

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 812 599**

51 Int. Cl.:

C12N 5/071 (2010.01)

C07K 14/805 (2006.01)

C12N 9/22 (2006.01)

C12N 15/85 (2006.01)

C07K 14/47 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **29.08.2013 PCT/US2013/057214**

87 Fecha y número de publicación internacional: **06.03.2014 WO14036219**

96 Fecha de presentación y número de la solicitud europea: **29.08.2013 E 13833658 (1)**

97 Fecha y número de publicación de la concesión europea: **24.06.2020 EP 2890780**

54 Título: **Procedimientos y composiciones para el tratamiento de una afección genética**

30 Prioridad:

29.08.2012 US 201261694693 P

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

17.03.2021

73 Titular/es:

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

72 Inventor/es:

**COST, GREGORY J.;
GREGORY, PHILIP D.;
GUSCHIN, DMITRY;
HOLMES, MICHAEL C.;
MILLER, JEFFREY C.;
PASCHON, DAVID;
REBAR, EDWARD J.;
REIK, ANDREAS;
URNOV, FYODOR y
ZHANG, LEI**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 812 599 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

Procedimientos y composiciones para el tratamiento de una afección genética

5 Campo técnico

La presente descripción se encuentra en el campo de la ingeniería genómica de las células madre hematopoyéticas, especialmente para el tratamiento de una hemoglobinopatía.

10 Antecedentes

La genoterapia encierra un potencial enorme para una nueva era en la medicina humana. Estas metodologías permitirán el tratamiento de afecciones que hasta ahora no se han podido abordar desde la práctica médica estándar. Un área que es especialmente prometedora es la capacidad de manipular genéticamente una célula para hacer que esa célula exprese un producto que no se producía previamente en esa célula. Los ejemplos de usos de esta tecnología incluyen la inserción de un gen que codifica una proteína terapéutica nueva, la inserción de una secuencia codificadora que codifica una proteína que falta en la célula o en el individuo, la inserción de un gen de tipo silvestre en una célula que contiene una secuencia génica mutada y la inserción de una secuencia que codifica un ácido nucleico estructural tal como un microARN o ARNip.

Los transgenes pueden administrarse a una célula mediante diversos medios, de modo que el transgén se integre en el propio genoma de la célula y se mantenga en el mismo. En los últimos años, se ha desarrollado una estrategia para la integración de transgenes que utiliza la escisión con nucleasas específicas del sitio para la inserción dirigida en un locus genómico elegido (véase, por ejemplo, la patente de Estados Unidos en copropiedad N° 7.888.121). Pueden utilizarse nucleasas específicas para genes diana de modo que el constructo del transgén se inserte por reparación dirigida por homología (HDR) o por captura final durante procesos conducidos por unión de extremos no homóloga (NHEJ). Los loci diana incluyen loci de "puerto seguro", por ejemplo, un gen CCR5, un gen CXCR4, un gen PPP1R12C (también conocido como AAVS1), un gen de albúmina o un gen *Rosa*. Véanse, por ejemplo, las publicaciones de patente de Estados Unidos N° 20080299580, 20080159996, 201000218264, 20110301073, 20130177983 y 20130177960 y la solicitud de patente provisional N° 61/823.689. La integración mediada por nucleasas ofrece la posibilidad de una mejor expresión del transgén, mayor seguridad y durabilidad de la expresión, en comparación con los enfoques de integración clásicos que dependen de la integración aleatoria del transgén, ya que permite el posicionamiento exacto del transgén para un riesgo mínimo de silenciamiento génico o activación de oncogenes cercanos.

Los glóbulos rojos (hematíes), o eritrocitos, son el principal componente celular de la sangre. De hecho, los glóbulos rojos representan una cuarta parte de las células de un ser humano. Los glóbulos rojos maduros carecen de núcleo y muchos otros orgánulos en los seres humanos, y están llenos de hemoglobina, una metaloproteína que se encuentra en los glóbulos rojos que funciona transportando oxígeno a los tejidos, así como recogiendo dióxido de carbono de los tejidos y dirigiéndose de vuelta a los pulmones para la extracción del mismo. La proteína constituye aproximadamente el 97% del peso seco de los glóbulos rojos y aumenta la capacidad de transporte de oxígeno de la sangre en aproximadamente setenta veces. La hemoglobina es un heterotetrámero que comprende dos cadenas de globina de tipo α y dos cadenas de globina de tipo β y 4 grupos hemo. En adultos, el tetrámero $\alpha_2\beta_2$ se denomina hemoglobina A (HbA) o hemoglobina adulta. Típicamente, las cadenas de alfa- y beta-globina se sintetizan en una relación aproximada de 1:1 y esta relación parece ser crítica con respecto a la hemoglobina y la estabilización de glóbulos rojos. De hecho, en algunos casos en los que un tipo de gen de globina se expresa de forma inadecuada (véase más adelante), reduciendo la expresión (por ejemplo, utilizando un ARNip específico) del otro tipo de globina, mediante la restauración de esta relación 1:1 se alivian algunos aspectos del fenotipo celular mutante (véase Voon et al. (2008) *Haematologica* 93 (8):1288). En un feto en desarrollo, se produce una forma diferente de hemoglobina, la hemoglobina fetal (HbF), que tiene una mayor afinidad de unión por el oxígeno que la hemoglobina A, de modo que el oxígeno puede ser suministrado al sistema del bebé a través del torrente sanguíneo de la madre. La hemoglobina fetal también contiene dos cadenas de α -globina, pero en lugar de las cadenas de β -globina de adultos, tiene dos cadenas de γ -globina fetal (es decir, la hemoglobina fetal es $\alpha_2\gamma_2$). Aproximadamente a las 30 semanas de gestación, la síntesis de γ -globina en el feto comienza a disminuir mientras aumenta la producción de β -globina. Aproximadamente a los 10 meses de edad, la hemoglobina del recién nacido es casi toda $\alpha_2\beta_2$, aunque parte de la HbF persiste hasta la edad adulta (aproximadamente el 1-3% de la hemoglobina total). La regulación del cambio de la producción de γ a β es bastante compleja, y principalmente implica una regulación a la baja de la expresión de γ -globina con una regulación al alza simultánea de la expresión de β -globina.

Los defectos genéticos en las secuencias que codifican las cadenas de hemoglobina pueden ser responsables de una serie de enfermedades conocidas como hemoglobinopatías, incluidas la anemia de células falciformes y las talasemias. En la mayor parte de los pacientes con hemoglobinopatías, los genes que codifican la γ -globina

permanecen presentes, pero la expresión es relativamente baja debido a la represión génica normal que tiene lugar alrededor del parto tal como se ha descrito anteriormente.

Se estima que 1 de cada 5000 personas en Estados Unidos padece la enfermedad de células falciformes (SCD), principalmente personas con ascendencia del África subsahariana. Parece existir un beneficio en la heterocigosidad de células falciformes para la protección contra la malaria, por lo que este rasgo puede haber sido seleccionado a lo largo del tiempo, de modo que se estima que en el África subsahariana un tercio de la población tiene el rasgo de células falciformes. La enfermedad de células falciformes está provocada por una mutación en el gen de la β -globina en la que la valina se sustituye por ácido glutámico en el aminoácido N° 6 (un GAG a GTG a nivel de ADN), denominándose la hemoglobina resultante "hemoglobina S" o "HbS". En condiciones de oxígeno más bajas, un cambio conformacional en la forma desoxi de HbS expone un parche hidrofóbico 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, haciendo que las moléculas de HbS se agreguen y formen precipitados fibrosos. Estos agregados a su vez causan la anomalía en los glóbulos rojos o formación de glóbulos rojos falciformes, lo que da como resultado una pérdida de flexibilidad de las células. Los glóbulos rojos falciformes ya no son capaces de comprimirse en los lechos capilares y pueden provocar una crisis vasooclusiva en pacientes con células falciformes. Además, los eritrocitos falciformes son más frágiles que los eritrocitos normales y tienden a hemólisis, lo que eventualmente conduce a anemia en el paciente.

El tratamiento y la gestión de los pacientes con células falciformes es una propuesta de por vida que involucra el tratamiento con antibióticos, la gestión del dolor y transfusiones durante episodios agudos. Un enfoque es el uso de hidroxiurea, que ejerce sus efectos en parte al aumentar la producción de γ -globina. Sin embargo, aún se desconocen los efectos secundarios a largo plazo del tratamiento crónico con hidroxiurea, y el tratamiento produce efectos secundarios no deseados y puede tener una eficacia variable de un paciente a otro. A pesar de un aumento en la eficacia de los tratamientos de células falciformes, la esperanza de vida de los pacientes aún es de mediados a finales del decenio de los 50 años de vida y las morbilidades asociadas de la enfermedad tienen un profundo impacto en la calidad de vida del paciente.

Las talasemias también son enfermedades relacionadas con la hemoglobina y típicamente implican una expresión reducida de las cadenas de globina. Esto puede suceder a través de mutaciones en las regiones reguladoras de los genes o a partir de una mutación en una secuencia codificante de globina que da como resultado una expresión reducida. Las alfa-talasemias están asociadas con personas con ascendencia del África occidental y del sur de Asia, y pueden conferir resistencia a la malaria. La beta-talasemia está asociada con personas de ascendencia mediterránea, típicamente de Grecia y las zonas costeras de Turquía e Italia. El tratamiento de las talasemias generalmente implica transfusiones de sangre y tratamiento de quelación del hierro. Los trasplantes de médula ósea también se están utilizando para el tratamiento de personas con talasemias graves si se puede identificar un donante apropiado, pero este procedimiento puede tener riesgos significativos.

Un enfoque para el tratamiento de SCD y beta-talasemias que se ha propuesto es aumentar la expresión de γ -globina con el objetivo de que la HbF reemplace funcionalmente a la hemoglobina adulta aberrante. Tal como se ha mencionado anteriormente, se cree que el tratamiento de pacientes con SCD con hidroxiurea es exitoso en parte debido a su efecto en el aumento de la expresión de γ -globina. El primer grupo de compuestos que se descubrió que afectaba a la actividad de reactivación de HbF fueron los fármacos citotóxicos. La capacidad para provocar la síntesis *de novo* de gamma-globina por manipulación farmacológica se demostró en primer lugar utilizando 5-azacitidina en animales experimentales (DeSimone (1982) *Proc Natl Acad Sci USA* 79 (14): 4428-31). Estudios posteriores confirmaron la capacidad de la 5-azacitidina para aumentar la HbF en pacientes con β -talasemia y enfermedad de células falciformes (Ley, et al., (1982) *N. Engl. J. Medicine*, 307: 1469-1475 y Ley, et al., (1983) *Blood* 62: 370-380). Además, se ha demostrado que ácidos grasos de cadena corta (por ejemplo, butirato y derivados) en sistemas experimentales aumentan la HbF (Constantoulakis et al., (1988) *Blood* 72 (6): 1961-1967). Además, existe un segmento de la población humana con una afección conocida como "persistencia hereditaria de la hemoglobina fetal" (HPFH) en la que persisten cantidades elevadas de HbF en la edad adulta (10-40% en heterocigotos HPFH (véase Thein et al. (2009) *Hum. Mol. Genet* 18 (R2): R216-R223). Esta es una afección rara, pero en ausencia de anomalías asociadas a la beta-globina, no está asociada con ninguna manifestación clínica significativa, incluso aunque el 100% de la hemoglobina del individuo sea HbF. Cuando las personas que tienen una beta-talasemia también tienen HPFH co-incidente, la expresión de HbF puede disminuir la gravedad de la enfermedad. Además, la gravedad del transcurso natural de la enfermedad de células falciformes puede variar significativamente de un paciente a otro, y esta variabilidad, en parte, puede atribuirse al hecho de que algunas personas con enfermedad más leve expresan niveles más altos de HbF.

Un enfoque para aumentar la expresión de HbF implica la identificación de genes cuyos productos desempeñan un papel en la regulación de la expresión de γ -globina. Uno de esos genes es BCL11A, identificado por primera vez por su papel en el desarrollo de linfocitos. BCL11A codifica una proteína de dedo de zinc que se cree que está involucrada

en la regulación específica de la etapa de la expresión de γ -globina. BCL11A se expresa en células precursoras eritroides adultas y la regulación a la baja de su expresión conduce a un aumento en la expresión de γ -globina. Además, parece que el corte y empalme del ARNm de BCL11A está regulado por el desarrollo. En las células embrionarias, parece que se expresan principalmente las variantes de ARNm de BCL11A más cortas, conocidas como BCL11A-S y BCL11A-XS, mientras que en las células adultas se expresan predominantemente las variantes de ARNm de BCL11A-L y BCL11A-XL más largas. Véase, Sankaran et al. (2008) *Science* 322 p. 1839. La proteína BCL11A parece interactuar con el locus de la β -globina para alterar su conformación y, por lo tanto, su expresión en diferentes etapas de desarrollo. Además, otra proteína reguladora KLF1 parece estar involucrada en la regulación de la expresión de la γ -globina. Se ha encontrado que los niveles de KLF1 son directamente proporcionales a los niveles de BCL11A, y ambos son inversamente proporcionales a los niveles de γ -globina. Por ejemplo, en una familia maltesa con expresión persistente de HbF, la familia porta una mutación heterocigótica del gen KLF1 (Borg et al. (2010) *Nat Genet*, 42 (9): 801-805). El producto del gen KLF1 parece unirse directamente al gen BCL11A *in vivo*, y por lo tanto puede ser responsable de su regulación al alza (véanse Borg et al., *ibidem*; Bieker (2010) *Nat Genet* 42 (9): 733-734; Zhou et al. (2010) *Nat Genet* 42 (9): 742-744). Por lo tanto, si KLF1 estimula la expresión de BCL11A, la acción de ese BCL11A inducido dará como resultado la supresión de la producción de γ -globina y HbF. Se ha propuesto el uso de un ARN inhibidor dirigido al gen BCL11A (véase, por ejemplo, la publicación de patente de Estados Unidos N° 20110182867) pero esta tecnología tiene varios inconvenientes potenciales, concretamente que no se puede lograr la desactivación completa, la administración de dichos ARN puede ser problemática y los ARN deben estar presentes de forma continua, lo que requiere múltiples tratamientos de por vida.

Las alfa-talasemias son también prevalentes en la población humana, especialmente en Asia, y se cree que algún tipo de aberración de alfa-globina es el trastorno genético más común en seres humanos. En las regiones tropicales y subtropicales del mundo, el trastorno de alfa-globina se encuentra en el 80-90% de la población (véase Hartevelde y Higgs (2010) *Orphanet Journal of Rare Diseases* 5:13).

Los seres humanos portan 2 copias del gen de la alfa-globina en tándem ($\alpha 1$ y $\alpha 2$) en el cromosoma 16, por lo que en una célula diploide normal hay 4 copias en total. El gen $\alpha 2$ normalmente supone 2-3 veces más ARNm de α -globina que el gen $\alpha 1$. La organización en tándem de estos dos genes puede estar asociada con la alta prevalencia de grandes deleciones en los genes de alfa-globina en pacientes con alfa-talasemia, relacionándose generalmente el número de genes de alfa-globina que no son funcionales directamente con la gravedad de cualquier alfa-talasemia (véase Chui et al. (2003) *Blood* 101 (3):791). La deleción de una copia parece ser bastante común (30% de los afroamericanos y 60-80% de las personas que viven en Arabia Saudita, India y Tailandia), y generalmente no es evidente en el individuo a menos que se realicen pruebas genéticas. La deleción de dos copias, ya sea en el mismo cromosoma (*cis*) o una de cada cromosoma (*trans*), puede causar que la persona afectada tenga anemia leve. Cuando se eliminan tres genes de α -globina, de modo que el individuo tenga solo un gen de α -globina en funcionamiento, se encuentra anemia moderada, pero lo más importante, se altera la relación crucial de α -globina con respecto a β -globina. Los tetrámeros $\beta 4$, que comprenden cuatro cadenas de beta-globina a menudo se observan en pacientes con un solo gen de alfa-globina funcional, una afección conocida como HbH. Los tetrámeros $\beta 4$ son capaces de unirse a oxígeno, pero no lo liberan en la periferia, provocando lo que se conoce como enfermedad de HbH. Las personas con enfermedad de HbH tienen glóbulos rojos con semividas más cortas y que experimentan hemólisis fácilmente, lo que produce un aumento de la anemia. La pérdida de los cuatro genes de α -globina suele ser mortal en el útero.

Bauer et al. ((2012) *Blood* 120 (15): 2945-2953) analizan las perspectivas de nuevos tratamientos para los trastornos de beta-globina.

Sebastiano et al. ((2011) *Stem Cells* 29 (11): 1717-1726) se refieren a la corrección genética de la mutación de células falciformes utilizando nucleasas que se dirigen al gen de beta-globina humana.

Por lo tanto, sigue existiendo la necesidad de procedimientos y composiciones adicionales que puedan utilizarse para la edición del genoma, para corregir un gen aberrante o para alterar la expresión de otros, por ejemplo, para tratar hemoglobinopatías tales como la enfermedad de células falciformes y la talasemia.

Sumario

La invención proporciona una célula precursora de glóbulos rojos genomanipulada caracterizada por una modificación genómica dentro del exón 2 o el exón 4 de BCL11A o dentro de BCL11A-XL realizada tras una escisión dentro del gen BCL11A por una nucleasa de dedo de zinc (ZFN), una nucleasa TALE (TALEN) o un sistema CRISPR/Cas que se une a un sitio diana dentro de cualquiera de las SEQ ID NO: 56, 63, 66, 71, 160, 170, 179, 183, 189, 193, 197, 200, 203, 207, 211 y 213, de forma que el gen BCL11A se inactive y aumente la expresión de gamma-globina.

La invención también proporciona una proteína de dedo de zinc que comprende 5 o 6 dominios de dedo de zinc que comprenden una región de hélice de reconocimiento, comprendiendo la proteína de dedo de zinc las regiones de

hélice de reconocimiento en el orden mostrado en una fila individual de la tabla 1A de las proteínas designadas SBS#39172, SBS#43490, SBS#44642, SBS#45148, SBS#45147, SBS#39145, SBS#44490, SBS#44489, SBS#45081, SBS#44493, SBS#29527, SBS#29528, SBS#29525, SBS#29526, SBS#34678, SBS#34642, SBS#44889, SBS#44888, SBS#44905, SBS#44904, SBS#44911, SBS#44910, SBS#44945, SBS#44944, SBS#44947 o SBS#44946.

La invención proporciona también una proteína de fusión que comprende la proteína de dedo de zinc de la invención y un dominio de escisión de tipo silvestre o genomanipulado o un semidominio de escisión de tipo silvestre o genomanipulado.

La invención proporciona un polinucleótido que codifica una o más proteínas de la invención.

La invención también proporciona una célula aislada que comprende una o más proteínas de fusión de la invención o uno o más polinucleótidos de la invención.

La invención proporciona además un kit que comprende la proteína de la invención, la proteína de fusión de la invención o el polinucleótido de la invención.

La invención proporciona un procedimiento *in vitro* para alterar la expresión del gen de globina en una célula, comprendiendo el procedimiento: introducir, en la célula, una o más proteínas de la invención o uno o más polinucleótidos de la invención, en condiciones tales que se expresen las, una o más, proteínas y se altere la expresión del gen de globina. La invención también proporciona una o más proteínas de la invención o uno o más polinucleótidos de la invención para su uso en un procedimiento de tratamiento de una hemoglobinopatía, comprendiendo el procedimiento: introducir, en la célula, una o más proteínas de la invención o uno o más polinucleótidos de la invención, en condiciones tales que se expresen una o más proteínas y se altere la expresión del gen de globina. La invención proporciona también una o más proteínas de la invención o uno o más polinucleótidos de la invención para su uso en un procedimiento de tratamiento de la enfermedad de células falciformes, comprendiendo el procedimiento: introducir, en la célula, una o más proteínas de la invención o uno o más polinucleótidos de la invención, en condiciones tales que se expresen una o más proteínas y se altere la expresión del gen de globina.

En el presente documento se divulgan procedimientos y composiciones para alterar la expresión o para corregir uno o más genes que codifican proteínas implicadas en una enfermedad genética (por ejemplo, que producen proteínas que faltan o que son deficientes o aberrantes en la enfermedad y/o proteínas que regulan estas proteínas), tales como la enfermedad de células falciformes o una talasemia. La alteración de dichas proteínas puede dar como resultado el tratamiento de estas enfermedades genéticas. En particular, la edición del genoma se utiliza para corregir un gen aberrante, insertar un gen de tipo silvestre o cambiar la expresión de un gen endógeno. A modo de ejemplo no limitante, un gen de tipo silvestre que codifica β -globina puede insertarse en una célula para producir una proteína de la que carece y/o tratar una hemoglobinopatía causada por β -globina defectuosa. En algunos casos, el gen de tipo silvestre puede insertarse en un locus de puerto seguro o en un locus que se sabe que está altamente expresado en un tejido de interés tal como el locus de β -globina en células eritroides. La edición del genoma puede utilizarse de forma similar para producir una proteína de la que carece (y por lo tanto tratar) una alfa-talasemia mediante la inserción de un gen de alfa-globina de tipo silvestre en un puerto seguro. Otro enfoque implica el uso de la corrección génica que se dirige a un gen de α - o β -globina endógeno defectuoso y se reemplaza la secuencia mutante. Alternativamente, un gen regulador involucrado en la represión de γ -globina puede alterarse o inactivarse (por ejemplo, para aumentar la expresión de γ -globina inactivando y/o reduciendo la cantidad de la proteína represiva) y/o el sitio de unión reguladora aguas arriba del gen de γ -globina o en otras zonas del locus de beta-globina puede alterarse para que los reguladores no puedan interactuar adecuadamente en el locus de γ -globina y se produzca HbF, anulando así los efectos (es decir SCD o β -talasemia) provocados por el gen aberrante de β -globina. Un enfoque implica además el uso de modificación de una célula madre (por ejemplo, células madre hematopoyéticas o precursores de glóbulos rojos), células madre que después pueden utilizarse para injertarse en un paciente para el tratamiento de una hemoglobinopatía.

En un aspecto, se describe en el presente documento una proteína de dedo de zinc (ZFP) que se une al sitio diana en una región de interés (por ejemplo, un gen de β -globina, de α -globina o de puerto seguro, o un gen regulador o su diana de ADN tal como BCL11A, γ -globina o KLF1) en un genoma, en el que la ZFP comprende uno o más dominios de unión de dedo de zinc genomanipulados. En un aspecto, la ZFP es una nucleasa de dedo de zinc (ZFN) que escinde una región genómica diana de interés, comprendiendo la ZFN uno o más dominios de unión de dedo de zinc genomanipulados y un dominio de escisión o semidominio de escisión de nucleasa. Los dominios de escisión y los semidominios de escisión pueden obtenerse, por ejemplo, a partir de diversas endonucleasas de restricción y/o endonucleasas de asentamiento (*homing*). En una forma de realización, los semidominios de escisión se derivan de una endonucleasa de restricción de tipo IIS (por ejemplo, *Fok I*). En determinados aspectos, el dominio del dedo de zinc reconoce un sitio diana en un gen de globina o de puerto seguro. En determinados aspectos, el dominio de dedo

de zinc comprende 5 o 6 dominios de dedo de zinc y reconoce un sitio diana en un gen de globina (por ejemplo, una proteína de dedo de zinc que tiene 5 o 6 dedos con las regiones de hélice de reconocimiento mostradas en la tabla 1A). En otro aspecto, el dominio del dedo de zinc reconoce un sitio diana en un gen BCL11A, KLF1, de α -, β - o γ -globina o sus elementos reguladores. En determinados aspectos, el dominio de dedo de zinc comprende 5 o 6 dominios de dedo de zinc y reconoce un sitio diana en un gen BCL11A, KLF1, de α -, β - o γ -globina o en sus elementos reguladores (por ejemplo, una proteína de dedo de zinc que tiene 5 o 6 dedos con las regiones de hélice de reconocimiento mostradas en la tabla 1A).

En otro aspecto, se describe en el presente documento una proteína TALE (similar a un activador de la transcripción) que se une al sitio diana en una región de interés (por ejemplo, un gen de α - o β -globina o de puerto seguro, o un gen regulador o su diana de ADN tal como BCL11A, γ -globina o KLF1) en un genoma, comprendiendo TALE uno o más dominios de unión de TALE genomanipulados. En un aspecto, la TALE es una nucleasa (TALEN) que escinde una región genómica diana de interés, en la que la TALEN comprende uno o más dominios de unión a ADN de TALE genomanipulados y un dominio de escisión o semidominio de escisión de nucleasa. Los dominios de escisión y los semidominios de escisión pueden obtenerse, por ejemplo, a partir de diversas endonucleasas de restricción y/o endonucleasas de asentamiento. En una forma de realización, los semidominios de escisión se derivan de una endonucleasa de restricción de tipo IIS (por ejemplo, *Fok I*). En determinados aspectos, el dominio de unión a ADN de TALE reconoce un sitio diana en un gen de globina o de puerto seguro. En otros aspectos, el dominio de unión a ADN de TALE reconoce un sitio diana en un gen BCL11A, KLF1, de α -, β - o γ -globina o en sus elementos reguladores (por ejemplo, una proteína TALEN ejemplificada en la tabla 3).

En otro aspecto, en el presente documento se describe un sistema CRISPR/Cas que se une al sitio diana en una región de interés (por ejemplo, un gen altamente expresado, un gen asociado a una enfermedad o un gen de puerto seguro) en un genoma, comprendiendo el sistema CRISPR/Cas una nucleasa CRISPR/Cas y un ARNcr/ARNtracr genomanipulado (o ARN guía único). En determinados aspectos, el sistema CRISPR/Cas reconoce un sitio diana en un gen altamente expresado, asociado a la enfermedad o de puerto seguro. En determinados aspectos, el sistema CRISPR/Cas reconoce una diana en un gen de globina, de albúmina, CCR5, CXCR4, AAVS1, *Rosa* o HPRT.

Las ZFN, las TALEN y/o el sistema CRISPR/Cas tal como se describen en el presente documento pueden unirse a, y/o escindir, la región de interés en una región codificante o no codificante dentro del gen o de forma adyacente al mismo, tal como, por ejemplo, una secuencia líder, secuencia de remolque o intrón, o dentro de una región no transcrita, ya sea aguas arriba o aguas abajo de la región codificante. En determinados aspectos, las ZFN, las TALEN y/o el sistema CRISPR/Cas se unen y/o escinden un gen de globina. En otros aspectos, las ZFN, las TALEN y/o el sistema CRISPR/Cas se unen y/o escinden un gen de puerto seguro, por ejemplo, un gen CCR5, un gen CXCR4, un gen PPP1R12C (también conocido como AAVS1), un gen de albúmina o un gen *Rosa*. Véanse, por ejemplo, las publicaciones de patente de Estados Unidos N° 20080299580, 20080159996, 201000218264, 20110301073, 20130177983 y 20130177960 y la solicitud provisional de Estados Unidos N° 61/823.689. Además, para ayudar en la selección, se puede utilizar el locus HPRT (véase la publicación de patente de Estados Unidos N° 20130122591). En otro aspecto, se describen en el presente documento composiciones que comprenden una o más de las nucleasas de dedo de zinc y/o TALE o el sistema CRISPR/Cas tal como se describen en el presente documento. En algunos aspectos, las ZFN, las TALEN y/o el sistema CRISPR/Cas se unen y escinden un gen BCL11A, KLF1, de α -, β - o γ -globina o escinden en sus elementos reguladores. En otro aspecto, se describen en el presente documento composiciones que comprenden una o más de las nucleasas de dedo de zinc, TALE o Cas tal como se describen en el presente documento.

En otro aspecto, se describe en el presente documento un polinucleótido que codifica uno o más de ZFN, TALEN y/o el sistema CRISPR/Cas tal como se describen en el presente documento. El polinucleótido puede ser, por ejemplo, ARNm. En algunos aspectos, el ARNm puede estar modificado químicamente (véase, por ejemplo, Kormann et al., (2011) *Nature Biotechnology* 29 (2): 154-157).

En otro aspecto, se describe en el presente documento un vector de expresión de ZFN, TALEN y/o el sistema CRISPR/Cas que comprende un polinucleótido, que codifica uno o más de ZFN, TALEN y/o el sistema CRISPR/Cas descritos en el presente documento, unido operativamente a un promotor. En una forma de realización, el vector de expresión es un vector vírico.

En un aspecto, se describe en el presente documento una proteína de ZFN, TALEN y/o el sistema CRISPR/Cas que se utiliza para escindir un ADN diana.

En otros aspectos, los precursores de glóbulos rojos modificados genéticamente (células madre hematopoyéticas conocidas como "HSC") se administran en un trasplante de médula ósea y los glóbulos rojos se diferencian y maduran *in vivo*. En algunas formas de realización, las HSC se aíslan después de la movilización inducida por G-CSF, y en otras, las células se aíslan a partir de médula ósea humana o de cordones umbilicales. En algunos aspectos, las HSC

se editan mediante el tratamiento con una nucleasa diseñada para inactivar un regulador de expresión de globina (por ejemplo, BCL11A o KLF1). En otros aspectos, las HSC se modifican con una nucleasa genomanipulada y un ácido nucleico donante de forma que un gen (por ejemplo, un gen de globina) de tipo silvestre se inserta y se expresa y/o se corrige un gen aberrante endógeno. En algunos casos, la secuencia del gen de tipo silvestre para la inserción codifica una β -globina de tipo silvestre o una α -globina de tipo silvestre. En otros casos, el gen aberrante endógeno es el gen de β -globina o de α -globina. En algunas formas de realización, las HSC modificadas se administran al paciente después de un precondicionamiento mieloablativo leve. En otros aspectos, las HSC se administran después de la mieloablación completa de forma que después del injerto, el 100% de las células hematopoyéticas se derivan de las HSC modificadas.

En otro aspecto se describe en el presente documento un procedimiento para escindir un gen endógeno (por ejemplo, un gen cuya inactivación da como resultado una mayor expresión de gamma-globina tal como BCL11A o KLF1) en una célula precursora de glóbulos rojos, comprendiendo el procedimiento: introducir, en la célula, uno o más polinucleótidos que codifican uno o más de ZFN, TALEN y/o un sistema CRISPR/Cas que se une a un sitio diana en los, uno o más, genes endógenos en condiciones tales que la o las ZFN, TALEN y/o el sistema CRISPR/Cas se expresan y se escinden los, uno o más, genes. En otro aspecto, se describe en el presente documento un procedimiento para escindir un gen BCL11A o KLF1 en una célula, comprendiendo el procedimiento: introducir, en la célula, uno o más polinucleótidos que codifican uno o más de ZFN, TALEN y/o sistemas CRISPR/Cas que se unen a un sitio diana en los, uno o más, genes BCL11A o KLF1 en condiciones tales que la o las ZFN, TALEN y/o el sistema CRISPR/Cas se expresan y se escinden uno o más genes BCL11A o KLF1. En determinados aspectos, el dominio de dedo de zinc comprende 5 o 6 dominios de dedo de zinc y reconoce un sitio diana en un gen de globina (por ejemplo, una proteína de dedo de zinc que tiene 5 a 6 dedos con las regiones de hélice de reconocimiento mostradas en la tabla 1A). En otros aspectos, la TALEN reconoce un sitio diana en una secuencia de β -globina, α -globina, gamma-globina, KLF o BCL11A (ejemplificada en la tabla 3). En otros aspectos, el sistema CRISPR/Cas reconoce un sitio diana en una secuencia de β -globina, α -globina, gamma-globina, KLF o BCL11A, estando el ARN guía único genomanipulado para reconocer un sitio diana deseado en el gen diana de interés. El o los genes escindidos pueden inactivarse (inactivación), por ejemplo, inactivando uno o más genes cuyo o cuyos productos pueden inhibir la expresión de un gen (por ejemplo, gen de globina), o alterando el sitio diana regulador en el ADN para dichas proteínas. En algunos aspectos, el o los genes inactivados o sus secuencias diana son los involucrados en la inhibición de la expresión de la hemoglobina fetal. Las células (por ejemplo, las células madre) cuando se diferencian contienen hemoglobina fetal y se pueden administrar a pacientes que lo necesitan. En algunos aspectos, se inactiva un gen de globina. Por ejemplo, un gen de alfa-globina puede inactivarse para restaurar la proporción de alfa-globina con respecto a beta-globina cuando una beta-globina se expresa de forma deficiente, o un gen de beta-globina que codifica HbS puede inactivarse junto con la inserción de una beta-globina de tipo silvestre. Las células (por ejemplo, las células madre) cuando se diferencian contendrán hemoglobina HbA y pueden administrarse a pacientes que lo necesitan.

