

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 900 317**

51 Int. Cl.:

**A61K 35/28** (2015.01)  
**A61K 38/00** (2006.01)  
**A61P 7/00** (2006.01)  
**A61P 7/06** (2006.01)  
**C12N 9/22** (2006.01)  
**C12N 15/90** (2006.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **24.08.2017** **PCT/US2017/048397**

87 Fecha y número de publicación internacional: **01.03.2018** **WO18039440**

96 Fecha de presentación y número de la solicitud europea: **24.08.2017** **E 17844404 (8)**

97 Fecha y número de publicación de la concesión europea: **13.10.2021** **EP 3504229**

54 Título: **Regulación de la expresión génica al usar nucleasas genomanipuladas**

30 Prioridad:

**24.08.2016 US 201662378978 P**  
**09.01.2017 US 201762443981 P**  
**15.08.2017 US 201762545778 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**16.03.2022**

73 Titular/es:

**SANGAMO THERAPEUTICS, INC. (100.0%)**  
**7000 Marina Blvd**  
**Brisbane, CA 94005, US**

72 Inventor/es:

**MILLER, JEFFREY C. y**  
**REBAR, EDWARD J.**

74 Agente/Representante:

**LEHMANN NOVO, María Isabel**

ES 2 900 317 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 génica al usar nucleasas genomanipuladas

## CAMPO TÉCNICO

La presente descripción se encuentra en el campo de la manipulación del genoma, particularmente la modificación dirigida del genoma de una célula hematopoyética.

## ANTECEDENTES

Cuando se considera que los esfuerzos de secuenciación genómica han revelado que el genoma humano contiene entre 20.000 y 25.000 genes, pero menos de 2000 reguladores transcripcionales, se torna evidente que varios factores deben interactuar para controlar la expresión génica en la totalidad de sus diversas manifestaciones temporales, del desarrollo y específicas del tejido. La expresión de los genes está controlada por una mezcla muy compleja de reguladores transcripcionales generales y específicos que interactúan con elementos de ADN. Estos elementos de ADN comprenden elementos de ADN local tales como el promotor central y sus sitios de unión a factores de transcripción asociados, así como elementos distales tales como potenciadores, silenciadores, aislantes y regiones de control del locus (LCR, por sus siglas en inglés) (ver Matson, et al. (2006) Ann Rev Genome Hum Genet 7: 29-50).

Se identificaron elementos potenciadores por primera vez en el genoma vírico de SV40 y después se hallaron en el locus de cadena pesada de la inmunoglobulina humana. Ahora se sabe que los potenciadores desempeñen funciones de regulación en la expresión de muchos genes y aparentan afectar principalmente los patrones temporales y especiales de la expresión génica. También se ha hallado que los potenciadores pueden tener la función de regular la expresión a grandes distancias desde el promotor central del gen diana y no dependen de ninguna orientación de secuencia específica con respecto al promotor. Los potenciadores pueden ubicarse varios cientos de kilobases gen arriba o gen abajo con respecto a una región promotora central, donde pueden ubicarse en una secuencia de intrón, o incluso más allá del extremo 3' de un gen.

Se han descrito diversos métodos y composiciones para la escisión dirigida de ADN genómico. Dichos eventos de escisión dirigida pueden usarse, por ejemplo, para inducir mutagénesis dirigida, inducir eliminaciones dirigidas de secuencias de ADN celular y facilitar la recombinación dirigida en un locus cromosómico predeterminado. Ver, p. ej., las patentes estadounidenses núms. 9.255.250; 9.200.266; 9.045.763; 9.005.973; 9.150.847; 8.956.828; 8.945.868; 8.703.489; 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; las publicaciones de patentes estadounidenses núms. 2003/0232410; 2005/0208489; 2005/0026157; 2005/0064474; 2006/0063231; 2008/0159996; 2010/00218264; 2012/0017290; 2011/0265198; 2013/0137104; 2013/0122591; 2013/0177983; 2013/0196373; 2015/0056705 y 2015/0335708.

Estos métodos a menudo implican el uso de sistemas genomanipulados de escisión para inducir una rotura de cadena doble (DSB, por su sigla en inglés) o un corte en la secuencia de ADN diana, de modo que la reparación de la rotura mediante un proceso que surja de un error, tal como la unión de extremos no homólogos (NHEJ, por sus siglas en inglés), captura de extremos dirigida sin homología de donantes o la reparación mediante el uso de una plantilla de reparación (reparación dirigida por homología o HDR, por sus siglas en inglés) pueda producir la inactivación de un gen o la inserción de una secuencia de interés (integración dirigida). Ver, p. ej., las patentes estadounidenses núms. 9.045.763; 9.200.266; 9.005.973; y 8.703.489. Estas técnicas también pueden usarse para introducir cambios específicos del sitio en la secuencia del genoma mediante el uso de un oligonucleótido donante, incluso la introducción de eliminaciones específicas de regiones genómicas, o de mutaciones puntuales específicas o alteraciones localizadas (también conocidas como corrección génica). La escisión puede producirse mediante el uso de nucleasas específicas tales como nucleasas de dedos de zinc (ZFN, por sus siglas en inglés) genomanipuladas, nucleasas efectoras similares al activador de la transcripción (TALEN, por sus siglas en inglés) o mediante el uso del sistema CRISPR/Cas con un ARNcr/ARNtracr genomanipulado ("ARN guía único") para guiar la escisión específica. Además, las nucleasas dirigidas se desarrollan en función del sistema Argonaute (p. ej., de *T. thermophilus*, conocido como "TtAgo", ver Swarts et al. (2014) Nature 507(7491): 258-261), que también puede tener el potencial de usarse en la edición del genoma y la terapia génica.

Los glóbulos rojos (RBC, por sus siglas en inglés), o eritrocitos, son el componente celular principal de la sangre y representan un cuarto de las células en un ser humano. Los RBC maduros carecen de un núcleo y muchos otros organelos y están llenos de hemoglobina, una metaloproteína que funciona para transportar oxígeno de los pulmones a los tejidos, así como para transportar dióxido de carbono hacia afuera de los tejidos y de regreso hacia los pulmones para su eliminación. Esta proteína constituye aproximadamente el 97 % del peso seco de los RBC y aumenta la capacidad de transporte de oxígeno de la sangre aproximadamente setenta veces. La hemoglobina es un heterotetramero que comprende dos cadenas de globina de tipo alfa ( $\alpha$ ), dos cadenas de globina de tipo beta ( $\beta$ ) y 4 grupos hemo. En adultos, el tetramero  $\alpha_2\beta_2$  se denomina Hemoglobina A (HbA) o hemoglobina del adulto. Típicamente, las cadenas de globina alfa y beta se sintetizan en una relación aproximada de 1:1 y esta relación aparenta ser fundamental 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 posee una afinidad de unión mayor por el oxígeno que la Hemoglobina A, de modo que puede administrarse oxígeno al sistema del bebé a través del torrente sanguíneo de la

madre. Hay dos genes que codifican la globina fetal cuyas secuencias son muy similares y se denominan HBG1 (también denominada Ggamma) y HBG2 (Agamma), según su orden de disposición en el locus génico de la globina beta. Como la hemoglobina del adulto, la proteína hemoglobina fetal contiene dos cadenas de globina  $\alpha$ , pero en lugar de las cadenas de globina  $\beta$  del adulto, tiene dos cadenas de globina gamma ( $\gamma$ ) fetal (es decir, la hemoglobina fetal es  $\alpha_2\gamma_2$ ). A las aproximadamente 30 semanas de gestación, la síntesis de la globina gamma en el feto comienza a disminuir, mientras que aumenta la producción de la globina beta. A los aproximadamente 10 meses de edad, la hemoglobina del recién nacido es casi toda  $\alpha_2\beta_2$ , aunque persiste algo de HbF en la vida adulta (aproximadamente 1-3 % de la hemoglobina total). La regulación del cambio de la producción de globina gamma a beta es bastante compleja, e implica principalmente una regulación por disminución de la transcripción de la globina gamma con una regulación por aumento simultánea de la transcripción de la globina beta.

Los defectos genéticos en las secuencias que codifican las cadenas de hemoglobina pueden ser responsables de un grupo de enfermedades conocidas como hemoglobinopatías, que incluyen anemia falciforme y las talasemias alfa y beta. En la mayoría de los pacientes con hemoglobinopatías, los genes que codifican la globina gamma permanecen presentes, pero la expresión es relativamente baja debido a la represión génica normal que se produce alrededor del parto según se describió anteriormente.

Se estima que 1 en 5000 personas en Estados Unidos padecen anemia de células falciformes (SCD, por sus siglas en inglés), mayormente en las personas con descendencia de África subsahariana (Roseff (2009) Immunohematology 25(2):67). Parece haber un beneficio para portadores heterocigóticos de la mutación falciforme para la protección contra la malaria, de modo que este rasgo puede haberse seleccionado de forma positiva con el tiempo, de modo que se estima que en África subsahariana, hasta el 28 % de la población posee el rasgo de las células falciformes (Elguero et al. (2015) PNAS USA 112 (22): 7051). La anemia de células falciformes es provocada por una mutación en el gen de la globina  $\beta$  como consecuencia de una sustitución de valina por ácido glutámico en el aminoácido núm. 6 (una mutación GAG a GTG a nivel del ADN), donde la hemoglobina resultante se denomina "hemoglobina S" o "HbS". En condiciones con menos oxígeno, un cambio conformacional de 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 de la hemoglobina pueden asociarse con el parche hidrófobo, lo que provoca que las moléculas de HbS se acumulen y formen precipitados fibrosos. Estos aglomerados, a su vez, provocan la anomalía o "formación falciforme" de los RBC, lo que da como resultado una pérdida de flexibilidad de las células. Los RBC falciformes ya no pueden caber en los lechos capilares y esto puede provocar una crisis vasooclusiva en los pacientes con anemia falciforme. Además, los RBC falciformes son más frágiles que los RBC normales, y tienden hacia la hemólisis, lo que finalmente lleva a la anemia en el paciente.

El tratamiento y el control de los pacientes con anemia falciforme es una propuesta que dura toda la vida e implica un tratamiento con antibióticos, gestión del dolor y transfusiones durante episodios agudos. Un abordaje es el uso de hidroxiurea, que ejerce sus efectos en parte al aumentar la producción de globina gamma. Sin embargo, todavía se desconocen los efectos secundarios a largo plazo de la terapia crónica con hidroxiurea y el tratamiento provoca efectos secundarios no deseados que llevan a un escaso cumplimiento del paciente y la eficacia es variable según el paciente (Brandow y Panepinto (2011) Am J Hematol 86(9):804-806). A pesar de un aumento de la eficacia de los tratamientos de células falciformes, la esperanza de vida de los pacientes todavía es solo de mediados a fines de los 50 y las morbilidades asociadas con la enfermedad tienen un gran impacto en la calidad de vida de un paciente.

Las talasemias también son enfermedades relacionadas con la hemoglobina y típicamente implican una reducción de la expresión de las cadenas de globina. Esto puede producirse a través de mutaciones en las regiones reguladoras de los genes o de una mutación en una secuencia codificante de una globina que produce una reducción de la expresión o una reducción de los niveles o proteína globina funcional. Las talasemias alfa, causadas por mutaciones en el locus de la globina alfa, se asocian principalmente con personas de descendencia de África occidental y Asia del sur, y pueden otorgar resistencia a la malaria. La talasemia beta, causada por mutaciones en el locus de la globina beta, se asocia principalmente con personas de descendencia mediterránea, típicamente de Grecia y las áreas costeras de Turquía e Italia. En la talasemia menor, solamente uno de los alelos de la globina  $\beta$  porta una mutación. Los individuos padecerán anemia microcítica y la detección normalmente implica un volumen corpuscular medio menor que el normal ( $<80\text{fL}$ ). Los alelos de los sujetos con talasemia menor son  $\beta^+/\beta$  o  $\beta^0/\beta$  (donde " $\beta^+$ " hace referencia a alelos que permiten que se produzca alguna cantidad de formación de la cadena  $\beta$ , " $\beta^0$ " hace referencia a los alelos de globina  $\beta$  natural, y " $\beta^0$ " hace referencia a mutaciones de globina  $\beta$  asociadas con una ausencia completa de la expresión de globina beta). Los sujetos con talasemia intermedia a menudo pueden vivir una vida normal, pero pueden precisar transfusiones ocasionales, especialmente en momentos de enfermedad o embarazo, dependiendo de la gravedad de su anemia. Los alelos de estos pacientes pueden ser  $\beta^+/\beta^+$  o  $\beta^0/\beta^+$ . La talasemia mayor se produce cuando ambos alelos poseen mutaciones de talasemia. Esta es una anemia gravemente microcítica e hipocrómica. Si no se trata, produce anemia, esplenomegalia y deformidad ósea grave y normalmente conduce a la muerte antes de los 20 años. El tratamiento consiste en transfusiones sanguíneas periódicas, esplenectomía para la esplenomegalia y la quelación de la sobrecarga de hierro provocada por la transfusión. También se usan trasplantes de médula ósea para el tratamiento de personas con talasemias graves si se puede identificar un donante adecuado, pero este procedimiento puede conllevar riesgos considerables.

Un abordaje que se ha propuesto para el tratamiento de SCD y talasemias beta es el de aumentar la expresión de la globina gamma con la intención de que HbF reemplace funcionalmente la hemoglobina del adulto aberrante. Según

se mencionó anteriormente, se considera que el tratamiento de los pacientes con SCD con hidroxurea es exitoso en parte debido a su efecto del aumento de la expresión de globina gamma. El primer grupo de compuestos que se descubrió que afectaban la expresión de globina gamma fueron fármacos citotóxicos. La capacidad de provocar la síntesis *de novo* de la globina gamma mediante manipulación farmacológica se demostró primero mediante el uso de 5-azacitidina en animales experimentales (DeSimone (1982) Proc Nat'l Acad Sci USA 79(14):4428-31). Los estudios posteriores confirmaron la capacidad de la 5-azacitidina de aumentar la HbF en los pacientes con talasemia  $\beta$  y anemia de células falciformes (Ley, et al., (1982) N. Engl. J. Medicine, 307: 1469-1475 y Ley, et al., (1983) Blood 62: 370-380). Además, se ha demostrado en sistemas experimentales que los ácidos grasos de cadena corta (p. ej., butirato y derivados) aumentan la HbF (Constantoulakis et al., (1988) Blood 72(6):1961-1967). A su vez, hay un segmento de la población humana con una afección conocida como "persistencia hereditaria de hemoglobina fetal" (PHHF) donde persisten cantidades elevadas de HbF en la adultez (10-40 % en heterocigóticos para PHHF (ver Thein et al. (2009) Hum. Mol. Genet 18 (R2): R216-R223). Esta es una afección poco común, pero ante la ausencia de anomalías asociadas de globina beta, no se asocia con ninguna manifestación clínica considerable, incluso cuando el 100 % de la hemoglobina del individuo es HbF. Cuando los individuos que padecen una talasemia beta también poseen una PHHF coincidente, la expresión de HbF puede disminuir la gravedad de la enfermedad (Potoka and Gladwin (2015) Am J Physiol Lung Cell Mol Physiol. 308(4): L314-L324). Además, la gravedad de la evolución natural de la anemia de células falciformes puede variar de manera significativa de paciente a paciente, y esta variabilidad, en parte, puede remitirse al hecho de que algunos individuos con una enfermedad más leve expresan niveles más elevados de HbF.

Un abordaje para aumentar la expresión de HbF implica la identificación de genes cuyos productos desempeñan una función en la regulación de la expresión de la globina gamma. Un gen de ese tipo es el BCL11A, identificado primero por su función en el desarrollo de los linfocitos. El BCL11A codifica una proteína de dedos de zinc que se cree que participa en la regulación específica de la etapa de desarrollo de la expresión de la globina gamma. El BCL11A se expresa en células precursoras eritroides adultas y la regulación por disminución de su expresión lleva a un aumento de la expresión de la globina gamma. Además, el empalme del ARNm del BCL11A se regula por medio del desarrollo. En las células embrionarias, las variantes más cortas del ARNm de BCL11A, conocidas como BCL11A-S y BCL11A-XS se expresan de forma primaria, mientras que, en las células adultas, las variantes de ARNm de BCL11A-L y BCL11A-XL más largas se expresan de forma predominante. Ver Sankaran et al. (2008) Science 322 pág. 1839. La proteína BCL11A aparenta interactuar con el locus de la globina beta para alterar su conformación y, por lo tanto, su expresión en distintas etapas del desarrollo. Se ha propuesto el uso de un ARN inhibidor dirigido al gen BCL11A (ver, p. ej., la publicación de patente estadounidense núm. 2011/0182867), pero esta tecnología tiene varias desventajas potenciales, a saber, que puede que no se logre la inactivación completa, el suministro de tales ARN puede ser problemática y los ARN deben estar presentes de forma continua, lo que requiere múltiples tratamientos de por vida.

El direccionamiento de secuencias potenciadoras de BCL11A proporciona un mecanismo para el aumento de HbF. Ver, p. ej., la patente estadounidense núm. 2015/0132269 y la publicación PCT núm. WO 2016/183298. Los estudios de asociación de genoma completo han identificado un conjunto de variaciones genéticas en el locus del gen BCL11A que se asocian con mayores niveles de HbF. Estas variaciones son una colección de polimorfismos de nucleótidos pequeños (SNP, por sus siglas en inglés) que se hallan en regiones no codificantes de BCL11A que funcionan como una región potenciadora específica de la etapa limitada por el linaje. Una investigación adicional reveló que este potenciador de BCL11A es necesario en las células eritroides para la expresión de BCL11A, pero no es necesario para su expresión en linfocitos B (ver Bauer et al. (2013) Science 343:253-257). La región potenciadora se halló dentro del intrón 2 del gen BCL11A, y se identificaron tres áreas de hipersensibilidad de ADN (que a menudo indica un estado de cromatina que se asocia con un potencial regulatorio) 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 tienen una longitud aproximada de 350 (+55); 550 (+58); y 350 (+62) nucleótidos (Bauer 2013, *ibid*).

Cuando se desarrolla una nucleasa para usar en tratamientos terapéuticos de seres humanos, es fundamental que la nucleasa tenga características de seguridad superiores. Específicamente, las nucleasas deben tener niveles muy bajos de escisión inespecífica. Cantidades significativas de cortes de cadena doble en lugares que no sean la diana especificada por el usuario pueden provocar la represión de genes inespecíficos y, en algunos casos poco frecuentes, la aparición de translocaciones cromosómicas (ver Hoban y Bauer (2016) Blood, 127(21):2525-2535 y Tsang et al. (2017) Nature Methods, en la prensa). Se pueden lograr mejoras en la especificidad eliminando las interacciones no específicas entre la nucleasa genomanipulada y el ADN genómico (ver las solicitudes de patente provisional estadounidenses núms. 62/378.978 y 62/443.981).

Por lo tanto, permanece una necesidad de métodos y composiciones altamente específicos para la alteración de la expresión génica de BCL11A, por ejemplo, para tratar hemoglobinopatías tales como la anemia de células falciformes y la talasemia beta.

#### COMPENDIO

En la presente memoria se describe composiciones y métodos altamente específicos para su uso en la terapia génica y genomanipulación del genoma. Específicamente, los métodos y composiciones descritos se refieren a la inactivación (p. ej., mediante la abolición parcial o completa de su expresión) de un gen BCL11A, por ejemplo, un gen que actúa como regulador de uno o más genes adicionales. En particular, se describen métodos y composiciones para interferir

- en la función potenciadora en un gen BCL11A para disminuir o inactivar su actividad en linajes celulares específicos (p. ej., eritroides). Además, se describen métodos y composiciones para interferir en las funciones potenciadoras de BCL11A, en donde las secuencias potenciadoras no se encuentran dentro de las secuencias codificantes del gen BCL11A, y en donde los reactivos proporcionados exhiben una actividad altamente específica. La regulación por
- 5 disminución resultante del gen BCL11A en estas circunstancias resulta en una mayor expresión de globina gamma y en la reducción de la cantidad de eventos de escisión inespecífica. La invención proporciona una nucleasa de dedos de zinc (ZFN) que comprende la secuencia de aminoácidos como se muestra en la SEQ ID NO:29 y la secuencia de aminoácidos como se muestra en la SEQ ID NO:31.
- 10 La invención también proporciona uno o más polinucleótidos que codifican una o más nucleasas de dedos de zinc según la invención, opcionalmente, en donde el uno o más polinucleótidos es(son) ARNm.
- La invención proporciona, además, uno o más vectores que comprenden secuencias que codifican una o más nucleasas de dedos de zinc de la invención.
- La invención proporciona una célula que comprende la nucleasa de dedos de zinc de la invención, el uno o más polinucleótidos de la invención o el uno o más vectores de la invención.
- 15 La invención también proporciona una célula o línea celular producida a partir de la célula de la invención, que comprende la nucleasa de dedos de zinc de la invención, el uno o más polinucleótidos de la invención o el uno o más vectores de la invención.
- 20 La invención proporciona, además, una célula parcialmente o completamente diferenciada que descende de la célula o línea celular de la invención, que comprende la nucleasa de dedos de zinc de la invención, el uno o más polinucleótidos de la invención o el uno o más vectores de la invención.
- La invención proporciona una composición farmacéutica que comprende la nucleasa de dedos de zinc de la invención, el uno o más polinucleótidos de la invención, el uno o más vectores de la invención y/o la célula o línea celular de la invención.
- 25 La invención también proporciona un método in vitro para modificar una secuencia potenciadora de BCL11A endógena en una célula, el método comprende administrar la nucleasa de dedos de zinc de la invención o el uno o más polinucleótidos de la invención a la célula, de modo que se modifique la secuencia potenciadora de BCL11A endógena, opcionalmente en donde: (i) el método comprende, además, introducir una secuencia exógena en la célula de modo que la secuencia exógena se inserte en la secuencia potenciadora de BCL11A endógena; o (ii) la modificación comprende una eliminación.
- 30 La invención proporciona, además, una composición farmacéutica de la invención para su uso en un método para tratar una hemoglobinopatía, tal como una talasemia (p. ej., talasemia beta) o anemia de células falciformes, en un paciente, el método comprende administrar la composición farmacéutica al paciente, opcionalmente, en donde el paciente es un ser humano.
- 35 La invención proporciona un método in vitro para producir una célula modificada genéticamente que comprende una modificación genómica dentro de una secuencia potenciadora de BCL11A endógena, el método comprende las etapas de: a) poner en contacto una célula con el uno o más polinucleótidos de la invención; b) someter a la célula a condiciones que lleven a la expresión de la proteína de fusión a partir del polinucleótido; y c) modificar la secuencia potenciadora de BCL11A endógena con la proteína de fusión expresada de manera suficiente para producir la célula modificada genéticamente, opcionalmente, en donde el método comprende, además, estimular la célula con al menos
- 40 una citocina.
- La invención también proporciona un kit que comprende la nucleasa de dedos de zinc de la invención, el uno o más polinucleótidos de la invención, el uno o más vectores de la invención y/o la célula o línea celular de la invención.
- La invención proporciona, además, una célula o línea celular de la invención para su uso en un método para tratar o prevenir una hemoglobinopatía, tal como una talasemia (p. ej., talasemia beta) o anemia de células falciformes.
- 45 En la presente memoria se describe una proteína de dedos de zinc de origen no natural que comprende una proteína de dedos de zinc (ZFP) que comprende 4, 5 o 6 dedos, cada dedo comprende una región de hélices de reconocimiento que reconoce un subsitio diana de ADN, en donde las regiones de hélices de reconocimiento comprenden las secuencias en el orden que se muestra en una fila individual de la Tabla 1. Dentro de cada dedo de zinc, la región de hélice de reconocimiento de 7 aminoácidos se enumera -1 a +6 dentro de la estructura del dedo de zinc (de
- 50 aproximadamente 30 residuos, incluidos residuos coordinantes de zinc). 1, 2, 3 o más de los dedos de zinc componente de las proteínas de dedos de zinc descritos en la presente memoria pueden comprender, además, mutaciones en uno o más residuos fuera de la región de hélice de reconocimiento, que incluyen, pero no se limitan a, mutaciones en aminoácidos en la posición -5, posición -14 o en ambas posiciones -5 y -14 (la numeración continúa desde la numeración -1 a +6 usada para la región de hélice de reconocimiento). Ver, p. ej., las mutaciones Qm4 y
- 55 Qm14 descritas en las solicitudes de patente provisional estadounidense 62/378.978 y 62/443.981. Los dedos de zinc componentes de la proteína de dedos de zinc se pueden enlazar con cualesquiera enlazadores, por ejemplo, como se

describen en la patente estadounidense núm. 8.772.453. La ZFP comprende las hélices de reconocimiento como se muestran en la Tabla 1 para las proteínas que se denominan de la siguiente manera: 63014 (que se une al sitio diana que se muestra en la SEQ ID NO:1) y 65722 (que se une al sitio diana que se muestra en la SEQ ID NO:2).

En ciertos aspectos de la descripción, las proteínas de dedos de zinc descritas en la presente memoria se pueden fusionar con un dominio funcional (p. ej., dominio de activación transcripcional, dominio de represión transcripcional, dominio de escisión (para formar una nucleasa de dedos de zinc), etc.). Se puede usar cualquier enlazador para enlazar funcionalmente el dominio de escisión y la proteína de dedos de zinc, incluso, pero sin limitarse a, los enlazadores como se describen en las patentes estadounidenses núm. 9.394.531 y 9.567.609. Además, cuando se usa un dominio de escisión FokI, puede haber más mutaciones en el dominio catalítico, dominio de dimerización, en residuos de contacto de fosfato (fuera del dominio de dimerización o catalítico), y combinaciones de mutaciones en uno cualquiera del dominio catalítico, dominio de dimerización y en residuos de contacto de fosfato, incluso, pero sin limitarse a, mutaciones ELD o KKR en el dominio de dimerización, mutaciones en residuos 525 (K a S) del dominio FokI y combinaciones de mutaciones de ELD o KKR en el dominio de dimerización y mutaciones en los residuos 525 (K a S) del dominio FokI, numerado con respecto al natural. Ver, las patentes estadounidenses núms. 7.888.121; 7.914.796; 8.034.598; 8.623.618 y la publicación de patente estadounidense núm. 2011/0201055 y las solicitudes de patente provisional estadounidenses núms. 62/378.978 y 62/443.981.

Las nucleasas de dedos de zinc (ZFN) pueden usarse en pares dimerizantes para escindir en o cerca de uno o ambos sitios diana para las ZFN del par, por ejemplo, el "compañero izquierdo" de la Tabla 1 (p. ej., 63014) puede formar un dímero con el "compañero derecho" de la Tabla 1 (p. ej., 65722) para escindir secuencias potenciadoras de BCL11A. Por ejemplo, un par de ZFN puede comprender las siguientes secuencias de aminoácidos: 63014:

MDYKDHGDYKDHDIDYKDDDDKMAPKKRKVGIIHGVPAAAMAERPFQCRI  
CMQNFSDQSNLRAHIRTHTGEKPFACDICGRKFARNFSLTMEKTIHTGSQKPF  
QCRICMQNFSSGTNLNTHRTHTGEKPFACDICGRKFATSGSLTRHTKIHTHPR  
APIPKPFQCRICMQNFSDQSNLRAHIRTHTGEKPFACDICGRKFQAAQCCLFHHT  
ELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGK  
HLGGSRKPDGAIYTVGSPIDYGVIVDTKA  
TRDKHLNPNEWVKVYPSSVTEFKFLFVSGHFSGNYKAQLTRLNHNITNCNGA  
VLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF (SEQ ID NO:3);  
MDYKDHGDYKDHDIDYKDDDDKMAPKKRKVGIIHGVPAAAMAERPFQCRI  
CMQKFARNDRHTTHTKIHTGEKPFQCRICMQNFSDQKAHLIRHIRTHTGEKPF  
CDICGRKFAQKGTLEHTHTGSQKPFQCRICMQNFSGRDLRSHIRTHTGE  
KPFACDICGRKFARRDNLSHTKIH-Enlazador-  
ELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGK  
HLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQ  
YSGGYNLPIGQADEMERYVEENQ  
TRNKHINPNNEWVKVYPSSVTEFKFLFVSGHFSGNYKAQLTRLNRKTNCSNGA  
VLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF (SEQ ID NO:4), en donde la secuencia del Enlazador puede ser cualquier secuencia de enlazador conocida en la técnica, por ejemplo, como se describen en las patentes estadounidenses núms. 9.394.531 y 9.567.609. El enlazador para 63014 puede comprender o consistir en el enlazador L7c5 (LRGSISRARPLNPHP (SEQ ID NO:5)) y el enlazador usado en 65722 comprende o consiste en el enlazador L0 (LRGSQLVKS (SEQ ID NO:6), ver la patente estadounidense núm. 9.567.609). La secuencia del dominio de escisión FokI de extremo C al enlazador de las secuencias mostradas anteriormente también puede comprender dominios de FokI alternativos enlazados funcionalmente con la proteína de dedos de zinc. El dominio de escisión de FokI puede incluir mutaciones alternativas o adicionales en el dominio catalítico, el dominio de dimerización, residuos de contacto de fosfato, y combinaciones de mutaciones en uno cualquiera del dominio catalítico, dominio de dimerización y residuos de contacto de fosfato.

En la presente memoria se describe el suministro de al menos una nucleasa (p. ej., una nucleasa que se une a una secuencia potenciadora de BCL11A) a una célula madre humana o célula precursora (HSC/PC) para lograr la genomanipulación del genoma. En ciertos aspectos de la descripción, la nucleasa comprende una proteína de dedos de zinc (ZFP) que comprende 4, 5 o 6 dedos, cada dedo comprende una región de hélices de reconocimiento que reconoce un subsitio diana, en donde las regiones de hélices de reconocimiento comprenden las secuencias en el orden que se muestra en una fila individual de la Tabla 1. La nucleasa de ZFN de la invención comprende el par de nucleasas designadas 63014/65722. La(s) nucleasa(s) como se describe(n) en la presente memoria puede(n) comprender, además, un enlazador (p. ej., entre el dominio de unión a ADN y el dominio de escisión), por ejemplo, un enlazador tal como se muestra en 9.567.609, que incluye, pero no se limita a, LRGSISRARPLNPHP (SEQ ID NO:5) o (LRGSQLVKS (SEQ ID NO:6)).

En algunas realizaciones, la nucleasa se suministra como un péptido, mientras que en otras se suministra 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 algunas instancias, el ARNm está protegido. En algunos aspectos, el ARNm se puede modificar químicamente (Ver, p. ej., Kormann et al. (2011) Nature Biotechnology 29(2): 154-157). En otros aspectos, el ARNm puede comprender un casquete ARCA (ver las patentes estadounidenses núms. 7.074.596 y 8.153.773). En realizaciones adicionales, el ARNm puede comprender una mezcla de nucleótidos no modificados y modificados (ver la publicación de patente estadounidense núm. 2012/0195936). En una realización preferida, el ácido nucleico que codifica la(s) nucleasa(s) se suministra a

HSC/PC a través de electroporación. En algunas realizaciones, la nucleasa escinde en o cerca del sitio de unión de un factor de transcripción. En algunos aspectos, el factor de transcripción es GATA-1.

En la presente memoria se describe una célula o línea celular en la que una secuencia potenciadora de BCL11A endógena se modifica genéticamente mediante una nucleasa como se describe en la presente memoria (p. ej., como se muestra en la Tabla 1), por ejemplo, en comparación con la secuencia natural de la célula. La modificación genética al potenciador de BCL11A resulta en la modificación de la expresión génica de globina (beta y gamma). Las células o las líneas celulares modificadas por nucleasas como se describen en la presente memoria se pueden distinguir por estructura, función y combinaciones de estructura y función de las naturales. Las líneas celulares o célula modificadas genéticamente pueden ser heterocigóticas u homocigóticas con respecto a la modificación. Las modificaciones pueden comprender inserciones (p. ej., inserción transgénica), eliminaciones y combinaciones de inserciones y eliminaciones; tales inserciones, eliminaciones y combinaciones de inserciones y eliminaciones se denominan comúnmente "indels". En algunas realizaciones preferidas, las indels producen la destrucción de un sitio de unión a factor de transcripción. En ciertas realizaciones, la modificación es en o cerca del(de los) sitio(s) de unión a nucleasa(s), sitio(s) de escisión y combinaciones de sitios de unión y escisión, por ejemplo, dentro de 1-300 pares de bases (o cualquier valor entre estos) anteriores o posteriores al(a los) sitio(s) de escisión, más preferiblemente, dentro de 1-100 pares de bases (o cualquier valor entre estos) de cualquier lado del(de los) sitio(s) de unión, sitio(s) de escisión y combinaciones de sitios de unión y escisión que se muestran en la Tabla 1, incluso más preferiblemente, dentro de 1 a 50 pares de bases (o cualquier valor entre estos) a cualquier lado del(de los) sitio(s) de unión, sitio(s) de escisión y combinaciones de sitios de unión y escisión. En ciertas realizaciones, la modificación genética de la secuencia potenciadora de BCL11A está dentro y/o entre las secuencias mostradas en la Tabla 1 (sitios diana). La modificación puede incluir también modificaciones a uno o más nucleótidos en los sitios de escisión. La modificación puede incluir también modificaciones a uno o más nucleótidos en los sitios de unión. La modificación puede incluir, además, modificaciones a uno o más nucleótidos en los sitios de escisión y en uno o más de los sitios de unión. En ciertas realizaciones, uno o más del(de los) sitio(s) diana de nucleasa no está(n) modificado(s). En otras realizaciones, al menos uno de los sitios diana para la(s) nucleasa(s) está(n) modificado(s). En ciertas realizaciones, la modificación es en o cerca de la región "+58" del potenciador de BCL11A, por ejemplo, en o cerca de un sitio de unión a nucleasa que se muestra en cualquiera de la SEQ ID NO: 1 y la SEQ ID NO:2. Cualquier célula o línea celular puede modificarse mediante las nucleasas tales como se describen en la presente memoria, por ejemplo, una célula madre (célula madre hematopoyética tal como una célula madre hematopoyética CD34+) o célula precursora de glóbulos rojos (RBC).

También se describen células o líneas celulares obtenidas después de la modificación por parte de una nucleasa tal como se describe en la presente memoria, por ejemplo, células o líneas celulares descendientes de una célula o línea celular modificada por nucleasa como se describe en la presente memoria. También se proporcionan células diferenciadas parcialmente o completamente descendientes de las células madre modificadas, tales como se describen en la presente memoria (p. ej., RBC o células precursoras de RBC). Las células descendientes de las células modificadas por nucleasas pueden propagarse, diferenciarse y combinaciones de ambos (cultivos) propagados y diferenciados *in vitro* o pueden diferenciarse dentro de un sujeto vivo, por ejemplo, después de la administración *ex vivo* de una célula madre modificada por nucleasa. Cualquiera de las células o líneas celulares modificadas genéticamente descritas en la presente memoria pueden mostrar una mayor expresión de globina gamma. También se proporcionan composiciones tales como composiciones farmacéuticas que comprenden las células modificadas genéticamente tales como se describen en la presente memoria.

El suministro de un ácido nucleico donante a una célula diana puede proporcionar una célula modificada genéticamente en la cual el donante está integrado en la célula. El donante puede suministrarse antes, después o junto con el ácido nucleico que codifica la(s) nucleasa(s) de la Tabla 1. El ácido nucleico donante puede comprender una secuencia exógena (transgén) que se va a integrar en el genoma de la célula, por ejemplo, un locus endógeno. En algunas realizaciones, el donante puede comprender un gen de longitud completa o fragmento de este, flanqueado por regiones de homología con el sitio de escisión al que se dirige. En algunas realizaciones, el donante carece de regiones homólogas y se integra en un locus diana a través de un mecanismo independiente de homología (es decir, NHEJ). El donante puede comprender cualquier secuencia de ácido nucleico, por ejemplo, un ácido nucleico que, al usarse como sustrato para la reparación dirigida por homología de la rotura de cadena doble inducida por nucleasas, lleva a que se genere una eliminación especificada por el donante en el locus cromosómico endógeno (p. ej., región potenciadora de BCL11A) o, de manera alternativa (o adicional), formas alélicas novedosas (p. ej., mutaciones puntuales que extraen un sitio de unión a factor de transcripción) del locus endógeno que se va a crear. En algunos aspectos, el ácido nucleico donante es un oligonucleótido en donde la integración lleva a un evento de corrección génica o una eliminación dirigida.

En otros aspectos, la nucleasa, el donante y las combinaciones de la nucleasa y el donante se suministran mediante métodos de transferencia génica víricos, no víricos y combinaciones de víricos y no víricos. En realizaciones preferidas, el donante se suministra a la célula a través de un virus adenoasociado (AAV, por sus siglas en inglés). En algunas instancias, el AAV comprende LTR que son de un serotipo heterólogo en comparación con el serotipo de la cápside.

