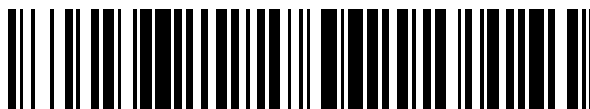


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 721 936**

51 Int. Cl.:

C12N 15/09 (2006.01)
A61K 38/46 (2006.01)
A61P 7/00 (2006.01)
A61K 35/12 (2015.01)
A61K 35/28 (2015.01)
C07K 14/47 (2006.01)
C12N 9/22 (2006.01)
C12N 15/113 (2010.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **13.11.2014** **PCT/US2014/065473**

87 Fecha y número de publicación internacional: **21.05.2015** **WO15073683**

96 Fecha de presentación y número de la solicitud europea: **13.11.2014** **E 14861930 (7)**

97 Fecha y número de publicación de la concesión europea: **27.02.2019** **EP 3068881**

54 Título: **Regulación de la expresión genética mediada por las nucleasas**

30 Prioridad:

13.11.2013 US 201361903823 P
26.08.2014 US 201462042075 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
06.08.2019

73 Titular/es:

CHILDREN'S MEDICAL CENTER CORPORATION
(50.0%)
55 Shattuck Street
Boston, MA 02115, US y
SANGAMO THERAPEUTICS, INC. (50.0%)

72 Inventor/es:

ORKIN, STUART H.;
REIK, ANDREAS y
URNOV, FYODOR

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 721 936 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

Regulación de la expresión genética mediada por las nucleasas

5 CAMPO TÉCNICO

[0001] La presente descripción está en el campo de la ingeniería del genoma, en particular la modificación dirigida del genoma de una célula hematopoyética.

10 ANTECEDENTES

[0002] Cuando se considera que los esfuerzos de secuenciación del genoma han revelado que el genoma humano contiene entre 20.000 y 25.000 genes, pero menos de 2.000 reguladores transcripcionales, se hace evidente que una serie de factores debe interactuar para controlar la expresión génica en todas sus diversas manifestaciones específicas temporales, del desarrollo y tisulares. La expresión de los genes se controla mediante una mezcla altamente compleja de reguladores transcripcionales generales y específicos, y la expresión también puede controlarse mediante elementos de ADN que actúan en cis. Estos elementos de ADN comprenden tanto elementos de ADN locales como el promotor central y sus sitios asociados de unión al factor de transcripción, así como elementos distales como potenciadores, silenciadores, aisladores y regiones de control de locus (LCR) (ver Matson et al (2006) Ann Rev Genome Hum Genet 7: 29-50).

[0003] Los elementos potenciadores se identificaron por primera vez en el genoma viral SV40, y luego que se encuentra en el locus de cadena pesada de inmunoglobulina humana. Ahora que se sabe que desempeñan funciones reguladoras en la expresión de muchos genes, los potenciadores parecen influir principalmente en los patrones temporales y espaciales de la expresión génica. También se ha encontrado que los potenciadores funcionan de una manera que no depende de la distancia del promotor central de un gen, y no depende de ninguna orientación de secuencia específica con respecto al promotor. Los potenciadores pueden ubicarse a varios cientos de kilobases hacia arriba o hacia abajo de una región promotora del núcleo, donde pueden ubicarse en una secuencia de intrones, o incluso más allá del extremo 3' de un gen.

[0004] Se han descrito varios métodos y composiciones para la escisión específica de ADN genómico. Dichos eventos de escisión dirigida pueden usarse, p. ej., para inducir mutagénesis dirigida, inducir delecciones dirigidas de secuencias de ADN celulares y facilitar la recombinación dirigida en un locus cromosómico predeterminado. Ver, p. ej., patentes de EE.UU. n° 8.623.618; 8.034.598; 8.586.526; 6.534.261; 6.599.692; 6.503.717; 6.689.558; 7.067.317; 7.262.054; 7.888.121; 7.972.854; 7.914.796; 7.951.925; 8.110.379; 8.409.861; Publicaciones de Patentes de Estados Unidos 20030232410; 20050208489; 20050026157; 20060063231; 20080159996; 201000218264; 20120017290; 20110265198; 20130137104; 20130122591; 20130177983 y 20130177960 y la solicitud de EE.UU. N° 14/278.903. Estos métodos a menudo implican el uso de sistemas de escisión diseñados para inducir una rotura de doble hebra (DSB) o una muesca en una secuencia de ADN diana, de modo que la reparación de la rotura por un proceso de nacimiento como la unión de extremo no homóloga (NHEJ) o la reparación usando una plantilla de reparación (reparación dirigida por homología o HDR) puede resultar en la desactivación de un gen o la inserción de una secuencia de interés (integración dirigida). Esta técnica también se puede utilizar para introducir cambios específicos del sitio en la secuencia del genoma mediante el uso de un oligonucleótido donante, incluida la introducción de delecciones específicas de regiones genómicas, o mutaciones puntuales específicas o alteraciones localizadas (también conocidas como corrección de genes). La escisión puede ocurrir mediante el uso de nucleasas específicas, como las nucleasas de zinc (ZFN), activadores de la transcripción, como las nucleasas efectoras (TALEN), o el uso del sistema CRISPR/Cas con un ARNcr/ARN tracr modificado ("ARN de guía única") para guiar la escisión específica. Además, nucleasas específicas se están desarrollando en base al sistema Argonaute (P. ej., a partir de *T. thermophilus*, conocido como 'TtAgo', ver Swarts et al (2014) Nature 507 (7491): 258-261), que también puede tener el potencial para usos en edición de genoma y terapia génica.

[0005] Las células rojas de la sangre (RBC), o eritrocitos, son el componente celular principal de la sangre. De hecho, los glóbulos rojos representan una cuarta parte de las células en un ser humano. Los RBC maduros carecen de un núcleo y muchos otros orgánulos en los seres humanos, y están llenos de hemoglobina, una metaloproteína que funciona para transportar oxígeno a los tejidos, así como para transportar dióxido de carbono de los tejidos y de regreso a los pulmones para su eliminación. Esta proteína constituye aproximadamente el 97% del peso seco de los glóbulos rojos y aumenta la capacidad de transporte de oxígeno de la sangre en aproximadamente setenta veces. La hemoglobina es un heterotetramero que comprende dos cadenas de globina similares a alfa (α) y dos cadenas de globina similares a beta (β) y 4 grupos hemo. En los adultos, el tetramero $\alpha_2\beta_2$ se conoce como hemoglobina A (HbA) o hemoglobina adulta. Típicamente, las cadenas de globina alfa y beta se sintetizan en una proporción aproximada de 1:1 y esta proporción parece ser crítica en términos de hemoglobina y estabilización de RBC. En un feto en desarrollo, se produce una forma diferente de hemoglobina, la hemoglobina fetal (HbF), que tiene una afinidad de enlace más alta para el oxígeno que la hemoglobina A, de manera que el oxígeno puede ser entregado al sistema del bebé a través del torrente sanguíneo de la madre. La hemoglobina fetal también contiene dos cadenas de globina α , pero en lugar de las cadenas de globina β adulta, tiene dos cadenas de gamma (γ)-globina fetal (es decir, la hemoglobina fetal es $\alpha_2\gamma_2$). Aproximadamente a las 30 semanas de gestación, la síntesis de

gamma globina en el feto comienza a disminuir a medida que aumenta la producción de beta globina. Aproximadamente a los 10 meses de edad, la hemoglobina del recién nacido es casi toda $\alpha_2\beta_2$, aunque algo de HbF persiste hasta la edad adulta (aproximadamente 1-3% de la hemoglobina total). La regulación del cambio de la producción de gamma a beta globina es bastante compleja, y principalmente implica una regulación a la baja de la transcripción de gamma globina con una regulación simultánea de la transcripción de la globina beta.

[0006] Los defectos genéticos en las secuencias que codifican las cadenas de hemoglobina pueden ser responsables de un número de enfermedades conocidas como hemoglobinopatías, incluyendo anemia de células falciformes y talasemias. En la mayoría de los pacientes con hemoglobina en las inopatías, los genes que codifican la gamma globina permanecen presentes, pero la expresión es relativamente baja debido a la represión génica normal que se produce alrededor del parto como se describió anteriormente.

[0007] Se estima que 1 de cada 5.000 personas en los EE.UU. tiene la enfermedad de células falciformes (ECF), sobre todo en personas de origen africano subsahariano. Parece haber un beneficio para los portadores heterocigotos de la mutación de células falciformes para la protección contra la malaria, por lo que este rasgo puede haber sido seleccionado positivamente con el tiempo, de modo que se estima que en el África subsahariana, un tercio de la población tiene el rasgo depreanocítico. La enfermedad de células falciformes es causada por una mutación en el gen de la globina β como consecuencia de la cual la valina se sustituye por el ácido glutámico en el aminoácido nº 6 (un GAG a GTG a nivel de ADN), donde la hemoglobina resultante se conoce como "hemoglobina S" o "HbS". En condiciones de oxígeno más bajas, un cambio conformacional en la forma desoxi de HbS expone un parche hidrófobo en la proteína entre las hélices E y F. Los residuos hidrófobos de la valina en la posición 6 de la cadena beta en la hemoglobina pueden asociarse con el parche hidrófobo, lo que hace que las moléculas de HbS se agreguen y formen precipitados fibrosos. Estos agregados, a su vez, causan la anomalía o la "enfermedad" de los glóbulos rojos, lo que resulta en una pérdida de flexibilidad de las células. Los eritrocitos no son capaces de apretar los lechos capilares y pueden ocasionar una crisis vaso-oclusiva en pacientes con células falciformes. Además, los eritrocitos falciformes son más frágiles que los eritrocitos normales y tienden a la hemólisis, lo que finalmente lleva a la anemia en el paciente.

[0008] El tratamiento y manejo de pacientes con células falciformes es una proposición de por vida que involucra tratamiento con antibióticos, manejo del dolor y transfusiones durante episodios agudos. Un enfoque es el uso de la hidroxiurea, que ejerce sus efectos en parte al aumentar la producción de gamma globina. Sin embargo, aún se desconocen los efectos secundarios a largo plazo de la terapia crónica con hidroxiurea, y el tratamiento produce efectos secundarios no deseados y puede tener una eficacia variable de un paciente a otro. A pesar de un aumento en la eficacia de los tratamientos con células falciformes, la expectativa de vida de los pacientes todavía es de mediados a finales de los 50 y la morbilidad asociada de la enfermedad tiene un impacto profundo en la calidad de vida de un paciente.

[0009] Las talasemias también son enfermedades relacionadas con la hemoglobina y típicamente implican una expresión reducida de las cadenas de globina. Esto puede ocurrir a través de mutaciones en las regiones reguladoras de los genes o de una mutación en una secuencia codificadora de globina que produce una expresión reducida o niveles reducidos o proteína globina funcional. Las talasemias alfa se asocian principalmente con personas de África occidental y de ascendencia del sur de Asia, y pueden conferir resistencia a la malaria. La talasemia beta está asociada principalmente con personas de ascendencia mediterránea, típicamente de Grecia y las zonas costeras de Turquía e Italia. El tratamiento de las talasemias generalmente involucra transfusiones de sangre y terapia de quelación de hierro. Los trasplantes de médula ósea también se están utilizando para el tratamiento de personas con talasemias graves si se puede identificar un donante adecuado, pero este procedimiento puede tener riesgos significativos.

[0010] Un enfoque que se ha propuesto para el tratamiento tanto de ECF y talasemias beta es aumentar la expresión de globina gamma con el objetivo de que HbF reemplace funcionalmente la hemoglobina adulta aberrante. Como se mencionó anteriormente, se cree que el tratamiento de los pacientes con ECF con hidroxiurea es exitoso en parte debido a su efecto en el aumento de la expresión de gamma globina. El primer grupo de compuestos que se descubrió que afectaban la actividad de reactivación de la globina gamma fueron los fármacos citotóxicos. La capacidad para causar la *síntesis de novo* de globina gamma por manipulación farmacológica se demostró por primera vez utilizando 5-azacitidina en animales experimentales (DeSimone (1982) Proc Nat'l Acad Sci EE.UU. 79 (14): 4428-31). Los estudios subsiguientes confirmaron la capacidad de la 5-azacitidina para aumentar la HbF en pacientes con β -talasemia y enfermedad de células falciformes (Ley, et al., (1982) N. Engl. J. Medicine, 307:1469-1475, y Ley, et al., (1983) Blood 62: 370-380). Además, en sistemas experimentales se ha demostrado que los ácidos grasos de cadena corta (p. ej., butirato y derivados) aumentan la HbF (Constantoulakis et al., (1988) Blood 72 (6): 1961-1967). Además, hay un segmento de la población humana con una condición conocida como "persistencia hereditaria de la hemoglobina fetal" (HPFH) donde persisten cantidades elevadas de HbF en la edad adulta (10-40% en los heterocigotos HPFH (ver Thein et al (2009) Hum. Mol. Genet 18 (R2): R216-R223. Se trata de una condición rara, pero en ausencia de cualquier anomalía asociada de la globina beta, no se asocia con ninguna manifestación clínica significativa, incluso cuando el 100% de la hemoglobina del individuo es HbF. Cuando las personas que tienen una beta talasemia también tienen HPFH coincidente, la expresión de HbF puede disminuir la gravedad de la enfermedad. Además, la gravedad del curso natural de la enfermedad de células falciformes puede variar

significativamente de un paciente a otro, y esta variabilidad, en parte, puede atribuirse al hecho de que algunos individuos con enfermedad más leve expresan niveles más altos de HbF.

[0011] Un enfoque para aumentar la expresión de HbF implica la identificación de genes cuyos productos desempeñan un papel en la regulación de la expresión de globina gamma. Uno de estos genes es BCL11A, que se identificó por primera vez debido a su papel en el desarrollo de linfocitos. BCL11A codifica una proteína de dedo de zinc que se cree que está involucrada en la regulación específica de la etapa de desarrollo de la expresión de gamma globina. BCL11A se expresa en células precursoras eritroides adultas y la regulación a la baja de su expresión conduce a un aumento en la expresión de gamma globina. Además, parece que el empalme del ARNm de BCL11A está regulado por el desarrollo. En las células embrionarias, parece que las variantes de ARNm BCL11A más cortas, conocidas como BCL11A-S y BCL11A-XS se expresan de forma primaria, mientras que en las células adultas, las variantes de ARNm más largas BCL11A-L y BCL11A-XL se expresan predominantemente. Ver, Sankaran *et al* (2008) Science 322 p. 1839. La proteína BCL11A parece interactuar con el locus de la globina beta para alterar su conformación y, por lo tanto, su expresión en diferentes etapas de desarrollo. Se ha propuesto el uso de un inhibidor ARN dirigido al gen BCL11A (véase, p. ej., Publicación de Patente de EE.UU. 20110182867) pero esta tecnología tiene varios inconvenientes potenciales, a saber, que el derribo completo no se puede conseguir, la entrega de tales ARN puede ser problemática y los ARN deben estar presentes continuamente, requiriendo múltiples tratamientos de por vida.

[0012] La orientación de las secuencias potenciadoras BCL11A puede proporcionar un mecanismo para aumentar la HbF. P. ej., los estudios de asociación amplia del genoma han identificado un conjunto de variaciones genéticas en BCL11A que están asociadas con un aumento de los niveles de HbF. Estas variaciones son una colección de SNP que se encuentran en las regiones no codificantes de BCL11A que funcionan como una región potenciadora de linaje restringido, específica de la etapa. Una investigación adicional reveló que este potenciador BCL11A se requiere en células eritroides para la expresión de BCL11A, pero no se requiere para su expresión en células B (consulte Bauer *et al*, (2013) Science 343: 253-257). La región potenciadora se encontró dentro del intrón 2 del gen BCL11A, y se identificaron tres áreas de hipersensibilidad a ADNaseI (a menudo indicativas de un estado de cromatina que está asociado con un potencial regulador) en el intrón 2. Estas tres áreas se identificaron como "+62", "+58" y "+55" de acuerdo con la distancia en kilobases desde el sitio de inicio de la transcripción de BCL11A. Estas regiones potenciadoras son aproximadamente 350 (+55); 550 (+58); y 350 (+62) nucleótidos de longitud (Bauer 2013, *ibid*).

[0013] Por lo tanto, sigue habiendo una necesidad de métodos y composiciones adicionales que pueden utilizar estos estudios de asociación en todo el genoma para la edición de genoma y la alteración de la expresión génica por ejemplo para tratar hemoglobinopatías tales como la enfermedad de células falciformes y la talasemia beta.

RESUMEN

[0014] La presente invención se define en las reivindicaciones y se refiere a composiciones y métodos para uso en terapia génica y la ingeniería del genoma. Específicamente, los métodos y composiciones descritos se refieren a la inactivación (p. ej., mediante la supresión completa o parcialmente su expresión) de un gen, por ejemplo un gen que actúa como regulador de uno o más genes. En particular, la invención se refiere a métodos y composiciones para interferir con la función potenciadora en un gen BCL11A para disminuir o eliminar su actividad en linajes celulares específicos. Además, la invención se refiere a métodos y composiciones para interferir con las funciones potenciadoras de BCL11A en las que las secuencias potenciadoras no están localizadas dentro del gen BCL11A. La regulación descendente resultante del gen BCL11A en estas circunstancias a su vez da como resultado un aumento de la expresión de gamma globina.

La invención proporciona una célula modificada genéticamente que comprende una modificación genómica realizada por una nucleasa de dedo de cinc (ZFN) o una nucleasa de dominio efector TAL (TALEN), en la que la modificación genómica se encuentra dentro de cualquiera de las secuencias diana de nucleasas de las SEQ ID NO: 4, 80, 143-184, 232-251 y 276 de una secuencia potenciadora BCL11a endógena, y además en la que la modificación genómica es una inserción y/o eliminación. En algunas realizaciones de la invención, la célula es una célula madre, tal como una célula madre hematopoyética (p. ej., una célula madre hematopoyética CD34+).

La invención también proporciona una célula diferenciada modificada genéticamente que desciende de una célula madre de la invención, opcionalmente en la que la célula es un glóbulo rojo (RBC). La invención proporciona además una composición farmacéutica que comprende una célula modificada genéticamente de la invención. La invención proporciona una proteína de unión a ADN que comprende una proteína de dedo de zinc o una proteína efectora de TALE (TALE), en donde (i) la proteína de dedo de zinc comprende 4, 5 o 6 dominios de dedo de zinc que comprenden una región de hélice de reconocimiento, en donde la proteínas de dedo de zinc comprende las regiones de hélice de reconocimiento en el orden mostrado en una sola fila de la Tabla 3 o la Tabla 6; y (ii) la proteína TALE comprende una pluralidad de unidades de repetición TALE, cada unidad de repetición que comprende una región dresídica hipervariable (RVD), en la que los RVD de las unidades de repetición TALE se muestran en una sola fila de la Tabla 1, Tabla 2 o Tabla 4.

La invención también proporciona una nucleasa que comprende una proteína de dedo de cinc o proteína TALE de la invención y un dominio de escisión de tipo salvaje o de ingeniería o un semidominio de escisión.

La invención proporciona además un polinucleótido que codifica una o más proteínas de unión a ADN o una

nucleasa de la invención. La invención proporciona una célula aislada que comprende una o más proteínas de unión a ADN o uno o más polinucleótidos de la invención, opcionalmente en donde la célula es una célula madre hematopoyética.

La invención también proporciona un kit que comprende al menos uno de: i) un polinucleótido que codifica una proteína de unión a ADN de la invención; o ii) una proteína de unión al ADN de la invención.

La invención proporciona además un método *in vitro* para alterar la expresión génica de globina en una célula, comprendiendo el método: introducir, en la célula, uno o más polinucleótidos de la invención, en condiciones tales que se expresan una o más proteínas y se altera la expresión del gen de la globina, opcionalmente, en donde aumenta la expresión del gen de la globina y/o el gen de la globina es un gen de globina gamma o globina beta.

La invención proporciona una composición farmacéutica de la invención para uso en un método para tratar o prevenir una hemoglobinopatía en un paciente que necesita un aumento en la expresión génica de globina, comprendiendo el método administrar al paciente la composición farmacéutica en una cantidad suficiente para aumentar la expresión génica de globina en el paciente, opcionalmente en donde la hemoglobinopatía es una talasemia o enfermedad de células falciformes, como la β -talasemia.

La invención también proporciona una o más nucleasas de la invención, o un ácido nucleico que codifica una o más nucleasas que se unen a una secuencia diana de cualquiera de las SEQ ID NO: 4-80, 143-184, 232-251 y 276, para uso en el tratamiento o prevención de una hemoglobinopatía, en donde la nucleasa se dirige a una secuencia potenciadora de BCL11A en una célula, donde la una o más nucleasas es una ZFN o TALEN.

[0015] En algunos aspectos, al menos una nucleasa (es decir, una nucleasa que se une a una secuencia de promotor de BCL11A) se entrega a una célula de células madre o precursor humano (HSC/PC) con el propósito de la ingeniería del genoma. La nucleasa puede reconocer una secuencia diana que comprende al menos 9 (p. ej., 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 o incluso más..) pares de bases contiguos de la SEQ ID NO: 1 o SEQ ID NO: 2. Las secuencias diana se muestran en las Tablas 1, 2, 3, 4 y 6. En ciertas realizaciones, la nucleasa comprende una proteína de unión a ADN que comprende una proteína de dedo de zinc que comprende 4, 5 o 6 dominios de dedo de zinc que comprenden una región de hélice de reconocimiento, en la que las proteínas de dedos de zinc comprenden las regiones de hélice de reconocimiento en el orden mostrado en una sola fila de la Tabla 3 o la Tabla 6. En otras realizaciones, la nucleasa comprende una proteína TALE que comprende una pluralidad de unidades de repetición TALE, comprendiendo cada unidad de repetición una región dieresidua hipervariable (RVD), donde los RVD de las unidades de repetición TALE se muestran en una sola fila de la Tabla 1, la Tabla 2 o la Tabla 4. La(s) nucleasa(s) como se describen en este documento pueden comprender además un enlazador (p. ej., entre el dominio de unión al ADN y el dominio de escisión), p. ej., un enlazador como se muestra en las Figuras 14 y 17.

[0016] En algunas realizaciones, la nucleasa se administra como un péptido, mientras que en otras se administra como un ácido nucleico que codifica la al menos una nucleasa. En algunas realizaciones, se usa más de una nucleasa. En algunas realizaciones preferidas, el ácido nucleico que codifica la nucleasa es un ARNm, y en algunos casos, el ARNm está protegido. En algunos aspectos, el ARNm puede modificarse químicamente (véase, p. ej., Kormann et al, (2011) Nature Biotechnology 29 (2): 154-157). En otros aspectos, el ARNm puede comprender una tapa ARCA (véanse las patentes de EE.UU. 7.074.596 y 8.153.773). En otras realizaciones, el ARNm puede comprender una mezcla de nucleótidos modificados y no modificados (véase la Publicación de Patente de Estados Unidos 2012-0195936). La nucleasa puede comprender una nucleasa de dedo de zinc (ZFN) o una nucleasa TALE (TALEN) como se define en las reivindicaciones, o una combinación de la nucleasa de dedo de zinc (ZFN), la nucleasa TALE (TALEN) o un sistema de nucleasa CRISPR/Cas. En una realización preferida, el ácido nucleico que codifica la(s) nucleasa(s) se administra a la HSC/PC a través de la electroporación. En algunas realizaciones, la nucleasa se escinde en o cerca del sitio de unión del factor de transcripción. En algunos aspectos, el factor de transcripción es GATA-1,

[0017] Se divulga una célula o línea celular en la que se modifica una secuencia potenciadora de BCL11A endógena, por ejemplo en comparación con la secuencia de tipo salvaje de la célula. La célula o las líneas celulares pueden ser heterocigotas u homocigotas para la modificación. Las modificaciones pueden comprender inserciones, deleciones y/o combinaciones de las mismas. En algunas realizaciones preferidas de la divulgación, las inserciones, deleciones y/o combinaciones de las mismas dan como resultado la destrucción de un sitio de unión al factor de transcripción. En ciertas realizaciones de la divulgación, la secuencia potenciadora BCL11A es modificada por una nucleasa (p. ej., ZFN, TALEN, sistema CRISPR/Cas, sistema Ttago, etc.). En ciertas realizaciones de la divulgación, el potenciador BCL11A se modifica en cualquier lugar entre el exón 2 y el exón 3. En otras realizaciones de la divulgación, el potenciador BCL11A se modifica en las regiones mostradas en la SEQ ID NO:1, SEQ ID NO: 2 o SEQ. ID NO: 3 (Figura 11). En ciertas realizaciones de la divulgación, la modificación es en o cerca de los sitios de unión y/o escisión de las nucleasas, p. ej., dentro de 1-300 (o cualquier valor entre ellos) pares de bases corriente arriba o corriente abajo del (de los) sitio(s) de escisión, más preferiblemente dentro de 1-100 pares de bases (o cualquier valor entre ellos) de cualquiera de los lados de los sitios de unión y/o escisión, incluso más preferiblemente dentro de 1 a 50 pares de bases (o cualquier valor entre ellos) a cada lado de los sitios de unión y/o escisión. En ciertas realizaciones de la divulgación, la modificación es en o cerca de la región "+58" del potenciador BCL11A, p. ej., en o cerca de un sitio de unión a nucleasa mostrado en cualquiera de las SEQ ID NO: 4 a 80 y 362. En otras realizaciones de la divulgación, la modificación es en o cerca de la región "+55" del potenciador BCL11A, p. ej., en o cerca de un sitio de nucleasa mostrado en cualquiera de las SEQ ID NO: 143 a 184 y 232-251. En otras realizaciones adicionales de la divulgación, la modificación se produce en otras secuencias potenciadoras BCL11A.

Cualquier célula o línea celular puede modificarse, p. ej., una célula madre (célula madre hematopoyética). Las células parcial o completamente diferenciadas descendientes de las células madre modificadas como se describe en el presente documento también se proporcionan (p. ej., eritrocitos o células precursoras de glóbulos rojos). Cualquiera de las células modificadas o líneas celulares descritas en el presente documento puede mostrar un aumento en la expresión de globina gamma. También se proporcionan composiciones tales como composiciones farmacéuticas que comprenden las células modificadas genéticamente como se describe en el presente documento.

[0018] En otros aspectos, un ácido nucleico donante puede ser administrado a una célula diana. El donante puede administrarse antes, después o junto con el ácido nucleico que codifica la(s) nucleasa(s). El ácido nucleico donante puede comprender una secuencia exógena (transgén) para integrarse en el genoma de la célula, p. ej., un locus endógeno. En algunas realizaciones, el donante puede comprender un gen de longitud completa o un fragmento del mismo flanqueado por regiones de homología con el sitio de escisión dirigido. En algunas realizaciones, el donante carece de regiones homólogas y está integrado en un locus diana a través del mecanismo independiente de homología (es decir, NHEJ). El donante puede comprender cualquier secuencia de ácido nucleico, p. ej., un ácido nucleico que, cuando se usa como un sustrato para la reparación dirigida por homología de la rotura de doble cadena inducida por nucleasa, conduce a una delección especificada por el donador que se generará en el locus de cromosoma endógeno (p. ej., región potenciadora BCL11A) o, alternativamente (o además de), nuevas formas alélicas de (p. ej., mutaciones puntuales que anulan un sitio de unión al factor de transcripción) el locus endógeno que se creará. En algunos aspectos, el ácido nucleico donante es un oligonucleótido en el que la integración conduce a un evento de corrección génica, o una eliminación dirigida.

[0019] En otros aspectos, la nucleasa y/o donante es (son) entregada(s) por métodos de transferencia génica viral y/o no viral. En realizaciones preferidas, el donante se administra a la célula a través de un virus asociado a adeno (AAV). En algunos casos, el AAV comprende LTR que son de un serotipo heterólogo en comparación con el serotipo de la cápside.

[0020] Los métodos y composiciones descritos en este documento comprenden una o más nucleasas (es decir, ZFNs y/o TALENs) dirigidas a regiones específicas de la región BCL11A potenciadora. En algunas realizaciones, el uno o más pares de nucleasas se dirigen a secuencias que resultan en la modificación de la región potenciadora mediante su eliminación en su totalidad, mientras que en otras realizaciones, subsecciones del potenciador se eliminan. En algunas realizaciones, la eliminación comprende una de las regiones de hipersensibilidad a +55 o +58 ADNasal de la región potenciadora, o más de una de las regiones de hipersensibilidad a +55, +58 o +62 ADNasal. En algunas realizaciones, solo se eliminan los +55 o solo los +58. En otras realizaciones, se eliminan dos de las regiones (p. ej., +55 y +58; +58 y +62; o +55 y +62).

[0021] En algunos aspectos, se hacen delecciones que comprenden regiones dentro de las regiones hipersensibles ADNasal del potenciador. Estas delecciones pueden comprender desde aproximadamente 1 nucleótido hasta aproximadamente 551 nucleótidos. Por lo tanto, las eliminaciones pueden comprender, 1, 5, 10, 15, 20, 25, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 550, 550 nucleótidos, o cualquier valor entre ellos. En algunas realizaciones, las eliminaciones comprenden regiones de unión para uno o más factores de transcripción. En algunas realizaciones preferidas, las eliminaciones comprenden un sitio de unión a GATA-1, o el sitio de unión para GATA-1 en combinación con otros factores.

[0022] Algunos aspectos de la invención se refieren a proteínas de unión de ADN (no natural) que se unen a la(s) secuencia(s) de BCL11A potenciador pero no escinden la misma. En algunos casos, los dominios de unión a ADN diseñados bloquean la interacción de los factores de transcripción activos en la actividad potenciadora de la unión a sus secuencias potenciadoras afines.

[0023] Dominios de unión a ADN pueden estar fusionados a un dominio funcional, tal como la fusión de los dominios de unión con dominios capaces de regular la expresión de un gen de ADN. Las proteínas de fusión pueden comprender un dominio de unión a ADN fusionado a un dominio modulador de la expresión génica en el que el modulador reprime la expresión génica.

[0024] En algunas realizaciones, las células HSC/PC ponen en contacto con las nucleasas y/o de unión al ADN proteínas de la invención. En algunas realizaciones, las nucleasas y/o proteínas de unión a ADN se administran como ácidos nucleicos y en otras realizaciones, se administran como proteínas. En algunas realizaciones, los ácidos nucleicos son ARNm que codifican las nucleasas y/o proteínas de unión a ADN, y en realizaciones adicionales, los ARNm pueden estar protegidos. En algunas realizaciones, el ARNm puede modificarse químicamente, puede comprender un límite ARCA y/o puede comprender una mezcla de nucleótidos modificados y no modificados.

[0025] En algunos aspectos, las HSC/PC se ponen en contacto con las nucleasas y/o de unión al ADN de proteínas de las invenciones *ex vitro*, tras la aféresis de la HSC/PC de un sujeto, o la purificación de la médula ósea cosechadas. En algunas realizaciones, las nucleasas causan modificaciones dentro de las regiones potenciadoras BCL11A. En realizaciones adicionales, el HSC/PC que contiene las modificaciones de la región potenciadora de BCL11A se introducen de nuevo en el sujeto. En algunos casos, el HSC/PC que contiene las modificaciones de la región del potenciador BCL11A se amplía antes de la introducción. En otros aspectos, el HSC/PC modificado

genéticamente se administra al sujeto en un trasplante de médula ósea en el que el injerto de HSC/PC se diferencia y madura *in vivo*. En algunas realizaciones, las HSC/PC se aíslan del sujeto después de la movilización inducida por G-CSF y/o plerixa, y en otras, las células se aíslan de médula ósea humana o cordones umbilicales humanos. En algunos aspectos, el sujeto es tratado con un procedimiento mieloablativo leve antes de la introducción del injerto que comprende el HSC/PC modificado, mientras que en otros aspectos, el sujeto se trata con un régimen de acondicionamiento mieloablativo vigoroso. En algunas realizaciones, los métodos y composiciones de la invención se usan para tratar o prevenir una hemoglobinopatía. En algunos aspectos, la hemoglobinopatía es una beta talasemia, mientras que en otros aspectos, la hemoglobinopatía es una enfermedad de células falciformes.

[0026] En algunas realizaciones, las HSC/PC se ponen en contacto adicionalmente con una molécula donante. En algunas realizaciones, la molécula donante se administra mediante un vector viral. La molécula donante puede comprender una o más secuencias que codifican un polipéptido funcional (p. ej., un ADNc o fragmento del mismo), con o sin un promotor. Se pueden incluir secuencias adicionales (secuencias codificantes o no codificantes) cuando se usa una molécula donante para la inactivación, incluidas, entre otras, secuencias que codifican un péptido 2A, un sitio SA, IRES, etc.

[0027] Se describen los métodos y composiciones para poner en contacto la HSC/PC *in vivo*. Las nucleasas y/o las proteínas de unión al ADN se administran a HSC/PC *in situ* por métodos conocidos en la técnica. Las nucleasas y/o proteínas de unión al ADN de la invención pueden comprender una partícula viral que se administra al sujeto en necesidad, o las nucleasas y/o proteínas de unión al ADN pueden comprender una nanopartícula (p. ej., liposoma). Las partículas virales y/o las nanopartículas pueden suministrarse al órgano (p. ej., médula ósea) en el que reside la HSC/PC.