En otro aspecto, se describe en el presente documento un procedimiento para insertar una secuencia en un gen endógeno (por ejemplo, un gen de beta-globina, de alfa-globina y/o de puerto seguro) en una célula (por ejemplo, célula madre), comprendiendo el procedimiento escindir el gen endógeno utilizando una o más nucleasas e insertar una secuencia en el sitio de escisión. En determinados aspectos, se reemplaza una secuencia genómica en cualquier gen diana, por ejemplo, utilizando un par de ZFN o TALEN, o un sistema CRISPR/Cas (o un vector que codifica dicha ZFN, TALEN y/o dicho sistema CRISPR/Cas) tal como se describe en el presente documento y una secuencia "donante" (también conocida como "transgén") que se inserta en el gen después de la escisión dirigida con la ZFN, TALEN y/o un sistema CRISPR/Cas. La secuencia donante puede estar presente en el vector ZFN o TALEN, presente en un vector separado (por ejemplo, Ad, AAV o LV) o, como alternativa, puede introducirse en la célula utilizando un mecanismo diferente de administración de ácido nucleico. Dicha inserción de una secuencia de nucleótidos donante en el locus diana (por ejemplo, gen de globina, otro gen de puerto seguro, etc.) da como resultado la expresión del transgén bajo el control de elementos de control genético del locus diana (por ejemplo, de globina). En algunas formas de realización, el transgén codifica un ARN no codificante (por ejemplo, un ARNsh). La expresión del transgén antes de la maduración del glóbulo rojo dará como resultado un glóbulo rojo que contiene el ARN no codificante de interés.

En otras formas de realización, el transgén comprende una proteína funcional, por ejemplo, una proteína globina (por ejemplo, beta y/o gamma de tipo silvestre). En algunas formas de realización, la inserción del transgén de interés en un gen endógeno (por ejemplo, un gen de globina) da como resultado la expresión de una secuencia de proteína exógena intacta y carece de cualquier secuencia codificada por el gen endógeno. En otras formas de realización, la proteína exógena expresada es una proteína de fusión y comprende aminoácidos codificados por el transgén y por un gen de globina (por ejemplo, del locus diana endógeno o, alternativamente, de secuencias que codifican la globina en el transgén). En algunos casos, el gen de globina es una beta-globina, mientras que, en otros casos, el gen de la globina es una alfa-globina. En otros casos, el gen de globina es un gen de gamma-globina. Cuando están presentes, las secuencias de globina endógenas pueden estar presentes en la porción amino-terminal (N) de la proteína exógena y/o la porción carboxi-terminal (C) de la proteína exógena. Las secuencias de globina pueden incluir secuencias de globina de tipo silvestre o mutantes de longitud completa o, como alternativa, pueden incluir secuencias codificantes

de globina parciales. En algunas formas de realización, la fusión globina-transgén se localiza en el locus endógeno dentro de la célula, mientras que, en otras formas de realización, la secuencia codificante globina-transgén se inserta en un puerto seguro dentro de un genoma. En algunos aspectos, el puerto seguro se selecciona de un gen CCR5, un gen CXCR4, un gen PPP1R12C (también conocido como AAVS1), un gen de albúmina o un gen *Rosa*. Véanse, por ejemplo, las publicaciones de patente de Estados Unidos N° 20080299580, 20080159996, 201000218264, 20110301073, 20130177983 y 20130177960 y la solicitud provisional de Estados Unidos N° 61/823.689. Además, para ayudar en la selección, se puede utilizar el locus HPRT (véase la publicación de patente de Estados Unidos N° 20130122591).

En otro aspecto más, en el presente documento se divulgan líneas celulares y/o modelos (sistemas) animales transgénicos. En algunos aspectos, la célula y/o el animal transgénico incluye un transgén que codifica un gen humano. En algunos casos, el animal transgénico comprende una inactivación en el locus endógeno correspondiente al transgén exógeno (por ejemplo, el gen de globina de ratón se inactiva y el gen de globina humana se inserta en un ratón), lo que permite el desarrollo de un sistema *in vivo* en el que la proteína humana puede estudiarse de forma aislada. Dichos modelos transgénicos pueden utilizarse con fines de cribado para identificar moléculas pequeñas o biomoléculas grandes u otras entidades que pueden interactuar con, o modificar, la proteína humana de interés. En algunos aspectos, el transgén se integra en el locus seleccionado (por ejemplo, globina o puerto seguro) en una célula madre (por ejemplo, una célula madre embrionaria, una célula madre pluripotente inducida, una célula madre hematopoyética, etc.) o un embrión animal obtenido por cualquiera de los procedimientos descritos en el presente documento, y después el embrión se implanta de forma que nazca un animal vivo. En otros aspectos, las células madre contienen alteraciones genómicas en loci endógenos tales como los genes BCL11A, KLF1 o de γ -globina, o combinaciones de los mismos, de modo que la expresión de γ -globina es elevada. En algunas formas de realización, la elevación de la expresión de γ -globina altera la relación de γ -globina con respecto a β -globina en la célula en comparación con la célula madre no editada. Después, el animal se cría hasta la madurez sexual y se le permite producir descendencia en la que al menos parte de la descendencia comprende una secuencia genética endógena editada o el transgén integrado.

En otro aspecto adicional, en el presente documento se divulga un procedimiento para la integración específica del sitio de una secuencia de ácido nucleico en un locus endógeno (por ejemplo, gen de globina o de puerto seguro) de un cromosoma, por ejemplo, en el cromosoma de un embrión. En determinados aspectos de la divulgación, el procedimiento comprende: (a) inyectar al embrión (i) al menos un vector de ADN, en el que el vector de ADN comprende una secuencia aguas arriba y una secuencia aguas abajo que flanquea la secuencia de ácido nucleico que se va a integrar, y (ii) al menos una molécula de ARN que codifica una nucleasa de dedo de zinc, TALE o Cas9. En el caso de utilizar una proteína Cas9, también se introduce un ARNsg genomanipulado. La nucleasa o el sistema de nucleasa reconoce el sitio diana en el locus diana (por ejemplo, globina o puerto seguro), y después (b) el embrión se cultiva para permitir la expresión de la nucleasa de dedo de zinc o TALE y/o el sistema CRISPR/Cas, en el que una rotura bicatenaria en la diana introducida por la nucleasa de dedo de zinc, TALEN o el sistema CRISPR/Cas se repara después, mediante recombinación homóloga con el vector de ADN, para integrar la secuencia de ácido nucleico en el cromosoma.

En cualquiera de los procedimientos descritos en el presente documento, el polinucleótido que codifica la o las nucleasas de dedo de zinc, la o las TALEN y/o el sistema CRISPR/Cas puede comprender ADN, ARN o combinaciones de los mismos. En determinadas formas de realización, el polinucleótido comprende un plásmido. En otras formas de realización, el polinucleótido que codifica la nucleasa comprende ARNm.

También se describe un kit que comprende las ZFP, TALEN y/o el sistema CRISPR/Cas de la invención. El kit puede comprender ácidos nucleicos que codifican las ZFP, TALEN o el sistema CRISPR/Cas, (por ejemplo, moléculas de ARN o genes que codifican ZFP, TALEN o Cas9 contenidos en un vector de expresión adecuado) y ARNsg genomanipulado si es necesario, o partes alícuotas de las proteínas nucleasas, moléculas donantes, líneas de células huésped adecuadas, instrucciones para llevar a cabo los procedimientos de la invención y similares.

Estos y otros aspectos serán fácilmente evidentes para el experto en la técnica a la luz de la divulgación en su conjunto.

Breve descripción de los dibujos

La figura 1 representa un alineamiento de la secuencia del gen de la globina de tipo β en la región que rodea la mutación de enfermedad de células falciformes (indicada en la figura). Se muestran (de la línea superior a la línea inferior) las secuencias beta de hemoglobina con la mutación de células falciformes (falciforme HBB, SEQ ID NO: 1); hemoglobina beta (HBB, SEQ ID NO: 2); hemoglobina delta (HBD, SEQ ID NO: 3); un pseudogén de hemoglobina beta (HBBP1, SEQ ID NO: 4); hemoglobina épsilon (HBE1, SEQ ID NO: 5); hemoglobina gamma 1 (HBG1, SEQ ID NO: 6) y hemoglobina gamma 2 (HBG2, SEQ ID NO: 7). Los resultados del análisis de actividad Cel 1 se muestran debajo del alineamiento para los cinco pares de ZFN indicados.

La figura 2, paneles A y B, son geles que representan la inserción de una secuencia especificada por un donante de β -globina en células CD34+. La figura 2A (RFLP) representa la inserción de una secuencia especificada por un donante de β -globina en la que la inserción se verifica por la presencia de un nuevo sitio de restricción presente en el ADN donante. La figura 2B representa los resultados de un ensayo de desajuste Cel-1 (Surveyor™, Transgenomic) que demuestra la presencia de nuevas secuencias que crean un desajuste. El porcentaje de alelos que portan la mutación (% de NHEJ) se indica en el texto a la derecha de los geles (la primera columna corresponde al número de carril en los geles). Los números se refieren a combinaciones de ZFN; unt: control no transfectado.

La figura 3 es un gráfico que representa los papeles que desempeñan KLF1 y BCL11A en la regulación de la expresión génica de β - y gamma-globina. La expresión de KLF1 estimula la expresión de los genes BCL11a y de β -globina. La proteína BCL11A reprime la expresión de gamma-globina.

La figura 4, paneles A a E, representan geles que muestran los resultados de un ensayo Cel 1 tal como se ha descrito anteriormente después del tratamiento de HSC con las ZFN específicas de BCL11A (figura 4A), de KLF1 (figuras 4C y 4D) o de HPRT (figura 4B) indicadas y ya sea tratando las células transducidas con un breve choque hipotérmico (30°) o en condiciones estándar (37°). El ADN se recolectó 3 días después de la transfección. La figura 4E representa el mismo tipo de análisis de Cel 1 realizado con muestras recogidas 3 días después de la transfección de las HSC o después de 17 días de diferenciación eritroide. El porcentaje de alelos que portan la mutación (% de NHEJ) se indica en la parte inferior de los carriles, y la identidad de los pares de ZFN utilizados se indica en cada figura.

La figura 5, paneles A y B, representa la expresión de gamma-globina en comparación con beta-globina (figura 5A) o ARNm de gamma-globina corregida con el patrón de ARN 18s (figura 5B) o bien 7 o bien 17 días después de la diferenciación según se analiza por un procedimiento Taqman®. El porcentaje de ARNm de gamma-globina en comparación con el ARNm de gamma + beta-globina se muestra encima de cada barra en la figura 5A. En la figura 5B, el nivel relativo de gamma-globina normalizado por el ARN 18s se representa encima de las barras, y demuestra que el nivel de ARNm de gamma-globina con respecto a 18S es mayor en las células que han sido tratadas con las ZFN específicas de BCL11A.

La figura 6 representa la cantidad de ARNm de gamma-globina en colonias en metilcelulosa derivadas de HSC dependiendo del genotipo de las células. Las células en las que ambos genes BCL11A son de tipo silvestre ("BB") producen la menor cantidad de ARNm de gamma-globina en comparación con las células que han tenido un único alelo de inactivación de BCL11A ("Bb") o han tenido ambos alelos inactivados ("inactivación"). Los números dispuestos encima de las barras indican el porcentaje de gamma-globina producida a partir de la beta-globina total.

La figura 7 muestra una serie de secuencias de ADN (SEQ ID NO: 140 a 148) de la región aguas arriba del gen de gamma-globina después del tratamiento con ZFN específicas de gamma-globina en células K562. Las secuencias tienen una serie de inserciones y deleciones que incluyen una deleción de 13 pb ("Δ13 pb") que es idéntica a uno de los genotipos humanos asociados con HPFH. La secuencia de 'referencia' en la parte superior es la secuencia de la región reguladora de tipo silvestre 5' para la gamma-globina. Los sitios de unión del par de ZFN están resaltados en rojo, la deleción de 13 pb que tiene lugar de forma natural está subrayada.

La figura 8, paneles A a C, representa el análisis de Taqman de colonias eritroides derivadas de HSC tratadas con ZFN dirigidas al promotor de la gamma-globina y después sembradas en colonias de metilcelulosa. Los números de las colonias se indican en la parte inferior de cada barra, como el genotipo. La figura 8A muestra las relaciones relativas de ARNm de gamma/beta-globina; la figura 8B muestra los niveles de ARNm de gamma-globina corregidos mediante los niveles de ARN 18s y la figura 8C muestra el análisis correspondiente de los niveles de beta-globina corregidos mediante 18s. La comparación de los promedios de las relaciones para colonias de tipo silvestre y mutadas indica que los niveles de gamma-globina en las colonias con mutaciones inducidas por ZFN en el promotor de gamma-globina son elevados.

La figura 9 muestra la región promotora del gen de gamma-globina (SEQ ID NO: 149-152). Dos alelos de gamma-globina están alineados (HBG1 y HBG2). Las diferencias en las secuencias de los dos alelos se indican con cuadros grises. Además, las mutaciones asociadas con HPFH se indican con contornos negros. El ATG inicial se indica al igual que los límites del exón 1. El aumento en los niveles de globina fetal asociados con cada mutación se indica mediante un número dispuesto encima de la misma.

La figura 10, paneles A y B, representa la cantidad de NHEJ (es decir, alteración del locus diana que es resultado de un evento de reparación basado en NHEJ de la rotura dirigida por nucleasa) y la corrección génica detectada para el gen de beta-globina en células CD34+ utilizando la nucleasa de dedo de zinc indicada y donantes oligonucleotídicos.

La figura 11 representa la cantidad de NHEJ y la integración dirigida de un nucleótido donante en células CD34+, siendo los brazos de homología en el donante variados.

5 La figura 12 representa la persistencia de la edición de genes en derivados eritroides de células madre que han sido tratadas con ZFN y donantes oligonucleotídicos. La modificación génica se analizó en cuatro tipos de poblaciones celulares derivadas de la diferenciación, unidades formadoras de colonias, eritroides ("CFU-E"), unidades formadoras de estallidos, eritroides ("BFU-E"), unidades formadoras de colonias, granulocitos/macrófagos ("CFU-GM") y unidades formadoras de colonias, granulocitos/eritrocitos/monocitos/macrófagos ("CFU-GEMM").

10 La figura 13 representa la estabilidad de la modificación génica del gen de beta-globina a lo largo del tiempo.

Descripción detallada

15 En el presente documento se divulgan procedimientos y composiciones para estudiar y tratar una enfermedad genética, tal como una hemoglobinopatía. La divulgación describe la edición genómica de una célula diana de forma que haya un cambio favorable en la expresión de uno o más genes de globina, que a su vez da como resultado el tratamiento de hemoglobinopatías tales como la enfermedad de células falciformes o una talasemia en un sujeto con necesidad de ello. Los cambios favorables en la expresión de un gen de globina incluyen, pero sin limitación, la provisión de un gen de γ -globina en un sujeto con β -globina aberrante; y/o la corrección de una secuencia aberrante de genes de α - o β -globina. Además, la administración de células madre hematopoyéticas alteradas en un trasplante alterado para expresar un producto proteico deseado puede ser igualmente beneficioso en el tratamiento de hemoglobinopatías tales como la anemia de células falciformes o la talasemia. También se describen líneas celulares y animales con expresión de globina alterada.

25 Por lo tanto, los procedimientos y las composiciones divulgados pueden utilizarse para alterar la expresión de uno o más genes de globina (por ejemplo, γ , α y/o β) en una célula (por ejemplo, una célula precursora eritroide). Estos procedimientos y estas composiciones se pueden utilizar para alterar genes involucrados en la represión de γ -globina (por ejemplo, BCL11A o KLF1), de forma que, después de la edición, las células expresarán γ -globina a niveles más altos, y se puede producir HbF. Alternativamente, la edición puede utilizarse para interrumpir el sitio de unión en un gen (por ejemplo, interrumpir la unión de BCL11A en el locus de beta-globina, interrumpir la unión del represor de la transcripción de la gamma-globina en el promotor de la gamma-globina) para desactivar la represión de un gen. Como alternativa o adicionalmente a estas alteraciones, los procedimientos y las composiciones se pueden utilizar para corregir un gen de α - y/o β -globina endógeno aberrante o insertar un gen de tipo silvestre en una ubicación deseada en el genoma de una célula (por ejemplo, en un HSC). Las células precursoras pueden derivarse de sujetos con necesidad, modificarse *ex vivo*, y después devolverse al sujeto en un injerto de médula ósea.

General

40 La puesta en práctica de los procedimientos, así como la preparación y el uso de las composiciones descritas en el presente documento emplean, a menos que se indique lo contrario, técnicas convencionales en biología molecular, bioquímica, estructura y análisis de cromatina, química computacional, cultivo celular, ADN recombinante y campos relacionados que se encuentran dentro de la capacidad de la técnica. Estas técnicas se explican completamente en la literatura. Véanse, 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; las series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, tercera edición, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P.M. Wassarman y A. P. Wolffe, eds.), Academic Press, San Diego, 1999; y METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P.B. Becker, ed.) Humana Press, Totowa, 1999.

Definiciones

55 Los términos "ácido nucleico", "polinucleótido" y "oligonucleótido" se utilizan indistintamente y se refieren a un polímero de desoxirribonucleótidos o ribonucleótidos, en conformación lineal o circular, y en forma monocatenaria o bicatenaria. Para los fines de la presente divulgación, estos términos no deben considerarse 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 y/o fosfato (por ejemplo, cadenas principales de fosforotioato). En general, un análogo de un nucleótido particular tiene la misma especificidad de emparejamiento de bases; es decir, un análogo de A se emparejará por bases con T.

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

5 "Unión" se refiere a una interacción no covalente específica de secuencia entre macromoléculas (por ejemplo, entre una proteína y un ácido nucleico). No todos los componentes de una interacción de unión deben ser específicos de secuencia (por ejemplo, contactos con residuos de fosfato en una cadena principal de ADN), siempre que la interacción en su conjunto sea específica de secuencia. Dichas interacciones se caracterizan generalmente por una constante de disociación (K_d) de 10^{-6} M^{-1} o inferior. "Afinidad" se refiere a la fuerza de la unión: una afinidad de unión aumentada se correlaciona con una K_d inferior.

10 Una "proteína de unión" es una proteína que es capaz de unirse de forma no covalente a otra molécula. Una proteína de unión puede unirse a, por ejemplo, una molécula de ADN (una proteína de unión a ADN), una molécula de ARN (una proteína de unión a ARN) y/o una molécula de proteína (una proteína de unión a proteína). En el caso de una proteína de unión a proteína, puede unirse a sí misma (para formar homodímeros, homotrímeros, etc.) y/o puede unirse a una o más moléculas de una proteína o de proteínas diferentes. Una proteína de unión puede tener más de un tipo de actividad de unión. Por ejemplo, las proteínas de dedo de zinc tienen actividad de unión a ADN, de unión a ARN y de unión a proteína.

15 Una "proteína de unión a ADN de dedo de zinc" (o dominio de unión) es una proteína, o un dominio dentro de una proteína más grande, que se une al ADN de una forma específica de 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 dedo de zinc a menudo se abrevia como proteína de dedo de zinc o ZFP.

20 Un "dominio de unión a ADN de TALE" o "TALE" es un polipéptido que comprende uno o más dominios/unidades de repetición TALE. Los dominios de repetición están involucrados en la unión del TALE a su secuencia de ADN diana asociada. Una sola "unidad de repetición" (también denominada "repetición") tiene típicamente una longitud de 33-35 aminoácidos y muestra al menos alguna homología de secuencia con otras secuencias de repetición TALE dentro de una proteína TALE natural. Véase, por ejemplo, la publicación de patente de Estados Unidos N° 20110301073.

25 Los dominios de unión de dedo de zinc y TALE se pueden "genomanipular" para que se unan a una secuencia de nucleótidos predeterminada, por ejemplo, mediante genomanipulación (alterando uno o más aminoácidos) de la región de hélice de reconocimiento de una proteína de dedo de zinc o TALE de origen natural. Por lo tanto, las proteínas (de dedo de zinc o TALE) de unión a ADN genomanipuladas son proteínas que no tienen un origen natural. Ejemplos no limitantes de procedimientos 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 resultado, principalmente, de criterios racionales. Los criterios racionales para el diseño incluyen la aplicación de reglas de sustitución y algoritmos computarizados para procesar información en una base de datos que almacena información de diseños y datos de unión de ZFP y/o TALE existentes. Véanse, por ejemplo, las patentes de Estados Unidos 6.140.081, 6.453.242 y 6.534.261; véanse también los documentos WO 98/53058, WO 98/53059, WO 98/53060, WO 02/016536 y WO 03/016496 y la publicación de Estados Unidos N° 20110301073.

35 Una proteína de dedo de zinc o TALE "seleccionada" es una proteína que no se encuentra en la naturaleza cuya producción es resultado, principalmente, de un proceso empírico tal como presentación en fagos, trampa de interacción o selección de híbridos. Véanse, por ejemplo, los documentos US 5.789.538, US 5.925.523, US 6.007.988, US 6.013.453, US 6.200.759, WO 95/19431, WO 96/06166, WO 98/53057, WO 98/54311, WO 00/27878, WO 01/60970, WO 01/88197, WO 02/099084 y la publicación de Estados Unidos N° 20110301073.

40 "Recombinación" se refiere a un proceso de intercambio de información genética entre dos polinucleótidos. Para los fines de la presente divulgación, "recombinación homóloga" (HR) se refiere a la forma especializada de dicho intercambio que tiene lugar, por ejemplo, durante la reparación de roturas bicatenarias en células mediante mecanismos de reparación dirigidos por homología. Este proceso requiere una homología de secuencia de nucleótidos, utiliza una molécula "donante" para la reparación asistida por plantilla de una molécula "diana" (es decir, la que experimentó la rotura bicatenaria), y recibe diversas denominaciones tales como "conversión génica sin entrecruzamiento" o "conversión génica de tramo corto", porque da lugar a la transferencia de información genética del donante a la diana. Sin pretender vincularse a ninguna teoría en particular, dicha transferencia puede implicar una corrección de desajuste del ADN heterodúplex que se forma entre la diana rota y el donante, y/o la "hibridación de cadenas dependiente de síntesis", en la que el donante se utiliza para volver a sintetizar información genética que se convertirá en parte de la diana, y/o procesos relacionados. Dicha HR especializada da como resultado a menudo una alteración de la secuencia de la molécula diana de tal forma que parte o toda la secuencia del polinucleótido donante se incorpora en el polinucleótido diana.

- En los procedimientos de la divulgación, una o más nucleasas dirigidas tal como se describen en el presente documento crean una rotura bicatenaria en la secuencia diana (por ejemplo, cromatina celular) en un sitio predeterminado, y un polinucleótido "donante", que tiene homología con la secuencia de nucleótidos en la región de la rotura, puede introducirse en la célula. Se ha demostrado que la presencia de la rotura bicatenaria facilita la integración de la secuencia donante. La secuencia donante puede integrarse físicamente o, como alternativa, el polinucleótido donante se utiliza como plantilla para reparar la rotura mediante recombinación homóloga, lo que 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, una primera secuencia en la cromatina celular puede alterarse y, en determinadas formas de realización, puede convertirse en una secuencia presente en un polinucleótido donante. Por lo tanto, el uso de los términos "reemplazar" o "reemplazo" puede entenderse que representa el reemplazo de una secuencia de nucleótidos por otra, (es decir, el reemplazo de una secuencia en un sentido informativo), y no necesariamente requiere el reemplazo físico o químico de un polinucleótido por otro.
- En cualquiera de los procedimientos descritos en el presente documento, se pueden utilizar pares adicionales de proteínas de dedo de zinc o TALEN para la escisión bicatenaria adicional de sitios diana adicionales dentro de la célula. Además, puede emplearse un sistema CRISPR/Cas de forma similar para inducir roturas bicatenarias adicionales.
- En procedimientos para la recombinación y/o el reemplazo y/o la alteración dirigidos de una secuencia en una región de interés en la cromatina celular, una secuencia cromosómica se altera mediante recombinación homóloga con una secuencia de nucleótidos "donante" exógena. Dicha recombinación homóloga es estimulada por la presencia de una rotura bicatenaria en la cromatina celular, si están presentes secuencias homólogas a la región de la rotura.
- En cualquiera de los procedimientos descritos en el presente documento, la secuencia de nucleótidos exógenos (la "secuencia donante" o "transgén") puede contener secuencias que son homólogas, pero no idénticas, a secuencias genómicas en la región de interés, estimulando así la recombinación homóloga para insertar una secuencia no idéntica presente en la región de interés. Por lo tanto, en determinadas formas de realización, partes de la secuencia donante que son homólogas a secuencias presentes en la región de interés muestran entre aproximadamente el 80 y el 99% (o cualquier número entero intermedio) de identidad de secuencia con la secuencia genómica que se reemplaza. En otras formas de realización, la homología entre la secuencia donante y la genómica es superior al 99%, por ejemplo, si solo 1 nucleótido difiere entre las secuencias donante y genómica de más de 100 pares de bases contiguos. En determinados casos, una parte no homóloga de la secuencia donante puede contener secuencias que no están presentes en la región de interés, de forma que se introducen nuevas secuencias en la región de interés. En estos casos, la secuencia no homóloga generalmente está flanqueada por secuencias de 50-1.000 pares de bases (o cualquier valor entero intermedio) o cualquier número de pares de bases superior a 1.000, que son homólogas o idénticas a las secuencias presentes en la región de interés. En otras formas de realización, la secuencia donante no es homóloga a la primera secuencia, y se inserta en el genoma mediante mecanismos de recombinación no homóloga.
- Puede utilizarse cualquiera de los procedimientos descritos en el presente documento para la inactivación parcial o completa de una o más secuencias diana en una célula mediante la integración dirigida de una secuencia donante que interrumpe la expresión del o de los genes de interés. También se divulgan líneas celulares con genes parcialmente o completamente inactivados.
- Además, los procedimientos de integración dirigida tal como se describen en el presente documento también se pueden utilizar para integrar una o más secuencias exógenas. La secuencia de ácido nucleico exógeno puede comprender, por ejemplo, uno o más genes o moléculas de ADNc, o cualquier tipo de secuencia codificante o no codificante, así como uno o más elementos de control (por ejemplo, promotores). Además, la secuencia de ácido nucleico exógeno puede producir una o más moléculas de ARN (por ejemplo, ARN de horquilla pequeña (ARNsh), ARN inhibidores (ARNi), microARN (miRNA), etc.).
- "Escisión" se refiere a la rotura covalente de la cadena principal de una molécula de ADN. La escisión puede iniciarse mediante una diversidad de procedimientos que incluyen, pero sin limitación, 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 del ADN puede dar como resultado la producción de extremos romos o extremos escalonados. En determinadas formas de realización, se utilizan polipéptidos de fusión para la escisión dirigida de ADN bicatenario.
- Un "semidominio de escisión" es una secuencia de polipéptidos que, junto con un segundo polipéptido (ya sea idéntico o diferente) forma un complejo que tiene actividad de escisión (preferentemente actividad de escisión bicatenaria). Los términos "primer y segundo semidominios de escisión", "semidominios de escisión + y -" y "semidominios de escisión derecho e izquierdo" se utilizan indistintamente para referirse a pares de semidominios de escisión que se dimerizan.

Un "semidominio de escisión genomanipulado" es un semidominio de escisión que se ha modificado para que forme heterodímeros obligados con otro semidominio de escisión (por ejemplo, otro semidominio de escisión genomanipulado). Véanse también las publicaciones de patente de Estados Unidos N° 2005/0064474, 2007/0218528, 2008/0131962 y 2011/0201055.

El término "secuencia" se refiere a una secuencia de nucleótidos de cualquier longitud, que puede ser 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 tener cualquier longitud, por ejemplo, de entre 2 y 10.000 nucleótidos de longitud (o cualquier valor entero intermedio o superior), preferentemente de entre aproximadamente 100 y 1.000 nucleótidos de longitud (o cualquier número entero intermedio), de forma más preferida de entre aproximadamente 200 y 500 nucleótidos de longitud.

Un "gen asociado a enfermedad" es uno que es defectuoso de alguna manera en una enfermedad monogénica. Los ejemplos no limitantes de enfermedades monogénicas incluyen inmunodeficiencia combinada grave, fibrosis quística, enfermedades de almacenamiento lisosómico (por ejemplo, de Gaucher, de Hurler, de Hunter, de Fabry, de Neimann-Pick, de Tay-Sach, etc.), anemia de células falciformes y talasemia.

La "cromatina" es la estructura de nucleoproteínas que comprende el genoma celular. La cromatina celular comprende ácido nucleico, principalmente ADN y proteínas, incluidas histonas y proteínas cromosómicas distintas de histonas. La mayor parte de la cromatina celular eucariota existe en forma de nucleosomas, en los que un núcleo de nucleosoma comprende aproximadamente 150 pares de bases de ADN asociados con un octamero que comprende dos de las histonas H2A, H2B, H3 y H4; y ADN enlazador (de longitud variable dependiendo del organismo) que se extiende entre núcleos de nucleosomas. Una molécula de histona H1 generalmente se asocia con el ADN enlazador. Para los fines de la presente divulgación, se pretende que el término "cromatina" abarque todos los tipos de nucleoproteína celular, tanto procariotas como eucariotas. La cromatina celular incluye cromatina tanto cromosómica como episómica.

Un "cromosoma" es un complejo de cromatina que comprende la totalidad o una parte 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, un complejo de nucleoproteína u otra estructura que comprende un ácido nucleico que no forma parte del cariotipo cromosómico de una célula. Los ejemplos de episomas incluyen plásmidos y determinados genomas víricos.

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

Una molécula "exógena" es una molécula que normalmente no está presente en una célula, pero que puede introducirse en una célula mediante uno o más procedimientos genéticos, bioquímicos o de otro tipo. La "presencia normal en la célula" se determina con respecto al estadio de desarrollo particular y las condiciones ambientales de la célula. Así, por ejemplo, una molécula que está presente solo durante el desarrollo embrionario del músculo es una molécula exógena con respecto a una célula muscular adulta. De forma similar, una molécula inducida por choque térmico es una molécula exógena con respecto a una célula que no ha experimentado un choque térmico. Una molécula exógena puede comprender, por ejemplo, una versión funcional de una molécula endógena disfuncional o una versión disfuncional de una molécula endógena normalmente funcional.