En algunos aspectos, las eliminaciones que comprenden regiones dentro de las regiones hipersensibles de ADNsal del potenciador (p. ej., la región +58 del potenciador de BCL11A) se realizan mediante el uso de una o más nucleasas que se muestran en la Tabla 1. Estas eliminaciones pueden comprender de aproximadamente 1 nucleótido a 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, 500, 550 nucleótidos o cualquier valor entre estos. 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.

- 5 En algunos aspectos de la descripción, los dominios de unión a ADN de la Tabla 1 se fusionan con un dominio funcional. Algunos aspectos incluyen la fusión de los dominios de unión al ADN con dominios capaces de regular la expresión de un gen. En algunos aspectos de la descripción, las proteínas de fusión comprenden el dominio de unión al ADN de la Tabla 1 fusionado con un dominio modulador de la expresión génica donde el modulador reprime la expresión génica.

- 10 En algunos aspectos de la descripción, las células HSC/PC se ponen en contacto con las nucleasas, las proteínas de unión a ADN y combinaciones de las nucleasas y proteínas de unión a ADN de la descripción (es decir, las ZFP como se muestran en la Tabla 1). Las nucleasas, las proteínas de unión a ADN y combinaciones de las nucleasas y proteínas de unión a ADN pueden suministrarse como ácidos nucleicos o como proteínas. Los ácidos nucleicos pueden ser ARNm que codifican las proteínas de unión a ADN, y combinaciones de las nucleasas y proteínas de unión a ADN y, en instancias adicionales, los ARNm pueden estar protegidos. El ARNm puede modificarse químicamente, puede comprender un casquete ARCA, una mezcla de nucleótidos no modificados o modificados y combinaciones de un casquete ARCA y una mezcla de nucleótidos modificados y no modificados. También se proporcionan células o líneas celulares derivadas de estas células, incluso células parcialmente o completamente diferenciadas.

- 20 En algunos aspectos, las HSC/PC se ponen en contacto con las nucleasas, las proteínas de unión a ADN y combinaciones de nucleasas y proteínas de unión a ADN de la descripción *ex vivo*, después de la aféresis de las HSC/PC a partir de un sujeto, o purificación de médula ósea cosechada. En algunas realizaciones, las nucleasas descritas en la presente memoria provocan modificaciones dentro de las regiones potenciadoras de BCL11A, por ejemplo, que dan como resultado una célula modificada genéticamente que se diferencia de manera estructural, funcional y combinaciones de diferencias estructurales y funcionales con respecto a las células naturales, otras células modificadas (p. ej., modificadas por nucleasa) y combinaciones de células naturales y otras modificadas. En realizaciones adicionales, las HSC/PC que contienen las modificaciones de la región potenciadora de BCL11A se introducen nuevamente en el sujeto. En algunas instancias, las HSC/PC que contienen las modificaciones de la región potenciadora de BCL11A se expanden antes de la introducción. En otros aspectos, las HSC/PC modificadas genéticamente se administran al sujeto en un trasplante de médula ósea, en donde las HSC/PC se injertan, diferencian y maduran *in vivo*. En algunas realizaciones, las HSC/PC se aíslan a partir del sujeto después de la movilización inducida por G-CSF, movilización inducida por plerixafor, y combinaciones de movilización inducida por G-CSF y plerixafor y, en otras, las células se aíslan a partir de la médula ósea humana o los cordones umbilicales humanos. En algunos aspectos, se trata al sujeto con un procedimiento mieloablato leve antes de la introducción del injerto que comprende las HSC/PC modificadas, mientras que, en otros aspectos, se trata al sujeto con un régimen de acondicionamiento mieloablato vigoroso. En algunas realizaciones, las composiciones de la invención se usan para tratar o prevenir una hemoglobinopatía. En algunos aspectos, la hemoglobinopatía es una talasemia. En algunos aspectos, la hemoglobinopatía es una talasemia beta, mientras que, en otros aspectos, la hemoglobinopatía es anemia de células falciformes.

- 40 En algunas realizaciones, las HSC/PC se ponen en contacto, además, con una molécula donante. En algunas realizaciones, la molécula donante se suministra mediante un vector vírico. La molécula donante puede comprender una o más secuencias que codifican un polipéptido funcional (p. ej., un ADNc o fragmento de este), con o sin un promotor. Pueden incluirse otras secuencias adicionales (secuencias codificantes o no codificantes) cuando se usa una molécula donante para la inactivación, incluso, pero sin limitarse a, secuencias que codifican un péptido 2A, sitio SA, IRES, etc.

- 45 En un aspecto, las HSC/PC se ponen en contacto mediante las composiciones de la invención *in vivo*. Las nucleasas, las proteínas de unión a ADN o las combinaciones de nucleasas y proteínas de unión a ADN se pueden suministrar a las HSC/PC *in situ* mediante métodos conocidos en la técnica. En algunas realizaciones, las nucleasas de la invención comprenden una partícula vírica que se administra al sujeto que lo necesita, mientras que, en otras, las nucleasas, proteínas de unión a ADN o la combinación de nucleasas y proteínas de unión a ADN comprenden una nanopartícula (p. ej., liposoma). En algunas realizaciones, las partículas víricas, nanopartículas o combinación de partículas víricas y nanopartículas se suministran al órgano (p. ej., médula ósea) en donde se encuentran las HSC/PC.

- 50 En otro aspecto, se describen en la presente memoria métodos para integrar un ácido nucleico donante en el genoma de una célula a través de mecanismos independientes de homología. Los métodos comprenden crear una rotura de cadena doble (DSB) en el genoma de una célula y escindir la molécula donante mediante el uso de una nucleasa como se describe en la presente memoria, de modo que se integre el ácido nucleico donante en el sitio de la DSB. En ciertas realizaciones, el ácido nucleico donante se integra a través de métodos no dependientes de homología (p. ej., NHEJ). Como se indicó anteriormente, tras la escisión *in vivo*, las secuencias donantes pueden integrarse de forma dirigida en el genoma de una célula en la ubicación de una DSB. La secuencia donante puede incluir uno o más de los mismos sitios diana para una o más de las nucleasas usadas para crear la DSB. Por lo tanto, la secuencia donante puede escindirse mediante una o más de las mismas nucleasas usadas para escindir el gen endógeno en el cual se desea realizar la integración. En ciertas realizaciones, la secuencia donante incluye diferentes sitios diana de nucleasa de las nucleasas usadas para inducir la DSB. Las DSB en el genoma de la célula diana pueden crearse mediante



cualquier mecanismo. En ciertas realizaciones, la DSB se crea mediante una o más nucleasas de dedos de zinc (ZFN), proteínas de fusión que comprenden un dominio de unión a dedo de zinc, que se genomanipula para unirse a una secuencia dentro de la región de interés, y un dominio de escisión o un dominio medio de escisión.

En un aspecto, el donante puede codificar una proteína reguladora de interés (p. ej., TF de ZFP, TF de TALE o una TF de CRISPR/Cas) que se une, modula la expresión o se une y 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 la proteína reguladora puede modular (es decir, inducir o reprimir) la expresión de un ADN diana.

En algunas realizaciones, la célula HSC/PC transgénica, animal transgénico o combinación de célula HSC/PC transgénica y animal incluye un transgén que codifica un gen humano. En algunas instancias, el animal transgénico comprende una inactivación en el locus endógeno y el remplazo del gen endógeno con su contraparte humano, lo que permite el desarrollo de un sistema *in vivo* donde la proteína humana puede estudiarse de forma aislada. Dichos modelos transgénicos pueden usarse con fines de análisis para identificar moléculas pequeñas o biomoléculas grandes, u otras entidades que pueden interactuar con la proteína humana de interés o modificarla. En algunos aspectos, el transgén se integra en el locus seleccionado (p. ej., seguro) en una célula madre (p. ej., una célula madre embrionaria, una célula madre pluripotente inducida, una célula madre hematopoyética, etc.) o el embrión animal no humano obtenido mediante cualquiera de los métodos descritos en la presente memoria y después el embrión se implanta de modo que nazca un animal no humano vivo. Después, se cría al animal hasta alcanzar su madurez sexual y se le permite tener crías, en donde al menos algunas de las crías comprenden la secuencia génica endógena editada o el transgén integrado.

En otro aspecto, se proporciona en la presente memoria un método para alterar la expresión génica (p. ej., BCL11A, un gen de globina y combinaciones de BCL11A y un gen de globina) en una célula, el método comprende: introducir, en la célula, una o más nucleasas descritas en la presente memoria (que se muestran en la Tabla 1), en condiciones tales que se exprese la una o más proteínas y se altere la expresión del gen. En ciertas realizaciones, se altera (p. ej., aumenta) la expresión de un gen de globina (p. ej., globina gamma o globina beta). Cualquiera de los métodos descritos en la presente memoria puede comprender, además, integrar una secuencia donante (p. ej., transgén o fragmento de este bajo el control de un promotor exógeno o endógeno) en el genoma de la célula, por ejemplo, integrar un donante en o cerca del sitio de escisión por nucleasa en el gen de BCL11A. La secuencia donante se introduce en la célula al usar un vector vírico, como un oligonucleótido, en un plásmido y combinaciones de uno o más métodos seleccionados del uso de un vector vírico, como un oligonucleótido o en un plásmido. La célula en la que se altera la expresión génica puede ser, por ejemplo, una célula precursora de glóbulos rojos (RBC), una célula madre hematopoyética (p. ej., célula CD34+) y combinaciones de la célula precursora de RBC y una célula madre hematopoyética.

En la presente memoria se describe un método para producir una célula modificada genéticamente que comprende una modificación genómica dentro de una secuencia potenciadora de BCL11A endógena (una modificación en la secuencia de nucleótidos de la secuencia potenciadora de BCL11A), el método comprende las etapas de: a) poner en contacto una célula con un polinucleótido (p. ej., ADN o ARNm) que codifica una nucleasa de dedos de zinc que comprende 4, 5 o 6 dominios de dedos de zinc en los que cada uno de los dominios de dedos de zinc comprende una región de hélices de reconocimiento en el orden que se muestra en una fila individual de la Tabla 1; b) someter a la célula a condiciones que lleven a la expresión de la proteína de dedos de zinc del polinucleótido; y c) modificar la secuencia potenciadora de BCL11A endógena con la proteína de dedos de zinc expresada de manera suficiente para producir la célula modificada genéticamente. Las células se pueden estimular con al menos una citocina (p. ej., antes de la etapa (a)). El polinucleótido puede ponerse en contacto con la célula mediante el uso de cualquier método adecuado, que incluye, pero no se limita a, mediante transfección, mediante el uso de un vector no vírico, mediante el uso de un vector vírico, a través de medios químicos o mediante exposición a un campo eléctrico (p. ej., electroporación).

También se proporcionan células que comprenden una o una combinación de las modificaciones genómicas descritas en la presente memoria, incluidas células que descenden de las células producidas mediante los métodos descritos en la presente memoria.

También se describe un método para tratar a un paciente que necesita un aumento de la expresión génica de globina, el método comprende administrar al paciente la preparación farmacéutica, en donde la preparación farmacéutica comprende células genéticamente modificadas, proteínas, polinucleótidos y combinaciones de uno o más seleccionados de células genéticamente modificadas, proteínas y polinucleótidos, como se describen en la presente memoria, en una cantidad suficiente para aumentar la expresión génica de globina en el paciente. Se puede saber que el paciente padece, se sospecha que padece o está en riesgo de desarrollar una talasemia o anemia de células falciformes.

También se proporciona un kit, que comprende ácidos nucleicos, proteínas, células genéticamente modificadas, y combinaciones de uno o más seleccionados de ácidos nucleicos, proteínas y células genéticamente modificadas de la invención. El kit puede comprender ácidos nucleicos que codifican las nucleasas, (p. ej., moléculas de ARN o genes que codifican el sistema de ZFN, TALEN o CRISPR/Cas contenidos en un vector de expresión adecuado), alícuotas

de las proteínas de nucleasa, moléculas donantes, modificadores adecuados de la autorrenovación de células madre ("capacidad de célula madre"), células, tampones, instrucciones (p. ej., para llevar a cabo los métodos de la invención) y similares, incluidas varias combinaciones de estos componentes del kit. Se describe una célula modificada genéticamente (p. ej., célula madre tal como una célula madre hematopoyética (CD34+) o célula precursora de RBC) que comprende al menos una modificación genómica realizada mediante una nucleasa (p. ej., como se muestra en una fila individual de la Tabla 1), en donde la modificación genómica se produce dentro de una secuencia potenciadora de BCL11A endógena y, además, en donde la modificación genómica se selecciona del grupo que consiste en inserciones, eliminaciones y combinaciones de estas y comprende una modificación en, cera o entre cualquiera de la SEQ ID NO: 1 y la SEQ ID NO:2. En ciertas realizaciones, la célula es una célula diferenciada modificada genéticamente que descende de una célula madre como se describe en la presente memoria (p. ej., un RBC que descende de una célula madre hematopoyética o célula precursora de RBC).

La nucleasa puede comprender al menos una nucleasa de dedos de zinc (ZFN) (p. ej., como se muestra en la Tabla 1), al menos un TALEN, y combinaciones de al menos una ZFN y al menos un TALEN. La(s) nucleasa(s) puede(n) introducirse en la célula en forma de proteína, como un polinucleótido que codifica la(s) nucleasa(s), o como una combinación de forma de proteína y polinucleótido que codifica la(s) nucleasa(s). En ciertas realizaciones, la modificación genómica comprende una inserción que comprende la integración de un polinucleótido donante que codifica un transgén. También se describen composiciones farmacéuticas que comprenden una o más de las células modificadas genéticamente como se describen en la presente memoria.

También se describe una proteína de unión a ADN que comprende una proteína de dedos de zinc que comprende 4, 5 o 6 dominios de dedos de zinc que comprenden una región de hélice de reconocimiento, en donde las proteínas de dedos de zinc comprenden las regiones hélice de reconocimiento en el orden que se muestra en una fila individual de la Tabla 1. También se describe una proteína TALE que comprende una pluralidad de repeticiones que se unen a una secuencia que comprende una porción (p. ej., al menos 4, 5, 6 o más) de pares de bases de los sitios diana que se muestran en la Tabla 1. También se describe una proteína de fusión que comprende una proteína de dedos de zinc o proteína TALE como se describe en la presente memoria y un dominio de escisión o dominio medio de escisión natural o genomanipulado, así como los polinucleótidos que codifican las proteínas (ZFP, TALE, ZFN, TALEN) como se describen en la presente memoria. También se describen células (p. ej., células madre aisladas tales como células madre hematopoyéticas (CD34+)) que comprenden uno o más polinucleótidos, proteínas y combinaciones de polinucleótidos y proteínas como se describen en la presente memoria. También se describen kits que comprenden una o más proteínas, polinucleótidos, células o combinaciones de estos como se describen en la presente memoria.

También se describe un método para alterar la expresión génica de globina en una célula (p. ej., célula precursora de RBC, célula madre hematopoyética y combinaciones de célula precursora de RBC y célula madre hematopoyética), el método comprende: introducir, en la célula, uno o más polinucleótidos que codifican una o más nucleasas como se describen en la presente memoria, en condiciones tales que se exprese la una o más proteínas y se altere (p. ej., aumente) la expresión del gen de globina (p. ej., globina gamma, globina beta y combinaciones de globina gamma y beta). Los métodos pueden comprender, además, integrar una secuencia donante en el genoma de la célula, por ejemplo, mediante el uso de un vector vírico, como un oligonucleótido o en un plásmido. La secuencia donante puede comprender un transgén bajo el control de un promotor endógeno o exógeno.

También se describe un método para producir una célula modificada genéticamente que comprende una modificación genómica dentro de una secuencia potenciadora de BCL11A endógena (p. ej., un sitio diana como se muestra en la Tabla 1), el método comprende las etapas de: (a) poner en contacto una célula con un polinucleótido que codifica una proteína de fusión que comprende una nucleasa de dedos de zinc que comprende 4, 5 o 6 dominios de dedos de zinc en los que cada uno de los dominios de dedos de zinc comprende una región de hélices de reconocimiento en el orden que se muestra en una fila individual de la Tabla 1; (b) someter a la célula a condiciones que lleven a la expresión de la proteína de fusión a partir del polinucleótido; y (c) modificar la secuencia potenciadora de BCL11A endógena con la proteína de fusión expresada de manera suficiente para producir la célula modificada genéticamente. El método puede comprender, además, estimular a las células con al menos una citocina. El(los) polinucleótido(s) puede(n) suministrarse dentro de la célula, por ejemplo, al usar un sistema de suministro no vírico, un sistema de suministro vírico, un vehículo de suministro y combinaciones que se seleccionan de un sistema de suministro no vírico, un sistema de suministro vírico y un vehículo de suministro y puede comprender someter las células a un campo eléctrico o emplear la alteración de la membrana celular como un mecanismo de suministro (llamado "Tecnología Squeeze", ver p. ej., Sharei et al. (2015) PLOS ONE doi: 10.1371/journal.pone.0118803).

También se describen métodos para tratar a un paciente que necesita un aumento de la expresión génica de globina (p. ej., un paciente que se sabe que padece, se sospecha que padece o está en riesgo de desarrollar una hemoglobinopatía tal como una talasemia (p. ej.,  $\beta$ -talasemia) o anemia de células falciformes, el método comprende administrar al paciente la composición farmacéutica como se describe en la presente memoria (p. ej., proteínas, polinucleótidos, células o una combinación seleccionada de proteínas, polinucleótidos y células) en una cantidad suficiente para aumentar la expresión génica de globina en el paciente.

Estos y otros aspectos serán fácilmente evidentes para el experto en la técnica en virtud de la descripción en su totalidad.

## BREVE DESCRIPCIÓN DE LOS DIBUJOS

Las **Figuras 1A y 1B** son esquemas que representan una vista general del ensayo de sitio de integración de estructura doble oligonucleotídica. La Figura 1 muestra la primera etapa del ensayo, en la cual las células se tratan con ZFN en presencia de ADN de estructura doble oligonucleotídica complementario, que se captura en una fracción de los eventos de escisión resultantes. La Figura 1B muestra la etapa posterior en la cual las células se cultivan durante siete días, el ADN genómico se aísla y los segmentos de los sitios de integración de donantes flanqueadores del genoma se amplifican a través de PCR mediada por adaptadores al usar un cebador para la estructura doble oligonucleotídica integrada. Después se secuencian los amplicones para revelar los sitios de escisión candidatos.

La **Figura 2** representa la ubicación de 455 loci inespecíficos potenciales identificados mediante el ensayo de sitio de integración de estructura doble oligonucleotídica al usar el par de ZNF 51857/51949 original o principal. Los 63 loci principales están resaltados en gris y se analizaron en análisis de indel de seguimiento. Cada locus se identifica por el cromosoma y el número de base que indica la ubicación más probable de escisión, así como una indicación de la cantidad de integrantes detectados en dicho locus.

Las **Figuras 3A a 3C** representan residuos de contacto con fosfato ilustrativos dentro de la estructura de los dedos de zinc. La Figura 3A representa una interacción de dedos de zinc con una molécula de ADN y muestra la ubicación de la cadena lateral de arginina natural y cómo interactúa con la estructura de fosfatos de la molécula de ADN. La Figura 3B (SEQ ID NOs:7 a 17) representa secuencias de ZFN ilustrativas que muestran dónde está ubicada la arginina en la secuencia primaria de cada dedo de zinc (fila vertical de residuos "R" indicada con una flecha en negrita) y también destaca las argininas que se sustituyeron con glutamina en la estructura de ZFN para eliminar el contacto de fosfato correspondiente (residuos "R" recuadrados individualmente). Las secuencias también muestran las regiones de hélices de reconocimiento (donde los residuos -1 a +3, +5 y +6 están en recuadros y sombreados), así como una porción del enlazador entre el dominio de dedos de zinc del extremo C y el dominio de escisión (el dominio de escisión no se muestra). La Figura 3C representa, además, la ubicación espacial de otra estructura potencial que entra en contacto con la cadena lateral de lisina presente en el dominio de escisión de FokI, que durante la optimización de la especificidad puede sustituirse con una serina.

La **Figura 4** es una gráfica que compara los niveles de modificación a través del tratamiento de células CD34 con el par 51857/51949 original o con el par 63014/65722 optimizado en varios loci analizados. A lo largo de la parte inferior o en el eje x están los loci identificados como dianas de escisión potenciales, con el porcentaje de indels para cada sitio indicado en el eje vertical o y. Cabe señalar que el eje y se muestra en una escala logarítmica. Las barras en gris oscuro muestran los loci escindidos por el par 51857/51949 y la cantidad de escisión detectada, mientras que las barras en gris claro son los loci escindidos por el par optimizado donde casi toda la escisión medida está en la secuencia diana de BCL11a diana.

Las **Figuras 5A y 5B** son gráficas que representan las relaciones relativas de ARNm de globina elaborados en células hCD34+ después del tratamiento con ZFN específicos de BCL11A y diferenciación eritroide. Las células CD34+ derivadas de dos donantes humanos sanos (PB-MR-003 o PB-MR-004) se trataron o no se trataron con el par de ZFN y después se analizó la expresión de globina  $\alpha$ ,  $\beta$  y  $\gamma$ . El mejor método para determinar la cantidad de ARNm de globina  $\gamma$  encontrada después del tratamiento con ZFN es expresar el cambio en la expresión como una relación entre globina  $\gamma$  y globina  $\beta$  (Figura 5A) o entre ARNm de globina  $\gamma$  y globina  $\alpha$  (Figura 5B).

La **Figura 6** es una gráfica que representa las cantidades relativas de proteína globina  $\gamma$  producida en las células CD34+ tratadas. Como se indicó anteriormente, se usaron dos lotes de células CD34+ derivadas de donantes humanos sanos (PB-MR-003 y PB-MR-004). En este experimento, se observó una elevación de aproximadamente 3-4 veces de los porcentajes de proteína globina fetal hasta niveles de aproximadamente 15 %-20 % en la progenie eritroide de HSPC después de la alteración mediada por 63014/65722 del potenciador de BCL11A en ambos lotes de donantes.

La **Figura 7** es una gráfica que indica el quimerismo humano relativo en ratones injertados con lotes de donantes tratados con 63014/65722 ("ZFN") como se describió anteriormente. El quimerismo humano se midió a través de la detección de células con un marcador hCD45 en su superficie al usar FACS. Se indican los porcentajes de células hCD45+ humanas en la sangre periférica recogida en las 8 o 12 semanas después del trasplante. Los datos mostraron buenos niveles de injerto en este estudio con un quimerismo humano semejante después del injerto de HSPC testigo sin transfectar ("(-)") y transfectada con ZFN ("ZFN"). Los círculos y triángulos abiertos representan animales individuales.

La **Figura 8** muestra el porcentaje de quimerismo detectado en la médula ósea de los ratones injertados donde las células humanas se identificaron mediante la presencia de hCD45 en sus superficies celulares. Las muestras se analizaron 12 semanas después del injerto.

Las **Figuras 9A a 9D** son gráficas que representan la reconstitución de varios linajes de células hematopoyéticas analizadas por análisis FACS de células de médula ósea en los ratones injertados obtenidos en la semana 12 con anticuerpos que reconocen marcadores de superficie celular específicos del linaje. Los datos mostraron una representación semejante de todos los linajes hematopoyéticos humanos analizados en la médula ósea en la semana

12 después de la inyección entre la progenie de células CD34+ tratadas con ARNm de ZFN específico para BCL11A ("14/22") y la de células sin transfectar ("(-)"). Se muestran datos de linfoides, mieloides, eritroides y HSPC (Figuras 9A a 9D, respectivamente) para células derivadas de ambos donantes ("003" y "004"). Los datos mostraron una representación semejante de todos los linajes hematopoyéticos humanos analizados en la médula ósea en la semana 12 después de la inyección entre la progenie de células CD34+ tratadas con ARNm de ZFN de Bcl11A y la de células sin transfectar.

La **Figura 10** es una gráfica que representa el nivel de modificación génica en la diana BCL11A en ADN aislado a partir de sangre periférica del ratón injertado, sometido a ensayo mediante secuenciación profunda. Se muestran los datos para las células de entrada (2 días después de la transfección de ZFN, ("+")), y después para células sanguíneas 8 o 12 semanas después del injerto y demostraron una buena retención de la modificación génica. Las células sin transfectar se representan mediante "(-)" en la línea de Tratamiento.

La **Figura 11** es una gráfica que representa la cantidad de modificación génica en la diana de BCL11A para las muestras de célula de médula ósea después del injerto de las células tratadas con ZFN ("+"). Las células sin tratar se representan mediante "(-)" en la línea de Tratamiento. Se observó una modificación semejante en los linajes dependientes de BCL11A (linfocitos B, que expresan el marcador CD19; progenitores primitivos que expresan CD45 y niveles altos de CD38) y linajes independientes de BCL11A (mieloides). Si bien los niveles de modificación génica de entrada fueron mayores en la muestra de donante PB-MR-003 que en la muestra de donante PB-MR-004, las células derivadas de PB-MR-004 mostraron de manera constante niveles de modificación más altos, es decir, una mejor retención de la modificación, en ratones que los derivados de PB-MR-003.

La **Figura 12** es una gráfica que representa la cantidad de modificación génica en células eritroides derivadas de células de médula ósea en la semana 12 después de la diferenciación *in vitro* durante 14 días. Los datos fueron de ratones injertados originalmente con los dos donantes diferentes descritos anteriormente, y demostraron que la modificación de BCL11A mediada por el tratamiento con ZFN ("+" ZFN) no se altera de manera marcada durante la diferenciación eritroide. Las células que no se trataron con ZFN se indican mediante "(-) ZFN".

Las **Figuras 13A y 13B** son gráficas que representan la cantidad relativa de ARNm que codifica globina  $\gamma$ , donde la concentración de ARNm de globina  $\gamma$  se representa como una relación entre el ARNm de globina  $\gamma$ /globina  $\beta$  (Figura 9A) o como una relación entre el ARNm de globina  $\gamma$ /globina  $\alpha$  (Figura 9B). Las relaciones entre ARNm de globina  $\gamma$  y globina  $\beta$  o globina  $\gamma$  y globina  $\alpha$  de las muestras tratadas con ZFN ("+" ZFN) y las sin transfectar ("(-) ZFN") difirieron ampliamente entre las progenies eritroides de ratones individuales del mismo grupo. Sin embargo, a pesar de esta variabilidad y la variabilidad introducida por el uso de dos donantes humanos diferentes, los promedios de la muestra tratada con 63014/65722 muestran un aumento de ~1,5-2 veces en los niveles de ARNm de globina  $\gamma$  en comparación con sus contrapartes respectivas no transfectadas.

La **Figura 14** es una gráfica que representa la diferencia en la cantidad de proteína globina  $\gamma$  (expresado como una relación entre globina  $\gamma$ /globina  $\alpha$  o globina  $\gamma$ /proteína total tipo  $\beta$ ) en células derivadas de médula ósea a partir de ratones injertados donde las células de médula ósea se sometieron a un protocolo de diferenciación *in vitro*. Se midieron los niveles de proteína a los 16 días de iniciada la diferenciación. Se determinaron las relaciones entre globina gamma ( $\gamma$ ) (suma de los picos de Agamma y Ggamma) y globina alfa ( $\alpha$ ), así como las relaciones entre globina gamma (suma de los picos de Agamma y Ggamma)/con respecto a la globina tipo beta (suma de los picos de Agamma, Ggamma, globina beta y delta). En consonancia con la escasa diferenciación de eritroides de las muestras derivadas de PB-MR-003, los niveles de globina gamma en las células no transfectadas derivadas de este donante fueron muy altos (~30 %) y, por lo tanto, el tratamiento con ZFN ("+" ZFN) dio como resultado un aumento de solo 1,2 veces en los niveles de globina gamma en comparación con las células no tratadas ("(-) ZFN"). PB-MR-004 mostró niveles más típicos no transfectados (~9 %) y exhibió un aumento de ~2 veces en los niveles de proteína globina gamma después de 12 semanas de pasaje por el ratón.

#### DESCRIPCIÓN DETALLADA

Se describen en la presente memoria composiciones y métodos para genomanipular el genoma para la modulación de BCL11A, globina gamma, y combinaciones de BCL11A y la expresión de globina gamma y para el tratamiento, prevención o tratamiento y prevención de hemoglobinopatías. En particular, a través del direccionamiento con nucleasas que comprenden las ZFP que poseen las regiones de hélices de reconocimiento como se muestran en una fila individual de la Tabla 1, se logra de forma eficaz la alteración de un potenciador de BCL11A en HSC/PC y se produce un cambio en la expresión de globina gamma relativa durante la eritrocitopoyesis posterior. Esta modulación de BCL11A y la expresión de globina gamma particularmente útil para el tratamiento de hemoglobinopatías (p. ej., talasemias beta, anemia de células falciformes), en donde hay una expresión insuficiente de globina beta o expresión de una forma mutada de globina beta. Mediante el uso de las composiciones de la invención, las complicaciones y secuelas relacionadas con la enfermedad provocadas por la globina beta aberrante pueden superarse mediante la alteración de la expresión de la globina gamma en células precursoras de eritrocitos.

## General

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

## Definiciones

Los términos "ácido nucleico", "polinucleótido" y "oligonucleótido" se usan de manera intercambiable y se refieren a un polímero de desoxirribonucleótidos o ribonucleótidos, en conformación lineal o circular, y en forma monocatenaria o bicatenaria. A los efectos de la presente descripción, estos términos no se deben interpretar como limitantes con respecto a la longitud de un polímero. Los términos pueden abarcar análogos conocidos de nucleótidos naturales, así como nucleótidos que están modificados en los restos de base, azúcar, fosfato (p. ej., estructuras de fosforotioato) y combinaciones seleccionadas de restos de base, azúcar y fosfato. En general, un análogo de un nucleótido particular posee la misma especificidad de pares de base; es decir, un análogo de A formará un par de bases con T.

Los términos "polipéptido", "péptido" y "proteína" se usan de manera intercambiable para hacer referencia a un polímero de residuos aminoacídicos. El término también se aplica a polímeros de aminoácidos en los cuales uno o más aminoácidos son análogos químicos o derivados modificados de aminoácidos de origen natural correspondientes.

"Unión" se refiere a una interacción no covalente específica de la secuencia entre macromoléculas (p. ej., entre una proteína y un ácido nucleico). No todos los componentes de una interacción de unión deben ser específicos para la secuencia (p. ej., los residuos de contacto con fosfato en una estructura de ADN), siempre y cuando la interacción en su totalidad sea específica para la secuencia. Dichas interacciones se caracterizan generalmente por una constante de disociación (Kd) de  $10^{-6}$  M<sup>-1</sup> o menor. "Afinidad" se refiere a la intensidad de la unión: una mayor afinidad de unión se correlaciona con una Kd menor.

Una "proteína de unión" es una proteína capaz de unirse a otra molécula. Una proteína de unión se puede unir a, por ejemplo, una molécula de ADN (una proteína de unión a ADN), una molécula de ARN (una proteína de unión a ARN), una molécula de proteína (una proteína de unión a proteína) o se puede unir a una combinación de moléculas seleccionadas de una molécula de ADN, una molécula de ARN o una proteína. En el caso de una proteína de unión a proteína, se puede unir a sí misma (para formar homodímeros, homotrímeros, etc.), se puede unir a una o más moléculas de una proteína o proteínas diferentes o se puede unir a sí misma y 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. Por ejemplo, las proteínas de dedos de zinc tienen actividad de unión al ADN, unión al ARN y unión a proteínas.

Una "proteína de unión a ADN de dedos de zinc" (o dominio de unión) es una proteína o un dominio dentro de una proteína mayor que se une a ADN de una forma 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 mediante la coordinación de un ion zinc. El término proteína de unión a ADN de dedos de zinc a menudo se abrevia como proteína de dedos de zinc o ZFP.

Un "dominio de unión a ADN de TALE" o "TALE" es un polipéptido que comprende uno o más dominios/unidades de repetición de TALE. Los dominios de repetición están implicados en la unión de TALE a su secuencia de ADN diana cognada. Una "unidad de repetición" individual (que también se denomina "repetición") típicamente tiene 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.

Los dominios de unión de dedo de zinc y TALE pueden "genomanipularse" para unirse a una secuencia de nucleótidos predeterminada, por ejemplo, a través de la genomanipulación (alteración de uno o más aminoácidos) de la región de hélices de reconocimiento de una proteína TALE o dedo de zinc de origen natural. Por lo tanto, las proteínas de unión al ADN (dedos de zinc o TALE) genomanipuladas son proteínas de origen no natural. Los ejemplos no limitantes de métodos para genomanipular proteínas de unión a ADN son el diseño y la selección. Una proteína de unión a ADN diseñada es una proteína que no se produce en la naturaleza cuyo diseño/composición es principalmente el resultado de criterios racionales. Los criterios racionales para el diseño incluyen la aplicación de reglas de sustitución y algoritmos computarizados para procesar la información en una base de datos que almacena la información de diseños y datos de unión de TALE y/o ZFP existentes. Ver, por ejemplo, las patentes estadounidenses núms. 6.140.081; 6.453.242; 6.534.261 y 8.585.526; ver también las publicaciones de PCT núms. WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 y WO 03/016496.

Una proteína de dedos de zinc o TALE "seleccionada" es una proteína que no se encuentra en la naturaleza, cuya producción es principalmente el resultado de un proceso empírico, tal como visualización en fagos, trampa de interacción o selección híbrida. Ver *p. ej.*, las patentes estadounidenses núms. 5.789.538; 5.925.523; 6.007.988; 6.013.453; 6.200.759; 8.586.526; publicación de PCT núms. WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197, WO 02/099084.

"TtAgo" es una proteína Argonauta procariota que se considera que participa en el silenciamiento génico. TtAgo se deriva de la bacteria *Thermus thermophilus*. Ver, *p. ej.*, Swarts *et al.*, *ibid*, G. Sheng *et al.* (2013) Proc. Natl. Acad. Sci. U.S.A. 111, 652). Un "sistema de TtAgo" implica todos los componentes necesarios que incluyen, por ejemplo, ADN guías para escisión mediante una enzima TtAgo.

"Recombinación" se refiere a un proceso de intercambio de información genética entre dos polinucleótidos, que incluye, pero no se limita a, la captura de donantes mediante la unión de extremos no homóloga (NHEJ) y la recombinación homóloga. A los efectos de la presente descripción, "recombinación homóloga (HR, por sus siglas en inglés)" se refiere a la forma especializada de dicho intercambio que se produce, por ejemplo, durante la reparación de roturas de cadena doble en células mediante mecanismos de reparación dirigidos por homología. Este proceso requiere homología de secuencia de nucleótidos, usa una molécula "donante" para reparar con plantilla una molécula "diana" (es decir, la que experimentó la rotura de cadena doble), y se conoce, de manera diversa, como "conversión génica no cruzada" o "conversión génica de tramo corto", porque lleva a la transferencia de información genética desde el donante a la diana. Sin ánimo de ceñirse a ninguna teoría en particular, dicha transferencia puede implicar la corrección de un malapareamiento de ADN de estructura heterodoble que se forma entre la diana rota y el donante, "hibridación de cadena dependiente de la síntesis", donde se usa el donante para volver a sintetizar la información genética que se volverá parte de la diana, procesos relacionados o combinaciones de estos. Dicha HR especializada a menudo produce una alteración de la secuencia de la molécula diana de forma tal que parte o la totalidad de la secuencia del polinucleótido donante se incorpora al polinucleótido diana.

En los métodos de la descripción, una o más nucleasas dirigidas como se describen en la presente memoria crean una rotura de cadena doble (DSB) en la secuencia diana (*p. ej.*, cromatina celular) en un sitio predeterminado. La DSB puede producir indels mediante reparación dirigida por homología o mediante mecanismos de reparación no dirigida por homología. Las eliminaciones pueden incluir cualquier cantidad de pares de bases. De manera similar, las inserciones pueden incluir cualquier cantidad de pares de bases que incluyen, por ejemplo, la integración de un polinucleótido "donante", que opcionalmente posee homología con respecto a la secuencia de nucleótidos en la región de la rotura. La secuencia donante puede integrarse físicamente o, de manera alternativa, se usa el polinucleótido donante como plantilla para reparar la rotura por medio de recombinación homóloga, lo cual da como resultado la introducción de la totalidad o parte de la secuencia de nucleótidos, como en el donante en la cromatina celular. Por lo tanto, puede alterarse una primera secuencia en la cromatina celular 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" representan el reemplazo de una secuencia de nucleótidos con otra (es decir, el reemplazo de una secuencia en sentido informativo) y no requiere necesariamente el reemplazo físico o químico de un polinucleótido con otro.

En cualquiera de los métodos descritos en la presente memoria, pueden usarse pares adicionales de proteínas de dedos de zinc o TALEN para la escisión adicional de cadena doble de sitios diana adicionales dentro de la célula.

