

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 023 563**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12N 15/63 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **14.10.2019** **PCT/US2019/056083**

87 Fecha y número de publicación internacional: **23.04.2020** **WO20081438**

96 Fecha de presentación y número de la solicitud europea: **14.10.2019** **E 19795751 (7)**

97 Fecha y número de publicación de la concesión europea: **02.04.2025** **EP 3867372**

54 Título: **Métodos de inserción dirigida de ADN en genes**

30 Prioridad:

16.10.2018 US 201862746497 P

08.04.2019 US 201962830654 P

20.06.2019 US 201962864432 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
02.06.2025

73 Titular/es:

BLUEALLELE CORPORATION (100.00%)

988 Inwood Avenue N.

Oakdale MN 55128, US

72 Inventor/es:

BALTES, NICHOLAS

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

Observaciones:

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

ES 3 023 563 T3

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

DESCRIPCIÓN

Métodos de inserción dirigida de ADN en genes

5 CAMPO TÉCNICO

El presente documento se encuentra en el campo de la edición del genoma. Más específicamente, este documento se refiere a un transgén para la modificación dirigida de genes endógenos. Las referencias a métodos para el tratamiento del cuerpo humano o animal mediante cirugía o terapia deben interpretarse como referencias a los compuestos, composiciones farmacéuticas y medicamentos de la presente invención para su uso en tales métodos.

ANTECEDENTES

Los trastornos monogénicos están provocados por una o más mutaciones en un único gen, ejemplos de los cuales incluyen la anemia de células falciformes (gen de hemoglobina beta), la fibrosis quística (gen regulador de la conductancia transmembrana de la fibrosis quística) y la enfermedad de Tay-Sachs (gen de la beta-hexosaminidasa A). Los trastornos monogénicos han sido de interés para la terapia génica, ya que la sustitución del gen defectuoso por una copia funcional podría aportar beneficios terapéuticos. Sin embargo, uno de los obstáculos para generar terapias eficaces es el tamaño de la copia funcional del gen. Muchos métodos de administración, incluyendo aquellos que usan virus, tienen limitaciones de tamaño que dificultan la administración de transgenes grandes. Además, muchos genes tienen patrones alternativos de corte y empalme que dan como resultado un único gen que codifica múltiples proteínas. Los métodos para corregir regiones parciales de un gen defectuoso pueden proporcionar un medio alternativo para tratar trastornos monogénicos.

El documento WO 2015/089351 describe composiciones de CRISPR/Cas y métodos para su uso en trastornos de repetición de nucleótidos. Robert, 2017, *Current Genetics*, 64(2) describe trabajos que caracterizan terminadores bidireccionales en *Saccharomyces cerevisiae*. El documento WO 2016/109840 describe líneas celulares para la edición del genoma usando sistemas CRISPR/Cas. Ouyang et al., 2018 (*Stem Cells and Development*, 27(11)) describe la delección dirigida mediante CRISPR/Cas9 de poliglutamina en células madre.

SUMARIO

La edición de genes es prometedora para corregir mutaciones que se encuentran en trastornos genéticos; sin embargo, sigue habiendo muchos desafíos para crear terapias eficaces para trastornos individuales, incluyendo los provocados por mutaciones de ganancia de función o en los que se requiere una reparación precisa. Estos desafíos se observan en trastornos como la ataxia espinocerebelosa 3 y la ataxia espinocerebelosa 6, en los que el trastorno está provocado por mutaciones de ganancia de función (repetición trinucleotídica expandida) en el extremo 3' de los genes.

La invención proporciona un transgén que comprende:

una primera secuencia aceptora de corte y empalme y una segunda secuencia aceptora de corte y empalme situadas dentro del transgén en direcciones opuestas;
una primera secuencia codificante enlazada operativamente con la primera secuencia aceptora de corte y empalme y una segunda secuencia codificante enlazada operativamente con la segunda secuencia aceptora de corte y empalme, en donde la primera y la segunda secuencias codificantes se sitúan dentro del transgén en direcciones opuestas; y
un primer terminador enlazado operativamente con la primera secuencia codificante y un segundo terminador enlazado operativamente con la segunda secuencia codificante, en donde el primer y el segundo terminadores están situados dentro del transgén en direcciones opuestas, o un terminador bidireccional enlazado operativamente con la primera y la segunda secuencias codificantes y situado dentro del transgén;
en donde la primera secuencia aceptora de corte y empalme flanquea la primera secuencia codificante y el primer terminador o el terminador bidireccional, en donde la segunda secuencia aceptora de corte y empalme flanquea la segunda secuencia codificante y el segundo terminador o el terminador bidireccional, y
en donde la primera y la segunda secuencias codificantes son las mismas secuencias de ácidos nucleicos, o en donde la primera y la segunda secuencias codificantes son secuencias de ácidos nucleicos diferentes que codifican para la misma proteína usando la degeneración de codones.

Los métodos descritos en la presente proporcionan enfoques novedosos para corregir las mutaciones que se encuentran en el extremo 3' de los genes. La divulgación de la presente se basa, por lo menos en parte, en el diseño de transgenes bimodulares compatibles con la integración a través de múltiples vías de reparación. Los transgenes descritos en la presente pueden integrarse en genes por la vía de recombinación homóloga, la vía de unión de extremos no homólogos, o tanto la vía de recombinación homóloga como la de unión de extremos no homólogos, o mediante transposición. Además, el resultado de la integración en cualquier caso (HR, NHEJ directa, NHEJ inversa; transposición directa o transposición inversa) puede dar como resultado una corrección/alteración precisa del producto

proteico del gen diana. Los transgenes descritos en la presente pueden usarse para fijar o introducir mutaciones en la región 3' de genes de interés. Los métodos son particularmente útiles en casos en los que es necesaria una edición precisa de los genes, o cuando el gen endógeno mutado al que se apunta no puede ser "sustituido" por una copia sintética porque excede la capacidad de tamaño de los vectores estándar o los vectores virales. Los métodos descritos en la presente pueden usarse para la investigación aplicada (por ejemplo, la terapia génica) o la investigación básica (por ejemplo, la creación de modelos animales o la comprensión de la función génica).

Los métodos descritos en la presente son compatibles con los vehículos de administración in vivo actuales (por ejemplo, vectores de virus adenoasociados y nanopartículas lipídicas), y abordan varios desafíos a la hora de lograr una alteración precisa de los productos génicos.

En la presente se describe un método para integrar un transgén de la invención en un gen endógeno. El método puede incluir la administración de un transgén, donde el transgén comprende una primera secuencia aceptora de corte y empalme y una segunda secuencia aceptora de corte y empalme situadas dentro del transgén en direcciones opuestas; una primera secuencia codificante enlazada operativamente con la primera secuencia aceptora de corte y empalme y una segunda secuencia codificante enlazada operativamente con la segunda secuencia aceptora de corte y empalme, en donde la primera y la segunda secuencias codificantes están situadas dentro del transgén en direcciones opuestas; y un primer terminador enlazado operativamente con la primera secuencia codificante y un segundo terminador enlazado operativamente con la segunda secuencia codificante, en donde el primer y el segundo terminadores están situados dentro del transgén en direcciones opuestas, o un terminador bidireccional enlazado operativamente con la primera y la segunda secuencias codificantes y situado dentro del transgén; en donde la primera secuencia aceptora de corte y empalme flanquea la primera secuencia codificante y el primer terminador o el terminador bidireccional, en donde la segunda secuencia aceptora de corte y empalme flanquea la segunda secuencia codificante y el segundo terminador o el terminador bidireccional, y en donde la primera y la segunda secuencias codificantes son las mismas secuencias de ácido nucleico, o en donde la primera y la segunda secuencias codificantes son secuencias de ácido nucleico diferentes que codifican para la misma proteína usando la degeneración de codones. El método incluye además la administración de una o más endonucleasas de corte raras dirigidas a un sitio dentro del gen endógeno, donde el transgén se integra luego en el gen endógeno. El transgén puede dirigirse a un sitio dentro de un intrón o en una unión intrón-exón. La primera y la segunda secuencias codificantes parciales están en una orientación de cola a cola, de tal manera que la integración del transgén en cualquier dirección (es decir, directa o inversa) por NHEJ puede dar como resultado una alteración precisa del producto proteico del gen. En otras realizaciones, el transgén puede incluir un brazo de homología izquierdo y derecho para permitir la integración por HR. Estos transgenes pueden albergarse en un vector de virus adenoasociado (AAV), en el que el transgén puede integrarse mediante HR (a través de los brazos de homología) o mediante NHEJ de dirección directa o NHEJ de dirección inversa (a través de la integración directa del vector AAV en una rotura de cadena doble). En una realización, los vectores con una primera y una segunda secuencias codificantes y un brazo de homología izquierdo y derecho pueden incluir además un primer y un segundo sitio para la escisión por una o más endonucleasas de corte raro. La escisión por las una o más endonucleasas de corte raro puede dar como resultado la liberación de un transgén lineal con brazos de homología, capaz de integrarse en el genoma mediante HR o NHEJ. En otra realización, los vectores con una primera y una segunda secuencia codificante pueden estar flanqueados por un primer y un segundo sitio de corte por una o más endonucleasas de corte raro. La escisión por las una o más endonucleasas de corte raro puede dar como resultado la liberación de un transgén lineal, capaz de integrarse en el genoma mediante NHEJ. En otra realización, los vectores con una primera y una segunda secuencias codificantes pueden estar flanqueados por un extremo izquierdo y derecho de transposón. La administración de una transposasa asociada a CRISPR (por ejemplo, Cas6/7/8 junto con TniQ, TnsA, TnsB y TnsC) puede dar como resultado la integración del transgén mediante transposición.

El transgén puede usarse para alterar el extremo C-terminal de proteínas producidas por genes endógenos. En algunas realizaciones, el gen endógeno puede incluir el gen ATXN3 o el gen CACNA1A. El ATXN3 es un gen que codifica la enzima ataxina-3. La ataxina-3 es un miembro del sistema del proteasoma de ubiquitina que facilita la destrucción de proteínas en exceso o dañadas. La ataxia espinocerebelosa tipo 3 es un trastorno genético provocado por una expansión de repetición de trinucleótidos en el extremo 3' del gen ATXN3. El CACNA1A es un gen que codifica proteínas implicadas en la formación de canales de calcio. La ataxia espinocerebelosa tipo 6 es un trastorno genético provocado por mutaciones en el gen CACNA1A. Las mutaciones que provocan la SCA6 incluyen una expansión de repetición de trinucleótidos en el extremo 3' del gen CACNA1A. En algunas realizaciones, los transgenes proporcionados en la presente pueden usarse para alterar el extremo 3' del gen ATXN3 endógeno o del gen CACNA1A. En realizaciones específicas, la diana para la integración de los transgenes descritos en la presente puede ser el intrón 9 del gen ATXN3 o el intrón 46 del gen CACNA1A.

A menos que se defina de otro modo, todos los términos técnicos y científicos usados en la presente tienen el mismo significado que el entendido comúnmente por un experto en la técnica a la que pertenece esta invención.

DESCRIPCIÓN DE LOS DIBUJOS

La FIG. 1 es una ilustración de los transgenes para la inserción dirigida en genes endógenos. TS1, sitio diana 1;

SA1, sitio aceptor de corte y empalme 1, CDS1, secuencia codificante 1; T1, terminador 1, TS2, sitio diana 2; SA2, sitio aceptor de corte y empalme 2, CDS2, secuencia codificante 2; T2, terminador 2; HA1, brazo de homología 1; HA2, brazo de homología 2; BT1, terminador bidireccional 1; AS1, secuencia adicional 1; AS2, secuencia adicional 2.

La FIG. 2 es una ilustración que muestra la integración de un transgén en un gen ejemplar. El transgén comprende dos sitios diana para una o más endonucleasas de corte raro, dos secuencias aceptoras de corte y empalme, dos secuencias codificantes (3.1 y 3.2) y dos terminadores (T). La integración se realiza mediante la unión de extremos no homólogos (NHEJ).

La FIG. 3 es una ilustración que muestra la integración de un transgén en un gen ejemplar. El transgén comprende dos brazos de homología, dos sitios diana para una o más endonucleasas de corte raro, dos secuencias aceptoras de corte y empalme, dos secuencias codificantes (3.1 y 3.2) y dos terminadores. La integración se produce mediante recombinación homóloga (HR) o unión de extremos no homólogos (NHEJ).

La FIG. 4 es una ilustración del exón 46, el intrón 46 y el intrón 47 del gen CACNA1A. También se muestra el transgén pB1011-D1 para su integración en el gen CACNA1A.

La FIG. 5 es una ilustración de los resultados de la integración del transgén pB1011-D1 dentro del gen CACNA1A.

La FIG. 6 es una ilustración del exón 9, el intrón 9, el exón 10, el intrón 10 y el exón 11 del gen ATXN3. También se muestra el transgén pB1012-D1 para su integración en el gen ATXN3.

La FIG. 7 es una ilustración de los resultados de la integración del transgén pB1012-D1 dentro del gen ATXN3.

La FIG. 8 son imágenes de geles que detectan la integración de transgenes en el gen ATXN3. 1, escalera de 100 bp con banda superior de 1.517 bp; 2, unión 5' de pBA1135; 3, unión 5' de pBA1136; 4, unión 5' de pBA1137, 5, unión 3' de pBA1135; 6, unión 3' de pBA1136; 7, unión 3' de pBA1137; 8, escalera de 1kb con bandas más oscuras discurriendo en 500 bp, 1.000 bp y 3.000 pb; 9, escalera de 1kb con bandas más oscuras discurriendo en 500 bp, 1.000 bp y 3.000 pb; 10, unión 5' invertida de pBA1135; 11, escalera de 1kb con bandas más oscuras discurriendo en 500 bp, 1.000 bp y 3.000 pb; 12, unión 5' invertida de pBA1136; 13, escalera de 1kb con bandas más oscuras discurriendo en 500 bp, 1.000 bp y 3.000 pb; 14, par de cebadores oNJB156+oNJB113; 15, par de cebadores 114+162; 16, par de cebadores oNJB116+oNJB113; 17, par de cebadores oNJB14+oNJB170; 18, par de cebadores oNJB167+oNJB170; 19, escalera de 100 bp con la banda oscura discurriendo en 500 pb; 20, ADN genómico procedente de la transfección con pBA1135 y nucleasa; 21, ADN genómico de la transfección con pBA1136 y nucleasa; 22, ADN genómico de la transfección con pBA1137 y nucleasa; 23, ADN genómico de la transfección con agua; 24, sin control de ADN.

DESCRIPCIÓN DETALLADA

La invención proporciona un transgén que comprende:

una primera secuencia aceptora de corte y empalme y una segunda secuencia aceptora de corte y empalme situadas dentro del transgén en direcciones opuestas;

una primera secuencia codificante enlazada operativamente con la primera secuencia aceptora de corte y empalme y una segunda secuencia codificante enlazada operativamente con la segunda secuencia aceptora de corte y empalme, en donde la primera y la segunda secuencias codificantes están situadas dentro del transgén en direcciones opuestas; y

un primer terminador enlazado operativamente con la primera secuencia codificante y un segundo terminador enlazado operativamente con la segunda secuencia codificante, en donde el primer y segundo terminadores están situados dentro del transgén en direcciones opuestas, o un terminador bidireccional enlazado operativamente con la primera y la segunda secuencias codificantes y situado dentro del transgén;

en donde la primera secuencia aceptora de corte y empalme flanquea la primera secuencia codificante y el primer terminador o el terminador bidireccional, en donde la segunda secuencia aceptora de corte y empalme flanquea la segunda secuencia codificante y el segundo terminador o el terminador bidireccional, y

en donde la primera y la segunda secuencias codificantes son las mismas secuencias de ácido nucleico, o en donde la primera y la segunda secuencias codificantes son secuencias de ácido nucleico diferentes que codifican para la misma proteína usando la degeneración de codones.

En la presente se divulgan métodos y composiciones para modificar la secuencia codificante de genes endógenos. Los métodos incluyen insertar un transgén en un gen endógeno, en donde el transgén proporciona una secuencia codificante parcial que sustituye a la secuencia codificante del gen endógeno.

Este documento presenta un método de integración de un transgén en un gen endógeno, el método incluyendo administrar un transgén, en donde el transgén comprende una primera y una segunda secuencia aceptora

de corte y empalme, una primera y una segunda secuencia codificante parcial, y un terminador bidireccional o un primer y un segundo terminador, y administrar una o más endonucleasas de corte raro dirigidas a un sitio dentro del gen endógeno, en donde el transgén se integra dentro del gen endógeno. El método incluye diseñar el transgén para que tenga el primer aceptor de corte y empalme enlazado operativamente con la primera secuencia codificante parcial y el segundo aceptor de corte y empalme enlazado operativamente con la segunda secuencia codificante parcial. La disposición también incluye tener la primera secuencia codificante parcial enlazada operativamente con el primer terminador, y la segunda secuencia codificante parcial enlazada operativamente con el segundo terminador. Los transgenes con el primer y el segundo aceptores de corte y empalme, la primera y la segunda secuencias codificantes parciales, y el primer y el segundo terminadores están orientados en una orientación de cola a cola. Los transgenes con una orientación de secuencias de cola a cola pueden comprender además un primer y un segundo sitio diana para una o más endonucleasas de corte raro, en donde los sitios diana flanquean el primer y el segundo aceptores de corte y empalme. En otra realización, los transgenes pueden comprender un brazo de homología izquierdo y derecho que flanquean el primer y el segundo aceptores de corte y empalme. En esta realización, el transgén puede albergarse dentro de un vector viral adenoasociado. En otra realización, el transgén puede comprender además un primer y un segundo sitio diana para las una o más endonucleasas de corte raro, en donde los sitios diana flanquean el primer y el segundo aceptores de corte y empalme. El primer y el segundo sitios diana pueden flanquear el primer y el segundo brazos de homología. En realizaciones, los transgenes descritos en la presente pueden integrarse dentro de un intrón del gen endógeno o en una unión intrón-exón. Los transgenes pueden integrarse dentro de un intrón o en la unión intrón-exón del gen ATXN3 o del gen CACNA1A. El transgén puede comprender una primera y una segunda secuencias codificantes parciales que codifiquen el péptido producido por el exón 10 de un gen ATXN3 no patógeno y puede dirigirse al intrón 9, o a la unión intrón 9 exón 10, de un gen ATXN3 patógeno. El transgén puede comprender una primera y una segunda secuencias codificantes parciales que codifiquen el péptido producido por el exón 47 de un gen CACNA1A no patógeno y puede dirigirse al intrón 46, o a la unión intrón 46 exón 47, de un gen CACNA1A patógeno. En ciertas realizaciones, la endonucleasa de corte raro puede ser una nucleasa CRISPR/Cas12a o una nucleasa CRISPR/Cas9. La primera y la segunda secuencias codificantes parciales codifican los mismos aminoácidos. La primera y la segunda secuencias codificantes difieren en la secuencia de ácido nucleico pero codifican los mismos aminoácidos. El transgén puede albergarse en un vector, en donde el formato del vector se selecciona entre ADN lineal de cadena doble, ADN circular de cadena doble o un vector viral. El vector viral puede incluir un vector de adenovirus, un vector de virus adenoasociado o un vector de lentivirus. Los métodos descritos en la presente pueden usarse con un transgén igual o menor de 4,7 kb. El transgén puede comprender una primera y una segunda secuencias codificantes parciales que codifiquen un péptido parcial de una proteína funcional producida por el gen endógeno diana. El gen endógeno diana puede ser aberrante.

