

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 979 222**

51 Int. Cl.:

**A61K 38/46** (2006.01)

**C12N 9/22** (2006.01)

**C12N 15/10** (2006.01)

**C12N 15/113** (2010.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **18.04.2017 PCT/IB2017/000577**

87 Fecha y número de publicación internacional: **26.10.2017 WO17182881**

96 Fecha de presentación y número de la solicitud europea: **18.04.2017 E 17727373 (7)**

97 Fecha y número de publicación de la concesión europea: **17.04.2024 EP 3445388**

54 Título: **Materiales y métodos para el tratamiento de hemoglobinopatías**

30 Prioridad:

**18.04.2016 US 201662324024 P**

**01.09.2016 US 201662382522 P**

**02.12.2016 US 201662429428 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**24.09.2024**

73 Titular/es:

**VERTEX PHARMACEUTICALS INCORPORATED**  
**(100.0%)**

**50 Northern Avenue**  
**Boston, MA 02210, US**

72 Inventor/es:

**COWAN, CHAD, ALBERT;**  
**LUNDBERG, ANTE, SVEN;**  
**CHAKRABORTY, TIRTHA;**  
**LIN, MICHELLE, I-CHING;**  
**MISHRA, BIBHU, PRASAD;**  
**PAIK, ELIZABETH, JAE-EUN;**  
**KERNYTSKY, ANDREW y**  
**BORLAND, TODD, DOUGLASS**

74 Agente/Representante:

**ARIAS SANZ, Juan**

### Observaciones:

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

ES 2 979 222 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

Materiales y métodos para el tratamiento de hemoglobinopatías

5 **Campo**

La presente descripción proporciona materiales y métodos para tratar a pacientes con hemoglobinopatías, tanto *ex vivo* como *in vivo*. Además, la presente descripción proporciona materiales y métodos para delecionar, modular o inactivar una secuencia de control transcripcional de un gen de linfoma de células B 11A (BCL11A) en una célula mediante edición genómica.

**Lista de secuencias**

Esta patente contiene una lista de secuencias en formato legible por ordenador (nombre del archivo: 160077PCT Sequence Listing; 14.446.299 bytes – archivo de texto ASCII; creado el 7 de abril de 2017).

**Antecedentes**

Las hemoglobinopatías incluyen anemias de origen genético, que dan como resultado una producción reducida y/o destrucción aumentada de glóbulos rojos. Estos trastornos también incluyen defectos genéticos, que dan como resultado la producción de hemoglobinas anómalas con una incapacidad asociada de mantener la concentración de oxígeno. Muchos de estos trastornos se denominan  $\beta$ -hemoglobinopatías debido a que no logran producir proteína  $\beta$ -globina normal en cantidades suficientes o no logran producir en absoluto proteína  $\beta$ -globina normal. Por ejemplo, las  $\beta$ -talasemias resultan de un defecto parcial o completo en la expresión del gen de  $\beta$ -globina, conduciendo a hemoglobina de adulto deficiente o ausente (HbA). La anemia de células falciformes resulta de una mutación puntual en el gen estructural de  $\beta$ -globina, conduciendo a la producción de una hemoglobina anómala (HbS) (Atweh, Semin. Hematol. 38(4):367-73 (2001)). Las hemoglobinopatías dan como resultado una reducción de la capacidad de transporte de oxígeno de la sangre, que puede conducir a síntomas tales como cansancio, mareo y dificultad para respirar, particularmente al hacer ejercicio.

Para pacientes con diagnóstico de hemoglobinopatía, actualmente solo están disponibles unos pocos tratamientos sintomáticos, tales como una transfusión de sangre, para aumentar los niveles de oxígeno en sangre.

La ingeniería genómica se refiere a las estrategias y técnicas para la modificación específica dirigida de la información genética (genoma) de organismos vivos. La ingeniería genómica es un campo de investigación muy activo debido a la alta gama de aplicaciones posibles, particularmente en los sectores de la salud humana; la corrección de un gen que porta una mutación perjudicial, por ejemplo, o para explorar la función de un gen. Las tecnologías iniciales desarrolladas para insertar un transgén en una célula viva estaban con frecuencia limitadas por la naturaleza aleatoria de la inserción de la nueva secuencia en el genoma. Las inserciones aleatorias en el genoma pueden dar como resultado la perturbación de la regulación normal de genes contiguos, conduciendo a graves efectos no deseados. Además, las tecnologías de integración aleatoria ofrecen poca reproducibilidad, ya que no hay ninguna garantía de que la secuencia se inserte en el mismo lugar en dos células diferentes. Recientes estrategias de ingeniería genómica, tales como ZFN, TALEN, HE y MegaTAL, permiten modificar una zona específica del ADN, aumentando de ese modo la precisión de la corrección o inserción en comparación con tecnologías iniciales. Estas plataformas más recientes ofrecen un grado mucho mayor de reproducibilidad, pero todavía tienen sus limitaciones.

A pesar de esfuerzos por parte de investigadores y profesionales médicos en todo el mundo que han estado intentando tratar hemoglobinopatías, todavía existe una necesidad crítica de desarrollar tratamientos seguros y eficaces para hemoglobinopatías.

**Sumario**

La presente divulgación presenta un enfoque para abordar la base genética de hemoglobinopatías. Usando herramientas de ingeniería genómica para crear cambios permanentes en el genoma que pueden delecionar, modular o inactivar una secuencia de control transcripcional del gen de BCL11A con un único tratamiento, la terapia resultante puede mejorar los efectos de hemoglobinopatías.

La presente invención se define mediante las reivindicaciones adjuntas.

En el presente documento se proporcionan métodos celulares, *ex vivo* e *in vivo* para crear cambios permanentes en el genoma mediante delección, modulación o inactivación de una secuencia de control transcripcional del gen de BCL11A, que pueden usarse para tratar hemoglobinopatías. En el presente documento también se proporcionan componentes, kits y composiciones para realizar tales métodos. También se describen células producidas mediante tales métodos. Ejemplos de hemoglobinopatías pueden ser anemia de células falciformes y talasemia ( $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\gamma$  y combinaciones de las mismas).

En el presente documento se proporciona un método para editar un gen de linfoma de células B 11A (BCL11A) en una célula humana mediante edición genómica, comprendiendo el método la etapa de introducir en la célula humana una o más ácido desoxirribonucleico (ADN) endonucleasas para realizar una o más roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB), dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A, que da como resultado una delección, modulación o inactivación permanente de una secuencia de control transcripcional del gen de BCL11A. La secuencia de control transcripcional puede estar ubicada dentro de un segundo intrón del gen de BCL11A. La secuencia de control transcripcional puede estar ubicada dentro de un sitio hipersensible de ADN (DHS) +58 del gen de BCL11A.

También se describe en el presente documento un método *ex vivo* para tratar a un paciente (por ejemplo, un ser humano) con una hemoglobinopatía, comprendiendo el método las etapas de: crear una célula madre pluripotente inducida (iPSC) específica de paciente; editar dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A de la iPSC; diferenciar la iPSC sometida a edición genómica para dar una célula progenitora hematopoyética; e implantar la célula progenitora hematopoyética en el paciente.

La etapa de crear una célula madre pluripotente inducida (iPSC) específica de paciente puede comprender: aislar una célula somática a partir del paciente; e introducir un conjunto de genes asociados con pluripotencia en la célula somática para inducir la célula somática para que se convierta en una célula madre pluripotente. La célula somática puede ser un fibroblasto. El conjunto de genes asociados con pluripotencia pueden ser uno o más de los genes seleccionados del grupo que consiste en OCT4, SOX2, KLF4, Lin28, NANOG y cMYC.

La etapa de editar dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A de la iPSC puede comprender introducir en la iPSC una o más ácido desoxirribonucleico (ADN) endonucleasas para realizar una o más roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB) dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que da como resultado una delección, modulación o inactivación permanente de una secuencia de control transcripcional del gen de BCL11A.

La etapa de diferenciar la iPSC sometida a edición genómica para dar una célula progenitora hematopoyética puede comprender uno o más de lo siguiente: tratamiento con una combinación de moléculas pequeñas, suministro de factores de transcripción (por ejemplo, factores de transcripción maestros) o suministro de ARNm que codifica para factores de transcripción (por ejemplo, factores de transcripción maestros).

La etapa de implantar la célula progenitora hematopoyética en el paciente puede comprender implantar la célula progenitora hematopoyética en el paciente mediante trasplante, inyección local, infusión sistémica o combinaciones de los mismos.

También se describe en el presente documento un método *ex vivo* para tratar a un paciente (por ejemplo, un ser humano) con una hemoglobinopatía, comprendiendo el método las etapas de: aislar una célula madre mesenquimatosa a partir del paciente; editar dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A de la célula madre mesenquimatosa; diferenciar la célula madre mesenquimatosa sometida a edición genómica para dar una célula progenitora hematopoyética; e implantar la célula progenitora hematopoyética en el paciente.

La célula madre mesenquimatosa puede aislarse a partir de la médula ósea o sangre periférica del paciente. La etapa de aislar una célula madre mesenquimatosa a partir del paciente puede comprender la aspiración de médula ósea y el aislamiento de células mesenquimatosas usando medios de centrifugación por gradiente de densidad.

La etapa de editar dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A de la célula madre mesenquimatosa puede comprender introducir en la célula madre mesenquimatosa una o más ácido desoxirribonucleico (ADN) endonucleasas para realizar una o más roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB) dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que da como resultado una delección, modulación o inactivación permanente de una secuencia de control transcripcional del gen de BCL11A.

La etapa de diferenciar la célula madre mesenquimatosa sometida a edición genómica para dar una célula progenitora hematopoyética puede comprender uno o más de lo siguiente: tratamiento con una combinación de moléculas pequeñas, suministro de factores de transcripción (por ejemplo, factores de transcripción maestros) o suministro de ARNm que codifica para factores de transcripción (por ejemplo, factores de transcripción maestros).

La etapa de implantar la célula progenitora hematopoyética en el paciente puede comprender implantar la célula progenitora hematopoyética en el paciente mediante trasplante, inyección local, infusión sistémica o combinaciones de los mismos.

También se describe en el presente documento un método *ex vivo* para tratar a un paciente (por ejemplo, un ser

humano) con una hemoglobinopatía, comprendiendo el método las etapas de: aislar una célula progenitora hematopoyética a partir del paciente; editar dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A de la célula progenitora hematopoyética; e implantar la célula progenitora hematopoyética sometida a edición genómica en el paciente.

El método puede comprender además tratar al paciente con factor estimulante de colonias de granulocitos (GCSF) antes de la etapa de aislar una célula progenitora hematopoyética a partir del paciente. La etapa de tratar al paciente con factor estimulante de colonias de granulocitos (GCSF) puede realizarse en combinación con plerixafor.

La etapa de aislar una célula progenitora hematopoyética a partir del paciente puede comprender aislar células CD34+.

La etapa de editar dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A de la célula progenitora hematopoyética puede comprender introducir en la célula progenitora hematopoyética una o más ácido desoxirribonucleico (ADN) endonucleasas para realizar una o más roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB) dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que da como resultado una delección, modulación o inactivación permanente de una secuencia de control transcripcional del gen de BCL11A.

La etapa de implantar la célula progenitora hematopoyética sometida a edición genómica en el paciente puede comprender implantar la célula progenitora hematopoyética sometida a edición genómica en el paciente mediante trasplante, inyección local, infusión sistémica o combinaciones de los mismos.

También se describe en el presente documento un método *in vivo* para tratar a un paciente (por ejemplo, un ser humano) con una hemoglobinopatía, comprendiendo el método la etapa de editar un gen de BCL11A en una célula del paciente.

La etapa de editar un gen de BCL11A en una célula del paciente puede comprender introducir en la célula una o más ácido desoxirribonucleico (ADN) endonucleasas para realizar una o más roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB) dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que da como resultado una delección, modulación o inactivación permanente de un control transcripcional del gen de BCL11A. La célula puede ser una médula ósea célula, una célula progenitora hematopoyética, una célula CD34+ o combinaciones de las mismas.

La una o más ADN endonucleasas pueden ser una endonucleasa Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (también conocida como Csn1 y Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 o Cpf1; un homólogo de las mismas, una recombinación de la molécula que se produce de manera natural de las mismas, optimización de codones de las mismas o versiones modificadas de las mismas y combinaciones de las mismas.

El método puede comprender introducir en la célula uno o más polinucleótidos que codifican para la una o más ADN endonucleasas. El método puede comprender introducir en la célula uno o más ácidos ribonucleicos (ARN) que codifican para la una o más ADN endonucleasas. El uno o más polinucleótidos o uno o más ARN pueden ser uno o más polinucleótidos modificados o uno o más ARN modificados. La una o más ADN endonucleasas pueden ser una o más proteínas o polipéptidos. La una o más proteínas o polipéptidos pueden estar flanqueados en el extremo N-terminal, el extremo C-terminal o tanto el extremo N-terminal como el extremo C-terminal por una o más señales de localización nuclear (NLS). La una o más proteínas o polipéptidos pueden estar flanqueados por dos NLS, una NLS ubicada en el extremo N-terminal y la segunda NLS ubicada en el extremo C-terminal. La una o más NLS pueden ser una NLS de SV40.

El método puede comprender además introducir en la célula uno o más ácidos ribonucleicos guía (ARNg). El uno o más ARNg pueden ser ARN guía de molécula simple (ARNgs). El uno o más ARNg o uno o más ARNg pueden ser uno o más ARNg modificados, uno o más ARNg modificados o combinaciones de los mismos. El uno o más ARNg modificados pueden comprender tres residuos 2'-O-metil-fosforotioato en o cerca de cada uno de sus extremos 5' y 3'. El ARNg modificado puede ser la secuencia de ácido nucleico de SEQ ID NO: 71.959. La una o más ADN endonucleasas pueden complejarse previamente con uno o más ARNg, uno o más ARNg o combinaciones de los mismos para formar una o más ribonucleoproteínas (RNP). La razón en peso de ARNg con respecto a ADN endonucleasa en la RNP puede ser de 1:1. El ARNg puede comprender la secuencia de ácido nucleico de SEQ ID NO: 71.959, la ADN endonucleasa puede ser una Cas9 de *S. pyogenes* que comprende una NLS de SV40 en el extremo N-terminal y una NLS de SV40 en el extremo C-terminal, y la razón en peso de ARNg con respecto a ADN endonucleasa puede ser de 1:1. El documento US2015044772 que se refiere a una proteína Cas de fusión quimérica da a conocer la secuencia de ácido nucleico de SEQ ID NO: 71.959. Dicha secuencia también se da a conocer en el documento WO2015183025 que se refiere a un método para analizar un genotipo usando un sistema de CRISPR/Cas.

El método puede comprender además introducir en la célula un molde donante de polinucleótido que comprende un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada.



El método puede comprender además introducir en la célula un ácido ribonucleico guía (ARNg) y un molde donante de polinucleótido que comprende un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada. La una o más ADN endonucleasas pueden ser una o más endonucleasas Cas9 o Cpf1 que realizan una rotura de cadena sencilla (SSB) o rotura de cadena doble (DSB) en un locus dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que facilita la inserción de una nueva secuencia a partir del molde donante de polinucleótido en el ADN cromosómico en el locus que da como resultado una inserción, modulación o inactivación permanente de la secuencia de control transcripcional del ADN cromosómico proximal con respecto al locus. El ARNg puede comprender una secuencia espaciadora que es complementaria a un segmento del locus. Proximal puede significar nucleótidos tanto en sentido de 5' como en sentido de 3' del locus.

El método puede comprender además introducir en la célula uno o más ácidos ribonucleicos guía (ARNg) y un molde donante de polinucleótido que comprende un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada. La una o más ADN endonucleasas pueden ser una o más endonucleasas Cas9 o Cpf1 que realizan o crean un par de roturas de cadena sencilla (SSB) y/o roturas de cadena doble (DSB), la primera rotura en un locus en 5' y la segunda rotura en un locus en 3', dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A, que facilita la inserción de una nueva secuencia a partir del molde donante de polinucleótido en el ADN cromosómico entre el locus en 5' y el locus en 3' que da como resultado una inserción, modulación o inactivación permanente de la secuencia de control transcripcional del ADN cromosómico entre el locus en 5' y el locus en 3'. Un ARNg guía puede crear un par de SSB o DSB. El ARNg guía puede comprender una secuencia espaciadora que es complementaria o bien al locus en 5' o bien al locus en 3'. Alternativamente, el método puede comprender un primer ARNg guía y un segundo ARNg guía. El primer ARNg guía puede comprender una secuencia espaciadora que es complementaria a un segmento del locus en 5' y el segundo ARNg guía puede comprender una secuencia espaciadora que es complementaria a un segmento del locus en 3'. El molde donante puede ser o bien monocatenario o bien bicatenario. La secuencia de control transcripcional modificada puede estar ubicada dentro de un segundo intrón del gen de BCL11A. La secuencia de control transcripcional modificada puede estar ubicada dentro de un sitio hipersensible de ADN (DHS) +58 del gen de BCL11A.

El uno o dos ARNg pueden ser uno o dos ARNg guía de molécula simple (ARNgs). El uno o dos ARNg o uno o dos ARNg pueden ser uno o dos ARNg modificados o uno o dos ARNg modificados. El ARNg modificado puede comprender tres residuos 2'-O-metil-fosforotioato en o cerca de cada uno de sus extremos 5' y 3'. El ARNg modificado puede ser la secuencia de ácido nucleico de SEQ ID NO: 71.959. La una o más endonucleasas Cas9 pueden complejarse previamente con uno o dos ARNg o uno o dos ARNg para formar una o más ribonucleoproteínas (RNP). La una o más endonucleasas Cas9 pueden estar flanqueadas en el extremo N-terminal, el extremo C-terminal o tanto el extremo N-terminal como el extremo C-terminal por una o más señales de localización nuclear (NLS). La una o más endonucleasas Cas9 pueden estar flanqueadas por dos NLS, una NLS ubicada en el extremo N-terminal y la segunda NLS ubicada en el extremo C-terminal. La una o más NLS pueden ser una NLS de SV40. La razón en peso de ARNg con respecto a endonucleasa Cas9 en la RNP puede ser de 1:1. El ARNg puede comprender la secuencia de ácido nucleico de SEQ ID NO: 71.959, la endonucleasa Cas9 puede ser una Cas9 de *S. pyogenes* que comprende una NLS de SV40 en el extremo N-terminal y una NLS de SV40 en el extremo C-terminal, y la razón en peso de ARNg con respecto a endonucleasa Cas9 puede ser de 1:1.

La inserción puede realizarse mediante reparación dirigida por homología (HDR).

La SSB, DSB, locus en 5' y/o locus en 3' pueden estar ubicados dentro de un segundo intrón del gen de BCL11A. La SSB, DSB, locus en 5' y/o locus en 3' pueden estar ubicados dentro de un sitio hipersensible de ADN (DHS) +58 del gen de BCL11A.

El método puede comprender además introducir en la célula uno o más ácidos ribonucleicos guía (ARNg). La una o más ADN endonucleasas pueden ser una o más endonucleasas Cas9 o Cpf1 que realizan o crean un par de roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB), la primera SSB o DSB en un locus en 5' y una segunda SSB o DSB en un locus en 3', dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que provoca una delección del ADN cromosómico entre el locus en 5' y el locus en 3' que da como resultado una delección, modulación o inactivación permanente de la secuencia de control transcripcional del ADN cromosómico entre el locus en 5' y el locus en 3'. El primer ARNg guía puede comprender una secuencia espaciadora que es complementaria a un segmento del locus en 5' y el segundo ARNg guía comprende una secuencia espaciadora que es complementaria a un segmento del locus en 3'. Un ARNg guía puede crear un par de SSB o DSB. El ARNg guía puede comprender una secuencia espaciadora que es complementaria o bien al locus en 5' o bien al locus en 3'. Alternativamente, el método puede comprender un primer ARNg guía y un segundo ARNg guía. El primer ARNg guía puede comprender una secuencia espaciadora que es complementaria a un segmento del locus en 5' y el segundo ARNg guía puede comprender una secuencia espaciadora que es complementaria a un segmento del locus en 3'.

El uno o más ARNg pueden ser uno o más ARNg guía de molécula simple (ARNgs). El uno o más ARNg o uno o más ARNg pueden ser uno o más ARNg modificados o uno o más ARNg modificados. El ARNg modificado puede

comprender tres residuos 2'-O-metil-fosforotioato en o cerca de cada uno de sus extremos 5' y 3'. El ARNg modificado puede ser la secuencia de ácido nucleico de SEQ ID NO: 71.959. La una o más endonucleasas Cas9 pueden complejarse previamente con uno o más ARNg o uno o más ARNGs para formar una o más ribonucleoproteínas (RNP). La una o más endonucleasas Cas9 pueden estar flanqueadas en el extremo N-terminal, el extremo C-terminal o tanto el extremo N-terminal como el extremo C-terminal por una o más señales de localización nuclear (NLS). La una o más endonucleasas Cas9 pueden estar flanqueadas por dos NLS, una NLS ubicada en el extremo N-terminal y la segunda NLS ubicada en el extremo C-terminal. La una o más NLS pueden ser una NLS de SV40. La razón en peso de ARNGs con respecto a endonucleasa Cas9 en la RNP puede ser de 1:1. El ARNg puede comprender la secuencia de ácido nucleico de SEQ ID NO: 71.959, la endonucleasa Cas9 puede ser una Cas9 de *S. pyogenes* que comprende una NLS de SV40 en el extremo N-terminal y una NLS de SV40 en el extremo C-terminal, y la razón en peso de ARNGs con respecto a endonucleasa Cas9 puede ser de 1:1.

El locus en 5' y/o el locus en 3' pueden estar ubicados dentro de un segundo intrón del gen de BCL11A. El locus en 5' y/o el locus en 3' pueden estar ubicados dentro de un sitio hipersensible de ADN (DHS) +58 del gen de BCL11A.

El ARNm de Cas9 o Cpf1, ARNg y molde donante pueden formularse en nanopartículas lipídicas independientes o formularse conjuntamente en una nanopartícula lipídica.

El ARNm de Cas9 o Cpf1 puede formularse en una nanopartícula lipídica, y el ARNg y el molde donante pueden suministrarse a la célula mediante un vector de virus adenoasociado (AAV).

El ARNm de Cas9 o Cpf1 puede formularse en una nanopartícula lipídica, y el ARNg puede suministrarse a la célula mediante electroporación y el molde donante puede suministrarse a la célula mediante un vector de virus adenoasociado (AAV).

La una o más RNP pueden suministrarse a la célula mediante electroporación.

La edición dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A puede reducir la expresión de gen de BCL11A.

El gen de BCL11A puede estar ubicado en el cromosoma 2: 60.451.167 – 60.553.567 (Consortio de Referencia del Genoma – GRCh38).

En el presente documento también se proporcionan uno o más ácidos ribonucleicos guía (ARNg) para editar un gen de BCL11A en una célula de un paciente con una hemoglobinopatía. El uno o más ARNg pueden comprender una secuencia espaciadora seleccionada del grupo que consiste en secuencias de ácido nucleico en SEQ ID NO: 1 – 71.947 de la lista de secuencias. El uno o más ARNg pueden ser uno o más ARN guía de molécula simple (ARNgs). El uno o más ARNg o uno o más ARNGs pueden ser uno o más ARNg modificados o uno o más ARNGs modificados. El uno o más ARNGs modificados pueden comprender tres residuos 2'-O-metil-fosforotioato en o cerca de cada uno de sus extremos 5' y 3'. El uno o más ARNGs modificados pueden comprender la secuencia de ácido nucleico de SEQ ID NO: 71.959. En el presente documento también se proporciona un ARN guía de molécula simple (ARNgs) que comprende la secuencia de ácido nucleico de SEQ ID NO: 71.959.

Se entiende que las invenciones descritas en esta memoria descriptiva no se limitan a los ejemplos resumidos en este sumario. En el presente documento se describen y se muestran a modo de ejemplo varios otros aspectos.

### Breve descripción de los dibujos

Diversos aspectos de materiales y métodos para el tratamiento de hemoglobinopatías dados a conocer y descritos en esta memoria descriptiva pueden entenderse mejor mediante referencia a las figuras adjuntas, en las que:

Las figuras 1A-C muestran plásmidos que comprenden un gen con optimización de codones para endonucleasa Cas9 de *S. pyogenes*.

La figura 1A es un plásmido (CTx-1) que comprende un gen con optimización de codones para endonucleasa Cas9 de *S. pyogenes*. El plásmido CTx-1 también comprende una secuencia de armazón de ARNg, que incluye una secuencia espaciadora de 20 pb de las secuencias indicadas en SEQ ID NO: 1 - 29.482 de la lista de secuencias.

La figura 1B es un plásmido (CTx-2) que comprende un gen con optimización de codones diferente para endonucleasa Cas9 de *S. pyogenes*. El plásmido CTx-2 también comprende una secuencia de armazón de ARNg, que incluye una secuencia espaciadora de 20 pb de las secuencias indicadas en SEQ ID NO: 1 - 29.482 de la lista de secuencias.

La figura 1C es un plásmido (CTx-3) que comprende aún otro gen con optimización de codones diferente para endonucleasa Cas9 de *S. pyogenes*. El plásmido CTx-3 también comprende una secuencia de armazón de ARNg, que incluye una secuencia espaciadora de 20 pb de las secuencias indicadas en SEQ ID NO: 1 - 29.482 de la lista de secuencias.

Las figuras 2A-B representan el sistema de sistema de CRISPR/Cas de tipo II.

La figura 2A representa el sistema de CRISPR/Cas de tipo II que incluye ARNg.

La figura 2B representa el sistema de CRISPR/Cas de tipo II que incluye ARNg.

La figura 3 muestra la tasa de edición de ADN en células progenitoras y madre hematopoyéticas CD34+ (HSPC) y cada uno de los diferentes genotipos de HPFH resultantes.

Las figuras 4A-C muestran la regulación por incremento de la expresión de  $\gamma$ -globina en eritrocitos diferenciados a partir de HSPC CD34+ humanas sometidas a edición globales a partir de sangre periférica movilizada (mPB).

La figura 4A representa hematopoyesis de HSPC CD34+ humanas para dar eritrocitos.

La figura 4B muestra la razón de  $\gamma$ /ARN 18s para cada una de la delección/modificación.

La figura 4C muestra la razón de  $\gamma/\alpha$  para cada una de la delección/modificación.

Las figuras 5A-B muestran la regulación por incremento de la expresión de  $\gamma$ -globina en eritrocitos diferenciados a partir de todas las colonias sometidas a edición génica a partir de HSPC CD34+ humanas.

La figura 5A muestra la razón de ARNm de  $\gamma/\alpha$ -globina (%) para cada una de las colonias sometidas a edición génica.

La figura 5B muestra la razón de ARNm de  $\gamma/\alpha$ -globina promedio (%) para cada una de las modificaciones génicas.

La figura 6 muestra la tasa de intrón de BCL11A (SPY101) de edición de ADN en HSPC CD34+ humana derivada de colonias eritroides.

Las figuras 7A-B muestran la correlación entre el genotipo de SPY101 y la expresión de  $\gamma$ -globina en colonias de células individuales diferenciadas a partir de HSPC CD34+ de mPB humanas sometidas a edición génica.

La figura 7A muestra el porcentaje de  $\gamma$ -globina con respecto a  $\alpha$ -globina (HBG/HBA) para cada una de las colonias sometidas a edición génica.

La figura 7B muestra el porcentaje de globinas de tipo  $\beta$  (HBG/(HBB+HBG)) para cada una de las colonias sometidas a edición génica.

La figura 8 muestra la eficacia de edición específica de varios ARNg en células CD34+ de mPB humanas.

Las figuras 9A-B muestran el ensayo de captura de híbridos usado para detectar edición inespecífica y resultados generados usando el ensayo de captura de híbridos a partir de HSPC CD34+ de mPB humanas sometidas a edición.

La figura 9A muestra un esquema de un ensayo de captura de híbridos usado para detectar la actividad de edición en posibles sitios inespecíficos.

La figura 9B muestra actividad inespecífica observada mediante secuenciación de captura de híbridos.

Las figuras 10A-B muestran razones de niveles de ARNm de globina medidos en células a partir de pacientes con SCD, un paciente con  $\beta$ -talasemia y donantes sanos.

La figura 10A muestra razones de niveles de ARNm de globina medidos en células a partir de pacientes con SCD en comparación con donantes sanos.

La figura 10B muestra razones de niveles de ARNm de globina medidos en células a partir de un paciente con  $\beta$ -talasemia en comparación con donantes sanos.

Las figuras 11A-C muestran la estrategia de citometría de flujo usada para detectar diversas poblaciones de células sometidas a edición génica y resultados generados usando la estrategia de citometría de flujo.

La figura 11A muestra subpoblaciones de HSPC CD34+ de mPB humanas, marcadores de superficie asociados y estrategia de selección por citometría de flujo.

La figura 11B muestra una distribución similar de tipos de célula en las condiciones simulada y sometida a edición.

La figura 11C muestra altas eficiencias de edición similares a lo largo de las subpoblaciones en comparación con la global.

La figura 12 muestra el análisis de poblaciones de células CD45RA<sup>+</sup> humanas en ratones NSG 8 semanas tras el injerto de HSPC CD34<sup>+</sup> de mPB humanas. Los puntos de datos representan animales individuales y representan el porcentaje de células vivas que eran células vivas CD45RA<sup>+</sup> humanas.

La figura 13 muestra la eficiencia de edición promedio de un ARNg de SPY101 y proteína Cas9 en HSPC CD34<sup>+</sup> de mPB humanas a escalas relevantes de laboratorio y clínica.

La figura 14 muestra un resumen del diseño de estudio de GLP/toxicología.

La figura 15 muestra un resumen de un enfoque experimental para análisis de colonias globales y de células individuales de niveles de ARNm y proteína de hemoglobina en poblaciones de células eritroides derivadas a partir de HSPC CD34<sup>+</sup> de mPB humanas sometidas a edición génica de CRISPR/Cas9.

Las figuras 16A-B muestran la regulación por incremento de ARNm y proteína de  $\gamma$ -globina en HSPC CD34<sup>+</sup> de mPB humanas diferenciadas globales modificadas con diferentes ediciones dirigidas.

La figura 16A muestra la regulación por incremento de ARNm de  $\gamma$ -globina en HSPC CD34<sup>+</sup> de mPB humanas diferenciadas globales modificadas con diferentes ediciones dirigidas.

La figura 16B muestra la regulación por incremento de proteína de  $\gamma$ -globina en HSPC CD34<sup>+</sup> de mPB humanas diferenciadas globales modificadas con diferentes ediciones dirigidas.

La figura 17 muestra la regulación por incremento de  $\gamma$ -globina promedio en colonias individuales de HSPC CD34<sup>+</sup> de mPB humanas diferenciadas modificadas con diferentes ediciones dirigidas.

Las figuras 18A-B muestran una correlación entre genotipo y fenotipo en colonias sometidas a edición con diana 5 y diana 6 de HSPC CD34<sup>+</sup> de mPB humanas diferenciadas eritroides.

La figura 18A incluye gráficos en el lado izquierdo que muestran el % de colonias con cada genotipo, y gráficos en el lado derecho que muestran el porcentaje de colonias con cada nivel de regulación por incremento de  $\gamma$ -globina (expresado como razón de ARNm de  $\gamma/(\gamma+\beta)$ -globina).

La figura 18B muestra niveles de transcritos de ARNm, para grupos de colonias con genotipos similares.

La figura 19 muestra un resumen de un enfoque experimental para análisis global de la eficiencia de edición a partir de ADN genómico, expresión de hemoglobina mediante ARNm, y proteína en poblaciones de células diferenciadas eritroides derivadas a partir de HSPC CD34<sup>+</sup> de mPB humanas sometidas a edición génica de CRISPR/Cas9.

Las figuras 20A-B muestran el porcentaje de edición génica mantenido a lo largo de la totalidad de la diferenciación eritroide ex vivo de HSPC CD34<sup>+</sup> de mPB sometidas a edición con ARNg de SPY101 o ARNg de SD2.

La figura 20A muestra el porcentaje de edición génica mantenido a lo largo de la totalidad de la diferenciación eritroide ex vivo de HSPC CD34<sup>+</sup> de mPB sometidas a edición con ARNg de SPY101.

La figura 20B muestra el porcentaje de edición génica mantenido a lo largo de la totalidad de la diferenciación eritroide ex vivo de HSPC CD34<sup>+</sup> de mPB sometidas a edición con ARNg de SD2.

Las figuras 21A-D muestran el aumento de transcrito de  $\gamma$ -globina representado como  $\gamma/\alpha$  o  $\gamma/(\gamma+\beta)$  en HSPC CD34<sup>+</sup> de mPB sometidas a edición génica en los días 11 o 15 tras la diferenciación eritroide.

La figura 21A muestra el aumento de transcrito de  $\gamma$ -globina ( $\gamma/\alpha$ ) en HSPC CD34<sup>+</sup> de mPB sometidas a edición génica en el día 11 tras la diferenciación.

La figura 21B muestra el aumento de transcrito de  $\gamma$ -globina ( $\gamma/\alpha$ ) en HSPC CD34<sup>+</sup> de mPB sometidas a edición génica en el día 15 tras la diferenciación.

La figura 21C muestra el aumento de transcrito de  $\gamma$ -globina ( $\gamma/(\gamma+\beta)$ ) en HSPC CD34<sup>+</sup> de mPB sometidas a edición génica en el día 11 tras la diferenciación.

La figura 21D muestra el aumento de transcrito de  $\gamma$ -globina ( $\gamma/(\gamma+\beta)$ ) en HSPC CD34<sup>+</sup> de mPB sometidas a edición génica en el día 15 tras la diferenciación.

Las figuras 22A-B es análisis de FACS y análisis de intensidad de fluorescencia media (MFI) que muestran la regulación por incremento de  $\gamma$ -globina en HSPC CD34+ de mPB sometidas a edición génica en el día 15 tras la diferenciación eritroide.

5 La figura 22A es análisis de FACS que muestra la regulación por incremento de  $\gamma$ -globina en HSPC CD34+ de mPB sometidas a edición génica 15 días tras la diferenciación eritroide.

La figura 22B es análisis de MFI que muestra la regulación por incremento promedio de  $\gamma$ -globina en células CD34+ de mPB sometidas a edición génica a partir de 4 donantes tras la diferenciación eritroide.

10 La figura 23A-D son datos de cromatografía de líquidos-espectrometría de masas (CL-EM) globales que muestran la regulación por incremento de  $\gamma$ -globina, representada como  $\gamma/\alpha$  o  $\gamma/(\gamma+\beta)$  en HSPC CD34+ de mPB sometidas a edición génica en el día 15 tras la diferenciación eritroide.

15 La figura 23A son datos de cromatografía de líquidos-espectrometría de masas (CL-EM) globales que muestran la regulación por incremento de  $\gamma$ -globina ( $\gamma/\alpha$ ) en HSPC CD34+ de mPB sometidas a edición génica en el día 15 tras la diferenciación.

20 La figura 23B son datos de cromatografía de líquidos-espectrometría de masas (CL-EM) globales que muestran la regulación por incremento de  $\gamma$ -globina ( $\gamma/\alpha$ ) en HSPC CD34+ de mPB sometidas a edición génica en el día 15 tras la diferenciación normalizada con respecto a  $\gamma$ -globina ( $\gamma/\alpha$ ) en HSPC CD34+ de mPB transfectadas con ARNg de GFP.

25 La figura 23C son datos de cromatografía de líquidos-espectrometría de masas (CL-EM) globales que muestran la regulación por incremento de  $\gamma$ -globina ( $\gamma/(\gamma+\beta)$ ) en HSPC CD34+ de mPB sometidas a edición génica en el día 15 tras la diferenciación.

30 La figura 23D son datos de cromatografía de líquidos-espectrometría de masas (CL-EM) globales que muestran la regulación por incremento de  $\gamma$ -globina ( $\gamma/(\gamma+\beta)$ ) en HSPC CD34+ de mPB sometidas a edición génica en el día 15 tras la diferenciación normalizada con respecto a  $\gamma$ -globina ( $\gamma/\alpha$ ) en HSPC CD34+ de mPB transfectadas con ARNg de GFP.

La figura 24 representa el diseño de cebo de captura de híbridos.

35 La figura 25 muestra un gráfico que representa la potencia del método de captura de híbridos para detectar indeles.

La figura 26 muestra un sumario de los datos generados a partir de experimentos de captura de híbridos usando ARNg de SPY101.

40 La figura 27 muestra un sumario de los datos generados a partir de experimentos de captura de híbridos usando ARNg de SD2.

La figura 28 muestra un plan de estudio para los experimentos de injerto.

45 Las figuras 29A-E muestran datos de análisis de sangre intermedio de 8 semanas para ratones sin tratar y ratones a los que se les inyectaron células sometidas a edición de manera simulada, células sometidas a edición con ARNg de GFP, células sometidas a edición con ARNg de SPY101 o células sometidas a edición con ARNg de SD2.

La figura 29A muestra datos de análisis de sangre intermedio de 8 semanas para ratones sin tratar (UnTx).

50 La figura 29B muestra datos de análisis de sangre intermedio de 8 semanas para ratones a los que se les inyectaron células sometidas a edición de manera simulada.

La figura 29C muestra datos de análisis de sangre intermedio de 8 semanas para ratones a los que se les inyectaron células sometidas a edición con ARNg de GFP.

55 La figura 29D muestra datos de análisis de sangre intermedio de 8 semanas para ratones a los que se les inyectaron células sometidas a edición con ARNg de SPY101.

60 La figura 29E muestra datos de análisis de sangre intermedio de 8 semanas para ratones a los que se les inyectaron células sometidas a edición con ARNg de SD2.

La figura 30 muestra datos de análisis de sangre intermedio de 8 semanas promedio.

65 La figura 31 muestra el % de indel para HSPC CD34+ de mPB humanas sometidas a electroporación con diversos ARNm de Cas9 y ARNg de SPY101 (ARNm 1-8) en comparación con HSPC CD34+ de mPB humanas sometidas a

electroporación con proteína Cas9 complejada con ARNg de SPY101 (un complejo de ribonucleoproteína, RNP).

Las figuras 32A-B muestran el recuento celular normalizado y la viabilidad celular de HSPC CD34+ de mPB humanas sometidas a electroporación con diversos ARNm de Cas9 y ARNg de SPY101 (ARNm 1-8) en comparación con HSPC CD34+ de mPB humanas sometidas a electroporación con proteína Cas9 complejada con ARNg de SPY101 (RNP).

La figura 32A muestra el aumento en veces del recuento celular a las 48 horas tras la electroporación para HSPC CD34+ de mPB humanas sometidas a electroporación con diversos ARNm de Cas9 y ARNg de SPY101 (ARNm 1-8) en comparación con HSPC CD34+ de mPB humanas sometidas a electroporación con proteína Cas9 complejada con ARNg de SPY101 (RNP).

La figura 32B muestra la viabilidad celular a las 48 horas tras la electroporación para HSPC CD34+ de mPB humanas sometidas a electroporación con diversos ARNm de Cas9 y ARNg de SPY101 (ARNm 1-8) en comparación con HSPC CD34+ de mPB humanas sometidas a electroporación con proteína Cas9 complejada con ARNg de SPY101 (RNP).

Las figuras 33A-C muestran varios constructos de RNP de Cas9 usados para la optimización de RNP de Cas9 y el % de indel asociado con cada uno de los constructos de RNP de Cas9.

La figura 33A muestra varios constructos de RNP de Cas9.

La figura 33B muestra el % de indel para cada uno de los constructos de RNP de Cas9 usando 1 µg de Cas9 : 1 µg de ARNg de SPY101.

La figura 33C muestra el % de indel para cada uno de los constructos de RNP de Cas9 usando 3 µg de Cas9 : 3 µg de ARNg de SPY101.

Las figuras 34A-B muestran la eficiencia de edición génica (%) para HSPC CD34+ de mPB humanas tratadas o bien con ARNm de Cas9 o bien con proteína Cas9 (Feldan o Aldevron) a escala no clínica y clínica.

La figura 34A muestra la eficiencia de edición génica (%) para HSPC CD34+ derivadas de médula ósea (BM) o mPB humanas tratadas o bien con ARNm de Cas9 o bien con proteína Cas9 (Feldan o Aldevron) a escala no clínica.

La figura 34B muestra la eficiencia de edición génica (%) para HSPC CD34+ de mPB humanas tratadas con proteína Cas9 (Aldevron) a escala clínica.

Las figuras 35A-B muestran la eficacia de SPY101 en HSPC CD34+ de mPB humanas presentando la razón de ARNm de  $\gamma/\alpha$ -globina en % y la razón de ARNm de  $\gamma/(\gamma+\beta)$ -globina en % para células tratadas o bien con ARNm de Cas9 y ARNg de SPY101 o bien con proteína Cas9 (Feldan o Aldevron) complejada con ARNg de SPY101.

La figura 35A muestra la razón de ARNm de  $\gamma/\alpha$ -globina en % para HSPC CD34+ de mPB humanas tratadas o bien con ARNm de Cas9 y ARNg de SPY101 o bien con proteína Cas9 (Feldan o Aldevron) complejada con ARNg de SPY101.

La figura 35B muestra la razón de ARNm de  $\gamma/(\gamma+\beta)$ -globina en % para HSPC CD34+ de mPB humanas tratadas o bien con ARNm de Cas9 y ARNg de SPY101 o bien con proteína Cas9 (Feldan o Aldevron) complejada con ARNg de SPY101.

Las figuras 36A-B muestran la eficacia de SPY101 en HSPC CD34+ derivadas de médula ósea presentando la razón de ARNm de  $\gamma/\alpha$ -globina en % y la razón de ARNm de  $\gamma/(\gamma+\beta)$ -globina en % para células tratadas con proteína Cas9 (Aldevron, optimizada técnicamente frente a no optimizada) complejada con ARNg de SPY101.

La figura 36A muestra la razón de ARNm de  $\gamma/\alpha$ -globina en % para HSPC CD34+ derivadas de médula ósea tratadas con proteína Cas9 complejada con ARNg de SPY101.

La figura 36B muestra la razón de ARNm de  $\gamma/(\gamma+\beta)$ -globina en % para HSPC CD34+ derivadas de médula ósea tratadas con proteína Cas9 complejada con ARNg de SPY101.

Las figuras 37A-B muestran la eficacia de SPY101 en muestras de pacientes con SCD y  $\beta$ -talasemia.

La figura 37A muestra la razón de ARNm de  $\gamma/(\gamma+\beta)$ -globina promedio en % para células diferenciadas eritroides a partir de seis pacientes con SCD y dos donantes sanos que se trataron con ARNg de SPY101 y proteína Cas9. Se restaron todos los valores de sus muestras de control respectivas tratadas con ARNg de GFP y proteína Cas9.

La figura 37B muestra la razón de ARNm de  $\gamma/\alpha$ -globina en % para células diferenciadas eritroides a partir de un paciente con  $\beta$ -talasemia y dos donantes sanos que se trataron con ARNg de SPY101 y proteína Cas9. Se restaron

todos los valores de sus muestras de control respectivas tratadas con ARNg de GFP y proteína Cas9.

Las figuras 38A-B muestran la tasa de intrón de Bcl11a (SPY101) de edición de ADN cuando se usa ARNm de Cas9 o RNP de Cas9.

La figura 38A muestra la tasa de intrón de BCL11A (SPY101) de edición de ADN cuando se usa ARNm de Cas9.

La figura 38B muestra la tasa de intrón de BCL11A (SPY101) de edición de ADN cuando se usa RNP de Cas9.

Las figuras 39A-B muestran que perturbaciones del sitio de unión a GATA1 (GBS) provocadas por SPY101/RNP de Cas9 en colonias de células individuales derivadas de HSPC CD34+ de mPB humanas diferenciadas eritroides están vinculadas a una expresión de  $\gamma$ -globina aumentada.

La figura 39A muestra la razón de ARNm de  $\gamma/\alpha$ -globina de colonias sometidas a edición con SPY101 sin perturbación de GBS, con perturbaciones de GBS monoalélicas o con perturbaciones de GBS bialélicas.

La figura 39B muestra la razón de ARNm de  $\gamma/(\gamma+\beta)$ -globina de colonias sometidas a edición con SPY101 sin perturbación de GBS, con perturbaciones de GBS monoalélicas o con una perturbación de GBS bialélica.

Las figuras 40A-E muestra una expresión de  $\gamma$ -globina aumentada en HSPC CD34+ de mPB humanas sometidas a edición con SPY101/RNP de Cas9 diferenciadas eritroides mediante análisis de citometría de flujo.

La figura 40A es un análisis de citometría de flujo que muestra la expresión de  $\alpha$ -globina en HSPC CD34+ de mPB humanas diferenciadas eritroides sometidas a edición con SPY101/RNP de Cas9 en comparación con la expresión de  $\alpha$ -globina en HSPC CD34+ de mPB humanas diferenciadas eritroides tratadas con ARNg de GFP/RNP de Cas9.

La figura 40B es un análisis de citometría de flujo que muestra la expresión de  $\beta$ -globina en HSPC CD34+ de mPB humanas diferenciadas eritroides sometidas a edición con SPY101/RNP de Cas9 en comparación con la expresión de  $\beta$ -globina en HSPC CD34+ de mPB humanas diferenciadas eritroides tratadas con ARNg de GFP/RNP de Cas9.

La figura 40C es un análisis de citometría de flujo que muestra la expresión de  $\gamma$ -globina en HSPC CD34+ de mPB humanas diferenciadas eritroides sometidas a edición con SPY101/RNP de Cas9 en comparación con la expresión de  $\gamma$ -globina en HSPC CD34+ de mPB humanas diferenciadas eritroides tratadas con ARNg de GFP/RNP de Cas9.

La figura 40D muestra el porcentaje de células positivas para  $\gamma$ -globina en HSPC CD34+ de mPB humanas diferenciadas eritroides sometidas a edición con SPY101/RNP de Cas9 en comparación con HSPC CD34+ de mPB humanas diferenciadas eritroides tratadas con ARNg de GFP/RNP de Cas9.

La figura 40E muestra la intensidad de fluorescencia media (MFI) en HSPC CD34+ de mPB humanas diferenciadas eritroides sometidas a edición con SPY101/RNP de Cas9 en comparación con HSPC CD34+ de mPB humanas diferenciadas eritroides tratadas con ARNg de GFP/RNP de Cas9.

#### Breve descripción de la lista de secuencias

SEQ ID NO: 1 - 29.482 son secuencias espaciadoras de 20 pb para el direccionamiento dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una endonucleasa Cas9 de *S. pyogenes*.

SEQ ID NO: 29.483 - 32.387 son secuencias espaciadoras de 20 pb para el direccionamiento dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una Endonucleasa Cas9 de *S. aureus*.

SEQ ID NO: 32.388 - 33.420 son secuencias espaciadoras de 20 pb para el direccionamiento dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una endonucleasa Cas9 de *S. thermophilus*.

SEQ ID NO: 33.421 - 33.851 son secuencias espaciadoras de 20 pb para el direccionamiento dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una endonucleasa Cas9 de *T. denticola*.

SEQ ID NO: 33.852 - 36.731 son secuencias espaciadoras de 20 pb para el direccionamiento dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una endonucleasa Cas9 de *N. meningitidis*.

SEQ ID NO: 36.732 - 71.947 son secuencias espaciadoras de 22 pb para el direccionamiento dentro o cerca de un

gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una endonucleasa Cpf1 de *Acidominococcus*, *Lachnospiraceae* y *Franciscella Novicida*.

SEQ ID NO: 71.948 es un ARN guía (ARNg) de muestra para una endonucleasa Cas9 de *S. pyogenes*.

SEQ ID NO: 71.949 muestra una familia conocida de endonucleasa de asentamiento, según se clasifica mediante su estructura.

SEQ ID NO: 71.950 es ARNg A (CLO1).

SEQ ID NO: 71.951 es ARNg B (CLO8).

SEQ ID NO: 71.952 es ARNg C (CSO2).

SEQ ID NO: 71.953 es ARNg D (CSO6).

SEQ ID NO: 71.954 es ARNg E (HPFH-15).

SEQ ID NO: 71.955 es ARNg F (HPFH-4).

SEQ ID NO: 71.956 es ARNg G (Kenya02).

SEQ ID NO: 71.957 es ARNg H (Kenya17).

SEQ ID NO: 71.958 es ARNg I (SD2).

SEQ ID NO: 71.959 es ARNg J (SPY101).

SEQ ID NO: 71.960 - 71.962 muestran secuencias de ARNGs de muestra.

### Descripción detallada

#### Hemoglobina fetal

La hemoglobina fetal (HbF,  $\alpha_2\gamma_2$ ) es la principal proteína de transporte de oxígeno en un feto humano e incluye subunidades alfa ( $\alpha$ ) y gamma ( $\gamma$ ). La expresión de HbF cesa aproximadamente 6 meses tras el nacimiento. La hemoglobina del adulto (HbA,  $\alpha_2\beta_2$ ) es la principal proteína de transporte de oxígeno en un ser humano después de ~34 semanas desde el nacimiento e incluye subunidades alfa ( $\alpha$ ) y beta ( $\beta$ ). Después de 34 semanas, un cambio del desarrollo da como resultado una transcripción reducida de los genes de  $\gamma$ -globina y una transcripción aumentada de los genes de  $\beta$ -globina. Dado que muchas de las formas de hemoglobinopatías son resultado de no lograr producir proteína  $\beta$ -globina normal en cantidades suficientes o no lograr producir en absoluto proteína  $\beta$ -globina normal, una expresión aumentada de  $\gamma$ -globina (es decir, HbF) mejorará la gravedad de la enfermedad de  $\beta$ -globina.

#### Linfoma de células B 11A (BCL11A)

El linfoma de células B 11A (BCL11A) es un gen ubicado en el cromosoma 2 y oscila desde 60.451.167 - 60.553.567 pb (GRCh38). BCL11A es un factor de transcripción de dedos de cinc que reprime la hemoglobina fetal (HbF) y regula por disminución la expresión de HbF empezando aproximadamente a las 6 semanas tras el nacimiento. El gen de BCL11A contiene 4 exones, que abarcan 102,4 kb de ADN genómico. BCL11A también está bajo regulación de la transcripción, incluyendo un dominio de unión en el intrón 2 para el factor de transcripción maestro GATA-1. La unión a GATA-1 potencia la expresión de BCL11A que, a su vez, reprime la expresión de HbF. El intrón 2 contiene múltiples sitios hipersensibles a ADNasa (DHS), incluyendo sitios denominados +55, +58 y +62 basándose en la distancia en kilobases desde el sitio de inicio de la transcripción. A continuación se comentan diversas estrategias de edición para delecionar, modular o inactivar las secuencias de control transcripcional de BCL11A. Los SNP que se producen de manera natural dentro de esta región se han asociado con una expresión de BCL11A reducida y niveles de Hb fetal aumentados (Orkin *et al.* 2013 GWAS study). Estos SNP están organizados alrededor de 3 sitios de hipersensibilidad de ADN, DHS +55, DHS +58 y DHS +62. De las 3 regiones, la región de DHS +58 parece ser la región clave asociada con niveles de Hb fetal aumentados y también alberga una región de control transcripcional de GATA1.

#### Enfoque terapéutico

Puede usarse la unión de extremos no homólogos (NHEJ) para delecionar segmentos de la secuencia de control transcripcional de BCL11A, o bien directamente o bien alterando sitios donantes o aceptores de corte y empalme mediante escisión por un ARNg que se dirige a varias ubicaciones, o varios ARNg.



La secuencia de control transcripcional del gen de BCL11A también puede modularse o inactivarse insertando un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada. Por ejemplo, el donante para la modulación o inactivación mediante reparación dirigida por homología (HDR) contiene la secuencia de control transcripcional modificada del gen de BCL11A con ramas de homología flanqueantes pequeñas o grandes para permitir el apareamiento. HDR es esencialmente un mecanismo libre de error que usa una secuencia de ADN homóloga suministrada como molde durante la reparación de DSB. La tasa de reparación dirigida por homología (HDR) depende de la distancia entre la secuencia de control transcripcional y el sitio de corte, de modo que resulta importante elegir sitios diana solapantes o cercanos. Los moldes pueden incluir secuencias adicionales flanqueadas por las regiones homólogas o pueden contener una secuencia que difiere de la secuencia genómica, permitiendo por tanto edición de secuencia.

Además de delecionar, modular o inactivar la secuencia de control transcripcional del gen de BCL11A mediante NHEJ o HDR, son posibles una gama de otras opciones. Si hay delecciones pequeñas o grandes, puede insertarse un ADNc que contiene una secuencia de control transcripcional modificada del gen de BCL11A. Puede insertarse un ADNc de longitud completa en cualquier "ubicación segura" (es decir, punto de inserción no perjudicial que no es el propio gen de BCL11A), con o sin secuencias reguladoras adecuadas. Si se inserta este constructo cerca de los elementos reguladores de BCL11A, debe tener control fisiológico, similar al gen normal. Pueden usarse dos o más (por ejemplo, un par) nucleasas para delecionar regiones de secuencia de control transcripcional, aunque habitualmente tendrá que proporcionarse un donante para modular o inactivar la función. En este caso, se suministrarán dos ARNg y una secuencia donante.

En el presente documento se proporcionan métodos celulares, *ex vivo* e *in vivo*, para usar herramientas de ingeniería de genoma para crear cambios permanentes en el genoma mediante: 1) modular o inactivar la secuencia de control transcripcional del gen de BCL11A, mediante delecciones que surgen debido a la ruta de NHEJ; 2) modular o inactivar la secuencia de control transcripcional del gen de BCL11A, mediante HDR; 3) modular o inactivar la secuencia de control transcripcional del gen de BCL11A, mediante delecciones de al menos una porción de la secuencia de control transcripcional y/o inserción de un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada en el locus de gen o un locus de ubicación segura. Tales métodos usan endonucleasas, tales como nucleasas asociadas con CRISPR (Cas9, Cpf1 y similares), para delecionar, insertar o editar de manera permanente la secuencia de control transcripcional dentro o cerca del locus genómico del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A. De esta manera, ejemplos expuestos en la presente divulgación pueden ayudar a delecionar, modular o inactivar la secuencia de control transcripcional del gen de BCL11A con un único tratamiento o un número limitado de tratamientos (en vez de suministrar posibles terapias durante toda la vida del paciente).

En el presente documento se describen métodos para tratar a un paciente con una hemoglobinopatía. Un aspecto de tal método es una terapia *ex vivo* basada en células. Por ejemplo, puede crearse una célula madre pluripotente inducida (iPSC) específica de paciente. Después, puede editarse el ADN cromosómico de estas células iPSC usando los materiales y métodos descritos en el presente documento. A continuación, pueden diferenciarse las iPSC sometidas a edición genómica para dar células progenitoras hematopoyéticas. Finalmente, pueden implantarse las células progenitoras hematopoyéticas en el paciente.

Aún otro aspecto de tales método es una terapia *ex vivo* basada en células. Por ejemplo, puede aislarse una célula madre mesenquimatosa a partir del paciente, que puede aislarse a partir de la médula ósea o sangre periférica del paciente. A continuación, puede editarse el ADN cromosómico de estas células madre mesenquimatosas usando los materiales y métodos descritos en el presente documento. A continuación, pueden diferenciarse las células madre mesenquimatosas sometidas a edición genómica para dar células progenitoras hematopoyéticas. Finalmente, pueden implantarse estas células progenitoras hematopoyéticas en el paciente.

Un aspecto adicional de tal método es una terapia *ex vivo* basada en células. Por ejemplo, puede aislarse una célula progenitora hematopoyética a partir del paciente. A continuación, puede editarse el ADN cromosómico de estas células usando los materiales y métodos descritos en el presente documento. Finalmente, pueden implantarse las células progenitoras hematopoyéticas sometidas a edición genómica en el paciente.

Una ventaja de un enfoque de terapia celular *ex vivo* es la capacidad de llevar a cabo un análisis exhaustivo del producto terapéutico antes de la administración. Los productos terapéuticos basados en nucleasa pueden tener cierto nivel de efectos inespecíficos. Realizar corrección genética *ex vivo* permite caracterizar la población celular corregida antes de la implantación. La presente divulgación incluye secuenciar todo el genoma de las células corregidas para garantizar que los efectos inespecíficos, si los hay, pueden estar en ubicaciones genómicas asociadas con un riesgo mínimo para el paciente. Además, pueden aislarse poblaciones de células específicas, incluyendo poblaciones clonales, antes de la implantación.

Otra ventaja de la terapia celular *ex vivo* está relacionada con a la corrección genética en iPSC en comparación con otras fuentes de células primarias. Las iPSC son prolíficas, haciendo que sea fácil obtener el gran número de células que se requerirán para una terapia basada en células. Además, las iPSC son un tipo de célula ideal para realizar aislamientos clonales. Esto permite seleccionar la corrección genómica correcta, sin riesgo de una reducción de la

viabilidad. En cambio, otras células primarias solo son viables durante unos pocos pases y son difíciles de expandir mediante clonación. Por tanto, la manipulación de iPSC para el tratamiento de una hemoglobinopatía puede ser mucho más fácil y puede acortar la cantidad de tiempo necesario para realizar la corrección genética deseada.

Para la terapia *ex vivo*, el trasplante requiere el aclaramiento de nichos de médula ósea o las HSC de donante que van a injertarse. Los métodos actuales se basan en radiación y/o quimioterapia. Debido a las limitaciones que imponen, se han desarrollado y están desarrollándose regímenes de acondicionamiento más seguros, tales como inmunodepleción de células de médula ósea mediante anticuerpos o conjugados de anticuerpo-toxina dirigidos contra marcadores de superficie de células hematopoyéticas, por ejemplo CD117, c-kit y otros. El éxito del trasplante de HSC depende de un asentamiento eficiente en la médula ósea, posterior injerto y repoblación de médula ósea. El nivel de células sometidas a edición génica injertadas es importante, al igual que la capacidad del injerto de linaje múltiple de las células.

Las células madre hematopoyéticas (HSC) son una diana importante para terapia génica *ex vivo* ya que proporcionan una fuente prolongada de las células corregidas. Las células CD34+ tratadas se devolverán al paciente.

Los métodos también pueden incluir una terapia de base *in vivo*. Se edita ADN cromosómico de las células en el paciente usando los materiales y métodos descritos en el presente documento. Las células pueden ser células de médula ósea, células progenitoras hematopoyéticas o células CD34+.

Aunque las células sanguíneas presentan una diana atractiva para el tratamiento y terapia *ex vivo*, una eficacia aumentada en el suministro puede permitir un suministro *in vivo* directo a las células madre hematopoyéticas (HSC) y/u otros progenitores de células B y T, tales como células CD34+. De manera ideal, el direccionamiento y la edición estarán dirigidos a las células relevantes. También puede prevenirse la escisión en otras células mediante el uso de promotores activos únicamente en determinadas células y/o fases de desarrollo. Los promotores adicionales son inducibles y, por tanto, pueden controlarse temporalmente si la nucleasa se suministra como un plásmido. La cantidad de tiempo que permanece el ARN y proteína suministrados en la célula también puede ajustarse usando tratamientos o dominios añadidos para cambiar la semivida. El tratamiento *in vivo* eliminará varias etapas de tratamiento, pero una tasa inferior de suministro puede requerir tasas superiores de edición. El tratamiento *in vivo* puede eliminar problemas y pérdidas del tratamiento *ex vivo* y el injerto.

Una ventaja de la terapia génica *in vivo* puede ser la facilidad de producción terapéutica y administración. El mismo enfoque terapéutico y terapia tendrá la posibilidad de usarse para tratar a más de un paciente, por ejemplo varios pacientes que comparten el mismo genotipo o alelo o uno similar. En cambio, la terapia celular *ex vivo* requiere normalmente usar las células del propio paciente, que se aíslan, se manipulan y se devuelven al mismo paciente.

También se describe en el presente documento un método celular para editar el gen de BCL11A en una célula mediante edición genómica. Por ejemplo, puede aislarse una célula a partir de un paciente o animal. Después, puede editarse el ADN cromosómico de la célula usando los materiales y métodos descritos en el presente documento.

Los métodos proporcionados en el presente documento, independientemente de si es un método celular o *ex vivo* o *in vivo*, pueden implicar una o una combinación de lo siguiente: 1) modular o inactivar la secuencia de control transcripcional del gen de BCL11A, mediante delecciones que surgen debido a la ruta de NHEJ, 2) modular o inactivar la secuencia de control transcripcional del gen de BCL11A, mediante HDR, o 3) modular o inactivar la secuencia de control transcripcional del gen de BCL11A, mediante delección de al menos una porción de la secuencia de control transcripcional y/o inserción de un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada en el locus de gen o una ubicación heteróloga en el genoma (tal como un sitio de ubicación segura, tal como AAVS1). Las estrategias tanto de HDR como de inserción usan un molde de ADN donante en reparación dirigida por homología (HDR). La HDR en cualquier estrategia puede lograrse realizando una o más roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB) en sitios específicos en el genoma usando una o más endonucleasas.

Por ejemplo, la estrategia de NHEJ puede implicar deleccionar al menos una porción de la secuencia de control transcripcional del gen de BCL11A induciendo una rotura de cadena sencilla o rotura de cadena doble dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una o más endonucleasas CRISPR y un ARNg (por ejemplo, ARNcr + ARNtracr, o ARNGs), o dos o más roturas de cadena sencilla o roturas de cadena doble dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con dos o más endonucleasas CRISPR y dos o más ARNGs. Este enfoque puede requerir el desarrollo y la optimización de ARNGs para la secuencia de control transcripcional del gen de BCL11A.

Por ejemplo, la estrategia de HDR puede implicar modular o inactivar la secuencia de control transcripcional del gen de BCL11A induciendo una rotura de cadena sencilla o rotura de cadena doble dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una o más endonucleasas CRISPR y un ARNg (por ejemplo, ARNcr + ARNtracr, o ARNGs), o dos o más roturas de cadena sencilla o roturas de cadena doble dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador

del gen de BCL11A con una o más endonucleasas CRISPR y dos o más ARNg, en presencia de un molde de ADN donante introducido de manera exógena para dirigir la respuesta de DSB celular frente a la reparación dirigida por homología (el molde de ADN donante puede ser un oligonucleótido monocatenario corto, un oligonucleótido bicatenario corto, una molécula de ADN monocatenario o bicatenario larga). Este enfoque puede requerir el desarrollo y la optimización de ARNg y moléculas de ADN donante que comprenden un gen de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada.

Por ejemplo, la estrategia de inserción implica insertar un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada en el locus del gen de BCL11A usando un ARNg (por ejemplo, ARNcr + ARNtracr, o ARNGs) o un par de ARNg dirigidos en sentido de 5', o en la secuencia de control transcripcional del gen de BCL11A, o en un sitio de ubicación segura (tal como AAVS1). El ADN donante puede ser ADN monocatenario o bicatenario y comprende un gen de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada.

Las ventajas de las estrategias anteriores (deleción/modulación/inactivación e inserción) son similares, incluyendo en principio efectos beneficiosos clínicos y de laboratorio tanto a corto como a largo plazo.

Además de las opciones de edición indicadas anteriormente, pueden usarse proteínas Cas9 o similares para dirigir dominios efectores a los mismos sitios diana que pueden identificarse para la edición, o sitios diana adicionales dentro del alcance del dominio efector. Puede usarse una gama de enzimas modificadoras de la cromatina, metilasas o desmetilasas, para alterar la expresión del gen diana. Estos tipos de regulación epigenética tienen ciertas ventajas, particularmente porque presentan posibles efectos inespecíficos limitados.

