

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 808 687**

51 Int. Cl.:

**C12N 9/22** (2006.01)

**C12N 15/62** (2006.01)

**C12N 15/01** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **06.08.2004 E 15159259 (9)**

97 Fecha y número de publicación de la concesión europea: **20.05.2020 EP 2927318**

54 Título: **Métodos y composiciones para escisión dirigida y recombinación**

30 Prioridad:

**08.08.2003 US 493931 P**

**07.11.2003 US 518253 P**

**18.12.2003 US 530541 P**

**05.02.2004 US 542780 P**

**26.03.2004 US 556831 P**

**01.06.2004 US 575919 P**

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

**01.03.2021**

73 Titular/es:

**SANGAMO THERAPEUTICS, INC. (100.0%)**

**Point Richmond Tech. Center, 501 Canal**

**Boulevard, Suite A100**

**Richmond, CA 94804, US**

72 Inventor/es:

**PABO, CARL O.;**

**URNOV, FYODOR;**

**HOLMES, MICHAEL C.;**

**MILLER, JEFFREY C. y**

**GUSCHIN, DMITRY**

74 Agente/Representante:

**IZQUIERDO BLANCO, María Alicia**

ES 2 808 687 T3

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

## DESCRIPCIÓN

Métodos y composiciones para escisión dirigida y recombinación

5 CAMPO TÉCNICO

**[0001]** La presente descripción se encuentra en el campo de la ingeniería del genoma y la recombinación homóloga.

10 ANTECEDENTES

**[0002]** Un campo de interés en la biología de genoma, especialmente a la luz de la determinación de las secuencias de nucleótidos completas de un número de genomas, es la alteración dirigida de secuencias del genoma. Para proporcionar solo un ejemplo, la anemia falciforme es causada por la mutación de un solo par de nucleótidos en el gen de la  $\beta$ -globina humana. Por lo tanto, la capacidad de convertir la copia genómica endógena de este par de nucleótidos mutantes a la secuencia de tipo silvestre de manera estable y producir  $\beta$ -globina normal proporcionaría una cura para la anemia falciforme.

**[0003]** Se han hecho intentos para alterar las secuencias genómicas en células cultivadas mediante el aprovechamiento del fenómeno natural de la recombinación homóloga. Ver, p. ej., Capecchi (1989) Science 244:1288-1292; Patentes de EE.UU. N<sup>os</sup> 6,528,313 y 6,528,314. Si un polinucleótido tiene suficiente homología con la región genómica que contiene la secuencia a alterar, es posible que parte o la totalidad de la secuencia del polinucleótido reemplace la secuencia genómica por recombinación homóloga. Sin embargo, la frecuencia de recombinación homóloga en estas circunstancias es extremadamente baja. Además, la frecuencia de inserción del polinucleótido exógeno en ubicaciones genómicas que carecen de homología de secuencia excede la frecuencia de recombinación homóloga en varios órdenes de magnitud.

**[0004]** La introducción de una rotura de doble cadena en el ADN genómico, en la región de la homología de cojinete de genoma de un polinucleótido exógeno, se ha demostrado para estimular la recombinación homóloga en este sitio por varios mil veces en células cultivadas. Rouet y col. (1994) Mol. Cell. Biol. 14: 8096-8106; Choulika y col. (1995) Mol. Cell. Biol. 15: 1968-1973; Donoho y col. (1998) Mol. Cell. Biol. 18: 4070-4078. Ver también Johnson et al. (2001) Biochem. Soc. Trans. 29: 196-201; y Yáñez et al. (1998) Gene Therapy 5: 149-159. En estos métodos, la escisión del ADN en la región genómica deseada se logró insertando un sitio de reconocimiento para una meganucleasa (es decir, una endonucleasa cuya secuencia de reconocimiento es tan grande que no ocurre, o ocurre raramente, en el genoma de interés) en la región genómica deseada. WO 95/17911 y Koller et al. (Genetics (1989) 86: 8932-8935) se refieren a la inactivación de un gen HLA usando recombinación homóloga. Bibikova y col. (Science (2003) 300 (5620): 764) describe un método para alterar un gen en *Drosophila* usando nucleasas de dedos de zinc, cuyos detalles adicionales se proporcionan en Bibikova et al. (Genética (2002) 161 (3): 1169-1175). Bicknell (Proc. Natl. Acad. Sci. EE.UU. (1994) 91: 4751-4755) y Bicknell (Current Biology (1996) 6 (12): 1695-1697) se relacionan con mutaciones del gen  $\beta$ 2-microglobulina identificadas en líneas celulares colorrectales y tumores. El documento US 5,792,640 describe métodos para clonar endonucleasas de restricción híbridas, y Porteus y Baltimore (Science (2003) 300 (5620): 763) se refieren al uso de nucleasas quiméricas para estimular la selección genética en células humanas.

**[0005]** Sin embargo, la recombinación homóloga estimulada por escisión de meganucleasa se basa en ya sea la presencia fortuita de, o la inserción dirigida de un sitio de reconocimiento de meganucleasa adecuado en el entorno de la región genómica que se altera. Dado que los sitios de reconocimiento de meganucleasas son raros (o inexistentes) en un genoma de mamífero típico, y la inserción de un sitio de reconocimiento de meganucleasas adecuado está plagado de las mismas dificultades que las asociadas con otras alteraciones genómicas, estos métodos no son ampliamente aplicables.

**[0006]** Por lo tanto, sigue habiendo una necesidad de composiciones y métodos para la alteración dirigida de secuencias en cualquier genoma.

SUMARIO

**[0007]** La invención proporciona un método in vitro para la escisión específica de un gen HLA de clase I, un gen HLA que codifica un complejo de histocompatibilidad mayor (MHC) de proteína de clase 1 o un gen HLA que codifica un gen de subunidad MHC  $\beta$  (microglobulina  $\beta$ 2), en una célula madre hematopoyética, el método comprende el uso de una o más nucleasas dirigidas para crear una ruptura bicatenaria en el gen HLA de modo que el gen HLA esté inactivado, en donde la ruptura bicatenaria en el gen HLA es seguida por una unión de extremo no homólogo (NHEJ), y en donde la una o más nucleasas dirigidas es una proteína de fusión que comprende un dominio de unión de dedos de zinc diseñado y un dominio de escisión o medio dominio de escisión.

**[0008]** La invención también proporciona una célula madre hematopoyética aislada producida por el método in vitro de la invención, en el que se inactiva el gen HLA.

**[0009]** La presente descripción proporciona composiciones y métodos para la escisión selectiva de la cromatina celular

en una región de interés y/o la recombinación homóloga en una región predeterminada de interés en células. Las células incluyen células cultivadas, células en un organismo y células que se han eliminado de un organismo para su tratamiento en los casos en que las células y/o sus descendientes serán devueltas al organismo después del tratamiento. Una región de interés en la cromatina celular puede ser, p. ej., una secuencia genómica o una porción de la misma. Las composiciones incluyen polipéptidos de fusión que comprenden un dominio de unión de dedos de zinc diseñado (p. ej., un dominio de unión de dedos de zinc que tiene una especificidad nueva) y un dominio de escisión, y polipéptidos de fusión que comprenden un dominio de unión de dedos de zinc diseñado y un medio dominio de escisión. Los dominios de división y los medios dominios de división pueden obtenerse, p. ej., a partir de diversas endonucleasas de restricción y/o endonucleasas de referencia.

**[0010]** La cromatina celular puede estar presente en cualquier tipo de célula incluyendo, pero no limitado a, procariotas y eucariotas, células fúngicas, células vegetales, células animales, células de mamífero, células de primates y células humanas.

**[0011]** Un método para la escisión de la cromatina celular en una región de interés (p. ej., un método para la escisión selectiva de se describe secuencias genómicas), comprendiendo el método: (a) seleccionar una primera secuencia en la región de interés; (b) diseñar un primer dominio de unión de dedos de zinc para unirse a la primera secuencia; y (c) expresar una primera proteína de fusión en la célula, la primera proteína de fusión comprende el primer dominio de unión de dedos de zinc modificado por ingeniería genética y un dominio de escisión; en donde la primera proteína de fusión se une a la primera secuencia y la cromatina celular se escinde en la región de interés. El sitio de escisión puede coincidir con la secuencia a la que se une la proteína de fusión, o puede ser adyacente (p. ej., separado del borde cercano del sitio de unión por 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o más nucleótidos). Una proteína de fusión se puede expresar en una célula, p. ej., entregando la proteína de fusión a la célula o entregando un polinucleótido que codifica la proteína de fusión a una célula, en donde el polinucleótido, si es ADN, se transcribe, y una molécula de ARN se entrega a la célula o una transcripción de una molécula de ADN se entrega a la célula, para generar la proteína de fusión. Los métodos para el suministro de polinucleótidos y polipéptidos a las células se presentan en otra parte de esta descripción.

**[0012]** En ciertas realizaciones, el dominio de escisión puede comprender dos medios dominios de escisión que están covalentemente enlazados en el mismo polipéptido. Los dos medios dominios de escisión pueden derivarse de la misma endonucleasa o de diferentes endonucleasas.

**[0013]** En realizaciones adicionales, la escisión selectiva de la cromatina celular en una región de interés se consigue mediante la expresión de dos proteínas de fusión en una célula, cada proteína de fusión que comprende un dominio de unión de dedos de cinc y un medio dominio de escisión. Uno o ambos dominios de unión de dedos de zinc de las proteínas de fusión se pueden diseñar para unirse a una secuencia diana en la vecindad del sitio de escisión deseado. Si la expresión de las proteínas de fusión es por suministro de polinucleótidos, cada una de las dos proteínas de fusión puede ser codificada por un polinucleótido separado, o un solo polinucleótido puede codificar ambas proteínas de fusión.

**[0014]** Por consiguiente, un método para escindir la cromatina celular en una región de interés puede comprender (a) seleccionar una primera secuencia en la región de interés; (b) diseñar un primer dominio de unión de dedos de zinc para unirse a la primera secuencia; (c) expresar una primera proteína de fusión en la célula, comprendiendo la primera proteína de fusión el primer dominio de unión de dedos de zinc y un primer medio dominio de escisión; y (d) expresar una segunda proteína de fusión en la célula, comprendiendo la segunda proteína de fusión un segundo dominio de unión de dedos de zinc y un segundo medio dominio de escisión, en donde la primera proteína de fusión se une a la primera secuencia, y la segunda proteína de fusión se une a una segunda secuencia ubicada entre 2 y 50 nucleótidos de la primera secuencia, colocando así los medios dominios de escisión de modo que la cromatina celular se escinde en la región de interés.

**[0015]** En ciertas realizaciones, la unión de las proteínas de fusión primera y segunda posiciona los dominios medios de escisión tales que se reconstituye un dominio de escisión funcional.

**[0016]** En ciertas realizaciones, el segundo dominio de unión de dedos de cinc está diseñado para unirse a la segunda secuencia. En realizaciones adicionales, el primer y segundo medio dominio de escisión se derivan de la misma endonucleasa, que puede ser, p. ej., una endonucleasa de resticción (p. ej., una endonucleasa de restricción de Tipo IIS tal como *Fok I*) o una endonucleasa de referencia.

**[0017]** Cualquiera de los métodos descritos en el presente documento puede comprender (a) seleccionar secuencias primera y segunda en una región de interés, en donde las secuencias primera y segunda son entre 2 y 50 nucleótidos de diferencia; (b) diseñar un primer dominio de unión de dedos de zinc para unirse a la primera secuencia; (c) diseñar un segundo dominio de unión de dedos de zinc para unirse a la segunda secuencia; (d) expresar una primera proteína de fusión en la célula, la primera proteína de fusión comprende el primer dominio de unión de dedos de zinc modificado por ingeniería genética y un primer medio dominio de escisión; (e) expresar una segunda proteína de fusión en la célula, comprendiendo la segunda proteína de fusión el segundo dominio de unión de dedos de zinc modificado por ingeniería genética y un segundo medio dominio de escisión; en donde la primera proteína de fusión se une a la

primera secuencia y la segunda proteína de fusión se une a la segunda secuencia, colocando así el primer y segundo medio dominio de escisión de modo que la cromatina celular se escinde en la región de interés.

5 **[0018]** En ciertas realizaciones, los medios dominios primero y segundo de escisión se derivan de la misma endonucleasa, p. ej., una endonucleasa de restricción de tipo IIS, p. ej., *Fok I*. En realizaciones adicionales, la cromatina celular se escinde en uno o más sitios entre la primera y la segunda secuencia a las que se unen las proteínas de fusión.