Este documento proporciona polinucleótidos de ADN con una primera y una segunda secuencias aceptoras de corte y empalme, una primera y una segunda secuencias codificantes parciales, un terminador bidireccional o un primer y un segundo terminadores, opcionalmente, un primer y un segundo brazos de homología y, opcionalmente, un primer y un segundo sitios diana de endonucleasa de corte raro. Los polinucleótidos de ADN incluyen un diseño que tiene el primer aceptor de corte y empalme enlazado operativamente a la primera secuencia codificante parcial y el segundo aceptor de corte y empalme enlazado operativamente a la segunda secuencia codificante. La disposición también incluye tener la primera secuencia codificante parcial enlazada operativamente con el primer terminador, y la segunda secuencia codificante parcial enlazada operativamente con el segundo terminador. Los polinucleótidos de ADN con el primer y el segundo aceptores de corte y empalme, la primera y la segunda secuencias codificantes, y el primer y el segundo terminadores están orientados en una orientación de cola a cola. Los polinucleótidos de ADN con una orientación de las secuencias de cola a cola pueden comprender además un primer y un segundo sitios diana para una o más endonucleasas de corte raro, en donde los sitios diana flanquean el primer y el segundo aceptores de corte y empalme. En otra realización, los polinucleótidos de ADN pueden comprender un brazo de homología izquierdo y derecho que flanquean el primer y el segundo aceptores de corte y empalme. En esta realización, el polinucleótido de ADN puede alojarse dentro de un vector viral adenoasociado. En otra realización, los polinucleótidos de ADN pueden comprender además un primer y un segundo sitios diana para una o más endonucleasas de corte raro, en donde los sitios diana flanquean el primer y el segundo aceptores de corte y empalme. El primer y el segundo sitios diana pueden flanquear el primer y el segundo brazos de homología. En algunas realizaciones, los polinucleótidos de ADN descritos en la presente pueden integrarse dentro de un intrón del gen endógeno o en una unión intrón-exón. Los polinucleótidos de ADN pueden integrarse dentro de un intrón, o en la unión intrón-exón del gen ATXN3 o del gen CACNA1A. El polinucleótido de ADN puede comprender una primera y una segunda secuencias codificantes parciales que codifiquen el péptido producido por el exón 10 de un gen ATXN3 no patógeno. El polinucleótido de ADN puede comprender una primera y una segunda secuencias codificantes parciales que codifiquen el péptido producido por el exón 47 de un gen CACNA1A no patógeno. La primera y la segunda secuencias codificantes parciales codifican los mismos aminoácidos. La primera y la segunda secuencias codificantes difieren en la secuencia de ácido nucleico pero codifican los mismos aminoácidos. Los polinucleótidos de ADN pueden alojarse en un vector, en donde el formato del vector se selecciona entre ADN lineal de cadena doble, ADN circular de cadena doble o un vector viral. El vector viral puede seleccionarse entre un vector de adenovirus, un vector de virus adenoasociado o un vector de lentivirus. Los polinucleótidos de ADN descritos en la presente pueden ser iguales o menores de 4,7 kb.

Este documento presenta un método de integración de un transgén en un gen endógeno, el método incluyendo administrar un transgén, en donde el transgén comprende un extremo de transposón izquierdo y derecho,

una primera y una segunda secuencias aceptoras de corte y empalme, una primera y una segunda secuencias codificantes parciales, y un terminador bidireccional o un primer y un segundo terminadores, y administrar una transposasa dirigida al gen endógeno, en donde el transgén se integra en el gen endógeno. El método incluye diseñar el transgén para que tenga el primer aceptor de corte y empalme enlazado operativamente con la primera secuencia codificante parcial y el segundo aceptor de corte y empalme enlazado operativamente con la segunda secuencia codificante. La disposición también incluye tener la primera secuencia codificante parcial enlazada operativamente con el primer terminador, y la segunda secuencia codificante parcial enlazada operativamente con el segundo terminador. Los transgenes con el primer y el segundo aceptores de corte y empalme, la primera y la segunda secuencias codificantes, y el primer y el segundo terminadores están orientados en una orientación de cola a cola. Los transgenes con una orientación de secuencias de cola a cola pueden comprender además un extremo de transposón izquierdo y derecho que flanquea el primer y el segundo aceptores de corte y empalme. En realizaciones, los transgenes descritos en la presente pueden integrarse dentro de un intrón del gen endógeno o en una unión intrón-exón. Los transgenes pueden integrarse dentro de un intrón o en la unión intrón-exón del gen ATXN3 o del gen CACNA1A. El transgén puede comprender una primera y una segunda secuencias codificantes parciales que codifiquen el péptido producido por el exón 10 de un gen ATXN3 no patógeno y puede dirigirse al intrón 9, o a la unión intrón 9 exón 10, de un gen ATXN3 patógeno. El transgén puede comprender una primera y una segunda secuencias codificantes parciales que codifican el péptido producido por el exón 47 de un gen CACNA1A no patógeno y puede dirigirse al intrón 46, o a la unión intrón 46 exón 47, de un gen CACNA1A patógeno. La transposasa puede ser una transposasa CRISPR, donde la transposasa CRISPR comprende la proteína Cas12k o Cas6. La primera y la segunda secuencias codificantes parciales codifican los mismos aminoácidos. La primera y la segunda secuencias codificantes difieren en la secuencia de ácido nucleico pero codifican los mismos aminoácidos. El transgén puede estar albergado en un vector, en donde el formato del vector se selecciona entre ADN lineal de cadena doble, ADN circular de cadena doble o un vector viral. El vector viral puede incluir un vector de adenovirus, un vector de virus adenoasociado o un vector de lentivirus. Los métodos descritos en la presente pueden usarse con un transgén igual o menor de 4,7 kb. El extremo izquierdo puede comprender la secuencia mostrada en la SEQ ID NO:41, y el extremo derecho puede comprender la secuencia mostrada en la SEQ ID NO: 13.

Este documento proporciona polinucleótidos de ADN con una primera y una segunda secuencias aceptoras de corte y empalme, una primera y una segunda secuencias codificantes parciales, un terminador bidireccional o un primer y un segundo terminadores, y un extremo izquierdo y derecho de transposón. Los polinucleótidos de ADN incluyen un diseño que tiene el primer aceptor de corte y empalme enlazado operativamente a la primera secuencia codificante parcial y el segundo aceptor de corte y empalme enlazado operativamente a la segunda secuencia codificante. La disposición también incluye tener la primera secuencia codificante parcial enlazada operativamente con el primer terminador, y la segunda secuencia codificante parcial enlazada operativamente con el segundo terminador. Los polinucleótidos de ADN con el primer y el segundo aceptores de corte y empalme, la primera y la segunda secuencias codificantes, y el primer y el segundo terminadores están orientados en una orientación de cola a cola. Los polinucleótidos de ADN con una orientación de secuencias de cola a cola pueden comprender además un extremo de transposón izquierdo y derecho que flanquean el primer y el segundo aceptores de corte y empalme. En realizaciones, los polinucleótidos de ADN descritos en la presente pueden integrarse dentro de un intrón del gen endógeno o en una unión intrón-exón. Los polinucleótidos de ADN pueden integrarse dentro de un intrón, o en la unión intrón-exón del gen ATXN3 o del gen CACNA1A. El polinucleótido de ADN puede comprender una primera y una segunda secuencias codificantes parciales que codifica el péptido producido por el exón 10 de un gen ATXN3 no patógeno. El polinucleótido de ADN puede comprender una primera y una segunda secuencias codificantes parciales que codifican el péptido producido por el exón 47 de un gen CACNA1A no patógeno. La primera y la segunda secuencias codificantes parciales codifican los mismos aminoácidos. La primera y la segunda secuencias codificantes difieren en la secuencia de ácido nucleico pero codifican los mismos aminoácidos. Los polinucleótidos de ADN pueden estar alojados en un vector, en donde el formato del vector se selecciona entre ADN lineal de cadena doble, ADN circular de cadena doble o un vector viral. El vector viral puede seleccionarse entre un vector de adenovirus, un vector de virus adenoasociado o un vector de lentivirus. Los polinucleótidos de ADN descritos en la presente pueden ser iguales o menores de 4,7 kb. El extremo izquierdo puede comprender la secuencia mostrada en la SEQ ID NO:41, y el extremo derecho puede comprender la secuencia mostrada en la SEQ ID NO: 13.

Este documento presenta un método para integrar un transgén en un gen endógeno, el método incluyendo administrar un transgén, en donde el transgén comprende una primera y una segunda secuencias aceptoras de corte y empalme, una primera y una segunda secuencias codificantes, un terminador bidireccional o un primer y un segundo terminadores, y un primer y un segundo brazos de homología, en donde el transgén se integra dentro del gen endógeno. El método incluye diseñar el transgén para que tenga el primer aceptor de corte y empalme enlazado operativamente con la primera secuencia codificante parcial y el segundo aceptor de corte y empalme enlazado operativamente con la segunda secuencia codificante. La disposición también incluye tener la primera secuencia codificante parcial enlazada operativamente con el primer terminador, y la segunda secuencia codificante parcial enlazada operativamente con el segundo terminador. Los brazos de homología pueden flanquear la primera y la segunda secuencias aceptoras de corte y empalme, la primera y la segunda secuencias codificantes, y el primer y el segundo terminadores. La secuencia codificante puede codificar una secuencia codificante completa o una secuencia codificante parcial. Los transgenes con el primer y el segundo aceptores de corte y empalme, la primera y la segunda secuencias codificantes, y el primer y el segundo terminadores pueden orientarse en una orientación de cola a cola.

Los transgenes con una orientación de secuencias de cola a cola pueden comprender además un primer y un segundo sitios diana para una o más endonucleasas de corte raro, en donde los sitios diana flanquean el primer y el segundo aceptores de corte y empalme. En otra realización, los transgenes pueden comprender un brazo de homología izquierdo y uno derecho que flanquean el primer y el segundo aceptores de corte y empalme. En esta realización, el transgén puede estar albergado dentro de un vector viral adenoasociado. En otra realización, el transgén puede comprender además un primer y un segundo sitios diana para las una o más endonucleasas de corte raro, en donde los sitios diana flanquean el primer y el segundo aceptores de corte y empalme. El primer y el segundo sitios diana pueden flanquear el primer y el segundo brazos de homología. En realizaciones, los transgenes descritos en la presente pueden integrarse dentro de un intrón del gen endógeno o en una unión intrón-exón.

La puesta práctica de los métodos, así como la preparación y el uso de las composiciones divulgadas en la presente emplean, a menos que se indique lo contrario, técnicas convencionales de biología molecular, bioquímica, estructura y análisis de la cromatina, química computacional, cultivo celular, ADN recombinante y campos relacionados, como están dentro de la capacidad de la técnica. Estas técnicas se explican detalladamente en la bibliografía. Véase, por ejemplo, Sambrook et al. *MOLECULAR CLONING: A LABORATORY MANUAL*, Segunda edición, Cold Spring Harbor Laboratory Press, 1989 y Tercera edición, 2001; Ausubel et al, *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, John Wiley & Sons, Nueva York, 1987 y actualizaciones periódicas; la serie *METHODS IN ENZYMOLOGY*, Academic Press, San Diego; Wolffe, *CHROMATIN STRUCTURE AND FUNCTION*, Tercera edición, Academic Press, San Diego, 1998; *METHODS IN ENZYMOLOGY*, Vol. 304, "Chromatin" (P. M. Wassarman y A P. Wolffe, eds.), Academic Press, San Diego, 1999; y *METHODS IN MOLECULAR BIOLOGY*, Vol. 119, "Chromatin Protocols" (P. B. Becker, ed.) Humana Press, Totowa, 1999.

Como se usa en la presente, los términos "ácido nucleico" y "polinucleótido" pueden usarse indistintamente. Ácido nucleico y polinucleótido pueden referirse a un polímero de desoxirribonucleótido o ribonucleótido, en conformación lineal o circular, y en forma de cadena sencilla o de cadena doble. Estos términos no deben interpretarse como limitativos con respecto a la longitud de un polímero. Los términos pueden abarcar análogos conocidos de nucleótidos naturales, así como nucleótidos modificados en la fracción de base, azúcar y/o fosfato.

Los términos "polipéptido", "péptido" y "proteína" pueden usarse indistintamente para referirse a residuos de aminoácidos enlazados covalentemente entre sí. El término también se aplica a las proteínas en las que uno o más aminoácidos son análogos químicos o derivados modificados de los aminoácidos de origen natural correspondientes.

Los términos "enlazado operativamente" o "enlazado operablemente" se usan indistintamente y se refieren a una yuxtaposición de dos o más componentes (como elementos de secuencia), en la que los componentes están dispuestos de tal manera que ambos componentes funcionan normalmente y permiten la posibilidad de que por lo menos uno de los componentes pueda mediar en una función que se ejerce sobre por lo menos uno de los otros componentes. A modo de ilustración, una secuencia reguladora transcripcional, como un promotor, está enlazada operativamente con una secuencia codificante si la secuencia reguladora transcripcional controla el nivel de transcripción de la secuencia codificante en respuesta a la presencia o ausencia de uno o más factores reguladores transcripcionales. Por lo general, una secuencia reguladora transcripcional está enlazada operativamente en cis con una secuencia codificante, pero no es necesario que sea directamente adyacente a ella. Por ejemplo, un potenciador es una secuencia reguladora transcripcional que está enlazada operativamente con una secuencia codificante, incluso aunque no sean contiguas. Además, a modo de ejemplo, un aceptor de corte y empalme puede estar enlazado operativamente con una secuencia codificante parcial si el aceptor de corte y empalme permite la delineación del límite 3' de un intrón, y si la traducción del ARNm maduro resultante da como resultado la incorporación de la secuencia peptídica codificada por la secuencia codificante parcial en el producto proteico final.

Como se usa en la presente, el término "escisión" se refiere a la ruptura de la estructura principal covalente de una molécula de ácido nucleico. La escisión puede iniciarse mediante una variedad de métodos que incluyen, pero no se limitan a, la hidrólisis enzimática o química de un enlace fosfodiéster. La escisión puede referirse tanto a un corte de cadena sencilla como a una rotura de cadena doble. Una rotura de cadena doble puede producirse como resultado de dos cortes de cadena sencilla distintos. La escisión del ácido nucleico puede dar como resultado la producción de extremos romos o extremos escalonados. En ciertas realizaciones, se usan endonucleasas de corte raro para la escisión dirigida de ADN cadena doble o de cadena sencilla.

Una molécula "exógena" puede referirse a una molécula pequeña (por ejemplo, azúcares, lípidos, aminoácidos, ácidos grasos, compuestos fenólicos, alcaloides), o una macromolécula (por ejemplo, proteína, ácido nucleico, carbohidrato, lípido, glicoproteína, lipoproteína, polisacárido), o cualquier derivado modificado de las moléculas anteriores, o cualquier complejo que comprenda una o más de las moléculas anteriores, generado o presente fuera de una célula, o que no esté normalmente presente en una célula. Las moléculas exógenas pueden introducirse en las células. Los métodos para la introducción o "administración" de moléculas exógenas en las células pueden incluir la transferencia mediada por lípidos, la electroporación, la inyección directa, la fusión celular, el bombardeo de partículas, la coprecipitación de fosfato cálcico, transferencia mediada por DEAE-dextrano y transferencia mediada por vectores virales. Como se define en la presente, "administrar" puede referirse a la entrega, la provisión o la introducción de moléculas exógenas en una célula. Si se administra un transgén o una endonucleasa

de corte raro a una célula, entonces el transgén o la endonucleasa de corte raro se entrega, proporciona o introduce en la célula. La endonucleasa de corte raro puede administrarse como proteína purificada, ácido nucleico o una mezcla de proteína purificada y ácido nucleico. El ácido nucleico (es decir, ARN o ADN), puede codificar para la endonucleasa de corte raro, o una parte de una endonucleasa de corte raro (por ejemplo, un ARNg). La administración puede lograrse mediante métodos como la transferencia mediada por lípidos, la electroporación, la inyección directa, la fusión celular, el bombardeo de partículas, la coprecipitación de fosfato cálcico, la transferencia mediada por DEAE-dextrano, la transferencia mediada por vectores virales, o cualquier medio adecuado para administrar proteínas o ácidos nucleicos purificados, o una mezcla de proteínas y ácidos nucleicos purificados, a una célula.