La regulación de la transcripción y la traducción implica varias clases diferentes de sitios que interaccionan con proteínas o nucleótidos celulares. Con frecuencia, los sitios de unión a ADN de factores de transcripción u otras proteínas pueden seleccionarse como diana para la mutación o deleción para estudiar la función del sitio, aunque también pueden seleccionarse como diana para cambiar la expresión génica. Pueden añadirse sitios mediante unión de extremos no homólogos, NHEJ, o edición genómica directa mediante reparación dirigida por homología (HDR). El uso aumentado de secuenciación genómica, expresión de ARN y estudios de todo el genoma de unión a factor de transcripción han aumentado la capacidad de identificar cómo conducen los sitios a la regulación génica de desarrollo o temporal. Estos sistemas de control pueden ser directos o pueden implicar una extensa regulación cooperativa que puede requerir la integración de actividades a partir de múltiples potenciadores. Los factores de transcripción se unen normalmente a secuencias de ADN degeneradas de 6-12 pb de longitud. El bajo nivel de especificidad proporcionado por sitios individuales sugiere que complejas interacciones y reglas están implicadas en la unión y el resultado funcional. Los sitios de unión con menos degeneración pueden proporcionar medios de regulación más sencillos. Pueden diseñarse factores de transcripción artificiales para especificar secuencias más largas que tienen secuencias menos similares en el genoma y tienen una posibilidad inferior de escisión inespecífica. Cualquiera de estos tipos de sitios de unión pueden mutarse, deleccionarse o incluso crearse para permitir cambios en la regulación o expresión génica (Canver, M.C. *et al.*, Nature (2015)). Los factores de transcripción de GATA son una familia de factores de transcripción caracterizados por su capacidad para unirse a la secuencia de unión a ADN de GATA. Una secuencia de unión a GATA está ubicada en el sitio hipersensible de ADN (DHS) +58 del gen de BCL11A.

Otra clase de regiones reguladoras génicas que tienen estas características son los sitios de unión a microARN (miARN). Los miARN son ARN no codificantes que desempeñan papeles en la regulación génica postranscripcional. El miARN puede regular la expresión del 30 % de todos los genes que codifican para proteínas de mamíferos. Se descubrió un silenciamiento génico específico y potente mediante ARN bicatenario (iARN), más ARN no codificante pequeño adicional (Canver, M.C. *et al.*, Nature (2015)). La clase más grande de ARN no codificantes importantes para el silenciamiento génico son los miARN. En mamíferos, los miARN se transcriben en primer lugar como transcritos de ARN largos, que pueden ser unidades transcripcionales independientes, parte de intrones de proteína u otros transcritos. Los transcritos largos se denominan miARN primario (pri-miARN) que incluyen estructuras de tipo horquilla con emparejamiento de bases imperfecto. Estos pri-miARN pueden escindirse para dar uno o más miARN precursores más cortos (pre-miARN) mediante microprocesador, un complejo de proteína en el núcleo, que implica Drosha.

Los pre-miARN son bucles de tallo cortos de ~70 nucleótidos de longitud con una proyección en 3' de 2 nucleótidos que se exportan en los dúplex de miARN:miARN\* maduros de 19-25 nucleótidos. La cadena de miARN con estabilidad de emparejamiento de bases inferior (la cadena guía) puede estar ubicada en el complejo de silenciamiento inducido por ARN (RISC). La cadena guía pasajera (marcada con \*) puede ser funcional, pero habitualmente está degradada. El miARN maduro ancla RISC a motivos de secuencia parcialmente complementaria en ARNm diana predominantemente encontrados dentro de las regiones no traducidas en 3' (UTR) e induce silenciamiento génico postranscripcional (Bartel, D.P. Cell 136, 215-233 (2009); Saj, A. y Lai, E.C. Curr Opin Genet Dev 21, 504-510 (2011)).

Los miARN pueden ser importantes en el desarrollo, diferenciación, control de ciclo celular y crecimiento, y en prácticamente todas las rutas biológicas en mamíferos y otros organismos multicelulares. Los miARN también pueden estar implicados en el control de ciclo celular, apoptosis y diferenciación de células madre, hematopoyesis, hipoxia, desarrollo muscular, neurogénesis, secreción de insulina, metabolismo de colesterol, envejecimiento, replicación viral y respuestas inmunitarias.

Un único miARN puede seleccionar como diana cientos de transcritos de ARNm diferentes, mientras que un transcrito individual puede seleccionarse como diana por muchos miARN diferentes. Se han indicado más de 28645 microARN en la última versión de miRBase (v.21). Algunos miARN pueden codificarse por múltiples loci, algunos de los cuales pueden expresarse a partir de agrupaciones transcritas conjuntamente en tándem. Las características permiten complejas redes reguladoras con múltiples rutas y controles de realimentación. Los miARN pueden ser partes integrales de estos circuitos de realimentación y reguladores y pueden ayudar a regular la expresión génica manteniendo la producción de proteínas dentro de límites (Herranz, H. y Cohen, S.M. *Genes Dev* 24, 1339-1344 (2010); Posadas, D.M. y Carthew, R.W. *Curr Opin Genet Dev* 27, 1-6 (2014)).

Los miARN también pueden ser importantes en un gran número de enfermedades de seres humanos que están asociadas con una expresión de miARN anómala. Esta asociación destaca la importancia de la ruta reguladora de miARN. Recientes estudios de deleción de miARN han vinculado miARN con la regulación de las respuestas inmunitarias (Stern-Ginossar, N. *et al.*, *Science* 317, 376-381 (2007)).

Los miARN también tienen un fuerte vínculo con el cáncer y pueden desempeñar una función en diferentes tipos de cáncer. Se ha encontrado que los miARN están regulados por disminución en varios tumores. Los miARN pueden ser importantes en la regulación de rutas clave relacionadas con cáncer, tales como control del ciclo celular y respuesta al daño de ADN, y, por tanto, pueden usarse en el diagnóstico y pueden seleccionarse clínicamente como diana. Los microARN pueden regular de manera delicada el equilibrio de angiogénesis, de tal manera que experimentos que producen la depleción de todos los microARN suprimen la angiogénesis tumoral (Chen, S. *et al.*, *Genes Dev* 28, 1054-1067 (2014)).

Tal como se ha mostrado para genes que codifican para proteínas, también pueden someterse genes de miARN a cambios epigenéticos que se producen con cáncer. Muchos loci de miARN pueden estar asociados con islas de CpG que aumentan su oportunidad de regulación mediante metilación de ADN (Weber, B., Stresemann, C., Brueckner, B. y Lyko, F. *Cell Cycle* 6, 1001-1005 (2007)). La mayoría de los estudios han usado tratamiento con fármacos de remodelación de cromatina para revelar miARN silenciados de manera epigenética.

Además de su función en el silenciamiento del ARN, los miARN también pueden activar la traducción (Posadas, D.M. y Carthew, R.W. *Curr Opin Genet Dev* 27, 1-6 (2014)). La desactivación de estos sitios puede conducir a una expresión reducida del gen seleccionado como diana, mientras que la introducción de estos sitios puede aumentar la expresión.

Pueden desactivarse miARN individuales de la manera más eficaz mutando la secuencia semilla (bases 2-8 del microARN), lo cual puede ser importante para la especificidad de unión. La escisión en esta región, seguida por reparación errónea mediante NHEJ, puede eliminar eficazmente la función de miARN bloqueando la unión a sitios diana. También pueden inhibirse miARN mediante selección como diana específica de la región de bucle especial adyacente a la secuencia palindrómica. También puede usarse Cas9 inactiva de manera catalítica para inhibir la expresión de ARNhp (Zhao, Y. *et al.*, *Sci Rep* 4, 3943 (2014)). Además de seleccionar como diana el miARN, también pueden seleccionarse como diana los sitios de unión y mutarse para prevenir el silenciamiento mediante miARN.

#### Células humanas

Para mejorar las hemoglobinopatías, tal como se describe y se ilustra en el presente documento, las principales dianas para la edición génica son células humanas. Por ejemplo, en los métodos *ex vivo*, las células humanas pueden ser células somáticas, que, tras modificarse usando las técnicas tal como se describe, pueden dar lugar a células progenitoras. Por ejemplo, en los métodos *in vivo*, las células humanas pueden ser una célula de médula ósea, una célula progenitora hematopoyética o una célula CD34+.

Realizando edición génica en células autólogas que se derivan a partir de, y por tanto ya coinciden completamente con, el paciente que lo necesita, es posible generar células que pueden volver a introducirse de manera segura en el paciente, y dar lugar eficazmente a una población de células que pueden ser eficaces para mejorar uno o más estados clínicos asociados con la enfermedad del paciente.

Las células progenitoras (también denominadas células madre en el presente documento) pueden tanto proliferar como dar lugar a más células progenitoras, teniendo éstas a su vez la capacidad de generar un gran número de células madre que a su vez pueden dar lugar a células hija diferenciadas o diferenciables. Las propias células hija pueden inducirse para proliferar y producir progenie que posteriormente se diferencia para dar uno o más tipos de células maduras, al tiempo que también conservan una o más células con potencial de desarrollo parental. Entonces, el término "célula madre" se refiere a una célula con la capacidad o el potencial, en circunstancias particulares, de diferenciarse para dar un fenotipo más especializado o diferenciado, y que conserva la capacidad, en determinadas circunstancias, de proliferar sin diferenciarse sustancialmente. En un aspecto, el término célula progenitora o madre se refiere a una célula madre generalizada cuyos descendientes (progenie) se especializan, con frecuencia en diferentes sentidos, mediante diferenciación, por ejemplo, adquiriendo caracteres completamente individuales, tal como sucede en la diversificación progresiva de células y tejidos embrionarios. La diferenciación celular es un proceso complejo que se produce normalmente a lo largo de muchas divisiones celulares. Una célula diferenciada puede

derivarse de una célula multipotente que a su vez se deriva de una célula multipotente y así sucesivamente. Aunque puede considerarse que cada una de estas células multipotentes es una célula madre, la gama de tipos de células a los que puede dar lugar cada una puede variar considerablemente. Algunas células diferenciadas también tienen la capacidad de dar lugar a células con mayor potencial de desarrollo. Tal capacidad puede ser natural o puede inducirse artificialmente tras el tratamiento con diversos factores. En muchos casos biológicos, las células madre también pueden ser “multipotentes” porque pueden producir progenie de más de un tipo de célula distinto, pero esto no se requiere para la cualidad de ser “células madre”.

La autorrenovación puede ser otro aspecto importante de la célula madre. En teoría, la autorrenovación puede producirse mediante cualquiera de dos mecanismos principales. Las células madre pueden dividirse de manera asimétrica, conservando una hija el estado de célula madre y expresando la otra hija alguna función y fenotipo específico distinto. Alternativamente, algunas de las células madre en una población pueden dividirse de manera simétrica para dar dos madres, por tanto manteniendo algunas células madre en la población en su conjunto, mientras que otras células en la población dan lugar únicamente a progenie diferenciada. Generalmente, las “células progenitoras” tienen un fenotipo celular que es más primitivo (es decir, está en una etapa anterior a lo largo de una ruta de desarrollo o progresión de lo que lo está una célula completamente diferenciada). Con frecuencia, las células progenitoras también tienen un potencial proliferativo significativo o muy alto. Las células progenitoras pueden dar lugar a múltiples tipos de células diferenciadas distintos o a un único tipo de célula diferenciada, dependiendo de la ruta de desarrollo y del entorno en el que se desarrollan y se diferencian las células.

En el contexto de ontogenia celular, el adjetivo “diferenciado” o “diferenciante” es un término relativo. Una “célula diferenciada” es una célula que ha avanzado por la ruta de desarrollo más que la célula con la que está comparándose. Por tanto, las células madre pueden diferenciarse para dar células precursoras de linaje restringido (tales como una célula progenitora hematopoyética), que a su vez pueden diferenciarse para dar otros tipos de células precursoras más avanzadas en la ruta (tales como un precursor hematopoyético) y después hasta una célula diferenciada de fase final, tal como un eritrocito, que desempeña un papel característico en un determinado tipo de tejido, y puede o no conservar la capacidad de proliferar adicionalmente.

El término “célula progenitora hematopoyética” se refiere a células de un linaje de células madre que dan lugar a todos los tipos de células sanguíneas, incluyendo eritroides (eritrocitos o glóbulos rojos (RBC)), mieloides (monocitos y macrófagos, neutrófilos, basófilos, eosinófilos, megacariocitos / plaquetas y células dendríticas) y linfoides (células T, células B, células NK).

Una “célula del linaje eritroide” indica que la célula con la que está poniéndose en contacto es una célula que experimenta eritropoyesis, de tal manera que, tras la diferenciación final, forma un eritrocito o glóbulo rojo. Tales células se originan a partir de células progenitoras hematopoyéticas de médula ósea. Tras la exposición a factores de crecimiento específicos y otros componentes del microentorno hematopoyético, las células progenitoras hematopoyéticas pueden madurar lo suficiente a través de una serie de tipos celulares de diferenciación intermedia, todos tipos intermedios del linaje eritroide, para dar RBC. Por tanto, las células del “linaje eritroide” comprenden células progenitoras hematopoyéticas, rubriblastos, prorubricitos, eritroblastos, metarubricitos, reticulocitos y eritrocitos.

La célula progenitora hematopoyética puede expresar al menos uno de los siguientes marcadores de superficie celular característicos de células progenitoras hematopoyéticas: CD34+, CD59+, Thy1/CD90+, CD381o/- y C-kit/CD117+. En algunos ejemplos proporcionados en el presente documento, los progenitores hematopoyéticos pueden ser CD34+.

La célula progenitora hematopoyética puede ser una célula madre de sangre periférica obtenida a partir del paciente tras haberse tratado el paciente con uno o más factores tales como factor estimulante de colonias de granulocitos (opcionalmente en combinación con plerixafor). Las células CD34+ pueden enriquecerse usando sistema de selección celular CliniMACS® (Miltenyi Biotec). Las células CD34+ pueden estimularse in medio libre de suero (por ejemplo, medio CellGrow SCGM, CellGenix) con citocinas (por ejemplo, SCF, rhTPO, rhFLT3) antes de la edición genómica. Se contempla que la adición de SR1 y dmPGE2 y/u otros factores mejora el injerto a largo plazo.

Las células progenitoras hematopoyéticas del linaje eritroide pueden tener un marcador de superficie celular característico del linaje eritroide: tal como CD71 y Ter119.

Las células madre hematopoyéticas (HSC) pueden ser una diana importante para la terapia génica puesto que proporcionan una fuente prolongada de las células corregidas. Las HSC dan lugar a linajes tanto mieloide como linfóide de células sanguíneas. Las células sanguíneas maduras tienen una vida finita y deben remplazarse continuamente a lo largo de la vida. Las células sanguíneas se producen continuamente mediante la proliferación y diferenciación de una población de HSC pluripotentes que pueden reponerse mediante autorrenovación. La médula ósea (BM) es el sitio principal de hematopoyesis en seres humanos y una buena fuente de células progenitoras y madre hematopoyéticas (HSPC). Las HSPC pueden encontrarse en pequeños números en la sangre periférica (PB). En algunas indicaciones o tratamientos, sus números aumentan. La progenie de HSC madura a lo largo de fases, generando células progenitoras con potencial múltiple y linaje dedicado incluyendo las células progenitoras linfoides que dan lugar a las células que expresan BCL11A. Las células progenitoras B y T son las dos poblaciones celulares que requieren la actividad de BCL11A, de modo que pueden editarse en las fases antes de la nueva disposición,

aunque corregir progenitores tienen la ventaja de continuar siendo una fuente de células corregidas. Las células tratadas, tales como células CD34+, se devolverán al paciente. El nivel de injerto puede ser importante, al igual que la capacidad del injerto de linaje múltiple de las células de células sometidas a edición génica tras la infusión de CD34+ *in vivo*.

## Células madre pluripotentes inducidas

Las células humanas modificadas por ingeniería genética descritas en el presente documento pueden ser células madre pluripotentes inducidas (iPSC). Una ventaja de usar iPSC es que las células pueden derivarse del mismo sujeto al que se le tienen que administrar las células progenitoras. Es decir, puede obtenerse una célula somática a partir de un sujeto, reprogramarse para dar una célula madre pluripotente inducida, y después volver a diferenciarse para dar una célula progenitora que va a administrarse al sujeto (por ejemplo, células autólogas). Dado que los progenitores se derivan esencialmente de una fuente autóloga, puede reducirse el riesgo de rechazo de injerto o respuesta alérgica en comparación con el uso de células de otro sujeto o grupo de sujetos. Además, el uso de iPSC elimina la necesidad de células obtenidas a partir de una fuente embrionaria. Por tanto, en un aspecto, las células madre usadas en los métodos dados a conocer no son células madre embrionarias.

Aunque la diferenciación es generalmente irreversible en contextos fisiológicos, recientemente se han desarrollado varios métodos para reprogramar células somáticas para dar iPSC. Los expertos en la técnica conocen métodos a modo de ejemplo y se describen de manera resumida a continuación en el presente documento.

El término "reprogramar" se refiere a un proceso que altera o revierte el estado de diferenciación de una célula diferenciada (por ejemplo, una célula somática). Dicho de otro modo, reprogramar se refiere a un proceso de impulsar la diferenciación de una célula hacia atrás para dar un tipo de célula más indiferenciado o más primitivo. Debe observarse que colocar muchas células primarias en cultivo puede conducir a cierta pérdida de características completamente diferenciadas. Por tanto, cultivar simplemente tales células incluidas en el término células diferenciadas no hace que estas células se conviertan en células diferenciadas (por ejemplo, células diferenciadas) o células pluripotentes. La transición de una célula diferenciada a pluripotencia requiere un estímulo de reprogramación más allá de los estímulos que conducen a la pérdida parcial de carácter diferenciado en cultivo. Las células reprogramadas también tienen la característica de la capacidad de someterse a fases extendidas sin pérdida de potencial de crecimiento, con respecto a progenitores de células primarias, que generalmente tienen únicamente capacidad para un número limitado de divisiones en cultivo.

La célula que va a reprogramarse puede estar o bien parcial o bien terminalmente diferenciada antes de la reprogramación. La reprogramación puede abarcar una reversión completa del estado de diferenciación de una célula diferenciada (por ejemplo, una célula somática) hasta un estado pluripotente o un estado multipotente. La reprogramación puede abarcar una reversión completa o parcial del estado de diferenciación de una célula diferenciada (por ejemplo, una célula somática) para dar una célula no diferenciada (por ejemplo, una célula de tipo embrionario). La reprogramación puede dar como resultado la expresión de genes particulares por las células, cuya expresión contribuye adicionalmente a la reprogramación. En determinados ejemplos descritos en el presente documento, la reprogramación de una célula diferenciada (por ejemplo, una célula somática) puede hacer que la célula diferenciada adopte un estado no diferenciado (por ejemplo, es una célula no diferenciada). Las células resultantes se denominan "células reprogramadas" o "células madre pluripotentes inducidas (iPSC o células iPS)".

La reprogramación puede implicar la alteración, por ejemplo, inversión, de al menos algunos patrones heredables de modificación de ácido nucleico (por ejemplo, metilación), condensación de cromatina, cambios epigenéticos, impresión genómica, etc., que se producen durante la diferenciación celular. La reprogramación es distinta de mantener simplemente el estado no diferenciado existente de una célula que ya es pluripotente o mantener el estado existente diferenciado de manera menos que completa de una célula que ya es una célula multipotente (por ejemplo, una célula madre hematopoyética). La reprogramación también es distinta de fomentar la autorrenovación o proliferación de células que ya son pluripotentes o multipotentes, aunque las composiciones y métodos descritos en el presente documento también pueden usarse para tales fines, en algunos ejemplos.

En la técnica se conocen muchos métodos que pueden usarse para generar células madre pluripotentes a partir de células somáticas. Cualquier método de este tipo que reprograma una célula somática para dar el fenotipo pluripotente será apropiado para su uso en los métodos descritos en el presente documento.

Se han descrito metodologías de reprogramación para generar células pluripotentes usando combinaciones definidas de factores de transcripción. Pueden convertirse células somáticas de ratón en células de tipo célula ES con potencial de desarrollo expandido mediante la transducción directa de Oct4, Sox2, Klf4 y c-Myc; véase, por ejemplo, Takahashi y Yamanaka, *Cell* 126(4): 663-76 (2006). Las iPSC se asemejan a células ES, ya que restauran el circuito transcripcional asociado con pluripotencia y gran parte del entorno epigenético. Además, las iPSC de ratón satisfacen todos los ensayos convencionales para determinar pluripotencia: específicamente, diferenciación *in vitro* para dar tipos de células de las tres capas germinales, formación de teratoma, contribución a quimeras, transmisión de líneas germinales [véase, por ejemplo, Maherali y Hochedlinger, *Cell Stem Cell*. 3(6):595-605 (2008)] y complementación de tetraploide.

Pueden obtenerse iPSC humanas usando métodos de transducción similares, y se ha establecido el trío de factores de transcripción, OCT4, SOX2 y NANOG como conjunto principal de factores de transcripción que rigen la pluripotencia; véanse, por ejemplo, Budnatzky y Gepstein, *Stem Cells Transl Med.* 3(4):448-57 (2014); Barrett *et al.*, *Stem Cells Trans Med* 3:1-6 sctm.2014-0121 (2014); Focosi *et al.*, *Blood Cancer Journal* 4: e211 (2014); y referencias citadas en los mismos. La producción de iPSC puede lograrse mediante la introducción de secuencias de ácido nucleico que codifican para genes asociados con células madre en una célula somática de adulto, históricamente usando vectores virales.

Pueden generarse iPSC o derivarse a partir de células somáticas diferenciadas de manera terminal, así como de células madre de adulto o células madre somáticas. Es decir, una célula progenitora no pluripotente puede volverse pluripotente o multipotente mediante reprogramación. En tales casos, puede ser necesario incluir tantos factores de programación como se requiera para reprogramar una célula diferenciada de manera terminal. Además, puede inducirse reprogramación mediante la introducción no viral de factores de reprogramación, por ejemplo, mediante introducción de las propias proteínas, o mediante introducción de ácidos nucleicos que codifican para los factores de reprogramación, o mediante introducción de ARN mensajeros que, con la traducción, producen los factores de reprogramación (véase, por ejemplo, Warren *et al.*, *Cell Stem Cell*, 7(5):618-30 (2010)). Puede lograrse la reprogramación introduciendo una combinación de ácidos nucleicos que codifican para genes asociados con células madre, incluyendo, por ejemplo, Oct-4 (también conocido como Oct-3/4 o Pou5f1), Sox1, Sox2, Sox3, Sox 15, Sox 18, NANOG, Klf1, Klf2, Klf4, Klf5, NR5A2, c-Myc, 1-Myc, n-Myc, Rem2, Tert y LIN28. La reprogramación usando los métodos y composiciones descritos en el presente documento puede comprender además introducir uno o más de Oct-3/4, un miembro de la familia de Sox, un miembro de la familia de Klf y un miembro de la familia de Myc, en una célula somática. Los métodos y composiciones descritos en el presente documento pueden comprender además introducir uno o más de cada uno de Oct-4, Sox2, Nanog, c-MYC y Klf4 para la reprogramación. Tal como se indicó anteriormente, el método exacto usado para la reprogramación no es necesariamente crítico para los métodos y composiciones descritos en el presente documento. Sin embargo, cuando tienen que usarse células diferenciadas a partir de las células reprogramadas, por ejemplo, en terapia para seres humanos, en un aspecto la reprogramación no se realiza mediante un método que altera el genoma. Por tanto, en tales ejemplos, la reprogramación puede lograrse, por ejemplo, sin el uso de vectores virales o de plásmido.

La eficiencia de la reprogramación (es decir, el número de células reprogramadas) derivada a partir de una población de células de partida puede potenciarse mediante la adición de diversos agentes, por ejemplo, moléculas pequeñas, tal como se muestra por Shi *et al.*, *Cell-Stem Cell* 2:525-528 (2008); Huangfu *et al.*, *Nature Biotechnology* 26(7):795-797 (2008) y Marson *et al.*, *Cell-Stem Cell* 3: 132-135 (2008). Por tanto, puede usarse un agente o combinación de agentes que potencian la eficiencia o tasa de producción de células madre pluripotentes inducidas en la producción de iPSC específicas de paciente o específicas de enfermedad. Algunos ejemplos no limitativos de agentes que potencian la eficiencia de reprogramación incluyen Wnt soluble, medios condicionados con Wnt, BIX-01294 (una G9a histona metiltransferasa), PD0325901 (un inhibidor de MEK), inhibidores de ADN metiltransferasa, inhibidores de histona desacetilasa (HDAC), ácido valproico, 5'-azacitidina, dexametasona, suberoilanolida, ácido hidroxámico (SAHA), vitamina C y tricostatina (TSA), entre otros.

Otros ejemplos no limitativos de agentes potenciadores de la reprogramación incluyen: ácido suberoilanolida-hidroxámico (SAHA (por ejemplo, MK0683, vorinostat) y otros ácidos hidroxámicos), BML-210, depudecina (por ejemplo, (-)-depudecina), toxina HC, Nullscript (4-(1,3-dioxo-1H,3H-benzo[des]isoquinolin-2-il)-N-hidrobutanamida), fenilbutirato (por ejemplo, fenilbutirato de sodio) y ácido valproico ((VPA) y otros ácidos grasos de cadena corta), scriptaid, suramina de sodio, tricostatina A (TSA), compuesto APHA 8, apicidina, butirato de sodio, butirato de pivaloiloximetilo (Pivanex, AN-9), trapoxina B, clamidocina, depsipéptido (también conocido como FR901228 o FK228), benzamidas (por ejemplo, CI-994 (por ejemplo, N-acetil-dinalina) y MS-27-275), MGCD0103, NVP-LAQ-824, CBHA (ácido m-carboxicinámico, ácido bishidroxámico), JNJ16241199, tubacina, A-161906, proxamida, oxamflatina, 3-CI-UCHA (por ejemplo, ácido 6-(3-clorofenilureido)caproico-hidroxámico), AOE (ácido 2-amino-8-oxo-9,10-epoxidodecanoico), CHAP31 y CHAP 50. Otros agentes potenciadores de la reprogramación incluyen, por ejemplo, formas negativas dominantes de los HDAC (por ejemplo, formas inactivas de manera catalítica), inhibidores de ARNip de los HDAC, y anticuerpos que se unen específicamente a los HDAC. Tales inhibidores están disponibles, por ejemplo, de BIOMOL International, Fukasawa, Merck Biosciences, Novartis, Gloucester Pharmaceuticals, Titan Pharmaceuticals, MethylGene y Sigma Aldrich.

Para confirmar la inducción de células madre pluripotentes para su uso con los métodos descritos en el presente documento, pueden someterse a prueba clones aislados para determinar la expresión de un marcador de células madre. Tal expresión en una célula derivada a partir de una célula somática identifica las células como células madre pluripotentes inducidas. Los marcadores de células madre pueden seleccionarse del grupo no limitativo que incluye SSEA3, SSEA4, CD9, Nanog, Fbx15, Ecattl, Esgl, Eras, Gdf3, Fgf4, Cripto, Daxl, Zpf296, Slc2a3, Rexl, Utf1 y Natl. En un caso, por ejemplo, una célula que expresa Oct4 o Nanog se identifica como pluripotente. Los métodos para detectar la expresión de tales marcadores pueden incluir, por ejemplo, RT-PCR y métodos inmunológicos que detectan la presencia de los polipéptidos codificados, tales como inmunotransferencias de tipo Western o análisis por citometría de flujo. La detección puede implicar no solo RT-PCR, sino que también puede incluir la detección de marcadores de proteínas. Los marcadores intracelulares pueden identificarse mejor mediante RT-PCR o métodos de detección de

proteínas tales como inmunocitoquímica, mientras que los marcadores de superficie celular se identifican fácilmente, por ejemplo, mediante inmunocitoquímica.

El carácter de célula madre pluripotente de células aisladas puede confirmarse mediante pruebas que evalúan la capacidad de las iPSC para diferenciarse para dar células de cada una de las tres capas germinales. Como un ejemplo, puede usarse la formación de teratoma en ratones desnudos para evaluar el carácter pluripotente de los clones aislados. Las células pueden introducirse en ratones desnudos y puede realizarse histología y/o inmunohistoquímica en un tumor que surge a partir de las células. El crecimiento de un tumor que comprende células de las tres capas germinales, por ejemplo, indica adicionalmente que las células son células madre pluripotentes.

Crear iPSC específicas de paciente

Una etapa de los métodos *ex vivo* de la presente divulgación puede implicar crear una célula iPS específica de paciente, células iPS específicas de paciente o una línea de células iPS específicas de paciente. Hay muchos métodos establecidos en la técnica para crear células iPS específicas de paciente, tal como se describe en Takahashi y Yamanaka 2006; Takahashi, Tanabe *et al.* 2007. Por ejemplo, la etapa de creación puede comprender: a) aislar una célula somática, tal como una célula de la piel o fibroblasto, a partir del paciente; y b) introducir un conjunto de genes asociados con pluripotencia en la célula somática con el fin de inducir la célula para que se convierta en una célula madre pluripotente. El conjunto de genes asociados con pluripotencia pueden ser uno o más de los genes seleccionados del grupo que consiste en OCT4, SOX2, KLF4, Lin28, NANOG y cMYC.

Realizar una biopsia o aspirado de la médula ósea del paciente

Una biopsia o aspirado es una muestra de tejido o líquido extraída del cuerpo. Hay muchas clases diferentes de biopsias o aspirados. Casi todas implican usar una herramienta afilada para extraer una pequeña cantidad de tejido. Si la biopsia se realiza en la piel u otra zona sensible, en primer lugar puede aplicarse medicamento insensibilizante. Puede realizarse una biopsia o aspirado según cualquiera de los métodos conocidos en la técnica. Por ejemplo, en un aspirado de médula ósea, se usa una aguja grande para penetrar en el hueso de la pelvis para extraer médula ósea.

Aislar una célula madre mesenquimatosas

Pueden aislarse células madre mesenquimatosas según cualquier método conocido en la técnica, tal como a partir de la médula ósea o sangre periférica de un paciente. Por ejemplo, puede extraerse aspirado de médula al interior de una jeringa con heparina. Pueden lavarse las células y centrifugarse en un gradiente de densidad de Percoll™. Pueden separarse células, tales como células sanguíneas, células hepáticas, células intersticiales, macrófagos, mastocitos y timocitos, usando Percoll™. Pueden cultivarse las células en medio de Eagle modificado por Dulbecco (DMEM) (bajo contenido en glucosa) que contiene el 10 % de suero bovino fetal (FBS) (Pittinger MF, Mackay AM, Beck SC *et al.*, Science 1999; 284:143-147).

Tratar a un paciente con GCSF

Opcionalmente puede tratarse a un paciente con factor estimulante de colonias de granulocitos (GCSF) según cualquier método conocido en la técnica. El GCSF puede administrarse en combinación con plerixafor.

Aislar una célula progenitora hematopoyética a partir de un paciente

Puede aislarse una célula progenitora hematopoyética a partir de un paciente mediante cualquier método conocido en la técnica. Pueden enriquecerse células CD34+ usando sistema de selección celular CliniMACS® (Miltenyi Biotec). Pueden estimularse débilmente células CD34+ en medio libre de suero (por ejemplo, medio CellGrow SCGM, CellGenix) con citocinas (por ejemplo, SCF, rhTPO, rhFLT3) antes de la edición genómica.

Edición genómica

La edición genómica se refiere de manera general al procedimiento de modificar la secuencia de nucleótidos de un genoma, preferiblemente de una manera precisa o predeterminada. Los ejemplos de métodos de edición genómica descritos en el presente documento incluyen métodos de usar nucleasas dirigidas al sitio para cortar ácido desoxirribonucleico (ADN) en ubicaciones diana precisas en el genoma, creando de ese modo roturas de ADN de cadena sencilla o de cadena doble en ubicaciones particulares dentro del genoma. Tales roturas pueden repararse, y habitualmente se reparan, mediante procesos celulares endógenos naturales, tales como reparación dirigida por homología (HDR) y NHEJ, tal como se revisó recientemente en Cox *et al.*, Nature Medicine 21(2), 121-31 (2015). Estos dos procesos de reparación de ADN principales consisten en una familia de rutas alternativas. NHEJ une directamente los extremos de ADN resultantes de una rotura de cadena doble, algunas veces con la pérdida o adición de secuencia de nucleótidos, lo cual puede perturbar o potenciar la expresión génica. HDR usa una secuencia homóloga, o secuencia donante, como molde para insertar una secuencia de ADN definida en el punto de rotura. La secuencia homóloga puede estar en el genoma endógeno, tal como una cromátida hermana. Alternativamente, el donante puede ser un ácido nucleico exógeno, tal como un plásmido, un oligonucleótido monocatenario, un oligonucleótido



bicatenario, un oligonucleótido dúplex o un virus, que tiene regiones de alta homología con el locus escindido por nucleasa, pero que también pueden contener secuencias o cambios de secuencia adicionales incluyendo delecciones que pueden incorporarse en el locus diana escindido. Un tercer mecanismo de reparación puede ser unión de extremos mediada por microhomología (MMEJ), también denominada “NHEJ alternativa”, en la que el resultado genético es similar a NHEJ en cuanto a que pueden producirse pequeñas delecciones e inserciones en el sitio de escisión. MMEJ puede usar secuencias homólogas de unos pocos pares de bases que flanquean el sitio de rotura de ADN para impulsar un resultado de reparación por unión de extremos de ADN más favorecido, y recientes informes han esclarecido adicionalmente el mecanismo molecular de este proceso; véase, por ejemplo, Cho y Greenberg, *Nature* 518, 174-76 (2015); Kent *et al.*, *Nature Structural and Molecular Biology*, Adv. Online doi: 10.1038/nsmb.2961 (2015); Mateos-Gomez *et al.*, *Nature* 518, 254-57 (2015); Ceccaldi *et al.*, *Nature* 528, 258-62 (2015). En algunos casos, puede ser posible predecir resultados de reparación probables basándose en análisis de posibles microhomologías en el sitio de la rotura de ADN.

Cada uno de estos mecanismos de edición genómica puede usarse para crear alteraciones genómicas deseadas. Una etapa en el procedimiento de edición genómica puede ser crear una o dos roturas de ADN, estas últimas como roturas de cadena doble o como dos roturas de cadena sencilla, en el locus diana lo más cerca posible del sitio de mutación prevista. Esto puede lograrse mediante el uso de polipéptidos dirigidos al sitio, tal como se describe y se ilustra en el presente documento.

Los polipéptidos dirigidos al sitio, tales como una ADN endonucleasa, pueden introducir roturas de cadena doble o roturas de cadena sencilla en ácidos nucleicos, por ejemplo, ADN genómico. La rotura de cadena doble puede estimular las rutas de reparación de ADN endógenas de una célula (por ejemplo, reparación dependiente de homología o unión de extremos no homólogos o unión de extremos no homólogos alternativa (A-NHEJ) o unión de extremos mediada por microhomología). NHEJ puede reparar ácido nucleico diana escindido sin necesidad de un molde homólogo. Algunas veces esto puede dar como resultado pequeñas delecciones o inserciones (indeles) en el ácido nucleico diana en el sitio de escisión, y puede conducir a la perturbación o alteración de expresión génica. Puede producirse HDR cuando se dispone de un molde de reparación homólogo, o donante. El molde donante homólogo puede comprender secuencias que pueden ser homólogas con respecto a secuencias que flanquean el sitio de escisión de ácido nucleico diana. La célula puede usar la cromátida hermana como molde de reparación. Sin embargo, para los fines de edición genómica, el molde de reparación puede suministrarse como un ácido nucleico exógeno, tal como un plásmido, oligonucleótido dúplex, oligonucleótido monocatenario, oligonucleótido bicatenario o ácido nucleico viral. Con moldes donantes exógenos, puede introducirse una secuencia de ácido nucleico (tal como un transgén) o modificación (tal como un cambio o delección de una única o de múltiples bases) adicional entre las regiones flanqueantes de homología de modo que la secuencia de ácido nucleico adicional o alterada también se incorpora en el locus diana. MMEJ puede dar como resultado un resultado genético que es similar a NHEJ en cuanto a que pueden producirse pequeñas delecciones e inserciones en el sitio de escisión. MMEJ puede usar secuencias homólogas de unos pocos pares de bases que flanquean el sitio de escisión para impulsar un resultado de reparación de ADN por unión de extremos favorecido. En algunos casos, puede ser posible predecir resultados de reparación probables basándose en análisis de posibles microhomologías en las regiones diana de nucleasa.

Por tanto, en algunos casos, puede usarse recombinación homóloga para insertar una secuencia de polinucleótido exógena en el sitio de escisión de ácido nucleico diana. Una secuencia de polinucleótido exógena se denomina polinucleótido donante (o donante o secuencia donante o molde donante de polinucleótido) en el presente documento. El polinucleótido donante, una porción del polinucleótido donante, una copia del polinucleótido donante o una porción de una copia del polinucleótido donante puede insertarse en el sitio de escisión de ácido nucleico diana. El polinucleótido donante puede ser una secuencia de polinucleótido exógena, es decir, una secuencia que no se produce de manera natural en el sitio de escisión de ácido nucleico diana.

Las modificaciones del ADN diana debido a NHEJ y/o HDR pueden conducir, por ejemplo, a mutaciones, delecciones, alteraciones, integraciones, corrección génica, sustitución génica, marcado génico, inserción de transgenes, delección de nucleótidos, perturbación génica, translocaciones y/o mutación génica. Los procedimientos de delección de ADN genómico e integración de ácido nucleico no nativo en ADN genómico son ejemplos de edición genómica.

#### Sistema de endonucleasa CRISPR

Un locus genómico de CRISPR (repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas) puede encontrarse en los genomas de muchas procariotas (por ejemplo, bacterias y arqueas). En procariotas, el locus de CRISPR codifica para productos que funcionan como un tipo de sistema inmunitario para ayudar a defender a las procariotas frente a invasores foráneos, tales como virus y fagos. Hay tres fases de función de locus de CRISPR: integración de nuevas secuencias en el locus de CRISPR, expresión de ARN de CRISPR (ARNcr) y silenciamiento de ácido nucleico invasor foráneo. Se han identificado cinco tipos de sistemas de CRISPR (por ejemplo, tipo I, tipo II, tipo III, tipo U y tipo V).

Un locus de CRISPR incluye varias secuencias de repetición cortas denominadas “repeticiones”. Cuando se expresan, las repeticiones pueden formar estructuras secundarias (por ejemplo, horquillas) y/o comprenden secuencias monocatenarias no estructuradas. Las repeticiones se producen habitualmente en agrupaciones y con frecuencia

divergen entre especies. Las repeticiones están regularmente interespaciadas con secuencias intermedias únicas denominadas “espaciadores”, dando como resultado una arquitectura de locus de repetición-espaciador-repetición. Los espaciadores son idénticos a, o tienen una alta homología con, secuencias invasoras foráneas conocidas. Una unidad de espaciador-repetición codifica para un ARNcrispr (ARNcr), que se procesa para dar una forma madura de la unidad de espaciador-repetición. Un ARNcr comprende una “semilla” o secuencia espaciadora que participa en la selección como diana de un ácido nucleico diana (en la forma que se produce de manera natural en procariotas, la secuencia espaciadora selecciona como diana el ácido nucleico invasor foráneo). Una secuencia espaciadora está ubicada en el extremo 5' o 3' del ARNcr.

Un locus de CRISPR también comprende secuencias de polinucleótido que codifican para genes asociados con CRISPR (Cas). Los genes de Cas codifican para endonucleasas implicadas en la biogénesis y las fases de interferencia de la función de ARNcr en procariotas. Algunos genes de Cas comprenden estructuras secundarias y/o terciarias homólogas.

#### Sistemas de CRISPR de tipo II

La biogénesis de ARNcr en un sistema de CRISPR de tipo II en la naturaleza requiere un ARN de CRISPR de transactivación (ARNtracr). El ARNtracr puede modificarse mediante ARNasa III endógena, y después se hibrida con una repetición de ARNcr en la matriz de pre-ARNcr. La ARNasa III endógena puede reclutarse para escindir el pre-ARNcr. Los ARNcr escindidos pueden someterse a recorte por exorribonucleasa para producir la forma de ARNcr maduro (por ejemplo, recorte en 5'). El ARNtracr puede permanecer hibridado al ARNcr, y el ARNtracr y el ARNcr se asocian con un polipéptido dirigido al sitio (por ejemplo, Cas9). El ARNcr del complejo de ARNcr-ARNtracr-Cas9 puede guiar el complejo a un ácido nucleico diana con el que puede hibridarse el ARNcr. La hibridación del ARNcr con el ácido nucleico diana puede activar Cas9 para la escisión de ácido nucleico diana. El ácido nucleico diana en un sistema de CRISPR de tipo II se denomina motivo adyacente de protoespaciador (PAM). En la naturaleza, el PAM es esencial para facilitar la unión de un polipéptido dirigido al sitio (por ejemplo, Cas9) al ácido nucleico diana. Los sistemas de tipo II (también denominados Nmeni o CASS4) se subdividen adicionalmente en tipo II-A (CASS4) y II-B (CASS4a). Jinek *et al.*, Science, 337(6096):816-821 (2012) mostraron que el sistema de CRISPR/Cas9 es útil para la edición genómica programable por ARN y la publicación de solicitud de patente internacional número WO2013/176772 proporciona numerosos ejemplos y aplicaciones del sistema de endonucleasa CRISPR/Cas para edición génica específica del sitio.

#### Sistemas de CRISPR de tipo V

Los sistemas de CRISPR de tipo V tienen varias diferencias importantes con respecto a sistemas de tipo II. Por ejemplo, Cpf1 es una endonucleasa guiada por ARN individual que, a diferencia de los sistemas de tipo II, carece de ARNtracr. De hecho, las matrices de CRISPR asociadas con Cpf1 pueden procesarse en la naturaleza para dar ARNcr maduros sin el requisito de un ARNtracr de transactivación adicional. La matriz de CRISPR de tipo V puede procesarse para dar ARNcr maduros cortos de 42-44 nucleótidos de longitud, comenzando cada ARNcr maduro con 19 nucleótidos de repetición directa seguidos por 23-25 nucleótidos de secuencia espaciadora. En cambio, los ARNcr maduros en los sistemas de tipo II pueden comenzar con 20-24 nucleótidos de secuencia espaciadora seguidos por aproximadamente 22 nucleótidos de repetición directa. Además, Cpf1 puede usar un motivo adyacente de protoespaciador rico en T de tal manera que complejos de Cpf1-ARNcr escinden de manera eficiente ADN diana precedido por un PAM rico en T corto, lo cual contrasta con el PAM rico en G que sigue al ADN diana para los sistemas de tipo II. Por tanto, los sistemas de tipo V escinden en un punto que está alejado del PAM, mientras que los sistemas de tipo II escinden en un punto que está adyacente al PAM. Además, a diferencia de los sistemas de tipo II, Cpf1 escinde ADN mediante una rotura de cadena doble de ADN escalonada con una proyección en 5' de 4 o 5 nucleótidos. Los sistemas de tipo II escinden mediante una rotura de cadena doble de extremos romos. De manera similar a los sistemas de tipo II, Cpf1 contiene un dominio de endonucleasa de tipo RuvC predicho, pero carece de un segundo dominio de endonucleasa de HNH, lo cual contrasta con los sistemas de tipo II.

#### Genes/polipéptidos de Cas y motivos adyacentes de protoespaciador

Los polipéptidos de CRISPR/Cas a modo de ejemplo incluyen los polipéptidos de Cas9 en la figura 1 de Fonfara *et al.*, Nucleic Acids Research, 42: 2577-2590 (2014). El sistema de nomenclatura de genes de CRISPR/Cas ha experimentado una extensa reescritura desde que se descubrieron los genes de Cas. La figura 5 de Fonfara, citado anteriormente, proporciona secuencias de PAM para los polipéptidos de Cas9 de diversas especies.

#### Polipéptidos dirigidos al sitio

Un polipéptido dirigido al sitio es una nucleasa usada en la edición genómica para escindir ADN. La nucleasa o polipéptido dirigido al sitio puede administrarse a una célula o a un paciente: o bien como uno o más polipéptidos, o bien como uno o más ARNm que codifican para el polipéptido.

En el contexto de un sistema de CRISPR/Cas o CRISPR/Cpf1, el polipéptido dirigido al sitio puede unirse a un ARN guía que, a su vez, especifica el sitio en el ADN diana al que se dirige el polipéptido. En los sistemas de CRISPR/Cas

o CRISPR/Cpf1 dados a conocer en el presente documento, el polipéptido dirigido al sitio puede ser una endonucleasa, tal como una ADN endonucleasa.

Un polipéptido dirigido al sitio puede comprender una pluralidad de dominios de escisión de ácido nucleico (es decir, nucleasa). Dos o más dominios de escisión de ácido nucleico pueden estar unidos entre sí mediante un ligador. Por ejemplo, el ligador puede comprender un ligador flexible. Los ligadores pueden comprender 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40 o más aminoácidos de longitud.

Las enzimas Cas9 de tipo natural que se producen de manera natural comprenden dos dominios de nucleasa, un dominio de nucleasa de HNH y un dominio de RuvC. En el presente documento, "Cas9" se refiere a Cas9 tanto que se producen de manera natural como recombinantes. Las enzimas Cas9 contempladas en el presente documento pueden comprender un dominio de nucleasa de HNH o de tipo HNH y/o un dominio de nucleasa de RuvC o de tipo RuvC.

Los dominios de HNH o de tipo HNH comprenden un pliegue de tipo McrA. Los dominios de HNH o de tipo HNH comprenden dos cadenas  $\beta$  antiparalelas y una hélice  $\alpha$ . Los dominios de HNH o de tipo HNH comprenden un sitio de unión a metal (por ejemplo, un sitio de unión a catión divalente). Los dominios de HNH o de tipo HNH pueden escindir una cadena de un ácido nucleico diana (por ejemplo, la cadena complementaria de la cadena dirigida a ARNcr).

Los dominios de RuvC o de tipo RuvC comprenden un pliegue de ARNasa H o de tipo ARNasa H. Los dominios de RuvC/ARNasa H están implicados en un conjunto diverso de funciones basadas en ácido nucleico incluyendo actuar tanto sobre ARN como sobre ADN. El dominio de ARNasa H comprende 5 cadenas  $\beta$  rodeadas por una pluralidad de hélices  $\alpha$ . Los dominios de RuvC/ARNasa H o de tipo RuvC/ARNasa H comprenden un sitio de unión a metal (por ejemplo, un sitio de unión a catión divalente). Los dominios de RuvC/ARNasa H o de tipo RuvC/ARNasa H pueden escindir una cadena de un ácido nucleico diana (por ejemplo, la cadena no complementaria de un ADN diana bicatenario).

Los polipéptidos dirigidos al sitio pueden introducir roturas de cadena doble o roturas de cadena sencilla en ácidos nucleicos, por ejemplo, ADN genómico. La rotura de cadena doble puede estimular las rutas de reparación de ADN endógenas de una célula (por ejemplo, reparación dependiente de homología (HDR) o NHEJ o unión de extremos no homólogos alternativa (A-NHEJ) o unión de extremos mediada por microhomología (MMEJ)). NHEJ puede reparar ácido nucleico diana escindido sin necesidad de un molde homólogo. Algunas veces esto puede dar como resultado pequeñas deleciones o inserciones (indeles) en el ácido nucleico diana en el sitio de escisión y puede conducir a la perturbación o alteración de la expresión génica. Puede producirse HDR cuando se dispone de un molde de reparación homólogo, o donante. El molde donante homólogo puede comprender secuencias que son homólogas con respecto a secuencias que flanquean el sitio de escisión de ácido nucleico diana. La célula puede usar la cromátida hermana como molde de reparación. Sin embargo, para los fines de edición genómica, el molde de reparación puede suministrarse como un ácido nucleico exógeno, tal como un plásmido, oligonucleótido dúplex, oligonucleótido monocatenario o ácido nucleico viral. Con moldes donantes exógenos, puede introducirse una secuencia de ácido nucleico (tal como un transgén) o modificación (tal como un cambio o deleción de una única o de múltiples bases) adicional entre las regiones flanqueantes de homología de modo que la secuencia de ácido nucleico adicional o alterada también se incorpora en el locus diana. MMEJ puede dar como resultado un resultado genético que es similar a NHEJ en cuanto a que pueden producirse pequeñas deleciones e inserciones en el sitio de escisión. MMEJ puede usar secuencias homólogas de unos pocos pares de bases que flanquean el sitio de escisión para impulsar un resultado de reparación de ADN por unión de extremos favorecido. En algunos casos, puede ser posible predecir resultados de reparación probables basándose en análisis de posibles microhomologías en las regiones diana de nucleasa.

Por tanto, en algunos casos, puede usarse recombinación homóloga para insertar una secuencia de polinucleótido exógena en el sitio de escisión de ácido nucleico diana. Una secuencia de polinucleótido exógena se denomina polinucleótido donante (o donante o secuencia donante) en el presente documento. El polinucleótido donante, una porción del polinucleótido donante, una copia del polinucleótido donante o una porción de una copia del polinucleótido donante puede insertarse en el sitio de escisión de ácido nucleico diana. El polinucleótido donante puede ser una secuencia de polinucleótido exógena, es decir, una secuencia que no se produce de manera natural en el sitio de escisión de ácido nucleico diana.

Las modificaciones del ADN diana debido a NHEJ y/o HDR pueden conducir, por ejemplo, a mutaciones, deleciones, alteraciones, integraciones, corrección génica, sustitución génica, marcado génico, inserción de transgenes, deleción de nucleótidos, perturbación génica, translocaciones y/o mutación génica. Los procedimientos de deleción de ADN genómico e integración de ácido nucleico no nativo en ADN genómico son ejemplos de edición genómica.

El polipéptido dirigido al sitio puede comprender una secuencia de aminoácidos que tiene al menos el 10 %, al menos el 15 %, al menos el 20 %, al menos el 30 %, al menos el 40 %, al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 75 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 99 % o el 100 % de identidad de secuencia de aminoácidos con respecto a un polipéptido dirigido al sitio a modo de ejemplo de tipo natural [por ejemplo, Cas9 de *S. pyogenes*, ID de secuencia n.º 8 del documento US2014/0068797 o

Sapranaukas *et al.*, Nucleic Acids Res, 39(21): 9275-9282 (2011)], y varios otros polipéptidos dirigidos al sitio. El polipéptido dirigido al sitio puede comprender al menos el 70, 75, 80, 85, 90, 95, 97, 99 o 100 % de identidad con respecto a un polipéptido dirigido al sitio de tipo natural (por ejemplo, Cas9 de *S. pyogenes*, citado anteriormente) a lo largo de 10 aminoácidos contiguos. El polipéptido dirigido al sitio puede comprender como máximo: el 70, 75, 80, 85, 90, 95, 97, 99 o 100 % de identidad con respecto a un polipéptido dirigido al sitio de tipo natural (por ejemplo, Cas9 de *S. pyogenes*, citado anteriormente) a lo largo de 10 aminoácidos contiguos. El polipéptido dirigido al sitio puede comprender al menos: el 70, 75, 80, 85, 90, 95, 97, 99 o 100 % de identidad con respecto a un polipéptido dirigido al sitio de tipo natural (por ejemplo, Cas9 de *S. pyogenes*, citado anteriormente) a lo largo de 10 aminoácidos contiguos en un dominio de nucleasa de HNH del polipéptido dirigido al sitio. El polipéptido dirigido al sitio puede comprender como máximo: el 70, 75, 80, 85, 90, 95, 97, 99 o 100 % de identidad con respecto a un polipéptido dirigido al sitio de tipo natural (por ejemplo, Cas9 de *S. pyogenes*, citado anteriormente) a lo largo de 10 aminoácidos contiguos en un dominio de nucleasa de HNH del polipéptido dirigido al sitio. El polipéptido dirigido al sitio puede comprender al menos: el 70, 75, 80, 85, 90, 95, 97, 99 o 100 % de identidad con respecto a un polipéptido dirigido al sitio de tipo natural (por ejemplo, Cas9 de *S. pyogenes*, citado anteriormente) a lo largo de 10 aminoácidos contiguos en un dominio de nucleasa de RuvC del polipéptido dirigido al sitio. El polipéptido dirigido al sitio puede comprender como máximo: el 70, 75, 80, 85, 90, 95, 97, 99 o 100 % de identidad con respecto a un polipéptido dirigido al sitio de tipo natural (por ejemplo, Cas9 de *S. pyogenes*, citado anteriormente) a lo largo de 10 aminoácidos contiguos en un dominio de nucleasa de RuvC del polipéptido dirigido al sitio.

El polipéptido dirigido al sitio puede comprender una forma modificada de un polipéptido dirigido al sitio a modo de ejemplo de tipo natural. La forma modificada del polipéptido dirigido al sitio a modo de ejemplo de tipo natural puede comprender una mutación que reduce la actividad de escisión de ácido nucleico del polipéptido dirigido al sitio. La forma modificada del polipéptido dirigido al sitio a modo de ejemplo de tipo natural puede tener menos del 90 %, menos del 80 %, menos del 70 %, menos del 60 %, menos del 50 %, menos del 40 %, menos del 30 %, menos del 20 %, menos del 10 %, menos del 5 % o menos del 1 % de la actividad de escisión de ácido nucleico del polipéptido dirigido al sitio a modo de ejemplo de tipo natural (por ejemplo, Cas9 de *S. pyogenes*, citado anteriormente). La forma modificada del polipéptido dirigido al sitio puede no tener ninguna actividad de escisión de ácido nucleico sustancial. Cuando un polipéptido dirigido al sitio es una forma modificada que no tiene ninguna actividad de escisión de ácido nucleico sustancial, se denomina en el presente documento “enzimáticamente inactivo”.

La forma modificada del polipéptido dirigido al sitio puede comprender una mutación de tal manera que puede inducir una rotura de cadena sencilla (SSB) en un ácido nucleico diana (por ejemplo, cortando únicamente una de las estructuras principales de fosfato de azúcar de un ácido nucleico diana bicatenario). La mutación puede dar como resultado menos del 90 %, menos del 80 %, menos del 70 %, menos del 60 %, menos del 50 %, menos del 40 %, menos del 30 %, menos del 20 %, menos del 10 %, menos del 5 % o menos del 1 % de la actividad de escisión de ácido nucleico en uno o más de la pluralidad de dominios de escisión de ácido nucleico del polipéptido dirigido al sitio de tipo natural (por ejemplo, Cas9 de *S. pyogenes*, citado anteriormente). La mutación puede dar como resultado que uno o más de la pluralidad de dominios de escisión de ácido nucleico conserven la capacidad para escindir la cadena complementaria del ácido nucleico diana, pero reduzcan su capacidad para escindir la cadena no complementaria del ácido nucleico diana. La mutación puede dar como resultado que uno o más de la pluralidad de dominios de escisión de ácido nucleico conserven la capacidad para escindir la cadena no complementaria del ácido nucleico diana, pero reduzcan su capacidad para escindir la cadena complementaria del ácido nucleico diana. Por ejemplo, residuos en el polipéptido de Cas9 de *S. pyogenes* a modo de ejemplo de tipo natural, tales como Asp10, His840, Asn854 y Asn856, se mutan para inactivar uno o más de la pluralidad de dominios de escisión de ácido nucleico (por ejemplo, dominios de nucleasa). Los residuos que van a mutarse pueden corresponder a residuos Asp10, His840, Asn854 y Asn856 en el polipéptido de Cas9 de *S. pyogenes* a modo de ejemplo de tipo natural (por ejemplo, tal como se determina mediante alineación de secuencias y/o estructural). Los ejemplos no limitativos de mutaciones incluyen D10A, H840A, N854A o N856A. Un experto en la técnica reconocerá que mutaciones distintas de sustituciones de alanina pueden ser adecuadas.

Puede combinarse una mutación D10A con una o más de las mutaciones H840A, N854A o N856A para producir un polipéptido dirigido al sitio que carece sustancialmente de actividad de escisión de ADN. Puede combinarse una mutación H840A con una o más de las mutaciones D10A, N854A o N856A para producir un polipéptido dirigido al sitio que carece sustancialmente de actividad de escisión de ADN. Puede combinarse una mutación N854A con una o más de las mutaciones H840A, D10A o N856A para producir un polipéptido dirigido al sitio que carece sustancialmente de actividad de escisión de ADN. Puede combinarse una mutación N856A con una o más de las mutaciones H840A, N854A o D10A para producir un polipéptido dirigido al sitio que carece sustancialmente de actividad de escisión de ADN. Los polipéptidos dirigidos al sitio que comprenden un dominio de nucleasa sustancialmente inactivo se denominan “enzimas melladoras”.

Pueden usarse variantes de enzimas melladoras de endonucleasas guiadas por ARN, por ejemplo Cas9, para aumentar la especificidad de la edición genómica mediada por CRISPR. Cas9 de tipo natural se guía normalmente mediante un único ARN guía diseñado para hibridarse con una secuencia de ~20 nucleótidos especificada en la secuencia diana (tal como un locus genómico endógeno). Sin embargo, pueden tolerarse varios apareamientos erróneos entre el ARN guía y el locus diana, reduciendo eficazmente la longitud de homología requerida en el sitio diana, por ejemplo, hasta tan solo 13 nt de homología, y dando por tanto como resultado un potencial elevado de unión

y escisión de ácido nucleico bicatenario por el complejo de CRISPR/Cas9 en otra parte en el genoma diana (también conocido como escisión inespecífica). Dado que las variantes de enzima melladora de Cas9 solo cortan, cada una, una cadena, con el fin de crear una rotura de cadena doble es necesario que un par de enzimas melladoras se unan en estrecha proximidad y en cadenas opuestas del ácido nucleico diana, creando de ese modo un par de mellas, lo cual es el equivalente de una rotura de cadena doble. Esto requiere que dos ARN guía independientes (uno por cada enzima melladora) deben unirse en estrecha proximidad y en cadenas opuestas del ácido nucleico diana. Este requisito duplica esencialmente la longitud mínima de homología necesaria para que se produzca la rotura de cadena doble, reduciendo de ese modo la probabilidad de que se produzca un acontecimiento de escisión de cadena doble en otra parte en el genoma, en el que es poco probable que los dos sitios de ARN guía (si existen) estén lo suficientemente cerca uno de otro como para permitir que se forme la rotura de cadena doble. Tal como se describe en la técnica, también pueden usarse enzimas melladoras para fomentar HDR frente a NHEJ. Puede usarse HDR para introducir cambios seleccionados en sitios diana en el genoma mediante el uso de secuencias donantes específicas que median eficazmente en los cambios deseados.

Las mutaciones contempladas pueden incluir sustituciones, adiciones y deleciones, o cualquier combinación de las mismas. La mutación convierte el aminoácido mutado en alanina. La mutación convierte el aminoácido mutado en otro aminoácido (por ejemplo, glicina, serina, treonina, cisteína, valina, leucina, isoleucina, metionina, prolina, fenilalanina, tirosina, triptófano, ácido aspártico, ácido glutámico, asparaginas, glutamina, histidina, lisina o arginina). La mutación convierte el aminoácido mutado en un aminoácido no natural (por ejemplo, selenometionina). La mutación convierte el aminoácido mutado en agentes miméticos de aminoácido (por ejemplo, agentes fosfomiméticos). La mutación puede ser una mutación conservativa. Por ejemplo, la mutación convierte el aminoácido mutado en aminoácidos que se asemejan en cuanto a tamaño, forma, carga, polaridad, conformación y/o rotámeros de los aminoácidos mutados (por ejemplo, mutación de cisteína/serina, mutación de lisina/asparagina, mutación de histidina/fenilalanina). La mutación puede provocar un desplazamiento en el marco de lectura y/o la creación de un codón de terminación prematuro. Las mutaciones pueden provocar cambios en regiones reguladoras de genes o loci que afectan a la expresión de uno o más genes.

El polipéptido dirigido al sitio (por ejemplo, polipéptido dirigido al sitio variante, mutado, enzimáticamente inactivo y/o enzimáticamente inactivo de manera condicional) puede seleccionar como diana ácido nucleico. El polipéptido dirigido al sitio (por ejemplo, endorribonucleasa variante, mutada, enzimáticamente inactiva y/o enzimáticamente inactiva de manera condicional) puede seleccionar como diana ADN. El polipéptido dirigido al sitio (por ejemplo, endorribonucleasa variante, mutada, enzimáticamente inactiva y/o enzimáticamente inactiva de manera condicional) puede seleccionar como diana ARN.

El polipéptido dirigido al sitio puede comprender una o más secuencias no nativas (por ejemplo, el polipéptido dirigido al sitio es una proteína de fusión).

El polipéptido dirigido al sitio puede comprender una secuencia de aminoácidos que comprende al menos el 15 % de identidad de aminoácidos con respecto a una Cas9 de una bacteria (por ejemplo, *S. pyogenes*), un dominio de unión a ácido nucleico y dos dominios de escisión de ácido nucleico (es decir, un dominio de HNH y un dominio de RuvC).

El polipéptido dirigido al sitio puede comprender una secuencia de aminoácidos que comprende al menos el 15 % de identidad de aminoácidos con respecto a una Cas9 de una bacteria (por ejemplo, *S. pyogenes*) y dos dominios de escisión de ácido nucleico (es decir, un dominio de HNH y un dominio de RuvC).

El polipéptido dirigido al sitio puede comprender una secuencia de aminoácidos que comprende al menos el 15 % de identidad de aminoácidos con respecto a una Cas9 de una bacteria (por ejemplo, *S. pyogenes*) y dos dominios de escisión de ácido nucleico, en el que uno o ambos de los dominios de escisión de ácido nucleico comprenden al menos el 50 % de identidad de aminoácidos con respecto a un dominio de nucleasa de Cas9 de una bacteria (por ejemplo, *S. pyogenes*).

El polipéptido dirigido al sitio puede comprender una secuencia de aminoácidos que comprende al menos el 15 % de identidad de aminoácidos con respecto a una Cas9 de una bacteria (por ejemplo, *S. pyogenes*), dos dominios de escisión de ácido nucleico (es decir, un dominio de HNH y un dominio de RuvC) y secuencia no nativa (por ejemplo, una señal de localización nuclear) o un ligador que une el polipéptido dirigido al sitio a una secuencia no nativa.

El polipéptido dirigido al sitio puede comprender una secuencia de aminoácidos que comprende al menos el 15 % de identidad de aminoácidos con respecto a una Cas9 de una bacteria (por ejemplo, *S. pyogenes*), dos dominios de escisión de ácido nucleico (es decir, un dominio de HNH y un dominio de RuvC), en el que el polipéptido dirigido al sitio comprende una mutación en uno o ambos de los dominios de escisión de ácido nucleico que reduce la actividad de escisión de los dominios de nucleasa en al menos el 50 %.

El polipéptido dirigido al sitio puede comprender una secuencia de aminoácidos que comprende al menos el 15 % de identidad de aminoácidos con respecto a una Cas9 de una bacteria (por ejemplo, *S. pyogenes*) y dos dominios de escisión de ácido nucleico (es decir, un dominio de HNH y un dominio de RuvC), en el que uno de los dominios de nucleasa comprende una mutación de ácido aspártico 10 y/o en el que uno de los dominios de nucleasa puede

comprender una mutación de histidina 840 y en el que la mutación reduce la actividad de escisión del/de los dominio(s) de nucleasa en al menos el 50 %.

El uno o más polipéptidos dirigidos al sitio, por ejemplo ADN endonucleasas, pueden comprender dos enzimas melladoras que, juntas, realizan una rotura de cadena doble en un locus específico en el genoma, o cuatro enzimas melladoras que, juntas, realizan o provocan dos roturas de cadena doble en loci específicos en el genoma. Alternativamente, un polipéptido dirigido al sitio, por ejemplo ADN endonucleasa, puede realizar o provocar una rotura de cadena doble en un locus específico en el genoma.

El polipéptido dirigido al sitio puede estar flanqueado en el extremo N-terminal, el extremo C-terminal o tanto el extremo N-terminal como el extremo C-terminal por una o más señales de localización nuclear (NLS). Por ejemplo, una endonucleasa Cas9 puede estar flanqueada por dos NLS, una NLS ubicada en el extremo N-terminal y la segunda NLS ubicada en el extremo C-terminal. La NLS puede ser cualquier NLS conocida en la técnica, tal como una NLS de SV40.