Cualquiera de los métodos descritos en la presente memoria puede usarse para la inserción de un donante de cualquier tamaño o la inactivación parcial o completa de una o más secuencias diana en una célula, mediante la integración dirigida de la secuencia donante que altera la expresión del(de los) gen(es) de interés. También se describen líneas celulares con genes inactivados parcialmente o completamente.

En cualquiera de los métodos descritos en la presente memoria, la secuencia de nucleótidos exógena (la "secuencia donante" o "transgén") puede contener secuencias que son homólogas, pero no idénticas, a secuencias genómicas en la región de interés, mediante lo cual se estimula 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 donante que son homólogas a secuencias en la región de interés exhiben una identidad de secuencia entre aproximadamente 80 a 99 % (o cualquier número entero entre estos) con la secuencia genómica que se reemplaza. En otras realizaciones, la homología entre la secuencia donante y la genómica es mayor que 99 %, por ejemplo, si solo 1 nucleótido difiere entre las secuencias donante y genómica de más de 100 pares de base contiguos. En ciertos casos, una porción no homóloga de la secuencia donante puede contener secuencias que no están presentes en la región de interés, de manera que se introducen secuencias nuevas en la región de interés. En estas instancias, la secuencia no homóloga está generalmente flanqueada por secuencias de 50-1000 pares de bases (o cualquier valor entero entre estos) o cualquier cantidad de pares de bases mayor que 1000, que son homólogas o idénticas a las secuencias en la región de interés. En otras realizaciones, la secuencia donante no es homóloga con respecto a la primera secuencia y se inserta en el genoma mediante mecanismos de recombinación no homóloga.

La "escisión" hace referencia a la ruptura de la estructura covalente de una molécula de ADN. La escisión puede iniciarse mediante una variedad de métodos que incluyen, pero no se limitan a, la hidrólisis enzimática o química de un enlace fosfodiéster. Son posibles tanto la escisión monocatenaria como la escisión bicatenaria, y la escisión

bicatenaria puede producirse como resultado de dos eventos de escisión monocatenaria distintos. La escisión de ADN puede generar la producción de extremos romos o extremos escalonados. En ciertas realizaciones, se usan polipéptidos de fusión para la escisión de ADN bicatenario dirigida.

5 Un "medio dominio de escisión" es una secuencia polipeptídica que, junto con un segundo polipéptido (ya sean idénticos o distintos) forma un complejo con actividad de escisión (preferiblemente actividad de escisión bicatenaria). Los términos "primer y segundo dominios medios de escisión"; "dominios medios de escisión + y -" y "dominios medios de escisión derecho e izquierdo" se usan de manera intercambiable para hacer referencia a los pares de dominios medios de escisión que se dimerizan.

10 Un "dominio medio de escisión genomanipulado" es un dominio medio de escisión que se modificó para formar heterodímeros estrictos con otro dominio medio de escisión (p. ej., otro dominio medio de escisión genomanipulado). Ver, también, las patentes estadounidenses núms. 7.888.121; 7.914.796; 8.034.598; 8.623.618 y la publicación de patente estadounidense núm. 2011/0201055.

15 El término "secuencia" se refiere a una secuencia de nucleótidos de cualquier longitud, que puede ser de ADN o ARN; puede ser lineal, circular o ramificada y puede ser monocatenaria o bicatenaria. 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, por ejemplo, entre 2 y 100.000.000 nucleótidos de longitud (o cualquier valor entero entre estos o por encima de estos), preferiblemente, entre aproximadamente 100 y 100.000 nucleótidos de longitud (o cualquier entero entre estos), más preferiblemente, entre aproximadamente 2000 y 20.000 nucleótidos de longitud (o cualquier valor entre estos) e incluso más preferiblemente, entre aproximadamente 5 y 15 kb (o cualquier valor entre estos).

20 "Cromatina" es la estructura de nucleoproteína que comprende el genoma celular. La cromatina celular comprende ácido nucleico, principalmente ADN, y proteína, incluidas histonas y proteínas cromosómicas distintas de histonas. La mayor parte 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 de cada una de las histonas H2A, H2B, H3 y H4; y ADN enlazador (de longitud variable, dependiendo del organismo) que se  
25 extiende entre los núcleos de nucleosoma. Una molécula de histona H1 generalmente se asocia con el ADN enlazador. A los efectos de la presente descripción, se pretende que el término "cromatina" abarque todos los tipos de nucleoproteína celular, tanto procariota como eucariota. La cromatina celular incluye cromatina cromosómica y episómica.

30 Un "cromosoma" es un complejo de cromatina que comprende la totalidad 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 comprenden el genoma de la célula. El genoma de una célula puede comprender uno o más cromosomas.

Un "episoma" es un ácido nucleico replicante, complejo de nucleoproteína u otra estructura que comprende un ácido nucleico que no es parte del cariotipo cromosómico de una célula. Los ejemplos de episomas incluyen plásmidos y ciertos genomas víricos.

35 Una "región accesible" es un sitio en la cromatina celular en el cual un sitio diana presente en el ácido nucleico puede unirse mediante una molécula exógena que reconoce el sitio diana. Sin ánimo de ceñirse a ninguna teoría en particular, se cree que una región accesible es una que no se empaqueta en una estructura nucleosómica. La estructura distintiva de una región accesible a menudo puede detectarse por su sensibilidad a sondas químicas y enzimáticas, por ejemplo, nucleasas.

40 Un "sitio diana" o "secuencia diana" es una secuencia de ácido nucleico que define una porción de un ácido nucleico a la cual se unirá una molécula de unión, siempre y cuando existan condiciones suficientes para la unión.

45 Una molécula "exógena" es una molécula que no está normalmente presente en una célula, pero puede introducirse en una célula mediante uno o más métodos genéticos, bioquímicos o de otro tipo. La "presencia normal en la célula" se determina con respecto a la etapa de desarrollo y condiciones ambientales particulares de la célula. Por lo tanto, por ejemplo, una molécula que está presente solo durante el desarrollo embrionario muscular es una molécula exógena con respecto a una célula muscular adulta. De manera similar, una molécula inducida por un choque térmico es una molécula exógena con respecto a una célula que no es de choque térmico. Una molécula exógena puede comprender, por ejemplo, una versión funcional de una molécula endógena con mal funcionamiento o una versión con mal funcionamiento de una molécula endógena que funciona normalmente.

50 Una molécula exógena puede ser, entre otras cosas, una molécula pequeña, tal como una generada 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 mono o bicatenarios; pueden ser lineales, ramificados o circulares; y pueden ser de cualquier longitud. Los ácidos  
55 nucleicos incluyen aquellos capaces de formar estructuras dobles, así como ácidos nucleicos que forman estructuras triples. Ver, por ejemplo, las patentes estadounidenses núms. 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, desmetilasas, acetilasas, desacetilasas, cinasas, fosfatasas, integrasas, recombinasas, ligasas, topoisomerasas, girasas y helicasas.

Una molécula exógena puede ser el mismo tipo de molécula que una molécula endógena, p. ej., una proteína o ácido nucleico exógeno. Por ejemplo, un ácido nucleico exógeno puede comprender un genoma vírico infectante, un plásmido o episoma introducido en una célula, o un cromosoma que no está normalmente presente en la célula. Los métodos para la introducción de moléculas exógenas en células son conocidos por los expertos en la técnica e incluyen, pero no se limitan a, transferencia mediada por lípidos (es decir, liposomas, incluidos lípidos neutros y catiónicos), electroporación, inyección directa, fusión celular, bombardeo de partículas, coprecipitación de fosfato cálcico, transferencia mediada por DEAE-dextrano y transferencia mediada por vector vírico. Una molécula exógena también puede ser el mismo tipo de molécula que una molécula endógena, pero derivada de una especie diferente de la que deriva la célula. Por ejemplo, se puede introducir una secuencia de ácido nucleico humana en una línea celular derivada originalmente de un ratón o hámster.

En cambio, una molécula "endógena" es una que normalmente está presente en una célula particular en una etapa de desarrollo particular en condiciones ambientales particulares. Por ejemplo, un ácido nucleico endógeno puede comprender un cromosoma, el genoma de una mitocondria, cloroplasto u otro organelo, o un ácido nucleico episómico de origen natural. Las moléculas endógenas adicionales pueden incluir proteínas, por ejemplo, factores de transcripción y enzimas.

Según se usa en la presente memoria, el término "producto de un ácido nucleico exógeno" incluye productos tanto de polinucleótidos como de polipéptidos, por ejemplo, productos de transcripción (polinucleótidos tales como ARN) y productos de traducción (polipéptidos).

Una molécula "de fusión" es una molécula en la cual se enlazan dos o más moléculas de subunidad, preferiblemente, de manera covalente. Las moléculas de la subunidad pueden ser del mismo tipo químico que la molécula o pueden ser diferentes tipos químicos de moléculas. Los ejemplos del primer tipo de molécula de fusión incluyen, pero no se limitan a, proteínas de fusión (por ejemplo, una fusión entre un dominio de unión a ADN de ZFP o TALE y uno o más dominios de activación) y ácidos nucleicos de fusión (por ejemplo, un ácido nucleico que codifica la proteína de fusión descrita anteriormente). Los ejemplos del segundo tipo de molécula de fusión incluyen, pero no se limitan a, una fusión entre un ácido nucleico que forma una estructura triple y un polipéptido, y una fusión entre un ligando de unión a la ranura menor y un ácido nucleico.

La expresión de una proteína de fusión en una célula puede ser el resultado de suministrar la proteína de fusión a la célula o de suministrar un polinucleótido que codifica la proteína de fusión a una célula, en donde se transcribe el polinucleótido y se traduce el transcrito para generar la proteína de fusión. El transempalme, la escisión del polipéptido y la ligación del polipéptido también pueden participar en la expresión de una proteína en una célula. Los métodos para el suministro de polinucleótidos y polipéptidos a células se presentan en otra parte de la presente descripción.

Un "gen", a los efectos de la presente descripción, incluye una región de ADN que codifica un producto génico (ver más adelante), así como todas las regiones de ADN que regulan la producción del producto génico, ya sea que estas secuencias reguladoras sean adyacentes o no a secuencias codificantes, secuencias transcritas y combinaciones de secuencias codificantes y transcritas. Por consiguiente, un gen incluye, pero no se limita necesariamente a, secuencias promotoras, terminadores, secuencias reguladoras de la traducción, tales como sitios de unión a ribosoma y sitios de entrada al ribosoma interno, potenciadores, silenciadores, aisladores, elementos de límite, orígenes de replicación, sitios de acoplamiento a la matriz y regiones de control del locus.

La "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 de transcripción directo de un gen (p. ej., ARNm, ARNt, ARNr, ARN antisentido, microARN, ribozima, ARN estructural o cualquier otro tipo de ARN) o una proteína producida mediante la traducción de un mRNA. Los productos génicos también incluyen ARN que están modificados, mediante procesos tales como la formación de casquetes, poliadenilación, metilación y edición, y proteínas modificadas, por ejemplo, mediante 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, activación génica y represión génica. La edición genómica (p. ej., escisión, alteración, inactivación y mutación aleatoria) puede usarse para modular la expresión. La inactivación génica se refiere a toda reducción de la expresión génica en comparación con una célula que no incluye un sistema de ZFP, TALE o CRISPR/Cas como se describe en la presente memoria. Por lo tanto, la inactivación génica puede ser parcial o completa.

Un ARNm "protegido" es uno en el que se alteró el ARNm de alguna forma para aumentar la estabilidad o la traducción del ARNm. Los ejemplos de protecciones incluyen el uso del reemplazo de hasta 25 % de los residuos de citidina y uridina con 2-tiouridina (s2U) y 5-metilcitidina (m5C). El ARNm resultante exhibe inmunogenicidad menor y más estabilidad en comparación con su contraparte no modificada (ver Karikó et al. ((2012), Molecular Therapy, Vol. 16, No. 11, páginas 1833 - 1844). Otros cambios incluyen la adición de un denominado casquete ARCA, que aumenta la capacidad de traducción del ARNm producido *in vitro* (ver la patente estadounidense núm. 7.074.596).



Una "región de interés" es cualquier región de cromatina celular, tal como, por ejemplo, un gen o una secuencia no codificante dentro de o adyacente a un gen, en la cual es deseable unir una molécula exógena. La unión puede ser a efectos de una escisión de ADN dirigida, recombinación dirigida y combinaciones de escisión de ADN dirigida y recombinación dirigida. Una región de interés puede estar presente en un cromosoma, un episoma, un genoma organular (p. ej., mitocondrial, cloroplasto) o un genoma vírico infeccioso, por ejemplo. Una región de interés puede estar dentro de la región codificante de un gen, dentro de regiones no codificantes transcritas tales como, por ejemplo, secuencias líderes, secuencias traseras o intrones, o dentro de regiones no transcritas, ya sea anteriores o posteriores a la región codificante. Una región de interés puede ser tan pequeña como un único par de nucleótidos o de hasta 2000 pares de nucleótidos de longitud, o cualquier valor entero de pares de nucleótidos.

Las células "eucariotas" incluyen, pero no se limitan a, células fúngicas (tales como levadura), células vegetales, células animales, células de mamífero y células humanas (p. ej., linfocitos T).

Los términos "enlace funcional" y "enlazado/a funcionalmente" (o "enlazado/a operativamente") se usan de manera intercambiable con referencia a una yuxtaposición de dos o más componentes (tales como elementos de secuencia), en los que los componentes se disponen de manera que ambos componentes funcionen de manera normal y permitan la posibilidad de que al menos uno de los componentes pueda mediar una función que se ejerce sobre al menos uno de los otros componentes. A modo de ilustración, una secuencia reguladora transcripcional, tal como un promotor, se enlaza funcionalmente a una secuencia codificante si la secuencia reguladora transcripcional controla el nivel de transcripción de la secuencia codificante como respuesta a la presencia o ausencia de uno o más factores reguladores transcripcionales. Una secuencia reguladora transcripcional generalmente se enlaza funcionalmente en *cis* con una secuencia codificante, pero no necesita estar directamente adyacente a esta. Por ejemplo, un potenciador es una secuencia reguladora transcripcional que se enlaza funcionalmente con una secuencia codificante, incluso cuando estas no son contiguas.

Con respecto a los polipéptidos de fusión, el término "enlazado/a funcionalmente" puede referirse al hecho de que cada uno de los componentes lleva a cabo la misma función al estar enlazado con el otro componente que tendría si no estuvieran enlazados. Por ejemplo, con respecto a un polipéptido de fusión en el cual un dominio de unión a ADN de ZFP, TALE o Cas se fusiona con un dominio de activación, el dominio de unión a ADN de ZFP, TALE o Cas y el dominio de activación tienen un enlace funcional 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, su sitio de unión y combinaciones de su sitio diana y sitio de unión, mientras que el dominio de activación es capaz de regular por aumento la expresión génica. Cuando un polipéptido de fusión en el cual un dominio de unión a ADN de ZFP, TALE o Cas se fusiona con un dominio de escisión, el dominio de unión a ADN de ZFP, TALE o Cas y el dominio de escisión tienen un enlace funcional si, en el polipéptido de fusión, la porción del dominio de unión a ADN de ZFP, TALE o Cas es capaz de unirse a su sitio diana, su sitio de unión y combinaciones de su sitio de unión y sitio diana, mientras que el dominio de escisión es capaz de escindir ADN en la proximidad del sitio diana.

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, polipéptido o ácido nucleico de longitud completa, pero que mantiene la misma función que la proteína, polipéptido o ácido nucleico de longitud completa. Un fragmento funcional puede tener más, menos o la misma cantidad de residuos que la molécula natural correspondiente, puede contener una o más sustituciones de aminoácidos o nucleótidos y pueden ser combinaciones que tienen más, menos o la misma cantidad de residuos que la molécula natural correspondiente y que contienen una o más sustituciones de aminoácidos o nucleótidos. Se conocen en la técnica métodos para determinar la función de un ácido nucleico (p. ej., función codificante, capacidad de hibridación con otro ácido nucleico). De manera similar, se conocen métodos para determinar la función proteica. Por ejemplo, la función de unión a ADN de un polipéptido se puede determinar, por ejemplo, mediante ensayos de unión a filtro, desplazamiento de movilidad electroforética o inmunoprecipitación. La escisión de ADN puede someterse a ensayo mediante electroforesis en gel. Ver Ausubel *et al.*, *supra*. La capacidad de una proteína para interactuar con otra proteína se puede determinar, por ejemplo, mediante coimmunoprecipitación, ensayos con dos híbridos o complementación, tanto genética como bioquímica.

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

Los términos "sujeto" y "paciente" se usan de manera intercambiable y hacen referencia a mamíferos, tales como pacientes humanos y primates no humanos, así como también animales experimentales, tales como cerdos, vacas, conejos, perros, gatos, ratas, ratones y otros animales. Por consiguiente, el término "sujeto" o "paciente", según se usa en la presente memoria, significa cualquier sujeto o paciente mamífero al cual se le pueden administrar las células madre de la invención. Los sujetos de la presente invención incluyen aquellos que se han expuesto a una o más toxinas químicas, que incluyen, por ejemplo, una toxina nerviosa.

"Capacidad de célula madre" se refiere a la capacidad relativa de cualquier célula para actuar de forma similar a una célula madre, es decir, el grado de totipotencia, pluripotencia, multipotencia u oligopotencia o autorrenovación expandida o indefinida que puede tener cualquier célula madre particular.

## Nucleasas

Se describen en la presente memoria composiciones, particularmente nucleasas, que son útiles para la escisión *in vivo* de una molécula donante que porta un transgén y nucleasas para la escisión del genoma de una célula, de modo que el transgén se integre en el genoma de forma dirigida. Una o más de las nucleasas pueden ser de origen natural. Una o más de las nucleasas pueden ser de origen no natural, es decir, genomanipuladas en el dominio de unión a ADN, el dominio de escisión y una combinación del dominio de unión a ADN y el dominio de escisión. Por ejemplo, puede alterarse el dominio de unión a ADN de una nucleasa de origen natural para que se una a un sitio diana seleccionado (p. ej., una meganucleasa que se ha genomanipulado para unirse a un sitio diferente del sitio de unión cognado). En otros aspectos de la descripción, la nucleasa comprende dominios de unión a ADN y de escisión heterólogos (p. ej., nucleasas de dedos de zinc; proteínas de unión a ADN del dominio efector de TAL; dominios de unión a ADN de meganucleasa con dominios de escisión heterólogos).

A. Dominios de unión a ADN

El dominio de unión a ADN de una o más de las nucleasas usadas para la escisión *in vivo*, escisión dirigida del genoma de una célula y combinaciones de escisión *in vivo* y escisión dirigida del genoma de una célula pueden comprender una proteína de dedos de zinc. Una proteína de dedos de zinc única está compuesta por múltiples dominios de dedos de zinc (p. ej., 3, 4, 5, 6 o más dominios de dedos de zinc). Cada dominio de dedos de zinc tiene aproximadamente 30 aminoácidos de longitud que contienen un giro beta (que contienen los dos residuos coordinantes de zinc y una hélice alfa (que contiene los dos residuos coordinantes de zinc invariantes), que se mantienen en una conformación particular que permite la unión de la proteína a la secuencia diana. Se pueden usar dominios de dedos de zinc canónicos (C2H2) que tienen dos residuos coordinantes de zinc cisteína (Cys) en el giro beta y dos residuos coordinantes de zinc histidina (His) en la hélice alfa o no canónicos (CH3). Ver, p. ej., la patente estadounidense núm. 9.234.187. Hay una hélice de reconocimiento de 7 aminoácidos contenida entre los residuos coordinantes de zinc del giro beta y los residuos coordinantes de zinc de la hélice alfa. La región de hélice de reconocimiento se enumera -1 a +6 dentro del dominio de dedos de zinc y los aminoácidos fuera de esta región de reconocimiento (y con la exclusión de los residuos coordinantes de zinc se denominan residuos de la estructura).

Preferiblemente, la proteína de dedos de zinc es de origen no natural en el sentido de que la hélice de reconocimiento está genomanipulada para unirse a un sitio diana de elección. Ver, por ejemplo, Ver, por ejemplo, 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 estadounidenses núms. 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 la publicación de patentes estadounidenses núms. 2005/0064474; 2007/0218528; 2005/0267061.

Un dominio de unión de dedo de zinc genomanipulado puede tener una especificidad de unión nueva en comparación con la proteína de dedo de zinc de origen natural. Los métodos de genomanipulación incluyen, pero no se limitan a, diseño racional y varios tipos de selección. El diseño racional incluye, por ejemplo, el uso de bases de datos que comprenden secuencias de nucleótidos en triplete (o cuadruplete) y secuencias de aminoácidos de dedos de zinc individuales, en las que cada secuencia de nucleótidos en triplete o cuadruplete se asocia con una o más secuencias de aminoácidos de dedos de zinc que se unen a la secuencia en triplete o cuadruplete particular. Ver, por ejemplo, las patentes estadounidenses núms. 6.453.242 y 6.534.261 de copropiedad.

Los dominios de unión a ADN de ZFP pueden comprender, además, una o más modificaciones en la estructura de uno o más de los dominios de dedos de zinc componentes. La especificidad de una ZFP para una secuencia de ADN diana depende de los contactos específicos de secuencias entre los dominios de dedos de zinc y bases de ADN específicas, en particular, entre la región de hélice de reconocimiento y el sitio diana (típicamente cada hélice de reconocimiento se une a un subsitio diana de 3 nucleótidos). Además, los dominios de dedos de zinc comprenden también residuos aminoacídicos que participan en interacciones no específicas con los fosfatos de la estructura de ADN. Elrod-Erickson et al. ((1996) Structure 4:1171) demostró a través de la cocrystalización de una proteína de dedos de zinc y su diana de ADN cognada que hay aminoácidos específicos capaces de interactuar con los fosfatos de la estructura de ADN a través de la formación de enlaces de hidrógeno. Las proteínas de dedos de zinc que emplean la estructura Zif268 bien conocida típicamente tienen una arginina como el residuo de extremo amínico de su segunda cadena de lámina  $\beta$ , que también es el extremo carboxílico en segunda posición con respecto a la segunda cisteína invariante. Esta posición se puede denominar (-5) dentro de cada dominio de dedos de zinc, ya que es el 5.º residuo que precede al inicio de la hélice  $\alpha$  (y la posición -5 en relación con la hélice de reconocimiento numerada -1 a +6). La arginina en esta posición puede interactuar con un fosfato en la estructura de ADN a través de la formación de un enlace de hidrógeno cargado con su grupo guanidinio de cadena lateral. Las proteínas de dedos de zinc en la estructura Zif268 también tienen frecuentemente una lisina en una posición que está a 4 residuos hacia el extremo amínico con respecto a la primera cisteína invariante. Esta posición se puede denominar (-14) dentro de cada dedo, dado que es el 14.º residuo que precede al inicio de la hélice  $\alpha$  para los dedos de zinc con dos residuos entre los residuos de cisteína coordinantes de zinc (y la posición -14 en relación con la región de hélice de reconocimiento numerada -1 a +6). La lisina puede interactuar con un fosfato en la estructura de ADN a través de la formación de un enlace de hidrógeno cargado mediado por agua con su grupo amino de cadena lateral. Dado que se encuentran grupos fosfato en toda la estructura de ADN, este tipo de interacción entre el dedo de zinc y una molécula de ADN se

considera generalmente como no específica de la secuencia (J. Miller, Massachusetts Institute of Technology Ph.D. Thesis, 2002).

Estudios recientes han hipotetizado que las cadenas laterales de contacto con fosfato no específicas en algunas nucleasas pueden ser responsables de parte de la no especificidad de dichas nucleasas (Kleinstiver et al. (2016) Nature 529(7587):490-5; Guilinger et al. (2014) Nat Meth: 429-435). Los investigadores han propuesto que estas nucleasas pueden tener "energía de unión a ADN en exceso", lo que significa que las nucleasas pueden tener una mayor afinidad por su diana de ADN que la necesaria para unirse sustancialmente y escindir el sitio diana. Por lo tanto, se hicieron intentos de reducir las cargas catiónicas en el dominio de unión a ADN de TALE (Guilinger, *ibid*) o el dominio de unión a ADN de Cas9 (Kleinstiver, *ibid*) para reducir la energía de unión a ADN de estas nucleasas, lo que resultó en un aumento de la especificidad de escisión *in vitro*. Sin embargo, estudios adicionales (Sternberg et al. (2015) Nature 527(7576):110-113) sugieren también una función en el plegado y la activación adecuados del dominio de la nucleasa Cas9 para algunos de los aminoácidos catiónicos que se mutaron en el estudio Kleinstiver del dominio de unión a ADN de Cas9. Por lo tanto, la función exacta de estos aminoácidos en la actividad de Cas9 se desconoce.

Se describen mutaciones en aminoácidos dentro del dominio de unión a ADN de ZFP ("estructura de ZFP") que pueden interactuar de manera no específica con fosfatos en la estructura de ADN, pero que no comprenden cambios en las hélices de reconocimiento de ADN. Se describen mutaciones de residuos aminoacídicos catiónicos en la estructura de ZFP que no son necesarios para la especificidad diana del nucleótido. Estas mutaciones en la estructura de ZFP pueden comprender: mutar un residuo aminoacídico catiónico en un residuo aminoacídico neutro o aniónico; o mutar un residuo aminoacídico polar en un residuo aminoacídico neutro o no polar. Se prefieren las mutaciones hechas en la posición (-5), posición (-9), posición (-14) y combinaciones de mutaciones seleccionadas de mutaciones hechas en la posición (-5), posición (-9) y posición (-14) con respecto a la hélice de unión a ADN. Un dedo de zinc puede comprender una o más mutaciones en (-5), (-9), (-14), y combinaciones de mutaciones seleccionadas de mutaciones en (-5), (-9) y (-14). Uno o más dedo(s) de zinc en una proteína de múltiples dedos de zinc puede comprender mutaciones en (-5), (-9), (-14) y combinaciones seleccionadas de (-5), (-9) y (-14). Los aminoácidos en (-5), (-9), (-14) y combinaciones seleccionadas de (-5), (-9) y (-14) (p. ej., una arginina (R) o lisina (K)) se pueden mutar en una alanina (A), leucina (L), Ser (S), Asp (D), Glu (E), Tyr (Y) y/o glutamina (Q).

En cualquiera de estos polipéptidos de fusión descritos en la presente memoria, los compañeros de ZFP pueden comprender, además, mutaciones en el dominio de unión a ADN de dedos de zinc en las posiciones (-5), (-9), (-14), y combinaciones de mutaciones seleccionadas de mutaciones en (-5), (-9) y (-14). La Arg (R) en la posición -5 se puede cambiar a una Tyr (Y), Asp (D), Glu (E), Leu (L), Gln (Q) o Ala (A). La Arg (R) en la posición (-9) se puede reemplazar con Ser (S), Asp (D) o Glu (E). La Arg (R) en la posición (-14) se puede reemplazar con Ser (S) o Gln (Q). Los polipéptidos de fusión pueden comprender mutaciones en el dominio de unión a ADN de dedos de zinc donde los aminoácidos en las posiciones (-5), (-9), (-14), y combinaciones de mutaciones seleccionadas de mutaciones en (-5), (-9), (-14) se cambian a cualesquiera de los aminoácidos indicados anteriormente en cualquier combinación.

Los ejemplos de métodos de selección, incluidos sistemas de visualización en fagos y de dos híbridos, se describen en las patentes estadounidenses 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 también WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 y GB 2.338.237. Además, potenciar la especificidad de unión para dominios de unión de dedos de zinc se ha descrito, por ejemplo, en la publicación internacional núm. WO 02/077227 de copropiedad.

Los expertos en la técnica conocen la selección de sitios diana; ZFP y métodos para el diseño y la construcción de proteínas de fusión (y polinucleótidos que las codifican) y se describen de forma detallada en las patentes estadounidenses núms. 6.140.081; 5.789.538; 6.453.242; 6.534.261; 5.925.523; 6.007.988; 6.013.453; 6.200.759; la publicación de PCT núm. 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.

Puede usarse casi cualquier enlazador (espaciador) 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 a ADN, entre el dominio de unión a ADN y el dominio funcional (p. ej., nucleasa) y entre uno o más dominios de unión a ADN y entre el dominio de unión a ADN y el dominio funcional. Los ejemplos no limitantes de secuencias de enlazador adecuadas incluyen las patentes estadounidenses núms. 8.772.453; 7.888.121; 6.479.626; 6.903.185; y 7.153.949; y las publicaciones de patentes estadounidenses núms. 2009/0305419; 2015/0064789 y 2015/0132269. Por lo tanto, las proteínas descritas en la presente memoria pueden incluir cualquier combinación de enlazadores adecuados entre los componentes individuales de unión a ADN, entre el dominio de unión a ADN y el dominio funcional o entre uno o más dominios de unión a ADN y entre el dominio de unión a ADN y el dominio funcional de las composiciones descritas en la presente memoria.

#### B. Dominios de escisión

Cualquier dominio de escisión adecuado puede enlazarse funcionalmente con los dominios de unión a ADN como se describen en la presente memoria para formar una nucleasa. El dominio de escisión puede ser heterólogo con respecto al dominio de unión a ADN, por ejemplo, un dominio de unión a ADN de dedos de zinc y un dominio de escisión de una nucleasa. Los dominios de escisión heterólogos pueden obtenerse de cualquier endonucleasa o exonucleasa. Loas endonucleasas ilustrativas de las cuales se puede derivar un dominio de escisión incluyen, pero no se limitan a,

endonucleasas de restricción y endonucleasas de asentamiento. Ver, por ejemplo, 2002-2003 Catalogue, 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 del poroto chino; DNasa pancreática I; nucleasa microcócica; endonucleasa HO de levadura; ver también Linn et al. (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993). Se pueden usar una o más de estas enzimas (o fragmentos funcionales de estas) como una fuente de dominios de escisión y dominios medios de escisión.

De manera similar, se puede derivar un dominio medio de escisión de cualquier nucleasa o porción de esta, tal como se indicó anteriormente, que requiera dimerización para la actividad de escisión. En general, se necesitan dos proteínas de fusión para la escisión si las proteínas de fusión comprenden dominios medios de escisión. De manera alternativa, puede usarse una proteína individual que comprende dos dominios medios de escisión. Los dos dominios medios de escisión pueden derivarse de la misma endonucleasa (o fragmentos funcionales de esta) o cada dominio medio de escisión puede derivarse de una endonucleasa diferente (o fragmentos funcionales de esta). Además, los sitios diana para las dos proteínas de fusión preferiblemente se disponen, uno con respecto a otro, de modo que la unión de las dos proteínas de fusión a sus respectivos sitios diana ubique los dominios medios de escisión en una orientación espacial entre sí que permita que los dominios medios de escisión formen un dominio de escisión funcional, p. ej., mediante dimerización. Por lo tanto, los bordes cercanos de los sitios diana están separados por 5-8 nucleótidos o por 15-18 nucleótidos. Sin embargo, puede intervenir cualquier número entero de nucleótidos o pares de nucleótidos 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 diana.

Las endonucleasas de restricción (enzimas de restricción) están presentes en muchas especies y son capaces de unión específica de secuencia a ADN (en un sitio de reconocimiento) y de escindir el ADN en o cerca del sitio de unión. Ciertas enzimas de restricción (p. ej., de Tipo IIS) escinden el ADN en sitios apartados del sitio de reconocimiento y tienen dominios de escisión y de unión que se pueden separar. Por ejemplo, la enzima de tipo IIS *Fok I* cataliza la escisión bicatenaria del ADN a 9 nucleótidos de su sitio de reconocimiento en una hebra y a 13 nucleótidos de su sitio de reconocimiento en la otra. Ver, por ejemplo, las patentes estadounidenses núms. 5.356.802; 5.436.150 y 5.487.994; así como Li et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:4275-4279; Li et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:2764-2768; Kim et al. (1994a) *Proc. Natl. Acad. Sci. USA* 91:883-887; Kim et al. (1994b) *J. Biol. Chem.* 269:31,978-31,982. Por lo tanto, las proteínas de fusión pueden comprender el dominio de escisión (o dominio medio de escisión) de al menos una enzima de restricción de Tipo IIS y uno o más dominios de unión de dedos de zinc, que pueden o no estar genomanipulados.

Para la especificidad de escisión óptima por una nucleasa (artificial) selectiva de secuencia, es deseable arreglar las condiciones de modo que la unión a la diana y la actividad no sean saturantes. En condiciones de saturación, por definición, se usa un exceso de nucleasa con respecto a lo necesario para lograr la actividad en la diana completa. Este exceso no proporciona beneficios en la diana, pero puede resultar, sin embargo, en una mayor escisión en los sitios inespecíficos. Para las nucleasas monoméricas, las condiciones de saturación se pueden evitar fácilmente al llevar a cabo un estudio de respuesta a la dosis simple para identificar y evitar la meseta de saturación en una curva de titulación. Sin embargo, para una nucleasa dimérica tal como ZFN, TALEN o dCas-Fok, identificar y evitar las condiciones de saturación puede ser más complicado si las afinidades de unión de los monómeros individuales son disímiles. En tales casos, un estudio de respuesta a la dosis que usa una relación de nucleasa 1:1 simple solo revelará el punto de saturación del monómero de unión más débil. En tal situación, si, por ejemplo, las afinidades de monómero difieren en un factor de 10, entonces en el punto de saturación identificado en un estudio de titulación 1:1, el monómero con mayor afinidad estará presente en una concentración que es 10 veces más alta que la necesaria. El exceso resultante del monómero de afinidad más alta puede, a su vez, aumentar la actividad inespecífica sin proporcionar un aumento beneficioso en la escisión en la diana prevista, lo que lleva potencialmente a una especificidad reducida general para cualquier par de nucleasas dado.

*Fok I* es una enzima de restricción de Tipo IIS ilustrativa cuyo dominio de escisión puede separarse del dominio de unión. Esta enzima particular es activa como dímero. Bitinaite et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:10,570-10,575. Por consiguiente, a los efectos de la presente descripción, la porción de la enzima *Fok I* usada en las proteínas de fusión descritas se considera un dominio medio de escisión. Por lo tanto, para la escisión bicatenaria dirigida, el remplazo dirigido de las secuencias celulares al usar fusiones de dedo de zinc-*Fok I* y combinaciones de escisión bicatenaria dirigida y remplazo dirigido de secuencias celulares al usar fusiones de dedos de zinc-*Fok I*, dos proteínas de fusión, donde cada una comprende un dominio medio de escisión de *Fok I*, se pueden usar para reconstituir un dominio de escisión catalíticamente activo. De manera alternativa, también puede usarse una sola molécula de polipéptido que contiene un dominio de unión de dedos de zinc y dos dominios medios de escisión de *Fok I*. Se proporcionan en otra parte de la descripción los parámetros para la escisión dirigida y alteración de secuencia dirigida mediante el uso de fusiones de dedos de zinc-*Fok I*.

Un dominio de escisión o dominio medio de escisión puede ser cualquier porción de una proteína que conserva la actividad de escisión, o que conserva la capacidad de multimerizarse (p. ej., dimerizarse) para formar un dominio de escisión funcional.

Se describen enzimas de restricción de Tipo IIS ilustrativas en la patente estadounidense núm. 7.888.121. Las enzimas de restricción adicionales también contienen dominios de unión y de escisión que se pueden separar, y estos se contemplan en la presente descripción. Ver, por ejemplo, Roberts et al. (2003) *Nucleic Acids Res.* 31:418-420.

El dominio de escisión puede comprender uno o más dominios medios de escisión genomanipulados (también denominados mutantes de dominio de dimerización) que minimizan o evitan la homodimerización, como se describe, por ejemplo, en ver p. ej., las patentes estadounidenses núms. 7.914.796; 8.034.598 y 8.623.618. Los residuos aminoacídicos 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 dianas para influir sobre la dimerización de los dominios medios de escisión de *FokI*, donde la numeración es con respecto a las estructuras de cristal 1FOK.pdb y 2FOK.pdb (ver Wah et al. (1997) *Nature* 388:97-100) que tiene la secuencia que se muestra a continuación: Dominio medio de escisión de *FokI* natural (SEQ ID NO:18)

QLVKSELEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVY  
GYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRY  
VEENQTRNKHINPNEWKVPSSVTEFKFLFVSGHFKGNKYAQLTRLNHNITN  
CNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

Los dominios medios de escisión de *FokI* genomanipulados ilustrativos que forman heterodímeros estrictos incluyen un par en el cual un primer dominio medio de escisión incluye mutaciones en los residuos aminoacídicos en las posiciones 490 y 538 de *FokI* y un segundo dominio medio de escisión incluye mutaciones en los residuos aminoacídicos 486 y 499.