[0028] En otro aspecto, se describen en el presente documento métodos de integración de un ácido nucleico donante en el genoma de una célula a través de mecanismos independientes de homología. Los métodos comprenden crear una ruptura de doble cadena (DSB) en el genoma de una célula y dividir la molécula donante utilizando una nucleasa, de manera que el ácido nucleico donante se integre en el sitio de la DSB. El ácido nucleico donador puede integrarse a través de métodos no dependientes de homología (p. ej., NHEJ). Como se señaló anteriormente, tras la escisión *in vivo*, las secuencias del donante pueden integrarse de manera dirigida en el genoma de una célula en la ubicación de un DSB. La secuencia del donante puede incluir uno o más de los mismos sitios de destino para una o más de las nucleasas utilizadas para crear el DSB. Por lo tanto, la secuencia del donante puede ser escindida por una o más de las mismas nucleasas usadas para escindir el gen endógeno en el que se desea la integración. La secuencia del donante puede incluir diferentes sitios diana de nucleasas de las nucleasas usadas para inducir la DSB. Los DSB en el genoma de la célula diana pueden ser creados por cualquier mecanismo. El DSB puede ser creado por una o más nucleasas de dedos de zinc (ZFN), proteínas de fusión que comprenden un dominio de unión a dedos de zinc, que está diseñado para unirse a una secuencia dentro de la región de interés, y un dominio de escisión o un semi-dominio de escisión. El DSB puede ser creado por uno o más dominios de unión al ADN TALE (de origen natural o no natural) fusionados a un dominio de nucleasa (TALEN). El DSB se puede crear utilizando un sistema de nucleasa CRISPR/Cas donde se utiliza un ARN guía único diseñado o su equivalente funcional para guiar la nucleasa a un sitio específico en un genoma.

[0029] En un aspecto, el donante puede codificar una proteína reguladora de interés (p. ej. ZFP TFS, TALE TFS o un CRISPR/Cas TF) que se une a y/o modula la expresión de un gen de interés. En una realización, las proteínas reguladoras se unen a una secuencia de ADN y evitan la unión de otros factores reguladores. En otra realización, la unión de una proteína reguladora puede modular (es decir, inducir o reprimir) la expresión de un ADN diana.

[0030] Se describe una célula HSC/PC transgénica y/o animal, incluyendo un transgén que codifica un gen humano. El animal transgénico puede comprender un knock out en el locus endógeno correspondiente al transgén exógeno, lo que permite el desarrollo de un sistema *in vivo* en el que la proteína humana puede estudiarse de forma aislada. Dichos modelos transgénicos pueden usarse para propósitos de detección para identificar moléculas pequeñas o biomoléculas grandes u otras entidades que pueden interactuar con o modificar la proteína humana de interés. En algunos aspectos de la divulgación, el transgén se integra en el locus seleccionado (p. ej., puerto seguro) en una célula madre (p. ej., una célula madre embrionaria, una célula madre pluripotente inducida, una célula madre hematopoyética, etc.) o un embrión animal obtenido por cualquiera de los métodos descritos aquí, y luego el embrión se implanta de manera tal que nace un animal vivo. El animal se eleva luego a la madurez sexual y se le permite producir descendencia en la que al menos algunas de las descendientes comprenden la secuencia del gen endógeno editado o el transgén integrado.

[0031] Se describe aquí un método para alterar la expresión génica (p. ej., BCL11A y/o un gen de globina) en una célula, comprendiendo el método: la introducción, en la célula, de una o más nucleasas tal como se describe en el presente documento, en condiciones de tal manera que se expresan una o más proteínas y se altera la expresión del gen. La expresión de un gen de globina (p. ej., globina gamma o globina beta) puede ser alterada (p. ej., aumenta). Cualquiera de los métodos descritos en este documento pueden además comprender la integración de una secuencia donante (p. ej., transgén o fragmento del mismo bajo el control de un promotor exógeno o endógeno) en el genoma de la célula, por ejemplo la integración de un donante en o cerca del sitio de la segmentación de nucleasas en el gen BCL11a. La secuencia del donante se introduce en la célula utilizando un vector viral, como un

oligonucleótido y/o en un plásmido. La célula en la que la expresión de genes se altera puede ser, p. ej., un glóbulo de células rojas (RBC) precursor y/o una célula madre hematopoyética (p. ej., células CD34+).

[0032] En este documento se describe un método para producir una célula modificada genéticamente que comprende una modificación genómica dentro de una secuencia reforzadora BCL11a endógena, comprendiendo el método las etapas de: a) poner en contacto una célula con un polinucleótido (p. ej., ADN o ARNm) que codifica una nucleasa de dedo de cinc que comprende 4, 5 o 6 dominios de dedo de cinc en la que cada uno de los dominios de dedo de cinc comprende una región de hélice de reconocimiento en el orden mostrado en una sola fila de la Tabla 3 o la Tabla 6; b) someter la célula a condiciones conducentes a la expresión de la proteína de dedo de zinc desde el polinucleótido; y c) modificando la secuencia del potenciador BCL11A endógeno con la proteína del dedo de zinc expresada suficiente para producir la célula modificada genéticamente. Las células pueden ser estimuladas con al menos una citoquina (p. ej., antes del paso (a)). El polinucleótido puede ponerse en contacto con la célula usando cualquier método adecuado, incluyendo pero no limitado, a través de la transfección, usando un vector no viral, utilizando un vector viral, por medios químicos o por la exposición a un campo eléctrico (p. ej., electroporación).

[0033] También se describe un método de tratamiento de un paciente en necesidad de un aumento en la expresión de genes de globina, comprendiendo el método administrar al paciente la preparación farmacéutica como se describe en el presente documento en una cantidad suficiente para aumentar la expresión de genes de globina en el paciente. Se sabe que el paciente tiene, se sospecha que tiene, o puede tener riesgo de desarrollar una talasemia o enfermedad de células falciformes.

[0034] Un kit, que comprende los ácidos nucleicos, proteínas y/o células de la invención, es también revelado. El kit puede comprender ácidos nucleicos que codifican las nucleasas (p. ej., moléculas de ARN o ZFN, TALEN o CRISPR/Cas que codifican genes contenidos en un vector de expresión adecuado), o partes alícuotas de las proteínas de nucleasa, moléculas donantes, modificadores de potencia adecuados, células, tampones e/o instrucciones (p. ej., para realizar los métodos de la invención) y similares.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0035]

La Figura 1 muestra un diagrama de la región de codificación BCL11A, que indica la posición de los intrones y de las regiones potenciadoras. Las derivaciones de los diferentes productos de empalme también están indicadas (ver Liu et al. (2006) Molecular Cancer 5:18).

La Figura 2 representa la región genómica que codifica las diversas isoformas Bcl11a (buscador de genoma de la Universidad de California Santa Cruz, coordenadas enumeradas en el ensamblaje hg19 del genoma humano), la región potenciadora en el intrón 2 BCL11A (coordenadas enumeradas en el ensamblaje hg19 del genoma humano), y define las tres subregiones de la región potenciadora representadas por sitios hipersensibles por ADNasa I (enumerados en kb por distancia aproximada del sitio de inicio de la transcripción): +55, +58 y +62. Las ubicaciones diana de nucleasa dentro de las tres subregiones se indican a continuación: las nucleasas se diseñaron para dividirse en el extremo izquierdo de cada subregión (los sitios 'L'), en el centro de cada uno (los sitios 'M') y en el extremo derecho (los sitios 'R'). La escisión del locus *in vivo* con un par de nucleasas da como resultado una delección de la región intermedia en una fracción significativa de las células.

La Figura 3 muestra los resultados de la escisión en la célula por los conjuntos de pares TALEN (indicados en la tabla en el panel superior de la figura) en la región del potenciador BCL11A según se mide por un ensayo basado en PCR para la eliminación de varias regiones del potenciador. Los pares de TALEN utilizados fueron diseñados para crear eliminaciones en HSPC humanas ya sea dentro de la región +55 ("55L-R"), dentro de la región +62 ("62L-R"), o dentro de la región +58 ("58L-R", "58M-R"). El gel muestra los productos de PCR producidos mediante el aislamiento del ADN genómico de las HSPC humanas transfectadas con construcciones de expresión que codifican los TALEN indicados y amplifican la región que rodea la región diana. Estos datos demuestran la generación de delecciones (banda indicada por el símbolo Δ) en la región seleccionada luego de la división por los conjuntos de pares TALEN indicados.

La Figura 4 es una gráfica que muestra los resultados de un análisis en tiempo real de RT-qPCR ("Taqman®") diseñado para detectar un cambio en la expresión de gamma fetal - ARNm de globina después de la edición dirigida del potenciador BCL11A. Después de la electroporación de células CD34 de voluntarios humanos sanos con ARNm que codifican las nucleasas designadas (ver FIG. 3), se erigieron eritrocitos *in vitro*, después de lo cual se recogió el ARN total de alfa globina y se determinó el ARNm de globina gamma para cada muestra en un análisis RT-PCR Taqman®, y se representó la relación relativa de ARNm de globina gamma/ARNm de alfa globina. Por lo tanto, el aumento de la expresión de globina gamma en las muestras tratadas con nucleasa conduce a un aumento en la relación gamma/alfa normalizada en comparación con los controles. La Figura muestra los resultados del tratamiento de células CD34 con pares TALEN únicos y para los conjuntos de pares TALEN descritos en las Figuras 2 y 3, El nivel de gamma/alfa se incrementa para los conjuntos de pares 58L-R y 58M-R. Tenga en cuenta que la relación en el control de transfección de GFP fue 3,4 en este experimento.

La Figura 5 es un gráfico que muestra los resultados de un análisis en tiempo real de RT-qPCR ("Taqman®") diseñado para detectar un cambio en la expresión del ARNm de globina gamma fetal después de la edición dirigida del potenciador BCL11A. Después de la electroporación de células CD34 de voluntarios humanos sanos

con ARNm que codifican las nucleasas designadas (ver FIG. 3), los eritrocitos se generaron *in vitro*, después de lo cual se recolectó el ARN total. Los niveles relativos de ARNm de globina beta y globina gamma para cada muestra se determinaron en un análisis de RT-PCR Taqman®, y se representó la relación relativa de ARNm de globina gamma/mRNA de globina beta. Los resultados demuestran que si bien los pares únicos TALEN utilizados en estos experimentos no pudieron inducir un cambio en la expresión de globina gamma, la creación de eliminaciones mediante el uso de conjuntos de pares causó un aumento en la expresión gamma relativa, corregida por la expresión de globina beta adulta en esta instancia. En particular, la eliminación del ADN secuenciado englobado por la globina gamma elevada hipersensible ADNasa I +55 o +58, mientras que tal elevación después de una eliminación de la secuencia englobada por el sitio hipersensible del ADNasa I en +62 no se detectó en este experimento. Tenga en cuenta que la proporción en el control de transfección de GFP fue de 0,5 en este experimento.

La Figura 6 muestra los resultados de un análisis Taqman® para los niveles de globina fetal como se describe en la Figura 4. En este conjunto de experimentos, se midieron los niveles de ARNm de globina beta y globina gamma, y por lo tanto, los datos descritos son la relación de gamma a beta y en comparación con la misma proporción gamma/beta en las células tratadas de control. Se utilizó una serie de pares TALEN individuales para "caminar" a través de la región +58 del potenciador BCL11A. En contraste con los experimentos anteriores en los que se requirió una eliminación de 400-900 pares de bases para ver un aumento en la expresión gamma, los resultados representados en este experimento demostraron un sitio único (indicado por una flecha) que cuando se escindió causó ese aumento relativo. Los resultados de los pares de borrados se incluyen en este gráfico para comparación. Consulte la Figura 7 para conocer la ubicación de los sitios de escisión de los pares TALEN en la región de interés.

La Figura 7 muestra los resultados de un análisis Taqman® como se describe en la Figura 4 utilizando pares de ZFN dirigidos a la región potenciadora de +58. Los niveles de beta globina y globina gamma se caracterizaron y utilizaron para expresar la proporción de globina gamma a beta en comparación con la misma proporción en las células tratadas de control. Los datos demuestran que la interrupción impulsada por ZFN de la misma región identificada en la pantalla TALEN (ver Figura 6 y Figura 8 a continuación) resultó en un aumento de la expresión de globina gamma. Tenga en cuenta que la relación en el control de transfección de GFP fue 2,6 en este experimento.

La Figura 8 muestra una representación de los sitios de unión (SEQ ID NO: 264 y SEQ ID NO: 265) de los TALEN específicos de la región potenciadora de +58 (102852 y 102853) y ZFN (45843 y 45844), cuyo uso en las HSPC humanas aumenta. La expresión relativa de globina gamma después de la eritropoyesis *in vitro*. La secuencia mostrada es la forma de doble cadena de la secuencia de ADN que abarca la región +58 del potenciador BCL11A (SEQ ID NO: 363), y el sistema de numeración se relaciona con el propio +58. También se indica en la figura la ubicación de una coincidencia con el sitio de unión del factor de transcripción GATA-1 (secuencia de locus, gtGATAAag, sitio GATA-1 de consenso - swGATAAvv). Además, los sitios de escisión de los pares TALEN utilizados en la "caminata" +58 se indican donde los números corresponden a las muestras utilizadas en los conjuntos de datos presentados en la Figura 6.

La Figura 9 muestra los resultados de un análisis Taqman® como se describe en la Figura 4, donde se realizó una serie de TALEN para apuntar a la región +55 del potenciador BCL11A. En este conjunto de experimentos, se caracterizaron los niveles de globina beta y globina gamma y, por lo tanto, los datos representados son la proporción de gamma a beta y se compararon con la misma proporción en las células tratadas de control (que, en este experimento, fue de 0,8). Los datos confirman que las mutaciones generadas en posiciones específicas dentro de la región +55 pueden aumentar la expresión relativa de la globina gamma (ver flechas).

La Figura 10 es una representación de los sitios de escisión de los TALEN específicos de la región del potenciador +55 como se muestra en la Figura 9. La secuencia que se muestra es la forma bicatenaria de la secuencia de ADN que abarca la región +55 del potenciador BCL11A (SEQ ID NO: 254). Los números que resaltan regiones cortas de nucleótidos indican los probables sitios de escisión inducidos por las muestras de TALEN enumeradas en la Figura 10. También se indican en la Figura 10 dos coincidencias con el sitio de unión de consenso para el factor de transcripción GATA-1.

Las Figuras 11A a 11C muestran la secuencia de ADN de los tres sitios hipersensibles a ADNasa I dentro de la secuencia potenciadora de BCL11A. Debido a que su identificación se realizó mediante sondeo de regiones de cromatina accesible en células (ver Bauer *et al*, (2013) *ibid*), los límites exactos de las regiones no se conocen y se muestran los límites aproximados. La Figura 11A muestra la secuencia de la región +55 (SEQ ID NO: 1), la Figura 11B muestra la secuencia de la región +58 (SEQ ID NO: 2) y la Figura 11C muestra la secuencia de la región +62 (SEQ ID NO: 3).

Las Figuras 12A a 12C demuestran que la división impulsada por ZFN en células más cercanas al núcleo del consenso GATA-1 eleva los niveles de globina fetal en un grado incluso mayor que la división más cercana al extremo 3' del motivo. La Figura 12A muestra un diagrama que muestra los sitios de unión de los pares de ZFN específicos de BCL11A en relación con la secuencia de consenso GATA-1 (Figura 12A) y representa una secuencia de ADN dentro de la región +58 que comprende la secuencia de consenso GATA-1 (SEQ ID NO: 255). Las barras por encima y por debajo de la secuencia de ADN indican los sitios de unión de los ZFN. La Figura 12B muestra la expresión relativa de globina gamma y globina beta medida por la expresión de ARNm después de HSPC humanas con ARNm que codifica los ZFN indicados (ver Figura 12A), seguida de eritropoyesis *in vitro* y medición de los niveles de globina fetal (ver Figura 4). La relación observada cuando se transfeció un ARNm que expresaba GFP en las células CD34+ fue de 0,97. La Figura 12C representa en forma de "gráfico circular" las formas alélicas del potenciador BCL11A (específicamente, la región dividida por los ZFN

mostrados en la Figura 12A) que se encuentra en las HSPC humanas después de la electroporación con los ZFN indicados. Si bien se observan niveles comparables de cromátidas no modificadas (tipo salvaje) en las dos muestras, la muestra tratada con ZFN que se cortan más cerca del motivo GATA-1 contiene un número mayor de cromátidas que eliminan el consenso de GATA-1 (p. ej., el alelo "-15", que representa una eliminación de 15 pares de bases). Los datos demuestran que la división por los dos pares de ZFN que están más cerca del centro de la secuencia de consenso GATA-1 (pares 46801/46880 y 46923/46999) se asocia con un aumento de la expresión de globina gamma.

La Figura 13 demuestra que la alteración del enlazador entre el dedo de zinc y el resto FokI en las ZFN utilizadas para la edición del genoma del potenciador BCL11A afecta los niveles de globina fetal después de la eritropoyesis *in vitro* a pesar de niveles comparables en las cromátidas. Las HSPC humanas se sometieron a electroporación con los ZFN indicados (el conector utilizado en cada monómero ZFN se indica entre paréntesis), e inmediatamente antes y después de la diferenciación eritroide *in vitro*, se midió el porcentaje de alelos alterados (se muestra a continuación cada muestra en forma "X/Y", con el primer número correspondiente al % de indeles de tipo no salvaje después de la electroporación, y el segundo número muestra los resultados después de 14 días de diferenciación eritroide *in vitro*). Se recogió el ARNm completo y se midieron los niveles de globina fetal (normalizada a globina alfa).

La Figura 14 representa las secuencias de aminoácidos y ADN para cuatro enlazadores (L0 (SEQ ID NO: 256 y 257), L7a (SEQ ID NO: 258 y 259), L7C₅ (SEQ ID NO: 260 y 261) y L8C₅ (SEQ ID NO: 262 y 263) utilizados en los diseños de ZFP. Las secuencias con un subrayado sólido indican la región carboxi terminal del dominio de unión al ADN de ZFP, mientras que las secuencias indicadas con el subrayado discontinuo indican la región amino terminal del dominio de nucleasa FokI. Las secuencias en negrita indican las secuencias novedosas agregadas al enlazador L0 estándar.

Las Figuras 15A y 15B representan el porcentaje de células de origen humano y la modificación genética dirigida en el sitio diana de la nucleasa en estas células humanas, respectivamente, encontradas en la sangre periférica de ratones después del trasplante de células CD34⁺ humanas editadas. La Figura 15A es un gráfico que representa el porcentaje de células humanas en la periferia del ratón después del trasplante de células CD34⁺ humanas que se habían editado con dos conjuntos diferentes de ZFN 4 semanas después del trasplante. La Figura 15B muestra el nivel de los indeles detectados en esas células humanas. Cada símbolo representa datos obtenidos de un ratón individual.

Las Figuras 16A a 16D, representan el porcentaje de indeles inducidos por las nucleasas en células humanas que se diferenciaron de las células CD34⁺ trasplantadas originales 16 semanas después del trasplante. La Figura 16A muestra el nivel de actividad indel en células panmieloides, identificada por la presencia del marcador CD33. La Figura 16B muestra la actividad en células B CD19⁺. La Figura 16C muestra la actividad en glyA⁺ o células eritroides, mientras que la Figura 16D muestra la actividad en células madre. Cada símbolo representa datos obtenidos de un ratón individual.

La Figura 17 muestra una serie de secuencias enlazadoras (SEQ ID NOs: 265-275). Estos enlazadores pueden servir para enlazar el dominio de unión al ADN del dedo de zinc con el dominio de nucleasa FokI,

DESCRIPCIÓN DETALLADA

[0036] En este documento se describen composiciones y métodos para la ingeniería del genoma para la modulación de la expresión de BCL11A y/o globina gamma y para el tratamiento y/o prevención de hemoglobinopatías. En particular, la eliminación dirigida por nucleasa (es decir, ZFN, TALEN o CRISPR/Cas o TtAgo) de sitios específicos en una región potenciadora de BCL11A se logra de manera eficiente en HSC/PC y da como resultado un cambio en la expresión de globina gamma relativa durante eritropoyesis. Esta modulación de BCL11A y expresión globina gamma es particularmente útil para el tratamiento de hemo-globinopathies (p. ej., talasemias beta, enfermedad de células falciformes) en las que no hay suficiente expresión de globina beta o la expresión de una forma mutada de beta-globina. Usando los métodos y composiciones de la invención, las complicaciones y las secuelas relacionadas con la enfermedad causadas por la globulina beta aberrante se pueden superar mediante la alteración de la expresión de globina gamma en células precursoras de eritrocitos.

General

[0037] La práctica de los métodos, así como la preparación y el uso de las composiciones descritas en este documento emplean, a menos que se indique lo contrario, técnicas convencionales en biología molecular, bioquímica, estructura y análisis de cromatina, química computacional, cultivo celular, ADN recombinante y campos relacionados están dentro de la habilidad de la técnica. Estas técnicas están completamente explicadas en la literatura. Ver, p. ej., 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" (PM Wassarman y AP Wolffe, eds.), Academic Press, San Diego, 1999; y METHODS IN MOLECULAR BIOLOGY, vol. 119, "Chromatin Protocols" (PB Becker, ed.) Humana Press, Totowa, 1999.

Definiciones

[0038] Los términos "ácido nucleico", "polinucleótido" y "oligonucleótido" se usan indistintamente y se refieren a un polímero deoxiribonucleótido o ribonucleótido, en conformación lineal o circular, y en forma ya sea monocatenaria o bicatenaria. Para los propósitos de la presente divulgación, estos términos no deben interpretarse como limitantes con respecto a la longitud de un polímero. Los términos pueden abarcar análogos conocidos de nucleótidos naturales, así como nucleótidos que se modifican en la base, azúcar y/o restos fosfato (p. ej., esqueletos de fosforotioato). En general, un análogo de un nucleótido particular tiene la misma especificidad de apareamiento de bases; es decir, un análogo de A se emparejará con T.

[0039] Los términos "polipéptido", "péptido" y "proteína" se utilizan indistintamente para referirse a un polímero de residuos de aminoácidos. El término también se aplica a los polímeros de aminoácidos en los que uno o más aminoácidos son análogos químicos o derivados modificados de los aminoácidos naturales correspondientes.

[0040] "Unión" se refiere a una interacción específica de secuencia, no covalente entre macromoléculas (p. ej., entre una proteína y un ácido nucleico). No todos los componentes de una interacción de unión tienen que ser específica de la secuencia (p. ej., los contactos con los residuos de fosfato en una cadena principal de ADN), siempre que la interacción como un todo es específica de la secuencia. Tales interacciones se caracterizan generalmente por una constante de disociación (K_d) de 10^{-6} M⁻¹ o inferior. "Afinidad" se refiere a la fuerza de unión: una mayor afinidad de unión se correlaciona con una K_d menor.

[0041] Una "proteína de unión" es una proteína que es capaz de unirse a otra molécula. Una proteína de unión puede unirse a, p. ej., una molécula de ADN (una proteína de unión a ADN), una molécula de ARN (una proteína de unión a ARN) y/o una molécula de proteína (una proteína de unión a proteína). En el caso de una proteína de unión a proteínas, puede unirse a sí misma (para formar homodímeros, homotrímeros, etc.) y/o puede unirse a una o más moléculas de una proteína o proteínas diferentes. Una proteína de unión puede tener más de un tipo de actividad de unión. P. ej., las proteínas de dedos de zinc tienen actividad de unión a ADN, unión a ARN y unión a proteínas.

[0042] Una "proteína de unión de ADN de dedo de zinc" (o dominio de unión) es una proteína, o un dominio dentro de una proteína más grande, que se une al ADN de una manera específica de la secuencia a través de uno o más dedos de zinc, que son regiones de secuencia de aminoácidos dentro del dominio de unión cuya estructura se estabiliza a través de la coordinación de un ión de zinc. El término proteína de unión al ADN del dedo de zinc a menudo se abrevia como proteína de dedo de zinc o ZFP.

[0043] Un "dominio de unión de ADN TALE" o "TALE" es un polipéptido que comprende uno o más dominios/unidades de repetición TALE. Los dominios de repetición están involucrados en la unión del TALE a su secuencia de ADN diana relacionada. Una sola "unidad de repetición" (también denominada "repetición") tiene típicamente una longitud de 33-35 aminoácidos y exhibe al menos alguna homología de secuencia con otras secuencias de repetición TALE dentro de una proteína TALE de origen natural.

[0044] Los dominios de unión de dedo de zinc y TALE pueden ser "diseñados" para unirse a una secuencia de nucleótidos predeterminada, por ejemplo a través de ingeniería (alterando uno o más aminoácidos) de la región de hélice de reconocimiento de un dedo de zinc natural o proteína TALE. Por lo tanto, las proteínas de unión al ADN diseñadas por ingeniería (dedos de zinc o TALE) son proteínas que no ocurren de forma natural. Ejemplos no limitativos de métodos para diseñar proteínas de unión a ADN son el diseño y la selección. Una proteína de unión a ADN diseñada es una proteína que no aparece en la naturaleza cuyo diseño/composición resulta principalmente de criterios racionales. Los criterios racionales para el diseño incluyen la aplicación de reglas de sustitución y algoritmos computarizados para procesar información en una base de datos que almacena información de diseños existentes de ZFP y/o TALE y datos de enlace. Ver, p. ej., Patentes de Estados Unidos 6.140.081; 6.453.242; 6.534.261 y 8.585.526; véase también el documento WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 y WO 03/016496.

[0045] Una proteína con dedos de zinc "seleccionada" o TALE es una proteína que no se encuentra en la naturaleza cuyos resultados principalmente de un proceso empírico tal como presentación de fagos, la trampa de interacción o selección de híbridos de producción. Ver p. ej., patentes de EE.UU. n° 5.789.538; 5.925.523; 6.007.988; 6.013.453; 6.200.759; 8.586.526; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197, WO 02/099084.

[0046] "TtAgo" es una proteína argonauta procariótica que se cree que está involucrada en el silenciamiento de genes. TtAgo se deriva de la bacteria *Thermus thermophilus*. Ver, p. ej., Swarts *et al*, *ibid*, G. Sheng *et al*, (2013) *Proc. Natl Acad Sci. EE.UU.* 111, 652). Un "sistema TtAgo" es todos los componentes necesarios que incluyen, p. ej., guías de ADN para la división por una enzima de TtAgo

[0047] "Recombinación" se refiere a un proceso de intercambio de información genética entre dos polinucleótidos, incluyendo pero no limitado a, capturas de donante por extremos no homólogos (NHEJ) y recombinación homóloga. Para los fines de esta descripción, "recombinación homóloga (HR)" se refiere a la forma especializada de dicho intercambio que tiene lugar, p. ej., durante la reparación de roturas de doble cadena en células mediante mecanismos de reparación dirigidos por homología. Este proceso requiere homología de secuencia de nucleótidos,

utiliza una molécula "donante" para reparar la plantilla de una molécula "diana" (es decir, la que experimentó la rotura de doble cadena), y se conoce de varias maneras como "conversión de genes no cruzados" o "conversión del gen del tracto corto", porque conduce a la transferencia de información genética del donante diana. Sin desear estar limitado por ninguna teoría particular, dicha transferencia puede implicar la corrección de desajustes del ADN heterodúplex que se forma entre el objetivo roto y el donante, y/o el "recocido de la cadena dependiente de la síntesis", en el que el donante se utiliza para resintetizar la información genética. que se convertirá en parte del objetivo, y/o procesos relacionados. Este HR especializado a menudo resulta en una alteración de la secuencia de la molécula diana, de manera que parte o toda la secuencia del polinucleótido donante se incorpora al polinucleótido diana.

[0048] En los métodos de la descripción, una o más nucleasas apuntadas como se describe en la presente memoria crean una rotura de doble cadena (DSB) en la secuencia diana (p. ej., la cromatina celular) en un sitio predeterminado. El DSB puede dar lugar a eliminaciones y/o inserciones por reparación dirigida por homología o por mecanismos de reparación no homologados. Las eliminaciones pueden incluir cualquier número de pares de bases. De manera similar, las inserciones pueden incluir cualquier número de pares de bases que incluyen, p. ej., la integración de un polinucleótido "donante", que tiene opcionalmente una homología con la secuencia de nucleótidos en la región de la ruptura. La secuencia del donante se puede integrar físicamente o, alternativamente, el polinucleótido del donante se usa como plantilla para la reparación de la rotura a través de la recombinación homóloga, lo que resulta en la introducción de toda o parte de la secuencia de nucleótidos como en el donante en la cromatina celular. Por lo tanto, una primera secuencia en la cromatina celular puede alterarse y, en ciertas realizaciones, puede convertirse en una secuencia presente en un polinucleótido donante. Por lo tanto, se puede entender que el uso de los términos "reemplazar" o "reemplazo" representa la sustitución de una secuencia de nucleótidos por otra, (es decir, la sustitución de una secuencia en el sentido informativo), y no necesariamente requiere un reemplazo físico o químico de un polinucleótido por otro.

[0049] En cualquiera de los métodos descritos en el presente documento, los pares adicionales de proteínas de dedos de cinc o TALEN se puede utilizar para la escisión de doble cadena adicional de sitios diana adicionales dentro de la célula.

[0050] Cualquiera de los métodos descritos en este documento puede ser utilizado para la inserción de un donante de cualquier tamaño y/o la inactivación parcial o completa de una o más secuencias diana en una célula mediante integración dirigida de la secuencia donante que interrumpe la expresión del gen de interés. También se proporcionan líneas celulares con genes parcialmente o completamente inactivados.

[0051] En cualquiera de los métodos descritos en el presente documento, la secuencia de nucleótidos exógena (la "secuencia donadora" o "transgen") pueden contener secuencias que son homólogas, pero no idénticas, a las secuencias genómicas en la región de interés, estimulando así la recombinación homóloga para insertar una secuencia no idéntica en la región de interés. Por lo tanto, en ciertas realizaciones, las porciones de la secuencia del donante que son homólogas a las secuencias en la región de interés exhiben entre aproximadamente el 80 y el 99% (o cualquier número entero entre ellas) la identidad de la secuencia con la secuencia genómica que se reemplaza. En otras realizaciones, la homología entre el donante y la secuencia genómica es superior al 99%, p. ej., si solo 1 nucleótido difiere entre el donante y las secuencias genómicas de más de 100 pares de bases contiguas. En ciertos casos, una parte no homóloga de la secuencia del donante puede contener secuencias que no están presentes en la región de interés, de manera que se introducen nuevas secuencias en la región de interés. En estos casos, la secuencia no homóloga generalmente está flanqueada por secuencias de 50-1.000 pares de bases (o cualquier valor integral entre ellas) o cualquier número de pares de bases mayores que 1.000, que sean homólogas o idénticas a las secuencias en la región de interés. En otras realizaciones, la secuencia del donante no es homóloga a la primera secuencia, y se inserta en el genoma por mecanismos de recombinación no homóloga.

[0052] "Escisión" se refiere a la rotura del esqueleto covalente de una molécula de ADN. La escisión puede iniciarse mediante una variedad de métodos que incluyen, pero no se limitan a, hidrólisis enzimática o química de un enlace fosfodiéster. Tanto la escisión monocatenaria como la escisión bicatenaria son posibles, y la escisión bicatenaria puede ocurrir como resultado de dos eventos distintos de escisión monocatenaria. La escisión del ADN puede resultar en la producción de extremos romos o extremos escalonados. En ciertas realizaciones, los polipéptidos de fusión se usan para la escisión de ADN de doble hebra dirigida.

[0053] Un "semi-dominio de escisión" es una secuencia de polipéptido que, en conjunción con un segundo polipéptido (idénticos o diferentes) forma una actividad de escisión que tiene complejo (actividad de escisión de preferencia de doble cadena). Los términos "primer y segundo medio dominio de división"; Los "semi-dominios de escisión + y" y los "semi-dominios de escisión derecha e izquierda" se usan indistintamente para referirse a los pares de semi-dominios de escisión que dimerizan.

[0054] Un "medio-dominio de escisión" es un medio-dominio de escisión que ha sido modificado a fin de formar heterodímeros obligados con otro medio-dominio de escisión (p. ej., otra escisión por medio-dominio modificado). Ver, también, la Publicación de Patente de Estados Unidos N° 2005/0064474, 20070218528, 20080131962 y 20110201055.

[0055] El término "secuencia" se refiere a una secuencia de nucleótidos de cualquier longitud, que puede ser ADN o ARN; puede ser lineal, circular o ramificado y puede ser monocatenario o bicatenario. El término "secuencia donante" se refiere a una secuencia de nucleótidos que se inserta en un genoma. Una secuencia donante puede ser de cualquier longitud, p. ej., entre 2 y 100.000.000 nucleótidos en longitud (o cualquier valor entero entre ellos o más), preferiblemente entre aproximadamente 100 y 100.000 nucleótidos en longitud (o cualquier número entero entre ellos), más preferiblemente entre aproximadamente 2.000 y 20.000 nucleótidos en longitud (o cualquier valor entre ellos) e incluso más preferible, entre aproximadamente 5 y 15 kb (o cualquier valor entre ellos).

[0056] La "cromatina" es la estructura de la nucleoproteína que comprende el genoma celular. La cromatina celular comprende ácido nucleico, principalmente ADN y proteínas, incluidas histonas y proteínas cromosómicas no histónicas. La mayoría de la cromatina celular eucariota existe en forma de nucleosomas, en donde un núcleo de nucleosoma comprende aproximadamente 150 pares de bases de ADN asociado con un octámero que comprende dos histonas H2A, H2B, H3 y H4; y el ADN enlazador (de longitud variable dependiendo del organismo) se extiende entre los núcleos de los nucleosomas. Una molécula de histona H1 se asocia generalmente con el ADN enlazador. Para los fines de la presente descripción, el término "cromatina" pretende abarcar todos los tipos de nucleoproteínas celulares, tanto procariotas como eucariotas. La cromatina celular incluye cromatina cromosómica y episomal.

[0057] Un "cromosoma", es un complejo de la cromatina que comprende toda o una porción del genoma de una célula. El genoma de una célula a menudo se caracteriza por su cariotipo, que es la colección de todos los cromosomas que forman el genoma de la célula. El genoma de una célula puede comprender uno o más cromosomas.

[0058] Un "episoma" es un ácido nucleico replicante, complejo de nucleoproteína u otra estructura comprende un ácido nucleico que no es parte del cariotipo cromosómico de una célula. Ejemplos de episomas incluyen plásmidos y ciertos genomas virales.