#### Ácido nucleico de direccionamiento al genoma

La presente divulgación proporciona un ácido nucleico de direccionamiento al genoma que puede dirigir las actividades de un polipéptido asociado (por ejemplo, un polipéptido dirigido al sitio) a una secuencia diana específica dentro de un ácido nucleico diana. El ácido nucleico de direccionamiento al genoma puede ser un ARN. Un ARN de direccionamiento al genoma se denomina "ARN guía" o "ARNg" en el presente documento. Un ARN guía puede comprender al menos una secuencia espaciadora que se hibrida con una secuencia de ácido nucleico diana de interés, y una secuencia de repetición CRISPR. En sistemas de tipo II, el ARNg también comprende un segundo ARN denominado secuencia de ARNtracr. En el ARN guía (ARNg) de tipo II, la secuencia de repetición CRISPR y la secuencia de ARNtracr se hibridan entre sí para formar un dúplex. En el ARN guía (ARNg) de tipo V, el ARNtracr forma un dúplex. En ambos sistemas, el dúplex puede unirse a un polipéptido dirigido al sitio, de tal manera que el ARN guía y el polipéptido dirigido al sitio forman un complejo. El ácido nucleico de direccionamiento al genoma puede proporcionar especificidad de diana al complejo gracias a su asociación con el polipéptido dirigido al sitio. Por tanto, el ácido nucleico de direccionamiento al genoma puede dirigir la actividad del polipéptido dirigido al sitio.

Los ARN guía a modo de ejemplo incluyen las secuencias espaciadoras en SEQ ID NO: 1 - 71.947 y las secuencias de ARNGs en SEQ ID NO: 71.950-71.959 de la lista de secuencias. Tal como entiende el experto habitual en la técnica, cada ARN guía puede estar diseñado para incluir una secuencia espaciadora complementaria a su secuencia diana genómica. Por ejemplo, cada una de las secuencias espaciadoras en SEQ ID NO: 1 - 71.947 de la lista de secuencias puede ponerse en una única quimera de ARN o un ARNtracr (junto con un ARNtracr correspondiente). Véase Jinek *et al.*, Science, 337, 816-821 (2012) y Deltcheva *et al.*, Nature, 471, 602-607 (2011) o la tabla 1.

El ácido nucleico de direccionamiento al genoma puede ser un ARN guía de molécula doble. El ácido nucleico de direccionamiento al genoma puede ser un ARN guía de molécula simple.

Un ARN guía de molécula doble puede comprender dos cadenas de ARN. La primera cadena comprende, en el sentido de 5' a 3', una secuencia de extensión espaciadora opcional, una secuencia espaciadora y una secuencia de repetición de CRISPR mínima. La segunda cadena puede comprender una secuencia de ARNtracr mínima (complementaria a la secuencia de repetición de CRISPR mínima), una secuencia de ARNtracr en 3' y una secuencia de extensión de ARNtracr opcional.

Un ARN guía de molécula simple (ARNgs) en un sistema de tipo II puede comprender, en el sentido de 5' a 3', una secuencia de extensión espaciadora opcional, una secuencia espaciadora, una secuencia de repetición de CRISPR mínima, un ligador guía de molécula simple, una secuencia de ARNtracr mínima, una secuencia de ARNtracr en 3' y una secuencia de extensión de ARNtracr opcional. La extensión de ARNtracr opcional puede comprender elementos que aportan funcionalidad adicional (por ejemplo, estabilidad) al ARN guía. El ligador guía de molécula simple puede unir la repetición de CRISPR mínima y la secuencia de ARNtracr mínima para formar una estructura de tipo horquilla. La extensión de ARNtracr opcional puede comprender una o más horquillas.

Un ARN guía de molécula simple (ARNgs) en un sistema de tipo V puede comprender, en el sentido de 5' a 3', una secuencia de repetición de CRISPR mínima y una secuencia espaciadora.

El ARNGs puede comprender una secuencia espaciadora de 20 nucleótidos en el extremo 5' de la secuencia de ARNGs. El ARNGs puede comprender una secuencia espaciadora de menos de 20 nucleótidos en el extremo 5' de la secuencia de ARNGs. El ARNGs puede comprender una secuencia espaciadora de más de 20 nucleótidos en el extremo 5' de la secuencia de ARNGs. El ARNGs puede comprender una secuencia espaciadora de longitud variable con 17-30 nucleótidos en el extremo 5' de la secuencia de ARNGs (véase la tabla 1).

El ARNGs puede no comprender ningún uracilo en el extremo 3' de la secuencia de ARNGs, tal como en SEQ ID NO: 71.961 de la tabla 1. El ARNGs puede comprender uno o más uracilos en el extremo 3' de la secuencia de ARNGs, tal como en SEQ ID NO: 71.962 en la tabla 1. Por ejemplo, el ARNGs puede comprender 1 uracilo (U) en el extremo 3'

de la secuencia de ARNGs. El ARNGs puede comprender 2 uracilos (UU) en el extremo 3' de la secuencia de ARNGs. El ARNGs puede comprender 3 uracilos (UUU) en el extremo 3' de la secuencia de ARNGs. El ARNGs puede comprender 4 uracilos (UUUU) en el extremo 3' de la secuencia de ARNGs. El ARNGs puede comprender 5 uracilos (UUUUU) en el extremo 3' de la secuencia de ARNGs. El ARNGs puede comprender 6 uracilos (UUUUUU) en el extremo 3' de la secuencia de ARNGs. El ARNGs puede comprender 7 uracilos (UUUUUUU) en el extremo 3' de la secuencia de ARNGs. El ARNGs puede comprender 8 uracilos (UUUUUUUU) en el extremo 3' de la secuencia de ARNGs.

El ARNGs puede no estar modificado o estar modificado. Por ejemplo, los ARNGs modificados pueden comprender uno o más nucleótidos 2'-O-metil-fosforotioato.

Tabla 1

SEQ ID NO.	Secuencia de ARNs
71.960	nnnnnnnnnnnnnnnnnnnnnguuuuagagcuagaaauagcaaguuaaaa <u>uaaggcuaguccguua</u> <u>ucaacuugaaaaaguggcaccgagucggugc</u> uuuu
71.961	nnnnnnnnnnnnnnnnnnnnnguuuuagagcuagaaauagcaaguuaaaa <u>uaaggcuaguccguua</u> <u>ucaacuugaaaaaguggcaccgagucggugc</u>
71.962	n <sub>(17-30)</sub> guuuuagagcuagaaauagcaaguuaaaa <u>uaaggcuaguccguua</u> ucaacuugaaaa <u>agu</u> ggcaccgagucggugc <sub>(1-8)</sub>

A modo de ilustración, los ARN guía usados en el sistema de CRISPR/Cas/Cpf1, u otros ARN más pequeños, pueden sintetizarse fácilmente mediante medios químicos, tal como se ilustra a continuación y se describe en la técnica. Aunque los procedimientos de síntesis química están evolucionando continuamente, las purificaciones de tales ARN mediante procedimientos tales como cromatografía de líquidos de alto rendimiento (HPLC, que evita el uso de geles tales como PAGE) tiende a volverse más difícil a medida que las longitudes de polinucleótidos aumentan significativamente más allá de aproximadamente cien nucleótidos. Un enfoque usado para generar ARN de mayor longitud es producir dos o más moléculas que se ligan entre sí. ARN mucho más largos, tales como los que codifican para una endonucleasa Cas9 o Cpf1, se generan más fácilmente de manera enzimática. Pueden introducirse diversos tipos de modificaciones de ARN durante o después de la síntesis química y/o generación enzimática de ARN, por ejemplo, modificaciones que potencian la estabilidad, reducen la probabilidad o el grado de respuesta inmunitaria innata y/o potencian otros atributos, tal como se describe en la técnica.

Secuencia de extensión espaciadora

En algunos ejemplos de ácidos nucleicos de direccionamiento al genoma, una secuencia de extensión espaciadora puede modificar la actividad, proporcionar estabilidad y/o proporcionar una ubicación para modificaciones de un ácido nucleico de direccionamiento al genoma. Una secuencia de extensión espaciadora puede modificar la actividad específica o inespecífica o la especificidad. En algunos ejemplos, puede proporcionarse una secuencia de extensión espaciadora. La secuencia de extensión espaciadora puede tener una longitud de más de 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 1000, 2000, 3000, 4000, 5000, 6000 o 7000 o más nucleótidos. La secuencia de extensión espaciadora puede tener una longitud de menos de 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 1000, 2000, 3000, 4000, 5000, 6000, 7000 o más nucleótidos. La secuencia de extensión espaciadora puede tener menos de 10 nucleótidos de longitud. La secuencia de extensión espaciadora puede tener entre 10-30 nucleótidos de longitud. La secuencia de extensión espaciadora puede tener entre 30-70 nucleótidos de longitud.

La secuencia de extensión espaciadora puede comprender otro resto (por ejemplo, una secuencia de control de la estabilidad, una secuencia de unión a endorribonucleasa, una ribozima). El resto puede reducir o aumentar la estabilidad de un ácido nucleico de direccionamiento a ácido nucleico. El resto puede ser un segmento terminador de la transcripción (es decir, una secuencia de terminación de la transcripción). El resto puede funcionar en una célula eucariota. El resto puede funcionar en una célula procariota. El resto puede funcionar en células tanto eucariotas como procariotas. Los ejemplos no limitativos de restos adecuados incluyen: una caperuza en 5' (por ejemplo, una caperuza de 7-metilguanilato (m7G)), una secuencia de ribointerruptor (por ejemplo, para permitir una estabilidad regulada y/o accesibilidad regulada por proteínas y complejos de proteínas), una secuencia que forma un dúplex de ARNbc (es decir, una horquilla), una secuencia que dirige el ARN a una ubicación subcelular (por ejemplo, núcleo, mitocondrias, cloroplastos y similares), una modificación o secuencia que proporciona el seguimiento (por ejemplo, conjugación directa a una molécula fluorescente, conjugación a un resto que facilita la detección fluorescente, una secuencia que permite la detección fluorescente, etc.) y/o una modificación o secuencia que proporciona un sitio de unión para proteínas (por ejemplo, proteínas que actúan sobre ADN, incluyendo activadores transcripcionales, controles transcripcionales, ADN metiltransferasas, ADN desmetilasas, histona acetiltransferasas, histona desacetilasas y similares).

## Secuencia espaciadora

La secuencia espaciadora se hibrida con una secuencia en un ácido nucleico diana de interés. El espaciador de un ácido nucleico de direccionamiento al genoma puede interactuar con un ácido nucleico diana de una manera específica de secuencia mediante hibridación (es decir, emparejamiento de bases). La secuencia de nucleótidos del espaciador puede variar dependiendo de la secuencia del ácido nucleico diana de interés.

En un sistema de CRISPR/Cas en el presente documento, la secuencia espaciadora puede estar diseñada para hibridarse con un ácido nucleico diana que está ubicado en 5' de un PAM de la enzima Cas9 usada en el sistema. El espaciador puede aparearse perfectamente con la secuencia diana o puede tener apareamientos erróneos. Cada enzima Cas9 tiene una secuencia de PAM particular que reconoce en un ADN diana. Por ejemplo, *S. pyogenes* reconoce en un ácido nucleico diana un PAM que comprende la secuencia 5'-NRG-3', en la que R comprende o bien A o bien G, en la que N es cualquier nucleótido y N está inmediatamente en 3' de la secuencia de ácido nucleico diana seleccionada como diana por la secuencia espaciadora.

La secuencia de ácido nucleico diana puede comprender 20 nucleótidos. El ácido nucleico diana puede comprender menos de 20 nucleótidos. El ácido nucleico diana puede comprender más de 20 nucleótidos. El ácido nucleico diana puede comprender al menos: 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 o más nucleótidos. El ácido nucleico diana puede comprender como máximo: 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 o más nucleótidos. La secuencia de ácido nucleico diana puede comprender 20 bases inmediatamente en 5' del primer nucleótido del PAM. Por ejemplo, en una secuencia que comprende 5'-NNNNNNNNNNNNNNNNNNNNNRG-3' (SEQ ID NO: 71.948), el ácido nucleico diana puede comprender la secuencia que corresponde a las N, en la que N es cualquier nucleótido, y la secuencia NRG subrayada es el PAM de *S. pyogenes*.

La secuencia espaciadora que se hibrida con el ácido nucleico diana puede tener una longitud de al menos aproximadamente 6 nucleótidos (nt). La secuencia espaciadora puede tener al menos aproximadamente 6 nt, al menos aproximadamente 10 nt, al menos aproximadamente 15 nt, al menos aproximadamente 18 nt, al menos aproximadamente 19 nt, al menos aproximadamente 20 nt, al menos aproximadamente 25 nt, al menos aproximadamente 30 nt, al menos aproximadamente 35 nt o al menos aproximadamente 40 nt, desde aproximadamente 6 nt hasta aproximadamente 80 nt, desde aproximadamente 6 nt hasta aproximadamente 50 nt, desde aproximadamente 6 nt hasta aproximadamente 45 nt, desde aproximadamente 6 nt hasta aproximadamente 40 nt, desde aproximadamente 6 nt hasta aproximadamente 35 nt, desde aproximadamente 6 nt hasta aproximadamente 30 nt, desde aproximadamente 6 nt hasta aproximadamente 25 nt, desde aproximadamente 6 nt hasta aproximadamente 20 nt, desde aproximadamente 6 nt hasta aproximadamente 19 nt, desde aproximadamente 10 nt hasta aproximadamente 50 nt, desde aproximadamente 10 nt hasta aproximadamente 45 nt, desde aproximadamente 10 nt hasta aproximadamente 40 nt, desde aproximadamente 10 nt hasta aproximadamente 35 nt, desde aproximadamente 10 nt hasta aproximadamente 30 nt, desde aproximadamente 10 nt hasta aproximadamente 25 nt, desde aproximadamente 10 nt hasta aproximadamente 20 nt, desde aproximadamente 10 nt hasta aproximadamente 19 nt, desde aproximadamente 19 nt hasta aproximadamente 25 nt, desde aproximadamente 19 nt hasta aproximadamente 30 nt, desde aproximadamente 19 nt hasta aproximadamente 35 nt, desde aproximadamente 19 nt hasta aproximadamente 40 nt, desde aproximadamente 19 nt hasta aproximadamente 45 nt, desde aproximadamente 19 nt hasta aproximadamente 50 nt, desde aproximadamente 19 nt hasta aproximadamente 60 nt, desde aproximadamente 20 nt hasta aproximadamente 25 nt, desde aproximadamente 20 nt hasta aproximadamente 30 nt, desde aproximadamente 20 nt hasta aproximadamente 35 nt, desde aproximadamente 20 nt hasta aproximadamente 40 nt, desde aproximadamente 20 nt hasta aproximadamente 45 nt, desde aproximadamente 20 nt hasta aproximadamente 50 nt, o desde aproximadamente 20 nt hasta aproximadamente 60 nt. En algunos ejemplos, la secuencia espaciadora puede comprender 20 nucleótidos. En algunos ejemplos, el espaciador puede comprender 19 nucleótidos.

En algunos ejemplos, la complementariedad en porcentaje entre la secuencia espaciadora y el ácido nucleico diana es de al menos aproximadamente el 30 %, al menos aproximadamente el 40 %, al menos aproximadamente el 50 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 97 %, al menos aproximadamente el 98 %, al menos aproximadamente el 99 % o el 100 %. En algunos ejemplos, la complementariedad en porcentaje entre la secuencia espaciadora y el ácido nucleico diana es como máximo aproximadamente del 30 %, como máximo aproximadamente del 40 %, como máximo aproximadamente del 50 %, como máximo aproximadamente del 60 %, como máximo aproximadamente del 65 %, como máximo aproximadamente del 70 %, como máximo aproximadamente del 75 %, como máximo aproximadamente del 80 %, como máximo aproximadamente del 85 %, como máximo aproximadamente del 90 %, como máximo aproximadamente del 95 %, como máximo aproximadamente del 97 %, como máximo aproximadamente del 98 %, como máximo aproximadamente del 99 % o el 100 %. En algunos ejemplos, la complementariedad en porcentaje entre la secuencia espaciadora y el ácido nucleico diana es del 100 % a lo largo de los seis nucleótidos contiguos más en 5' de la secuencia diana de la cadena complementaria del ácido nucleico diana. La complementariedad en porcentaje entre la secuencia espaciadora y el ácido nucleico diana puede ser de al menos el 60 % a lo largo de aproximadamente 20 nucleótidos contiguos. La longitud de la secuencia espaciadora y el ácido nucleico diana pueden diferir en de 1 a



6 nucleótidos, lo cual puede considerarse como protuberancia o protuberancias.

La secuencia espaciadora puede diseñarse o elegirse usando un programa informático. El programa informático puede usar variables, tales como temperatura de fusión predicha, formación de estructura secundaria, temperatura de apareamiento predicha, identidad de secuencia, contexto genómico, accesibilidad de cromatina, % de GC, frecuencia de aparición genómica (por ejemplo, de secuencias que son idénticas o son similares pero varían en uno o más puntos como resultado de apareamiento erróneo, inserción o delección), estado de metilación, presencia de SNP y similares.

#### Secuencia de repetición de CRISPR mínima

Una secuencia de repetición de CRISPR mínima puede ser una secuencia con al menos aproximadamente el 30 %, aproximadamente el 40 %, aproximadamente el 50 %, aproximadamente el 60 %, aproximadamente el 65 %, aproximadamente el 70 %, aproximadamente el 75 %, aproximadamente el 80 %, aproximadamente el 85 %, aproximadamente el 90 %, aproximadamente el 95 % o el 100 % de identidad de secuencia con respecto a una secuencia de repetición de CRISPR de referencia (por ejemplo, ARNcr de *S. pyogenes*).

Una secuencia de repetición de CRISPR mínima puede comprender nucleótidos que pueden hibridarse a una secuencia de ARNtracr mínima en una célula. La secuencia de repetición de CRISPR mínima y una secuencia de ARNtracr mínima pueden formar un dúplex, es decir una estructura bicaenaria de bases emparejadas. En conjunto, la secuencia de repetición de CRISPR mínima y la secuencia de ARNtracr mínima pueden unirse al polipéptido dirigido al sitio. Al menos una parte de la secuencia de repetición de CRISPR mínima puede hibridarse con la secuencia de ARNtracr mínima. Al menos una parte de la secuencia de repetición de CRISPR mínima puede comprender al menos aproximadamente el 30 %, aproximadamente el 40 %, aproximadamente el 50 %, aproximadamente el 60 %, aproximadamente el 65 %, aproximadamente el 70 %, aproximadamente el 75 %, aproximadamente el 80 %, aproximadamente el 85 %, aproximadamente el 90 %, aproximadamente el 95 % o el 100 % de complementariedad con respecto a la secuencia de ARNtracr mínima. Al menos una parte de la secuencia de repetición de CRISPR mínima puede comprender como máximo aproximadamente el 30 %, aproximadamente el 40 %, aproximadamente el 50 %, aproximadamente el 60 %, aproximadamente el 65 %, aproximadamente el 70 %, aproximadamente el 75 %, aproximadamente el 80 %, aproximadamente el 85 %, aproximadamente el 90 %, aproximadamente el 95 % o el 100 % de complementariedad con respecto a la secuencia de ARNtracr mínima.

La secuencia de repetición de CRISPR mínima puede tener una longitud de desde aproximadamente 7 nucleótidos hasta aproximadamente 100 nucleótidos. Por ejemplo, la longitud de la secuencia de repetición de CRISPR mínima es de desde aproximadamente 7 nucleótidos (nt) hasta aproximadamente 50 nt, desde aproximadamente 7 nt hasta aproximadamente 40 nt, desde aproximadamente 7 nt hasta aproximadamente 30 nt, desde aproximadamente 7 nt hasta aproximadamente 25 nt, desde aproximadamente 7 nt hasta aproximadamente 20 nt, desde aproximadamente 7 nt hasta aproximadamente 15 nt, desde aproximadamente 8 nt hasta aproximadamente 40 nt, desde aproximadamente 8 nt hasta aproximadamente 30 nt, desde aproximadamente 8 nt hasta aproximadamente 25 nt, desde aproximadamente 8 nt hasta aproximadamente 20 nt, desde aproximadamente 8 nt hasta aproximadamente 15 nt, desde aproximadamente 15 nt hasta aproximadamente 100 nt, desde aproximadamente 15 nt hasta aproximadamente 80 nt, desde aproximadamente 15 nt hasta aproximadamente 50 nt, desde aproximadamente 15 nt hasta aproximadamente 40 nt, desde aproximadamente 15 nt hasta aproximadamente 30 nt o desde aproximadamente 15 nt hasta aproximadamente 25 nt. En algunos ejemplos, la secuencia de repetición de CRISPR mínima puede tener aproximadamente 9 nucleótidos de longitud. La secuencia de repetición de CRISPR mínima puede tener aproximadamente 12 nucleótidos de longitud.

La secuencia de repetición de CRISPR mínima puede ser idéntica en al menos aproximadamente el 60 % a una secuencia de repetición de CRISPR mínima de referencia (por ejemplo, ARNcr de tipo natural de *S. pyogenes*) a lo largo de un tramo de al menos 6, 7 u 8 nucleótidos contiguos. Por ejemplo, la secuencia de repetición de CRISPR mínima puede ser idéntica en al menos aproximadamente el 65 %, idéntica en al menos aproximadamente el 70 %, idéntica en al menos aproximadamente el 75 %, idéntica en al menos aproximadamente el 80 %, idéntica en al menos aproximadamente el 85 %, idéntica en al menos aproximadamente el 90 %, idéntica en al menos aproximadamente el 95 %, idéntica en al menos aproximadamente el 98 %, idéntica en al menos aproximadamente el 99 % o idéntica en el 100 % a una secuencia de repetición de CRISPR mínima de referencia a lo largo de un tramo de al menos 6, 7 u 8 nucleótidos contiguos.

#### Secuencia de ARNtracr mínima

Una secuencia de ARNtracr mínima puede ser una secuencia con al menos aproximadamente el 30 %, aproximadamente el 40 %, aproximadamente el 50 %, aproximadamente el 60 %, aproximadamente el 65 %, aproximadamente el 70 %, aproximadamente el 75 %, aproximadamente el 80 %, aproximadamente el 85 %, aproximadamente el 90 %, aproximadamente el 95 % o el 100 % de identidad de secuencia con respecto a una secuencia de ARNtracr de referencia (por ejemplo, ARNtracr de tipo natural de *S. pyogenes*).

Una secuencia de ARNtracr mínima puede comprender nucleótidos que se hibridan con una secuencia de repetición de CRISPR mínima en una célula. Una secuencia de ARNtracr mínima y una secuencia de repetición de CRISPR

mínima forman un dúplex, es decir una estructura bicatenaria de bases emparejadas. En conjunto, la secuencia de ARNtracr mínima y la repetición de CRISPR mínima pueden unirse a un polipéptido dirigido al sitio. Al menos una parte de la secuencia de ARNtracr mínima puede hibridarse con la secuencia de repetición de CRISPR mínima. La secuencia de ARNtracr mínima puede tener al menos aproximadamente el 30 %, aproximadamente el 40 %, aproximadamente el 50 %, aproximadamente el 60 %, aproximadamente el 65 %, aproximadamente el 70 %, aproximadamente el 75 %, aproximadamente el 80 %, aproximadamente el 85 %, aproximadamente el 90 %, aproximadamente el 95 % o el 100 % de complementariedad con respecto a la secuencia de repetición de CRISPR mínima.

La secuencia de ARNtracr mínima puede tener una longitud de desde aproximadamente 7 nucleótidos hasta aproximadamente 100 nucleótidos. Por ejemplo, la secuencia de ARNtracr mínima puede tener una longitud de desde aproximadamente 7 nucleótidos (nt) hasta aproximadamente 50 nt, desde aproximadamente 7 nt hasta aproximadamente 40 nt, desde aproximadamente 7 nt hasta aproximadamente 30 nt, desde aproximadamente 7 nt hasta aproximadamente 25 nt, desde aproximadamente 7 nt hasta aproximadamente 20 nt, desde aproximadamente 7 nt hasta aproximadamente 15 nt, desde aproximadamente 8 nt hasta aproximadamente 40 nt, desde aproximadamente 8 nt hasta aproximadamente 30 nt, desde aproximadamente 8 nt hasta aproximadamente 25 nt, desde aproximadamente 8 nt hasta aproximadamente 20 nt, desde aproximadamente 8 nt hasta aproximadamente 15 nt, desde aproximadamente 15 nt hasta aproximadamente 100 nt, desde aproximadamente 15 nt hasta aproximadamente 80 nt, desde aproximadamente 15 nt hasta aproximadamente 50 nt, desde aproximadamente 15 nt hasta aproximadamente 40 nt, desde aproximadamente 15 nt hasta aproximadamente 30 nt o desde aproximadamente 15 nt hasta aproximadamente 25 nt. La secuencia de ARNtracr mínima puede tener aproximadamente 9 nucleótidos de longitud. La secuencia de ARNtracr mínima puede tener aproximadamente 12 nucleótidos. El ARNtracr mínimo puede consistir en los nt 23-48 de ARNtracr descrito en Jinek *et al.*, citado anteriormente.

La secuencia de ARNtracr mínima puede ser idéntica en al menos aproximadamente el 60 % a una secuencia de ARNtracr mínima de referencia (por ejemplo, ARNtracr de tipo natural de *S. pyogenes*) a lo largo de un tramo de al menos 6, 7 u 8 nucleótidos contiguos. Por ejemplo, la secuencia de ARNtracr mínima puede ser idéntica en al menos aproximadamente el 65 %, idéntica en aproximadamente el 70 %, idéntica en aproximadamente el 75 %, idéntica en aproximadamente el 80 %, idéntica en aproximadamente el 85 %, idéntica en aproximadamente el 90 %, idéntica en aproximadamente el 95 %, idéntica en aproximadamente el 98 %, idéntica en aproximadamente el 99 % o idéntica en el 100 % a una secuencia de ARNtracr mínima de referencia a lo largo de un tramo de al menos 6, 7 u 8 nucleótidos contiguos.

El dúplex entre el ARN de CRISPR mínimo y el ARNtracr mínimo puede comprender una doble hélice. El dúplex entre el ARN de CRISPR mínimo y el ARNtracr mínimo puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más nucleótidos. El dúplex entre el ARN de CRISPR mínimo y el ARNtracr mínimo puede comprender como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más nucleótidos.

El dúplex puede comprender un apareamiento erróneo (es decir, las dos cadenas del dúplex no son complementarias al 100 %). El dúplex puede comprender al menos aproximadamente 1, 2, 3, 4 o 5 o apareamientos erróneos. El dúplex puede comprender como máximo aproximadamente 1, 2, 3, 4 o 5 o apareamientos erróneos. El dúplex puede comprender no más de 2 apareamientos erróneos.

#### Protuberancias

En algunos casos, puede haber una "protuberancia" en el dúplex entre el ARN de CRISPR mínimo y el ARNtracr mínimo. Una protuberancia es una región no emparejada de nucleótidos dentro del dúplex. Una protuberancia puede contribuir a la unión del dúplex al polipéptido dirigido al sitio. La protuberancia puede comprender, en un lado del dúplex, un 5'-XXXY-3' no emparejado en el que X es cualquier purina e Y comprende un nucleótido que puede formar un par oscilante con un nucleótido en la cadena opuesta, y una región de nucleótidos no emparejados en el otro lado del dúplex. El número de nucleótidos no emparejados en los dos lados del dúplex puede ser diferente.

En un ejemplo, la protuberancia puede comprender una purina no emparejada (por ejemplo, adenina) en la cadena de repetición de CRISPR mínima de la protuberancia. En algunos ejemplos, la protuberancia puede comprender un 5'-AAGY-3' no emparejado de la cadena de secuencia de ARNtracr mínima de la protuberancia, en el que Y comprende un nucleótido que puede formar un emparejamiento oscilante con un nucleótido en la cadena de repetición de CRISPR mínima.

Una protuberancia en el lado de repetición de CRISPR mínima del dúplex puede comprender al menos 1, 2, 3, 4 o 5 o más nucleótidos no emparejados. Una protuberancia en el lado de repetición de CRISPR mínima del dúplex puede comprender como máximo 1, 2, 3, 4 o 5 o más nucleótidos no emparejados. Una protuberancia en el lado de repetición de CRISPR mínima del dúplex puede comprender 1 nucleótido no emparejado.

Una protuberancia en el lado de secuencia de ARNtracr mínima del dúplex puede comprender al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más nucleótidos no emparejados. Una protuberancia en el lado de secuencia de ARNtracr mínima del dúplex puede comprender como máximo 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más nucleótidos no emparejados. Una

protuberancia en un segundo lado del dúplex (por ejemplo, el lado de secuencia de ARNtracr mínima del dúplex) puede comprender 4 nucleótidos no emparejados.

Una protuberancia puede comprender al menos un emparejamiento oscilante. En algunos ejemplos, una protuberancia puede comprender como máximo emparejamiento oscilante. Una protuberancia puede comprender al menos un nucleótido de purina. Una protuberancia puede comprender al menos 3 nucleótidos de purina. Una protuberancia secuencia puede comprender al menos 5 nucleótidos de purina. Una secuencia de protuberancia puede comprender al menos un nucleótido de guanina. En algunos ejemplos, una secuencia de protuberancia puede comprender al menos un nucleótido de adenina.

#### Horquillas

En diversos ejemplos, una o más horquillas pueden estar ubicadas en 3' con respecto al ARNtracr mínimo en la secuencia de ARNtracr en 3'.

La horquilla puede empezar al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 o 20 o más nucleótidos en 3' desde el último nucleótido emparejado en el dúplex de secuencia de repetición de CRISPR mínima y de ARNtracr mínima. La horquilla puede empezar como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más nucleótidos en 3' desde el último nucleótido emparejado en el dúplex de secuencia de repetición de CRISPR mínima y de ARNtracr mínima.

La horquilla puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 o 20 o más nucleótidos consecutivos. La horquilla puede comprender como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 o más nucleótidos consecutivos.

La horquilla puede comprender un dinucleótido de CC (es decir, dos nucleótidos de citosina consecutivos).

La horquilla puede comprender nucleótidos en dúplex (por ejemplo, nucleótidos en una horquilla, hibridados entre sí). Por ejemplo, una horquilla puede comprender un dinucleótido de CC que se hibrida a un dinucleótido de GG en un dúplex de horquilla de la secuencia de ARNtracr en 3'.

Una o más de las horquillas pueden interaccionar con regiones de interacción con ARN guía de un polipéptido dirigido al sitio.

En algunos ejemplos, hay dos o más horquillas, y en otros ejemplos hay tres o más horquillas.

#### Secuencia de ARNtracr en 3'

Una secuencia de ARNtracr en 3' puede comprender una secuencia con al menos aproximadamente el 30 %, aproximadamente el 40 %, aproximadamente el 50 %, aproximadamente el 60 %, aproximadamente el 65 %, aproximadamente el 70 %, aproximadamente el 75 %, aproximadamente el 80 %, aproximadamente el 85 %, aproximadamente el 90 %, aproximadamente el 95 % o el 100 % de identidad de secuencia con respecto a una secuencia de ARNtracr de referencia (por ejemplo, un ARNtracr de *S. pyogenes*).

La secuencia de ARNtracr en 3' puede tener una longitud de desde aproximadamente 6 nucleótidos hasta aproximadamente 100 nucleótidos. Por ejemplo, la secuencia de ARNtracr en 3' puede tener una longitud de desde aproximadamente 6 nucleótidos (nt) hasta aproximadamente 50 nt, desde aproximadamente 6 nt hasta aproximadamente 40 nt, desde aproximadamente 6 nt hasta aproximadamente 30 nt, desde aproximadamente 6 nt hasta aproximadamente 25 nt, desde aproximadamente 6 nt hasta aproximadamente 20 nt, desde aproximadamente 6 nt hasta aproximadamente 15 nt, desde aproximadamente 8 nt hasta aproximadamente 30 nt, desde aproximadamente 8 nt hasta aproximadamente 25 nt, desde aproximadamente 8 nt hasta aproximadamente 20 nt, desde aproximadamente 8 nt hasta aproximadamente 15 nt, desde aproximadamente 15 nt hasta aproximadamente 100 nt, desde aproximadamente 15 nt hasta aproximadamente 80 nt, desde aproximadamente 15 nt hasta aproximadamente 50 nt, desde aproximadamente 15 nt hasta aproximadamente 40 nt, desde aproximadamente 15 nt hasta aproximadamente 30 nt o desde aproximadamente 15 nt hasta aproximadamente 25 nt. La secuencia de ARNtracr en 3' puede tener una longitud de aproximadamente 14 nucleótidos.

La secuencia de ARNtracr en 3' puede ser idéntica en al menos aproximadamente el 60 % a una secuencia de ARNtracr en 3' de referencia (por ejemplo, secuencia de ARNtracr en 3' de tipo natural de *S. pyogenes*) a lo largo de un tramo de al menos 6, 7 u 8 nucleótidos contiguos. Por ejemplo, la secuencia de ARNtracr en 3' puede ser idéntica en al menos aproximadamente el 60 %, idéntica en aproximadamente el 65 %, idéntica en aproximadamente el 70 %, idéntica en aproximadamente el 75 %, idéntica en aproximadamente el 80 %, idéntica en aproximadamente el 85 %, idéntica en aproximadamente el 90 %, idéntica en aproximadamente el 95 %, idéntica en aproximadamente el 98 %, idéntica en aproximadamente el 99 % o idéntica en el 100 %, a una secuencia de ARNtracr en 3' de referencia (por ejemplo, secuencia de ARNtracr en 3' de tipo natural de *S. pyogenes*) a lo largo de un tramo de al menos 6, 7 u

8 nucleótidos contiguos.

La secuencia de ARNtracr en 3' puede comprender más de una región en dúplex (por ejemplo, región hibridada de tipo horquilla). La secuencia de ARNtracr en 3' puede comprender dos regiones en dúplex.

La secuencia de ARNtracr en 3' puede comprender una estructura de tallo-bucle. La estructura de tallo-bucle en el ARNtracr en 3' puede comprender al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 o 20 o más nucleótidos. La estructura de tallo-bucle en el ARNtracr en 3' puede comprender como máximo 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más nucleótidos. La estructura de tallo-bucle puede comprender un resto funcional. Por ejemplo, la estructura de tallo-bucle puede comprender un aptámero, una ribozima, una horquilla de interacción con proteína, una matriz de CRISPR, un intrón o un exón. La estructura de tallo-bucle puede comprender al menos aproximadamente 1, 2, 3, 4 o 5 o más restos funcionales. La estructura de tallo-bucle puede comprender como máximo aproximadamente 1, 2, 3, 4 o 5 o más restos funcionales.

La horquilla en la secuencia de ARNtracr en 3' puede comprender un dominio P. En algunos ejemplos, el dominio P puede comprender una región bicatenaria en la horquilla.

#### Secuencia de extensión de ARNtracr

Puede proporcionarse una secuencia de extensión de ARNtracr tanto si el ARNtracr está en el contexto de guías de molécula simple como guías de molécula doble. La secuencia de extensión de ARNtracr puede tener una longitud de desde aproximadamente 1 nucleótido hasta aproximadamente 400 nucleótidos. La secuencia de extensión de ARNtracr puede tener una longitud de más de 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380 o 400 nucleótidos. La secuencia de extensión de ARNtracr puede tener una longitud de desde aproximadamente 20 hasta aproximadamente 5000 o más nucleótidos. La secuencia de extensión de ARNtracr puede tener una longitud de más de 1000 nucleótidos. La secuencia de extensión de ARNtracr puede tener una longitud de menos de 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400 o más nucleótidos. La secuencia de extensión de ARNtracr puede tener una longitud de menos de 1000 nucleótidos. La secuencia de extensión de ARNtracr puede comprender menos de 10 nucleótidos de longitud. La secuencia de extensión de ARNtracr puede tener 10-30 nucleótidos de longitud. La secuencia de extensión de ARNtracr puede tener 30-70 nucleótidos de longitud.

La secuencia de extensión de ARNtracr puede comprender un resto funcional (por ejemplo, una secuencia de control de la estabilidad, ribozima, secuencia de unión a endorribonucleasa). El resto funcional puede comprender un segmento terminador de la transcripción (es decir, una secuencia de terminación de la transcripción). El resto funcional puede tener una longitud total de desde aproximadamente 10 nucleótidos (nt) hasta aproximadamente 100 nucleótidos, desde aproximadamente 10 nt hasta aproximadamente 20 nt, desde aproximadamente 20 nt hasta aproximadamente 30 nt, desde aproximadamente 30 nt hasta aproximadamente 40 nt, desde aproximadamente 40 nt hasta aproximadamente 50 nt, desde aproximadamente 50 nt hasta aproximadamente 60 nt, desde aproximadamente 60 nt hasta aproximadamente 70 nt, desde aproximadamente 70 nt hasta aproximadamente 80 nt, desde aproximadamente 80 nt hasta aproximadamente 90 nt o desde aproximadamente 90 nt hasta aproximadamente 100 nt, desde aproximadamente 15 nt hasta aproximadamente 80 nt, desde aproximadamente 15 nt hasta aproximadamente 50 nt, desde aproximadamente 15 nt hasta aproximadamente 40 nt, desde aproximadamente 15 nt hasta aproximadamente 30 nt, o desde aproximadamente 15 nt hasta aproximadamente 25 nt. El resto funcional puede funcionar en una célula eucariota. El resto funcional puede funcionar en una célula procariota. El resto funcional puede funcionar en células tanto eucariotas como procariotas.

Los ejemplos no limitativos de restos funcionales de extensión de ARNtracr adecuados incluyen una cola poliadenilada en 3', una secuencia de ribointerruptor (por ejemplo, para permitir una estabilidad regulada y/o accesibilidad regulada mediante proteínas y complejos de proteínas), una secuencia que forma un dúplex de ARNbc (es decir, una horquilla), una secuencia que dirige el ARN a una ubicación subcelular (por ejemplo, núcleo, mitocondrias, cloroplastos y similares), una modificación o secuencia que proporciona el seguimiento (por ejemplo, conjugación directa a una molécula fluorescente, conjugación a un resto que facilita la detección fluorescente, una secuencia que permite la detección fluorescente, etc.) y/o una modificación o secuencia que proporciona un sitio de unión para proteínas (por ejemplo, proteínas que actúan sobre ADN, incluyendo activadores transcripcionales, controles transcripcionales, ADN metiltransferasas, ADN desmetilasas, histona acetiltransferasas, histona desacetilasas y similares). La secuencia de extensión de ARNtracr puede comprender un sitio de unión a cebador o un índice molecular (por ejemplo, secuencia de código de barras). La secuencia de extensión de ARNtracr puede comprender una o más etiquetas de afinidad.

#### Secuencia de ligador guía de molécula simple

La secuencia de ligador de un ácido nucleico guía de molécula simple puede tener una longitud de desde aproximadamente 3 nucleótidos hasta aproximadamente 100 nucleótidos. En Jinek *et al.*, citado anteriormente, por ejemplo, se usó un "tetrabucle" de 4 nucleótidos simple (-GAAA-), Science, 337(6096):816-821 (2012). Un ligador ilustrativo tiene una longitud de desde aproximadamente 3 nucleótidos (nt) hasta aproximadamente 90 nt, desde aproximadamente 3 nt hasta aproximadamente 80 nt, desde aproximadamente 3 nt hasta aproximadamente 70 nt,

desde aproximadamente 3 nt hasta aproximadamente 60 nt, desde aproximadamente 3 nt hasta aproximadamente 50 nt, desde aproximadamente 3 nt hasta aproximadamente 40 nt, desde aproximadamente 3 nt hasta aproximadamente 30 nt, desde aproximadamente 3 nt hasta aproximadamente 20 nt, desde aproximadamente 3 nt hasta aproximadamente 10 nt. Por ejemplo, el ligador puede tener una longitud de desde aproximadamente 3 nt hasta aproximadamente 5 nt, desde aproximadamente 5 nt hasta aproximadamente 10 nt, desde aproximadamente 10 nt hasta aproximadamente 15 nt, desde aproximadamente 15 nt hasta aproximadamente 20 nt, desde aproximadamente 20 nt hasta aproximadamente 25 nt, desde aproximadamente 25 nt hasta aproximadamente 30 nt, desde aproximadamente 30 nt hasta aproximadamente 35 nt, desde aproximadamente 35 nt hasta aproximadamente 40 nt, desde aproximadamente 40 nt hasta aproximadamente 50 nt, desde aproximadamente 50 nt hasta aproximadamente 60 nt, desde aproximadamente 60 nt hasta aproximadamente 70 nt, desde aproximadamente 70 nt hasta aproximadamente 80 nt, desde aproximadamente 80 nt hasta aproximadamente 90 nt o desde aproximadamente 90 nt hasta aproximadamente 100 nt. El ligador de un ácido nucleico guía de molécula simple puede tener entre 4 y 40 nucleótidos. El ligador puede tener al menos aproximadamente 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500 o 7000 o más nucleótidos. El ligador puede tener como máximo aproximadamente 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500 o 7000 o más nucleótidos.

Los ligadores pueden comprender cualquiera de una variedad de secuencias, aunque en algunos ejemplos el ligador no comprenderá secuencias que tienen regiones extensas de homología con otras porciones del ARN guía, que pueden provocar unión intramolecular que puede interferir con otras regiones funcionales de la guía. En Jinek *et al.*, citado anteriormente, se usó una secuencia de 4 nucleótidos simple -GAAA-, Science, 337(6096):816-821 (2012), pero pueden usarse igualmente numerosas otras secuencias, incluyendo secuencias más largas.

La secuencia de ligador puede comprender un resto funcional. Por ejemplo, la secuencia de ligador puede comprender una o más características, incluyendo un aptámero, una ribozima, una horquilla de interacción con proteína, un sitio de unión a proteína, una matriz de CRISPR, un intrón o un exón. La secuencia de ligador puede comprender al menos aproximadamente 1, 2, 3, 4 o 5 o más restos funcionales. En algunos ejemplos, la secuencia de ligador puede comprender como máximo aproximadamente 1, 2, 3, 4 o 5 o más restos funcionales.

Una etapa de los métodos *ex vivo* de la presente divulgación puede comprender editar las células iPSC específicas de paciente usando ingeniería genómica. Alternativamente, una etapa de los métodos *ex vivo* de la presente divulgación puede comprender editar célula madre mesenquimatosas o célula progenitora hematopoyética. Asimismo, una etapa de los métodos *in vivo* de la presente divulgación puede comprender editar las células en un paciente que tiene hemoglobinopatía usando ingeniería genómica. De manera similar, una etapa en los métodos celulares de la presente divulgación puede comprender editar dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A en una célula humana mediante ingeniería genómica.

Diferentes pacientes con hemoglobinopatía requerirán generalmente diferentes estrategias de delección, modulación o inactivación. Puede usarse cualquier endonucleasa CRISPR en los métodos de la presente divulgación, teniendo cada endonucleasa CRISPR su propio PAM asociado, que puede ser específico de enfermedad o no. Por ejemplo, se han identificado secuencias espaciadoras de ARNg para el direccionamiento dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una endonucleasa CRISPR/Cas9 de *S. pyogenes* en SEQ ID NO: 1- 29.482 de la lista de secuencias. Se han identificado secuencias espaciadoras de ARNg para el direccionamiento dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una endonucleasa CRISPR/Cas9 de *S. aureus* en SEQ ID NO: 29.483-32.387 de la lista de secuencias. Se han identificado secuencias espaciadoras de ARNg para el direccionamiento dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una endonucleasa CRISPR/Cas9 de *S. thermophilus* en SEQ ID NO: 32.388-33.420 de la lista de secuencias. Se han identificado secuencias espaciadoras de ARNg para el direccionamiento dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una endonucleasa CRISPR/Cas9 de *T. denticola* en SEQ ID NO: 33.421-33.851 de la lista de secuencias. Se han identificado secuencias espaciadoras de ARNg para el direccionamiento dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una endonucleasa CRISPR/Cas9 de *N. meningitidis* en SEQ ID NO: 33.852-36.731. Se han identificado secuencias espaciadoras de ARNg para el direccionamiento dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A con una endonucleasa CRISPR/Cpf1 de *Acidimicrococcus*, *Lachnospiraceae* y *Franciscella Novicida* en SEQ ID NO: 36.732-71.947.

Por ejemplo, la secuencia de control transcripcional del gen de BCL11A puede modularse o inactivarse mediante delecciones que surgen debido a la ruta de NHEJ. Puede usarse NHEJ para deleccionar segmentos de la secuencia de control transcripcional del gen de BCL11A, o bien directamente o bien alterando sitios donantes o aceptores de corte y empalme mediante escisión por un ARNg que selecciona como diana varias ubicaciones, o varios ARNg.

La secuencia de control transcripcional del gen de BCL11A también puede modularse o inactivarse insertando un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada. Por ejemplo, el donante para modulación o activación mediante HDR contiene la secuencia de control transcripcional modificada del gen de BCL11A con ramas de homología flanqueantes pequeñas o grandes para permitir el apareamiento. HDR

es esencialmente un mecanismo libre de error que usa una secuencia de ADN homóloga suministrada como molde durante la reparación de DSB. La tasa de reparación dirigida por homología (HDR) es una función de la distancia entre la secuencia de control transcripcional y el sitio de corte de modo que resulta importante elegir sitios diana solapantes o cercanos. Los moldes pueden incluir secuencias adicionales flanqueadas por las regiones homólogas o pueden

contener una secuencia que difiere de la secuencia genómica, permitiendo por tanto edición de secuencia.

Además de modular o inactivar la secuencia de control transcripcional del gen de BCL11A mediante NHEJ o HDR, son posibles una gama de otras opciones. Si hay deleciones pequeñas o grandes, puede insertarse un ADNc que contiene una secuencia de control transcripcional modificada. Puede insertarse un ADNc de longitud completa en cualquier "ubicación segura", pero debe usar un promotor suministrado u otro. Si se inserta este constructo en la ubicación correcta, tendrá control fisiológico, similar al gen normal. Pueden usarse pares de nucleasas para delecionar regiones de genes, aunque habitualmente tendrá que proporcionarse un donante para modular o inactivar la función. En este caso, se suministrarán dos ARNg y una secuencia donante.

Algunas estrategias de ingeniería genómica implican modular o inactivar una secuencia de control transcripcional del gen de BCL11A delecionando al menos una porción de la secuencia de control transcripcional del gen de BCL11A y/o insertando un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada en el locus del gen correspondiente o un locus de ubicación segura mediante reparación dirigida por homología (HDR), que también se conoce como recombinación homóloga (HR). Esta estrategia puede modular o inactivar la secuencia de control transcripcional del gen de BCL11A y revertir, tratar y/o mitigar el estado patológico. Con frecuencia los nucleótidos donantes para modular/inactivar secuencias de control transcripcional son pequeños (< 300 pb). Esto resulta ventajoso, ya que las eficiencias de HDR pueden estar inversamente relacionadas con el tamaño de la molécula donante. Además, se espera que los moldes donantes puedan ajustarse en moléculas de virus adenoasociado (AAV) de tamaño restringido, que se ha mostrado que son un medio eficaz de suministro de molde donante.

La reparación dirigida por homología es un mecanismo celular para reparar roturas de cadena doble (DSB). La forma más común es la recombinación homóloga. Hay rutas adicionales para HDR, incluyendo apareamiento de cadena sencilla y HDR alternativa. Las herramientas de ingeniería genómica permiten a los investigadores manipular las rutas de recombinación homóloga celulares para crear modificaciones específicas de sitio en el genoma. Se ha encontrado que las células pueden reparar una rotura de cadena doble usando una molécula donante sintética proporcionada en trans. Por tanto, introduciendo una rotura de cadena doble cerca de una mutación específica y proporcionando un donante adecuado, pueden realizarse cambios dirigidos en el genoma. La escisión específica aumenta la tasa de HDR más de 1.000 veces por encima de la tasa de 1 de cada  $10^6$  células que reciben un donante homólogo solo. La tasa de reparación dirigida por homología (HDR) en un nucleótido particular es una función de la distancia hasta el sitio de corte, de modo que resulta importante elegir sitios diana solapantes o cercanos. La edición génica ofrece ventajas con respecto a la adición génica, ya que la corrección *in situ* deja sin perturbar el resto del genoma.

Los donantes suministrados para edición mediante HDR varían notablemente, pero pueden contener la secuencia pretendida con ramas de homología flanqueantes pequeñas o grandes para permitir el apareamiento con el ADN genómico. Las regiones de homología que flanquean los cambios genéticos introducidos pueden ser de 30 pb o más pequeñas, o de hasta un casete de múltiples kilobases que puede contener promotores, ADNc, etc. Se han usado donantes de oligonucleótidos tanto monocatenarios como bicatenarios. El tamaño de estos oligonucleótidos oscila desde menos de 100 nt hasta más de muchas kb, aunque también pueden generarse y usarse ADNmc más largos. Pueden usarse donantes bicatenarios, incluyendo amplicones de PCR, plásmidos y minicírculos. En general, se ha encontrado que un vector de AAV puede ser un medio muy eficaz de suministro de un molde donante, aunque el límite de empaquetamiento para donantes individuales es de <5 kb. La transcripción activa del donante aumentó la HDR tres veces, indicando que la inclusión de promotor puede aumentar la conversión. A la inversa, la metilación de CpG del donante redujo la expresión génica y HDR.

Además de endonucleasas de tipo natural, tales como Cas9, existen variantes de enzima melladora que tienen uno u otro dominio de nucleasa inactivado, dando como resultado el corte de tan solo una cadena de ADN. Puede dirigirse HDR a partir de enzimas melladoras Cas individuales o usando pares de enzimas melladoras que flanquean la zona diana. Los donantes pueden ser monocatenarios, mellados o ADNbc.

El ADN donante puede suministrarse con la nucleasa o de manera independiente mediante una variedad de métodos diferentes, por ejemplo mediante transfección, nanopartículas, microinyección o transducción viral. Se ha propuesto una gama de opciones de anclaje para aumentar la disponibilidad de los donantes para HDR. Los ejemplos incluyen unión del donante a la nucleasa, unión a proteínas de unión a ADN que se unen cerca, o unión a proteínas que están implicadas en la reparación o unión de extremos de ADN.

La elección de ruta de reparación puede estar guiarse mediante varias condiciones de cultivo, tales como las que influyen en los ciclos celulares, o mediante selección como diana de proteínas de reparación de ADN y asociadas. Por ejemplo, para aumentar HDR, pueden suprimirse moléculas clave de NHEJ, tales como KU70, KU80 o ADN ligasa IV.

Sin un donante presente, los extremos de una rotura de ADN o extremos de diferentes roturas pueden unirse usando

las diversas rutas de reparación no homólogas en las que los extremos de ADN se unen con poco o ningún emparejamiento de bases en la unión. Además de NHEJ canónica, hay mecanismos de reparación similares, tales como alt-NHEJ. Si hay dos roturas, el segmento intermedio puede delecionarse o invertirse. Las rutas de reparación de NHEJ pueden conducir a inserciones, delecciones o mutaciones en las uniones.

Se usó NHEJ para insertar un casete de expresión génica inducible de 15 kb en un locus definido en líneas de células humanas tras escisión con nucleasa. Maresca, M., Lin, V.G., Guo, N. y Yang, Y., *Genoma Res* 23, 539-546 (2013).

Además de edición genómica mediante NHEJ o HDR, se han realizado inserciones génicas específicas de sitio que usan tanto la ruta de NHEJ como HR. Un enfoque de combinación puede ser aplicable en determinados entornos, posiblemente incluyendo límites de intrones/exones. NHEJ puede demostrar ser eficaz para la ligación en el intrón, mientras que la HDR libre de error puede ser más adecuada en la región codificante.

Como alternativa adicional, puede insertarse un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada en el locus del gen correspondiente o insertarse en un sitio de ubicación segura, tal como AAVS1. En algunos ejemplos, los métodos pueden proporcionar un ARNg o un par de ARNg que pueden usarse para facilitar la incorporación de una nueva secuencia a partir de un molde donante de polinucleótido para insertar una parte o la totalidad del gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada.

Los métodos pueden proporcionar pares de ARNg que realizan una delección cortando el gen dos veces, un corte de ARNg en el extremo 5' de una o más mutaciones y el otro corte de ARNg en el extremo 3' de una o más mutaciones lo que facilita la inserción de una nueva secuencia a partir de un molde donante de polinucleótido para sustituir a la secuencia de control transcripcional del gen de BCL11A. El corte puede lograrse mediante un par de ADN endonucleasas que realizan, cada una, una DSB en el genoma, o mediante múltiples enzimas melladoras que, juntas, realizan una DSB en el genoma.

Alternativamente, los métodos pueden proporcionar un ARNg para realizar un corte de cadena doble alrededor de una secuencia de control transcripcional del gen de BCL11A que facilita la inserción de una nueva secuencia a partir de un molde donante de polinucleótido para sustituir la secuencia de control transcripcional del gen de BCL11A con un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada. El corte de cadena doble puede realizarse mediante una única ADN endonucleasa o múltiples enzimas melladoras que, juntas, realizan una DSB en el genoma.

Las modificaciones ilustrativas dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A incluyen sustituciones dentro o cerca de (proximales a) la secuencia de control transcripcional del gen de BCL11A anteriormente mencionada, tal como dentro de la región de menos de 3 kb, menos de 2kb, menos de 1 kb, menos de 0,5 kb en el sentido de 5' o en el sentido de 3' de la secuencia de control transcripcional.

Tales variantes pueden incluir sustituciones que son más grandes en el sentido de 5' y/o 3' que la sustitución específica en cuestión, o más pequeñas en cualquier sentido. Por consiguiente, por "cerca" o "proximal" con respecto a sustituciones específicas, quiere decirse que el locus de SSB o DSB asociado con un límite de sustitución deseado (también denominado en el presente documento punto final) puede estar dentro de una región que está a menos de aproximadamente 3 kb desde el locus de referencia indicado. El locus de SSB o DSB puede ser más proximal y estar dentro de 2 kb, dentro de 1 kb, dentro de 0,5 kb o dentro de 0,1 kb. En el caso de sustitución pequeña, el punto final deseado puede estar en o "adyacente al" locus de referencia, mediante lo cual quiere decirse que el punto final puede estar dentro de 100 pb, dentro de 50 pb, dentro de 25 pb o a menos de aproximadamente 10 pb a 5 pb desde el locus de referencia.

Puede esperarse que los ejemplos que comprenden sustituciones más grandes o más pequeñas proporcionen el mismo beneficio, siempre que se module o inactive la actividad de control transcripcional. Por tanto, se espera que muchas variaciones de las sustituciones descritas e ilustradas en el presente documento pueden ser eficaces para mejorar hemoglobinopatías.

Otra estrategia de ingeniería genómica implica la delección de exones o intrones. La delección dirigida de exones o intrones específicos puede ser una estrategia atractiva para tratar un gran subconjunto de pacientes con un único cóctel terapéutico. Las delecciones pueden ser o bien delecciones de un único exón o intrón o bien delecciones de múltiples exones o intrones. Aunque las delecciones de múltiples exones pueden alcanzar a un número mayor de pacientes, para delecciones más grandes la eficiencia de delección se reduce en gran medida con el aumento del tamaño. Por tanto, los intervalos de tamaño de delecciones pueden ser de desde 40 hasta 10.000 pares de bases (pb). Por ejemplo, el tamaño de las delecciones puede oscilar desde 40-100; 100-300; 300-500; 500-1.000; 1.000-2.000; 2.000-3.000; 3.000-5.000; o 5.000-10.000 pares de bases. Puede ser deseable deleccionar un intrón si el intrón contiene un elemento regulador, tal como un elemento de control transcripcional (por ejemplo, un sitio de unión a factor de transcripción).

Con el fin de garantizar que el pre-ARNm se procesa de manera apropiada tras la delección, pueden deleccionarse las señales de corte y empalme circundantes. Los donantes y aceptores de corte y empalme están generalmente dentro de 100 pares de bases del intrón contiguo. Por tanto, en algunos ejemplos, los métodos pueden proporcionar todos los ARNg que cortan aproximadamente +/- 100-3100 pb con respecto a cada unión de exón/intrón de interés.

Para cualquiera de las estrategias de edición genómica, puede confirmarse la edición génica mediante secuenciación o análisis de PCR.

#### Selección de secuencia diana

Pueden usarse desplazamientos en la ubicación del límite en 5' y/o el límite en 3' con respecto a loci de referencia particulares para facilitar o potenciar aplicaciones particulares de edición génica, que dependen en parte del sistema de endonucleasa seleccionado para la edición, tal como se describe y se ilustra adicionalmente en el presente documento.

En un primer ejemplo no limitativo de tal selección de secuencia diana, muchos sistemas de endonucleasa tienen reglas o criterios que pueden guiar la selección inicial de posibles sitios diana para la escisión, tales como el requisito de un motivo de secuencia de PAM en una posición particular adyacente a los sitios de escisión de ADN en el caso de endonucleasas CRISPR de tipo II o tipo V.

En otro ejemplo no limitativo de selección u optimización de secuencia diana, la frecuencia de actividad inespecífica para una combinación particular de secuencia diana y endonucleasa de edición génica (es decir, la frecuencia de DSB que se producen en sitios distintos de la secuencia diana seleccionada) puede evaluarse con respecto a la frecuencia de actividad específica. En algunos casos, células que se han sometido correctamente a edición en el locus deseado pueden tener una ventaja selectiva con respecto a otras células. Ejemplos ilustrativos, pero no limitativos, de una ventaja selectiva incluyen la adquisición de atributos tales como tasas potenciadas de replicación, persistencia, resistencia a determinadas condiciones, tasas potenciadas de injerto satisfactorio o persistencia *in vivo* tras la introducción en un paciente, y otros atributos asociados con el mantenimiento o aumento de números o viabilidad de tales células. En otros casos, células que se han sometido correctamente a edición en el locus deseado pueden seleccionarse positivamente mediante uno o más métodos de examen usados para identificar, clasificar o seleccionar de otro modo células que se han sometido correctamente a edición. Tanto la ventaja selectiva como los métodos de selección dirigida pueden aprovechar el fenotipo asociado con la corrección. En algunos casos, pueden editarse células dos o más veces con el fin de crear una segunda modificación que crea un nuevo fenotipo que se usa para seleccionar o purificar la población pretendida de células. Una segunda modificación de este tipo puede crearse añadiendo un segundo ARNg para un marcador seleccionable o que puede examinarse. En algunos casos, pueden editarse correctamente células en el locus deseado usando un fragmento de ADN que contiene el ADNc y también un marcador seleccionable.

Tanto si puede aplicarse cualquier ventaja selectiva como si tiene que aplicarse cualquier selección dirigida en un caso particular, la selección de secuencia diana también puede guiarse teniendo en cuenta frecuencias inespecíficas con el fin de potenciar la eficacia de la aplicación y/o reducir la posibilidad de alteraciones no deseadas en sitios distintos de la diana deseada. Tal como se describe adicionalmente y se ilustra en el presente documento y en la técnica, la aparición de actividad inespecífica puede verse influida por varios factores incluyendo similitudes y diferencias entre el sitio diana y diversos sitios inespecíficos, así como la endonucleasa particular usada. Hay herramientas de bioinformática disponibles que ayudan en la predicción de actividad inespecífica, y con frecuencia tales herramientas también pueden usarse para identificar los sitios de actividad inespecífica más probables, que entonces pueden evaluarse en entornos experimentales para evaluar frecuencias relativas de actividad inespecífica con respecto a específica, permitiendo de ese modo la selección de secuencias que tienen actividades específicas relativas superiores. En el presente documento se proporcionan ejemplos ilustrativos de tales técnicas, y otros se conocen en la técnica.

Otro aspecto de selección de secuencia diana se refiere a acontecimientos de recombinación homóloga. Las secuencias que comparten regiones de homología pueden servir como puntos focales para acontecimientos de recombinación homóloga que dan como resultado la delección de secuencias intermedias. Tales acontecimientos de recombinación se producen durante el transcurso normal de la replicación de cromosomas y otras secuencias de ADN, y también en otros momentos cuando están sintetizándose secuencias de ADN, tal como en el caso de reparaciones de roturas de cadena doble (DSB), que se producen de manera regular durante el ciclo de replicación celular normal pero también pueden potenciarse mediante la aparición de diversos acontecimientos (tales como luz UV y otros factores de inducción de rotura de ADN) o la presencia de determinados agentes (tales como diversos factores de inducción químicos). Muchos de tales factores de inducción provocan que se produzcan DSB indiscriminadamente en el genoma, y las DSB pueden inducirse de manera regular y repararse en células normales. Durante la reparación, la secuencia original puede reconstruirse con completa fidelidad, sin embargo, en algunos casos, se introducen pequeñas inserciones o delecciones (denominadas "indeles") en el sitio de DSB.

Las DSB también pueden inducirse específicamente en ubicaciones particulares, como en el caso de los sistemas de endonucleasas descritos en el presente documento, que pueden usarse para provocar acontecimientos de



modificación génica dirigidos o preferibles en ubicaciones cromosómicas seleccionadas. La tendencia para que se sometan secuencias homólogas a recombinación en el contexto de reparación de ADN (así como replicación) puede aprovecharse en varias circunstancias y es la base para una aplicación de sistemas de edición génica, tales como CRISPR, en los que se usa reparación dirigida por homología para insertar una secuencia de interés, proporcionada mediante el uso de un polinucleótido “donante”, en una ubicación cromosómica deseada.

También pueden usarse regiones de homología entre secuencias particulares, que pueden ser pequeñas regiones de “microhomología” que pueden comprender tan solo diez pares de bases o menos, para provocar deleciones deseadas. Por ejemplo, puede introducirse una única DSB en un sitio que muestra microhomología con una secuencia cercana. Durante el transcurso normal de la reparación de tal DSB, un resultado que se produce con alta frecuencia es la deleción de la secuencia intermedia como resultado de facilitarse la recombinación mediante la DSB y el proceso de reparación celular concomitante.

Sin embargo, en algunas circunstancias, seleccionar secuencias diana dentro de regiones de homología también puede dar lugar a deleciones mucho más grandes, incluyendo fusiones génicas (cuando las deleciones están en regiones codificantes), que pueden o no desearse dadas las circunstancias particulares.

Los ejemplos proporcionados en el presente documento ilustran adicionalmente la selección de diversas regiones diana para la creación de DSB diseñadas para inducir sustituciones que dan como resultado la modulación o inactivación de actividad de proteína de control transcripcional, así como la selección de secuencias diana específicas dentro de tales regiones que están diseñadas para minimizar acontecimientos inespecíficos con respecto a acontecimientos específicos.

#### Modificaciones de ácido nucleico

En algunos casos, los polinucleótidos introducidos en células pueden comprender una o más modificaciones que pueden usarse de manera individual o en combinación, por ejemplo, para potenciar la actividad, estabilidad o especificidad, tras el suministro, reducir respuestas inmunitarias innatas en células huésped, o para otras potenciaciones, tal como se describe adicionalmente en el presente documento y se conoce en la técnica.

En determinados ejemplos, pueden usarse polinucleótidos modificados en el sistema de CRISPR/Cas9/Cpf1, en cuyo caso pueden modificarse los ARN guía (o bien guías de molécula sencilla o bien guías de molécula doble) y/o un ADN o un ARN que codifica para una endonucleasa Cas o Cpf1 introducidos en una célula, tal como se describe y se ilustra a continuación. Tales polinucleótidos modificados pueden usarse en el sistema de CRISPR/Cas9/Cpf1 para editar uno cualquiera o más loci genómicos.

Usar el sistema de CRISPR/Cas9/Cpf1 con los fines de ilustraciones no limitativas de tales usos, pueden usarse modificaciones de ARN guía para potenciar la formación o estabilidad del complejo de edición genómica de CRISPR/Cas9/Cpf1 que comprende ARN guía, que pueden ser guías de molécula sencilla o molécula doble, y una endonucleasa Cas o Cpf1. Además o alternativamente pueden usarse modificaciones de ARN guía para potenciar la iniciación, estabilidad o cinética de interacciones entre el complejo de edición genómica con la secuencia diana en el genoma, lo cual puede usarse, por ejemplo, para potenciar la actividad específica. Además o alternativamente pueden usarse modificaciones de ARN guía para potenciar la especificidad, por ejemplo, las tasas relativas de edición genómica en el sitio específico en comparación con efectos en otros sitios (inespecíficos).