Una molécula exógena puede ser, entre otras cosas, una molécula pequeña, tal como la que se genera mediante un proceso químico combinatorio, o una macromolécula tal como una proteína, un ácido nucleico, un carbohidrato, un lípido, una glicoproteína, una lipoproteína, un 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 monocatenarios o bicatenarios; puede ser lineales, ramificados o circulares; y pueden tener cualquier longitud. Los ácidos nucleicos incluyen aquellos capaces de formar dúplex, así como los ácidos nucleicos que forman tríplex. Véanse, por ejemplo, las patentes de Estados Unidos N° 5.176.996 y 5.422.251. Las proteínas incluyen, pero sin limitación, 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, quinasas, fosfatasas, integrasas, recombinasas, ligasas, topoisomerasas, girasas y helicasas.

Una molécula exógena puede ser del mismo tipo de molécula que una molécula endógena, por ejemplo, una proteína o ácido nucleico exógenos. Por ejemplo, un ácido nucleico exógeno puede comprender un genoma vírico infeccioso, un plásmido o episoma introducido en una célula o un cromosoma que normalmente no está presente en la célula. Los expertos en la técnica conocen los procedimientos para la introducción de moléculas exógenas en células y estos

incluyen, pero sin limitación, transferencia mediada por lípidos (es decir, liposomas, incluidos lípidos neutros y catiónicos), electroporación, inyección directa, fusión celular, bombardeo de partículas, coprecipitación con fosfato de calcio, transferencia mediada por DEAE-dextrano y transferencia mediada por vectores víricos. Una molécula exógena también puede ser del mismo tipo de molécula que una molécula endógena pero derivada de una especie diferente de la que se deriva la célula. Por ejemplo, una secuencia de ácido nucleico humano puede introducirse en una línea celular originariamente derivada de un ratón o un hámster.

Por el contrario, una molécula "endógena" es una que normalmente está presente en una célula particular en un estadio 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 orgánulo, o un ácido nucleico episómico de origen natural. Las moléculas endógenas adicionales pueden incluir proteínas, por ejemplo, factores de transcripción y enzimas.

Una molécula de "fusión" es una molécula en la que dos o más moléculas de subunidades están unidas, preferentemente de forma covalente. Las moléculas de subunidades pueden ser del mismo tipo químico de molécula, o pueden ser de diferentes tipos químicos de moléculas. Los ejemplos del primer tipo de molécula de fusión incluyen, pero sin limitación, 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 nucleico ácido que codifica la proteína de fusión descrita anteriormente). Los ejemplos del segundo tipo de molécula de fusión incluyen, pero sin limitación, una fusión entre un ácido nucleico que forma tríplex y un polipéptido, y una fusión entre un ligante de surco menor y un ácido nucleico.

La expresión de una proteína de fusión en una célula puede ser el resultado de la administración de la proteína de fusión a la célula o de la administración de un polinucleótido que codifica la proteína de fusión a una célula, en la que el polinucleótido se transcribe y el transcrito se traduce, para generar la proteína de fusión. El corte y empalme en trans, la escisión de polipéptidos y la ligadura de polipéptidos también pueden estar implicados en la expresión de una proteína en una célula. Los procedimientos para la administración de polinucleótidos y polipéptidos a las células se presentan en otra parte de la presente divulgación.

Un "gen", para los fines de la presente divulgación, incluye una región de ADN que codifica un producto génico (véase más adelante), así como todas las regiones de ADN que regulan la producción del producto génico, independientemente de si dichas secuencias reguladoras son adyacentes o no a las secuencias codificadoras y/o transcritas. En consecuencia, un gen incluye, pero no se limita necesariamente a, secuencias promotoras, terminadores, secuencias reguladoras traduccionales tales como sitios de unión a ribosoma y sitios de entrada al ribosoma internos, potenciadores, silenciadores, aisladores, elementos de límite, orígenes de replicación, sitios de unión a la matriz y regiones de control de locus.

"Expresión génica" se refiere a la conversión de la información, contenida en un gen, en un producto génico. Un producto génico puede ser el producto transcripcional directo de un gen (por ejemplo, ARNm, ARNt, ARNr, ARN antisentido, ribozima, ARN estructural o cualquier otro tipo de ARN) o una proteína producida por la traducción de un ARNm. Los productos génicos también incluyen ARN que se han modificados mediante procesos tales como encapsulación terminal, poliadenilación, metilación y edición, y proteínas modificadas, por ejemplo, por metilación, acetilación, fosforilación, ubiquitinación, ADP-ribosilación, miristilación y glicosilación.

"Modulación" de la expresión génica se refiere a un cambio en la actividad de un gen. La modulación de la expresión puede incluir, pero sin limitación, activación génica y represión génica. La edición (por ejemplo, escisión, alteración, inactivación, mutación aleatoria) genómica puede utilizarse para modular la expresión. La inactivación génica se refiere a cualquier reducción en la expresión génica en comparación con una célula que no incluye una ZFP o TALEN tal como se describe en el presente documento. Por lo tanto, la inactivación genética puede ser parcial o completa.

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 que es deseable unir una molécula exógena. La unión puede ser para fines de escisión de ADN dirigida y/o recombinación dirigida. Una región de interés puede estar presente en un cromosoma, un episoma, un genoma de orgánulo (por ejemplo, mitocondrial, de 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 de remolque o intrones, o dentro de regiones no transcritas, ya sea aguas arriba o aguas abajo de la región codificante. Una región de interés puede ser tan pequeña como de un solo par de nucleótidos o de hasta 2.000 pares de nucleótidos de longitud, o cualquier valor entero de pares de nucleótidos.

Las células "eucariotas" incluyen, pero sin limitación, células fúngicas (tales como levadura), células vegetales, células animales, células de mamíferos y células humanas (por ejemplo, células T).

5 Los "glóbulos rojos" (hematíes), o eritrocitos, son células diferenciadas de forma terminal derivadas de células madre hematopoyéticas. Carecen de una nucleasa y de la mayor parte de los orgánulos celulares. Los glóbulos rojos contienen hemoglobina para transportar oxígeno desde los pulmones a los tejidos periféricos. De hecho, el 33% de un glóbulo rojo individual es hemoglobina. También transportan el CO₂ producido por las células durante el metabolismo fuera de los tejidos y de nuevo a los pulmones para liberarlo durante la exhalación. Los glóbulos rojos se producen en la médula ósea en respuesta a la hipoxia sanguínea que está mediada por la liberación de eritropoyetina (EPO) por el riñón. La EPO provoca un aumento en el número de proeritroblastos y acorta el tiempo requerido para la maduración completa de los glóbulos rojos. Después de aproximadamente 120 días, dado que los glóbulos rojos no contienen un núcleo ni ninguna otra capacidad regenerativa, las células son eliminadas de la circulación por las actividades fagocíticas de los macrófagos en el hígado, el bazo y los ganglios linfáticos (~90%) o por hemólisis en el plasma (~10%). Después de ser engullidas por los macrófagos, los componentes químicos de los glóbulos rojos se degradan dentro de las vacuolas de los macrófagos debido a la acción de las enzimas lisosómicas.

15 Los "tejidos secretores" son aquellos tejidos en un animal que secretan productos fuera de la célula individual en un lumen de algún tipo que generalmente se deriva del epitelio. Los ejemplos de tejidos secretores que se localizan en el tubo gastrointestinal incluyen las células que recubren el intestino, el páncreas y la vesícula biliar. Otros tejidos secretores incluyen el hígado, tejidos asociados con el ojo y membranas mucosas, tales como las glándulas salivales, las glándulas mamarias, la próstata, la glándula pituitaria y otros elementos del sistema endocrino. Además, los tejidos secretores incluyen células individuales de un tipo de tejido que son capaces de realizar una secreción.

25 Los términos "unión operativa" y "unido operativamente" (o "unido de forma operativa") se utilizan indistintamente con referencia a una yuxtaposición de dos o más componentes (tales como elementos de secuencia), en la que los componentes están dispuestos de forma que ambos componentes funcionen normalmente 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 de la transcripción, tal como un promotor, está operativamente unida a una secuencia codificante si la secuencia reguladora de la transcripción controla el nivel de transcripción de la secuencia codificante en respuesta a la presencia o la ausencia de uno o más factores reguladores de la transcripción. Una secuencia reguladora de la transcripción generalmente está operativamente unida en *cis* con una secuencia codificante, pero no es necesario que esté directamente adyacente a la misma. Por ejemplo, un potenciador es una secuencia reguladora de la transcripción que está operativamente unida a una secuencia codificante, aunque no sean contiguas.

35 Con respecto a los polipéptidos de fusión, el término "unido operativamente" puede referirse al hecho de que cada uno de los componentes realiza la misma función en unión con el otro componente tal como lo haría si no estuvieran unidos de esta forma. Por ejemplo, con respecto a un polipéptido de fusión en el que un dominio de unión a ADN de ZFP, TALE o Cas está fusionado con un dominio de activación, el dominio de unión a ADN de ZFP, TALE o Cas y el dominio de activación se encuentran unidos operativamente si, en el polipéptido de fusión, la parte del dominio de unión a ADN de ZFP, TALE o Cas es capaz de unirse a su sitio diana y/o su sitio de unión, mientras el dominio de activación es capaz de regular al alza la expresión génica. Cuando un polipéptido de fusión en el que un dominio de unión a ADN de ZFP o TALE está fusionado con un dominio de escisión, el dominio de unión a ADN de ZFP o TALE y el dominio de escisión se encuentran unidos operativamente si, en el polipéptido de fusión, la parte de dominio de unión a ADN de ZFP o TALE es capaz de unirse a su sitio diana y/o su sitio de unión, mientras el dominio de escisión es capaz de escindir ADN presente en las proximidades del sitio diana.

45 Un "fragmento funcional" de una proteína, un polipéptido o un ácido nucleico es una proteína, un polipéptido o un ácido nucleico cuya secuencia no es idéntica a la proteína, el polipéptido o el ácido nucleico de longitud completa, pero que aún conserva la misma función que la proteína, el polipéptido o el ácido nucleico de longitud completa. Un fragmento funcional puede poseer más, menos o el mismo número de residuos que la molécula nativa correspondiente, y/o puede contener una o más sustituciones de aminoácidos o nucleótidos. Los procedimientos para determinar la función de un ácido nucleico (por ejemplo, función de codificación, capacidad para hibridarse con otro ácido nucleico) son bien conocidos en la técnica. Del mismo modo, los procedimientos para determinar la función de las proteínas son bien conocidos. Por ejemplo, la función de unión a ADN de un polipéptido puede determinarse, por ejemplo, mediante ensayos de unión a filtros, desplazamiento de la movilidad electroforética o inmunoprecipitación. La escisión del ADN se puede analizar mediante electroforesis en gel. Véase Ausubel et al., anteriormente. La capacidad de una proteína para interactuar con otra proteína puede determinarse, por ejemplo, mediante coinmunoprecipitación, ensayos de dos híbridos o complementación, tanto genética como bioquímica. Veáanse, por ejemplo, Fields et al. (1989) *Nature* 340: 245-246, la patente de Estados Unidos N° 5.585.245 y el documento PCT WO 98/44350.

60 Un "vector" es capaz de transferir secuencias génicas a células diana. Típicamente, "constructo de vector", "vector de expresión" y "vector de transferencia génica" significan cualquier constructo 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 de expresión, así como vectores de integración.

Un "gen informador" o "secuencia informadora" se refiere a cualquier secuencia que produce un producto proteico que se mide fácilmente, preferentemente, aunque no necesariamente, en un ensayo rutinario. Los genes informadores adecuados incluyen, pero sin limitación, secuencias que codifican proteínas que median la resistencia a los antibióticos (por ejemplo, resistencia a ampicilina, resistencia a neomicina, resistencia a G418, resistencia a puromicina), secuencias que codifican proteínas coloreadas, fluorescentes o luminiscentes (por ejemplo, proteína fluorescente verde, proteína fluorescente verde potenciada, proteína fluorescente roja, luciferasa) y proteínas que median el crecimiento celular potenciado y/o la amplificación génica (por ejemplo, dihidrofolato reductasa). Las etiquetas de epítipo incluyen, por ejemplo, una o más copias de FLAG, His, myc, Tap, HA o cualquier secuencia de aminoácidos detectable. Las "etiquetas de expresión" incluyen secuencias que codifican informadores que pueden estar operativamente unidos a una secuencia génica deseada con el fin de controlar la expresión del gen de interés.

Los términos "sujeto" y "paciente" se utilizan indistintamente y se refieren a mamíferos tales como pacientes humanos y primates no humanos, así como a animales experimentales tales como conejos, perros, gatos, ratas, ratones y otros animales. Por consiguiente, el término "sujeto" o "paciente", tal como se utiliza en el presente documento, significa cualquier paciente o sujeto mamífero al que se puedan administrar los glóbulos rojos (o células madre) alterados de la invención. Los sujetos de la presente invención incluyen aquellos que se han expuesto a una o más toxinas químicas, incluida, por ejemplo, una toxina nerviosa.

Nucleasas

En el presente documento se describen composiciones, particularmente nucleasas, que son útiles para dirigirlas a un gen para su uso con hemoglobinopatías. En determinadas formas de realización, la nucleasa es de origen natural. En otras formas de realización, la nucleasa es de origen no natural, es decir, está genomanipulada en el dominio de unión a ADN y/o el dominio de escisión. Por ejemplo, el dominio de unión a ADN de una nucleasa de origen natural puede alterarse para que se una a un sitio diana seleccionado (por ejemplo, una meganucleasa que se ha sido genomanipulado para que se una a un sitio diferente a su sitio de unión asociado). En otras formas de realización, la nucleasa comprende dominios de unión a ADN y de escisión heterólogos (por ejemplo, nucleasas de dedo de zinc; nucleasas efectoras TAL; dominios de unión a ADN de meganucleasa con dominios de escisión heterólogos), o una nucleasa genérica guiada por un ARN guía específico (por ejemplo, un CRISPR/Cas).

A. dominios de unión a ADN

En determinados aspectos de la divulgación, la nucleasa es una meganucleasa (endonucleasa de asentamiento). Las meganucleasas de origen natural reconocen los sitios de escisión de 15-40 pares de bases y se agrupan comúnmente en cuatro familias: la familia LAGLIDADG, la familia GIY-YIG, la familia de cajas His-Cyst y la familia HNH. Los ejemplos de endonucleasas de asentamiento incluyen I-Scel, I-Ceul, PI-Pspl, PI-Sce, I-ScelV, I-Csml, I-Panl, I-Scell, I-Ppol, I-Scelll, I-Crel, I-Tevl, I-TeVl y I-TeVll. Sus secuencias de reconocimiento son conocidas. Véanse, también, patente de Estados Unidos Nº 5.420.032; patente de Estados Unidos Nº 6.833.252; Belfort et al. (1997) *Nucleic Acids Res.* 25:3379-3388; Dujon et al. (1989) *Gene* 82:115-118; Perler et al. (1994) *Nucleic Acids Res.* 22, 1125-1127; Jasin (1996) *Trends Genet.* 12:224-228; Gimble et al. (1996) *J. Mol. Biol.* 263:163-180; Argast et al. (1998) *J. Mol. Biol.* 280:345-353 y el catálogo de New England Biolabs.

En determinados aspectos de la divulgación, la nucleasa comprende una endonucleasa de asentamiento (meganucleasa) genomanipulada (de origen no natural). Las secuencias de reconocimiento de endonucleasas de asentamiento y meganucleasas tales como I-Scel, I-Ceul, PI-Pspl, PI-Sce, I-ScelV, I-Csml, I-Panl, I-Seell, I-Ppol, I-Scelll, I-Crel, I-Tevl, I-TeVl y I-TeVll son conocidas. Véanse, también, patente de Estados Unidos Nº 5.420.032; patente de Estados Unidos Nº 6.833.252; Belfort et al. (1997) *Nucleic Acids Res.* 25:3379-3388; Dujon et al. (1989) *Gene* 82:115-118; Perler et al. (1994) *Nucleic Acids Res.* 22, 1125-1127; Jasin (1996) *Trends Genet.* 12:224-228; Gimble et al. (1996) *J. Mol. Biol.* 263:163-180; Argast et al. (1998) *J. Mol. Biol.* 280:345-353 y el catálogo de New England Biolabs. Además, la especificidad de unión a ADN de endonucleasas de asentamiento y meganucleasas puede manipularse genéticamente para que se unan a sitios diana no naturales. Véanse, por ejemplo, Chevalier et al. (2002) *Molec. Cell* 10:895-905; Epinat et al. (2003) *Nucleic Acids Res.* 31:2952-2962; Ashworth et al. (2006) *Nature* 441:656-659; Paques et al. (2007) *Current Gene Therapy* 7:49-66; publicación de patente de Estados Unidos Nº 20070117128. Los dominios de unión a ADN de endonucleasas de asentamiento y meganucleasas pueden alterarse en el contexto de la nucleasa en su conjunto (es decir, de forma que la nucleasa incluya el dominio de escisión asociado) o pueden fusionarse a un dominio de escisión heterólogo.

En otras formas de realización, el dominio de unión a ADN comprende un dominio de unión a ADN efector TAL de origen natural o genomanipulado (de origen no natural). Véase, por ejemplo, la publicación de patente de Estados

Unidos Nº 20110301073. Se sabe que las bacterias fitopatógenas del género *Xanthomonas* provocan muchas enfermedades en plantas de cultivo importantes. La patogenicidad de *Xanthomonas* depende de un sistema de secreción conservado de tipo III (T3S) que inyecta más de 25 proteínas efectoras diferentes en la célula vegetal. Entre estas proteínas inyectadas se encuentran efectores similares a activadores de la transcripción (TALE) que imitan los activadores transcripcionales de la planta y manipulan el transcriptoma de la planta (véase, Kay et al. (2007) *Science* 318: 648-651) Estas proteínas contienen un dominio de unión a ADN y un dominio de activación de la transcripción. Uno de los TALE mejor caracterizados es AvrBs3 de *Xanthomonas campestris* pv. *Vesicatoria* (véanse Bonas et al. (1989) *Mol Gen Genet* 218: 127-136 y el documento WO2010079430). Los TALE contienen un dominio centralizado de repeticiones en tándem, conteniendo cada repetición aproximadamente 34 aminoácidos, que son clave para la especificidad de unión a ADN de estas proteínas. Además, contienen una secuencia de localización nuclear y un dominio de activación transcripcional ácido (para una revisión, véase Schornack S, et al. (2006) *J Plant Physiol* 163 (3): 256-272). Además, en las bacterias fitopatógenas *Ralstonia solanacearum* se han encontrado dos genes, designados brg11 y hpx17, que son homólogos a la familia AvrBs3 de *Xanthomonas* en la cepa de *R. solanacearum* biovar 1 GMI1000 y en la cepa de biovar 4 RS1000 (véase Heuer et al. (2007) *Appl y Envir Micro* 73 (13): 4379-4384). Estos genes son el 98,9% idénticos en la secuencia de nucleótidos entre sí, pero difieren en una delección de 1.575 pb en el dominio de repetición de hpx17. Sin embargo, ambos productos génicos tienen menos del 40% de identidad de secuencia con las proteínas de la familia AvrBs3 de *Xanthomonas*.

Por lo tanto, en algunas formas de realización, el dominio de unión a ADN que se une a un sitio diana en un locus diana (por ejemplo, globina o puerto seguro) es un dominio genomanipulado de un efector TAL similar a los derivados de los patógenos de plantas *Xanthomonas* (véanse Boch et al., (2009) *Science* 326: 1509-1512 y Moscou y Bogdanove, (2009) *Science* 326: 1501) y *Ralstonia* (véanse Heuer et al. (2007) *Applied and Environmental Microbiology* 73(13): 4379-4384); patentes de Estados Unidos Nº 8.420.782 y 8.440.431 y la publicación de patente de Estados Unidos Nº 20110301073.

En determinadas formas de realización, el dominio de unión a ADN comprende una proteína de dedo de zinc (por ejemplo, una proteína de dedo de zinc que se une a un sitio diana en un gen de globina o puerto seguro). Preferentemente, la proteína de dedo de zinc no es de origen natural, ya que está genomanipulada para que se una al sitio diana de elección. Véanse, 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; patentes de Estados Unidos Nº 6.453.242, 6.534.261, 6.599.692, 6.503.717, 6.689.558, 7.030.215, 6.794.136, 7.067.317, 7.262.054, 7.070.934, 7.361.635, 7.253.273; y las publicaciones de patente de Estados Unidos Nº 2005/0064474, 2007/0218528, 2005/0267061.

Un dominio de unión de dedo de zinc o TALE genomanipulado puede tener una especificidad de unión novedosa, en comparación con una proteína de dedo de zinc natural. Los procedimientos de genomanipulación incluyen, pero sin limitación, 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 tripletes (o cuadrupletes) y secuencias de aminoácidos de dedo 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. Véanse, por ejemplo, las patentes de Estados Unidos en copropiedad 6.453.242 y 6.534.261.

Los ejemplos de procedimientos de selección, que incluyen los sistemas de presentación en fagos y de dos híbridos, se describen en las patentes de Estados Unidos Nº 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 en los documentos WO 98/37186, WO 98/53057, WO 00/27878, WO 01/88197 y GB 2.338.237. Además, la potenciación de la especificidad de unión para dominios de unión de dedo de zinc se ha descrito, por ejemplo, en el documento en copropiedad WO 02/077227.

Además, tal como se divulga en estas y otras referencias, los dominios de ADN (por ejemplo, los dominios de proteínas de dedo de zinc con múltiples dedos o de TALE) pueden unirse entre sí utilizando cualquier secuencia enlazadora adecuada, incluyendo, por ejemplo, enlazadores de 5 o más aminoácidos de longitud. Véanse también las patentes de Estados Unidos Nº 6.479.626, 6.903.185 y 7.153.949 para secuencias enlazadoras ejemplares de 6 o más aminoácidos de longitud. Las proteínas de unión a ADN descritas en el presente documento pueden incluir cualquier combinación de enlazadores adecuados entre los dedos de zinc individuales de la proteína. Además, la mejora de la especificidad de unión para dominios de unión de dedo de zinc se ha descrito, por ejemplo, en el documento en copropiedad WO 02/077227.

La selección de sitios diana, los dominios de unión a ADN y los procedimientos para el diseño y la construcción de proteínas de fusión (y polinucleótidos que los codifican) son conocidos por los expertos en la técnica y se describen en detalle en las patentes de Estados Unidos Nº 6.140.0815, 789.538, 6.453.242, 6.534.261, 5.925.523, 6.007.988, 6.013.453, 6.200.759, los documentos 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 y la publicación de Estados Unidos N° 20110301073.

Además, tal como se divulga en estas y otras referencias, los dominios de unión a ADN (por ejemplo, proteínas de dedo de zinc con múltiples dedos) pueden unirse entre sí utilizando cualquier secuencia enlazadora adecuada, incluyendo, por ejemplo, enlazadores de 5 o más aminoácidos de longitud. Véanse también, las patentes de Estados Unidos N° 6.479.626, 6.903.185 y 7.153.949 para secuencias enlazadoras ejemplares de 6 o más aminoácidos de longitud. Las proteínas descritas en el presente documento pueden incluir cualquier combinación de enlazadores adecuados entre los dedos de zinc individuales de la proteína.

B. Dominios de escisión

Cualquier dominio de escisión adecuado puede unirse operativamente a un dominio de unión a ADN para formar una nucleasa. Por ejemplo, los dominios de unión a ADN de ZFP se han fusionado con dominios de nucleasa para crear ZFN, una entidad funcional que es capaz de reconocer su ácido nucleico diana previsto a través de su dominio de unión de ADN genomanipulado (ZFP) y hacer que el ADN se corte cerca del sitio de unión de ZFP por medio de la actividad de nucleasa. Véase, por ejemplo, Kim et al. (1996) *Proc Nat'l Acad Sci USA* 93 (3): 1156-1160. Más recientemente, las ZFN se han utilizado para la modificación genómica en una diversidad de organismos. Véanse, por ejemplo, las publicaciones de patente de Estados Unidos 20030232410, 20050208489, 20050026157, 20050064474, 20060188987, 20060063231, y la publicación internacional WO 07/014275. Del mismo modo, los dominios de unión a ADN de TALE se han fusionado con dominios de nucleasa para crear TALEN. Véase, por ejemplo, la publicación de Estados Unidos N° 20110301073.

Tal como se ha indicado anteriormente, el dominio de escisión puede ser heterólogo con respecto al dominio de unión a ADN, por ejemplo, un dominio de unión a ADN de dedo de zinc y un dominio de escisión de una nucleasa o un dominio de unión a ADN de TALEN y un dominio de escisión, o dominio de unión a ADN de meganucleasa y dominio de escisión de una nucleasa diferente. Los dominios de escisión heterólogos se pueden obtener a partir de cualquier endonucleasa o exonucleasa. Los ejemplos de endonucleasas a partir de las cuales se puede derivar un dominio de escisión incluyen, pero sin limitación, endonucleasas de restricción y endonucleasas de asentamiento. Véanse, por ejemplo, Catálogo 2002-2003 de New England Biolabs, Beverly, MA; y Belfort et al. (1997) *Nucleic Acids Res.* 25: 3379-3388. Se conocen enzimas adicionales que escinden el ADN (por ejemplo, nucleasa S1, nucleasa de frijol mungo, DNasa I pancreática, nucleasa micrococcal, endonucleasa HO de levadura; véase también Linn et al. (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993). Una o más de estas enzimas (o fragmentos funcionales de las mismas) pueden utilizarse como fuente de dominios de escisión y semidominios de escisión.

De forma similar, un semidominio de escisión puede derivarse de cualquier nucleasa o parte de la misma, tal como se ha establecido anteriormente, lo que requiere la dimerización para la actividad de escisión. En general, se requieren dos proteínas de fusión para la escisión si las proteínas de fusión comprenden semidominios de escisión. Alternativamente, se puede utilizar una única proteína que comprende dos semidominios de escisión. Los dos semidominios de escisión pueden derivarse de la misma endonucleasa (o fragmentos funcionales de la misma), o cada semidominio de escisión puede derivarse de una endonucleasa diferente (o fragmentos funcionales de la misma). Además, los sitios diana para las dos proteínas de fusión se disponen preferentemente, uno con respecto al otro, de forma que la unión de las dos proteínas de fusión a sus respectivos sitios diana dispone los semidominios de escisión en una orientación espacial, uno con respecto a otro, que permite que los semidominios de escisión formen un dominio de escisión funcional, por ejemplo, mediante dimerización. Por lo tanto, en determinadas formas de realización, los bordes cercanos de los sitios diana están separados por 5-8 nucleótidos o por 15-18 nucleótidos. No obstante, puede interponerse cualquier número entero de nucleótidos o pares de nucleótidos entre dos sitios diana (por ejemplo, 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 unirse al ADN de una forma específica de secuencia (en un sitio de reconocimiento), y escindir el ADN en o cerca del sitio de unión. Determinadas enzimas de restricción (por ejemplo, tipo IIS) escinden el ADN en sitios eliminados del sitio de reconocimiento y tienen dominios de unión y de escisión separables. Por ejemplo, la enzima de tipo IIS *Fok I* cataliza la escisión bicatenaria de ADN, a 9 nucleótidos desde su sitio de reconocimiento en una cadena y a 13 nucleótidos desde su sitio de reconocimiento en la otra. Véanse, por ejemplo, las patentes de Estados Unidos N° 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, en una forma de realización, las proteínas de fusión comprenden el dominio de escisión (o semidominio de escisión) de al menos una enzima de restricción de tipo IIS y uno o más dominios de unión de dedo de zinc, que pueden o no estar genomanipulados.

- Un ejemplo de enzima de restricción de tipo IIS, cuyo dominio de escisión es separable del dominio de unión, es *Fok I*. Esta enzima particular es activa como un dímero. Bitinaite et al. (1998) *Proc. Natl. Acad. Sci. USA* 95: 10.570-10.575. En consecuencia, para los fines de la presente divulgación, la parte de la enzima *Fok I* utilizada en las proteínas de fusión descritas se considera un semidominio de escisión. Por lo tanto, para la escisión bicatenaria dirigida y/o el reemplazo dirigido de secuencias celulares utilizando fusiones de dedo de zinc-*Fok I*, pueden utilizarse dos proteínas de fusión, cada una con un semidominio de escisión de *Fok I*, para reconstituir un dominio de escisión catalíticamente activo. Alternativamente, también puede utilizarse una única molécula de polipéptido que contiene un dominio de unión a ADN y dos semidominios de escisión de *Fok I*.
- Un dominio de escisión o un semidominio de escisión puede ser cualquier parte de una proteína que conserva la actividad de escisión, o que conserva la capacidad de multimerizarse (por ejemplo, dimerizarse) para formar un dominio de escisión funcional.
- Ejemplos de enzimas de restricción de tipo IIS se describen en la publicación internacional WO 07/014275. Las enzimas de restricción adicionales también contienen dominios de unión y de escisión separables, y estos se contemplan por la presente divulgación. Véase, por ejemplo, Roberts et al. (2003) *Nucleic Acids Res.* 31: 418-420.
- En determinadas formas de realización, el dominio de escisión comprende uno o más semidominios de escisión genomanipulados (también denominados mutantes del dominio de dimerización) que minimizan o evitan la homodimerización, tal como se describe, por ejemplo, en las publicaciones de patente de Estados Unidos N° 20050064474, 20060188987 y 20080131962. Los residuos de aminoácidos en las posiciones 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537 y 538 de *Fok I* son todos dianas para influir en la dimerización de los semidominios de escisión de *Fok I*.
- Los ejemplos de semidominios de escisión de *Fok I* genomanipulados que forman heterodímeros obligados incluyen un par en el que un primer semidominio de escisión incluye mutaciones en los residuos de aminoácidos en las posiciones 490 y 538 de *Fok I* y un segundo semidominio de escisión incluye mutaciones en los residuos de aminoácidos 486 y 499.
- Por lo tanto, en una forma de realización, una mutación en 490 reemplaza Glu (E) por Lys (K); la mutación en 538 reemplaza Iso (I) por Lys (K); la mutación en 486 reemplaza Gln (Q) por Glu (E); y la mutación en la posición 499 reemplaza Iso (I) por Lys (K). Específicamente, los semidominios de escisión genomanipulados descritos en el presente documento se prepararon mutando las posiciones 490 (E → K) y 538 (I → K) en un semidominio de escisión para producir un semidominio de escisión genomanipulado designado "E490K:I538K" y mutando las posiciones 486 (Q → E) y 499 (I → L) en otro semidominio de escisión para producir un semidominio de escisión genomanipulado designado "Q486E:I499L". Los semidominios de escisión genomanipulados descritos en el presente documento son mutantes heterodímeros obligados en los que la escisión aberrante se minimiza o se suprime. Véase, por ejemplo, la publicación de patente de Estados Unidos N° 2008/0131962.
- En determinadas formas de realización, el semidominio de escisión genomanipulado comprende mutaciones en las posiciones 486, 499 y 496 (numeración con respecto a *Fok I* de tipo silvestre), por ejemplo, mutaciones que reemplazan el residuo de tipo silvestre Gln (Q) en la posición 486 por un residuo Glu (E), el residuo de tipo silvestre Iso (I) en la posición 499 por un residuo Leu (L) y el residuo de tipo silvestre Asn (N) en la posición 496 por un residuo Asp (D) o Glu (E) (también denominados dominios "ELD" y "ELE", respectivamente). En otras formas de realización, el semidominio de escisión genomanipulado comprende mutaciones en las posiciones 490, 538 y 537 (numeración con respecto a *Fok I* de tipo silvestre), por ejemplo, mutaciones que reemplazan el residuo de tipo silvestre Glu (E) en la posición 490 por un residuo Lys (K), el residuo de tipo silvestre Iso (I) en la posición 538 por un residuo Lys (K), y el residuo de tipo silvestre His (H) en la posición 537 por un residuo Lys (K) o un residuo Arg (R) (también denominados dominios "KKK" y "KKR", respectivamente). En otras formas de realización, el semidominio de escisión genomanipulado comprende mutaciones en las posiciones 490 y 537 (numeración con respecto a *Fok I* de tipo silvestre), por ejemplo mutaciones que reemplazan el residuo de tipo silvestre Glu (E) en la posición 490 por un residuo Lys (K) y el residuo de tipo silvestre His (H) en la posición 537 por un residuo Lys (K) o un residuo Arg (R) (también denominados dominios "KIK" y "KIR", respectivamente). (Véase la publicación de patente de Estados Unidos N° 20110201055). Los semidominios de escisión genomanipulados descritos en el presente documento pueden prepararse utilizando cualquier procedimiento adecuado, por ejemplo, mediante mutagénesis dirigida a sitio de semidominios de escisión de tipo silvestre (*Fok I*) tal como se describe en las publicaciones de patente de Estados Unidos N° 20050064474, 20080131962 y 20110201055.
- Alternativamente, las nucleasas pueden ensamblarse *in vivo* en el sitio diana del ácido nucleico utilizando la tecnología denominada "enzima fragmentada" ("*split-enzyme*") (véase, por ejemplo, la publicación de patente de Estados Unidos N° 20090068164). Los componentes de dichas enzimas fragmentadas se pueden expresar en constructos de expresión separados, o se pueden unir en un marco de lectura abierto en el que los componentes individuales están

separados, por ejemplo, por un péptido 2A autoescindible o una secuencia IRES. Los componentes pueden ser dominios de unión de dedo de zinc individuales o dominios de un dominio de unión de ácido nucleico de meganucleasa.