Una molécula "endógena" es una molécula que está presente en una célula particular en una etapa de desarrollo particular bajo condiciones ambientales particulares. Una molécula endógena puede ser un ácido nucleico, un cromosoma, el genoma de una mitocondria, cloroplasto u otro orgánulo, o un ácido nucleico episomal natural. Moléculas endógenas adicionales pueden incluir proteínas, por ejemplo, factores de transcripción y enzimas.

Como se usa en la presente, un "gen" se refiere a una región de ADN que codifica un producto génico, incluyendo todas las regiones de ADN que regulan la producción del producto génico. Por consiguiente, un gen incluye, pero no se limita necesariamente a, secuencias promotoras, terminadoras, secuencias reguladoras de la traducción, como sitios de unión a ribosomas y sitios internos de entrada de ribosomas, potenciadores, silenciadores, aisladores, elementos de frontera, orígenes de replicación, sitios de unión a la matriz y regiones de control de locus. Como se usa en la presente, un "gen de tipo salvaje" se refiere a una forma del gen que está presente en una población particular con la frecuencia más alta.

Un "gen endógeno" se refiere a una región de ADN presente normalmente en una célula particular que codifica un producto génico, así como a todas las regiones de ADN que regulan la producción del producto génico.

"Expresión génica" se refiere a la conversión de la información, contenida en un gen, en un producto génico. Un producto génico puede ser el producto transcripcional directo de un gen. Por ejemplo, el producto génico puede ser, pero no se limita a, ARNm, ARNt, ARNr, ARN antisentido, ribozima, ARN estructural o una proteína producida por traducción de un ARNm. Los productos génicos también incluyen ARN que se hayan modificado mediante procesos como la adición de caperuza, la poliadenilación, la metilación y la edición, y proteínas modificadas, por ejemplo, por metilación, acetilación, fosforilación, ubiquitinación, ADP-ribosilación, miristilación y glicosilación.

"Codificación" se refiere a la conversión de la información contenida dentro de un ácido nucleico, en un producto, en donde el producto puede resultar del producto transcripcional directo de una secuencia de ácido nucleico. Por ejemplo, el producto puede ser, pero no se limita a, ARNm, ARNt, ARNr, ARN antisentido, ribozima, ARN estructural o una proteína producida por la traducción de un ARNm. Los productos génicos también incluyen ARN que se hayan modificado por procesos como la adición de caperuza, la poliadenilación, la metilación y la edición, y proteínas modificadas, por ejemplo, por metilación, acetilación, fosforilación, ubiquitinación, ADP-ribosilación, miristilación y glicosilación.

Un "sitio diana" o "secuencia diana" define una porción de un ácido nucleico a la que se unirá una endonucleasa de corte rara o una transposasa asociada a CRISPR, siempre que existan condiciones suficientes para la unión.

Como se usa en la presente, el término "recombinación" se refiere a un proceso de intercambio de información genética entre dos polinucleótidos. El término "recombinación homóloga (HR)" se refiere a una forma especializada de recombinación que puede tener lugar, por ejemplo, durante la reparación de roturas de cadena doble. La recombinación homóloga requiere una homología de secuencia de nucleótidos presente en una molécula "donante". La molécula donante puede ser usada por la célula como plantilla para reparar una rotura de cadena doble. La información contenida en la molécula donante que difiere de la secuencia genómica en o cerca de la rotura de cadena doble puede incorporarse de manera estable en el ADN genómico de la célula.

El término "integrar", como se usa en la presente, se refiere al proceso de añadir ADN a una región diana del ADN. Como se describe en la presente, la integración puede facilitarse mediante varios medios diferentes, incluyendo la unión de extremos no homólogos, la recombinación homóloga o la transposición dirigida. A modo de ejemplo, la integración de una molécula de ADN suministrada por el usuario en un gen diana puede facilitarse mediante la unión de extremos no homólogos. En este caso, se produce una rotura de cadena doble dirigida dentro del gen diana y se administra una molécula de ADN suministrada por el usuario. La molécula de ADN suministrada por el usuario puede comprender extremos de ADN expuestos para facilitar la captura durante la reparación del gen diana mediante la unión de extremos no homólogos. Los extremos expuestos pueden estar presentes en la molécula de ADN tras la administración (es decir, la administración de una molécula de ADN lineal) o crearse tras la administración a la célula (es decir, una endonucleasa de corte raro escinde la molécula de ADN suministrada por el usuario dentro de la célula para exponer los extremos). Además, la molécula de ADN suministrada por el usuario puede estar albergada en un vector viral, incluyendo un vector de virus adenoasociado. En otro ejemplo, la integración se produce mediante recombinación homóloga. En este caso, el ADN suministrado por el usuario puede albergar un brazo homólogo

izquierdo y uno derecho. En otro ejemplo, la integración se produce por transposición. En este caso, el ADN suministrado por el usuario alberga un extremo izquierdo y derecho del transposón.

El término "transgén", como se usa en la presente, se refiere a una secuencia de ácidos nucleicos que puede transferirse a un organismo o célula. El transgén puede comprender un gen o secuencia de ácidos nucleicos que normalmente no están presentes en el organismo o célula diana. Además, el transgén puede comprender una copia de un gen o secuencia de ácidos nucleicos que normalmente está presente en el organismo o célula diana. Un transgén puede ser una secuencia de ADN exógena introducida en el citoplasma o núcleo de una célula diana. En una realización, los transgenes descritos en la presente contienen secuencias codificantes parciales, en donde las secuencias codificantes parciales codifican una porción de una proteína producida por un gen en la célula huésped.

Como se usa en la presente, el término "patógeno" se refiere a cualquier cosa que pueda provocar una enfermedad. Una mutación patógena puede referirse a una modificación en un gen que provoca enfermedad. Un gen patógeno se refiere a un gen que comprende una modificación que provoca enfermedad. A modo de ejemplo, un gen ATXN3 patógeno en pacientes con ataxia espinocerebelosa 3 se refiere a un gen ATXN3 con una repetición de trinucleótidos CAG expandida, en donde la repetición de trinucleótidos CAG expandida provoca la enfermedad.

Como se usa en la presente, el término "de cola a cola" se refiere a una orientación de dos unidades en direcciones opuestas e inversas. Las dos unidades pueden ser dos secuencias en una única molécula de ácido nucleico, donde el extremo 3' de cada secuencia se colocan adyacentes entre sí. Por ejemplo, un primer ácido nucleico que tiene los elementos, en dirección 5' a 3', [aceptor de corte y empalme 1] - [secuencia codificante parcial 1] - [terminador 1] y un segundo ácido nucleico que tiene los elementos [aceptor de corte y empalme 2] - [secuencia codificante parcial 2] - [terminador 2] pueden colocarse en orientación de cola a cola dando como resultado [aceptor de corte y empalme 1] - secuencia codificante parcial 1] - [terminador 1] - [terminador 2 RC] - [secuencia codificante parcial 2 RC] - [aceptor de corte y empalme 2 RC], donde RC se refiere al complemento inverso.

El término "unión intrón-exón" se refiere a una localización específica dentro de un gen. La ubicación específica se encuentra entre el último nucleótido de un intrón y el primer nucleótido del exón siguiente. Al integrar un transgén descrito en la presente, el transgén puede integrarse dentro de la "unión intrón-exón". Si el transgén comprende carga, la carga se integrará inmediatamente después del último nucleótido en el intrón. En algunos casos, la integración de un transgén dentro de la unión intrón-exón puede dar como resultado la eliminación de la secuencia dentro del exón (por ejemplo, integración mediante HR y sustitución de la secuencia dentro del exón por la carga dentro del transgén).

El término "homólogo" se refiere a una secuencia de ácidos nucleicos o aminoácidos que tiene similitud con una segunda secuencia de ácidos nucleicos o aminoácidos. En algunas realizaciones, las secuencias homólogas pueden tener por lo menos un 80% de identidad de secuencia (por ejemplo, un 81%, 85%, 90%, 95%, 96%, 97%, 98% o un 99% de identidad de secuencia) entre sí.

El término "secuencia codificante parcial", como se usa en la presente, se refiere a una secuencia de ácidos nucleicos que codifica una proteína parcial. La secuencia codificante parcial puede codificar una proteína que comprenda uno o menos aminoácidos en comparación con la proteína de tipo salvaje o la proteína funcional. La secuencia codificante parcial puede codificar una proteína parcial con homología con la proteína de tipo salvaje o proteína funcional. El término "secuencia codificante parcial" cuando se refiere a ATXN3 se refiere a una secuencia de ácidos nucleicos que codifica una proteína ATXN3 parcial. La proteína ATXN3 parcial tiene uno o menos aminoácidos en comparación con una proteína ATXN3 de tipo salvaje. Si se modifica el extremo 3' del gen, el uno o menos aminoácidos pueden ser del extremo N-terminal de la proteína. Si el gen ATXN3 tiene 11 exones, entonces la secuencia codificante parcial puede comprender la secuencia que codifica el péptido producido por los exones 2-11, o 3-11 o 4-11, o 5-11, o 6-11, o 7-11, o 8-11, o 9-11, o 10-11, o 11.

Los métodos y composiciones descritos en este documento pueden usar transgenes que tienen una secuencia de carga. El término "carga" puede referirse a elementos como la secuencia codificante completa o parcial de un gen, una secuencia parcial de un gen que alberga polimorfismos de un solo nucleótido con respecto a la diana WT o alterada, un aceptor de corte y empalme, un terminador, un elemento regulador transcripcional, etiquetas de purificación (por ejemplo, glutatión-S-transferasa, poli(His)), proteína de unión a maltosa, etiqueta de Strep, etiqueta de Myc, etiqueta AviTag, etiqueta de HA, o proteína de unión a quitina) o genes reporteros (por ejemplo, GFP, RFP, lacZ, cat, luciferasa, puro, neomicina). Como se define en la presente, "carga" puede referirse a la secuencia dentro de un transgén que se integra en un sitio diana. Por ejemplo, "carga" puede referirse a la secuencia en un transgén entre dos brazos de homología, dos sitios diana de endonucleasas de corte raro, o un extremo izquierdo y derecho de transposón.

El término "secuencia de homología" se refiere a una secuencia de ácidos nucleicos que comprende homología con un segundo ácido nucleico. La secuencia de homología, por ejemplo, puede estar presente en una molécula donante como un "brazo de homología". Un brazo de homología puede ser una secuencia de ácidos nucleicos dentro de una molécula donante que facilita la recombinación homóloga con el segundo ácido nucleico.

Como se define en la presente, un brazo de homología también puede denominarse "brazo". En una molécula donante con dos brazos de homología, los brazos de homología pueden denominarse "brazo 1" y "brazo 2". En un aspecto, una secuencia de carga puede estar flanqueada por un primer y por un segundo brazo de homología.

El término "terminador bidireccional" se refiere a un terminador que puede terminar la transcripción de la ARN polimerasa en la dirección de sentido o antisentido. Al contrario que con dos terminadores unidireccionales en orientación de cola a cola, un terminador bidireccional puede comprender una secuencia no quimérica de ADN. Ejemplos de terminadores bidireccionales incluyen el terminador ARO4, TRP1, TRP4, ADH1, CYC1, GAL1, GAL7, y GAL10.

El extremo 5' o 3' de una molécula de ácido nucleico hace referencia a la direccionalidad y orientación química del ácido nucleico. Como se define en la presente, el "extremo 5' de un gen" puede comprender el exón con el codón de inicio, pero no el exón con el codón de parada. Como se define en la presente, el "extremo 3' de un gen" puede comprender el exón con el codón de parada, pero no el exón con el codón de inicio.

El término gen "ATXN3" se refiere a un gen que codifica la enzima ataxina-3. Una secuencia representativa del gen ATXN3 puede encontrarse con la Secuencia de Referencia de NCBI: NG_008198.2 y la SEQ ID NO:42 correspondiente. Los límites del exón y el intrón pueden definirse con la secuencia proporcionada en la SEQ ID NO:42. Específicamente, el exón 1 incluye la secuencia de 1 a 54. El exón 2 incluye la secuencia de 9745 a 9909. El exón 3 incluye la secuencia de 10446 a 10490. El exón 4 incluye la secuencia de 12752 a 12837. El exón 5 incluye la secuencia de 13265 a 13331. El exón 6 incluye la secuencia de 17766 a 17853. El exón 7 incluye la secuencia de 23325 a 23457. El exón 8 incluye la secuencia de 24117 a 24283. El exón 9 incluye la secuencia de 25522 a 25618. El exón 10 incluye la secuencia de 35530 a 35648. El exón 11 incluye la secuencia de 42169 a 48031. El intrón 1 incluye la secuencia de 55 a 9744. El intrón 2 incluye la secuencia de 9910 a 10445. El intrón 3 incluye la secuencia de 10491 a 12751. El intrón 4 incluye la secuencia de 12838 a 13264. El intrón 5 incluye la secuencia de 13332 a 17765. El intrón 6 incluye la secuencia de 17854 a 23324. El intrón 7 incluye la secuencia de 23458 a 24116. El intrón 8 incluye la secuencia de 24284 a 25521. El intrón 9 incluye la secuencia de 25619 a 35529. El intrón 10 incluye la secuencia de 35649 a 42168.

El término gen "CACNA1A" se refiere a un gen que codifica la proteína alfa1A de la subunidad del canal de calcio dependiente de voltaje. Una secuencia representativa del gen CACNA1A puede encontrarse con la Secuencia de Referencia de NCBI: NG_011569.1 y la SEQ ID NO:43 correspondiente. Los límites del exón y del intrón pueden definirse con la secuencia proporcionada en la SEQ ID NO:43. Específicamente, el exón 1 incluye la secuencia de 1 a 529. El exón 2 incluye la secuencia de 51249 a 51354. El exón 3 incluye la secuencia de 53446 a 53585. El exón 4 incluye la secuencia de 134682 a 134773. El exón 5 incluye la secuencia de 140992 a 141144. El exón 6 incluye la secuencia de 146662 a 146855. El exón 7 incluye la secuencia de 170552 a 170655. El exón 8 incluye la secuencia de 171968 a 172083. El exón 9 incluye la secuencia de 173536 a 173592. El exón 10 incluye la secuencia de 176125 a 176217. El exón 11 incluye la secuencia de 189140 a 189349. El exón 12 incluye la secuencia de 193680 a 193792. El exón 13 incluye la secuencia de 197933 a 198045. El exón 14 incluye la secuencia de 198210 a 198341. El exón 15 incluye la secuencia de 198607 a 198679. El exón 16 incluye la secuencia de 202577 a 202694. El exón 17 incluye la secuencia de 202848 a 202915. El exón 18 incluye la secuencia de 205805 a 205911. El exón 19 incluye la secuencia de 207108 a 207917. El exón 20 incluye la secuencia de 219495 a 219958. El exón 21 incluye la secuencia de 221255 a 221393. El exón 22 incluye la secuencia de 223065 a 223194. El exón 23 incluye la secuencia de 229333 a 229392. El exón 24 incluye la secuencia de 230505 a 230611. El exón 25 incluye la secuencia de 243628 a 243727. El exón 26 incluye la secuencia de 244851 a 245011. El exón 27 incluye la secuencia de 246760 a 246897. El exón 28 incluye la secuencia de 248910 a 249111. El exón 29 incluye la secuencia de 251202 a 251366. El exón 30 incluye la secuencia de 253360 a 253470. El exón 31 incluye la secuencia de 261196 a 261279. El exón 32 incluye la secuencia de 270731 a 270847. El exón 33 incluye la secuencia de 271187 a 271252. El exón 34 incluye la secuencia de 271425 a 271540. El exón 35 incluye la secuencia de 274601 a 274751. El exón 36 incluye la secuencia de 276252 a 276379. El exón 37 incluye la secuencia de 277666 a 277762. El exón 38 incluye la secuencia de 281689 a 281794. El exón 39 incluye la secuencia de 291853 a 291960. El exón 40 incluye la secuencia de 292128 a 292228. El exón 41 incluye la secuencia de 293721 a 293830. El exón 42 incluye la secuencia de 293939 a 294077. El exón 43 incluye la secuencia de 294245 a 294358. El exón 44 incluye la secuencia de 295809 a 295844. El exón 45 incluye la secuencia de 296963 a 297149. El exón 46 incluye la secuencia de 297452 a 297705. El exón 47 incluye la secuencia de 298413 a 300019. El intrón 1 incluye la secuencia de 530 a 51248. El intrón 2 incluye la secuencia de 51355 a 53445. El intrón 3 incluye la secuencia de 53586 a 134681. El intrón 4 incluye la secuencia de 134774 a 140991. El intrón 5 incluye la secuencia de 141145 a 146661. El intrón 6 incluye la secuencia de 146856 a 170551. El intrón 7 incluye la secuencia de 170656 a 171967. El intrón 8 incluye la secuencia de 172084 a 173535. El intrón 9 incluye la secuencia de 173593 a 176124. El intrón 10 incluye la secuencia de 176218 a 189139. El intrón 11 incluye la secuencia de 189350 a 193679. El intrón 12 incluye la secuencia de 193793 a 197932. El intrón 13 incluye la secuencia de 198046 a 198209. El intrón 14 incluye la secuencia de 198342 a 198606. El intrón 15 incluye la secuencia de 198680 a 202576. El intrón 16 incluye la secuencia de 202695 a 202847. El intrón 17 incluye la secuencia de 202916 a 205804. El intrón 18 incluye la secuencia de 205912 a 207107. El intrón 19 incluye la secuencia de 207918 a 219494. El intrón 20 incluye la secuencia de 219959 a 221254. El intrón 21 incluye la secuencia de 221394 a 223064. El intrón 22 incluye la secuencia de 223195 a 229332. El intrón 23 incluye la secuencia de 229393 a 230504. El intrón 24 incluye la

secuencia de 230612 a 243627. El intrón 25 incluye la secuencia de 243728 a 244850. El intrón 26 incluye la secuencia de 245012 a 246759. El intrón 27 incluye la secuencia de 246898 a 248909. El intrón 28 incluye la secuencia de 249112 a 251201. El intrón 29 incluye la secuencia de 251367 a 253359. El intrón 30 incluye la secuencia de 253471 a 261195. El intrón 31 incluye la secuencia de 261280 a 270730. El intrón 32 incluye la secuencia de 270848 a 271186. El intrón 33 incluye la secuencia de 271253 a 271424. El intrón 34 incluye la secuencia de 271541 a 274600. El intrón 35 incluye la secuencia de 274752 a 276251. El intrón 36 incluye la secuencia de 276380 a 277665. El intrón 37 incluye la secuencia de 277763 a 281688. El intrón 38 incluye la secuencia de 281795 a 291852. El intrón 39 incluye la secuencia de 291961 a 292127. El intrón 40 incluye la secuencia de 292229 a 293720. El intrón 41 incluye la secuencia de 293831 a 293938. El intrón 42 incluye la secuencia de 294078 a 294244. El intrón 43 incluye la secuencia de 294359 a 295808. El intrón 44 incluye la secuencia de 295845 a 296962. El intrón 45 incluye la secuencia de 297150 a 297451. El intrón 46 incluye la secuencia de 297706 a 298412.