Además o alternativamente pueden usarse modificaciones para aumentar la estabilidad de un ARN guía, por ejemplo, aumentando su resistencia frente a la degradación mediante ribonucleasas (ARNasas) presentes en una célula, provocando de ese modo que se aumente su semivida en la célula. Las modificaciones que potencian la semivida de ARN guía pueden ser particularmente útiles en aspectos en los que se introduce una endonucleasa Cas o Cpf1 en la célula para editarse mediante un ARN que se necesita traducir con el fin de generar endonucleasa, porque el aumento de la semivida de ARN guía introducidos al mismo tiempo que el ARN que codifica para la endonucleasa puede usarse para aumentar el tiempo durante el cual coexisten los ARN guía y la endonucleasa Cas o Cpf1 codificada en la célula.

Además o alternativamente pueden usarse modificaciones para reducir la probabilidad o el grado en el que ARN introducidos en células provocan respuestas inmunitarias innatas. Tales respuestas, que se han caracterizado correctamente en el contexto de interferencia de ARN (iARN), incluyendo ARN de interferencia pequeños (ARNip), tal como se describe a continuación y en la técnica, tienden a estar asociados con una semivida reducida del ARN y/o la provocación de citocinas u otros factores asociados con respuestas inmunitarias.

También pueden realizarse uno o más tipos de modificaciones en ARN que codifican para una endonucleasa que se introducen en una célula, incluyendo, sin limitación, modificaciones que potencian la estabilidad del ARN (tal como aumentando su degradación mediante ARNasas presentes en la célula), modificaciones que potencian la traducción del producto resultante (es decir, la endonucleasa), y/o modificaciones que reducen la probabilidad o el grado en el que los ARN introducidos en células provocan respuestas inmunitarias innatas.

También pueden usarse combinaciones de modificaciones, tales como las anteriores y otras. En el caso de

CRISPR/Cas9/Cpf1, por ejemplo, pueden realizarse uno o más tipos de modificaciones en ARN guía (incluyendo los mostrados anteriormente a modo de ejemplo) y/o pueden realizarse uno o más tipos de modificaciones en ARN que codifican para endonucleasa Cas (incluyendo los mostrados anteriormente a modo de ejemplo).

A modo de ilustración, los ARN guía usados en el sistema de CRISPR/Cas9/Cpf1 u otros ARN más pequeños pueden sintetizarse fácilmente mediante medios químicos, permitiendo incorporar fácilmente varias modificaciones, tal como se ilustra a continuación y se describe en la técnica. Aunque los procedimientos de síntesis química están evolucionando continuamente, las purificaciones de tales ARN mediante procedimientos tales como cromatografía de líquidos de alto rendimiento (HPLC, que evita el uso de geles tales como PAGE) tiende a volverse más difícil a medida que las longitudes de polinucleótidos aumentan significativamente más allá de aproximadamente cien nucleótidos. Un enfoque que puede usarse para generar ARN químicamente modificados de mayor longitud es producir dos o más moléculas que se ligan entre sí. ARN mucho más largos, tales como los que codifican para una endonucleasa Cas9, se generan más fácilmente de manera enzimática. Aunque hay menos tipos de modificaciones disponibles para su uso en ARN producidos de manera enzimática, todavía hay modificaciones que pueden usarse, por ejemplo, para potenciar la estabilidad, reducir la probabilidad o el grado de respuesta inmunitaria innata y/o potenciar otros atributos, tal como se describe adicionalmente a continuación y en la técnica; y están desarrollándose de manera regular nuevos tipos de modificaciones.

A modo de ilustración de diversos tipos de modificaciones, especialmente las usadas con frecuencia con ARN químicamente sintetizados más pequeños, las modificaciones pueden comprender uno o más nucleótidos modificados en la posición 2' del azúcar, en algunos aspectos un nucleótido modificado con 2'-O-alquilo, 2'-O-alquil-O-alquilo o 2'-fluoro. En algunos ejemplos, las modificaciones de ARN pueden comprender modificaciones con 2'-fluoro, 2'-amino o 2'-O-metilo en la ribosa de pirimidinas, residuos abásicos o una base invertida en el extremo 3' del ARN. Tales modificaciones pueden incorporarse de manera rutinaria en oligonucleótidos y se ha mostrado que estos oligonucleótidos tienen una Tf mayor (es decir, afinidad de unión a diana mayor) que 2'-desoxioligonucleótidos frente a una diana dada.

Se ha mostrado que varias modificaciones de nucleótidos y nucleósidos hacen que el oligonucleótido en el que se incorporan sea más resistente frente a digestión con nucleasa que el oligonucleótido nativo; estos oligonucleótidos modificados sobreviven intactos durante más tiempo que los oligonucleótidos no modificados. Los ejemplos específicos de oligonucleótidos modificados incluyen los que comprenden estructuras principales modificadas, por ejemplo, fosforotioatos, fosfotriésteres, metil-fosfonatos, uniones entre azúcares de cicloalquilo o alquilo de cadena corta o uniones entre azúcares heterocíclicas o heteroatómicas de cadena corta. Algunos oligonucleótidos son oligonucleótidos con estructuras principales de fosforotioato y aquellos con estructuras principales de heteroátomos, particularmente estructuras principales de  $\text{CH}_2\text{-NH-O-CH}_2$ ,  $\text{CH}_2\text{-N(CH}_3\text{)-O-CH}_2$  (conocido como estructura principal de metileno (metilimino) o MMI),  $\text{CH}_2\text{-O-N(CH}_3\text{)-CH}_2$ ,  $\text{CH}_2\text{-N(CH}_3\text{)-N(CH}_3\text{)-CH}_2$  y  $\text{O-N(CH}_3\text{)-CH}_2\text{-CH}_2$ , en las que la estructura principal de fosfodiéster nativa se representa como  $\text{O-P-O-CH}$ ; estructuras principales de amida [véase De Mesmaeker *et al.*, *Ace. Chem. Res.*, 28:366-374 (1995)]; estructuras principales de morfolino (véase Summerton y Weller, patente estadounidense n.º 5.034.506); estructura principal de ácido nucleico peptídico (PNA) (en la que la estructura principal de fosfodiéster del oligonucleótido se sustituye por una estructura principal de poliamida, estando los nucleótidos unidos directa o indirectamente a los átomos de nitrógeno de aza de la estructura principal de poliamida, véase Nielsen *et al.*, *Science* 1991, 254, 1497). Las uniones que contienen fósforo incluyen, pero no se limitan a, fosforotioatos, fosforotioatos quirales, fosforoditioatos, fosfotriésteres, aminoalquilfosfotriésteres, fosfonatos de metilo y otros alquilo que comprenden 3'-alquilen-fosfonatos y fosfonatos quirales, fosfinatos, fosforamidatos que comprenden 3'-amino-fosforamidato y aminoalquilfosforamidatos, tionofosforamidatos, tionoalquilfosfonatos, tionoalquilfosfotriésteres y boranofosfatos que tienen uniones en 3'-5' normales, análogos unidos en 2'-5' de los mismos, y los que tienen polaridad invertida en los que los pares adyacentes de unidades de nucleósido están unidas de 3'-5' a 5'-3' o de 2'-5' a 5'-2'; véanse las patentes estadounidenses n.ºs 3.687.808; 4.469.863; 4.476.301; 5.023.243; 5.177.196; 5.188.897; 5.264.423; 5.276.019; 5.278.302; 5.286.717; 5.321.131; 5.399.676; 5.405.939; 5.453.496; 5.455.233; 5.466.677; 5.476.925; 5.519.126; 5.536.821; 5.541.306; 5.550.111; 5.563.253; 5.571.799; 5.587.361; y 5.625.050.

Se describen compuestos oligoméricos basados en morfolino en Braasch y David Corey, *Biochemistry*, 41(14): 4503-4510 (2002); Genesis, volumen 30, número 3, (2001); Heasman, *Dev. Biol.*, 243: 209-214 (2002); Nasevicius *et al.*, *Nat. Genet.*, 26:216-220 (2000); Lacerra *et al.*, *Proc. Natl. Acad. Sci.*, 97: 9591-9596 (2000); y la patente estadounidense n.º 5.034.506, expedida el 23 de julio de 1991.

Se describen agentes miméticos de oligonucleótidos de ácido nucleico de ciclohexenilo en Wang *et al.*, *J. Am. Chem. Soc.*, 122: 8595-8602 (2000).

Las estructuras principales de oligonucleótido modificadas que no incluyen un átomo de fósforo en las mismas tienen estructuras principales que están formadas por uniones entre nucleósidos de cicloalquilo o alquilo de cadena corta, uniones entre nucleósidos de cicloalquilo o alquilo y de heteroátomos mixtas, o una o más uniones entre nucleósidos heterocíclicas o heteroatómicas de cadena corta. Estas comprenden las que tienen uniones de morfolino (formadas en parte a partir de la porción de azúcar de un nucleósido); estructuras principales de siloxano; sulfuro, sulfóxido y estructuras principales de sulfona; estructuras principales de formacetilo y tioformacetilo; estructuras principales de

metilen-formaceto y tioformaceto; estructuras principales que contienen alqueno; estructuras principales de sulfamato; estructuras principales de metilenoimino y metilenohidrazino; estructuras principales de sulfonato y sulfonamida; estructuras principales de amida; y otras que tienen partes de componentes de N, O, S y CH<sub>2</sub> mixtas; véanse las patentes estadounidenses n.ºs 5.034.506; 5.166.315; 5.185.444; 5.214.134; 5.216.141; 5.235.033; 5.264.562; 5.264.564; 5.405.938; 5.434.257; 5.466.677; 5.470.967; 5.489.677; 5.541.307; 5.561.225; 5.596.086; 5.602.240; 5.610.289; 5.602.240; 5.608.046; 5.610.289; 5.618.704; 5.623.070; 5.663.312; 5.633.360; 5.677.437; y 5.677.439.

También pueden incluirse uno o más restos de azúcar sustituido, por ejemplo, uno de los siguientes en la posición 2': OH, SH, SCH<sub>3</sub>, F, OCN, OCH<sub>3</sub>OCH<sub>3</sub>, OCH<sub>3</sub>O(CH<sub>2</sub>)<sub>n</sub>CH<sub>3</sub>, O(CH<sub>2</sub>)<sub>n</sub>NH<sub>2</sub>, u O(CH<sub>2</sub>)<sub>n</sub>CH<sub>3</sub>, en los que n es desde 1 hasta aproximadamente 10; alquilo inferior C1 a C10, alcoxicolcoxilo, alquilo inferior sustituido, alcarilo o aralquilo; Cl; Br; CN; CF<sub>3</sub>; OCF<sub>3</sub>; O-, S- o N-alquilo; O-, S- o N-alqueno; SOCH<sub>3</sub>; SO<sub>2</sub>CH<sub>3</sub>; ONO<sub>2</sub>; NO<sub>2</sub>; N<sub>3</sub>; NH<sub>2</sub>; heterocicloalquilo; heterocicloalcarilo; aminoalquilamino; polialquilamino; sililo sustituido; un grupo de escisión de ARN; un grupo notificador; un intercalador; un grupo para mejorar las propiedades farmacocinéticas de un oligonucleótido; o un grupo para mejorar las propiedades farmacodinámicas de un oligonucleótido y otros sustituyentes que tienen propiedades similares. En algunos aspectos, una modificación incluye 2'-metoxietoxilo (2'-O-CH<sub>2</sub>CH<sub>2</sub>OCH<sub>3</sub>, también conocido como 2'-O-(2-metoxietilo)) (Martin *et al.*, *Helv. Chim. Acta*, 1995, 78, 486). Otras modificaciones incluyen 2'-metoxilo (2'-O-CH<sub>3</sub>), 2'-propoxilo (2'-OCH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>) y 2'-fluoro (2'-F). También pueden realizarse modificaciones similares en otras posiciones en el oligonucleótido, particularmente la posición 3' del azúcar en el nucleótido 3'-terminal y la posición 5' del nucleótido 5'-terminal. Los oligonucleótidos también pueden tener agentes miméticos de azúcar, tales como ciclobutilos en lugar del grupo pentofuranosilo.

En algunos ejemplos, tanto un azúcar como una unión entre nucleósidos, es decir, la estructura principal, de las unidades de nucleótido pueden sustituirse por grupos novedosos. Las unidades de base pueden mantenerse para la hibridación con un compuesto diana de ácido nucleico apropiado. Un compuesto oligomérico de este tipo, un agente mimético de oligonucleótido que se ha mostrado que tiene excelentes propiedades de hibridación, se denomina ácido nucleico peptídico (PNA). En compuestos de PNA, la estructura principal de azúcar de un oligonucleótido puede sustituirse por una estructura principal que contiene amida, por ejemplo, una estructura principal de aminoetilglicina. Las nucleobases pueden conservarse y unirse directa o indirectamente a átomos de nitrógeno de aza de la porción de amida de la estructura principal. Patentes estadounidenses representativas que enseñan la preparación de compuestos de PNA comprenden, pero no se limitan a, las patentes estadounidenses n.ºs 5.539.082; 5.714.331; y 5.719.262. Pueden encontrarse enseñanzas adicionales de compuestos de PNA en Nielsen *et al.*, *Science*, 254: 1497-1500 (1991).

Los ARN guía también pueden incluir, adicional o alternativamente, modificaciones o sustituciones de nucleobases (con frecuencia denominadas en la técnica simplemente "bases"). Tal como se usa en el presente documento, las nucleobases "no modificadas" o "naturales" incluyen adenina (A), guanina (G), timina (T), citosina (C) y uracilo (U). Las nucleobases modificadas incluyen nucleobases encontradas tan solo de manera poco frecuente o transitoria en ácidos nucleicos naturales, por ejemplo, hipoxantina, 6-metiladenina, 5-Me-pirimidinas, particularmente 5-metilcitosina (también denominada 5-metil-2'-desoxicitosina y con frecuencia denominada en la técnica as 5-Me-C), 5-hidroximetilcitosina (HMC), glicosil-HMC y gentobiosil-HMC, así como nucleobases sintéticas, por ejemplo, 2-aminoadenina, 2-(metilamino)adenina, 2-(imidazolilalquil)adenina, 2-(aminoalquilamino)adenina u otras alquiladeninas sustituidas con heteroátomos, 2-tiouracilo, 2-tiotimina, 5-bromouracilo, 5-hidroximetiluracilo, 8-azaguanina, 7-desazaguanina, N6-(6-aminoheptil)adenina y 2,6-diaminopurina. Kornberg, A., *DNA Replication*, W. H. Freeman & Co., San Francisco, págs. 75-77 (1980); Gebeyehu *et al.*, *Nucl. Acids Res.* 15: 4513 (1997). También puede incluirse una base "universal" conocida en la técnica, por ejemplo, inosina. Se ha mostrado que las sustituciones con 5-Me-C aumentan la estabilidad de dúplex de ácido nucleico en 0,6-1,2°C (Sanghvi, Y. S., en Crooke, S. T. and Lebleu, B., eds., *Antisense Research and Applications*, CRC Press, Boca Raton, 1993, págs. 276-278) y son aspectos de sustituciones de bases.

Las nucleobases modificadas pueden comprender otras nucleobases sintéticas y naturales, tales como 5-metilcitosina (5-me-C), 5-hidroximetil-citosina, xantina, hipoxantina, 2-aminoadenina, derivados de 6-metilo y otros alquilo de adenina y guanina, derivados de 2-propilo y otros alquilo de adenina y guanina, 2-tiouracilo, 2-tiotimina y 2-tiocitosina, 5-halouracilo y citosina, 5-propinil-uracilo y citosina, 6-azo-uracilo, citosina y timina, 5-uracilo (pseudo-uracilo), 4-tiouracilo, 8-halo, 8-amino, 8-tiol, 8-tioalquil, 8-hidroxi y otras adeninas y guaninas sustituidas en 8, 5-halo, particularmente 5-bromo, 5-trifluorometil y otros uracilos y citosinas sustituidos en 5, 7-metilguanina y 7-metiladenina, 8-azaguanina y 8-aza-adenina, 7-desazaguanina y 7-desaza-adenina, y 3-desazaguanina y 3-desaza-adenina.

Además, las nucleobases pueden comprender las dadas a conocer en la patente estadounidense n.º 3.687.808, las dadas a conocer en "The Concise Encyclopedia of Polymer Science And Engineering", páginas 858-859, Kroschwitz, J.I., ed. John Wiley & Sons, 1990, las dadas a conocer por Englisch *et al.*, *Angewandte Chemie*, International Edition, 1991, 30, página 613, y las dadas a conocer por Sanghvi, Y. S., capítulo 15, *Antisense Research and Applications*, páginas 289-302, Crooke, S.T. and Lebleu, B. ea., CRC Press, 1993. Algunas de estas nucleobases son particularmente útiles para aumentar la afinidad de unión de los compuestos oligoméricos de la invención. Estas incluyen pirimidinas sustituidas en 5, 6-azapirimidinas y purinas sustituidas en N-2, N-6 y O-6, que comprenden 2-aminopropiladenina, 5-propiniluracilo y 5-propinilcitosina. Se ha mostrado que las sustituciones con 5-metilcitosina

aumentan la estabilidad de dúplex de ácido nucleico en 0,6-1,2°C (Sanghvi, Y.S., Crooke, S.T. and Lebleu, B., eds, "Antisense Research and Applications", CRC Press, Boca Raton, 1993, págs. 276-278) y son aspectos de sustituciones de bases, incluso más particularmente cuando se combinan con modificaciones de azúcar con 2'-O-metoxietil. Se describen nucleobases modificadas en las patentes estadounidenses n.ºs 3.687.808, así como 4.845.205; 5.130.302; 5.134.066; 5.175.273; 5.367.066; 5.432.272; 5.457.187; 5.459.255; 5.484.908; 5.502.177; 5.525.711; 5.552.540; 5.587.469; 5.596.091; 5.614.617; 5.681.941; 5.750.692; 5.763.588; 5.830.653; 6.005.096; y la publicación de solicitud de patente estadounidense 2003/0158403.

Por tanto, el término "modificado" se refiere a un azúcar, fosfato o base no natural que se incorpora en un ARN guía, una endonucleasa, o secuencia de control transcripcional de BCL11A o tanto un ARN guía como una endonucleasa. No es necesario que todas las posiciones en un oligonucleótido dado estén modificadas de manera uniforme y, de hecho, más de una de las modificaciones anteriormente mencionadas pueden incorporarse en un único oligonucleótido, o incluso en un único nucleósido dentro de un oligonucleótido.

Los ARN guía y/o ARNm (o ADN) que codifican para una endonucleasa pueden estar químicamente unidos a uno o más restos o conjugados que potencian la actividad, distribución celular o captación celular del oligonucleótido. Tales restos comprenden, pero no se limitan a, restos de lípidos tales como un resto de colesterol [Letsinger *et al.*, Proc. Natl. Acad. Sci. USA, 86: 6553-6556 (1989)]; ácido cólico [Manoharan *et al.*, Bioorg. Med. Chem. Lett., 4: 1053-1060 (1994)]; un tioéter, por ejemplo, hexil-S-tritilitol [Manoharan *et al.*, Ann. N. Y. Acad. Sci., 660: 306-309 (1992) y Manoharan *et al.*, Bioorg. Med. Chem. Lett., 3: 2765-2770 (1993)]; un tiocolesterol [Oberhauser *et al.*, Nucl. Acids Res., 20: 533-538 (1992)]; una cadena alifática, por ejemplo, residuos dodecanodiol o undecilo [Kabanov *et al.*, FEBS Lett., 259: 327-330 (1990) y Svinarchuk *et al.*, Biochimie, 75: 49-54 (1993)]; un fosfolípido, por ejemplo, di-hexadecil-rac-glicerol o 1,2-di-O-hexadecil-rac-glicerol-3-H-fosfonato de trietilamonio [Manoharan *et al.*, Tetrahedron Lett., 36: 3651-3654 (1995) y Shea *et al.*, Nucl. Acids Res., 18: 3777-3783 (1990)]; una poliamina o una cadena de polietilenglicol [Manoharan *et al.*, Nucleosides & Nucleotides, 14: 969-973 (1995)]; ácido adamantano-acético [Manoharan *et al.*, Tetrahedron Lett., 36: 3651-3654 (1995)]; un resto palmitilo [(Mishra *et al.*, Biochim. Biophys. Acta, 1264: 229-237 (1995)]; o un resto octadecilamina o hexilamino-carbonil-t-oxicolessterol [Crooke *et al.*, J. Pharmacol. Exp. Ther., 277: 923-937 (1996)]. Véanse también las patentes estadounidenses n.ºs 4.828.979; 4.948.882; 5.218.105; 5.525.465; 5.541.313; 5.545.730; 5.552.538; 5.578.717; 5.580.731; 5.580.731; 5.591.584; 5.109.124; 5.118.802; 5.138.045; 5.414.077; 5.486.603; 5.512.439; 5.578.718; 5.608.046; 4.587.044; 4.605.735; 4.667.025; 4.762.779; 4.789.737; 4.824.941; 4.835.263; 4.876.335; 4.904.582; 4.958.013; 5.082.830; 5.112.963; 5.214.136; 5.082.830; 5.112.963; 5.214.136; 5.245.022; 5.254.469; 5.258.506; 5.262.536; 5.272.250; 5.292.873; 5.317.098; 5.371.241; 5.391.723; 5.416.203; 5.451.463; 5.510.475; 5.512.667; 5.514.785; 5.565.552; 5.567.810; 5.574.142; 5.585.481; 5.587.371; 5.595.726; 5.597.696; 5.599.923; 5.599.928 y 5.688.941.

Pueden usarse azúcares y otros restos para dirigir proteínas y complejos que comprenden nucleótidos, tales como polisomas catiónicos y liposomas, a sitios particulares. Por ejemplo, la transferencia dirigida por células hepáticas puede mediar mediante receptores de asialoglicoproteína (ASGPR); véase, por ejemplo, Hu, *et al.*, Protein Pept Lett. 21(10):1025-30 (2014). Pueden usarse otros sistemas conocidos en la técnica y desarrollados de manera regular para dirigir biomoléculas de uso en el presente caso y/o complejos de las mismas a células de interés diana particulares.

Estos restos o conjugados de direccionamiento pueden incluir grupos conjugados unidos de manera covalente a grupos funcionales, tales como grupos hidroxilo primario o secundario. Los grupos conjugados de la divulgación incluyen intercaladores, moléculas notificadoras, poliaminas, poliamidas, polietilenglicoles, poliéteres, grupos que potencian las propiedades farmacodinámicas de oligómeros, y grupos que potencian las propiedades farmacocinéticas de oligómeros. Los grupos conjugados típicos incluyen colesterol, lípidos, fosfolípidos, biotina, fenazina, folato, fenantridina, antraquinona, acridina, fluoresceínas, rodaminas, cumarinas y colorantes. Los grupos que potencian las propiedades farmacodinámicas, en el contexto de esta divulgación, incluyen grupos que mejoran la captación, potencian la resistencia a la degradación, y/o refuerzan la hibridación específica de secuencia con el ácido nucleico diana. Los grupos que potencian las propiedades farmacocinéticas, en el contexto de esta divulgación, incluyen grupos que mejoran la captación, distribución, metabolismo o excreción de los compuestos de la presente divulgación. Se dan a conocer grupos conjugados representativos en la patente estadounidense n.º 6.287.860. Los restos conjugados incluyen, pero no se limitan a, restos de lípidos tales como un resto de colesterol, ácido cólico, un tioéter, por ejemplo, hexil-5-tritilitol, un tiocolesterol, una cadena alifática, por ejemplo, residuos dodecanodiol o undecilo, un fosfolípido, por ejemplo, di-hexadecil-rac-glicerol o 1,2-di-O-hexadecil-rac-glicerol-3-H-fosfonato de trietilamonio, una poliamina o una cadena de polietilenglicol, o ácido adamantano-acético, un resto palmitilo, o un resto octadecilamina o hexilamino-carbonil-oxi-colessterol. Véanse, por ejemplo, las patentes estadounidenses n.ºs 4.828.979; 4.948.882; 5.218.105; 5.525.465; 5.541.313; 5.545.730; 5.552.538; 5.578.717; 5.580.731; 5.580.731; 5.591.584; 5.109.124; 5.118.802; 5.138.045; 5.414.077; 5.486.603; 5.512.439; 5.578.718; 5.608.046; 4.587.044; 4.605.735; 4.667.025; 4.762.779; 4.789.737; 4.824.941; 4.835.263; 4.876.335; 4.904.582; 4.958.013; 5.082.830; 5.112.963; 5.214.136; 5.082.830; 5.112.963; 5.214.136; 5.245.022; 5.254.469; 5.258.506; 5.262.536; 5.272.250; 5.292.873; 5.317.098; 5.371.241; 5.391.723; 5.416.203; 5.451.463; 5.510.475; 5.512.667; 5.514.785; 5.565.552; 5.567.810; 5.574.142; 5.585.481; 5.587.371; 5.595.726; 5.597.696; 5.599.923; 5.599.928 y 5.688.941.

Polinucleótidos más largos, que son menos propensos a síntesis química y normalmente se producen mediante

síntesis enzimática, también pueden modificarse mediante diversos medios. Tales modificaciones pueden incluir, por ejemplo, la introducción de determinados análogos de nucleótido, la incorporación de secuencias particulares u otros restos en los extremos 5' o 3' de moléculas, y otras modificaciones. A modo de ilustración, el ARNm que codifica para Cas9 tiene aproximadamente 4 kb de longitud y puede sintetizarse mediante transcripción *in vitro*. Pueden aplicarse modificaciones al ARNm, por ejemplo, para aumentar su traducción o estabilidad (tal como aumentando su resistencia a la degradación con una célula), o para reducir la tendencia del ARN a provocar una respuesta inmunitaria innata que se observa con frecuencia en células tras la introducción de ARN exógenos, particularmente ARN más largos tales como el que codifica para Cas9.

Se han descrito numerosas de tales modificaciones en la técnica, tales como colas de poliA, análogos de caperuza en 5' (por ejemplo, análogo de caperuza anti-inverso (ARCA) o m7G(5')ppp(5')G (mCAP)), regiones no traducidas (UTR) en 5' o 3' modificadas, uso de bases modificadas (tal como pseudo-UTP, 2-tio-UTP, 5-metilcitidina-5'-trifosfato (5-metil-CTP) o N6-metil-ATP), o tratamiento con fosfatasa para eliminar fosfatos terminales en 5'. Estas y otras modificaciones se conocen en la técnica, y nuevas modificaciones de ARN están desarrollándose de manera regular.

Hay numerosos proveedores comerciales de ARN modificados, incluyendo, por ejemplo, TriLink Biotech, AxoLabs, Bio-Synthesis Inc., Dharmacon y muchos otros. Tal como se describe por TriLink, por ejemplo, puede usarse 5-metil-CTP para conferir características deseables, tales como aumento de la estabilidad frente a nucleasa, aumento de la traducción o interacción reducida de receptores inmunitarios innatos con ARN transcrito *in vitro*. También se ha mostrado que 5-metilcitidina-5'-trifosfato (5-metil-CTP), N6-metil-ATP, así como pseudo-UTP y 2-tio-UTP, reducen la estimulación inmunitaria innata en cultivo e *in vivo* al tiempo que potencian la traducción, tal como se ilustra en publicaciones de Kormann *et al.* y de Warren *et al.* mencionadas a continuación.

Se ha mostrado que puede usarse ARNm químicamente modificado suministrado *in vivo* para lograr efectos terapéuticos mejorados; véase, por ejemplo, Kormann *et al.*, Nature Biotechnology 29, 154-157 (2011). Tales modificaciones pueden usarse, por ejemplo, para aumentar la estabilidad de la molécula de ARN y/o reducir su inmunogenicidad. Usando modificaciones químicas tales como pseudo-U, N6-metil-A, 2-tio-U y 5-metil-C, se encontró que sustituir tan solo una cuarta parte de los residuos de uridina y citidina por 2-tio-U y 5-metil-C, respectivamente, daba como resultado una disminución significativa del reconocimiento mediado por receptor de tipo Toll (TLR) del ARNm en ratones. Reduciendo la activación del sistema inmunitario innato, pueden usarse estas modificaciones para aumentar eficazmente la estabilidad y longevidad del ARNm *in vivo*; véase, por ejemplo, Kormann *et al.*, citado anteriormente.

También se ha mostrado que la administración repetida de ARN mensajeros sintéticos que incorporan modificaciones diseñadas para evitar respuestas antivirales innatas pueden reprogramar células humanas diferenciadas para dar pluripotencia. Véase, por ejemplo, Warren, *et al.*, Cell Stem Cell, 7(5):618-30 (2010). Tales ARNm modificados que actúan como proteínas de reprogramación primarias pueden ser unos medios eficientes de reprogramación de múltiples tipos de células humanas. Tales células se denominan células madre de pluripotencia inducidas (iPSC), y se encontró que puede usarse ARN enzimáticamente sintetizado que incorpora 5-metil-CTP, pseudo-UTP y un análogo de caperuza anti-inverso (ARCA) para evadir eficazmente la respuesta antiviral de la célula; véase, por ejemplo, Warren *et al.*, citado anteriormente.

Otras modificaciones de polinucleótidos descritas en la técnica incluyen, por ejemplo, el uso de colas de poliA, la adición de análogos de caperuza en 5' (tales como m7G(5')ppp(5')G (mCAP)), modificaciones de regiones no traducidas (UTR) en 5' o 3', o tratamiento con fosfatasa para eliminar fosfatos terminales en 5', y nuevos enfoques están desarrollándose de manera regular.

Se han desarrollado varias composiciones y técnicas aplicables a la generación de ARN modificados para su uso en el presente documento en relación con la modificación de interferencia de ARN (iARN), incluyendo ARN de interferencia pequeños (ARNip). Los ARNip presentan dificultades particulares *in vivo* porque sus efectos sobre el silenciamiento génico mediante interferencia de ARNm son generalmente transitorios, lo cual puede requerir una administración repetida. Además, los ARNip son ARN bicatenarios (ARNbc) y las células de mamífero tienen respuestas inmunitarias que han evolucionado para detectar y neutralizar ARNbc, que con frecuencia es un subproducto de la infección viral. Por tanto, hay enzimas de mamífero tales como PKR (cinasa sensible a ARNbc), y posiblemente gen inducible por ácido retinoico I (RIG-I), que pueden mediar en respuestas celulares frente a ARNbc, así como receptores de tipo Toll (tales como TLR3, TLR7 y TLR8) que pueden desencadenar la inducción de citocinas en respuesta a tales moléculas; véanse, por ejemplo, las revisiones de Angart *et al.*, Pharmaceuticals (Basel) 6(4): 440-468 (2013); Kanasty *et al.*, Molecular Therapy 20(3): 513-524 (2012); Burnett *et al.*, Biotechnol J. 6(9):1130-46 (2011); Judge y MacLachlan, Hum Gene Ther 19(2):111-24 (2008); y referencias citadas en los mismos.

Se ha desarrollado y aplicado una gran variedad de modificaciones para potenciar la estabilidad de ARN, reducir las respuestas inmunitarias innatas y/o lograr otros beneficios que pueden ser útiles en relación con la introducción de polinucleótidos en células humanas, tal como se describe en el presente documento; véanse, por ejemplo, las revisiones de Whitehead KA *et al.*, Annual Review of Chemical and Biomolecular Engineering, 2: 77-96 (2011); Gaglione y Messere, Mini Rev Med Chem, 10(7):578-95 (2010); Chernolovskaya *et al.*, Curr Opin Mol Ther., 12(2):158-67 (2010); Deleavey *et al.*, Curr Protoc Nucleic Acid Chem Chapter 16: Unit 16.3 (2009); Behlke, Oligonucleotides

18(4):305-19 (2008); Fucini *et al.*, *Nucleic Acid Ther* 22(3): 205-210 (2012); Bremsen *et al.*, *Front Genet* 3:154 (2012).

Tal como se indicó anteriormente, hay varios proveedores comerciales de ARN modificados, muchos de los cuales se han especializado en modificaciones diseñadas para mejorar la eficacia de los ARNip. Se ofrece una variedad de enfoques basándose en diversos hallazgos notificados en la bibliografía. Por ejemplo, Dharmacon indica que la sustitución de un oxígeno que no forma puente por azufre (fosforotioato, PS) se ha usado extensamente para mejorar la resistencia a nucleasa de ARNip, tal como se notifica por Kole, *Nature Reviews Drug Discovery* 11:125-140 (2012). Se ha notificado que modificaciones de la posición 2' de la ribosa mejoran la resistencia a nucleasa del enlace fosfato entre nucleótidos al tiempo que aumenta la estabilidad de dúplex (Tf), lo cual también se ha mostrado que proporciona protección frente a activación inmunitaria. Una combinación de modificaciones de estructura principal de PS moderadas con pequeñas sustituciones en 2' bien toleradas (2'-O-metilo, 2'-fluoro, 2'-hidro) se ha asociado con ARNip altamente estables para aplicaciones *in vivo*, tal como se notifica por Soutschek *et al.* *Nature* 432:173-178 (2004); y se ha notificado que las modificaciones con 2'-O-metilo son eficaces para mejorar la estabilidad tal como se notifica por Volkov, *Oligonucleotides* 19:191-202 (2009). Con respecto a la reducción de la inducción de respuestas inmunitarias innatas, se ha notificado que modificar secuencias específicas con 2'-O-metilo, 2'-fluoro, 2'-hidro reduce la interacción TLR7/TLR8 al tiempo que conserva generalmente la actividad de silenciamiento; véanse, por ejemplo, Judge *et al.*, *Mol. Ther.* 13:494-505 (2006); y Cekaite *et al.*, *J. Mol. Biol.* 365:90-108 (2007). También se ha mostrado que modificaciones adicionales, tales como 2-tiouracilo, pseudouracilo, 5-metilcitosina, 5-metiluracilo y N6-metiladenosina minimizan los efectos inmunitarios mediados por TLR3, TLR7 y TLR8; véase, por ejemplo, Kariko, K. *et al.*, *Immunity* 23:165-175 (2005).

Tal como se conoce en la en la técnica, y está comercialmente disponible, pueden aplicarse varios conjugados a polinucleótidos, tales como ARN, para su uso en el presente documento que pueden potenciar su suministro y/o captación por células, incluyendo, por ejemplo, colesterol, tocoferol y ácido fólico, lípidos, péptidos, polímeros, ligadores y aptámeros; véase, por ejemplo, la revisión de Winkler, *Ther. Deliv.* 4:791-809 (2013), y referencias citadas en el mismo.

#### Optimización de codones

Un polinucleótido que codifica para un polipéptido dirigido al sitio puede someterse a optimización de codones según métodos convencionales en la técnica para la expresión en la célula que contiene el ADN diana de interés. Por ejemplo, si el ácido nucleico diana previsto está en una célula humana, se contempla un polinucleótido humano sometido a optimización de codones que codifica para Cas9 para su uso para la producción del polipéptido de Cas9.

Complejos de un ácido nucleico de direccionamiento al genoma y un polipéptido dirigido al sitio

Un ácido nucleico de direccionamiento al genoma interacciona con un polipéptido dirigido al sitio (por ejemplo, una nucleasa guiada por ácido nucleico tal como Cas9), formando de ese modo un complejo. El ácido nucleico de direccionamiento al genoma guía el polipéptido dirigido al sitio a un ácido nucleico diana.

#### RNP

Cada uno del polipéptido dirigido al sitio y el ácido nucleico de direccionamiento al genoma pueden administrarse por separado a una célula o a un paciente. Por otro lado, el polipéptido dirigido al sitio puede complejarse previamente con uno o más ácidos nucleicos de direccionamiento al genoma (ARN guía, ARNGs, o ARNcr junto con un ARNtracr). Entonces puede administrarse el material previamente complejoado a una célula o a un paciente. Tal material previamente complejoado se conoce como partícula de ribonucleoproteína (RNP). El polipéptido dirigido al sitio en la RNP puede ser, por ejemplo, una endonucleasa Cas9 o una endonucleasa Cpf1. El polipéptido dirigido al sitio puede estar flanqueado en el extremo N-terminal, el extremo C-terminal o tanto el extremo N-terminal como el extremo C-terminal por una o más señales de localización nuclear (NLS). Por ejemplo, una endonucleasa Cas9 puede estar flanqueada por dos NLS, una NLS ubicada en el extremo N-terminal y la segunda NLS ubicada en el extremo C-terminal. La NLS puede ser cualquier NLS conocida en la técnica, tal como una NLS de SV40. La razón en peso de ácido nucleico de direccionamiento al genoma con respecto a polipéptido dirigido al sitio en la RNP puede ser de 1:1. Por ejemplo, la razón en peso de ARNGs con respecto a endonucleasa Cas9 en la RNP puede ser de 1:1. Por ejemplo, el ARNGs puede comprender la secuencia de ácido nucleico de SEQ ID NO: 71.959, la endonucleasa Cas9 puede ser una Cas9 de *S. pyogenes* que comprende una NLS de SV40 en el extremo N-terminal y una NLS de SV40 en el extremo C-terminal, y la razón en peso de ARNGs con respecto a endonucleasa Cas9 puede ser de 1:1.

#### Ácidos nucleicos que codifican para componentes de sistema

La presente divulgación da a conocer un ácido nucleico que comprende una secuencia de nucleótidos que codifica para un ácido nucleico de direccionamiento al genoma de la divulgación, un polipéptido dirigido al sitio de la divulgación y/o cualquier molécula de ácido nucleico o proteica necesaria para llevar a cabo los aspectos de los métodos de la divulgación.

El ácido nucleico que codifica para un ácido nucleico de direccionamiento al genoma de la divulgación, un polipéptido

dirigido al sitio de la divulgación y/o cualquier molécula de ácido nucleico o proteica necesaria para llevar a cabo los aspectos de los métodos de la divulgación puede comprender un vector (por ejemplo, un vector de expresión recombinante).

El término “vector” se refiere a una molécula de ácido nucleico que puede transportar otro ácido nucleico al que se ha unido. Un tipo de vector es un “plásmido”, que se refiere a un bucle de ADN bicatenario circular en el que pueden ligarse segmentos de ácido nucleico adicionales. Otro tipo de vector es un vector viral, en el que pueden ligarse segmentos de ácido nucleico adicionales en el genoma viral. Determinados vectores pueden realizar replicación autónoma en una célula huésped en la que se introducen (por ejemplo, vectores bacterianos que tienen un origen de replicación bacteriano y vectores de mamífero episomales). Otros vectores (por ejemplo, vectores de mamífero no episomales) se integran en el genoma de una célula huésped tras la introducción en la célula huésped y, de ese modo, se replican junto con el genoma huésped.

En algunos ejemplos, los vectores pueden ser capaces de dirigir la expresión de ácidos nucleicos a los que están operativamente unidos. Tales vectores se denominan en el presente documento “vectores de expresión recombinante”, o más simplemente, “vectores de expresión”, que sirven para funciones equivalentes.

El término “operativamente unido” significa que la secuencia de nucleótidos de interés está unida a secuencia(s) reguladora(s) de una manera que permite la expresión de la secuencia de nucleótidos. Se pretende que el término “secuencia reguladora” incluya, por ejemplo, promotores, potenciadores y otros elementos de control de la expresión (por ejemplo, señales de poliadenilación). Tales secuencias reguladoras se conocen bien en la técnica y se describen, por ejemplo, en Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Las secuencias reguladoras incluyen las que dirigen la expresión constitutiva de una secuencia de nucleótidos en muchos tipos de células huésped y las que dirigen la expresión de la secuencia de nucleótidos únicamente en determinadas células huésped (por ejemplo, secuencias reguladoras específicas de tejido). Los expertos en la técnica apreciarán que el diseño del vector de expresión puede depender de factores tales como la elección de la célula diana, el nivel de expresión deseado y similares.

Los vectores de expresión contemplados incluyen, pero no se limitan a, vectores virales basados en virus vaccinia, poliovirus, adenovirus, virus adenoasociado, SV40, virus del herpes simple, virus de la inmunodeficiencia humana, retrovirus (por ejemplo, virus de leucemia murina, virus de necrosis del bazo y vectores derivados de retrovirus tales como virus del sarcoma de Rous, virus del sarcoma de Harvey, virus de la leucosis aviar, lentivirus, virus de la inmunodeficiencia humana, virus del sarcoma mieloproliferativo y virus de tumor de mama) y otros vectores recombinantes. Otros vectores contemplados para células diana eucariotas incluyen, pero no se limitan a, los vectores pXT1, pSG5, pSVK3, pBPV, pMSG y pSVLSV40 (Farmacia). Vectores adicionales contemplados para células diana eucariotas incluyen, pero no se limitan a, los vectores pCTx-1, pCTx-2, y pCTx-3, que se describen en las figuras 1A a 1C. Pueden usarse otros vectores siempre que sean compatibles con la célula huésped.

En algunos ejemplos, un vector puede comprender uno o más elementos de control de la transcripción y/o la traducción. Dependiendo del sistema de huésped/vector usado, puede usarse cualquiera de varios elementos de control de la transcripción y la traducción adecuados, incluyendo promotores constitutivos e inducibles, elementos potenciadores de la transcripción, terminadores de la transcripción, etc., en el vector de expresión. El vector puede ser un vector autoinactivante que o bien inactiva las secuencias virales o bien los componentes de la maquinaria de CRISPR u otros elementos.

Los ejemplos no limitativos de promotores eucariotas adecuados (es decir, promotores funcionales en una célula eucariota) incluyen los de citomegalovirus (CMV) inmediato temprano, timidina cinasa de virus del herpes simple (HSV), SV40 temprano y tardío, repeticiones terminales largas (LTR) de retrovirus, promotor de factor de elongación 1 (EF1) humano, un constructo híbrido que comprende el potenciador de citomegalovirus (CMV) fusionado al promotor de beta-actina de pollo (CAG), promotor de virus de células madre murino (MSCV), promotor de locus de fosfoglicerato cinasa 1 (PGK) y metalotioneína I de ratón.

Para expresar ARN pequeños, incluyendo ARN guía usados en relación con endonucleasa Cas, diversos promotores tales como promotores de ARN polimerasa III, incluyendo, por ejemplo, U6 y H1, pueden ser ventajosos. En la técnica se conocen descripciones, y parámetros para potenciar el uso, de tales promotores, e información y enfoques adicionales se describen de manera regular; véase, por ejemplo, Ma, H. *et al.*, *Molecular Therapy - Nucleic Acids* 3, e161 (2014) doi: 10.1038/mtna.2014.12.

El vector de expresión también puede contener un sitio de unión al ribosoma para el inicio de la traducción y un terminador de la transcripción. El vector de expresión también puede comprender secuencias apropiadas para amplificar la expresión. El vector de expresión también puede incluir secuencias de nucleótidos que codifican para etiquetas no nativas (por ejemplo, etiqueta de histidina, etiqueta de hemaglutinina, proteína verde fluorescente, etc.) que se fusionan al polipéptido dirigido al sitio, dando por tanto como resultado una proteína de fusión.

Un promotor puede ser un promotor inducible (por ejemplo, un promotor de choque térmico, promotor regulado por tetraciclina, promotor regulado por esteroides, promotor regulado por metal, promotor regulado por receptor de

estrógenos, etc.). El promotor puede ser un promotor constitutivo (por ejemplo, promotor de CMV, promotor de UBC). En algunos casos, el promotor puede ser un promotor espacialmente restringido y/o temporalmente restringido (por ejemplo, un promotor específico de tejido, un promotor específico de tipo de célula, etc.).

El ácido nucleico que codifica para un ácido nucleico de direccionamiento al genoma de la divulgación y/o un polipéptido dirigido al sitio puede empaquetarse en o sobre la superficie de vehículos de suministro para el suministro a células. Los vehículos de suministro contemplados incluyen, pero no se limitan a, nanoesferas, liposomas, puntos cuánticos, nanopartículas, partículas de polietilenglicol, hidrogeles y micelas. Tal como se describe en la técnica, puede usarse una variedad de restos de direccionamiento para potenciar la interacción preferente de tales vehículos con ubicaciones o tipos de células deseados.

La introducción de los complejos, polipéptidos y ácidos nucleicos de la divulgación en células puede producirse mediante infección viral o por bacteriófago, transfección, conjugación, fusión de protoplasto, lipofección, electroporación, nucleofección, precipitación con fosfato de calcio, transfección mediada por polietilénimina (PEI), transfección mediada por DEAE-dextrano, transfección mediada por liposoma, tecnología de pistola de partículas, precipitación con fosfato de calcio, micro-inyección directa, suministro de ácido nucleico mediado por nanopartículas y similares.

#### Suministro

Pueden suministrarse polinucleótidos de ARN guía (ARN o ADN) y/o polinucleótido(s) de endonucleasa (ARN o ADN) mediante vehículos de suministro virales o no virales conocidos en la técnica, tales como electroporación, fuerza mecánica, deformación celular (SQZ Biotech) y péptidos de penetración en células. Alternativamente, puede(n) suministrarse polipéptido(s) de endonucleasa mediante vehículos de suministro virales o no virales conocidos en la técnica, tales como electroporación o nanopartículas lipídicas. En aspectos alternativos adicionales, la ADN endonucleasa puede suministrarse como uno o más polipéptidos, o bien solos o bien previamente complejados con uno o más ARN guía, o uno o más ARNcr junto con un ARNtracr.

La electroporación es una técnica de suministro en la que se aplica un campo eléctrico a una o más células con el fin de aumentar la permeabilidad de la membrana celular, lo cual permite introducir sustancias tales como fármacos, ácidos nucleicos (ácidos nucleicos de direccionamiento al genoma), proteínas (polipéptidos dirigidos al sitio) o RNP, en la célula. En general, la electroporación funciona haciendo pasar miles de voltios a través de una distancia de uno a dos milímetros de células en suspensión en una cubeta de electroporación (1,0 - 1,5 kV, 250 - 750 V/cm).

Pueden suministrarse polinucleótidos mediante vehículos de suministro no virales incluyendo, pero sin limitarse a, nanopartículas, liposomas, ribonucleoproteínas, péptidos con carga positiva, conjugados de ARN de molécula pequeña, quimeras de aptámero-ARN y complejos de ARN-proteína de fusión. Algunos vehículos de suministro no virales a modo de ejemplo se describen en Peer y Lieberman, *Gene Therapy*, 18: 1127-1133 (2011) (que se centra en vehículos de suministro no virales para ARNip que también son útiles para el suministro de otros polinucleótidos).

Pueden suministrarse polinucleótidos, tales como ARN guía, ARNGs y ARNm que codifica para una endonucleasa, a una célula o a un paciente mediante una nanopartícula lipídica (LNP).

Una LNP se refiere a cualquier partícula que tiene un diámetro de menos de 1000 nm, 500 nm, 250 nm, 200 nm, 150 nm, 100 nm, 75 nm, 50 nm o 25 nm. Alternativamente, el tamaño de una nanopartícula puede oscilar desde 1-1000 nm, 1-500 nm, 1-250 nm, 25-200 nm, 25-100 nm, 35-75 nm o 25-60 nm.

Las LNP pueden realizarse a partir de lípidos catiónicos, aniónicos o neutros. Los lípidos neutros, tales como el fosfolípido fusogénico DOPE o el componente de membrana colesterol, pueden incluirse en LNP como "lípidos auxiliares" para potenciar la actividad de transfección y estabilidad de nanopartículas. Las limitaciones de los lípidos catiónicos incluyen baja eficacia debido a una escasa estabilidad y rápido aclaramiento, así como la generación de respuestas inflamatorias o antiinflamatorias.

Las LNP también pueden estar compuestas por lípidos hidrófobos, lípidos hidrófilos o lípidos tanto hidrófobos como hidrófilos.

Puede usarse cualquier lípido o combinación de lípidos que se conocen en la técnica para producir una LNP. Ejemplos de lípidos usados para producir LNP son: DOTMA, DOSPA, DOTAP, DMRIE, DC-colesterol, DOTAP-colesterol, GAP-DMORIE-DPyPE, y GL67A-DOPE-DMPE-polietilenglicol (PEG). Ejemplos de lípidos catiónicos son: 98N12-5, C12-200, DLin-KC2-DMA (KC2), DLin-MC3-DMA (MC3), XTC, MD1 y 7C1. Ejemplos de lípidos neutros son: DPSC, DPPC, POPC, DOPE y SM. Ejemplos de lípidos modificados con PEG son: PEG-DMG, PEG-CerC14 y PEG-CerC20.

Los lípidos pueden combinarse en cualquiera de varias razones molares para producir una LNP. Además, el/los polinucleótido(s) puede(n) combinarse con lípido(s) en un amplio intervalo de razones molares para producir una LNP.

Tal como se indicó anteriormente, el polipéptido dirigido al sitio y el ácido nucleico de direccionamiento al genoma



pueden administrarse, cada uno, por separado a una célula o a un paciente. Por otro lado, el polipéptido dirigido al sitio puede complejarse previamente con uno o más ARN guía, o uno o más ARNcr junto con un ARNtracr. Entonces puede suministrarse el material previamente complejoado a una célula o a un paciente. Tal material previamente complejoado se conoce como partícula de ribonucleoproteína (RNP).

ARN puede formar interacciones específicas con ARN o ADN. Aunque esta propiedad se aprovecha en muchos procesos biológicos, también presenta el riesgo de interacciones promiscuas en un entorno celular rico en ácido nucleico. Una solución a este problema es la formación de partículas de ribonucleoproteína (RNP), en las que el ARN está previamente complejoado con una endonucleasa. Otro beneficio de la RNP es la protección del ARN frente a la degradación.

La endonucleasa en la RNP puede estar modificada o no modificada. Asimismo, el ARNg, ARNcr, ARNtracr o ARNGs puede estar modificado o no modificado. Se conocen numerosas modificaciones en la técnica y pueden usarse.

La endonucleasa y el ARNGs pueden combinarse generalmente en una razón molar de 1:1. Alternativamente, la endonucleasa, ARNcr y ARNtracr pueden combinarse generalmente en una razón molar de 1:1:1. Sin embargo, puede usarse un amplio intervalo de razones molares para producir una RNP.

Puede usarse un vector recombinante de virus adenoasociado (AAV) para el suministro. Las técnicas para producir partículas de partículas rAAV, en las que un genoma de AAV que va a empaquetarse que incluye el polinucleótido que va a suministrarse, genes de rep y caperuza (cap), y funciones de virus auxiliar se proporcionan a una célula, son convencionales en la técnica. La producción de rAAV requiere normalmente que los siguientes componentes estén presentes dentro de una única célula (denominada en el presente documento célula de empaquetamiento): un genoma de rAAV, genes de rep y caperuza de AAV separados del (es decir, no en el) genoma de rAAV, y funciones de virus auxiliar. Los genes de rep y caperuza de AAV pueden ser de cualquier serotipo de AAV para el que puede suministrarse virus recombinante, y pueden ser de un serotipo de AAV diferente de las ITR de genoma de rAAV, incluyendo, pero sin limitarse a, serotipos de AAV AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-7, AAV-8, AAV-9, AAV-10, AAV-11, AAV-12, AAV-13 y AAV rh.74. La producción de rAAV pseudotipado se da a conocer, por ejemplo, en la publicación de solicitud de patente internacional número WO 01/83692. Véase la tabla 2.

Tabla 2

Serotipo de AAV	N.º de registro de Genbank
AAV-1	NC_002077.1
AAV-2	NC_001401.2
AAV-3	NC_001729.1
AAV-3B	AF028705.1
AAV-4	NC_001829.1
AAV-5	NC_006152.1
AAV-6	AF028704.1
AAV-7	NC_006260.1
AAV-8	NC_006261.1
AAV-9	AX753250.1
AAV-10	AY631965.1
AAV-11	AY631966.1
AAV-12	DQ813647.1
AAV-13	EU285562.1

Un método de generación de una célula de empaquetamiento implica crear una línea celular que expresa de manera estable todos los componentes necesarios para la producción de partículas de AAV. Por ejemplo, se integra un plásmido (o múltiples plásmidos) que comprende un genoma de rAAV que carece de genes de rep y caperuza de AAV, genes de rep y caperuza de AAV separados del genoma de rAAV, y un marcador seleccionable, tal como un gen de resistencia a neomicina, en el genoma de una célula. Se han introducido genomas de AAV en plásmidos bacterianos mediante procedimientos tales como colas de GC (Samulski *et al.*, 1982, Proc. Natl. Acad. Sci. USA, 79:2077-2081), adición de ligadores sintéticos que contienen sitios de escisión con endonucleasa de restricción (Laughlin *et al.*, 1983, Gene, 23:65-73) o mediante ligación de extremos romos directa (Senapathy & Carter, 1984, J. Biol. Chem., 259:4661-4666). Entonces puede infectarse la línea de células de empaquetamiento con un virus auxiliar, tal como adenovirus. Las ventajas de este método son que las células son seleccionables y son adecuadas para la producción a gran escala de rAAV. Otros ejemplos de métodos adecuados emplean adenovirus o baculovirus, en vez de plásmidos, para introducir genomas de rAAV y/o genes de rep y caperuza en células de empaquetamiento.

Los principios generales de producción de rAAV se revisan, por ejemplo, en Carter, 1992, Current Opinions in Biotechnology, 1533-539; y Muzyczka, 1992, Curr. Topics in Microbial. and Immunol., 158:97-129). Se describen diversos enfoques en Ratschin *et al.*, Mol. Cell. Biol. 4:2072 (1984); Hermonat *et al.*, Proc. Natl. Acad. Sci. USA, 81:6466 (1984); Tratschin *et al.*, Mol. Cell. Biol. 5:3251 (1985); McLaughlin *et al.*, J. Virol., 62:1963 (1988); y Lebkowski

*et al.*, 1988 Mol. Cell. Biol., 7:349 (1988). Samulski *et al.* (1989, J. Virol., 63:3822-3828); patente estadounidense n.º 5.173.414; documento WO 95/13365 y patente estadounidense correspondiente n.º 5.658.776 ; documentos WO 95/13392; WO 96/17947; PCT/US98/18600; WO 97/09441 (PCT/US96/14423); WO 97/08298 (PCT/US96/13872); WO 97/21825 (PCT/US96/20777); WO 97/06243 (PCT/FR96/01064); WO 99/11764; Perrin *et al.* (1995) Vaccine 13:1244-1250; Paul *et al.* (1993) Human Gene Therapy 4:609-615; Clark *et al.* (1996) Gene Therapy 3:1124-1132; patente estadounidense n.º 5.786.211; patente estadounidense n.º 5.871.982; y patente estadounidense n.º 6.258.595.

Pueden hacerse coincidir serotipos de vector de AAV con tipos de células diana. Por ejemplo, pueden transducirse los siguientes tipos de células a modo de ejemplo mediante los serotipos de AAV indicados, entre otros. Véase la tabla 3.

Tabla 3

Tipo de tejido/célula	Serotipo
Hígado	AAV8, AA3, AA5, AAV9
Músculo esquelético	AAV1, AAV7, AAV6, AAV8, AAV9
Sistema nervioso central	AAV5, AAV1, AAV4
RPE	AAV5, AAV4
Células fotorreceptoras	AAV5
Pulmón	AAV9
Corazón	AAV8
Páncreas	AAV8
Riñón	AAV2, AA8

Además de vectores virales adenoasociados, pueden usarse otros vectores virales. Tales vectores virales incluyen, pero no se limitan a, lentivirus, alfavirus, enterovirus, pestivirus, baculovirus, herpesvirus, virus de Epstein Barr, papovavirus, poxvirus, virus vaccinia y virus del herpes simple.

En algunos casos, cada uno de ARNm de Cas9, ARNGs que selecciona como diana uno o dos loci dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A, y ADN donante pueden formularse por separado en nanopartículas lipídicas, o todos ellos se formulan conjuntamente en una nanopartícula lipídica.

En algunos casos, puede formularse ARNm de Cas9 en una nanopartícula lipídica, mientras que ARNGs y ADN donante pueden suministrarse en un vector de AAV.

Hay opciones disponibles para suministrar la nucleasa Cas9 como plásmido de ADN, como ARNm o como proteína. El ARN guía puede expresarse a partir del mismo ADN o también puede suministrarse como ARN. El ARN puede modificarse químicamente para alterar o mejorar su semivida, o reducir la probabilidad o el grado de respuesta inmunitaria. La proteína endonucleasa puede complejarse con el ARNG antes de suministrarse. Los vectores virales permiten un suministro eficiente; pueden empaquetarse versiones divididas de Cas9 y ortólogos más pequeños de Cas9 en AAV, al igual que donantes para HDR. También existe una gama de métodos de suministro no virales que pueden suministrar cada uno de estos componentes, o pueden emplearse métodos no virales y virales en tándem. Por ejemplo, pueden usarse nanopartículas para suministrar la proteína y el ARN guía, mientras que puede usarse AAV para suministrar un ADN donante.

#### Células genéticamente modificadas

El término “célula genéticamente modificada” se refiere a una célula que comprende al menos una modificación genética introducida mediante edición genómica (por ejemplo, usando el sistema de CRISPR/Cas9/Cpf1). En algunos ejemplos *ex vivo* en el presente documento, la célula genéticamente modificada puede ser célula progenitora genéticamente modificada. En algunos ejemplos *in vivo* en el presente documento, la célula genéticamente modificada puede ser una célula progenitora hematopoyética genéticamente modificada. Una célula genéticamente modificada que comprende un ácido nucleico de direccionamiento al genoma exógeno y/o un ácido nucleico que codifica para un ácido nucleico de direccionamiento al genoma exógeno se contempla en el presente documento.

El término “población tratada con control” describe una población de células que se ha tratado con medios, inducción viral, secuencias de ácido nucleico, temperatura, confluencia, tamaño de matraz, pH, etc., idénticos con la excepción de la adición de los componentes de edición genómica. Puede usarse cualquier método conocido en la técnica para medir la modulación o inactivación de la secuencia de control transcripcional del gen de BCL11A o expresión o actividad de proteína, por ejemplo análisis de inmunotransferencia de tipo Western de la secuencia de control transcripcional de la proteína de gen de BCL11A o cuantificación de la secuencia de control transcripcional del ARNm de gen de BCL11A.

El término “célula aislada” se refiere a una célula que se ha retirado de un organismo en el que se encontraba originalmente, o un descendiente de una célula de este tipo. Opcionalmente, la célula puede cultivarse *in vitro*, por

ejemplo, en condiciones definidas o en presencia de otras células. Opcionalmente, la célula puede introducirse posteriormente en un segundo organismo o reintroducirse en el organismo a partir del cual se aisló (o la célula de la que descende).

El término “población aislada” con respecto a una población aislada de células se refiere a una población de células que se han retirado y separado a partir de una población mixta o heterogénea de células. En algunos casos, la población aislada puede ser una población sustancialmente pura de células, en comparación con la población heterogénea a partir de la cual se aislaron o enriquecieron las células. En algunos casos, la población aislada puede ser una población aislada de células progenitoras humanas, por ejemplo, una población sustancialmente pura de células progenitoras humanas, en comparación con una población heterogénea de células que comprenden células progenitoras humanas y células de las que se derivaron las células progenitoras humanas.

El término “sustancialmente potenciado”, con respecto a una población celular particular, se refiere a una población de células en la que la aparición de un tipo particular de célula está aumentada con respecto a niveles previamente existentes o de referencia, en al menos 2 veces, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 20, al menos 50, al menos 100, al menos 400, al menos 1000, al menos 5000, al menos 20000, al menos 100000 o más veces dependiendo, por ejemplo, de los niveles deseados de tales células para mejorar la hemoglobinopatía.

El término “sustancialmente enriquecido” con respecto a una población celular particular, se refiere a una población de células que es de al menos aproximadamente el 10 %, aproximadamente el 20 %, aproximadamente el 30 %, aproximadamente el 40 %, aproximadamente el 50 %, aproximadamente el 60 %, aproximadamente el 70 % o más con respecto a las células que constituyen una población celular total.

El término “sustancialmente puro” con respecto a una población celular particular, se refiere a una población de células que es pura en al menos aproximadamente el 75 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 % o al menos aproximadamente el 95 %, con respecto a las células que constituyen una población celular total. Es decir, los términos “sustancialmente puro” o “esencialmente purificado”, con respecto a una población de células progenitoras, se refieren a una población de células que contienen menos de aproximadamente el 20 %, aproximadamente el 15 %, aproximadamente el 10 %, aproximadamente el 9 %, aproximadamente el 8 %, aproximadamente el 7 %, aproximadamente el 6 %, aproximadamente el 5 %, aproximadamente el 4 %, aproximadamente el 3 %, aproximadamente el 2 %, aproximadamente el 1 % o menos del 1 %, de células que no son células progenitoras tal como se define mediante los términos en el presente documento.

Diferenciación de iPSC sometidas a edición genómica para dar células progenitoras hematopoyéticas

Otra etapa de los métodos *ex vivo* de la presente divulgación puede comprender diferenciar las iPSC sometidas a edición genómica para dar células progenitoras hematopoyéticas. La etapa de diferenciación puede realizarse según cualquier método conocido en la técnica.

Diferenciación de células madre mesenquimatosas sometidas a edición genómica para dar células progenitoras hematopoyéticas

Otra etapa de los métodos *ex vivo* de la presente divulgación puede comprender diferenciar las células madre mesenquimatosas sometidas a edición genómica para dar células progenitoras hematopoyéticas. La etapa de diferenciación puede realizarse según cualquier método conocido en la técnica.

Implantación de células en pacientes

Otra etapa de los métodos *ex vivo* de la presente divulgación puede comprender implantar las células en pacientes. Esta etapa de implantación puede lograrse usando cualquier método de implantación conocido en la técnica. Por ejemplo, las células genéticamente modificadas pueden inyectarse directamente en la sangre del paciente o administrarse de otro modo al paciente. Las células genéticamente modificadas pueden purificarse *ex vivo* usando un marcador seleccionado.

Portadores farmacéuticamente aceptables

Los métodos *ex vivo* de administración de células progenitoras a un sujeto contemplados en el presente documento pueden implicar el uso de composiciones terapéuticas que comprenden células progenitoras.

Las composiciones terapéuticas pueden contener un portador fisiológicamente tolerable junto con la composición de células, y opcionalmente al menos un agente bioactivo adicional tal como se describe en el presente documento, disuelto o dispersado en la misma como principio activo. En algunos casos, la composición terapéutica no es sustancialmente inmunogénica cuando se administra a un paciente mamífero o humano con fines terapéuticos, a menos que se desee.

En general, las células progenitoras descritas en el presente documento pueden administrarse como una suspensión con un portador farmacéuticamente aceptable. Un experto en la técnica reconocerá que un portador farmacéuticamente aceptable que va a usarse en una composición de células no incluirá tampones, compuestos, agentes de criopreservación, conservantes u otros agentes en cantidades que interfieran sustancialmente con la viabilidad de las células que van a suministrarse al sujeto. Una formulación que comprende células puede incluir, por ejemplo, tampones osmóticos que permiten mantener la integridad de la membrana celular, y, opcionalmente, nutrientes para mantener la viabilidad celular o potenciar el injerto tras la administración. Tales formulaciones y suspensiones las conocen los expertos en la técnica y/o pueden adaptarse para su uso con las células progenitoras, tal como se describe en el presente documento, usando experimentación rutinaria.

Una composición de células también puede emulsionarse o presentarse como composición de liposoma, siempre que el procedimiento de emulsificación no afecte de manera adversa a la viabilidad celular. Las células y cualquier otro principio activo pueden mezclarse con excipientes que son farmacéuticamente aceptables y compatibles con el principio activo, y en cantidades adecuadas para su uso en los métodos terapéuticos descritos en el presente documento.

Los agentes adicionales incluidos en una composición de células pueden incluir sales farmacéuticamente aceptables de los componentes en la misma. Las sales farmacéuticamente aceptables incluyen las sales de adición de ácido (formadas con los grupos amino libres del polipéptido) que se forman con ácidos inorgánicos, tales como, por ejemplo, ácidos clorhídrico o fosfórico, o ácidos orgánicos tales como acético, tartárico, mandélico y similares. Sales formadas con los grupos carboxilo libres también pueden derivarse a partir de bases inorgánicas, tales como, por ejemplo, hidróxidos de sodio, potasio, amonio, calcio o férricos, y bases orgánicas tales como isopropilamina, trimetilamina, 2-etilaminoetanol, histidina, procaína y similares.