10 **[0019]** En realizaciones adicionales de la divulgación, un método para la escisión de la cromatina celular en una región de interés comprende (a) la selección de la región de interés; (b) diseñar un primer dominio de unión de dedos de zinc para unirse a una primera secuencia en la región de interés; (c) proporcionar un segundo dominio de unión de dedos de zinc que se une a una segunda secuencia en la región de interés, en donde la segunda secuencia está ubicada entre 2 y 50 nucleótidos de la primera secuencia; (d) expresar una primera proteína de fusión en la célula, comprendiendo la primera proteína de fusión el primer dominio de unión de dedos de zinc y un primer medio dominio de escisión; y (e) expresar una segunda proteína de fusión en la célula, comprendiendo la segunda proteína de fusión el segundo dominio de unión de dedos de zinc y un segundo medio dominio de escisión; en donde la primera proteína de fusión se une a la primera secuencia, y la segunda proteína de fusión se une a la segunda secuencia, posicionando así los medios dominios de escisión de modo que la cromatina celular se escinde en la región de interés.

20 **[0020]** En cualquiera de los métodos descritos en el presente documento, los medios dominios primero y segundo de escisión se pueden derivar de la misma endonucleasa o de diferentes endonucleasas. En realizaciones adicionales, el segundo dominio de unión de dedos de zinc está diseñado para unirse a la segunda secuencia.

25 **[0021]** Si uno o más polinucleótidos que codifican las proteínas de fusión se introducen en la célula, un método de ejemplo para la escisión selectiva de la cromatina celular en una región de interés comprende (a) seleccionar la región de interés; (b) diseñar un primer dominio de unión de dedos de zinc para unirse a una primera secuencia en la región de interés; (c) proporcionar un segundo dominio de unión de dedos de zinc que se une a una segunda secuencia en la región de interés, en donde la segunda secuencia está ubicada entre 2 y 50 nucleótidos de la primera secuencia; y (d) poner en contacto una célula con (i) un primer polinucleótido que codifica una primera proteína de fusión, comprendiendo la proteína de fusión el primer dominio de unión de dedos de zinc y un primer medio dominio de escisión, y (ii) un segundo polinucleótido que codifica una segunda proteína de fusión, comprendiendo la proteína de fusión el segundo dominio de unión de dedos de zinc y un segundo medio dominio de escisión; en donde las proteínas de fusión primera y segunda se expresan, la primera proteína de fusión se une a la primera secuencia y la segunda proteína de fusión se une a la segunda secuencia, colocando así los medios dominios de escisión de modo que la cromatina celular se escinde en la región de interés. En una variación de este método, una célula se pone en contacto con un único polinucleótido que codifica ambas proteínas de fusión.

35 **[0022]** Para cualquiera de los métodos mencionados anteriormente, la cromatina celular puede estar en un cromosoma, episoma o genoma organular. Además, en cualquiera de los métodos descritos aquí, al menos un dominio de unión de dedos de zinc está diseñado, p. ej., por métodos de diseño o selección.

40 **[0023]** Del mismo modo, para cualquiera de los métodos mencionados anteriormente, el dominio de un medio de escisión se pueden derivar de, p. ej., una endonucleasa homing o una endonucleasa de restricción, p. ej., una endonucleasa de restricción de tipo IIS. Un ejemplar de endonucleasa de restricción de tipo IIS es *Fok I*.

45 **[0024]** Para cualquiera de los métodos de escisión dirigida, mutagénesis dirigida y/o recombinación dirigida da a conocer en la presente memoria la utilización de proteínas de fusión que comprenden un medio dominio de escisión, los bordes cerca de los sitios de unión de las proteínas de fusión pueden ser separados por 5 o 6 de pares base. En estas realizaciones, el dominio de unión y el dominio de escisión de las proteínas de fusión se pueden separar mediante un conector de 4 residuos de aminoácidos.

50 **[0025]** En ciertas realizaciones, es posible obtener una mayor especificidad de escisión mediante la utilización de proteínas de fusión en donde uno o ambos de medios dominios de escisión contiene una alteración en la secuencia de aminoácidos de la interfaz de dimerización.

55 **[0026]** La mutagénesis dirigida de una región de interés en la cromatina celular puede ocurrir cuando un evento de escisión dirigido, como se describió anteriormente, es seguido por extremos no homólogos (NHEJ). Por consiguiente, se describen métodos para la alteración de una primera secuencia de nucleótidos en una región de interés en la cromatina celular, en donde los métodos comprenden los pasos de (a) diseñar un primer dominio de unión de dedos de zinc para unirse a una segunda secuencia de nucleótidos en la región de interés, en donde la segunda secuencia comprende al menos 9 nucleótidos; (b) proporcionar un segundo dominio de unión de dedos de zinc para unirse a una tercera secuencia de nucleótidos, en donde la tercera secuencia comprende al menos 9 nucleótidos y está ubicada entre 2 y 50 nucleótidos de la segunda secuencia; (c) expresar una primera proteína de fusión en la célula, la primera proteína de fusión comprende el primer dominio de unión de dedos de zinc y un primer medio dominio de escisión; y (d) expresar una segunda proteína de fusión en la célula, comprendiendo la segunda proteína de fusión el segundo dominio de unión de dedos de zinc y un segundo medio dominio de escisión; en donde la primera proteína de fusión

se une a la segunda secuencia, y la segunda proteína de fusión se une a la tercera secuencia, colocando así los medios dominios de escisión de modo que la cromatina celular se escinde en la región de interés y el sitio de escisión se somete a unión homóloga final.

5 **[0027]** Mutaciones dirigidas resultantes del método antes mencionado incluyen, pero no se limitan a, mutaciones puntuales (es decir, conversión de un solo par de bases de un par de bases diferente), sustituciones (es decir, la conversión de una pluralidad de pares de bases a una diferente secuencia de longitud idéntica), inserciones o uno o más pares de bases, eliminaciones de uno o más pares de bases y cualquier combinación de las alteraciones de secuencia mencionadas anteriormente.

10 **[0028]** También se describen métodos para la recombinación dirigida (para, p. ej., alteración o sustitución de una secuencia en un cromosoma o una región de interés en la cromatina celular). Por ejemplo, una secuencia genómica mutante puede reemplazarse por una secuencia de tipo silvestre, p. ej., para el tratamiento de enfermedades genéticas o trastornos hereditarios. Además, una secuencia genómica de tipo silvestre puede ser reemplazada por una secuencia mutante, p. ej., para evitar la función de un producto oncogénico o un producto de un gen involucrado en una respuesta inflamatoria inapropiada. Además, un alelo de un gen puede ser reemplazado por un alelo diferente.

15 **[0029]** En dichos métodos, una o más nucleasas específicas crean una rotura de doble cadena en la cromatina celular en un sitio predeterminado, y un polinucleótido de donantes, que tiene homología con la secuencia de nucleótidos de la cromatina celular en la región de la ruptura, se introduce en la célula. Los procesos de reparación del ADN celular se activan por la presencia de la ruptura bicatenaria y el polinucleótido del donante se usa como plantilla para la reparación de la ruptura, lo que resulta en la introducción de toda o parte de la secuencia de nucleótidos del donante en la cromatina celular. Así, una primera secuencia en la cromatina celular puede alterarse y, en ciertos casos, puede convertirse en una secuencia presente en un polinucleótido donante.

20 **[0030]** En este contexto, se puede entender el uso de los términos "reemplazar" o "sustitución" para representar la sustitución de una secuencia de nucleótido por otro, (es decir, la sustitución de una secuencia en el sentido de información), y no requiere necesariamente reemplazo físico o químico de un polinucleótido por otro.

25 **[0031]** Por consiguiente, un método para la sustitución de una región de interés en la cromatina celular (*p. ej.*, una secuencia genómica) se da a conocer con una primera secuencia de nucleótidos, comprendiendo el método: (a) diseñar un dominio de unión de dedos de zinc para unirse a una segunda secuencia en la región de interés; (b) expresar una proteína de fusión en una célula, comprendiendo la proteína de fusión el dominio de unión de dedos de zinc y un dominio de escisión; y (c) poner en contacto la célula con un polinucleótido que comprende la primera secuencia de nucleótidos; en donde la proteína de fusión se une a la segunda secuencia de tal manera que la cromatina celular se escinde en la región de interés y una secuencia de nucleótidos en la región de interés se reemplaza con la primera secuencia de nucleótidos. Generalmente, la cromatina celular se escinde en la región de interés en o adyacente a la segunda secuencia. El dominio de escisión puede comprender dos medios dominios de escisión, que pueden derivarse de la misma o de nucleasas diferentes.

30 **[0032]** Además, se describe un método para la sustitución de una región de interés en la cromatina celular (*p. ej.*, una secuencia genómica) con una primera secuencia de nucleótidos, comprendiendo el método: (a) la ingeniería de un primer dominio de unión al dedo de zinc para unirse a una segunda secuencia en la región de interés; (b) proporcionar un segundo dominio de unión de dedos de zinc para unirse a una tercera secuencia en la región de interés; (c) expresar una primera proteína de fusión en una célula, comprendiendo la primera proteína de fusión el primer dominio de unión de dedos de zinc y un primer medio dominio de escisión; (d) expresar una segunda proteína de fusión en la célula, comprendiendo la segunda proteína de fusión el segundo dominio de unión de dedos de zinc y un segundo medio dominio de escisión; y (e) poner en contacto la célula con un polinucleótido que comprende la primera secuencia de nucleótidos; en donde la primera proteína de fusión se une a la segunda secuencia y la segunda proteína de fusión se une a la tercera secuencia, colocando así los medios dominios de escisión de modo que la cromatina celular se escinde en la región de interés y se produce una secuencia de nucleótidos en la región de interés reemplazada con la primera secuencia de nucleótidos. Generalmente, la cromatina celular se escinde en la región de interés en un sitio entre la segunda y la tercera secuencia.

35 **[0033]** Los métodos adicionales para la sustitución de una región de interés en la cromatina celular (*p. ej.*, una secuencia genómica) con una primera secuencia de nucleótidos comprenden: (a) seleccionar una segunda secuencia, en donde la segunda secuencia está en la región de interés y tiene una longitud de al menos 9 nucleótidos; (b) diseñar un primer dominio de unión de dedos de zinc para unirse a la segunda secuencia; (c) seleccionar una tercera secuencia, en donde la tercera secuencia tiene una longitud de al menos 9 nucleótidos y está ubicada entre 2 y 50 nucleótidos de la segunda secuencia; (d) proporcionar un segundo dominio de unión de dedos de zinc para unirse a la tercera secuencia; (e) expresar una primera proteína de fusión en una célula, la primera proteína de fusión comprende el primer dominio de unión de dedos de zinc y un primer medio dominio de escisión; (f) expresar una segunda proteína de fusión en la célula, comprendiendo la segunda proteína de fusión el segundo dominio de unión de dedos de zinc y un segundo medio dominio de escisión; y (g) poner en contacto la célula con un polinucleótido que comprende la primera secuencia de nucleótidos; en donde la primera proteína de fusión se une a la segunda secuencia y se une la segunda proteína de fusión a la tercera secuencia, posicionando de ese modo los medios dominios de

escisión de tal manera que la cromatina celular se escinde en la región de interés y una secuencia de nucleótidos en la región de interés es reemplazada con la primera secuencia de nucleótidos. Generalmente, la cromatina celular se escinde en la región de interés en un sitio entre la segunda y la tercera secuencia.

5 **[0034]** En otro aspecto, se describen métodos para la recombinación específica en los que una primera secuencia de nucleótidos, situada en una región de interés en la cromatina celular, se sustituye por una segunda secuencia de nucleótidos. Los métodos comprenden (a) diseñar un primer dominio de unión de dedos de zinc para unirse a una tercera secuencia en la región de interés; (b) proporcionar un segundo dominio de unión de dedos de zinc para unirse a una cuarta secuencia; (c) expresar una primera proteína de fusión en una célula, la proteína de fusión comprende el primer dominio de unión de dedos de zinc y un primer medio dominio de escisión; (d) expresar una segunda proteína de fusión en la célula, comprendiendo la segunda proteína de fusión el segundo dominio de unión de dedos de zinc y un segundo medio dominio de escisión; y (e) poner en contacto una célula con un polinucleótido que comprende la segunda secuencia de nucleótidos; en donde la primera proteína de fusión se une a la tercera secuencia y la segunda proteína de fusión se une a la cuarta secuencia, colocando así los medios dominios de escisión de modo que la cromatina celular se escinde en la región de interés y la primera secuencia de nucleótidos se reemplaza por la segunda secuencia de nucleótidos.