El porcentaje de identidad de secuencia entre una secuencia de ácido nucleico o aminoácidos particular y una secuencia referenciada por un número de identificación de secuencia particular se determina de la siguiente manera. En primer lugar, se compara una secuencia de ácido nucleico o aminoácidos con la secuencia expuesta en un número de identificación de secuencia particular usando el programa BLAST 2 Sequences (BI2seq) de la versión independiente de BLASTZ que contiene BLASTN versión 2.0.14 y BLASTP versión 2.0.14. Esta versión independiente de BLASTZ puede obtenerse en línea en fr.com/blast o en ncbi.nlm.nih.gov. En el archivo readme que acompaña a BLASTZ pueden encontrarse instrucciones que explican cómo usar el programa BI2seq. BI2seq realiza una comparación entre dos secuencias usando el algoritmo BLASTN o BLASTP. El BLASTN se usa para comparar secuencias de ácidos nucleicos, mientras que el BLASTP se usa para comparar secuencias de aminoácidos. Para comparar dos secuencias de ácidos nucleicos, las opciones se configuran de la siguiente manera: -i se establece en un archivo que contiene la primera secuencia de ácido nucleico que se va a comparar (por ejemplo, C:\seq1.txt); -j se establece en un archivo que contiene la segunda secuencia de ácido nucleico que se va a comparar (por ejemplo, C:\seq2.txt); -p se establece en blastn; -o se establece en cualquier nombre de archivo deseado (por ejemplo, C:\output.txt); -q se establece en -1; -r se establece en 2; y todas las demás opciones se dejan en su configuración por defecto. Por ejemplo, puede usarse el siguiente comando para generar un archivo de salida que contenga una comparación entre dos secuencias: C:\BI2seq -i c:\seq1.txt -j c:\seq2.txt -p blastn -o c:\output.txt -q -1 -r 2. Para comparar dos secuencias de aminoácidos, las opciones de BI2seq se configuran de la siguiente manera: -i se establece en un archivo que contiene la primera secuencia de aminoácidos que se va a comparar (por ejemplo, C:\seq1.txt); -j se establece en un archivo que contiene la segunda secuencia de aminoácidos que se va a comparar (por ejemplo, C:\seq2.txt); -p se establece en blastp; -o se establece en cualquier nombre de archivo deseado (por ejemplo, C:\output.txt); y todas las demás opciones se dejan en su configuración por defecto. Por ejemplo, puede usarse el siguiente comando para generar un archivo de salida que contenga una comparación entre dos secuencias de aminoácidos: C:\BI2seq -i c:\seq1.txt -j c:\seq2.txt -p blastp -o c:\output.txt. Si las dos secuencias comparadas comparten homología, entonces el archivo de salida designado presentará esas regiones de homología como secuencias alineadas. Si las dos secuencias comparadas no comparten homología, entonces el archivo de salida designado no presentará secuencias alineadas.

Una vez alineadas, el número de coincidencias se determina contando el número de posiciones en las que hay un residuo de nucleótido o aminoácido idéntico en ambas secuencias. El porcentaje de identidad de secuencia se determina dividiendo el número de coincidencias por la longitud de la secuencia establecida en la secuencia identificada, o por una longitud articulada (por ejemplo, 100 nucleótidos o residuos de aminoácidos consecutivos de una secuencia establecida en una secuencia identificada), y multiplicando el valor resultante por 100. El valor del porcentaje de identidad de secuencia se redondea a la décima más próxima.

Este documento presenta métodos para modificar el extremo 3' de genes endógenos, donde los genes endógenos tienen por lo menos un intrón entre dos exones codificantes. El intrón puede ser cualquier intrón que sea eliminado del ARN mensajero precursor por la maquinaria normal de procesamiento del ARN mensajero. El intrón puede tener entre 20 bp y >500 kb y comprender elementos que incluyan un sitio donante de corte y empalme, una secuencia de ramificación y un sitio aceptor. Los transgenes divulgados en la presente para la modificación del extremo 3' de genes endógenos pueden comprender múltiples elementos funcionales, incluyendo sitios diana para endonucleasas de corte raro, brazos de homología, secuencias aceptoras de corte y empalme, secuencias codificantes y terminadores de transcripción (FIG. 1).

En una realización, el transgén comprende dos sitios diana para una o más endonucleasas de corte raro. Los sitios diana pueden tener una secuencia y longitud adecuadas para la escisión por una endonucleasa de corte raro. El sitio diana puede ser susceptible de escisión por sistemas CRISPR, nucleasas efectoras TAL, nucleasas de dedos de zinc o meganucleasas, o una combinación de sistemas CRISPR, nucleasas TALE, nucleasas de dedos de zinc o meganucleasas, o cualquier otra nucleasa específico del sitio. Los sitios diana pueden colocarse de tal manera que la escisión por la endonucleasa de corte raro dé como resultado la liberación de un transgén a partir de un vector. El vector puede incluir vectores virales (por ejemplo, vectores adenoasociados) o vectores no virales (por ejemplo, plásmidos, vectores minicirculares). Si el transgén comprende dos sitios diana, los sitios diana pueden ser la misma secuencia (es decir, dirigidos por la misma endonucleasa de corte raro) o pueden ser secuencias diferentes (es decir, dirigidos por dos o más endonucleasas de corte raro diferentes).

En una realización, el transgén comprende un primer y un segundo sitio diana para una o más endonucleasas de corte raro junto con un primer y un segundo brazos de homología. El primer y el segundo brazos de homología pueden incluir una secuencia que sea homóloga con una secuencia genómica en o cerca del lugar deseado de integración. Los brazos de homología pueden tener una longitud adecuada para participar en la recombinación homóloga con la secuencia en o cerca del lugar deseado de integración. La longitud de cada brazo de homología puede estar entre 20 nt y 10.000 nt (por ejemplo, 20 nt, 30 nt, 40 nt, 50 nt, 100 nt, 200 nt, 300 nt, 400 nt, 500 nt, 600 nt, 700 nt, 800 nt, 900 nt, 1.000 nt, 2.000 nt, 3.000 nt, 4.000 nt, 5.000 nt, 6.000 nt, 7.000 nt, 8.000 nt, 9.000 nt, 10.000 nt). En una realización, un brazo de homología puede comprender elementos funcionales, incluyendo un sitio diana para una endonucleasa de corte raro y/o una secuencia aceptora de corte y empalme. En una realización, un primer brazo de homología (por ejemplo, un brazo de homología izquierdo) puede comprender una secuencia homóloga al intrón objetivo, que incluye el sitio aceptor de corte y empalme del intrón objetivo. En otra realización, un segundo brazo de homología puede comprender una secuencia homóloga a la secuencia genómica en sentido descendente del intrón objetivo (por ejemplo, secuencia de exón, secuencia 3' UTR). Sin embargo, el segundo brazo de homología no debe poseer funciones de aceptor de corte y empalme en la dirección inversa del complemento. Para determinar si una secuencia posee funciones de aceptor de corte y empalme, pueden seguirse varios pasos, incluyendo análisis in silico y pruebas experimentales. Para determinar si existe potencial para funciones de aceptor de corte y empalme, puede buscarse en la secuencia deseada para el segundo brazo de homología secuencias de ramificación de consenso (por ejemplo, YTRAC) y sitios de aceptor de corte y empalme (por ejemplo, NCAGG rico en Y). Si hay secuencias ramificadas o aceptoras de corte y empalme, pueden introducirse polimorfismos de un solo nucleótido para destruir la función, o puede seleccionarse una secuencia diferente pero adyacente que no comprenda dichas secuencias. Preferiblemente, la ventana de secuencia que puede usarse para un segundo brazo de homología se extiende de 1 bp a 10kb en sentido descendente del intrón que es el objetivo de la integración. Para determinar experimentalmente si la segunda homología posee una función de aceptor de corte y empalme, puede construirse un constructo sintético que comprenda el segundo brazo de homología dentro de un intrón de un gen reportero. A continuación, el constructo puede administrarse a un tipo de célula apropiado y monitorizarse para determinar la función de corte y empalme.

El transgén comprende dos secuencias aceptoras de corte y empalme, denominadas en la presente como primera y segunda secuencias aceptoras de corte y empalme. La primera y la segunda secuencias aceptoras de corte y empalme se sitúan dentro del transgén en direcciones opuestas (es decir, en orientaciones de cola a cola) y flanqueando secuencias internas (es decir, secuencias codificantes y terminadores). Cuando el transgén se integra en un intrón en dirección directa o inversa, las secuencias aceptoras de corte y empalme facilitan la eliminación de la secuencia del intrón adyacente/en sentido ascendente durante el procesamiento del ARNm. La primera y la segunda secuencias aceptoras de corte y empalme pueden ser las mismas secuencias o secuencias diferentes. Una o ambas secuencias aceptores de corte y empalme pueden ser la secuencia aceptor de corte y empalme del intrón en el que se va a integrar el transgén. Una o ambas secuencias aceptores de corte y empalme pueden ser una secuencia aceptor de corte y empalme sintética o una secuencia aceptor de corte y empalme de un intrón de un gen diferente.

El transgén comprende una primera y una segunda secuencias codificantes enlazadas operativamente a la primera y a la segunda secuencias aceptoras de corte y empalme. La primera y la segunda secuencias codificantes están situadas dentro del transgén en direcciones opuestas (es decir, en orientaciones de cola a cola). Cuando el transgén se integra en un gen endógeno en dirección directa o inversa, el promotor del gen endógeno transcribe la primera o la segunda secuencia codificante en ARNm. Las secuencias codificantes pueden diseñarse para que corrijan secuencias codificantes defectuosas, introduzcan mutaciones o introduzcan nuevas secuencias peptídicas. La primera y la segunda secuencias codificantes pueden ser la misma secuencia de ácido nucleico y codificar la misma proteína. Alternativamente, la primera y la segunda secuencias codificantes pueden ser diferentes secuencias de ácido nucleico y codificar para la misma proteína (es decir, usando la degeneración de codones). La secuencia codificante puede codificar etiquetas de purificación (por ejemplo, glutatión-S-transferasa, poli(His), proteína de unión a maltosa, etiqueta de Strep, etiqueta de Myc, etiqueta AviTag, etiqueta de HA, o proteína de unión a quitina) o proteínas reporteras (por ejemplo, GFP, RFP, lacZ, cat, luciferasa, puro, neomicina). En una realización, el transgén comprende una primera y una segunda secuencias codificantes parciales enlazadas operativamente a una primera y una segunda secuencias aceptoras de corte y empalme, y el transgén no comprende un promotor.

El transgén comprende un terminador bidireccional o un primer y un segundo terminadores, enlazados operativamente a una primera y una segunda secuencias codificantes. El primero y el segundo terminadores están situados dentro del transgén en direcciones opuestas (es decir, en orientaciones de cola a cola). Cuando el transgén se integra en un gen endógeno en dirección directa o inversa, el un terminador bidireccional o el primero y el segundo terminadores terminan la transcripción a partir del promotor del gen endógeno. El primer y el segundo terminadores pueden ser los mismos terminadores o terminadores diferentes.

En una realización, este documento proporciona un transgén que comprende un primer y un segundo sitios diana de endonucleasas de corte raro, una primera y una segunda secuencias aceptoras de corte y empalme, una primera y una segunda secuencias codificantes, y un terminador bidireccional o un primer y un segundo terminadores. El transgén puede integrarse en genes endógenos mediante métodos no dependientes de la homología, incluyendo

la unión de extremos no homólogos y la unión de extremos no homólogos alternativa o mediante la unión de extremos mediada por microhomología. En un aspecto, el transgén se integra en un intrón dentro del gen endógeno (FIG. 2).

En otra realización, este documento proporciona un transgén que comprende un primer y segundo brazos de homología, un primer y un segundo sitios diana de endonucleasa de corte raro, una primera y una segunda secuencias aceptoras de corte y empalme, una primera y una segunda secuencias codificantes, y un terminador bidireccional o un primer y un segundo terminadores. El transgén puede integrarse en genes endógenos tanto mediante métodos dependientes de homología (por ejemplo, apareamiento de cadena dependiente de síntesis y unión de extremos mediada por microhomología) como mediante métodos no dependientes de homología (por ejemplo, unión de extremos no homólogos y unión de extremos no homólogos alternativa). En un aspecto, el transgén se integra en un intrón dentro del gen endógeno (FIG. 3). En otro aspecto, el transgén se integra al final del intrón o al comienzo del exón en sentido descendente (FIG. 3).

En otra realización, este documento proporciona un transgén que comprende un primer y un segundo brazos de homología, una primera y una segunda secuencias codificantes, una primera y una segunda secuencias aceptoras de corte y empalme, y un terminador bidireccional o un primer y un segundo terminadores (FIG. 1). En otra realización, este documento proporciona un transgén que comprende, una primera y una segunda secuencias codificantes, una primera y una segunda secuencias aceptoras de corte y empalme, y un terminador bidireccional o un primer y un segundo terminadores.

En otra realización, este documento proporciona un transgén que comprende un primer y un segundo brazos de homología, una primera y una segunda secuencias codificantes, una primera y una segunda secuencias aceptoras de corte y empalme, un primer y un segundo terminadores, y un terminador bidireccional o una primera y una segunda secuencias adicionales (FIG. 1). En ciertas realizaciones, la secuencia adicional puede ser cualquier secuencia adicional que esté presente en el transgén en los extremos 5' y 3', sin embargo, la secuencia adicional no debe comprender ningún elemento que funcione como aceptor de corte y empalme. La secuencia adicional puede ser, por ejemplo, repeticiones terminales invertidas del genoma de un virus. La secuencia adicional puede estar presente en un transgén que tenga un formato lineal. El formato lineal permite la integración por NHEJ. Por ejemplo, un transgén albergado en un vector de virus adenoasociado, en donde la secuencia adicional son las repeticiones terminales invertidas, puede integrarse directamente mediante NHEJ en un sitio diana después de la escisión por una endonucleasa de corte raro (es decir, no se requiere procesamiento del transgén). En otro ejemplo, la secuencia adicional es un extremo izquierdo y derecho del transposón.

En otra realización, este documento proporciona transgenes dentro de vectores virales, incluyendo virus adenoasociados y adenovirus, donde el transgén comprende una primera y una segunda secuencias aceptoras de corte y empalme, una primera y una segunda secuencias codificantes, y un terminador bidireccional o un primer y un segundo terminadores. Debido a las repeticiones terminales invertidas de los vectores virales, los transgenes también comprenden una primera y una segunda secuencias adicionales.

En otra realización, este documento proporciona transgenes dentro de vectores virales, incluyendo virus adenoasociados y adenovirus, donde el transgén comprende un primer y un segundo brazos de homología, una primera y una segunda secuencias aceptoras de corte y empalme, una primera y una segunda secuencias codificantes, y un terminador bidireccional o un primer y un segundo terminadores. Debido a las repeticiones terminales invertidas de los vectores virales, los transgenes también comprenden una primera y una segunda secuencias adicionales.

En algunas realizaciones, los transgenes proporcionados en la presente pueden integrarse con transposasas. Las transposasas pueden incluir transposasas CRISPR (Strecker et al., *Science* 10.11126/science.aax9181, 2019; Klompe et al., *Nature*, 10.1038/s41586-019-1323-z, 2019). Las transposasas pueden usarse en combinación con un transgén que comprende, una primera y una segunda secuencias aceptoras de corte y empalme, una primera y una segunda secuencias codificantes, y un terminador bidireccional o un primer y un segundo terminadores (FIG. 1), y un extremo izquierdo y un extremo derecho del transposón. Las transposasas CRISPR pueden incluir la proteína CRISPR TipoV-U5, C2C5, Cas12k, junto con las proteínas tnsB, tnsC y tniQ. En algunas realizaciones, la Cas12k puede ser de *Scytonema hofmanni* (SEQ ID NO:30) o *Anabaena cylindrica* (SEQ ID NO:31). En una realización, los transgenes descritos en la presente que comprenden un extremo de transposón izquierdo (SEQ ID NO:32) y uno derecho (SEQ ID NO:33) pueden administrarse a las células junto con ShCas12k, tnsB, tnsC, TniQ y un ARNg (SEQ ID NO: 14).

Alternativamente, la transposasa CRISPR puede incluir la proteína Cas6, junto con proteínas auxiliares que incluyen Cas7, Cas8 y TniQ. En una realización, los transgenes descritos en la presente que comprenden un extremo de transposón izquierdo (SEQ ID NO:41) y uno derecho (SEQ ID NO: 13) pueden administrarse a células eucariotas junto con Cas6 (SEQ ID NO:37), Cas7 (SEQ ID NO:37), Cas8 (SEQ ID NO:37), TniQ (SEQ ID NO:37), TnsA (SEQ ID NO:37), TnsB (SEQ ID NO:37), TnsC (SEQ ID NO:37) y un ARNg (SEQ ID NO: 12). Las proteínas pueden administrarse a las células directamente como proteína purificada o codificadas en ARN o ADN. Si se codifican en ARN o ADN, la secuencia puede optimizarse en codones para su expresión en células eucariotas. El ARNg (SEQ ID NO: 12) puede colocarse en sentido descendente de un promotor de ARN polIII y terminarse con un terminador poli(T).