Las nucleasas pueden cribarse para detectar su actividad antes de su uso, por ejemplo, en un sistema cromosómico basado en levadura tal como se describe en los documentos WO 2009/042163 y 20090068164. Los constructos de expresión de nucleasas pueden diseñarse fácilmente utilizando procedimientos conocidos en la técnica. Véanse, por ejemplo, las publicaciones de patente de Estados Unidos 20030232410, 20050208489, 20050026157, 20050064474, 20060188987, 20060063231, y la publicación internacional WO 07/014275. La expresión de la nucleasa puede encontrarse bajo el control de un promotor constitutivo o un promotor inducible, por ejemplo, el promotor de galactocinasa que se activa (desreprime) en presencia de rafinosa y/o galactosa y se reprime en presencia de glucosa.

El sistema CRISPR/Cas

Recientemente ha surgido una evidencia convincente de la existencia de una ruta de defensa del genoma mediada por ARN en arqueas y muchas bacterias sobre la que se ha planteado la hipótesis de que es paralela a la ruta eucariota de ARNi (para revisiones, véanse Godde y Bickerton, 2006. *J. Mol. Evol.* 62: 718-729; Lillestol et al., 2006. *Archaea* 2: 59-72; Makarova et al., 2006. *Biol. Direct* 1: 7.; Sorek et al., 2008. *Nat. Rev. Microbiol.* 6: 181-186). Conocido como el sistema CRISPR-Cas o ARNi procariota (pARNi), se propone que la ruta surja de dos loci génicos vinculados evolutivamente y a menudo físicamente: el locus CRISPR (repeticiones palindrómicas cortas agrupadas regularmente interespaciadas), que codifica los componentes de ARN del sistema, y el locus cas (asociado a CRISPR), que codifica proteínas (Jansen et al., 2002. *Mol. Microbiol* 43: 1565-1575; Makarova et al., 2002. *Nucleic Acids Res.* 30: 482-496; Makarova et al., 2006. *Biol. Direct* 1: 7; Haft et al., 2005. *PLoS Comput. Biol.* 1: e60). Los loci CRISPR en huéspedes microbianos contienen una combinación de genes asociados a CRISPR (Cas), así como elementos de ARN no codificantes capaces de programar la especificidad de la escisión de ácido nucleico mediada por CRISPR. Las proteínas Cas individuales no comparten una similitud de secuencia significativa con los componentes proteicos de la maquinaria de ARNi eucariota, pero tienen funciones predichas análogas (por ejemplo, ARN de unión, nucleasa, helicasa, etc.) (Makarova et al., 2006. *Biol. Direct* 1: 7). Los genes asociados a CRISPR (cas) a menudo se asocian con matrices de repetición de CRISPR-espaciadores. Se han descrito más de cuarenta familias diferentes de proteínas Cas. De estas familias de proteínas, Cas1 parece ser omnipresente entre los diferentes sistemas CRISPR/Cas. Se han utilizado combinaciones particulares de genes cas y estructuras de repetición para definir 8 subtipos CRISPR (Ecoli, Ypest, Nmeni, Dvulg, Tneap, Hmari, Aperi y Mtube), algunos de los cuales están asociados con un módulo génico adicional que codifica proteínas misteriosas asociadas a la repetición (RAMP). Puede estar presente más de un subtipo de CRISPR en un solo genoma. La distribución esporádica de los subtipos de CRISPR/Cas sugiere que el sistema está sujeto a la transferencia horizontal de genes durante la evolución microbiana.

El CRISPR de tipo II (ejemplificado por Cas9) es uno de los sistemas mejor caracterizados y lleva a cabo una rotura bicatenaria de ADN dirigida en cuatro etapas secuenciales. En primer lugar, dos ARN no codificantes, la matriz pre-ARNcr y el ARNtracr, se transcriben a partir del locus CRISPR. En segundo lugar, el ARNtracr se hibrida con las regiones de repetición del pre-ARNcr y media el procesamiento del pre-ARNcr en ARNcr maduros que contienen secuencias espaciadoras individuales. En tercer lugar, el complejo ARNcr:ARNtracr maduro dirige Cas9 al ADN diana a través del emparejamiento de bases de Watson-Crick entre el espaciador en el ARNcr y el protoespaciador en el ADN diana junto al motivo adyacente protoespaciador (PAM), un requisito adicional para el reconocimiento de la diana. Finalmente, Cas9 media la escisión del ADN diana para crear una rotura bicatenaria dentro del protoespaciador. La actividad del sistema CRISPR/Cas comprende tres etapas: (i) inserción de secuencias de ADN extrañas en la matriz CRISPR para prevenir futuros ataques, en un proceso denominado "adaptación", (ii) expresión de las proteínas relevantes, así como expresión y procesamiento de la matriz, seguidos de (iii) interferencia mediada por ARN con el ácido nucleico extraño. Por lo tanto, en la célula bacteriana, varias de las denominadas proteínas 'Cas' están involucradas con la función natural del sistema CRISPR/Cas.

Los productos principales de los loci CRISPR parecen ser ARN cortos que contienen las secuencias de direccionamiento del invasor, y se denominan ARN guía o ARN silenciadores procariotas (pARNsi) en función de su papel hipotético en la ruta (Makarova et al. (2006) *Biol. Direct* 1:7; Hale et al. (2008) *RNA* 14: 2572-2579). El análisis de ARN indica que los transcritos de locus CRISPR se escinden dentro de las secuencias de repetición para liberar intermedios de ARN de ~60 a 70 nt que contienen secuencias de direccionamiento invasoras individuales y fragmentos de repetición flanqueantes (Tang et al. (2002) *Proc. Natl. Acad. Sci.* 99: 7536-7541; Tang et al. (2005) *Mol. Microbiol* 55: 469-481; Lillestol et al. (2006) *Archaea* 2: 59-72; Brouns et al. (2008) *Science* 321: 960-964; Hale et al. (2008) *ARN* 14: 2572-2579) En el arqueón *Pyrococcus furiosus*, estos ARN intermedios se procesan adicionalmente para dar pARNsi maduros de ~35 a 45 nt estables abundantes (Hale et al. (2008) *RNA* 14: 2572-2579).

Proteínas Cas

El polipéptido "Cas1" se refiere a la proteína (Cas) asociada a CRISPR1. Cas1 (COG1518 en las agrupaciones de grupos ortólogos del sistema de clasificación de proteínas) es el mejor marcador de los sistemas asociados a CRISPR

(CASS). Según las comparaciones filogenéticas, se han identificado siete versiones distintas del sistema inmunitario asociado a CRISPR (CASS1-7).

5 El polipéptido Cas1 utilizado en los procedimientos descritos en el presente documento puede ser cualquier polipéptido Cas1 presente en cualquier procarionta. En determinadas formas de realización, un polipéptido Cas1 es un polipéptido Cas1 de un microorganismo arquea. En determinadas formas de realización, un polipéptido Cas1 es un polipéptido Cas1 de un microorganismo Euryarchaeota. En determinadas formas de realización, un polipéptido Cas1 es un polipéptido Cas1 de una bacteria. En determinadas formas de realización, un polipéptido Cas1 es un polipéptido Cas1 de una bacteria gram negativa o gram positiva. En determinadas formas de realización, un polipéptido Cas1 es un polipéptido Cas1 de *Pseudomonas aeruginosa*. En determinadas formas de realización, un polipéptido Cas1 es un polipéptido Cas1 de *Aquifexaeolicus*. En determinadas formas de realización, un polipéptido Cas1 es un polipéptido Cas1 que es miembro de uno de CASS1-7. En determinadas formas de realización, el polipéptido Cas1 es un polipéptido Cas1 que es miembro de CASS3. En determinadas formas de realización, un polipéptido Cas1 es un polipéptido Cas1 que es miembro de CASS7. En determinadas formas de realización, un polipéptido Cas1 es un polipéptido Cas1 que es miembro de CASS3 o CASS7.

20 En algunas formas de realización, un polipéptido Cas1 está codificado por una secuencia de nucleótidos proporcionada en GenBank, por ejemplo, con el número GenelD: 2781520, 1006874, 9001811, 947228, 3169280, 2650014, 1175302, 3993120, 4380485, 906625, 3165126, 905808, 1454460, 1445886, 1485099, 4274010, 888506, 3169526, 997745, 897836 o 1193018 y/o una secuencia de aminoácidos que muestra homología (por ejemplo, más del 80%, del 90 al 99%, incluidos el 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99%) con los aminoácidos codificados por estos polinucleótidos y cuyos polipéptidos funcionan como polipéptidos Cas1.

25 Cas6 es otro polipéptido Cas, y la actividad de endorribonucleasa se denomina en el presente documento actividad de endorribonucleasa de Cas6. Ejemplos no limitantes de polipéptidos Cas6 adecuados se representan en N° de acceso de GenBank AAL81255. Un polipéptido Cas6 puede enriquecerse, aislarse o purificarse a partir de un microbio que tiene un locus CRISPR y el locus cas (asociado a CRISPR), tal como, pero sin limitación, *Pyrococcus furiosus*, o puede producirse utilizando técnicas recombinantes, o sintetizarse químicamente o enzimáticamente utilizando procedimientos rutinarios. En algunos aspectos, un polipéptido Cas6 puede enriquecerse, aislarse o purificarse a partir de un microbio que no tiene loci CRISPR. Un polipéptido Cas6 contiene al menos un residuo que puede desempeñar un papel en la catálisis o la sustitución conservadora del mismo. Un polipéptido Cas6 puede contener otros residuos que también pueden desempeñar un papel en la catálisis o la sustitución conservadora del mismo. El o los residuos que se espera que desempeñen un papel en la catálisis pueden ubicarse cerca del bucle rico en G que contiene el motivo distintivo de Cas6 en la estructura 3D de la proteína. Los polipéptidos Cas6 pueden incluir dominios presentes en la base de datos TIGRFAM con los números de acceso TIGR01877 y PF01881. La base de datos TIGRFAM incluye familias de polipéptidos para los que se conserva la función (Haft et al. (2003) *Nucl. Acidos Res.* 31: 371-373, Bateman y Haft (2002) *Briefings Bioinformatics*, 3: 236-245y Haft et al. (2005) *PLoS Computational Biol.* 1(6):e60).

40 Otros ejemplos de polipéptidos Cas6 divulgados en el presente documento incluyen los presentes en microbios procariontas que tienen un locus CRISPR y un locus cas. Los polipéptidos Cas6 se pueden identificar fácilmente en cualquier microbio que incluya un locus CRISPR. Una región codificante que codifica un polipéptido Cas6 se encuentra típicamente en un locus cas ubicada muy cerca de un locus CRISPR. Haft et al. (2005) *PLoS Computational Biol.* 1 (6): e60) revisan la familia de proteínas Cas y crean reglas para la identificación de subtipos específicos del sistema CRISPR/Cas. Haft et al. describen la región codificante que codifica los polipéptidos Cas6 como que se encuentra asociada a al menos cuatro subtipos de CRISPR/Cas separados (Tneap, Hmari, Aperi y Mtube) y como que es la región codificante cas más distal al locus CRISPR. Los polipéptidos Cas6 pueden identificarse utilizando los recursos disponibles en JCVI Comprehensive Microbial Resource. Por lo tanto, los polipéptidos Cas6 que son útiles en los procedimientos descritos en el presente documento pueden identificarse por parte del experto utilizando procedimientos rutinarios.

55 Los ejemplos de microbios procariontas con secuencias genómicas completas conocidas que contienen regiones codificantes que se espera que codifiquen un polipéptido Cas6 incluyen *Thermotogamaritima* MSB8, *Campylobacter fetus* subsp. fetus 82-40, *Fusobacteriumnucleatum* ATCC 25586, *Streptococcus thermophilus* LMG 18311, *Thermoanaerobactertengcongensis* MB4(T), *Moorellathermoacetica* ATCC 39073, *Desulfitobacteriumhafniense* Y51, *Clostridium tetani* E88, *Clostridium perfringens* SM101, *Clostridium difficile* QCD-32g58, *Clostridium botulinum* Hall A Sanger, *Clostridium botulinum* F Langeland, *Clostridium botulinum* B1 cepa Okra, *Clostridium botulinum* A3 cepa Loch Maree, *Clostridium botulinum* A Hall, *Clostridium botulinum* A ATCC 19397, *Carboxydothemushydrogenoformans* Z-2901, *Staphylococcus epidermidis* RP62A, *Thermusthermophilus* HB8, *Thermusthermophilus* HB27, *Nostoc* sp. PCC 7120, *Anabaena variabilis* ATCC 29413, *Synechococcus* sp. OS tipo B prime, *Synechococcus* sp. OS tipo A, *Porphyromonasgingivalis* W83, *Bacteroidesfragilis* YCR46, *Bacteroidesfragilis* NCTC9343, *Aquifexaeolicus* VF5, *Rubrobacterxylophilus* DSM 9941, *Mycobacterium tuberculosis* H37Rv (cepa de laboratorio), *Mycobacterium*

tuberculosis CDC1551, *Mycobacterium bovis* subsp. *bovis* AF2122/97, *Frankiaalni* ACN14a, *Thermoplasma volcanium* GSS1, *Picrophilustorridus* DSM 9790, *Thermococcus kodakarensis* KOD1, *Pyrococcus horikoshii* OT3, *Pyrococcus furiosus* DSM 3638, *Pyrococcus abyssi* GE5, *Methanosarcinabarkerifusaro*, *Methanosarcina acetivorans* C2A, *Methanococcus burtonii* DSM 6242, *Methanococcus jannaschii* DSM2661, *Methanobacterium thermoautotrophicum* delta H, *Haloarcula marismortui* ATCC 43049, *Archaeoglobus fulgidus* DSM4304, *Pyrobaculum aerophilum* 1M2, *Sulfolobus tokodaii* cepa 7, *Sulfolobus solfataricus* P2, *Sulfolobus acidocaldarius* DSM 639, *Aeropyrum pernix* K1. Los expertos en la técnica conocen otros ejemplos de polipéptidos Cas6; véanse, por ejemplo, los miembros del grupo de polipéptidos COG1583 (disponible en la página web Clusters of Orthologous Groups of protein (COGs) a través del sitio web del National Center for Biotechnology Information, véanse además Tatusov et al. (1997) *Science* 278: 631-637 y Tatusov et al. (2003) *BMC Bioinformatics* 4 (1):41), miembros de la familia InterPro con número de acceso IPRO10156, Makarova et al. (2002) *Nuc. Acidos Res.* 30: 482-496 y Haft et al. (2005) *PLoS Comput. Biol.* 1 (6): e60, 474-483).

Hay tres tipos de sistemas CRISPR/Cas que incorporan ARN y proteínas Cas. Los tipos I y III tienen endonucleasas Cas que procesan los pre-ARNcr, que, cuando se procesan por completo dando ARNcr, ensamblan un complejo de proteínas multi-Cas que es capaz de escindir ácidos nucleicos que son complementarios al ARNcr.

En los sistemas CRISPR/Cas de tipo II, los ARNcr se producen utilizando un mecanismo diferente en el que un ARN de activación en trans (ARNtracr) complementario a secuencias de repetición en el pre-ARNcr, desencadena el procesamiento por una RNasa III específica de doble cadena en presencia de proteína Cas9. Cas9 es capaz de escindir un ADN diana que es complementario al ARNcr maduro; sin embargo, la escisión por parte de Cas 9 depende tanto del emparejamiento de bases entre el ARNcr y el ADN diana como de la presencia de un motivo corto en el ARNcr denominado la secuencia PAM (*protospacer adyacente motif* (motivo adyacente protoespaciador)) (véase Qi et al. (2013) *Cell* 152:1173). Además, el ARNtracr también debe estar presente ya que se empareja con el ARNcr en su extremo 3', y esta asociación desencadena la actividad de Cas9.

La proteína Cas9 tiene al menos dos dominios de nucleasa: un dominio de nucleasa es similar a una endonucleasa HNH, mientras que el otro se asemeja a un dominio de endonucleasa Ruv. El dominio de tipo HNH parece ser responsable de escindir la cadena de ADN que es complementaria al ARNcr mientras que el dominio Ruv escinde la cadena no complementaria.

El requisito del complejo ARNcr-ARNtracr puede evitarse mediante el uso de un "ARN guía único" (ARNsg) genomanipulado que comprende la horquilla que normalmente se forma por el alineamiento del ARNcr y el ARNtracr (véanse, Jinek et al. (2012) *Science* 337:816 y Cong et al. (2013) *Scienceexpress*/10.1126/science.1231143). En *S. pyrogenes* la fusión de ARNcr-ARNtracr genomanipulada, o el ARNsg, guía a Cas9 para escindir el ADN diana cuando se forma un heterodímero de ARN:ADN bicatenario entre los ARN asociados a Cas y el ADN diana. Este sistema que comprende la proteína Cas9 y un ARNsg genomanipulado que contiene una secuencia PAM se ha utilizado para la edición del genoma guiada por ARN (véase Ramalingam, ibidem) y ha sido útil para la edición genómica de embriones de pez cebra *in vivo* (véase Hwang et al. (2013) *Nature Biotechnology* 31 (3):227) con eficacias de edición similares a ZFN y TALEN.

En determinadas formas de realización, la proteína Cas puede ser un "derivado funcional" de una proteína Cas de origen natural. Un "derivado funcional" de un polipéptido de secuencia nativa es un compuesto que tiene una propiedad biológica cualitativa en común con un polipéptido de secuencia nativa. Los "derivados funcionales" incluyen, pero sin limitación, fragmentos de una secuencia nativa y derivados de un polipéptido de secuencia nativa y sus fragmentos, siempre que tengan una actividad biológica en común con un polipéptido de secuencia nativa correspondiente. Una actividad biológica contemplada en el presente documento es la capacidad del derivado funcional para hidrolizar un sustrato de ADN en fragmentos. El término "derivado" abarca tanto variantes de secuencia de aminoácidos del polipéptido como modificaciones covalentes y fusiones de las mismas.

El "polipéptido Cas" abarca un polipéptido Cas de longitud completa, un fragmento enzimáticamente activo de un polipéptido Cas y derivados enzimáticamente activos de un polipéptido Cas o fragmento de los mismos. Los derivados adecuados de un polipéptido Cas o un fragmento de los mismos incluyen, pero sin limitación, mutantes, fusiones, modificaciones covalentes de la proteína Cas o un fragmento de los mismos.

Las proteínas Cas y los polipéptidos Cas pueden obtenerse a partir de una célula o sintetizarse químicamente o mediante una combinación de estos dos procedimientos. La célula puede ser una célula que produce de forma natural proteína Cas, o una célula que produce de forma natural proteína Cas y está genomanipulada para producir la proteína Cas endógena a un nivel de expresión más alto o para producir una proteína Cas a partir de un ácido nucleico introducido de forma exógena, ácido nucleico que codifica un Cas que es igual o diferente al Cas endógeno. En algunos casos, la célula no produce de forma natural la proteína Cas y está genomanipulada para producir una proteína Cas.

El sistema CRISPR/Cas también se puede utilizar para inhibir la expresión génica. Lei et al. (2013) *Cell* 152 (5): 1173-1183) han demostrado que una Cas9 catalíticamente muerta que carece de actividad de endonucleasa, cuando se coexpresa con un ARN guía, genera un complejo de reconocimiento de ADN que puede interferir específicamente con el alargamiento transcripcional, la unión de la ARN polimerasa o la unión del factor de transcripción. Este sistema, llamado CRISPR interferencia (CRISPRi), puede reprimir eficazmente la expresión de genes específicos.

Además, se han desarrollado proteínas Cas que comprenden mutaciones en sus dominios de escisión para que sean incapaces de inducir un DSB y, en su lugar, introducen una mella en el ADN diana ("enzima melladora Cas9", véase Cong et al., *ibídem*).

Las proteínas Cas de la invención pueden mutarse para alterar la funcionalidad. Los ejemplos de procedimientos de selección, que incluyen presentación en fagos y sistemas de dos híbridos, se divulgan en las patentes de Estados Unidos N° 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 en los documentos WO 98/37186, WO 98/53057, WO 00/27878, WO 01/88197 y GB 2.338.237.

Componentes de ARN de CRISPR/Cas

El sistema CRISPR/Cas relacionado con Cas9 comprende dos componentes no codificantes de ARN: ARNtracr y una matriz de pre-ARNcr que contiene secuencias guía de nucleasas (espaciadores) interespaciadas por repeticiones directas (DR) idénticas. Para utilizar un sistema CRISPR/Cas para lograr la genomaniplación del genoma, ambas funciones de estos ARN deben estar presentes (véase Cong et al, (2013) *Scienceexpress* 1/10.1126/science 1231143). En algunas formas de realización, el ARNtracr y los pre-ARNcr se suministran a través de constructos de expresión separados o como ARN separados. En otras formas de realización, se construye un ARN quimérico en el que un ARNcr maduro genomaniplado (que confiere especificidad a la diana) se fusiona con un ARNtracr (que proporciona interacción con el Cas9) para crear un híbrido ARNcr-ARNtracr quimérico (también denominado ARN guía único). (Véanse Jinek, *ibídem* y Cong, *ibídem*).

Los quiméricos o ARNsg pueden genomaniplarse para que comprendan una secuencia complementaria a cualquier diana deseada. Los ARN comprenden 22 bases de complementariedad con una diana y de la forma G[n19], seguidos de un motivo adyacente protoespaciador (PAM) de la forma NGG. Por lo tanto, en un procedimiento, los ARNsg pueden diseñarse mediante la utilización de una diana de ZFN conocida en un gen de interés (i) alineando la secuencia de reconocimiento del heterodímero ZFN con la secuencia de referencia del genoma relevante (humano, de ratón o de una especie de planta particular); (ii) identificando la región espaciadora entre los semisitios ZFN; (iii) identificando la ubicación del motivo G[N20]GG que está más cerca de la región espaciadora (cuando más de un motivo se superpone al espaciador, se elige el motivo centrado con respecto al espaciador); (iv) utilizando ese motivo como el núcleo del ARNsg. Este procedimiento se basa ventajosamente en dianas de nucleasa probadas. Alternativamente, los ARNsg pueden diseñarse para que se dirijan a cualquier región de interés simplemente identificando una secuencia diana adecuada que se ajuste a la fórmula G[n20]GG.

Sitios diana

Tal como se ha descrito en detalle anteriormente, los dominios de unión a ADN pueden genomaniplarse para que se unan a cualquier secuencia de elección en un locus, por ejemplo, un gen de globina o de puerto seguro. Un dominio de unión a ADN genomaniplado puede tener una especificidad de unión novedosa, en comparación con un dominio de unión a ADN de origen natural. Los procedimientos de genomaniplación incluyen, pero sin limitación, 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 individuales (por ejemplo, de dedo de zinc), en las que cada secuencia de nucleótidos en triplete o cuadruplete está asociada con una o más secuencias de aminoácidos del dominio de unión a ADN que se une a la secuencia en triplete o cuadruplete particular. Véanse, por ejemplo, las patentes de Estados Unidos en copropiedad 6.453.242 y 6.534.261. También se puede realizar el diseño racional de los dominios efectores TAL. Véase, por ejemplo, la publicación de patente de Estados Unidos N° 20110301073.

Ejemplos de procedimientos de selección aplicables a dominios de unión a ADN, que incluyen sistemas de presentación en fagos y de dos híbridos, se describen en las patentes de Estados Unidos N° 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 los documentos WO 98/37186, WO 98/53057, WO 00/27878, WO 01/88197 y GB 2.338.237.

La selección de sitios diana, las nucleasas y los procedimientos para el diseño y la construcción de proteínas de fusión (y los polinucleótidos que codifican las mismas) son conocidos por los expertos en la técnica y se describen en detalle en las publicaciones de solicitud de patente de Estados Unidos N° 20050064474 y 20060188987.

Además, tal como se divulga en estas y otras referencias, los dominios de unión a ADN (por ejemplo, proteínas de dedos de zinc con múltiples dedos) pueden unirse entre sí utilizando cualquier secuencia enlazadora adecuada, incluyendo, por ejemplo, enlazadores de 5 o más aminoácidos. Véanse, por ejemplo, las patentes de Estados Unidos N° 6.479.626, 6.903.185 y 7.153.949 para ejemplos de secuencias enlazadoras de 6 o más aminoácidos de longitud.

5 Las proteínas descritas en el presente documento pueden incluir cualquier combinación de enlazadores adecuados entre los dominios de unión a ADN individuales de la proteína. Véase también la publicación de patente de Estados Unidos N° 20110287512.

Donantes

10 Tal como se ha indicado anteriormente, la inserción de una secuencia exógena (también denominada "secuencia donante" o "donante" o "transgén"), por ejemplo, para la corrección de un gen mutante o para la expresión aumentada de un gen de tipo silvestre. Será fácilmente evidente que la secuencia donante generalmente no es idéntica a la secuencia genómica en la que se dispone. Una secuencia donante puede contener una secuencia no homóloga

15 flanqueada por dos regiones de homología para permitir una HDR eficaz en la ubicación de interés. Además, las secuencias donantes pueden comprender una molécula 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. Por ejemplo, para la inserción dirigida de secuencias que normalmente no están presentes en una región de interés, dichas secuencias pueden estar presentes en una molécula de ácido

20 nucleico donante y estar flanqueadas por regiones de homología con respecto a la secuencia presente en la región de interés.

En el presente documento se describen procedimientos de inserción dirigida de cualquier polinucleótido para su inserción en una ubicación elegida. Los polinucleótidos para inserción también pueden denominarse polinucleótidos

25 "exógenos", polinucleótidos o moléculas "donantes" o "transgenes". El polinucleótido donante puede ser ADN o ARN, monocatenario y/o bicatenario y puede introducirse en una célula en forma lineal o circular. Véanse, por ejemplo, las publicaciones de patente de Estados Unidos N° 20100047805, 20110281361, 20110207221 y la solicitud de Estados Unidos N° 13/889.162. La o las secuencias donantes pueden estar contenidas dentro de un ADN MC, que puede introducirse en la célula en forma circular o lineal. Si se introduce en forma lineal, los extremos de la secuencia donante

30 pueden protegerse (por ejemplo, de la degradación exonucleolítica) mediante procedimientos conocidos por los expertos en la técnica. Por ejemplo, se añaden uno o más residuos de didesoxinucleótidos al extremo 3' de una molécula lineal y/o se unen oligonucleótidos autocomplementarios a uno o ambos extremos. Véanse, por ejemplo, Chang et al. (1987) *Proc. Natl. Acad. Sci. USA* 84: 4959-4963; Nehls et al. (1996) *Science* 272: 886-889. Los procedimientos adicionales para proteger polinucleótidos exógenos de la degradación incluyen, pero sin limitación, la

35 adición de un grupo o grupos amino terminales y el uso de enlaces internucleotídicos modificados tales como, por ejemplo, fosforitoatos, fosforamidatos y residuos de O-metilribosa o desoxirribosa.

Se puede introducir un polinucleótido en una célula como parte de una molécula vector que tiene secuencias adicionales tales como, por ejemplo, orígenes de replicación, promotores y genes que codifican resistencia a

40 antibióticos. Además, los polinucleótidos donantes pueden introducirse como ácido nucleico desnudo, como ácido nucleico complejado con un agente tal como un liposoma o un poloxámero, o pueden administrarse por medio de virus (por ejemplo, adenovirus, AAV, herpesvirus, retrovirus, lentivirus y lentivirus deficientes en integrasa (IDLV).

En determinadas formas de realización, el donante bicatenario incluye secuencias (por ejemplo, secuencias codificantes, también denominadas transgenes) de más de 1 kb de longitud, por ejemplo, de entre 2 y 200 kb, de entre

45 2 y 10 kb (o cualquier valor intermedio). El donante bicatenario también incluye al menos un sitio diana de nucleasa, por ejemplo. En determinadas formas de realización, el donante incluye al menos 1 sitio diana, por ejemplo, para su uso con un CRISPR/Cas, o 2 sitios diana, por ejemplo, para un par de ZFN o TALEN. Típicamente, los sitios diana de nucleasa se encuentran fuera de las secuencias transgénicas, por ejemplo, 5' y/o 3' con respecto a las secuencias transgénicas, para la escisión del transgén. El o los sitios de escisión de la nucleasa pueden ser para cualquier nucleasa. En determinadas formas de realización, los sitios diana de nucleasa contenidos en el donante bicatenario son para la o las mismas nucleasas utilizadas para escindir la diana endógena en la que se integra el donante escindido

50 por medio de procedimientos independientes de la homología.

El donante generalmente se inserta de forma que su expresión esté dirigida por el promotor endógeno en el sitio de la integración, es decir, el promotor que dirige la expresión del gen endógeno en el que se inserta el donante (por ejemplo, globina, AAVS1, etc.). Sin embargo, será evidente que el donante puede comprender un promotor y/o potenciador,

55 por ejemplo, un promotor constitutivo o un promotor inducible o específico de tejido.

La molécula donante puede insertarse en un gen endógeno de forma que se exprese la totalidad, parte o nada del gen endógeno. Por ejemplo, un transgén tal como se describe en el presente documento puede insertarse en un locus de globina de forma que algunas o ninguna de las secuencias de globina endógena se expresen, por ejemplo, como

60

una fusión con el transgén. En otras formas de realización, el transgén (por ejemplo, con o sin secuencias codificadoras de globina) se integra en cualquier locus endógeno, por ejemplo, un locus de puerto seguro. Véanse, por ejemplo, las publicaciones de patente de Estados Unidos N° 20080299580, 20080159996 y 201000218264.