[0059] Una "región accesible" es un sitio de la cromatina celular en el que un sitio diana presente en el ácido nucleico puede estar unido por una molécula exógena que reconoce el sitio de destino. Sin desear estar limitado por ninguna teoría particular, se cree que una región accesible es aquella que no está empaquetada en una estructura nucleosómica. La estructura distinta de una región accesible a menudo se puede detectar por su sensibilidad a las sondas químicas y enzimáticas, p. ej., las nucleasas.

[0060] Un "sitio diana" o "secuencia diana" es una secuencia de ácido nucleico que define una parte de un ácido nucleico a la cual se unirá una molécula de unión, siempre que existan condiciones suficientes para unión.

[0061] Una molécula "exógena" es una molécula que normalmente no está presente en una célula, pero puede ser introducida en una célula por uno o más métodos genéticos, bioquímicos u otros métodos. La "presencia normal en la célula" se determina con respecto a la etapa de desarrollo particular y las condiciones ambientales de la célula. Así, p. ej., una molécula que está presente solo durante el desarrollo embrionario del músculo es una molécula exógena con respecto a una célula muscular adulta. De manera similar, una molécula inducida por choque térmico es una molécula exógena con respecto a una célula sin choque térmico. Una molécula exógena puede comprender, p. ej., una versión funcional de una molécula endógena que funciona mal o una versión defectuosa de una molécula endógena que funciona normalmente.

[0062] Una molécula exógena puede ser, entre otras cosas, una molécula pequeña, como se genera por un proceso de química combinatoria, o una macromolécula tal como una proteína, ácido nucleico, carbohidrato, lípido, glicoproteína, lipoproteína, polisacárido, cualquier derivado modificado de las moléculas anteriores, o cualquier complejo que comprenda una o más de las moléculas anteriores. Los ácidos nucleicos incluyen ADN y ARN, pueden ser de cadena sencilla o doble; Puede ser lineal, ramificado o circular; y puede ser de cualquier longitud. Los ácidos nucleicos incluyen aquellos capaces de formar dúplex, así como ácidos nucleicos formadores de triplex. Véanse, p. ej., las patentes de EE.UU. nº 5.176.996 y 5.422.251. Las proteínas incluyen, pero no se limitan a, proteínas de unión a ADN, factores de transcripción, factores de remodelación de la cromatina, proteínas de unión a ADN metiladas, polimerasas, metilasas, demetilasas, acetilasas, desacetilasas, quinasas, fosfatasas, integrasas, recombinasas, topoisomasas, girasas y helicasas.

[0063] Una molécula exógena puede ser el mismo tipo de molécula como una molécula endógena, p. ej., una proteína exógena o ácido nucleico. P. ej., un ácido nucleico exógeno puede comprender un genoma viral infectante, un plásmido o episoma introducido en una célula, o un cromosoma que normalmente no está presente en la célula. Los métodos para la introducción de moléculas exógenas en las células son conocidos por los expertos en la técnica e incluyen, entre otros, la transferencia mediada por lípidos (es decir, liposomas, incluidos los lípidos neutros y catiónicos), electroporación, inyección directa, células fusión, carga de partículas, co-precipitación con fosfato de calcio, transferencia mediada por DEAE-dextrano y transferencia mediada por vector viral. Una molécula exógena también puede ser el mismo tipo de molécula que una molécula endógena pero derivada de una especie diferente de la que se deriva la célula. P. ej., una secuencia de ácido nucleico humano puede introducirse en una línea celular originaria originalmente de un ratón o hámster. Los métodos para la introducción de moléculas exógenas en células

vegetales son conocidos por los expertos en la técnica e incluyen, entre otros, transformación de protoplastos, carburo de silicio (p. ej., WHISKERS™), transformación mediada por *Agrobacterium*, transferencia mediada por lípidos (es decir, liposomas, incluidos los lípidos neutros y catiónicos, electroporación, inyección directa, fusión celular, bombardeo de partículas (p. ej., usando una "pistola de genes"), co-precipitación con fosfato de calcio, transferencia mediada por DEAE-dextrano y transferencia mediada vector viral.

[0064] Por el contrario, una molécula "endógena" es una que está normalmente presente en una célula particular en una etapa de desarrollo particular bajo condiciones ambientales particulares. P. ej., un ácido nucleico endógeno puede comprender un cromosoma, el genoma de una mitocondria, un cloroplasto u otro orgánulo, o un ácido nucleico episomal de origen natural. Moléculas endógenas adicionales pueden incluir proteínas, p. ej., factores de transcripción y enzimas.

[0065] Como se usa en el presente documento, el término "producto de un ácido nucleico exógeno" incluye productos tanto de polinucleótidos como de polipéptidos, p. ej., productos de transcripción (polinucleótidos tales como ARN) y productos de traducción (polipéptidos).

[0066] Una molécula de "fusión" es una molécula en la que dos o más moléculas de subunidades están vinculadas, preferentemente de forma covalente. Las moléculas de la subunidad pueden ser del mismo tipo de molécula química, o pueden ser diferentes tipos de moléculas químicas. Los ejemplos del primer tipo de molécula de fusión incluyen, pero no se limitan a, proteínas de fusión (p. ej., una fusión entre un dominio de unión al ADN de ZFP o TALE y uno o más dominios de activación) y ácidos nucleicos de fusión (p. ej., un ácido codificante de la proteína de fusión descrito anteriormente. Los ejemplos del segundo tipo de molécula de fusión incluyen, pero no se limitan a, una fusión entre un ácido nucleico formador de triplex y un polipéptido, y una fusión entre un ligante de surco menor y un ácido nucleico.

[0067] La expresión de una proteína de fusión en una célula puede resultar de la entrega de la proteína de fusión a la célula o mediante la entrega de un polinucleótido que codifica la proteína de fusión a una célula, en donde el polinucleótido se transcribe, y la transcripción se traduce, para generar la proteína de fusión. El empalme trans, la escisión de polipéptidos y la ligadura de polipéptidos también pueden estar implicados en la expresión de una proteína en una célula. Los métodos para el suministro de polinucleótidos y polipéptidos a las células se presentan en otra parte de esta descripción.

[0068] Un "gen", para los fines de la presente descripción, incluye una región de ADN que codifica un producto génico (véase *infra*), así como todas las regiones de ADN que regulan la producción del producto del gen, si tales secuencias reguladoras o no son adyacentes a las secuencias de codificación y/o transcritas. Por consiguiente, un gen incluye, pero no está necesariamente limitado a, secuencias promotoras, terminadores, secuencias reguladoras de la traducción tales como sitios de unión a ribosomas y sitios de entrada de ribosomas internos, potenciadores, silenciadores, aisladores, elementos de frontera, orígenes de replicación, sitios de unión a matriz y control de locus regiones.

[0069] "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 (p. ej., ARNm, ARNt, ARNr, ARN antisentido, ribozima, ARN estructural o cualquier otro tipo de ARN) o una proteína producida por traducción de un ARNm. Los productos genéticos también incluyen ARN que se modifican, mediante procesos tales como capping, poliadenilación, metilación y edición, y proteínas modificadas, p. ej., por metilación, acetilación, fosforilación, ubiquitinación, ADP-ribosilación, miristilación y glicosilación. La "modulación" de la expresión génica se refiere a un cambio en la actividad de un gen. La modulación de la expresión puede incluir, pero no se limita a, la activación de genes y la represión de genes. La edición del genoma (p. ej., división, alteración, inactivación, mutación aleatoria) se puede usar para modular la expresión. La inactivación de genes se refiere a cualquier reducción en la expresión de genes en comparación con una célula que no incluye un sistema ZFP, TALE o CRISPR/Cas como se describe en este documento. Por lo tanto, la inactivación de genes puede ser parcial o completa.

[0071] Un "protegido" ARNm es uno en el que el ARNm se ha alterado de alguna manera para aumentar la estabilidad o traducción del ARNm. Los ejemplos de protecciones incluyen el uso de la sustitución de hasta el 25% de los residuos de citidina y uridina con 2-tiouridina (S2U) y 5-metilcitidina (m5C). El ARNm resultante exhibe menos inmunogenicidad y más estabilidad en comparación con su contraparte no modificada. (ver Kariko et al. ((2012), Molecular Therapy, Vol. 16, No. 11, páginas 1833 - 1844). Otros cambios incluyen la adición de la llamada tapa ARCA, que aumenta la capacidad de traducción de la ARNm producida *in vitro* (véase la patente estadounidense US7074596).

[0072] Una "región de interés" es cualquier región de la cromatina celular, tal como, p. ej., un gen o una secuencia no codificante dentro o adyacentes a un gen, en donde es deseable que se una una molécula exógena. La unión puede ser para los fines de la segmentación de ADN dirigida y/o la recombinación dirigida. Una región de interés puede estar presente en un cromosoma, un episoma, un genoma de organoides (p. ej., mitocondrial, de cloroplastos), o un genoma viral que infecta, por ejemplo. Una región de interés puede estar dentro de la región codificante de un gen, dentro de regiones no codificadas transcritas, tales como, p. ej., secuencias líder, secuencias

o intrones finales, o dentro de regiones no transcritas, corriente arriba o corriente abajo de la región codificante. Una región de interés puede ser tan pequeña como un solo par de nucleótidos o hasta 2.000 pares de nucleótidos de longitud, o cualquier valor integral de pares de nucleótidos.

5 **[0073]** Las células "eucariotas" incluyen, pero no se limitan a, células de hongos (tales como levadura), células vegetales, células animales, células de mamíferos y células humanas (p. ej., células T).

10 **[0074]** Los términos "unión operativa" y "unido operativamente" (o "unido operativamente") se usan indistintamente con referencia a una yuxtaposición de dos o más componentes (tales como elementos de secuencia), en donde los componentes están dispuestos de tal manera que ambos componentes funcionan normalmente y permiten la posibilidad de que al menos uno de los componentes pueda mediar una función que se ejerce sobre al menos uno de los otros componentes. A modo de ilustración, una secuencia reguladora de la transcripción, como un promotor, está unida operativamente a una secuencia de codificación si la secuencia reguladora de la transcripción controla el nivel de transcripción de la secuencia de codificación en respuesta a la presencia o ausencia de uno o más factores reguladores de la transcripción. Una secuencia reguladora de la transcripción generalmente está unida operativamente en *cis* con una secuencia de codificación, pero no necesita ser directamente adyacente a ella. P. ej., un potenciador es una secuencia reguladora de la transcripción que está unida operativamente a una secuencia de codificación, aunque no sean contiguas.

20 **[0075]** Con respecto a los polipéptidos de fusión, el término "operativamente enlazado" puede referirse al hecho de que cada uno de los componentes realiza la misma función en vinculación con el otro componente como lo haría si no fuera así vinculado. P. ej., con respecto a un polipéptido de fusión en el que un dominio de unión a ADN ZFP, TALE o Cas se fusiona a un dominio de activación, el dominio de unión a ADN ZFP, TALE o Cas y el dominio de activación están en enlace operativo si, en el polipéptido de fusión, la ZFP, TALE de la porción del dominio de unión a ADN de Cas es capaz de unirse a su sitio diana y/o su sitio de unión, mientras que el dominio de activación es capaz de regular la expresión de genes. Cuando un polipéptido de fusión en el cual un dominio de unión a ZFP, TALE o Cas se une a un dominio de escisión, el dominio de unión a ZFP, TALE o Cas de ADN y el dominio de escisión están en enlace operativo si, en el polipéptido de fusión, la porción del dominio de unión al ADN de ZFP, TALE o Cas es capaz de unirse a su sitio diana y/o su sitio de unión, mientras que el dominio de escisión es capaz de escindir el ADN en la vecindad del sitio diana.

35 **[0076]** Un "fragmento funcional" de una proteína, polipéptido o ácido nucleico es una proteína, polipéptido o ácido nucleico cuya secuencia no es idéntica a la proteína de longitud completa, polipéptido o ácido nucleico, sin embargo, conserva la misma función que la completa proteína de longitud, polipéptido o ácido nucleico. Un fragmento funcional puede poseer más, menos o el mismo número de residuos que la molécula nativa correspondiente, y/o puede contener una o más sustituciones de aminoácidos o nucleótidos. Los métodos para determinar la función de un ácido nucleico (p. ej., la función de codificación, la capacidad de hibridar con otro ácido nucleico) son bien conocidos en la técnica. De manera similar, los métodos para determinar la función de la proteína son bien conocidos. P. ej., la función de unión al ADN de un polipéptido se puede determinar, p. ej., mediante ensayos de unión al filtro, desplazamiento de la movilidad electroforética o inmunoprecipitación. La escisión del ADN se puede analizar mediante electroforesis en gel. Ver Ausubel *et al.*, *Supra*. La capacidad de una proteína para interactuar con otra proteína se puede determinar, p. ej., mediante co-inmunoprecipitación, ensayos de dos híbridos o complementación, tanto genética como bioquímica. Ver, p. ej., Fields *et al.* (1989) *Nature* **340**: 245-246; Patente de EE.UU. Nº 5.585.245 y PCT WO 98/44350,

45 **[0077]** Un "vector" es capaz de transferir secuencias génicas a células diana. Típicamente, "construcción de vector", "vector de expresión" y "vector de transferencia de genes" significa cualquier construcción de ácido nucleico capaz de dirigir la expresión de un gen de interés y que puede transferir secuencias de genes a células diana. Por lo tanto, el término incluye la clonación y los vehículos de expresión, así como la integración de vectores.

50 **[0078]** Los términos "sujeto" y "paciente" se usan indistintamente y se refieren a mamíferos tales como pacientes humanos y primates no humanos, así como animales experimentales tales como conejos, perros, gatos, ratas, ratones y otros animales. En consecuencia, el término "sujeto" o "paciente", tal como se usa en el presente documento, significa cualquier paciente o sujeto de mamífero al que se pueden administrar las células madre o las células madre de la invención. Los sujetos de la presente invención incluyen aquellos que han sido expuestos a una o más toxinas químicas, incluyendo, p. ej., una toxina nerviosa.

60 **[0079]** "Tallo" se refiere a la capacidad relativa de cualquier célula para actuar de manera similar a una célula madre, es decir, el grado de toti-, pluri- u oligopotencia y auto-renovación expandida o indefinida que puede tener cualquier célula madre particular.

Nucleasas

65 **[0080]** En el presente documento se describen composiciones, particularmente nucleasas, que son útiles para la escisión *in vivo* de una molécula donadora que porta un transgén y nucleasas para la escisión del genoma de una

célula de manera que el transgén se integra en el genoma de una manera específica. En ciertas realizaciones, una o más de las nucleasas son naturales. En otras realizaciones, una o más de las nucleasas son de origen no natural, es decir, diseñadas en el dominio de unión al ADN y/o en el dominio de escisión. P. ej., el dominio de unión a ADN de una nucleasa de origen natural puede ser alterada para unirse a un sitio diana seleccionado (p. ej., una meganucleasa que ha sido diseñada para unirse al sitio diferente que el sitio de unión reconocido). En otras formas de realización, la nucleasa comprende dominios de unión a ADN y la escisión heterólogas (p. ej., nucleasas de dedos de zinc; proteínas de unión de dominio de ADN de efector TAL; dominios de unión ADN de meganucleasa con los dominios de escisión heterólogos).

A. Dominios de unión al ADN

[0081] En ciertas realizaciones, la composición y los métodos descritos en el presente documento emplean una meganucleasa (endonucleasas de anidación) de dominio de unión a ADN para la unión a la molécula donante y/o la unión a la región de interés en el genoma de la célula. Las meganucleasas naturales reconocen los sitios de escisión de 15-40 pares de bases y generalmente se agrupan en cuatro familias: la familia LAGLIDADG, la familia GIY-YIG, la familia de cajas His-Cyst y la familia HNH. Las endonucleasas homing ejemplares incluyen I-SceI, I-CeuI, PI-PspI, PI-Sce, I-SceIV, I-Csm I, I-Pan I, I-Scell, I-PpoI, I-ScellI, I-CreI, I-TevI, I-TevII y I-TevIII. Sus secuencias de reconocimiento son conocidas. Véase también la Patente de Estados Unidos N° 5.420.032; Patente de Estados Unidos N° 6.833.252; Belfort y otros (1997) Nucleic Acids Res. 25: 3379-3388; Dujon et al. (1989) Gene 82: 115-118; Perler y otros (1994) Nucleic Acids Res. 22, 1125-1127; Jasin (1996) Trends Genet. 12: 224-228; Gimble et al. (1996) J. Mol. Biol. 263: 163-180; Argast et al. (1998) J. Mol. Biol. 280: 345-353 y el catálogo de New England Biolabs.

[0082] En ciertas realizaciones, los métodos y composiciones descritos en este documento hacen uso de una nucleasa que comprende una endonucleasa (meganucleasa) modificada (de origen no natural). Las secuencias de reconocimiento de endonucleasas de la vivienda y meganucleasas tales como I-SceI, I-CeuI, PI-PspI, PI-Sce, I-SceIV, I-CsmI, I-PanI, I-Scell, I-PpoI, I-Sce III, I-CreI, I-TevI, I-TevII y I-TevIII son conocidos. Véase también la Patente de Estados Unidos N° 5.420.032; Patente de Estados Unidos N° 6.833.252; Belfort y otros (1997) Nucleic Acids Res. 25: 3379-3388; Dujon et al. (1989) Gene 82: 115-118; Perler y otros (1994) Nucleic Acids Res. 22, 1125-1127; Jasin (1996) Trends Genet. 12: 224-228; Gimble et al. (1996) J. Mol. Biol. 263: 163-180; Argast et al. (1998) J. Mol. Biol. 280: 345-353 y el catálogo de New England Biolabs. Además, la especificidad de unión al ADN de las endonucleasas y meganucleasas homing puede diseñarse para unirse a sitios de destino no naturales. Ver, p. ej., Chevalier et al. (2002) Molec. Cell 10: 895-905; Epinat y otros (2003) Nucleic Acids Res. 31: 2952-2962; Ashworth et al. (2006) Nature 441: 656-659; Paques et al. (2007) Current Gene Therapy 7: 49-66; Publicación de Patente de Estados Unidos N° 20070117128. Los dominios de unión al ADN de las endonucleasas y meganucleasas de referencia pueden alterarse en el contexto de la nucleasa como un todo (es decir, de tal manera que la nucleasa incluya el dominio de escisión cognado) o pueden fusionarse a un dominio de escisión heterólogo.

[0083] En otras realizaciones, el dominio de unión a ADN de una o más de las nucleasas usadas en los métodos y composiciones descritos en el presente documento comprende un dominio de unión de ADN efector TAL de origen natural o diseñados (de origen no natural). Véase, p. ej., la Patente de Estados Unidos N° 8.586.526. Se sabe que las bacterias patógenas para las plantas del género *Xanthomonas* causan muchas enfermedades en importantes plantas de cultivo. La patogenicidad de *Xanthomonas* depende de un sistema conservado de secreción tipo III (T3S) que inyecta más de 25 proteínas efectoras diferentes en la célula vegetal. Entre estas proteínas inyectadas se encuentran los efectores similares a activadores de transcripción (TAL) que imitan a los activadores transcripcionales de las plantas y manipulan el transcriptoma de la planta (ver Kay et al (2007) Science 318: 648-651). Estas proteínas contienen un dominio de unión a ADN y un dominio de activación transcripcional. Uno de los efectores TAL mejor caracterizados es AvrBs3 de *Xanthomonas campestris* pv. *Vesicatoria* (ver Bonas et al (1989) Mol Gen Genet 218: 127-136 y WO2010079430). Los efectores de TAL contienen un dominio centralizado de repeticiones en tándem, cada una de las cuales contiene aproximadamente 34 aminoácidos, que son clave para la especificidad de unión a ADN de estas proteínas. Además, contienen una secuencia de localización nuclear y un dominio de activación transcripcional ácido (para una revisión, véase Schornack S, et al. (2006) J Plant Physiol 163 (3): 256-272). Además, en la bacteria fitopatógena *Ralstonia solanacearum* se han encontrado dos genes, denominados brg11 y hpx17, que son homólogos a la familia AvrBs3 de *Xanthomonas* en la cepa *R. solanacearum* biovar 1 GMI1.000 y en la cepa biovar 4 RS1.000 (véase Heuer et al (2007) Appl y Envir Micro 73 (13): 4379-4384). Estos genes son 98,9% idénticos en secuencia de nucleótidos entre sí pero difieren por una eliminación de 1,575 pares de bases en el dominio de repetición de hpx17. Sin embargo, ambos productos génicos tienen menos de 40% de identidad de secuencia con las proteínas de la familia AvrBs3 de *Xanthomonas*. Ver, p. ej., Patente de Estados Unidos N° 8.586.526.

[0084] La especificidad de estos efectores TAL depende de las secuencias encontradas en las repeticiones en tándem. La secuencia repetida comprende aproximadamente 102 pares de bases (pb) y las repeticiones son típicamente 91-100% homólogas entre sí (Bonas et al, *ibid*). El polimorfismo de las repeticiones suele ubicarse en las posiciones 12 y 13 y parece haber una correspondencia uno a uno entre la identidad de los díresiduos hipervariable (RVD) en las posiciones 12 y 13 con la identidad de los nucleótidos contiguos en secuencia TAL diana del efector (véase Moscou y Bogdanove, (2009) Science 326: 1501 y Boch et al (2009) Science 326: 1509-1512).

Experimentalmente, el código natural para el reconocimiento de ADN de estos efectores TAL se ha determinado de tal manera que una secuencia HD en las posiciones 12 y 13 conduce a una unión a la citosina (C), NG se une a T, NI a A, C, G o T, NN se une a A o G, e ING se une a T. Estas repeticiones de unión a ADN se han ensamblado en proteínas con nuevas combinaciones y números de repeticiones, para hacer factores de transcripción artificial que puedan interactuar con nuevas secuencias y activar la expresión de un gen reportero no endógeno en células vegetales (Boch *et al, ibid*). Las proteínas TAL diseñadas se han vinculado a un medio dominio de escisión *FokI* para producir una fusión de nucleasa de dominio efector TAL (TALEN) que muestra actividad en un ensayo de indicador de levadura (diana basada en plásmidos). Ver, p. ej., Patente de Estados Unidos N° 8.586.526; Christian *et al* ((2010) < *Genetics* epub 10,1534/genetics,110,120717). En ciertas realizaciones, el dominio TALE comprende una N-cap y/o C-cap como se describe en la Patente de Estados Unidos número 8.586.526.

[0085] En ciertas realizaciones, el dominio de unión de ADN de uno o más de las nucleasas utilizadas para la unión *in vivo* de escisión y/o escisión dirigida del genoma de una célula comprende una proteína de dedo de zinc. Preferiblemente, la proteína de dedo de zinc no se produce de forma natural ya que está diseñada para unirse a un sitio diana de elección. Ver, p. ej., Ver, p. ej., Beerli *et al.* (2002) *Nature Biotechnol.*, 20:135-141; Pabo *et al.* (2001) *Ann. Rev. Biochem.*, 70: 313-340; Isalan *et al.* (2001) *Nature Biotechnol.*, 19: 656-660; Segal *et al.* (2001) *Curr. Opin. Biotechnol.*, 12: 632-637; Choo *et al.* (2000) *Curr. Opin. Struct Biol.*, 10: 411-416; Las patentes de EE.UU. n° 6.453.242; 6.534.261; 6.599.692; 6.503.717; 6.689.558; 7.030.215; 6.794.136; 7.067.317; 7.262.054; 7.070.934; 7.361.635; 7.253.273; y las Publicaciones de Patentes de Estados Unidos N° 2005/0064474; 2007/0218528; 2005/0267061,

[0086] Un dominio de unión de dedos de cinc puede tener una especificidad de unión novedosa, en comparación con una proteína de dedos de cinc de origen natural. Los métodos de ingeniería incluyen, pero no se limitan a, diseño racional y varios tipos de selección. El diseño racional incluye, p. ej., el uso de bases de datos que comprenden secuencias de nucleótidos triplete (o cuadruplete) y secuencias de aminoácidos de dedo de zinc individuales, en las que cada secuencia de nucleótido triplete o cuadruplete está asociada con una o más secuencias de aminoácidos de dedos de zinc que se unen al triplete particular o secuencia cuadruplete. Ver, p. ej., co-propiedad de las patentes de EE.UU. 6.453.242 y 6.534.261,

[0087] Los métodos de selección ejemplares, que incluyen la visualización de fagos y los sistemas de dos híbridos, se describen en las patentes de EE.UU. 5.789.538; 5.925.523; 6.007.988; 6.013.453; 6.410.248; 6.140.466; 6.200.759; y 6.242.568; así como el documento WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 y GB 2.338.237. Además, la mejora de la especificidad de unión para los dominios de unión al dedo de zinc se ha descrito, p. ej., en el documento de propiedad conjunta WO 02/077227.

[0088] Además, como se describe en estas y otras referencias, dominios de dedos de zinc y/o proteínas de dedos de zinc de múltiples dedos pueden unirse entre sí utilizando cualquier secuencia de enlazador adecuado, incluyendo p. ej., enlazadores de 5 o más aminoácidos de longitud. Véanse, también, las patentes de EE.UU. n° 6.479.626; 6.903.185; y 7.153.949 para las secuencias de engarce ejemplares 6 o más aminoácidos en longitud. Las proteínas descritas en el presente documento pueden incluir cualquier combinación de enlazadores adecuados entre los dedos de zinc individuales de la proteína.

[0089] La selección de sitios de destino; Los ZFP y los métodos para el diseño y construcción de proteínas de fusión (y polinucleótidos que los codifican) son conocidos por los expertos en la técnica y se describen en detalle en las patentes de EE.UU. N° 6.140.081; 5.789.538; 6.453.242; 6.534.261; 5.925.523; 6.007.988; 6.013.453; 6.200.759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197; WO 02/099084; WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 y WO 03/016496.

[0090] Casi cualquier enlazador (espaciador) puede utilizarse entre uno o más de los componentes del dominio de unión a ADN (p. ej., dedos de zinc), entre uno o más dominios de unión al ADN y/o entre el ADN dominio de unión y el dominio funcional (p. ej., nucleasa). Los ejemplos no limitantes de secuencias de engarce adecuadas incluyen las patentes de EE.UU. N° 8.772.453; 7.888.121; 6.479.626; 6.903.185; y 7.153.949; Publicación de los Estados Unidos N° 20090305419 y Solicitud de los Estados Unidos N° 14/471.782. Por lo tanto, las proteínas descritas en este documento pueden incluir cualquier combinación de enlazadores adecuados entre los componentes de unión a ADN individuales y/o entre el dominio de unión a ADN y el dominio funcional de las composiciones descritas en este documento.

[0091] Por lo tanto, la nucleasa puede comprender un dominio de unión a ADN que se une específicamente a un sitio diana en cualquier gen en el que se desea insertar un donante (transgén).

B. Dominios de escisión

[0092] Cualquier dominio de escisión adecuado puede ser ligado operativamente a un dominio de unión al ADN para formar una nucleasa. P. ej., los dominios de unión al ADN de ZFP se han fusionado con los dominios de nucleasa para crear ZFN, una entidad funcional que es capaz de reconocer su objetivo de ácido nucleico deseado a través de su dominio de unión al ADN diseñado (ZFP) y hacer que el ADN se corte cerca del sitio de unión ZFP a través de la actividad nucleasa. Ver, p. ej., Kim *et al.* (1996) *Proc Nat'l Acad Sci EE.UU.* 93 (3): 1156-1160, Más recientemente,

los ZFN se han utilizado para la modificación del genoma en una variedad de organismos. Ver, p. ej., las Publicaciones de Patentes de Estados Unidos 20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20060063231; y la publicación internacional WO 07/014275. Del mismo modo, los dominios de unión al ADN TALE se han fusionado con los dominios de nucleasa para crear TALENs. Ver, p. ej., Publicación de Estados Unidos N° 20110301073.

[0093] Como se señaló anteriormente, el dominio de escisión puede ser heterólogo al dominio de unión a ADN, por ejemplo un dominio de unión a ADN de dedos de cinc y un dominio de escisión de una nucleasa o un dominio TALEN de unión a ADN y un dominio de escisión, o dominio de unión al ADN de meganucleasa y dominio de escisión de una nucleasa diferente. Los dominios de escisión heteróloga se pueden obtener a partir de cualquier endonucleasa o exonucleasa. Las endonucleasas ejemplares a partir de las cuales se puede derivar un dominio de escisión incluyen, entre otras, endonucleasas de restricción y endonucleasas homing. Ver, p. ej., 2002-2003 Catalog, New England Biolabs, Beverly, MA; y Belfort et al. (1997) *Nucleic Acids Res.* 25: 3379-3388. Se conocen enzimas adicionales que escinden el ADN (p. ej., nucleasa S1; nucleasa de frijol mungo; ADNasa I pancreática; nucleasa microcócica; endonucleasa HO de levadura; véase también Linn y otros (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993). Una o más de estas enzimas (o fragmentos funcionales de las mismas) se pueden usar como una fuente de dominios de escisión y semidominios de escisión.

[0094] De manera similar, un semi-dominio de escisión se puede derivar de cualquier nucleasa o porción del mismo, como se expuso anteriormente, que requiere dimerización para la actividad de escisión. En general, se requieren dos proteínas de fusión para la escisión si las proteínas de fusión comprenden semi-dominios de escisión. Alternativamente, se puede usar una proteína única que comprende dos semi-dominios de escisión. Los dos semi-dominios de escisión pueden derivarse de la misma endonucleasa (o fragmentos funcionales de la misma), o cada semidominio de escisión se puede derivar de una endonucleasa diferente (o fragmentos funcionales de la misma). Además, los sitios diana para las dos proteínas de fusión se disponen preferiblemente, uno con respecto al otro, de manera que la unión de las dos proteínas de fusión a sus sitios diana respectivos coloca los semi-dominios de escisión en una orientación espacial entre sí que permite la dividir los semi-dominios para formar un dominio de escisión funcional, p. ej., por dimerización. Por lo tanto, en ciertas realizaciones, los bordes cercanos de los sitios objetivo están separados por 5-8 nucleótidos o por 15-18 nucleótidos. Sin embargo, cualquier número entero de nucleótidos o pares de nucleótidos puede intervenir entre dos sitios diana (p. ej., de 2 a 50 pares de nucleótidos o más). En general, el sitio de escisión se encuentra entre los sitios de destino.

[0095] Las endonucleasas de restricción (enzimas de restricción) están presentes en muchas especies y son capaces de una unión específica de secuencia al ADN (en un sitio de reconocimiento), y de escindir el ADN en el sitio de unión o cerca de él. Ciertas enzimas de restricción (p. ej., tipo IIS) ADN escinden en sitios retirados del sitio de reconocimiento y que tienen dominios de unión separable y de escisión. P. ej., la enzima tipo IIS *FokI* cataliza la escisión bicatenaria del ADN, a 9 nucleótidos desde su sitio de reconocimiento en una cadena y 13 nucleótidos desde su sitio de reconocimiento en la otra. Ver, p. ej., las patentes de EE.UU. 5.356.802; 5.436.150 y 5.487.994; así como Li et al. (1992) *Proc. Natl Acad Sci. EE.UU.* 89: 4275-4279; Li et al. (1993) *Proc. Natl Acad Sci. EE.UU.* 90: 2764-2768; Kim et al. (1994a) *Proc. Natl Acad Sci. EE.UU.* 91: 883-887; Kim et al. (1994b) *J. Biol. Chem.* 269: 31,978-31,982. Por lo tanto, en una realización, las proteínas de fusión comprenden el dominio de escisión (o semidominio de escisión) de al menos una enzima de restricción de Tipo IIS y uno o más dominios de unión a los dedos de zinc, que pueden o no estar diseñados.

[0096] Una enzima de restricción de tipo IIS ejemplar, cuyo dominio de escisión es separable del dominio de unión, es *FokI*. Esta enzima particular es activa como un dímero. Bitinaite et al. (1998) *Proc. Natl Acad Sci. EE.UU.* 95:10,570-10,575. Por consiguiente, para los fines de la presente divulgación, la porción de la enzima *FokI* usada en las proteínas de fusión descritas se considera un semi-dominio de escisión. Por lo tanto, para la segmentación de doble cadena dirigida y/o la sustitución selectiva de secuencias celulares que utilizan fusiones *FokI* con dedos de zinc, se pueden usar dos proteínas de fusión, cada una de las cuales comprende un semi-dominio de escisión *FokI*, para reconstituir un dominio de escisión catalíticamente activo. Alternativamente, también se puede usar una única molécula polipeptídica que contiene un dominio de unión a los dedos de zinc y dos semi-dominios de escisión de *FokI*. Los parámetros para la escisión dirigida y la alteración de la secuencia dirigida utilizando fusiones de zinc dedo-*FokI* se proporcionan en otra parte de esta descripción.

[0097] Un dominio de escisión o escisión semi-dominio puede ser cualquier porción de una proteína que conserva la actividad de escisión, o que conserva la capacidad de multimerizar (p. ej., dimerizan) para formar un dominio de escisión funcional.

[0098] Enzimas de restricción de tipo IIS ejemplar se describen en la publicación internacional WO 07/014275. Las enzimas de restricción adicionales también contienen dominios de unión y escisión separables, y estos están contemplados por la presente descripción. Ver, p. ej., Roberts et al. (2003) *Nucleic Acids Res.* 31: 418-420,

[0099] En ciertas realizaciones, el dominio de escisión comprende uno o más semi-dominios de escisión (también denominados mutantes de dominio de dimerización) que minimizan o evitan la homodimerización, como se describe, p. ej., en Ver, p. ej., patentes de EE.UU. n° 7.914.796; 8.034.598 y 8.623.618. Los residuos de aminoácidos en las

posiciones 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537 y 538 de *FokI* son todos objetivos para influir en la dimerización de la *FokI* se divide en medio dominios.

