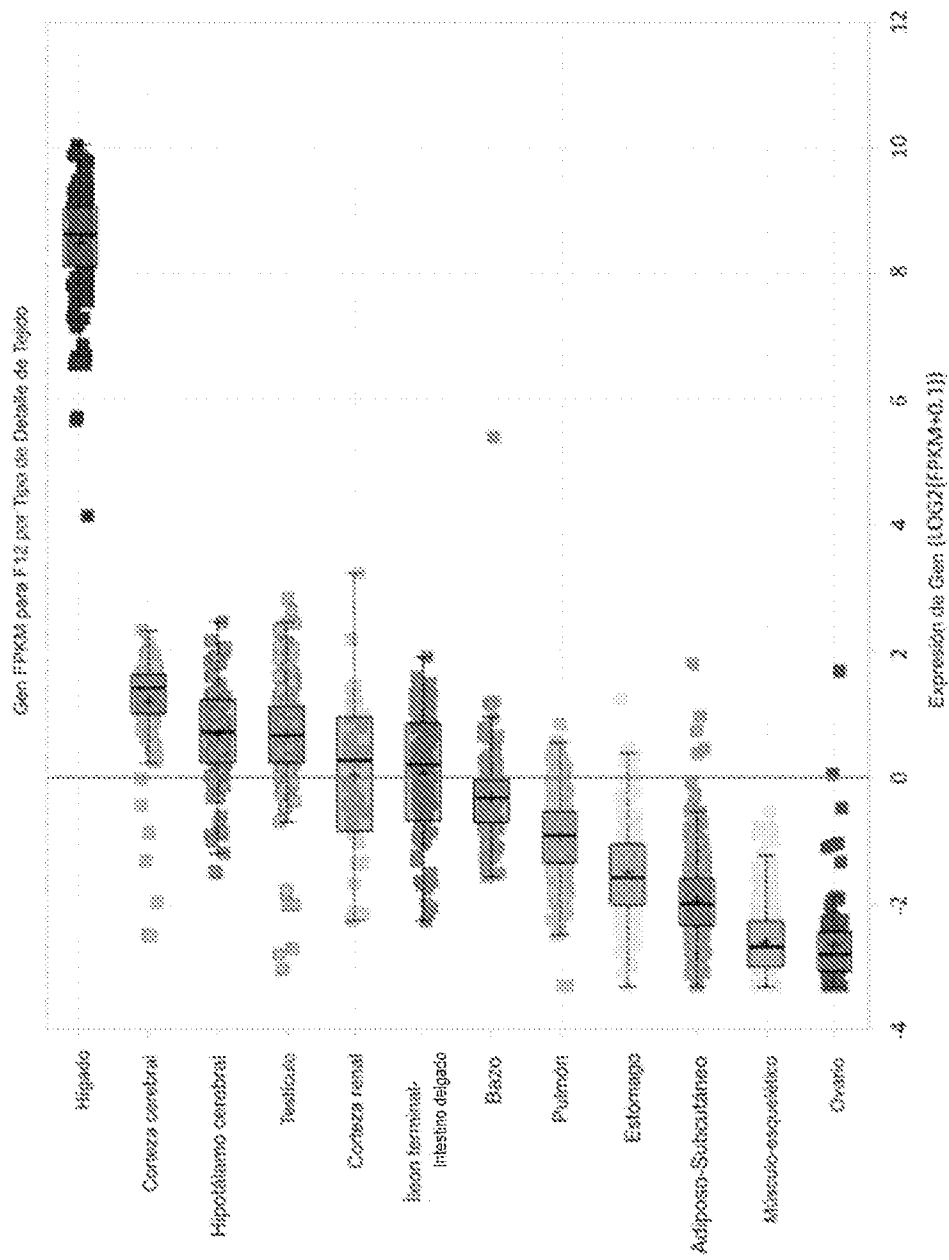