5 Cuando se expresan secuencias adicionales (por ejemplo, secuencias de globina, endógenas o parte del transgén) con el transgén, las secuencias adicionales (por ejemplo, globina) pueden ser secuencias de longitud completa (de tipo silvestre o mutante) o secuencias parciales. Preferentemente, las secuencias adicionales son funcionales. Los ejemplos no limitantes de la función de estas secuencias adicionales de longitud completa o parciales, por ejemplo, secuencias que codifican la globina, incluyen aumentar la semivida en suero del polipéptido expresado por el transgén (por ejemplo, gen terapéutico) y/o actuar como vehículo.

Además, aunque no se requieren para la expresión, las secuencias exógenas también pueden incluir secuencias reguladoras de la transcripción o de la traducción, por ejemplo, promotores, potenciadores, aislantes, sitios internos de entrada al ribosoma, secuencias que codifican péptidos 2A y/o señales de poliadenilación.

15 Los transgenes portados en las secuencias donantes descritas en el presente documento pueden aislarse a partir de plásmidos, células u otras fuentes utilizando técnicas estándar conocidas en la técnica, tales como PCR. Los donantes para su uso pueden incluir diversos tipos de topología, incluidos superenrollados circulares, relajados circulares, lineales y similares. Alternativamente, pueden sintetizarse químicamente utilizando técnicas de síntesis de oligonucleótidos estándar. Además, los donantes pueden estar metilados o carecer de metilación. Los donantes pueden estar en forma de cromosomas bacterianos o de levadura artificial (BAC o YAC).

Los polinucleótidos donantes bicatenarios descritos en el presente documento pueden incluir una o más bases y/o cadenas principales no naturales. En particular, la inserción de una molécula donante con citosinas metiladas se puede llevar a cabo utilizando los procedimientos descritos en el presente documento para lograr un estado de inactividad transcripcional en una región de interés.

El polinucleótido exógeno (donante) puede comprender cualquier secuencia de interés (secuencia exógena). Los ejemplos de secuencias exógenas incluyen, pero sin limitación, cualquier secuencia codificante de polipéptidos (por ejemplo, ADNc), secuencias promotoras, secuencias potenciadoras, etiquetas de epítipo, genes marcadores, sitios de reconocimiento de enzimas de escisión y diversos tipos de constructos de expresión. Los genes marcadores incluyen, pero sin limitación, secuencias que codifican proteínas que median la resistencia a antibióticos (por ejemplo, resistencia a ampicilina, resistencia a neomicina, resistencia a G418, resistencia a puromicina), secuencias que codifican proteínas coloreadas, fluorescentes o luminiscentes (por ejemplo, proteína fluorescente verde, proteína fluorescente verde potenciada, proteína fluorescente roja, luciferasa) y proteínas que median el crecimiento celular mejorado y/o la amplificación de genes (por ejemplo, dihidrofolato reductasa). Las etiquetas de epítipo incluyen, por ejemplo, una o más copias de FLAG, His, myc, Tap, HA o cualquier secuencia de aminoácidos detectable.

En una forma de realización preferida, la secuencia exógena (transgén) comprende un polinucleótido que codifica cualquier polipéptido cuya expresión en la célula se desea, incluidos, pero sin limitación, anticuerpos, antígenos, enzimas, receptores (de superficie celular o nucleares), hormonas, linfocinas, citocinas, polipéptidos informadores, factores de crecimiento y fragmentos funcionales de cualquiera de los anteriores. Las secuencias de codificación pueden ser, por ejemplo, ADNc.

En determinadas formas de realización, las secuencias exógenas pueden comprender un gen marcador (descrito anteriormente), que permite la selección de células que se han sometido a integración dirigida, y una secuencia única que codifica una funcionalidad adicional. Ejemplos no limitantes de genes marcadores incluyen GFP, marcador(es) de selección de fármacos y similares.

Las secuencias de genes adicionales que pueden insertarse pueden incluir, por ejemplo, genes de tipo silvestre para reemplazar secuencias mutadas. Por ejemplo, una secuencia del gen de la beta-globina de tipo silvestre puede insertarse en el genoma de una célula madre en la que la copia endógena del gen está mutada. La copia de tipo silvestre puede insertarse en el locus endógeno, o puede dirigirse alternativamente a un locus de puerto seguro.

La construcción de dichos casetes de expresión, siguiendo las enseñanzas de la presente especificación, utiliza metodologías bien conocidas en la técnica de la biología molecular (véase, por ejemplo, Ausubel o Maniatis). Antes de utilizar el casete de expresión para generar un animal transgénico, se puede analizar la capacidad de respuesta del casete de expresión al inductor de estrés asociado con elementos de control seleccionados introduciendo el casete de expresión en una línea celular adecuada (por ejemplo, células primarias, células transformadas o líneas celulares inmortalizadas).

Además, aunque no se requiere para la expresión, las secuencias exógenas también pueden ser secuencias reguladoras de la transcripción o de la traducción, por ejemplo, promotores, potenciadores, aisladores, sitios internos de entrada al ribosoma, secuencias que codifican péptidos 2A y/o señales de poliadenilación. Además, los elementos de control de los genes de interés se pueden unir operativamente a genes informadores para crear genes quiméricos (por ejemplo, casetes de expresión informadores).

También se puede lograr la inserción dirigida de la secuencia de ácido nucleico no codificante. Las secuencias que codifican ARN antisentido, ARNi, ARNsh y microARN (miARN) también pueden utilizarse para inserciones dirigidas.

En formas de realización adicionales, el ácido nucleico donante puede comprender secuencias no codificantes que son sitios diana específicos para diseños de nucleasa adicionales. Posteriormente, se pueden expresar nucleasas adicionales en células de modo que la molécula donante original se escinda y se modifique mediante la inserción de otra molécula donante de interés. De esta forma, se pueden generar integraciones reiterativas de moléculas donantes que permitan el apilamiento de rasgos en un lugar de interés particular o en un lugar de puerto seguro.

Administración

Las nucleasas, los polinucleótidos que codifican estas nucleasas, los polinucleótidos donantes y las composiciones que comprenden las proteínas y/o los polinucleótidos descritos en el presente documento pueden administrarse *in vivo* o *ex vivo* por cualquier medio adecuado.

Los procedimientos para administrar nucleasas tal como se describen en el presente documento se describen, por ejemplo, en las patentes de Estados Unidos N° 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.

Las nucleasas y/o los constructos donantes tal como se describen en el presente documento también pueden administrarse utilizando vectores que contienen secuencias que codifican uno o más proteínas de dedos de zinc o TALEN. Se puede utilizar cualquier sistema de vectores que incluye, pero sin limitación, vectores plasmídicos, vectores retrovirales, vectores lentivirales, vectores de adenovirus, vectores de poxvirus; vectores de herpesvirus y vectores de virus adenoasociados, etc. Véanse también las patentes de Estados Unidos N° 6.534.261, 6.607.882, 6.824.978, 6.933.113, 6.979.539, 7.013.219 y 7.163.824. Además, será evidente que cualquiera de estos vectores puede comprender una o más de las secuencias necesarias para el tratamiento. Por lo tanto, cuando una o más nucleasas y un constructo donante se introducen en la célula, las nucleasas y/o el polinucleótido donante pueden transportarse en el mismo vector o en diferentes vectores. Cuando se utilizan múltiples vectores, cada vector puede comprender una secuencia que codifica una o múltiples nucleasas y/o constructos donantes.

Se pueden utilizar procedimientos convencionales de transferencia de genes basados en virus y no basados en virus para introducir ácidos nucleicos que codifican nucleasas y constructos donantes en las células (por ejemplo, células de mamíferos) y tejidos diana. Los sistemas de administración de vectores no víricos incluyen plásmidos de ADN, ácido nucleico desnudo y ácido nucleico complejado con un vehículo de administración tal como un liposoma o un poloxámero. Los sistemas de administración de vectores víricos incluyen virus de ADN y ARN, que tienen genomas episómicos o integrados después de la administración a la célula. Para una revisión de los procedimientos de genoterapia, véanse Anderson, *Science* 256: 808-813 (1992); Nabel y Felgner, *TIBTECH* 11: 211-217 (1993); Mitani y Caskey, *TIBTECH* 11: 162-166 (1993); Dillon, *TIBTECH* 11: 167-175 (1993); Miller, *Nature* 357: 455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer y Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada et al., en *Current Topics in Microbiology and Immunology*, Doerfler y Böhm (eds.) (1995); y Yu et al., *Gene Therapy* 1:13-26 (1994).

Los procedimientos de administración no vírica de ácidos nucleicos incluyen electroporación, lipofección, microinyección, biolística, virosomas, liposomas, inmunoliposomas, polimerización o conjugados de lípido:ácido nucleico, ADN desnudo, viriones artificiales y captación de ADN potenciada por agentes. También puede utilizarse sonoporación usando, por ejemplo, el sistema Sonitron 2000 (Rich-Mar) para la administración de ácidos nucleicos.

Los sistemas de administración de ácido nucleico ejemplares adicionales incluyen los proporcionados por Amaxa Biosystems (Colonia, Alemania), Maxcyte, Inc. (Rockville, Maryland), BTX Molecular Delivery Systems (Holliston, MA) y Copernicus Therapeutics Inc. (véase, por ejemplo, el documento US6008336). La lipofección se describe en, por ejemplo, las patentes de Estados Unidos N° 5.049.386, 4.946.787 y 4.897.355) y los reactivos de lipofección se venden comercialmente (por ejemplo, Transfectam™ y Lipofectin™). Los lípidos catiónicos y neutros que son adecuados para la lipofección eficaz de polinucleótidos por reconocimiento de receptor incluyen los de Felgner, documento WO 91/17424, documento WO 91/16024.

La preparación de complejos de lípido:ácido nucleico, incluidos liposomas específicos tales como complejos de inmunolípidos, es bien conocida por un experto en la técnica (véanse, por ejemplo, Crystal, *Science* 270: 404-410 (1995); Blaese et al., *Cancer Gene Ther.* 2: 291-297 (1995); Behr et al., *Bioconjugate Chem.* 5: 382-389 (1994); Remy et al., *Bioconjugate Chem.* 5: 647-654 (1994); Gao et al., *Gene Therapy* 2: 710-722 (1995); Ahmad et al., *Cancer Res.* 52: 4817-4820 (1992); patentes de Estados Unidos N° 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).

Los procedimientos adicionales de administración incluyen el uso de empaquetamiento de los ácidos nucleicos para su administración en vehículos de administración EnGeneIC (EDV). Estos EDV se administran específicamente a tejidos diana utilizando anticuerpos biespecíficos en los que un brazo del anticuerpo tiene especificidad por el tejido diana y el otro tiene especificidad por el EDV. El anticuerpo lleva los EDV a la superficie celular diana y después el EDV se lleva dentro de la célula por endocitosis. Una vez en la célula, se liberan los contenidos. (Véase 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 ZFP genomanipulados aprovecha procesos altamente evolucionados para dirigir un virus a células específicas en el cuerpo y transportar la carga vírica hacia el núcleo. Los vectores víricos se pueden administrar directamente a los sujetos (*in vivo*) o pueden utilizarse para tratar células *in vitro* y las células modificadas se administran a sujetos (*ex vivo*). Los sistemas convencionales basados en virus para el suministro de ZFP incluyen, pero sin limitación, vectores de virus retrovirales, lentivirus, adenovirales, adenoasociados, vacuna y de herpes simple para la transferencia de genes. La integración en el genoma del huésped es posible con los procedimientos de transferencia génica de retrovirus, lentivirus y virus adenoasociados, lo que a menudo da como resultado la expresión a largo plazo del transgén insertado. Además, se han observado altas eficacias de transducción en muchos tipos diferentes de células y tejidos diana.

El tropismo de un retrovirus puede alterarse incorporando proteínas de envoltura extrañas, expandiendo la población diana potencial de células diana. Los vectores lentivirales son vectores retrovirales que son capaces de transducir o infectar células que no están dividiendo y típicamente producen altos títulos víricos. La selección de un sistema de transferencia génica retroviral depende del tejido diana. Los vectores retrovirales están compuestos por repeticiones terminales largas que accionan en *cis* con capacidad de empaquetamiento de hasta 6-10 kb de secuencia extraña. Los LTR mínimos que actúan en *cis* son suficientes para la replicación y el empaquetamiento de los vectores, que después se utilizan para integrar el gen terapéutico en la célula diana para proporcionar una expresión transgénica permanente. Los vectores retrovirales ampliamente utilizados incluyen aquellos basados en el virus de la leucemia murina (MuLV), el virus de la leucemia del mono gibón (GaLV), el virus de la inmunodeficiencia de simios (SIV), el virus de la inmunodeficiencia humana (VIH) y combinaciones de los mismos (véanse, por ejemplo, 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); documento PCT/US94/05700).

En las aplicaciones en las que se prefiere la expresión transitoria, se pueden utilizar sistemas basados en adenovirus. Los vectores basados en adenovirus son capaces de una eficacia de transducción muy alta en muchos tipos de células y no requieren división celular. Con dichos vectores se han obtenido un alto título y altos niveles de expresión. Este vector se puede producir en grandes cantidades en un sistema relativamente simple. Los vectores de virus adenoasociados ("AAV") también se utilizan para transducir células con ácidos nucleicos diana, por ejemplo, en la producción *in vitro* de ácidos nucleicos y péptidos, y para procedimientos *in vivo* y *ex vivo* de genoterapia (véanse, por ejemplo, West et al., *Virology* 160: 38-47 (1987); la patente de Estados Unidos N° 4.797.368; el documento 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 una serie de publicaciones, que incluyen la patente de Estados Unidos N° 5.173.414; Tratschin et al., *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, et al., *Mol. Cell. Biol.* 4: 2072-2081 (1984); Hermonat y Muzyczka, *PNAS* 81: 6466-6470 (1984); y Samulski et al., *J. Virol.* 63: 03822-3828 (1989).

Actualmente, al menos seis enfoques de vectores víricos están disponibles para la transferencia de genes en ensayos clínicos, que utilizan enfoques que implican la complementación de vectores defectuosos por medio de genes insertados en líneas celulares auxiliares para generar el agente de transducción.

pLASN y MFG-S son ejemplos de vectores retrovirales que se han utilizado en ensayos clínicos (Dunbar et al., *Blood* 85: 3048-305 (1995); Kohn et al., *Nat. Med.* 1: 1017-102 (1995); Malech et al., *PNAS* 94:22 12133-12138 (1997)). PA317/pLASN fue el primer vector terapéutico utilizado en un ensayo de genoterapia (Blaese et al., *Science* 270: 475-480 (1995)). Se han observado eficacias de transducción del 50% o más para los vectores empaquetados 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 sistemas prometedores de administración de genes alternativos basados en el virus parvovirus adenoasociado de tipo 2 defectuoso y no patógeno. Todos los vectores se

derivan de un plásmido que conserva solo las repeticiones terminales invertidas AAV de 145 pb que flanquean el casete de expresión del transgén. La transferencia eficaz de genes y la administración estable de transgenes 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)). Otros serotipos de AAV, incluidos AAV1, AAV3, AAV4, AAV5, AAV6, AAV8, AAV9 y AAVrh10, y todas sus variantes, también pueden utilizarse según la presente invención.

Los vectores adenovirales (Ad) recombinantes con replicación deficiente pueden producirse con un título alto e infectar fácilmente varios tipos de células diferentes. La mayor parte de los vectores de adenovirus están genomanipulados de tal forma que un transgén reemplaza los genes de Ad E1a, E1b y/o E3; posteriormente, el vector con replicación deficiente se propaga en células 293 humanas que suministran la función del gen eliminado en trans. Los vectores Ad pueden transducir múltiples tipos de tejidos *in vivo*, incluidas células diferenciadas que no se dividen, tales como las que se encuentran en el hígado, los riñones y los músculos. Los vectores Ad convencionales tienen una gran capacidad de carga. Un ejemplo del uso de un vector Ad en un ensayo clínico incluyó el tratamiento con polinucleótidos para la inmunización antitumoral con inyección intramuscular (Sterman et al., *Hum. Gene Ther.* 7: 1083-9 (1998)). Los ejemplos adicionales del uso de vectores de adenovirus para la transferencia de genes en ensayos clínicos incluyen Rosenecker et al., *Infection* 24:1 5-10 (1996); Sterman et al., *Hum. Gene Ther.* 9:7 1083-1089 (1998); Welsh et al., *Hum. Gene Ther.* 2:205-18 (1995); Álvarez et al., *Hum. Gene Ther.* 5:597-613 (1997); Topf et al., *Gene Ther.* 5:507-513 (1998); Sterman et al., *Hum. Gene Ther.* 7:1083-1089 (1998).

Las células de empaquetamiento se utilizan para formar partículas de virus que son capaces de infectar una célula huésped. Dichas células incluyen células 293, que empaquetan adenovirus, y células ψ 2 o células PA317, que empaquetan retrovirus. Los vectores víricos utilizados en la genoterapia generalmente se generan por una línea celular productora que empaqueta un vector de ácido nucleico en una partícula vírica. Los vectores típicamente contienen las secuencias víricas mínimas requeridas para el empaquetamiento y la posterior integración en un huésped (si corresponde), siendo reemplazadas otras secuencias víricas por un casete de expresión que codifica la proteína que se va a expresar. Las funciones víricas que faltan se suministran en trans por la línea celular de empaquetamiento. Por ejemplo, los vectores de AAV utilizados en genoterapia típicamente solo poseen secuencias de repetición terminal invertida (ITR) del genoma de AAV que se requieren para el empaquetamiento y la integración en el genoma del huésped. El ADN vírico se empaqueta en una línea celular, que contiene un plásmido auxiliar que codifica los otros genes AAV, concretamente rep y cap, pero carece de secuencias ITR. La línea celular también se infecta con adenovirus como ayudante. El virus auxiliar promueve la replicación del vector AAV y la expresión de genes AAV a partir del plásmido auxiliar. El plásmido auxiliar no se empaqueta en cantidades significativas debido a la falta de secuencias ITR. La contaminación con adenovirus puede reducirse mediante, por ejemplo, tratamiento térmico al que el adenovirus es más sensible que el AAV.

En muchas aplicaciones de genoterapia, es deseable que el vector de genoterapia se administre con un alto grado de especificidad a un tipo de tejido particular. Por consiguiente, un vector vírico puede modificarse para que tenga especificidad para un tipo de célula dado mediante la expresión de un ligando como una proteína de fusión con una proteína de la cubierta vírica en la superficie externa del virus. El ligando se elige para que tenga afinidad por un receptor que se sabe que está presente en el tipo de célula de interés. Por ejemplo, Han et al., *Proc. Natl. Acad. Sci. USA* 92: 9747-9751 (1995), informaron que el virus de la leucemia murina de Moloney se puede modificar para expresar heregulina humana fusionada con gp70, y el virus recombinante infecta determinadas células de cáncer de mama humano que expresan el receptor del factor de crecimiento epidérmico humano. Este principio puede extenderse a otras parejas de virus-célula diana, en las que la célula diana expresa un receptor y el virus expresa una proteína de fusión que comprende un ligando para el receptor de la superficie celular. Por ejemplo, se puede genomanipular un fago filamentoso para que muestre fragmentos de anticuerpos (por ejemplo, FAB o Fv) que tienen afinidad de unión específica por prácticamente cualquier receptor celular elegido. Aunque la descripción anterior se aplica principalmente a vectores víricos, los mismos principios se pueden aplicar a vectores no víricos. Dichos vectores pueden genomanipularse para que contengan secuencias de captación específicas que favorecen la captación por células diana específicas.

Se pueden administrar vectores de genoterapia *in vivo* mediante administración a un sujeto individual, típicamente por administración sistémica (por ejemplo, infusión intravenosa, intraperitoneal, intramuscular, subdérmica o intracraneal) o aplicación tópica, tal como se describe a continuación. Alternativamente, los vectores pueden administrarse a las células *ex vivo*, tales como células explantadas de un paciente individual (por ejemplo, linfocitos, aspirados de médula ósea, biopsia de tejido) o células madre hematopoyéticas de donantes universales, operación seguida de reimplantación de las células en un paciente, generalmente después de la selección de las células que han incorporado el vector.

También se pueden administrar vectores (por ejemplo, retrovirus, adenovirus, liposomas, etc.) que contienen nucleasas y/o constructos donantes directamente a un organismo para la transducción de células *in vivo*. Alternativamente, se puede administrar ADN desnudo. La administración se realiza por cualquiera de las rutas que

normalmente se utilizan para introducir una molécula en contacto final con las células sanguíneas o tisulares, incluidas, pero sin limitación, inyección, infusión, aplicación tópica y electroporación. Los procedimientos adecuados para administrar dichos ácidos nucleicos están disponibles y son bien conocidos por los expertos en la técnica, y, aunque se puede utilizar más de una ruta para administrar una composición particular, una ruta particular a menudo puede proporcionar una reacción más inmediata y más eficaz que otra ruta.

Los vectores adecuados para la introducción de polinucleótidos descritos en el presente documento incluyen vectores de lentivirus que no se integran (IDLV). Véanse, 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; publicación de patente de Estados Unidos N° 2009/054985.

Los vehículos farmacéuticamente aceptables se determinan en parte por la composición particular que se administra, así como por el procedimiento particular utilizado para administrar la composición. En consecuencia, existe una amplia diversidad de formulaciones adecuadas de composiciones farmacéuticas disponibles, tal como se describe a continuación (véase, por ejemplo, *Remington's Pharmaceutical Sciences*, 17^a ed., 1989).

Será evidente que las secuencias que codifican la nucleasa y los constructos donantes pueden administrarse utilizando el mismo sistema o sistemas diferentes. Por ejemplo, un polinucleótido donante puede ser transportado por un plásmido, mientras que las, una o más, nucleasas pueden ser transportadas por un vector AAV. Además, los diferentes vectores pueden administrarse por la misma o por diferentes rutas (inyección intramuscular, inyección en la vena de la cola, otra inyección intravenosa, administración intraperitoneal y/o inyección intramuscular). Los vectores pueden administrarse simultáneamente o en cualquier orden secuencial.

Por lo tanto, la presente divulgación incluye el tratamiento *in vivo* o *ex vivo* de enfermedades y afecciones que son susceptibles de inserción de transgenes que codifican una proteína terapéutica, por ejemplo, el tratamiento de hemoglobinopatías mediante la integración mediada por nucleasas de un gen que codifica una proteína globina. 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 ser por cualquier medio en el que los polinucleótidos se suministren a las células diana deseadas. Por ejemplo, se contemplan procedimientos tanto *in vivo* como *ex vivo*. La inyección intravenosa en la vena porta es un procedimiento preferido de administración. Otros modos de administración *in vivo* incluyen, por ejemplo, inyección directa en los lóbulos del hígado o el conducto biliar e inyección intravenosa distal al hígado, incluso a través de la arteria hepática, inyección directa en el parénquima hepático, inyección a través de la arteria hepática y/o inyección retrógrada a través del árbol biliar. Los modos de administración *ex vivo* incluyen transducción *in vitro* de los hepatocitos resecados u otras células del hígado, seguida de la infusión de los hepatocitos resecados y transducidos nuevamente en la vasculatura portal, el parénquima hepático o el árbol biliar del paciente humano, véase, por ejemplo, Grossman et al., (1994) *Nature Genetics*, 6: 335-341.

La cantidad eficaz de nucleasa(s) y donante que se va a administrar variará de paciente a paciente y según el polipéptido terapéutico de interés. Por consiguiente, las cantidades eficaces las determina mejor el médico que administra las composiciones y las dosis apropiadas las puede determinar fácilmente un experto en la técnica. Después de permitir suficiente tiempo para la integración y la expresión (por lo general, de 4 a 15 días, por ejemplo), el análisis del suero u otros niveles tisulares del polipéptido terapéutico y la comparación con el nivel inicial antes de la administración determinarán si la cantidad administrada es demasiado baja, se encuentra dentro del intervalo correcto o es demasiado alta. Los regímenes adecuados para las administraciones iniciales y posteriores también son variables, pero están tipificados por una administración inicial seguida de administraciones posteriores si es necesario. Las administraciones posteriores pueden administrarse a intervalos variables, que van desde diariamente a anualmente a cada varios años. Un experto en la técnica apreciará que se pueden recomendar técnicas inmunosupresoras apropiadas para evitar la inhibición o el bloqueo de la transducción por inmunosupresión de los vectores de administración; véase, por ejemplo, Vilquin et al., (1995) *Human Gene Ther.* 6: 1391-1401.

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

Aplicaciones

Los procedimientos y las composiciones que se divulgan en el presente documento son para modificar la expresión de proteínas, o corregir una secuencia de genes aberrantes que codifica una proteína expresada en una enfermedad genética, tal como una enfermedad de células falciformes o una talasemia. Por lo tanto, los procedimientos y las

composiciones prevén el tratamiento y/o la prevención de dichas enfermedades genéticas. La edición del genoma, por ejemplo, de células madre, se utiliza para corregir un gen aberrante, insertar un gen de tipo silvestre o cambiar la expresión de un gen endógeno. A modo de ejemplo no limitante, un gen de tipo silvestre, por ejemplo, que codifica al menos una globina (por ejemplo, α - y/o β -globina), se pueden insertar en una célula para proporcionar las proteínas globinas que son deficientes y/o que faltan en la célula y, por lo tanto, tratar una enfermedad genética, por ejemplo, una hemoglobinopatía, causada por una expresión de globina defectuosa. Como alternativa o adicionalmente, la edición genómica con o sin la administración del donante apropiado puede corregir el gen endógeno defectuoso, por ejemplo, corrigiendo la mutación puntual en la hemoglobina α o β , para restaurar la expresión del gen y/o tratar una enfermedad genética, por ejemplo la enfermedad de células falciformes, y/o la inactivación o la alteración (sobrexpresión o represión) de cualquier gen regulador de globina directo o indirecto (por ejemplo inactivación del gen regulador de γ -globina BCL11A o del regulador de BCL11A KLF1).

Los procedimientos y las composiciones de la invención también se pueden utilizar en cualquier circunstancia en la que se desee suministrar un transgén que codifique uno o más productos terapéuticos de forma que el o los productos terapéuticos se produzcan en un glóbulo rojo y/o una célula madre hematopoyética de modo que los glóbulos rojos maduros derivados de estas células contengan el agente terapéutico.

Los siguientes ejemplos se refieren a aspectos ejemplares de la presente divulgación en los que la nucleasa comprende una nucleasa de dedo de zinc (ZFN) o una TALEN.

Ejemplos

Ejemplo 1: Diseño, construcción y caracterización general de las nucleasas de proteína de dedo de zinc (ZFN)

Las proteínas de dedos de zinc se diseñaron e incorporaron en plásmidos, AAV o vectores adenovirales esencialmente tal como se describe por Urnov et al. (2005) *Nature* 435 (7042): 646-651, Perez et al (2008) *Nature Biotechnology* 26 (7): 808-816 y tal como se describe en la patente de Estados Unidos N° 6.534.261. Para ZFN y TALEN específicas para el locus de beta_globina humana y el locus HPRT humano, véanse la patente de Estados Unidos en copropiedad N° 7.888.121 y las publicaciones de patente de Estados Unidos N° 20130137104 y 20130122591. Para las nucleasas específicas para AAVS1 humano, véase la patente de Estados Unidos en copropiedad N° 8.110.379. Para las nucleasas específicas para CCR5, véase la patente de Estados Unidos en copropiedad N° 7.951.925. Para las nucleasas específicas para albúmina, véanse las publicaciones de patente de Estados Unidos N° 20130177983 y 20139177968.

Ejemplo 2: Actividad de ZFN específicas de globina

Se utilizaron pares de ZFN dirigidos al locus de globina humana o reguladores de la expresión del gen de la globina de tipo beta para analizar la capacidad de estos ZFN para inducir DSB en un sitio diana específico. Las secuencias de aminoácidos de las regiones de hélice de reconocimiento de cada dedo de las ZFN indicadas se muestran a continuación en la tabla 1A junto con todos los sitios diana (sitios diana de ADN indicados en letras mayúsculas; nucleótidos no contactados indicados en minúsculas).