[0100] Los semi-dominios de escisión por ingeniería ejemplares de *FokI* que forman heterodímeros obligados incluyen un par en el que un primer semi-dominio de escisión incluye mutaciones en los residuos de aminoácidos en las posiciones 490 y 538 de *FokI* y un segundo semi-dominio de escisión incluye mutaciones en los residuos de aminoácidos 486 y 499.

[0101] Por lo tanto, en una realización, una mutación en 490 sustituye Glu (E) con Lys (K); la mutación en 538 reemplaza Iso (I) con Lys (K); la mutación en 486 reemplazó a Gln (Q) con Glu (E); y la mutación en la posición 499 reemplaza Iso (I) con Lys (K). Específicamente, los semi-dominios de escisión de ingeniería descritos aquí se prepararon mutando las posiciones 490 (E → K) y 538 (I → K) en un semi-dominio de escisión para producir un semi-dominio de escisión de ingeniería denominado "E490K: I538K" y por las posiciones de mutación 486 (Q → E) y 499 (I → L) en otro semi-dominio de escisión para producir un semi-dominio de escisión diseñado como "Q486E: I499L". Los semi-dominios de escisión de ingeniería descritos aquí son mutantes heterodímeros obligados en los que se minimiza o anula la escisión aberrante. Ver, p. ej., Publicación de Patente de Estados Unidos N° 2008/0131962. En ciertas realizaciones, el semi-dominio de escisión por ingeniería comprende mutaciones en las posiciones 486, 499 y 496 (numeradas con relación a *FokI* de tipo salvaje), por ejemplo mutaciones que reemplazan el residuo Gln (Q) de tipo salvaje en la posición 486 con un Glu (E), el residuo Iso (I) de tipo salvaje en la posición 499 con un residuo Leu (L) y el residuo Asn (N) de tipo salvaje en la posición 496 con un residuo Asp (D) o Glu (E) (también denominado como un dominio "ELD" y "ELE", respectivamente). En otras realizaciones, el semi-dominio de escisión por ingeniería comprende mutaciones en las posiciones 490, 538 y 537 (numeradas con relación a *FokI* de tipo salvaje), por ejemplo mutaciones que reemplazan el residuo de Glu (E) de tipo salvaje en la posición 490 con un Lys (K) residuo, el residuo Iso (I) de tipo salvaje en la posición 538 con un residuo Lys (K), y el residuo His (H) de tipo salvaje en la posición 537 con un residuo Lys (K) o un residuo Arg (R) (también conocidos como dominios "KKK" y "KKR", respectivamente). En otras realizaciones, el semi-dominio de escisión por ingeniería comprende mutaciones en las posiciones 490 y 537 (numeradas con relación a *FokI* de tipo salvaje), por ejemplo mutaciones que reemplazan el residuo de Glu (E) de tipo salvaje en la posición 490 con un residuo de Lys (K) y el residuo His (H) de tipo salvaje en la posición 537 con un residuo Lys (K) o un residuo Arg (R) (también conocidos como dominios "KKK" y "KKR", respectivamente). Ver, p. ej., la patente de EE.UU. N°s 7.914.796; 8.034.598 y 8.623.618. En otras realizaciones, el medio dominio de escisión por ingeniería comprende las mutaciones "Sharkey" y/o "Sharkey" (véase Guo et al, (2010) J. Mol. Biol. 400 (1): 96-107).

[0102] Los semi-dominios de escisión de ingeniería descritos en el presente documento pueden prepararse usando cualquier método adecuado, p. ej., mediante mutagénesis dirigida de semi-dominios de escisión de tipo salvaje (*FokI*) como se describe en las publicaciones de patente de EE.UU. Números 20050064474; 20080131962; y 20110201055.

[0103] Alternativamente, las nucleasas pueden ensamblarse *in vivo* en el sitio nucleico diana de ácido usando la llamada tecnología de "enzima escindida" (ver, p. ej., la Publicación de Patente de Estados Unidos N° 20090068164). Los componentes de tales enzimas divididas se pueden expresar en construcciones de expresión separadas, o se pueden vincular en un marco de lectura abierto donde los componentes individuales se separan, p. ej., por un péptido 2A auto-escindible o una secuencia IRES. Los componentes pueden ser dominios de unión a dedos de zinc individuales o dominios de un dominio de unión a ácido nucleico de meganucleasa.

[0104] Las nucleasas pueden seleccionarse para determinar su actividad antes de su uso, p. ej., en un sistema cromosómico basado en levadura como se describe en los documentos WO2009/042163 y 20090068164. Los constructos de expresión de nucleasas se pueden diseñar fácilmente usando métodos conocidos en la técnica. Ver, p. ej., Publicaciones de Patentes de Estados Unidos 20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20060063231; y la publicación internacional WO 07/014275. La expresión de la nucleasa puede estar bajo el control de un promotor constitutivo o un promotor inducible, p. ej., el promotor galactoquinasa que se activa (desrepresión) en presencia de rafinosa y/o galactosa y se reprime en presencia de glucosa.

Sitios de destino

[0105] Como se describe en detalle anteriormente, los dominios de ADN se pueden diseñar para unirse a cualquier secuencia de elección. Un dominio de unión a ADN diseñado puede tener una nueva especificidad de unión, en comparación con un dominio de unión a ADN natural. En ciertas realizaciones de la divulgación, los dominios de unión a ADN se unen a una secuencia dentro de una secuencia potenciadora BCL11A, por ejemplo un sitio diana (típicamente 9, 10, 11, 12, 13, 14, 15, 16, 17, 18), 19, 20, 21 o incluso más pares de bases) se encuentra entre el exón 2 y el exón 3 de BCL11A, incluidos los dominios de unión al ADN que se unen a una secuencia dentro de un sitio hipersensible a ADNasa en la secuencia del potenciador de BCL11A (p. ej., +55, +58, +62; ver Figura 11). Los métodos de ingeniería incluyen, pero no se limitan a, diseño racional y varios tipos de selección. El diseño racional incluye, p. ej., el uso de bases de datos que comprenden secuencias de nucleótidos triplete (o cuadrupleto) y secuencias de aminoácidos de dedo de zinc individuales, en las que cada secuencia de nucleótido triplete o cuadrupleto está asociada con una o más secuencias de aminoácidos de dedos de zinc que se unen al triplete

particular o secuencia cuádruple. Ver, p. ej., co-propiedad de las patentes de EE.UU. 6.453.242 y 6.534.261, También se puede realizar un diseño racional de los dominios de efector TAL. Ver, p. ej., Publicación de Estados Unidos N° 20110301073.

[0106] Los métodos de selección a modo de ejemplo aplicables a los dominios de unión al ADN, que incluyen la presentación de fagos y los sistemas de dos híbridos, se describen en las patentes de EE.UU. 5.789.538; 5.925.523; 6.007.988; 6.013.453; 6.410.248; 6.140.466; 6.200.759; y 6.242.568; así como el documento WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 y GB 2,338,237. Además, la mejora de la especificidad de unión para los dominios de unión al dedo de zinc se ha descrito, p. ej., en el documento de propiedad conjunta WO 02/077227.

[0107] La selección de sitios de destino; nucleasas y métodos para el diseño y construcción de proteínas de fusión (y polinucleótidos que codifican las mismas) son conocidas por los expertos en la técnica y se describen en detalle en la solicitud de patente US N°s 20050064474 y 20060188987.

[0108] Además, como se describe en estas y otras referencias, los dominios de unión a ADN (p. ej., proteínas de dedos de zinc de múltiples dedos) y/o fusiones de dominio(s) de unión a ADN y dominio(s) funcional(es) se pueden vincular entre sí utilizando cualquier secuencia de enlace adecuada, incluyendo P. ej., enlazadores de 5 o más aminoácidos. Las patentes de EE.UU. n° 8.772.453; 7.888.121 (p. ej., enlazador "ZC"); 6.479.626; 6.903.185; y 7.153.949; Publicación de los Estados Unidos N° 20090305419) y la Solicitud de los Estados Unidos N° 14/471.782. Las proteínas descritas en el presente documento pueden incluir cualquier combinación de enlazadores adecuados entre los dominios de unión a ADN individuales de la proteína. Véase, *asimismo*, la patente de EE.UU. n° 8.586.526.

Donantes

[0109] En ciertas realizaciones, la presente descripción se refiere a la integración dirigida mediada por la nucleasa de una secuencia exógena en el genoma de una célula usando las moléculas de unión de región potenciadora de BCL11A descrita en este documento. Como se señaló anteriormente, la inserción de una secuencia exógena (también llamada "secuencia donante" o "donante" o "transgén"), p. ej., para la eliminación de una región específica y/o la corrección de un gen mutante o para la expresión incrementada de un gen de tipo silvestre. Será fácilmente evidente que la secuencia del donante generalmente no es idéntica a la secuencia genómica donde se coloca. Una secuencia donante puede contener una secuencia no homóloga flanqueada por dos regiones de homología para permitir una HDR eficiente en la ubicación de interés o puede integrarse a través de mecanismos de reparación no homologados. Además, las secuencias del donante pueden comprender una molécula de vector que contiene secuencias que no son homólogas a la región de interés en la cromatina celular. Una molécula donante puede contener varias regiones discontinuas de homología con la cromatina celular y, p. ej., conducir a la eliminación de una región potenciadora de Bel11a (o un fragmento de la misma) cuando se usa como un sustrato para la reparación de una DBS inducida por una de las nucleasas se describen aquí. Además, para la inserción dirigida de secuencias que normalmente no están presentes en una región de interés, dichas secuencias pueden estar presentes en una molécula de ácido nucleico donante y flanqueadas por regiones de homología a secuencia en la región de interés.

[0110] Los polinucleótidos para la inserción también se pueden denominar polinucleótidos "exógenos", polinucleótidos o moléculas "donantes" o "transgenes". El polinucleótido donante puede ser ADN o ARN, monocatenario y/o bicatenario y puede introducirse en una célula en forma lineal o circular. Ver, p. ej., Publicaciones de Patente de EE.UU. N°s 20100047805 y 20110207221, Las secuencias donantes están contenidas preferiblemente dentro de un ADN MC, que puede introducirse en la célula en forma circular o lineal. Si se introducen en forma lineal, los extremos de la secuencia del donante pueden protegerse (p. ej., de la degradación exonucleolítica) mediante métodos conocidos por los expertos en la técnica. P. ej., uno o más residuos de didesoxinucleótidos se agregan al extremo 3' de una molécula lineal y/o los oligonucleótidos autocomplementarios se ligan a uno o ambos extremos. Ver, p. ej., Chang et al. (1987) ProC. Natl Acad Sci. EE.UU., 84: 4959-4963; Nehls et al. (1996) Science 272: 886-889. Los métodos adicionales para proteger los polinucleótidos exógenos de la degradación incluyen, entre otros, la adición de uno o más grupos amino terminales y el uso de enlaces internucleótidos modificados como, p. ej., fosforotioatos, fosforamidatos y residuos de O-metilo ribosa o desoxirribosa. Si se introduce en forma de doble cadena, el donante puede incluir uno o más sitios diana de nucleasa, p. ej., sitios diana de nucleasa que flanquean el transgén para integrarse en el genoma de la célula. Ver, p. ej., Publicación de Patente de EE.UU. N° 20130326645.

[0111] Un polinucleótido se puede introducir en una célula como parte de una molécula de vector que tiene secuencias adicionales, tales como, p. ej., orígenes de replicación, promotores y genes que codifican resistencia a los antibióticos. Por otra parte, los polinucleótidos donantes pueden introducirse como ácido nucleico desnudo, como el ácido nucleico complejo con un agente tal como un liposoma o poloxámero, o puede ser entregado por los virus (p. ej., adenovirus, AAV, herpesvirus, retrovirus, lentivirus y lentivirus defectuoso por integrasa (IDLV)).

[0112] En ciertas realizaciones, el donante de doble cadena incluye secuencias (p. ej., secuencias de codificación, también llamadas transgenes) de más de 1 kb de longitud, por ejemplo entre 2 y 200 kb, entre 2 y 10kb (o cualquier valor entre ellos). El donante de doble cadena también incluye al menos un sitio diana de nucleasa, por ejemplo. En

ciertas realizaciones, el donante incluye al menos 2 sitios diana, p. ej., para un par de ZFN o TALEN. Típicamente, los sitios diana de nucleasas están fuera de las secuencias transgénicas, p. ej., 5' y/o 3' respecto a las secuencias transgénicas, para la escisión del transgén. El (los) sitio(s) de escisión de las nucleasas puede ser para cualquier nucleasa. En ciertas realizaciones, el (los) sitio(s) diana de la nucleasa contenido(s) en el donante bicatenario son para la misma(s) nucleasa(s) utilizada(s) para escindir la diana endógena en el cual el donante cortado se integra a través de métodos independientes de la homología.

[0113] El donante se inserta generalmente de modo que su expresión está dirigida por el promotor endógeno en el sitio de integración, es decir, el promotor que conduce la expresión del gen endógeno en el que se inserta el donante (p. ej., globina, AAVS1, etc.). Sin embargo, será evidente que el donante puede comprender un promotor y/o potenciador, por ejemplo un promotor constitutivo o un promotor inducible o específico de tejido.

[0114] La molécula donante se puede insertar en un gen endógeno de tal manera que todos, algunos o ninguno del gen endógeno se expresa. En otras realizaciones, el transgén (p. ej., con o sin secuencias de codificación de globina) está integrado en cualquier locus endógeno, por ejemplo un locus puerto seguro. Ver, p. ej., publicaciones de patentes estadounidenses 20080299580; 20080159996 y 201000218264.

[0115] Además, aunque no se requiere para la expresión, las secuencias exógenas también pueden incluir secuencias de transcripción o traducción de regulación, p. ej., promotores, potenciadores, aisladores, sitios de entrada de ribosoma interno, secuencias que codifican péptidos 2A y/o señales de poliadenilación.

[0116] Los transgenes realizados en las secuencias donantes descritas en este documento pueden aislarse a partir de plásmidos, células u otras fuentes utilizando técnicas estándar conocidas en la técnica, tales como PCR. Los donantes para su uso pueden incluir diversos tipos de topología, incluyendo circular superenrollado, circular relajado, lineal y similares. Alternativamente, pueden sintetizarse químicamente usando técnicas de síntesis de oligonucleótidos estándar. Además, los donantes pueden estar metilados o carecer de metilación. Los donantes pueden estar en forma de cromosomas bacterianos o de levadura artificial (BAC o YAC).

[0117] Los polinucleótidos de doble cadena donantes descritos en este documento pueden incluir una o más bases y/o cadenas principales no naturales. En particular, la inserción de una molécula donante con citosinas metiladas se puede llevar a cabo utilizando los métodos descritos aquí para lograr un estado de inactividad transcripcional en una región de interés.

[0118] El polinucleótido (donante) exógeno puede comprender cualquier secuencia de interés (secuencia exógena). Las secuencias exógenas ejemplares incluyen, pero no se limitan a cualquier secuencia de codificación de polipéptido (p. ej., ADNc), secuencias promotoras, secuencias potenciadoras, etiquetas de epítipo, genes marcadores, los sitios de reconocimiento de enzimas de escisión y diversos tipos de construcciones de expresión. Los genes marcadores incluyen, pero no se limitan a, secuencias que codifican proteínas que median la resistencia a antibióticos (p. ej., resistencia a la ampicilina, resistencia a la neomicina, resistencia a G418, resistencia a puromicina), secuencias que codifican coloreado o proteínas fluorescentes o luminiscentes (p. ej., proteína fluorescente verde, proteína fluorescente roja, luciferasa), y las proteínas que median la mejora del crecimiento celular y/o amplificación de genes (p. ej., reductasa de dihidrofolato). Las etiquetas de epítipo incluyen, p. ej., una o más copias de FLAG, HiS, myc, Tap, HA o cualquier secuencia de aminoácidos detectable.

[0119] En una realización preferida, la secuencia exógena (transgén) comprende un polinucleótido que codifica cualquier polipéptido de la cual se desea la expresión en la célula, incluyendo, pero no limitado a anticuerpos, antígenos, enzimas, receptores (de la superficie celular o nucleares), hormonas, linfocinas, citoquinas, polipéptidos informadores, factores de crecimiento y fragmentos funcionales de cualquiera de los anteriores. Las secuencias codificantes pueden ser, p. ej., ADNc.

[0120] P. ej., la secuencia exógena puede comprender una secuencia que codifica un polipéptido que carece o no es funcional en el sujeto que tiene una enfermedad genética, que incluye pero no se limita a cualquiera de las siguientes enfermedades genéticas: acondroplasia, acromatopsia, deficiencia de maltasa ácida, deficiencia de adenosina desaminasa (OMIM n° 102700), adrenoleucodistrofia, síndrome de aicardi, deficiencia de antitripsina alfa-1, talasemia alfa, síndrome de insensibilidad a los andrógenos, síndrome de Apert, ventrículo derecho arritmogénico, displasia, telangiectasia atáutica, síndrome de Barth, beta-talasemia, síndrome de Bean, enfermedad de Canavan, enfermedades de granulomatosis crónicas (CGD), síndrome Cri du Chat, fibrosis quística, enfermedad de Dercum, displasia ectodérmica, anemia de Fanconi, síndrome progresivo fibrodysplasia ossificans, síndrome frágil X, galactosemia, enfermedad de Gaucher, gangliosidosis generalizadas (p. ej., GM1), la hemocromatosis, la mutación de la hemoglobina C en el 6° codón de beta-globina (HBC), hemofilia, enfermedad de Huntington, síndrome de Hurler, hipofosfatasa, síndrome de Klinefelter, enfermedad de Krabbes, síndrome de Langer-Giedion, deficiencia de adhesión de leucocitos (LAD, OMIM No. 116920), leucodistrofia, síndrome de QT largo, síndrome de Marfan, síndrome de Moebius, mucopolisacaridosis (MPS), síndrome de 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, enfermedad de células falciformes (anemia de células falciformes), síndrome de Smith-Magenis, síndrome de Stickler, enfermedad de Tay-Sachs, síndrome de Rausus de ausencia de trombocitopenia (TAR), síndrome de Treacher Collins, trisomía, esclerosis tuberosa, síndrome de Turner, trastorno del ciclo de urea, enfermedad de von Hippel-Landau, síndrome de Waardenburg, síndrome de Williams, enfermedad de Wilson, síndrome de Wiskott-Aldrich, síndrome linfoproliferativo ligado a X (XLP, OMIM N°. 308240).

[0121] Enfermedades ejemplares adicionales que pueden ser tratadas mediante integración dirigida incluyen inmunodeficiencias adquiridas, enfermedades de almacenamiento lisosomal (p. ej., enfermedad de Gaucher, GM1, la enfermedad de Fabry y enfermedad de Tay-Sachs), mucopolisacahidosi (p. ej., enfermedad de Hunter, enfermedad de Hurler), hemoglobinopatías (p. ej., enfermedades de células falciformes, HbC, α -talasemia, β -talasemia) y hemofilias.

[0122] En ciertas realizaciones, las secuencias exógenas pueden comprender un gen marcador (descrito anteriormente), lo que permite la selección de células que han experimentado la integración dirigida, y una secuencia de ligado que codifican una funcionalidad adicional. Los ejemplos no limitantes de genes marcadores incluyen GFP, marcadores de selección de fármacos y similares.

[0123] Secuencias de genes adicionales que pueden ser insertadas pueden incluir, p. ej., genes de tipo salvaje para reemplazar las secuencias mutadas. P. ej., una secuencia del gen del Factor IX de tipo salvaje puede insertarse en el genoma de una célula madre en la que se muta la copia endógena del gen. La copia de tipo salvaje puede insertarse en el locus endógeno, o alternativamente puede dirigirse a un locus de puerto seguro.

[0124] La construcción de tales casetes de expresión, siguiendo las enseñanzas de la presente memoria, utiliza metodologías bien conocidas en la técnica de biología molecular (véase, p. ej., Ausubel o Maniatis). Antes del uso del casete de expresión para generar un animal transgénico, la capacidad de respuesta del casete de expresión al inductor de tensión asociado con los elementos de control seleccionados se puede ensayar mediante la introducción del casete de expresión en una línea celular apropiada (p. ej., células primarias, células transformadas, o líneas celulares inmortalizadas).

[0125] Además, aunque no se requiere para la expresión, las secuencias exógenas pueden también ser secuencias reguladoras transcripcionales o traduccionales, p. ej., promotores, potenciadores, aisladores, sitios de entrada de ribosoma interno, secuencias que codifican péptidos 2A y/o señales de poliadenilación. Además, los elementos de control de los genes de interés pueden ser unidos operativamente a genes informadores para crear genes quiméricos (p. ej., casetes de expresión de reportero).

[0126] También se puede lograr la inserción dirigida de secuencia de ácido nucleico no codificante. Las secuencias que codifican ARN antisentido, ARNi, ARNs y micro ARN (miARN) también pueden usarse para inserciones dirigidas.

[0127] En realizaciones adicionales, el ácido nucleico donante puede comprender secuencias no codificantes que son sitios diana específicos para diseños de nucleasa adicionales. Posteriormente, las nucleasas adicionales pueden expresarse en células de modo que la molécula donante original se escinda y se modifica mediante la inserción de otra molécula donadora de interés. De esta manera, se pueden generar integraciones reiterativas de moléculas donadoras que permitan el apilamiento de rasgos en un lugar de interés particular o en un lugar de puerto seguro.

Entrega

[0128] Las nucleasas, los polinucleótidos que codifican estas nucleasas, polinucleótidos donantes y composiciones que comprenden las proteínas y/o polinucleótidos descritos en el presente documento se pueden administrar *in vivo* o *ex vitro* mediante cualquier medio adecuado en cualquier tipo de célula.

[0129] Las células adecuadas incluyen eucariota (p. ej., animal) y células procariotas y/o líneas celulares. Ejemplos no limitativos de tales células o líneas celulares generadas a partir de tales células incluyen COS, CHO (p. ej., CHO-S, CHO-K1, CHO-DG44, CHO-DUXB11, CHO-DUKX, CHOK1SV), VERO, MDCK, WI38, V79, B14AF28-G3, BHK, HaK, NS0, SP2/0-Ag14, HeLa, HEK293 (p. ej., HEK293-F, HEK293-H, HEK293-T), y células perC6, así como células de insectos como *Spodoptera frugiperda* (Sf), o células fúngicas como *Saccharomyces*, *Pichia* y *Schizosaccharomyces*. En ciertas realizaciones, la línea celular es una línea celular CHO, MDCK o HEK293. Las células adecuadas también incluyen células madre tales como, a modo de ejemplo, células madre embrionarias, células madre pluripotentes inducidas, células madre hematopoyéticas, células madre neuronales y células madre mesenquimáticas.

[0130] Los métodos de administración de nucleasas como se describen en el presente documento se describen, p. ej., en las patentes de EE.UU. N°s 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; y 7,163,824.

[0131] Las nucleasas y/o construcciones de donantes como se describen en el presente documento también pueden administrarse utilizando vectores que contienen secuencias que codifican uno o más de los sistemas ZFN(s), TALEN(s) o CRISPR/Cas. Se puede usar cualquier sistema de vector, incluidos, entre otros, vectores plasmídicos, vectores retrovirales, vectores lentivirales, vectores de adenovirus, vectores de poxvirus; vectores de virus del herpes y vectores de virus adeno-asociados, etc. Ver, también, la Patente de Estados Unidos N° 6.534.261; 6.607.882; 6.824.978; 6.933.113; 6.979.539; 7.013.219; y 7.163.824. Además, será evidente que cualquiera de estos vectores puede comprender una o más de las secuencias necesarias para el tratamiento. Por lo tanto, cuando una o más nucleasas y un constructo donante se introducen en la célula, las nucleasas y/o el polinucleótido donante pueden transportarse en el mismo vector o en diferentes vectores (ADN(s) MC(s)). Cuando se usan múltiples vectores, cada vector puede comprender una secuencia que codifica una o varias nucleasas y/o construcciones de donantes. Métodos convencionales de transferencia de genes basados en virus y no virales pueden utilizarse para introducir ácidos nucleicos que codifican nucleasas y/o construcciones de donantes en las células (p. ej., células de mamífero) y tejidos diana. Los sistemas de administración de vectores no virales incluyen plásmidos de ADN o ARN, ADN de MC, ácido nucleico desnudo y ácido nucleico complejo con un vehículo de administración tal como un liposoma o poloxámero. Los vectores no virales adecuados incluyen vectores de nanotaxis, incluyendo vectores disponibles comercialmente en InCellArt (Francia). Los sistemas de administración de vectores virales incluyen virus de ADN y ARN, que tienen genomas episomales o integrados después de la entrega a la célula. Para una revisión del suministro *in vivo* de proteínas de unión a ADN diseñadas y proteínas de fusión que comprenden estas proteínas de unión, consulte, p. ej., Rebar (2004) Expert Opinion Invest. Drogas 13 (7): 829-839; Rossi et al. (2007) Nature Biotech. 25 (12): 1444-1454, así como referencias generales de administración de genes, tales como Anderson, Science 256: 808-813 (1992); Nabel y Felgner, TIBTECH 11: 211-217 (1993); Mitani y Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357: 455-460 (1992); Van Brunt, Biotechnology 6 (10): 1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8: 35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51 (1): 31-44 (1995); Haddada et al., En Current Topics in Microbiology and Immunology Doerfler and Böhm (eds.) (1995); y Yu et al., Gene Therapy 1:13-26 (1994).

[0134] Los métodos de administración no viral de ácidos nucleicos incluyen electroporación, lipofección, microinyección, biolística, virosomas, liposomas, inmunoliposomas, policación o lípidos: conjugados de ácido nucleico, ADN desnudo, viriones artificiales y captación de ADN mejorada por el agente. La sonoporación que utiliza, p. ej., el sistema Sonitron 2.000 (Rich-Mar) también se puede usar para la administración de ácidos nucleicos.

[0133] Los sistemas de administración de ácido nucleico ejemplares adicionales incluyen los proporcionados por Amaxa Biosystems (Colonia, Alemania), Maxcyte, Inc. (Rockville, Maryland), BTX Molecular Delivery Systems (Holliston, MA) y Copernicus Therapeutics Inc., (ver, por ejemplo, US6008336). La lipofección se describe en p. ej., patentes de EE.UU. n° 5,049,386; 4,946,787; y 4,897,355) y lipofección reactivos se venden comercialmente (p. ej., Transfectam™ y Lipofectin™). Los lípidos catiónicos y neutros que son adecuados para la lipofección de polinucleótidos de reconocimiento de receptor eficiente incluyen los de Felgner, WO 91/17424, WO 91/16024.

[0134] La preparación de complejos de lípido: ácido nucleico, incluidos liposomas dirigidos tales como complejos de inmunolípido, es bien conocido para un experto en la técnica (ver, p. ej., Crystal, Science 270: 404-410 (1995); Blaese et al., Cancer Gene Ther. 2: 291-297 (1995); Behr et al., Bioconjugate Chem. 5: 382-389 (1994); Remy et al., Bioconjugate Chem. 5: 647-654 (1994); Gao et al., Gene Therapy 2: 710-722 (1995); Ahmad et al., Cancer Res. 52: 4817-4820 (1992); Patentes de Estados Unidos números 4.186.183. 4.217.344. 4.235.871. 4.261.975. 4.485.054. 4.501.728. 4.774.085. 4.837.028 y 4.946.787).

[0135] Los métodos adicionales de administración incluyen el uso de envasado de los ácidos nucleicos a ser entregados en vehículos de reparto EnGeneIC (VED). Estos EDV se suministran específicamente a los tejidos diana utilizando anticuerpos biespecíficos donde un brazo del anticuerpo tiene especificidad para el tejido diana y el otro tiene especificidad para el EDV. El anticuerpo lleva los EDV a la superficie de la célula diana y luego el EDV se introduce en la célula por endocitosis. Una vez en la célula, los contenidos se publican (ver MacDiarmid et al (2009) Nature Biotechnology 27 (7): 643).

[0136] El uso de sistemas basados en virus de ARN o ADN para la entrega de ácidos nucleicos que codifican por ingeniería ZFPS, TALEs y/o CRISPR/sistemas Cas aprovechan los procesos altamente evolucionados para dirigir un virus a células específicas en el cuerpo y el tráfico de la carga viral útil al núcleo. Los vectores virales pueden administrarse directamente a los pacientes (*in vivo*) o pueden usarse para tratar células *in vitro* y las células modificadas se administran a los pacientes (*ex vitro*). Los sistemas basados en virus convencionales para la administración de ZFP incluyen, entre otros, vectores de virus retroviral, lentivirus, adenoviral, adenoasociado, virus de la vacuna y herpes simple para la transferencia de genes. La integración en el genoma del huésped es posible con los métodos de transferencia de genes de retrovirus, lentivirus y virus adenoasociados, que a menudo dan como resultado la expresión a largo plazo del transgen insertado. Además, se han observado altas eficiencias de transducción en diferentes tipos de células y tejidos diana.

[0137] El tropismo de un retrovirus puede alterarse por incorporación de proteínas de la envoltura exterior, la ampliación de la potencial población diana de células diana. Los vectores lentivíricos son vectores retrovirales que

son capaces de transducir o infectar células que no se dividen y, por lo general, producen títulos virales elevados. La selección de un sistema de transferencia de genes retrovirales depende del tejido diana. Los vectores retrovirales están formados por repeticiones terminales largas que actúan en *cis* con capacidad de empaquetamiento de hasta 6-10 kb de secuencia extraña. Los LTR mínimos que actúan en *cis* son suficientes para la replicación y el empaquetamiento de los vectores, que luego se utilizan para integrar el gen terapéutico en la célula diana para proporcionar una expresión transgenular permanente. Los vectores retrovirales ampliamente utilizados incluyen aquellos que se basan en el virus de la leucemia murina (MuLV), el virus de la leucemia del mono gibón (GaLV), el virus de la inmunodeficiencia de Simio (SIV), el virus de la inmunodeficiencia humana (VIH) y sus combinaciones (ver, p. ej. Buchscher y otros, J. Virol. 66: 2731-2739 (1992); Johann y otros, J. Virol. 66: 1635-1640 (1992); Sommerfelt y otros, Virol. 176: 58-59 (1990).; Wilson et al., J. Virol. 63: 2374-2378 (1989); Miller et al., J. Virol. 65: 2220-2224 (1991); PCT/US94/05700).

[0138] En aplicaciones en las que se prefiere la expresión transitoria, los sistemas basados adenovirales pueden ser utilizados. 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, se han obtenido altos títulos y altos niveles de expresión. Este vector se puede producir en grandes cantidades en un sistema relativamente simple. Los vectores de virus adenoasociados ("AAV") también se usan para transducir células con ácidos nucleicos diana, p. ej., en la producción *in vitro* de ácidos nucleicos y péptidos, y para procedimientos de terapia génica *in vivo* y *ex vitro* (ver, p. ej., West et al., Virology 160: 38-47 (1987); Patente de EE.UU. N° 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5: 793-801 (1994); Muzyczka, J. Clin. Invest. 94: 1351 (1994). La construcción de vectores de AAV recombinantes se describe en varias publicaciones, incluida la Patente N° 5,173,414; Tratschin et al., Mol. Cell. Biol. 5: 3251-3260 (1985); Tratschin, et al., Mol. Célula. Biol. 4: 2072-2081 (1984); Hermonat y Muzyczka, PNAS 81: 6466-6470 (1984); y Samulski et al., J. Virol. 63: 03822-3828 (1989).

[0139] Actualmente, al menos seis enfoques de vectores virales están disponibles para la transferencia de genes en ensayos clínicos, que utilizan enfoques que implican la complementación de vectores defectuosos por genes insertados en líneas celulares auxiliares para generar el agente transductor.

[0140] pLASN y MFG-S son ejemplos de vectores retrovirales que se han utilizado en ensayos clínicos (Dunbar et al., Blood 85: 3048-305 (1995); Kohn et al., Nat. Med. 1:1017-102 (1995); Malech et al., PNAS 94:22 12133-12138 (1997)). PA317/pLASN fue el primer vector terapéutico utilizado en un ensayo de terapia génica. (Blaese et al., Science 270: 475-480 (1995)). Se han observado eficiencias de transducción de 50% o más para los vectores empaquetados de MFG-S. (Ellem y otros, Immunol Immunother. 44 (1): 10-20 (1997); Dranoff y otros, Hum. Gene Ther. 1:111-2 (1997).

[0141] Los vectores de virus adeno-asociados recombinantes (rAAV) son prometedores sistemas de suministro de genes alternativos basados en el virus tipo 2 adenoasociado defectuoso y no patógeno de parvovirus. Todos los vectores se derivan de un plásmido que retiene solo las repeticiones terminales invertidas de AAV de 145 pb que flanquean el casete de expresión del transgén. La transferencia eficiente de genes y el suministro estable de transgenes debido a la integración en los genomas de las células transducidas son características clave de este sistema de vectores. (Wagner y otros, Lancet 351: 9117 1702-3 (1998), Kearns y otros, Gene Ther. 9: 748-55 (1996)). De acuerdo con la presente invención, también se pueden usar otros serotipos de AAV, incluyendo AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9 y AAVrh,10 y cualquier nuevo serotipo de AAV.

[0142] Los vectores adenovirales recombinantes deficientes en la replicación (Ad) se pueden producir en títulos altos e infectan fácilmente varios tipos de células diferentes. La mayoría de los vectores de adenovirus están diseñados de manera tal que un transgén reemplaza los genes Ad E1a, E1b y/o E3; posteriormente, el vector defectuoso de replicación se propaga en células 293 humanas que suministran la función del gen eliminado en *trans*. Los vectores de anuncios pueden transducir múltiples tipos de tejidos *in vivo*, incluidas las células diferenciadas no divisorias, como las que se encuentran en el hígado, los riñones y los músculos. Los vectores de anuncios convencionales tienen una gran capacidad de carga. Un ejemplo del uso de un vector Ad en un ensayo clínico involucró la terapia con polinucleótidos para la inmunización antitumoral con inyección intramuscular (Sternan et al., Hum. Gene Ther. 7:1083-9 (1998)). Ejemplos adicionales del uso de vectores de adenovirus para la transferencia de genes en ensayos clínicos incluyen Rosenecker et al., Infección 24: 1 5-10 (1996); Sternan et al., Hum. Gene Ther. 9: 7 1083-1089 (1998); Welsh y otros, Hum. Gene Ther. 2: 205-18 (1995); Alvarez et al., Hum. Gene Ther. 5: 597 - 613 (1997); Topf et al., Gene Ther. 5: 507-513 (1998); Sternan et al., Hum. Gene Ther. 7:1083-1089 (1998).