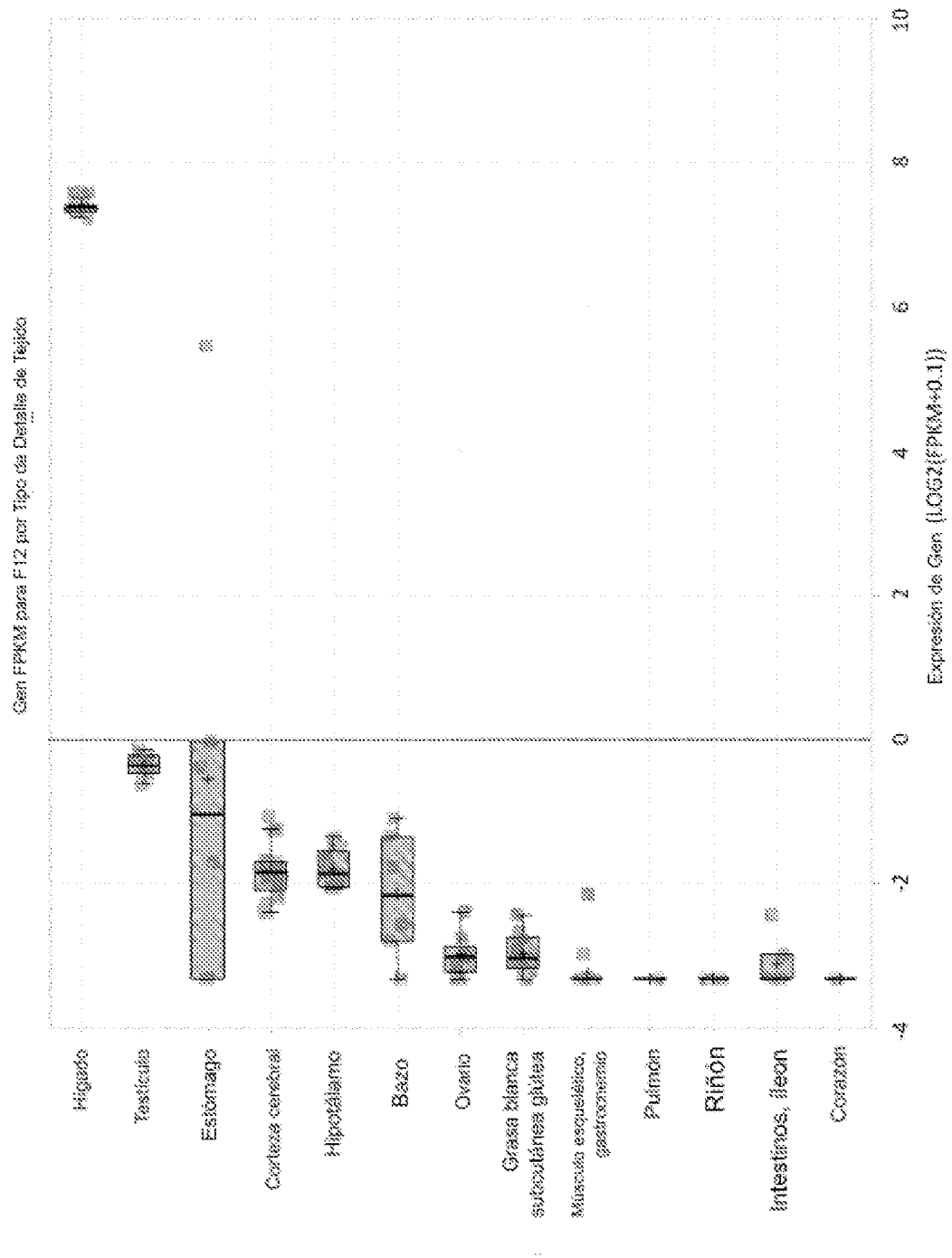


FIG. 9D

FIG. 9C