Por lo tanto, una mutación en 490 puede reemplazar Glu (E) con Lys (K); la mutación en 538 puede reemplazar Ile (I) con Lys (K); la mutación en 486 puede reemplazar Gln (Q) con Glu (E); y la mutación en la posición 499 puede reemplazar Ile (I) con Lys (K). Específicamente, los dominios medios de escisión genomanipulados descritos en la presente memoria se prepararon al mutar las posiciones 490 (E→K) y 538 (I→K) en un dominio medio de escisión para producir un dominio medio de escisión genomanipulado designado "E490K:I538K" y al mutar las posiciones 486 (Q→E) y 499 (I→L) en otro dominio medio de escisión para producir un dominio medio de escisión genomanipulado designado "Q486E:I499L". Los dominios medios de escisión genomanipulados descritos en la presente memoria son mutantes de heterodímeros estrictos en los que se minimiza o elimina la escisión aberrante a través de los homodímeros de ZFN. Ver, p. ej., la publicación de patente estadounidense núm. 2008/0131962. El dominio medio de escisión genomanipulado puede comprender mutaciones en las posiciones 486, 499 y 496 (numeradas con respecto a *FokI* natural), por ejemplo, mutaciones que reemplazan el residuo Gln (Q) natural en la posición 486 con un residuo Glu (E), el residuo Ile (I) natural en la posición 499 con un residuo Leu (L), y el residuo Asn (N) natural en la posición 496 con un residuo Asp (D) o un residuo Glu (E) (también denominados dominios "ELD" y "ELE", respectivamente). El dominio medio de escisión genomanipulado puede comprender mutaciones en las posiciones 490, 538 y 537 (numeradas con respecto a *FokI* natural), por ejemplo, mutaciones que reemplazan el residuo Glu (E) natural en la posición 490 con un residuo Lys (K), el residuo Ile (I) natural en la posición 538 con un residuo Lys (K), y el residuo His (H) natural en la posición 537 con un residuo Lys (K) o un residuo Arg (R) (a los que también se denomina dominios "KKK" y "KKR", respectivamente). El dominio medio de escisión genomanipulado puede comprender mutaciones en las posiciones 490 y 537 (numeradas con respecto a *FokI* natural), por ejemplo, mutaciones que reemplazan el residuo Glu natural (E) en la posición 490 con un residuo Lys (K) y el residuo His (H) natural en la posición 537 con un residuo Lys (K) o un residuo Arg (R) (a los que también se denomina dominios "KIK" y "KIR", respectivamente). Ver, p. ej., las patentes estadounidenses núms. 7.914.796; 8.034.598 y 8.623.618. El dominio medio de escisión genomanipulado puede comprender "Sharkey", las mutaciones de "Sharkey" y combinaciones de "Sharkey" y mutaciones de "Sharkey" (ver Guo et al. (2010) *J. Mol. Biol.* 400(1):96-107).

Por lo tanto, los dominios medios de escisión derivados de *FokI* pueden comprender una mutación en uno o más residuos aminoacídicos como se muestra en la SEQ ID NO:18, incluso mutaciones en el dominio de dimerización como se describió anteriormente; mutaciones en el dominio catalítico, mutaciones en los otros residuos aminoacídicos tales como residuos de contacto con fosfato, y cualquier combinación de mutaciones seleccionadas de mutaciones en el dominio de dimerización, mutaciones en el dominio catalítico y mutaciones en otros residuos aminoacídicos tales como residuos de contacto con fosfato. Las mutaciones incluyen sustituciones (de un residuo aminoacídico natural con un residuo diferente), inserciones (de uno o más residuos aminoacídicos), eliminaciones (de uno o más residuos aminoacídicos) y cualquier combinación de mutaciones seleccionadas de sustituciones, inserciones y eliminaciones. Uno o más de los residuos 414-426, 443-450, 467-488, 501-502, 521-531 (numerados en relación con la SEQ ID NO:18) y cualquier combinación de tales residuos se puede mutar ya que estos residuos están ubicados cerca de la estructura de ADN en un modelo molecular de una ZFN unida a su sitio diana descrito en Miller et. al. ((2007) *Nat Biotechnol* 25:778-784). Uno o más residuos en las posiciones 416, 422, 447, 448 y 525 se pueden mutar. La mutación puede comprender una sustitución de un residuo natural con un residuo diferente, por ejemplo, un residuo serina (S). El dominio de escisión de *FokI* de las nucleasas descritas en la presente memoria comprenden una mutación de dominio de dimerización de ELD, una mutación de dominio de dimerización de KKR, una mutación K525S, y cualquier combinación seleccionada de una mutación de dominio de dimerización de ELD o una mutación de dominio de dimerización de KKR y una mutación K525S.

Los dominios de escisión genomanipulados descritos en la presente memoria pueden prepararse mediante el uso de cualquier método adecuado, por ejemplo, mediante mutagénesis dirigida al sitio de dominios medios de escisión natural (*FokI*) tal como se describe en las patentes estadounidenses núms. 7.888.121; 7.914.796; 8.034.598 y

8.623.618. Además, los dominios de escisión descritos en la presente memoria se pueden fusionar con un dominio de unión a ADN (p. ej., ZFP) al usar cualquier enlazador adecuado que incluye, pero no se limita a, los enlazadores descritos en las patentes estadounidenses núms. 9.394.531 y 9.567.609.

Alternativamente, las nucleasas se pueden ensamblar *in vivo* en el sitio diana de ácido nucleico al usar la llamada tecnología de "enzima de separación" (ver, p. ej., la publicación de patente estadounidense núm. 2009/0068164). Los componentes de dichas enzimas de separación pueden expresarse ya sea en construcciones de expresión separadas o pueden enlazarse en un marco de lectura abierto donde se separan los componentes individuales, por ejemplo, mediante un péptido 2A o una secuencia IRES autoescindible. Los componentes pueden ser dominios de unión de dedo de zinc individuales o dominios de un dominio de unión a ácido nucleico de meganucleasa.

Las nucleasas pueden analizarse para determinar la actividad antes de su uso, por ejemplo, en un sistema cromosómico a base de levadura como se describe en WO 2009/042163 y 20090068164. La expresión de la nucleasa puede estar bajo el control de un promotor constitutivo o de un promotor inducible, por ejemplo, el promotor de galactocinasa que se activa (desreprime) en presencia de rafinosa, galactosa y una combinación de rafinosa y galactosa y se reprime en presencia de glucosa.

La(s) nucleasa(s) como se describe(n) en la presente memoria pueden hacer uno o más cortes bicatenarios, uno o más cortes monocatenarios y combinaciones de uno o más cortes bicatenarios y uno o más monocatenarios en el sitio diana. En ciertos aspectos de la descripción, la nucleasa comprende un dominio de escisión catalíticamente inactivo (p. ej., FokI, una proteína Cas y combinaciones de FokI y una proteína Cas). Ver, p. ej., las patentes estadounidenses núms. 9.200.266; 8.703.489 y Guillingier et al. (2014) Nature Biotech. 32(6):577-582. El dominio de escisión catalíticamente inactivo, en combinación con un dominio catalíticamente activo, puede actuar como una nickasa para hacer un corte monocatenario. Por lo tanto, pueden usarse dos nickasas en combinación para hacer un corte bicatenario en una región específica. También se conocen nickasas adicionales en la técnica, por ejemplo, McCaffery et al. (2016) Nucleic Acids Res. 44(2):e11. doi: 10.1093/nar/gkv878. Epub 2015 Oct 19.

#### Sitios diana

Como se describió anteriormente de manera detallada, los dominios de ADN pueden genomanipularse para que se unan a cualquier secuencia elegida. Un dominio de unión a ADN genomanipulado puede tener una nueva especificidad de unión en comparación con un dominio de unión a ADN de origen natural. Los dominios de unión a ADN pueden unirse a una secuencia dentro de una secuencia potenciadora de 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) que está entre el exón 2 y el exón 3 de BCL11A, incluidos los dominios de unión a ADN que se unen a una secuencia dentro de un sitio hipersensible a DNasaI en la secuencia potenciadora de BCL11A (p. ej., +58) como se muestra en la Tabla 1. Los métodos de genomanipulación incluyen, pero no se limitan a, diseño racional y varios tipos de selección. El diseño racional incluye, por ejemplo, el uso de bases de datos que comprenden secuencias de nucleótidos en triplete (o cuadruplete) y secuencias de aminoácidos de dedos de zinc individuales, en las que cada secuencia de nucleótidos en triplete o cuadruplete se asocia con una o más secuencias de aminoácidos de dedos de zinc que se unen a la secuencia en triplete o cuadruplete particular. Ver, por ejemplo, las patentes estadounidenses núms. 6.453.242 y 6.534.261 de copropiedad. También puede llevarse a cabo el diseño racional de dominios efectores de TAL. Ver, p. ej., la publicación de patente estadounidense núm. 2011/0301073.

Los métodos de selección ilustrativos aplicables a los dominios de unión a ADN, incluidos sistemas de visualización en fagos y de dos híbridos, se describen en las patentes estadounidenses núms. 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 también las publicaciones de PCT núms. WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 y la patente del RU núm. GB 2.338.237. Además, potenciar la especificidad de unión para dominios de unión de dedos de zinc se ha descrito, por ejemplo, en la publicación internacional núm. WO 02/077227 de copropiedad.

Los expertos en la técnica conocen la selección de sitios diana; nucleasas y métodos para diseñar y construir proteínas de fusión (y polinucleótidos que las codifican) y se describen en detalle en las publicaciones de patente estadounidense núms. 2005/0064474 y 2006/0188987.

Además, como se describe en estas y otras referencias, los dominios de unión a ADN (p. ej., las proteínas de dedos de zinc de múltiples dedos) y fusiones de dominio(s) de unión a ADN y dominio(s) funcional(es) pueden enlazarse entre sí mediante el uso de cualesquiera secuencias enlazadoras adecuadas, que incluyen, por ejemplo, enlazadores de 5 o más aminoácidos. Las patentes estadounidenses núms. 8.772.453; 7.888.121 (p. ej., el enlazador "ZC"); 6.479.626; 6.903.185; y 7.153.949; la publicación estadounidense núm. 2009/0305419 y 2015/0064789. Las proteínas descritas en la presente memoria pueden incluir cualquier combinación de enlazadores adecuados entre los dominios de unión a ADN individuales de la proteína. Ver, también, la patente estadounidense núm. 8.586.526.

#### Donantes

La presente descripción se relaciona con la integración dirigida mediada por nucleasa de una secuencia exógena en el genoma de una célula mediante el uso de moléculas de unión a la región potenciadora de BCL11A descrita en la presente memoria. Como se describió anteriormente, la inserción de una secuencia exógena (también denominada

"secuencia donante" o "donante" o "transgén"), por ejemplo, para la eliminación de una región especificada, para la corrección de un gen mutante, para una combinación de eliminación de una región especificada y corrección de un gen mutante o para una mayor expresión de un gen natural. Será fácilmente evidente que la secuencia donante típicamente 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 eficaz en la ubicación de interés o puede integrarse mediante mecanismos de reparación no dirigida por homología, tales como NHEJ. Además, las secuencias donantes 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 respecto a la cromatina celular y, por ejemplo, puede llevar a la eliminación de una región potenciadora de BCL11A (o un fragmento de esta) cuando se usa como sustrato para la reparación de una DSB inducida por una de las nucleasas descritas en la presente. Además, para una inserción de secuencias diana que no se encuentran normalmente en una región de interés, dichas secuencias pueden estar presentes en una molécula de ácido nucleico donante y flanqueadas por regiones de homología con respecto a la secuencia en la región de interés.

También se puede hacer referencia a los polinucleótidos para inserción como polinucleótidos "exógenos", polinucleótidos o moléculas "donantes" o "transgenes". El polinucleótido donante puede ser ADN o ARN, monocatenario o bicatenario y puede introducirse en una célula de forma lineal o circular. Ver, por ejemplo, la publicación de solicitud de patente estadounidense núms. 2010/0047805 y 2011/0207221. La(s) secuencia(s) donante(s) preferiblemente está(n) contenida(s) dentro de un MC de ADN, que puede introducirse en la célula de forma lineal o circular. Si se introduce en forma lineal, los extremos de la secuencia donante se pueden proteger (p. ej., contra la degradación exonucleolítica) mediante métodos conocidos para los expertos en la técnica. Por ejemplo, se agregan uno o más residuos de didesoxinucleótido en el extremo hacia 3' de una molécula lineal y se ligan, opcionalmente, oligonucleótidos autocomplementarios a uno o ambos extremos. Ver, por ejemplo, Chang et al. (1987) Proc. Natl. Acad. Sci. USA 84:4959-4963; Nehls et al. (1996) Science 272:886-889. Los métodos adicionales para proteger polinucleótidos exógenos contra la degradación incluyen, pero no se limitan a, adición de grupo(s) amino terminal(es) y el uso de ligaduras internucleotídicas modificadas tales como, por ejemplo, fosforotioatos, fosforamidatos y residuos O-metil ribosa o desoxirribosa. Si se introduce en forma bicatenaria, el donante puede incluir uno o más sitios diana de nucleasa, por ejemplo, sitios diana de nucleasa que flanquean al transgén que se va a integrar en el genoma de la célula. Ver, p. ej., la publicación de patente estadounidense núm. 2013/0326645.

Un polinucleótido puede introducirse en una célula como parte de una molécula de vector que tiene secuencias adicionales tales como, por ejemplo, orígenes de replicación, promotores y genes que codifican la resistencia antibiótica. Además, los polinucleótidos donantes pueden introducirse como ácido nucleico no conjugado, como ácido nucleico en complejo con un agente tal como un liposoma o poloxámero, o pueden suministrarse a través de virus (p. ej., adenovirus, AAV, virus del herpes, retrovirus, lentivirus y lentivirus deficiente en integrasa (IDLV, por sus siglas en inglés)).

En ciertas realizaciones, el donante bicatenario incluye secuencias (p. ej., secuencias codificantes, también denominadas transgenes) con una longitud mayor que 1 kb, por ejemplo, entre 2 y 200 kb, entre 2 y 10 kb (o cualquier valor entre estos). El donante bicatenario también incluye al menos un sitio diana de nucleasa, por ejemplo. En ciertas realizaciones, el donante incluye al menos 2 sitios diana, por ejemplo, para un par de ZFN o TALEN. Típicamente, los sitios diana de nucleasa se encuentran fuera de las secuencias de transgén, por ejemplo, hacia 5' y/o hacia 3' con respecto a las secuencias de transgén, para la escisión del transgén. El(los) sitio(s) de escisión de nucleasa puede ser para cual(es)quier(a) nucleasa(s). En ciertas realizaciones, el(los) sitio(s) diana de nucleasa contenido(s) en el donante bicatenario son para la(s) misma(s) nucleasa(s) usada(s) para escindir la diana endógena en la cual se integra el donante escindido a través de métodos independientes de homología.

El donante generalmente se inserta para que su expresión se accione mediante el promotor endógeno en el sitio de integración, a saber, el promotor que acciona 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, un potenciador y combinaciones de ambos, promotor y potenciador, por ejemplo, un promotor constitutivo o un promotor inducible o específico de un tejido.

La molécula donante se puede insertar en un gen endógeno, de modo que se exprese la totalidad, parte o nada del gen endógeno. En otras realizaciones, el transgén (p. ej., con o sin secuencias que codifican globina) se integra en cualquier locus endógeno, por ejemplo, un locus seguro. Ver, p. ej., las publicaciones de patente estadounidenses núms. 2008/0299580; 2008/0159996 y 2010/00218264.

Además, si bien no es necesario para la expresión, las secuencias exógenas también pueden incluir secuencias reguladoras de transcripción o traducción, por ejemplo, promotores, potenciadores, aislantes, sitios de entrada de ribosomas internos, secuencias que codifican péptidos 2A, señales de poliadenilación y combinaciones de estos.

Los transgenes que llevan las secuencias donantes descritas en la presente memoria pueden aislarse a partir de plásmidos, células u otras fuentes mediante el uso de técnicas estándar conocidas en la técnica tales como PCR. Los donantes para uso pueden incluir diversos tipos de topología, incluidas superenrollados, circulares relajados, lineales y similares. Alternativamente, pueden sintetizarse químicamente mediante el uso de técnicas estándar de síntesis de

oligonucleótidos. Además, los donantes pueden estar metilados o carecer de metilación. Los donantes pueden encontrarse en forma de cromosomas artificiales bacterianos o de levadura (BAC o YAC).

Los polinucleótidos donantes bicatenarios descritos en la presente memoria pueden incluir una o más bases no naturales, una o más estructuras y combinaciones de una o más bases no naturales y una o más estructuras. En particular, la inserción de una molécula donante con citocinas metiladas puede llevarse a cabo mediante el uso de los métodos descritos en la presente memoria para lograr un estado de quiescencia transcripcional en una región de interés.

El polinucleótido exógeno (donante) puede comprender cualquier secuencia de interés (secuencia exógena). Las secuencias exógenas ilustrativas incluyen, pero no se limitan a, cualquier secuencia codificante de polipéptidos (p. ej., ADNc), secuencias promotoras, secuencias potenciadoras, etiquetas de epítomos, genes marcadores, 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 kanamicina, resistencia a la neomicina, resistencia a G418, resistencia a puromicina, resistencia a la higromicina, resistencia a la blasticidina), secuencias que codifican proteínas coloreadas o fluorescentes o luminiscentes (p. ej., proteína verde fluorescente, proteína verde fluorescente mejorada, proteína roja fluorescente, luciferasa) y proteínas que median el crecimiento celular mejorado, la amplificación génica (p. ej., dihidrofolato reductasa) y combinaciones de crecimiento celular mejorado y amplificación génica. Las etiquetas de epítomos incluyen, por ejemplo, una o más copias de FLAG, His, Myc, purificación de afinidad en tándem (TAP), HA, péptido biotinilable o cualquier secuencia de aminoácidos detectable.

En una realización preferida, la secuencia exógena (transgén) comprende un polinucleótido que codifica cualquier polipéptido del cual se desea la expresión en la célula, que incluye, pero no se limita a, anticuerpos, antígenos, enzimas, receptores (de superficie celular o nucleares), hormonas, linfocinas, citocinas, polipéptidos indicadores, factores de crecimiento y fragmentos funcionales de cualquiera de los anteriores. Las secuencias codificantes pueden ser, por ejemplo, ADNc.

Por ejemplo, la secuencia exógena puede comprender una secuencia que codifica un polipéptido que falta o no es funcional en el sujeto que padece 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 núm. 102700), adrenoleucodistrofia, síndrome de Aicardi, deficiencia de antitripsina alfa-1, talasemia alfa, síndrome de insensibilidad a los andrógenos, síndrome de Apert, arritmia ventricular derecha, displasia, ataxia telangiectasia, síndrome de Barth, talasemia beta, síndrome del nevus azul, enfermedad de Canavan, enfermedades granulomatosas crónicas (CGD, por sus siglas en inglés), síndrome del maullido de gato, fibrosis quística, enfermedad de Dercum, displasia ectodérmica, anemia de Fanconi, fibrodisplasia osificante progresiva, síndrome de X frágil, galactosemia, enfermedad de Gaucher, gangliosidosis generalizada (p. ej., GM1), hemocromatosis, la mutación de hemoglobina C en el 6º codón de globina beta (HbC), hemofilia, enfermedad de Huntington, síndrome de Hurler, hipofosfatasa, síndrome de Klinefelter, enfermedad de Krabbes, síndrome de Langer-Giedion, deficiencia de adhesión leucocitaria (LAD, OMIM núm. 116920), leucodistrofia, síndrome de QT largo, síndrome de Marfan, síndrome de Moebius, mucopolisacaridosis (MPS), síndrome de uña-rótula, diabetes insípida nefrogénica, neurofibromatosis, enfermedad de Neimann-Pick, osteogénesis imperfecta, porfiria, síndrome de Prader-Willi, progeria, síndrome de Proteo, retinoblastoma, síndrome de Rett, síndrome de Rubinstein-Taybi, síndrome de Sanfilippo, inmunodeficiencia combinada grave (SCID, por sus siglas en inglés), síndrome de Shwachman, anemia de células falciformes (anemia falciforme), síndrome de Smith-Magenis, síndrome de Stickler, enfermedad de Tay-Sachs, síndrome de trombocitopenia + aplasia radial (TAR, por sus siglas en inglés), síndrome de Treacher Collins, trisomía, esclerosis tuberosa, síndrome de Turner, trastorno del ciclo de la urea, enfermedad de von Hippel-Landau, síndrome de Waardenburg, síndrome de Williams, enfermedad de Wilson, síndrome de Wiskott-Aldrich, síndrome linfoproliferativo ligado al cromosoma X (XLP, por sus siglas en inglés, OMIM núm. 308240).

Las enfermedades ilustrativas adicionales que pueden tratarse mediante integración dirigida incluyen inmunodeficiencias adquiridas, enfermedades de almacenamiento lisosómico (p. ej., enfermedad de Gaucher, GM1, enfermedad de Fabry y enfermedad de Tay-Sachs), mucopolisacaridosis (p. ej., enfermedad de Hunter, enfermedad de Hurler), hemoglobinopatías (p. ej., anemia de células falciformes, HbC, talasemia  $\alpha$ , talasemia  $\beta$ ) y hemofilias.

En ciertas realizaciones, las secuencias exógenas pueden comprender un gen marcador (descrito anteriormente), que permite la selección de células que se sometieron a integración dirigida, y una secuencia enlazada que codifica una funcionalidad adicional. Los ejemplos no limitantes de genes marcadores incluyen GFP, marcador(es) de selección de fármaco y similares.

Otras secuencias génicas adicionales que pueden insertarse pueden incluir, por ejemplo, genes naturales para reemplazar secuencias mutadas. Por ejemplo, una secuencia génica del Factor IX natural puede insertarse en el genoma de una célula madre donde la copia endógena del gen está mutada. La copia natural puede insertarse en el locus endógeno o, alternativamente, puede dirigirse a un locus seguro.

La construcción de dichos casetes de expresión, que siguen las indicaciones de la presente memoria descriptiva, utiliza metodologías conocidas en la técnica de biología molecular (ver, por ejemplo, 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 estrés asociado con los elementos de control seleccionados puede analizarse al introducir el casete de expresión en una línea celular adecuada (p. ej., células primarias, células transformadas o líneas celulares inmortalizadas).

5 Además, si bien no se necesitan para la expresión, las secuencias exógenas también pueden ser secuencias reguladoras de la transcripción o traducción, por ejemplo, promotores, potenciadores, aislantes, sitios de entrada a ribosomas internos, secuencias que codifican péptidos 2A, señales de poliadenilación y combinaciones de polipéptidos 2A y señales de poliadenilación. Además, los elementos de control de los genes de interés pueden enlazarse funcionalmente con genes indicadores para crear genes quiméricos (p. ej., casetes de expresión indicadores).

10 También puede lograrse la inserción dirigida de la secuencia de ácido nucleico no codificante. También pueden usarse las secuencias que codifican ARN, iARN, ARNhc y microARN (miARN) de antisentido para inserciones dirigidas.

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 escinde y modifica mediante la inserción de otra molécula donante de interés. De esta manera, pueden generarse integraciones repetidas de moléculas donantes, lo que permite el apilamiento de rasgos en un locus particular de interés o en un locus seguro.

15

### Suministro

Las nucleasas como se describen en la presente memoria (Tabla 1), los polinucleótidos que codifican estas nucleasas, los polinucleótidos donantes y composiciones que comprenden las proteínas, polinucleótidos y combinaciones de proteínas y polinucleótidos descritos en la presente memoria pueden suministrarse *in vivo* o *ex vivo* a través de cualquier medio adecuado a cualquier tipo de célula.

20

Las células adecuadas incluyen células eucariotas (p. ej., animales) y procariotas y líneas celulares eucariotas y procariotas. Los ejemplos no limitantes de dichas células o líneas celulares generadas a partir de dichas 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 también células de insectos tales como *Spodoptera frugiperda* (Sf), o células fúngicas tales 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 mesenquimatosas.

25

30

Se describen métodos para suministrar nucleasas como se describen en la presente memoria, por ejemplo, en las patentes estadounidenses núms. 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.

También pueden suministrarse nucleasas, construcciones donantes y combinaciones de nucleasas y construcciones donantes como se describen en la presente memoria mediante el uso de vectores que contienen secuencias que codifican una o más de la(s) ZFN descrita(s) en la presente memoria. Se pueden usar cualesquiera sistemas de vectores que incluyen, pero no se limitan a, vectores plasmídicos, vectores retrovíricos, vectores lentivíricos, vectores de adenovirus, vectores de virus de la viruela; vectores de virus del herpes y vectores de virus adenoasociados, etc. Ver, también, las patentes estadounidenses núms. 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 una construcción donante se introducen en la célula, las nucleasas, polinucleótidos donantes y combinaciones de nucleasas y polinucleótidos donantes pueden llevarse en el mismo vector o en vectores diferentes (MC de ADN). Cuando se usan múltiples vectores, cada vector puede comprender una secuencia que codifica una o múltiples nucleasas, una o más construcciones donantes y combinaciones de una o más nucleasas y una o más construcciones donantes. Los métodos de transferencia génica convencionales basados en virus o no basados en virus se pueden usar para introducir ácidos nucleicos que codifican nucleasas, construcciones donantes y combinaciones de nucleasas y construcciones donantes en células (p. ej., células de mamífero) y tejidos diana. Los sistemas de suministro de vectores no víricos incluyen plásmidos de ADN o ARN, MC de ADN, ácido nucleico no conjugado y ácido nucleico en complejo con un vehículo de suministro tal como un liposoma o poloxámero. Los vectores no víricos adecuados incluyen vectores de nanotaxis, que incluyen vectores disponibles comercialmente de InCellArt (Francia). Los sistemas de suministro de vectores víricos incluyen virus de ADN y ARN que tienen ya sea genomas episómicos o integrados después del suministro a la célula. Para un análisis del suministro *in vivo* de proteínas de unión a ADN y proteínas de fusión genomanipuladas que comprenden estas proteínas de unión, ver, p. ej., Rebar (2004) Expert Opinion Invest. Drugs 13(7):829-839; Rossi et al. (2007) Nature Biotech. 25(12): 1444-1454, así como referencias de suministro génico generales tales como Anderson, Science 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & 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

35

40

45

50

55

51(1):31-44 (1995); Haddada et al., in *Current Topics in Microbiology and Immunology* Doerfler and Böhm (eds.) (1995); y Yu et al., *Gene Therapy* 1:13-26 (1994).

Los métodos de suministro no vírico de ácidos nucleicos incluyen electroporación, lipofección, microinyección, biolística, virosomas, liposomas, inmunoliposomas, conjugados de lípidos:ácidos nucleicos o policonjugados, ADN no conjugado, viriones artificiales, deformación de la membrana y absorción de ADN potenciada por agente. También se puede usar sonoporación, p. ej., con el sistema Sonitron 2000 (Rich-Mar) para el suministro de ácidos nucleicos.

Los sistemas de suministro de ácido nucleico ilustrativos adicionales incluyen los proporcionados por Amara 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., las patentes estadounidenses núms. 5.049.386; 4.946.787; y 4.897.355) y los reactivos de lipofección se venden comercialmente (p. ej., Transfectam™ y Lipofectin™). Los lípidos catiónicos y neutros que son adecuados para la lipofección de reconocimiento de receptor eficaz de polinucleótidos incluyen los de Felgner, WO 91/17424, WO 91/16024.

Un experto en la técnica conoce la preparación de complejos de lípido:ácido nucleico, que incluyen liposomas dirigidos tales como complejos de inmunolípidos (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); las patentes estadounidenses núms. 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). Otros complejos de lípido:ácido nucleico incluyen los que comprenden lípidos catiónicos nuevos, lípidos pegilados nuevos y combinaciones de lípidos catiónicos nuevos y lípidos pegilados nuevos (ver, p. ej., la solicitud de patente provisional estadounidense núm. 62/432.042 y 62/458.373).

Los métodos de suministro adicionales incluyen el uso de envolver los ácidos nucleicos para suministrarlos en vehículos de suministro EnGeneIC (EDV). Estos EDV se suministran específicamente a tejidos diana mediante el uso de 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 celular diana y después, se introduce el EDV en la célula mediante endocitosis. Una vez que se encuentra en la célula, se libera el contenido (ver MacDiarmid et al. (2009) *Nature Biotechnology* 27(7):643).

El uso de sistemas basados en virus de ARN o ADN para el suministro de ácidos nucleicos que codifican sistemas de CRISPR/Cas, ZFP o TALE genomanipulados aprovechan los procesos muy evolucionados para dirigir un virus a células específicas en el cuerpo y transferir la carga vírica al núcleo. Los vectores víricos pueden administrarse directamente a los pacientes (*in vivo*) o pueden usarse para tratar células *in vitro* y administrar las células modificadas a pacientes (*ex vivo*). Los sistemas convencionales a base de virus para el suministro de ZFP incluyen, pero no se limitan a, vectores retrovíricos, lentivíricos, adenovíricos, adenoasociados, del virus de la vacuola de la viruela y del virus del herpes simple para la transferencia génica. La integración en el genoma hospedante es posible con los métodos de transferencia génica de retrovirus, lentivirus y virus adenoasociados, lo que frecuentemente genera la expresión a largo plazo del transgén insertado. De manera adicional, se han observado eficacias de transducción elevadas en muchos tipos de células y tejidos diana diferentes.

El tropismo de un retrovirus se puede alterar al incorporar proteínas de envoltura externas que expanden la posible población diana de células diana. Los vectores lentivíricos son vectores retrovíricos que son capaces de transducir o infectar células que no se dividen y típicamente producen títulos víricos elevados. La selección de un sistema de transferencia génica retrovírico depende del tejido diana. Los vectores retrovíricos están compuestos por repeticiones terminales largas que actúan en *cis* con capacidad de envoltura de hasta 6-10 kb de secuencia externa. Las LTR que actúan en *cis* mínimas son suficientes para la replicación y envoltura de los vectores, que después se usan para integrar el gen terapéutico en la célula diana para proporcionar una expresión de transgén permanente. Los vectores retrovíricos ampliamente usados incluyen los basados en el virus de leucemia murina (MuLV), virus de leucemia de mono gibbon (GaLV), virus de inmunodeficiencia en simios (SIV, por sus siglas en inglés), virus de la inmunodeficiencia humana (VIH), y combinaciones de estos (ver, p. ej., Buchscher et al., *J. Virol.* 66:2731-2739 (1992); Johann et al., *J. Virol.* 66:1635-1640 (1992); Sommerfelt et al., *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).

En las aplicaciones en las que se prefiere la expresión transitoria, se pueden usar sistemas basados en adenovirus. Los vectores adenovíricos son capaces de una eficacia de transducción muy elevada en muchos tipos de células y no requieren división celular. Con dichos vectores se han obtenido titulaciones elevadas y niveles elevados de expresión. Este vector puede producirse en grandes cantidades en un sistema relativamente simple. Los vectores de virus adenoasociado ("AAV") también se usan para la transducción de 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 vivo* (ver, p. ej., West et al., *Virology* 160:38-47 (1987); la patente estadounidense núm. 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 que incluyen la patente estadounidense núm. 5.173.414; Tratschin et al., *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, et al., *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); y Samulski et al., *J. Virol.* 63:03822-3828 (1989).

En la actualidad, se encuentran disponibles al menos seis abordajes con vectores víricos para la transferencia génica en ensayos clínicos, los cuales utilizan abordajes que implican la complementación de vectores defectuosos mediante genes insertados en líneas celulares auxiliares para generar el agente de transducción.

Tanto pLASN como MFG-S son ejemplos de vectores retrovíricos que se han usado 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 usado en un ensayo de terapia génica. (Blaese et al., Science 270:475-480 (1995)). Se han observado eficacias de transducción de 50 % o mayores para vectores envueltos con MFG-S (Ellem et al., Immunol Immunother. 44(1): 10-20 (1997); Dranoff et al., Hum. Gene Ther. 1:111-2 (1997)).

Los vectores de virus adenoasociados recombinantes (rAAV) son una alternativa prometedora de sistemas de suministro génico a base del virus adenoasociado de parvovirus de tipo 2 no patógeno y defectuoso. Todos los vectores derivan de un plásmido que conserva solamente las repeticiones terminales invertidas de AAV de 145 bp que flanquean el casete de expresión del transgén. La transferencia génica eficaz y el suministro estable del transgén debido a la integración en los genomas de la célula transducida son características clave para este sistema de vectores. (Wagner et al., Lancet 351:9117 1702-3 (1998); Kearns et al., Gene Ther. 9:748-55 (1996)). También pueden usarse otros serotipos de AAV, incluidos AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9 y AAVrh.10, y cualquier serotipo nuevo de AAV según la presente invención.

Los vectores adenovíricos (Ad) recombinantes con deficiencia de replicación se pueden producir con títulos elevados e infectar fácilmente múltiples tipos de células diferentes. La mayoría de los vectores de adenovirus están genomanipulados de modo que un transgén reemplaza los genes E1a, E1b o E3 de Ad; posteriormente, el vector de replicación defectuosa se propaga en células 293 humanas que suplen la función génica eliminada en *trans*. Los vectores de Ad pueden transducir múltiples tipos de tejidos *in vivo*, incluso células diferenciadas que no se dividen, tales como las que se hallan en el hígado, riñones y músculos. Los vectores de Ad convencionales tienen una gran capacidad de carga. Un ejemplo del uso de un vector de Ad en un ensayo clínico implicó la terapia con polinucleótidos para la inmunización antitumoral con inyección intramuscular (Stermann et al., Hum. Gene Ther. 7:1083-9 (1998)). Los ejemplos adicionales del uso de vectores de adenovirus para la transferencia génica en ensayos clínicos incluyen Rosenecker et al., Infection 24:1 5-10 (1996); Stermann et al., Hum. Gene Ther. 9:7 1083-1089 (1998); Welsh et al., 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); Stermann et al., Hum. Gene Ther. 7:1083-1089 (1998).

Se utilizan células de envoltura para formar partículas de virus que son capaces de infectar una célula hospedante. Dichas células incluyen células HEK293 y Sf9, que se pueden usar para envolver AAV y adenovirus, y células  $\psi$ 2 o células PA317, que envuelven retrovirus. Los vectores víricos usados en terapia génica usualmente se generan mediante una línea celular productora que envuelve un vector de ácido nucleico en una partícula vírica. Los vectores típicamente contienen las secuencias víricas mínimas necesarias para la envoltura y la posterior integración en un hospedante (si corresponde), donde otras secuencias víricas se reemplazan por un casete de expresión que codifica la proteína que se va a expresar. La línea celular de envoltura suple las funciones víricas faltantes en *trans*. Por ejemplo, los vectores de AAV usados en terapia génica típicamente poseen solamente secuencias de repeticiones terminales invertidas (ITR) del genoma de AAV que son necesarias para la envoltura y la integración en el genoma hospedante. El ADN vírico se envuelve en una línea celular que contiene un plásmido auxiliar que codifica los otros genes de AAV, a saber, *rep* y *cap*, pero que carece de secuencias de ITR. La línea celular también se infecta el adenovirus como auxiliar. El virus auxiliar promueve la replicación del vector de AAV y la expresión de genes de AAV a partir del plásmido auxiliar. El plásmido auxiliar no se envuelve en cantidades significativas debido a la falta de secuencias de ITR. La contaminación con adenovirus se puede reducir mediante, p. ej., tratamiento con calor al cual el adenovirus es más sensible que el AAV. En algunas realizaciones, el AAV se produce al usar un sistema de expresión de baculovirus (ver, p. ej., las patentes estadounidenses núms. 6.723.551 y 7.271.002).

La purificación de partículas de AAV a partir de un sistema de 293 o baculovirus típicamente implica el crecimiento de células que producen el virus, con posterior recolección de partículas víricas del sobrenadante celular o el lisado de células y la recolección de virus a partir del lisado bruto. Después, se purifica el AAV mediante métodos conocidos en la técnica que incluyen cromatografía de intercambio iónico (p. ej., ver las patentes estadounidenses núms. 7.419.817 y 6.989.264) cromatografía de intercambio iónico y centrifugación de densidad de CsCl (p. ej., la publicación de PCT WO2011094198A10), cromatografía por inmunoafinidad (p. ej., WO2016128408) o purificación mediante el uso de AVB Sepharose (p. ej., GE Healthcare Life Sciences).

En muchas aplicaciones de terapia génica, es deseable que el vector de terapia génica se suministre con un alto grado de especificidad a un tipo de tejido particular. Por consiguiente, se puede modificar un vector vírico para que tenga una especificidad para un tipo de célula dado al expresar un ligando como una proteína de fusión con una proteína de recubrimiento vírico en la superficie externa del virus. El ligando se selecciona de forma que tenga afinidad por un receptor que se sabe que está presente en el tipo de célula de interés. Por ejemplo, Han et al., Proc. Natl. Acad. Sci. USA 92:9747-9751 (1995), indicaron que el virus de leucemia murina de 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 virus-célula diana, en los cuales la célula diana expresa un receptor y el virus expresa una proteína de fusión que comprende un ligando para el receptor de superficie celular. Por ejemplo, se puede genomanipular el fago filamentoso

para visualizar en él fragmentos de anticuerpo (p. ej., FAB o Fv) que tienen afinidad de unión específica por prácticamente cualquier receptor celular elegido. Si bien la descripción que antecede se aplica principalmente a vectores víricos, los mismos principios se pueden aplicar a vectores no víricos. Dichos vectores se pueden genomanipular para contener secuencias de captación específicas que favorecen la captación por parte de células diana específicas.