Los transgenes descritos en la presente pueden tener una combinación de elementos que incluyen aceptores de corte y empalme, secuencias codificantes parciales, terminadores, brazos de homología, extremos izquierdo y derecho de transposasa, y sitios para la escisión por endonucleasas de corte raro. En una realización, la combinación puede ser, de 5' a 3', [aceptor de corte y empalme 1] - [secuencia codificante parcial 1] - [terminador 1] - [terminador 2 RC] - [secuencia codificante parcial 2 RC] - [aceptor de corte y empalme 2 RC], donde RC significa complemento inverso. Esta combinación puede estar albergada en una molécula lineal de ADN o en una molécula AAV y puede integrarse mediante NHEJ a través de una rotura dirigida en el gen diana. En otra realización, la combinación puede ser, de 5' a 3', [sitio de escisión de la endonucleasa de corte raro 1] - [aceptor de corte y empalme 1] - [secuencia codificante parcial 1] - [terminador 1] - [terminador 2 RC] - [secuencia codificante parcial 2 RC] - [aceptor de corte y empalme 2 RC] - [sitio de escisión de la endonucleasa de corte raro 1]. En otra realización, la combinación puede ser, de 5' a 3', [sitio de escisión de la endonucleasa de corte raro 1] - [brazo de homología 1] - [aceptor de corte y empalme 1] - [secuencia codificante parcial 1] - [terminador 1] - [terminador 2 RC] - [secuencia codificante parcial 2 RC] - [aceptor de corte y empalme 2 RC] - [brazo de homología 2] - [sitio de escisión de la endonucleasa de corte raro 2]. En esta combinación pueden usarse una o más endonucleasas de corte raro para facilitar la HR y la NHEJ. Por ejemplo, una única nucleasa de corte raro puede escindir el gen diana (es decir, un intrón deseado) y los sitios de escisión que flanquean los brazos de homología pueden diseñarse para ser la misma secuencia diana dentro del intrón. En otra realización, la combinación puede ser, de 5' a 3', [brazo de homología 1 + sitio de escisión de la endonucleasa de corte raro 1] - [aceptor de corte y empalme 1] - [secuencia codificante parcial 1] - [terminador 1] - [terminador 2 RC] - [secuencia codificante parcial 2 RC] - [aceptor de corte y empalme 2 RC] - [brazo de homología 2] - [sitio de escisión de la endonucleasa de corte raro 1]. En esta combinación, una o más endonucleasas de corte raro pueden facilitar la RH y la NHEJ. Por ejemplo, una sola nucleasa de corte raro puede escindir dentro del brazo de homología 1, en sentido descendente del brazo de homología 2, y en el sitio genómico diana (es decir, en el sitio con homología con la secuencia en el brazo de homología 1). En otra realización, la combinación puede ser de 5' a 3', [extremo izquierdo para una transposasa] - [aceptor de corte y empalme 1] - [secuencia codificante parcial 1] - [terminador 1] - [terminador 2 RC] - [secuencia codificante parcial 2 RC] - [aceptor de corte y empalme 2 RC] - [extremo derecho para una transposasa]. En todas las realizaciones, el aceptor de corte y empalme 1 y el aceptor de corte y empalme 2 pueden ser secuencias iguales o diferentes; la secuencia codificante parcial 1 y la secuencia codificante parcial 2 pueden ser secuencias iguales o diferentes; el terminador 1 y el terminador 2 pueden ser secuencias iguales o diferentes.

En algunas realizaciones, puede integrarse en el ADN un transgén que comprende la estructura [sitio de escisión de la endonucleasa de corte raro 1] - [brazo de homología 1] - [aceptor de corte y empalme 1] - [secuencia codificante parcial 1] - [terminador 1] - [terminador 2 RC] - [secuencia codificante parcial 2 RC] - [aceptor de corte y empalme 2 RC] - [brazo de homología 2] - [sitio de escisión de la endonucleasa de corte raro 2] mediante la administración de una o más endonucleasas de corte raro. Si se administra una endonucleasa de corte raro, la endonucleasa de corte raro puede liberar el transgén mediante escisión en los sitios de corte de la endonucleasa de corte raro 1 y 2. Además, la misma endonucleasa de corte raro puede crear una rotura dentro del gen diana, simulando la inserción mediante HR o NHEJ.

En otras realizaciones, puede integrarse en el ADN un transgén que comprende la estructura [brazo de homología 1 + sitio de escisión de endonucleasa de corte raro 1] - [aceptor de corte y empalme 1] - [secuencia codificante parcial 1] - [terminador 1] - [terminador 2 RC] - [secuencia codificante parcial 2 RC] - [aceptor de corte y empalme 2 RC] - [brazo de homología 2] - [sitio de escisión de endonucleasa de corte raro 1] mediante la administración de una o más endonucleasas de corte raro. Si se administra una endonucleasa de corte raro, la endonucleasa de corte raro puede liberar el transgén por escisión en el sitio de escisión de la endonucleasa de corte raro 1 y 2. Además, la misma endonucleasa de corte raro puede crear una rotura dentro del gen diana, simulando la inserción mediante HR o NHEJ. La integración por HR puede producirse cuando la escisión se produce en sentido ascendente del sitio de integración (es decir, dentro de un brazo de homología).

En algunas realizaciones, la ubicación para la integración de los transgenes puede ser un intrón o una unión de intrón-exón. Cuando se dirige a un intrón, la secuencia codificante parcial puede comprender la secuencia que codifica el péptido producido por los siguientes exones dentro del gen endógeno. Por ejemplo, si el transgén está diseñado para integrarse en el intrón 9 de un gen endógeno con 11 exones, entonces la secuencia codificante parcial puede comprender la secuencia que codifica el péptido producido por los exones 10 y 11 del gen endógeno. Cuando se dirige a una unión intrón-exón, el transgén puede diseñarse para que comprenda brazos de homología con secuencia homóloga a la 3' de dicho intrón.

En algunas realizaciones, las secuencias codificantes parciales pueden ser secuencias codificantes completas. La secuencia codificante completa puede codificar un gen endógeno (por ejemplo, Factor VIII, Factor IX, o INS), o genes reporteros (por ejemplo, RFP, GFP, cat, lacZ, luciferasa). Las secuencias codificantes completas pueden enlazarse operativamente con aceptores de corte y empalme y terminadores y colocarse en un transgén en una orientación de cola a cola.

Los métodos y composiciones proporcionados en la presente pueden usarse para modificar genes endógenos dentro de las células. Los genes endógenos pueden incluir, fibrinógeno, protrombina, factor tisular, Factor V, Factor

VII, Factor VIII, Factor IX, Factor X, Factor XI, Factor XII (factor de Hageman), Factor XIII (factor estabilizador de fibrina), factor de von Willebrand, precalicreína, cininógeno de alto peso molecular (factor de Fitzgerald), fibronectina, antitrombina III, cofactor II de la heparina, proteína C, proteína S, proteína Z, inhibidor de la proteasa relacionado con la proteína Z, plasminógeno, alfa 2-antiplasmina, activador tisular del plasminógeno, uroquinasa, inhibidor 1 del activador del plasminógeno, inhibidor 2 del activador del plasminógeno, glucocerebrosidasa (GBA), α -galactosidasa A (GLA), iduronato sulfatasa (IDS), iduronidasa (IDUA), esfingomielinasa ácida (SMPD1), MMAA, MMAB, MMACHC, MMADHC (C2orf25), MTRR, LMBRD1, MTR, propionil-CoA carboxilasa (PCC) (subunidades PCCA y/o PCCB), una proteína transportadora de glucosa-6-fosfato (G6PT) o glucosa-6-fosfatasa (G6Pasa), un receptor de LDL (LDLR), ApoB, LDLRAP-1, una PCSK9, una proteína mitocondrial como la NAGS (N-acetilglutamato sintetasa), CPS1 (carbamoil fosfato sintetasa I), y OTC (ornitina transcarbamilasa), ASS (ácido argininosuccínico sintetasa), ASL (ácido argininosuccinasa liasa) y/o ARG1 (arginasa), y/o una proteína de la familia 25 de transportadores de solutos (SLC25A13, un transportador de aspartato/glutamato), un UGT1A1 o polipéptido A1 de UDP glucuronosiltransferasa, una fumarilacetacetato hidrolasa (FAH), una proteína alanina-gloxilato aminotransferasa (AGXT), una proteína glioxilato reductasa/hidroxipiruvato reductasa (GRHPR), una proteína del gen de la transtiretina (TTR), una proteína de ATP7B, una proteína fenilalanina hidroxilasa (PAH), una proteína USH2A, una proteína ATXN y una proteína lipoproteína liasa (LPL).

El transgén puede incluir una secuencia para modificar la secuencia que codifica un polipéptido que carece o no es funcional o que tiene una mutación de ganancia de función en el sujeto que padece una enfermedad genética, incluyendo, entre otras, las siguientes enfermedades genéticas: acondroplasia, acromatopsia, deficiencia de maltasa ácida, deficiencia de adenosina deaminasa, adrenoleucodistrofia, síndrome de aicardi, deficiencia de alfa-1 antitripsina, alfa-talasemia, síndrome de insensibilidad a los andrógenos, síndrome de pert, displasia arritmogénica del ventrículo derecho, ataxia telangictasia, síndrome de Barth, beta-talasemia, síndrome del nevus azul en tetina de goma, enfermedad de Canavan, enfermedades granulomatosas crónicas (EGC), síndrome de Cri du Chat, fibrosis quística, enfermedad de Dercum, displasia ectodérmica, anemia de Fanconi, fibrodisplasia osificante progresiva, síndrome del cromosoma X frágil, galactosemia, enfermedad de Gaucher, gangliosidosis generalizada (por ejemplo, GM1), hemocromatosis, mutación de la hemoglobina C en el 6º codón de la beta-globina (HbC), hemofilia, enfermedad de Huntington, síndrome de Hurler, hipofosfatasa, síndrome de Klinefelter, enfermedad de Krabbes, síndrome de Langer-Giedion, deficiencia de adhesión leucocitaria, leucodistrofia, síndrome de QT largo, síndrome de Marfan, síndrome de Moebius, mucopolisacaridosis (MPS), síndrome de la rótula ungueal, diabetes insípida nefrogénica, neurofibromatosis, enfermedad de Neimann-Pick, osteogénesis imperfecta, porfiria, síndrome de Prader-Willi, progeria, síndrome de Proteus, retinoblastoma, síndrome de Rett, síndrome de Rubinstein-Taybi, síndrome de Sanfilippo, inmunodeficiencia combinada grave (SCID), síndrome de Shwachman, drepanocitosis (anemia de células falciformes), síndrome de Smith-Magenis, síndrome de Stickler, enfermedad de Tay-Sachs, síndrome de trombocitopenia de radio ausente (TAR), síndrome de Treacher Collins, trisomía, esclerosis tuberosa, síndrome de Turner, trastorno del ciclo de la urea, enfermedad de von Rippel-Landau, síndrome de Waardenburg, síndrome de Williams, enfermedad de Wilson, síndrome de Wiskott-Aldrich, síndrome linfoproliferativo ligado al cromosoma X, enfermedades por almacenamiento lisosómico (por ejemplo, enfermedad de Gaucher, GM1, enfermedad de Fabry y enfermedad de Tay-Sachs), mucopolisacaridosis (por ejemplo, enfermedad de Hunter, enfermedad de Hurler), hemoglobinopatías (por ejemplo, enfermedad de células falciformes, HbC, α -talasemia, β -talasemia) y hemofilias.

Enfermedades adicionales que pueden tratarse mediante integración dirigida son la enfermedad de von Willebrand, el síndrome de Usher, enfermedad renal poliquística, la ataxia espinocerebelosa de tipo 3 y la ataxia espinocerebelosa de tipo 6.

En una realización, la modificación genómica es la inserción de un transgén en la secuencia genómica endógena de CACNA1A. El transgén puede incluir una secuencia codificante sintética y parcial para la proteína CACNA1A. La secuencia codificante parcial puede ser homóloga a la secuencia codificante dentro de un gen CACNA1A de tipo salvaje, o una variante funcional del gen CACNA1A de tipo salvaje, o un mutante del gen CACNA1A de tipo salvaje. En una realización, el transgén que codifica la proteína CACNA1A parcial se inserta en el intrón 46 o al comienzo del exón 47.

En otra realización, la modificación genómica es la inserción de un transgén en la secuencia genómica endógena de ATXN3. El transgén puede incluir una secuencia codificante sintética y parcial para la proteína ATXN3. La secuencia codificante parcial puede ser homóloga a la secuencia codificante dentro de un gen ATXN3 de tipo salvaje, o una variante funcional del gen ATXN3 de tipo salvaje, o un mutante del gen ATXN3 de tipo salvaje. En una realización, el transgén que codifica la proteína ATXN3 parcial se inserta en el intrón 9 o al principio del exón 10.

Los métodos y composiciones descritos en la presente pueden usarse para modificar el extremo 3' de un gen endógeno, dando como resultado por tanto la modificación del extremo C-terminal de la proteína codificada por el gen endógeno. La modificación del extremo 3' de la secuencia codificante del gen endógeno puede incluir la sustitución del exón codificante final (es decir, el exón que comprende el codón de parada), hasta un exón que se encuentre entre el exón con el inicio de la codificación y el exón final. Como se define en la presente, "sustitución" se refiere a la inserción de ADN en un gen, en donde el ADN insertado proporciona la información para producir el ARNm y la proteína de 1 o más exones. La sustitución puede producirse mediante la integración de un transgén en el gen

endógeno, en donde el transgén comprende una o más secuencias codificantes enlazadas operativamente con un aceptor de corte y empalme. La inserción puede dar como resultado o no la delección de secuencia dentro del gen endógeno (por ejemplo, delección de intrones y exones). Por ejemplo, si un gen comprende 72 exones, y el codón de inicio está dentro del exón 1, la modificación puede incluir la sustitución de los exones 2-72, 3-72, 4-72, 5-72, 6-72, 7-72, 8-72, 9-72, 10-72, 11-72, 12-72, 13-72, 14-72, 15-72, 16-72, 17-72, 18-72, 19-72, 20-72, 21-72, 22-72, o 23-72, o 24-72, o 25-72, o 26-72, o 27-72, o 28-72, o 29-72, o 30-72, o 31-72, o 32-72, o 33-72, o 34-72, o 35-72, o 36-72, o 37-72, o 38-72, o 39-72, o 40-72, o 41-72, o 42-72, o 43-72, o 44-72, o 45-72, o 46-72, o 47-72, o 48-72, o 49-72, o 50-72, o 51-72, o 52-72, o 53-72, o 54-72, o 55-72, o 56-72, o 57-72, o 58-72, o 59-72, o 60-72, o 61-72, o 62-72, o 63-72, o 64-72, o 65-72, o 66-72, o 67-72, o 68-72, o 69-72, o 70-72, o 71-72 o 72. En una realización, los exones del gen endógeno pueden sustituirse integrando un transgén en el gen endógeno, en donde el transgén comprende una primera y una segunda secuencias codificantes parciales, en donde la primera y la segunda secuencias codificantes parciales codifican un péptido producido por los exones de los genes endógenos. Por ejemplo, la primera y la segunda secuencias codificantes del transgén pueden codificar un péptido que se produce por los exones 2-72, 3-72, 4-72, 5-72, 6-72, 7-72, 8-72, 9-72, 10-72, 11-72, 12-72, 13-72, 14-72, 15-72, 16-72, 17-72, 18-72, 19-72, 20-72, 21-72, 22-72, o 23-72, o 24-72, o 25-72, o 26-72, o 27-72, o 28-72, o 29-72, o 30-72, o 31-72, o 32-72, o 33-72, o 34-72, o 35-72, o 36-72, o 37-72, o 38-72, o 39-72, o 40-72, o 41-72, o 42-72, o 43-72, o 44-72, o 45-72, o 46-72, o 47-72, o 48-72, o 49-72, o 50-72, o 51-72, o 52-72, o 53-72, o 54-72, o 55-72, o 56-72, o 57-72, o 58-72, o 59-72, o 60-72, o 61-72, o 62-72, o 63-72, o 64-72, o 65-72, o 66-72, o 67-72, o 68-72, o 69-72, o 70-72, o 71-72 o 72 del gen endógeno. El transgén puede integrarse dentro del gen endógeno en el intrón en sentido ascendente o al principio del exón correspondiente al primer exón dentro de la secuencia codificante parcial del transgén (FIG. 2). El transgén puede diseñarse para que tenga 4,7kb o menos, e incorporarse en un vector y una partícula AAV, y administrarse in vivo a las células diana.

En una realización, el transgén es una secuencia de ADN que alberga una primera y una segunda secuencias codificantes parciales, en donde las secuencias codificantes parciales codifican una proteína parcial, en donde la proteína parcial es homóloga a una región correspondiente en una proteína funcional producida a partir de un gen de tipo salvaje. El gen huésped o gen endógeno es aquel en el que la expresión de la proteína es aberrante, en otras palabras, no se expresa, se expresa a niveles bajos, o se expresa pero el ARNm o el producto proteico o una porción del mismo no es funcional, tiene una función reducida, o tiene una ganancia de función, lo que provoca un trastorno en el huésped.

Como se describe en la presente, la molécula donante puede estar en un vector viral o no viral. Los vectores pueden estar en forma de ADN circular o lineal de cadena doble o de cadena sencilla. La molécula donante puede conjugarse o asociarse con un reactivo que facilite la estabilidad o la actualización celular. El reactivo puede ser lípidos, fosfato cálcico, polímeros catiónicos, DEAE-dextrano, dendrímeros, péptidos de penetración celular de polietilenglicol (PEG), microburbujas encapsuladas en gas o perlas magnéticas. La molécula donante puede incorporarse en una partícula vírica. El virus puede ser retroviral, adenoviral, vectores adenoasociados (AAV), herpes simplex, virus de la viruela, vector adenoviral híbrido, virus de Epstein-bar, lentivirus o virus herpes simplex.