[0143] Células de empaquetamiento se utilizan para formar partículas de virus que son capaces de infectar una célula huésped. Estas células incluyen células 293, que empaquetan adenovirus, y células ψ 2 o células PA317, que empaquetan retrovirus. Los vectores virales utilizados en la terapia génica generalmente son generados por una línea celular productora que empaqueta un vector de ácido nucleico en una partícula viral. Los vectores contienen típicamente las secuencias virales mínimas requeridas para el empaquetamiento y la posterior integración en un huésped (si corresponde), y otras secuencias virales se reemplazan por un casete de expresión que codifica la proteína a expresar. Las funciones virales faltantes se suministran en *trans* por la línea celular de empaquetamiento. P. ej., los vectores de AAV utilizados en la terapia génica típicamente solo poseen secuencias de repetición terminal invertida (ITR) del genoma de AAV que se requieren para el empaquetamiento y la integración en el genoma del

huésped. El ADN viral se empaqueta en una línea celular, que contiene un plásmido auxiliar que codifica los otros genes AAV, a saber, *rep* y *caP*, pero que carecen de secuencias ITR. La línea celular también está infectada con adenovirus como ayudante. El virus auxiliar promueve la replicación del vector AAV y la expresión de los genes AAV del plásmido auxiliar. El plásmido auxiliar no está empaquetado en cantidades significativas debido a la falta de secuencias ITR. La contaminación con adenovirus se puede reducir mediante, p. ej., un tratamiento térmico al que el adenovirus es más sensible que el AAV.

[0144] En muchas aplicaciones de terapia génica, es deseable que el vector de terapia génica puede entregar con un alto grado de especificidad para un tipo de tejido particular. Por consiguiente, un vector viral puede modificarse para tener especificidad para un tipo de célula dado expresando un ligando como una proteína de fusión con una proteína de la cubierta viral en la superficie exterior del virus. El ligando se elige para que tenga afinidad por un receptor que se sabe que está presente en el tipo de célula de interés. P. ej., Han et al., *Proc. Natl Acad Sci. EE.UU.* 92: 9747-9751 (1995), informó que el virus de la leucemia murina Moloney puede modificarse para expresar la heregulina humana fusionada a gp70, y el virus recombinante infecta ciertas células de cáncer de mama humano que expresan el receptor del factor de crecimiento epidérmico humano. Este principio puede extenderse a otros pares de células virus-diana, en las que la célula diana expresa un receptor y el virus expresa una proteína de fusión que comprende un ligando para el receptor de la superficie celular. P. ej., el fago filamentoso puede diseñarse para mostrar fragmentos de anticuerpos (p. ej., Fab o Fv) que tienen afinidad de unión específica para virtualmente cualquier receptor celular seleccionado. Aunque la descripción anterior se aplica principalmente a los vectores virales, los mismos principios se pueden aplicar a los vectores no virales. Dichos vectores pueden diseñarse para contener secuencias de captación específicas que favorecen la captación por células diana específicas.

[0145] Los vectores de terapia génica pueden administrarse *in vivo* mediante la administración a un paciente individual, típicamente por administración sistémica (p. ej., intravenosa, intraperitoneal, intramuscular, subdérmica, o infusión intracraneal) o aplicación tópica, como se describe a continuación. Alternativamente, los vectores pueden administrarse a células *ex vitro*, tales como células explantadas de un paciente individual (p. ej., linfocitos, aspirados de médula ósea, biopsia tisular) o células madre hematopoyéticas donantes universales, seguido por reimplantación de las células en un paciente, generalmente después de la selección de células que han incorporado el vector.

[0146] Los vectores (p. ej., retrovirus, adenovirus, liposomas, etc.) que contienen nucleasas y/o construcciones de donantes también se pueden administrar directamente a un organismo para transducción de células *in vivo*. Alternativamente, puede administrarse ADN desnudo. La administración se realiza mediante cualquiera de las rutas que se utilizan normalmente para introducir una molécula en contacto final con células de la sangre o tejidos, incluidas, entre otras, inyección, infusión, aplicación tópica y electroporación. Los métodos adecuados para administrar tales ácidos nucleicos están disponibles y son bien conocidos por los expertos en la técnica, y, aunque se puede usar más de una vía para administrar una composición particular, una vía particular a menudo puede proporcionar un método más inmediato y reacción más efectiva que otra vía.

[0147] Los vectores adecuados para la introducción de polinucleótidos (p. ej. donantes de codificación de nucleasa y/o doble cadena) se describe la presente invención incluyen vectores de lentivirus no integradores (IDLV). Ver, p. ej., Ory et al. (1996) *Proc. Natl Acad Sci. EE.UU.* 93: 11382-11388; Dull et al. (1998) *J. Virol.* 72: 8463-8471; Zuffery et al. (1998) *J. Virol.* 72: 9873-9880; Follenzi et al. (2000) *Nature Genetics* 25: 217-222; Publicación de Patente de Estados Unidos N° 2009/0117617.

[0148] Los vehículos farmacéuticamente aceptables se determinan en parte por la composición particular que se administra, así como por el método particular usado para administrar la composición. En consecuencia, existe una amplia variedad de formulaciones adecuadas de composiciones farmacéuticas disponibles, como se describe a continuación (ver, p. ej., Remington Pharmaceutical Sciences, 17ª ed., 1989).

[0149] Será evidente que las secuencias que codifican la nucleasa y las construcciones del donante se pueden administrar utilizando el mismo o diferentes sistemas. P. ej., las nucleasas y los donantes pueden ser transportados por el mismo ADN MC. Alternativamente, un polinucleótido donador puede ser transportado por un MC, mientras que una o más nucleasas pueden ser transportadas por un plásmido estándar o un vector AAV. Además, los diferentes vectores pueden administrarse por la misma vía o por diferentes vías (inyección intramuscular, inyección en la vena de la cola, otra inyección intravenosa, administración intraperitoneal y/o inyección intramuscular. Los vectores pueden administrarse simultáneamente o en cualquier orden secuencial.

[0150] Por lo tanto, la presente divulgación incluye el tratamiento *in vivo* o *ex vitro* de enfermedades y afecciones susceptibles de inserción de los transgenes que codifican una proteína terapéutica. Las composiciones se administran a un paciente humano en una cantidad eficaz para obtener la concentración deseada del polipéptido terapéutico en el suero o el órgano o células diana. La administración puede realizarse por cualquier medio en el que los polinucleótidos se administren a las células diana deseadas. P. ej., se contemplan métodos tanto *in vivo* como *ex vitro*. La inyección intravenosa en la vena porta es un método de administración preferido. Otros modos de administración *in vivo* incluyen, p. ej., la inyección directa en los lóbulos del hígado o el conducto biliar y la inyección intravenosa distal al hígado, incluida la arteria hepática, la inyección directa en el parénquima hepático, la inyección a través de la arteria hepática. y/o inyección retrógrada a través del árbol biliar. Los modos de administración *ex vitro*

incluyen la transducción *in vitro* de hepatocitos resecados u otras células del hígado, seguido de la infusión de los hepatocitos resucitados y transducidos de nuevo en la vasculatura portal, parénquima hepático o árbol biliar del paciente humano, ver, p. ej., Grossman et al. al., (1994) Nature Genetics, 6: 335-341,

[0151] La cantidad eficaz de nucleasa(s) y de donantes a ser administrada variará de paciente a paciente y de acuerdo con el polipéptido terapéutico de interés. Por consiguiente, las cantidades efectivas se determinan mejor por el médico que administra las composiciones y las dosis apropiadas pueden determinarse fácilmente por un experto en la técnica. Después de dejar suficiente tiempo para la integración y la expresión (típicamente 4-15 días, por ejemplo), el análisis del suero u otros niveles tisulares del polipéptido terapéutico y la comparación con el nivel inicial antes de la administración determinarán si la cantidad administrada es demasiado baja, dentro del rango correcto o demasiado alto. Los regímenes adecuados para las administraciones iniciales y posteriores también son variables, pero están tipificados por una administración inicial seguida de administraciones subsiguientes si es necesario. Las administraciones subsiguientes pueden administrarse a intervalos variables, desde diario hasta anualmente hasta cada varios años. Un experto en la materia apreciará que pueden recomendarse técnicas inmunosupresoras apropiadas para evitar la inhibición o el bloqueo de la integración por inmunosupresión de los vectores de administración, véase p. ej., Vilquin et al., (1995) Human Gene Ther., 6: 1391-1401,

[0152] Las formulaciones para administraciones tanto *ex vitro* como *in vivo* incluyen suspensiones en líquidos o líquidos emulsionados. Los ingredientes activos a menudo se mezclan con excipientes que son farmacéuticamente aceptables y compatibles con el ingrediente activo. Los excipientes adecuados incluyen, p. ej., agua, solución salina, dextrosa, glicerol, etanol o similares, y combinaciones de los mismos. Además, la composición puede contener cantidades menores de sustancias auxiliares, como agentes humectantes o emulsionantes, agentes de tamponamiento del pH, agentes estabilizantes u otros reactivos que mejoran la efectividad de la composición farmacéutica.

Células

[0153] También descritas en el presente documento están las células y/o líneas celulares en las que se modifica una secuencia de BCL11A potenciador endógeno. La modificación puede ser, p. ej., en comparación con la secuencia de tipo salvaje de la célula. La célula o las líneas celulares pueden ser heterocigotas u homocigotas para la modificación. Las modificaciones a la secuencia BCL11A pueden comprender inserciones, deleciones y/o combinaciones de las mismas.

[0154] De acuerdo con la divulgación del presente documento, la secuencia de promotor de BCL11A puede ser modificada por una nucleasa (p. ej., ZFN, TALEN, CRISPR/sistema Cas, sistema Ttgo, etc.), por ejemplo una nucleasa como se describe en el presente documento. En ciertas realizaciones de la divulgación, el potenciador BCL11A se modifica en cualquier lugar entre el exón 2 y el exón 3. En otras realizaciones de la divulgación, el potenciador BCL11A se modifica en las regiones mostradas en la SEQ ID NO:1, SEQ ID NO: 2 o SEQ. ID NO: 3 (Figura 11). La modificación es preferiblemente en o cerca de los sitios de unión y/o escisión de las nucleasas, p. ej., dentro de 1-300 (o cualquier valor entre ellos) pares de bases corriente arriba o corriente abajo del sitio o sitios de escisión, más preferiblemente dentro de 1-100 pares de bases (o cualquier valor entre ellos) de cualquiera de los lados de los sitios de unión y/o escisión, aún más preferiblemente dentro de 1 a 50 pares de bases (o cualquier valor entre ellos) en cualquier lado de la unión y/o sitio(s) de escisión. En ciertas realizaciones de la divulgación, la modificación es en o cerca de la región "+58" del potenciador BCL11A, p. ej., en o cerca de un sitio de unión a nucleasa mostrado en cualquiera de las SEQ ID NO: 4 a 80 y 362. En otras realizaciones de la divulgación, la modificación es en o cerca de la región "+55" del potenciador BCL11A, p. ej., en o cerca de un sitio de nucleasa mostrado en cualquiera de las SEQ ID NO: 143 a 184 y 232-251,

[0155] Cualquier célula o línea celular puede modificarse, p. ej., una célula madre, por ejemplo una célula madre embrionaria, una célula madre pluripotente inducida, una célula madre hematopoyética, una célula madre neuronal y una célula madre mesenquimática. Otros ejemplos no limitativos de células como se describe en el presente documento incluyen células T (p. ej., CD4+, CD3+, CD8+, etc.); células dendríticas; Células B. Un descendiente de una célula madre, incluyendo una célula diferenciada parcial o totalmente, también se proporciona (p. ej., Un RBC o célula precursora RBC). Ejemplos no limitantes de otras líneas celulares, incluyendo una secuencia BCL11A modificada, incluyen COS, CHO (p. ej., CHO-S, CHO-K1, CHO-DG44, CHO-DUXB11, CHO-DUKX, CHOK1SV), VERO, MDCK, WI38, V79, B14AF28-G3, BHK, HaK, NS0, SP2/0- Ag14, HeLa, HEK293 (p. ej., HEK293-F, HEK293-H, HEK293-T), y células perC6, así como células de insecto tales como *Spodoptera frugiperda* (Sf), o células fúngicas como *Saccharomyces*, *Pichia* y *Schizosaccharomyces*.

[0156] Las células como se describe en el presente documento son útiles para tratar y/o prevenir un trastorno, p. ej., por terapias *ex vitro*. Las células modificadas con nucleasa se pueden expandir y luego reintroducir en el paciente utilizando técnicas estándar. Ver, p. ej., Tebas et al (2014) New Eng J Med 370 (10): 901, En el caso de las células madre, después de la infusión en el sujeto, también se produce la diferenciación *in vivo* de estos precursores en células que expresan el transgén funcional. También se proporcionan composiciones farmacéuticas que comprenden las células como se describe en el presente documento. Además, las células pueden ser crioconservadas antes de la administración a un paciente.

[0157] Cualquiera de las células modificadas o líneas celulares descritas en el presente documento puede mostrar una expresión incrementada de globina gamma. También se proporcionan composiciones tales como composiciones farmacéuticas que comprenden las células modificadas genéticamente como se describe en el presente documento.

Aplicaciones

[0158] Los métodos y composiciones descritos en este documento son para modificar la expresión de proteínas, o la corrección de una secuencia de gen aberrante que codifica una proteína expresada en una enfermedad genética, tal como una enfermedad de células falciformes o una talasemia. Por lo tanto, los métodos y composiciones proporcionan el tratamiento y/o prevención de tales enfermedades genéticas. La edición del genoma, por ejemplo de células madre, se puede usar para corregir un gen aberrante, insertar un gen de tipo salvaje o cambiar la expresión de un gen endógeno. A modo de ejemplo no limitativo, un gen de tipo salvaje, *e. g.* que codifica al menos una globina (p. ej., globina α y/o β), puede insertarse en una célula (*p. ej.*, en una secuencia endógena potenciadora de BCL11a usando una o más nucleasas como se describe aquí) para proporcionar las proteínas globinas deficientes y/o carentes de la célula y por lo tanto tratar una enfermedad genética, p. ej., una hemoglobinopatía, causada por expresión de globina defectuosa. Alternativamente o además, la edición genómica con o sin la administración del donante apropiado, puede corregir el gen endógeno defectuoso, p. ej., corrigiendo la mutación puntual en la hemoglobina α o β , para restablecer la expresión del gen y/o tratar una enfermedad genética, *e. g.* enfermedad de células falciformes y/o knock out o alteración (sobreexpresión o represión) de cualquier gen regulador globina directa o indirecta (p. ej. inactivación del gen regulador de globina γ BCL11A o KLF1 regulador de BCL11A). Específicamente, los métodos y composiciones de la invención tienen uso en el tratamiento o prevención de hemoglobinopatías.

[0159] Las nucleasas de la invención están dirigidas a la región potenciadora BCL11A, conocida por ser necesaria para la expresión de BCL11A, y por lo tanto la regulación a la baja de la expresión de globina gamma. La modificación de esta región potenciadora puede dar lugar a eritrocitos con un aumento de la expresión de globina gamma, y por lo tanto puede ser útil para el tratamiento o la prevención de la enfermedad de células falciformes o la talasemia beta.

[0160] Los siguientes ejemplos se refieren a formas de realización ejemplares de la presente descripción en donde la nucleasa comprende una nucleasa de dedo de zinc (ZFN) o TALEN.

EJEMPLOS

Ejemplo 1: Ensamblaje de nucleasas de dedos de zinc y nucleasas TALEN

[0161] Los ZFN se ensamblaron contra el gen BCL11A humano y se ensayaron mediante ensayos CEL1 como se describe en Miller et al. (2007) Nat. Biotechnol. 25: 778-785. Los TALEN se ensamblaron como se describe en Miller et al (2011) Nature Biotechnology 29 (2): 143-151, Además, consulte las solicitudes de patente de EE.UU. 14/013.236 y la patente de EE.UU. N° 8.586.526. Los TALENs se ensamblaron con la arquitectura +63.

Ejemplo 2: Introducción de eliminaciones en las regiones potenciadoras BCL11A +55, +58 y +62

[0162] Para ensayar qué regiones de la región potenciadora BCL11A intrón 2 (Figura 1) se requerían para la represión de globina gamma durante la eritropoyesis, se hicieron una serie de TALENs para apuntar secciones de estas regiones (Figura 2). Los pares de TALEN se muestran a continuación en la Tabla 1, Los nucleótidos en el sitio diana que son contactados por la nucleasa se indican en letras mayúsculas; nucleótidos no contactados indicados en minúsculas.

Tabla 1: TALENs orientadas a la región potenciadora de BCL11A

Muestra	SBSn°	Diana 5'→3'	N° de RVD	SEQ ID NO:	Secuencia N→C RVD
55R	102740	ctACATAGAGGCCCTTCCTgc	17	362	NI-HD-NI-NG-NI-NN-NI-NN-NN-HD-HD-HD-NG-NG-HD-HD-NG
	102741	gtGGAGGGGATAACTGGGTca	17	4	NN-NN-NI-NN-NN-NN-NN-NI-NG-NI-NI-HD-NG-NN-NN-NN-NG
55M	102736	ttGTGTGCTTGGTCGGCActg	17	5	NN-NG-NN-NG-NN-HD-NG-NG-NN-NN-NG-HD-NN-NN-HD-NI-HD
	102737	gtGCCGACAACCTCCCTACCgc	17	6	NN-HD-HD-NN-NI-HD-NI-NI-HD-NG-HD-HD-HD-NG-NI-HD-HD
58R	102756	gtGCCGACAACCTCCCTACCgc	17	6	HD-NG-NG-NN-NN-NG-NN-NI-NG-NN-NN-NI-NN-NI-NI-NG-NG
	102757	atTATTTTCATTCCCATTGAg	17	7	NG-NI-NG-NG-NG-HD-NI-NG-NG-HD-HD-HD-NI-NG-NG-NN-NI
58M	102752	atAGGCCAGAAAAGAGATAtg	17	8	NI-NN-NN-HD-HD-NI-NN-NI-NI-NI-NI-NN-NI-NN-NI-NG-NI
	102753	ctGGTGTGTTATGTCTAAGag	17	9	NN-NN-NG-NN-NG-NN-NG-NG-NI-NG-NN-NG-HD-NG-NI-NI-NK

Muestra	SBSn°	Diana 5'→3'	N° de RVD	SEQ ID NO:	Secuencia N→C RVD
58L	102750	ctAGTTTATAGGGGGTTCTac	17	10	NI-NN-NG-NG-NG-NI-NG-NI-NN-NN-NN-NN-NG-NG-HD-NG
	102751	atAGCACCCAAGGTCCATCag	17	11	NI-NN-HD-NI-HD-HD-HD-NI-NI-NN-NN-NG-HD-HD-NI-NG-HD
62R	102775	atTCAACAAATAGCATATAaa	17	12	NG-HD-NI-NI-HD-NI-NI-NI-NG-NI-NN-HD-NI-NG-NI-NG-NI
	102774	ctTCCCTTTTAGGAAGGTAAa	17	13	NG-HD-HD-HD-NG-NG-NG-NG-NI-NN-NN-NI-NI-NN-NN-NG-NI
62L	102795	atGCCAGAGGGCAGCAAACat	17	14	NN-HD-HD-NI-NN-NI-NN-NN-NN-HD-NI-NN-HD-NI-NI-NI-HD
	102794	ctTAATAGCTGAAGGGGGCca	17	15	NG-NI-NI-NG-NI-NN-HD-NG-NN-NI-NI-NN-NN-NN-NN-HD

[0163] Las células K562 humanas se cultivaron en DMEM suplementado con 10% de FBS y 200.000 células fueron transfectadas con 800 ng de ADN de plásmido que codifica la TALENs por Amaxa Nucleofector® siguiendo las instrucciones del fabricante. El ensayo Cel-I (Surveyor™, Transgenomics) como se describe en Perez et al. (2008) Nat. Biotechnol. 26: 808-816 y Guschin et al. (2010) Methods Mol Biol. 649: 247-56), se usó para detectar modificaciones inducidas por TALEN del gen diana. En este ensayo, la amplificación por PCR del sitio diana fue seguida por la cuantificación de las inserciones y/o deleciones (indels) utilizando la enzima de detección de desajuste Cel-I (Yang et al. (2000) Biochemistry 39: 3533-3541) que proporcionó una estimación de límite inferior de la frecuencia DSB. La secuenciación profunda en la plataforma Illumina ("miSEQ") se usó de acuerdo con las instrucciones del fabricante para medir la eficiencia de edición, así como la naturaleza de los alelos generados por la edición. Para detectar las deleciones después del tratamiento celular con más de un par de nucleasas, se usó un ensayo basado en PCR en el que el ADN genómico en masa se amplifica con cebadores que flanquean la región que se eliminará, y se usa un gel para separar el producto de PCR derivado de alelo de tipo salvaje y el derivado del alelo que contiene deleción. Todos los diseños mostrados en la Tabla 1 estaban activos.

[0164] Tres días después de la transfección del vector de expresión TALEN en condiciones estándar (37°C), se aisló ADN genómico de células K562 usando el kit DNeasy (Qiagen) o QuickExtract (Epicentro) y se sometió a amplificación por PCR.

[0165] Los resultados del ensayo Cel-I demostraron que las TALENs fueron capaces de inducir la escisión en sus respectivos sitios diana.

[0166] Para ensayar el efecto sobre la expresión relativa de globina gamma, los ARNm que codifican los pares TALEN se introdujeron en células CD34+ (obtenidas a partir de voluntarios donante sano) por BTX nucleofección de acuerdo con las instrucciones del fabricante. Las células se diferenciaron entonces en eritrocitos. Brevemente, las células CD34+ se purificaron utilizando micropelotas Ficol-Paque (GE Healthcare) y CD34+ (MiltényiBiotec) de acuerdo con las instrucciones del fabricante. Las células CD34+ se cultivaron en MDM de Iscove con BIT 95000 (StemCell Technologies) en presencia de factores de crecimiento. Las células se diferenciaron hacia el linaje eritroide utilizando un modelo de cultivo líquido de 3 pasos. Durante los primeros 6 días (primera fase), las células CD34+ se expandieron con SCF (100 ng/ml), Flt3-L (100 ng/ml) e IL-3 (20 ng/ml). Las células expandidas se comprometieron y diferenciaron hacia el linaje eritroide (segunda fase) con Epo (2 U/ml) y SCF (50 ng/ml). Ver, Giarratana et al. (2011) Blood 118 (19): 5071-9.

[0167] Para analizar la expresión relativa de globina gamma, las relaciones de los ARNm que codifican globina gamma y globina beta después del tratamiento con TALEN se determinaron 14 días después de la introducción de TALEN mediante el análisis Taqman®. El análisis se realizó mediante un análisis estándar de Taqman®, siguiendo el protocolo y utilizando ensayos específicos de genes suministrados por el fabricante (Applied Biosystems) y los conjuntos de cebadores suministrados. Los niveles relativos de globina gamma se normalizaron por el nivel de expresión de globina alfa o beta donde la proporción se comparó con la relación alfa/beta o gamma/beta en células no tratadas. Los resultados (Figuras 3, 4 y 5) demostraron que las deleciones de regiones dentro del sitio hipersensible ADNasal BCL11A +58 y +55 dieron como resultado un aumento en los niveles relativos de expresión de globina gamma en estos experimentos.

Ejemplo 3: TALEN "atraviesa" el sitio hipersensible +58 ADNasa I en BCL11A

[0168] Para definir aún más el área requerida para la actividad potenciador en la región 58, se hicieron una serie de TALENs para crear una serie de DSBs y supresiones a través de este tramo de ADN. Los TALEN utilizados en este experimento se muestran a continuación en la Tabla 2.

Tabla 2: TALENs usados en la caminata del reforzador +58

Muestra	SBSn°	Diana 5'→3'	N° de RVD	SEQ ID NO	Secuencia N→C RVD
5	1	102830gtGTGCATAAGTAAGAGCAga	17	16	NN-NG-NN-HD-NI-NG-NI-NI- NN-NG-NI-NI-NN-NI-NN-HD- NI
		102831ctGTATGGACTTTGCACTGga	17	17	NN-NG-NI-NG-NN-NN-NI-HD- NG-NG-NG-NN-HD-NI-HD-NG- NK
10	2	102832gtAAGAGCAGATAGCTGATtc	17	18	NI-NI-NN-NI-NN-HD-NI-NN- NI-NG-NI-NN-HD-NG-NN-NI- NG
		102833atGTTATTACCTGTATGGAAct	17	19	NN-NG-NG-NI-NG-NG-NI-HD- HD-NG-NN-NG-NI-NG-NN-NN- NI
20	3	102834atAGCTGATTCCAGTGCAAag	17	20	NG-NG-HD-NG-NN-NN-HD- HD-NG-NI-NG-NN-NG-NG-NI- NG-NG
		102835ttTCTGGCCTATGTTATTac	17	21	NN-HD-NI-NI-NI-NN-NG-HD- HD-NI-NG-NI-HD-NI-NN-NN- NG
25	4	102836gtGCAAAGTCCATACAGGTaa	17	22	NN-HD-NI-NI-NI-NN-NG-HD- HD-NI-NG-NI-HD-NI-NN-NN- NG
		102837atGCCATATCTCTTTTCTGgc	17	23	NN-HD-HD-NI-NG-NI-NG-HD- NG-HD-NG-NG-NG-NG-HD- NG-NK
30	5	102838atACAGGTAATAACATAGGcc	17	24	NI-HD-NI-NN-NN-NG-NI-NI- NG-NI-NI-HD-NI-NG-NI-NN-NK
		102839ctAAGAGTAGATGCCATATct	17	25	NI-NI-NN-NI-NN-NG-NI-NN- NI-NG-NN-HD-HD-NI-NG-NI- NG
35	6	102840atAACATAGGCCAGAAAAGag	17	26	NI-NI-HD-NI-NG-NI-NN-NN- HD-HD-NI-NN-NI-NI-NI-NI-NK
		102841gtGTTATGTCTAAGAGTAGat	17	27	NN-NG-NG-NI-NG-NN-NG-HD- NG-NI-NI-NN-NI-NN-NG-NI- NK
40	7	102842ctCTTAGACATAACACACCag	17	28	HD-NG-NG-NI-NN-NI-HD-NI- NG-NI-NI-HD-NI-HD-NI-HD-HD
		102843ctAGACTAGCTTCAAAGTTgt	17	29	NI-NN-NI-HD-NG-NI-NN-HD- NG-NG-HD-NI-NI-NI-NN-NG- NG
45	8	102844atAACACACCAGGGTCAATac	17	30	NI-NI-HD-NI-HD-NI-HD-HD-NI- NN-NN-NN-NG-HD-NI-NI-NG
		102845gtTAGCTTGCACTAGACTAgc	17	31	NG-NI-NN-HD-NG-NG-NN-HD- NI-HD-NG-NI-NN-NI-HD-NG- NI
50	9	102846gtCAATACAACCTTTGAAGCta	17	32	HD-NI-NI-NG-NI-HD-NI-NI-HD- NG-NG-NG-NN-NI-NI-NN-HD
		102847atAAAAGCAACTGTTAGCtt	17	33	NI-NI-NI-NI-NN-HD-NI-NI-HD- NG-NN-NG-NG-NI-NN-HD
55	10	102848ttGAAGCTAGTCTAGTGCAag	17	34	NN-NI-NI-NN-HD-NG-NI-NN- NG-HD-NG-NI-NN-NG-NN-HD- NI
		102849ctGGAGCCTGTGATAAAAGca	17	35	NN-NN-NI-NN-HD-HD-NG-NN- NG-NN-NI-NG-NI-NI-NI-NI-NK

Muestra	SBSn°	Diana 5'-'>3'	N° de RVD	SEQ ID NO:	Secuencia N->C RVD	
5	11	102850ctAGTCTAGTGCAAGCTAac	17	36	NI-NN-NG-HD-NG-NI-NN-NG- NN-HD-NI-NI-NN-HD-NG-NI	
		102851ctTCCTGGAGCCTGTGATAaa	17	37	NG-HD-HD-NG-NN-NN-NI-NN- HD-HD-NG-NN-NG-NN-NI-NG- NI	
10	12	102852gtGCAAGCTAACAGTTGCTtt	17	38	NN-HD-NI-NI-NN-HD-NG-NI- NI-HD-NI-NN-NG-NG-NN-HD- NG	
		102853atCAGAGGCCAAACCCTTCct	17	39	HD-NI-NN-NI-NN-NN-HD-HD-	
15					NI-NI-NI-HD-HD-HD-NG-NG- HD	
20	13	102854ctAACAGTTGCTTTTATCAca	17	40	NI-NI-HD-NI-NN-NG-NG-NN- HD-NG-NG-NG-NG-NI-NG-HD- NI	
		102855ctAATCAGAGGCCAAACCCTt	17	41	NI-NI-NG-HD-NI-NN-NI-NN- NN-HD-HD-NI-NI-NI-HD-HD- HD	
25	14	102856atCACAGGCTCCAGGAAGGgt	17	42	HD-NI-HD-NI-NN-NN-HD-NG- HD-HD-NI-NN-NN-NI-NI-NN- NK	
		102857ctACCCACCCACGCCCCCCac	17	43	NI-HD-HD-HD-HD-NI-HD-HD- HD-NI-HD-NN-HD-HD-HD-HD- HD	
30	15	102858ctCCAGGAAGGGTTTGGCCtc	17	44	HD-HD-NI-NN-NN-NI-NI-NN- NN-NN-NG-NG-NG-NN-NN- HD-HD	
		102859ctACCCACCCACGCCCCCCac	17	45	NI-HD-HD-HD-HD-NI-HD-HD- HD-NI-HD-NN-HD-HD-HD-HD- HD	
40	16	102860ttGGCCTCTGATTAGGGTGgg	17	46	NN-NN-HD-HD-NG-HD-NG- NN-NI-NG-NG-NI-NN-NN-NN- NG-NK	
		102861ctGCCAGTCCTCTTCTACCcc	17	47	NN-HD-HD-NI-NN-NG-HD-HD- NG-HD-NG-NG-HD-NG-NI-HD- HD	
45	17	102862atTAGGGTGGGGGCGTGGGtg	17	48	NG-NI-NN-NN-NN-NG-NN- NN-NN-NN-NN-HD-NN-NG- NN-NN-NK	
		102863atGGAGAGGTCTGCCAGTCct	17	49	NN-NN-NI-NN-NI-NN-NN-NG- HD-NG-NN-HD-HD-NI-NN-NG- HD	
50	18	102864gtGGGGTAGAAGAGGACTGgc	17	50	NN-NN-NN-NN-NG-NI-NN-NI- NI-NN-NI-NN-NN-NI-HD-NG- NK	
		102865ctGGGCAAACGGCCACCGAtg	17	51	NN-NN-NN-HD-NI-NI-NI-HD- NN-NN-HD-HD-NI-HD-HD-NN- NI	
60	19	102866ctGGCAGACCTCTCCATCGgt	17	52	NN-NN-HD-NI-NN-NI-HD-HD- NG-HD-NG-HD-HD-NI-NG-HD- NK	
		102867ctTCCGAAAGAGGCCCCCTg	17	53	NG-HD-HD-NN-NI-NI-NI-NN- NI-NN-NN-HD-HD-HD-HD-HD- HD	

Muestra	SBSn°	Diana 5'→3'	Nº de RVD	SEQ ID NO:	Secuencia N→C RVD
20	102868	atCGGTGGCCGTTTGCCAg	17	54	HD-NN-NN-NG-NN-NN-HD- HD-NN-NG-NG-NG-HD
	102869	atCACCAAGAGAGCCTTCCga	17	55	HD-NI-HD-HD-NI-NI-NN-NI- NN-NI-NN-HD-HD-NG-NG-HD- HD
21	102870	gtTTGCCCAGGGGGGCCTCtt	17	56	NG-NG-NN-HD-HD-HD-NI-NN- NN-NN-NN-NN-NN-HD-HD- NG-HD
	102871	atTCTCCATCACCAAGAGAgc	17	57	NG-HD-NG-HD-HD-NI-NG-HD- NI-HD-HD-NI-NI-NN-NI-NN-NI
22	102872	ttGCCCAGGGGGGCCTCTTtc	17	58	NN-HD-HD-HD-NI-NN-NN-NN- NN-NN-NN-HD-HD-NG-HD- NG-NG
	102873	atAAAAATCCAATTCTCCATca	17	59	NI-NI-NI-NI-NG-HD-HD-NI-NI- NG-NG-HD-NG-HD-HD-NI-NG
23	102874	ctTTCGGAAGGCTCTCTTGgt	17	60	NG-NG-HD-NN-NN-NI-NI-NN- NN-HD-NG-HD-NG-HD-NG- NG-NK
25	102875	atTGAGAAATAAAATCCAAtt	17	61	NG-NN-NI-NN-NI-NI-NI-NG- NI-NI-NI-NI-NG-HD-HD-NI-NI
58R	102756	ctCTTGGTGATGGAGAATTgg	17	62	HD-NG-NG-NN-NN-NG-NN-NI- NG-NN-NN-NI-NN-NI-NI-NG- NG
	102757	atTATTTCAATCCCATTGAga	17	7	NG-NI-NG-NG-NG-HD-NI-NG- NG-HD-HD-HD-NI-NG-NG-NN- NI
58M	102752	atAGGCCAGAAAAGAGATAtg	17	8	NI-NN-NN-HD-HD-NI-NN-NI- NI-NI-NI-NN-NI-NN-NI-NG-NI
	102753	ctGGTGTGTTATGTCTAAGag	17	9	NN-NN-NG-NN-NG-NN-NG- NG-NI-NG-NN-NG-HD-NG-NI- NI-NK
58L	102750	ctAGTTTATAGGGGGTTCTac	17	10	NI-NN-NG-NG-NG-NI-NG-NI- NN-NN-NN-NN-NN-NG-NG- HD-NG
	102751	atAGCACCCAAGGTCCATCag	17	11	NI-NN-HD-NI-HD-HD-HD-NI- NI-NN-NN-NG-HD-HD-NI-NG- HD

[0169] En esta tabla, 'muestra' se refiere a las muestras mostradas en la Figura 6. Los resultados demuestran que el par TALEN 102853/102852 (indicado por la flecha en la figura) fue capaz de aumentar la expresión de gamma relativa. Además, grandes eliminaciones introducidas por algunos conjuntos de TALENs (Muestra 24: par de Muestra 22 + par de Muestra 6; Muestra 25: par de muestra 16 + par de muestra 6; Muestra 26: par de muestra 22 + par de muestra 16) también fueron capaces de aumentar la expresión gamma relativa. Los TALEN se diseñaron en este estudio para sondear toda la región +58 (ver la Figura 8 que muestra la secuencia del potenciador y los sitios de escisión de TALEN). Todos los diseños mostrados en la Tabla 2 fueron activos.