20 **[0035]** Se describe un método para la alteración de una primera secuencia de nucleótidos en una región de interés en la cromatina celular, comprendiendo el método las etapas de (a) la ingeniería de un primer dominio de unión de dedo de zinc a una segunda secuencia de nucleótidos en la región de interés, en donde la segunda secuencia comprende al menos 9 nucleótidos; (b) proporcionar un segundo dominio de unión de dedos de zinc para unirse a una tercera secuencia de nucleótidos, en donde la tercera secuencia comprende al menos 9 nucleótidos y está ubicada entre 2 y 50 nucleótidos de la segunda secuencia; (c) expresar una primera proteína de fusión en la célula, comprendiendo la primera proteína de fusión el primer dominio de unión de dedos de zinc y un primer medio dominio de escisión; (d) expresar una segunda proteína de fusión en la célula, comprendiendo la segunda proteína de fusión el segundo dominio de unión de dedos de zinc y un segundo medio dominio de escisión; y (e) poner en contacto la célula con un polinucleótido que comprende una cuarta secuencia de nucleótidos, en donde la cuarta secuencia de nucleótidos es homóloga pero no idéntica con la primera secuencia de nucleótidos; en donde la primera proteína de fusión se une a la segunda secuencia, y la segunda proteína de fusión se une a la tercera secuencia, colocando así los medios dominios de escisión de modo que la cromatina celular se escinde en la región de interés y se altera la primera secuencia de nucleótidos. La primera secuencia de nucleótidos puede convertirse en la cuarta secuencia de nucleótidos. Además, las secuencias de nucleótidos segunda y tercera (es decir, los sitios de unión para las proteínas de fusión) pueden estar presentes en el polinucleótido que comprende la cuarta secuencia de nucleótidos (es decir, el polinucleótido donante) y se escinde el polinucleótido que comprende la cuarta secuencia de nucleótidos.

35 **[0036]** En los métodos anteriormente mencionados para la recombinación específica, los sitios de unión para las proteínas de fusión (es decir, las secuencias tercera y cuarta) pueden comprender cualquier número de nucleótidos. Preferiblemente, tienen al menos nueve nucleótidos de longitud, pero también pueden ser más grandes (*p. ej.*, 10, 11, 12, 13, 14, 15, 16, 17, 18 y hasta 100 nucleótidos, incluido cualquier valor integral entre 9 y 100 nucleótidos); además, las secuencias tercera y cuarta no necesitan tener la misma longitud. La distancia entre los sitios de unión (es decir, la longitud de la secuencia de nucleótidos entre la tercera y la cuarta secuencia) puede ser cualquier número integral de pares de nucleótidos entre 2 y 50, (*p. ej.*, 5 o 6 pares de bases) medidos desde el extremo cercano de un sitio de unión al extremo cercano del otro sitio de unión.

45 **[0037]** En los métodos anteriormente mencionados para la recombinación dirigida, la cromatina celular se puede escindir en un sitio ubicado entre los sitios de unión de las dos proteínas de fusión. Los sitios de unión pueden estar en cadenas de ADN opuestas. Además, la expresión de las proteínas de fusión en la célula se puede lograr mediante la introducción de las proteínas en la célula o mediante la introducción de uno o más polinucleótidos en la célula, que se transcriben opcionalmente (si el polinucleótido es ADN), y el (los) transcrito(s) traducido(s), para producir las proteínas de fusión. Por ejemplo, dos polinucleótidos, cada uno de los cuales comprende secuencias que codifican una de las dos proteínas de fusión, se pueden introducir en una célula. Alternativamente, un único polinucleótido que comprende secuencias que codifican ambas proteínas de fusión se puede introducir en la célula.

55 **[0038]** Por lo tanto, un método para la sustitución de una región de interés en la cromatina celular (*p. ej.*, una secuencia genómica) con una primera secuencia de nucleótidos comprende: (a) la ingeniería de un primer dominio de unión de dedo de zinc a una segunda secuencia en la región de interés; (b) proporcionar un segundo dominio de unión de dedos de zinc para unirse a una tercera secuencia; y (c) poner en contacto una célula con:

- 60 (i) un primer polinucleótido que comprende la primera secuencia de nucleótidos;  
 (ii) un segundo polinucleótido que codifica una primera proteína de fusión, comprendiendo la primera proteína de fusión el primer dominio de unión de dedos de zinc y un primer medio dominio de escisión; y  
 (iii) un tercer polinucleótido que codifica una segunda proteína de fusión, comprendiendo la segunda proteína de fusión el segundo dominio de unión de dedos de zinc y un segundo medio dominio de escisión;

65 en donde las proteínas de fusión primera y segunda se expresan, la primera proteína de fusión se une a la segunda secuencia y la segunda proteína de fusión se une a la tercera secuencia, colocando así los medios dominios de

escisión de modo que la cromatina celular se escinde en la región de interés; y la región de interés se reemplaza con la primera secuencia de nucleótidos.

**[0039]** En métodos preferidos de recombinación dirigida y/o sustitución y/o alteración de una secuencia en una región de interés en la cromatina celular, una secuencia cromosómica se altera por recombinación homóloga con una secuencia de nucleótidos de "donante" exógeno. Dicha recombinación homóloga es estimulada por la presencia de una ruptura bicatenaria en la cromatina celular, si están presentes secuencias homólogas a la región de la ruptura. Las roturas de doble cadena en la cromatina celular también pueden estimular los mecanismos celulares de unión final no homóloga.

**[0040]** En cualquiera de los métodos descritos en el presente documento, la primera secuencia de nucleótidos (la "secuencia donadora") puede contener secuencias que son homólogas, pero no idénticas, a las secuencias genómicas en la región de interés, estimulando así la recombinación homóloga para insertar una secuencia no idéntica en la región de interés. Por lo tanto, las porciones de la secuencia donante que son homólogas a las secuencias en la región de interés pueden exhibir entre aproximadamente el 80 y el 99% (o cualquier número entero entre ellas) de identidad de secuencia con la secuencia genómica que se reemplaza. La homología entre el donante y la secuencia genómica puede ser superior al 99%, p. ej., si solo 1 nucleótido difiere entre el donante y las secuencias genómicas de más de 100 pares de bases contiguas. En ciertos casos, una porción no homóloga de la secuencia donante puede contener secuencias que no están presentes en la región de interés, de modo que se introducen nuevas secuencias en la región de interés. En estos casos, la secuencia no homóloga generalmente está flanqueada por secuencias de 50-1000 pares de bases (o cualquier valor integral entre ellas) o cualquier número de pares de bases mayor que 1000, que son homólogas o idénticas a las secuencias en la región de interés. En otros casos, la secuencia donante puede ser no homóloga a la primera secuencia y se inserta en el genoma mediante mecanismos de recombinación no homóloga.

**[0041]** En los métodos de recombinación dirigida y/o sustitución y/o alteración de una secuencia de interés en cromatina celular, los medios dominios primero y segundo de escisión se pueden derivar de la misma endonucleasa o de diferentes endonucleasas. Las endonucleasas incluyen, pero no se limitan a, endonucleasas dirigidas y endonucleasas de restricción. Las endonucleasas de restricción ejemplares son las endonucleasas de restricción de Tipo IIS; una endonucleasa de restricción de tipo IIS ejemplar es *Fok I*.

**[0042]** La región de interés puede estar en un cromosoma, episoma o genoma organular. La región de interés puede comprender una mutación, que puede reemplazarse por una secuencia de tipo silvestre (o por una secuencia mutante diferente), o la región de interés puede contener una secuencia de tipo silvestre que se reemplaza por una secuencia mutante o un alelo diferente. Las mutaciones incluyen, pero no se limitan a, mutaciones puntuales (transiciones, transversiones), inserciones de uno o más pares de nucleótidos, deleciones de uno o más pares de nucleótidos, reordenamientos, inversiones y translocaciones. Las mutaciones pueden cambiar la secuencia de codificación, introducir codones de parada prematuros y/o modificar la frecuencia de un motivo de secuencia repetitiva (p. ej., repetición de trinucleótidos) en un gen. Para aplicaciones en las que la recombinación dirigida se usa para reemplazar una secuencia mutante, la cromatina celular generalmente se escinde en un sitio ubicado dentro de 100 nucleótidos a cada lado de la mutación, aunque también se pueden usar sitios de escisión ubicados hasta 6-10 kb del sitio de una mutación.

**[0043]** En cualquiera de los métodos descritos en el presente documento, el segundo dominio de unión de dedos de cinc puede ser diseñado, por ejemplo diseñado y/o seleccionado.

**[0044]** Además, el polinucleótido donante puede ser ADN o ARN, puede ser lineal o circular, y puede ser de cadena sencilla o de doble cadena. Puede administrarse a la célula como ácido nucleico desnudo, como un complejo con uno o más agentes de administración (p. ej., liposomas, poloxámeros) o contenido en un vehículo de administración viral, como, p. ej., un adenovirus o un virus adenoasociado (AAV). Las secuencias del donante pueden variar en longitud de 10 a 1000 nucleótidos (o cualquier valor integral de nucleótidos entre ellos) o más.

**[0045]** De manera similar, los polinucleótidos que codifican las fusiones entre un dominio de unión de dedos de cinc y un dominio de escisión o medio-dominio puede ser ADN o ARN, puede ser lineal o circular, y puede ser de cadena sencilla o de doble cadena. Pueden administrarse a la célula como ácido nucleico desnudo, como un complejo con uno o más agentes de administración (p. ej., liposomas, poloxámeros) o contenidos en un vehículo de administración viral, como, p. ej., un adenovirus o un virus adenoasociado. (AAV) Un polinucleótido puede codificar una o más proteínas de fusión.

**[0046]** En los métodos para la recombinación dirigida descritos aquí, al igual que con los métodos para la escisión específica, de un dominio de escisión o medio-dominio pueden derivarse de cualquier nucleasa, p. ej., una endonucleasa homing o una endonucleasa de restricción, en particular, una endonucleasa de restricción de tipo IIS. Los medios dominios de escisión pueden derivar de las mismas o de diferentes endonucleasas. Una fuente ejemplar, de la que un medio en el dominio de escisión se puede derivar, es la endonucleasa de restricción de tipo IIS *Fok I*.

**[0047]** La frecuencia de recombinación homóloga se puede mejorar mediante la detención de las células en la fase G2 del ciclo celular y/o mediante la activación de la expresión de una o más moléculas (proteínas, ARN) implicadas

en la recombinación homóloga y/o mediante la inhibición de la expresión o actividad de proteínas implicadas en la unión final no homóloga.

#### BREVE DESCRIPCIÓN DE LOS DIBUJOS

5

[0048]

10

**La Figura 1** muestra la secuencia de nucleótidos, en forma bicatenaria, de una porción del gen hSMC1L1 humano que codifica la porción amino terminal de la proteína (SEQ ID NO:1) y la secuencia de aminoácido codificada (SEQ ID NO:2). Las secuencias diana para las ZFP específicas de hSMC1 están subrayadas (una en cada cadena de ADN).

**La Figura 2** muestra un diagrama esquemático de un plásmido que codifica una fusión ZFP-FokI para la escisión dirigida del gen hSMC1.

15

**La Figura 3 A-D** muestra un diagrama esquemático del gen hSMC1. La Figura 3A muestra un esquema de una porción del cromosoma X humano que incluye el gen hSMC1. La Figura 3B muestra un esquema de una porción del gen hSMC1 que incluye la región aguas arriba (izquierda de +1), el primer exón (entre +1 y el extremo derecho de la flecha etiquetada "secuencia de codificación SMC1") y una porción del primer intrón. También se proporcionan ubicaciones de secuencias homólogas a los cebadores de amplificación iniciales y al cebador específico de cromosomas (ver Tabla 3). La Figura 3C muestra la secuencia de nucleótidos del cromosoma X humano en la región del codón de iniciación SMC1 (SEQ ID NO:3), la secuencia de aminoácidos codificada (SEQ ID NO:4) y los sitios diana para las proteínas de los dedos zinc específicos de SMC1. La Figura 3D muestra la secuencia de la región correspondiente de la molécula donante (SEQ ID NO:5), con las diferencias entre las secuencias donantes y cromosómicas subrayadas. Las secuencias contenidas en el cebador de amplificación específico del donante (Tabla 3) se indican mediante doble subrayado.

20

**La Figura 4** muestra un diagrama esquemático de la construcción del donante hSMC1.

**La Figura 5** muestra el análisis por PCR del ADN de células HEK293 transfectadas. Desde la izquierda, los carriles muestran resultados de células transfectadas con un plásmido que codifica GFP (plásmido de control), células transfectadas con dos plásmidos, cada uno de los cuales codifica una de las dos proteínas de fusión ZFP-FokI específicas de hSMC1 (solo ZFP), células transfectadas con dos concentraciones del plásmido donante hSMC1 (solo donante) y células transfectadas con los dos plásmidos que codifican ZFP y el plásmido donante (ZFP + donante). Vea el Ejemplo 1 para más detalles.