En ciertas realizaciones, los vectores AAV como se describen en la presente pueden derivarse de cualquier AAV. En ciertas realizaciones, el vector AAV se deriva del virus parvovirus adenoasociado tipo 2, defectuoso y no patógeno. Todos estos vectores se derivan de un plásmido que conserva únicamente las repeticiones terminales invertidas de 145 bp del AAV que flanquean el casete de expresión del transgén. La transferencia génica eficiente y la administración estable de transgén debidas a la integración en los genomas de la célula transducida son características clave de este sistema vectorial. (Wagner et al., Lancet 351:9117 1702-3, 1998; Kearns et al., Gene Ther. 9:748-55, 1996). De acuerdo con la presente invención también pueden usarse otros serotipos de AAV, incluyendo AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9 y AAVrh.10 y cualquier serotipo de AAV novedoso. En algunas realizaciones, se usan AAV quiméricos en los que los orígenes virales de las secuencias de repetición terminal larga (LTR) del ácido nucleico viral son heterólogos al origen viral de las secuencias de la cápside. Ejemplos no limitativos incluyen virus quiméricos con LTR derivadas de AAV2 y cápsides derivadas de AAV5, AAV6, AAV8 o AAV9 (es decir, AAV2/5, AAV2/6, AAV2/8 y AAV2/9, respectivamente).

Los constructos descritos en la presente también pueden incorporarse en un sistema de vectores adenovirales. Los vectores basados en adenovirus son capaces de una eficiencia de transducción muy alta en muchos tipos de células y no requieren división celular. Con tales vectores, pueden obtenerse títulos y niveles de expresión elevados.

Los métodos y composiciones descritos en la presente son aplicables a cualquier organismo eucariota en el que se desee alterar el organismo mediante modificación genómica. Los organismos eucariotas incluyen plantas, algas, animales, hongos y protistas. Los organismos eucariotas también pueden incluir células vegetales, células de algas, células animales, células de hongos y células de protistas.

Células de mamífero ejemplares incluyen, pero no se limitan a, ovocitos no humanos, células K562, células CHO (ovario de hámster chino), células HEP-G2, células BaF-3, células Schneider, células COS (células de riñón de mono que expresan el antígeno T SV40), células CV-1, células HuTu80, células NTERA2, células NB4, células HL-60

y células HeLa, células 293 (véase, por ejemplo, Graham et al. (1977) J. Gen. Virol. 36:59), y células de mieloma como SP2 o NS0 (véase, por ejemplo, Galfre y Milstein (1981) Meth. Enzymol. 73(B):3 46). También pueden usarse mononucleocitos de sangre periférica (PBMC) o células T, así como células madre embrionarias y adultas. Por ejemplo, las células madre que pueden usarse incluyen células madre embrionarias (ES), células madre pluripotentes inducidas (iPSC), células madre mesenquimales, células madre hematopoyéticas, células madre hepáticas, células madre cutáneas y células madre neuronales.

Los métodos y composiciones descritos en la presente pueden usarse en la producción de organismos modificados. Los organismos modificados pueden ser mamíferos pequeños, animales de compañía, ganado y primates. Ejemplos no limitativos de roedores pueden incluir ratones, ratas, hámsteres, jerbos y cobayas. Ejemplos no limitativos de animales de compañía pueden incluir gatos, perros, conejos, erizos y hurones. Ejemplos no limitativos de ganado: caballos, cabras, ovejas, cerdos, llamas, alpacas y vacas. Ejemplos no limitativos de primates pueden incluir monos capuchinos, chimpancés, lémures, macacos, titíes, tamarinos, monos araña, monos ardilla y monos vervet. Los métodos y composiciones descritos en la presente pueden usarse en humanos.

Las plantas y células vegetales ejemplares que pueden modificarse usando los métodos descritos en la presente incluyen, entre otros, plantas monocotiledóneas (por ejemplo, trigo, maíz, arroz, mijo, cebada, caña de azúcar), plantas dicotiledóneas (por ejemplo, soja, patata, tomate, alfalfa), cultivos frutales (por ejemplo, tomate, manzana, pera, fresa, naranja), cultivos forrajeros (por ejemplo, alfalfa), cultivos de raíces vegetales (por ejemplo, zanahoria, patata, remolacha azucarera, batata), cultivos de verduras de hoja (por ejemplo, lechuga, espinaca); cultivos vegetales para el consumo (por ejemplo, soja y otras legumbres, calabaza, pimiento, berenjena, apio, etc.), plantas de flor (por ejemplo, petunia, rosa, crisantemo), coníferas y pinos (por ejemplo, abeto, picea); álamos (por ejemplo, *P. tremula* x *P. alba*); cultivos de fibras (algodón, yute, lino, bambú); plantas usadas en fitorremediación (por ejemplo, plantas acumuladoras de metales pesados); cultivos oleaginosos (por ejemplo, girasol, colza) y plantas usadas con propósitos experimentales (por ejemplo, *Arabidopsis*). Los métodos divulgados en la presente pueden usarse dentro de los géneros *Asparagus*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Cucurbita*, *Daucus*, *Erigeron*, *Glycine*, *Gossypium*, *Hordeum*, *Lactuca*, *Lolium*, *Lycopersicon*, *Malus*, *Manihot*, *Nicotiana*, *Orychophragmus*, *Oryza*, *Persea*, *Phaseolus*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Secale*, *Solanum*, *Sorghum*, *Triticum*, *Vitis*, *Vigna* y *Zea*. El término células vegetales incluye células vegetales aisladas, así como plantas enteras o partes de plantas enteras, como semillas, callos, hojas y raíces. La presente divulgación también abarca las semillas de las plantas descritas anteriormente en las que la semilla ha sido modificada usando las composiciones y/o métodos descritos en la presente. La presente divulgación abarca además la progenie, clones, líneas celulares o células de las plantas transgénicas descritas anteriormente en las que dicha progenie, clon, línea celular o célula tiene el transgén o constructo génico. Las especies de algas ejemplares incluyen microalgas, diatomeas, *Botryococcus braunii*, *Chlorella*, *Dunaliella tertiolecta*, *Gracilaria*, *Pleurochrysis carterae*, *Sargassum* y *Ulva*.

Los métodos descritos en este documento pueden incluir el uso de endonucleasas de corte raro para estimular la recombinación homóloga o la integración no homóloga de una molécula de transgén en un gen endógeno. La endonucleasa de corte raro puede incluir CRISPR, TALEN o nucleasas de dedos de zinc (ZFN). El sistema CRISPR puede incluir CRTSPR/Cas9 o CRTSPR/Cas 12a (Cpf1). El sistema CRTSPR puede incluir variantes que muestren una capacidad de PAM amplia (Hu et al., *Nature* 556, 57-63, 2018; Nishimasu et al., *Science* DOI: 10.1126, 2018) o una actividad de unión o escisión en el objetivo más alta (Kleinstiver et al., *Nature* 529:490-495, 2016). El reactivo de edición génica puede tener el formato de una nucleasa (Mali et al., *Science* 339:823-826, 2013; Christian et al., *Genetics* 186:757-761, 2010), nicasa (Cong et al., *Science* 339:819-823, 2013; Wu et al., *Biochemical and Biophysical Research Communications* 1:261-266, 2014), dímeros CRISPR-FokI (Tsai et al., *Nature Biotechnology* 32:569-576, 2014) o nicasas CRISPR emparejadas (Ran et al., *Cell* 154:1380-1389, 2013).

Los métodos y composiciones descritos en este documento pueden usarse en una circunstancia en la que se desee modificar el extremo 3' de la secuencia codificante de un gen endógeno. Por ejemplo, los pacientes con SCA3 o SCA6 tienen repeticiones CAG expandidas en los exones 10 (penúltimo exón) y 47 (último exón), respectivamente. Los pacientes con SCA3 o SCA6 pueden beneficiarse de la sustitución de los exones 10-11 y 47, respectivamente. En otros ejemplos, los pacientes con trastornos genéticos debidos a mutaciones de pérdida de función dentro del extremo 3' de un gen endógeno podrían beneficiarse de la sustitución de los exones finales de dicho gen.

La invención se describirá con más detalle en los siguientes ejemplos, que no limitan el alcance de la invención descrita en las reivindicaciones.

EJEMPLOS

Ejemplo 1: Integración dirigida de ADN en el gen ATXN3

Se construyeron tres plásmidos con transgenes diseñados para integrarse en el gen ATXN3 en células humanas. Todos los transgenes se diseñaron para ser insertados en el intrón 9 o en la unión del intrón 9 y el exón 10

del gen ATXN3 y todos los transgenes se diseñaron para insertar por lo menos un aceptor de corte y empalme y por lo menos una secuencia codificante funcional para los exones 10 y 11 del gen ATXN3. El primer plásmido, designado pBA1135, comprendía un brazo homólogo izquierdo y derecho con secuencia homóloga al extremo 3' del intrón 9 y al extremo 5' del intrón 10 (es decir, el direccionamiento del gen con éxito daría como resultado la eliminación del exón 10 y su sustitución por la secuencia de carga dentro de pBA1135). Entre los brazos de homología, de 5' a 3', había un aceptor de corte y empalme (aceptor de corte y empalme del intrón 9 de ATXN3), la secuencia de codificación para los exones 10 y 11 de ATXN3, el terminador de SV40, el terminador de BGH inverso, la secuencia de codificación inversa de los exones 10 y 11 (codón ajustado) y el aceptor de corte y empalme inverso. La secuencia para el transgén pBA1135 se muestra en la SEQ ID NO: 17. Se diseñó una nucleasa Cas9 correspondiente para escindir i) dentro del intrón 9 del gen ATXN3, ii) dentro del brazo de homología izquierdo de pBA1135, y iii) en el extremo 3' del brazo de homología derecho de pBA1135. Se esperaba que la escisión con éxito del plásmido liberara el transgén, permitiendo de este modo que la secuencia se usase como plantilla para la HR o para la integración mediante NHEJ. El sitio diana del ARNg de Cas9 se muestra en SEQ ID NO: 18. Los elementos individuales dentro de pBA1135 se muestran en las SEQ ID NO:44-51. La SEQ ID NO:44 comprende el brazo de homología izquierdo, el sitio diana de la nucleasa y el aceptor de corte y empalme. La SEQ ID NO:45 comprende la secuencia codificante parcial (exón 10 y 11) de un gen ATXN3 no patógeno. La SEQ ID NO:46 comprende la secuencia terminadora de SV40 p(A). La SEQ ID NO:47 comprende el terminador de BGH en complemento inverso. La SEQ ID NO:48 comprende la secuencia codificante parcial ajustada por codones, del complemento inverso, (exón 10 y 11) de un gen ATXN3 no patógeno. La SEQ ID NO:49 comprende la secuencia para el aceptor de corte y empalme. La SEQ ID NO:50 comprende la secuencia del brazo de homología derecho. La SEQ ID NO: 51 comprende la secuencia del sitio diana para la nucleasa. El segundo plásmido, denominado pBA1136, comprendía la misma carga que el pBA1135, pero se eliminaron los brazos de homología. Se mantuvieron los sitios diana de la nucleasa para facilitar la liberación del transgén del plásmido. Se esperaba que la escisión con éxito del plásmido liberase el transgén, permitiendo de este modo que la secuencia se usase para la integración por NHEJ en el gen ATXN3. La secuencia del pBA1136 se muestra en la SEQ ID NO: 19. El tercer plásmido, designado pBA1137, comprendía la misma secuencia que pBA1135, excepto por las secuencias inversas y el sitio diana de la nucleasa (es decir, terminador inverso, secuencia codificante inversa y aceptor de corte y empalme inverso). Para los métodos convencionales basados en HR se usó el plásmido pBA1137 como control. La secuencia de pBA1137 se muestra en la SEQ ID NO:20.

La transfección se realizó usando células HEK293T. Las células HEK293T se mantuvieron a 37°C y 5% de CO₂ en DMEM alto suplementado con 10% de suero fetal bovino (FBS). Las células HEK293T se transfectaron con 2 ug de donante, 2 ug de ARN guía (formato ARN) y 2 ug de Cas9 (formato ARN). Las transfecciones se realizaron usando electroporación. El ADN genómico se aisló 72 horas después de la transfección y se evaluó para detectar eventos de integración. En la Tabla 1 se muestra una lista de los cebadores usados para detectar la integración o el ADN genómico.

Tabla 1: Cebadores para detectar la integración de transgenes en ATXN3.

| Nombre de cebador | Secuencia (de 5' a 3') | SEQ ID NO: |
|-------------------|---------------------------|------------|
| oNJB043 | CAAAGGTGCCCTTGAGGTT | 21 |
| oNJB044 | AGGAGAAGTCTGCCGTTACT | 22 |
| oNJB113 | GGACAAACCACAACCTAGAATGC | 23 |
| oNJB114 | TAGGAAAGGACAGTGGGAGT | 24 |
| oNJB116 | CCATTATGTCTCAGTTGTTTCAGTG | 25 |
| oNJB156 | CCAGACCATCTCAGACACC | 26 |
| oNJB162 | GGCTGGGCTTCCAATTAC | 27 |
| oNJB167 | GTGGTTTGTCCAAACTCATCAA | 28 |
| oNJB170 | AGTAACTCTGCACTTCCCATTG | 29 |

Para detectar la integración de pBA1135, pBA1136 y pBA1137, se realizaron PCR en el ADN genómico. Respecto a pBA1137, el transgén se diseñó para que se integrase con precisión mediante HR. Por consiguiente, se detectaron bandas en las PCR de las uniones 5' y 3', que indican una inserción precisa en el exón 10 (FIG. 8 carriles 4 y 7). Los tamaños de banda esperados eran de 1.520 bp para la unión 5' y 786 bp para la unión 3'. Para la PCR de la unión 5' se usaron los cebadores oNJB113 y oNJB116. Para la PCR de la unión 3' se usaron los cebadores oNJB167 y oNJB170. En referencia a pBA1136, como no había brazos de homología, se predijo que el transgén se insertaría mediante inserción NHEJ. Se observaron bandas de tamaño adecuado para el transgén integrándose en las direcciones directa e inversa. La integración en la dirección directa puede observarse en la FIG. 8, carriles 3 (tamaño esperado de aproximadamente 1.520 pb) y 6 (tamaño esperado de aproximadamente 1.519 pb). La integración en la dirección inversa puede verse en la FIG. 8, carril 12 (tamaño esperado de aproximadamente 1.520 pb). Para la PCR de la unión 5' se usaron los cebadores oNJB113 y oNJB116. Para la PCR de la unión 3' se usaron

los cebadores oNJB114 y oNJB170. Para la PCR de la unión 5' inversa se usaron los cebadores oNJB116 y oNJB114. En referencia a ppBA1135, en el transgén había tanto brazos de homología como sitios de escisión por nucleasas. La integración por HR se observó mediante la detección de bandas en las PCR de las uniones 5' y 3' (FIG. 8 carriles 2 y 5). Además, la integración por NHEJ se observó mediante la detección de bandas en una PCR de unión 5' inversa (FIG. 8, carril 10). El tamaño esperado para la PCR de la unión 5' era de 1.520 bp. El tamaño esperado para la PCR de la unión 3' era de 1.157 bp. El tamaño esperado para la PCR de la unión 5' inversa era de aproximadamente 1.520 bp. Para la PCR de la unión 5' se usaron los cebadores oNJB113 y oNJB116. Para la PCR de la unión 3' se usaron los cebadores oNJB114 y oNJB170. Para la PCR de la unión 5' inversa se usaron los cebadores oNJB116 y oNJB114.

Los resultados muestran que los transgenes descritos que comprenden secuencias codificantes parciales bidireccionales pueden integrarse en el ADN genómico a través de múltiples vías de reparación diferentes.

Ejemplo 2: Integración dirigida de ADN en el gen CACNA1A

Se diseña un transgén dirigido a CACNA1A para sustituir el extremo 3' de la secuencia codificante de CACNA1A. Se construye un plásmido con un transgén diseñado para integrar la secuencia codificante WT en el intrón 46 o en el inicio del exón 47 (FIG. 4). El transgén comprende un primer brazo de homología que es homólogo a la secuencia inmediatamente después del sitio donante de corte y empalme en el intrón 46. El primer brazo de homología también comprende el sitio diana para una nucleasa (SEQ ID NO:9) y una secuencia aceptora de corte y empalme. El primer brazo de homología va seguido de una primera secuencia codificante que comprende el exón 47 de CACNA1A y una secuencia de repetición CAG no expandida (SEQ ID NO:3). Después de la primera secuencia codificante hay una secuencia de terminación poli(A) de SV40 (SEQ ID NO:4). En una orientación de cola a cola, hay un segundo conjunto de elementos funcionales. El comienzo del segundo conjunto de elementos comprende un sitio diana para la nucleasa (SEQ ID NO:9) seguido de un segundo brazo de homología. El segundo brazo de homología alberga 446 bp que es homólogo a la secuencia inmediatamente posterior a la codificación de parada (SEQ ID NO:8). Mediante análisis in silico, se determinó que esta secuencia no contenía secuencias aceptores de corte y empalme o de bifurcación de consenso. Después del segundo brazo de homología hay un segundo aceptor de corte y empalme del intrón 1 de beta-actina de carpa (SEQ ID NO:7). Después del aceptor de corte y empalme hay una versión optimizada del codón del exón 47 de CACNA1A (SEQ ID NO:6) y un terminador poli(A) de bGH (SEQ ID NO:5).

Se diseña una nucleasa Cas12a correspondiente para crear tres roturas de cadena doble después de la transfección del plásmido: i) dentro del intrón 46 del gen CACNA1A endógeno, 2) dentro del primer brazo de homología en el transgén pBA1011-D1, y 3) después del segundo brazo de homología en el transgén pBA1011-D1. En la SEQ ID NO:9 se muestra la secuencia diana para la nucleasa Cas12a.