Ejemplo 4: ZFN dirigidos a la región potenciadora de +58 de BCL11A

[0170] En paralelo a los pares TALEN descritos en el Ejemplo 3, se hicieron pares ZFN para orientar la región 58. Los ZFN utilizados se muestran a continuación en la Tabla 3. Las nucleasas se identifican por su número "SBS", un identificador numérico único para cada proteína.

Tabla 3: Pares de ZFN específicos para la región potenciadora de +58 BCL11A

SBS nº (sitio diana, 5'-3')	Diseño					
	F1	F2	F3	F4	F5	F6
45796 atGGCTGAAA AGCGATACAG ggctggct (SEQ ID NO: 63)	RSDNLSE (SEQ ID NO: 81)	TRSPLRN (SEQ ID NO: 82)	RSDDLTR (SEQ ID NO: 83)	QKSNLSS (SEQ ID NO: 84)	QSAHRKN (SEQ ID NO: 85)	DSSHRTR (SEQ ID NO: 86)
45795 tcACTACAGA TAACTCCaa gtcctgtc (SEQ ID NO: 64)	DSSDRKK (SEQ ID NO: 87)	DRSNRTT (SEQ ID NO: 88)	TNSNRKR (SEQ ID NO: 89)	QSGDLTR (SEQ ID NO: 90)	LKDTLRR (SEQ ID NO: 91)	N/A
45802 caTAAGTAAG AGCAGATAGC tgattcca (SEQ ID NO: 65)	GYCCLRD (SEQ ID NO: 92)	TSGNLTR (SEQ ID NO: 93)	QSGDLTR (SEQ ID NO: 90)	QRTHLKA (SEQ ID NO: 94)	QSGALAR (SEQ ID NO: 95)	QSANRTK (SEQ ID NO: 96)
45800 caCCTGGGGC AtAGAGCCag ccctgtat (SEQ ID NO: 66)	DSSDRKK (SEQ ID NO: 87)	QNAHRKT (SEQ ID NO: 97)	QSGDLTR (SEQ ID NO: 90)	RSDHLSR (SEQ ID NO: 98)	QQWDRKQ (SEQ ID NO: 99)	N/A
45812 tcCATAACAGG TAATaACATA Ggccagaa (SEQ ID NO: 67)	RSDYLSK (SEQ ID NO: 100)	TSSVRTT (SEQ ID NO: 101)	TNQNLTV (SEQ ID NO: 102)	TSGHLSR (SEQ ID NO: 103)	RSADLTR (SEQ ID NO: 104)	TNQNRIT (SEQ ID NO: 105)
45813 acTTTGCACCT GGAAtcagct atctgctc (S EQ ID NO: 68)	QSGALAR (SEQ ID NO: 95)	RLDWLPM (SEQ ID NO: 106)	QSGDLTR (SEQ ID NO: 90)	HKWVLRQ (SEQ ID NO: 107)	N/A	N/A
45816 aaGTAAGAGC AGATAGCtga ttccagtg (S EQ ID NO: 69)	GYCCLRD (SEQ ID NO: 92)	TSGNLTR (SEQ ID NO: 93)	QSGDLTR (SEQ ID NO: 90)	QRTHLKA (SEQ ID NO: 94)	QSGALAR (SEQ ID NO: 95)	N/A
45815 caCACCTGGG GCATAGAGCC agccctgt (SEQ ID NO: 70)	DSSDRKK (SEQ ID NO: 87)	QNAHRKT (SEQ ID NO: 97)	LKQNLDA (SEQ ID NO: 108)	RSAHLSR (SEQ ID NO: 109)	RSDVLST (SEQ ID NO: 110)	DTRNLRA (SEQ ID NO: 111)

5	45844 tcACAGGCTC CAGGAAGGGT ttggcctc (SEQ ID NO:71)	LRHHLTR (SEQ ID NO:112)	RRDNLHS (SEQ ID NO: 113)	RSDHLSN (SEQ ID NO:114)	DSRSRIN (SEQ ID NO: 115)	DRSHLTR (SEQ ID NO: 116)	QSGTRKT (SEQ ID NO: 117)
10							
15	45843 aaGCAACTGT TAGCTTGAC tagactag (SEQ ID NO:72)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO: 119)	TGYNLTN (SEQ ID NO:120)	TSGSLTR (SEQ ID NO: 121)	QHQLVLR (SEQ ID NO: 122)	QNATRTK (SEQ ID NO: 123)
20	45849 caCAGGCTCC AGGAAGGGTT tggcctct (S EQ ID NO:73)	TSGSLSR (SEQ ID NO:124)	RSDHLTQ (SEQ ID NO: 125)	QSGHLAR (SEQ ID NO:126)	QKGTIGE (SEQ ID NO: 127)	QSSDLRS (SEQ ID NO: 128)	RRDNLHS (SEQ ID NO: 113)
25							
30	45848 aaAGCAACtG TTAGCTTGCA Ctagacta (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO: 119)	TGYNLTN (SEQ ID NO:120)	TSGSLTR (SEQ ID NO: 121)	DQSNLRA (SEQ ID NO: 118)	AQCCLFH (SEQ ID NO: 129)
35	45872 gtGGGGGCGT GGGTGGGGTA gaagagga (SEQ ID NO:75)	QSGALAR (SEQ ID NO:95)	RSDHLSR (SEQ ID NO: 98)	TSGHLSR (SEQ ID NO:103)	RSDALAR (SEQ ID NO: 130)	DRSHLTR (SEQ ID NO: 116)	RSDHLSR (SEQ ID NO: 98)
40	45871 ctAATCAGAG GCCAaaccct tcctggag (SEQ ID NO:76)	QSNLDSN (SEQ ID NO: 131)	RSHHLKA (SEQ ID NO: 132)	RSDNLSE (SEQ ID NO:81)	TSSNRKT (SEQ ID NO: 133)	N/A	N/A
45							
50	45881 tgGCCGTTtG CCCAGGGGGG Cctctttc (SEQ ID NO:77)	DRSHLTR (SEQ ID NO: 116)	RSDHLSR (SEQ ID NO: 98)	RSDNLSE (SEQ ID NO:81)	ASKTRKN (SEQ ID NO: 134)	TSGSLSR (SEQ ID NO : 124)	QWKSRRAR (SEQ ID NO: 135)
55	45880 cgATGGAGaG GTCTGCCAGT Cctctttc (SEQ ID NO:78)	DRSALSR (SEQ ID NO:136)	QSGDLTR (SEQ ID NO: 90)	RSDVLSE (SEQ ID NO:137)	TSGHLSR (SEQ ID NO: 103)	RSANLAR (SEQ ID NO: 138)	RSDALTQ (SEQ ID NO: 139)
60							
65	45889 ttGCCCAGGG GGGCctcttt cggaaggc (SEQ ID NO:79)	DRSHLTR (SEQ ID NO:116)	RSDHLSR (SEQ ID NO: 98)	RSDNLSE (SEQ ID NO:81)	ASKTRKN (SEQ ID NO: 134)	N/A	N/A

5	45888 ccACCGATGG AGAGGTCTgc cagtcctc (SEQ ID NO:80)	DRSALSR (SEQ ID NO:136)	RSDNLTR (SEQ ID NO: 140)	QSGHLSR (SEQ ID NO:141)	TSGNLTR (SEQ ID NO: 93)	DLTTLRK (SEQ ID NO : 142)	N/A
10	48117 aaAGCAACTG TTAGCTTGCA Ctagacta (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO: 119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO: 121)	DQSNLRA (SEQ ID NO: 118)	AQCCLFH (SEQ ID NO: 129)
15	48037 aaAGCAACTG TTAGCTTGCA Ctagacta (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO: 119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO: 121)	DQSNLRA (SEQ ID NO: 118)	AQCCLFH (SEQ ID NO: 129)
20							

[0171] Estos ZFNs se utilizaron para escindir la región potenciadora BCL11A 58 en las células CD34 y se analizaron las células transfectadas para la expresión gamma relativa después de la diferenciación de eritrocitos. Se identificó un par de ZFN, 45844/45843 que causó un aumento en la expresión gamma relativa en comparación con las células tratadas con un control de transducción de GFP, o las células tratadas con otros ZFN. La Figura 8 muestra que el sitio seleccionado por este par de ZFN se superpone parcialmente al apuntado por el par TALEN más activo que se describe en el Ejemplo 3. Una inspección más detallada de la secuencia que se escinde revela que contiene un sitio de consenso "GATA-1" (A/T GATA A/G), conocido por ser una de las secuencias unidas por el GATA1 y factores de transcripción relacionados. Ver, p. ej., Martin y Orkin (1990) Genes Dev 4:1886-1898; Fujiwara et al. (2009) Molecular Cell 36: 667-681; Tijssen et al. (2011) Developmental Cell 20: 597-609; May et al. (2013) Cell Stem Cell 13: 1-15. Todos los diseños mostrados en la Tabla 3 estaban activos.

[0172] Una región que comprende el sitio de escisión se amplificó mediante PCR, y después de la amplificación, el producto PCR se secuenció a través de análisis de secuenciación MiSeq de alto rendimiento de acuerdo con las instrucciones del fabricante (Illumina) tanto para los pares de ZFN 45843/45844 y TALEN 102852/102853. Los 5 genotipos más comunes se muestran a continuación en las Tablas 5a y 5b, muestran la escisión en o cerca de los sitios diana de la nucleasa y revelan una pérdida mediada por nucleasa de la secuencia de consenso GATA-1 (en el recuadro) en ambos casos (la Tabla 5a muestra la SEQ ID NO: 201-205 de arriba a abajo; la Tabla 5b muestra las SEQ ID NO: 206-210 de arriba a abajo). El par TALEN se escinde ligeramente hacia abajo de la secuencia de consenso, lo que potencialmente resulta en una menor incidencia de la eliminación de esta secuencia.

Tabla 5a: Análisis MiSeq de la región de eliminación para el par ZFN 45843/45844

Conteo	Longitud de Sec.	Longitud Vs	Amplicon	% general	Alineación
18 669	123	NA	0	53.565	ACACACCAGGGTCAATA- CAACTTTGAAGCTAGTCTAG TGCAAGCTAACAGTTGCTTT- TATCACAGGCTCCAG- GAAGG
2187	108	Más corta	-15	6.275	ACACACCAGGGTCAATA- CAACTTT- GAAGCTAGTCTAGT- GAAAGCTAACAG----- GCTCAAGGGAAGG
1415	110	Más corta	-13	4.060	ACACACCAGGGTCAATA- CAACTTT- GAAGCTAGTCTAGT- GCAAGCTAACAGTTGCT----- -----CCAGGGAAGG
1374	122	Más corta	-1	3.942	ACACACCAGGGTCAATA- CAACTTTGAAGCTAGTCTAG TGCAAGCTAACAGTTGCTTT- ATCACAGGCTCCGGGAAGG
818	118	Más corta	-5	2.347	ACACACCAGGGTCAATA- CAACTTT- GGAGCTAGTCTAGT- GCAAGCTAACAGTTGCC----- CACAGGCTCCAGGGAAGG

Tabla 5b: Análisis MiSeq de la región de eliminación para el par TALEN 102852/102853

Conteo	Longitud de Sec.	Longitud Vs	Amplicon	% general	Alineación
19238	123	NA	0	68.511	ACACACCCAGGGTCAATA- CAACTTTGAAGCTAGTCTAG TGCAAGCTAACAGTTGCTTT- TATCACAGCCTCCAG- GAAGG
1265	115	Más corta	-8	4.505	ACACACCCAGGGTCAATA- CAATTGAAGCTAGTCTAGT- GCAAGCTACCAGTTGCTTT- TATC-----CAGGAAGG
893	110	Más corta	-13	3.180	ACACACCCAGGGTCAATA- CAACTTT- GAAGCTAGTCTAGT- GCAAGCTACCAGTTGCT----- -----CCAGGAAGG
524	121	Más corta	-2	1.866	ACACACCCAGGGTCAATA- CAACTTTGAAGCTAGTCTAG TGCAAGCTAACAGTTGCTTT- TCATCA--GGCTCCAG- GAAGG
449	122	Más corta	-1	1.599	ACACACCCAGGGTCAATA- CAACTTTGAAGCTAGTCTAG TGCAAGCTAACAGTT- GCTTTTGTC-CAGGCTCCAG- GAAGG

Ejemplo 5: Caminata TALEN a través del sitio hipersensible de +55 ADNasal en la región del potenciador de BCL11A

[0173] Para refinar aún más el área requerida para actividad de potenciador en la región 55, se hicieron una serie de pares TALEN para crear una serie de mutaciones a través de este tramo de ADN. Los pares de TALEN hechos en este experimento se muestran a continuación en la Tabla 4.

Tabla 4: pares de TALEN que reconocen la región potenciadora +55 BCL11A

Muestra	SBS nº	Diana 5' ->3'	Nº de RVDs	SEQ ID NO:	N ->C Secuencia RVD	
1	102876	enAATGAATGTCCCAGGCCaa	17	143	NI-NI-NG-NN-NI-NI-NG-NN-NG-HD-HD-HD-NI-NN-NN-HD-HD	
	102877	ctGCCCCATACCCACTTCcc	16	144	NN-HD-HD-HD-HD-NI-NG-NI-HD-HD-HD-NI-HD-NG-NG-HD	
2	102878	atTCTAGGAAGGGAAGTGGgt	17	145	NG-HD-NG-NI-NN-NN-NI-NI-NN-NN-NN-NI-NI-NN-NG-NN-NK	
	102879	gtACCAGGAAGGCAATGGGct	17	146	NI-HD-HD-NI-NN-NN-NI-NI-NN-NN-HD-NI-NI-NG-NN-NN-NK	
3	102880	gtGGGTATGGGGCAGCCCAtt	17	147	NN-NN-NN-NG-NI-NG-NN-NN-NN-NN-HD-NI-NN-HD-HD-HD-NI	
	102881	atTGCATCATCCTGGTACca	16	148	NG-NN-HD-NI-NG-HD-NI-NG-HD-HD-NG-NN-HD	
4	102882	ctTCCTGGTACCAGGATGAtg	17	149	NG-HD-HD-NG-NN-NN-NG-NI-HD-HD-NI-NN-NN-NI-NG-NN-NI	
	102883	gtGGGGAGCTCACAGCCTCca	17	150	NN-NN-NN-NN-NI-NN-HD-NG-HD-NI-HD-NI-NN-HD-HD-NG-HD	

ES 2 721 936 T3

5	Muestra	SBS nº	Diana 5' ->3'	Nº de RVDs	SEQ ID NO:	N ->C Secuencia RVD	
10	5	102884	atGATGCAATGCTTGGAGGct	17	151	NN-NI-NG-NN-HD- NI-NI-NG-NN-HD- NG-NG-NN-NN-NI- NN-NK	
15		102885	gtGTGCCCTGAGAAGGTGGgg	17	152	NN-NG-NN-HD-HD- HD-NG-NN-NI-NN- NI-NI-NN-NN-NG- NN-NK	
20	6	102886	atGCTTGGAGGCTGTGAGCtc	17	153	NN-HD-NG-NG-NN- NN-NI-NN-NN-HD- NG-NN-NG-NN-NI- NN-HD	
25		102887	atCACAGGGTGTGCCCTGAga	17	154	HD-NI-HD-NI-NN-NN- NN-NG-NN-NG-NN- HD-HD-HD-NG-NN-NI	
30	7	102888	ctCCCCACCTTCTCAGGGCac	17	155	HD-HD-HD-HD-NI- HD-HD-NG-NG-HD- NG-HD-NI-NN-NN- NN-HD	
35		102889	ctGGACAGAGGGGTCCCAc	17	156	NN-NN-NI-HD-NI- NN-NI-NN-NN-NN- NN-NG-HD-HD-HD- NI-HD	
40	8	102890	ctTCTCAGGGCACACCCTGtg	17	157	NG-HD-NG-HD-NI- NN-NN-NN-HD-NI- HD-NI-HD-HD-HD- NG-NK	
45		102891	ctGGGCTGGACAGAGGGGTcc	17	158	NN-NN-NN-HD-NG- NN-NN-NI-HD-NI- NN-NI-NN-NN-NN- NN-NG	
50	9	102892	ctGTGATCTTGTGGGACCcc	16	159	NN-NG-NN-NI-NG- HD-NG-NG-NN-NG- NN-NN-NN-NI-HD- HD	
55		102893	atGCACACCCAGGCTGGGct	16	160	NN-HD-NI-HD-NI-HD- HD-HD-NI-NN-NN- HD-NG-NN-NN-NK	
60							

65

Muestra	SBS nº	objetivo 5' ->3'	nº de RVDs	SEQ NO:	ID	N ->C Secuencia RVD	
10	102894	ctTGTGGGACCCCTCTGTCCa	17	161		NG-NN-NG-NN-NN- NN-NI-HD-HD-HD- HD-NG-HD-NG-NN- NG-HD	
	102895	gtGCCGACCAAGCACACAga	17	162		NN-HD-HD-NN-NI- HD-HD-NI-NI-NN-HD- NI-HD-NI-HD-NI-NI	
11	102896	ctGTCCAGCCCAGCCTGGGtg	17	163		NN-NG-HD-HD-NI- NN-HD-HD-HD-NI- NN-HD-HD-NG-NN- NN-NK	
	102897	enCAGTGCCGACCAAGCACac	17	164		HD-NI-NN-NG-NN- HD-HD-NN-NI-HD- HD-NI-NI-NN-HD-NI- HD	
12	102898	ctGGGTGTGCATCTTGTGTgc	17	165		NN-NN-NN-NG-NN- NG-NN-HD-NI-NG- HD-NG-NG-NN-NG- NN-NG	
	102899	ctACCGCGACCCCTATCAGtg	17	166		NI-HD-HD-NN-HD- NN-NI-HD-HD-HD- HD-NG-NI-NG-HD-NI- NK	
13	102902	gtAGGGAGTTGTGGGCACa	17	167		NI-NN-NN-NN-NI- NN-NG-NG-NN-NG- HD-NN-NN-HD-NI- HD-NI	
	102903	ttGGGGACCGCTCACAGGA	17	168		NN-NN-NN-NN-NI- HD-HD-NN-HD-NG- HD-NI-HD-NI-NN-NN- NI	
14	102904	ctGCTGCATGTCCTGTGA	16	169		NN-HD-NG-NN-HD- NI-NG-NN-NG-HD- HD-NG-NN-NG-NN- NI	
	102905	ctGAAGGCTGGGCACAGCct	17	170		NN-NI-NI-NN-NN- HD-NG-NN-NN-NN- HD-NI-HD-NI-NN-HD- HD	

ES 2 721 936 T3

Muestra	SBS nº	objetivo 5' ->3'	nº de RVDs	SEQ ID NO:	N ->C Secuencia RVD
15	102906	gtCCCCAAGGCTGTGCCCAgc	17	171	HD-HD-HD-HD-NI-NI- NN-NN-HD-NG-NN- NG-NN-HD-HD-HD-NI
	102907	ctGTCAGAAGAGGCCCTGGac	17	172	NN-NG-HD-NI-NN- NI-NI-NN-NI-NN-NN- HD-HD-HD-NG-NN- NK
dieciséis	102912	ttCTGACAGGCCCTGCTGGtt	17	173	HD-NG-NN-NI-HD-NI- NN-NN-HD-HD-HD- NG-NN-HD-NG-NN- NK
	102913	gtGGTGCGTGGAGATAATGcc	17	174	NN-NN-NG-NN-HD- NN-NG-NN-NN-NI- NN-NI-NG-NI-NI-NG- NK
17	102914	ctGCTGGTTATCACTGTTGgc	17	175	NN-HD-NG-NN-NN- NG-NG-NI-NG-HD-NI- HD-NG-NN-NG-NG- NK
	102915	ctGGGCACAGAAGTGGTGcgt	17	176	NN-NN-NN-HD-NI- HD-NI-NN-NI-NI-NN- NG-NN-NN-NG-NN- HD
18	102916	ttGGCATTATCTCCACGCAcc	17	177	NN-NN-HD-NI-NG- NG-NI-NG-HD-NG- HD-HD-NI-HD-NN- HD-NI
	102917	gtGACCCAGCAGCCCTGGGca	17	178	NN-NI-HD-HD-HD- NI-NN-HD-NI-NN-HD- HD-HD-NG-NN-NN- NK
19	102918	atCTCCACGCACCACTTCTgt	17	179	HD-NG-HD-HD-NI- HD-NN-HD-NI-HD- HD-NI-HD-NG-NG- HD-NG
	102919	ctCCTTAAGGTGACCCAGCag	17	180	HD-HD-NG-NG-NI-NI- NN-NN-NG-NN-NI- HD-HD-HD-NI-NN-HD

Muestra	SBS nº	objetivo 5' ->3'	nº de RVDs	SEQ ID NO:	N ->C Secuencia RVD
20	102920	gtGCCCAGGGCTGCTGGGTca	17	181	NN-HD-HD-HD-NI- NN-NN-NN-HD-NG- NN-HD-NG-NN-NN- NN-NG
	102921	ctATGTAGACGGGTGTGTGgc	17	182	NI-NG-NN-NG-NI- NN-NI-HD-NN-NN- NN-NG-NN-NG-NN- NG-NK
21	102922	gtCACCTTAAGGAGCCACa	17	183	HD-NI-HD-HD-NG- NG-NI-NI-NN-NN-NI- NN-HD-HD-NI-HD-NI
	102923	gtCAGACCCAAGCAGGAagg	17	184	HD-NI-NN-NI-HD-HD- HD-HD-NI-NI-NN-HD- NI-NN-NN-NI-NI

[0174] Las TALENs se introdujeron en células CD34+ como se describe anteriormente y se indujeron las células para diferenciarse en células eritroides como se describe anteriormente. El análisis Taqman® se realizó como se describe y se identificaron varios sitios que causaron un aumento en la expresión gamma relativa (Figura 9). Los sitios de escisión se muestran en la Figura 10. Todos los diseños mostrados en la Tabla 4 fueron activos. Interesantemente, uno de los pares TALEN que provocó un aumento en los ARNm de globina gamma se escinde en otra secuencia de consenso GATA-1 (sitio de escisión 17 en la Figura 10).

Ejemplo 6: ZFNs dirigidas al sitio hipersensible +55 ADNasa I en BCL11A

[0175] Al igual que en el Ejemplo 4, se hizo un conjunto de ZFNs para sondear la región hipersensible +55 ADNasa I. Los ZFNs se muestran a continuación en la Tabla 6.

Tabla 6: ZFN específicos de región potenciadora +55: diseños y dianas

SBS nº (sitio de destino, 5'-3')	Diseño					
	F1	F2	F3	F4	F5	F6
46156 (tgGCCTGGG ACATTCATTA Tttagccac; SEQ ID NO: 232)	TSSNRKT (SEQ ID NO: 133)	AACNRNA (SEQ ID NO: 185)	WKCQLPI (SEQ ID NO: 186)	DRSNLTR (SEQ ID NO: 187)	RSDHLSQ (SEQ ID NO: 188)	DSSTRKK (SEQ ID NO: 189)
46158 (tcTAGGAAG GGAAGTGGGT Atggggcag; SEQ ID NO: 233)	QSGALAR (SEQ ID NO: 95)	RKYYLAK (SEQ ID NO: 190)	RSDNLSV (SEQ ID NO: 191)	RSAHLR (SEQ ID NO: 109)	QSGNLAR (SEQ ID NO: 192)	ARWSLGK (SEQ ID NO: 193)

SBS N° (sitio de destino 5'-3')	Diseño					
	F1	F2	F3	F4	F5	F6
46163 (taGAATTGG CTGGGACAT Tcattattt; SEQ ID NO:234)	WKCQLPI (SEQ ID NO: 186)	DRSNLTR (SEQ ID NO: 187)	RSDHLSQ (SEQ ID NO: 188)	DSSTRKK (SEQ ID NO: 189)	RPYTLRL (SEQ ID NO: 119)	QSGNLAR (SEQ ID NO: 192)
46164 (gaAGGGAAG TGGGTATGGG Gcagcccat; SEQ ID NO:235)	RSAHLR (SEQ ID NO: 109)	RSDALTQ (SEQ ID NO: 139)	TSGHLR (SEQ ID NO: 103)	RSDALAR (SEQ ID NO: 130)	QSGNLAR (SEQ ID NO: 192)	RQEHRVA (SEQ ID NO: 194)
46180 (tcATCCTGg TACCAGGAAG GCaatgggc; SEQ ID NO:236)	DRSHLTR (SEQ ID NO: 116)	QSGNLAR (SEQ ID NO: 192)	RSDSLA (SEQ ID NO: 195)	DNSNRIK (SEQ ID NO: 196)	RSDVLSE (SEQ ID NO: 137)	SPSSRRT (SEQ ID NO: 197)
46181 (gcAATGCTt GGAGGCTGTG AGctcccca; SEQ ID NO:237)	RSDNLAR (SEQ ID NO: 198)	WQSSLIV (SEQ ID NO: 199)	DRSHLTR (SEQ ID NO: 116)	QSGHLR (SEQ ID NO: 141)	QSSDLR (SEQ ID NO: 128)	LKWNLRT (SEQ ID NO: 200)
46188 (ccTGAGAAG GTGGGGAGCT Cacagcctc; SEQ ID NO:238)	PCRYRLD (SEQ ID NO: 201)	RSANLTR (SEQ ID NO: 202)	RSDHLR (SEQ ID NO: 98)	TSGHLR (SEQ ID NO: 103)	QSGNLAR (SEQ ID NO: 192)	QKPWRTP (SEQ ID NO: 203)
46189 (acACCCTGt GATCTTGIGG GAcccctct; SEQ ID NO:239)	QSSHLTR (SEQ ID NO: 204)	RSDALAR (SEQ ID NO: 130)	YRSSLKE (SEQ ID NO: 205)	TSGNLTR (SEQ ID NO: 93)	RSDTLA (SEQ ID NO: 206)	DKSTRTK (SEQ ID NO: 207)
46208 (ggGCTGGAc AGAGGGGTc cacaagatc; SEQ ID NO:240)	DRSALAR (SEQ ID NO: 208)	RSDHLR (SEQ ID NO: 98)	QGAHLGA (SEQ ID NO: 209)	QSSHLTR (SEQ ID NO: 204)	QSSDLTR (SEQ ID NO: 210)	N/A

SBS N° (sitio de destino 5'-3')	destino					
	F1	F2	F3	F4	F5	F6
46209 (gcCTGGGTG TGCACTTTGT Gtgcttggt; SEQ ID NO:241)	RSDSLR (SEQ ID NO: 211)	SASARWW (SEQ ID NO: 212)	TQSNLRM (SEQ ID NO: 213)	RNASRTR (SEQ ID NO: 214)	DRSHLTR (SEQ ID NO: 116)	RLDWLPM (SEQ ID NO: 106)
46216 (caGGCTGGG CTGGAcAGAG GGgtccac; SEQ ID NO:242)	RSDHLSR (SEQ ID NO: 98)	QGAHLGA (SEQ ID NO: 209)	QSSHLTR (SEQ ID NO: 204)	QSSDLTR (SEQ ID NO: 210)	RSDHLSQ (SEQ ID NO: 188)	DSSHRTTR (SEQ ID NO: 86)
46217 (gtGTGCATC TTGTGTGCTT GGtcggcac; SEQ ID NO: 243)	RSDHLSQ (SEQ ID NO: 188)	RRSDLKR (SEQ ID NO: 215)	RSDSLR (SEQ ID NO: 211)	SASARWW (SEQ ID NO: 212)	TQSNLRM (SEQ ID NO: 213)	RNASRTR (SEQ ID NO: 214)
46226 (gtGCCGACC AAGCACACAA Gatgcacac; SEQ ID NO:244)	RSDNLST (SEQ ID NO: 216)	DNSNRIN (SEQ ID NO: 217)	QSGDLTR (SEQ ID NO: 90)	QSGNLHV (SEQ ID NO: 218)	DRSDLSR (SEQ ID NO: 219)	DSSTRRR (SEQ ID NO: 220)
46228 (atAGGGGTc GCGGTAGGGA GTtgtcggc; SEQ ID NO:245)	LKQNLDA (SEQ ID NO: 108)	RSAHLSR (SEQ ID NO: 109)	QSGALAR (SEQ ID NO: 95)	RSDDLTR (SEQ ID NO: 83)	LKQNLDA (SEQ ID NO: 108)	RSHHLKA (SEQ ID NO: 132)
46229 (ccTATCAGt GCCGACCAAG Cacacaaga; SEQ ID NO:246)	QSGDLTR (SEQ ID NO: 90)	QSGNLHV (SEQ ID NO: 218)	DRSDLSR (SEQ ID NO: 219)	DSSTRRR (SEQ ID NO: 220)	RSDNLSE (SEQ ID NO: 81)	TSSNRKT (SEQ ID NO: 133)
46230 (tcGCGGTAg GGAGITGTCG GCacacact; SEQ ID NO:247)	DRSHLSR (SEQ ID NO: 221)	DRSALAR (SEQ ID NO: 208)	TSGSLSR (SEQ ID NO: 124)	QAGHLAK (SEQ ID NO: 222)	QSGALAR (SEQ ID NO: 95)	RSDDLTR (SEQ ID NO: 83)

5

10

15

20

25

30

35

SBS N° (sitio de destino 5'-3')	destino					
	F1	F2	F3	F4	F5	F6
46240 (ctCACAGGa CATGCAGCAG TGtgtgccg; SEQ ID NO:248)	RSDSLSV (SEQ ID NO: 223)	QSGDLTR (SEQ ID NO: 90)	QSGDLTR (SEQ ID NO: 90)	TSHNRNA (SEQ ID NO: 224)	RSDHLSQ (SEQ ID NO: 188)	DNSNRIN (SEQ ID NO: 217)
46241 (tcCCCAAGG CTGTGCCAg ccttcagtg; SEQ ID NO:249)	DRSNLSS (SEQ ID NO: 225)	RSHSLLR (SEQ ID NO: 226)	QSSDLSR (SEQ ID NO: 128)	RSDNLSV (SEQ ID NO: 191)	DNRDRIK (SEQ ID NO: 227)	N/A
46246 (gtGTGGCTc CTTAAGGTGA CCcagcagc; SEQ ID NO:250)	ASKTRTN (SEQ ID NO: 228)	RNASRTR (SEQ ID NO: 214)	RSDNLSV (SEQ ID NO: 191)	YSSTRNS (SEQ ID NO: 229)	QSSDLSR (SEQ ID NO: 128)	RSDALAR (SEQ ID NO: 130)
46247 (ccGTCTACA TAGAGgccct tcctgcttg; SEQ ID NO:251)	RSDNLAR (SEQ ID NO: 198)	QSTPRNT (SEQ ID NO: 230)	WPDYLPT (SEQ ID NO: 231)	DRSALAR (SEQ ID NO: 208)	N/A	N/A

40 **[0176]** Los ZFNs específicos a la región +55 se ensayaron para determinar la actividad de escisión usando el ensayo Cel-I y se encontró que eran activos en las células K562. Todos los diseños mostrados en la Tabla 6 fueron activos y la modificación del gen inducida por ZFN, descrita como %NHEJ (unión de extremo no homóloga) encontrada para cada par, se muestra a continuación en la Tabla 7,

45 **Tabla 7: Actividad de escisión de BCL11A +55 ZFN específica de la región potenciadora en células K562**

50

55

60

ZFN 1 (SBS N°)	ZFN 2 (SBS N°)	%NHEJ
46158	46156	30,36
46164	46163	21,90
46181	46180	15,21
46189	46188	12,76
46209	46208	27,96
46217	46216	16,63
46228	46226	37,79
46230	46229	37,23
46241	46240	24,97
46247	46246	17,30
Control de transducción GFP		0,00

65 **[0177]** Las células CD34+ fueron transfectadas con los pares ZFN como se describe para el Ejemplo 4, y luego se diferenciaron en eritrocitos como anteriormente. Los ZFN mostrados en 6 se unieron y modificaron la región

potenciadora BCL11A +55.

Ejemplo 7: Incremento de la actividad de pares de ZFN específicos +58

[0178] Los ZFNs dirigidos a la región +58 se refinaron adicionalmente por desplazamiento de las secuencias diana, alterando la identidad de dedo y el uso de enlazadores alternativos entre el ADN de dedo de zinc de dominio y los dominios de escisión de FokI de unión.