30

**La Figura 6** muestra la secuencia de nucleótidos de un producto de amplificación derivado de un gen hSMC1 mutado (SEQ ID NO:6) generado por recombinación homóloga dirigida. Las secuencias derivadas del vector en donde se clonó el producto de amplificación están subrayadas individualmente, las secuencias cromosómicas que no están presentes en la molécula donante se indican mediante subrayado discontinuo (nucleótidos 32-97), las secuencias comunes al donante y el cromosoma no están subrayadas (nucleótidos 98-394 y 402-417), y las secuencias únicas del donante están doblemente subrayadas (nucleótidos 395-401). Las letras minúsculas representan secuencias que difieren entre el cromosoma y el donante.

35

**La Figura 7** muestra la secuencia de nucleótidos de una porción del gen IL2R $\gamma$  humano que comprende el extremo 3' del segundo intrón y el extremo 5' del tercer exón (SEQ ID NO:7) y la secuencia de aminoácidos codificada por la porción mostrada de el tercer exón (SEQ ID NO:8). Las secuencias diana para el segundo par de ZFP específicas de IL2R $\gamma$  están subrayadas. Vea el Ejemplo 2 para más detalles.

40

**La Figura 8** muestra un diagrama esquemático de un plásmido que codifica una fusión ZFP-FokI para la escisión dirigida del gen IL2R $\gamma$ .

45

**La Figura 9 A-D** muestra un diagrama esquemático del gen IL2R $\gamma$ . La Figura 9A muestra un esquema de una porción del cromosoma X humano que incluye el gen IL2R $\gamma$ . La Figura 9B muestra un esquema de una porción del gen IL2R $\gamma$  que incluye una porción del segundo intrón, el tercer exón y una porción del tercer intrón. También se proporcionan ubicaciones de secuencias homólogas a los cebadores de amplificación iniciales y al cebador específico de cromosomas (ver Tabla 5). La Figura 9C muestra la secuencia de nucleótidos del cromosoma X humano en la región del tercer exón del gen IL2R $\gamma$  (SEQ ID NO:9), la secuencia de aminoácidos codificada (SEQ ID NO:10) y los sitios diana para el primer par de IL2R gamma - proteínas específicas de los dedos de zinc. La Figura 9D muestra la secuencia de la región correspondiente de la molécula donante (SEQ ID NO:11), con las diferencias entre las secuencias donantes y cromosómicas subrayadas. Las secuencias contenidas en el cebador de amplificación específico del donante (Tabla 5) se indican mediante doble sobre-revestimiento.

50

**La Figura 10** muestra un diagrama esquemático de la construcción del donante IL2R $\gamma$ .

55

**La Figura 11** muestra el análisis por PCR del ADN de células K652 transfectadas. Desde la izquierda, los carriles muestran resultados de células transfectadas con dos plásmidos, cada uno de los cuales codifica uno de un par de proteínas de fusión ZFP-FokI específicas de IL2R $\gamma$  (solo ZFP, carril 1), células transfectadas con dos concentraciones del plásmido donante IL2R $\gamma$  (solo donante, carriles 2 y 3), y células transfectadas con los dos plásmidos que codifican ZFP y el plásmido donante (ZFP + donante, carriles 4-7). Se usaron cada uno de los dos pares de fusiones específicas de IL2R $\gamma$  ZFP-FokI (identificadas como "par 1" y "par 2") y el uso de ambos pares dio como resultado la producción del producto de amplificación de diagnóstico (etiquetado como "producto químico esperado" en la figura). Vea el ejemplo 2 para más detalles.

60

**La Figura 12** muestra la secuencia de nucleótidos de un producto de amplificación derivado de un gen IL2R $\gamma$

65

mutado (SEQ ID NO:12) generado por recombinación homóloga dirigida. Las secuencias derivadas del vector en donde se clonó el producto de amplificación están subrayadas individualmente, las secuencias cromosómicas que no están presentes en la molécula del donante se indican mediante subrayado discontinuo (nucleótidos 460-552), las secuencias comunes al donante y el cromosoma no están subrayadas (nucleótidos 32-42 y 59-459), y un tramo de secuencia que contiene nucleótidos que distinguen las secuencias de donantes de las secuencias cromosómicas está doblemente subrayado (nucleótidos 44-58). Las letras minúsculas representan nucleótidos cuya secuencia difiere entre el cromosoma y el donante.

**La Figura 13** muestra la secuencia de nucleótidos de una porción de los segmentos que codifican el gen de la beta-globina humana del promotor central, los dos primeros exones y el primer intrón (SEQ ID NO:13). Una mutación sin sentido que cambia una A (en negrita y subrayada) en la posición 5212541 en el cromosoma 11 (BLAT, sitio de bioinformática del genoma UCSC) a una T produce anemia falciforme. Se diseñó una primera proteína de fusión de dedo de zinc/*Fok I* de manera que los contactos primarios estuvieran con la secuencia subrayada de 12 nucleótidos AAGGTGAACGTG (nucleótidos 305-316 de la SEQ ID NO:13), y se diseñó una segunda proteína de fusión de dedo de zinc/*Fok I* tal que los contactos primarios estaban con el complemento de la secuencia subrayada de 12 nucleótidos CCGTTACTGCC (nucleótidos 325-336 de SEQ ID NO:13).

**La Figura 14** es un diagrama esquemático de un plásmido que codifica la fusión ZFP-*Fok I* para la escisión dirigida del gen de la beta globina humana.

**La Figura 15** es un diagrama esquemático del gen de la beta globina humana clonada que muestra la región aguas arriba, el primer y el segundo exón, el primer intrón y los sitios de unión del cebador.

**La Figura 16** es un diagrama esquemático del constructo donante de beta globina, donante pCR4-TOPO-HBB.

**La Figura 17** muestra el análisis por PCR del ADN de células transfectadas con dos pares de nucleasas de ZFP específicas de  $\beta$ -globina y un plásmido donador de beta globina. El panel de la izquierda es un control de carga en donde se utilizaron los cebadores iniciales amp 1 y amp 2 iniciales (Tabla 7) para la amplificación. En el experimento que se muestra en el panel derecho, se usaron los cebadores "específicos a cromosomas" y "específicos a donante" (Tabla 7) para la amplificación. El carril de la izquierda en cada panel contiene marcadores de peso molecular y el siguiente carril muestra los productos de amplificación obtenidos de células transfectadas por simulacro. Los carriles restantes, de izquierda a derecha, producto de amplificación espectáculo a partir de células transfectadas con: un plásmido que codifica GFP, 100 ng de cada plásmido de codificación ZFP/*Fok I*, 200 ng de cada plásmido de codificación ZFP/*Fok I*, 200 ng de plásmido donador, 600 ng de plásmido donador, 200 ng de plásmido donador + 100 ng de cada plásmido codificador ZFP/*Fok I*, y 600 ng de plásmido donador + 200 ng de cada plásmido codificador ZFP/*Fok I*.

**La Figura 18** muestra la secuencia de nucleótidos de un producto de amplificación derivado de un gen de beta-globina mutada (SEQ ID NO:14) generado por recombinación homóloga dirigida. Las secuencias cromosómicas que no están presentes en la molécula del donante se indican mediante subrayado discontinuo (nucleótidos 1-72), secuencias comunes al donante y cromosomas no están subrayados (nucleótidos 73-376), y un tramo de secuencia que contiene nucleótidos que distinguen las secuencias de donantes de las secuencias cromosómicas está doblemente subrayado (nucleótidos 377-408). Las letras minúsculas representan nucleótidos cuya secuencia difiere entre el cromosoma y el donante.

**La Figura 19** muestra la secuencia de nucleótidos de una porción del quinto exón del gen de la cadena gamma del receptor de interleucina-2 (IL-2R $\gamma$ ) (SEQ ID NO:15). También se muestran las secuencias diana (subrayadas) para las proteínas de fusión 5-8 y 5-10 ZFP/*Fok I*. Vea el Ejemplo 5 para más detalles.

**La Figura 20** muestra la secuencia de aminoácidos de la fusión 5-8 ZFP/*Fok I* dirigida al exón 5 del gen humano IL-2R $\gamma$  (SEQ ID NO:16). Los residuos de aminoácidos 1-17 contienen una secuencia de localización nuclear (NLS, subrayado); los residuos 18-130 contienen la porción ZFP, con las regiones de reconocimiento de los dedos de zinc componentes mostrados en negrita; el enlazador ZFP-*Fok I* (enlazador ZC, subrayado) se extiende desde los residuos 131 a 140 y el medio dominio de escisión *Fok I* comienza en el residuo 141 y se extiende hasta el final de la proteína en el residuo 336. El residuo que se alteró para generar la mutación Q486E se muestra subrayado y en negrita.

**La Figura 21** muestra la secuencia de aminoácidos de la fusión 5-10 ZFP/*Fok I* dirigida al exón 5 del gen humano IL-2R $\gamma$  (SEQ ID NO:17). Los residuos de aminoácidos 1-17 contienen una secuencia de localización nuclear (NLS, subrayado); los residuos 18-133 contienen la porción ZFP, con las regiones de reconocimiento de los dedos de zinc componentes mostrados en negrita; el enlazador ZFP-*Fok I* (enlazador ZC, subrayado) se extiende desde los residuos 134 a 143 y el medio dominio de escisión *Fok I* comienza en el residuo 144 y se extiende hasta el final de la proteína en el residuo 339. El residuo que fue alterado para generar la mutación E490K se muestra subrayada y en negrita.

**La Figura 22** muestra la secuencia de nucleótidos del gen de la proteína fluorescente verde mejorada (SEQ ID NO:18) derivado del gen *Aequorea victoria* GFP (Tsien (1998) Ann. Rev. Biochem. 67: 509-544). El codón de iniciación de ATG, así como la región que se mutagenizó, están subrayados.

**La Figura 23** muestra la secuencia de nucleótidos de un gen eGFP defectuoso mutante (SEQ ID NO:19). Los sitios de unión para las nucleasas ZFP están subrayados y la región entre los sitios de unión corresponde a la región que se modificó.

**La Figura 24** muestra las estructuras de plásmidos que codifican Nucleasas de Dedo de Zinc dirigidas al gen eGFP.

**La Figura 25** muestra un autorradiograma de un gel de acrilamida al 10% utilizado para analizar la escisión de ADN dirigida de un gen eGFP mutante por endonucleasas de dedo de zinc. Vea el Ejemplo 8 para más

detalles.

**La Figura 26** muestra la estructura del plásmido pcDNA4/TO/GFPmut (véase el Ejemplo 9).

**La Figura 27** muestra los niveles de ARNm de eGFPmut, normalizados a ARNm de GAPDH, en diversas líneas celulares obtenidas de la transfección de células HEK293 humanas. Las barras de luz muestran niveles en células no tratadas; Las barras oscuras muestran niveles en células que habían sido tratadas con doxiciclina 2 ng/ml. Vea el ejemplo 9 para más detalles.

**La Figura 28** muestra la estructura del plásmido pCR (R) 4-TOPO-GFPdonante5. Vea el Ejemplo 10 para más detalles.

**La Figura 29** muestra la secuencia de nucleótidos del inserto eGFP en pCR (R) 4-TOPO-GFPdonante5 (SEQ ID NO:20). El inserto contiene secuencias que codifican una porción de una proteína fluorescente verde mejorada no modificada, que carece de un codón de iniciación. Vea el Ejemplo 10 para más detalles.

**La Figura 30** muestra una traza FACS de células T18 transfectadas con plásmidos que codifican dos nucleasas ZFP y un plásmido que codifica una secuencia donante, que fueron detenidos en la fase G2 del ciclo celular 24 horas después de la transfección con 100 ng/ml de nocodazol durante 48 horas. Se reemplazó el medio y se permitió que las células se recuperaran durante 48 horas adicionales, y se midió la corrección génica mediante análisis FACS. Vea el ejemplo 11 para más detalles.

**La Figura 31** muestra una traza FACS de células T18 transfectadas con plásmidos que codifican dos nucleasas ZFP y un plásmido que codifica una secuencia donante, que fueron detenidas en la fase G2 del ciclo celular 24 horas después de la transfección con vinblastina 0,2  $\mu$ M durante 48 horas. Se reemplazó el medio y se permitió que las células se recuperaran durante 48 horas adicionales, y se midió la corrección génica mediante análisis FACS. Vea el Ejemplo 11 para más detalles.

**La Figura 32** muestra la secuencia de nucleótidos de un inserto de eGFP de 1527 nucleótidos en pCR(R)4-TOPO (SEQ ID NO:21). La secuencia codifica una proteína fluorescente verde mejorada no modificada que carece de un codón de iniciación. Vea el Ejemplo 13 para más detalles.

**La Figura 33** muestra un diagrama esquemático de un ensayo utilizado para medir la frecuencia de edición del gen endógeno humano IL-2R $\gamma$ . Vea el Ejemplo 14 para más detalles.