Los vectores de terapia génica se pueden suministrar *in vivo* mediante la administración a un paciente individual, típicamente mediante administración sistémica (p. ej., infusión intravenosa, intraperitoneal, intramuscular, subdérmica o intracraneal) o aplicación tópica, tal como se describe más adelante. Alternativamente, los vectores se pueden suministrar a células *ex vivo*, tales como células explantadas de un paciente individual (p. ej., linfocitos, aspirados de médula ósea, biopsia de tejido) o células madre hematopoyéticas de donantes universales, con posterior reimplante de las células en un paciente, habitualmente después de la selección de células que tienen incorporado el vector.

Los vectores (p. ej., retrovirus, adenovirus, liposomas, etc.) que contienen nucleasas, construcciones donantes y combinaciones de nucleasas y construcciones donantes también se pueden administrar directamente a un organismo para la transducción de células *in vivo*. Alternativamente, se puede administrar ADN sin conjugar. La administración se realiza mediante cualesquiera de las vías usadas normalmente para introducir una molécula en contacto en última instancia con sangre o células tisulares que incluyen, pero no se limitan a, inyección, infusión, aplicación tópica y electroporación. Los métodos adecuados para administrar dichos ácidos nucleicos se encuentran disponibles y son conocidos para los expertos en la técnica y, si bien se puede usar más de una vía para administrar una composición particular, una vía particular a menudo puede proporcionar una reacción más inmediata y más eficaz que otra vía.

Los vectores adecuados para la introducción de polinucleótidos (p. ej., codificantes de nucleasa, donantes bicatenarios, y combinaciones de codificantes de nucleasa y donantes bicatenarios) descritos en la presente memoria incluyen vectores de lentivirus sin integración (IDLV). Ver, por ejemplo, Ory et al. (1996) Proc. Natl. Acad. Sci. USA 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; la publicación de patente estadounidense núm. 2009/0117617.

Los portadores farmacéuticamente aceptables se determinan en parte por la composición particular que se va a administrar, así como mediante el método particular usado para administrar la composición. Por consiguiente, existe una amplia variedad de formulaciones adecuadas de composiciones farmacéuticas disponibles, tal como se describe más adelante (ver, p. ej., Remington's Pharmaceutical Sciences, 17<sup>a</sup> ed., 1989).

Será evidente que las secuencias que codifican nucleasa y las construcciones donantes se pueden suministrar mediante el uso de sistemas iguales o diferentes. Por ejemplo, las nucleasas y donantes pueden transportarse en el mismo MC de ADN. Alternativamente, un polinucleótido donante puede ser transportado por un MC, mientras que la una o más nucleasas pueden ser transportadas por un plásmido estándar o vector de AAV. Además, los diferentes vectores pueden administrarse mediante vías iguales o distintas (inyección intramuscular, inyección en la vena de la cola, otras inyecciones intravenosas, administración intraperitoneal o inyección intramuscular). Los vectores pueden suministrarse simultáneamente o en cualquier orden secuencial.

Por lo tanto, la presente descripción incluye tratamiento *in vivo* o *ex vivo* de enfermedades y afecciones que admiten la inserción de un transgén que codifica 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 producirse a través de cualquier medio en el que se suministran polinucleótidos a las células diana deseadas. Por ejemplo, se contemplan tanto métodos *in vivo* como *ex vivo*. La inyección intravenosa a la vena porta es un método de administración preferido. Otros modos de administración *in vivo* incluyen, por ejemplo, la inyección directa en los lóbulos del hígado o el ducto biliar y la inyección intravenosa distal al hígado, que incluye a través de la arteria hepática, inyección directa en el parénquima hepático, inyección a través de la arteria hepática e inyección retrógrada a través del árbol biliar. Los modos de administración *ex vivo* incluyen la transducción *in vitro* de hepatocitos extirpados u otras células del hígado, con posterior infusión de los hepatocitos transducidos extirpados nuevamente a la vasculatura portal, parénquima hepático o árbol biliar del paciente humano, ver, p. ej., Grossman et al., (1994) Nature Genetics, 6:335-341.

La cantidad eficaz de nucleasa(s) y donante que se va a administrar varía de paciente a paciente y según el polipéptido terapéutico de interés. Por consiguiente, el médico que administra las composiciones determinará mejor las cantidades eficaces y un experto en la técnica puede determinar fácilmente las dosificaciones apropiadas. Después de un tiempo suficiente para permitir la integración y expresión (típicamente 4-15 días, por ejemplo), el análisis del suero u otros niveles en tejido del polipéptido terapéutico y la comparación con el nivel inicial antes de la administración determinará si la cantidad que se administra es demasiado baja, si se encuentra dentro del intervalo correcto o es demasiado elevada. Los regímenes adecuados para las administraciones iniciales y posteriores también son variables, pero se tipifican mediante una administración inicial seguida por administraciones posteriores, si son necesarias. Las administraciones posteriores pueden administrarse en intervalos variables, que varían de intervalos diarios, a anuales, o cada varios años. Un experto en la técnica apreciará que pueden recomendarse técnicas inmunosupresoras adecuadas para evitar la inhibición o el bloqueo de la transducción mediante inmunosupresión de los vectores de suministro; ver, p. ej., Vilquin et al., (1995) Human Gene Ther., 6:1391-1401.

Las formulaciones de las administraciones *ex vivo* e *in vivo* incluyen las suspensiones en líquido o líquidos emulsionados. Los ingredientes activos se mezclan a menudo con excipientes que son farmacéuticamente aceptables y compatibles con el ingrediente activo. Los excipientes adecuados incluyen, por ejemplo, agua, disolución salina, dextrosa, glicerol, etanol o similares y combinaciones de estos. Además, la composición puede contener cantidades menores de sustancias auxiliares, tales como, agentes humectantes o emulsionantes, agentes tampón de pH, agentes estabilizadores u otros reactivos que aumentan la eficacia de la composición farmacéutica.

## Células

También se describen en la presente memoria células y líneas celulares en las que se modifica una secuencia potenciadora de BCL11A endógena mediante las nucleasas descritas en la presente memoria (Tabla 1). La modificación puede ser, por ejemplo, en comparación con la secuencia natural de la célula. La célula o líneas celulares pueden ser heterocigóticas u homocigóticas con respecto a la modificación. Las modificaciones a la secuencia de BCL11A puede comprender indels.

La modificación es preferiblemente en o cerca del(de los) sitio(s) de unión a nucleasa(s), sitio(s) de escisión y combinaciones de sitio(s) de unión y sitio(s) de escisión, por ejemplo, dentro de 1-300 pares de bases (o cualquier valor entre estos) anteriores o posteriores al(a los) sitio(s) de escisión, más preferiblemente, dentro de 1-100 pares de bases (o cualquier valor entre estos) de cualquier lado del(de los) sitio(s) de unión, sitio(s) de escisión o sitio(s) de unión y sitio(s) de escisión, incluso más preferiblemente, dentro de 1 a 50 pares de bases (o cualquier valor entre estos) de cualquier lado del(de los) sitio(s) de unión, sitio(s) de escisión o sitio(s) de unión y sitio(s) de escisión. En ciertas realizaciones, la modificación es en o cerca de la región "+58" del potenciador de BCL11A, por ejemplo, en o cerca de un sitio de unión a nucleasa que se muestra en cualquiera de la primera columna de la Tabla 1.

Puede modificarse cualquier célula o línea celular, por ejemplo, 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 mesenquimatosa. Otros ejemplos no limitantes de células como se describen en la presente memoria incluyen linfocitos T (p. ej., CD4+, CD3+, CD8+, etc.); células dendríticas; linfocitos B. También se proporciona un descendiente de una célula madre, que incluye una célula diferenciada parcialmente o completamente (p. ej., un RBC o célula precursora de RBC). Los ejemplos no limitantes de otras líneas celulares que incluyen una secuencia de 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 también células de insectos tales como *Spodoptera frugiperda* (Sf), o células fúngicas tales como *Saccharomyces*, *Pichia* y *Schizosaccharomyces*.

Las células como se describen en la presente memoria son útiles en el tratamiento o prevención de un trastorno, por ejemplo, mediante terapias *ex vivo*. Las células modificadas por nucleasa puede expandirse y después volver a introducirse en el paciente mediante el uso de 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 describen composiciones farmacéuticas que comprenden las células como se describen en la presente memoria. Además, las células pueden criopreservarse antes de administrarse a un paciente.

Cualquiera de las células o líneas celulares modificadas descritas en la presente memoria pueden mostrar una mayor expresión de globina gamma. También se describen composiciones tales como composiciones farmacéuticas que comprenden las células modificadas genéticamente como se describen en la presente memoria.

## Aplicaciones

Los métodos y composiciones descritos en la presente memoria son para modificar la expresión de la proteína, o corregir una secuencia génica aberrante que codifica una proteína expresada en una enfermedad genética, tal como una anemia de células falciformes o una talasemia. Por lo tanto, los métodos y composiciones proporcionan el tratamiento o prevención de dichas enfermedades genéticas. Se puede usar edición genómica, por ejemplo, de células madre, para corregir un gen aberrante, insertar un gen natural o cambiar la expresión de un gen endógeno. A modo de ejemplo no limitante, puede insertarse un gen natural, p. ej., que codifica al menos una globina (p. ej., globina  $\alpha$ , globina  $\gamma$ , globina  $\beta$  y combinaciones de estas), en una célula (p. ej., en una secuencia potenciadora de BCL11A endógena mediante el uso de una o más nucleasas como se describen en la presente memoria) para proporcionar las proteínas globina deficientes o que faltan en la célula y, de esta forma, tratar una enfermedad genética, p. ej., una hemoglobinopatía, provocada por una expresión defectuosa de globina. Alternativamente o además, la edición genómica con o sin la administración del donante apropiado puede corregir el gen endógeno defectuoso, p. ej., corregir la mutación puntual en una hemoglobina  $\alpha$  o  $\beta$  para restaurar la expresión del gen o tratar una enfermedad genética, p. ej., anemia de células falciformes, inactivar o alterar (sobrexpresar o reprimir) cualquier gen regulador de globina directo o indirecto (p. ej., la inactivación del gen BCL11A que regula la globina  $\gamma$  o del KLF1 que regula BCL11A). Específicamente, las composiciones de la invención pueden usarse en el tratamiento o prevención de hemoglobinopatías.

Las nucleasas de la invención se dirigen a la región potenciadora de BCL11A, que se conoce como necesaria para la expresión de BCL11A durante la eritrocitopoyesis y, por lo tanto, la regulación por disminución de la expresión de globina gamma. La modificación de esta región potenciadora puede producir eritrocitos con una mayor expresión de globina gamma y, por lo tanto, puede ser útil para el tratamiento o prevención de la anemia de células falciformes o talasemia beta.

Los siguientes Ejemplos se relacionan con realizaciones ilustrativas de la presente descripción donde la nucleasa comprende una nucleasa de dedos de zinc (ZFN). Se apreciará que esto es a los efectos de ejemplificación solamente y que se pueden usar otras nucleasas, por ejemplo, sistemas de TtAgo y CRISPR/Cas, endonucleasas de asentamiento (meganucleasas) con dominios de unión a ADN genomanipulados, fusiones de endonucleasas de asentamiento genomanipuladas de origen natural (meganucleasas), dominios de unión a ADN, incluidas combinaciones de endonucleasas de asentamiento (meganucleasas) con dominios de unión al ADN genomanipulados y fusiones de endonucleasas de asentamiento genomanipuladas de origen natural (meganucleasas) dominios de unión a ADN, y dominios de escisión heterólogos, fusiones de meganucleasas y proteínas TALE, que incluyen combinaciones de dominios de escisión heterólogos y fusiones de meganucleasas y proteínas TALE.

## EJEMPLOS

### Ejemplo 1: Ensamblaje de nucleasas de dedos de zinc

Se ensamblaron ZFN contra el gen de BCL11A humano y se sometió a prueba la actividad mediante el análisis de secuenciación profunda de ADN aislado a partir de células transfectadas como se describe a continuación. Las ZFN específicas para la región +58 de la región potenciadora se elaboraron como se describe. El par de ZFN 51857/51949 se ha descrito anteriormente (ver WO 2016/183298).

### Ejemplo 2: Análisis inespecífico

Para analizar la escisión inespecífica mediante los pares de ZFN, se llevó a cabo un análisis de especificidad imparcial de dos etapas. En la primera etapa (Figura 1), los sitios inespecíficos candidatos para cada ZFN se identificaron a través de un ensayo de sitio de integración de estructura doble oligonucleotídica al usar un procedimiento similar al descrito por Tsai et al. ((2015), Nat Biotechnol 33(2): 187-197. doi: 10.1038/nbt.3117).

El ensayo de sitio de integración de estructura doble oligonucleotídica está basado en la observación de que la reintroducción de una nucleasa y un segmento corto de ADN de estructura doble en una célula diana da como resultado una integración de estructura doble durante la reparación de una fracción de eventos de escisión del genoma mediante la vía de reparación de ADN NHEJ (Orlando et al., (2010), Nucleic Acids Res, 38(15) e152. doi: 10.1093/nar/gkq512; Gabriel et al., (2011), Nat Biotechnol. 2011 Aug 7;29(9):816-23. doi: 10.1038/nbt.1948; Tsai et al., *ibid*). Después de la integración, la estructura doble proporciona una etiqueta permanente del evento de escisión. Después se identifican los sitios de integración mediante la ligación de un adaptador oligonucleotídico al ADN genómico con cizallamiento, seguido de 2 pasadas de 25 ciclos de PCR anidado y secuenciación profunda de las uniones de donante-genoma resultantes. Este ensayo permite la evaluación de todos los sitios de integración potenciales dentro del genoma.

El ensayo de sitio de integración se llevó en células K562 para maximizar el suministro del donante, la expresión de ZFN y la integración del donante. Asimismo, dado que las células K562 se dividen rápidamente (tiempo duplicado en aproximadamente 24 horas), se espera que impongan restricciones epigenéticas mínimas en la capacidad de las ZFN para escindir dianas celulares. Las células ( $2 \times 10^5$ ) se sometieron a electroporación con 0,47 µg de donante de estructura doble oligonucleotídica y 400 ng de cada ARNm codificante de ZFN al usar un transportador Amaxa y las configuraciones se optimizaron para una actividad en la diana máxima de las ZFN. Se prepararon cuatro muestras por replicación para cada combinación de oligo y ARNm. A los días 7 de la transfección, se aisló el ADN genómico para cada muestra (kit de sangre y tejido Qiagen Dneasy) y se usó 400 ng (133000 genomas haploides) como entrada para el protocolo de amplificación descrito en la Figura 1. Después, las muestras se procesaron esencialmente como se describió (Tsai et al. *ibid*). Los productos finales se agruparon, se cuantificaron y se secuenciaron en un instrumento MiSeq (Illumina) mediante el uso de un kit de secuenciación v2 de 300 ciclos con lecturas de 150 bp de extremos apareados y lecturas de índices dobles de 8 bp/16 bp para detectar los códigos de barra de muestra en cada extremo del amplicón.

Para generar una lista de sitios inespecíficos candidatos, los datos de la secuenciación se filtraron para la secuencia de cebado correcta, seguido por el corte de secuencias de adaptador y el mapeo en el genoma. A continuación, se mapearon las coordenadas de unión y la unión de la estructura doble-genoma, así como la posición del corte causado por el cizallamiento del ADN, se usaron para identificar eventos de integración distintos. Después, se procesaron los eventos de integración para identificar aglomeraciones de integraciones próximas dentro del genoma (mínimo de 4 eventos de integración distintos dentro de 100 bp entre sí, en todas las replicaciones). Las aglomeraciones que estaban en los contigios que no se pudieron mapear en el ensamblaje hg38 (es decir, chrUn en hg38) se retiraron de los análisis adicionales. Las aglomeraciones que se mapearon en loci repetitivos (una mediana de tres o más resultados en el genoma en todas las secuencias en una aglomeración) también se eliminaron ya que la experiencia previa indica que son artefactos de amplificación. Las aglomeraciones restantes se puntuaron como sitios de escisión de ZFN

candidatos si derivaron de al menos 2 muestras tratadas con ZFN replicadas (de 4 en total) y presentaron un exceso de  $\geq 5$  veces de los eventos de integración en las muestras tratadas con ZFN en comparación con los testigos. Los sitios de escisión candidatos se clasificaron por el número total de integraciones únicas en las muestras tratadas con ZFN. Los loci candidatos identificados a través de este análisis se proporcionan en la Figura 2 para el par de ZFN 51857/51949, clasificados por recuento de integrantes.

### Ejemplo 3: Optimización de las ZFN

Para reducir la escisión inespecífica, se adoptó una estrategia para la optimización de nucleasas en la que los contactos con fosfato no específicos se eliminan selectivamente para causar la supresión global de la escisión inespecífica (Guilinger et al. (2014) Nat Methods. 11(4):429-35. doi: 10.1038/nmeth.2845; Kleinstiver et al. (2016) Nature 529(7587):490-5. doi: 10.1038/nature16526; Slaymaker et al. (2016) Science 351(6268):84-8. doi: 10.1126/science.aad5227) (ver las solicitudes provisionales estadounidenses núms. 62/443.981 y 62/378.978). Las sustituciones de aminoácidos se hicieron en una posición clave dentro del marco de dedos de zinc que interactúa con la estructura de fosfato del ADN (Pavletich and Pabo, (1991) Science 252(5007):809-17; Elrod-Erickson et al. (1996) Structure 4(10): 1171-80) (Figura 3A-3B), así como en una posición única en el dominio de FokI de ZFN derecho que también se predijo que hace un contacto con fosfato (Figura 3C).

La especificidad se mejoró, además, al permitir la expresión independiente de cada ZFN de dos ARNm separados, lo que posibilita la optimización de las relaciones de suministro. Estos esfuerzos proporcionaron pares de ZFN optimizados que están altamente relacionados con el original, y difieren en sustituciones que reducen la energética de la interacción con la estructura de fosfato de ADN, pero que afectan mínimamente o nada el reconocimiento de base específico de la secuencia. En consonancia con esto, el ensayo del sitio de integración proporcionó 455 loci para dianas potenciales de escisión de ZFN para el par original 51857/51949. Para el par optimizado, se identificó una cantidad de loci mucho menor para examinar adicionalmente como dianas potenciales de la escisión por ZFN (72 en total) mediante este análisis. Para ambos pares, la diana prevista dentro del potenciador de BCL11A fue el locus con mejor clasificación. Además, se advirtió una fracción mucho más alta de eventos de integración en el potenciador de BCL11A para el par optimizado, en consonancia con su mayor especificidad.

Es importante advertir que, al definir el canal de procesamiento de datos de secuencia, los parámetros clave se eligieron de manera conservadora, para incluir la mayor cantidad de loci inespecíficos candidatos posibles en vez de filtrarlos. Esto se hizo para garantizar la identificación y análisis de que cada locus que pueda representar un sitio de escisión confiable para las ZFN optimizadas en estudios de indel de seguimiento, incluso con el costo de aceptar una cantidad mucho mayor que podría convertirse en falsos positivos. Se esperó que la primera etapa del análisis proporcionara un conjunto grande de loci candidatos para cada par de ZFN, del cual, la gran mayoría (particularmente para las ZFN optimizadas) no representaría sitios de escisión inespecíficos reales, sino eventos de fondo que tendrían un resultado negativo para la escisión en estudios de indel de seguimiento.

En la segunda etapa del análisis, los loci inespecíficos candidatos identificados a través del ensayo de sitio de integración se analizaron para detectar evidencia de modificaciones (p. ej., la presencia de indels) en HSPC CD34+ tratadas con ZFN.

En particular, HSPC CD34+ humanas derivadas de sangre periférica movilizada se trataron con los pares de ZFN originales y optimizados al usar condiciones clínicas y de escala clínica para la transfección de ARN (120  $\mu\text{g}/\text{mL}$  de ARNm para el par de ZFN original y 100  $\mu\text{g}/\text{mL}$  de ARNm para el par optimizado). El ADN genómico se aisló 2 días después de la transfección, con posterior amplificación por PCR de los loci inespecíficos candidatos y secuenciación profunda para cuantificar los niveles de indel. Para los pares de ZFN originales y optimizados, se analizó el mismo conjunto de 137 loci inespecíficos candidatos en esta etapa, junto con una cantidad más pequeña de sitios inespecíficos candidatos que se habían identificado a través de otros métodos en estudios anteriores con los ZFN originales.

Los resultados mostraron que las ZFN optimizadas son marcadamente más específicas que el par original. Esto es aparente no solo por la cantidad de loci que se clasificaron como positivos para la evidencia de escisión por ZFN (52 para el par original frente a 3 para el par optimizado), sino también por los niveles de indel observados, que para el par optimizado fueron mucho menores. La Figura 4 muestra gráficas de valores de indel en cada locus que exhibe evidencia de escisión por ZFN en este estudio (cabe señalar la escala logarítmica del eje y). La agregación de indels inespecíficos en todos los loci indica una reducción en la actividad inespecífica de 300 veces (46,5 % de indels inespecíficos agregado para el par original, frente a 0,15 % de indels inespecíficos para el par optimizado). Esta reducción en la actividad inespecífica se logró sin pérdida de actividad en el sitio diana previsto (72,5 % de indels para el par original frente a 81,9 % para las ZFN optimizadas). En estos estudios, el par original (o par principal) fue 51857/51949, mientras que el par de ZFN optimizadas fue 63014/65722 (ver a continuación).

Los diseños de nucleasa se muestran a continuación en la Tabla 1:

**Tabla 1: Pares de ZFN específicos para la región potenciadora de BCL11A +58**

SBS núm. (sitio diana, 5'-3')	Diseño [Secuencia de hélice, SEQ ID]						Enlazador
	[Mutaciones en la estructura de dedo]						Mutantes de Fok
	F1	F2	F3	F4	F5	F6	
Compañero izquierdo							
<b>51857</b> aaAGCAACTG TTAGCTTGCA Ctagacta (SEQ ID NO:1)	DQSNLRA (SEQ ID NO:19)	RNFSLTM (SEQ ID NO:20)	STGNLTN (SEQ ID NO:21)	TGSLTR (SEQ ID NO:22)	DQSNLRA (SEQ ID NO:19)	AQCCLFH (SEQ ID NO:23)	L7c5
	ninguno	ninguno	ninguno	ninguno	ninguno	ninguno	ELD
<b>63014</b> aaAGCAACTG TTAGCTTGCA Ctagacta (SEQ ID NO:1)	DQSNLRA (SEQ ID NO:19)	RNFSLTM (SEQ ID NO:20)	STGNLTN (SEQ ID NO:21)	TGSLTR (SEQ ID NO:22)	DQSNLRA (SEQ ID NO:19)	AQCCLFH (SEQ ID NO:23)	L7c5
	Qm5	ninguno	Qm5	ninguno	Qm5	ninguno	ELD
<b>65459</b> aaAGCAACTG TTAGCTTGCA Ctagacta (SEQ ID NO:1)	DQSNLRA (SEQ ID NO:19)	RNFSLTM (SEQ ID NO:20)	STGNLTN (SEQ ID NO:21)	TGSLTR (SEQ ID NO:22)	DQSNLRA (SEQ ID NO:19)	AQCCLFH (SEQ ID NO:23)	L7c5
	Qm14Qm5	ninguno	Qm5	ninguno	Qm5	ninguno	ELD
Compañero derecho							
<b>51949</b> caCAGGCTCC AGGAAGGgtt tgccctct (SEQ ID NO:2)	RNDHRTT (SEQ ID NO:24)	QKAHLIR (SEQ ID NO:25)	QKGTIGE (SEQ ID NO:26)	RGRDLR (SEQ ID NO:26)	RRDNLHS (SEQ ID NO:27)	N/A	LO
	ninguno	ninguno	ninguno	ninguno	ninguno	N/A	KKR
<b>65722</b> caCAGGCTCC AGGAAGGgtt tgccctct (SEQ ID NO:2)	RNDHRTT (SEQ ID NO:24)	QKAHLIR (SEQ ID NO:25)	QKGTIGE (SEQ ID NO:26)	RGRDLR (SEQ ID NO:26)	RRDNLHS (SEQ ID NO:27)	N/A	LO
	Qm5	Qm5	ninguno	Qm5	ninguno	N/A	KKR K525S
<b>65526</b> caCAGGCTCC AGGAAGGgtt tgccctct (SEQ ID NO:2)	RNDHRTT (SEQ ID NO:24)	QKAHLIR (SEQ ID NO:25)	QKGTIGE (SEQ ID NO:26)	RGRDLR (SEQ ID NO:26)	RRDNLHS (SEQ ID NO:27)	N/A	LO



SBS núm. (sitio diana, 5'-3')	Diseño [Secuencia de hélice, SEQ ID]						Enlazador
	[Mutaciones en la estructura de dedo]						Mutantes de Fok
	F1	F2	F3	F4	F5	F6	
Compañero izquierdo							
	Qm5	Qm5	ninguno	Qm5	Qm5	N/A	KKR R416S
<b>65549</b> caCAGGCTCC AGGAAGGgtt tggcctct (SEQ ID NO:2)	RNDHRTT (SEQ ID NO:24)	QKAHLIR (SEQ ID NO:25)	QKGTIGE (SEQ ID NO:26)	RGRDLR (SEQ ID NO:26)	RRDNLHS (SEQ ID NO:27)	N/A	LO
	Qm5	Qm5	Qm5	Qm5	ninguno	N/A	KKR K525S
<b>65550</b> caCAGGCTCC AGGAAGGgtt tggcctct (SEQ ID NO:2)	RNDHRTT (SEQ ID NO:24)	QKAHLIR (SEQ ID NO:25)	QKGTIGE (SEQ ID NO:26)	RGRDLR (SEQ ID NO:26)	RRDNLHS (SEQ ID NO:27)	N/A	LO
	Qm5	Qm5	ninguno	Qm5	Qm5	N/A	KKR K525S

La Tabla 1 muestra información de caracterización relacionada con cada ZFN. Al empezar desde la izquierda, el número de SBS (p. ej., 51857) se muestra con la diana de ADN a la que se une la ZFN para mostrarse debajo del número de SBS. A continuación, se muestran los diseños de hélice de reconocimiento de aminoácidos para los dedos 1-6 o 1-5 (columna subdividida 2 de la Tabla 1). En la Tabla 1 también se muestra debajo de los diseños de hélices apropiados, las mutaciones hechas a las secuencias de estructuras de ZFN del dedo indicado, como se describe en las solicitudes de patente provisional estadounidenses núms. 62/378.978 y 62/443.981. En la notación usada en la Tabla 1, "Qm5" significa que en la posición **menos 5** (relativa a la hélice que se enumera -1 a +6) del dedo indicado, la arginina en esta posición se ha remplazado con una glutamina (**Q**), mientras "Qm14" significa que la arginina (R) normalmente presente en la posición **menos 14** se ha remplazado con una glutamina (**Q**). "Ninguno" indica sin cambios fuera de la región de hélice de reconocimiento. Por lo tanto, por ejemplo, SBS núm. 63014 incluye la mutación Qm5 en los dedos 1, 3 y 5 mientras que los dedos 2, 4 y 6 no tienen mutaciones en la estructura de dedo de zinc (p. ej., la secuencia de dedo de zinc fuera de la región de hélice de reconocimiento).

Finalmente, la columna más a la derecha de la Tabla 1 muestra el enlazador usado para enlazar el dominio de unión a ADN al dominio de escisión de FokI (p. ej., "L7c5" (LRGSISRARPLNPHP (SEQ ID NO:5), como se describe por ejemplo en la patente estadounidense núm. 9.567.609) se muestra en la línea superior de la columna, con los sitios de las mutaciones de contacto de fosfato de FokI y mutaciones de dimerización mostradas en la casilla debajo de la designación del enlazador. Específicamente, en la línea superior de la casilla de mutantes de FokI se indica el tipo de mutación encontrada en el dominio de dimerización (p. ej., ELD o KKR como se describe, por ejemplo, en la patente estadounidense núm. 8.962.281). Debajo de las designaciones de mutante de dimerización se muestran cualesquiera mutaciones presentes en el dominio de FokI hechas para eliminar el contacto de fosfato no específico mostrado en la parte inferior (p. ej., K525S o R416S donde los residuos serina en las posiciones de aminoácidos 525 o 416 se han sustituido para lisina o arginina, respectivamente como se describe en las solicitudes de patente provisional estadounidenses núms. 62/378.978 y 62/443.981). Por lo tanto, por ejemplo, en la SBS núm. 63014, el enlazador es un enlazador L7c5 y el dominio de escisión de FokI incluye los mutantes de dimerización de ELD y ninguna mutación de contacto de fosfato. Además, para la SBS núm. 65722, el enlazador es un enlazador L0 (LRGSQLVKS (SEQ ID NO:6), también denominado como el enlazador "estándar", ver patente estadounidense núm. 9.567.609) y el dominio de escisión de FokI incluye las mutaciones de dimerización y la mutación de contacto de fosfato de FokI K525S.

Todos las ZFN se analizaron para determinar su funcionalidad KICR (actividad de escisión como se determinó mediante el ensayo de indels como se describe en el Ejemplo 4 a continuación) y se halló que eran activas.

Además, para determinar qué diseños de ZFN fueron los más específicos, se realizaron los análisis de indel de los sitios conocidos de escisión inespecífica por el par de ZFN original en las HSPC CD34+ tratadas con ZFN. Para lograr esto, HSPC CD34+ humanas derivadas de sangre periférica movilizada se trataron con los pares de ZFN originales y optimizados al usar condiciones clínicas y concentraciones de ARNm (120 µg/mL de ARNm para el par de ZFN original y 100 µg/mL para el par optimizado). El ADN genómico se aisló 2 días después de la transfección de estas células y los testigos no tratados, con posterior amplificación por PCR de cada locus candidato y secuenciación profunda para cuantificar los niveles de indel.

Los niveles de modificación en cada locus se determinaron mediante una secuenciación profunda con extremos apareados en un Illumina MiSeq al usar un cartucho de 300 ciclos. Las secuencias apareadas se fusionaron, el adaptador se cortó con SeqPrep filtrado para una puntuación de calidad de ≥15 en todas las bases y después se mapeó en el genoma humano (ensamble hg38). Las secuencias que se mapearon en un locus incorrecto se descartaron. Las secuencias más cortas que el amplicón natural de >70 bp o >70 % se retiraron para minimizar los productos de cebador-dímero. Se llevó a cabo una alineación de Needleman-Wunsch (Needleman and Wunsch, (1970), J Mol Biol 48(3):443-53) entre el amplicón diana y cada lectura de MiSeq para mapear indels. Los indels en las secuencias alineadas se definieron como se describe en Gabriel *et al.* 2011 (*ibid*), excepto que los indels de 1 bp de longitud también se aceptaron para evitar contar de menos los eventos reales. Cabe señalar que una fracción de loci no se amplificó o no se secuenció, o se rechazaron del análisis debido a su fondo alto (>1 % de modificación en las muestras testigo) o profundidad de secuenciación insuficiente (<10000 lecturas). Los resultados de este análisis y la comparación con el par de ZFN 51857/51949 "original" se proporcionan a continuación en la Tabla 2.

**Tabla 2: Análisis de escisión inespecífica**

Dímero de ZFN				Inespecífico/a: núm. de loci			núm. de loci OT positivos para indel		
Izquierdo	Derecho	µg ARN I:D	% de indels de BCL11A	diana	por PCR	Analizable	P<0,05.	manual	¿En original? Captura/confirmado
51857	51949	60:60	73,0	23	21	17	15	17	-
63014	65722	60:15	82,2	31	24	14	0	4	4/3

Dímero de ZFN				Inespecífico/a: núm. de loci			núm. de loci OT positivos para indel		
Izquierdo	Derecho	µg ARN l:D	% de indels de BCL11A	diana	por PCR	Analizable	P<0,05.	manual	¿En original? Captura/confirmado
63014	65526	60:15	81,2	23	22	9	3	4	4/4
63014	65527	60:60	81,4	30	24	10	4	4	4/2
63014	65549	60:60	80,0	30	24	13	0	2	2/1
63014	65550	60:60	79,8	30	24	9	0	1	1/1
65459	65526	60:15	76,9	23	19	14	0	2	2/2

#### Ejemplo 4: Actividad de ZFN en células CD34+ humanas

Para las pruebas *in vitro*, se evaluaron las nucleasas en células CD34+. Las ZFN se suministraron como ARNm, donde los ARNm se elaboraron *in vitro* de la manera que sigue: los plásmidos que comprenden los genes que codifican las ZFN se linealizan y usan para la transcripción de ARNm *in vitro* al usar un Kit mMessage mMachine® T7 Ultra (Ambion/Applied Biosystems). El ARNm después se purificó al usar un kit mini de RNeasy® (Qiagen).

Las células CD34+ se aislaron a partir de la sangre periférica movilizada y se mantuvieron en el medio X-VIVO 10 complementado con penicilina, estreptomycinina y glutamina, así como StemSpan CC110 y se incubaron en 37 °C y CO<sub>2</sub> al 5 %. Las células se transfectaron 48 horas después del aislamiento o después de la descongelación. Se mezcló una pequeña alícuota 1:1 con disolución de azul de tripano al 0,4 % (p/v) en PBS (Corning) y se determinaron las cantidades de células en un contador celular automatizado TC20 (Bio-Rad).

Para las transfecciones a gran escala, las células se lavaron con tampón de electroporación MaxCyte (Maxcyte) y se resuspendieron a 3 a 5e7 células por mL en tampón de electroporación en 100 µL. Típicamente, se usaron concentraciones de ARNm entre 60 µg/mL y 120 µg/mL para analizar conjuntos de ZFN candidatos. Después, las células se cultivaron en medios de cultivo a 3e6 células por mL durante 18 horas a 30 °C y después se diluyeron hasta 1e6 células por mL durante 24 horas adicionales a 37 °C. Para la determinación de la actividad de escisión, se aisló el ADN genómico 2-3 días después de la transfección, y se midió el nivel de modificación génica en el locus de potenciador de BCL11A a través de secuenciación profunda en un secuenciador MiSeq (Illumina).

Los pares de ZFN de la Tabla 1 se analizaron en células CD34+ y los resultados de la actividad se muestran a continuación en la Tabla 3.

**Tabla 3: Actividad de pares de ZFN contra la diana BCL11A**

ZFN de la derecha	ZFN de la izquierda	Conc. de ZFN de la derecha (µg)	Conc. de ZFN de la izquierda (µg)	Indels (%)
51857	51949	60	60	72,98
63014	65722	15	60	80,62
63014	65722	60	15	82,19
63014	65526	60	15	81,22
63014	65527	60	60	81,43
63014	65549	60	60	79,82
63014	65550	60	60	79,96
Testigo con GFP				0,07

Además de analizar la actividad de nucleasa en las células CD34+ antes de la diferenciación eritroide, las células editadas también se diferenciaron *in vitro* en células eritroides. El protocolo que se siguió se basó en Giarratana et al., ((2011) Blood 120 (15):2945-53). En resumen, se siguió el protocolo que sigue a continuación:

- 5      Día 0 a Día 7: Se cultivaron células CD34+  $4 \times 10^4$  a una densidad de  $2 \times 10^4$  /mL en un medio de diferenciación (EDM) (Medio de Dulbecco modificado por Iscove [IMDM, por sus siglas en inglés], 330 µg/mL de Transferrina, 10 µg/mL de Insulina humana, 2 U/mL de Heparina sódica, plasma AB+ humano la 5 %) en presencia de  $10^{-6}$  M hidrocortisona, 100 ng/mL factor de células madre (SCF, por sus siglas en inglés), 5 ng/mL de IL 3 y 3 IU/mL eritropoyetina (EPO).

Día 4: Las células se resuspendieron en EDM nuevo que contenía SCF, IL-3, EPO e hidrocortisona.

- 10     Día 7 a Día 11: Las células se resuspendieron a una densidad de  $1,5 \times 10^5$  células por mL nuevo de EDM complementado con SCF y EPO.

Día 11 a Día 21: El día 11, las células se volvieron a poner en placas a  $1 \times 10^6$  /mL en EDM nuevo complementado con EPO. Las células posteriormente se volvieron a colocar en placas posteriormente en estos mismos medios a  $5 \times 10^6$  /mL el día 14. El crecimiento llegó a una meseta durante este período de tiempo entre el día 14 a 18, cuando la viabilidad celular empieza a caer hasta la terminación de los cultivos el día 21.

- 15

Los recuentos celulares se tomaron en el momento de la siembra y a lo largo de toda la diferenciación al medir la positividad de naranja de Acridina y la exclusión de yoduro de propidio (AOPI) al usar un Nexcelom Bioscience Cellometer K2 con el modo de ensayo eritroide AOPI con canal de fluorescencia 1 (AO) fijado en 700 milisegundos y canal de fluorescencia 2 (PI) fijado en 5000 milisegundos.