Tabla 1A: nucleasas de dedos de zinc

SBS #, Diana	Diseño					
	ZFN específicas de B-hemoglobina humana					
	F1	F2	F3	F4	F5	F6
SBS#33511 <u>ggGCAGTAACGGC</u> <u>AGACtctcctca</u> gg (SEQ ID NO:8)	DRSNLSR (SEQ ID NO:9)	QSSDLRR (SEQ ID NO:10)	RSDTLSA (SEQ ID NO:11)	QSGALAR (SEQ ID NO:12)	QSGDLTR (SEQ ID NO:13)	N/A
SBS#33533 <u>tgGGCAAGGTGA</u> <u>ACGTGGAtgaagt</u> tg	QSAHRKN (SEQ ID NO:15)	LKHHLTD (SEQ ID NO:16)	QRSNLVR (SEQ ID NO:17)	TSGHLSR (SEQ ID NO:18)	QSNHLTE (SEQ ID NO:19)	RSHHLKA (SEQ ID NO:20)

ES 2 812 599 T3

(SEQ ID NO:14)						
SBS#35256 agAGTCAGGTGCA CCATggtgtctgt tt (SEQ ID NO:21)	TNQNRI (SEQ ID NO:22)	DRSNRTT (SEQ ID NO:23)	RNASRTR (SEQ ID NO:24)	RSDNLSE (SEQ ID NO:25)	RSQHRKT (SEQ ID NO:26)	N/A
SBS#35263 gtGGAGAAGTCtG CCGTtactgcctc gt (SEQ ID NO:27)	TSGSLSR (SEQ ID NO:28)	DRSDLSR (SEQ ID NO:29)	DRSALAR (SEQ ID NO:30)	QSSNLAR (SEQ ID NO:31)	QSGHLR (SEQ ID NO:32)	N/A
SBS#34770 acAGGAGTCAGGT GCACcatggtgtc tg (SEQ ID NO:33)	DQSNLRA (SEQ ID NO:34)	RNASRTR (SEQ ID NO:24)	RSDNLSE (SEQ ID NO:25)	RSQHRKT (SEQ ID NO:26)	RSDHLTQ (SEQ ID NO:35)	N/A
SBS#34791 gaGAAGTCtGCCG TTACTgcctctgtg gg (SEQ ID NO:36)	ARSTRTN (SEQ ID NO:37)	TSGSLSR (SEQ ID NO:28)	DRSDLSR (SEQ ID NO:29)	DRSARTR (SEQ ID NO:38)	QSGNLAR (SEQ ID NO:39)	N/A
SBS#34805 taACGGCAGACTT CTCCAcaggagtc ag (SEQ ID NO:40)	QSGDLTR (SEQ ID NO:13)	SSSDRKK (SEQ ID NO:41)	DRSNLSR (SEQ ID NO:9)	QSADRTK (SEQ ID NO:42)	RSDTLA (SEQ ID NO:11)	N/A
SBS#34826 gcCCTGTGGGCA AGGTgaacgtgga tg (SEQ ID NO:43)	LRHHLTR (SEQ ID NO:44)	QSGNLHV (SEQ ID NO:45)	RSAHLR (SEQ ID NO:46)	RSDVLST (SEQ ID NO:47)	RKQDLRT (SEQ ID NO:48)	N/A
SBS#35301 ggGCAGTAACGGC AGACTtctcctca gg (SEQ ID NO:8)	DRSNLSR (SEQ ID NO:9)	QSGDLTR (SEQ ID NO:13)	RSDTLA (SEQ ID NO:11)	QSGALAR (SEQ ID NO:12)	QSGDLTR (SEQ ID NO:13)	N/A
SBS#35328 tgGGGCAAGGTGA ACGTg gatgaagt tg (SEQ ID NO:14)	MSHHLRD (SEQ ID NO:49)	QRSNLVR (SEQ ID NO:17)	TSGHLR (SEQ ID NO:18)	QSNHLTE (SEQ ID NO:19)	RSHHLKA (SEQ ID NO:20)	N/A
SBS#35497 caCAGGGCAGTAA CgGCAGACTtctc ct (SEQ ID NO:50)	DRSNLSR (SEQ ID NO:9)	QSGDLTR (SEQ ID NO:13)	DRSNLSR (SEQ ID NO:9)	LKHHLTD (SEQ ID NO:16)	DRSHLTR (SEQ ID NO:51)	RSDNLRE (SEQ ID NO:52)
SBS#35506 ggCAAGGTGAACG TGGAtgaagttgg tg (SEQ ID NO:53)	QSGHLAR (SEQ ID NO:54)	VSHHLRD (SEQ ID NO:55)	QSGNLAR (SEQ ID NO:39)	LRHHLTR (SEQ ID NO:44)	QSGNLHV (SEQ ID NO:45)	N/A
Beta-globina IVS.1						
SBS#43545 atCAAGGTACAA GACAGGTttaagg ag (SEQ ID NO:158)	LRHHLTR (SEQ ID NO:44)	QSGTRKT (SEQ ID NO:153)	RSDNLST (SEQ ID NO:154)	DSANRIK (SEQ ID NO:155)	LRHHLTR (SEQ ID NO:44)	QSGNLHV (SEQ ID NO:45)

ES 2 812 599 T3

SBS#43544 aaTCTGCCAGGG CCTCaccaccaac tt (SEQ ID NO:159)	AMQTLRV (SEQ ID NO:156)	DRSHLAR (SEQ ID NO:76)	RSDNLSE (SEQ ID NO:25)	ASKTRKN (SEQ ID NO:77)	RNSDRTK (SEQ ID NO:157)	N/A
ZFN específicas de BCL11A humano						
Exón 2						
SBS#39172 ctCCAGAAGGGGA TCATGACctcctc ac (SEQ ID NO:160)	DRSNLSR (SEQ ID NO:9)	LRQNLIM (SEQ ID NO:161)	TSANLTV (SEQ ID NO:162)	RSDHLSR (SEQ ID NO:94)	QSGNLAR (SEQ ID NO:39)	QRNDRKS (SEQ ID NO:163)
SBS#43490 ctCCAGAAGGGGA TCATGACctcctc ac (SEQ ID NO:160)	DRSNLSR (SEQ ID NO:9)	LRQNLIM (SEQ ID NO:161)	LQSQ LNR (SEQ ID NO:164)	RSDHLSR (SEQ ID NO:94)	QSGNLAR (SEQ ID NO:39)	QRNDRKS (SEQ ID NO:163)
SBS#44642 ctCCAGAAGGGGA TCATGACctcctc ac (SEQ ID NO:160)	DRANLSR (SEQ ID NO:165)	LRQNLIM (SEQ ID NO:161)	LQSQ LNR (SEQ ID NO:164)	RSDHLSR (SEQ ID NO:94)	QSGNLAR (SEQ ID NO:39)	QRNDRKS (SEQ ID NO:163)
SBS#45148 ctCCAGAAGGGGA TCATGACctcctc ac (SEQ ID NO:160)	DRSNLSR (SEQ ID NO:9)	TSSNRNH (SEQ ID NO:166)	HSGNLTK (SEQ ID NO:167)	RSDHLSR (SEQ ID NO:94)	QSGNLAR (SEQ ID NO:39)	QKVDLSR (SEQ ID NO:168)
SBS#45147 ctCCAGAAGGGGA TCATGACctcctc ac (SEQ ID NO:160)	DRSNLSR (SEQ ID NO:9)	TSSNRNH (SEQ ID NO:166)	QANNLKV (SEQ ID NO:169)	RSDHLSR (SEQ ID NO:94)	QSGNLAR (SEQ ID NO:39)	QKVDLSR (SEQ ID NO:168)
SBS#39145 ccCAACGGGCCGT GGTCTGGttcatc at (SEQ ID NO:170)	RSDHLSA (SEQ ID NO:59)	DRSALAR (SEQ ID NO:30)	RSDSLSR (SEQ ID NO:171)	DRSVRTK (SEQ ID NO:172)	RSDHLSA (SEQ ID NO:59)	QRSNLKV (SEQ ID NO:173)
SBS#44490 ccCAACGGGCCGT GGTCTGGttcatc at (SEQ ID NO:170)	RSDHLTQ (SEQ ID NO:35)	DRSALAR (SEQ ID NO:30)	RSDSLSR (SEQ ID NO:171)	DRSVRTK (SEQ ID NO:172)	RSDHLSA (SEQ ID NO:59)	QRSNLKV (SEQ ID NO:173)
SBS#44489 ccCAACGGGCCGT GGTCTGGttcatc at (SEQ ID NO:170)	RSDHLTT (SEQ ID NO:174)	DRSALAR (SEQ ID NO:30)	RSDSLSR (SEQ ID NO:171)	DRSVRTK (SEQ ID NO:172)	RSDHLSA (SEQ ID NO:59)	QRSNLKV (SEQ ID NO:173)
SBS#45081 ccCAACGGGCCGT GGTCTGGttcatc at (SEQ ID NO:170)	RSDHLSA (SEQ ID NO:59)	WATARDR (SEQ ID NO:175)	RSDSLSR (SEQ ID NO:171)	HTKSLSR (SEQ ID NO:176)	RSDHLSA (SEQ ID NO:59)	QRSNLKV (SEQ ID NO:173)
SBS#44493 ccCAACGGGCCGT GGTCTGGttcatc at (SEQ ID NO:170)	RSAHLTQ (SEQ ID NO:177)	DRSVLRR (SEQ ID NO:178)	RSDSLSR (SEQ ID NO:171)	DRSVRTK (SEQ ID NO:172)	RSDHLSA (SEQ ID NO:59)	QRSNLKV (SEQ ID NO:173)
SBS#29527 atCCCATGGAGAG GTGGCTGggaagg	RSDVLSE (SEQ ID NO:57)	RNQHRTK (SEQ ID NO:58)	RSDHLSA (SEQ ID NO:59)	RSANLTR (SEQ ID NO:60)	RSDVLSN (SEQ ID NO:61)	DRSTRIT (SEQ ID NO:62)

ES 2 812 599 T3

ac (SEQ ID NO:56)						
SBS#29528 atATTGCAGACAA TAACccctttaac ct (SEQ ID NO:63)	DRSNLSR (SEQ ID NO:9)	HRQHLVT (SEQ ID NO:64)	DRSNLTR (SEQ ID NO:65)	QSGDLTR (SEQ ID NO:13)	HRSSLN(SEQ ID NO:66)	N/A
SBS# 29525 caTCCAGGCGTG GGGAttagagctc ca (SEQ ID NO:66)	QSGHLSR (SEQ ID NO:32)	RSDHLST (SEQ ID NO:67)	RSADLSR (SEQ ID NO:68)	RSDNLSQ (SEQ ID NO:69)	ASNDRKK(SEQ ID NO:70)	N/A
SBS#29526 gtGCAGAAATATGC CCCGCAGggtatt tg (SEQ ID NO:71)	RSDNLSA (SEQ ID NO:72)	RNNDRKT (SEQ ID NO:73)	DRSDLSR (SEQ ID NO:29)	TSSNRTK (SEQ ID NO:74)	QSGNLAR(SEQ ID NO:39)	QSGDLTR (SEQ ID NO:13)
Exón 4						
SBS#34678 atATTGCAGACAA TAACccctttaac ct (SEQ ID NO:179)	DRSNLSR (SEQ ID NO:9)	HRQHLVT (SEQ ID NO:64)	DRSNLTR (SEQ ID NO:65)	QSGDLTR (SEQ ID NO:13)	HRWLRSN(SEQ ID NO:180)	N/A
SBS#34642 atCCCATGgAGAG GTGGCTGGgaagg ac (SEQ ID NO:56)	RSDHLSQ (SEQ ID NO:99)	DSSHRTR (SEQ ID NO:181)	LRHHLTR (SEQ ID NO:44)	QSAHLKA (SEQ ID NO:182)	RSDVLSN(SEQ ID NO:61)	DRSTRIT (SEQ ID NO:62)
Bc111a-XL						
SBS#44889 ctCACTGTCCACA GGAGaagccacac gg (SEQ ID NO:183)	RSANLAR (SEQ ID NO:184)	RLDNRTA (SEQ ID NO:185)	QSNLNS (SEQ ID NO:186)	WRSSLKT (SEQ ID NO:187)	DRSNRKT (SEQ ID NO:188)	N/A
SBS#44888 ttGCTACAGTTCT TGAAGACTttccc ac (SEQ ID NO:189)	DRSNLSR (SEQ ID NO:9)	QSGNLAR (SEQ ID NO:39)	YKHVLS (SEQ ID NO:190)	TSGSLTR (SEQ ID NO:191)	QSGDLTR (SEQ ID NO:13)	LKDTLRR (SEQ ID NO:192)
SBS#44905* gaGAAGCCACACG GGCGAAAggcctt at (SEQ ID NO:193)	QSGNLDS (SEQ ID NO:194)	RSADLSR (SEQ ID NO:68)	RSDHLSE (SEQ ID NO:78)	QNATRIN (SEQ ID NO:195)	WNSDLRK (SEQ ID NO:196)	QSGNLAR (SEQ ID NO:39)
SBS#44904* tgGACAGTGAGAT TGCTacagttctt ga (SEQ ID NO:197)	QSSDLRS (SEQ ID NO:88)	YKWTLRN (SEQ ID NO:198)	RSANLTR (SEQ ID NO:60)	TSTKLRT (SEQ ID NO:199)	DRSNLTR (SEQ ID NO:65)	N/A
SBS#44911 gcCACACGGGCGA AaGGCCTTataaa tg (SEQ ID NO:200)	AMQTLRV (SEQ ID NO:156)	DRSHLAR (SEQ ID NO:76)	QRSNLVR (SEQ ID NO:17)	DRSHLAR (SEQ ID NO:76)	RSDTLST (SEQ ID NO:201)	DSSNRIN (SEQ ID NO:202)
SBS#44910 ctCCTGTGGACAG TGAGATTgctaca gt (SEQ ID NO:203)	NDLFLYL (SEQ ID NO:204)	RSANLTR (SEQ ID NO:60)	TSTKLRT (SEQ ID NO:199)	DRSNLTR (SEQ ID NO:65)	RSDSLSV (SEQ ID NO:205)	HNDSRKN (SEQ ID NO:206)
SBS#44945	QSGNLAR	CRQNLAN	YQGVLTR	RSDNLRE	DRSNRTT	HRSSLRR

ES 2 812 599 T3

aaGCTCACCAGGC ACATGAAaacgca tg (SEQ ID NO:207)	(SEQ ID NO:39)	(SEQ ID NO:208)	(SEQ ID NO:209)	(SEQ ID NO:52)	(SEQ ID NO:23)	(SEQ ID NO:210)
SBS#44944 ctACTCTGgGCAC AGGCATAGttgca ca (SEQ ID NO:211)	RSDNLST (SEQ ID NO:154)	QSSDLRR (SEQ ID NO:10)	RSDALSE (SEQ ID NO:212)	QNATRTK (SEQ ID NO:115)	RSDTLSE (SEQ ID NO:84)	ARSTRTN (SEQ ID NO:37)
SBS#44947** ctCACCAGGCACA TGAAAACgcatgg cc (SEQ ID NO:213)	GSSALTQ (SEQ ID NO:214)	QSGNLR (SEQ ID NO:39)	TASHLKE (SEQ ID NO:215)	QNATRTK (SEQ ID NO:115)	RSDNLSE (SEQ ID NO:25)	SSRNLAS (SEQ ID NO:216)
SBS#44946** ctACTCTGgGCAC AGGCATAGttgca ca (SEQ ID NO:211)	RSDNLST (SEQ ID NO:154)	QSSDLRR (SEQ ID NO:10)	RSDALSE (SEQ ID NO:212)	QNATRTK (SEQ ID NO:115)	RSDTLSE (SEQ ID NO:84)	ARSTRTN (SEQ ID NO:37)
ZFN específicas de KLF1 humano						
KLF- Exón 1						
SBS#36004 ggGAAGGGGCCCA GGGCGGTcagtgt gc (SEQ ID NO:75)	TSGHLSR (SEQ ID NO:18)	DRSHLAR (SEQ ID NO:76)	RSDNLSQ (SEQ ID NO:69)	ASNDRKK (SEQ ID NO:70)	RSDHLSE (SEQ ID NO:78)	QSGNLR (SEQ ID NO:39)
SBS#36021 acACACAGGATGA Cttcctcaaggtg gg (SEQ ID NO:79)	DRSNLTR (SEQ ID NO:65)	TSANLSR (SEQ ID NO:217)	RSDHLSE (SEQ ID NO:78)	QSASRKN (SEQ ID NO:81)	NA	NA
SBS#33237 ggGAAGGGGCCCA GGGCGGTcagtgt gc (SEQ ID NO:75)	TSGHLSR (SEQ ID NO:18)	DRSHLAR (SEQ ID NO:76)	RSDNLSE (SEQ ID NO:25)	ASKTRKN (SEQ ID NO:77)	RSDHLSE (SEQ ID NO:78)	QSGNLR (SEQ ID NO:39)
SBS#33238 acACACAGGATGA Cttcctcaaggtg gg (SEQ ID NO:79)	DRSNLSR (SEQ ID NO:9)	TSGNLTR (SEQ ID NO:80)	RSDHLSE (SEQ ID NO:78)	QSASRKN (SEQ ID NO:81)	N/A	N/A
SBS#33257 cgCCACCGGGCTC CGGGcctcagaag tt (SEQ ID NO:82)	RSAHLSR (SEQ ID NO:46)	DSSDRKK (SEQ ID NO:83)	DRSHLAR (SEQ ID NO:76)	RSDTLSE (SEQ ID NO:84)	QSGDLTR (SEQ ID NO:13)	N/A
SBS#33258 ccCCAGACcTGCG CTCTGGCGcccag cg (SEQ ID NO:85)	RSDSLLR (SEQ ID NO:86)	RLDWLPV (SEQ ID NO:87)	QSSDLRS (SEQ ID NO:88)	AASNRSK (SEQ ID NO:89)	DRSNLSR (SEQ ID NO:9)	QSGDLTR (SEQ ID NO:13)
SBS#33269 ggCTCGGGgGCCG GGGCTGGAgccag gg (SEQ ID NO:91)	QSSHLTR (SEQ ID NO:91)	QSSDLTR (SEQ ID NO:92)	RSDHLSE (SEQ ID NO:78)	HSRTRTK (SEQ ID NO:93)	RSDHLRS (SEQ ID NO:94)	DRSARNS (SEQ ID NO:95)

ES 2 812 599 T3

NO:90)						
SBS#33270 aaGGCGCTGGCGC TgCAACCGgtgta cc (SEQ ID NO:96)	RSDTLSE (SEQ ID NO:84)	QSHNRK (SEQ ID NO:97)	QSSDLR (SEQ ID NO:88)	DRSHLAR (SEQ ID NO:76)	QSSDLR (SEQ ID NO:88)	DRSHLAR (SEQ ID NO:76)
SBS#33271 ttGCAGCGCCAGC GCCTTGgctcgg gg (SEQ ID NO:98)	RSDHLSQ (SEQ ID NO:99)	HRSSLGD (SEQ ID NO:100)	RSDDLTR (SEQ ID NO:101)	QRSTLSS (SEQ ID NO:102)	RSADLTR (SEQ ID NO:103)	QSGDLTR (SEQ ID NO:13)
SBS#33272 cgGTGTACCCGGG GCCCGgcccggc tc (SEQ ID NO:104)	DRSDLSR (SEQ ID NO:29)	RSTHLVR (SEQ ID NO:105)	RSDSLST (SEQ ID NO:106)	DSSDRTK (SEQ ID NO:107)	RSAALAR (SEQ ID NO:108)	N/A
KLF- Exón 2						
SBS#36071 ggTGAGGAGGAGA TCCAggtcccagg tg (SEQ ID NO:218)	NNRDLIN (SEQ ID NO:219)	TSSNLSR (SEQ ID NO:220)	QSGHLSR (SEQ ID NO:32)	QSGHLAR (SEQ ID NO:54)	QRTHLNS (SEQ ID NO:221)	N/A
SBS#36085 ctTCTCGGGCCCCG GaCCCCGgtggcg cg (SEQ ID NO:222)	RSDHLSQ (SEQ ID NO:78)	HSRTRTK (SEQ ID NO:93)	RSDHLSQ (SEQ ID NO:78)	HSRTRTK (SEQ ID NO:93)	RSDHLSQ (SEQ ID NO:78)	RKSDRIK (SEQ ID NO:223)
ZFN de la región reguladora 5' de gamma-globina humana						
región reguladora (-175)						
SBS#34360 ttGCATTGAGATA GTGTGGGgaagg gc (SEQ ID NO:109)	RSDHLSV (SEQ ID NO:110)	RSDVRKT (SEQ ID NO:111)	RSDYLSK (SEQ ID NO:112)	TSSVRTT (SEQ ID NO:113)	RPYTLRL (SEQ ID NO:114)	QNATRTK (SEQ ID NO:115)
SBS#34363 atCTGTCTGAAAC GGTCcctggctaa ac (SEQ ID NO:116)	DRSALAR (SEQ ID NO:30)	RRDILHQ (SEQ ID NO:117)	QSGNLAR (SEQ ID NO:39)	LAYDRRK (SEQ ID NO:118)	RSDVLSE (SEQ ID NO:57)	N/A
SBS#34398 ttGCATTGAGAT AGTggtgggaagg gg (SEQ ID NO:119)	RSDSLR (SEQ ID NO:86)	QSCARNV (SEQ ID NO:120)	RSDNLAR (SEQ ID NO:121)	HRNTLLG (SEQ ID NO:122)	MRNRLNR (SEQ ID NO:123)	N/A
SBS#34400 ctGTCTGAaACGG TCcCTGGCTaaac tc (SEQ ID NO:124)	QSSDLR (SEQ ID NO:88)	RRDALLM (SEQ ID NO:125)	DRSALAR (SEQ ID NO:30)	RRDILHQ (SEQ ID NO:117)	QNAHRKT (SEQ ID NO:126)	DRSALAR (SEQ ID NO:30)
SBS#31160 taTTGCATtGAG ATAGTGTGgggaa gg (SEQ ID NO:127)	RSDSLR (SEQ ID NO:86)	LQHHLTD (SEQ ID NO:128)	TSGNLTR (SEQ ID NO:80)	TSTHLHI (SEQ ID NO:129)	QSGDLTR (SEQ ID NO:13)	HKWVLRQ (SEQ ID NO:130)
SBS#34365 ctGTCTGAaACGG	QSSDLR (SEQ ID	RRDALLM (SEQ ID	DRSALAR (SEQ ID	RRDILHQ (SEQ ID	QNAHRKT (SEQ ID	DRSALAR (SEQ ID

TCcCTGGCTaaac tc (SEQ ID NO:124)	NO:88)	NO:131)	NO:30)	NO:117)	NO:126)	NO:30)
región reguladora (-110)						
SBS#34539 tgGTCAAGGCAAG GCTGgccaaccca tg (SEQ ID NO:224)	RSDVLSE (SEQ ID NO:57)	RNQHRTK (SEQ ID NO:58)	QSGDLTR (SEQ ID NO:13)	RSDHLST (SEQ ID NO:67)	DRSALAR (SEQ ID NO:30)	NA
SBS#34574 gcCTTGACAAGGC AAACttgaccaat ag (SEQ ID NO:225)	DRSNRTT (SEQ ID NO:23)	QSGSLTR (SEQ ID NO:226)	RSDNLSV (SEQ ID NO:227)	DRSNLSR (SEQ ID NO:9)	LKFALAN (SEQ ID NO:228)	NA
SBS#43865 gcCTTGACAAGGC AAACttgaccaat ag (SEQ ID NO:225)	NPANLTR (SEQ ID NO:229)	QNATRTK (SEQ ID NO:115)	RSDNLSV (SEQ ID NO:227)	DRSNLSR (SEQ ID NO:9)	LKFALAN (SEQ ID NO:228)	NA
SBS#43852 tgGTCAAGGCAAG GCTGgccaaccca tg (SEQ ID NO:224)	RSDVLSE (SEQ ID NO:57)	RNQHRTK (SEQ ID NO:58)	QSGDLTR (SEQ ID NO:13)	RSDNLST (SEQ ID NO:154)	DSSARKK (SEQ ID NO:230)	NA

Nota: Los pares de ZFN específicos de BCL11A XL marcados con un solo asterisco (*) o con un doble asterisco (**) contienen los nuevos enlazadores L7a y L8p, respectivamente. Véase el ejemplo 6.

- 5 El ensayo Cel-I (Surveyor™, Transgenomics) tal como se describe por Pérez et al. (2008) *Nat. Biotechnol.* 26: 808-816 y Guschin et al. (2010) *Methods Mol Biol.* 649:247-56), se utilizó para detectar modificaciones inducidas por ZFN del gen diana en K562 o HSC. En este ensayo, la amplificación por PCR del sitio diana vino seguida de la cuantificación de inserciones y deleciones (indeles) utilizando la enzima de detección de desajuste Cel-I (Yang et al. (2000) *Biochemistry* 39: 3533-3541) que proporcionó una estimación de límite inferior de frecuencia DSB. Tres días después de la transfección del vector de expresión de ZFN en condiciones estándar (37 °C) o utilizando un choque hipotérmico (30 °C, véase la publicación de patente de Estados Unidos en copropiedad N° 20110041195), se aisló ADN genómico de células K562 utilizando el kit DNeasy (Qiagen).

15 Los resultados del ensayo Cel-I demostraron que las ZFN eran capaces de inducir la escisión en sus sitios diana respectivos (véase, también, la solicitud provisional de Estados Unidos en copropiedad N° 61/556.691). Los resultados se muestran en la figura 1 e indican que se encontraron proteínas activas para la mayor parte de los loci diana en el gen de beta-globina.

20 **Ejemplo 3:** Edición del locus de beta-globina

25 Las ZFN específicas del gen de beta-globina humana (HBB) (tabla 1) se utilizaron para introducir un ADN donante en el locus de beta-globina de la forma siguiente. Los ADN donantes se diseñaron de modo que la secuencia que codifica las secuencias del gen HBB estuviera flanqueada por secuencias que eran homólogas (brazos de homología) a la región que rodea el sitio de escisión de ZFN en el gen de beta-globina. Los brazos de homología tienen aproximadamente 500-600 pares de bases de longitud. La secuencia donante de HBB carece de cualquier secuencia no codificante, de modo que cuando se inserta en el sitio diana de beta-globina, la expresión del donante está regulada por el promotor de beta-globina y cualquier otra secuencia reguladora de beta-globina. Cuando se inserta, el donante HBB se fusiona en marco con las secuencias endógenas de globina y da como resultado una proteína de fusión. Además, se diseñó un oligo donante HBB para la captura en el gen HBB escindido después del tratamiento con ZFN. El oligo contenía un sitio de restricción tal que, después de la inserción del oligo, se introdujo un nuevo sitio de restricción en el gen HBB que posteriormente podría escindirse.

Tal como se muestra en la figura 2A, el donante oligo de β-globina se insertó en el locus apropiado, tal como se verificó por la presencia del nuevo sitio de restricción presente en el ADN donante. Además, tal como se muestra en la figura

5' CGTTCACCTTGCCCCACAGGGCAGTAACAGCAGATTTTTTCTCAGGAGTCAGGTGCACCA
TGGTGTCTGTTTGAGGTTGCTAGTGAAC

SMS12, 88 pb, R (SEQ ID NO: 234):

5' CGTTCACCTTGCCCCACAGGGCAGTAACGGCAGATTTTTTCTCAGGAGTCAGGTGCACCA
TGGTGTCTGTTTGAGGTTGCTAGTGAAC

5

SMS124, 101 pb, R (SEQ ID NO: 235):

5' CTTTCATCCACGTTTACCTTGCCCCACAGGGCAGTAACAGCAGATTTTTTCTCAGGAGTCA
GGTGCACCATGGTGTCTGTTTGAGGTTGCTAGTGAACACAG

10 Para investigar la capacidad de diferenciación y la longevidad de la corrección génica durante la diferenciación de células CD34+, se indujo a grupos de células CD34+ modificadas con ZFN a diferenciarse utilizando el medio de metilcelulosa Methocult de Stemcell Technologies según las instrucciones del fabricante. La diferenciación se analizó mediante el ensayo de tipos de colonias que surgen de la diferenciación inducida por Methocult: unidades formadoras de colonias, eritroides ("CFU-E"); unidades formadoras de estallido, eritroides ("BFU-E"); unidades formadoras de colonias, granulocitos/macrófagos ("CFU-GM") y unidades formadoras de colonias; granulocitos/eritrocitos/monocitos/macrófagos ("CFU-GEMM"). Los resultados indicaron que las células tratadas con ZFN conservan la misma capacidad de diferenciarse que las células transfectadas de forma simulada. Las colonias individuales de BFU-E se recogieron de la placa y se genotiparon en HBB. Los resultados indicaron que las modificaciones inducidas por ZFN se mantuvieron durante la diferenciación de colonias (véase la figura 11). Además, la frecuencia de las colonias de BFU-E modificadas fue similar a la frecuencia de los alelos modificados en el grupo inicial, lo que demuestra que no hay sesgo contra las células editadas durante la formación de BFU-E. Además, se analizó la población celular en su conjunto para determinar la modificación génica a lo largo del transcurso de la diferenciación de glóbulos rojos *in vitro* en cultivo líquido. Las modificaciones fueron estables durante al menos el proceso de diferenciación de glóbulos rojos de 18 días (véase la figura 12).

25 Otra mutación común en el gen de beta-globina que está asociada con beta-talasemia se conoce como IVS1.1. Esta mutación G-> A se encuentra dentro del primer par de bases del intrón 1 del gen de beta-globina, y su presencia en el gen da como resultado un corte y empalme defectuoso del pre-ARNm de beta-globina. Por lo tanto, se genomanipuló un par de ZFN para reconocer y escindir la región, esencialmente recapitulando esta mutación para propósitos de modelo. El análisis de estas ZFN encontró que pudieron escindir el sitio en el gen de beta-globina dando como resultado el 52,63% de NHEJ en células CD34+.

30

Ejemplo 5: Inserción de un donante de beta-globina en un locus de puerto seguro

35 Para insertar un gen de beta-globina de tipo silvestre en un locus de puerto seguro, de modo que la expresión del transgén corrija un déficit de beta-globina en un HSC, las nucleasas específicas de ese locus de puerto seguro se introducen en la célula junto con un ácido nucleico donante. Nucleasas específicas para HPRT (véanse las publicaciones de patente de Estados Unidos N° 20130137104 y 20130122591), AAVS1 (véase la patente de Estados Unidos N° 8.110.379), CCR5 (véase la patente de Estados Unidos N° 7.951.925) o beta-globina (véase la tabla 1A) se introducen en una célula madre CD34+ derivada del paciente. La introducción puede realizarse a través de cualquier procedimiento conocido en la técnica, tal como electroporación de ARNm. El ADN donante está diseñado para contener el transgén, la beta-globina de tipo silvestre y las regiones de homología que flanquean el transgén con suficiente homología con la región que rodea la diana de puerto seguro para permitir HDR (típicamente 500 pb en cada lado). Alternativamente, se puede proporcionar un constructo donante que, tanto si contiene como si carece de regiones de homología, se integre en el locus dirigido a ZFN o TALEN a través de la captura final (véase la solicitud de Estados Unidos N°. 13/889.162). El donante se cointroduce en la célula CD34+ antes, durante o después de la introducción de la ZFN. Las células CD34+ modificadas se reintroducen en el paciente y, después del injerto, producen beta-hemoglobina a niveles suficientes para permitir que se produzca una cantidad terapéuticamente relevante de hemoglobina.

40

45

50 **Ejemplo 6:** Inactivación de BCL11A y KLF1

Nucleasas específicas para BCL11A y KLF1 (por ejemplo, ZFN tal como se muestran en la tabla 1A) se introdujeron en HSC tal como se ha descrito anteriormente para causar una regulación al alza de la expresión de gamma-globina (véase la figura 3) y el genoma de las células analizadas por el ensayo Cel 1 tal como se ha descrito anteriormente (Pérez et al (2008), *ibídem*).

55

Tal como se muestra en la figura 4, después del tratamiento de HSC con las ZFN específicas de KLF1 indicados, las ZFN modificaron con éxito el locus KLF1 (figuras 4C y 4D). Del mismo modo, las ZFN específicas de BCL11A

modificaron el locus BCL11A (figura 4A). Un par de ZFN dirigidas al locus HPRT (véase la solicitud provisional de Estados Unidos en copropiedad 61/552.309) se utilizaron como control y también demostraron una escisión exitosa (figura 4B). La comparación de la señal en el día 3 después de la transducción de células CD34+ con el día 17 de cultivo de diferenciación (figura 4E) demostró que el porcentaje de edición génica (% de NHEJ) es estable a lo largo del tiempo. En cada gel que se muestra en la figura 4E, los carriles que carecen de identificación son controles negativos.

Se analizaron de forma similar pares adicionales de ZFN, dirigidos al exón 2 o al exón 4 de 4BCL11A. Para estos estudios, los pares de ZFN candidatos en células K562 se introdujeron por Amaxa tal como se ha descrito anteriormente o se introdujeron en células CD34+. Para la transducción de CD34+, se utilizó un dispositivo BTX ECM830 con una cubeta de hueco de 2 mm. Los ARNm de las células se prepararon utilizando un kit mMessageMachine T7 Ultra (# AM1345, Ambion). Se cultivaron células CD34+ humanas en medio x-vivo10 (Lonza) con 1xCC110 (Stem cell Technology) en placas no tratadas con cultivo tisular. Las células se contaron y se recogieron por centrifugación a 1200 rpm durante 10 minutos a temperatura ambiente. Las células se lavaron 1-2 veces con PBS a temperatura ambiente. Se utilizaron 200.000 células para cada transfección, y se resuspendieron en 100 µl de solución BTexpress. Se añadieron 2-4 µg de ARNm por transfección y la mezcla se transfirió a la cubeta. Inmediatamente después de la transferencia, la mezcla se sometió a electroporación a 250 V durante 5 ms. Se añadieron a la cubeta medios precalentados y los medios más las células se transfirieron a placas no tratadas con cultivo tisular de 48 pocillos y después se incubaron a 37 °C.