La confirmación de la función del transgén y de los vectores CRISPR se logra mediante la transfección de células HEK293. Las células HEK293 se mantienen a 37°C y 5% de CO₂ en medio DMEM alto en glucosa sin L-glutamina sin piruvato sódico suplementado con 10% de suero fetal bovino (FBS) y 1% de solución de penicilina-estreptomicina (PS) 100X. Las células HEK293 se transfectan con cada uno de los constructos de plásmidos y combinaciones de los mismos usando Lipofectamine 3000. Dos días después de la transfección, se extrae el ADN y se evalúa en busca de mutaciones e inserciones dirigidas dentro del gen CACNA1A. La actividad de nucleasa se analiza usando el ensayo Cel-I o por secuenciación profunda de amplicones que comprenden la secuencia diana CRISPR/Cas12a. La integración con éxito del transgén se analiza mediante PCR (FIG. 5).

Ejemplo 3: Integración dirigida de ADN en el gen ATXN3

Se diseña un transgén dirigido a ATXN3 para sustituir el extremo 3' de la secuencia codificante de ATXN (exones 10 y 11). Se construye un plásmido con un transgén diseñado para integrar la secuencia codificante WT en el intrón 9 o en el inicio del exón 10 (FIG. 5). El transgén comprende un primer brazo de homología que es homólogo al intrón 9 de la secuencia (SEQ ID NO: 10). El primer brazo de homología también comprende el sitio diana para una nucleasa Cas12a y una secuencia aceptora de corte y empalme. El primer brazo de homología va seguido de una primera secuencia codificante que comprende los exones 10 y 11 de ATXN3 y una secuencia de repetición CAG no expandida. Después de la primera secuencia codificante hay una secuencia de terminación poli(A) de SV40. En una orientación de cola a cola, hay un segundo conjunto de elementos funcionales. El comienzo del segundo conjunto de elementos comprende un sitio diana para la nucleasa Cas12a seguido de un segundo brazo de homología. El segundo brazo de homología alberga 379 bp que son homólogos a la secuencia inmediatamente siguiente al final del exón 10 (es decir, el inicio del intrón 10). Se determinó mediante análisis in silico que esta secuencia tiene un número limitado de posibles secuencias de ramificación o aceptores de corte y empalme. Después del segundo brazo de homología hay un segundo aceptor de corte y empalme del intrón 1 de beta-actina de carpa. Después del aceptor de corte y empalme hay una versión con codón optimizado de los exones 10 y 11 de ATXN3 y un terminador poli(A) de bGH.

Se diseña una nucleasa Cas12a correspondiente para crear tres roturas de cadena doble después de la transfección del plásmido: i) dentro del intrón 9 del gen ATXN3 endógeno, 2) dentro del primer brazo de homología en el transgén pBA1012-D1, y 3) después del segundo brazo de homología en el transgén pBA1012-D1. En la SEQ ID NO: 11 se muestra la secuencia diana para la nucleasa Cas12a.

La confirmación de la función del transgén y de los vectores CRISPR se logra mediante la transfección de células HEK293. Las células HEK293 se mantienen a 37°C y 5% de CO₂ en medio DMEM alto en glucosa sin L-glutamina sin piruvato sódico suplementado con 10% de suero fetal bovino (FBS) y 1% de solución de penicilina-estreptomicina (PS) 100X. Las células HEK293 se transfectan con cada uno de los constructos de plásmidos y sus combinaciones usando Lipofectamine 3000. Dos días después de la transfección, se extrae el ADN y se evalúa en busca de mutaciones e inserciones dirigidas dentro del gen ATXN3. La actividad nucleasa se analiza usando el ensayo Cel-I o mediante secuenciación profunda de los amplicones que comprenden la secuencia diana CRISPR/Cas12a. La integración con éxito del transgén se analiza mediante PCR (FIG. 7).

Ejemplo 4: Integración dirigida de ADN en el gen ATXN3 usando transposasas Cas12k

Se diseña un transgén dirigido a ATXN3 para sustituir el extremo 3' de la secuencia codificante de ATXN (exones 10 y 11). Se construye un plásmido con un transgén diseñado para integrar la secuencia codificante WT en el intrón 9 o en el inicio del exón 10. El transgén comprende un extremo derecho y un extremo izquierdo del transposón, un primer y un segundo aceptores de corte y empalme, una primera y una segunda secuencias codificantes (que codifican aminoácidos de los exones 10 y 11), y un primer y un segundo terminadores. En la SEQ ID NO: 17 se muestra la secuencia entre los extremos derecho e izquierdo del transposón.

Los plásmidos se manipulan para que expresen los genes tnsB, tnsC, tniQ y Cas12k de *Scytonema hofmanni* (SEQ ID NO:30) usando promotores eucariotas. Un segundo plásmido está manipulado para que exprese el ARN guía Cas12k correspondiente (SEQ ID NO: 14). La secuencia diana del ARN guía es CCGCCGACCTTCACTTTC (SEQ ID NO: 15). Los plásmidos del transposón Cas12k se cotransforman en células HEK293 con un plásmido que alberga el transgén dirigido a ATXN3. Las células HEK293 se mantienen a 37°C y 5% de CO₂ en medio DMEM alto en glucosa sin L-glutamina sin piruvato sódico suplementado con 10% de suero fetal bovino (FBS) y 1% de solución de penicilina-estreptomicina (PS) 100X. Las células HEK293 se transfectan con cada uno de los constructos plasmídicos y sus combinaciones usando Lipofectamine 3000. Dos días después de la transfección, se extrae el ADN y se evalúa la presencia de inserciones dirigidas dentro del gen ATXN3. La integración del transgén se analiza usando PCR.

Ejemplo 5: Integración dirigida de ADN en el gen CACNA1A

Se diseña un transgén dirigido a CACNA1A para sustituir el extremo 3' de la secuencia codificante de CACNA1A. Se construye un plásmido con un transgén diseñado para integrar la secuencia codificante WT en el intrón 46 o en el inicio del exón 47. El transgén comprende un extremo derecho y un extremo izquierdo del transposón, un primer y un segundo aceptores de corte y empalme, una primera y una segunda secuencias codificantes (que codifican aminoácidos del exón 47), y un primer y un segundo terminadores.

Los plásmidos se manipulan para que expresen los genes tnsB, tnsC, tniQ y Cas12k de *Scytonema hofmanni* (SEQ ID NO:30) usando promotores eucariotas. Un segundo plásmido está diseñado para expresar el ARN guía Cas12k correspondiente (SEQ ID NO: 14). El ARN guía está diseñado para dirigirse a la secuencia CCCGGATCCCGGCTGTGACC (SEQ ID NO: 16). Los plásmidos del transposón Cas12k se cotransforman en células HEK293 con un plásmido que alberga el transgén dirigido a ATXN3. Las células HEK293 se mantienen a 37°C y 5% de CO₂ en medio DMEM alto en glucosa sin L-glutamina sin piruvato sódico suplementado con 10% de suero fetal bovino (FBS) y 1% de solución de penicilina-estreptomicina (PS) 100X. Las células HEK293 se transfectan con cada uno de los constructos plasmídicos y sus combinaciones usando Lipofectamine 3000. Dos días después de la transfección, se extrae el ADN y se evalúa la presencia de inserciones dirigidas en el gen ATXN3. La integración del transgén se analiza usando PCR.

REIVINDICACIONES

1. Un transgén que comprende:

- 5 una primera secuencia aceptora de corte y empalme y una segunda secuencia aceptora de corte y empalme situadas dentro del transgén en direcciones opuestas;
- 10 una primera secuencia codificante enlazada operativamente con la primera secuencia aceptora de corte y empalme y una segunda secuencia codificante enlazada operativamente con la segunda secuencia aceptora de corte y empalme, en donde la primera y la segunda secuencias codificantes están situadas dentro del transgén en direcciones opuestas; y
- 15 un primer terminador enlazado operativamente con la primera secuencia codificante y un segundo terminador enlazado operativamente con la segunda secuencia codificante, en donde el primer y el segundo terminadores están situados dentro del transgén en direcciones opuestas, o un terminador bidireccional enlazado operativamente con la primera y la segunda secuencias codificantes y situado dentro del transgén;
- 20 en donde la primera secuencia aceptora de corte y empalme flanquea la primera secuencia codificante y el primer terminador o el terminador bidireccional, en donde la segunda secuencia aceptora de corte y empalme flanquea la segunda secuencia codificante y el segundo terminador o el terminador bidireccional, y en donde la primera y la segunda secuencias codificantes son las mismas secuencias de ácido nucleico, o en donde la primera y la segunda secuencias codificantes son secuencias de ácido nucleico diferentes que codifican para la misma proteína usando la degeneración de codones.

25

30

35

40

45

50

55

60

65

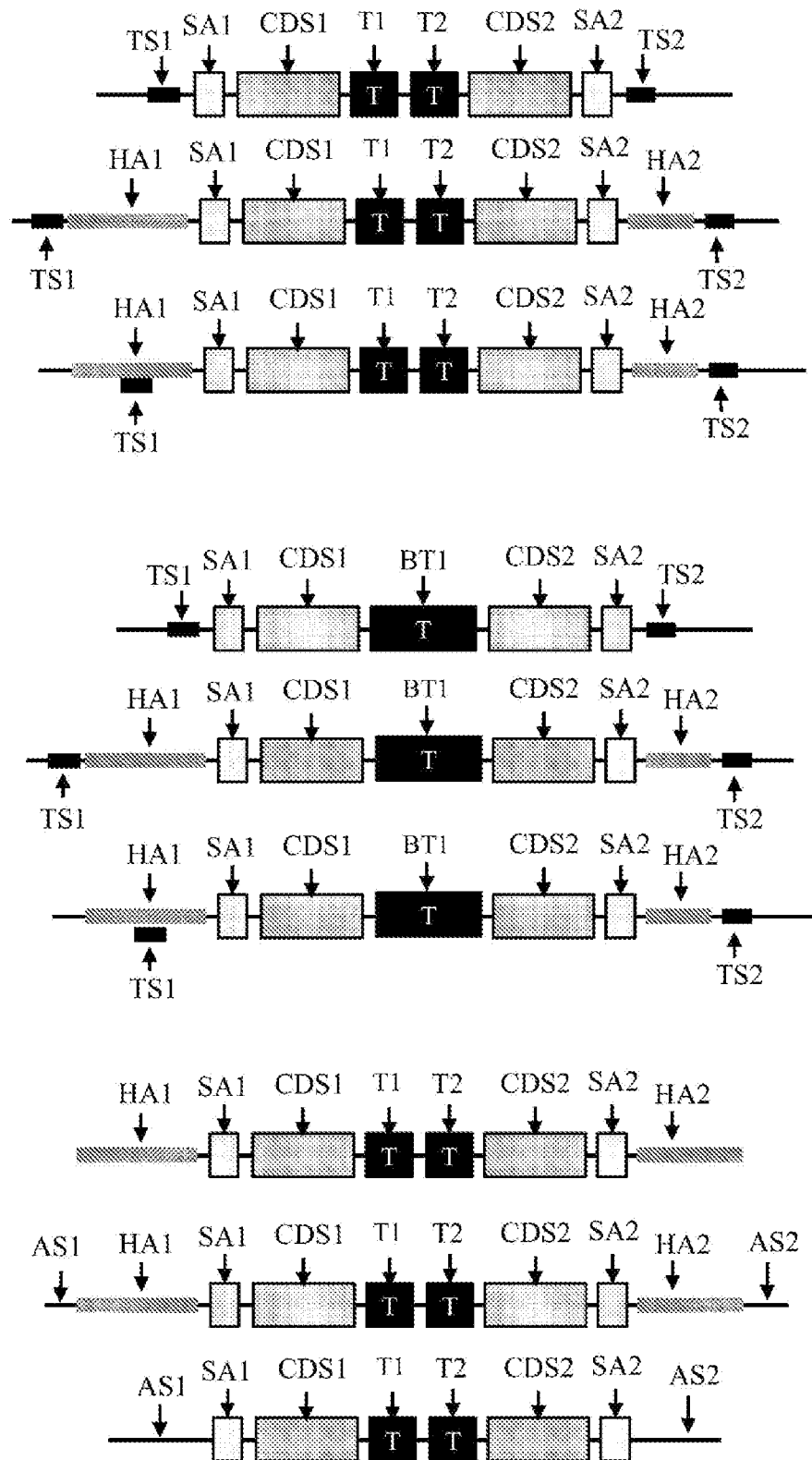


FIG. 1

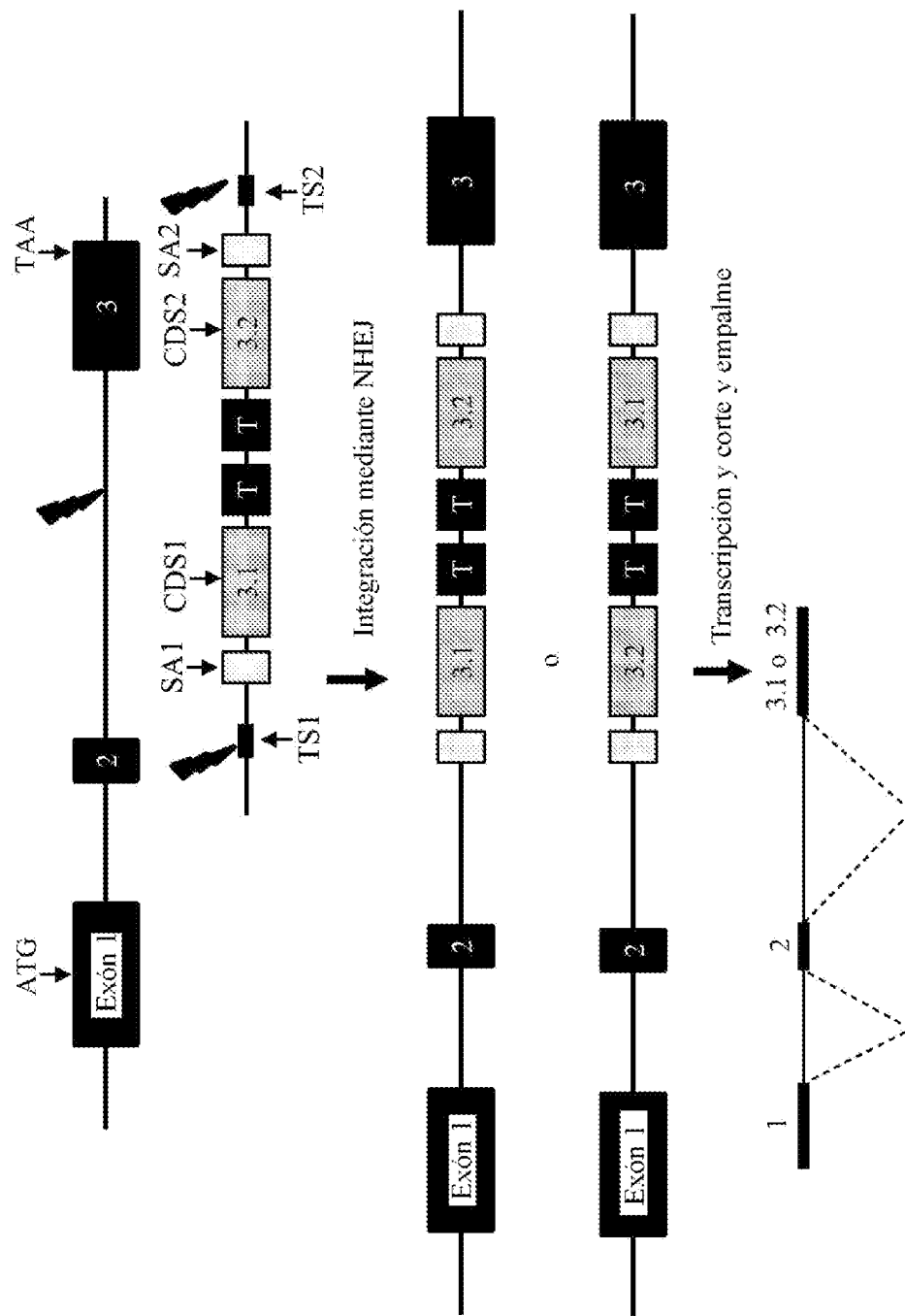
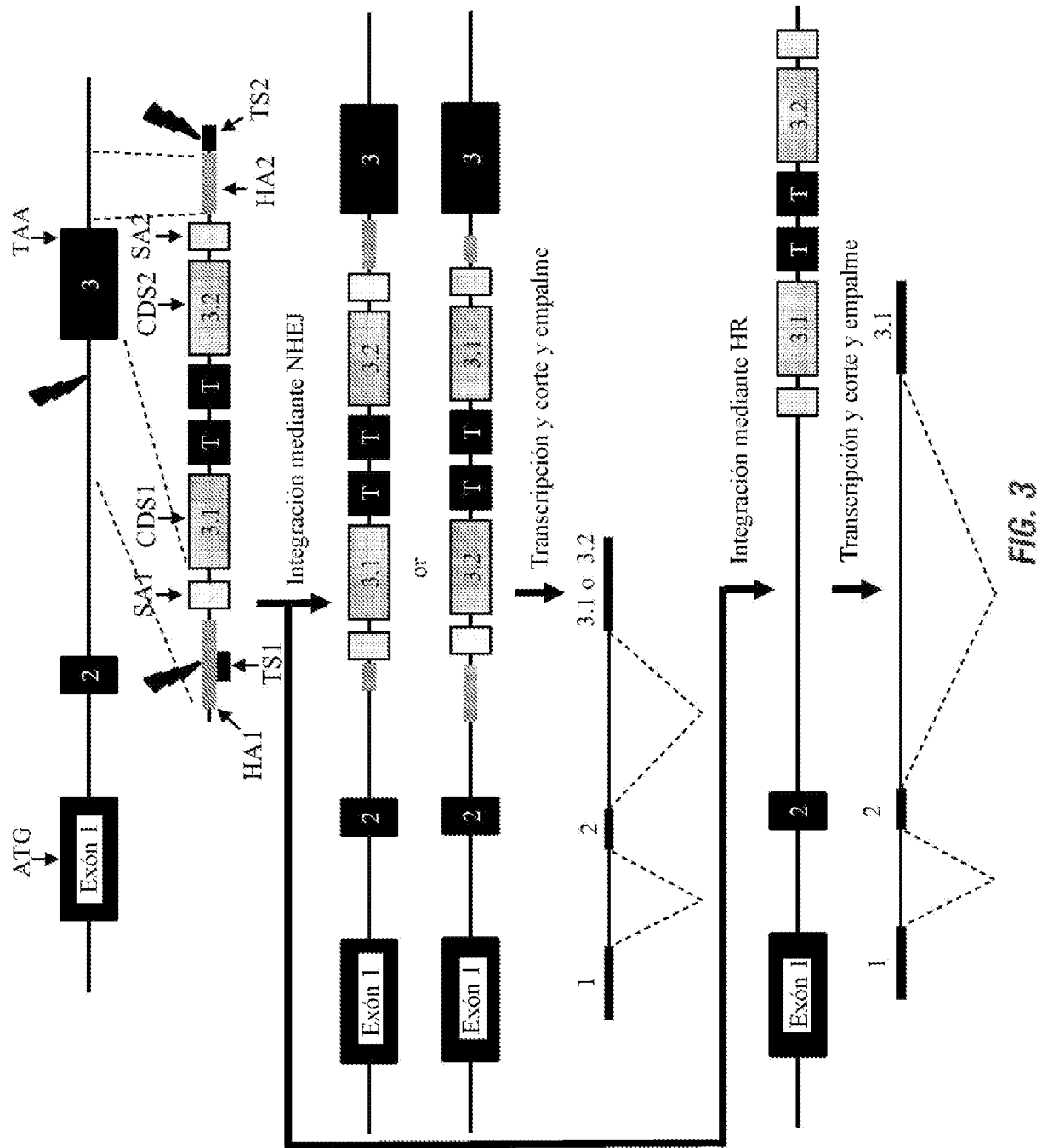