**La Figura 34** muestra autorradiogramas de geles de acrilamida utilizados en un ensayo para medir la frecuencia de edición de un gen celular endógeno mediante escisión dirigida y recombinación homóloga. El carril marcado como "GFP" muestra los resultados del ensayo de un control en donde las células se transfectaron con un vector que codifica eGFP; el carril marcado como "ZFPs solamente" muestra los resultados de otro experimento de control en donde las células se transfectaron con los dos plásmidos que codifican ZFP/nucleasa (50 ng de cada uno) pero no con una secuencia donante. Los carriles marcados como "solo donante" muestran resultados de un experimento de control en donde las células se transfectaron con 1  $\mu$ g de plásmido donante pero no con los plásmidos que codifican ZFP/nucleasa. En los carriles experimentales, 50Z se refiere a células transfectadas con 50 ng de cada plásmido de expresión de ZFP/nucleasa, 100Z se refiere a células transfectadas con 100 ng de cada plásmido de expresión de ZFP/nucleasa, 0,5D se refiere a células transfectadas con 0,5  $\mu$ g del donante plásmido, y 1D se refiere a células transfectadas con 1,0  $\mu$ g del plásmido donante. "+" se refiere a células que se expusieron a vinblastina 0,2  $\mu$ m; "-" se refiere a células que no fueron expuestas a vinblastina. "wt" se refiere al fragmento obtenido después de la digestión con *Bsr* BI de los productos de amplificación obtenidos a partir de cromosomas que contienen el gen cromosómico IL-2R $\gamma$  de tipo silvestre; "rflp" se refiere a los dos fragmentos (de peso molecular aproximadamente igual) obtenidos después de la digestión con *Bsr*BI de productos de amplificación obtenidos a partir de secuencias que contienen cromosomas del plásmido donante que se había integrado por recombinación homóloga.

**La Figura 35** muestra una imagen autorradiográfica de una exposición de cuatro horas de un gel utilizado en un ensayo para medir la recombinación dirigida en el locus de IL-2R $\gamma$  humano en células K562. "wt" identifica una banda que es diagnóstica para ADN cromosómico que contiene la secuencia nativa K562 IL-2R $\gamma$ ; "rflp" identifica un diagnóstico de doblete para ADN cromosómico que contiene la secuencia IL-2R $\gamma$  alterada presente en la molécula de ADN donante. El símbolo "+" encima de un carril indica que las células fueron tratadas con vinblastina 0,2  $\mu$ M; el símbolo "-" indica que las células no fueron tratadas con vinblastina. Los números en los carriles del "donante ZFP +" indican el porcentaje de secuencia que contiene el ADN cromosómico total originalmente presente en la molécula de ADN del donante, calculado utilizando el "buscador de picos", función de línea de base automática del software ImageQuant v. 5,1 de Molecular Dynamics como se describe en el cap. 8 del manual del fabricante (Guía del usuario de Molecular Dynamics ImageQuant; parte 218-415). "Untr" indica células no transfectadas. Vea el Ejemplo 15 para detalles adicionales.

**La Figura 36** muestra una imagen autorradiográfica de una exposición de cuatro horas de un gel utilizado en un ensayo para medir la recombinación dirigida en el locus de IL-2R $\gamma$  humano en células K562. "wt" identifica una banda que es diagnóstica para ADN cromosómico que contiene la secuencia K562 IL-2R $\gamma$  nativa; "rflp" identifica una banda que es diagnóstica de ADN cromosómico que contiene la secuencia de IL-2R $\gamma$  alterada presente en la molécula de ADN donante. El símbolo "+" encima de un carril indica que las células fueron tratadas con vinblastina 0,2  $\mu$ M; el símbolo "-" indica que las células no fueron tratadas con vinblastina. Los números debajo de los carriles de "donante ZFP +" indican el porcentaje de secuencia que contiene ADN cromosómico total presente originalmente en la molécula de ADN donante, calculada como se describe en el Ejemplo 35. Vea el Ejemplo 15 para detalles adicionales.

**La Figura 37** muestra un autorradiograma de una exposición de cuatro horas de una transferencia de ADN

sondeada con un fragmento específico del gen de IL-2R $\gamma$  humano. La flecha a la derecha de la imagen indica la posición de una banda correspondiente al ADN genómico cuya secuencia ha sido alterada por recombinación homóloga. El símbolo "+" encima de un carril indica que las células fueron tratadas con vinblastina 0,2  $\mu$ M; el símbolo "-" indica que las células no fueron tratadas con vinblastina. Los números debajo de los carriles de "donante ZFP +" indican el porcentaje de secuencia que contiene ADN cromosómico total presente originalmente en la molécula de ADN donante, calculada como se describe en el Ejemplo 35. Vea el Ejemplo 15 para detalles adicionales.

**La Figura 38** muestra imágenes autorradiográficas de geles utilizados en un ensayo para medir la recombinación dirigida en el locus de IL-2R $\gamma$  humano en células de médula ósea humana CD34 $^{+}$ . El panel izquierdo muestra un estándar de referencia en donde el porcentaje establecido de ADN genómico humano normal (que contiene un sitio *Mae* II) se agregó al ADN genómico de las células Jurkat (que carece de un sitio *Mae* II), la mezcla se amplificó por PCR para generar un radiomarcado producto de amplificación, y el producto de amplificación se digirió con *Mae* II. "wt" identifica una banda que representa ADN no digerido, y "rflp" identifica una banda resultante de la digestión con *Mae* II.

El panel derecho muestra los resultados de un experimento en donde las células CD34 $^{+}$  se transfectaron con ADN del donante que contenía un sitio *Bsr*BI y plásmidos que codifican endonucleasas de fusión de dedo de zinc *Fok* I. La región genómica relevante se amplificó y se marcó, y el producto de amplificación marcado se digirió con *Bsr*BI. "GFP" indica células de control que se transfectaron con un plásmido que codifica GFP; "Donante solamente" indica células de control que se transfectaron solo con ADN del donante, y "Donante ZFP +" indica células que se transfectaron con ADN del donante y con plásmidos que codifican las nucleasas de dedo de zinc/*Fok* I. "wt" identifica una banda que es diagnóstica para ADN cromosómico que contiene la secuencia de IL-2R $\gamma$  nativa; "rflp" identifica una banda que es diagnóstica de ADN cromosómico que contiene la secuencia de IL-2R $\gamma$  alterada presente en la molécula de ADN donante. El carril de la derecha contiene marcadores de tamaño de ADN. Vea el Ejemplo 16 para detalles adicionales.

**La Figura 39** muestra una imagen de una inmunotransferencia utilizada para analizar los niveles de proteína Ku70 en células transfectadas con ARNip dirigido a Ku70. La línea celular T7 (Ejemplo 9, Figura 27) se transfectó con dos concentraciones cada una de ARNip de dos agrupaciones de ARNip diferentes (véase el Ejemplo 18). Carril 1: 70 ng de grupo de ARNip D; Carril 2: 140 ng de ARNip reserva D; Carril 3: 70 ng de grupo de ARNip E; Carril 4: 140 ng de grupo de ARNip E. "Ku70" indica la banda que representa la proteína Ku70; "TFIIB" indica una banda que representa el factor de transcripción TFIIB, utilizado como control.

**La Figura 40** muestra las secuencias de aminoácidos de cuatro dominios de dedos de zinc dirigidos al gen de la  $\beta$ -globina humana: sca-29b (SEQ ID NO:22); sca-36a (SEQ ID NO:23); sca-36b (SEQ ID NO:24) y sca-36c (SEQ ID NO:25). El sitio diana para el dominio sca-29b está en una cadena de ADN, y los sitios diana para los dominios sca-36a, sca-36b y sca-36c están en la cadena opuesta. Véase el Ejemplo 20.

**La Figura 41** muestra los resultados de un ensayo *in vitro*, en donde se probaron diferentes combinaciones de nucleasas de fusión de dedo de zinc/*Fok* I (ZFN) para determinar la escisión de ADN de secuencia específica. El carril con la etiqueta "U" muestra una muestra de la plantilla de ADN. Los siguientes cuatro carriles muestran los resultados de la incubación de la plantilla de ADN con cada uno de los cuatro ZFN dirigidos a  $\beta$ -globina (ver el Ejemplo 20 para la caracterización de estos ZFN). Los tres carriles más a la derecha muestran los resultados de la incubación de la plantilla de ADN con el ZFN sca-29b y uno de los ZFN sca-36a, sca-36b o sca-36c (todos los cuales están dirigidos a la cadena opuesta a la que se dirige sca-29b).

**La Figura 42** muestra los niveles de ARNm de eGFP en células T18 (barras) en función de la concentración de doxiciclina (proporcionada en la abscisa). El número sobre cada barra representa el porcentaje de corrección de la mutación eGFP, en células transfectadas con ADN de donante y plásmidos que codifican nucleasas de dedos de zinc dirigidas a eGFP, en función de la concentración de doxiciclina.

#### DESCRIPCIÓN DETALLADA

**[0049]** En el presente documento se describen composiciones y métodos útiles para la escisión dirigida de cromatina celular y para la alteración dirigida de una secuencia de nucleótidos celular, p. ej., mediante escisión dirigida seguida de unión final no homóloga o escisión dirigida seguida de recombinación homóloga entre un polinucleótido exógeno (que comprende una o más regiones de homología con la secuencia de nucleótidos celulares) y una secuencia genómica. Las secuencias genómicas incluyen aquellas presentes en cromosomas, episomas, genomas orgánulos (p. ej., mitocondrias, cloroplastos), cromosomas artificiales y cualquier otro tipo de ácido nucleico presente en una célula como, p. ej., secuencias amplificadas, cromosomas de doble minuto y genomas endógenos o infectar bacterias y virus. Las secuencias genómicas pueden ser normales (es decir, de tipo silvestre) o mutantes; las secuencias mutantes pueden comprender, p. ej., inserciones, deleciones, translocaciones, reordenamientos y/o mutaciones puntuales. Una secuencia genómica también puede comprender uno de varios alelos diferentes.

**[0050]** Las composiciones útiles para la escisión específica y recombinación incluyen proteínas de fusión que comprenden una escisión de dominio (o una medio dominio de escisión) y un dominio de unión de dedos de cinc, los polinucleótidos que codifican estas proteínas y combinaciones de polipéptidos y polinucleótidos que codifican el polipéptido. Un dominio de unión de dedos de cinc puede comprender uno o más dedos de cinc (p. ej., 2, 3, 4, 5, 6, 7, 8, 9 o más dedos de cinc), y puede diseñarse para unirse a cualquier secuencia genómica. Así, al identificar una región genómica diana de interés en donde se desea la escisión o la recombinación, se puede, de acuerdo con los métodos

descritos aquí, construir una o más proteínas de fusión que comprenden un dominio de escisión (o medio dominio de escisión) y un dominio de dedo de zinc diseñado para reconocer una secuencia diana en dicha región genómica. La presencia de dicha proteína (o proteínas) de fusión en una célula dará como resultado la unión de la(s) proteína(s) de fusión a su(s) sitio(s) de unión y escisión dentro o cerca de dicha región genómica. Además, si un polinucleótido exógeno homólogo a la región genómica también está presente en dicha célula, la recombinación homóloga se produce a una velocidad alta entre la región genómica y el polinucleótido exógeno.

### General

**[0051]** La práctica de los métodos, así como la preparación y uso de las composiciones descritas en el presente documento emplean, a menos que se indique lo contrario, técnicas convencionales en biología molecular, bioquímica, estructura y análisis de cromatina, química computacional, cultivo celular, ADN recombinante y campos relacionados. como están dentro de la habilidad de la técnica. Estas técnicas se explican completamente en la literatura. *Ver, p. ej.,* Sambrook et al. *MOLECULAR CLONING: A LABORATORY MANUAL*, Segunda edición, Cold Spring Harbor Laboratory Press, 1989 y Tercera edición, 2001; Ausubel et al., *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, John Wiley & Sons, Nueva York, 1987 y actualizaciones periódicas; la serie *METHODS IN ENZYMOLOGY*, Academic Press, San Diego; Wolffe, *CHROMATIN STRUCTURE AND FUNCTION*, Tercera edición, Academic Press, San Diego, 1998; *METHODS IN ENZYMOLOGY*, vol. 304, "Chromatin" (PM Wassarman y AP Wolffe, eds.), Academic Press, San Diego, 1999; y *METHODS IN MOLECULAR BIOLOGY*, vol. 119, "Chromatin Protocols" (PB Becker, ed.) Humana Press, Totowa, 1999.