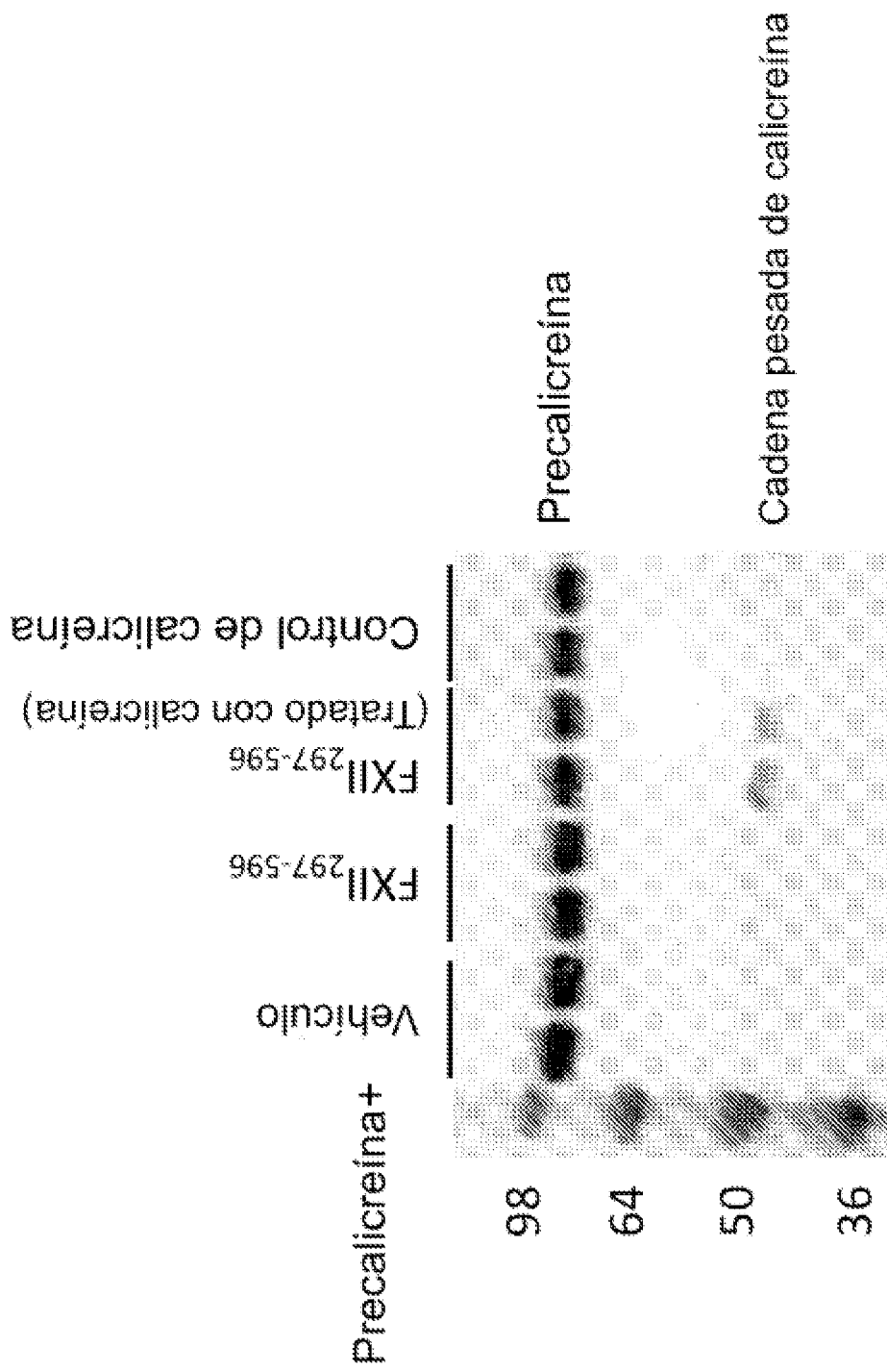


FIG. 11