[0179] Se hicieron pares ZFN para apuntar a una secuencia muy cerca del sitio de escisión del par 45843/45844 ZFN. Los pares se muestran a continuación en la Tabla 8, y la ubicación de sus sitios de unión se muestra en la Figura 12A.

Tabla 8: ZFNs dirigidos a la región potenciadora +58

SBS N° (sitio de destino 5'-3')	destino						
	F1	F2	F3	F4	F5	F6	linker
46880 caCAGGCTC CAGGAAGGg tttggcctc t (SEQ ID NO:73)	RSDHLTQ (SEQ ID NO:125)	QSGHLAR (SEQ ID NO:126)	QKGTIGE (SEQ ID NO:127)	QSSDLSR (SEQ ID NO:128)	RRDNLHS (SEQ ID NO:113)	N/A	L7a
47923 caCAGGCTC CAGGAAGGg tttggcctc t (SEQ ID NO:73)	RSDHLTQ (SEQ ID NO:125)	QSGHLAR (SEQ ID NO:126)	QKGTIGE (SEQ ID NO:127)	QSSDLSR (SEQ ID NO:128)	RRDNLHS (SEQ ID NO:113)	N/A	L0
50679 caCAGGCTC CAGGAAGGg tttggcctc t (SEQ ID NO:73)	RSPHLTQ (SEQ ID NO:125)	QSGHLAR (SEQ ID NO:126)	QKGTIGE (SEQ ID NO:127)	QSSDLSR (SEQ ID NO:128)	RRDNLHS (SEQ ID NO:113)	N/A	L0[-1]
50680 caCAGGCTC CAGGAAGGg tttggcctc t (SEQ ID NO:73)	RSDHLTQ (SEQ ID NO:125)	QSGHLAR (SEQ ID NO:126)	QKGTIGE (SEQ ID NO:127)	QSSDLSR (SEQ ID NO:128)	RRDNLHS (SEQ ID NO:113)	N/A	L0[-2]
46923 aaGCAACTG TTAGCttGC ACTAgacta g (SEQ ID NO:72)	QKGTIGE (SEQ ID NO:127)	QSGSLTR (SEQ ID NO:252)	TGYNLTN (SEQ ID NO:120)	TSGSLTR (SEQ ID NO:121)	QHQLVLR (SEQ ID NO:122)	QNATRTK (SEQ ID NO:123)	L7e4
46999 caCAGGCTC CAGGAAGGg tttggcctc t (SEQ ID NO:73)	RSDHLTQ (SEQ ID NO:125)	QSGHLAR (SEQ ID NO:126)	QKGTIGE (SEQ ID NO:127)	QSSDLSR (SEQ ID NO:128)	RRDNLHS (SEQ ID NO:113)	N/A	L7c5

SBS N° (sitio de destino 5'-3')	destino						
	F1	F2	F3	F4	F5	F6	linker
45844							L7a
tcACAGGCT CCAGGAAGG GTttggcct c (SEQ ID NO:71)	LRHHLTR (SEQ ID NO:112)	RRDNLHS (SEQ ID NO:113)	RSDHLSN (SEQ ID NO:114)	DSRSRIN (SEQ ID NO:115)	DRSHLTR (SEQ ID NO:116)	QSGTRKT (SEQ ID NO:117)	
47021							L7c5
tcACAGGCT CCAGGAAGG GTttggcct c (SEQ ID NO:71)	LRHHLTR (SEQ ID NO:112)	RRDNLHS (SEQ ID NO:113)	RSDHLSN (SEQ ID NO:114)	DSRSRIN (SEQ ID NO:115)	DRSHLTR (SEQ ID NO:116)	QSGTRKT (SEQ ID NO:117)	
45843							L7a
aaGCAACTG TTAGCTTGC ACtagacta g (SEQ ID NO:72)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	TGYNLTN (SEQ ID NO:120)	TSGSLTR (SEQ ID NO:121)	QHQLVLR (SEQ ID NO:122)	QNATRTK (SEQ ID NO:123)	
46801							L7a
aaAGCAACT GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	
46786							L0
aaAGCAACT GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	
46934							L7c5
aaAGCAACT GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	
46816							L8c4
aaAGCAACT GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	

SBS N° (sitio de destino 5'-3')	destino						
	F1	F2	F3	F4	F5	F6	linker
50670 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO: 74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L0 [+9]
50671 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO: 74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L0 [+7]
50672 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO: 74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L0 [+5]
48117 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO: 74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L7c5
50674 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO: 74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L0 [+11]
50676 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO: 74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L0 [+7]
48037 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO: 74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO: 129)	L7a

SBS N° (sitio de destino 5'-3')	destino						linker
	F1	F2	F3	F4	F5	F6	
45844 tcACAGGCT CCAGGAAGG GTttggcct c (SEQ ID NO:71)	LRHHLTR (SEQ ID NO:112)	RRDNLHS (SEQ ID NO:113)	RSDHLSN (SEQ ID NO:114)	DSRSRIN (SEQ ID NO:115)	DRSHLTR (SEQ ID NO:116)	QSGTRKT (SEQ ID NO:117)	L7a
47021 tcACAGGCT CCAGGAAGG GTttggcct c (SEQ ID NO:71)	LRHHLTR (SEQ ID NO:112)	RRDNLHS (SEQ ID NO:113)	RSDHLSN (SEQ ID NO:114)	DSRSRIN (SEQ ID NO:115)	DRSHLTR (SEQ ID NO:116)	QSGTRKT (SEQ ID NO:117)	L7c5
45843 aaGCAACTG TTAGCTTGC ACTagacta g (SEQ ID NO:72)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	TGYNLTN (SEQ ID NO:121) SBS N° (sitio de destino 5'-3')	TSGSLTR (SEQ ID NO:122)	QHQLVLR (SEQ ID NO:123)	QNATRTK (SEQ ID NO:124)	L7a
46801 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L7a
46786 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L0
46934 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L7c5
46816 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L8c4

5	SBS N° (sitio de destino 5'-3')	destino						
		F1	F2	F3	F4	F5	F6	linker
10	50670 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L0 [+9]
15	50671 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L0 [+7]
20	50672 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L0 [+5]
25	48117 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L7c5
30	50674 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L0 [+11]
35	50676 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L0 [+7]
40	48037 aaAGCAAct GTTAGCTTG CACTagact a (SEQ ID NO:74)	DQSNLRA (SEQ ID NO:118)	RPYTLRL (SEQ ID NO:119)	SGYNLEN (SEQ ID NO:253)	TSGSLTR (SEQ ID NO:121)	DQSNLRA (SEQ ID NO:118)	AQCCLFH (SEQ ID NO:129)	L7a
45								
50								

[0180] Como puede verse en la Figura 12, el sitio de unión de los nuevos pares se encuentra un par de bases más cerca del centro de la secuencia de consenso GATA-1. Todos los pares fueron activos para unir y dividir sus dianas en el genoma según lo medido por una pantalla inicial en células K562. Los ARNm que codifican las ZFN se sometieron a electroporación en células CD34+ y luego las células se diferenciaron en el linaje eritroide como se describe en el Ejemplo 2. Para analizar la expresión relativa de la globina gamma, se determinaron las proporciones de los ARNm que codifican globina gamma y globina beta después del tratamiento con ZFN mediante análisis Taqman® a los 14 días después de la introducción de ZFN. Los resultados (Figura 12B) demostraron que los pares de ZFN dirigidos a los sitios de unión desplazados tuvieron una mayor influencia en la expresión de la globina gamma.

[0181] A continuación, las proteínas se hicieron con un tipo de enlazadores alternativos para poner a prueba el efecto sobre las proteínas Bc111a. Se hicieron conjuntos similares de ZFN que comprendían las mismas hélices en los mismos dedos pero donde cada uno contenía diferentes enlazadores entre el dominio de unión al ADN de ZFP y

la nucleasa FokI. P. ej., ZFNs 46801, 46786, 46816 y 46934 tienen el mismo dominio de unión a ADN de ZFP, pero están vinculados al dominio de nucleasa usando los enlazadores L7a, L0, L8c4 y L7C5 respectivamente. De manera similar, los ZFN 45844 y 47021 tienen el mismo dominio de unión a ADN, pero 45844 tiene el enlazador L7a, mientras que 47021 usa el enlazador L7C5. Además, 46880, 47923, 50679 y 50680 tienen los mismos dominios de unión al ADN, pero 46880 usa el enlazador L7a; 47923 tiene el enlazador L7C5; 50679 usa L0 [-1] y 50680 tiene L0 [-3]. Los enlazadores se muestran en la Figura 14 y la Figura 17.

[0182] Como se muestra en la Figura 13, los ZFN se ensayaron en varios pares donde un conjunto de pares apuntó al sitio de unión tipificado por el apuntado por el par 45843/45844, y luego un segundo conjunto tipificado por el apuntado por 46801/46880, Los pares probados se muestran a continuación en la Tabla 9 como sigue junto con el porcentaje de escisión (NHEJ) medido por análisis de secuencia en el Día 0 (D0) o el Día 14 (D14). Cada punto de datos es una media de tres repeticiones.

Tabla 9: Efecto del enlazador sobre la actividad de escisión de ZFN

Par	Enlazadores	Tipo de diana	NHEJ, D0	NHEJ, D14	Relación HGB/HBA
45843/45844	L7a/L7a	45843/45844	78%	73%	7,65
45843/47021	L7a/L7c5	45843/45844	82%	72%	7,57
46801/46880	L7a/L7a	46801/46880	87%	64%	8,33
46801/47923	L7a/L0	46801/46880	82%	65%	12,07
46786/46880	L0/L7a	46801/46880	71%	40%	6,06
46934/47923	L7c5/L0	46801/46880	80%	58%	8,3
GFP					2,71

[0183] Pruebas adicionales fueron diseñadas para medir el número de indel que contiene las modificaciones que destruyeron el sitio GATA en la diana. En la Tabla 10 a continuación se muestran combinaciones de ZFN con varios enlazadores y el efecto que tienen los enlazadores en el aumento del porcentaje de los indeles en general y el aumento de los indeles que resultan en la pérdida del sitio de enlace GATA.

Tabla 10: Actividad del enlazador ejemplar

ZFN izquierdo		ZFN derecho		% indel, no queda GATA	% total_indels	% de no GATA/indel total
SBS n°	Tipo de vinculator	SBS n°	Tipo de vinculator			
50670	L0 [+9]	47923	L0	28,0	30,8	0,91
50671	L0 [+7]			24,5	26,7	0,92
50672	L0 [+5]			30,3	33,0	0,92
46801	L7a			27,6	30,4	0,91
46816	L8c4			27,7	29,9	0,93
46934	L7C5			43,8	47,8	0,92
50670	L0 [+9]	50679	L0 [-1]	30,1	32,6	0,92
50671	L0 [+7]			23,3	25,3	0,92
50672	L0 [+5]			31,1	34,2	0,91
46801	L7a			16,1	18,0	0,89
46816	L8c4			33,8	36,2	0,93
46934	L7C5			23,7	25,4	0,94
50670	L0 [+9]	50680	L0 [-2]	16,9	18,1	0,94
50671	L0 [+7]			16,9	17,9	0,94
50672	L0 [+5]			23,2	24,9	0,93
46801	L7a			24,2	26,0	0,93
46816	L8c4			21,3	23,2	0,92
46934	L7C5			25,1	27,0	0,93
46801		47923		16,5	19,1	0,86
		GFP		1,4	0,8	1,64

[0184] Como se puede ver en la tabla anterior, el refinado del par original, 46801 (L7a)/47923 (L0), cuya actividad se midió en este experimento fue de 19% de formación de indeles en general, con un 86% de esos indeles medidos que tienen un sitio de unión GATA destruido, puede llevar a un aumento general en la actividad de escisión (indel), y un aumento general en el porcentaje de indels que conducen a la destrucción de GATA. Ver, p. ej., 46934 (L7C₅)/47923 (L0) donde se observaron 47,8% de indels totales y 92% de esos indels tenían un sitio GATA destruido. Estos y otros enlazadores descritos (ver Figura 17) pueden incorporarse en los ZFN para aumentar y/o refinar la actividad.

[0185] Por lo tanto, los enlazadores alternativos se pueden usar con los pares ZFN descritos en este documento para escindir BCL11A y aumentar en hemoglobina de gamma con respecto a la hemoglobina alfa.

Ejemplo 8: Administración *in vivo*

[0186] Las composiciones que incluyen células (p. ej., HSC y/o células precursoras de RBC), proteínas (p. ej., nucleasas) y/o polinucleótidos (p. ej., codificando nucleasas) como se describe en este documento se administran a un sujeto, p. ej., un sujeto con una hemoglobinopatía, esencialmente como se describe en las patentes de EE.UU. N^{os} 7.837.668; 8.092.429; Publicación de Patente de Estados Unidos N^o 20060239966; las patentes de EE.UU. n^o 6.180.613; 6.503.888 y/o las patentes de EE.UU. N^{os} 6.998.118 y 7.101.540 para proporcionar terapia a un sujeto que la necesite.

[0187] Además, las células se estudian para su uso en la producción a gran escala de LT-HSC editadas. Las células CD34⁺ a granel se estimulan previamente con citoquinas que comprenden StemspanTM CC110, ligando de Flt-3, SCF y TPO y todas sus combinaciones en concentraciones de 10 ng/mL a 1.000 ng/mL. La estimulación previa puede requerir tiempos de exposición de 24, hasta 48 y hasta 72 horas. Para la transfección de HSPC a escala clínica, se puede usar cualquier sistema de alta capacidad (p. ej., el sistema de transfección Maxcyte GT Flow).

[0188] Para terapias *ex vitro*, las células editadas (p. ej., HSCs) se someten a ensayos de formación de colonias en medio de metilcelulosa para confirmar la frecuencia de células pluripotentes y para verificar que las colonias poseen la edición genética deseada en las frecuencias esperadas. Los estudios de metilcelulosa se llevan a cabo utilizando métodos conocidos en la técnica (ver, p. ej., Keller *et al* (1993) *Mol Cell Bio* 13 (1): 473).

[0189] Para garantizar aún más la capacidad de injerto de las células editadas con BCL11a (p. ej., HSC), las células se injertan en un modelo de ratón relevante y/o un modelo de primate no humano. El injerto en estos animales se realiza de acuerdo con los métodos conocidos en la técnica. Ver, p. ej., Holt *et al.* (2010) *Nat Biotech* 28, 839-847; Ho *et al* (2009) *Retrovirology* 6:65 y Peterson *et al* (2014) *J. Med Primatol* 42: 237. El injerto con (1020 cGy irradiación) o sin (200 cGy irradiation) precondicionamiento mieloablativo se utiliza para investigar las condiciones óptimas de injerto y expansión para el trasplante de células madre.

Ejemplo 9: Administración *in vivo* y injerto

[0190] Como se describió anteriormente, las células humanas CD34⁺ fueron tratadas con ARNm que codifican las ZFNs específicas +55 potenciadoras y luego se implantaron en ratones GSN. Las células CD34⁺ se obtuvieron de voluntarios humanos sanos. En algunos casos, se realizaron estrategias de movilización de CD34⁺, utilizando G-CSF (Neupogen®) o G-CSF + Plerixafor (Mozobil®) antes de la aféresis. El G-CSF se administró diariamente durante los cuatro días anteriores a la aféresis de acuerdo con las instrucciones del fabricante, y si se usó Plerixafor, se administró la última noche antes de la cosecha, nuevamente de acuerdo con las instrucciones del fabricante. La aféresis se realizó mediante métodos estándar. Las células CD34⁺ se enriquecieron a partir de los leukopaks de PBMC movilizados utilizando un sistema Miltenyi CliniMACs mediante métodos estándar y de acuerdo con las instrucciones del fabricante.

[0191] Se sintetizaron ARNm recubiertos y poli-adenilados que codifican las ZFN usando el kit de Ambion mMessage mMachine® T7 ultra según las instrucciones del fabricante y luego se sometieron a electroporación en las células CD34⁺ usando un sistema Maxcyte GT o un electroporador BTX ECM830, ambos de acuerdo con las instrucciones del fabricante.

[0192] Se usaron ratones NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ para recibir el trasplante CD34⁺. Un día (16-24 horas) antes de la implantación, los ratones se sometieron a irradiación subletal (300 RAD). Las células CD34⁺ tratadas con ZFN de arriba se trasplantaron en ratones irradiados a través de una inyección en la vena de la cola, donde se administraron 1 millón de células en 0,5 ml de PBS-BSA al 0,1% por ratón.

[0193] Para este experimento, las células CD34⁺ fueron electroporadas con ARNm que codifican ya sea el par 45843/45844 (electroporación n^o 1) o el par 46801/46880 (electroporación n^o 2). En ambos casos, se utilizó GFP como control. Después del trasplante en los ratones, se tomaron muestras a las 4 o 16 semanas después del trasplante para observar el nivel de marcaje en las células. En la semana 4, hasta aproximadamente el 4% de las células en la sangre periférica eran células humanas de ambas electroporaciones (consulte la Figura 15A). La

edición del genoma (indels) en estas células fue de alrededor del 30-40% (Figura 15B).

[0194] En la semana 16 después de trasplante, el nivel de edición se midió en pan-mieloide humano, células B, células madre y eritroides en los ratones. En estos experimentos, las células de ambas electroporaciones se agruparon. Los datos (Figura 16) indicaron que se detectó un 40-50% de edición del gen en todas las poblaciones de células humanas analizadas. Este experimento demuestra que las células CD34+ trasplantadas se mantienen y diferencian mientras que se mantiene la edición del gen en el locus BCL11A.

5
10
15
20
25
30
35
40
45
50
55
60
65

REIVINDICACIONES

1. Una célula modificada genéticamente que comprende una modificación genómica hecha por una nucleasa de dedo de cinc (ZFN) o una nucleasa de dominio efector TAL (TALEN), en donde la modificación genómica está dentro de cualquiera de las secuencias diana de nucleasas de las SEQ ID NO: 4-80, 143-184, 232-251 y 276 de una secuencia potenciadora BCL11a endógena, y además en la que la modificación genómica es una inserción y/o eliminación.
2. La célula modificada genéticamente de la reivindicación 1, en la que la célula es una célula madre, opcionalmente en donde la célula madre es una célula madre hematopoyética tal como una célula madre hematopoyética CD34+.
3. Una célula diferenciada genéticamente modificada descendiente de la célula madre de la reivindicación 2, opcionalmente en la que la célula es un glóbulo rojo (RBC).
4. La célula modificada genéticamente de cualquiera de las reivindicaciones 1 a 3, en la que:
 - (a) la nucleasa se introduce en la célula como un polinucleótido; y/o
 - (b) la inserción comprende la integración de un polinucleótido donante que codifica un transgén, opcionalmente en donde:
 - (i) el ZFN comprende una proteína de dedo de zinc, la proteína de dedo de zinc que comprende 4, 5 o 6 dominios de dedo de zinc que comprenden una región de hélice de reconocimiento y, además, en donde la proteína de dedo de zinc comprende las regiones de hélice de reconocimiento en el orden mostrado en una sola fila de la Tabla 3 o la Tabla 6; o
 - (ii) el TALEN comprende una proteína TALE, comprendiendo la proteína TALE una pluralidad de unidades de repetición TALE, comprendiendo cada unidad de repetición una región diresídica hipervariable (RVD), en donde los RVD de las unidades de repetición TALE se muestran en una sola fila de la Tabla 1, Tabla 2 o Tabla 4.
5. Una composición farmacéutica que comprende la célula modificada genéticamente de cualquiera de las reivindicaciones 1 a 4.
6. Una proteína de unión al ADN que comprende una proteína de dedo de zinc o una proteína efectora de TALE (TALE), en donde
 - (i) la proteína de dedo de zinc comprende 4, 5 o 6 dominios de dedo de zinc que comprenden una región de hélice de reconocimiento, en la que las proteínas de dedo de zinc comprenden las regiones de hélice de reconocimiento en el orden mostrado en una sola fila de la Tabla 3 o la Tabla 6; y
 - (ii) la proteína TALE comprende una pluralidad de unidades de repetición TALE, comprendiendo cada unidad de repetición una región diresidútica hipervariable (RVD), en la que los RVD de las unidades de repetición TALE se muestran en una sola fila de la Tabla 1, Tabla 2 o Tabla 4.
7. Una nucleasa que comprende una proteína de dedo de zinc o una proteína TALE de la reivindicación 6 y un dominio de escisión de tipo salvaje o de ingeniería o un semi-dominio de escisión.
8. Un polinucleótido que codifica una o más proteínas de la reivindicación 6 o la nucleasa según la reivindicación 7.
9. Una célula aislada que comprende una o más proteínas de la reivindicación 6 o la reivindicación 7 o uno o más polinucleótidos de la reivindicación 8, opcionalmente, en donde la célula es una célula madre hematopoyética.
10. Un kit que comprende al menos uno de: i) un polinucleótido que codifica la proteína de acuerdo con las reivindicaciones 6 o 7, o ii) una proteína de las reivindicaciones 6 o 7.
11. Un método *in vitro* para alterar la expresión del gen de globina en una célula, comprendiendo el método: introducir, en la célula, uno o más polinucleótidos de la reivindicación 8, en condiciones tales que la una o más proteínas se expresan y la expresión del gen de globina se altera, opcionalmente, en donde la expresión del gen de globina aumenta y/o el gen de globina es un gen de globina gamma o globina beta.
12. El método de la reivindicación 11, que comprende además integrar una secuencia donante en el genoma de la célula, opcionalmente en la que:
 - (a) la secuencia del donante se introduce en la célula utilizando un vector viral, como un oligonucleótido o en un plásmido;
 - (b) la célula se selecciona del grupo que consiste en una célula precursora de glóbulos rojos (eritrocitos) y una célula madre hematopoyética; y/o
 - (c) la secuencia del donante comprende un transgén bajo el control de un promotor endógeno o exógeno.

13. La composición farmacéutica de la reivindicación 5 para uso en un método para tratar o prevenir una hemoglobinopatía en un paciente que necesita un aumento en la expresión génica de globina, comprendiendo el método administrar al paciente la composición farmacéutica en una cantidad suficiente para aumentar la expresión génica de globina en el paciente, opcionalmente en donde la hemoglobinopatía es una talasemia o enfermedad de células falciformes, como la β -talasemia.

14. El método de la reivindicación 11 o la composición farmacéutica para el uso de acuerdo con la reivindicación 13, en donde la expresión génica de globina es el nivel de ARNm de globina o proteína de globina producida.

15. Una o más nucleasas de acuerdo con la reivindicación 7, o un ácido nucleotídico codificando una o más nucleasas que se unen a una secuencia diana de cualquiera de SEQ ID NOs: 4-80, 143-184, 232-251 y 276, para uso en el tratamiento o la prevención de una hemoglobinopatía, en donde la nucleasa se dirige a una secuencia potenciadora BCL11A en una célula, en donde una o más nucleasas es una ZFN o TALEN

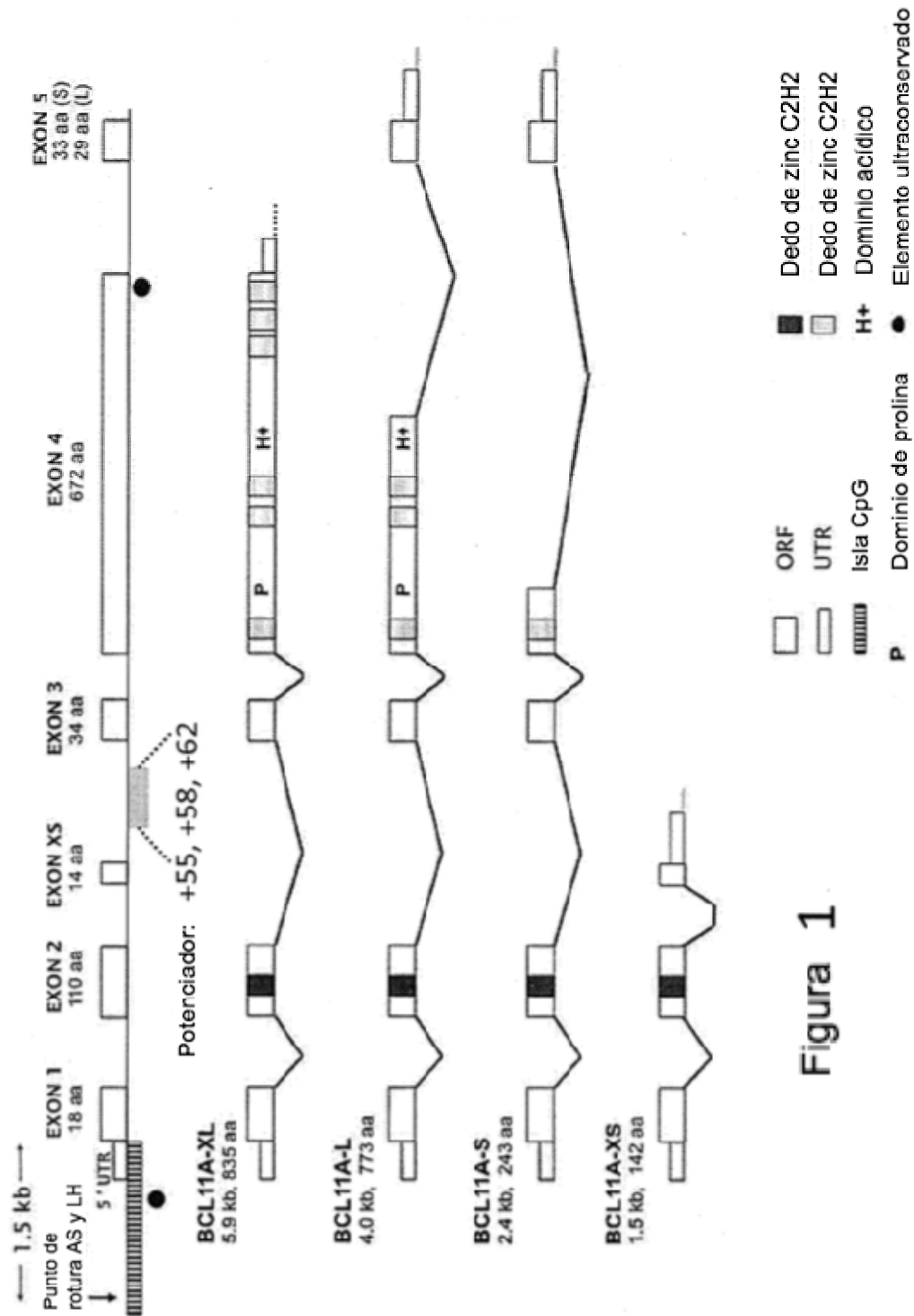


Figura 1

chr2:60,671,288-60,783,031:

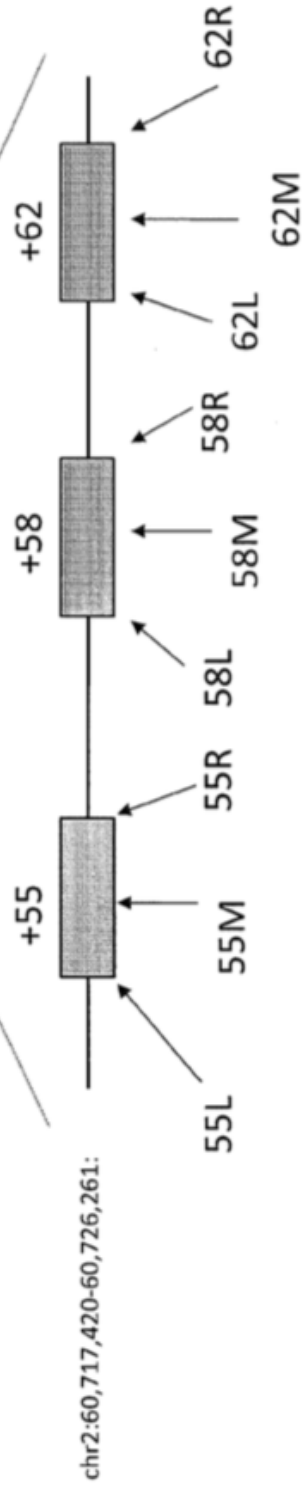
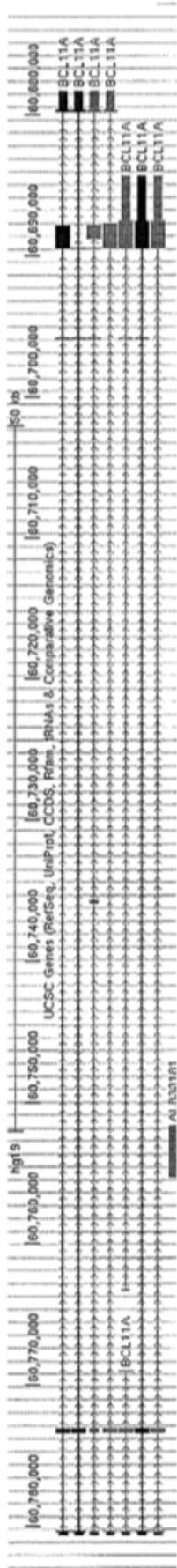
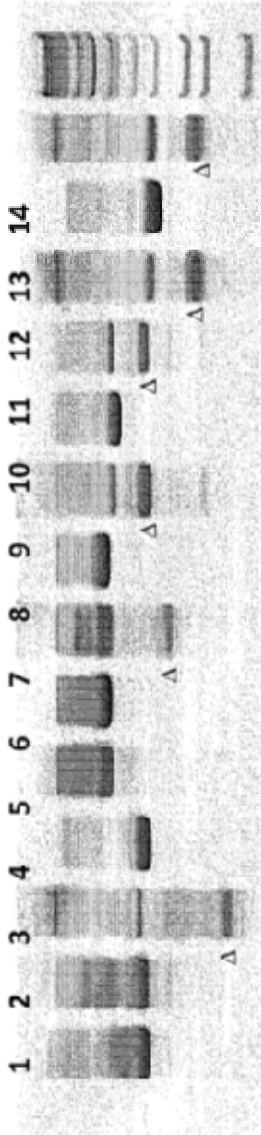


Figura 2

PCT/US

Carril	ARNm1 (SBS#)	ARNm2 (SBS#)	ARNm3 (SBS#)	ARNm4 (SBS#)	Tamaño de deleción (bp)
	55L		55R		
1	102783	102782	-	-	
2	-	-	102740	102741	
3	102783	102782	102740	102741	487
4	Control de transducción GFP				
	62L		62R		
5	102795	102794	-	-	
6	-	-	102775	102774	
7	102795	102794	102775	102774	505
8	Control de transducción GFP				
	58L		58R		
9	102750	102751	102756	102757	419
10	Control de transducción GFP				
11	102750	102751	102756	102757	
	58M		58R		
12	102752	102753	102756	102757	219
13	Control de transducción GFP				
14	102752	102753	102756	102757	