### Definiciones

**[0052]** Los términos "ácido nucleico", "polinucleótido" y "oligonucleótido" se usan indistintamente y se refieren a un polímero de desoxirribonucleótidos o ribonucleótidos, en conformación lineal o circular, y en forma monocatenaria o bicatenaria. Para los fines de la presente divulgación, estos términos no deben interpretarse como limitantes con respecto a la longitud de un polímero. Los términos pueden abarcar análogos conocidos de nucleótidos naturales, así como nucleótidos que se modifican en los restos base, azúcar y/o fosfato (*p. ej.,* esqueletos de fosforotioato). En general, un análogo de un nucleótido particular tiene la misma especificidad de emparejamiento de bases; *es decir,* un análogo de A se emparejará con T.

**[0053]** Los términos "polipéptido", "péptido" y "proteína" se utilizan indistintamente para referirse a un polímero de aminoácidos residuos. El término también se aplica a los polímeros de aminoácidos en los que uno o más aminoácidos son análogos químicos o derivados modificados de los aminoácidos de origen natural correspondientes.

**[0054]** "Unión" se refiere a una secuencia específica, la interacción no covalente entre macromoléculas (*p. ej.,* entre una proteína y un ácido nucleico). No todos los componentes de una interacción de unión deben ser específicos de secuencia (*p. ej.,* contactos con residuos de fosfato en una cadena principal de ADN), siempre que la interacción en su conjunto sea específica de secuencia. Tales interacciones se caracterizan generalmente por una constante de disociación ( $K_d$ ) de  $10^{-6} M^{-1}$  o inferior. "Afinidad" se refiere a la fuerza de unión: el aumento de la afinidad de unión se correlaciona con una  $K_d$  menor.

**[0055]** Una "proteína de unión" es una proteína que es capaz de unirse no covalentemente a otra molécula. Una proteína de unión puede unirse, *p. ej.,* a una molécula de ADN (una proteína de unión a ADN), una molécula de ARN (una proteína de unión a ARN) y/o una molécula de proteína (una proteína de unión a proteínas). En el caso de una proteína de unión a proteínas, puede unirse a sí misma (para formar homodímeros, homotrímeros, *etc.*) y/o puede unirse a una o más moléculas de una proteína o proteínas diferentes. Una proteína de unión puede tener más de un tipo de actividad de unión. Por ejemplo, las proteínas de dedo de zinc tienen actividad de unión al ADN, unión al ARN y unión a proteínas.

**[0056]** Una "proteína de unión a ADN de dedo de zinc" (o dominio de unión) es una proteína, o un dominio dentro de una proteína más grande, que se une al ADN de una manera específica de la secuencia a través de uno o más dedos de zinc, que son regiones de secuencia de aminoácido dentro del dominio de unión cuya estructura se estabiliza mediante la coordinación de un ion zinc. El término proteína de unión al ADN del dedo de zinc a menudo se abrevia como proteína de dedo de zinc o ZFP.

**[0057]** Dominios de unión a dedo de zinc pueden ser "modificados" para unirse a una secuencia de nucleótidos predeterminada. Ejemplos no limitantes de métodos para diseñar proteínas de dedo de zinc son diseño y selección. Una proteína de dedo de zinc diseñada es una proteína que no ocurre en la naturaleza cuyo diseño/composición resulta principalmente de criterios racionales. Los criterios racionales para el diseño incluyen la aplicación de reglas de sustitución y algoritmos computarizados para procesar información en una base de datos que almacena información de diseños ZFP existentes y datos vinculantes. Véanse, *p. ej.,* las patentes de los Estados Unidos 6,140,081; 6,453,242; y 6,534,261; ver también WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 y WO 03/016496.

**[0058]** Una proteína con dedos de zinc "seleccionada" es una proteína que no se encuentra en la naturaleza cuya producción resulta principalmente de un proceso empírico tal como presentación de fagos, la trampa de interacción o

selección de híbridos. Véanse, p. ej., los documentos US 5,789,538; US 5,925,523; US 6,007,988; US 6,013,453; US 6,200,759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197 y WO 02/099084.

5 **[0059]** El término "secuencia" se refiere a una secuencia de nucleótidos de cualquier longitud, que puede ser ADN o ARN; puede ser lineal, circular o ramificado y puede ser monocatenario o bicatenario. El término "secuencia donante" se refiere a una secuencia de nucleótidos que se inserta en un genoma. Una secuencia donante puede tener cualquier longitud, por ejemplo entre 2 y 10.000 nucleótidos de longitud (o cualquier valor entero entre ellos o por encima), preferiblemente entre aproximadamente 100 y 1000 nucleótidos de longitud (o cualquier número entero entre ellos),  
10 más preferiblemente entre aproximadamente 200 y 500 nucleótidos de longitud.

**[0060]** Una "secuencia no idéntica homóloga" se refiere a una primera secuencia que comparte un grado de identidad de secuencia con una segunda secuencia, pero cuya secuencia no es idéntica a la de la segunda secuencia. Por ejemplo, un polinucleótido que comprende la secuencia de tipo silvestre de un gen mutante es homólogo y no idéntico a la secuencia del gen mutante. En ciertas realizaciones, el grado de homología entre las dos secuencias es suficiente para permitir la recombinación homóloga entre ellas, utilizando mecanismos celulares normales. Dos secuencias homólogas no idénticas pueden tener cualquier longitud y su grado de no homología puede ser tan pequeño como un solo nucleótido (p. ej., para la corrección de una mutación puntual genómica por recombinación homóloga dirigida) o tan grande como 10 o más kilobases (p. ej., para la inserción de un gen en un sitio ectópico predeterminado en un cromosoma). No es necesario que dos polinucleótidos que comprenden las secuencias homólogas no idénticas tengan la misma longitud. Por ejemplo, se puede usar un polinucleótido exógeno (es decir, polinucleótido donante) de entre  
15 20 y 10.000 nucleótidos o pares de nucleótidos.

**[0061]** Las técnicas para determinar el ácido nucleico y de aminoácidos de identidad de secuencia de ácido son conocidos en la técnica. Típicamente, tales técnicas incluyen determinar la secuencia de nucleótidos del ARNm para un gen y/o determinar la secuencia de aminoácidos codificada por el mismo, y comparar estas secuencias con una segunda secuencia de nucleótidos o aminoácidos. Las secuencias genómicas también se pueden determinar y comparar de esta manera. En general, la identidad se refiere a una correspondencia exacta de nucleótido a nucleótido o de aminoácido a aminoácido de dos polinucleótidos o secuencias de polipéptidos, respectivamente. Se pueden comparar dos o más secuencias (polinucleótidos o aminoácidos) determinando su porcentaje de identidad. El porcentaje de identidad de dos secuencias, ya sea de ácido nucleico o de aminoácidos, es el número de coincidencias exactas entre dos secuencias alineadas divididas por la longitud de las secuencias más cortas y multiplicadas por 100. El alineamiento local aproximado para las secuencias de ácidos nucleicos es proporcionado por algoritmo de homología local de Smith y Waterman, *Advances in Applied Mathematics* 2: 482-489 (1981). Este algoritmo se puede aplicar a las secuencias de aminoácidos utilizando la matriz de puntuación desarrollada por Dayhoff, *Atlas of Protein Sequences and Structure*, MO Dayhoff ed., 5 supl. 3: 353-358, National Biomedical Research Foundation, Washington, DC, EE.UU., y normalizado por Gribskov, *Nucl. Acidos Res.* 14 (6): 6745-6763 (1986). Genetics Computer Group (Madison, WI) proporciona una implementación ejemplar de este algoritmo para determinar el porcentaje de identidad de una secuencia en la aplicación de utilidad "BestFit". Los parámetros predeterminados para este método se describen en el Wisconsin Sequence Analysis Package Program Manual, Versión 8 (1995) (disponible de Genetics Computer Group, Madison, WI). Un método preferido para establecer el porcentaje de identidad en el contexto de la presente divulgación es usar el paquete de programas MPSRCH con derechos de autor de la Universidad de Edimburgo, desarrollado por John F. Collins y Shane S. Sturrok, y distribuido por IntelliGenetics, Inc. (Mountain View, CA). A partir de este conjunto de paquetes, se puede utilizar el algoritmo Smith-Waterman donde se utilizan parámetros predeterminados para la tabla de puntuación (p. ej., penalización por apertura de hueco de 12, penalización de extensión de hueco de uno y un hueco de seis). A partir de los datos generados, el valor "Match" refleja la identidad de secuencia. Otros programas adecuados para calcular el porcentaje de identidad o similitud entre secuencias son generalmente conocidos en la técnica, p. ej., otro programa de alineación es BLAST, usado con parámetros predeterminados. Por ejemplo, BLASTN y BLASTP pueden usarse usando los siguientes parámetros predeterminados: código genético = estándar; filtro = ninguno; cadena = ambos; corte = 60; esperar = 10; Matriz = BLOSUM62; Descripciones = 50 secuencias; ordenar por = PUNTUACIÓN ALTA; Bases de datos = no redundantes, GenBank + EMBL + DDBJ + PDB + Traducciones de CDS GenBank + proteína suiza + Spupdate + PIR. Los detalles de estos programas se pueden encontrar en la siguiente dirección de Internet: <http://www.ncbi.nlm.gov/cgi-bin/BLAST>. Con respecto a las secuencias descritas en el presente documento, el intervalo de grados de identidad de secuencia deseados es aproximadamente del 80% al 100% y cualquier valor entero entre ellos. Típicamente, las identidades porcentuales entre secuencias son al menos 70-75%, preferiblemente 80-82%, más preferiblemente 85-90%, incluso más preferiblemente 92%, aún más preferiblemente 95%, y lo más preferiblemente 98% de identidad de secuencia.  
25 30 35 40 45 50 55

**[0062]** Alternativamente, el grado de similitud de secuencia entre los polinucleótidos puede determinarse por hibridación de polinucleótidos en condiciones que permiten la formación de dúplex estables entre regiones homólogas, seguido de digestión con nucleasa específica de cadena sencilla, y la determinación del tamaño de los fragmentos digeridos. Dos ácidos nucleicos, o dos secuencias de polipéptidos son sustancialmente homólogos entre sí cuando las secuencias exhiben al menos aproximadamente 70%-75%, preferiblemente 80%-82%, más preferiblemente 85%-90%, incluso más preferiblemente 92%, aún más preferiblemente 95%, y lo más preferiblemente 98% de identidad de secuencia sobre una longitud definida de las moléculas, según se determina usando los métodos anteriores. Como se usa en el presente documento, sustancialmente homólogo también se refiere a secuencias que muestran identidad  
60 65

completa con una secuencia de ADN o polipéptido especificada. Las secuencias de ADN que son sustancialmente homólogas pueden identificarse en un experimento de hibridación Southern en, p. ej., condiciones estrictas, como se define para ese sistema en particular. La definición de condiciones de hibridación apropiadas está dentro de la habilidad de la técnica. Ver, p. ej., Sambrook et al., *Supra*; Nucleic Acid Hybridization: A Practical Approach, editores BD Hames y S.J. Higgins, (1985) Oxford; Washington DC; IRL Press).

**[0063]** La hibridación selectiva de dos fragmentos de ácido nucleico se puede determinar como sigue. El grado de identidad de secuencia entre dos moléculas de ácido nucleico afecta la eficiencia y la fuerza de los eventos de hibridación entre dichas moléculas. Una secuencia de ácido nucleico parcialmente idéntica inhibirá al menos parcialmente la hibridación de una secuencia completamente idéntica a una molécula diana. La inhibición de la hibridación de una secuencia completamente idéntica se puede evaluar utilizando ensayos de hibridación que son bien conocidos en la técnica (p. ej., transferencia Southern (ADN), transferencia Northern (ARN), hibridación en solución o similares, ver Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, 2ª Edición, (1989) Cold Spring Harbor, NY). Dichos ensayos pueden llevarse a cabo utilizando diversos grados de selectividad, p. ej., utilizando condiciones que varían de baja a alta rigurosidad. Si se emplean condiciones de baja rigurosidad, la ausencia de unión no específica se puede evaluar utilizando una sonda secundaria que carece incluso de un grado parcial de identidad de secuencia (p. ej., una sonda que tiene menos del 30% de identidad de secuencia con la molécula diana), de modo que, en ausencia de eventos de unión no específicos, la sonda secundaria no se hibridará con la diana.

**[0064]** Cuando se utiliza un sistema de detección basado en hibridación, una sonda de ácido nucleico se elige que es complementaria a una secuencia de ácido nucleico de referencia, y luego mediante la selección de condiciones apropiadas la sonda y la referencia de secuencia selectivamente hibridan o se unen, el uno al otro para formar una molécula dúplex. Una molécula de ácido nucleico que es capaz de hibridarse selectivamente a una secuencia de referencia en condiciones de hibridación moderadamente estrictas, típicamente hibrida en condiciones que permiten la detección de una secuencia de ácido nucleico diana de al menos aproximadamente 10-14 nucleótidos de longitud que tiene al menos aproximadamente 70% de identidad de secuencia con la secuencia de la sonda de ácido nucleico seleccionada. Las condiciones de hibridación rigurosas típicamente permiten la detección de secuencias de ácido nucleico diana de al menos aproximadamente 10-14 nucleótidos de longitud que tienen una identidad de secuencia mayor de aproximadamente 90-95% con la secuencia de la sonda de ácido nucleico seleccionada. Las condiciones de hibridación útiles para la hibridación de la sonda/secuencia de referencia, donde la sonda y la secuencia de referencia tienen un grado específico de identidad de secuencia, se pueden determinar como se conoce en la técnica (véase, p. ej., *Nucleic Acid Hybridization: A Practical Approach*, editores BD Hames y SJ Higgins, (1985) Oxford; Washington, DC; IRL Press).