En la técnica se conocen bien portadores fisiológicamente tolerables. Los portadores líquidos a modo de ejemplo son disoluciones acuosas estériles que no contienen ningún material además de los principios activos y agua, o contienen un tampón tal como fosfato de sodio a un valor de pH fisiológico, solución salina fisiológica o ambos, tal como solución salina tamponada con fosfato. Todavía adicionalmente, los portadores acuosos pueden contener más de una sal de tampón, así como sales tales como cloruros de sodio y potasio, dextrosa, polietilenglicol y otros solutos. Las composiciones líquidas también pueden contener fases líquidas además de, y excluyendo, agua. A modo de ejemplo de tales fases líquidas adicionales se presentan glicerina, aceites vegetales tales como aceite de semilla de algodón, y emulsiones de agua-aceite. La cantidad de un compuesto activo usado en las composiciones de células que es eficaz en el tratamiento de un trastorno o estado particular puede depender de la naturaleza del trastorno o estado, y puede determinarse mediante técnicas clínicas convencionales.

#### Administración y eficacia

Los términos “administrar”, “introducir” y “trasplantar” se usan de manera intercambiable en el contexto de la colocación de células, por ejemplo, células progenitoras, en un sujeto, mediante un método o vía que da como resultado una localización al menos parcial de las células introducidas en un sitio deseado, tal como un sitio de lesión o reparación, de tal manera que se produce(n) un(os) efecto(s) deseado(s). Las células, por ejemplo, células progenitoras, o su progenie diferenciada, pueden administrarse mediante una vía apropiada que da como resultado el suministro a una ubicación deseada en el sujeto en la que al menos una porción de las células implantadas o componentes de las células permanecen viables. El periodo de viabilidad de las células tras la administración a un sujeto puede ser de tan solo unas pocas horas, por ejemplo, veinticuatro horas, hasta unos pocos días, hasta varios años o incluso toda la vida del paciente, es decir, injerto a largo plazo. Por ejemplo, en algunos aspectos descritos en el presente documento, se administra una cantidad eficaz de células progenitoras miogénicas mediante una vía de administración sistémica, tal como una vía intraperitoneal o intravenosa.

Los términos “individuo”, “sujeto”, “huésped” y “paciente” se usan de manera intercambiable en el presente documento y se refieren a cualquier sujeto para el que se desea un diagnóstico, tratamiento o terapia. En algunos aspectos, el sujeto es un mamífero. En algunos aspectos, el sujeto es un ser humano.

Cuando se proporcionan de manera profiláctica, pueden administrarse células progenitoras descritas en el presente documento a un sujeto antes de cualquier síntoma de una hemoglobinopatía, por ejemplo, antes del desarrollo de fatiga, dificultad para respirar, ictericia, crecimiento lento, pubertad tardía, dolor de articulaciones, huesos y tórax, esplenomegalia y hepatomegalia. Por consiguiente, la administración profiláctica de una población de células progenitoras hematopoyéticas sirve para prevenir una hemoglobinopatía, tal como B-talasemia o enfermedad de células falciformes.

Cuando se proporcionan de manera terapéutica, se proporcionan células progenitoras hematopoyéticas en (o después de) la aparición de un síntoma o indicación de hemoglobinopatía, por ejemplo, tras la aparición de la enfermedad.

La población de células progenitoras hematopoyéticas que está administrándose según los métodos descritos en el presente documento puede comprender células progenitoras hematopoyéticas alogénicas obtenidas a partir de uno o más donantes. “Alogénico” se refiere a una célula progenitora hematopoyética o muestras biológicas que comprenden

células progenitoras hematopoyéticas obtenidas a partir de uno o más donantes diferentes de la misma especie, en las que los genes en uno o más loci no son idénticos. Por ejemplo, una población de células progenitoras hematopoyéticas que está administrándose a un sujeto puede derivarse de uno o más sujetos donantes no relacionados, o de uno o más hermanos no idénticos. En algunos casos, pueden usarse poblaciones de células progenitoras hematopoyéticas singénicas, tales como las obtenidas a partir de animales genéticamente idénticos o a partir de gemelos idénticos. Las células progenitoras hematopoyéticas pueden ser células autólogas; es decir, las células progenitoras hematopoyéticas se obtienen o se aíslan a partir de un sujeto y se administran al mismo sujeto, es decir, el donante y el receptor son el mismo.

El término "cantidad eficaz" se refiere a la cantidad de una población de células progenitoras o a progenie necesaria para prevenir o aliviar al menos uno o más signos o síntomas de hemoglobinopatía, y se refiere a una cantidad suficiente de una composición para proporcionar el efecto deseado, por ejemplo, para tratar a un sujeto que tiene hemoglobinopatía. Por tanto, el término "cantidad terapéuticamente eficaz" se refiere a una cantidad de células progenitoras o una composición que comprende células progenitoras que es suficiente para fomentar un efecto particular cuando se administra a un sujeto típico, tal como uno que tiene o corre el riesgo de hemoglobinopatía. Una cantidad eficaz también incluirá una cantidad suficiente para prevenir o retardar el desarrollo de un síntoma de la enfermedad, alterar el transcurso de un síntoma de la enfermedad (por ejemplo, pero sin limitarse a, ralentizar la progresión de un síntoma de la enfermedad), o revertir un síntoma de la enfermedad. Se entiende que, para cualquier caso dado, una "cantidad eficaz" apropiada puede determinarse por un experto habitual en la técnica usando experimentación rutinaria.

Para su uso en los diversos aspectos descritos en el presente documento, una cantidad eficaz de células progenitoras comprende al menos  $10^2$  células progenitoras, al menos  $5 \times 10^2$  células progenitoras, al menos  $10^3$  células progenitoras, al menos  $5 \times 10^3$  células progenitoras, al menos  $10^4$  células progenitoras, al menos  $5 \times 10^4$  células progenitoras, al menos  $10^5$  células progenitoras, al menos  $2 \times 10^5$  células progenitoras, al menos  $3 \times 10^5$  células progenitoras, al menos  $4 \times 10^5$  células progenitoras, al menos  $5 \times 10^5$  células progenitoras, al menos  $6 \times 10^5$  células progenitoras, al menos  $7 \times 10^5$  células progenitoras, al menos  $8 \times 10^5$  células progenitoras, al menos  $9 \times 10^5$  células progenitoras, al menos  $1 \times 10^6$  células progenitoras, al menos  $2 \times 10^6$  células progenitoras, al menos  $3 \times 10^6$  células progenitoras, al menos  $4 \times 10^6$  células progenitoras, al menos  $5 \times 10^6$  células progenitoras, al menos  $6 \times 10^6$  células progenitoras, al menos  $7 \times 10^6$  células progenitoras, al menos  $8 \times 10^6$  células progenitoras, al menos  $9 \times 10^6$  células progenitoras o múltiplos de los mismos. Las células progenitoras pueden derivarse de uno o más donantes, o pueden obtenerse de una fuente autóloga. En algunos ejemplos descritos en el presente documento, las células progenitoras pueden expandirse en cultivo antes de la administración a un sujeto que lo necesita.

"Administrado" se refiere al suministro de una composición de células progenitoras en un sujeto mediante un método o ruta que da como resultado una localización al menos parcial de la composición de células en un sitio deseado. Una composición de células puede administrarse mediante cualquier ruta apropiada que dé como resultado un tratamiento eficaz en el sujeto, es decir, la administración da como resultado el suministro a una ubicación deseada en el sujeto en la que se suministra al menos una porción de la composición, es decir, se suministran al menos  $1 \times 10^4$  células al sitio deseado durante un periodo de tiempo. Los modos de administración incluyen inyección, infusión, instilación o ingestión. "Inyección" incluye, sin limitación, inyección e infusión intravenosa, intramuscular, intraarterial, intratecal, intraventricular, intracapsular, intraorbital, intracardiaca, intradérmica, intraperitoneal, transtraqueal, subcutánea, subcuticular, intraarticular, subcapsular, subaracnoidea, intrarraquídea, intracfalorraquídea e intraesternal. En algunos ejemplos, la ruta es intravenosa. Para el suministro de células, puede realizarse una administración mediante inyección o infusión.

Las células pueden administrarse de manera sistémica. Las expresiones "administración sistémica", "administrado de manera sistémica", "administración periférica" y "administrado de manera periférica" se refieren a la administración de una población de células progenitoras distinta de directamente en un sitio diana, tejido u órgano diana, de tal manera que entra, en su lugar, en el sistema circulatorio del sujeto y, por tanto, se somete a metabolismo y otros procesos similares.

La eficacia de un tratamiento que comprende una composición para el tratamiento de hemoglobinopatías puede determinarse por el médico experto. Sin embargo, se considera que un tratamiento es un "tratamiento eficaz", si uno cualquiera o la totalidad de los signos o síntomas de, tan solo como un ejemplo, niveles de BCL11A funcional y HbF funcional, se alteran de una manera beneficiosa (por ejemplo, se reducen en al menos el 10 % para BCL11A y/o se aumentan en al menos el 10 % para HbF), o se mejoran o alivian otros síntomas o marcadores de enfermedad clínicamente aceptados. La eficacia también puede medirse al no empeorar un individuo tal como se evalúa mediante hospitalización o necesidad de intervenciones médicas (por ejemplo, se detiene o al menos se ralentiza la progresión de la enfermedad). Métodos de medición de estos indicadores los conocen los expertos en la técnica y/o se describen en el presente documento. El tratamiento incluye cualquier tratamiento de una enfermedad en un individuo o un animal (algunos ejemplos no limitativos incluyen un ser humano o un mamífero) e incluye: (1) inhibir la enfermedad, por ejemplo, detener o ralentizar la progresión de síntomas; o (2) aliviar la enfermedad, por ejemplo, provocar la regresión de síntomas; y (3) prevenir o reducir la probabilidad de desarrollo de síntomas.

El tratamiento según la presente divulgación puede mejorar uno o más síntomas asociados con hemoglobinopatías

reduciendo la cantidad de BCL11A funcional y/o aumentando la cantidad de HbF funcional en el individuo. Los signos tempranos normalmente asociados con hemoglobinopatías incluyen, por ejemplo, fatiga, dificultad para respirar, ictericia, crecimiento lento, pubertad tardía, dolor de articulaciones, huesos y tórax, esplenomegalia y hepatomegalia.

## 5 Kits

La presente divulgación da a conocer kits para llevar a cabo los métodos descritos en el presente documento. Un kit puede incluir uno o más de un ácido nucleico de direccionamiento al genoma, un polinucleótido que codifica para un ácido nucleico de direccionamiento al genoma, un polipéptido dirigido al sitio, un polinucleótido que codifica para un polipéptido dirigido al sitio y/o cualquier molécula de ácido nucleico o proteica necesaria para llevar a cabo los aspectos de los métodos descritos en el presente documento, o cualquier combinación de los mismos.

Un kit puede comprender: (1) un vector que comprende una secuencia de nucleótidos que codifica para un ácido nucleico de direccionamiento al genoma, (2) el polipéptido dirigido al sitio o un vector que comprende una secuencia de nucleótidos que codifica para el polipéptido dirigido al sitio, y (3) un reactivo para reconstitución y/o dilución del/de los vector(s) y o polipéptido.

Un kit puede comprender: (1) un vector que comprende (i) una secuencia de nucleótidos que codifica para un ácido nucleico de direccionamiento al genoma, y (ii) una secuencia de nucleótidos que codifica para el polipéptido dirigido al sitio; y (2) un reactivo para reconstitución y/o dilución del vector.

En cualquiera de los kits anteriores, el kit puede comprender un ácido nucleico de direccionamiento al genoma guía de molécula simple. En cualquiera de los kits anteriores, el kit puede comprender un ácido nucleico de direccionamiento al genoma de molécula doble. En cualquiera de los kits anteriores, el kit puede comprender dos o más guías de molécula doble o guías de molécula simple. Los kits pueden comprender un vector que codifica para el ácido nucleico de direccionamiento a ácido nucleico.

En cualquiera de los kits anteriores, el kit puede comprender además un polinucleótido que va a insertarse para realizar la modificación genética deseada.

Los componentes de un kit pueden estar en recipientes independientes o combinados en un único recipiente.

Cualquier kit descrito anteriormente puede comprender además uno o más reactivos adicionales, en los que tales reactivos adicionales se seleccionan de un tampón, un tampón para introducir un polipéptido o polinucleótido en una célula, un tampón de lavado, un reactivo de control, un vector de control, un polinucleótido de ARN de control, un reactivo para la producción *in vitro* del polipéptido a partir de ADN, adaptadores para secuenciación y similares. Un tampón puede ser un tampón de estabilización, un tampón de reconstitución, un tampón de dilución o similar. Un kit también puede comprender uno o más componentes que pueden usarse para facilitar o potenciar la unión específica o la escisión de ADN mediante la endonucleasa, o mejorar la especificidad de direccionamiento.

Además de los componentes anteriormente mencionados, un kit puede comprender además instrucciones para usar los componentes del kit para poner en práctica los métodos. Las instrucciones para poner en práctica los métodos pueden grabarse en un medio de grabación adecuado. Por ejemplo, las instrucciones pueden imprimirse en un sustrato, tal como papel o plástico, etc. Las instrucciones pueden estar presentes en los kits como prospecto, en el etiquetado del recipiente del kit o componentes del mismo (es decir, asociadas con el acondicionamiento o acondicionamiento secundario), etc. Las instrucciones pueden estar presentes como archivo de datos de almacenamiento electrónico presente en un medio de almacenamiento legible por ordenador adecuado, por ejemplo CD-ROM, diskette, memoria USB, etc. En algunos casos, las instrucciones reales no están presentes en el kit, sino que pueden proporcionarse medios para obtener las instrucciones a partir de una fuente remota (por ejemplo, a través de Internet). Un ejemplo de este caso es un kit que comprende una dirección web en la que pueden visualizarse las instrucciones y/o de la cual pueden descargarse las instrucciones. Como con las instrucciones, estos medios para obtener las instrucciones pueden estar grabados en un sustrato adecuado.

## Formulación de ARN guía

Los ARN guía de la presente divulgación pueden formularse con excipientes farmacéuticamente aceptables tales como portadores, disolventes, estabilizantes, adyuvantes, diluyentes, etc., dependiendo del modo de administración y la forma de dosificación particulares. Pueden formularse composiciones de ARN guía para lograr un pH fisiológicamente compatible, y oscilar desde un pH de aproximadamente 3 hasta un pH de aproximadamente 11, de aproximadamente pH 3 a aproximadamente pH 7, dependiendo de la formulación y ruta de administración. En algunos casos, el pH puede ajustarse hasta un intervalo de desde aproximadamente pH 5,0 hasta aproximadamente pH 8. En algunos casos, las composiciones pueden comprender una cantidad terapéuticamente eficaz de al menos un compuesto tal como se describe en el presente documento, junto con uno o más excipientes farmacéuticamente aceptables. Opcionalmente, las composiciones pueden comprender una combinación de los compuestos descritos en el presente documento, o pueden incluir un segundo principio activo útil en el tratamiento o la prevención de crecimiento bacteriano (por ejemplo, y sin limitación, agentes antibacterianos o antimicrobianos) o pueden incluir una combinación de

reactivos de la presente divulgación.

Los excipientes adecuados incluyen, por ejemplo, moléculas portadoras que incluyen macromoléculas grandes que se metabolizan lentamente tales como proteínas, polisacáridos, poli(ácidos lácticos), poli(ácidos glicólicos), aminoácidos poliméricos, copolímeros de aminoácido y partículas de virus inactivas. Otros excipientes a modo de ejemplo pueden incluir antioxidantes (por ejemplo, y sin limitación, ácido ascórbico), agentes quelantes (por ejemplo, y sin limitación, EDTA), hidratos de carbono (por ejemplo, y sin limitación, dextrina, hidroxialquilcelulosa e hidroxialquilmethylcelulosa), ácido esteárico, líquidos (por ejemplo, y sin limitación, aceites, agua, solución salina, glicerol y etanol), agentes humectantes o emulsionantes, sustancias de tamponamiento del pH y similares.

Otros posibles enfoques terapéuticos

La edición génica puede llevarse a cabo usando nucleasas modificadas por ingeniería para seleccionar como diana secuencias específicas. Hasta ahora, hay cuatro tipos principales de nucleasas: meganucleasas y sus derivados, nucleasas de dedos de cinc (ZFN), nucleasas efectoras de tipo activador de la transcripción (TALEN) sistemas de nucleasas CRISPR-Cas9. Las plataformas de nucleasas varían en cuanto a la dificultad de diseño, densidad de direccionamiento y modo de acción, particularmente dado que la especificidad de ZFN y TALEN es mediante interacciones proteína-ADN, mientras que las interacciones ARN-ADN guían principalmente a Cas9. La escisión de Cas9 también requiere un motivo adyacente, el PAM, que difiere entre diferentes sistemas de CRISPR. Cas9 de *Streptococcus pyogenes* escinde usando un PAM de NGG, CRISPR de *Neisseria meningitidis* puede escindir en sitios con PAM que incluyen NNNNGATT, NNNNGTTT y NNNNGCTT. Varios otros ortólogos de Cas9 seleccionan como diana un protoespaciador adyacente a PAM alternativos.

Pueden usarse endonucleasas CRISPR, tales como Cas9, en los métodos de la presente divulgación. Sin embargo, las enseñanzas descritas en el presente documento, tales como sitios diana terapéuticos, pueden aplicarse a otras formas de endonucleasas, tales como ZFN, TALEN, HE o MegaTAL, o usando combinaciones de nucleasas. Sin embargo, con el fin de aplicar las enseñanzas de la presente divulgación a tales endonucleasas, se necesitará, entre otras cosas, modificar por ingeniería proteínas dirigidas a los sitios diana específicos.

Pueden fusionarse dominios de unión adicionales a la proteína Cas9 para aumentar la especificidad. Los sitios diana de estos constructos se mapearán en el sitio especificado de ARNg identificado, pero requerirán motivos de unión adicionales, tal como un dominio de dedos de cinc. En el caso de Mega-TAL, puede fusionarse una meganucleasa a un dominio de unión a ADN de TALE. El dominio de meganucleasa puede aumentar la especificidad y proporcionar la escisión. De manera similar, puede fusionarse Cas9 inactivada o muerta (dCas9) a un dominio de escisión y requerir el sitio diana de ARNg/Cas9 y sitio de unión adyacente para el dominio de unión a ADN fusionado. Es probable que esto requiera alguna modificación por ingeniería de proteína de la dCas9, además de la inactivación catalítica, para reducir la unión sin el sitio de unión adicional.

Nucleasas de dedos de cinc

Las nucleasas de dedos de cinc (ZFN) son proteínas modulares compuestas por un dominio de unión a ADN de dedos de cinc modificado por ingeniería unido al dominio catalítico de la endonucleasa de tipo II FokI. Dado que FokI solo funciona como dímero, debe modificarse por ingeniería un par de ZFN para unirse a secuencias de "semisito" diana relacionada en cadenas de ADN opuestas y con separación precisa entre las mismas para permitir que se forme el dímero de FokI catalíticamente activo. Tras la dimerización del dominio de FokI, que no tiene ninguna especificidad de secuencia en sí mismo, se genera una rotura de cadena doble de ADN entre los semisitos de ZFN como etapa de iniciación en la edición genómica.

El dominio de unión a ADN de cada ZFN está normalmente compuesto por 3-6 dedos de cinc con la arquitectura abundante de Cys2-His2, reconociendo principalmente cada dedo un triplete de nucleótidos en una cadena de la secuencia de ADN diana, aunque la interacción de cadena cruzada con un cuarto nucleótido también puede ser importante. La alteración de los aminoácidos de un dedo en posiciones que forman contactos clave con el ADN altera la especificidad de secuencia de un dedo dado. Por tanto, una proteína de dedos de cinc de cuatro dedos reconocerá de manera selectiva una secuencia diana de 12 pb, en la que la secuencia diana esa estructura compuesta de las preferencias de triplete aportadas por cada dedo, aunque la preferencia de triplete puede verse influida hasta grados variables por dedos contiguos. Un aspecto importante de las ZFN es que pueden redirigirse fácilmente a casi cualquier dirección genómica simplemente modificando dedos individuales, aunque se requiere una experiencia considerable para hacer esto correctamente. En la mayoría de las aplicaciones de ZFN, se usan proteínas de 4-6 dedos, que reconocen 12-18 pb respectivamente. Por tanto, normalmente un par de ZFN reconocerán una secuencia diana combinada de 24-36 pb, sin incluir el espaciador de 5-7 pb típico entre semisitos. Los sitios de unión pueden estar adicionalmente separados con espaciadores más grandes, incluyendo 15-17 pb. Es probable que una secuencia diana de esta longitud sea única en el genoma humano, suponiendo que se excluyen secuencias repetitivas u homólogos génicos durante el procedimiento de diseño. No obstante, las interacciones proteína de ZFN-ADN no son absolutas en cuanto a su especificidad, de modo que se producen acontecimientos de unión y escisión inespecíficos, o bien como heterodímero entre las dos ZFN, o bien como homodímero de una u otra de las ZFN. Esta última posibilidad se ha eliminado eficazmente mediante modificación por ingeniería de la superficie de contacto de dimerización del

dominio de FokI para crear variantes “positiva” y “negativa”, también conocidas como variantes de heterodímero obligado, que solo pueden dimerizarse entre sí, y no consigo mismas. Forzar los heterodímeros obligados previene la formación del homodímero. Esto ha potenciado en gran medida la especificidad de las ZFN, así como cualquier otra nucleasa que adopta estas variantes de FokI.

Se ha descrito una variedad de sistemas basados en ZFN en la técnica, se notifican modificaciones de las mismas de manera regular, y numerosas referencias describen reglas y parámetros que se usan para guiar el diseño de las ZFN; véanse, por ejemplo, Segal *et al.*, Proc Natl Acad Sci USA 96(6):2758-63 (1999); Dreier B *et al.*, J Mol Biol. 303(4):489-502 (2000); Liu Q *et al.*, J Biol Chem. 277(6):3850-6 (2002); Dreier *et al.*, J Biol Chem 280(42):35588-97 (2005); y Dreier *et al.*, J Biol Chem. 276(31):29466-78 (2001).

#### Nucleasas efectoras de tipo activador de la transcripción (TALEN)

Las TALEN representan otro formato de nucleasas modulares mediante las cuales, como con las ZFN, se une un dominio de unión a ADN modificado por ingeniería al dominio de nucleasa FokI, y un par de TALEN funcionan en tándem para lograr la escisión de ADN dirigida. La principal diferencia con respecto a las ZFN es la naturaleza del dominio de unión a ADN y las propiedades de reconocimiento de secuencia de ADN diana asociadas. El dominio de unión a ADN de TALEN se deriva de proteínas TALE, que originalmente se describieron en el patógeno bacteriano de plantas *Xanthomonas sp.* Las TALE están compuestas por matrices en tándem de repeticiones de 33-35 aminoácido, reconociendo cada repetición un único par de bases en la secuencia de ADN diana que normalmente tiene hasta 20 pb de longitud, dando una longitud de secuencia diana total de hasta 40 pb. La especificidad de nucleótidos de cada repetición está determinada por el residuo variable de repetición (RVD), que incluye tan solo dos aminoácidos en las posiciones 12 y 13. Las bases guanina, adenina, citosina y timina se reconocen de manera predominante por los cuatro RVD: Asn-Asn, Asn-Ile, His-Asp y Asn-Gly, respectivamente. Esto constituye un código de reconocimiento mucho más sencillo que para los dedos de cinc y, por tanto, representa una ventaja con respecto a estos últimos para el diseño de nucleasas. No obstante, como con las ZFN, las interacciones proteína-ADN de las TALEN no son absolutas en cuanto a su especificidad, y las TALEN también se han beneficiado del uso de variantes de heterodímeros obligados del dominio de FokI para reducir la actividad inespecífica.

Se han creado variantes adicionales del dominio de FokI que tienen su función catalítica desactivada. Si una mitad de o bien un par de TALEN o bien uno de ZFN contiene un dominio de FokI inactivo, entonces solo se producirá la escisión de ADN de cadena sencilla (mellado) en el sitio diana, en vez de una DSB. El resultado es comparable al uso de mutantes de “enzima melladora” de CRISPR/Cas9/Cpf1 en las que uno de los dominios de escisión de Cas9 se ha desactivado. Pueden usarse mellas de ADN para impulsar la edición genómica mediante HDR, pero con una eficiencia menor que con una DSB. El beneficio principal es que las mellas inespecíficas se reparan de manera rápida y precisa, a diferencia de la DSB, que es propensa a reparación errónea mediada por NHEJ.

Se ha descrito en la técnica una variedad de sistemas basados en TALEN, y se notifican modificaciones de los mismos de manera regular; véanse, por ejemplo, Boch, Science 326(5959):1509-12 (2009); Mak *et al.*, Science 335(6069):716-9 (2012); y Moscou *et al.*, Science 326(5959):1501 (2009). El uso de TALEN basadas en la plataforma “Golden Gate”, o esquema de clonación, se ha descrito por múltiples grupos; véanse, por ejemplo, Cermak *et al.*, Nucleic Acids Res. 39(12):e82 (2011); Li *et al.*, Nucleic Acids Res. 39(14):6315-25(2011); Weber *et al.*, PLoS One. 6(2):e16765 (2011); Wang *et al.*, J Genet Genomics 41(6):339-47, publicación electrónica, 17 de mayo de 2014 (2014); y Cermak T *et al.*, Methods MolBiol. 1239:133-59 (2015).

#### Endonucleasas de asentamiento

Las endonucleasas de asentamiento (HE) son endonucleasas específicas de secuencia que tienen largas secuencias de reconocimiento (14-44 pares de bases) y escinden el ADN con alta especificidad (con frecuencia en sitios únicos en el genoma). Se conocen al menos seis familias de HE clasificadas mediante su estructura, incluyendo LAGLIDADG (SEQ ID NO. 71.949), GIY-YIG, caja His-Cis, H-N-H, PD-(D/E)xK, y de tipo Vsr que se derivan de una amplia gama de huéspedes, incluyendo *Eukarya*, protistas, bacterias, arqueas, cianobacterias y fagos. Como con las ZFN y las TALEN, las HE pueden usarse para crear una DSB en un locus diana como etapa inicial en la edición genómica. Además, algunas HE naturales y modificadas por ingeniería solo cortan una única cadena de ADN, funcionando de ese modo como enzimas melladoras específicas de sitio. La gran secuencia diana de las HE y la especificidad que ofrecen han hecho que sean candidatos atractivos para crear DSB específicas de sitio.

Se ha descrito en la técnica una variedad de sistemas basados en HE, y se notifican modificaciones de los mismos de manera regular; véanse, por ejemplo, las revisiones de Steentoft *et al.*, Glycobiology 24(8):663-80 (2014); Belfort y Bonocora, Methods Mol Biol. 1123:1-26 (2014); Hafez y Hausner, Genome 55(8):553-69 (2012); y referencias citadas en los mismos.

#### MegaTAL / Tev-mTALEN / MegaTev

Como ejemplos adicionales de nucleasas híbridas, la plataforma MegaTAL y la plataforma Tev-mTALEN usan una fusión de dominios de unión a ADN de TALE y HE catalíticamente activas, aprovechando tanto la unión a ADN



ajustable como la especificidad de la TALE, así como la especificidad de secuencia de escisión de la HE; véanse, por ejemplo, Boissel *et al.*, NAR 42: 2591-2601 (2014); Kleinstiver *et al.*, G3 4:1155-65 (2014); y Boissel y Scharenberg, Methods Mol. Biol. 1239: 171-96 (2015).

En una variación adicional, la arquitectura de MegaTev es la fusión de una meganucleasa (Mega) con el dominio de nucleasa derivado de la endonucleasa de asentamiento I-Tevl (Tev) de GIY-YIG. Los dos sitios activos están posicionados separados ~30 pb en un sustrato de ADN y generan dos DSB con extremos cohesivos no compatibles; véase, por ejemplo, Wolfs *et al.*, NAR 42, 8816-29 (2014). Se prevé que evolucionarán otras combinaciones de enfoques basados en nucleasas existentes y serán útiles para lograr las modificaciones de genoma dirigidas descritas en el presente documento.

#### dCas9-FokI o dCpf1-FokI y otras nucleasas

Combinar las propiedades estructurales y funcionales de las plataformas de nucleasa descritas anteriormente ofrece un enfoque adicional para la edición genómica que puede superar posiblemente algunas de las deficiencias inherentes. Como un ejemplo, el sistema de edición genómica de CRISPR usa normalmente una única endonucleasa Cas9 para crear una DSB. La especificidad de direccionamiento está impulsada por una secuencia de 20 o 24 nucleótidos en el ARN guía que se somete a emparejamiento de bases de Watson-Crick con el ADN diana (más 2 bases adicionales en la secuencia de PAM de NAG o NGG adyacente en el caso de Cas9 de *S. pyogenes*). Una secuencia de este tipo es lo suficientemente larga como para ser única en el genoma humano, sin embargo, la especificidad de la interacción ARN/ADN no es absoluta, tolerándose algunas veces una promiscuidad significativa, particularmente en la mitad en 5' de la secuencia diana, reduciendo eficazmente el número de bases que impulsan la especificidad. Una solución a esto ha sido desactivar completamente la función catalítica de Cas9 o Cpf1 (conservando únicamente la función de unión a ADN guiada por ARN) y fusionando, en vez de eso, un dominio de FokI a la Cas9 desactivada; véanse, por ejemplo, Tsai *et al.*, Nature Biotech 32: 569-76 (2014); y Guilinger *et al.*, Nature Biotech. 32: 577-82 (2014). Dado que FokI debe dimerizarse para volverse catalíticamente activo, se requieren dos ARN guía para anclar dos fusiones de FokI en estrecha proximidad para formar el dímero y escindir ADN. Esto duplica esencialmente el número de bases en los sitios diana combinados, aumentando de ese modo la rigurosidad de direccionamiento mediante sistemas basados en CRISPR.

Como ejemplo adicional, la fusión del dominio de unión a ADN de TALE a una HE catalíticamente activa, tal como I-Tevl, aprovecha tanto la unión a ADN ajustable y especificidad de la TALE, así como la especificidad de secuencia de escisión de I-Tevl, con la expectativa de que pueda reducirse adicionalmente la escisión inespecífica.

#### Detección de mutaciones específicas e inespecíficas mediante secuenciación

Para secuenciar sitios específicos sitios inespecíficos supuestos, se identificaron los cebadores de amplificación apropiados y se establecieron reacciones con estos cebadores usando el ADN genómico recogido usando disolución de extracción de ADN de QuickExtract (Epicentre) a partir de células tratadas tres días tras la transfección. Los cebadores de amplificación contienen la porción específica de gen flanqueada por adaptadores. El extremo 5' del cebador directo incluye un sitio de unión a cebador directo modificado (lectura 1). El extremo 5' del cebador inverso contiene un sitio de unión a cebador de código de barras e inverso modificado combinado (lectura 2), en orientación opuesta. Se validaron las reacciones de PCR individuales mediante separación en geles de agarosa, después se purificaron y volvieron a amplificarse. Los cebadores directos de segunda ronda contienen la secuencia de Illumina P5, seguida por una porción del sitio de unión a cebador directo modificado (lectura 1). Los cebadores inversos de segunda ronda contienen la secuencia de Illumina P7 (en el extremo 5'), seguida por el código de barras de 6 bases y el sitio de unión a cebador de código de barras e inverso modificado combinado (lectura 2). También se comprobaron las amplificaciones de segunda ronda en geles de agarosa, después se purificaron y se cuantificaron usando un espectrofotómetro NanoDrop. Se combinaron los productos de amplificación para hacer coincidir la concentración y después se presentaron a Emory Integrated Genomic Core para la preparación de la biblioteca y secuenciación en una máquina Illumina Miseq.

Se clasificaron las lecturas de secuenciación mediante código de barras y después se alinearon con las secuencias de referencia suministradas mediante bioinformática para cada producto. Las tasas de inserción y delección en las lecturas de secuenciación alineadas se detectaron en la región de los sitios de corte supuestos usando software anteriormente descrito; véase, por ejemplo, Lin *et al.*, Nucleic Acids Res., 42: 7473-7485 (2014). Después se compararon los niveles de inserciones y delecciones detectados en este intervalo con el nivel observado en la misma ubicación en ADN genómico aislado a partir de células transfectadas de manera simulada para minimizar los efectos de artefactos de secuenciación.

#### Ensayos de detección de mutaciones

Se midieron las actividades de escisión específica e inespecífica de combinaciones de Cas9 y ARN guía usando las tasas de mutación resultantes de la reparación imperfecta de roturas de cadena doble mediante NHEJ.

Se amplificaron loci específicos usando ADN polimerasa Taq de alta fidelidad AccuPrime (Life Technologies, Carlsbad,

CA) siguiendo las instrucciones del fabricante durante 40 ciclos (94°C, 30 s; 52-60°C, 30 s; 68°C, 60 s) en reacciones de 50 µl que contenían 1 µl del lisado celular, y 1 µl de cada cebador de amplificación 10 µM. Se realizaron ensayos de detección de mutaciones T7EI, según el protocolo del fabricante [Reyon *et al.*, Nat. Biotechnol., 30: 460-465 (2012)], separándose las digestiones en los geles de agarosa al 2 % y cuantificándose usando ImageJ [Guschin *et al.*, Methods Mol. Biol., 649: 247-256 (2010)]. Los ensayos determinan el porcentaje de inserciones/deleciones ("indeles") en la población de células global.

#### Métodos y composiciones

Por consiguiente, la presente divulgación se refiere en particular a los siguientes aspectos no limitativos: en un primer método, método 1, la presente divulgación proporciona un método para editar un gen de BCL11A en una célula humana mediante edición genómica, comprendiendo el método la etapa de: introducir en la célula humana una o más ácido desoxirribonucleico (ADN) endonucleasas para realizar una o más roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB) dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que da como resultado una deleción, modulación o inactivación permanente de una secuencia de control transcripcional del gen de BCL11A.

En otro método, método 2, la presente divulgación proporciona un método para editar un gen de BCL11A en una célula humana mediante edición genómica, tal como se proporciona en el método 1, en el que la secuencia de control transcripcional está ubicada dentro de un segundo intrón del gen de BCL11A.

En otro método, método 3, la presente divulgación proporciona un método para editar un gen de BCL11A en una célula humana mediante edición genómica, tal como se proporciona en los métodos 1 o 2, en el que la secuencia de control transcripcional está ubicada dentro de un sitio hipersensible de ADN (DHS) +58 del gen de BCL11A.

En otro método, método 4, la presente divulgación describe un método *ex vivo* para tratar a un paciente con una hemoglobinopatía, comprendiendo el método las etapas de: crear una célula madre pluripotente inducida (iPSC) específica de paciente; editar dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A de la iPSC; diferenciar la iPSC sometida a edición genómica para dar una célula progenitora hematopoyética; e implantar la célula progenitora hematopoyética en el paciente.

En otro método, método 5, la presente divulgación describe un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se describe en el método 4, en el que la etapa de creación comprende: aislar una célula somática a partir del paciente; e introducir un conjunto de genes asociados con pluripotencia en la célula somática para inducir la célula somática para que se convierta en una célula madre pluripotente.

En otro método, método 6, la presente divulgación describe un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se describe en el método 5, en el que la célula somática es un fibroblasto.

En otro método, método 7, la presente divulgación describe un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se describe en los métodos 5 o 6, en el que el conjunto de genes asociados con pluripotencia es uno o más de los genes seleccionados del grupo que consiste en OCT4, SOX2, KLF4, Lin28, NANOG y cMYC.

En otro método, método 8, la presente divulgación describe un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se describe en uno cualquiera de los métodos 4-7, en el que la etapa de edición comprende introducir en la iPSC una o más ácido desoxirribonucleico (ADN) endonucleasas para realizar una o más roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB) dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que da como resultado una deleción, modulación o inactivación permanente de una secuencia de control transcripcional del gen de BCL11A.

En otro método, método 9, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se da a conocer en uno cualquiera de los métodos 4-8, en el que la etapa de diferenciación comprende uno o más de lo siguiente para diferenciar la iPSC sometida a edición genómica para dar una célula progenitora hematopoyética: tratamiento con una combinación de moléculas pequeñas, suministro de factores de transcripción maestros, suministro de ARNm que codifica para factores de transcripción maestros o suministro de ARNm que codifica para factores de transcripción.

En otro método, método 10, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se da a conocer en uno cualquiera de los métodos 4-9, en el que la etapa de implantación comprende implantar la célula progenitora hematopoyética en el paciente mediante trasplante, inyección local, infusión sistémica o combinaciones de los mismos.

En otro método, método 11, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía, comprendiendo el método las etapas de: aislar una célula madre mesenquimatosa a partir del paciente; editar dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador

del gen de BCL11A de la célula madre mesenquimatosa; diferenciar la célula madre mesenquimatosa sometida a edición genómica para dar una célula progenitora hematopoyética; e implantar la célula progenitora hematopoyética en el paciente.

5 En otro método, método 12, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se da a conocer en el método 11, en el que la célula madre mesenquimatosa se aísla a partir de la médula ósea o sangre periférica del paciente.

10 En otro método, método 13, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se da a conocer en los métodos 11 o 12, en el que la etapa de aislamiento comprende: aspiración de médula ósea y aislamiento de células mesenquimatosas usando medios de centrifugación por gradiente de densidad.

15 En otro método, método 14, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se da a conocer en uno cualquiera de los métodos 11-13, en el que la etapa de edición comprende introducir en la célula madre mesenquimatosa una o más ácido desoxirribonucleico (ADN) endonucleasas para realizar una o más roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB) dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que da como resultado una delección, modulación o inactivación permanente de una secuencia de control transcripcional del gen de BCL11A.

20 En otro método, método 15, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se da a conocer en uno cualquiera de los métodos 11-14, en el que la etapa de diferenciación comprende uno o más de lo siguiente para diferenciar la célula madre mesenquimatosa sometida a edición genómica para dar una célula progenitora hematopoyética: tratamiento con una combinación de moléculas pequeñas, suministro de factores de transcripción maestros, suministro de ARNm que codifica para factores de transcripción maestros o suministro de ARNm que codifica para factores de transcripción.

25 En otro método, método 16, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se da a conocer en uno cualquiera de los métodos 11-15, en el que la etapa de implantación comprende implantar la célula progenitora hematopoyética en el paciente mediante trasplante, inyección local, infusión sistémica o combinaciones de los mismos.

30 En otro método, método 17, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía, comprendiendo el método las etapas de: aislar una célula progenitora hematopoyética a partir del paciente; editar dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A de la célula progenitora hematopoyética; e implantar la célula progenitora hematopoyética sometida a edición genómica en el paciente.

35 En otro método, método 18, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se proporciona en el método 17, en el que el método comprende además tratar al paciente con factor estimulante de colonias de granulocitos (G-CSF) antes de la etapa de aislamiento.

40 En otro método, método 19, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se da a conocer en el método 18, en el que la etapa de tratamiento se realiza en combinación con plerixafor.

45 En otro método, método 20, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se da a conocer en uno cualquiera de los métodos 17-19, en el que la etapa de aislamiento comprende aislar células CD34+.

50 En otro método, método 21, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se da a conocer en uno cualquiera de los métodos 17-20, en el que la etapa de edición comprende introducir en la célula progenitora hematopoyética una o más ácido desoxirribonucleico (ADN) endonucleasas para realizar una o más roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB) dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que da como resultado una delección, modulación o inactivación permanente de una secuencia de control transcripcional del gen de BCL11A.

55 En otro método, método 22, la presente divulgación da a conocer un método *ex vivo* para tratar a un paciente con una hemoglobinopatía tal como se da a conocer en uno cualquiera de los métodos 17-21, en el que la etapa de implantación comprende implantar la célula progenitora hematopoyética sometida a edición genómica en el paciente mediante trasplante, inyección local, infusión sistémica o combinaciones de los mismos.

60 En otro método, método 23, la presente divulgación da a conocer un método *in vivo* para tratar a un paciente con una hemoglobinopatía, comprendiendo el método la etapa de editar un gen de BCL11A en una célula del paciente.

En otro método, método 24, la presente divulgación da a conocer un método *in vivo* para tratar a un paciente con una hemoglobinopatía tal como se da a conocer en el método 23, en el que la etapa de edición comprende introducir en la célula una o más ácido desoxirribonucleico (ADN) endonucleasas para realizar una o más roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB) dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que da como resultado una delección, modulación o inactivación permanente de a transcripcional control del gen de BCL11A.

En otro método, método 25, la presente divulgación da a conocer un método *in vivo* para tratar a un paciente con una hemoglobinopatía tal como se proporciona en los métodos 23 o 24, en el que la célula es una célula de médula ósea, una célula progenitora hematopoyética o una célula CD34+.

En otro método, método 26, la presente divulgación da a conocer un método según uno cualquiera de los métodos 1, 8, 14, 21 y 24, en el que la una o más ADN endonucleasas son una endonucleasa Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (también conocida como Csn1 y Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 o Cpf1; un homólogo de la misma, una recombinación de la molécula que se produce de manera natural de la misma, optimización de codones de la misma, o versiones modificadas de la misma, y combinaciones de la misma.

En otro método, método 27, la presente divulgación proporciona un método tal como se da a conocer en el método 26, en el que el método comprende introducir en la célula uno o más polinucleótidos que codifican para la una o más ADN endonucleasas.

En otro método, método 28, la presente divulgación proporciona un método tal como se proporciona en los métodos 26 o 27, en el que el método comprende introducir en la célula uno o más ácidos ribonucleicos (ARN) que codifican para la una o más ADN endonucleasas.

En otro método, método 29, la presente divulgación da a conocer un método tal como se da a conocer en los métodos 27 o 28, en el que el uno o más polinucleótidos o uno o más ARN son uno o más polinucleótidos modificados o uno o más ARN modificados.

En otro método, método 30, la presente divulgación da a conocer un método tal como se da a conocer en el método 26, en el que la una o más ADN endonucleasas son una o más proteínas o polipéptidos.

En otro método, método 31, la presente divulgación proporciona un método tal como se da a conocer en el método 30, en el que la una o más proteínas o polipéptidos están flanqueados en el extremo N-terminal, el extremo C-terminal o tanto el extremo N-terminal como el extremo C-terminal por una o más señales de localización nuclear (NLS).

En otro método, método 32, la presente divulgación proporciona un método tal como se proporciona en el método 31, en el que la una o más proteínas o polipéptidos están flanqueados por dos NLS, una NLS ubicada en el extremo N-terminal y la segunda NLS ubicada en el extremo C-terminal.

En otro método, método 33, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 31-32, en el que la una o más NLS son una NLS de SV40.

En otro método, método 34, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 1-33, en el que el método comprende además introducir en la célula uno o más ácidos ribonucleicos guía (ARNg).

En otro método, método 35, la presente divulgación proporciona un método tal como se proporciona en el método 34, en el que el uno o más ARNg son ARN guía de molécula simple (ARNgs).

En otro método, método 36, la presente divulgación proporciona un método tal como se proporciona en los métodos 34 o 35, en el que el uno o más ARNg o uno o más ARNg son uno o más ARNg modificados o uno o más ARNg modificados.

En otro método, método 37, la presente divulgación proporciona un método tal como se proporciona en el método 36, en el que el uno o más ARNg modificados comprenden tres residuos 2'-O-metil-fosforotioato en o cerca de cada uno de sus extremos 5' y 3'.

En otro método, método 38, la presente divulgación proporciona un método tal como se proporciona en el método 37, en el que el ARNg modificado es la secuencia de ácido nucleico de SEQ ID NO: 71.959.

En otro método, método 39, la presente divulgación proporciona un método tal como se proporciona en los métodos 34-38, en el que la una o más ADN endonucleasas se complejan previamente con uno o más ARNg o uno o más ARNg para formar una o más ribonucleoproteínas (RNP).

En otro método, método 40, la presente divulgación proporciona un método tal como se proporciona en el método 39, en el que la razón en peso de ARNGs con respecto a ADN endonucleasa en la RNP es de 1:1.

5 En otro método, método 41, la presente divulgación proporciona un método tal como se proporciona en el método 40, en el que el ARNGs comprende la secuencia de ácido nucleico de SEQ ID NO: 71.959, la ADN endonucleasa es una Cas9 de *S. pyogenes* que comprende una NLS de SV40 en el extremo N-terminal y una NLS de SV40 en el extremo C-terminal, en el que la razón en peso de ARNGs con respecto a ADN endonucleasa es de 1:1.

10 En otro método, método 42, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 1-41, en el que el método comprende además introducir en la célula un molde donante de polinucleótido que comprende un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada.

15 En otro método, método 43, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 1, 8, 14, 21 o 24, en el que el método comprende además introducir en la célula un ácido ribonucleico guía (ARNG) y un molde donante de polinucleótido que comprende un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada, y en el que la una o más ADN endonucleasas son una o más endonucleasas Cas9 o Cpf1 que realizan una rotura de cadena sencilla (SSB) o rotura de cadena doble (DSB) en un locus dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que facilita la inserción de una nueva secuencia a partir del molde donante de polinucleótido en el ADN cromosómico en el locus que da como resultado una inserción, modulación o inactivación permanente de la secuencia de control transcripcional del ADN cromosómico proximal al locus, y en el que el ARNG comprende una secuencia espaciadora que es complementaria a un segmento del locus.

25 En otro método, método 44, la presente divulgación proporciona un método tal como se proporciona en el método 43, en el que el proximal significa nucleótidos tanto en el sentido de 5' como en el sentido de 3' del locus.

30 En otro método, método 45, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 1, 8, 14, 21 o 24, en el que el método comprende además introducir en la célula uno o más ácidos ribonucleicos guía (ARNG) y un molde donante de polinucleótido que comprende un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada, y en el que la una o más ADN endonucleasas son una o más endonucleasas Cas9 o Cpf1 que realizan o crean un par de roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB), la primera rotura en un locus en 5' y la segunda rotura en un locus en 3', dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que facilita la inserción de una nueva secuencia a partir del molde donante de polinucleótido en el ADN cromosómico entre el locus en 5' y el locus en 3' que da como resultado una inserción, modulación o inactivación permanente de la secuencia de control transcripcional del ADN cromosómico entre el locus en 5' y el locus en 3'.

40 En otro método, método 46, la presente divulgación proporciona un método tal como se proporciona en el método 45, en el que un ARNG crea un par de SSB o DSB.

45 En otro método, método 47, la presente divulgación proporciona un método tal como se proporciona en el método 45, en el que un ARNG comprende una secuencia espaciadora que es complementaria o bien al locus en 5' o bien al locus en 3'.

50 En otro método, método 48, la presente divulgación proporciona un método tal como se proporciona en el método 45, en el que el método comprende un primer ARN guía y un segundo ARN guía, en el que el primer ARN guía comprende una secuencia espaciadora que es complementaria a un segmento del locus en 5' y el segundo ARN guía comprende una secuencia espaciadora que es complementaria a un segmento del locus en 3'.

En otro método, método 49, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 43-48, en el que el uno o dos ARNG son uno o dos ARN guía de molécula simple (ARNGs).

55 En otro método, método 50, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 43-49, en el que el uno o dos ARNG o uno o dos ARNGs son uno o dos ARNG modificados o uno o dos ARNGs modificados.

60 En otro método, método 51, la presente divulgación proporciona un método tal como se proporciona en el método 50, en el que el ARNGs modificado comprende tres residuos 2'-O-metil-fosforotioato en o cerca de cada uno de sus extremos 5' y 3'.

65 En otro método, método 52, la presente divulgación proporciona un método tal como se proporciona en el método 51, en el que el ARNGs modificado es la secuencia de ácido nucleico de SEQ ID NO: 71.959.

En otro método, método 53, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera

de los métodos 43-52, en el que la una o más endonucleasas Cas9 se complejan previamente con uno o dos ARNg o uno o dos ARNGs para formar una o más ribonucleoproteínas (RNP).

En otro método, método 54, la presente divulgación proporciona un método tal como se proporciona en el método 53, en el que la una o más endonucleasas Cas9 están flanqueadas en el extremo N-terminal, el extremo C-terminal o tanto el extremo N-terminal como el extremo C-terminal por una o más señales de localización nuclear (NLS).

En otro método, método 55, la presente divulgación proporciona un método tal como se proporciona en el método 54, en el que la una o más endonucleasas Cas9 están flanqueadas por dos NLS, una NLS ubicada en el extremo N-terminal y la segunda NLS ubicada en el extremo C-terminal.

En otro método, método 56, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 54-55, en el que la una o más NLS son una NLS de SV40.

En otro método, método 57, la presente divulgación proporciona un método tal como se proporciona en el método 53, en el que la razón en peso de ARNGs con respecto a endonucleasa Cas9 en la RNP es de 1:1.

En otro método, método 58, la presente divulgación proporciona un método tal como se proporciona en el método 53, en el que el ARNg comprende la secuencia de ácido nucleico de SEQ ID NO: 71.959, la endonucleasa Cas9 es una Cas9 de *S. pyogenes* que comprende una NLS de SV40 en el extremo N-terminal y una NLS de SV40 en el extremo C-terminal, en el que la razón en peso de ARNGs con respecto a endonucleasa Cas9 es de 1:1.

En otro método, método 59, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 43-58, en el que el molde donante es o bien monocatenario o bien bicatenario.

En otro método, método 60, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 42-59, en el que la secuencia de control transcripcional modificada está ubicada dentro de un segundo intrón del gen de BCL11A.

En otro método, método 61, tal como se proporciona en uno cualquiera de los métodos 42-59, en el que la secuencia de control transcripcional modificada está ubicada dentro de un sitio hipersensible de ADN (DHS) +58 del gen de BCL11A.

En otro método, método 62, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 42-61, en el que la inserción se realiza mediante reparación dirigida por homología (HDR).

En otro método, método 63, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 8, 14, 21, 24, 43 y 45, en el que la SSB, DSB, o el locus en 5' y el locus en 3' están ubicados dentro de un segundo intrón del gen de BCL11A.

En otro método, método 64, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 8, 14, 21, 24, 43 y 45, en el que la SSB, DSB, o DSB en 5' y DSB en 3' están ubicadas dentro de un sitio hipersensible de ADN (DHS) +58 del gen de BCL11A.

En otro método, método 65, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 1, 8, 14, 21 o 24, en el que el método comprende además introducir en la célula uno o más ácidos ribonucleicos guía (ARNg), y en el que la una o más ADN endonucleasas son una o más endonucleasas Cas9 o Cpf1 que realizan o crean un par de roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB), una primera SSB o DSB en un locus en 5' y una segunda SSB o DSB en un locus en 3', dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que provoca una delección del ADN cromosómico entre el locus en 5' y el locus en 3' que da como resultado una delección, modulación o inactivación permanente de la secuencia de control transcripcional del ADN cromosómico entre el locus en 5' y el locus en 3'.

En otro método, método 66, la presente divulgación proporciona un método tal como se proporciona en el método 65, en el que un ARNg crea un par de SSB o DSB.

En otro método, método 67, la presente divulgación proporciona un método tal como se proporciona en el método 65, en el que un ARNg comprende una secuencia espaciadora que es complementaria o bien al locus en 5' o bien al locus en 3'.

En otro método, método 68, la presente divulgación proporciona un método tal como se proporciona en el método 65, en el que el método comprende un primer ARN guía y un segundo ARN guía, en el que el primer ARN guía comprende una secuencia espaciadora que es complementaria a un segmento del locus en 5' y el segundo ARN guía comprende una secuencia espaciadora que es complementaria a un segmento del locus en 3'.

En otro método, método 69, la presente divulgación proporciona un método tal como se proporciona en los métodos

65-68, en el que el uno o más ARNg son uno o más ARN guía de molécula simple (ARNgs).

En otro método, método 70, la presente divulgación proporciona un método tal como se proporciona en los métodos 65-69, en el que el uno o más ARNg o uno o más ARNg son uno o más ARNg modificados o uno o más ARNg modificados.

En otro método, método 71, la presente divulgación proporciona un método tal como se proporciona en el método 70, en el que el ARNg modificado comprende tres residuos 2'-O-metil-fosforotioato en o cerca de cada uno de sus extremos 5' y 3'.

En otro método, método 72, la presente divulgación proporciona un método tal como se proporciona en el método 71, en el que el ARNg modificado es la secuencia de ácido nucleico de SEQ ID NO: 71.959.

En otro método, método 73, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 65-72, en el que la una o más endonucleasas Cas9 se complejan previamente con uno o más ARNg o uno o más ARNg para formar una o más ribonucleoproteínas (RNP).

En otro método, método 74, la presente divulgación proporciona un método tal como se proporciona en el método 73, en el que la una o más endonucleasas Cas9 están flanqueadas en el extremo N-terminal, el extremo C-terminal o tanto el extremo N-terminal como el extremo C-terminal por una o más señales de localización nuclear (NLS).

En otro método, método 75, la presente divulgación proporciona un método tal como se proporciona en el método 74, en el que la una o más endonucleasas Cas9 están flanqueadas por dos NLS, una NLS ubicada en el extremo N-terminal y la segunda NLS ubicada en el extremo C-terminal.

En otro método, método 76, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 74-75, en el que la una o más NLS son una NLS de SV40.

En otro método, método 77, la presente divulgación proporciona un método tal como se proporciona en el método 73, en el que la razón en peso de ARNg con respecto a endonucleasa Cas9 en la RNP es de 1:1.

En otro método, método 78, la presente divulgación proporciona un método tal como se proporciona en el método 73, en el que el ARNg comprende la secuencia de ácido nucleico de SEQ ID NO: 71.959, la endonucleasa Cas9 es una Cas9 de *S. pyogenes* que comprende una NLS de SV40 en el extremo N-terminal y una NLS de SV40 en el extremo C-terminal, en el que la razón en peso de ARNg con respecto a endonucleasa Cas9 es de 1:1.

En otro método, método 79, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 65-78, en el que tanto el locus en 5' como el locus en 3' están ubicados dentro de un segundo intrón del gen de BCL11A.

En otro método, método 80, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 65-78, en el que tanto el locus en 5' como el locus en 3' están ubicados dentro de un sitio hipersensible de ADN (DHS) +58 del gen de BCL11A.

En otro método, método 81, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 1, 8, 14, 21 o 24-80, en el que el ARNm de Cas9 o Cpf1, ARNg y molde donante se formulan, cada uno, en nanopartículas lipídicas independientes o todos ellos se formulan conjuntamente en una nanopartícula lipídica.

En otro método, método 82, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 1, 8, 14, 21 o 24-80, en el que el ARNm de Cas9 o Cpf1 se formula en una nanopartícula lipídica, y tanto el ARNg como el molde donante se suministran a la célula mediante un vector de virus adenoasociado (AAV).

En otro método, método 83, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 1, 8, 14, 21 o 24-80, en el que el ARNm de Cas9 o Cpf1 se formula en una nanopartícula lipídica, y el ARNg se suministra a la célula mediante electroporación y el molde donante se suministra a la célula mediante un vector de virus adenoasociado (AAV).

En otro método, método 84, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 1, 8, 14, 21 o 24-80, en el que la una o más RNP se suministran a la célula mediante electroporación.

En otro método, método 85, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 1-84, en el que el gen de BCL11A está ubicado en el cromosoma 2: 60.451.167-60.553.567 (Consorcio de Referencia del Genoma – GRCh38).

En otro método, método 86, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 1-85, en el que la hemoglobinopatía se selecciona de un grupo que consiste en anemia de células

falciformes y talasemia ( $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\gamma$  y combinaciones de las mismas).

En otro método, método 87, la presente divulgación proporciona un método tal como se proporciona en uno cualquiera de los métodos 1-86, en el que la edición dentro o cerca de un gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A puede reducir la expresión de gen de BCL11A.

En una primera composición, composición 1, la presente divulgación proporciona uno o más ácidos ribonucleicos guía (ARNg) para editar un gen de BCL11A en una célula a partir de un paciente con una hemoglobinopatía, comprendiendo el uno o más ARNg una secuencia espaciadora seleccionada del grupo que consiste en secuencias de ácido nucleico en SEQ ID NO: 1 - 71.947 de la lista de secuencias.

En otra composición, composición 2, la presente divulgación proporciona el uno o más ARNg de la composición 1, en la que el uno o más ARNg son uno o más ARN guía de molécula simple (ARNgs).

En otra composición, composición 3, la presente divulgación proporciona el uno o más ARNg o ARNgs de las composiciones 1 o 2, en la que el uno o más ARNg o uno o más ARNgs son uno o más ARNg modificados o uno o más ARNgs modificados.

En otra composición, composición 4, la presente divulgación proporciona el uno o más ARNgs de la composición 3, en la que el uno o más ARNgs modificados comprenden tres residuos 2'-O-metil-fosforotioato en o cerca de cada uno de sus extremos 5' y 3'.

En otra composición, composición 5, la presente divulgación proporciona el uno o más ARNgs de la composición 3, en la que el uno o más ARNgs modificados comprenden la secuencia de ácido nucleico de SEQ ID NO: 71.959.

En otra composición, composición 6, la presente divulgación proporciona un ARN guía de molécula simple (ARNgs) que comprende la secuencia de ácido nucleico de SEQ ID NO: 71.959.

#### Definiciones

El término "que comprende" o "comprende" se usa haciendo referencia a composiciones, métodos y componente(s) respectivo(s) de los mismos, que son esenciales para la invención, pero abierto a la inclusión de elementos no especificados, ya sean esenciales o no.

El término "que consiste esencialmente en" se refiere a los elementos requeridos para un aspecto dado. El término permite la presencia de elementos adicionales que no afectan sustancialmente a la(s) característica(s) básica(s) y novedosa(s) o funcional(es) de ese aspecto de la invención.

El término "que consiste en" se refiere a composiciones, métodos y componentes respectivos de los mismos tal como se describe en el presente documento, que excluyen cualquier elemento no mencionado en esa descripción del aspecto.

Las formas en singular "un", "una" y "el/la" incluyen referencias en plural, a menos que el contexto indique claramente lo contrario.

Cualquier intervalo numérico mencionado en esta memoria descriptiva describe todos los subintervalos con la misma precisión numérica (es decir, que tienen el mismo número de cifras especificados) incluidos dentro del intervalo mencionado. Por ejemplo, un intervalo mencionado de "1,0 a 10,0" describe todos los subintervalos entre (e incluyendo) el valor mínimo mencionado de 1,0 y el valor máximo mencionado de 10,0, tales como, por ejemplo, "de 2,4 a 7,6", aunque el intervalo de "2,4 a 7,6" no se mencione expresamente en el texto de la memoria descriptiva. Por consiguiente, el solicitante se reserva el derecho a modificar esta memoria descriptiva, incluyendo las reivindicaciones, para mencionar expresamente cualquier subintervalo con la misma precisión numérica incluido dentro de los intervalos expresamente mencionados en esta memoria descriptiva. Todos de tales intervalos quedan inherentemente descritos en esta memoria descriptiva de tal manera que la modificación para mencionar expresamente cualquiera de tales subintervalos cumplirá con los requisitos de descripción escrita, suficiencia de la descripción y materia añadida, incluyendo los requisitos según 35 U.S.C. § 112(a) y el artículo 123(2) del CPE. Además, a menos que se especifique expresamente o se requiera lo contrario por el contexto, todos los parámetros numéricos descritos en esta memoria descriptiva (tales como los que expresan valores, intervalos, cantidades, porcentajes y similares) pueden interpretarse como si estuvieran precedidos por el término "aproximadamente", aunque el término "aproximadamente" no aparezca expresamente antes de un número. Adicionalmente, parámetros numéricos descritos en esta memoria descriptiva deben interpretarse a la vista del número de cifras significativas notificadas, precisión numérica, y aplicando técnicas de redondeo habituales. También se entiende que los parámetros numéricos descritos en esta memoria descriptiva presentarán necesariamente la característica de variabilidad inherente de las técnicas de medición subyacentes usadas para determinar el valor numérico del parámetro.



## Ejemplos

La invención se entenderá más completamente mediante referencia a los siguientes ejemplos, que proporcionan aspectos ilustrativos, no limitativos, de la invención.

- 5 Los ejemplos describen el uso de los sistema de CRISPR como técnica de edición genómica ilustrativa para crear delecciones, inserciones o sustituciones genómicas definidas, denominadas “modificaciones genómicas” en el presente documento, dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A que conducen a una delección, modulación o inactivación permanente de una secuencia de control transcripcional del gen de BCL11A. La introducción de las modificaciones terapéuticas definidas representa una  
10 estrategia terapéutica novedosa para la posible mejora de una hemoglobinopatía, tal como se describe y se ilustra en el presente documento.

Ejemplo 1 – Sitios diana de CRISPR/SpCas9 para la secuencia de control transcripcional del gen de BCL11A

- 15 Se exploraron regiones de la secuencia de control transcripcional de 12,4 kb del gen de BCL11A para detectar sitios diana. Se exploró cada zona para detectar un motivo adyacente de protoespaciador (PAM) que tenía la secuencia NRG. Se identificaron secuencias espaciadoras de ARNg de 20 pb correspondientes al PAM, tal como se muestra en SEQ ID NO: 1 - 29.482 de la lista de secuencias.

- 20 Ejemplo 2 – Sitios diana de CRISPR/SaCas9 para la secuencia de control transcripcional del gen de BCL11A

- Se exploraron regiones de la secuencia de control transcripcional de 12,4 kb del gen de BCL11A para detectar sitios diana. Se exploró cada zona para detectar un motivo adyacente de protoespaciador (PAM) que tenía la secuencia NNGRRT. Se identificaron secuencias espaciadoras de ARNg de 20 pb correspondientes al PAM, tal como se muestra  
25 en SEQ ID NO: 29.483 – 32.387 de la lista de secuencias.

Ejemplo 3 – Sitios diana de CRISPR/StCas9 para la secuencia de control transcripcional del gen de BCL11A

- 30 Se exploraron regiones de la secuencia de control transcripcional de 12,4 kb del gen de BCL11A para detectar sitios diana. Se exploró cada zona para detectar un motivo adyacente de protoespaciador (PAM) que tenía la secuencia NNAGAAW. Se identificaron secuencias espaciadoras de ARNg de 20 pb correspondientes al PAM, tal como se muestra en SEQ ID NO: 32.388 – 33.420 de la lista de secuencias.

Ejemplo 4 – Sitios diana de CRISPR/TdCas9 para la secuencia de control transcripcional del gen de BCL11A

- 35 Se exploraron regiones de la secuencia de control transcripcional de 12,4 kb del gen de BCL11A para detectar sitios diana. Se exploró cada zona para detectar un motivo adyacente de protoespaciador (PAM) que tenía la secuencia NAAAAC. Se identificaron secuencias espaciadoras de ARNg de 20 pb correspondientes al PAM, tal como se muestra en SEQ ID NO: 33.421 – 33.851 de la lista de secuencias.

40 Ejemplo 5 – Sitios diana de CRISPR/NmCas9 para la secuencia de control transcripcional del gen de BCL11A

- Se exploraron regiones de la secuencia de control transcripcional de 12,4 kb del gen de BCL11A para detectar sitios diana. Se exploró cada zona para detectar un motivo adyacente de protoespaciador (PAM) que tenía la secuencia  
45 NNNNGHTT. Se identificaron secuencias espaciadoras de ARNg de 20 pb correspondientes al PAM, tal como se muestra en SEQ ID NO: 33.852 – 36.731 de la lista de secuencias.

Ejemplo 6 – Sitios diana de CRISPR/Cpf1 para la secuencia de control transcripcional del gen de BCL11A

- 50 Se exploraron regiones de la secuencia de control transcripcional de 12,4 kb del gen de BCL11A para detectar sitios diana. Se exploró cada zona para detectar un motivo adyacente de protoespaciador (PAM) que tenía la secuencia YTN. Se identificaron secuencias espaciadoras de ARNg de 22 pb correspondientes al PAM, tal como se muestra en SEQ ID NO: 36.732 – 71.947 de la lista de secuencias.