- 20     El porcentaje de células enucleadas se determinó el día 21 de la diferenciación al usar el siguiente protocolo. La tasa de enucleación fue semejante entre los testigos no transfectados y las muestras transfectadas con ZFN con porcentajes de 59-63 % de estos dos grupos:

1. Recuento celular

2. 100.000 células, centrifugación a 450 x g, 5 min, TA.

- 25     3. Resuspender en 50 µL de PBS-BSA + 1 µL de GlyA-FITC (DAKO).

4. Tinción durante 15 min en refrigerador.

5. Agregar 1 mL de PBS-BSA, agitación vorticial, centrifugación.

6. Resuspender en 250 µL de PBS-BSA-NucRed (2 gotas NucRed por mL).

7. Adquirir en FACS Canto al usar el canal APC para NucRed.

- 30     8. Las células eritroides nucleadas estarán en la fracción baja/negativa de NucRed positiva para GlyA y los eritoblastos estarán en la fracción positiva GlyA-NucRed doble.

La modificación génica de BCL11A se midió mediante secuenciación profunda en MiSeq en las muestras de ADN recogidas a) 48 horas después de la electroporación b) el día de descongelar las células, en el momento en que se inició la diferenciación *in vitro* y c) el día 14 de la diferenciación eritroide *in vitro*. Si bien la diferenciación se llevó a cabo durante 21 días, se eligió el punto temporal del día 14 para el análisis de ADN ya que es anterior a la enucleación de una fracción grande de las células eritroides, lo que resulta en una pérdida de recuperación de ADN. Los porcentajes de modificación observados en el potenciador de BCL11A están indicados en la Tabla 4 junto con los detalles de las condiciones de transfección.

- 35

**Tabla 4: Niveles de modificación del gen de BCL11A mediante análisis por MiSeq**

Prep de células CD34+	Transfección	Modificación del gen de BCL11A (%)		
		Día 2 después de TF	Posdescongelación	Día 14 de la diferenciación
Prep núm. 1	80 µg/mL 63014 + 20 µg/mL 65722	78,6	81,4	72,0
	Sin transfectar	0,1	0,2	0,2
Prep núm. 2	80 µg/mL 63014 + 20 µg/mL 65722	75,4	77,3	72,3
	Sin transfectar	0,1	0,1	0,0

Estos datos muestran que la transfección de células CD34+ con ARNm del par 63014 y 65722 optimizado conducen a una modificación génica muy eficaz en el sitio diana del potenciador de BCL11A (>75 % de los alelos modificados) y que la modificación se mantiene muy bien (>90 % de retención de la modificación) después de congelar y descongelar las células y después de la diferenciación eritroide.

Se seleccionó el par de ZFN 63014/65722 para un análisis adicional. Las secuencias de aminoácidos para estas ZFN se muestran a continuación, donde cada una comprende una señal de localización nuclear (NLS, Kalderon et al. (1984) Cell 39 (3 Pt 2):499-509) y un péptido hidrófilo (Hopp et al. (1988) Nat Biotechnol 6:1204-10) que potencia la actividad de ZFN en la diana, ambos fusionados en la secuencia codificante de extremo N. Por lo tanto, las secuencias de ARNm y aminoácidos de las ZFN son las siguientes:

ARNm de 63014 (1725 nt)  
5'gggagacaagcuuugaauuacaagcuugcuuguucuuuuugcagaagcucagaauaaacgcucaacuugg  
cagaucgaaauucgccaugggacuacaagacccaugacggugauuuauaaagaucaugacaucgaaauacaaggau  
acgaugacaagaugggccccaagaagaagaggaaggucggcauccacgggguacccggccgcuauggcugagag  
gcccuccagugucgaaucugcaugcagaacuucagugaccaguccaaccugcgcgccacauccgcacccaca  
ccggcgagaagccuuuugccugugacauuugugggaggaaauuugccgcaacuucuccugaccaugcauac  
caagauacacacgggcagccaaaagcccuuccagugucgaaucugcaugcagaacuucaguuccaccggcaacc  
ugaccaaccacaucgcacccacaccggcgagaagccuuuugccugugacauuugugggaggaaauuugccac  
cuccggcuuccugacccgccauaccaagauacacacgcacccgcgccccgauccgaagcccuuccaguguc

gaaucugcaugcagaacuucagugaccaguccaaccugcgcgcccacaucggcaccacaccggcgagaagccu  
uuugccugugacauuuuggggaggaaauuugccgcccaguguuugucuguuccaccauaccaagauacaccugc  
ggggauccaucagcagagccagaccaguccaaccgcccagcugggaggagaagaaguccagcugcgccac  
aagcugaaguacugccccacgaguacaucgagcugaucgagcaggaacagcaccagaccgcauccu  
ggagauagaaggugauggaguucuucaugaaggguacggcuacaggggaaagcaccugggcggaagcagaaa  
gccugacggcgccaucuauacagugggcagccccaucgauuacggcgugaucguggacacaaaggccuacagc  
ggcgguacaauucgcuacggccagggcagcagagauaggagagauacguggaggagaaccagaccgggga  
agcaccucaaccccaacgaguggggaaggguuacccuagcagcugaccgaguucaaguuccugugag  
cggccacuuaagggcaacuacaagggccagcugaccaggcugaaccacaucaaacugcaaugggcgugc  
ugagcugggaggagcugcugaucggcgcgagaucaaaagccggcaccugacacugggaggaggugcggc  
gcaaguucaacaaggcgagaucaacuucagauuagauaacucgagucuaagcucgcuuucugcugucc  
aaauucuaauaaagguuccuuuuguccuaaguccaacuacuaaacugggggauauuaugaaggccuugagc  
aucuggaauucgcucaauaaaaaacauuuuuuucuuugcugcgcuagaagcucgcuuucugcuguccaau  
ucuauuaaagguuccuuuuguccuaaguccaacuacuaaacugggggauauuaugaaggccuugagcaucu  
ggauucugccuauaaaaaacauuuuuuucuuugcugcgggacauucuaauuaaaaaaaaaaaaaaaaaa  
aaacuaag

(SEQ ID NO:28).

Secuencia de aminoácidos de 63014 (las regiones de hélice de reconocimiento están subrayadas; el enlazador se muestra en mayúsculas y cursiva; las mutaciones en los residuos de la estructura de los dedos 1, 3 y 5 se muestran con subrayado doble; las mutaciones del dominio de dimerización (ELD) se muestran en negrita y cursiva; el péptido hidrófilo está indicado en el texto en minúsculas; y la señal de localización nuclear (NLS) se muestra en minúsculas y cursiva):

MdykdhdgdykdhdidykdddkMApkkkrkvGIHGVPAAMAERPFCRICMQNFSDQS  
NLRAHIRTHTGEKPFACDICGRKFARNFSLTMHTKIHTGSQKPFQCRICMQNF  
SSTGNLTNHIRTHTGEKPFACDICGRKFATSGSLTRHTKIHTHPRAPIKPFQCR  
ICMQNFSDQSNLRAHIRTHTGEKPFACDICGRKFAAQCCLFHHTKIHLRGSISR  
ARPLNPHPELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMK  
VYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEME  
RYVEENQTRDKHLNPNEWVKVYPSSVTEFKFLVSGHFKGNKYAQLTRLNH  
ITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFRS (SEQ ID

NO:29).

ARNm de 65722 (1680 nucleótidos):

5'gggagacaagcuugaauacaagcuugcuuguucuuuuugcagaagcucagaauaaacgcuaacuuuggca  
gaucgaauucgcuagagauucggcgcgaggaggcagagggaagucucuaaaccugcgugacguggagga

gaaucccgcccuaggaccauggacuacaaagaccaugacggugauuuaaagaucgacaucaugauuacaagg  
 augacgaugacaagauggccccaagaagaagagggaaggucggcauucguggguacccgcccgaugggcuga  
 gagggccuuccagugucgaauucgcaugcagaaguuugcccgcacgaccaccgaccaccacauaccaagauac  
 acacgggcgagaagcccuuccagugucgaauucgcaugcagaacuucagucagaaggcccaccugauccggcac  
 auccgcacccacaccggcgagaagccuuuugccugugacauuugugggaggaaaauugcccagaaggggcacc  
 ugggcgagcauaccaagauacacacgggaucucagaagcccuuccagugucgaauucgcaugcagaacuucag  
 ucggcgccgagaccugucccgccacauccgcacccacaccggcgagaagccuuuugccugugacauuuguggg  
 aggaaaauugcccgcgacacaccugcacuccauaccaagauacaccugcggggaucacgucgugaagag  
 cgagcuggaggagaagaaguccgagcugcggcacaagcugaaguacgugcccacgaguacaucgagcugauc  
 gagaucgcccaggaacagcaccagaccgcauccuggagauagaaggugauggauguuucgagaaggguuacg  
 gcuacaggggaaagcaccugggcggaagcagaagccugacggcgccaucuacagugggcgagcccaucga  
 uuacggcgugaucguggacacaaaggccuacagcgggcgcuacaauucgcuacggccaggccgacgagaug  
 cagagauacgugaaggagaaccagaccgggaauaagcacaauaccccaacgagugggggaagguguacccuag  
 cagcgugaccgaguuaaguuccugugagcgccacuucagcggaacuacaaggcccagcugaccagg  
 cugaaccgcacaaacugcaauugcgccgugcugagcgugggaggagcugcugaucggcgcgagaucauca  
 aagccggcaccucgacuggaggaggugcgcgcaaguuaacaacggcgagaucaacuucgauaacucga  
 gucuagaagcugcuuuucugcuguccaaauucuaauuaaagguuccuuuguuccuaaguccaacuacuaaac  
 uggggggaauuaugaaggccuugagcaucuggaucugccuaauaaaaaacauuuuuuucgugcgcg  
 uagaagcugcguuuucugcuguccaaauucuaauuaaagguuccuuuguuccuaaguccaacuacuaaacug  
 gggauuuuaugaaggccuugagcaucuggaucugccuaauaaaaaacauuuuuuucgugcggggac  
 auucuuuaauuaaacuag  
 (SEQ ID NO:30).

Secuencia de aminoácidos de 65722 (las regiones de hélice de reconocimiento están subrayadas; el enlazador se muestra en mayúsculas y cursiva; el péptido hidrófilo está en minúsculas; la señal de localización de la nucleasa está en minúsculas y cursiva; las mutaciones a los residuos de la estructura de los dedos 1, 2 y 4 se muestran con subrayado doble; las mutaciones del dominio de dimerización (ELD) se muestran en negrita y cursiva; y la(s) mutación(es) de contacto con fosfato de FokI se muestra(n) con un subrayado ondulado):

MdykdhdgdykdhdidykdddkMApkkkrkvGIHGVPAMAERPFQCRICMQKFARND  
HTTHTKIHTGEKPFQCRICMQNFSQKAHLIRHIRTHTGEKPFACDICGRKFAQ  
KGTLGEHTKIHTGSQKPFQCRICMQNFSRGRDLSRHIRTHTGEKPFACDICGR  
KFARRDNLHSHTKIHLRGSQVLKSELEKKSELRHKLKYPHEYIELIEIARNST  
 QDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKA  
 YSGGYNLPIGQADEMQRYV**KENQ**TRNKHINPNEWWKVYPSSVTEFKFLFVS  
 GHFSGNYKAQLTRLN**RK**TNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFN

NGEINF (SEQ ID NO:31).

#### Ejemplo 5: Evaluación de los niveles de globina en la progenie eritroide

Los niveles de ARNm de globina  $\alpha$ ,  $\beta$  y  $\gamma$  en ARNm celular aislado el día 14 de la diferenciación (antes de que los niveles de ARNm generales bajen radicalmente en el transcurso de la enucleación y maduración eritroide) se determinaron mediante RT-qPCR para las dos prep celulares mostradas anteriormente en la Tabla 4. Los valores de ARNm de globina  $\gamma$  se muestran normalizados en relación con el ARNm de globina  $\beta$  (Figura 5A) o con los valores de

ARNm de globina  $\alpha$  (Figura 5B) de las mismas muestras (al usar unidades arbitrarias basadas en la relación en el estándar de RT-PCR sin transfectar definido como 1).

La HPLC de fase inversa de las muestras de proteína aisladas el día 21, el punto final de la diferenciación de eritroides, se usó para determinar si la modificación mediada por ZFN del potenciador eritroide de BCL11A aumenta la hemoglobina fetal a nivel de la proteína. Se determinaron las relaciones entre globina gamma (suma de los picos de Agamma y Ggamma) y globina alfa, así como las relaciones entre globina gamma (suma de los picos de Agamma y Ggamma)/ con respecto a globina tipo beta (suma de los picos de Agamma, Ggamma, globina beta y delta) y se muestran en la Figura 6.

En este experimento, se observó una evaluación de aproximadamente 3-4 veces de los porcentajes de proteína globina fetal hasta niveles de aproximadamente 15 %-20 % en la progenie eritroide de HSPC después de la alteración mediada por 63014/65722 del potenciador de BCL11A.

#### **Ejemplo 6: Injerto de células editadas en ratones NSG**

Después, se inyectaron las células CD34+ humanas editadas en ratones NSG para evaluar el injerto. La extensión del quimerismo humano (es decir, el porcentaje de células CD45+) se midió al usar clasificación de células activadas por fluorescencia (FACS, por sus siglas en inglés) en la sangre periférica recogida a las 8 y 12, 16 y 20 semanas después del trasplante, y en la médula ósea recogida en la semana 12 y en la semana 20. Además, para evaluar el nivel de injerto de las células modificadas por ZFN, se evaluó el nivel de alteración génica en el locus del potenciador de BCL11A mediante la secuenciación de alto rendimiento directa del locus diana de ZFN y se comparó con los niveles de modificación génica diana medida en el material de entrada.

Las HSPC de dos donantes sanos (llamados PB-MR-003 y PB-MR-004) se movilizaron con G-SCF y Plerixafor y se purificaron como se describe en Yannaki et al. ((2012) Mol Ther 20(1):230-8. doi: 10.1038/mt.2011). El agotamiento de plaquetas se llevó a cabo en el producto de leucaferesis al usar el dispositivo Fresenius-Kabi Lovo antes de enriquecerlo con células CD34+ al usar el instrumento Miltenyi Biotech CliniMACS Plus. Las células purificadas después se sembraron en el cultivo para la transfección.

Dos días después de la purificación de células CD34+, las células se sometieron a electroporación al usar el instrumento Maxcyte en la presencia de 120  $\mu\text{g/mL}$  de un mRNA simple que codifica el par de ZFN original, 63014/65722 o cantidades optimizadas de dos ARNm separados que codifican el par de ZFN optimizado 80  $\mu\text{g/mL}$  de 63014 y 20  $\mu\text{g/mL}$  de 65722. Antes de la transfección, se separó una alícuota de las células como el testigo sin transfectar. Se transfectaron 95 millones de células de PB-MR-003 y se transfectaron 120 millones de células de PB-MR-004.

Después de la electroporación, se llevó a cabo un cultivo transitorio durante la noche a 30 °C y las células después se cultivaron 24 horas adicionales a 37 °C. Dos días después de la electroporación, se tomaron alícuotas de células para el análisis de ADN y las células restantes se recogieron, crioconservaron y almacenaron en nitrógeno líquido.

Acondicionamiento: Los ratones se trataron con 10 mg/kg/día de agua Baytril 1-2 días antes de la irradiación y se irradiaron subletalmente con 300 RAD 16-24 horas antes del trasplante. El trasplante se llevó a cabo a través de una inyección en la vena de la cola (ver a continuación). Después, los ratones recibieron agua Baytril fresca. El agua Baytril se reemplazó una semana después y se interrumpió la adición de agua Baytril 14 días después del trasplante.

Trasplante: El día del trasplante, preparar cóctel de X-Vivo 10/ PSG al 1 % + 3 citocinas precalentado (Factor de célula madre humano recombinante (SCF), Trombopoyetina humana recombinante (TPO), y Ligando Flt-3 humano recombinante (Flt-3L)) a 37 °C, PBS nuevo/BSA al 0,1 % a temperatura ambiente (estéril/filtrado). Las células crioconservadas se descongelaron a 37 °C, se sedimentaron, se resuspendieron en un medio X-Vivo precalentado, se volvieron a sedimentaron, se resuspendieron en PBS/BSA al 0,1 % y se contaron. Después de otra sedimentación, el sedimento celular se resuspendió en 550  $\mu\text{L}$  por ratón de PBS/BSA al 0,1 % ( $2 \times 10^6$  células/mL basándose en el recuento celular). Las células después se inyectaron a temperatura ambiente en la vena de la cola del ratón con una aguja de calibre 25. Los grupos de estudio se muestran a continuación en la Tabla 5.



**Tabla 5: Grupos de dosificación para el injerto de células hCD34+ editadas**

Especie: Ratón				Sexo, edad, cantidad: 60 ratones hembra NSG			
Núm. de grupo	N / Grupo	Artículo de prueba	% de indels (Día 2)	Viabilidad 1d después de descong.	Dosis (células/ratón)	N / Sacrificio	
						Semana 12	Semana 20
1	10 (M1-10)	Células de donante PB-MR-003, tratadas con ARNm de 63014 y 65722.	79 %	83 %	1 millón	5	5
3	10 (M21-30)	PB-MR-003 sin transfectar	0,1 %	95 %	1 millón	5	5
4	10 (M31-40)	Células de donante PB-MR-004 tratadas con ARNm de 63014 y 65722.	75 %	77 %	1 millón	5	5
6	10 (M51-60)	PB-MR-004 sin transfectar	0,1 %	92 %	1 millón	5	5

Los animales se observaron a diario para determinar su salud general y se pesaron a diario durante las primeras 2 semanas y después se pesaron cada dos semanas. Se recogió sangre periférica de la vena submandibular (100 µL) a las 8, 12, 16 y 20 semanas después del trasplante o a través de punción cardíaca (1 mL) para los animales sacrificados a las 12 y 20 semanas después del trasplante. La mitad de los animales en cada grupo (5 ratones por grupo) se sacrificaron a las 12 semanas después del trasplante y se recogieron la médula ósea y la sangre terminal para su análisis. Los animales restantes en cada grupo (5 ratones por grupo) se sacrificaron a las 20 semanas después del trasplante.

Recolección de sangre, cultivo celular y procesamiento: Se recogió sangre periférica a través de la vena submandibular o punción cardíaca en tubos EDTA y se centrifugó a 500 x g durante 5 min para retirar el plasma. Después del lavado con seroalbúmina bovina (BSA) disolución salina tamponada con fosfato (PBS) y la centrifugación, se agregó un volumen de 10X de tampón hemolítico al sedimento, y la mezcla se incubó a 37 °C durante 15 min, se centrifugó y se volvió a lavar. La fracción sedimentada se reconstituyó en 1 mL de PBS BSA; se retiró una alícuota y se centrifugó a 1000 x g durante 5 min, con el sedimento resultante conservado para la genotipificación. La fracción de sobrenadante se utilizó para análisis de FACS.

Se recolectó médula ósea, huesos de fémur, tibia y pelvis en medio de Dulbecco modificado de Iscove (IMDM) que contenía suero fetal de ternero (FCS); la médula ósea total se enjuagó en una disolución de PBS BSA y se filtró al usar un colador de nylon de 70 µm. El volumen se ajustó hasta 10 mL con PBS BSA y se usó una alícuota para el recuento celular (Cellometer).

Se analizó la actividad de ZFN al usar secuenciación profunda por MiSeq. En resumen, se aisló el ADN genómico de ratones inyectados con HSPC CD34+ testigo sin transfectar o HSPC CD34+ transfectadas con ARNm de ZFN dirigido al potenciador a partir de muestras de sangre obtenidas a las 8 semanas y 12 semanas o a partir de médula ósea a las 12 semanas después de la inyección. La región de interés (que contiene el sitio de unión a ZFN dentro del locus de BCL11A) se amplificó mediante PCR y se determinó el nivel de modificación mediante secuenciación profunda de extremos apareados en la plataforma Illumina (MiSeq).

Para generar bibliotecas compatibles con la plataforma de secuenciación Illumina MiSeq, se acoplaron adaptadores, códigos de barra y aglutinante de células de flujo (secuencia de ADN corta) a los amplicones específicos diana mediante el uso de dos conjuntos de cebadores de fusión en PCR secuenciales. Para las evaluaciones por MiSeq de modificación del potenciador de BCL11A humano en las muestras de sangre y médula ósea de ratón, el protocolo tuvo que ajustarse debido a las cantidades bajas de ADN diana en estas muestras.

Se usaron los siguientes cebadores para la PCR de adaptador de MiSeq:

PRJIYLFN-f2: ACA CGA CGC TCT TCC GAT CTN NNN AGT CCT CTT CTA CCC CAC CCA (SEQ ID NO:32) y

PRJIYLFN-r4: GAC GTG TGC TCT TCC GAT CTC TAC TCT TAG ACA TAA CAC ACC AGG G (SEQ ID NO:33).

5 Para el análisis, se aisló el ADN a partir de las muestras de médula ósea de ratón con DNeasy y se usaron  
aproximadamente 100 ng de ADN en cada reacción de PCR. El ADN de las muestras de sangre de ratón se aisló con  
Tissue XS y se usó 10 µL de los 15 µL de ADN aislado en cada reacción. Además del ADN, se agregó lo siguiente a  
la reacción de PCR por MiSeq: 25 µL de mezcla Taq HotStar (Qiagen), 0,5 µL de cada uno de los cebadores del  
potenciador de BCL11A indicados anteriormente (en una concentración de 100 nM), y agua hasta un volumen total de  
10 reacción de 50 µL. Las condiciones de PCR MiSeq típicas fueron: desnaturalización a 95 °C durante 15' y 30 ciclos a  
94 °C durante 30", 62 °C durante 30" y 72 °C durante 40", con posterior alargamiento durante 10' a 72 °C. Después  
de la PCR con MiSeq, el producto de PCR se diluyó entre 1:50 y 1:200 con agua, o se dejó sin diluir para las muestras  
con cantidades de células de partida muy bajas. Se realizó la PCR de código de barras con 1 µL del producto de PCR  
con MiSeq diluido como se describió anteriormente, 25 µL de mezcla Taq HotStar, 1 µL de cebador de código de  
15 barras directo, 1 µL de cebador de código de barras inverso (ambos en una concentración de 10 nM) y agua hasta un  
volumen total de reacción de 50 µL. Las condiciones de PCR de código de barras fueron: desnaturalización a 95 °C  
durante 15' y 18 ciclos a 94 °C durante 30", 60 °C durante 30" y 72 °C durante 30", con posterior alargamiento durante  
10' a 72 °C. Los productos de PCR de código de barras se agruparon y se secuenciaron en el secuenciador Illumina  
MiSeq. Los resultados se muestran en la Tabla 5 anterior.

20 Análisis FACS para el quimerismo y la determinación de linaje de células. Para evaluar el grado de quimerismo humano,  
la fracción de células en la sangre periférica (a las 8, 12, 16 y 20 semanas después del injerto) y la médula ósea (a las  
12 y 20 semanas después del injerto) se tiñeron con anticuerpos hCD45-APC Cy7 (Biolegend) y hCD45-BV510 (BD  
Biosciences) respectivamente y se realizó el análisis FACS. Además, se realizó el análisis de linajes hematopoyéticos  
al teñir las células de médula ósea con los anticuerpos específicos descritos en la Tabla 6 a continuación:

**Tabla 6: Fuentes de anticuerpo para marcadores celulares**

<b>CD3-FITC: BD 561807 (clon UCHT1)</b>	BD 561807
<b>CD19-PE: BD 340364 (clon SJ25C1)</b>	BD 340364
<b>CD45-BV510</b>	BD 563204
<b>Lin-APC (CD3/UCHT1, CD14/HCD14, CD16/3G8, CD19/HIB19, CD20/2H7, CD56/HCD56)</b>	BIOLEGEND 348803
<b>CD33-PE-CF594</b>	BD 562492
<b>GlyA-FITC</b>	DAKO 0870
<b>CD38-PerCP Cy5,5</b>	BD 551400
<b>CD14-PE</b>	BD 555398
<b>CD34-PE Cy7</b>	BD 560710
<b>CD71-APC Cy7/H7</b>	BD 563671
<b>CD15-BV650</b>	BD 564232
<b>CD8-PerCP Cy5,5</b>	BD 341051
<b>CD4-PE Cy7</b>	BD 344612
<b>CD56-APC</b>	BD 318310
<b>IgM-APC Cy7</b>	BD 314520
<b>CD20-BV650</b>	BD 563780

Además, para purificar y clasificar las poblaciones de HSPC, se usó una estrategia de enriquecimiento/agotamiento al usar separación celular magnética (MACS, por sus siglas en inglés). Las células de médula ósea se tiñeron primero con CD19-biotina, CD3-biotina, B220-biotina, TER119-biotina y m-ckit-biotina (BD Biosciences) y después se incubaron con perlas antibiotina (Miltenyi Biotec). La fracción positiva y la fracción agotada se separaron al usar columnas de LS (Miltenyi Biotec) colocadas en el campo magnético de un MACS. Después de la separación, la fracción positiva se tiñó con estreptavidina-APC, CD3-FITC, CD19-PE, CD45-BV510 (BD Biosciences) y la fracción agotada con CD34-FITC (BD Biosciences), Gly-A-PE (DAKO), CD19-APC (BD), Lin-APC (Biolegend), estreptavidina-APC, CD45-BV510, CD33-PE-CF594 (BD) y CD38-PECy-7 (Biolegend).

Las HSPC sin transfectar y las HSPC transfectadas con 63014/65722 se injertaron en ratones NSG al usar procedimientos estándares como se describieron anteriormente. El grado de quimerismo humano en estos ratones después del injerto se evaluó al medir la fracción de células positivas para hCD45 al usar FACS.

La Figura 7 muestra los porcentajes de células CD45+ humanas en sangre periférica recogida a las 8 y 12 semanas después del trasplante y la Figura 8 muestra porcentajes en médula ósea recogida en la Semana 12. Como se muestra, los niveles de injerto en este estudio fueron de un quimerismo humano semejante después del injerto de HSPC no transfectadas y HSPC transfectadas con 63014/65722. Solo 3 ratones de 60 distribuidos a lo largo de los grupos no tenían células CD45+, lo que indica una falla en el injerto.

Se evaluó la reconstitución de varios linajes de células hematopoyéticas mediante análisis FACS de células de médula ósea obtenidas en la semana 12 con anticuerpos que reconocen marcadores de superficie celular específicos de linaje al usar procedimientos estándares. Como se muestra en la Figura 9, se observó una representación semejante de todos los linajes hematopoyéticos humanos en la médula ósea en la semana 12 después de la inyección entre la progenie de células CD34+ tratadas con ARNm de ZFN específica para BCL11A y la de células sin transfectar. La médula ósea de los ratones sacrificados en la Semana 12 después del injerto se aisló y se analizó la distribución de varios linajes hematopoyéticos mediante FACS al usar anticuerpos que reconocen los marcadores de linaje indicados. Todos los números se dan como la relación entre células que tienen una tinción positiva para el marcador de linaje indicado frente al porcentaje de células positivas para CD45 humano, excepto para las células que expresan el marcador eritroide Cd71+ (Ter119) en la Figura 9C, que se dan como el porcentaje de células con tinción positiva en toda la población, ya que las células eritroides no son positivas para CD45.

Los niveles de modificación génica en el potenciador eritroide de BCL11A (% de alelos con inserciones y eliminaciones [indels]) se evaluaron mediante secuenciación profunda de la región diana de ZFN al usar la plataforma de secuenciación MiSeq como se describió anteriormente. Los datos se muestran en la Figura 10 para muestras de sangre de la semana 8 y semana 12 y en la Figura 11 para las muestras de médula ósea de la semana 12 y linajes clasificados derivados de las muestras de células de médula ósea de la semana 12 de las células tratadas con 63014/65722. Para la comparación, los porcentajes de indel medidos 2 días después de la transfección (como se indicó en la Tabla 5) también se muestran en las gráficas de las Figuras 10 y 11.

Además, se halló buena retención de la modificación génica en el potenciador eritroide de BCL11A para conjuntos de donantes HSPC tratadas con 63014/65722 en varios puntos temporales y en varios linajes. Se observó una modificación semejante en linajes dependientes de BCL11A (linfocitos B, "CD19"; progenitores primitivos, "CD38H") y linajes independientes de BCL11A (mieloide "CD33"). Si bien los niveles de modificación de genes de entrada fueron más altos en la muestra de donante PB-MR-003 que en la muestra de donante PB-MR-004, las células derivadas de PB-MR-004 mostraron de manera constante niveles de modificación más altos, es decir, una mejor retención de la modificación, en ratones que los derivados de PB-MR-003.

En general, la retención observada de la modificación génica en el potenciador eritroide de BCL11A estuvo en consonancia con la observada en experimentos de ratón previos al usar varias ZFN que se dirigen a una variedad de dianas génicas.

Además, como los progenitores eritroides humanos no son capaces de diferenciarse en ratones, para determinar la cantidad de modificación génica diana de BCL11A que se produjo en estas células, las células de médula ósea se eliminaron de los ratones y se diferenciaron *in vitro*. En estos experimentos, las células humanas derivadas de médula ósea se eliminaron de ratones sacrificados en la semana 12 después del injerto y se diferenciaron *in vitro* como se describió anteriormente. La modificación del gen diana de BCL11A se midió con una secuenciación por MiSeq de alto rendimiento del ADN aislado a partir de células en el día 14 de la diferenciación.

Los datos de modificación (indels) se presentaron en la Figura 12, que muestra los niveles de modificación el día 14 de la diferenciación eritroide. Los porcentajes de indel en el día 14 de la diferenciación *in vitro* varían marcadamente para cada cultivo que se generó a partir de las células aisladas de un ratón, lo que refleja la naturaleza oligocelular de la expansión obtenida en estas condiciones. Los datos indican que la modificación del potenciador de BCL11A mediada por las ZFN 63014/65722 no se alteró marcadamente durante la diferenciación eritroide. Como se observó en las muestras de sangre y médula ósea, las muestras de progenie eritroide de las células derivadas de PB-MR-004 mostraron niveles promedio más altos de modificación que la progenie eritroide de células derivadas de PB-MR-003.

Los niveles relativos de varios ARNm de globina se determinaron mediante análisis de RT-PCR de ARN aislado a partir de células el día 14 de la diferenciación eritroide *in vitro*, y los datos se presentan en la Figura 13A donde las relaciones entre ARNm de globina  $\gamma$  y globina  $\beta$  y ARNm de globina  $\gamma$  y globina  $\alpha$  relativas (Figura 13B) se promediaron para los 5 cultivos eritroides de cada grupo. Las relaciones entre ARNm de globina  $\gamma$  y globina  $\beta$  o globina  $\gamma$  y globina  $\alpha$  de las muestras tratadas con 63014/65722 y las sin transfectar difirieron ampliamente entre las progenies eritroides de ratones individuales del mismo grupo. Los cultivos derivados de PB-MR-004 de donante muestran relaciones de globina  $\gamma$  en promedio más bajas que los del donante PB-MR-003, en consonancia con la mejor maduración observada para las muestras derivadas de PB-MR-004. Sin embargo, a pesar de esta variabilidad, los promedios de muestras tratadas con ZFN muestran un aumento de ~1,5-2 en los niveles de ARNm de globina  $\gamma$  en comparación con sus contrapartes sin transfectar respectivas.

Se evaluaron los niveles de proteína globina mediante análisis por HPLC. La Figura 14 muestra análisis de proteína globina de muestras recogidas el día 16 de la diferenciación. Se determinaron las relaciones entre globina gamma (suma de los picos de Agamma y Ggamma) y alfa globina, así como las relaciones entre globina gamma (suma de los picos de Agamma y Ggamma)/ con respecto a globina tipo beta (suma de los picos de Agamma, Ggamma, globina beta y delta) y los promedios para cada grupo se muestran sobre cada barra. En consonancia con la escasa diferenciación de eritroides de las muestras derivadas de PB-MR-003, los niveles de globina gamma en las células no transfectadas derivadas de este donante fueron muy altos (~30 %) y, por lo tanto, el tratamiento con ZFN dio como resultado un aumento de solo 1,2 veces en los niveles de globina gamma. PB-MR-004 mostró niveles más típicos no transfectados (~9 %) y exhibió un aumento de ~2 veces en los niveles de proteína globina gamma después de 12 semanas de pasaje por el ratón.

Se cree que los pacientes que tienen un >8,6 % de globina  $\gamma$  naturalmente están en ventaja en comparación con los pacientes con niveles de globina  $\gamma$  <8,6 % (Platt et al. (1994) N Engl J Med, 330:1639-44). De hecho, lograr un porcentaje de 10-20 % quimérico de RBC de células no falciformes a través del injerto de células editadas puede conducir a la mejora clínica (Chang et al. (2017) Mol Ther Methods Clin Dev 4:137-148. doi 10.1016/j.omtm.2016.12.009). Por lo tanto, a pesar de tener que pasar por un proceso de diferenciación eritroide *in vitro*, el porcentaje de células quiméricas y el nivel de proteína globina  $\gamma$  detectada indican la eficacia terapéutica.

## REIVINDICACIONES

1. Una nucleasa de dedos de zinc (ZFN) que comprende la secuencia de aminoácidos como se muestra en la SEQ ID NO:29 y la secuencia de aminoácidos como se muestra en la SEQ ID NO:31.
- 5 2. Uno o más polinucleótidos que codifican una o más nucleasas de dedos de zinc según la reivindicación 1, opcionalmente en donde el uno o más polinucleótidos es(son) ARNm.
3. El uno o más polinucleótidos de la reivindicación 2 que comprenden la SEQ ID NO: 28 y la SEQ ID NO:30.
4. Uno o más vectores que comprenden secuencias que codifican una o más nucleasas de dedos de zinc según la reivindicación 1.
- 10 5. Una célula que comprende la nucleasa de dedos de zinc de la reivindicación 1, el uno o más polinucleótidos de la reivindicación 2 o la reivindicación 3 o el uno o más vectores de la reivindicación 4.
6. La célula de la reivindicación 5, en donde:
  - (i) la célula es una célula madre o célula precursora, opcionalmente en donde la célula es una célula humana; y/o
  - (ii) el genoma de la célula se modifica mediante la nucleasa de dedos de zinc, opcionalmente en donde la modificación genómica se selecciona del grupo que consiste en inserciones, eliminaciones y combinaciones de esta, tal como una
- 15 7. Una célula o línea celular producida a partir de la célula de la reivindicación 5 o la reivindicación 6, que comprende la nucleasa de dedos de zinc de la reivindicación 1, el uno o más polinucleótidos de la reivindicación 2 o la reivindicación 3 o el uno o más vectores de la reivindicación 4.
8. Una célula parcialmente o completamente diferenciada que descende de la célula o línea celular de una cualquiera de las reivindicaciones 5 a 7, que comprende la nucleasa de dedos de zinc de la reivindicación 1, el uno o más polinucleótidos de la reivindicación 2 o la reivindicación 3 o el uno o más vectores de la reivindicación 4.
- 20 9. La célula o línea celular de una cualquiera de las reivindicaciones 5 a 8, en donde el genoma de la célula se modifica mediante la nucleasa de dedos de zinc y la célula o línea celular exhibe mayor expresión de globina gamma y/o beta en comparación con una célula o línea celular sin la modificación genómica.
- 25 10. Una composición farmacéutica que comprende la nucleasa de dedos de zinc según la reivindicación 1, el uno o más polinucleótidos de la reivindicación 2 o la reivindicación 3, el uno o más vectores de la reivindicación 4 y/o la célula o línea celular de una cualquiera de las reivindicaciones 5-9.
11. La composición farmacéutica de la reivindicación 10, que comprende la célula o línea celular de una cualquiera de las reivindicaciones 5-9 y que comprende además células propagadas y diferenciadas que descienden de la célula o
- 30 línea celular.
12. Un método *in vitro* para modificar una secuencia potenciadora de BCL11A endógena en una célula, el método comprende administrar la nucleasa de dedos de zinc según la reivindicación 1 o el uno o más polinucleótidos de la reivindicación 2 o reivindicación 3 a la célula de modo que se modifique la secuencia potenciadora de BCL11A endógena,
- 35 opcionalmente en donde:
  - (i) el método comprende además introducir una secuencia exógena en la célula, de modo que la secuencia exógena se inserte en la secuencia potenciadora de BCL11a endógena; o
  - (ii) la modificación comprende una eliminación.
13. La composición farmacéutica de la reivindicación 10 o la reivindicación 11 para su uso en un método para tratar una hemoglobinopatía, tal como una talasemia (p. ej., talasemia beta) o anemia de células falciformes, en un paciente, el método comprende administrar la composición farmacéutica al paciente, opcionalmente, en donde el paciente es un ser humano.
- 40 14. Un método *in vitro* para producir una célula modificada genéticamente que comprende una modificación genómica dentro de una secuencia potenciadora de BCL11A endógena, el método comprende las etapas de:
  - a) poner en contacto una célula con el uno o más polinucleótidos de la reivindicación 2 o reivindicación 3;
  - b) someter la célula a condiciones que lleven a la expresión de la proteína de fusión a partir del polinucleótido; y
  - c) modificar la secuencia potenciadora de BCL11A endógena con la proteína de fusión expresada de manera suficiente para producir la célula modificada genéticamente,
- 45

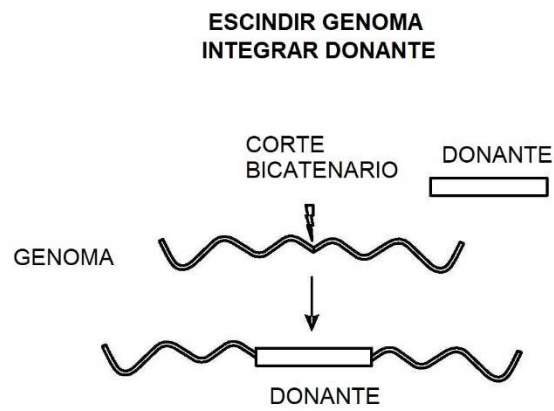
opcionalmente en donde el método comprende además estimular a la célula con al menos una citocina.