Después del número de días especificado, las células se sometieron a análisis genómico utilizando un Illumina MiSeq. Para cuantificar el porcentaje de alelos editados, la región genómica de interés se amplificó por PCR utilizando cebadores que añaden las secuencias del adaptador de secuenciación de Illumina estándar. Se realizó un segundo grupo de 13 rondas de PCR para añadir secuencias de código de barras y de adaptador de puente a ambos extremos. La secuenciación se realizó en un Illumina MiSeq según los protocolos del fabricante para la secuenciación de amplicones. El MiSeq genera lecturas de extremo emparejado, que se fusionan y se recortan del adaptador mediante un programa informático de alineación estándar. Después, las lecturas se desmultiplexaron por muestra a través de pares de secuencias de códigos de barras utilizando scripts personalizados. Las secuencias de amplicón se alinearon globalmente a una secuencia de referencia a través de una implementación del algoritmo Needleman-Wunsch (Needleman, Saul B.; y Wunsch, Christian D. (1970). *Jour Mol Bio* 48 (3): 443-53). Los huecos o las inserciones en la alineación se contaron como % de eventos de NHEJ y se compararon con una secuencia de muestra de control no tratada para determinar las tasas de fondo específicas de la secuencia.

Para el cálculo de la integración dirigida, las secuencias de amplicón se alinearon globalmente con una secuencia de referencia a través de una implementación por Biopython del algoritmo Needleman-Wunsch (Needleman, Saul B.; y Wunsch, Christian D. ibídem). Los cambios de secuencia generados a través de tratamientos experimentales se buscaron, se contaron y se compararon con los recuentos en las muestras de control. Los polimorfismos de una sola característica (SNP) conocidos pueden enmascarse durante este proceso y excluirse de otros recuentos (por ejemplo, SNP de eliminación de 1 pb cerca del sitio diana de ZFN). El % de NHEJ (también denominado indeles) se calculó determinando el porcentaje de secuencias que contienen inserciones o deleciones. Las muestras tratadas solo con el vector GFP se utilizaron para evaluar la PCR y la frecuencia de fondo de las inserciones y las deleciones basada en el error de secuenciación. Se observaron frecuencias de fondo inferiores al 1%.

Un conjunto de datos representativos se muestra a continuación en la tabla 1B y demostró que estas proteínas nucleasas son activas en la escisión de sus dianas. Además, se supervisó la expresión de gamma-globina en algunas de las células tratadas con nucleasa. Para realizar este análisis, se utilizó RT-qPCR en tiempo real ("Taqman") según el procedimiento estándar (véase más adelante). Los resultados de un conjunto de datos representativos se muestran como el aumento de la expresión de gamma-globina en comparación con las células de control tratadas con GFP. Los valores de gamma se calculan como una relación de gamma-globina con respecto a alfa-globina, por lo que cualquier aumento observado a continuación representa un aumento en la relación de gamma con respecto a alfa en las células tratadas con nucleasa en comparación con la relación de gamma con respecto a alfa en las células tratadas con el vector GFP.

Tabla 1B: Actividad de los pares de ZFN de exón 2 y exón 4 de BCL11A

Diana	Par ZFN	% de indeles, K562	% de indeles, CD34+	Veces de aumento en ARNm de gamma
Exón 2	39145/39172		69,78	3,65X
	39145/43490	19,88		nd
	39145/44642	38,52		nd

Diana	Par ZFN	% de indeles, K562	% de indeles, CD34+	Veces de aumento en ARNm de gamma
	39145/41548	42,26		nd
	39145/41547	35,63		nd
	44490/39172	29,38		nd
	44489/39172	24,34		nd
	45081/39172	27,80		nd
	44493/39172	25,68		nd
Exón 4	34678/34642		82,24	3,52X

También se produjeron TALEN en las regiones exón 2 y exón 4 de BCL11A. Las TALEN se construyeron tal como se ha descrito anteriormente, utilizando el código TALE canónico y la cadena principal '+17' TALEN (véase la publicación de patente de Estados Unidos en copropiedad N° 20110301073). La tabla 1C muestra la secuencia diana para las TALEN, así como la secuencia RVD en el dominio de unión a ADN.

5

Tabla 1C: pares de TALEN contra BCL11A

Número SBS (exón)	Secuencia diana 5' -> 3'	Secuencia RVD (N-> C)
101291 (exón 2)	ctGTGGGCAGTGCCAGATga (SEQ ID NO: 236)	NN-NG-NN-NN-NN-HD-NI-NN-NG-NN-HD- HD-NI-NN-NI-NG (SEQ ID NO:237)
101292 (exón 2)	ctCGATAAAAATAAGAATgt (SEQ ID NO: 238)	HD-NN-NI-NG-NI-NI-NI-NI-NI-NG-NI- NI-NN-NI-NI-NG (SEQ ID NO:239)
101301 (exón 4)	atGTCCTTCCCAGCCACCTct (SEQ ID NO: 240)	NN-NG-HD-HD-NG-NG-HD-HD-HD-NI-NN- HD-HD-NI-HD-HD-NG (SEQ ID NO:241)
101304 (exón 4)	gtTAAAGGGGTTATTGTct (SEQ ID NO: 242)	NG-NI-NI-NI-NN-NN-NN-NN-NG-NG-NI- NG-NG-NN-NG (SEQ ID NO:243)

10 Los pares de TALEN mostrados anteriormente se introdujeron en las células y mostraron actividad de escisión. El par 101291/101292 produjo un valor del 0,8% de indeles medido mediante el ensayo Cel-1 en células K562. El par de TALEN 101301/101304 dio un valor de formación de indeles del 35,7% en células CD34+, y se descubrió que mediante el ensayo de RT-PCR descrito anteriormente inducía un aumento en la expresión de ARNm de gamma-globina de aproximadamente 2,31 veces.

15 Los pares de ZFN también se produjeron para dirigirlos a la porción 'XL' de la variante de corte y empalme BCL11A-XL. Estas proteínas se analizaron en células K562 y a continuación se muestra un conjunto de datos representativos en la tabla 1D. La isoforma 'XL' de BCL11A contiene 3 dedos de zinc naturales adicionales (dedos 4-6), por lo que el enfoque adoptado implicó la interrupción del gen BCL11A en esta región para provocar el despliegue de los dedos potencialmente de zinc 4, 5 y/o 6 y combinaciones de los mismos (números 1 a 3 dentro de la región XL). Las ZFN también se diseñaron para evitar la escisión de la secuencia del gen BCL11B relacionado. Un par de ZFN, 44888/44889, se dirigió al cuarto dedo de zinc de BCL11A, mientras que dos pares 44904/44905 y 44910/44911 se dirigieron aguas arriba del cuarto dedo (número 1 dentro de la región XL) mientras que los otros dos pares, 44946/44947 y 44945/44944 se dirigieron al quinto dedo (número 2 dentro de la región XL). Estas proteínas se analizaron en células K562 y a continuación se muestra un conjunto de datos representativos en la tabla 1D. Dos de los pares de ZFN contenían nuevas secuencias de enlace entre el dominio de unión a ADN de ZFN y el dominio de nucleasa FokI. El par 44904/44905 contiene la secuencia de enlazador L7a (véase la solicitud de patente de Estados Unidos N° 20090305419) y el par 44946/44947 contenían la secuencia de enlazador L8p, que se muestran a continuación. Véase también la solicitud provisional de Estados Unidos N° 61/871.219:

25

L7a: HTKIHLRGSQVLVKSSEAAAR (SEQ ID NO: 244)

L8p: HTKIHLRGSYAPMPPLALASP (SEQ ID NO: 245).

5

Tabla 1D: Actividad de pares de ZFN específicos para BCL11A XL

Par de ZFN	% de indeles, K562
44889/44888	35,14
44905/44904	25,45
44911/44910	36,43
44945/44944	24,03
44947/44946	34,22

10 Los pares BCL11A XL se analizan en células CD34+ y están activos. La medición de la expresión de gamma-globina demuestra que la modificación de BCL11A XL produce un aumento de la expresión de gamma-globina con respecto a alfa-globina.

15 Se analizaron pares adicionales de ZFN específicos de KLF1 para determinar su actividad en células CD34+, y estas células se analizaron para detectar cualquier cambio en la expresión de gamma-globina. Un conjunto de datos representativos se muestra a continuación en la tabla 1E.

Tabla 1E: Actividad de pares de ZFN específicos de KLF

Diana	Par de ZFN	% de indeles, CD34+	Veces de aumento en ARNm gamma
KLF exón 1	36004/36021	44,4	2,2X
KLF exon 2	36071/36085	22,6	3,17X

20 Las relaciones de ARNm que codifican γ -globina y β -globina después del tratamiento de nucleasas específicas de BCL11A o KLF1 en HSC se determinaron en varios puntos temporales hasta 17 días después de la introducción de ZFN mediante análisis Taqman, y los niveles de ARNm de globina tipo beta también se normalizaron al nivel de ARNr 18S. Los niveles de expresión de gamma-globina aumentaron en aquellas células que habían sido tratadas con las nucleasas específicas de BCL11A o KLF1 (figura 5). El análisis se realizó mediante análisis estándar de Taqman, siguiendo el protocolo y utilizando ensayos específicos de genes suministrados por el fabricante (Applied Biosystems).

25

Las células BCL11A modificadas con ZFN también se analizaron para determinar las relaciones de ARNm γ/β entre poblaciones de células en las que un alelo se modificó por las ZFN ("Bb"), células en los ambos alelos se modificaron por las ZFN ("inactivados") y de tipo silvestre ("BB").

30

Tal como se muestra en la figura 6, las relaciones de ARNm γ/β son diferentes entre las células en las que se produjo la inactivación de BCL11A solo en un alelo (Bb, barras 6-10 desde la izquierda) o en las que ambos alelos se inactivaron (inactivación, más a la derecha 5 barras, barras 11-15 de la izquierda), y ambas agrupaciones de células difieren del tipo silvestre (BB, primeras 5 barras).

35

Ejemplo 7: Modificación de la región reguladora del gen de gamma-globina

En otro enfoque para aumentar la expresión de gamma-globina, se realizaron mutaciones en la región reguladora del gen de gamma-globina para imitar las mutaciones de HPFH (véase la figura 9). A continuación, se muestra la región de -202 a -102 con respecto al ATG en el gen de gamma-globina. En esta secuencia hay cuadros grises que indican áreas que se ha demostrado que están asociadas con HPFH, y una secuencia subrayada que, cuando se elimina, también se ha asociado con HPFH (véase *A Syllabus of Thalassemia Mutations (1997) por Titus H.J. Huisman, Marianne F.H. Carver y Erol Baysal, publicado por The Sickle Cell Anemia Foundation en Augusta, GA, Estados Unidos Copyright © 1997 por Titus H.J. Huisman*):

45

-202

CCCTTCCCACACTATCTCAATGCAAAATATGTCTCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGTT
GGCCTTGCCCTTGACCAATAGCCTTGAC(SEQ ID NO:132)

-102

5 Las nucleasas se diseñaron tal como se describe en el ejemplo 1 y se muestran en la tabla 1A para que se unan en la región de estas mutaciones asociadas a HPFH para inducir mutaciones en la región de tipo silvestre. El porcentaje de alelos editados detectados (% de NHEJ) en células K562 por análisis Cel I (véase Pérez et al. (2008), ibídem) se muestra a continuación en la tabla 2. Además, se sometieron a ensayo algunos pares en células CD34+ tal como se ha descrito anteriormente y se analizaron mediante secuenciación MiSeq tal como se ha descrito anteriormente. Para algunos pares, las células se analizaron para detectar cualquier cambio en la expresión de gamma-globina. La tabla 10 2 a continuación muestra conjuntos de datos representativos:

Tabla 2: Edición por pares de ZFN específicos de gammaglobina

Par de ZFN (ubicación)	% de NHEJ K562	% de NHEJ CD34+	Veces de aumento en ARNm gamma
34360/34363 (-175)	39		
34398/34400 (-175)	54		
31160/34365 (-175)	53	45,22	1,63X
34539/34574 (-110)		45,71	5,38X
43865/43852 (-110)		56,13	

15 Los primeros tres pares analizados en este ensayo se dirigieron a la región alrededor de -175 en la región promotora gamma, mientras que los dos últimos se dirigieron a la región -110 en el promotor de gamma-globina.

20 La región promotora gamma en las células K562 que había sido editada se secuenció para analizar las mutaciones creadas. La región se amplificó en primer lugar por PCR y después los productos de PCR se secuenciaron y se observaron varias mutaciones diferentes, incluidas deleciones e inserciones (figura 8). En este experimento, el 42% de los alelos estaban mutados y el 20% portaba la deleción de 13 pb de -114 a -102 asociada con HPFH.

25 También se utilizaron dos pares de ZFN dirigidos al promotor de gamma-globina para tratar las células en combinación con un donante de oligonucleótidos diseñado para recrear las mutaciones más comunes en sujetos con HPFH. Se siguió el mismo protocolo descrito anteriormente para su uso con el dispositivo BTX con la adición de 3 µl de una solución 100 µM del oligonucleótido donante. La secuencia de los donantes oligonucleotídicos se muestra a continuación. Típicamente, se utilizó en estos experimentos el donante oligonucleotídico directo, pero el donante 30 inverso también funcionó:

HBG_d13 directo :

acactatctcaatgcaaatatctgtctgaaacggtcctggctaaactccaccatg
ggttggccagccttgccctgacaaggcaacttgaccaatagtcttagagtatccag
tgaggccagg (124mer, SEQ ID NO:246)

HBG_d13_inverso :

cctggcctcactggatactctaagactattggtcaagtttgccctgtcaaggcaagg
ctggccaaccatgggtggagtttagccagggaccgtttcagacagatattgcat
gagatagtgt (124mer, SEQ ID NO:247)

35 Para el par de ZFN 34539/34574 en presencia del donante, la producción de ARNm del gen de gamma-globina aumentó 6,38 veces en comparación con las células tratadas con un vector GFP, mientras que para el par de ZFN 31160/34365, el ARNm de gamma aumentó en 6,13 veces en comparación a células tratadas con un vector GFP.

Las HSC tratadas con nucleasa se sembraron en metilcelulosa. Después de genotipar colonias individuales mediante secuenciación por PCR, medimos los niveles de ARNm para gamma-globina, beta-globina y el control de ARNr 18s

para colonias de tipo silvestre y mutadas mediante RT-PCR (figura 8). En promedio, los mutantes promotores de gamma-globina tenían una relación más alta de mensaje de gamma-globina a beta-globina que las células de tipo silvestre y la corrección por la señal de ARNr 18s indica que el aumento en la proporción de gamma-globina/beta-globina en las colonias mutadas viene provocado por un aumento en los niveles de ARNm de gamma-globina en estas colonias en lugar de una reducción de los niveles de ARNm de beta-globina.

Ejemplo 8: Nucleasas TALE dirigidas al promotor de la gammaglobina

También se produjeron nucleasas TALE que se dirigieran a la región -200 o la región -110 (descrita anteriormente) de la región promotora de gamma-globina. Las TALEN se construyeron tal como se ha descrito anteriormente, utilizando el código TALE canónico y la cadena principal TALEN '+17' (véase la publicación de patente de Estados Unidos en copropiedad 20110301073).

Tabla 3: TALEN específicas del promotor de gamma-globina

Número SBS	Secuencia 5'-> 3'	Secuencia RVD (N-> C)
102314	gtATCCTCTTGGGGGcc (SEQ ID NO:133)	NI-NG-HD-HD-NG-HD-NG-NG-NN-NN-NN-NN-NK (SEQ ID NO:134)
102318	atATTTGCATTGAGATAGT gt (SEQ ID NO:135)	NI-NG-NG-NG-NN-HD-NI-NG-NG-NN-NI-NN-NI- NG-NI-NN-NG (SEQ ID NO:136)
102315	gtATCCTCTTGGGGGcc (SEQ ID NO:133)	NI-NG-HD-HD-NG-HD-NG-NG-NN-NN-NN-NN-HD (SEQ ID NO:137)
102320	atATTTGCATTGAGATAgT (SEQ ID NO:135)	NI-NG-NG-NG-NN-HD-NI-NG-NG-NN-NI-NN-NI- NG-NI (SEQ ID NO:136)
102316	gtATCCTCTTGGGGGCCcc (SEQ ID NO:133)	NI-NG-HD-HD-NG-HD-NG-NG-NN-NN-NN-NN- HD-HD (SEQ ID NO:138)
102321	atATTTGCATTGAGATag (SEQ ID NO:135)	NI-NG-NG-NG-NN-HD-NI-NG-NG-NN-NI-NN-NI-NG (SEQ ID NO:139)
102566 (-110)	gtTGGCCAGCCTTGCCCTTG ac (SEQ ID NO:248)	NG-NN-NN-HD-HD-NI-NN-HD-HD-NG-NG-NN-HD- HD-NG-NG-NK (SEQ ID NO:249)
102568 (-110)	ttGGTCAAGTTTGCCTTGT ca (SEQ ID NO:250)	NN-NN-NG-HD-NI-NI-NN-NG-NG-NG-NN-HD-HD- NG-NG-NN-NG (SEQ ID NO:251)

Las TALEN se utilizaron después en pares para evaluar la escisión en células K562 y se analizaron mediante el ensayo Cel 1 tal como se ha descrito anteriormente y los resultados de los pares se muestran a continuación en la tabla 4. Además, el par de TALEN 102566/102568 se analizó frente a células CD34+ y se encontró que tenía el 51,39% de NHEJ medido por análisis MiSeq.

También se analizaron dos pares de las TALEN para determinar la expresión de ARNm de gamma-globina medida por la relación de ARNm de gamma-globina con respecto a alfa-globina. Se encontró que el par 102566/102568 aumentaba la expresión de gamma-globina en 6,25 veces en comparación con las células CD34+ tratadas con un vector GFP, y el par 102318/102314 aumentaba la gamma-globina en 2,14 veces en comparación con las células CD34+ tratadas con un vector GFP. El par 102566/102568 también se analizó con el oligo donante descrito anteriormente y se encontró que las células resultantes tenían un aumento en la expresión de gamma-globina de 9,13 veces en comparación con las células CD34+ tratadas con un vector GFP.

Tabla 4: Edición de la región promotora de gamma-globina con TALEN

Par de TALEN	% de NHEJ '+17'
102314:102318	41,6
102315:102320	47,9
102316:102321	46,6

5 **Ejemplo 9:** Edición de gamma-globina en células madre CD34+

Las nucleasas específicas para la región promotora de gamma-globina se utilizan después en células CD34+ derivadas del paciente. Las células se tratan con las nucleasas y después se analizan para una edición exitosa mediante el análisis Cel 1. Las células se analizan adicionalmente para examinar las relaciones de gamma-globina frente a beta-globina y demostrar una mayor expresión de gamma-globina. Los datos representativos encontrados para aumentar la expresión de la gammaglobina se encuentran en las secciones experimentales para los diferentes enfoques anteriores.

10

15 **Ejemplo 10:** Injerto de CD34+ editadas en ratones

Las células CD34+ tratadas con nucleasas (HSPC progenitoras de células madre humanas) conservaron su capacidad para injertar ratones NOD/SCID/IL2rgamma (nulo) y dar lugar a una progenie policlonal de múltiples linajes en la que los genes implicados en la regulación de la gamma-globina están alterados permanentemente (véase Holt et al., (2010) *Nat Biotechnol.* Agosto; 28 (8): 839-47). De forma similar, las CD34 + o HSPC editadas en el locus de beta-globina en el que se corrige una mutación, o se inserta un gen de beta-globina donante en un locus de puerto seguro, o se tratan con nucleasas para alterar la expresión de gamma-globina, pueden injertarse y dar lugar a la progenie de múltiples linajes que portan la edición del genoma deseada. La demostración de que una minoría de HSPC editadas puede poblar un animal con progenie editada respalda el uso de células madre hematopoyéticas autólogas modificadas con nucleasa como un enfoque clínico para tratar las hemoglobinopatías.

20

25

REIVINDICACIONES

- 5 1. Una célula precursora de glóbulos rojos genomanipulada caracterizada por una modificación genómica dentro del exón 2 o el exón 4 de BCL11A o dentro de BCL11A-XL realizada después de la escisión dentro del gen BCL11A por una nucleasa de dedo de zinc (ZFN), una nucleasa TALE (TALEN) o un sistema CRISPR/Cas que se une a un sitio diana dentro de cualquiera de las SEQ ID NO: 56, 63, 66, 71, 160, 170, 179, 183, 189, 193, 197, 200, 203, 207, 211 y 213 de forma que el gen BCL11A se inactiva y la expresión de gammaglobina se aumente.
- 10 2. La célula precursora de glóbulos rojos genomanipulada de la reivindicación 1, siendo la célula una célula madre hematopoyética.
3. La célula precursora de glóbulos rojos genomanipulada de la reivindicación 1, en la que:
- 15 (a) el gen BCL11A es escindido por una nucleasa de dedo de zinc;
- (b) la nucleasa se introduce en la célula como un polinucleótido; o
- (c) la modificación genómica comprende la integración de un polinucleótido donante que codifica un transgén.
- 20 4. La célula precursora de glóbulos rojos genomanipulada de la reivindicación 1, en la que la nucleasa de dedo de zinc comprende una proteína de dedo de zinc que comprende 5 o 6 dominios de dedo de zinc que comprenden una hélice de reconocimiento y además en la que la proteína de dedo de zinc comprende las regiones de hélice de reconocimiento en el orden mostrado en una fila individual de la tabla 1A de las proteínas designadas SBS#39172, SBS#43490, SBS#44642, SBS#45148, SBS#45147, SBS#39145, SBS#44490, SBS#44489, SBS#45081, SBS#44493, SBS#29527, SBS#29528, SBS#29525, SBS#29526, SBS#34678, SBS#34642, SBS#44889, SBS#44888, SBS#44905, SBS#44904, SBS#44911, SBS#44910, SBS#44945, SBS#44944, SBS#44947 o SBS#44946.
- 25 5. Una proteína de dedo de zinc que comprende 5 o 6 dominios de dedo de zinc que comprenden una región de hélice de reconocimiento, en la que la proteína de dedo de zinc comprende las regiones de hélice de reconocimiento en el orden que se muestra en una fila individual de la tabla 1A de las proteínas designadas SBS#39172, SBS#43490, SBS#44642, SBS#45148, SBS#45147, SBS#39145, SBS#44490, SBS#44489, SBS#45081, SBS#44493, SBS#29527, SBS#29528, SBS#29525, SBS#29526, SBS#34678, SBS#34642, SBS#44889, SBS#44888, SBS#44905, SBS#44904, SBS#44911, SBS#44910, SBS#44945, SBS#44944, SBS#44947 o SBS#44946.
- 30 6. Una proteína de fusión que comprende la proteína de dedo de zinc de la reivindicación 5 y un dominio de escisión de tipo silvestre o genomanipulado o un semidominio de escisión de tipo silvestre o genomanipulado.
- 35 7. Un polinucleótido que codifica una o más proteínas de la reivindicación 5 o la reivindicación 6.
- 40 8. Una célula aislada que comprende una o más proteínas de fusión de la reivindicación 6 o uno o más polinucleótidos de la reivindicación 7.
9. La célula de la reivindicación 8, en la que se selecciona la célula del grupo que consiste en un glóbulo rojo (eritrocito) o una célula precursora, tal como una célula madre hematopoyética CD34+.
- 45 10. Un kit que comprende la proteína de la reivindicación 5, la proteína de fusión de la reivindicación 6 o el polinucleótido de la reivindicación 7.
- 50 11. Un procedimiento *in vitro* para alterar la expresión del gen de globina en una célula, comprendiendo el procedimiento:
- introducir, en la célula, una o más proteínas de la reivindicación 5 o la reivindicación 6 o uno o más polinucleótidos de la reivindicación 7, en condiciones tales que se expresen una o más proteínas y se altere la expresión del gen de globina.
- 55 12. Las, una o más, proteínas de la reivindicación 5 o la reivindicación 6 o uno o más polinucleótidos de la reivindicación 7 para su uso en un procedimiento de tratamiento de una hemoglobinopatía, comprendiendo el procedimiento:
- 60 introducir, en la célula, las, una o más, proteínas de la reivindicación 5 o la reivindicación 6 o uno o más polinucleótidos de la reivindicación 7, en condiciones tales que se expresen una o más proteínas y se altere la expresión del gen de globina.

13. El procedimiento de la reivindicación 11, en el que:

(a) las proteínas aumentan la expresión del gen de globina, tal como un gen de gamma-globina o beta-globina;

5 (b) el procedimiento comprende además integrar una secuencia donante en el genoma de la célula, por ejemplo, introduciendo la secuencia donante en la célula utilizando un vector vírico, un oligonucleótido o un plásmido;

(c) la célula es una célula precursora de glóbulos rojos (eritrocitos), tal como una célula madre hematopoyética CD34+; y/o

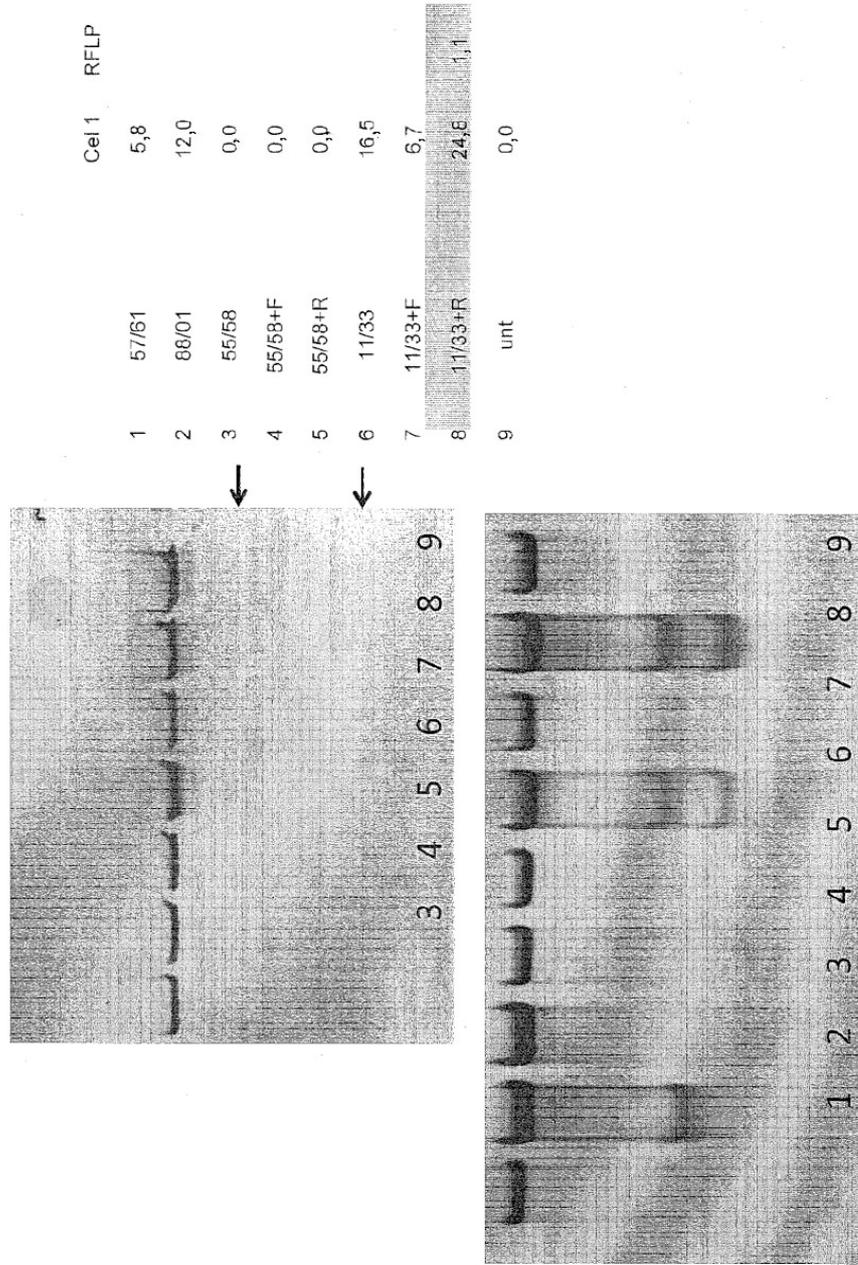
10 (d) la secuencia donante comprende un transgén bajo el control de un promotor endógeno o un promotor exógeno.

14. Una o más proteínas o uno o más polinucleótidos para su uso según la reivindicación 12, siendo la hemoglobinopatía una talasemia.

15 15. Las, una o más, proteínas de la reivindicación 5 o la reivindicación 6 o uno o más polinucleótidos de la reivindicación 7 para su uso en un procedimiento para tratar la enfermedad de células falciformes, comprendiendo el procedimiento:

20 introducir, en la célula, las, una o más, proteínas de la reivindicación 5 o la reivindicación 6 o uno o más polinucleótidos de la reivindicación 7, en condiciones tales que se expresen una o más proteínas y se altere la expresión del gen de globina.