FIG. 2



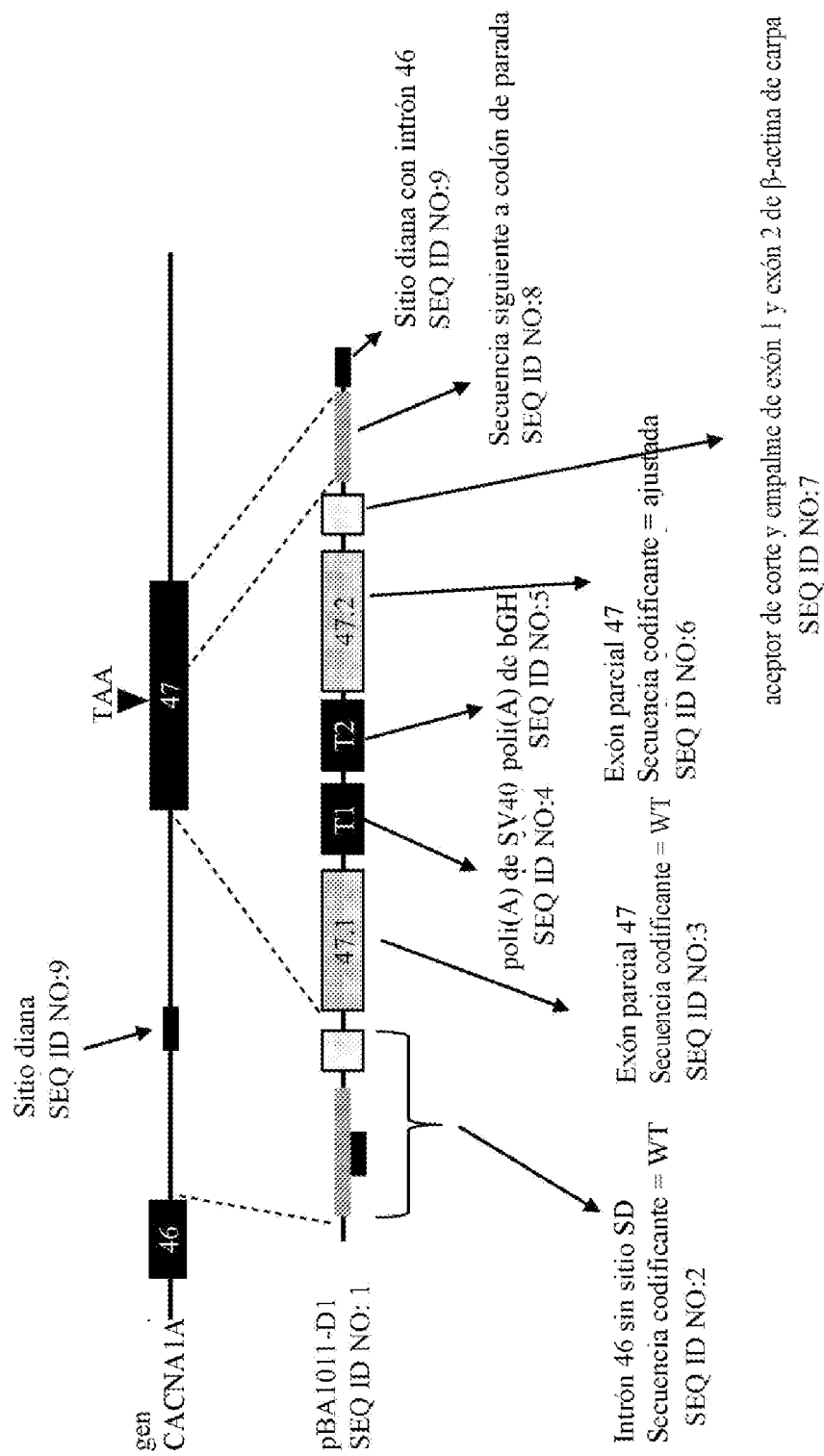


FIG. 4

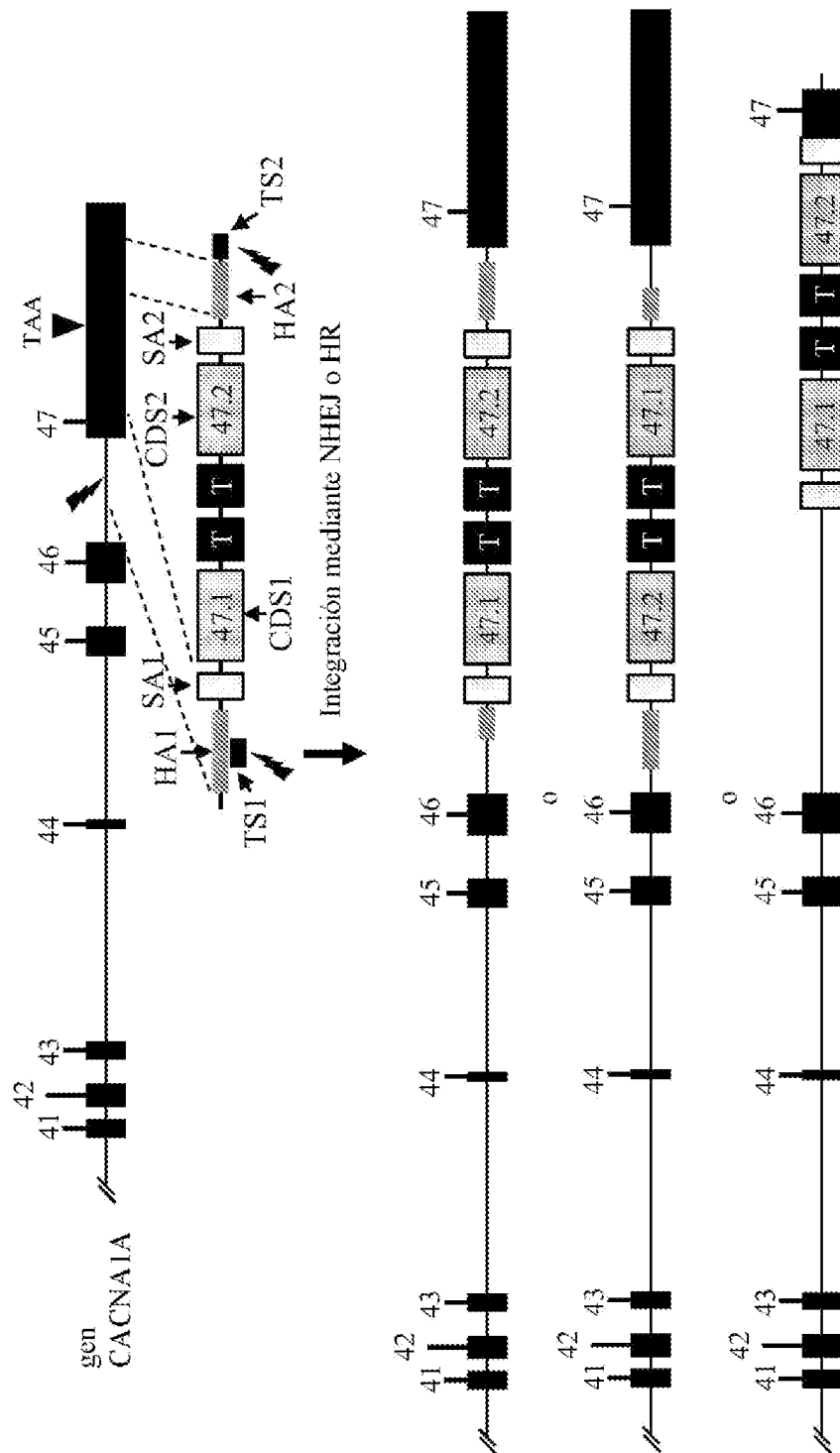


FIG. 5

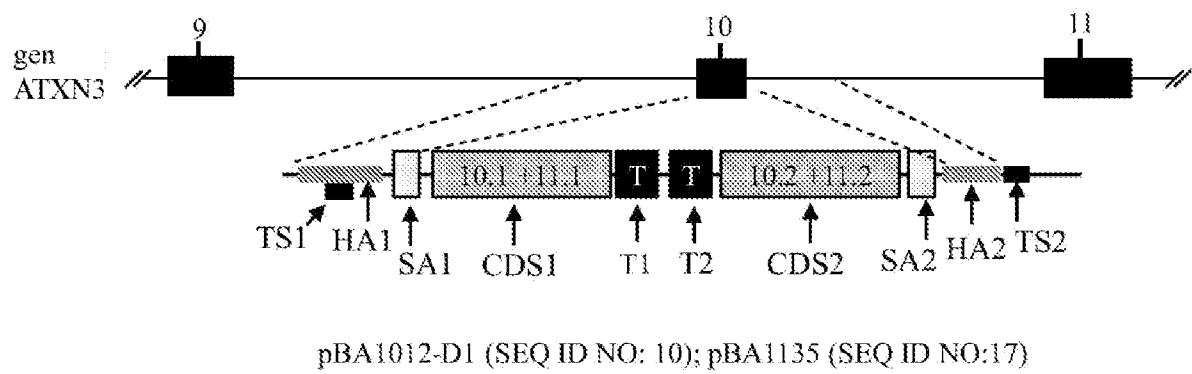


FIG. 6

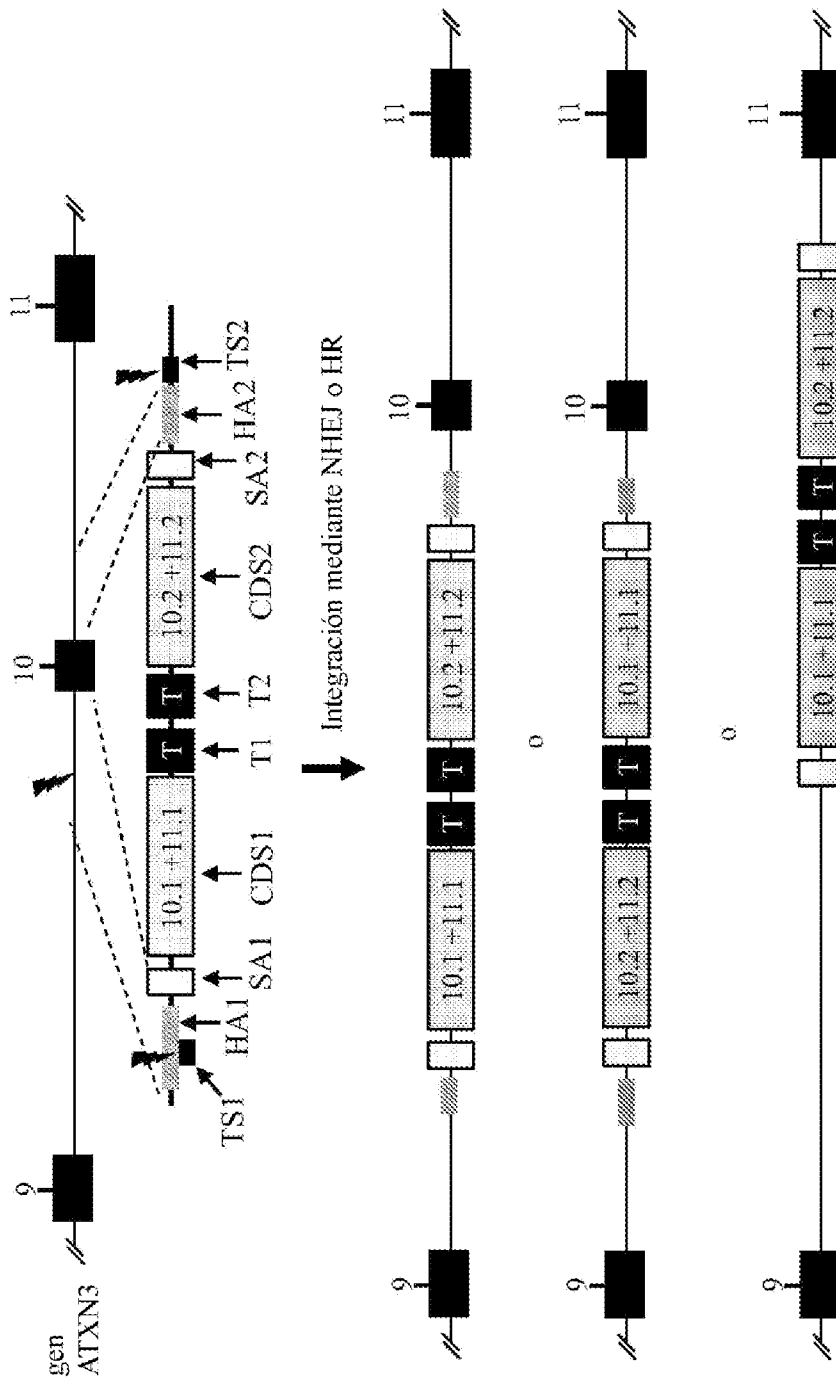


FIG. 7

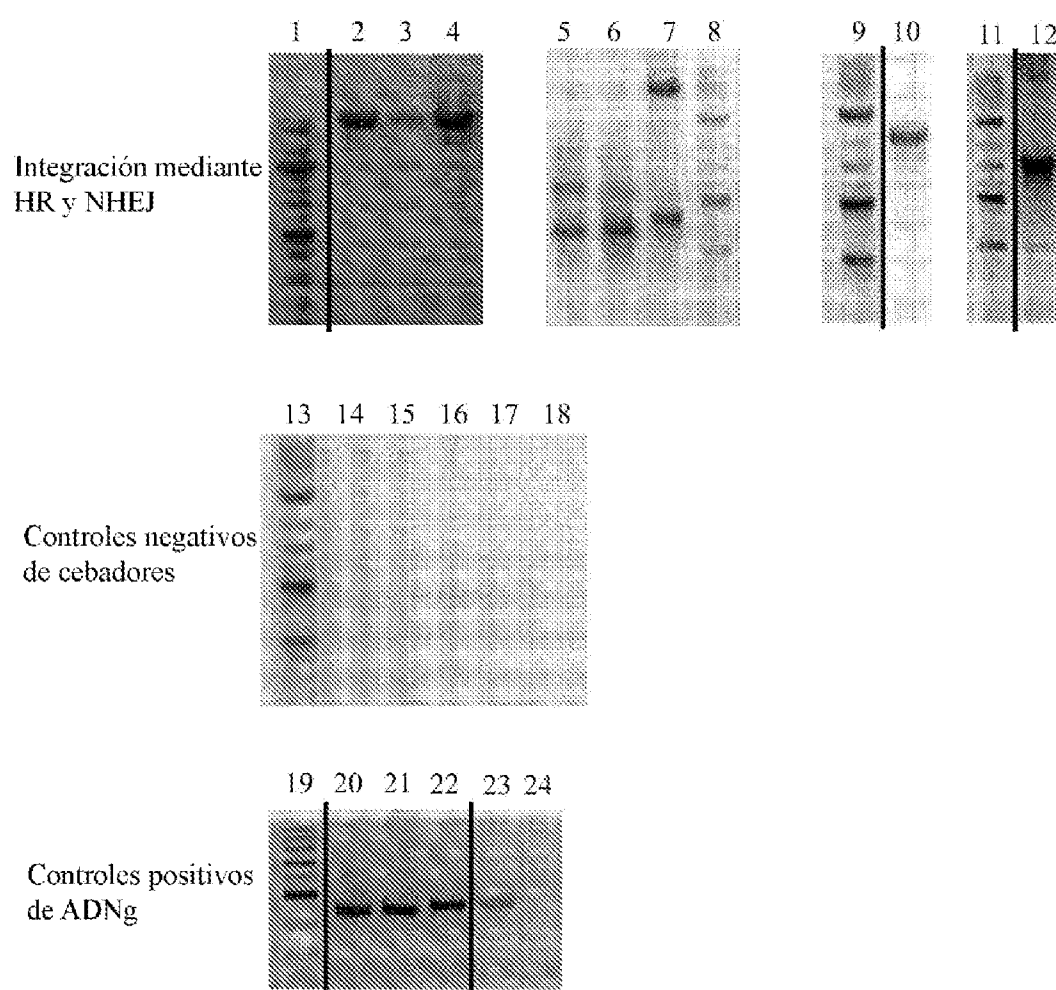


FIG. 8