15. Un kit que comprende la nucleasa de dedos de zinc de la reivindicación 1, el uno o más polinucleótidos de la reivindicación 2 o la reivindicación 3, el uno o más vectores de la reivindicación 4 y/o la célula o línea celular de una cualquiera de las reivindicaciones 5-9.

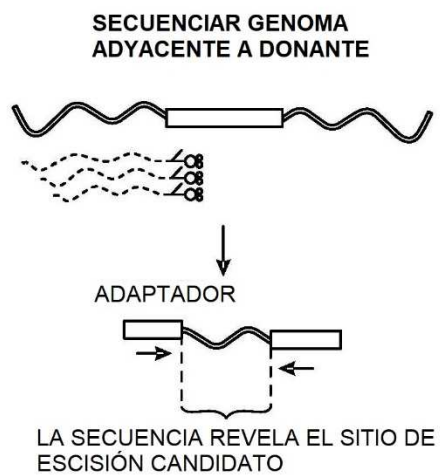
- 5      16. La célula o línea celular de cualquiera de las reivindicaciones 5-9 para su uso en un método para tratar o prevenir una hemoglobinopatía, tal como una talasemia (p. ej., talasemia beta) o anemia de células falciformes.

17. La célula o línea celular para su uso según la reivindicación 16, en donde la célula es una célula madre o célula progenitora humana y el genoma de la célula se modifica dentro de la secuencia potenciadora de BCL11A mediante la nucleasa de dedos de zinc, opcionalmente en donde:

- 10      (i) las células se expanden antes de la administración a un sujeto;
- (ii) las células se administran al sujeto en un trasplante de médula ósea, donde se injertan, se diferencian y maduran in vivo; y/o
- (iii) antes de la introducción de un injerto que comprende las células, se trata al sujeto con un procedimiento mieloablativo leve o un régimen de acondicionamiento mieloablativo vigoroso.



**FIG. 1A**



**FIG. 1B**

CROMOSOMA	N.º BASE	RECuento DE INTEGR.
2	60495266	1064
8	119856442	939
2	62164814	156
10	132654828	105
14	67422056	102
4	7829724	93
2	23702832	88
14	73740920	85
1	29182166	82
10	69571176	75
16	15141540	68
8	94988042	66
16	2122340	66
19	45406680	64
16	67184172	59
16	46394744	58
13	75549350	57
16	34593646	57
16	46390726	54
1	54506814	53
19	55115766	53
6	53805956	52
17	29624234	52
1	26938300	48
5	145864942	40
20	37707466	39
21	33308800	36
1	204975770	35
5	132776212	35
11	33894468	34
10	101133378	33
17	28854518	33
X	30280458	31
2	84459580	30
7	131503656	30
2	112583534	27
3	49724756	27
12	55928112	27

CROMOSOMA	N.º BASE	RECuento DE INTEGR.
22	20508058	15
1	167637598	14
1	211655462	14
9	129611194	14
16	46399310	14
19	38665086	14
14	50504	14
14	50604	14
2	96873178	13
2	107763998	13
2	127820122	13
6	1596312	13
6	35468034	13
6	144064118	13
8	107386748	13
10	98910092	13
12	104048816	13
14	89466992	13
1	21635648	12
2	135453242	12
6	7899998	12
8	13105966	12
9	128604032	12
11	2992880	12
13	91576712	12
14	18295748	12
16	34587146	12
21	25607680	12
22	15226546	12
1	70765014	11
1	120942294	11
1	200702950	11
3	19942238	11
5	2057482	11
6	3196958	11
9	98755462	11
11	30406776	11
1	177943758	10

CROMOSOMA	N.º BASE	RECuento DE INTEGR.
17	33940450	9
18	20578334	9
19	850496	9
20	29445862	9
1	51299472	8
1	201507258	8
2	16715068	8
2	20996616	8
2	36785314	8
2	226916064	8
2	237467876	8
6	49519624	8
6	160955686	8
10	26734304	8
10	79114636	8
12	50505884	8
14	73746866	8
17	43805774	8
17	82718744	8
18	20575274	8
20	5788510	8
21	25629778	8
22	19983370	8
X	47067946	8
1	19916860	7
1	143187676	7
1	148402426	7
1	223224616	7
2	119198990	7
3	45419400	7
4	39364694	7
4	99547294	7
5	38056212	7
6	6715482	7
6	15090744	7
6	89830096	7
6	8988012	7
7	113383212	7

FIG. 2



CROMOSOMA	N.º BASE	RECUENTO DE INTEGR.	CROMOSOMA	N.º BASE	RECUENTO DE INTEGR.	CROMOSOMA	N.º BASE	RECUENTO DE INTEGR.
8	66165666	26	2	218419648	10	8	37700430	7
11	119316888	26	6	136793200	10	8	88289848	7
X	134914730	26	8	25969948	10	9	101660500	7
8	95226212	25	10	46426956	10	9	112857552	7
12	124995298	25	10	47957704	10	10	42097318	7
16	10992070	25	10	52321784	10	10	91591618	7
3	47439806	24	10	52321864	10	10	120431936	7
11	70518294	23	11	102685874	10	11	5225468	7
16	10839466	22	12	54411358	10	12	155086	7
X	66004390	22	12	76936636	10	12	51951306	7
1	9267210	21	15	19782342	10	15	64356576	7
9	91430052	21	16	15220070	10	15	65185434	7
16	30692064	21	16	15220090	10	16	34573028	7
1	155276858	20	16	18131696	10	16	34573058	7
2	173018750	20	16	18131716	10	16	46401528	7
16	88559110	20	18	5447734	10	16	81738062	7
2	149709484	19	19	43779827	10	17	836804	7
8	124924782	19	19	51327816	10	17	75814568	7
13	112237576	19	21	33199996	10	19	17579762	7
20	44683654	19	1	26569750	9	19	38478442	7
1	228965770	18	1	57734650	9	19	40578628	7
7	13991304	18	1	115988958	9	19	41165952	7
7	101110944	18	1	149636460	9	20	58585134	7
11	108224088	18	3	51942092	9	22	36368580	7
12	89874464	18	3	149936952	9	22	36423722	7
1	208159440	17	4	84425848	9	1	2914438	6
5	150115792	17	4	138895946	9	1	91089046	6
12	7096136	16	6	3737388	9	1	115328550	6
15	28018428	16	6	164378486	9	1	125169466	6
19	1279936	16	9	93003938	9	1	125169552	6
20	48636010	16	9	104239026	9	1	155610614	6
1	165379386	15	10	67764450	9	1	155746194	6
3	71103574	15	10	69789200	9	1	201664794	6
5	103961570	15	13	98822320	9	1	131443890	6
7	128077200	15	15	31382604	9	2	4323698	6
8	29495584	15	15	34422898	9	3	58414862	6
16	46399748	15	15	34569124	9	3	66129874	6
16	57534378	15	15	78243618	9	3	87800708	6

FIG. 2 (cont.)

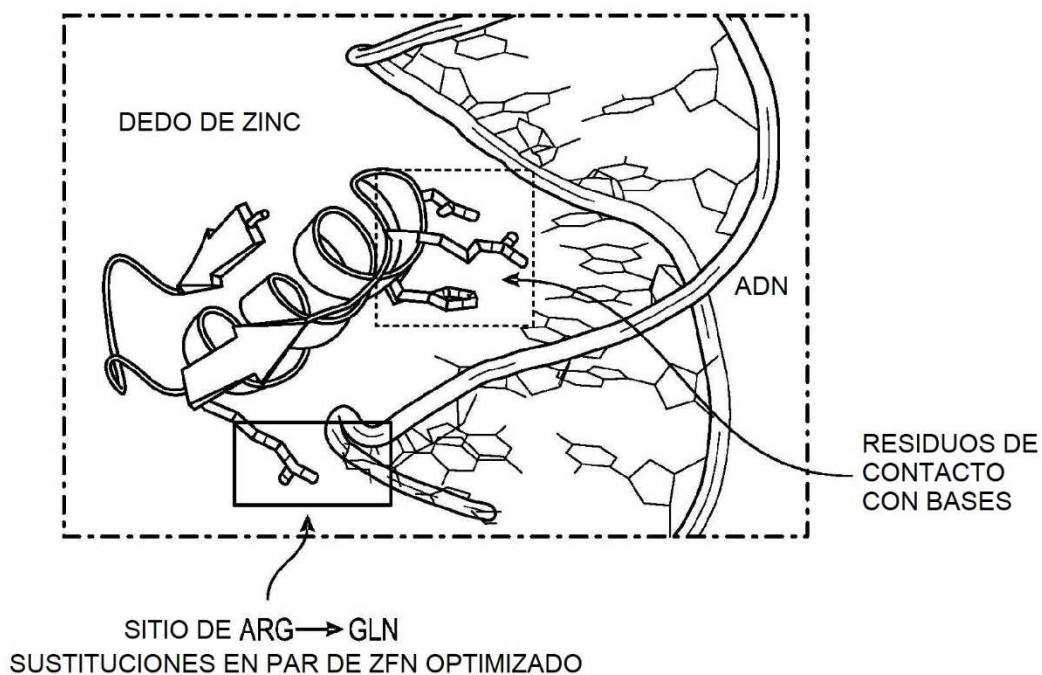
CROMOSOMA	N.º BASE	RECUENTO DE INTEGR.	CROMOSOMA	N.º BASE	RECUENTO DE INTEGR.	CROMOSOMA	N.º BASE	RECUENTO DE INTEGR.
3	14128736	6	13	79775976	5	9	92277906	4
4	151031930	6	15	40114388	5	9	97934932	4
5	178121148	6	15	68307028	5	9	121566378	4
6	28931180	6	15	82539916	5	9	124392210	4
6	108893514	6	15	86572492	5	9	126591182	4
6	136088656	6	16	21645824	5	10	138212048	4
6	163725808	6	16	53042972	5	10	25357460	4
7	162658	6	16	58873960	5	10	41847448	4
7	35016604	6	16	67841642	5	10	41847504	4
8	30408876	6	16	78808108	5	10	42103514	4
8	91876290	6	17	17025784	5	10	64508254	4
8	119945566	6	17	26884866	5	10	76825618	4
9	4090722	6	17	72902448	5	10	109893420	4
9	104968230	6	18	11456592	5	11	35944902	4
10	72293352	6	18	33538826	5	11	72431102	4
10	79866034	6	19	7159860	5	12	6232920	4
11	118997014	6	19	11922082	5	12	7242432	4
11	121069264	6	19	18462114	5	12	49404046	4
12	3073480	6	19	45686048	5	12	75511410	4
12	26186106	6	20	29881056	5	12	88031364	4
13	43878548	6	20	43558622	5	12	95073610	4
15	41284346	6	20	47575112	5	12	113153372	4
15	58555840	6	22	19545504	5	13	18964478	4
15	60478340	6	22	23284226	5	13	34779902	4
16	3106742	6	14	33684	5	13	113600728	4
16	4518046	6	1	6359676	4	14	92616966	4
16	29292606	6	1	18556502	4	15	33229354	4
16	77989246	6	1	24862574	4	15	64142544	4
16	88266652	6	1	25442520	4	15	67498752	4
16	88732066	6	1	30801948	4	15	75356538	4
17	36938710	6	1	51631166	4	15	76355710	4
19	267174	6	1	72774624	4	15	684294	4
22	39690408	6	1	109213944	4	15	14285910	4
X	40259106	6	1	153609522	4	15	85317250	4
X	40835932	6	1	175878760	4	17	9710280	4
1	1724636	5	1	201154009	4	17	21372082	4
1	6161778	5	1	224687886	4	17	21906776	4
1	17419988	5	2	137774808	4	17	26603948	4

FIG. 2 (cont.)

CROMOSOMA	N.º BASE	RECUENTO DE INTEGR.	CROMOSOMA	N.º BASE	RECUENTO DE INTEGR.	CROMOSOMA	N.º BASE	RECUENTO DE INTEGR.
1	36401004	5	2	166910124	4	17	26603992	4
1	143208542	5	2	234668446	4	17	40557500	4
1	150528982	5	3	49802584	4	17	43796376	4
1	154750274	5	3	195856500	4	17	49883740	4
2	86563402	5	3	198148214	4	18	13272880	4
2	105038552	5	4	1686620	4	18	30136646	4
2	117954094	5	4	3480800	4	18	30845084	4
2	130230372	5	4	99869180	4	18	74019902	4
2	176188548	5	4	138090616	4	19	40266218	4
3	2041446	5	4	158876704	4	19	42741976	4
3	16078092	5	5	10727038	4	19	42767202	4
3	75669968	5	5	36318878	4	19	42881190	4
3	178998880	5	5	36322678	4	19	49683046	4
3	184689238	5	5	65562414	4	20	4673782	4
4	49119414	5	5	100709408	4	20	24683868	4
4	73573948	5	5	142193948	4	20	44911634	4
4	78700116	5	5	149564992	4	20	48621130	4
5	7057496	5	5	164896712	4	20	49279476	4
5	11767140	5	6	176740968	4	20	49699412	4
5	56314802	5	6	2643624	4	20	56914278	4
5	132314286	5	6	27782516	4	20	62819166	4
6	58116958	5	6	45156450	4	21	10442640	4
6	60870558	5	6	146750040	4	22	22533154	4
7	128792974	5	6	169090710	4	22	22648222	4
7	133081754	5	7	1214740	4	22	22956152	4
9	92835866	5	7	37357350	4	22	23111094	4
9	131373156	5	7	37357420	4	22	23162790	4
10	50701042	5	7	62355868	4	22	24629320	4
10	87382350	5	7	71679460	4	22	32136074	4
11	49652236	5	7	106778616	4	22	45216526	4
11	61612468	5	7	133077880	4	X	6134666	4
11	62795536	5	8	5753270	4	X	39688532	4
11	67277754	5	8	5753290	4	X	52934730	4
11	82032970	5	8	8244554	4	X	71283192	4
11	88966978	5	8	8371950	4	X	90323678	4
12	6326346	5	8	101955394	4	X	106863278	4
12	104138422	5	8	119000446	4	Y	3900438	4
12	121500124	5	9	38068006	4			

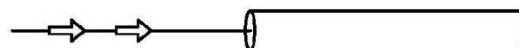
FIG. 2 (cont.)



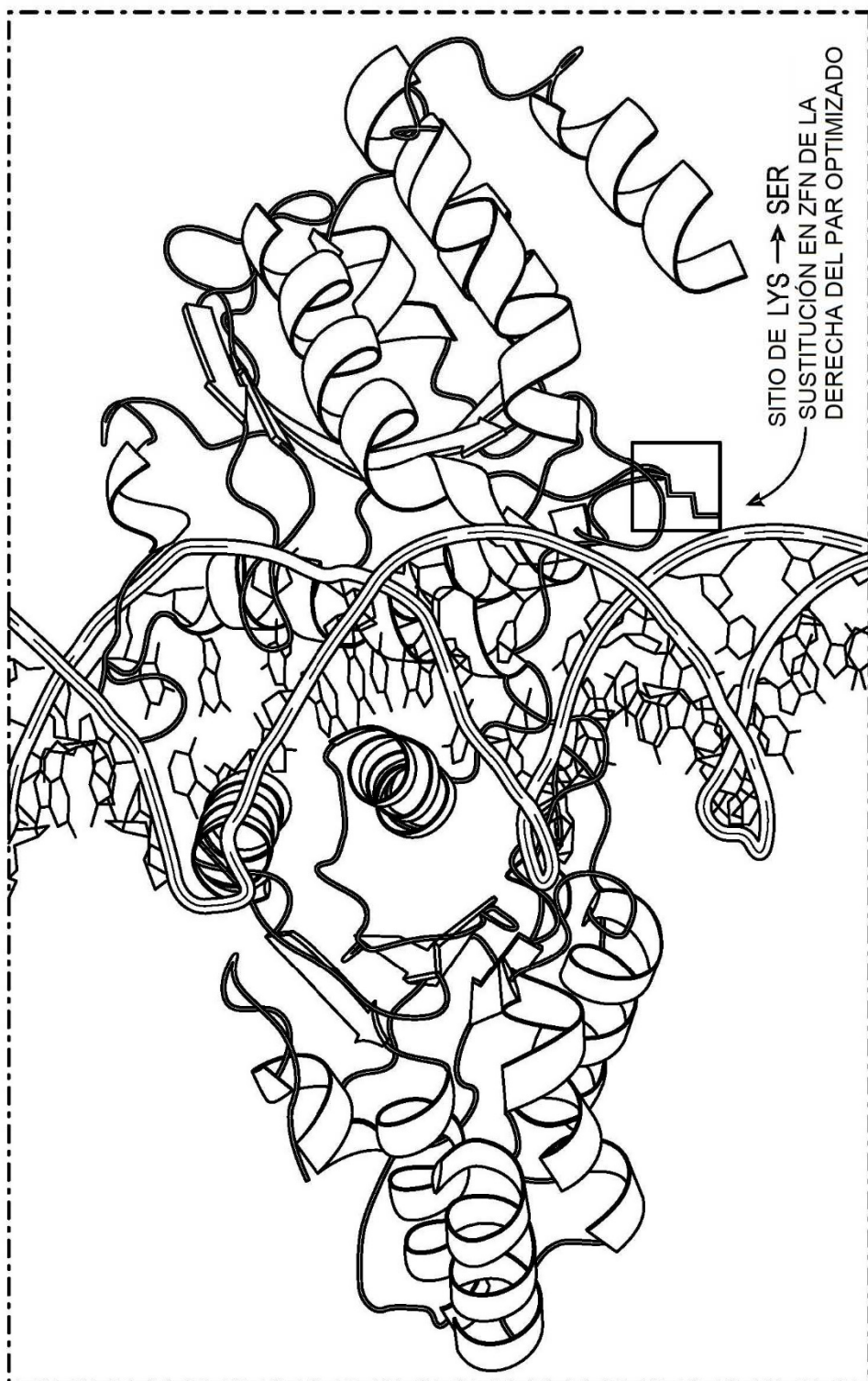


**FIG. 3A**

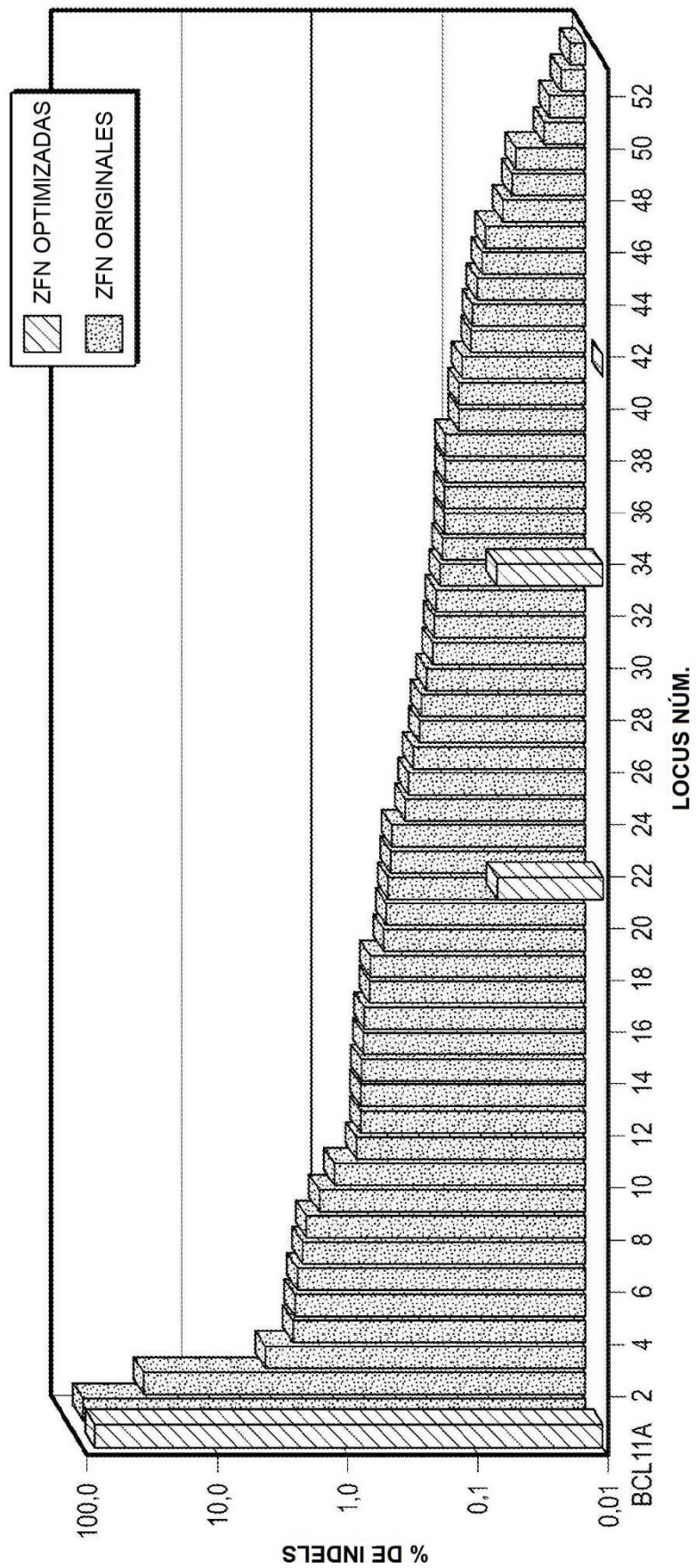
		SITIO DE ARG → GLN		RESIDUOS DE CONTACTO CON BASES	
		SUSTITUCIONES EN PAR DE ZFN OPTIMIZADO			
ZFP IZQ.	F1	VPAAMAERP	FQCRICMNF	S	DQSNLRHIRTH
	F2	TGEKP	FACDICGRKF	A	RNFSLTMTHTKIH
	F3	TGSQKP	FQCRICMNF	S	STGNLTNHIRTH
	F4	TGEKP	FACDICGRKF	A	TSGSLTRHTKIH
	F5	THPRAPIPKP	FQCRICMNF	S	DQSNLRHIRTH
	F6	TGEKP	FACDICGRKF	A	AQCCLFHTHTKIH
ZFP DER.	F1	VPAAMAERP	FQCRICMNF	A	RNDHRTHTHTKIH
	F2	TGEKP	FQCRICMNF	S	QKAHLIRHIRTH
	F3	TGEKP	FACDICGRKF	A	QKGTLGEHTKIH
	F4	TGSQKP	FQCRICMNF	S	RGRDLSRHIRTH
	F5	TGEKP	FACDICGRKF	A	RRDNLHSHHTKIH
				LRGS	



**FIG. 3B**



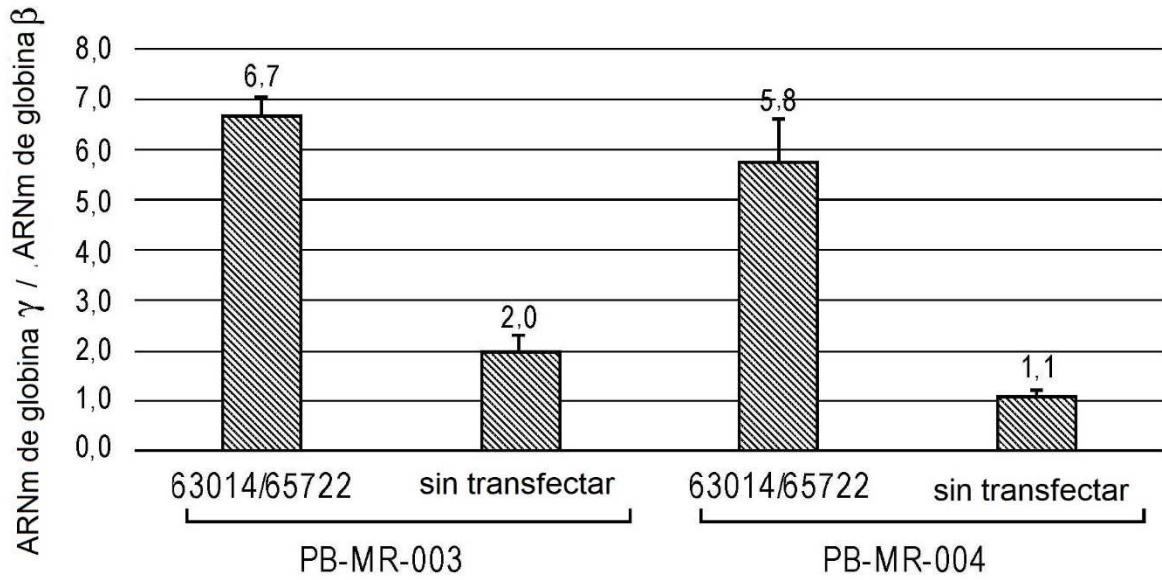
**FIG. 3C**



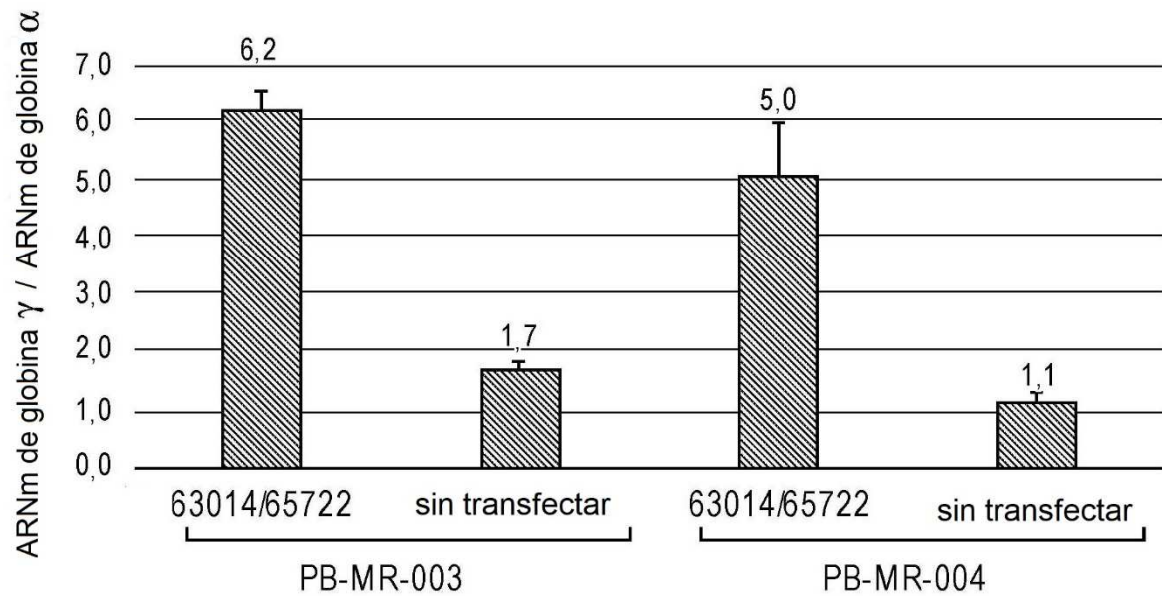
ZFN ORIGINALES: 51857/51949, ZFN OPTIMIZADAS: 63014/65722