- 55 Ejemplo 7 – Análisis bioinformático de las cadenas guía

- Se examinaron y seleccionaron guías candidatas en un procedimiento de múltiples etapas que implica tanto unión teórica como actividad evaluada de manera experimental. A modo de ilustración, guías candidatas que tienen secuencias que coinciden con un sitio específico particular, tal como un sitio dentro de la secuencia de control transcripcional del gen de BCL11A, con PAM adyacente pueden evaluarse para determinar su potencial para escindir  
60 en sitios inespecíficos que tienen una secuencias similar, usando una o más de una variedad de herramientas de bioinformática disponibles para evaluar la unión inespecífica, tal como se describe y se ilustra en más detalle a continuación, con el fin de evaluar la probabilidad de efectos en posiciones cromosómicas distintas de las pretendidas. Entonces, los candidatos que se predice que tienen un potencial relativamente inferior para actividad inespecífica pueden evaluarse de manera experimental para medir su actividad específica, y después actividades inespecíficas en  
65 diversos sitios. Las guías preferidas tienen una actividad específica suficientemente alta como para lograr niveles

deseados de edición génica en el locus seleccionado, y actividad inespecífica relativamente inferior para reducir la probabilidad de alteraciones en otros loci cromosómicos. La razón de actividad específica con respecto a inespecífica se denomina con frecuencia “especificidad” de una guía.

Para el examen inicial de actividades inespecíficas predichas, hay varias herramientas de bioinformática que se conocen y están públicamente disponibles que pueden usarse para predecir los sitios inespecíficos más probables; y, dado que la unión a sitios diana en el sistema de nucleasa CRISPR/Cas9/Cpf1 está impulsada por emparejamiento de bases de Watson-Crick entre secuencias complementarias, el grado de diferencia (y por tanto el potencial reducido para unión inespecífica) está esencialmente relacionado con diferencias de secuencia primaria: apareamientos erróneos y protuberancias, es decir bases que se cambian a una base no complementaria, e inserciones o deleciones de bases en el posible sitio inespecífico con respecto al sitio diana. Una herramienta de bioinformática a modo de ejemplo denominada COSMID (“*CRISPR Off-target Sites with Mismatches, Insertions and Deletions*”, sitios inespecíficos de CRISPR con apareamientos erróneos, inserciones y deleciones) (disponible en la web en [crispr.bme.gatech.edu](http://crispr.bme.gatech.edu)) compila tales similitudes. Otras herramientas de bioinformática incluyen, pero no se limitan a, GUIDO, autoCOSMID y CCTop.

Se usó bioinformática para minimizar la escisión inespecífica con el fin de reducir los efectos perjudiciales de mutaciones y reordenamientos cromosómicos. Estudios con sistemas de CRISPR/Cas9 sugirieron la posibilidad de una alta actividad inespecífica debido a la hibridación no específica de la cadena guía a secuencias de ADN con apareamientos erróneos de pares de bases y/o protuberancias, particularmente en posiciones distales con respecto a la región de PAM. Por tanto, es importante tener una herramienta de bioinformática que pueda identificar posibles sitios inespecíficos que tienen inserciones y/o deleciones entre la cadena guía de ARN y secuencias genómicas, además de apareamientos erróneos de pares de bases. Por tanto, se usó la herramienta basada en bioinformática, COSMID (sitios inespecíficos de CRISPR con apareamientos erróneos, inserciones y deleciones) para buscar genomas para detectar posibles sitios inespecíficos de CRISPR (disponible en la web en [crispr.bme.gatech.edu](http://crispr.bme.gatech.edu)). COSMID emitió listas clasificadas de los posibles sitios inespecíficos basándose en el número y la ubicación de apareamientos erróneos, permitiendo una elección más fundada de sitios diana, y evitando el uso de sitios con escisión inespecífica más probable.

Se emplearon segmentaciones bioinformáticas adicionales que ponderan la actividad específica y/o inespecífica estimada de ARNg que selecciona como diana sitios en una región. Otras características que pueden usarse para predecir la actividad incluyen información sobre el tipo de célula en cuestión, accesibilidad de ADN, estado de cromatina, sitios de unión a factor de transcripción, datos de unión a factor de transcripción y otros datos de CHIP-seq. Se ponderaron factores adicionales que predicen la eficiencia de edición, tales como posiciones relativas y direcciones de pares de ARNg, características de secuencia local y microhomologías.

#### Ejemplo 8 – Pruebas de guías preferidas en células para determinar actividad específica

Entonces se someterán a prueba los ARNg que se predice que tienen la actividad inespecífica más baja para determinar la actividad específica en células K562 y se evaluará su frecuencia de indeles usando TIDE.

TIDE es una herramienta en la web para evaluar rápidamente la edición genómica mediante CRISPR-Cas9 de un locus diana determinado mediante un ARN guía (ARNg o ARNGs). Basándose en datos de perfiles de secuencia cuantitativos a partir de dos reacciones de secuenciación capilar convencionales, el software TIDE cuantifica la eficacia de edición e identifica los tipos predominantes de inserciones y deleciones (indeles) en el ADN de una combinación de células seleccionadas como diana. Véase Brinkman *et al*, Nucl. Acids Res. (2014) para una explicación detallada y ejemplos. Un método alternativo es secuenciación de nueva generación (NGS), también conocida como secuenciación de alto rendimiento, que es el término general usado para describir varias tecnologías de secuenciación modernas diferentes incluyendo: secuenciación de Illumina (Solexa), secuenciación de Roche 454, secuenciación Ion torrent: Proton/PGM, y secuenciación SOLiD. Estas recientes tecnologías permiten secuenciar ADN y ARN de manera mucho más rápida y económica que la secuenciación de Sanger anteriormente usada y, como tal, han revolucionado el estudio de la genómica y biología molecular.

La transfección de células de cultivo tisular permite examinar diferentes constructos y unos medios robustos de someter a prueba la actividad y especificidad. Las líneas celulares de cultivo tisular, tales como K562 o HEK293T, se transfectan fácilmente y dan como resultado una alta actividad. Se evaluarán estas u otras líneas celulares para determinar las líneas celulares que coinciden con CD34+ y proporcionan el mejor sustituto. Entonces se usarán estas células para muchas pruebas de fase temprana. Por ejemplo, pueden transfectarse ARNg individuales para Cas9 de *S. pyogenes* en las células usando plásmidos, tales como, por ejemplo, CTx-1, CTx-2 o CTx-3 descritos en la figura 1A-1C, que son adecuados para la expresión en células humanas. Alternativamente, también pueden usarse vectores comercialmente disponibles. Para la evaluación de frecuencia de indeles de los ARNg de BCL11A descritos en el presente documento, se empleó un plásmido de expresión de Cas9 comercialmente disponible (GeneArt, Thermo Fisher). Varios días después (48 h para este experimento), se recogió el ADN genómico y se amplificó el sitio diana mediante PCR. Se midió la actividad de corte mediante la tasa de inserciones, deleciones y mutaciones introducidas mediante reparación de NHEJ de los extremos de ADN libres. Aunque este método no puede diferenciar secuencias correctamente reparadas de ADN no escindido, puede estimarse el nivel de corte mediante la cantidad de reparación

errónea. Puede observarse actividad inespecífica amplificando sitios inespecíficos supuestos identificados y usando métodos similares para detectar la escisión. También puede someterse a ensayo la translocación usando cebadores que flanquean sitios de corte, para determinar si se producen translocaciones y cortes específicos. Se han desarrollado ensayos no guiados que permiten pruebas complementarias de escisión inespecífica incluyendo secuencia guía. Entonces puede realizarse un seguimiento del ARNg o pares de ARNg con actividad significativa en células cultivadas para medir la modulación o inactivación del sitio hipersensible de ADN (DHS) +58 dentro de la secuencia de control transcripcional del gen de BCL11A. Puede realizarse de nuevo un seguimiento de los acontecimientos inespecíficos. De manera similar, pueden transfectarse células CD34+ y medirse el nivel de modulación o inactivación del sitio hipersensible de ADN (DHS) +58 dentro de la secuencia de control transcripcional del gen de BCL11A y posibles acontecimientos inespecíficos. Estos experimentos permiten la optimización del diseño y suministro de nucleasa y donante.

#### Ejemplo 9 – Pruebas de guías preferidas en células para determinar actividad inespecífica

Entonces, los ARNg que tienen la mejor actividad específica a partir de los estudios de TIDE y de secuenciación de nueva generación en el ejemplo anterior se someterán a prueba para determinar la actividad inespecífica usando secuenciación de genoma completo. Los ARNg candidatos se evaluarán de manera más completa en células CD34+ o iPSC.

#### Ejemplo 10 – Pruebas de combinaciones de ARNg preferidos en células

Los ARNg que tienen la mejor actividad específica a partir de los estudios de TIDE y de secuenciación de nueva generación y la actividad inespecífica más baja se someterán a prueba en combinaciones para evaluar el tamaño de la delección resultante del uso de cada combinación de ARNg. Se evaluarán posibles combinaciones de ARNg en células CD34+ humanas primarias.

Por ejemplo, se someterán a prueba combinaciones de ARNg para determinar la eficiencia de la delección de la totalidad o una porción de la secuencia de control transcripcional del gen de BCL11A. También se someterán a prueba las combinaciones de ARNg para determinar la eficiencia de la delección de la totalidad o una porción del sitio hipersensible de ADN (DHS) +58 del gen de BCL11A.

#### Ejemplo 11 – Pruebas de diferentes enfoques para edición génica mediante HDR

Tras someter a prueba los ARNg para detectar tanto actividad específica como actividad inespecífica, se someterán a prueba estrategias de modulación/inactivación e inserción para determinar la edición génica mediante HDR.

Para el enfoque de modulación/inactivación, se proporcionará molde de ADN donante como un oligonucleótido monocatenario corto, un oligonucleótido bicatenario corto (secuencia de PAM intacta/secuencia de PAM mutada), una molécula de ADN monocatenario larga (secuencia de PAM intacta/secuencia de PAM mutada) o una molécula de ADN bicatenario larga (secuencia de PAM intacta/secuencia de PAM mutada). El molde de ADN donante comprenderá o bien un gen o ADNc de BCL11A de tipo natural que comprende una secuencia de control transcripcional modificada o bien un gen o ADNc de BCL11A de tipo natural que comprende un sitio hipersensible de ADN (DHS) +58 modificado (por ejemplo, mutado). Además, el molde de ADN donante se suministrará mediante AAV.

Para el enfoque de inserción de ADNc, un ADN monocatenario o bicatenario puede incluir más de 40 nt de la secuencia de control transcripcional modificada del gen de BCL11A. El ADN monocatenario o bicatenario puede incluir más de 80 nt de la secuencia de control transcripcional modificada del gen de BCL11A. El ADN monocatenario o bicatenario puede incluir más de 100 nt de la secuencia de control transcripcional modificada del gen de BCL11A. El ADN monocatenario o bicatenario puede incluir más de 150 nt de la secuencia de control transcripcional modificada del gen de BCL11A. El ADN monocatenario o bicatenario puede incluir más de 300 nt de la secuencia de control transcripcional modificada del gen de BCL11A. El ADN monocatenario o bicatenario puede incluir más de 400 nt de la secuencia de control transcripcional modificada del gen de BCL11A. Alternativamente, el molde de ADN se suministrará mediante AAV.

Para el enfoque de inserción de ADNc, un ADN monocatenario o bicatenario puede incluir más de 40 nt del sitio hipersensible de ADN (DHS) +58 modificado del gen de BCL11A. El ADN monocatenario o bicatenario puede incluir más de 80 nt del sitio hipersensible de ADN (DHS) +58 modificado del gen de BCL11A. El ADN monocatenario o bicatenario puede incluir más de 100 nt del sitio hipersensible de ADN (DHS) +58 modificado del gen de BCL11A. El ADN monocatenario o bicatenario puede incluir más de 150 nt del sitio hipersensible de ADN (DHS) +58 modificado del gen de BCL11A. El ADN monocatenario o bicatenario puede incluir más de 300 nt del sitio hipersensible de ADN (DHS) +58 modificado del gen de BCL11A. El ADN monocatenario o bicatenario puede incluir más de 400 nt del sitio hipersensible de ADN (DHS) +58 modificado del gen de BCL11A. Alternativamente, el molde de ADN se suministrará mediante AAV.

#### Ejemplo 12 – Nueva evaluación de combinaciones principales de CRISPR-Cas9/donante de ADN

Tras someter a prueba las diferentes estrategias para edición génica mediante HDR, volverán a evaluarse las combinaciones principales de CRISPR-Cas9/donante de ADN en células humanas primarias para determinar la eficiencia de delección, recombinación y especificidad inespecífica. Se formulará ARNm de Cas9 o RNP en nanopartículas lipídicas para el suministro, se formularán ARNGs en nanopartículas o se suministrarán como AAV, y se formulará ADN donante en nanopartículas o se suministrará como AAV.

#### Ejemplo 13 – Pruebas *in vivo* en modelo animal relevante

Tras haberse evaluado de nuevo las combinaciones de CRISPR-Cas9/donante de ADN, se someterán a prueba las formulaciones principales *in vivo* en un modelo animal.

El cultivo en células humanas permite pruebas directas en la diana humana y el contexto de genoma humano, tal como se describió anteriormente.

Pueden observarse evaluaciones de eficacia y seguridad preclínicas mediante injerto de células CD34+ humanas o de ratón modificadas en ratones NSG o similares. Pueden observarse las células modificadas en los meses posteriores al injerto.

#### Ejemplo 14 – Edición de células con diversos ARNG

Se cultivaron células CD34+ de sangre periférica humanas movilizadas a partir de donantes humanos 1-3 en medio StemSpan libre de suero con suplemento de expansión CD34+ durante dos días. Se lavaron 100.000 células y se sometieron a electroporación usando ARNm de Cas9 con ARNG de Corfu grande (CLO), ARNG de Corfu pequeño (CSO), ARNG de HPFH5, ARNG de Kenya, ARNGs de SD2, o ARNGs de SPY101. Se dejaron recuperar las células durante dos días antes de cambiarse a un medio de diferenciación eritroide (IMDM+Glutamax suplementado con suero humano al 5 %, insulina 10 ug/ml, SCF 20 ng/ml, IL-3 5 ng/ml, EPO 3 U/ml, dexametasona 1 uM,  $\beta$ -estradiol 1 uM, holo-transferrina 330 ug/ml y heraprina 2 U/ml). Se determinó el porcentaje de inserciones/delecciones ("indeles") para cada una de las células sometidas a electroporación con ARNG de Corfu grande (CLO), células sometidas a electroporación con ARNG de Corfu pequeño (CSO), células sometidas a electroporación con ARNG de HPFH5, células sometidas a electroporación con ARNG de Kenya, células sometidas a electroporación con ARNGs de SD2 y las células sometidas a electroporación con ARNGs de SPY101 (figura 3), tal como se describe en las secciones de "Detección de mutaciones específicas e inespecíficas mediante secuenciación" y "Ensayos de detección de mutaciones" descritas en el presente documento. Tras diferenciar estas células durante 12 días en medio de diferenciación eritroide, se recogió ARN para evaluar los niveles de hemoglobina mediante PCR en tiempo real cuantitativa (figuras 4A-4C).

Se generaron progenitores eritroides individuales usando citometría de flujo un día después y se cultivaron en el medio de diferenciación eritroide para expandirse y crecer como colonias. Se dividió cada colonia y se recogió 12 días tras la clasificación para el análisis de ADN y ARN. Se recogieron las colonias hermanas 15 días tras la clasificación para el análisis de proteínas de hemoglobina. Se determinó la expresión de globina (razón de  $\gamma$ /ARN 18s o razón de  $\gamma/\alpha$ ) mediante PCR en tiempo real cuantitativa y se comparó para cada una de las colonias eritroides sometidas a edición (figuras 5A-5B).

#### Ejemplo 15 – Pruebas de ARNGs de SPY101

Pueden producirse tres resultados de edición génica posibles dentro del intrón 2 del gen de BCL11A cuando se usa ARNGs de SPY101. El primer resultado de edición génica que puede producirse cuando se usa ARNGs de SPY101 solo da como resultado indeles en ambos alelos (indel/indel, figura 6). El segundo resultado de edición génica que puede producirse cuando se usa ARNGs de SPY101 da como resultado un clon tanto con indeles como con secuencias de tipo natural en los dos alelos (indel/WT, figura 6). El tercer resultado de edición génica que puede producirse cuando se usa ARNGs de SPY101 da como resultado una colonia con secuencias de tipo natural en ambos alelos (WT/WT, figura 6).

Cuando se usa ARNGs de SPY101, el 92 % de las colonias eritroides se sometieron a edición. Por ejemplo, el 92 % de las colonias eritroides tenían alelos con indeles (figura 6).

Se midió la expresión de  $\gamma$ -globina (razón de ARNm de  $\gamma/\alpha$ -globina o razón de ARNm de  $\gamma/(\gamma+\beta)$ -globina) en colonias eritroides individuales sometidas a edición con SPY101 (figuras 7A-B). Las colonias eritroides individuales incluyeron colonias con indel bialélico u homocigótico (indel/indel), colonias con un indel monoalélico o heterocigótico (indel/WT) y colonias con secuencias de tipo natural en ambos alelos (WT/WT). Las colonias eritroides que tenían indeles podían expresar niveles superiores de gamma-globina en comparación con los clones con secuencias de tipo natural en ambos alelos (figuras 7A-B).

#### Ejemplo 16 – Estrategia terapéutica para enfermedad de células falciformes (SCD) y $\beta$ -talasemia

La siguiente tabla (tabla 4) proporciona información relacionada con los ARNG usados en los ejemplos 16-17.

Tabla 4

ARNg	Nombre	Secuencia	SEQ ID NO.
ARNg A	CL01	5'usgsusGUGCUGGCCCGCAACUUGUU UUAGAGCUAGAAAUAGCAAGUUAAAA UAAGGCUAGUCCGUUAUCAACUUGAA AAAGUGGCACCGAGUCGGUGCususus U3'	71950
ARNg B	CL08	5'cscscsACUCAAGAGAU AUGGUGGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71951
ARNg C	CS02	5'gsusasGACCACCAGUAAUCUGAGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71952
ARNg D	CS06	5'asgsusAUACCUC CCAUACCAUGGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71953
ARNg E	HPFH5-15	5'csusgsUCUUAUUACCCUGUCAUGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71954
ARNg F	HPFH5-4	5'ascsusGAGUUCUAAAAUCAUCGGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71955

ARNg G	Kenya 02	5'gsuscsUUCAGCCUACAACAUACGUUU UAGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71956
ARNg H	Kenya 17	5'gsususAAGUUCAUGUCAUAGGAGUU UUAGAGCUAGAAAUAGCAAGUUA AAAA UAAGGCUAGUCCGUUAUCAACUUGAA AAAGUGGCACCGAGUCGGUGCususus U3'	71957
ARNg I	SD2	5'csususGUCAAGGCUAUUGGUCAGUU UUAGAGCUAGAAAUAGCAAGUUA AAAA UAAGGCUAGUCCGUUAUCAACUUGAA AAAGUGGCACCGAGUCGGUGCususus U3'	71958
ARNg J	SPY	5'csusasACAGUUGCUUUUAUCACGUUU UAGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71959

a, g, u: residuos 2'-O-metilo

s: fosforotioato

5

A, C, G, U: residuos de ARN

La siguiente tabla (tabla 5) proporciona información relacionada con las dianas mencionadas en los ejemplos 16-17.

10 Tabla 5

Diana 1	Corfu grande
Diana 2	Corfu pequeño
Diana 3	HPFH5
Diana 4	KENYA
Diana 5	SD2
Diana 6	SPY101

Una estrategia terapéutica para SCD y  $\beta$ -talasemia usó CRISPR/Cas9 para recrear las mismas mutaciones genéticas que se producen de manera natural en pacientes con HPFH. Se aislaron células madre hematopoyéticas de los pacientes, se trataron esas células *ex vivo* con CRISPR/Cas9 para crear ediciones genéticas de HPFH y después volvieron a introducirse las células sometidas a edición en los pacientes. Las células madre genéticamente modificadas dieron lugar a eritrocitos que contienen niveles suficientes de HbF como para reducir significativamente la gravedad de los síntomas de la enfermedad. Se han priorizado varias ediciones genéticas basándose en el grado de regulación por incremento de HbF observado en la naturaleza, la capacidad para recrear estas ediciones con alta eficiencia usando CRISPR/Cas9 y la ausencia de edición inespecífica.

15

20

Se seleccionaron computacionalmente secuencias de ARN guía (ARNg) candidatas y después se examinaron para determinar la eficacia de edición específica en células CD34+. En la figura 8 se muestran los resultados de un examen de este tipo. Se identificaron ARNg con una edición específica sistemática alta (>70 %) a lo largo de múltiples muestras de donante. Cada donante de células CD34+ se representa mediante un símbolo único (▲, ★, ●) y se mide dos veces la eficiencia de edición específica para cada donante.

Se examinaron ARNg candidatos en células CD34+ para determinar la actividad inespecífica examinando cientos de sitios que se identificó computacionalmente que eran los más similar en cuanto a secuencia al sitio específico pretendido y, por tanto, tienen el mayor potencial de actividad inespecífica. Las figuras 9A-B muestran el enfoque experimental (figura 9A) y los resultados (figura 9B) para cada uno de los ARNg sometidos a prueba en la figura 8. La mayoría de los ARNg no presentaron ninguna actividad inespecífica detectable, ni siquiera en sitios predichos. Tan solo ARNg C y ARNg G muestran actividad inespecífica. Se usaron múltiples sondas para cada sitio predicho para aumentar la sensibilidad del ensayo.

Se usaron ARNg candidatos para recrear HPFH específica u otras modificaciones en células eritroides obtenidas a partir de pacientes con SCD y  $\beta$ -talasemia, así como de donantes sanos. Tras la diferenciación eritroide, se midieron los niveles de transcrito de globina para evaluar el aumento de  $\gamma$ -globina con respecto a  $\alpha$  o  $\beta$ -globina. Tal como se muestra en las figuras 10A-B, se observaron niveles de ARNm de  $\gamma$ -globina mayores del 30 % en células de paciente sometidas a edición con ARNg para recrear la diana 5 y 6 de HPFH. Las muestras de pacientes con SCD y  $\beta$ -talasemia mostraron un aumento absoluto más grande de  $\gamma$ -globina que las de donantes sanos, compatible con la observación de HbF superior en pacientes que en portadores heterocigóticos con HPFH. El nivel de fondo para células tratadas de manera simulada a partir de cada donante se restó de los valores mostrados. Los datos representan un único experimento, excepto por los datos de pacientes con SCD que representan la media de 3 muestras de donantes diferentes. La eficiencia de edición fue similar para todos los experimentos.

Para garantizar que las eficiencias de edición en la población CD34+ global eran representativas de aquellas en HSC de repoblación a largo plazo (LT-HSC), se clasificaron células CD34+ globales en subpoblaciones específicas y se sometieron a ensayo para determinar la eficiencia de edición específica tal como se muestra en las figuras 11A-C. Se observó una alta eficiencia de edición en la población de LT-HSC. Se realizaron experimentos usando proteína SPY101 y Cas9 en 4 donantes. Las barras representan la media  $\pm$  EEM. LT-HSC, célula madre hematopoyética a largo plazo; MPP, progenitor multipotente; MLP, progenitor multilinfoide; CMP, progenitor mieloide común; MEP, progenitor de eritrocitos de megacariocito; GMP, progenitor de macrófagos de granulocitos.

Se realizaron estudios de injerto *in vivo* en ratones inmunodeprimidos para confirmar que las HSPC sometidas a edición génica conservan el potencial para repoblación a largo plazo del sistema hematopoyético. Células CD34+ humanas a partir de donantes sanos no se trataron, no se sometieron a edición o se sometieron a edición génica usando ARNg de SPY101 y se introdujeron en ratones NSG. Tal como se muestra en la figura 12, la presencia de niveles similares de células hCD45RA+ (a las 8 semanas tras el injerto) en ratones a los que se les inyectaron HSPC sin tratar/sin someter a edición y ratones a los que se les inyectaron HSPC sometidas a edición génica con SPY101 confirmó que las células sometidas a edición con SPY101 conservaron potencial de injerto. Los puntos de datos representan animales individuales y representan el porcentaje de células vivas que eran CD45RA+ humanas. Media  $\pm$  DE. "Sin tratar" representa HSPC que no se sometieron a electroporación y se inyectaron en ratones inmunodeprimidos. "Sin someter a edición" representa HSPC que se sometieron a electroporación, pero no se sometieron a edición génica y se inyectaron en ratones inmunodeprimidos. "SPY101" representa HSPC que se sometieron a electroporación con Cas9 y ARNg de SPY101 y se inyectaron en ratones inmunodeprimidos.

El desarrollo del procedimiento se inició en una instalación con capacidad para GMP en preparación para estudios clínicos. Tal como se muestra en la figura 13, no se observó ninguna pérdida significativa de eficacia de edición génica a escala clínica en un procedimiento compatible con GMP. Se calculó el promedio de los datos a lo largo de 4 o más experimentos,  $\pm$  DE.

Se han iniciado estudios de GLP/toxicología para los principales candidatos, tal como se muestra en la figura 14. Dos estudios independientes en ratones NSG permitirán una caracterización exhaustiva de biodistribución y toxicología de células CD34+ sometidas a edición.

#### Ejemplo 17 – Estrategia terapéutica para enfermedad de células falciformes (SCD) y $\beta$ -talasemia

Los resultados de la recreación de seis variantes de HPFH diferentes, o "dianas" de edición, en células CD34+ de mPB humanas se muestran en las figuras 16A-B y la figura 17. Se trataron las células CD34+ con CRISPR/Cas9, se diferenciaron para dar eritrocitos y después se sometieron a ensayo para determinar la expresión de ARNm y proteína de HbF globales (figuras 16A-B) y en colonias (figura 17), usando un procedimiento experimental demostrado en la figura 15.

Los resultados presentados en las figuras 16A-B eran de 3 donantes diferentes para las dianas 1-3, y 7 donantes

diferentes para las dianas 4-6. Se restó el nivel de fondo para células tratadas de manera simulada. Los datos son la media  $\pm$  EEM. El análisis global confirmó la regulación por incremento de HBF y permitió la priorización de dianas que demostraron los niveles más altos de HBF.

5 El análisis clonal presentado en la figura 17 permitió la confirmación de que las ediciones genéticas provocadas por CRISPR/Cas9 eran realmente la causa del aumento de HBF a nivel de células individuales. Los resultados eran de un único donante, y 50-80 colonias por diana. Se midieron niveles de transcrito de ARNm mediante qRT-PCR. Los datos son la media  $\pm$  EEM.

10 Las dianas 5 y 6 presentaron los niveles de HBF más altos y se analizaron adicionalmente en las figuras 18A-B. Los datos son la media  $\pm$  EEM. WT designa colonias que no muestran evidencias de edición génica, Heterocigótico o Het designa colonias con un alelo sometido a edición, y Homocigótico u Homo designa colonias con ambos alelos sometidos a edición. Las evidencias en las figuras 16A-B, 17 y 18A-B respaldan la relación causal entre las ediciones genéticas producidas y la regulación por incremento deseada de HBF, proporcionando validación adicional para la estrategia terapéutica propuesta.

#### Ejemplo 18 – Pruebas de ARN guía preferidos en células para determinar actividad específica

20 Se cultivaron células CD34<sup>+</sup> de sangre periférica movilizadas (mPB) humanas de cuatro donantes independientes en medios CellGro® libres de suero que incluían factor de células madre humano recombinante (SCF) 100 ng/ml, ligando de Fit 3 humano recombinante (FLT3L) 100 ng/ml y trombopoietina (TPO) 100 ng/ml. Se lavaron 200.000 células por donante y se sometieron a electroporación usando un electroporador Lonza sin ningún componente de edición de CRISPR/Cas9 (muestra de electroporación simulada), con ARNg de GFP y proteína Cas9 como control negativo (GFP), con ARNg de SPY101 y proteína Cas9 (SPY), con ARNg de SD2 y proteína Cas9 (SD2) o ARNg dobles de exón 2 de BCL11A y proteína Cas9 (Ex2). La proteína Cas9 recombinante codifica para Cas9 de *S. pyogenes* flanqueada por dos secuencias de localización nuclear (NLS) de SV40. Estos experimentos se realizaron usando una ribonucleoproteína (RNP) con una razón en peso de ARNg con respecto a Cas9 de 1:1. El ARNg de SPY101 crea una perturbación de indel de sitio de unión a gata1 de DHS+58 en el intrón 2 del locus de BCL11a. El ARNg de SD2 crea indeles y una delección de 4,9 Kb en el locus de beta-globina humana. La delección de 4,9 Kb está ubicada en el sentido de 5' de HBG1 e incluye toda la secuencia de HBG2. La delección de 4,9 Kb empieza 168 pb en 5' con respecto a la secuencia que codifica para HBG2 y termina 168 pb en 5' con respecto a la secuencia que codifica para HBG1. Los ARNg de exón 2 crean una delección de 196 pb en el exón 2 del locus de BCL11A y sirvieron como control positivo. Células CD34<sup>+</sup> de mPB humanas que no se sometieron a electroporación sirvieron como control negativo (sin EP).

35 Tras la electroporación, se dejaron recuperar las células CD34<sup>+</sup> de mPB sometidas a edición génica durante dos días antes de cambiarse a un medio de diferenciación eritroide (IMDM+L-glutamina suplementada con suero humano al 5 %, insulina 10 ug/ml, SCF 20 ng/ml, IL-3 5 ng/ml, EPO 3 U/ml, dexametasona 1 uM, holo-transferrina 330 ug/ml y heparina 2 U/ml). Se diferenciaron las células CD34<sup>+</sup> de mPB sometidas a edición génica para dar eritrocitos y se sometieron adicionalmente a prueba mediante análisis de TIDE, análisis de ddPCR, análisis de PCR en tiempo real cuantitativa, FACS y CL-EM (figuras 20A-B, 21A-D, 22A-B y 23A-D). El procedimiento experimental global se demuestra en la figura 19.

#### Análisis de TIDE / análisis de ddPCR

45 Se aisló ADN genómico y se sometió a prueba para cada una de las muestras de células CD34<sup>+</sup> de mPB humanas sometidas a edición génica que se hicieron crecer en medio de diferenciación. Se aisló ADN genómico a partir de las células en los días 1, 11, 13 y 15 tras la diferenciación. Se analizó el ADN genómico mediante análisis de TIDE, que es una herramienta en la web para evaluar rápidamente la edición genómica mediante CRISPR-Cas9 de un locus diana determinado mediante un ARN guía (ARNg o ARNgS). Los resultados presentados en las figuras 20A-B eran de 4 donantes diferentes y demostraron que el porcentaje de edición génica se mantenía a lo largo de toda la diferenciación eritroide *ex vivo* de células CD34<sup>+</sup> de mPB sometidas a edición con ARNg de SD2 (figura 20B) y células CD34<sup>+</sup> de mPB sometidas a edición con ARNg de SPY101 (figura 20A). Los datos son la media  $\pm$  DE.

55 También se analizó el ADN genómico mediante análisis de ddPCR para detectar la frecuencia de delección de 4,9 kb con tratamiento con SD2. Los resultados presentados en la figura 20B eran de 4 donantes diferentes y demostraron que el porcentaje de edición génica se mantenía a lo largo de toda la diferenciación eritroide *ex vivo* de células CD34<sup>+</sup> de mPB sometidas a edición con ARNg de SD2 (figura 20B). Los datos son la media  $\pm$  DE.

#### Análisis de PCR en tiempo real cuantitativa

60 Se aisló ARNm y se sometió a prueba para cada una de las muestras de células CD34<sup>+</sup> de mPB humanas sometidas a edición génica que se hicieron crecer en medio de diferenciación. Se realizó el aislamiento de ARNm en los días 11 y 15 tras la diferenciación. Se determinó la expresión de globina (razón de  $\gamma/\alpha$  y razón de  $\gamma/(\gamma+\beta)$ ) mediante PCR en tiempo real cuantitativa y se comparó para cada una de las células CD34<sup>+</sup> de mPB humanas sometidas a edición con ARNg de SD2 y células CD34<sup>+</sup> de mPB humanas sometidas a edición con ARNg de SPY101 (figuras 21A-D). Los resultados presentados en las figuras 21A-D eran de 4 donantes diferentes y demostraron un aumento en el transcrito



de  $\gamma$ -globina en células CD34+ de mPB humanas sometidas a edición con ARNg de SD2 y células CD34+ de mPB humanas sometidas a edición con ARNg de SPY101 en comparación con el control negativo. Los datos son la media  $\pm$  DE.

## 5 FACS / CL-EM

Se hicieron crecer células CD34+ de mPB humanas sometidas a edición con ARNg de SD2 y células CD34+ de mPB humanas sometidas a edición con ARNg de SPY101 en medio de diferenciación durante 15 días. También se sometieron células CD34+ de mPB humanas a edición con ARNg dobles de exón 2 de BCL11A (Ex2) o ARNg de GFP y se hicieron crecer en medio de diferenciación durante 15 días. Algunas células CD34+ de mPB humanas no se sometieron a edición con ningún componente de edición de CRISPR/Cas9 (muestra de electroporación simulada) y algunas células CD34+ de mPB humanas no se sometieron a electroporación (sin EP). Se tiñeron las células vivas con glicoforina A, un marcador de maduración eritroide. Después se fijaron las células y se permeabilizaron. Se tiñeron las células fijadas con anticuerpo conjugado con fluoróforo para cada subunidad de globina. Después se analizaron las células teñidas mediante FACS, un ejemplo de  $\gamma$ -globina se representa en la figura 22A. La intensidad de fluorescencia media promedio para  $\gamma$ -globina a partir de 4 donantes diferentes se representa en la figura 22B (media  $\pm$  EEM) y demostró una regulación por incremento de  $\gamma$ -globina en células CD34+ de mPB humanas sometidas a edición con ARNg de SD2 y células CD34+ de mPB humanas sometidas a edición con ARNg de SPY101.

Se hicieron crecer células CD34+ de mPB humanas sometidas a edición con ARNg de SD2 y células CD34+ de mPB sometidas a edición con ARNg de SPY101 en medio de diferenciación durante 15 días. También se sometieron células CD34+ de mPB humanas a edición con ARNg dobles de exón 2 de BCL11A (Ex2) o ARNg de GFP y se hicieron crecer en medio de diferenciación durante 15 días. Algunas células CD34+ de mPB humanas no se sometieron a edición con ningún componente de edición de CRISPR/Cas9 (muestra de electroporación simulada) y algunas células CD34+ de mPB humanas no se sometieron a electroporación (sin EP). Se usó cromatografía de líquidos – espectrometría de masas (CL-EM) para detectar monómeros de globina desnaturalizados (figuras 23A-D). Los resultados presentados en las figuras 23A-D eran de 4 donantes diferentes y también demostraron una regulación por incremento de  $\gamma$ -globina en células CD34+ de mPB sometidas a edición con ARNg de SD2 y células CD34+ de mPB sometidas a edición con ARNg de SPY101. Los datos son la media  $\pm$  DE.

## 30 Ejemplo 19 – Pruebas de ARN guía preferidos en células para determinar actividad inespecífica

Aunque la edición específica del genoma es fundamental para una terapia satisfactoria, la detección de cualquier acontecimiento de edición inespecífica es un componente importante de garantizar seguridad del producto. Un método para detectar modificaciones en sitios inespecíficos implica el enriquecimiento de regiones del genoma que son las más similares al sitio específico mediante secuenciación de captura de híbridos y cuantificación de cualquier indel que se detecte.

La secuenciación de captura de híbridos es un método que cuantifica ediciones inespecíficas en células sometidas a edición con CRISPR-Cas9 y ADN. Los detalles relacionados con el método de secuenciación de captura de híbridos son los siguientes:

### Materiales y métodos

#### 45 Materiales y fuentes

##### 1.1.1. ADN genómico

Dado que el propósito de este método es determinar si se ha producido edición mediante CRISPR-Cas9 en sitios inespecíficos en el genoma, normalmente se usan al menos dos muestras de entrada, muestras tratada y de control (sin tratar, sometida a electroporación simulada, etc.). De cada muestra se extrae ADN genómico (ADNg) mediante un método apropiado y se hibrida ese ADNg con las bibliotecas de captura de híbridos (1.1.2) seguido por el resto del protocolo tal como se describe a continuación.

##### 55 1.1.2. Bibliotecas de captura de híbridos

Se generan bibliotecas de captura de híbridos tal como se describe en el punto (1.2.2) proporcionando una lista de hasta 57.000 secuencias de cebo de oligonucleótidos de 120 meros que después se sintetizan como kit de captura de híbridos personalizado de SureSelect XT.

#### 60 1.2 Métodos

##### 1.2.1. Algoritmos de detección de sitios inespecíficos

65 Para determinar los sitios que es más probable que tengan edición inespecífica, se usaron varios algoritmos con

diferentes características para garantizar que se cubría una amplia gama de sitios inespecíficos.

#### 1.1.1.1. CCTop

Para una secuencia guía dada, CCTop usa el algoritmo de mapeo de secuencias Bowtie 1 para buscar en el genoma para detectar sitios inespecíficos con hasta 5 apareamientos erróneos entre el sitio y la guía. A estos sitios se les denomina "sitios inespecíficos homólogos" (en vez de "sitios inespecíficos predichos") dado que solo se usa homología de secuencia para determinar los posibles sitios inespecíficos en el genoma. Estos 5 apareamientos erróneos están limitados a no más de 2 apareamientos erróneos en la región de semilla de alineación de 5 bases más cerca del extremo de PAM de las secuencias. El algoritmo CRISPOR (1.2.1.2) no tiene la limitación en la región de semilla y, por tanto, complementa a CCTop.

#### 1.2.1.1. COSMID

Dado que algunos sitios de escisión de Cas9 inespecíficos pueden tener indeles cortos (también denominados protuberancias) entre sí mismos y la guía, también se buscó con el algoritmo COSMID que puede detectar sitios inespecíficos con indeles (normalmente limitado hasta 2 indeles) y, por tanto, complementa la búsqueda realizada con CCTop.

#### 1.2.1.2. CRISPOR

CRISPOR es una herramienta que implementa muchas funciones de puntuación específicas e inespecíficas de CRISPR publicadas con el fin de comparar diversos métodos. Usa el algoritmo BWA para buscar secuencias guía con respecto al genoma para hallar sus sitios inespecíficos. Esto difiere del algoritmo Bowtie 1 usado en CCTop y permite una búsqueda que es ligeramente más permisiva ya que los apareamientos erróneos cerca de la región de PAM no están limitados a 2 de cada 5 bases como en CCTop.

#### 1.2.1.3. PAM

Por defecto, se realizan exámenes con una búsqueda de guías con PAM de NGG o NAG ya que tienen alguna de la mayor actividad. Exámenes de fase posterior pueden incluir más PAM para garantizar que no se omite ningún sitio inespecífico, ni siquiera aquellos con muy baja actividad.

#### 1.2.1.4. Combinación de algoritmos

Las guías emitidas por cada algoritmo se juntan entre sí para eliminar sitios inespecíficos idénticos y se alimentan al componente de diseño de cebo de captura de híbridos.

#### 1.2.2. Cebos de captura de híbridos

##### 1.2.2.1 Diseño

Después se usa la lista de sitios producidos por los algoritmos de detección de sitios inespecíficos (1.2.1.) para generar sondas de captura de híbridos que enriquecerán cada uno de los sitios inespecíficos en las muestras de ADN de entrada. Aunque un cebo puede ser suficiente para enriquecer satisfactoriamente una secuencia de ADN diana, generalmente se diseñan varios cebos y se colocan a lo largo del sitio diana (figura 24) con el fin de hacer que sea más probable que un cebo extraiga específicamente una región diana aunque esté flanqueada en un lado por una secuencia repetitiva que puede ser difícil que se una de manera específica. Los cebos de captura de híbridos (120 meros, porciones de color oscuro) se colocan sobre un cebo (20 meros, porción clara designada mediante \*) (figura 24).

#### 1.2.3. Secuenciación

Tras el enriquecimiento de captura de híbridos, se realiza la secuenciación en un secuenciador Illumina HiSeq con lecturas de 125 pb de extremos emparejados y un tamaño de inserción de 175 pb. La secuenciación se realiza normalmente para seleccionar como diana una profundidad de cobertura que selecciona como diana tener 5 lecturas detectadas a partir de un acontecimiento de frecuencia mínima. Para detectar, por ejemplo, acontecimientos del 0,5 %, se realiza una secuenciación con una cobertura de 1000x de modo que un acontecimiento del 0,5 % puede tener 5 lecturas.

#### 1.2.4. Eficacia de cebo

En un experimento típico, se encuentra que los cebos cubren la gran mayoría de los sitios diana con altos niveles de cobertura de secuenciación. Hay algunas limitaciones en cuanto a la cobertura de secuenciación que puede lograrse mediante métodos de secuenciación de nueva generación (NGS) debido a: alto o bajo % de GC, secuencias de baja complejidad, baja afinidad de cebo, ausencia de especificidad de cebo y otros motivos. La potencia real para detectar

indeles en un experimento se estima calculando la potencia de muestreo de diferente cobertura de secuenciación para sitios con diferentes frecuencias de indeles verdaderos. Generalmente, una cobertura de secuencia aumentada proporciona potencia aumentada para detectar sitios con indeles de baja frecuencia. Por ejemplo, si un sitio tiene una cobertura de secuenciación de 2500x, la captura de híbridos tendrá una potencia del 99 % para ver sitios con una frecuencia de indeles del 0,4 %, y una potencia del 94 % para ver sitios con una frecuencia de indeles del 0,3 % (figura 25).

#### 1.2.5. Cuantificación

Se alinean datos de secuenciación con el algoritmo BWA usando parámetros por defecto con respecto a la versión de genoma humano hg38. Para cada posible sitio inespecífico, se cuentan todos los indeles dentro de 3 pb del posible sitio de escisión de Cas9 y se dividen entre la cobertura en el sitio de corte y, por tanto, se proporciona una cantidad de indeles en un sitio de corte particular.

#### 1.2.6. Evaluación estadística de sitios de corte significativos

Diversos acontecimientos pueden conducir a indeles que no son un resultado de que se detecte CRISPR-Cas9 en sitios a lo largo de todo el genoma: polimorfismos o variantes de indeles de línea germinal, regiones propensas a roturas genómicas, regiones con tramos homopoliméricos y regiones que son de otro modo difíciles de secuenciar.

##### 1.2.6.1. Sitios excluidos del análisis

Se excluye del análisis: cualquier sitio con un indel "de línea germinal" según cada donante (el donante tiene una frecuencia de indeles de > 30 % en cada muestra), cualquier sitio del cromosoma Y en muestras de mujeres, y cualquier sitio con cobertura 0.

##### 1.2.6.2. Prueba estadística

Para evaluar si un indel observado en un posible sitio inespecífico es realmente un acontecimiento inducido por CRISPR-Cas9, se somete a prueba si las muestras tratadas con Cas9 y guía tienen una frecuencia de indeles significativamente superior a las muestras sin tratar usando tanto la prueba de Mann-Whitney-Wilcoxon como la prueba de la t de Student. Si cualquiera de estas pruebas es significativa ( $p < 0,05$ ) se considera que el sitio está etiquetado para su seguimiento con PCR para determinar si hay edición significativa. Para garantizar que se etiquetan sitios para su seguimiento de la manera más agresiva posible, no se realiza corrección de pruebas de múltiples hipótesis, lo cual reduciría el número de sitios que se encuentran significativos.

También se establece un análisis de control negativo, en el que se repite el análisis, excepto porque se buscan sitios con frecuencia de indeles en la muestra sin tratar superior a la muestra tratada. Desde el punto de vista biológico, no hay ningún motivo por el que se esperaría encontrar "resultados positivos verdaderos" en este análisis, lo cual proporciona información empírica sobre el número de falsos positivos que puede esperarse encontrar en este conjunto de datos que pueden ser atribuibles a ruido de fondo. Además, puede expandirse esto para dar una distribución nula empírica aprovechando dos muestras de control negativo adicionales, incluyendo células sometidas a electroporación sin Cas9 o guía, y células sometidas a electroporación con Cas9 y una guía de GFP. Sometiendo a prueba para detectar resultados positivos en muestras que están "menos tratadas" en comparación con muestras que están "más tratadas", se determina una distribución nula empírica conservativa de resultados falsos positivos, lo cual puede usarse para informar sobre la credibilidad de los resultados positivos en el análisis original para muestras tratadas frente a sin tratar.

Se analizaron células CD34+ de mPB humanas sometidas a edición con ARNg de SD2 y células CD34+ de mPB humanas sometidas a edición con ARNg de SPY101 mediante un método de secuenciación de captura de híbridos descrito en el presente documento. Los resultados presentados en las figuras 26-27 eran de 3-4 donantes diferentes y demostraron 0 sitios inespecíficos con evidencias de corte en células CD34+ de mPB sometidas a edición génica, que se sometieron a edición génica con ARNg de SD2 (figura 27) y células CD34+ de mPB sometidas a edición génica con ARNg de SPY101 (figura 26). La frecuencia de indeles para la ronda 1 es mayor del (>) 0,5 %. La frecuencia de indeles para la ronda 2 es mayor del (>) 0,2 %.

#### Ejemplo 20 – Experimentos de injerto

Se aislaron células CD34+ de sangre periférica movilizada (mPB) humanas a partir de donantes sanos usando microperlas CliniMACS CD34 con el instrumento CliniMACS Prodigy (Miltenyi Biotec) y se cultivaron en medios CellGro® libres de suero que incluían factor de células madre humano recombinante (SCF) 100 ng/ml, ligando de Fit 3 humano recombinante (FLT3L) 100 ng/ml y trombopoyetina (TPO) 100 ng/ml. Después se sometieron las células a electroporación usando un dispositivo Maxcyte® siguiendo las instrucciones del fabricante con uno de los siguientes: un vector vacío que no contiene ningún componente de edición de CRISPR/Cas9 (muestra de electroporación simulada), con ARNg de GFP y proteína Cas9 como control negativo (GFP), con ARNg de SPY101 y proteína Cas9 (SPY101) o con ARNg de SD2 y proteína Cas9 (SD2). La proteína Cas9 recombinante codifica para Cas9 de S.

*pyogenes* flanqueada por dos secuencias de localización nuclear (NLS) de SV40. Estos experimentos se realizaron usando una ribonucleoproteína (RNP) con una razón en peso de ARNg con respecto a Cas9 de 1:1.

Se inyectó cada una de las células CD34+ de mPB sometidas a edición génica humanas a través de la vena de la cola en 16 ratones inmunodeficientes ("NSG" o NOD scid gamma -NOD) para demostrar las capacidades de asentamiento y de injerto. NSG es una raza de ratones de laboratorio endogámicos y es una de las más inmunodeficientes descritas hasta ahora; véase, por ejemplo, Shultz *et al.*, Nat. Rev. Immunol. 7(2): 118-130 (2007). En la figura 28 se presentan detalles relacionados con el experimento de injerto. A las 8 semanas tras la inyección, se extrajo sangre de los ratones NSG y se analizó la sangre periférica mediante FACS para detectar células vivas CD45RA+ humanas y CD45+ de ratón. A las 16 semanas tras la inyección, se sacrificaron los ratones NSG y se analizaron la médula ósea, el bazo y sangre periférica mediante FACS para detectar células vivas CD45RA+ humanas y CD45+ de ratón. Se observó el injerto de células CD34+ de mPB humanas en ratones NSG irradiados en todos los grupos de tratamiento a partir de los tres donantes sanos. Se detectaron células CD45RA+ humanas usando FACS en los 3 órganos hematopoyéticos de todos los donantes. Las células de control CD34+ sin transfectar mostraban porcentajes de injerto ligeramente mejores. Todos los grupos de célula transfectadas tenían porcentajes de injerto similares, incluyendo el grupo transfectado de manera simulada en los 3 donantes sanos. En general, la adición de RNP de Cas9-ARNg no afectó al injerto en comparación con el control de transfección simulada (figuras 29A-E y figura 30). Los puntos de datos en las figuras 29A-E representan ratones individuales y representan el porcentaje de células vivas que eran células CD45RA+ humanas. Los datos son la media  $\pm$  EEM.

Ejemplo 21 – Evaluación de la eficiencia y eficacia de edición de SPY101 usando RNP de Cas9

Con el fin de lograr la mayor eficacia usando CRISPR-Cas9 para tratar SCD y  $\beta$ -talasemia usando SPY101, se evaluaron dos formatos de Cas9 diferentes, ARNm de Cas9 o proteína Cas9, para determinar su eficiencia de edición, eficacia y toxicidad en células CD34+ humanas a partir de sangre periférica movilizada (mPB). Se compararon diversas fuentes para ARNm de Cas9 con respecto a proteína Cas9 mediante electroporación de ARNm de Cas9 y ARNg de SPY101 o proteína Cas9 complejada con ARNg de SPY101 (como complejo de ribonucleoproteína (RNP)) en células CD34+ de mPB humanas y se evaluaron para determinar su eficiencia de edición y viabilidad celular a las 48 horas tras la electroporación. Se compararon diversas fuentes para ARNm de Cas9 con respecto a proteína Cas9 y se encontró que, aunque pueden lograrse niveles similares de eficiencia de edición entre algunos ARNm de Cas9 con respecto a proteína Cas9 (figura 31), la mayoría tenían una viabilidad celular significativamente inferior en comparación con muestras de control (controles sin electroporación (sin EP) o sin electroporación en sustrato (EP simulada)) mostradas en las figuras 32A-B. Esto indica que RNP de Cas9 es el mejor formato para usar para un suministro eficiente de Cas9 y ARNg al interior de células CD34+ de mPB humanas.

A continuación se compararon diferentes fuentes de proteína Cas9 así como proteína Cas9 con números variables de señales de localización nuclear (NLS) en el extremo o bien N-terminal o bien C-terminal, ya que esto puede afectar a la localización eficiente de Cas9 en el núcleo para permitir la edición. Tal como se muestra en las figuras 33A-C, se encontró que la proteína Cas9 de Aldevron con una NLS en el extremo tanto N-terminal como C-terminal proporcionó la mejor eficiencia de edición sin cambios en la viabilidad celular.

A continuación, se compiló la eficiencia de edición con SPY101 examinada en diversos donantes de CD34+ de mPB humanas usando o bien ARNm de Cas9 o bien proteína Cas9 (se eligieron Feldan o Aldevron del ejemplo anterior) y se observó que la proteína Cas9 de Aldevron daba como resultado la eficacia de edición más alta (figura 34A). Además, fue posible lograr tasas similares de eficiencia de edición en la fabricación compatible con GMP a escala clínica usando proteína Cas9 (figura 34B).

A continuación se examinó la eficacia de SPY101 en varios donantes de CD34+ derivados de mPB (figuras 35A-B) o médula ósea (BM, figuras 36A-B). Se observa que, a lo largo del procedimiento de optimización, se logra una mejor eficacia, medida como expresión de  $\gamma$ -globina con respecto a globinas de tipo  $\alpha$ -globina o  $\beta$ -globina ( $\beta$ -globina +  $\gamma$ -globina) mediante PCR en tiempo real cuantitativa en células CD34+ diferenciadas eritroides.

Después se investigó si SPY101 sería eficaz en células obtenidas a partir de pacientes con SCD o  $\beta$ -talasemia. Se sometieron células mononucleares de sangre periférica a partir de donantes sanos o pacientes a electroporación con RNP de SPY101-Cas9 y se sometieron a diferenciación eritroide de manera similar a los ejemplos anteriores antes de extraer ARN para medir la expresión de  $\gamma$ -globina. Se observa que, de hecho, SPY101 era eficaz en el aumento de  $\gamma$ -globina en muestras de pacientes (figuras 37A-B).

Con el fin de entender mejor una relación de genotipo con respecto a fenotipo en células eritroides sometidas a edición con SPY101, se realizó un análisis de colonias individuales similar al ejemplo 15 con RNP de Cas9 y se encontró que esto aumentó una fracción mayor de colonias sometidas a edición bialélica usando RNP de Cas9 en comparación con ARNm de Cas9 (figuras 38A-B). Además, una descomposición detallada de colonias no sometidas a edición, perturbación monoalélica de sitio de unión a GATA1 seleccionado como diana por SPY101 y perturbación bialélica de sitio de unión a GATA1, reveló una eficacia dependiente de la dosis de SPY101, medida como aumento de  $\gamma$ -globina en comparación con células tratadas con ARNg de GFP de control (figuras 39A-B).

5 Para examinar el porcentaje de células que expresan  $\gamma$ -globina, se realizó un análisis de FACS en células CD34+ sometidas a edición con RNP de SPY101-Cas9 a partir de mPB humana. En comparación con células tratadas con ARNg de GFP de control, se observa un porcentaje superior de células diferenciadas eritroides que expresan  $\gamma$ -globina (figuras 40A-D), así como un aumento de la expresión de  $\gamma$ -globina por cada célula (figura 40E) en células tratadas con SPY101.

## REIVINDICACIONES

1. Un ARN guía de molécula simple (ARNgs) que comprende la secuencia de ácido nucleico de SEQ ID NO: 71.959, en el que el ARNgs comprende tres residuos 2'-O-metil-fosforotioato en cada uno de sus extremos 5' y 3'.
2. Un ARN guía de molécula simple (ARNgs) que comprende la secuencia de ácido nucleico de SEQ ID NO: 71.959, en el que las bases 1-3 y 97-99 son nucleótidos de 2'-O-metil-fosforotioato modificados.
3. Un ácido nucleico que comprende una secuencia de nucleótidos que codifica para el ARNgs según la reivindicación 1 o 2.
4. Un vector que comprende el ácido nucleico según la reivindicación 3.
5. El ARNgs según la reivindicación 1 o 2, el ácido nucleico según la reivindicación 3, o el vector según la reivindicación 4, para su uso en el tratamiento de un paciente con hemoglobinopatía.
6. El ARNgs para su uso según la reivindicación 5, el ácido nucleico para su uso según la reivindicación 5, o el vector para su uso según la reivindicación 5, en los que la hemoglobinopatía es anemia de células falciformes o talasemia.
7. El ARNgs para su uso según la reivindicación 6, el ácido nucleico para su uso según la reivindicación 6, o el vector para su uso según la reivindicación 6, en los que la hemoglobinopatía es una talasemia y la talasemia se selecciona del grupo que consiste en  $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\gamma$ , y combinaciones de las mismas, de manera adicionalmente opcional en los que la talasemia es  $\beta$ -talasemia.
8. Un método *ex vivo* de edición de un gen de linfoma de células B 11A (BCL11A) en una célula humana mediante edición genómica, comprendiendo el método:
 

introducir en la célula humana una o más ácido desoxirribonucleico (ADN) endonucleasas o un ácido nucleico que comprende una secuencia de nucleótidos que codifica para la una o más ADN endonucleasas o un vector que comprende dicho ácido nucleico para realizar una o más roturas de cadena sencilla (SSB) o roturas de cadena doble (DSB) dentro o cerca del gen de BCL11A u otra secuencia de ADN que codifica para un elemento regulador del gen de BCL11A, que da como resultado una delección, modulación o inactivación permanente de una secuencia de control transcripcional del gen de BCL11A,

en el que el método comprende además introducir en la célula el ARNgs según la reivindicación 1 o la reivindicación 2, un ácido nucleico que comprende una secuencia de nucleótidos que codifica para la secuencia de ácido nucleico de dicho ARNgs, o un vector que comprende dicho ácido nucleico, y en el que la una o más ADN endonucleasas comprenden una endonucleasa Cas9.
9. El método según la reivindicación 8, en el que el ácido nucleico que comprende una secuencia de nucleótidos que codifica para la una o más ADN endonucleasas es uno o más ácidos ribonucleicos (ARN).
10. El método según la reivindicación 8 o 9, en el que la una o más ADN endonucleasas comprenden, cada una, en el extremo N-terminal, el extremo C-terminal o tanto el extremo N-terminal como el extremo C-terminal, una o más señales de localización nuclear (NLS), opcionalmente en el que la una o más NLS comprenden una NLS de SV40.
11. El método según la reivindicación 8 o 10, en el que la una o más ADN endonucleasas se complejan previamente con el ARNgs para formar una o más ribonucleoproteínas (RNP), opcionalmente en el que la una o más ADN endonucleasas se combinan con el ARNgs en una razón en peso de 1:1 de ADN endonucleasa con respecto a ARNgs para formar la una o más RNP.
12. El método según la reivindicación 11, en el que la una o más RNP se suministran a la célula humana mediante electroporación, opcionalmente en el que la una o más ADN endonucleasas son una Cas9 de *S. pyogenes* que comprende una NLS de SV40 en el extremo N-terminal y una NLS de SV40 en el extremo C-terminal.
13. El método según una cualquiera de las reivindicaciones 8-12, en el que la célula humana es una célula progenitora hematopoyética, opcionalmente en el que la célula progenitora hematopoyética es una célula CD34+.