FIGURA 2



A. RFLP

B. CEL-I

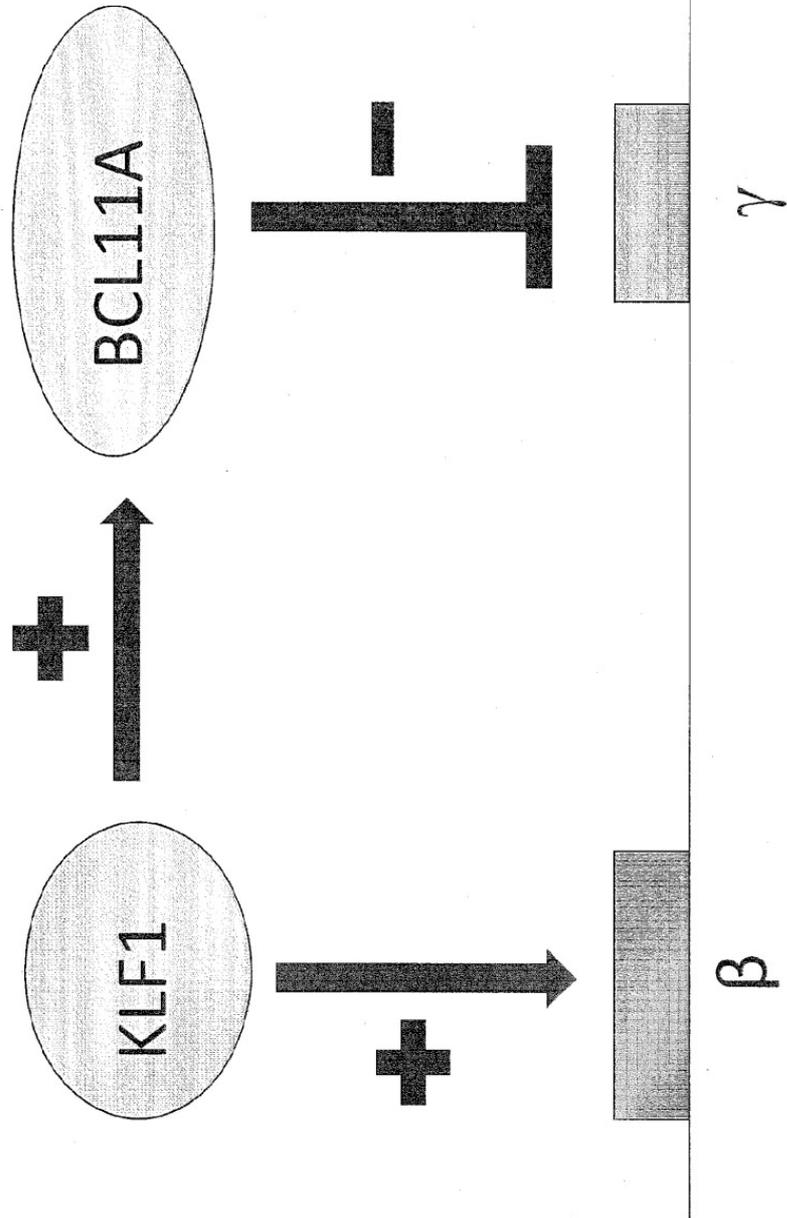


Figura 3

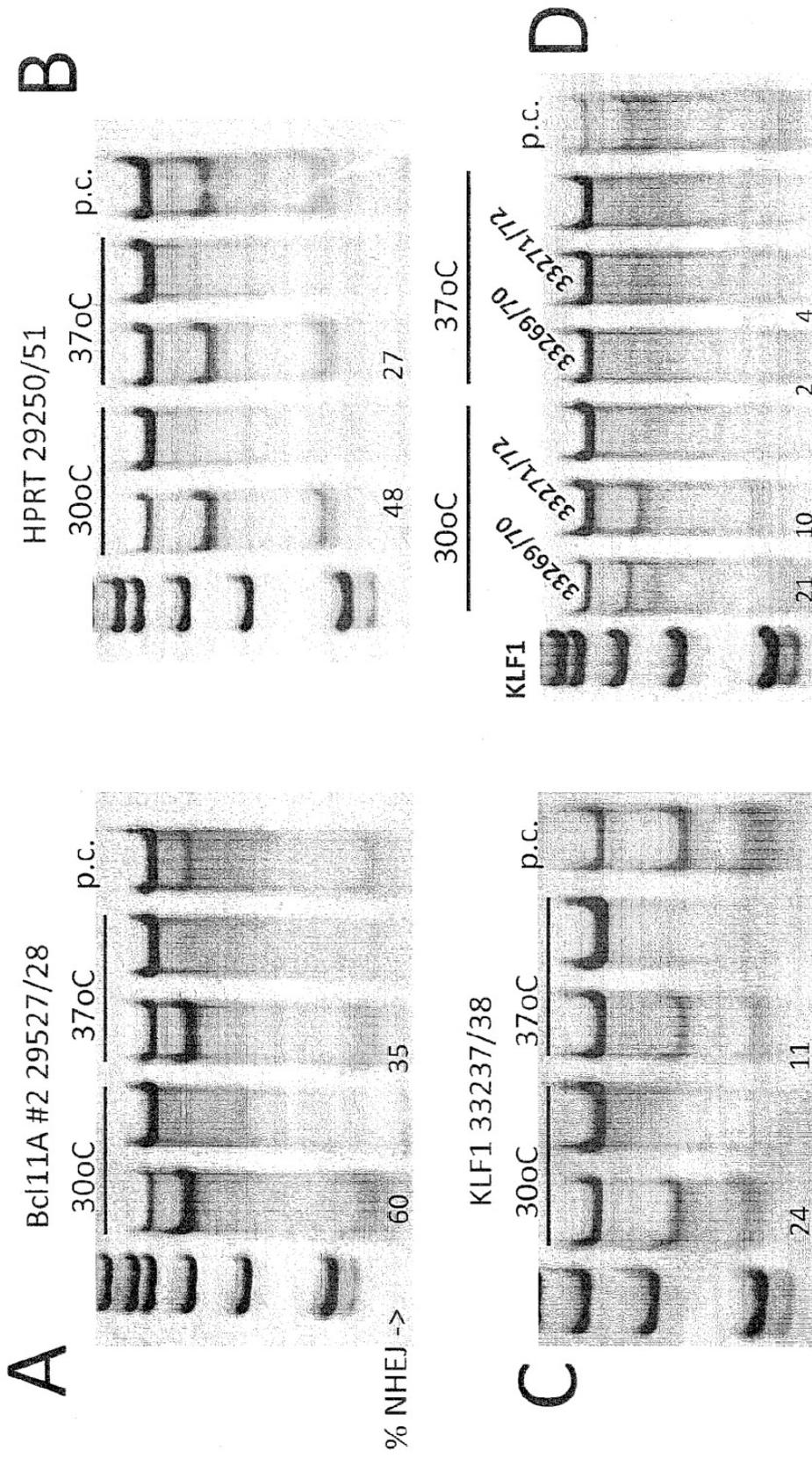


Figura 4

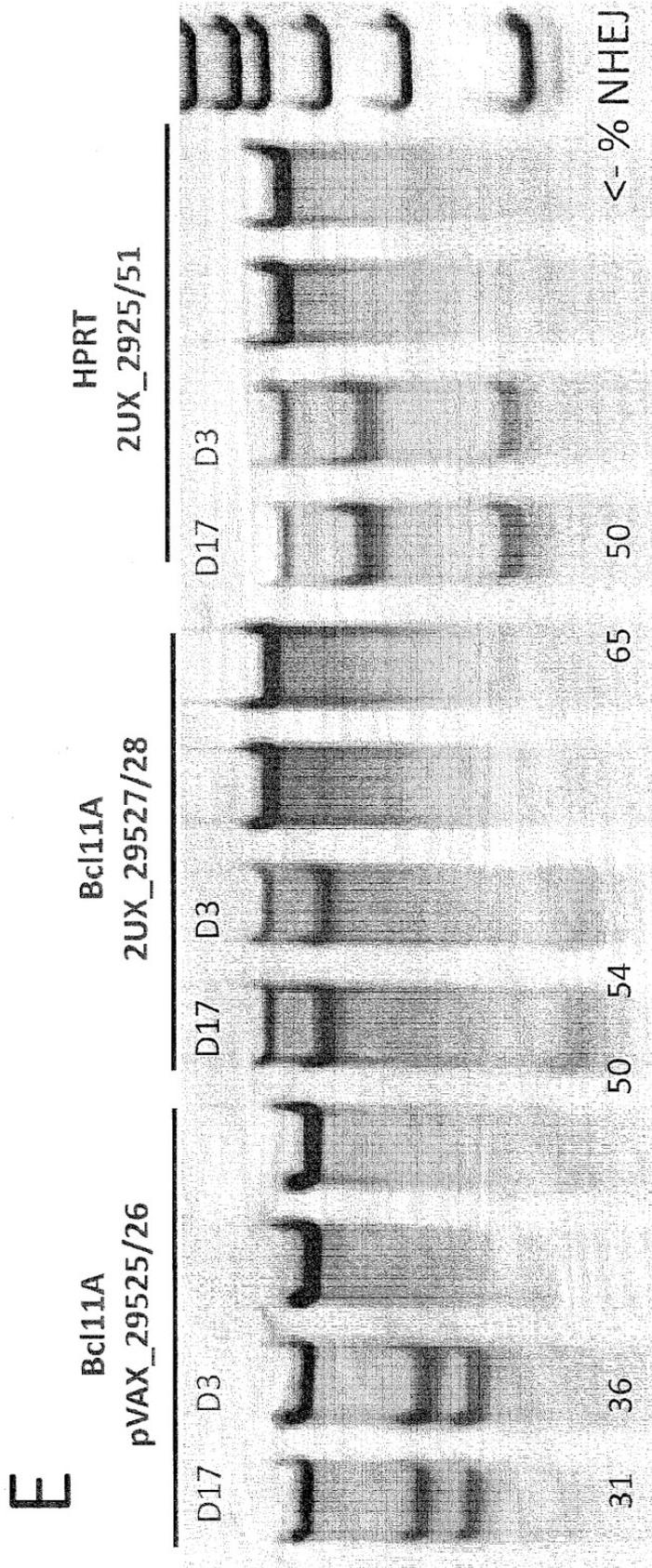


Figura 4

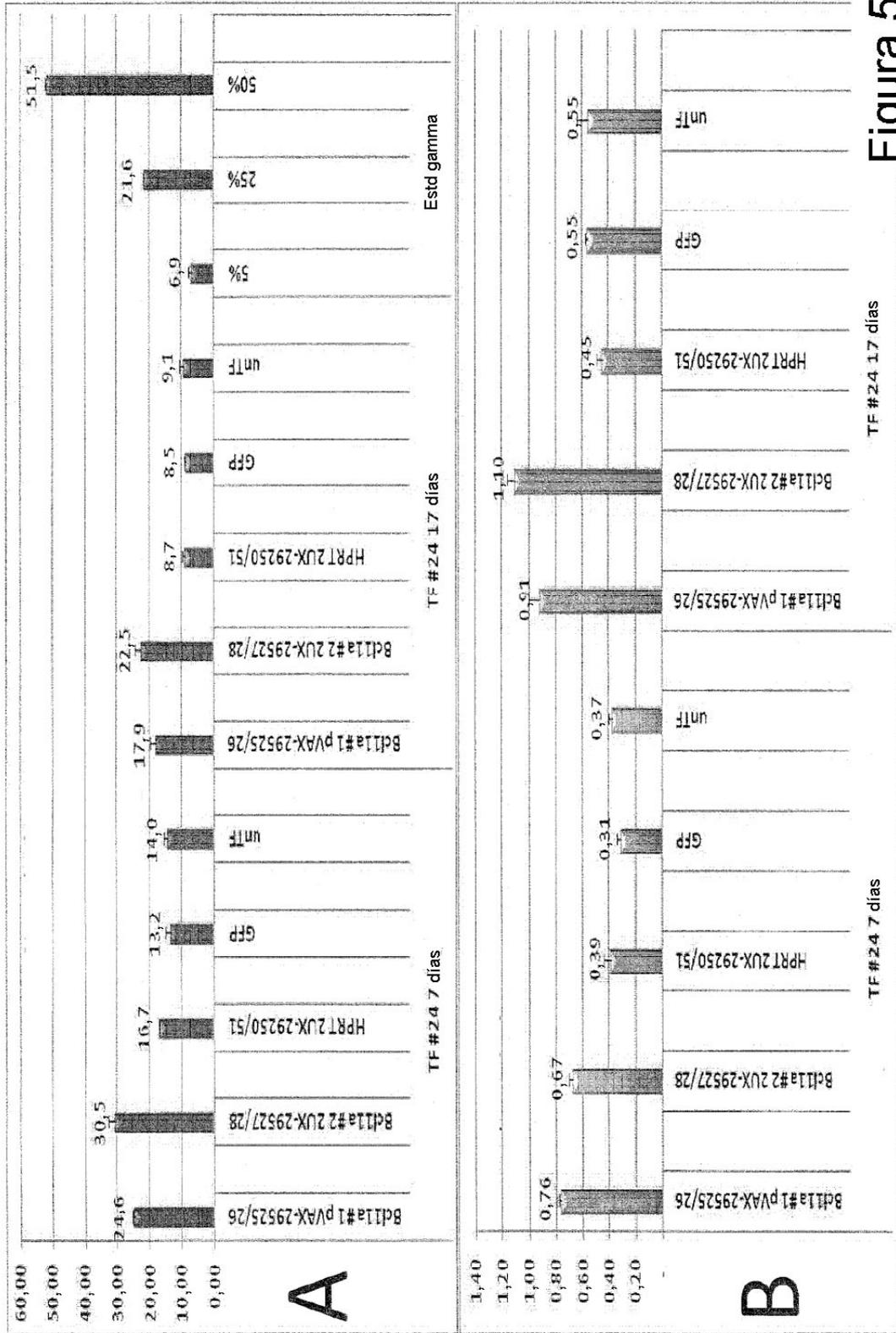
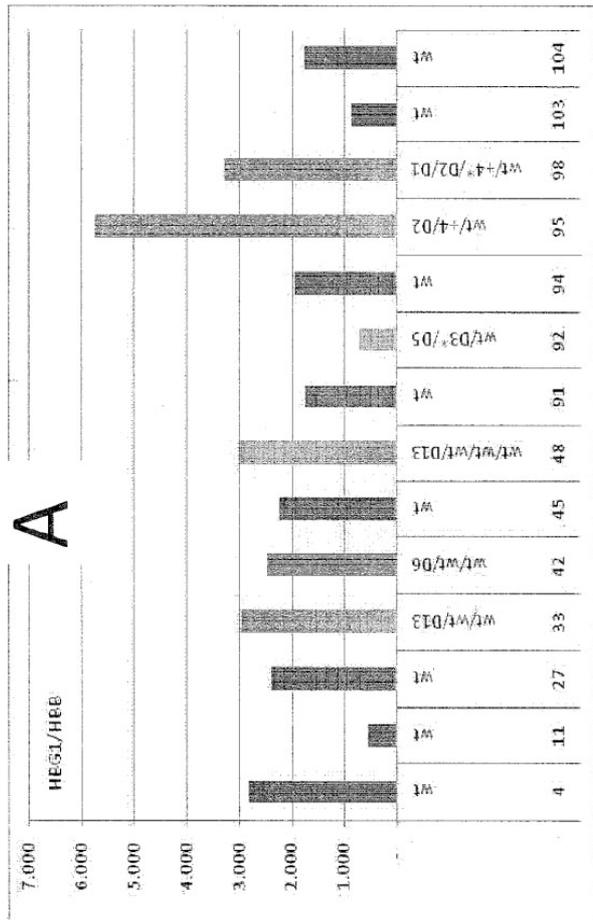


Figura 5

Referencia	GGCCAGCCCTTGCCCTTGACC <u>CAATAGCCCTTGACAAGGCCAAACTTGACC</u> CAATAG
wt (26)	GGCCAGCCCTTGCCCTTGACC <u>CAATAGCCCTTGACAAGGCCAAACTTGACC</u> CAATAG
Δ 1pb (1)	GGCCAGCCCTTGCCCTTGACC- <u>ATAGCCCTTGACAAGGCCAAACTTGACC</u> CAATAG
Δ 4pb (1)	GGCCAGCCCTTGCCCTT----- <u>AAATAGCCCTTGACAAGGCCAAACTTGACC</u> CAATAG
Δ 6pb (2)	GGCCAGCCCTTGCCCTTGAC----- <u>CCTTGACAAGGCCAAACTTGACC</u> CAATAG
Δ 6pb (1)	GGCCAGCCCTTGCCCTTGACC <u>CAATAGC</u> ----- <u>AAAGGCCAAACTTGACC</u> CAATAG
Δ 13pb (9)	GGCCAGCCCTTGCCCTTGAC----- <u>AAAGGCCAAACTTGACC</u> CAATAG
+1pb (1)	GGCCAGCCCTTGCCCTTGACC <u>CAATA</u> TAGCCCTTGACAAGGCCAAACTTGACC <u>CAATA</u>
+3pb (1)	GGCCAGCCCTTGCCCTTGACC <u>CAATAGCAGCCTTGACAAGGCCAAACTTGACC</u> CAA
+5pb (3)	GGCCAGCCCTTGCCCTTGACC <u>CAATAGAA</u> TAGCCCTTGACAAGGCCAAACTTGACC

Figura 7



promedio wt 1,62
promedio mut 2,01

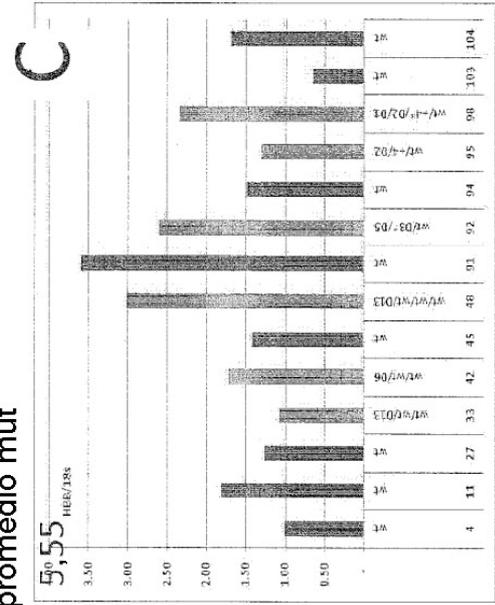
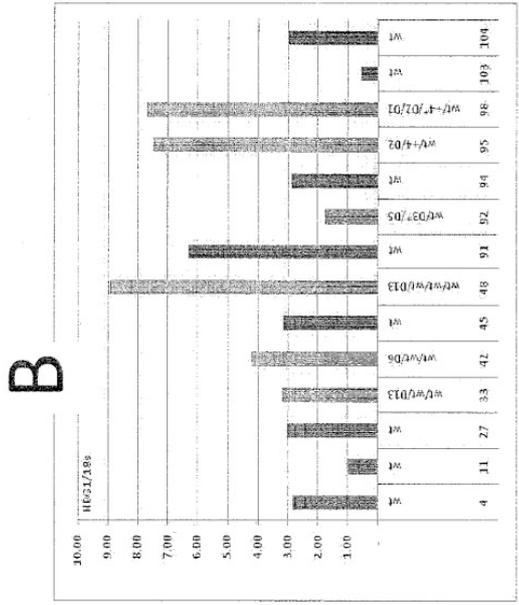
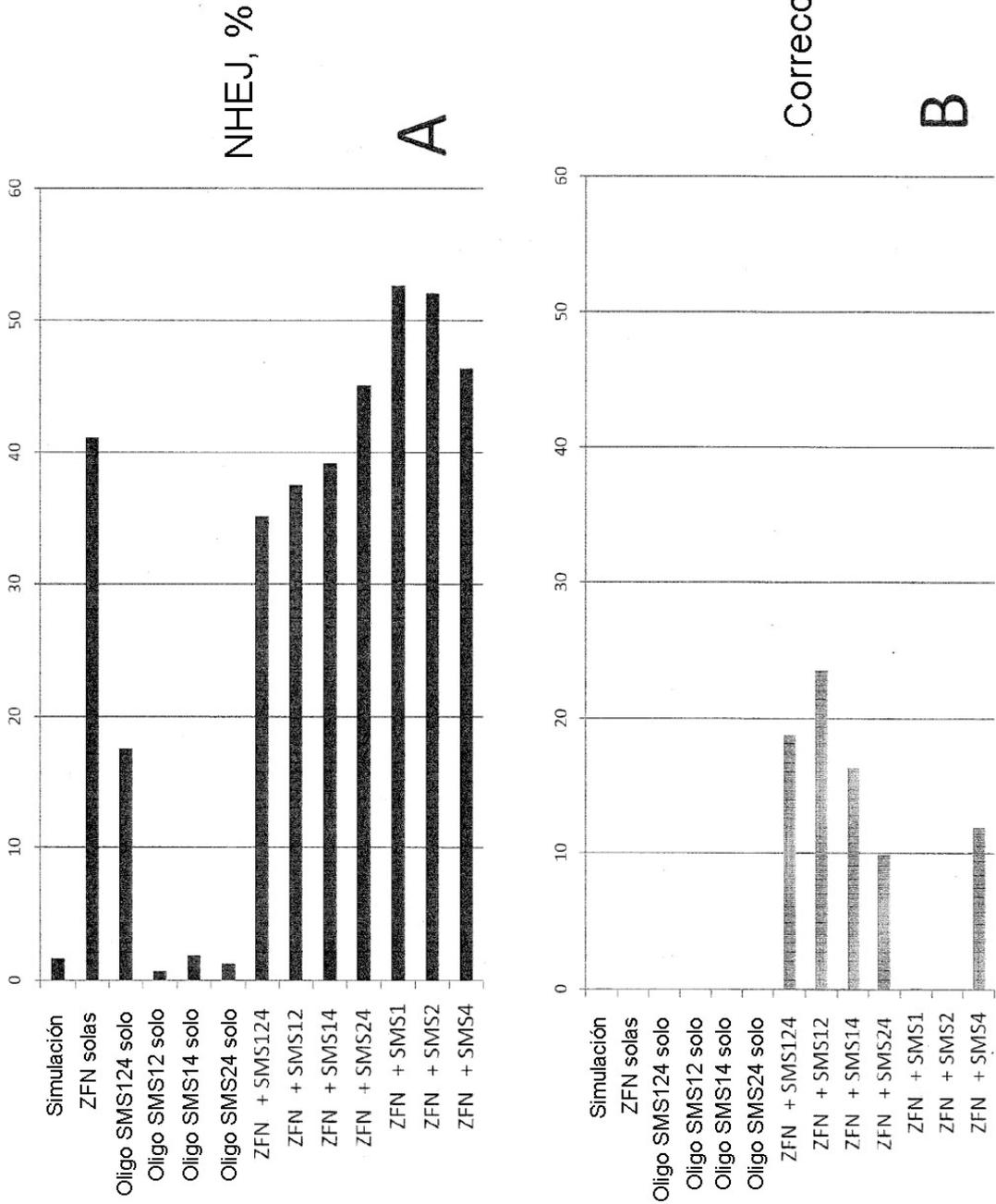


Figura 8

Figura 10



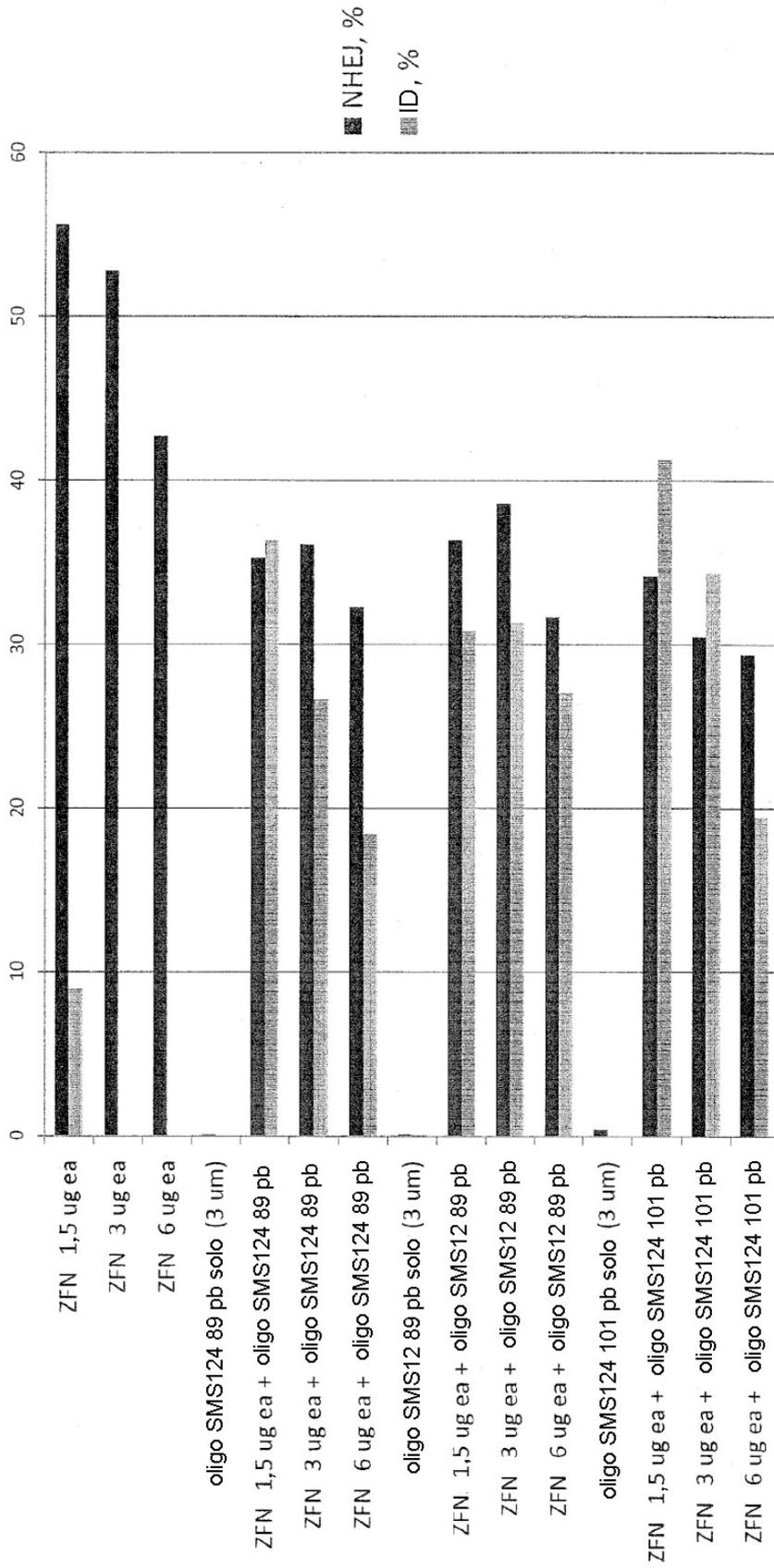


Figura 11

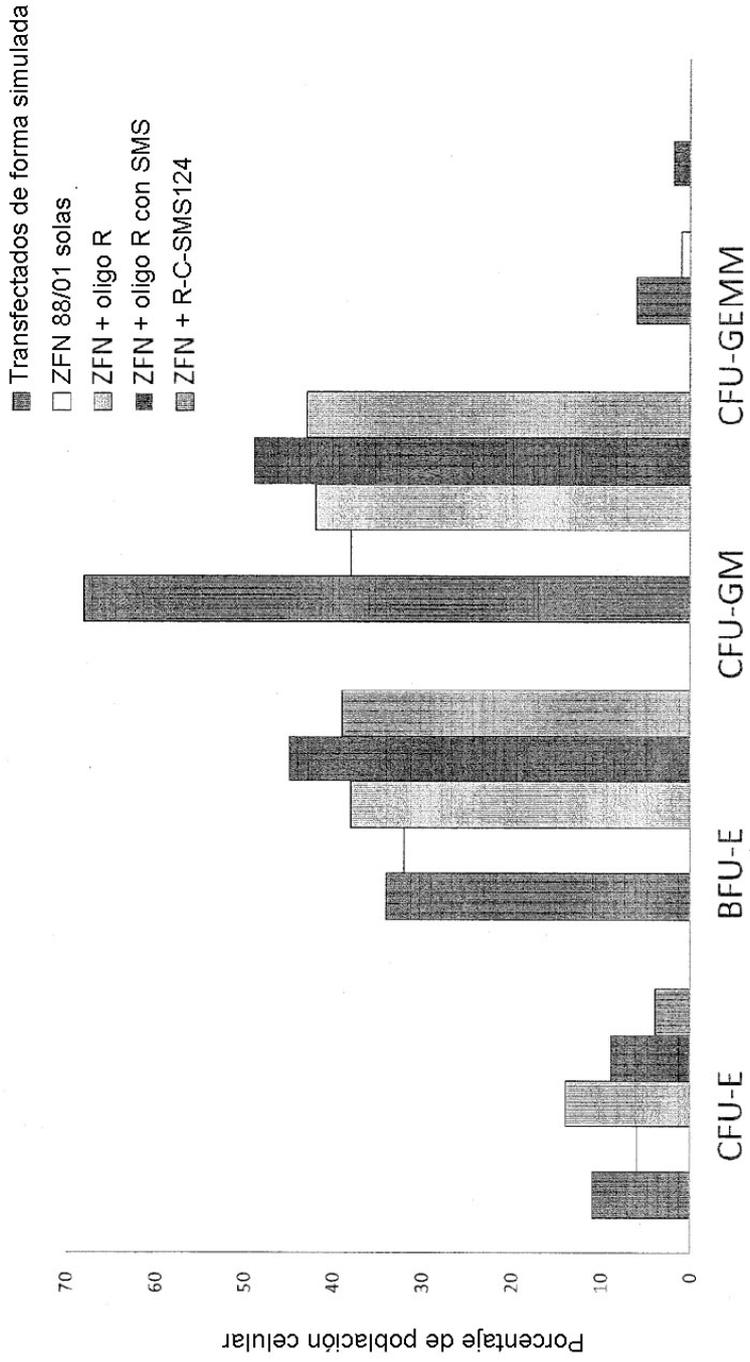


Figura 12

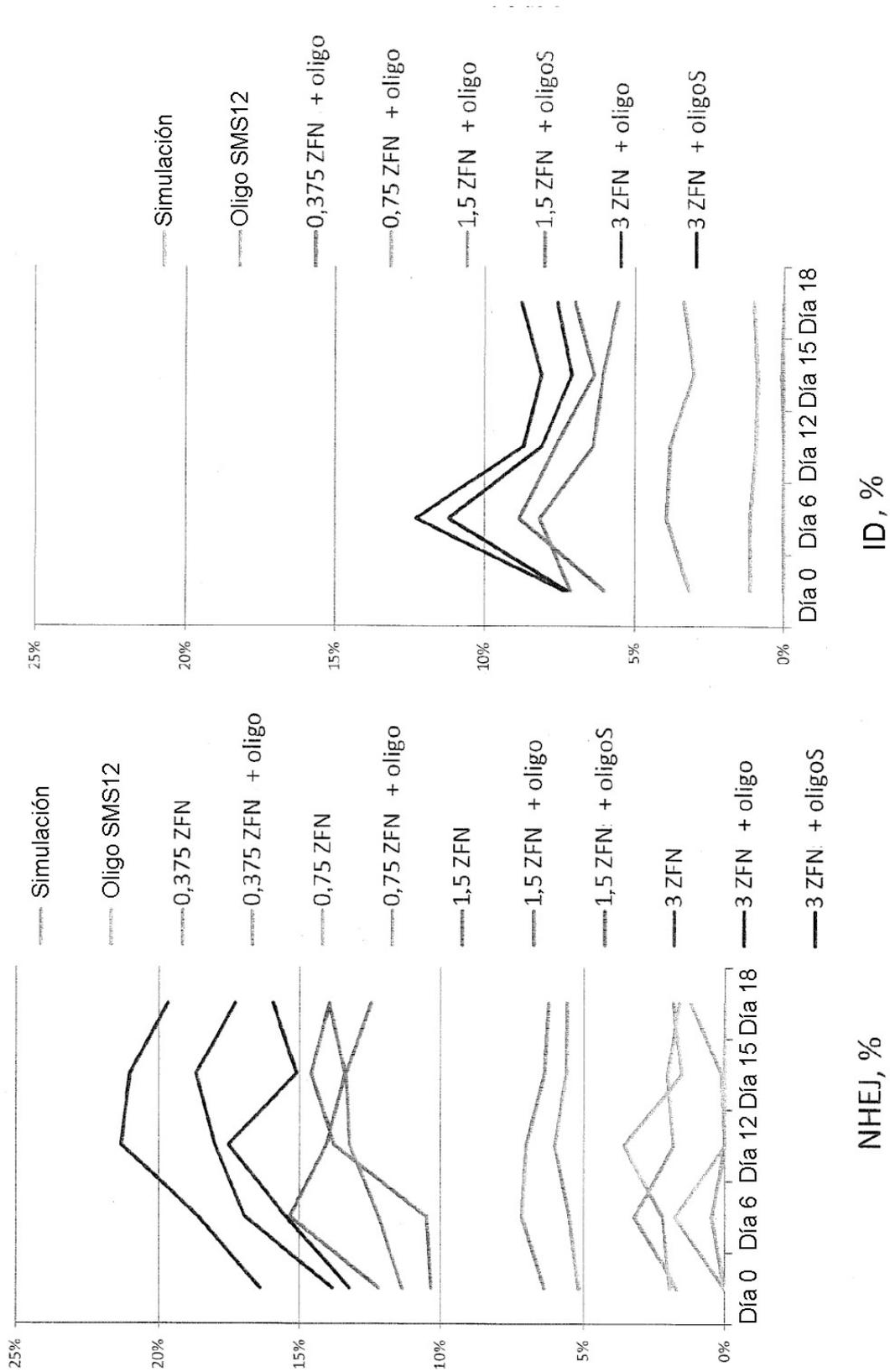


Figura 13