**FIG. 4**

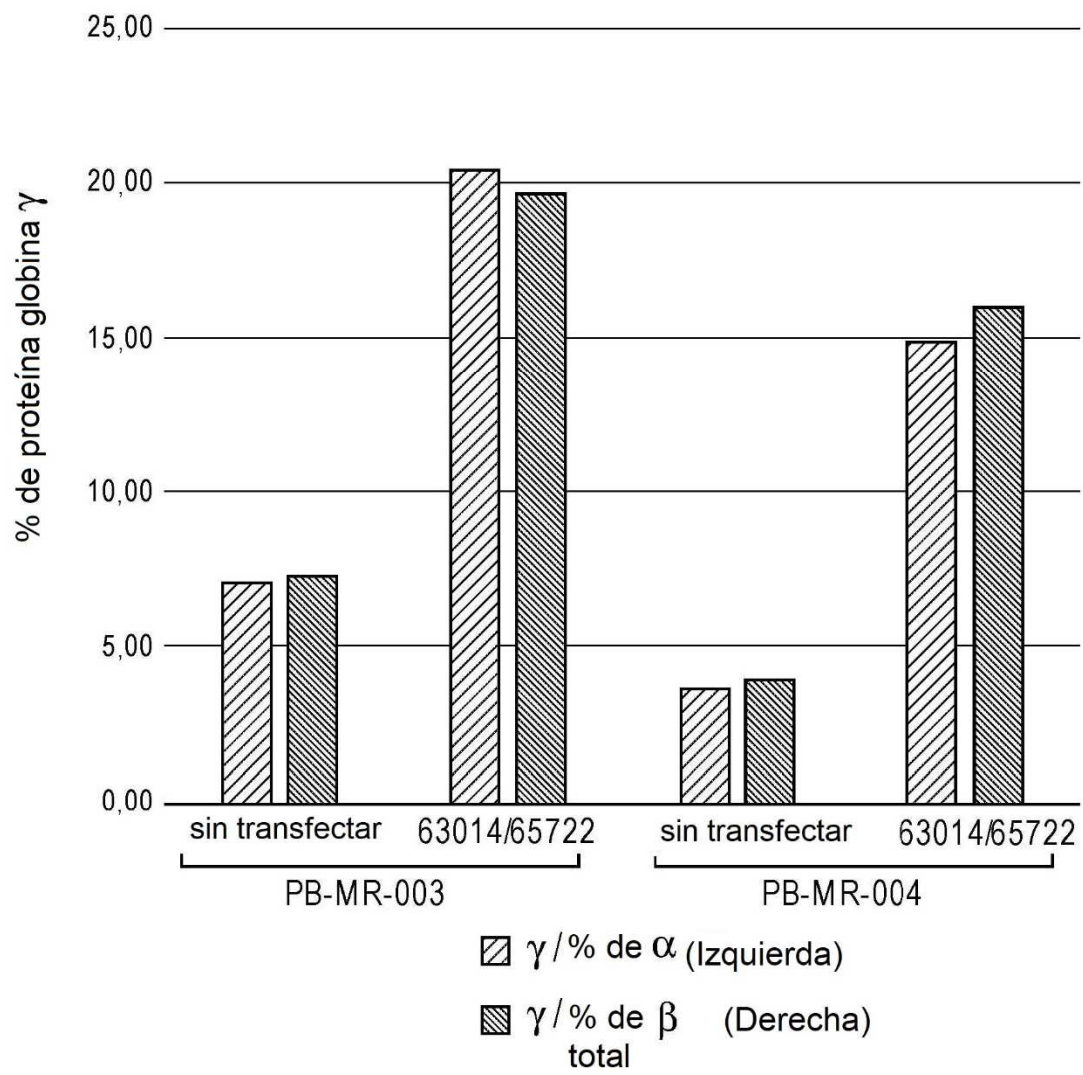




**FIG. 5A**

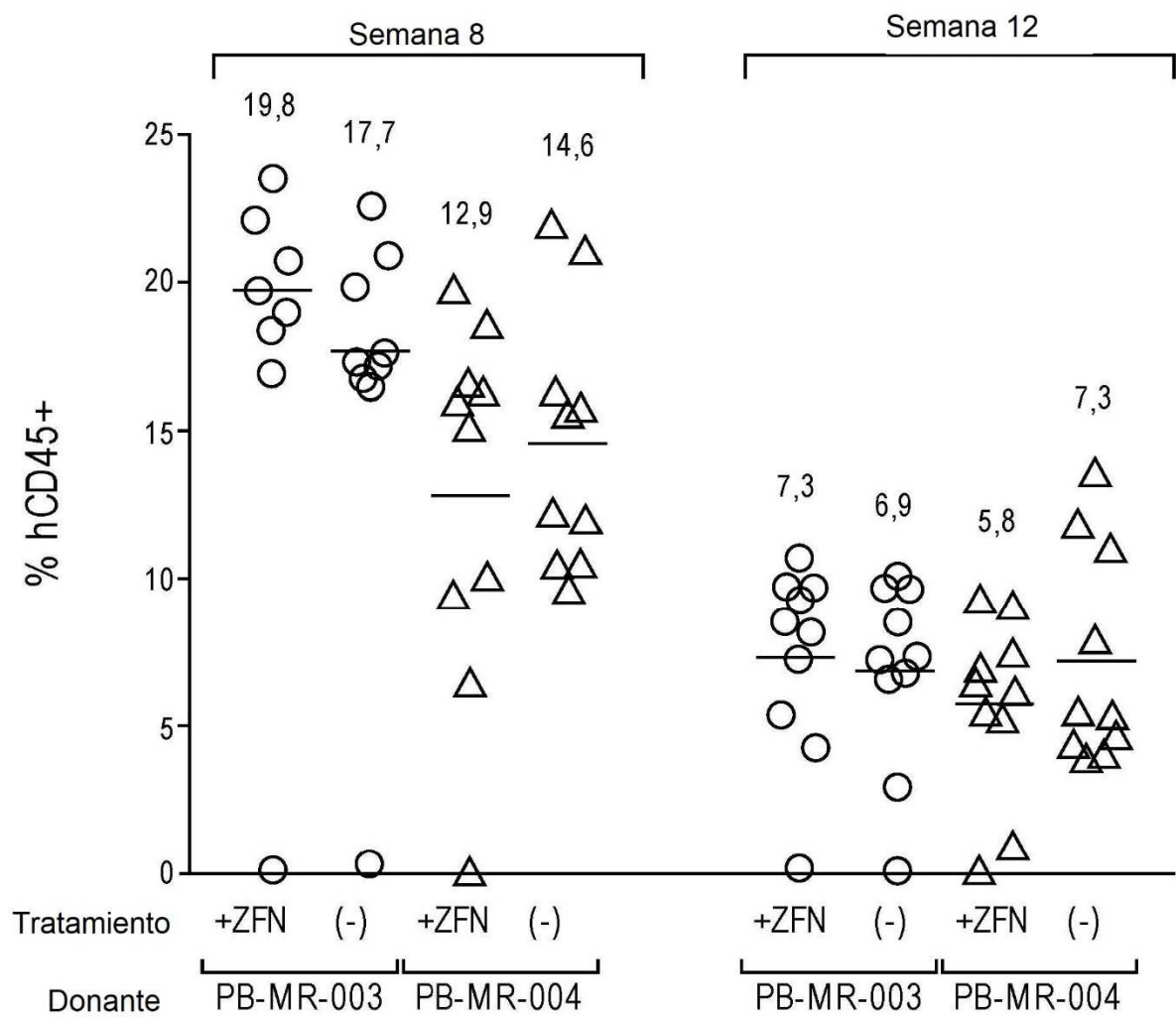


**FIG. 5B**

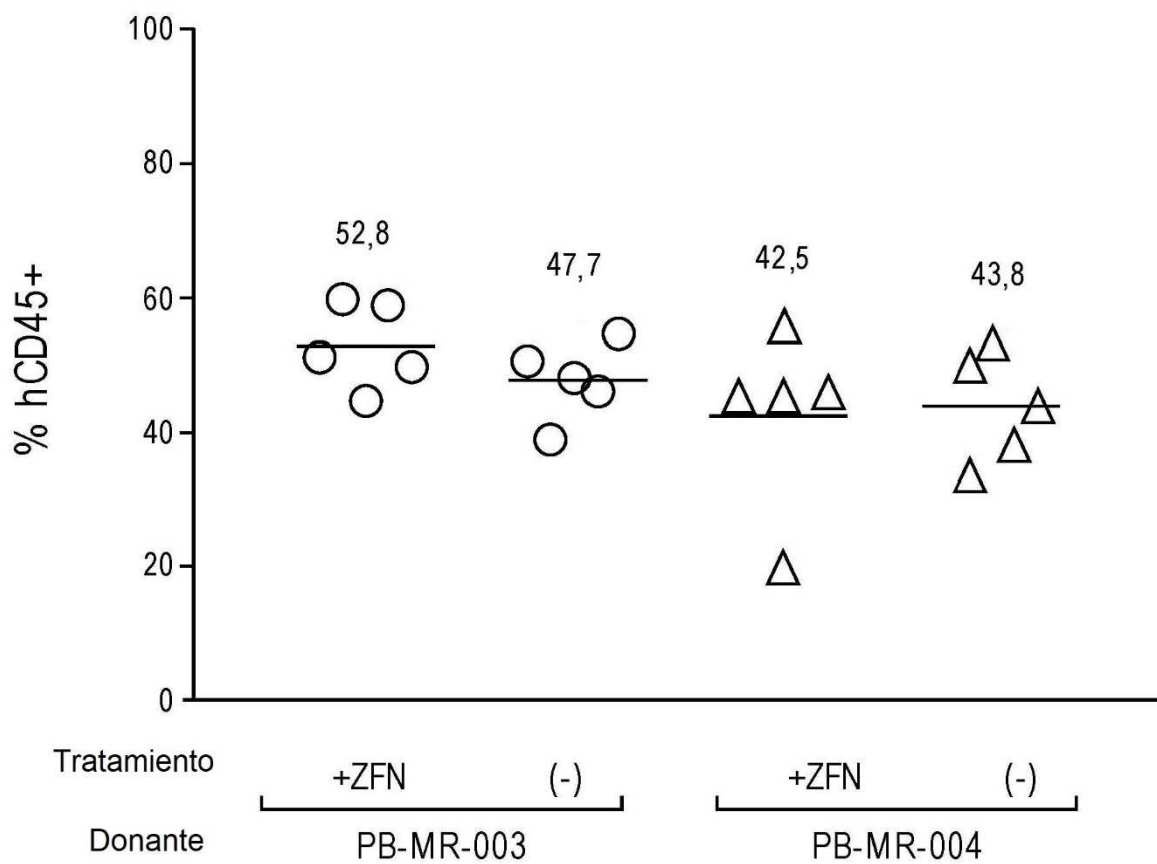


**FIG. 6**



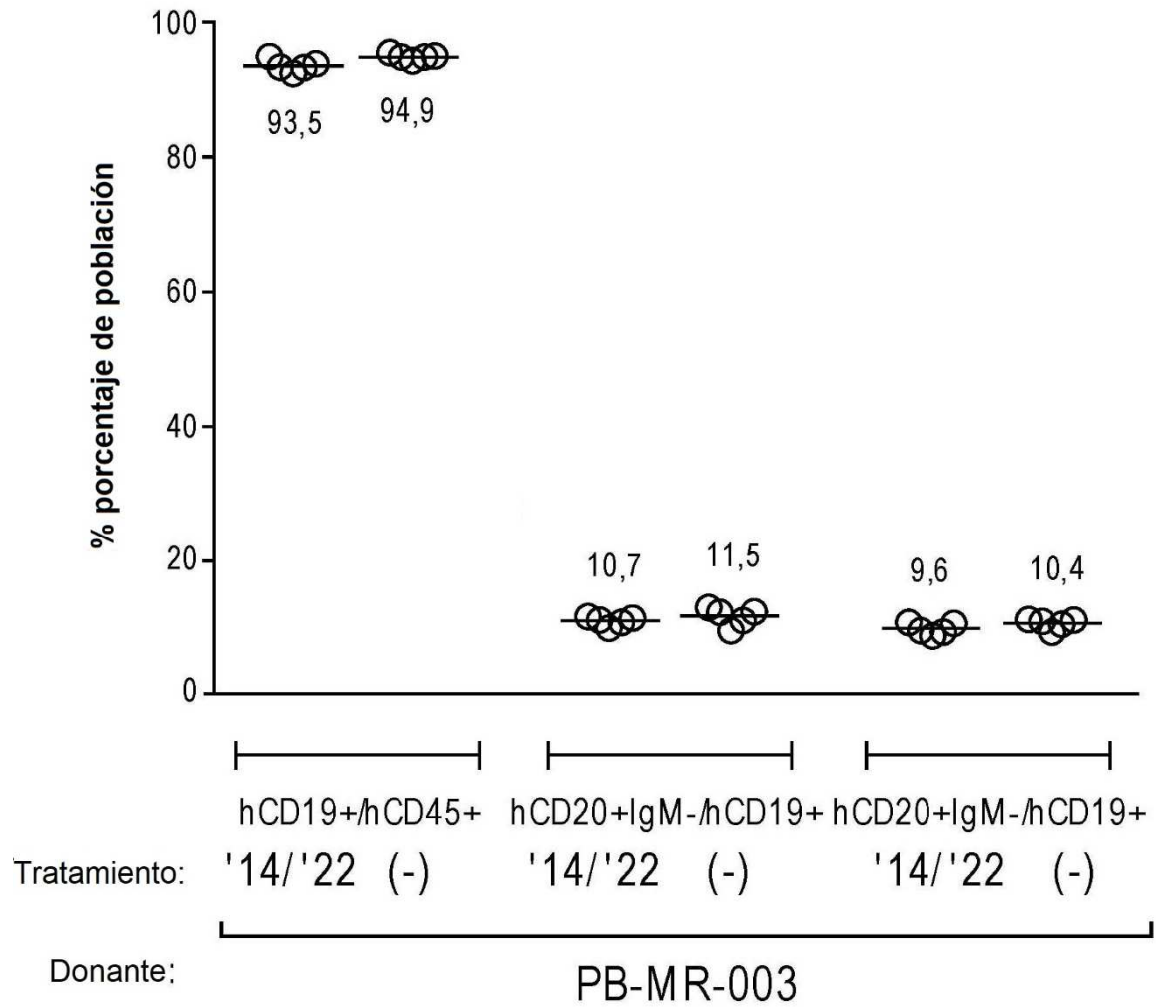


**FIG. 7**

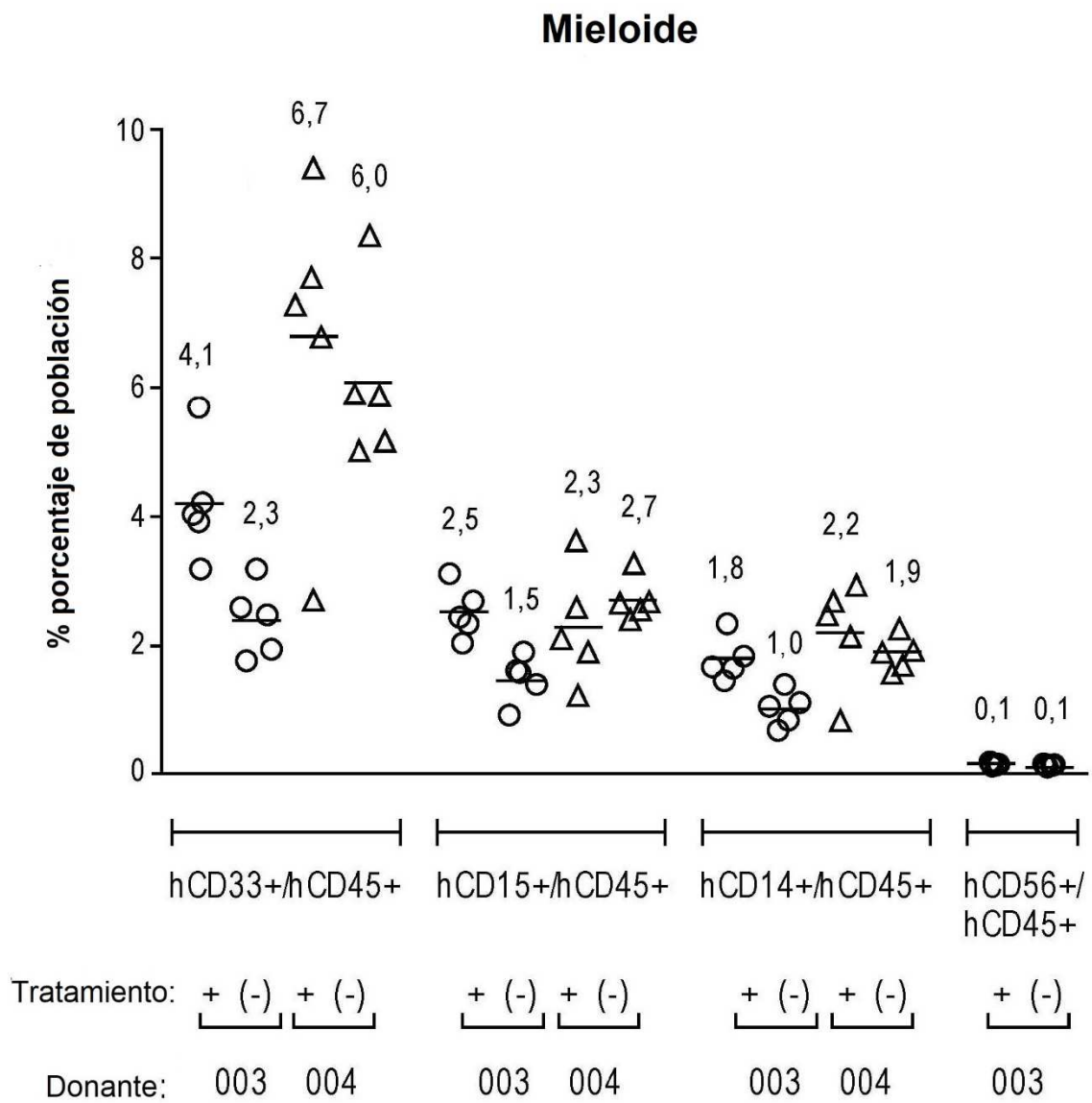


**FIG. 8**

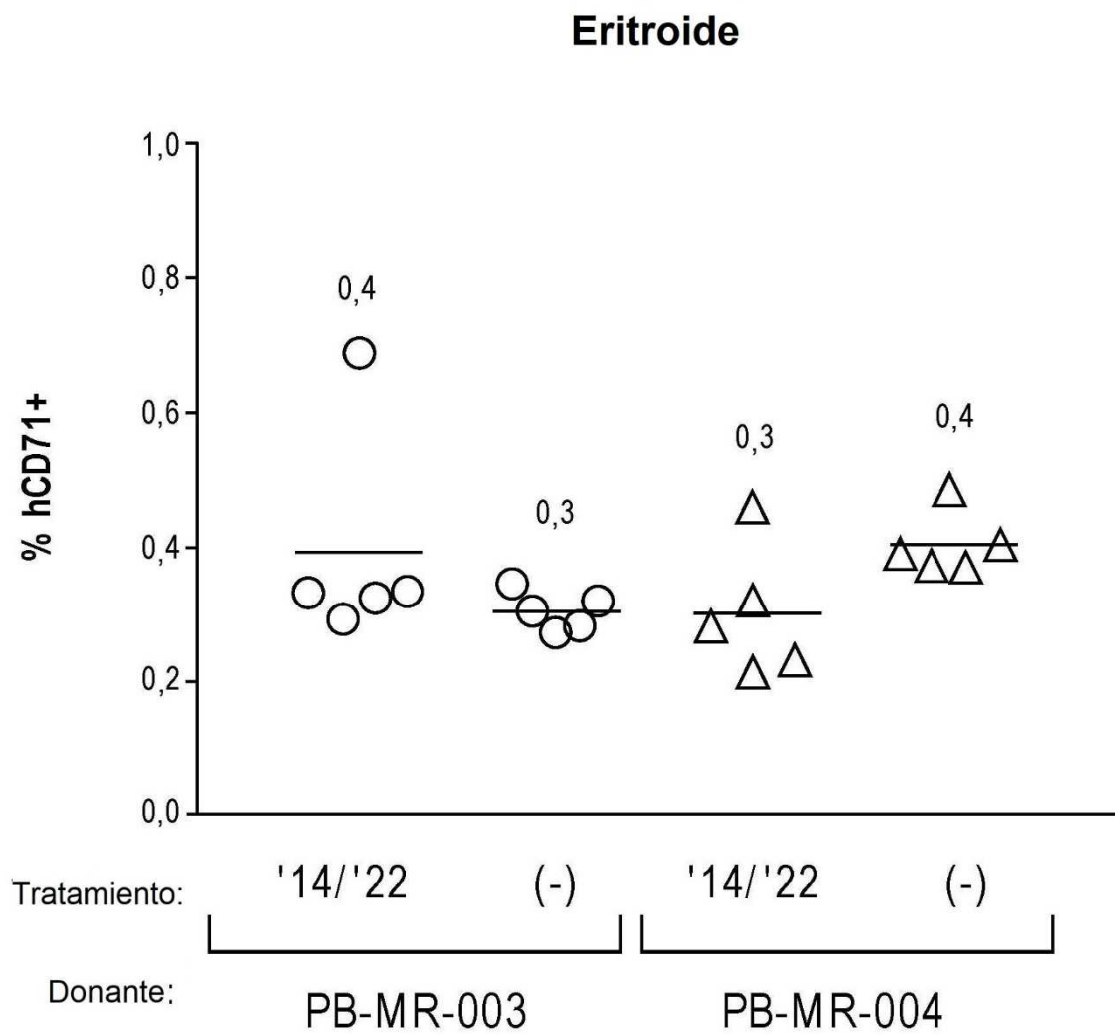
## Linfoide (linfocitos B)



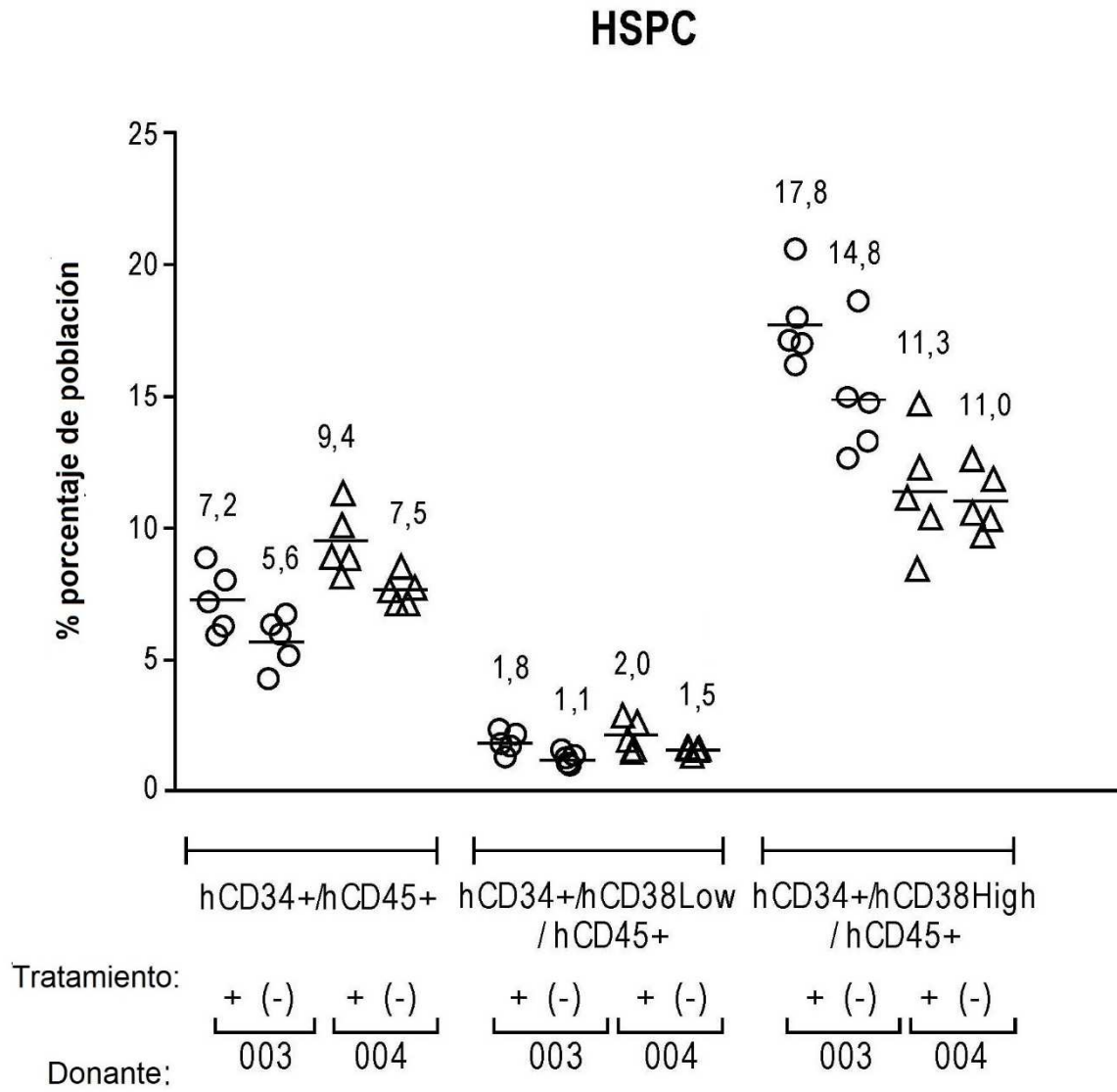
**FIG. 9A**



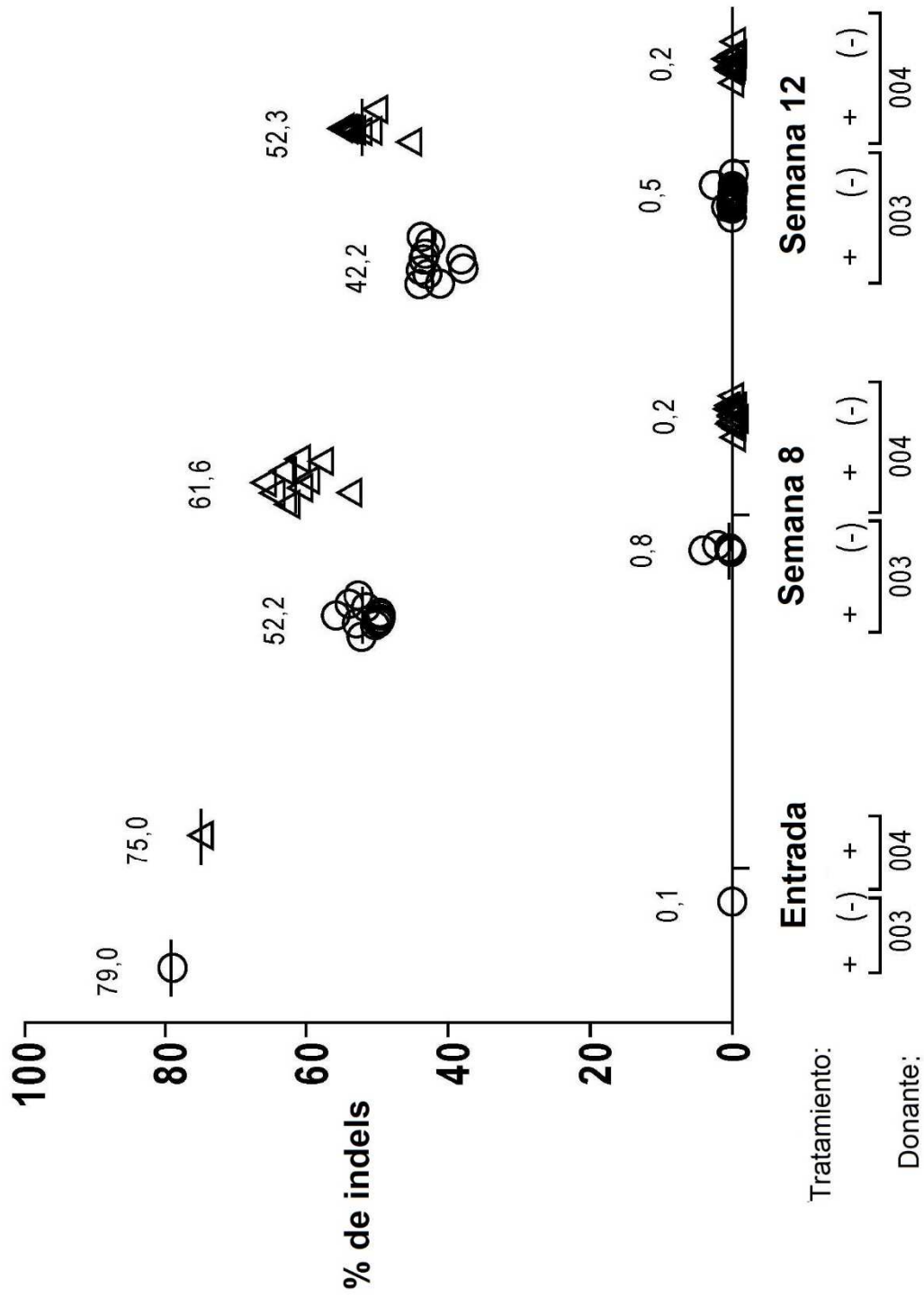
**FIG. 9B**



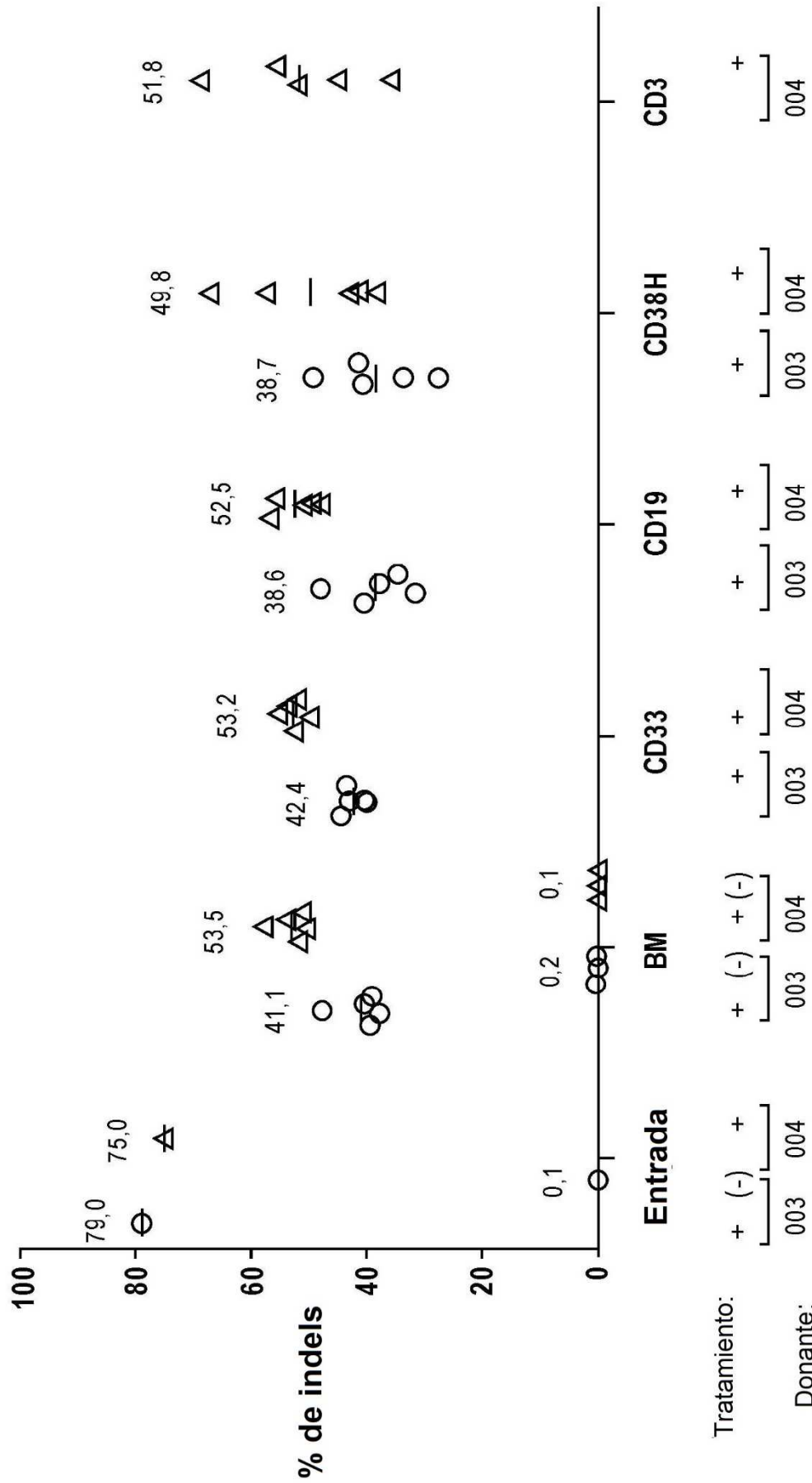
**FIG. 9C**



**FIG. 9D**

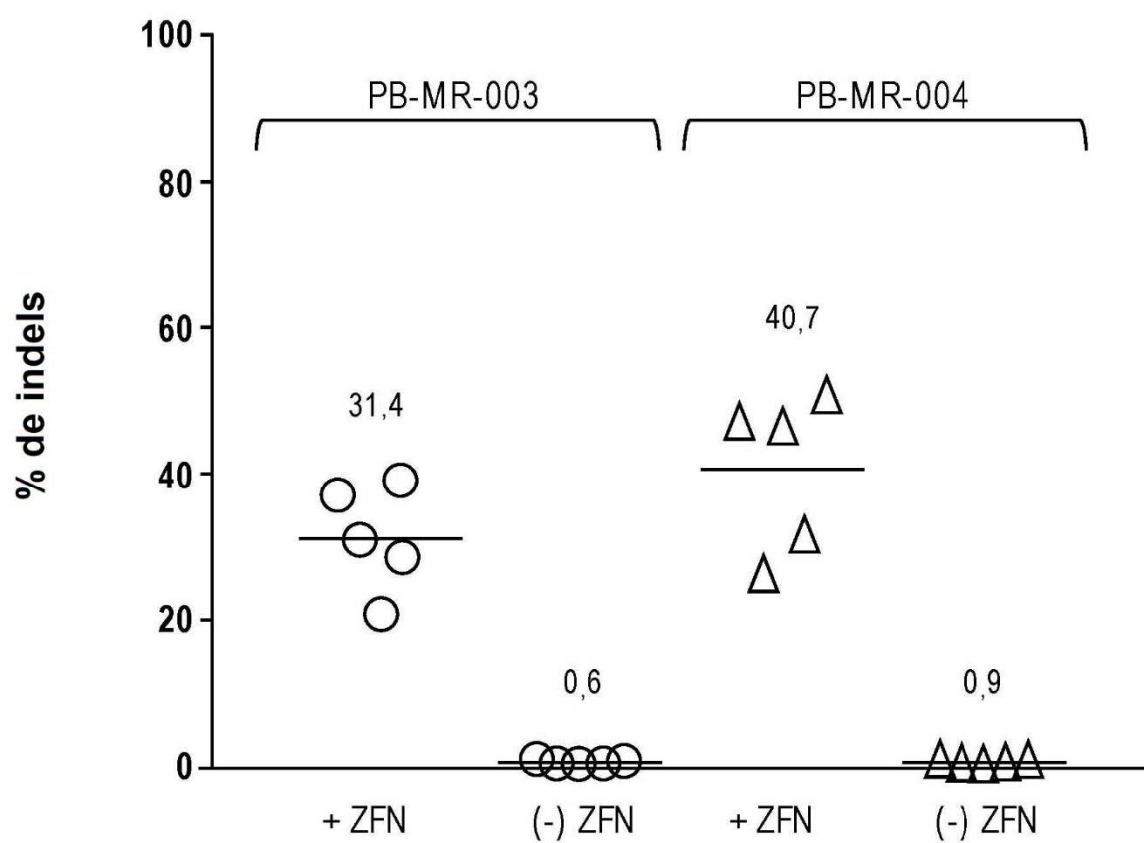


**FIG. 10**

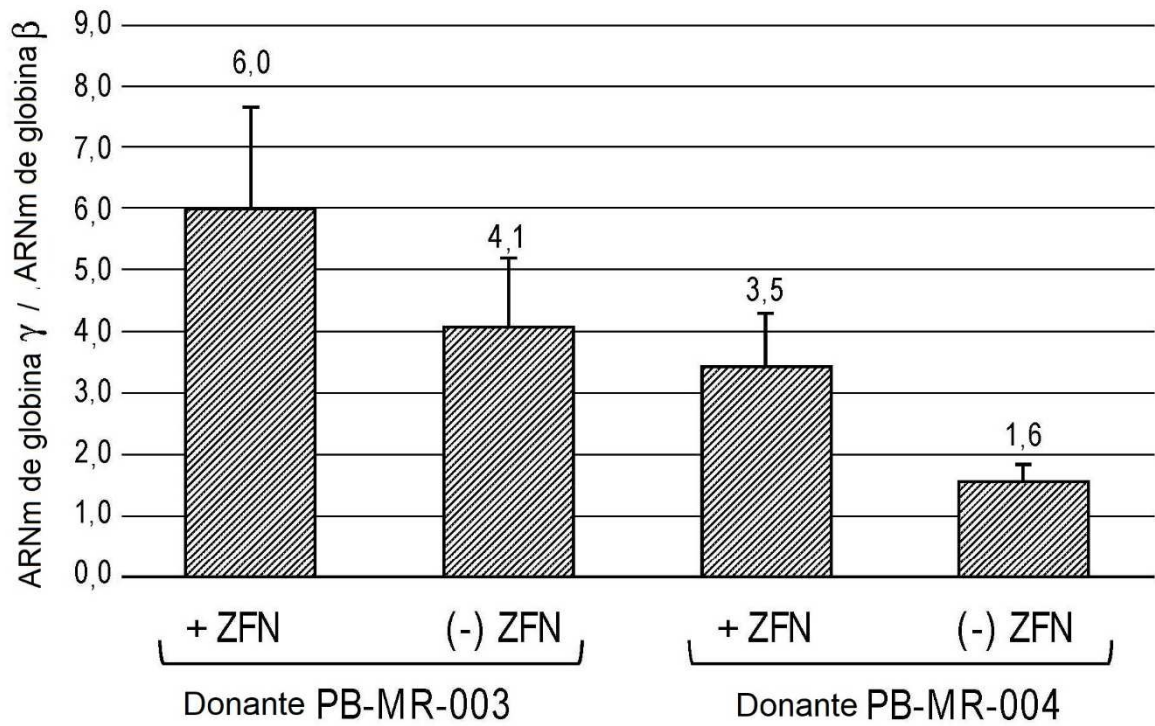


**FIG. 11**

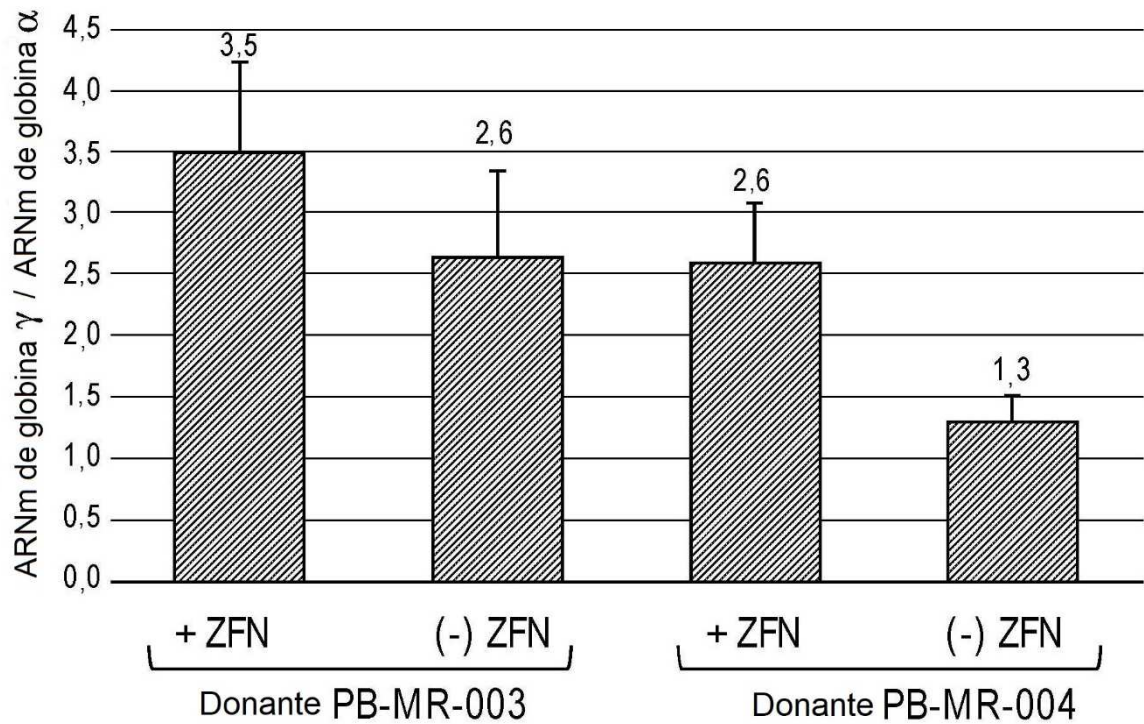




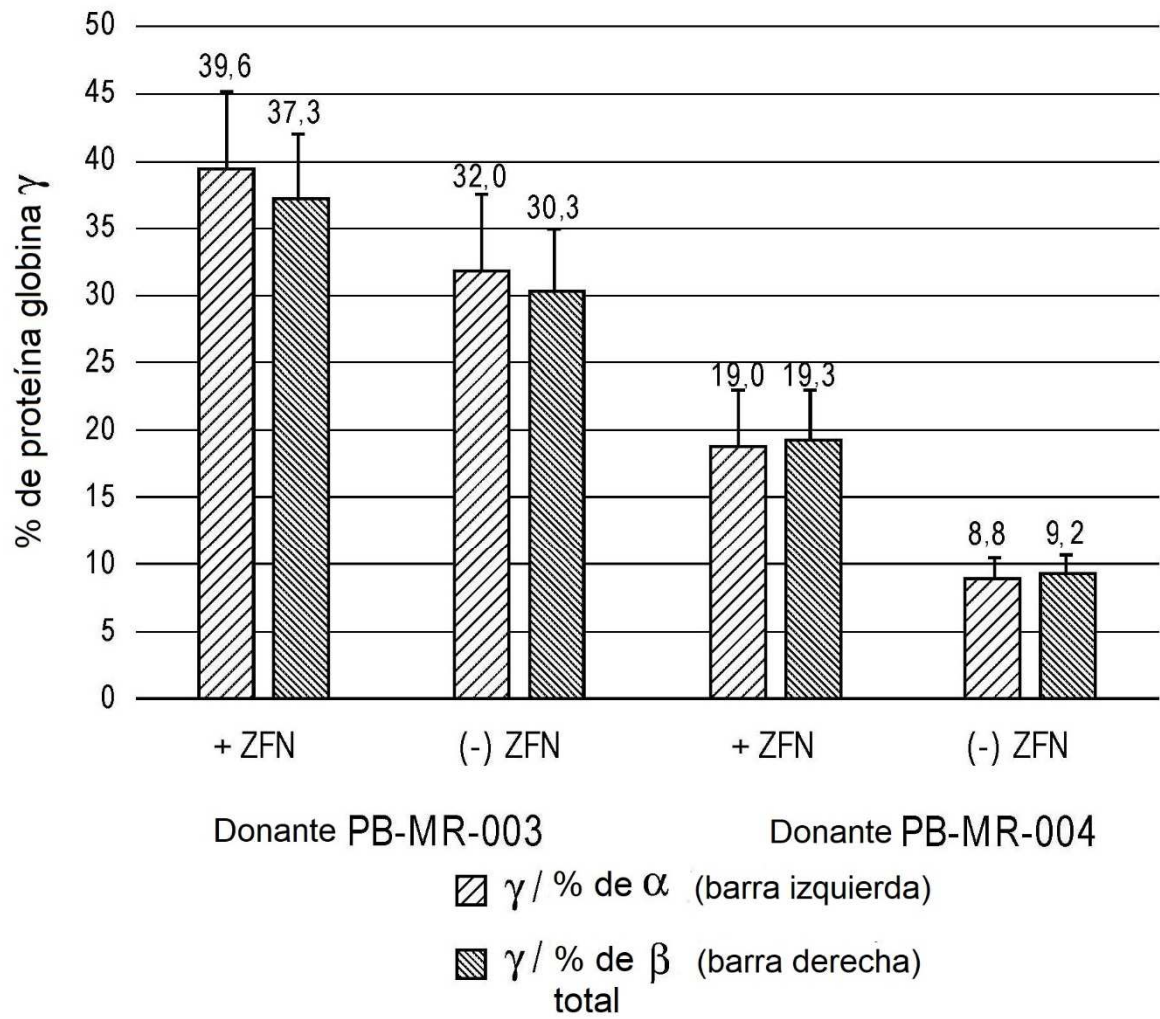
**FIG. 12**



**FIG. 13A**



**FIG. 13B**



**FIG. 14**