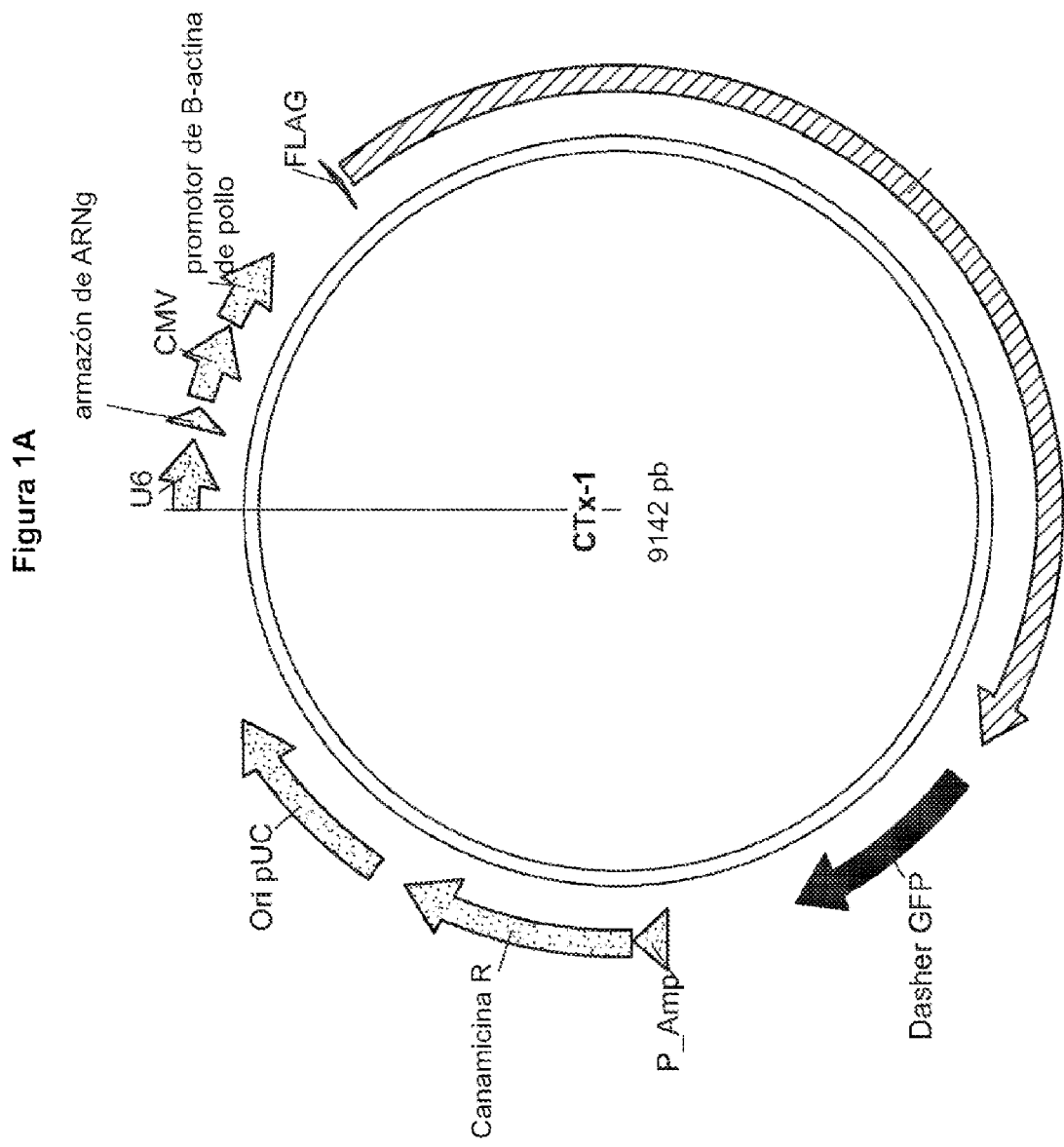


Figura 1B

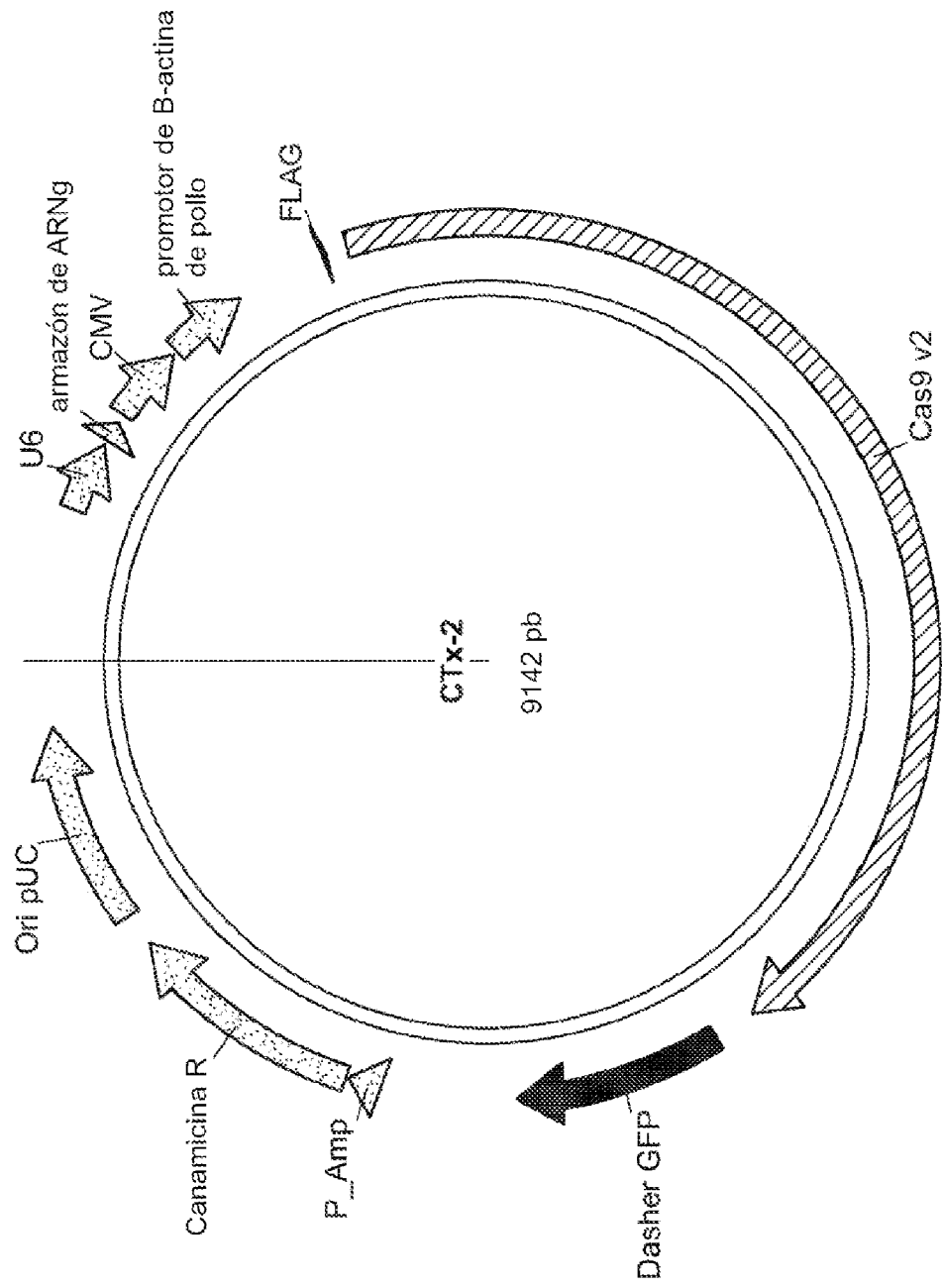
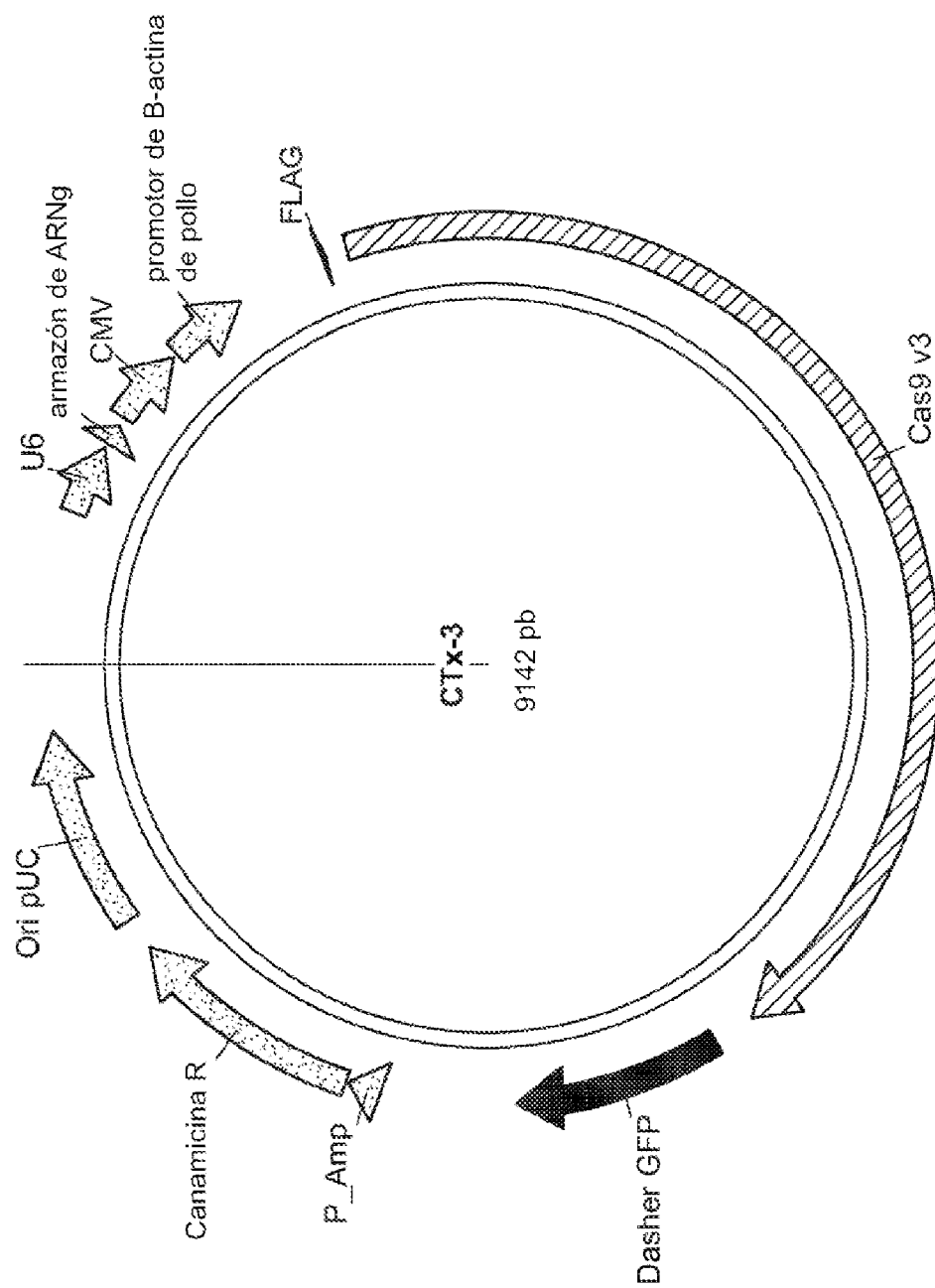
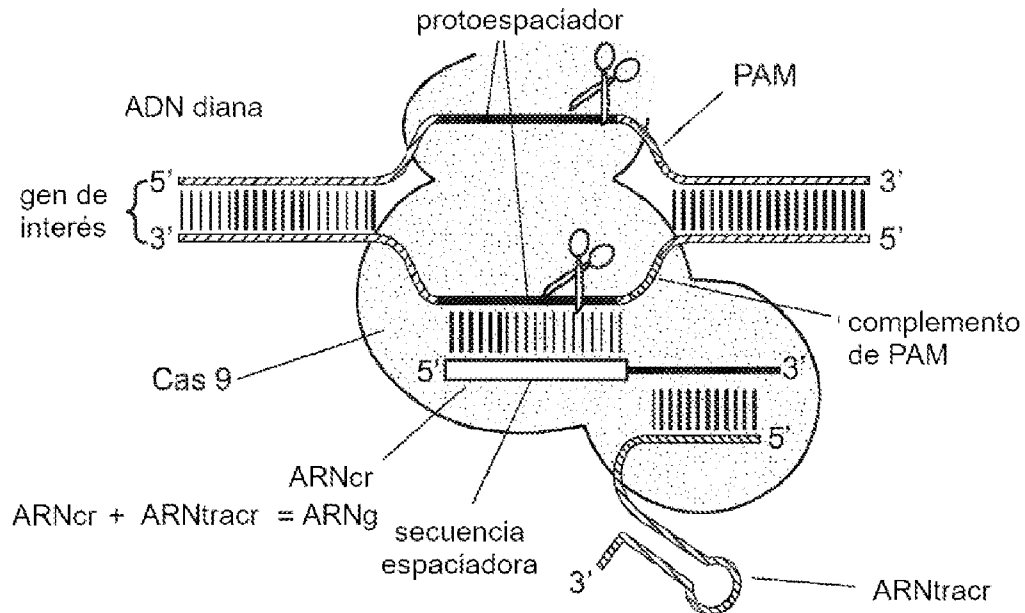




Figura 1C



**Figura 2A**



**Figura 2B**

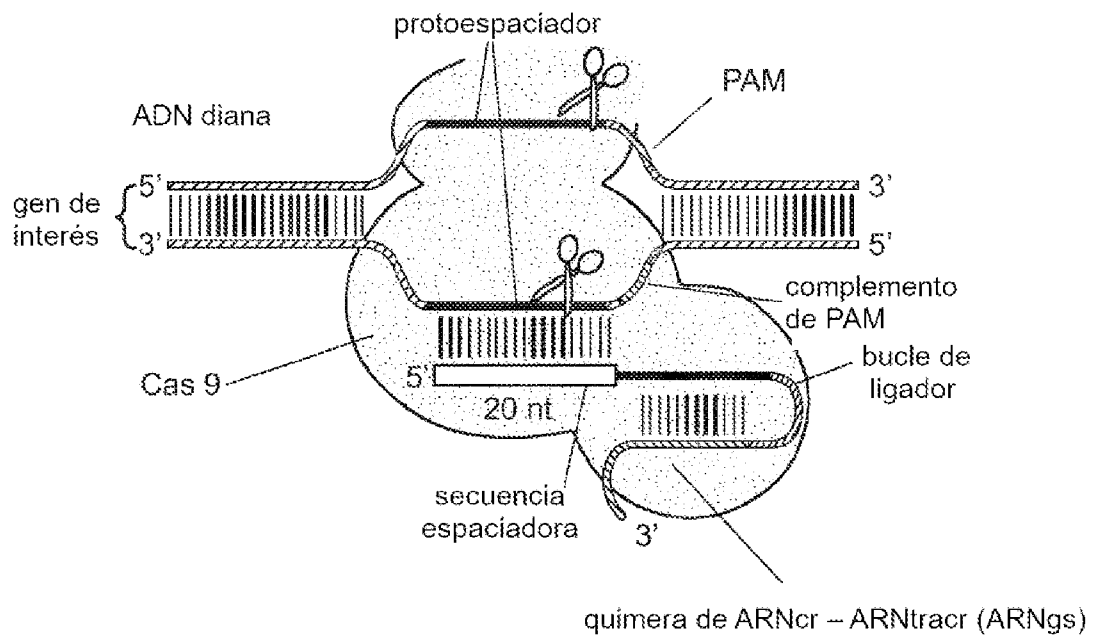


Figura 3

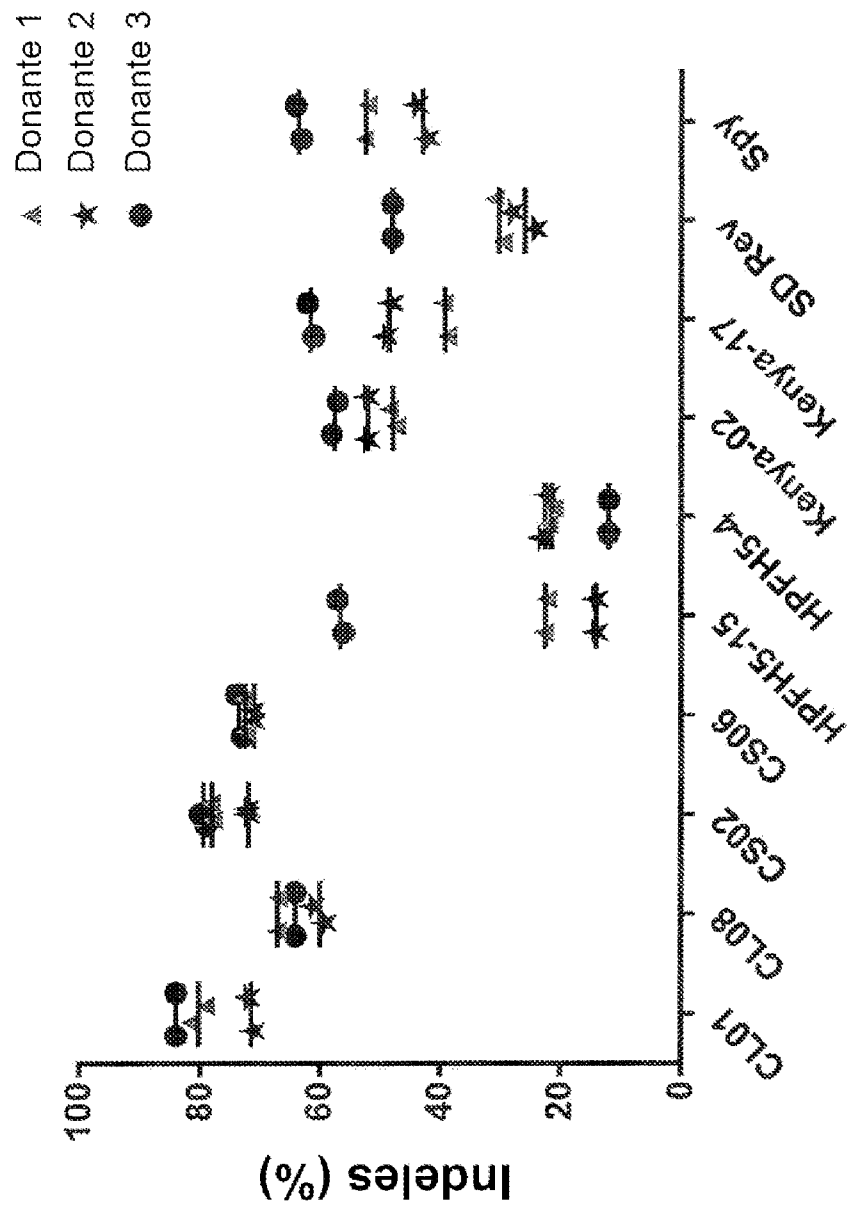


Figura 4A

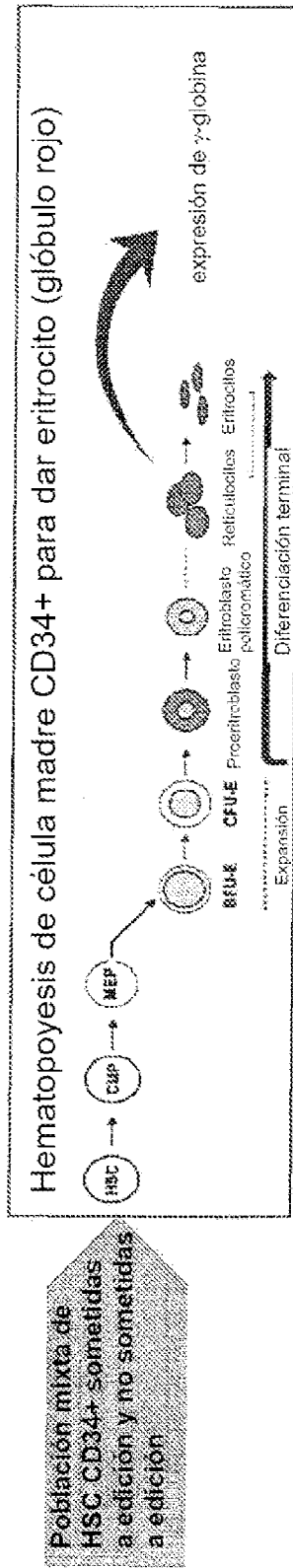


Figura 4B

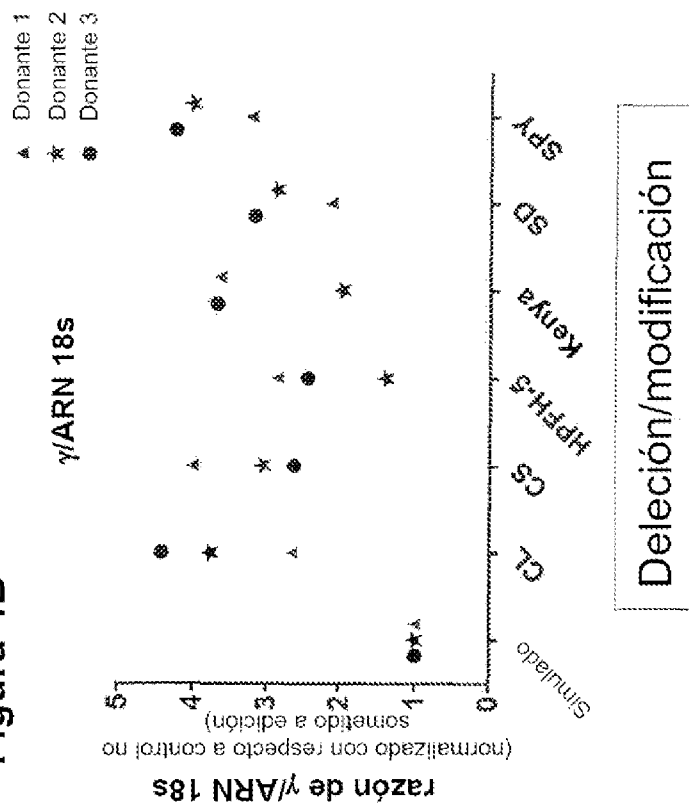
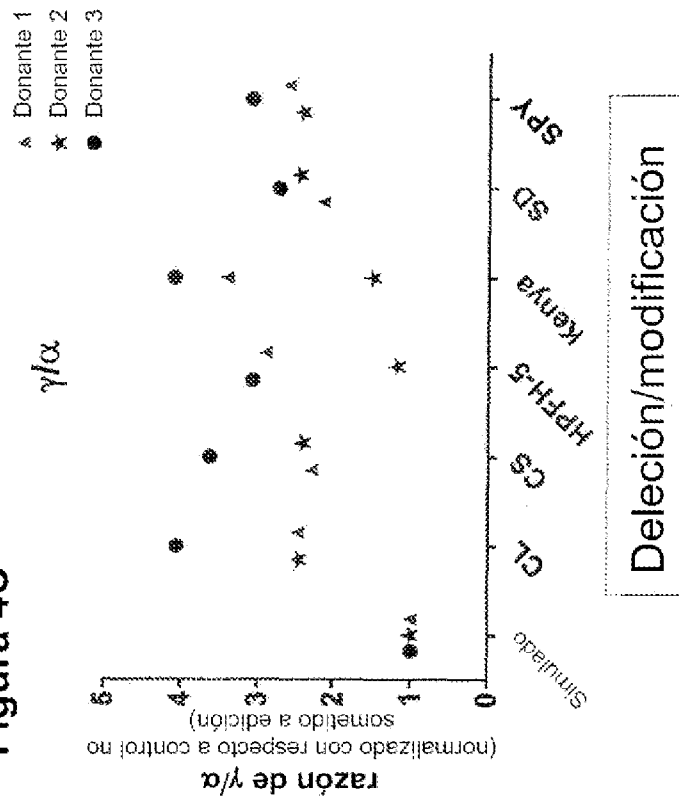
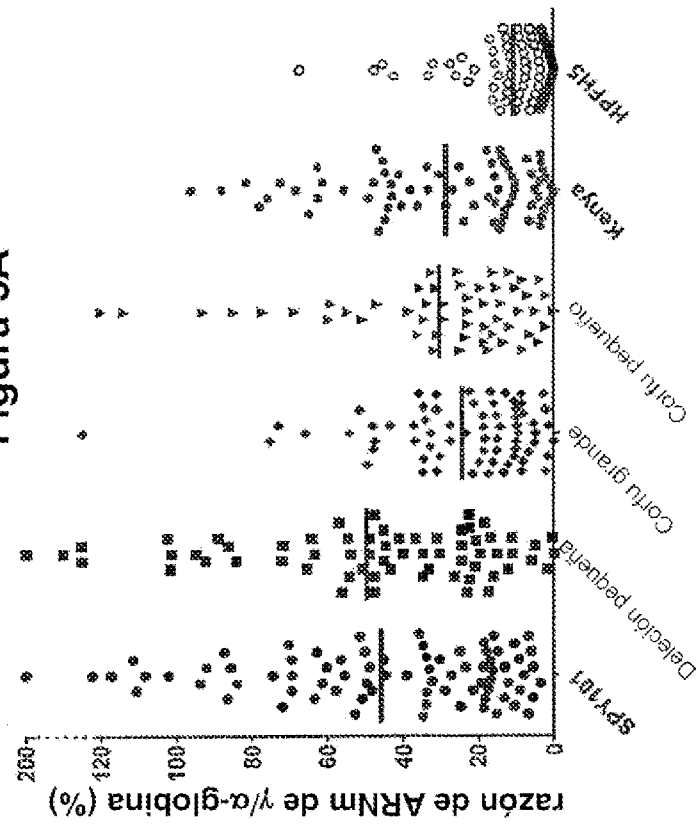


Figura 4C



Clones CD34+ tanto sometidos a edición como no sometidos a edición diferenciados para dar RBC

**Figura 5A**



**Figura 5B**

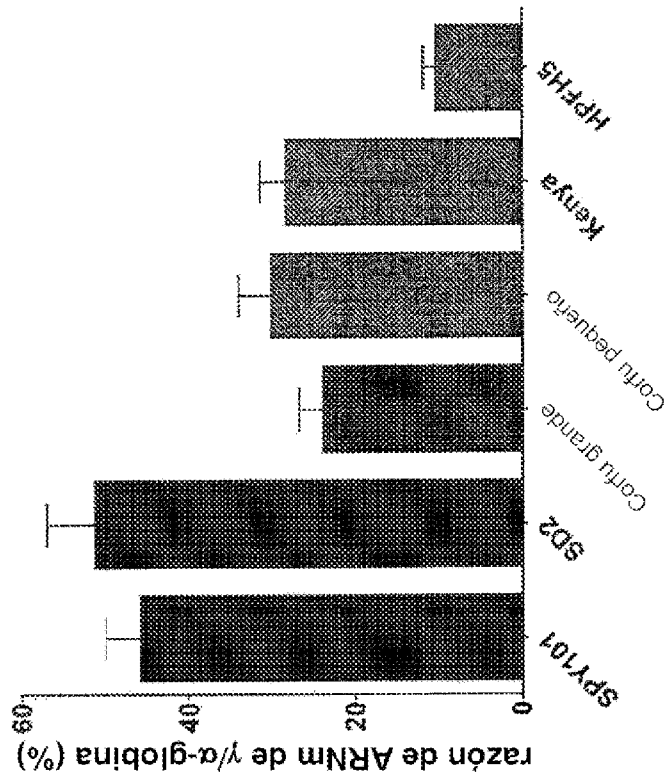
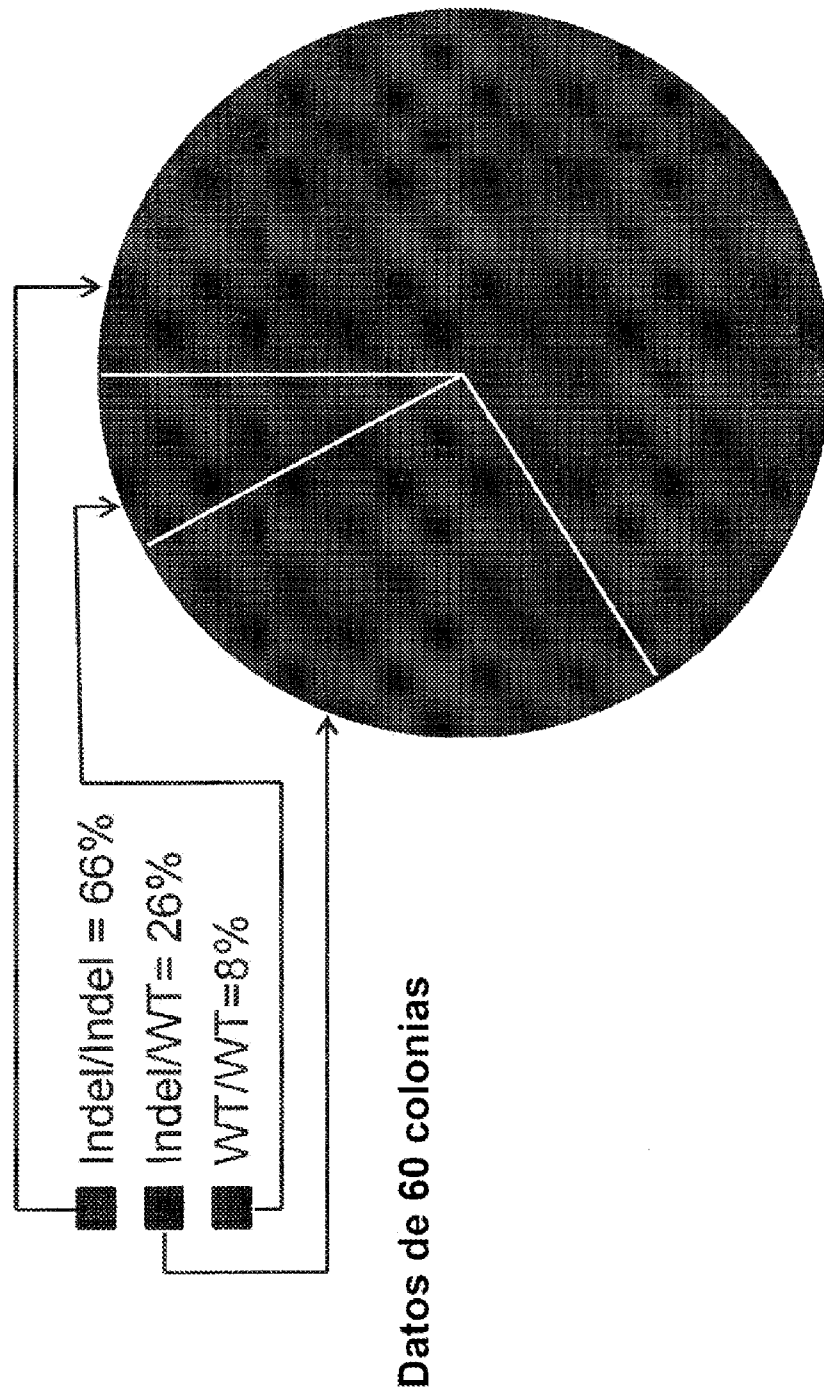


Figura 6

Tasa de edición de intrones de Bcl11a (SPY101)



# Edición de intrones de Bcl11a (SPY101)

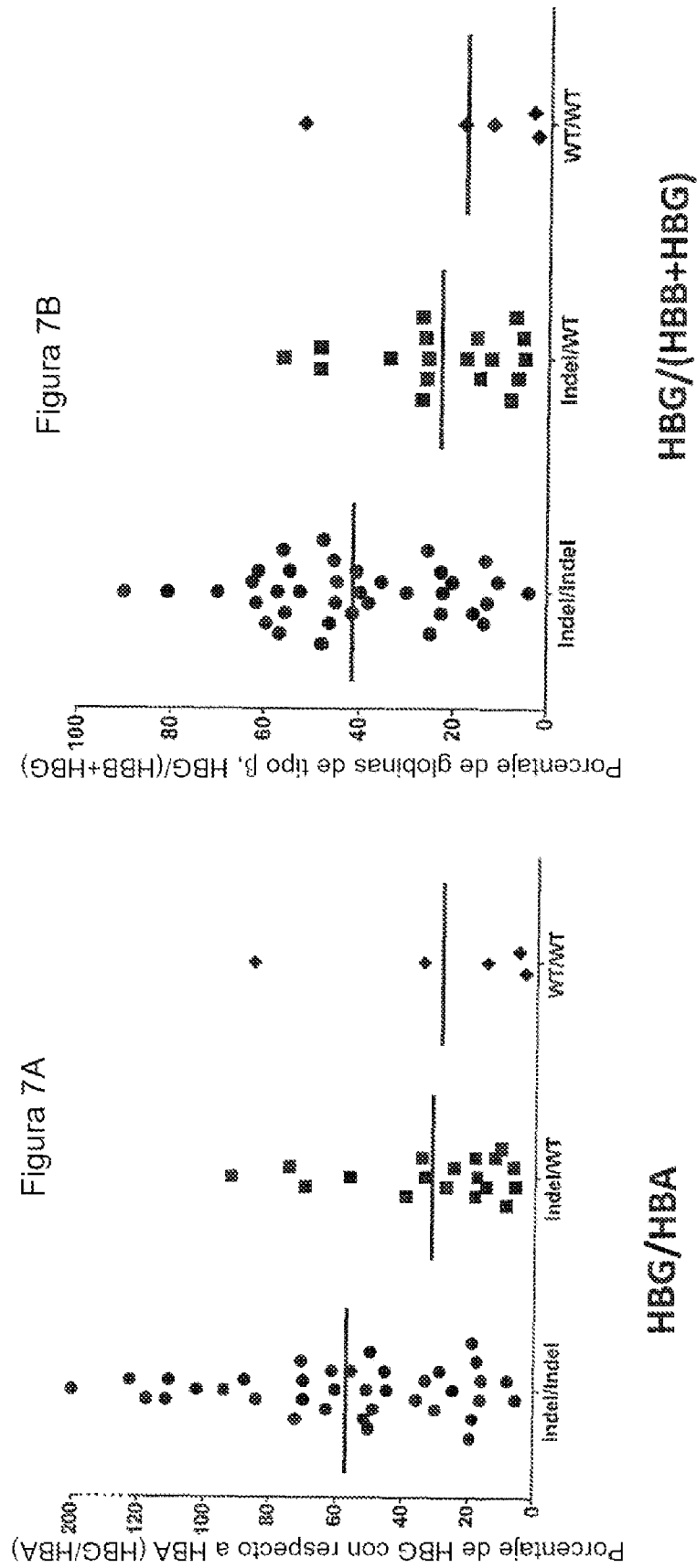


Figura 8

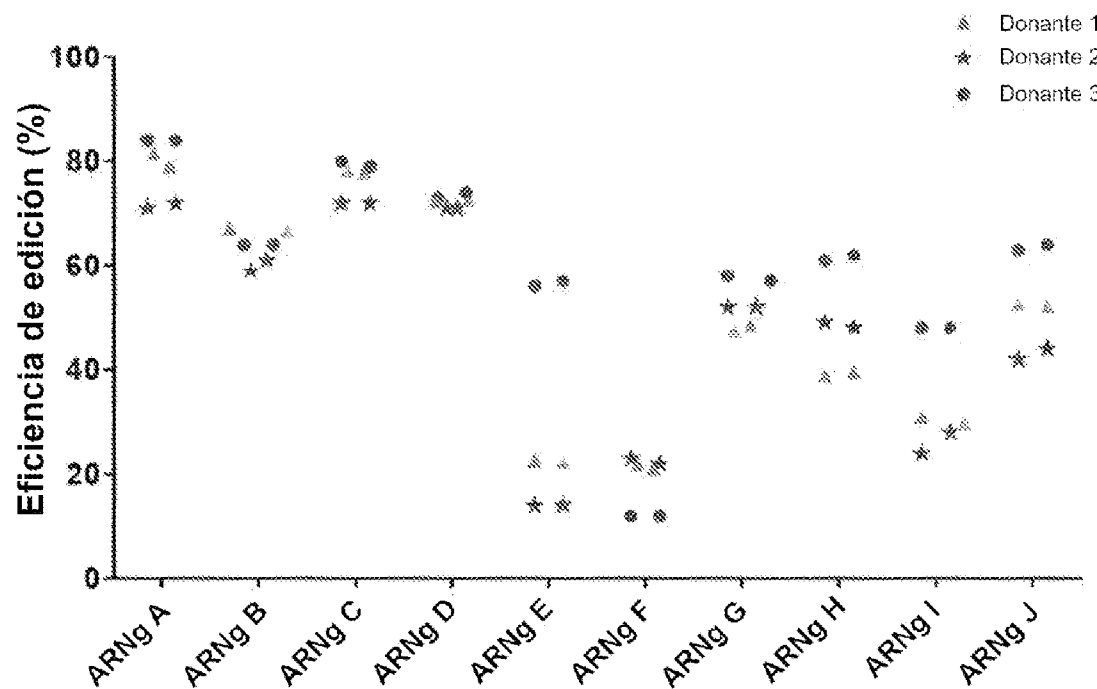




Figura 9A

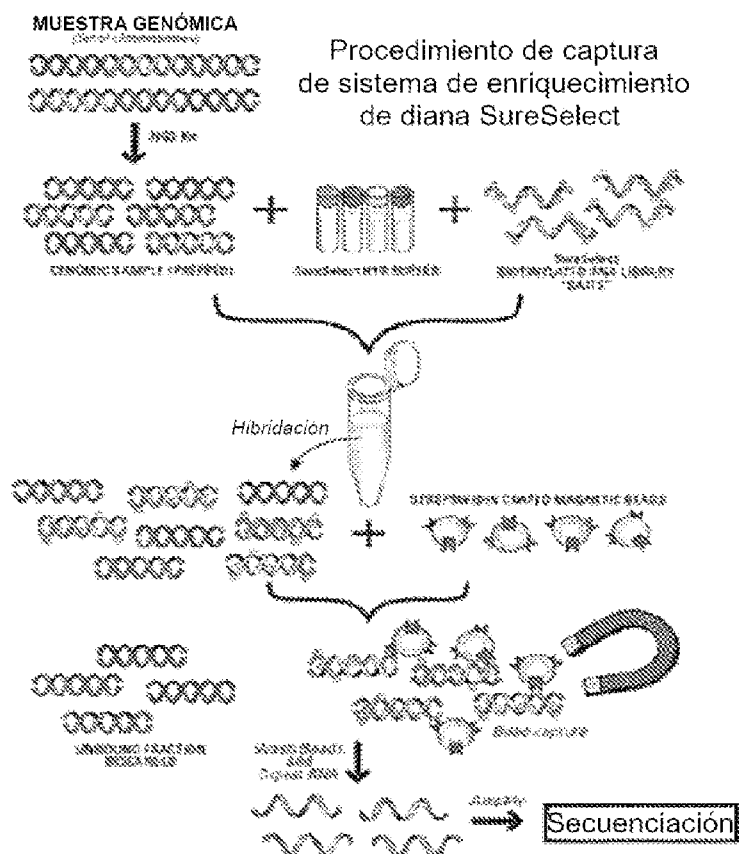


Figura 9B

ARNg	n.º de sitios de corte inespecíficos examinados	n.º de sitios de corte inespecíficos detectados	Eficiencia de escisión inespecífica máxima
A	365	0	-
B	576	0	-
D	330	0	-
E	696	0	-
F	483	0	-
H	676	0	-
I	677	0	-
J	840	0	-
C	419	8	3,6%
G	530	1	0,6%

Figura 10A

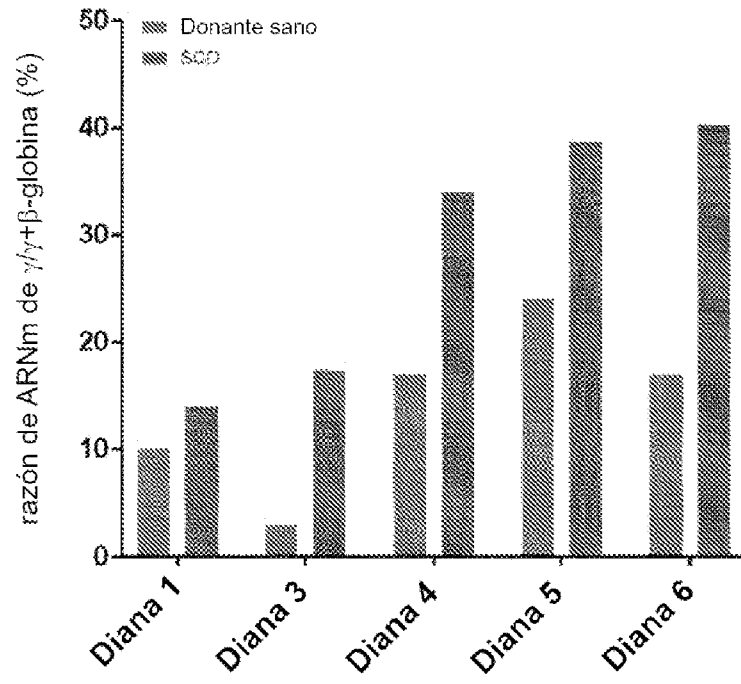


Figura 10B

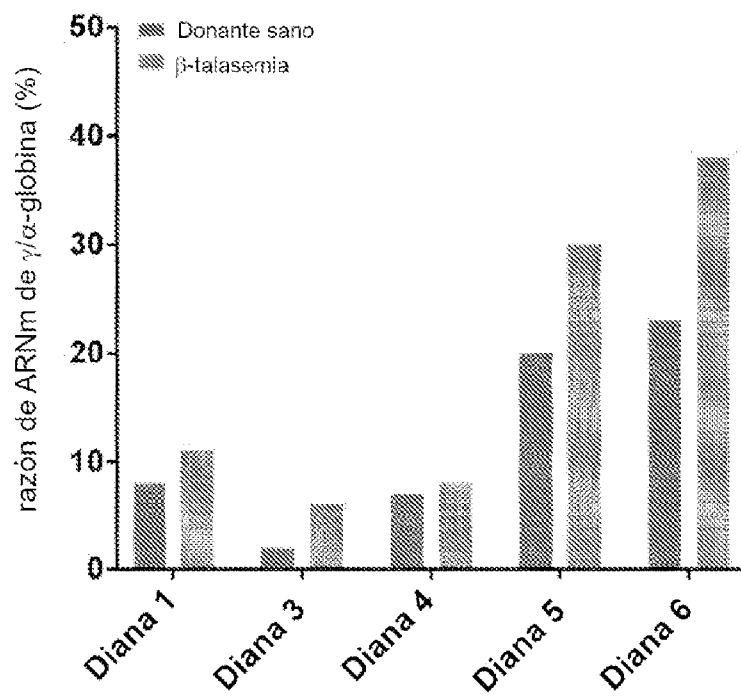


Figura 11A

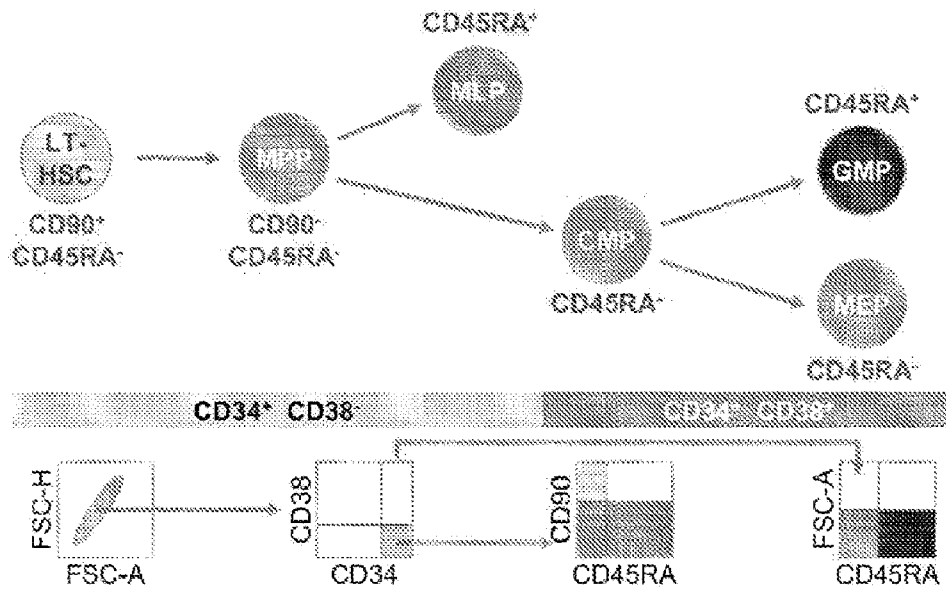


Figura 11B

La distribución de tipos de células permanece inalterada tras la edición

Población	Frecuencia	
	Simulado	Sometido a edición
Global	94,83	94,30
LT-HSC	7,82	7,98
MPP	18,97	16,38
MLP	13,23	8,47
CMP/MEP	11,23	11,32
GMP	8,39	13,26

Figura 11C

Altas eficiencias de edición en todos los tipos de células

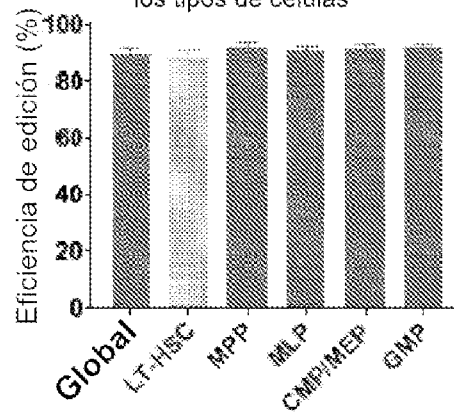


Figura 12

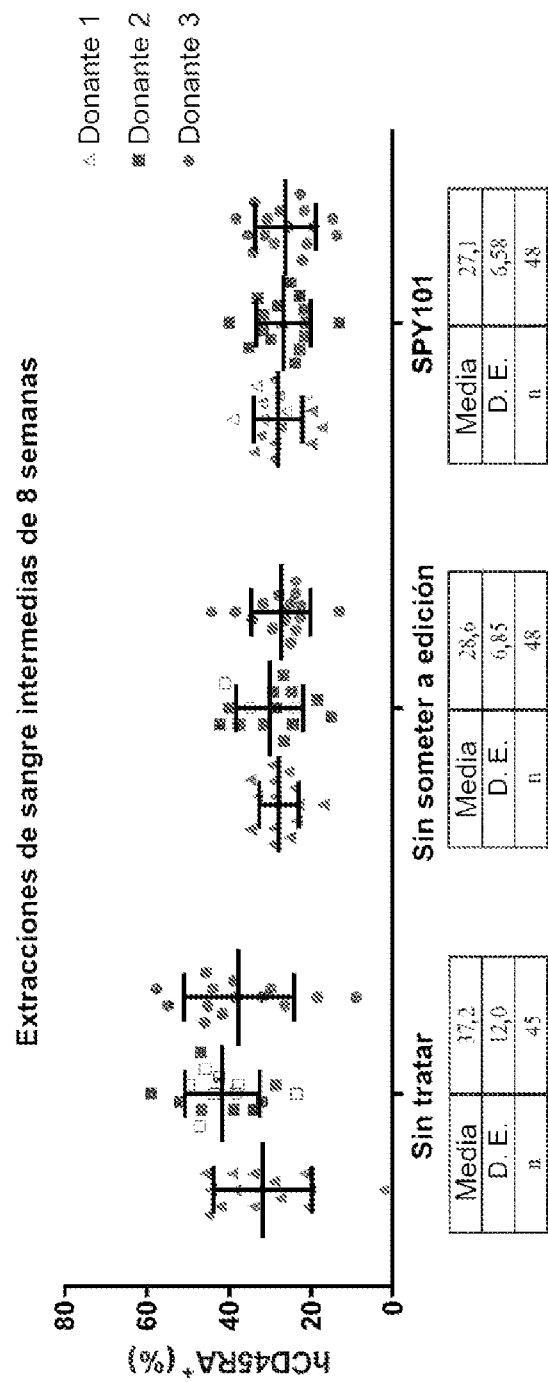


Figura 13

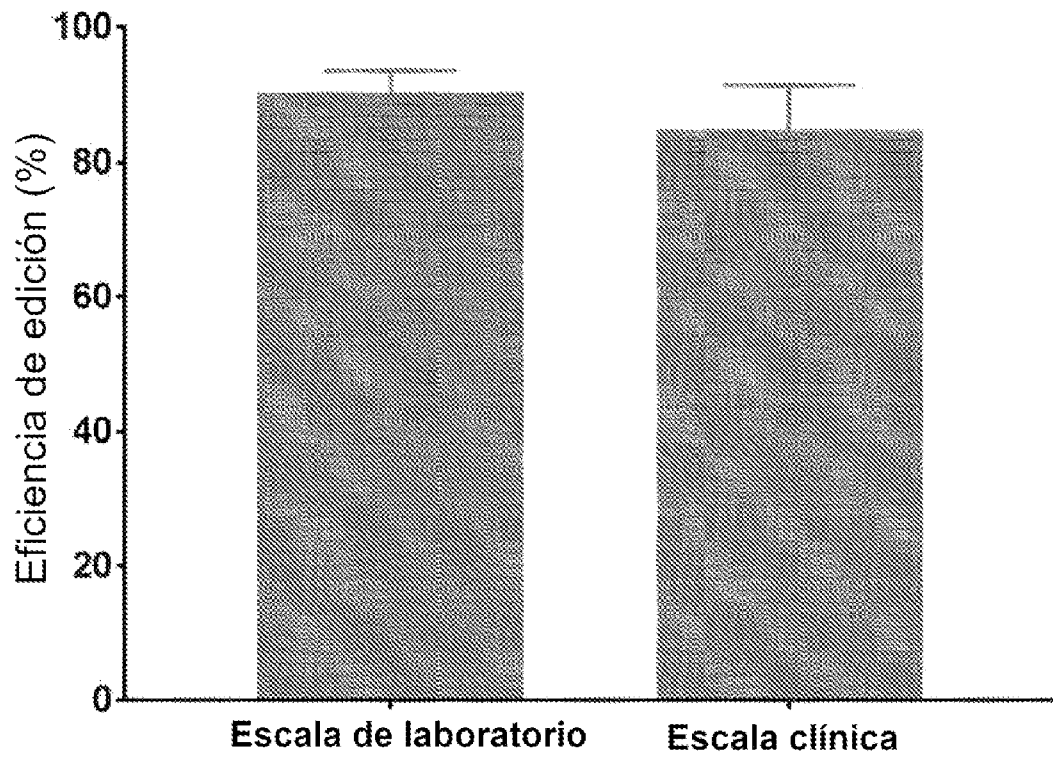
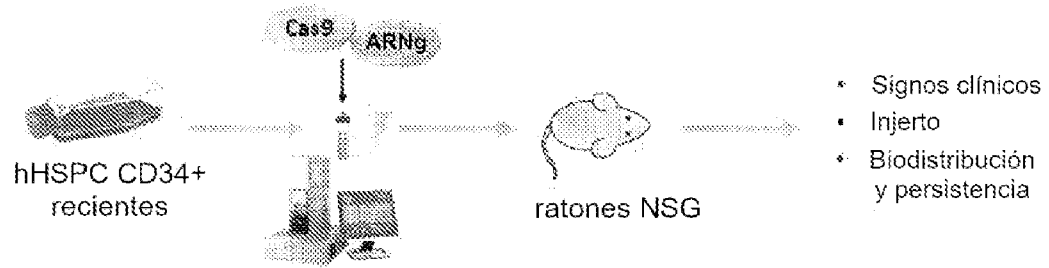


Figura 14

Estudio de biodistribución / persistencia



Estudio de tumorigenicidad / toxicología general

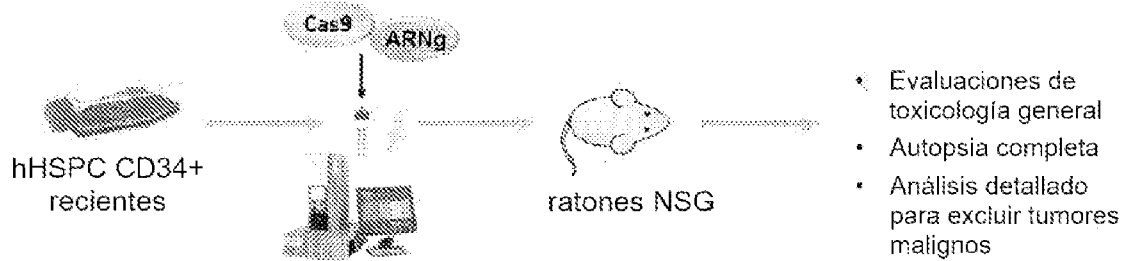


Figura 15

### Configuración experimental

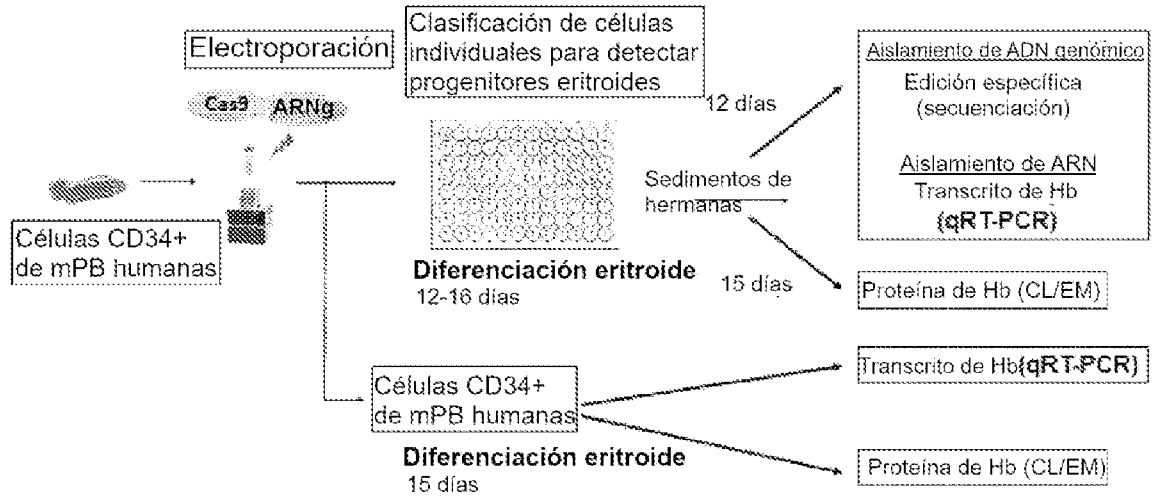


Figura 16A

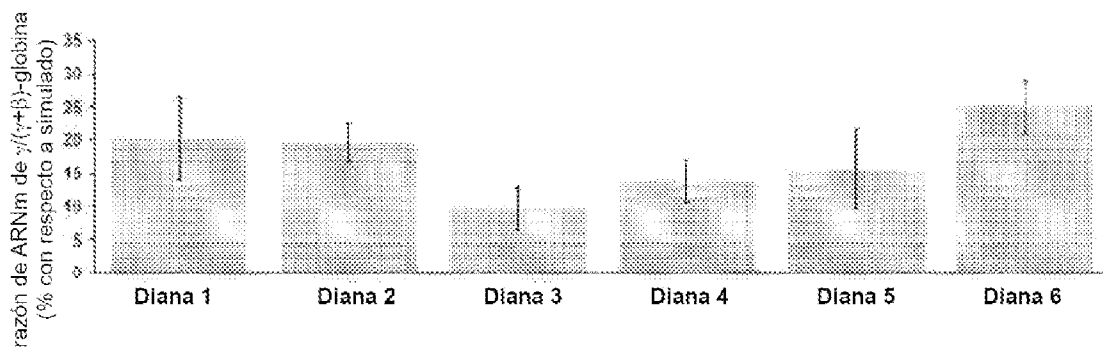


Figura 16B

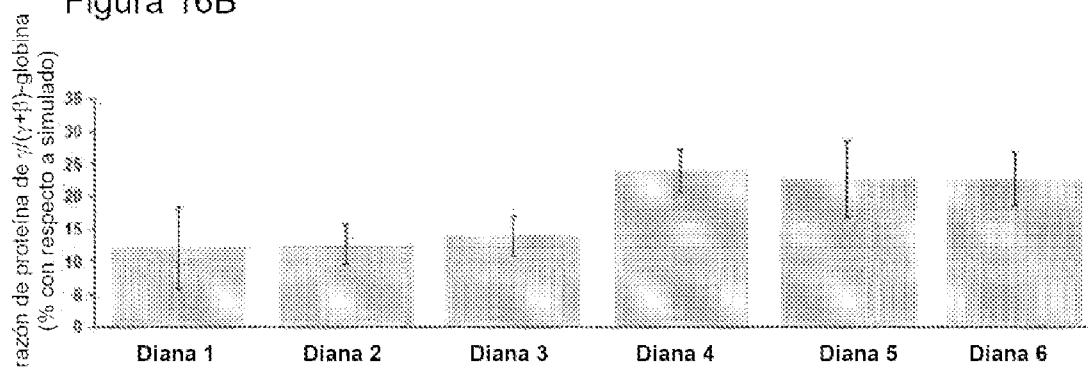




Figura 17

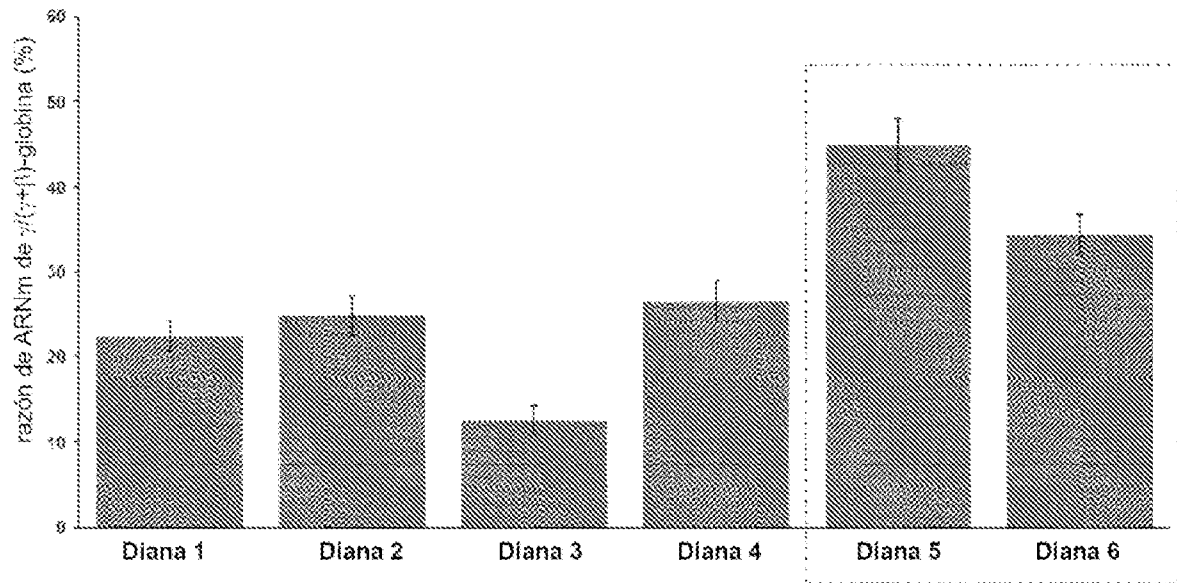


Figura 18B

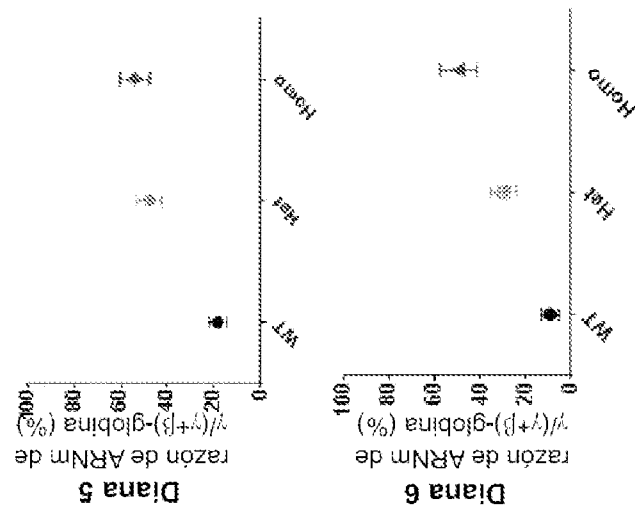


Figura 18A

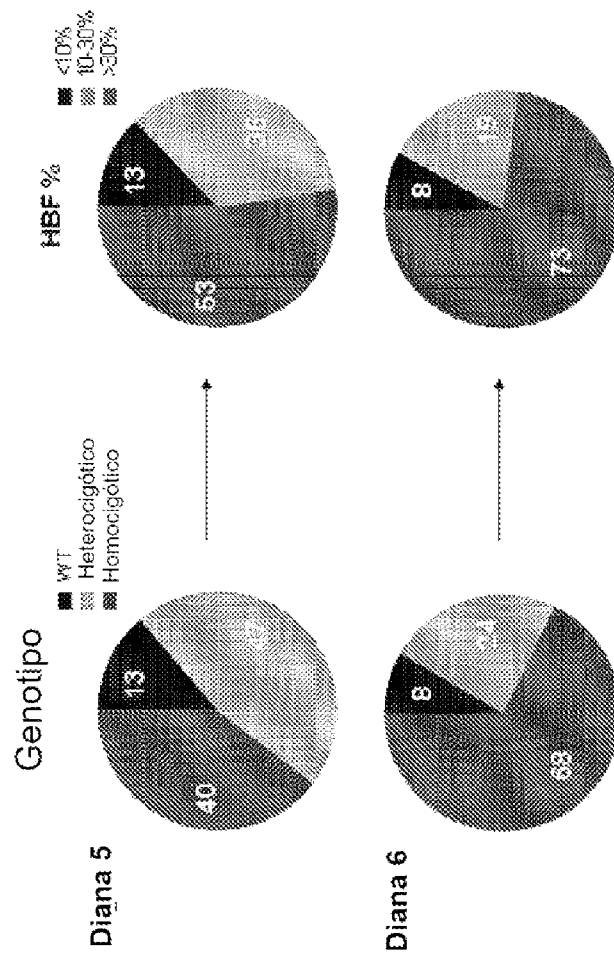
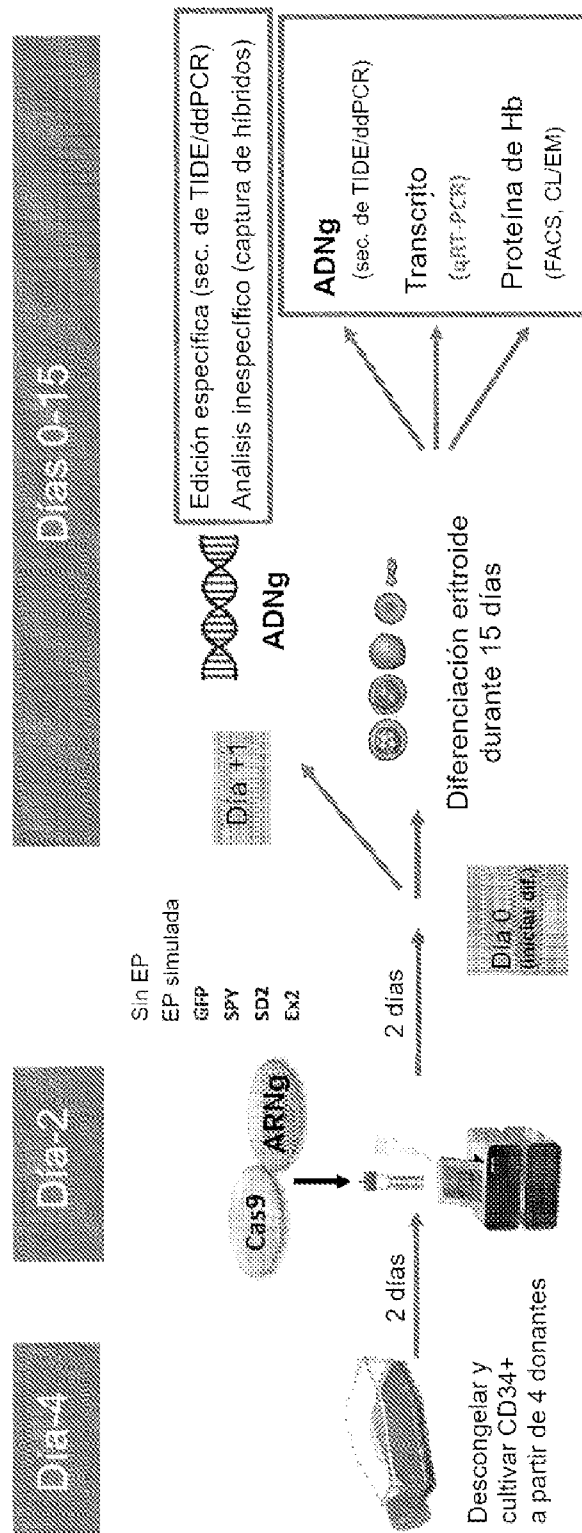


Figura 19  
Flujo de trabajo global y lista de datos globales



# % de edición mantenido a lo largo de la diferenciación eritroide ex vivo de células CD34+ de mPB

Figura 20A

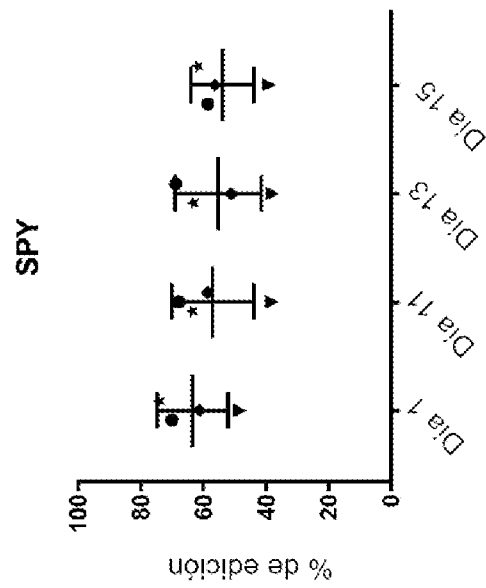
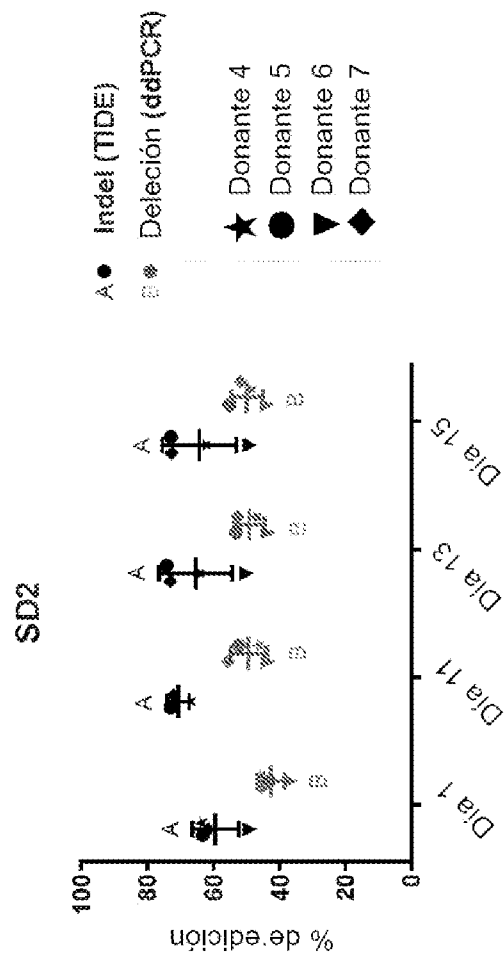


Figura 20B



# Aumento de transcrito de $\gamma$ -globina tras la edición

Figura 21A

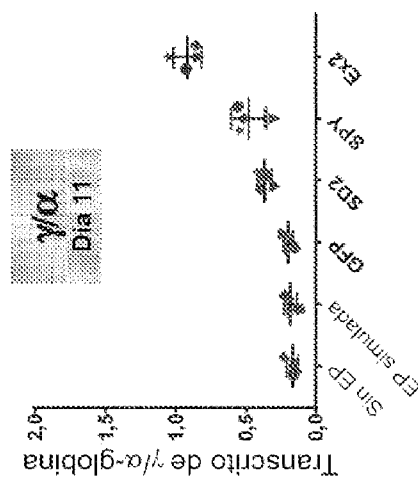


Figura 21B

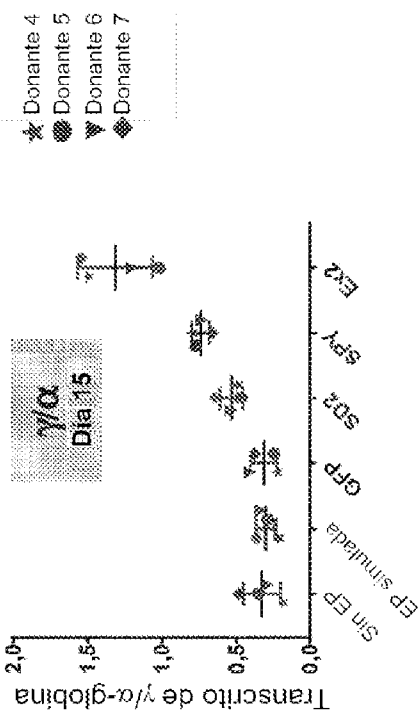


Figura 21C

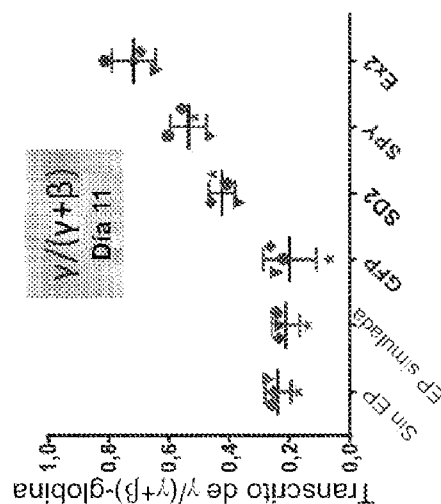
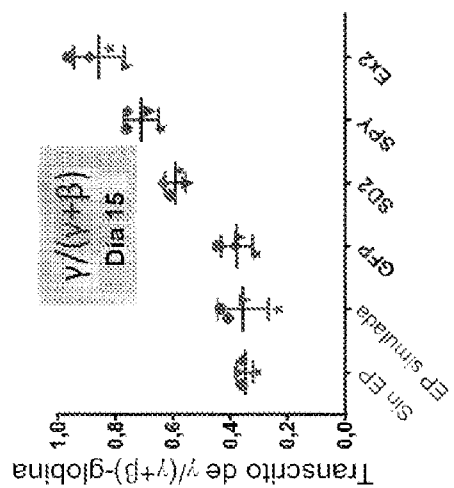


Figura 21D



El análisis de FACS muestra regulación por incremento de  $\gamma$ -globina en células sometidas a edición