**[0065]** Las condiciones para la hibridación son bien conocidas por los expertos en la técnica. La rigurosidad de la hibridación se refiere al grado en que las condiciones de hibridación favorecen la formación de híbridos que contienen nucleótidos no coincidentes, con una rigurosidad más alta correlacionada con una tolerancia más baja para los híbridos no coincidentes. Los factores que afectan la rigurosidad de la hibridación son bien conocidos por los expertos en la materia e incluyen, pero no se limitan a, temperatura, pH, fuerza iónica y concentración de disolventes orgánicos como, p. ej., formamida y dimetilsulfóxido. Como saben los expertos en la materia, la rigurosidad de la hibridación aumenta con temperaturas más altas, menor fuerza iónica y menores concentraciones de solvente.

**[0066]** Con respecto a las condiciones de rigurosidad para la hibridación, es bien sabido en la técnica que numerosas condiciones equivalentes se pueden emplear para establecer una rigurosidad en particular variando, p. ej., los siguientes factores: la longitud y naturaleza de las secuencias, la composición de bases de las diversas secuencias, concentraciones de sales y otros componentes de la solución de hibridación, la presencia o ausencia de agentes bloqueantes en las soluciones de hibridación (p. ej., sulfato de dextrano y polietilenglicol), temperatura de reacción de hibridación y parámetros de tiempo, así como condiciones de lavado variables. La selección de un conjunto particular de condiciones de hibridación se selecciona siguiendo métodos estándar en la técnica (véase, p. ej., Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, 2ª Edición, (1989) Cold Spring Harbor, NY).

**[0067]** "Recombinación" se refiere a un proceso de intercambio de información genética entre dos polinucleótidos. Para los fines de esta divulgación, "recombinación homóloga (RH)" se refiere a la forma especializada de dicho intercambio que tiene lugar, p. ej., durante la reparación de roturas de doble cadena en las células. Este proceso requiere homología de secuencia de nucleótidos, utiliza una molécula "donante" para reparar la plantilla de una molécula "diana" (es decir, la que experimentó la ruptura de doble cadena), y se conoce de diversas maneras como "conversión génica no cruzada" o "conversión del gen del tracto corto", porque conduce a la transferencia de información genética del donante a la diana. Sin desear estar sujeto a ninguna teoría en particular, dicha transferencia puede involucrar una corrección de desajuste del ADN heterodúplex que se forma entre la diana rota y el donante, y/o el "recocido de cadena dependiente de síntesis", en donde el donante se utiliza para volver a sintetizar información genética que se convertirá en parte de la diana y/o procesos relacionados. Dicha RH especializada a menudo da como resultado una alteración de la secuencia de la molécula diana de tal manera que parte o toda la secuencia del polinucleótido donante se incorpora al polinucleótido diana.

**[0068]** "Escisión" se refiere a la rotura de la columna vertebral covalente de una molécula de ADN. La escisión puede iniciarse mediante una variedad de métodos que incluyen, entre otros, la hidrólisis enzimática o química de un enlace

fosfodiéster. Son posibles tanto la escisión de cadena simple como la escisión de doble cadena, y la escisión de doble cadena puede ocurrir como resultado de dos eventos distintos de una sola hebra de escisión. La escisión del ADN puede dar como resultado la producción de extremos romos o extremos escalonados. Los polipéptidos de fusión se pueden usar para la escisión de ADN bicatenario dirigido.

5 **[0069]** Un "dominio de escisión" comprende una o más secuencias de polipéptidos que posee actividad catalítica para el ADN de escisión. Un dominio de escisión puede estar contenido en una sola cadena de polipéptidos o la actividad de escisión puede resultar de la asociación de dos (o más) polipéptidos.

10 **[0070]** Un "medio dominio de escisión" es una secuencia de polipéptido que, en conjunción con un segundo polipéptido (ya sea idéntico o diferente) forma una actividad de escisión que tiene complejo (preferiblemente actividad de la doble cadena de escisión).

15 **[0071]** "Cromatina" es la estructura de nucleoproteína que comprende el genoma celular. La cromatina celular comprende ácido nucleico, principalmente ADN y proteínas, incluidas las histonas y las proteínas cromosómicas no histonas. La mayoría de la cromatina celular eucariota existe en forma de nucleosomas, en donde un núcleo de nucleosomas comprende aproximadamente 150 pares de bases de ADN asociados con un octamero que comprende dos de las histonas H2A, H2B, H3 y H4; y el ADN conector (de longitud variable dependiendo del organismo) se extiende entre núcleos de nucleosomas. Una molécula de histona H1 generalmente se asocia con el ADN conector. Para los fines de la presente divulgación, el término "cromatina" pretende abarcar todos los tipos de nucleoproteína celular, tanto procariotas como eucariotas. La cromatina celular incluye cromatina cromosómica y episomal.

20 **[0072]** Un "cromosoma", es un complejo de cromatina que comprende la totalidad o una porción del genoma de una célula. El genoma de una célula a menudo se caracteriza por su cariotipo, que es la colección de todos los cromosomas que comprenden el genoma de la célula. El genoma de una célula puede comprender uno o más cromosomas.

25 **[0073]** Un "episoma" es un ácido nucleico replicante, complejo de nucleoproteína, u otra estructura que comprende un ácido nucleico que no es parte del cariotipo cromosómico de una célula. Los ejemplos de episomas incluyen plásmidos y ciertos genomas virales.

30 **[0074]** Una "región accesible" es un sitio de la cromatina celular en donde un sitio diana presente en el ácido nucleico puede ser obligado por una molécula exógena que reconoce el sitio diana. Sin desear limitarse a ninguna teoría en particular, se cree que una región accesible es aquella que no está empaquetada en una estructura nucleosómica. La estructura distinta de una región accesible a menudo se puede detectar por su sensibilidad a sondas químicas y enzimáticas, p. ej., nucleasas.

35 **[0075]** Un "sitio diana" o "secuencia diana" es una secuencia de ácido nucleico que define una porción de un ácido nucleico al cual se unirá una molécula de unión, siempre que existan las condiciones suficientes para unión. Por ejemplo, la secuencia 5'-GAATTC-3' es un sitio diana para la endonucleasa de restricción Eco RI.

40 **[0076]** Una molécula "exógena" es una molécula que normalmente no está presente en una célula, pero puede ser introducida en una célula por uno o más métodos genéticos, bioquímicos u otros métodos. La "presencia normal en la célula" se determina con respecto a la etapa de desarrollo particular y las condiciones ambientales de la célula. Así, p. ej., una molécula que está presente solo durante el desarrollo embrionario del músculo es una molécula exógena con respecto a una célula muscular adulta. De manera similar, una molécula inducida por choque térmico es una molécula exógena con respecto a una célula no sometida a choque térmico. Una molécula exógena puede comprender, p. ej., una versión funcional de una molécula endógena que funciona mal o una versión que funciona mal de una molécula endógena que funciona normalmente.

45 **[0077]** Una molécula exógena puede ser, entre otras cosas, una molécula pequeña, como se genera por un proceso químico combinatorio, o una macromolécula tal como una proteína, ácido nucleico, carbohidrato, lípido, glicoproteína, lipoproteína, polisacárido, cualquier derivado modificado de las moléculas anteriores, o cualquier complejo que comprenda una o más de las moléculas anteriores. Los ácidos nucleicos incluyen ADN y ARN, pueden ser monocatenarios o bicatenarios; pueden ser lineales, ramificados o circulares; y pueden ser de cualquier longitud. Los ácidos nucleicos incluyen aquellos capaces de formar dúplex, así como los ácidos nucleicos formadores de triplex. Véanse, p. ej., las Patentes de EE.UU. N<sup>os</sup> 5,176,996 y 5,422,251. Las proteínas incluyen, pero no se limitan a, proteínas de unión al ADN, factores de transcripción, factores de remodelación de la cromatina, proteínas de unión al ADN metiladas, polimerasas, metilasas, desmetilasas, acetilasas, quinasas, fosfatasa, integrasas, recombinasas, ligasas, topoisomerasas, girasas y helicasas.

50 **[0078]** Una molécula exógena puede ser el mismo tipo de molécula como una molécula endógena, p. ej., una proteína exógena o ácido nucleico. Por ejemplo, un ácido nucleico exógeno puede comprender un genoma viral infectante, un plásmido o episoma introducido en una célula o un cromosoma que normalmente no está presente en la célula. Los expertos en la técnica conocen métodos para la introducción de moléculas exógenas en las células e incluyen, entre otros, transferencia mediada por lípidos (es decir, liposomas, incluidos lípidos neutros y catiónicos), electroporación, inyección directa, fusión celular, bombardeo de partículas, coprecipitación con fosfato de calcio, transferencia mediada

55 60 65

por DEAE-dextrano y transferencia mediada por vectores virales.

**[0079]** Por el contrario, una molécula "endógena" es una que está normalmente presente en una célula particular en una etapa de desarrollo particular bajo condiciones ambientales particulares. Por ejemplo, un ácido nucleico endógeno puede comprender un cromosoma, el genoma de una mitocondria, cloroplasto u otro orgánulo, o un ácido nucleico episomal natural. Las moléculas endógenas adicionales pueden incluir proteínas, p. ej., factores de transcripción y enzimas.

**[0080]** Una molécula de "fusión" es una molécula en donde dos o más moléculas de subunidades están vinculadas, preferentemente de forma covalente. Las moléculas de subunidad pueden ser del mismo tipo químico de molécula, o pueden ser diferentes tipos químicos de moléculas. Los ejemplos del primer tipo de molécula de fusión incluyen, pero no se limitan a, proteínas de fusión (p. ej., una fusión entre un dominio de unión a ADN de ZFP y un dominio de escisión) y ácidos nucleicos de fusión (p. ej., un ácido nucleico que codifica la proteína de fusión descrita anteriormente). Los ejemplos del segundo tipo de molécula de fusión incluyen, entre otros, una fusión entre un ácido nucleico de formación triple y un polipéptido, y una fusión entre un aglutinante menor y un ácido nucleico.

**[0081]** La expresión de una proteína de fusión en una célula puede resultar de la administración de la proteína de fusión a la célula o mediante la entrega de un polinucleótido que codifica la proteína de fusión a una célula, en donde el polinucleótido se transcribe, y la transcripción está traducida, para generar la proteína de fusión. El corte y empalme trans, la escisión de polipéptidos y la ligadura de polipéptidos también pueden estar implicados en la expresión de una proteína en una célula. Los métodos para el suministro de polinucleótidos y polipéptidos a las células se presentan en otra parte de esta descripción.

**[0082]** Un "gen", para los propósitos de la presente descripción, incluye una región de ADN que codifica un producto génico (véase *infra*), así como todas las regiones de ADN que regulan la producción del producto del gen, si o no tales reguladoras secuencias son adyacentes a la codificación y/o secuencias transcritas. Por consiguiente, un gen incluye, pero no se limita necesariamente a, secuencias promotoras, terminadores, secuencias reguladoras de la traducción tales como sitios de unión a ribosomas y sitios de entrada de ribosomas internos, potenciadores, silenciadores, aisladores, elementos de límite, orígenes de replicación, sitios de unión a matriz y regiones de control de locus.

**[0083]** "Expresión génica" se refiere a la conversión de la información, contenida en un gen, en un producto de gen. Un producto génico puede ser el producto transcripcional directo de un gen (p. ej., ARNm, ARNt, ARNr, ARN antisentido, ribozima, ARN estructural o cualquier otro tipo de ARN) o una proteína producida por la traducción de un ARNm. Los productos genéticos también incluyen ARN que se modifican, mediante procesos como la limitación, la poliadenilación, la metilación y la edición, y proteínas modificadas por, por ejemplo, metilación, acetilación, fosforilación, ubiquitinación, ribosilación de ADP, miristilación y glicosilación.

**[0084]** "Modulación" de la expresión de gen se refiere a un cambio en la actividad de un gen. La modulación de la expresión puede incluir, pero no se limita a, activación y represión de genes.

**[0085]** Las células "eucariotas" incluyen, pero no se limitan a, células de hongos (tales como levadura), células vegetales, células animales, de mamífero y células humanas.