Figura 3



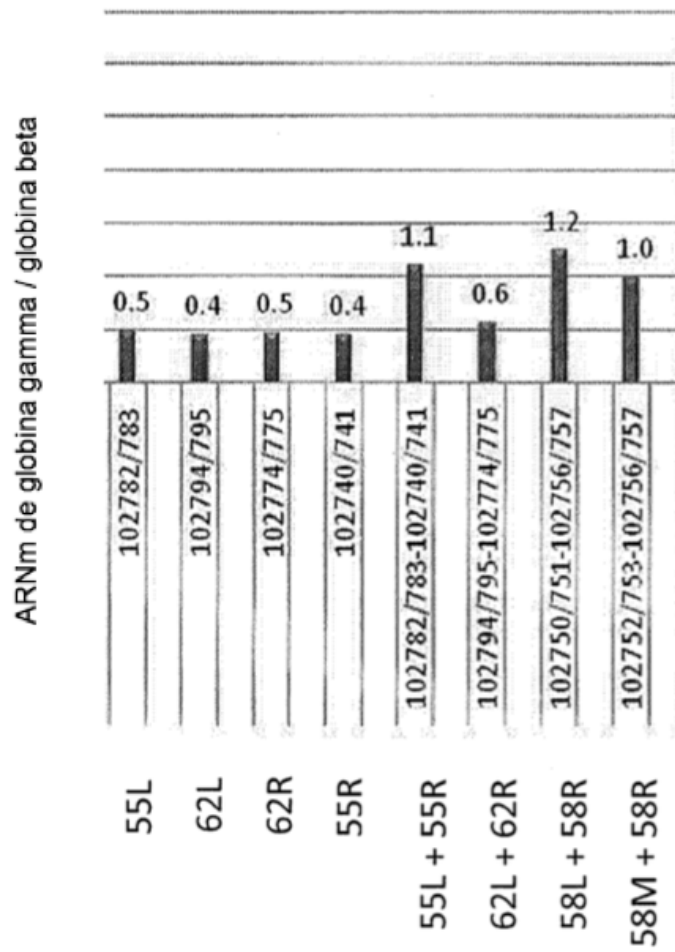


Figura 4

Figura 5

ARNm de globina gamma / globina alfa

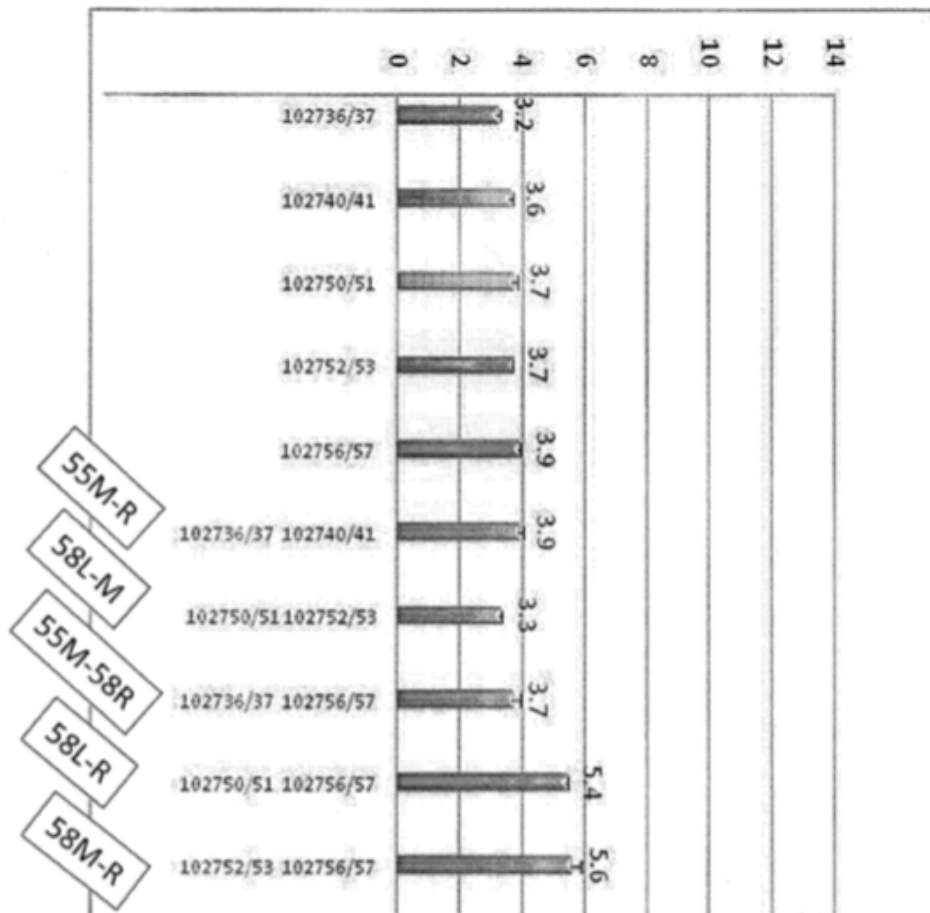


Figura 6

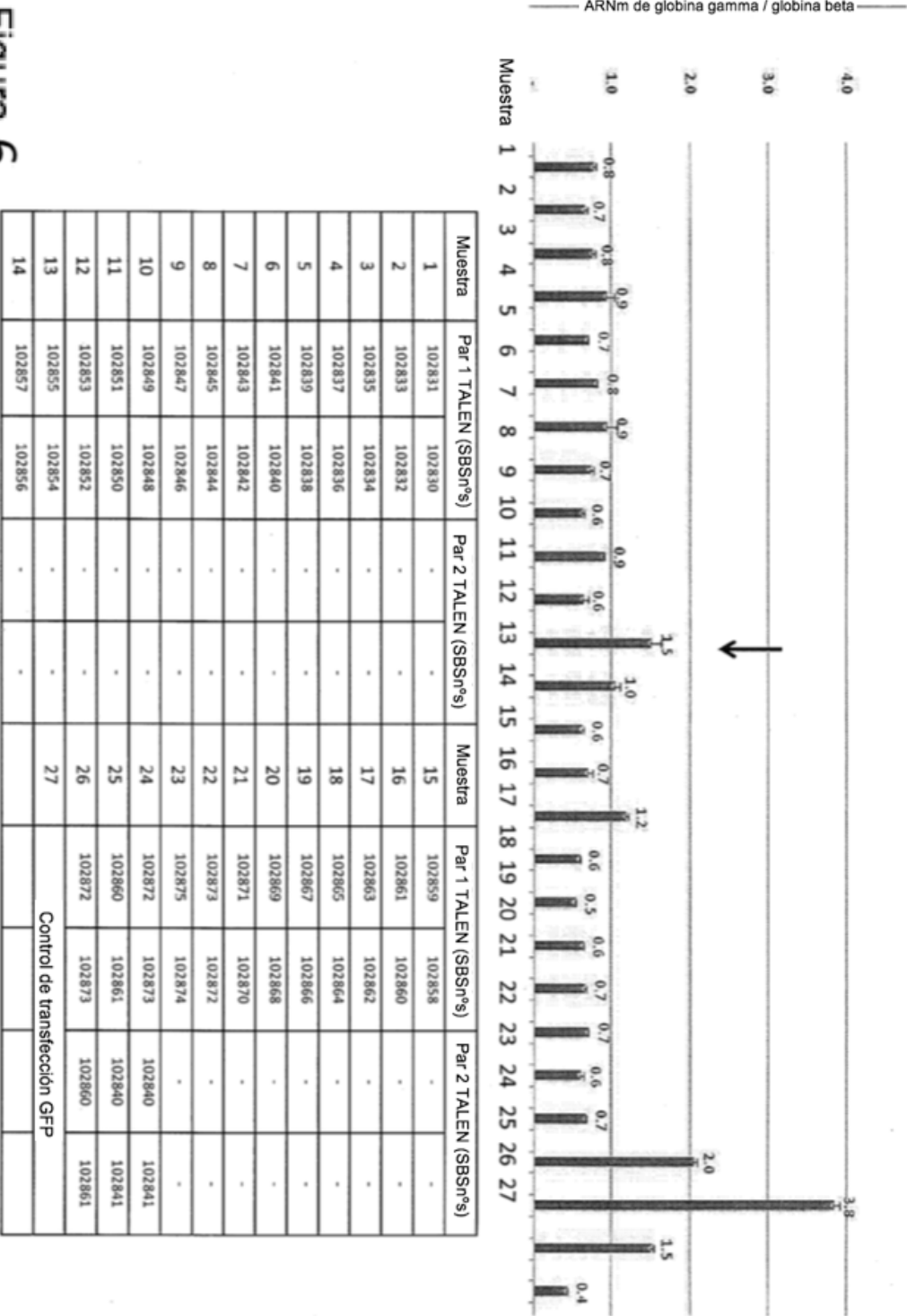
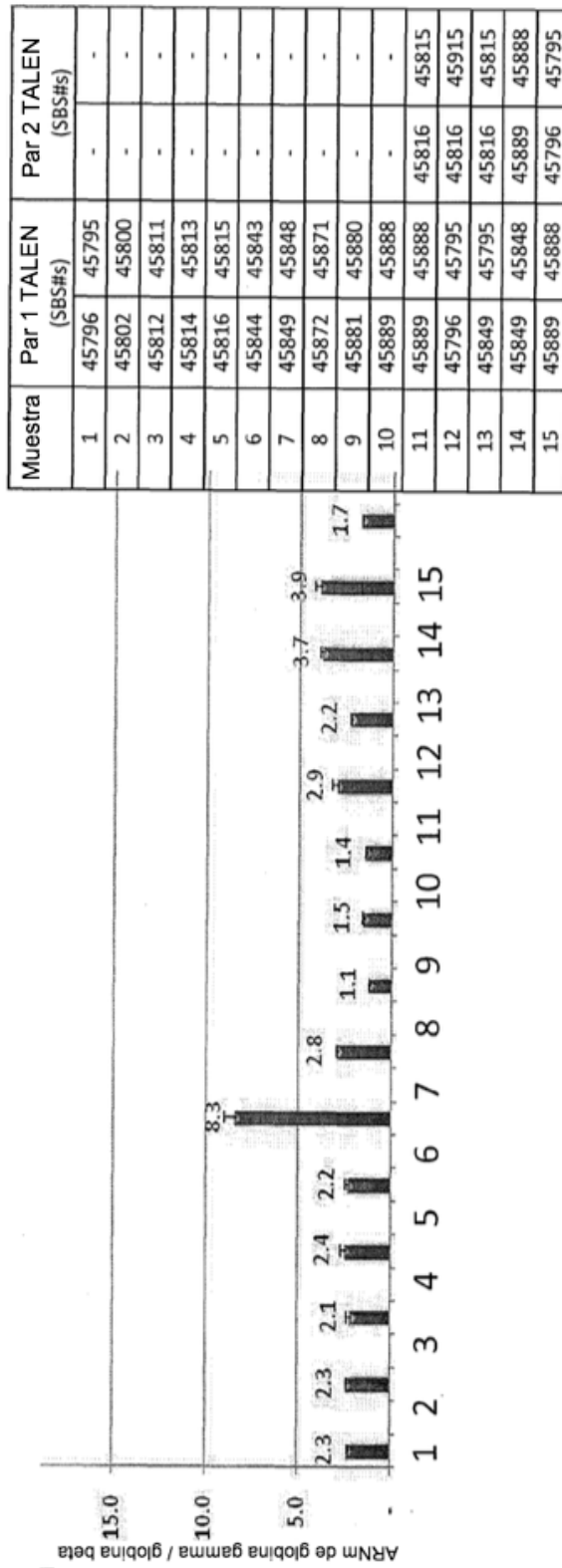


Figura 7





Bell, A. T. & Co.,

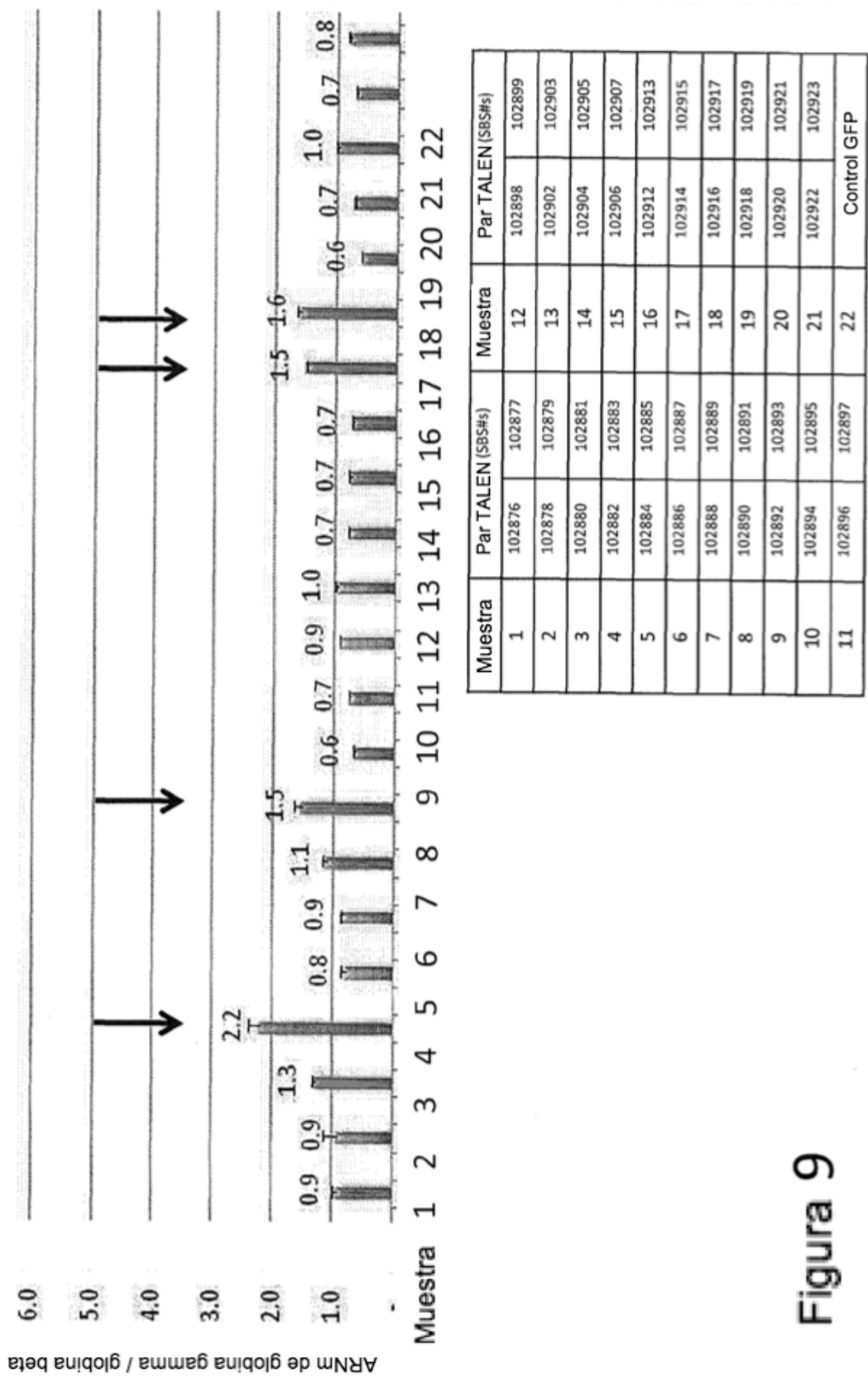


Figura 9

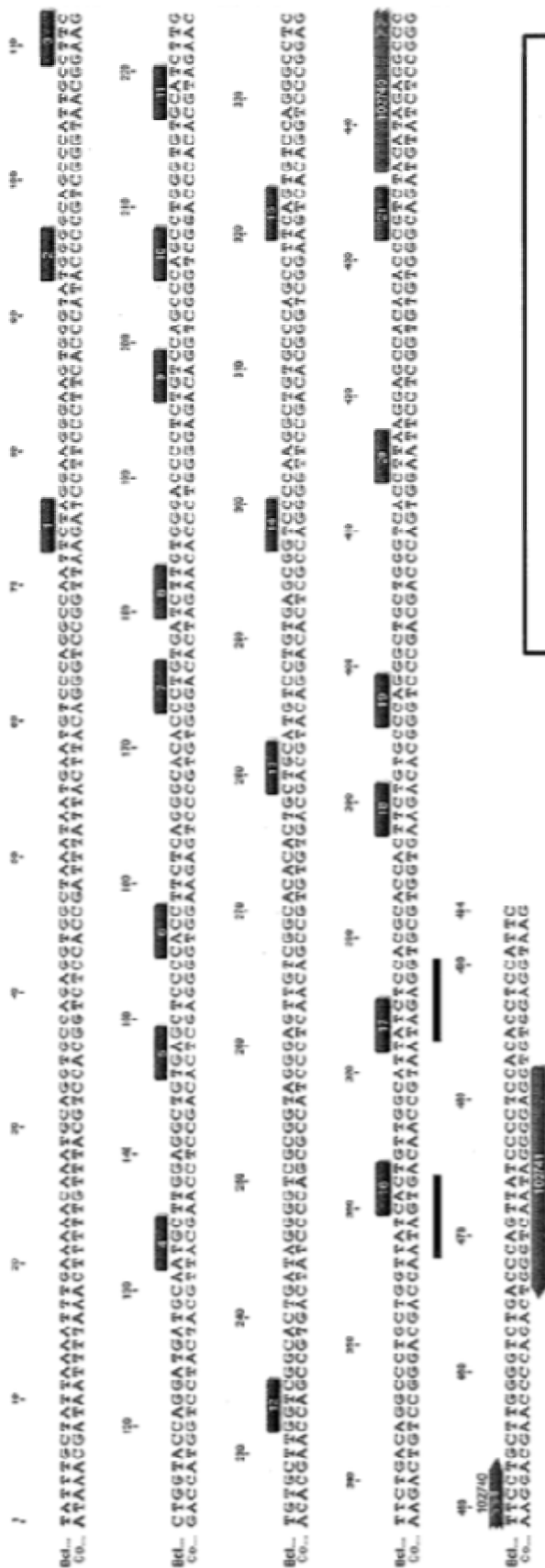


Figura 10

Figura 11A: Secuencia de potenciador +55 BCL11A (SEQ ID NO:1)

5'
 caattctag gaagggaagt gggatatggg cagccattg ccttcctggt accaggatga
 tgcaatgctt ggaggetgtg agtcccccac cttctcaggg cacacctgt gatottgtgg gacccctctg
 tccagcccag cctgggtgtg catcttgtgt gcttggtcgg cactgatagg ggtcgcggtg gggagttgtc
 ggcacacact gotgcatgtc ctgtgagcgg tccccaaggg tgtgccagc cttcagtgtc cagggcctct
 tctgacaggc cctgctggtt atcactgttg gcattatctc cagcaccac ttctgtgcc agggctgtg
 ggtcacctta aggagccaca caccgt 3'

Figura 11B: Secuencia de potenciador +58 BCL11A (SEQ ID NO:2)

5' gagg tactgatgga ccttggtgtc tattcctgtg ataaggaagg cagctagaca
 ggacttgga gttatctgta gtgagatggc tgaaaagcga tacagggctg gotctatgce
 ccaggtgtgc ataagtaaga gcagatagct gattccagtg caaagtccat acaggttaata
 acataggcca gaaaagagat atggcatcta ctcttagaca taacacacca gggatcaatac
 aactttgaag ctagtctagt gcaagctaac agttgctttt atcacaggct ccaggaaggg
 tttggcctct gallaggggtg ggggcgtggg tggggtagaa gaggactggc agacctctcc
 atcggtggcc gtttgcccag gggggcctct ttoggaaggc tctcttggtg atggagaatt
 ggattttatt tctcaatggg aatgaaataa tttgtatgce atgcegtgtg gactcccaaa
 attgtaaagg aggtgaagct tcccctgtct gactctccc ctctcataa ttgtccattt
 ttcactctgtc gggctgtcc 3'

Figura 11C: Secuencia de potenciador +62 BCL11A (SEQ ID NO:3)

5' cgtttttag aacttagctt tttgcattga ggatgcgcag gtggctgaga ctaactttctt
 lgcagatgac catggltgaa agtcagctat agagttgcac aaccacgtag ttgggcttca
 catatagaag atgttgtoat ttttggttaa ctctgtcaga ctttaccacac ctggcgcaca
 gtctgggttg cacataaaact tcacatttgc tctctccag ggtgtggggg ggtgtgttaa
 agaggggtga tattcatgct aatctttgtg tagcataaca tgtaactgca acttgctttt
 ttttttttat ctgaaagttc aagtagatat cagaaggga atgtttgtgg gtg 3'

Figura 12A

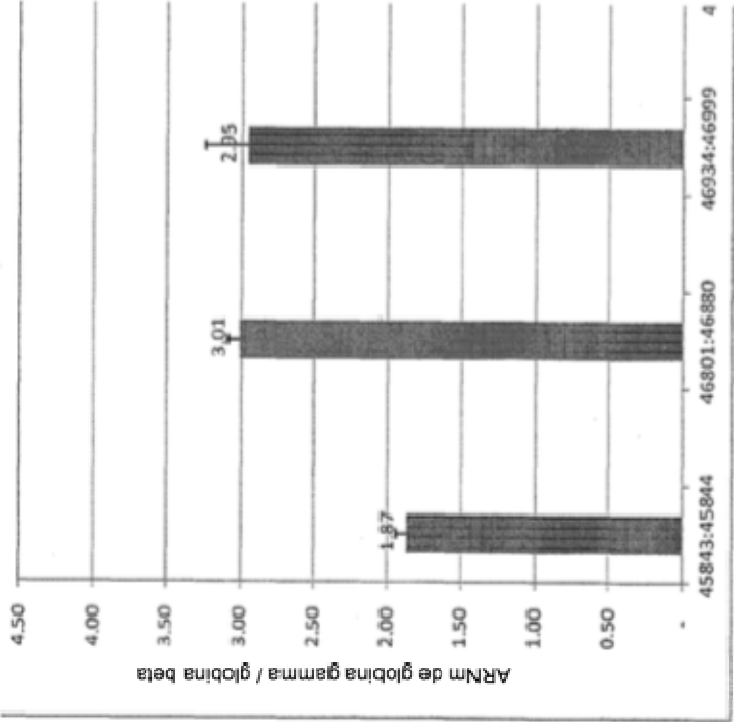
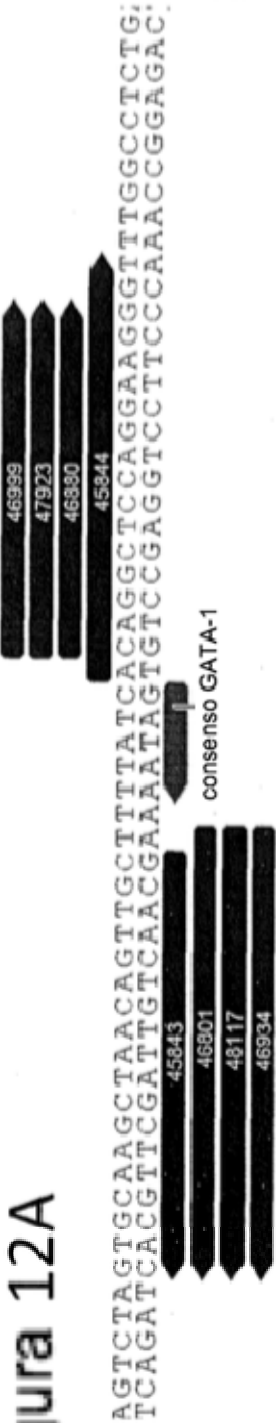


Figura 12B

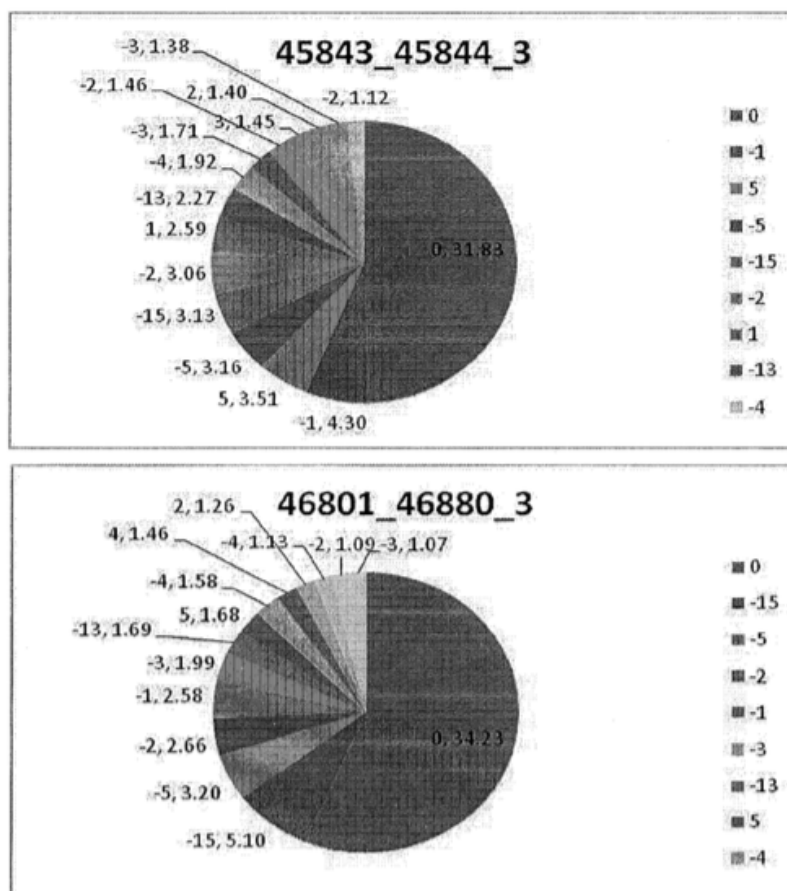
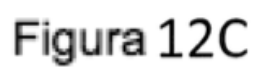
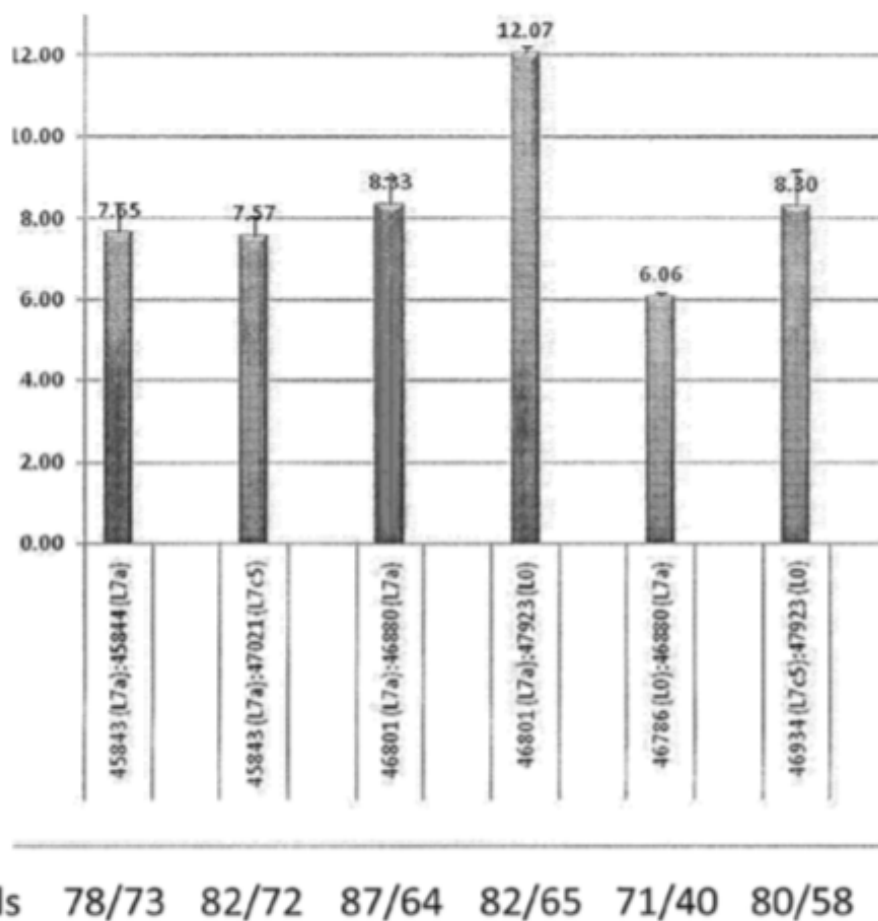


Figura 13



L0
 G S Q L V K S E L E E K (SEQ IN NO:256)
GGATCCAGCTGGTGAAGAGCCAGCTGGAGGAGAAG (SEQ IN NO:257)

L7a
 G S Q L V K S K S E A A R E L E E K (SEQ IN NO:258)
GGATCCAGCTGGTGAAGAGCAAGAGCGAGCCGCTGCCCCCGCGAGCTGGAGGAGAAG (SEQ IN NO:259)

L7c5
 G S I S R A R P L N P H P E L E E K (SEQ IN NO:260)
GGATCCATCAGCAGCCAGACCACCTGAACCCGACCCGAGCTGGAGGAGAAG (SEQ IN NO:261)

L8c4
 G S Y A P M P P L A L A S P E L E E K (SEQ IN NO:262)
GGATCCTACGCTCCAATGCCACCCCTGGCTCTGGCTTCCCCCAGAGCTGGAGGAGAAG (SEQ IN NO:263)

Figura 14

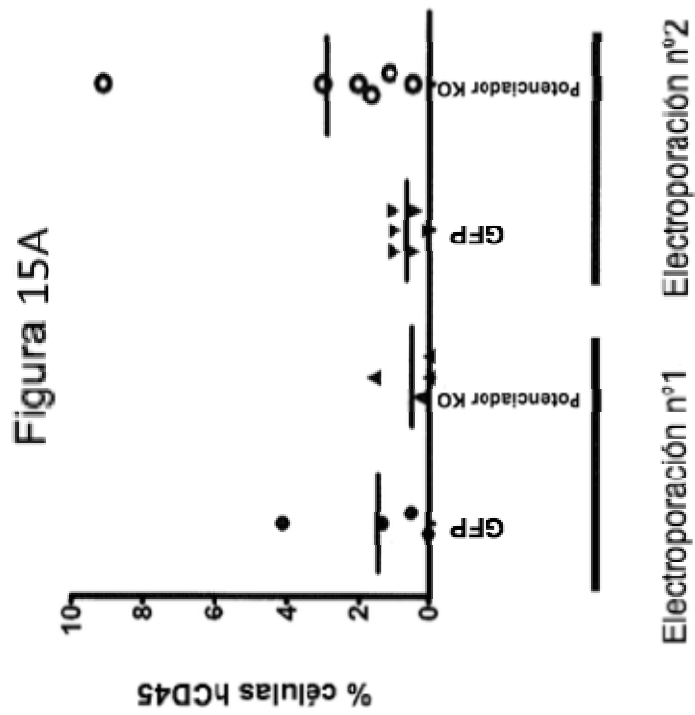
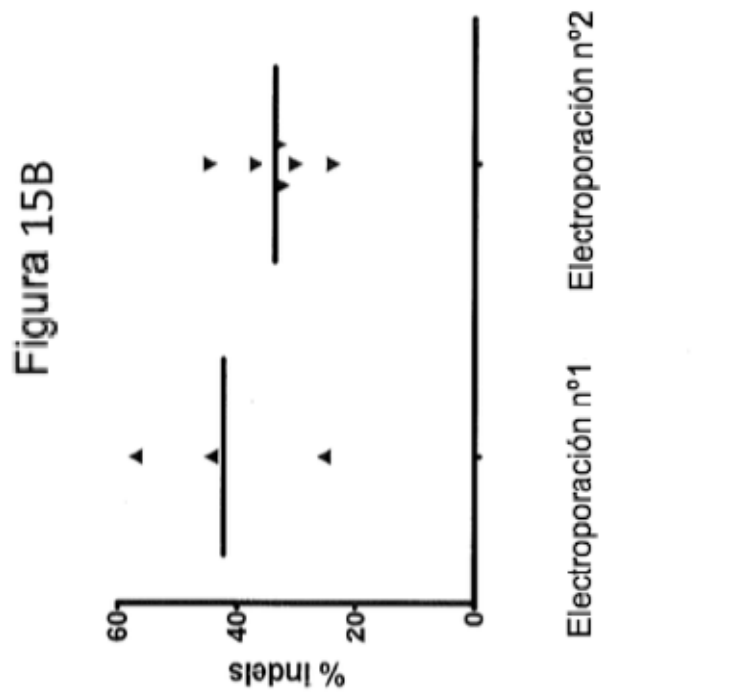


Figura 16B

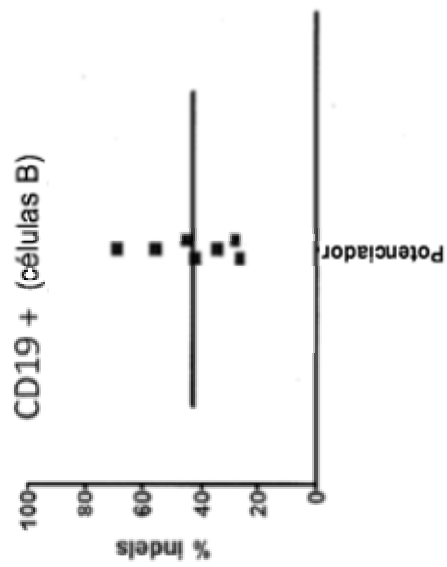


Figura 16A

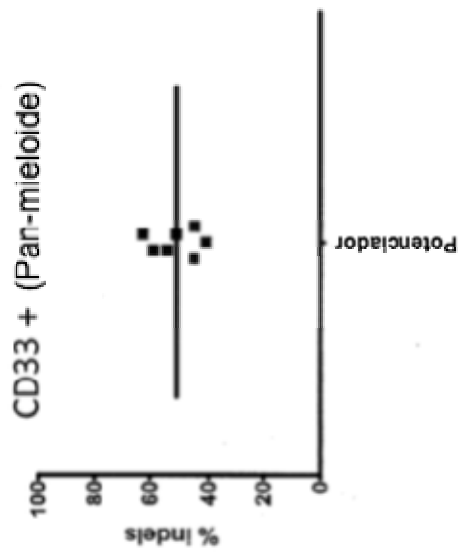


Figura 16D

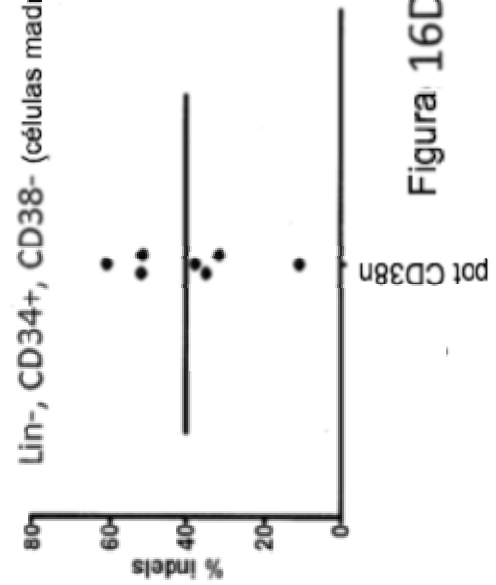
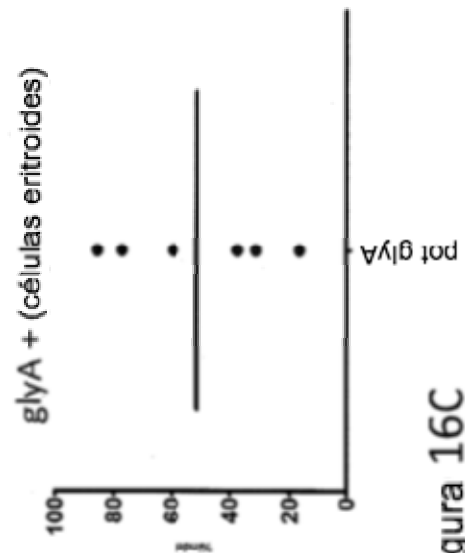


Figura 16C



Nombre de enlazador

L0 [+11]	LRGSQFVIPNRGVTKQLVKSELEEEKKSEL
L0 [+9]	LRGS--VIPNRGVTKQLVKSELEEEKKSEL
L0 [+7]	LRGS----PNRGVTKQLVKSELEEEKKSEL
L0 [+5]	LRGS-----RGVTKQLVKSELEEEKKSEL
L0 [+3]	LRGS-----VTKQLVKSELEEEKKSEL
L7a (+7)	LRGS----QLVKSKSEAAARELEEEKKSEL
L8c4 (+7)	LRGS----YAPMPPLALASPELEEEKKSEL
L7c5 (+6)	LRGS-----ISRARPLNPHPELEEEKKSEL
L0	LRGS-----QLVKSELEEEKKSEL
L0 [-1]	LRGS-----LVKSELEEEKKSEL
L0 [-2]	LRGS-----VKSELEEEKKSEL

Figura 17