Figura 22A

FACS

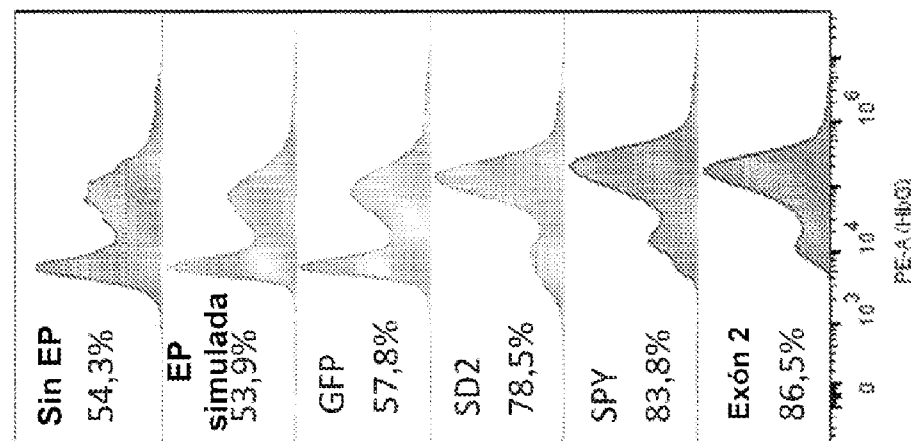
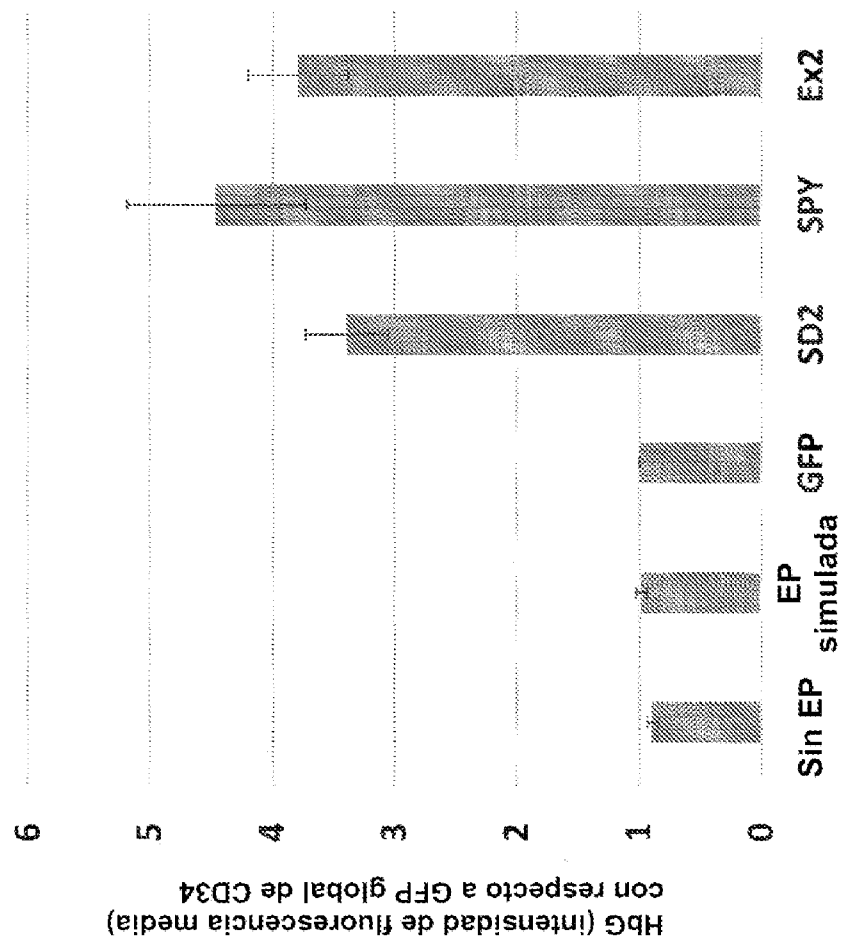


Figura 22B

MFI



Los datos de CL-EM globales validan adicionalmente la regulación por incremento de  $\gamma$ -globina tras la edición

Figura 23A

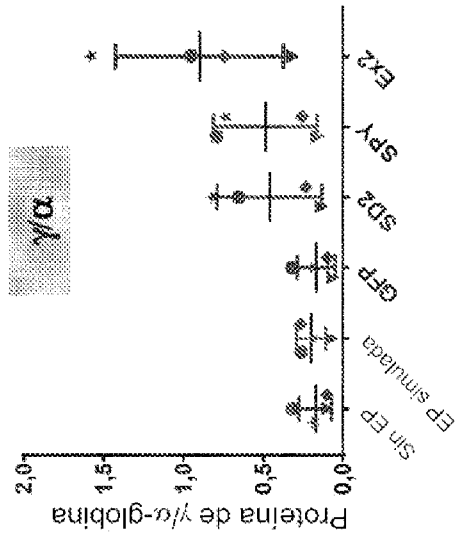


Figura 23B

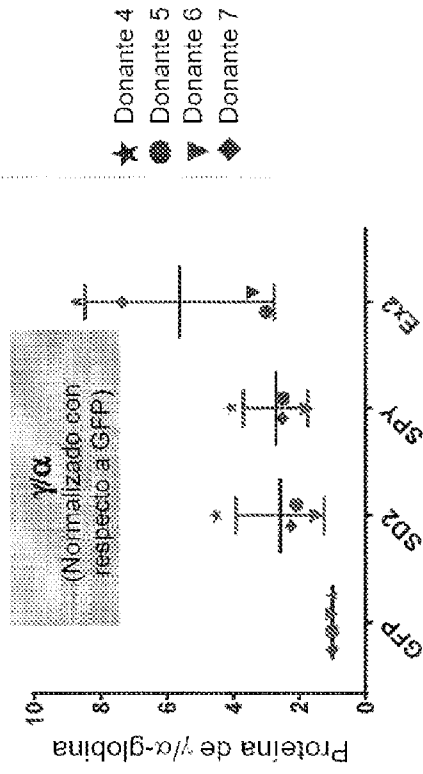


Figura 23C

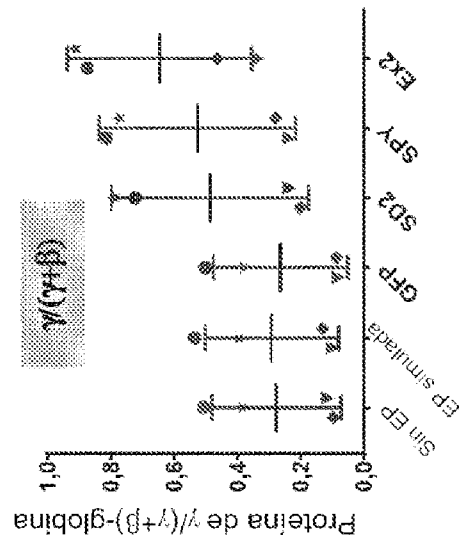
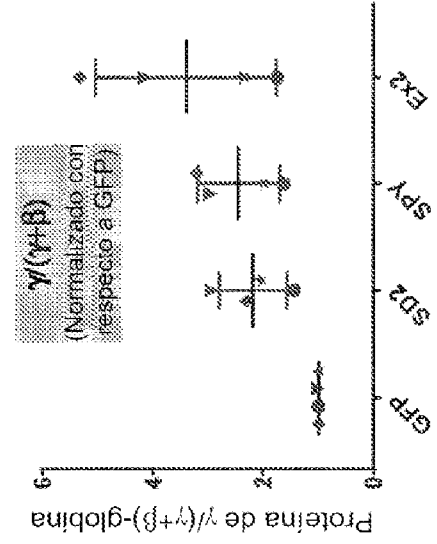
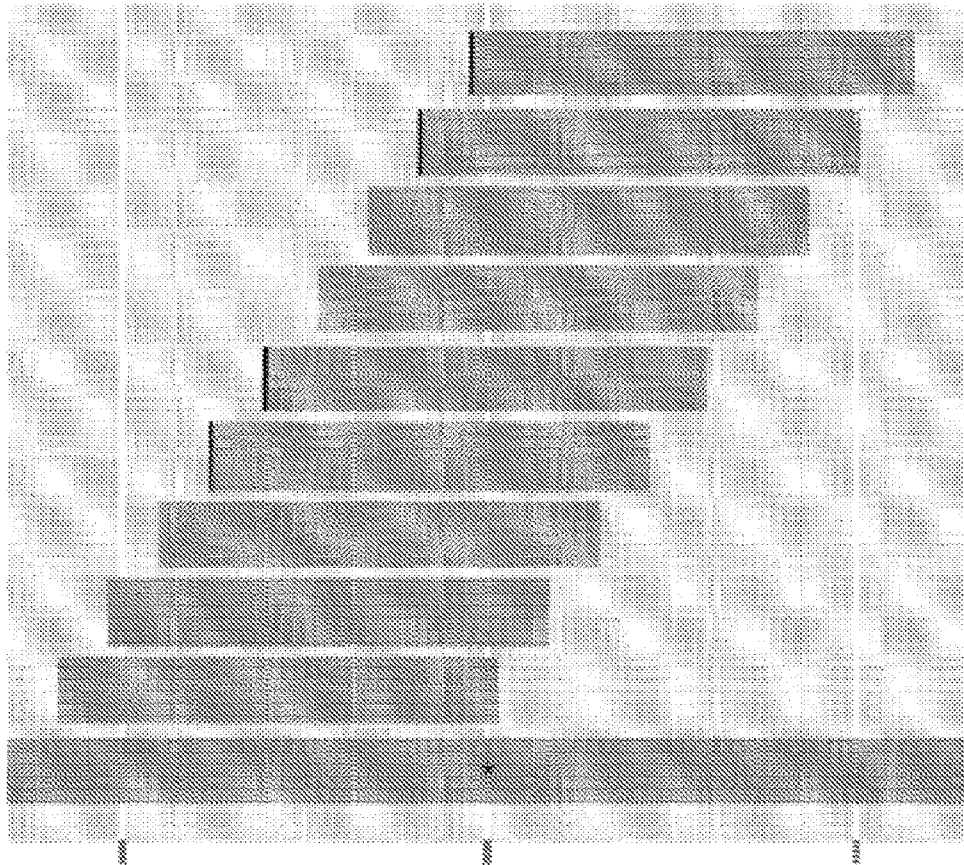


Figura 23D



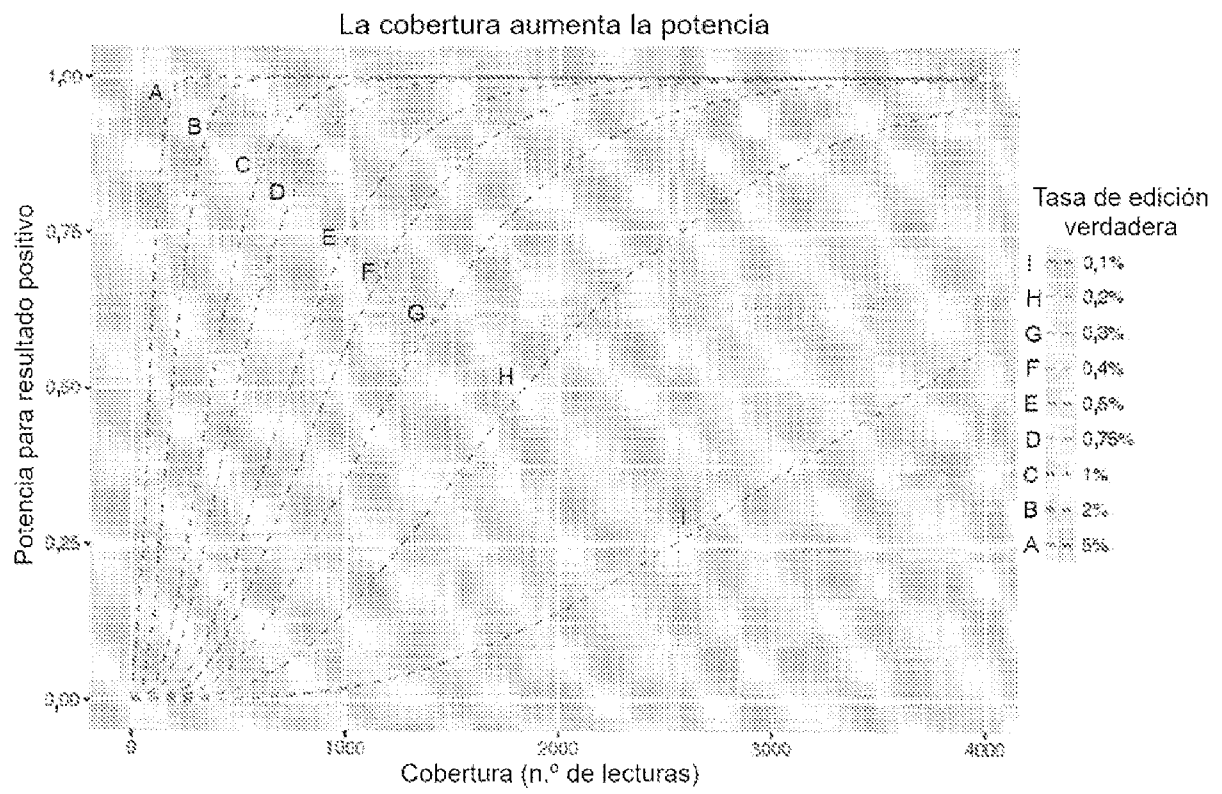
**Figura 24**



**Diseño de cebo de captura de híbridos**



**Figura 25**



**Figura 26**  
**Resumen de captura de híbridos (SPY101)**

SPY101	Ronda 1	Ronda 2
Número de donantes	3	4
Formato de Cas9	ARNm	RNP
Mediana de la cobertura de secuenciación	~800	2492
Número de sitios	840	5003
Número de apareamientos erróneos y tipos de PAM	4: NGG y NAG	5: NGG, NAG 4: NGA, NAA, NCG, NGC, NTG, NGT
N.º de sitios inespecíficos con evidencias de corte	0*	0 <sup>Δ</sup>
	*Frec. de indeles >0,5%	<sup>Δ</sup> Frec. de indeles >0,2%

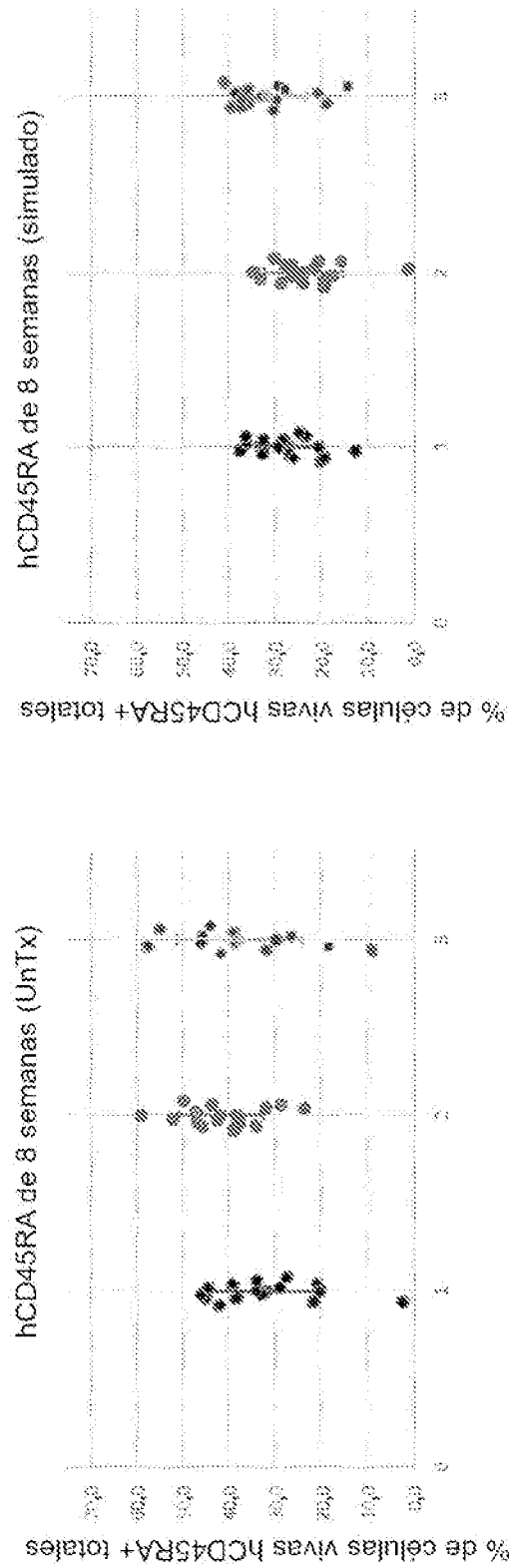
**Figura 27**  
**Resumen de captura de híbridos (SD2)**

SD2	Ronda 1	Ronda 2
Número de donantes	3	4
Formato de Cas9	ARNm	RNP
Mediana de la cobertura de secuenciación	~800	2775
Número de sitios	677	3272
Número de apareamientos erróneos y tipos de PAM	4: NGG y NAG	5: NGG, NAG 4: NGA, NAA, NCG, NGC, NTG, NCT
N.º de sitios inespecíficos con evidencias de corte*	0*	0 <sup>Δ</sup>
	*Frec. de indeles >0,5%	<sup>Δ</sup> Frec. de indeles >0,2%

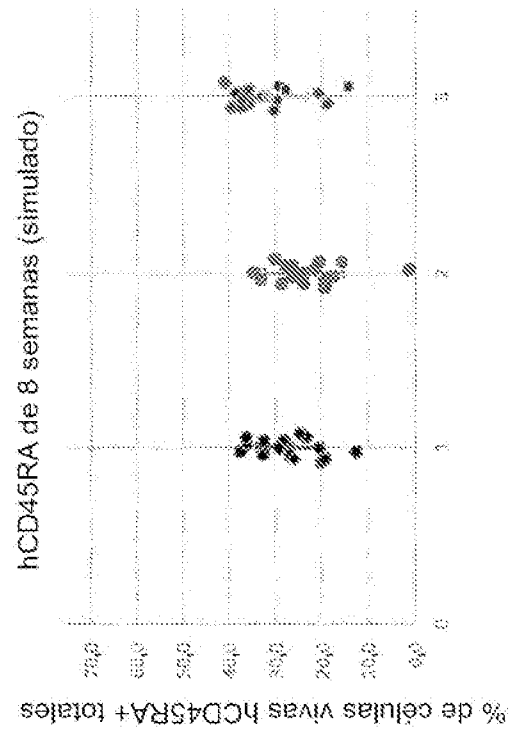
**Figura 28**  
**Plan de estudio de injerto piloto**

Grupo	Tratamiento	Dosis de radiación (cGy)	Tasa de dosis de radiación (cGy/min)	Número de animales hembra		
				Cohorte 1 célula: N.º de donante 1	Cohorte 2 célula: N.º de donante 2	Cohorte 3 origen de célula: N.º de donante 3
1	Células sin tratar	200		16	16	16
2	"Electroporación simulada" de células sin ningún componente de edición de CRISPR/Cas9	200		16	16	16
3	Electroporación de células con ARN guía de GFP y proteína Cas9	200	160	16	16	16
4	Electroporación de células con ARN guía de SPY101 y proteína Cas9	200		16	16	16
5	Electroporación de células con ARN guía de SD2 y proteína Cas9	200		16	16	16
6	Control sin tratar	200		10	10	10

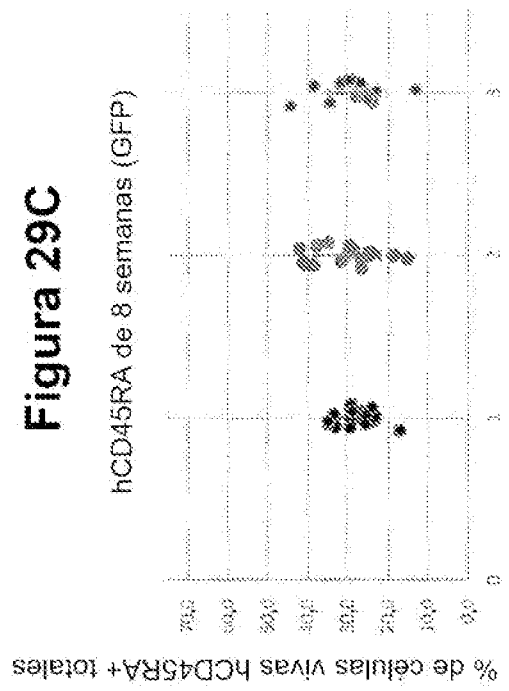
**Figura 29A**



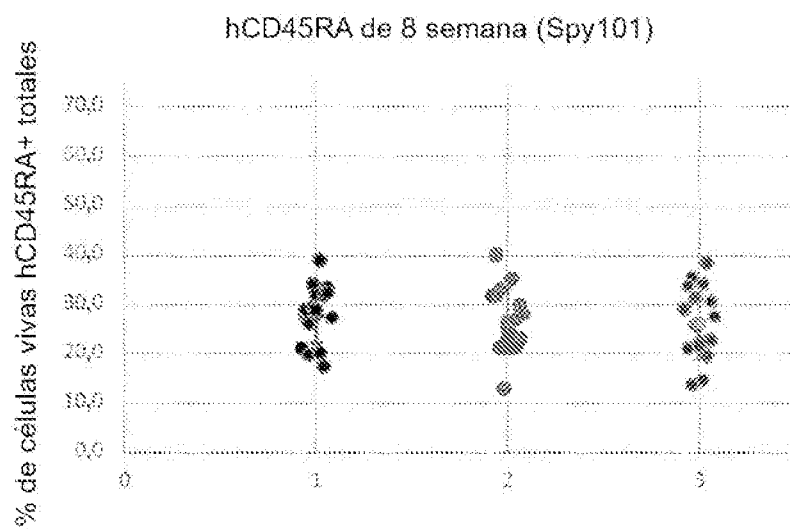
**Figura 29B**



**Figura 29C**



**Figura 29D**



**Figura 29E**

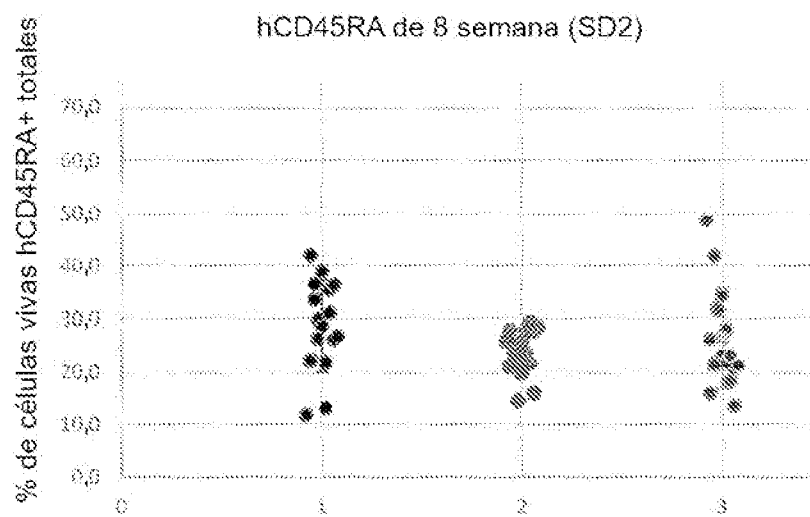


Figura 30

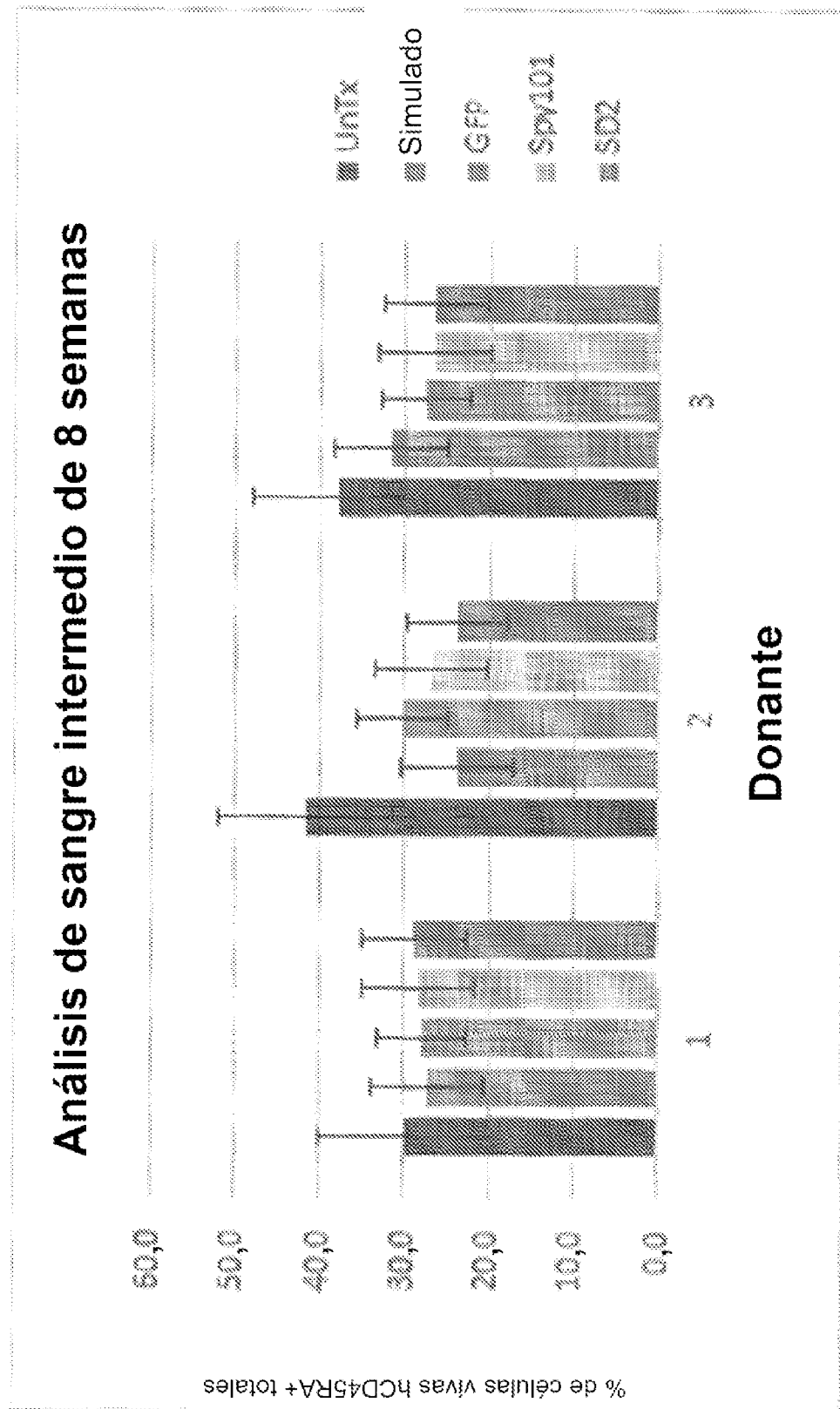
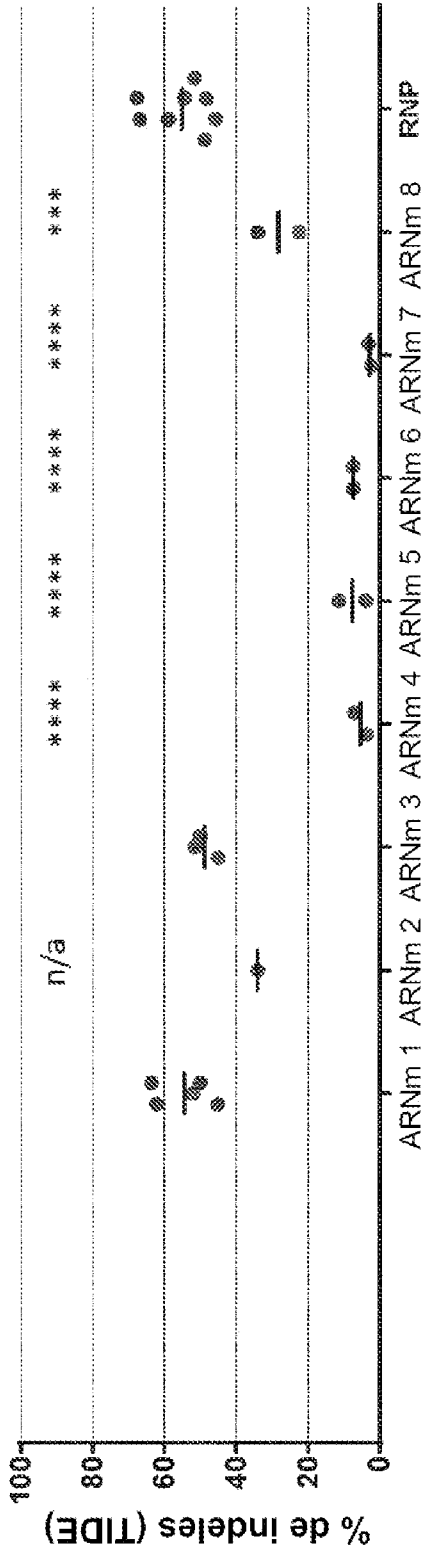


Figura 31

RNP de Cas9 es igual de eficiente en la edición que ARNm de Cas9, pero superior en la viabilidad tras la electroporación

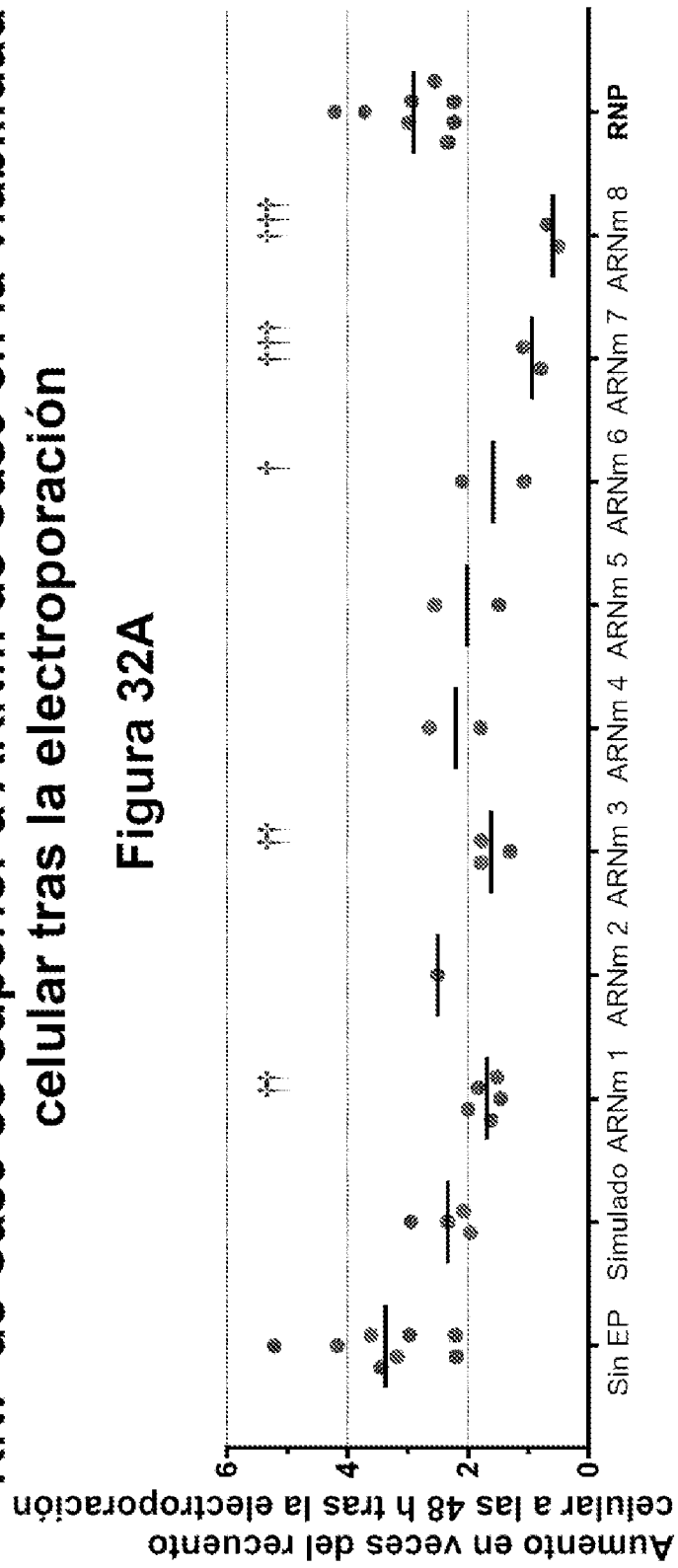


\*\*\*p<0,001 en comparación con RNP mediante ANOVA de un factor  
\*\*\*\*p<0,0001 en comparación con RNP mediante ANOVA de un factor



# RNP de Cas9 es superior a ARNm de Cas9 en la viabilidad celular tras la electroporación

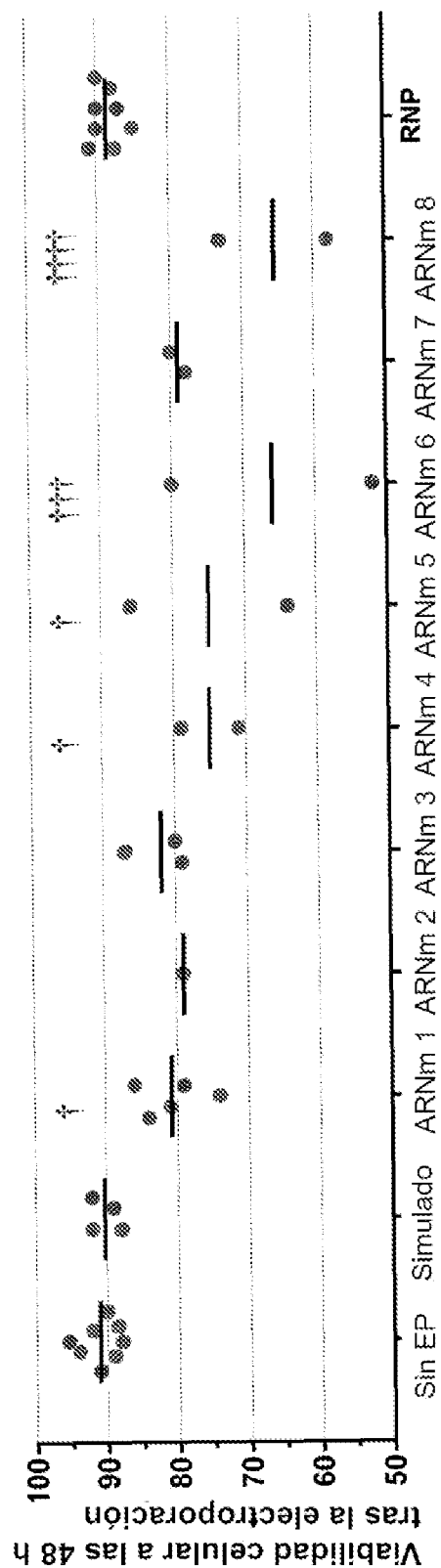
Figura 32A



† p<0,05, †† p<0,01, ††† p<0,001, †††† p<0,0001 en comparación con sin EP mediante ANOVA de un factor

## RNP de Cas9 es superior a ARNm de Cas9 en la viabilidad celular tras la electroporación

Figura 32B



†  $p < 0,05$ , ††  $p < 0,01$ , †††  $p < 0,001$ , ††††  $p < 0,0001$  en comparación con sin EP mediante ANOVA de un factor

Figura 33A Optimización de RNP de Cas9

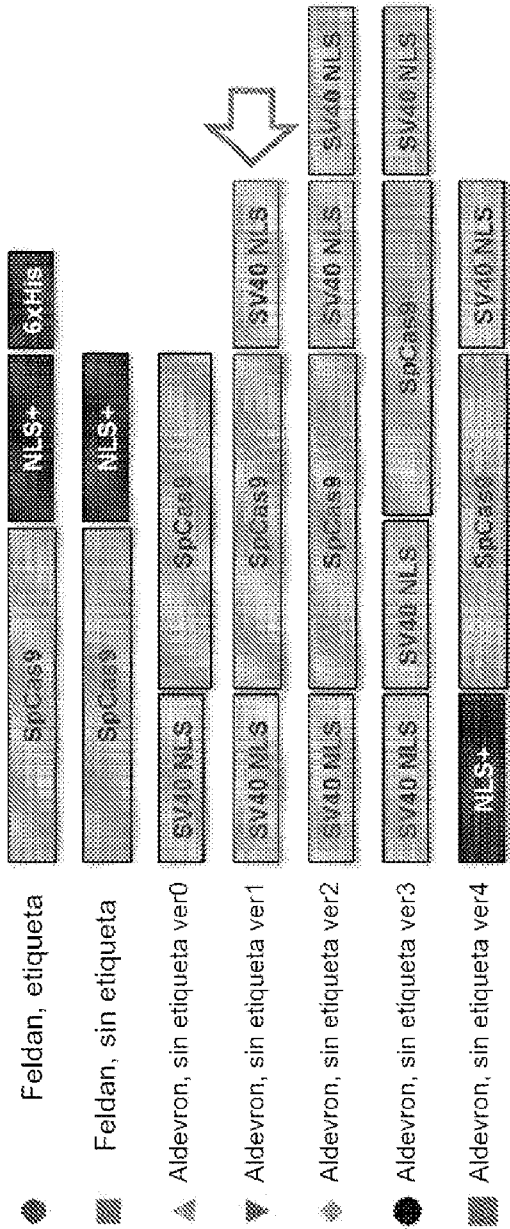


Figura 33B SPY101 (Cas9 1ug: gRNA 1ug)

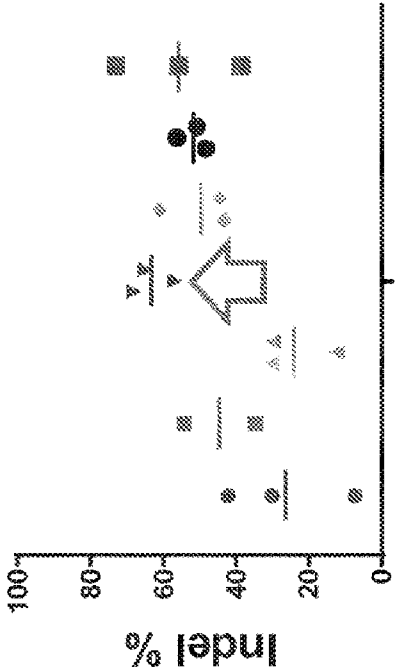
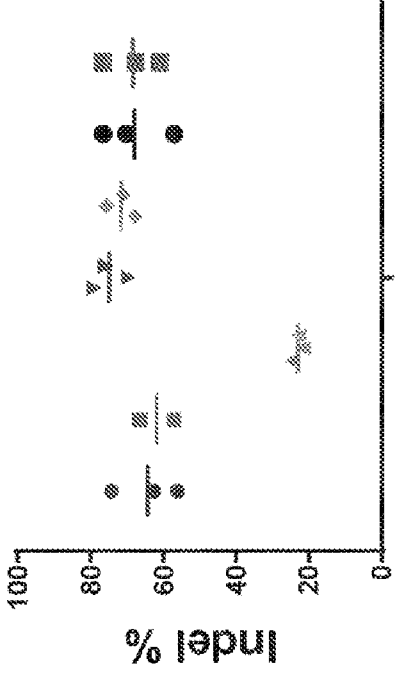


Figura 33C SPY101 Cas9 3ug: gRNA 3ug)



Eficiencia de edición global para SPY101

Figura 34A

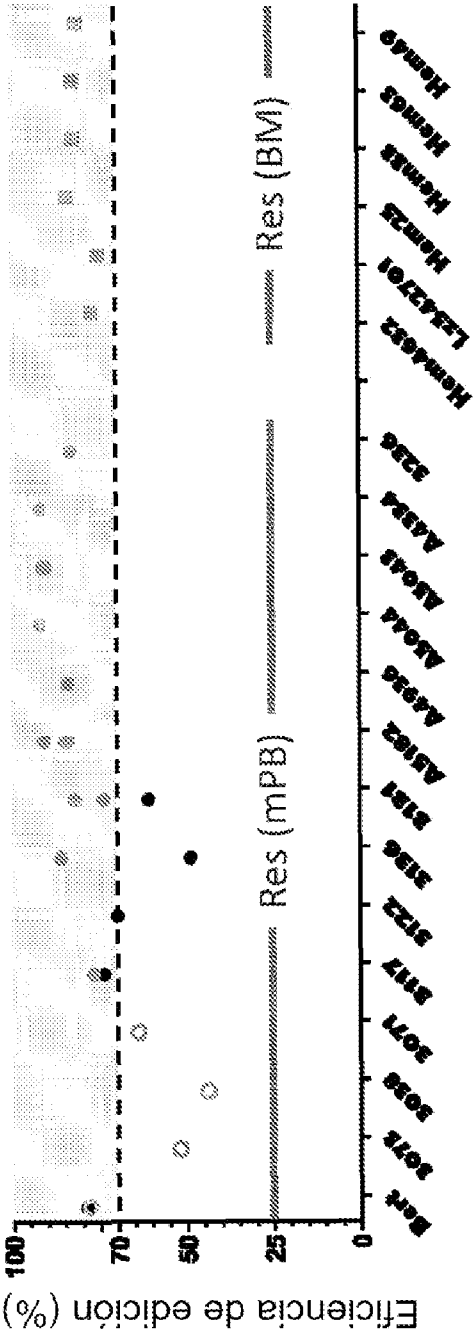
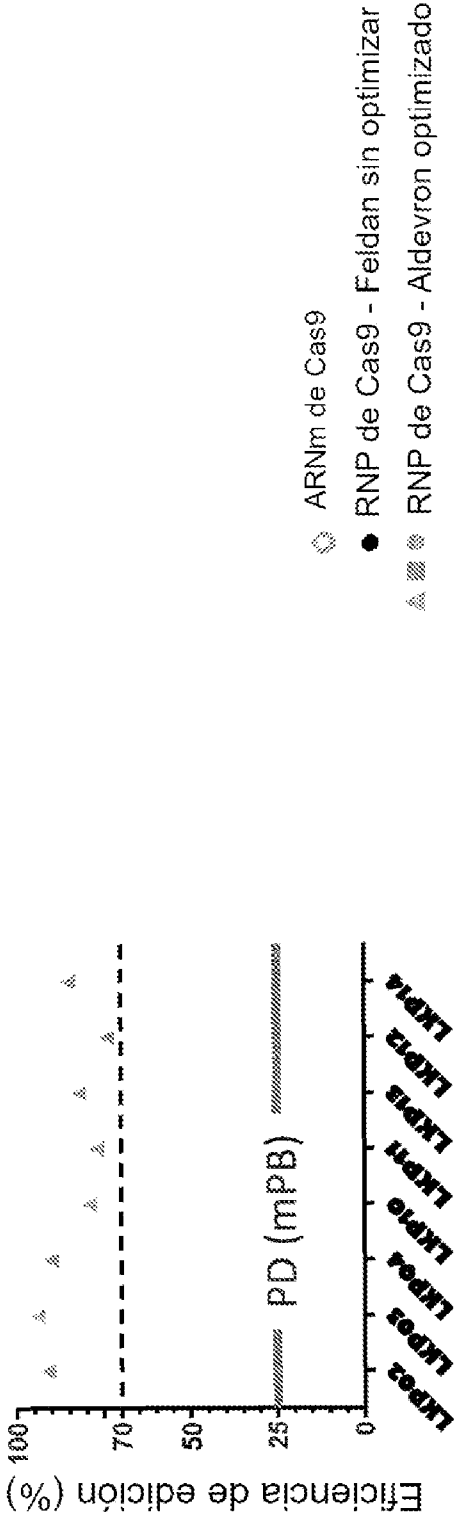


Figura 34B



# Eficacia de SPY101 en células CD34+ de mPB

Figura 35A

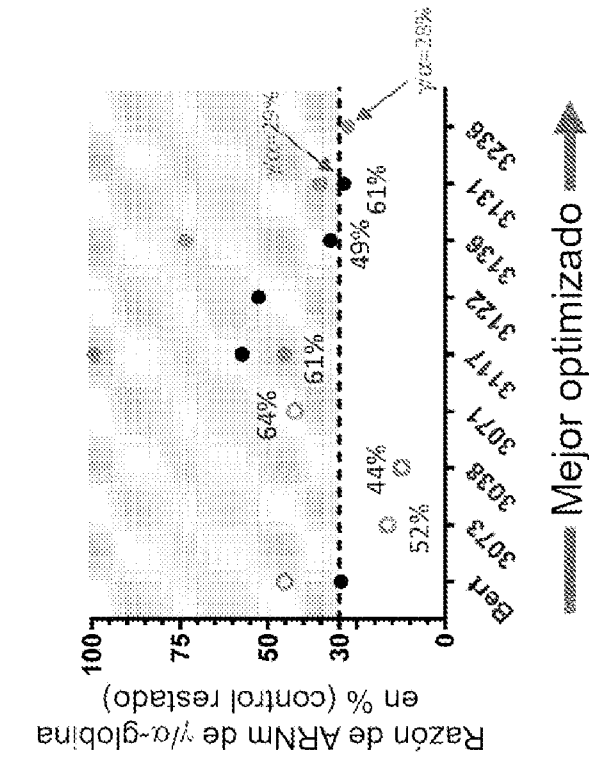
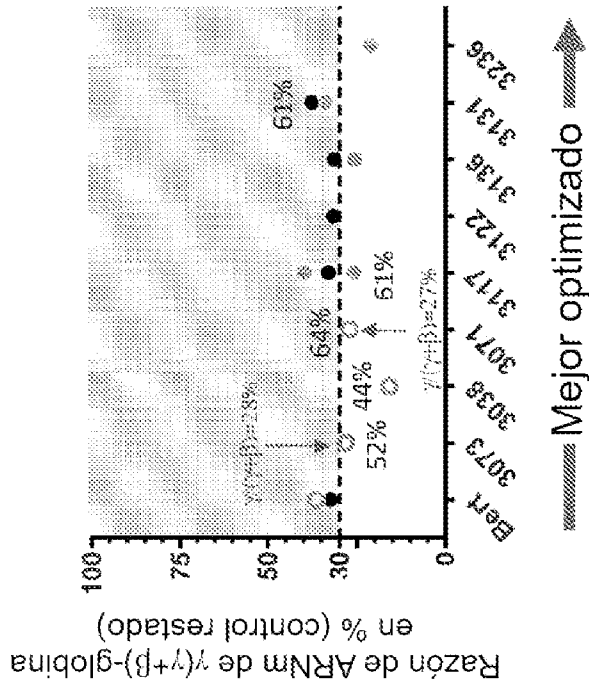


Figura 35B



% = Eficiencia de edición

Todos los puntos de datos sin marcar tenían >70% de eficiencia de edición

- ARNm de Cas9 - Diferenciación no optimizada
- RNP de Cas9 - Diferenciación no optimizada
- ▨ RNP de Cas9 - Diferenciación optimizada

# Eficacia de SPY101 en células CD34+ de BM

Figura 36A

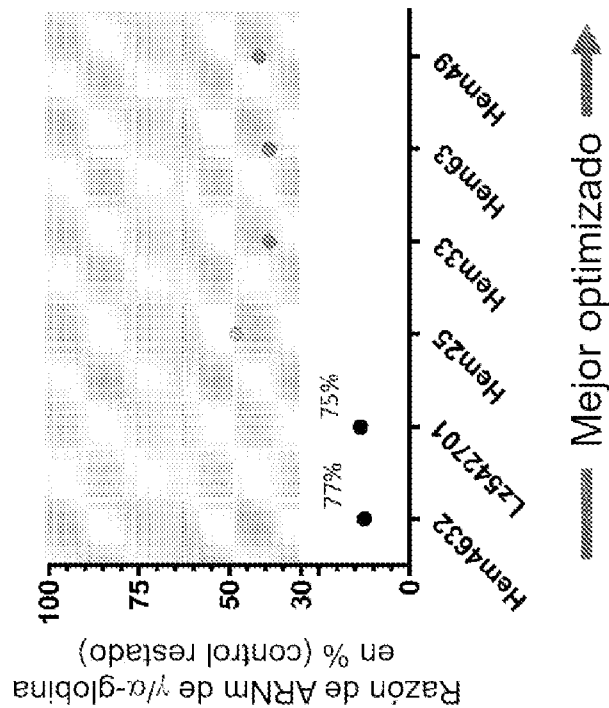
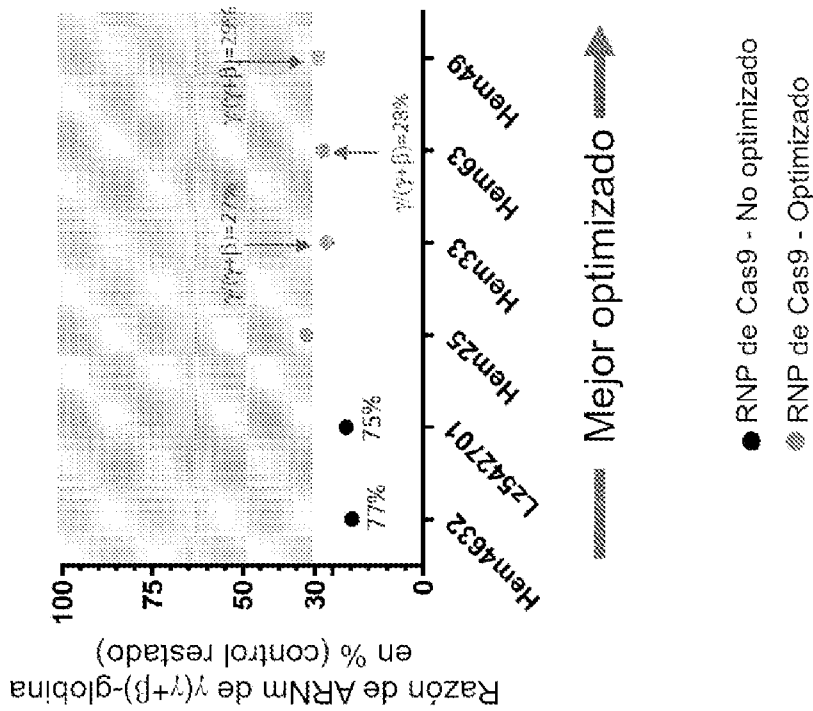


Figura 36B



% = Eficiencia de edición

Todos los puntos de datos sin marcar tenían >80% de eficiencia de edición

- RNP de Cas9 - No optimizado
- ◆ RNP de Cas9 - Optimizado

# Eficacia de SPY101 en muestras de pacientes

Figura 37A

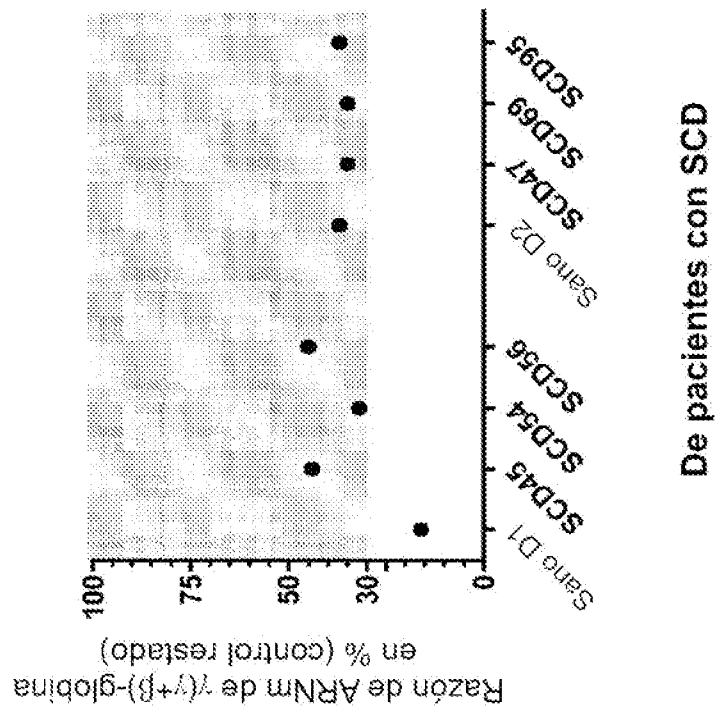
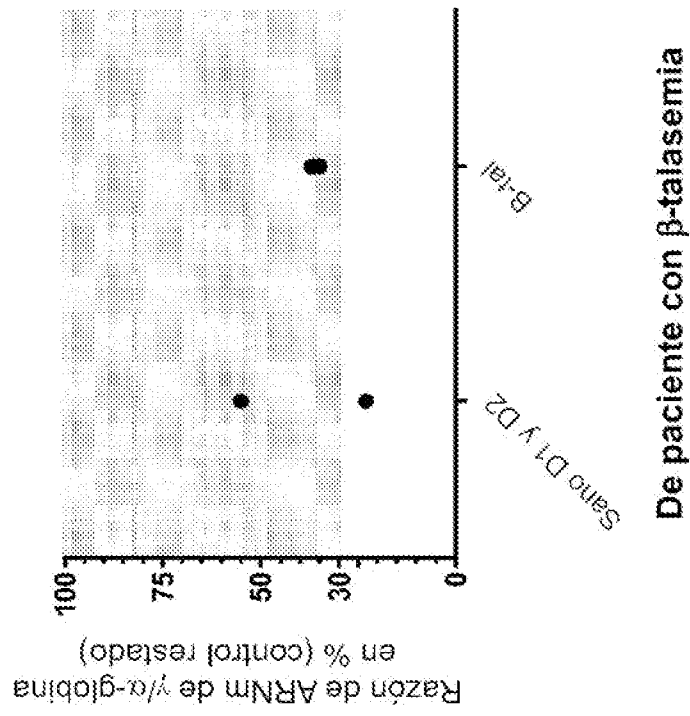


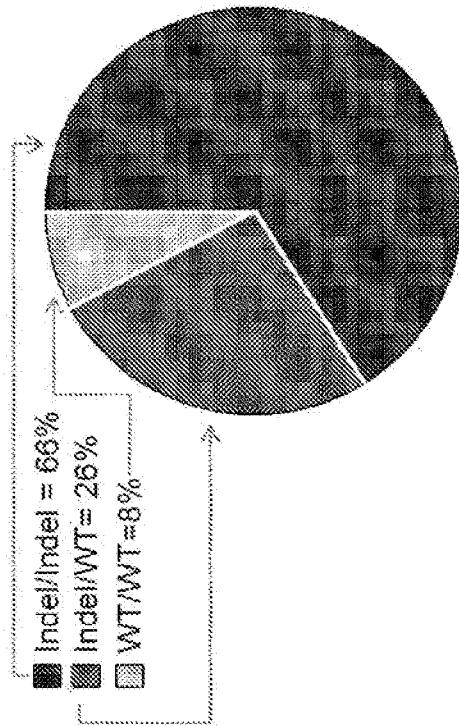
Figura 37B



# Descomposición de genotipo sometido a edición con SPY101

**Figura 38A**

ARNm de Cas9

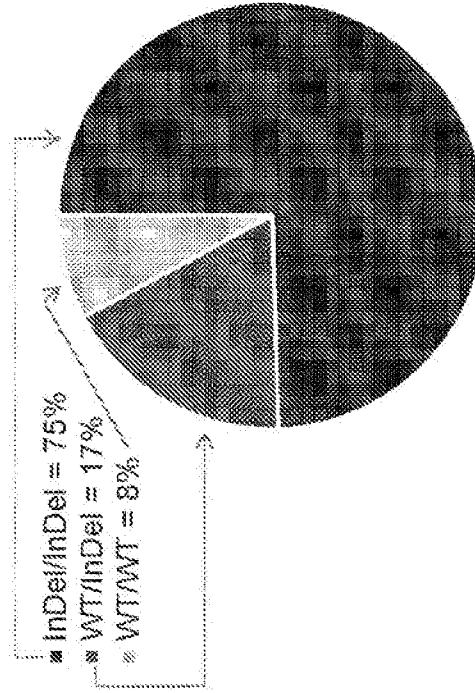


Frecuencia de indeles = 79%  
Colonias sometidas a edición = 92%

**Datos de 60 colonias**

**Figura 38B**

RNP de Cas9



Frecuencia de indeles = 83%  
Colonias sometidas a edición = 92%

**Datos de 164 colonias**



# Perturbación del sitio de unión a Gata1 mediante SPY101 vinculada a la expresión aumentada de $\gamma$ -globina

Figura 39A

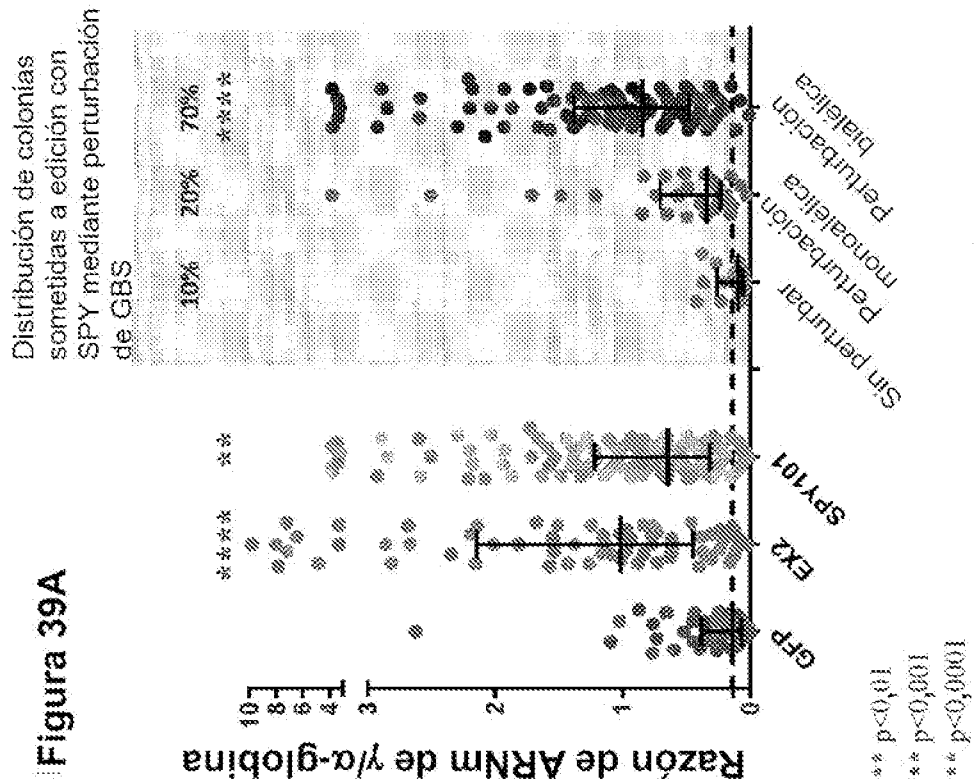
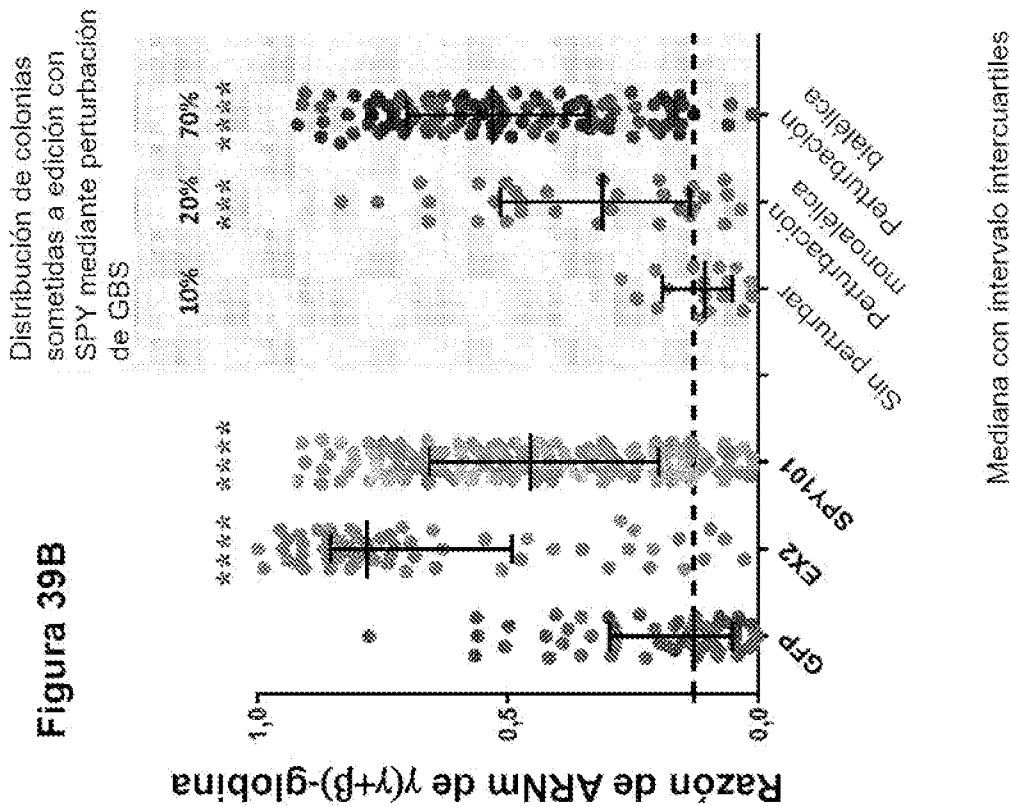


Figura 39B



# SPY101 aumentó la expresión de $\gamma$ -globina

Figura 40A

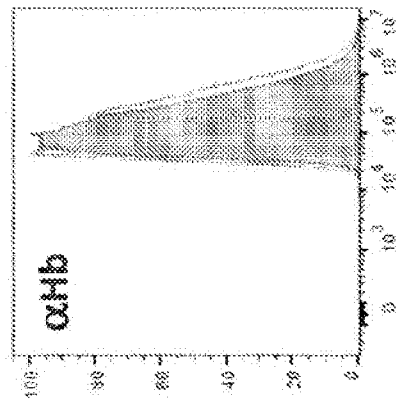


Figura 40B

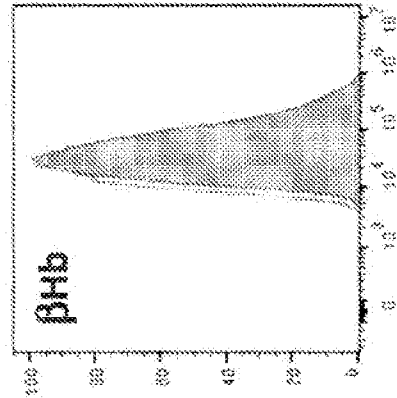


Figura 40C

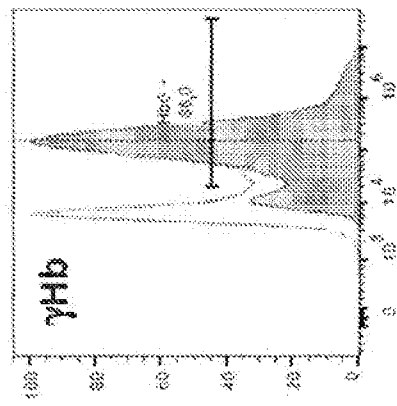


Figura 40D

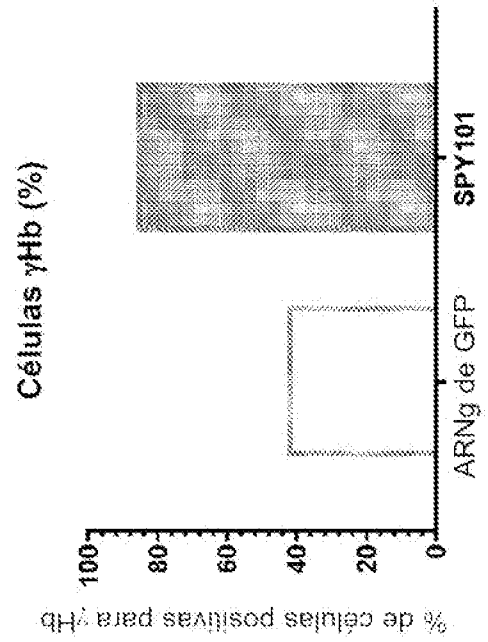


Figura 40E