**[0086]** Una "región de interés" es cualquier región de la cromatina celular, tal como, p. ej., un gen o una secuencia no codificante dentro o adyacentes a un gen, en donde es deseable unir una molécula exógena. La unión puede ser para fines de escisión de ADN dirigida y/o recombinación dirigida. Una región de interés puede estar presente en un cromosoma, un episoma, un genoma organular (p. ej., mitocondrial, cloroplasto) o un genoma viral infectante, por ejemplo. Una región de interés puede estar dentro de la región de codificación de un gen, dentro de regiones no codificantes transcritas tales como, p. ej., secuencias líder, secuencias de arrastre o intrones, o dentro de regiones no transcritas, ya sea aguas arriba o aguas abajo de la región de codificación. Una región de interés puede ser tan pequeña como un solo par de nucleótidos o hasta 2.000 pares de nucleótidos de longitud, o cualquier valor integral de pares de nucleótidos.

**[0087]** Los términos "unión operativa" y "unido operativamente" (o "unido operativamente") se usan indistintamente con referencia a una yuxtaposición de dos o más componentes (tales como elementos de secuencia), en donde los componentes están dispuestos de tal manera que tanto los componentes funcionan normalmente y permiten la posibilidad de que al menos uno de los componentes pueda mediar una función que se ejerce sobre al menos uno de los otros componentes. A modo de ilustración, una secuencia reguladora transcripcional, como un promotor, está operativamente unida a una secuencia codificante si la secuencia reguladora transcripcional controla el nivel de transcripción de la secuencia codificante en respuesta a la presencia o ausencia de uno o más factores reguladores transcripcionales. Una secuencia reguladora transcripcional generalmente está operativamente unida en *cis* con una secuencia de codificación, pero no necesita estar directamente adyacente a ella. Por ejemplo, un potenciador es una secuencia reguladora de la transcripción que está operativamente unida a una secuencia de codificación, aunque no son contiguas.

**[0088]** Con respecto a los polipéptidos de fusión, el término "unido operativamente" puede referirse al hecho de que

5 cada uno de los componentes realiza la misma función en vinculación con el otro componente como lo haría si no fuera así vinculado. Por ejemplo, con respecto a un polipéptido de fusión en donde un dominio de unión a ADN de ZFP se fusiona a un dominio de escisión, el dominio de unión a ADN de ZFP y el dominio de escisión están en enlace operativo si, en el polipéptido de fusión, la porción de dominio de unión a ADN de ZFP puede unir su sitio diana y/o su sitio de unión, mientras que el dominio de escisión puede escindir el ADN en las proximidades del sitio diana.

10 [0089] Un "fragmento funcional" de una proteína, polipéptido o ácido nucleico es una proteína, polipéptido o ácido nucleico cuya secuencia no es idéntica a la proteína de longitud completa, polipéptido o ácido nucleico, sin embargo, conserva la misma función que la proteína de longitud completa, polipéptido o ácido nucleico. Un fragmento funcional puede poseer más, menos o el mismo número de residuos que la molécula nativa correspondiente, y/o puede contener una o más sustituciones de aminoácidos o nucleótidos. Los métodos para determinar la función de un ácido nucleico (p. ej., la función de codificación, la capacidad de hibridarse con otro ácido nucleico) son bien conocidos en la técnica. Del mismo modo, los métodos para determinar la función de la proteína son bien conocidos. Por ejemplo, la función de unión al ADN de un polipéptido se puede determinar, p. ej., mediante la unión del filtro, el cambio de movilidad electroforética o los ensayos de inmunoprecipitación. La escisión del ADN se puede analizar mediante electroforesis en gel. Ver Ausubel *et al.*, *Supra*. La capacidad de una proteína para interactuar con otra proteína puede determinarse, p. ej., mediante inmunoprecipitación conjunta, ensayos de dos híbridos o complementación, tanto genética como bioquímica. Ver, p. ej., Fields *et al.* (1989) *Nature* 340: 245-246; Patente de Estados Unidos N° 5,585,245 y PCT WO 98/44350.

### 20 Sitios diana

25 [0090] Las composiciones y métodos descritos incluyen proteínas de fusión que comprenden un dominio de escisión (o un medio dominio de escisión) y un dominio de dedos de zinc, en donde el dominio del dedo de zinc, al unirse a una secuencia en la cromatina celular (p. ej., un sitio diana o un sitio de unión), dirige la actividad del dominio de escisión (o medio dominio de escisión) a la vecindad de la secuencia y, por lo tanto, induce escote en la vecindad de la secuencia diana. Como se establece en otra parte de esta descripción, un dominio de dedo de zinc se puede diseñar para unirse a prácticamente cualquier secuencia deseada. Por consiguiente, después de identificar una región de interés que contiene una secuencia en donde se desea la escisión o la recombinación, se pueden diseñar uno o más dominios de unión de dedos de zinc para que se unan a una o más secuencias en la región de interés. La expresión de una proteína de fusión que comprende un dominio de unión de dedos de zinc y un dominio de escisión (o de dos proteínas de fusión, cada una de las cuales comprende un dominio de unión de dedos de zinc y un medio dominio de escisión), en una célula, efectúa la escisión en la región de interés.

35 [0091] La selección de una secuencia en la cromatina celular para la unión por un dominio de dedos de zinc (p. ej., un sitio diana) se puede conseguir, p. ej., de acuerdo con los métodos descritos en Patente de Estados Unidos N° 6,453,242 en co-propiedad (17 de septiembre, 2002), que también describe métodos para diseñar ZFP para unirse a una secuencia seleccionada. Los expertos en la materia tendrán claro que la inspección visual simple de una secuencia de nucleótidos también se puede utilizar para la selección de un sitio diana. Por consiguiente, cualquier medio para la selección del sitio diana puede usarse en los métodos reivindicados.

45 [0092] Sitios diana se componen generalmente de una pluralidad de subsitios diana adyacentes. Un subsitio diana se refiere a la secuencia (generalmente un triplete de nucleótidos o un cuadruplete de nucleótidos que puede solaparse por un nucleótido con un cuadruplete adyacente) unida por un dedo de zinc individual. Véase, p. ej., el documento WO 02/077227. Si la cadena con la que una proteína de dedo de zinc hace la mayoría de los contactos se designa la cadena diana "cadena de reconocimiento primario" o "cadena de contacto primaria", algunas proteínas de dedo de zinc se unen a un triplete de tres bases en la cadena diana y una cuarta base en el filamento no diana. Un sitio diana generalmente tiene una longitud de al menos 9 nucleótidos y, en consecuencia, está unido por un dominio de unión de dedos de zinc que comprende al menos tres dedos de zinc. Sin embargo, la unión de, p. ej., un dominio de unión de 4 dedos a un sitio diana de 12 nucleótidos, un dominio de unión de 5 dedos a un sitio diana de 15 nucleótidos o un dominio de unión de 6 dedos a un sitio diana de 18 nucleótidos, también es posible. Como será evidente, también es posible la unión de dominios de unión más grandes (p. ej., 7, 8, 9 dedos y más) a sitios diana más largos.

55 [0093] No es necesario que un sitio diana sea un múltiplo de tres nucleótidos. Por ejemplo, en los casos en que se producen interacciones entre cadenas cruzadas (véase, p. ej., la patente de los Estados Unidos 6,453,242 y el documento WO 02/077227), uno o más de los dedos de zinc individuales de un dominio de unión de múltiples dedos pueden unirse a subsitios cuádruples superpuestos. Como resultado, una proteína de tres dedos puede unirse a una secuencia de 10 nucleótidos, en donde el décimo nucleótido es parte de un cuadruplete unido por un dedo terminal, una proteína de cuatro dedos puede unirse a una secuencia de 13 nucleótidos, en donde el decimotercer nucleótido es parte de un cuádruplo obligado por un terminal de dedo, etc.

65 [0094] La longitud y la naturaleza de secuencias de enlazadores de aminoácidos entre los dedos de zinc individuales en un dominio de unión multi-dedo también afecta a la unión a una secuencia diana. Por ejemplo, la presencia de un denominado "enlazador no canónico", "enlazador largo" o "enlazador estructurado" entre dedos de zinc adyacentes en un dominio de unión de múltiples dedos puede permitir que esos dedos se unan a subsitios que no son inmediatamente adyacentes. Ejemplos no limitativos de tales enlazadores se describen, p. ej., en la Patente de EE.UU.

Nº 6,479,626 y WO 01/53480. En consecuencia, uno o más subsitios, en un sitio diana para un dominio de unión de dedos de zinc, se pueden separar entre sí por 1, 2, 3, 4, 5 o más nucleótidos. Para proporcionar solo un ejemplo, un dominio de unión de cuatro dedos puede unirse a un sitio diana de 13 nucleótidos que comprende, en secuencia, dos subsitios contiguos de 3 nucleótidos, un nucleótido intermedio y dos subsitios de triplete contiguos.

**[0095]** Distancia entre secuencias (*p. ej.*, sitios diana) se refiere al número de nucleótidos o pares de nucleótidos que intervienen entre dos secuencias, como se mide a partir de los bordes de las secuencias más cercanas entre sí.

**[0096]** En ciertas realizaciones en las que la escisión depende de la unión de moléculas de fusión de dominio de dos dedos de zinc/medio dominio de escisión a sitios diana separados, los dos sitios diana pueden estar en cadenas de ADN opuestas. En otras realizaciones, ambos sitios diana están en la misma cadena de ADN.

### Dominios de unión de dedo de zinc

**[0097]** Un dominio de unión a dedo de zinc comprende uno o más dedos de zinc. Miller y col. (1985) EMBO J. 4:1609-1614; Rhodes (1993) Scientific American, febrero: 56-65; Patente de Estados Unidos Nº 6,453,242. Típicamente, un solo dominio de dedo de zinc tiene aproximadamente 30 aminoácidos de longitud. Los estudios estructurales han demostrado que cada dominio de dedo de zinc (motivo) contiene dos láminas beta (sostenidas en un giro beta que contiene los dos residuos de cisteína invariantes) y una hélice alfa (que contiene los dos residuos de histidina invariantes), que se mantienen en una conformación particular a través de la coordinación de un átomo de zinc por las dos cisteínas y las dos histidinas.

**[0098]** Dedos de zinc incluyen tanto dedos de zinc canónicos  $C_2H_2$  (es decir, aquellos en los que el ion de zinc está coordinado por dos residuos de cisteína y dos de histidina) como dedos de zinc no canónicos tales como, *p. ej.*, dedos de zinc  $C_3H$  (aquellos en los que el ion zinc está coordinado por tres residuos de cisteína y un residuo de histidina) y dedos de zinc  $C_4$  (aquellos en los que el ion zinc está coordinado por cuatro residuos de cisteína). Véase también el documento WO 02/057293.

**[0099]** Dominios de unión de dedo de zinc se pueden diseñar para unirse a una secuencia de elección. *Ver, p. ej.*, Beerli et al. (2002) Nature Biotechnol. 20: 135-141; Pabo y col. (2001) Ann. Rev. Biochem. 70: 313-340; Isalan y col. (2001) Nature Biotechnol. 19: 656-660; Segal y col. (2001) Curr. Opin. Biotechnol. 12: 632-637; Choo y col. (2000) Curr. Opin. Struct. Biol. 10: 411-416. Un dominio de unión de dedos de zinc diseñado puede tener una especificidad de unión novedosa, en comparación con una proteína de dedos de zinc que se produce naturalmente. Los métodos de ingeniería incluyen, pero no se limitan a, diseño racional y varios tipos de selección. El diseño racional incluye, *p. ej.*, el uso de bases de datos que comprenden secuencias de nucleótidos de triplete (o cuádruple) y secuencias de aminoácidos de dedo de zinc individuales, en las que cada secuencia de nucleótido de triplete o cuádruple se asocia con una o más secuencias de aminoácidos de dedos de zinc que se unen al triplete particular o secuencia cuádruple. Véanse, *p. ej.*, las patentes de los Estados Unidos de propiedad conjunta 6,453,242 y 6,534,261.

**[0100]** Métodos de selección ejemplares, que incluyen visualización de fagos y sistemas de dos híbridos, se describen en las patentes de los Estados Unidos 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; y 6,242,568; así como WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 y GB 2,338,237.

**[0101]** La mejora de la especificidad de unión para dominios de unión de dedos de zinc se ha descrito, *p. ej.*, en el documento WO 02/077227.

**[0102]** Dado que un dedo de zinc individual se une a una secuencia de tres nucleótidos (es decir, triplete) (o una secuencia de cuatro nucleótidos que puede solaparse, por un nucleótido, con el sitio de unión de cuatro nucleótidos de un dedo de zinc adyacente), la longitud de una secuencia a la que un dominio de unión de dedos de zinc está diseñado para unirse (*p. ej.*, una secuencia diana) determinará el número de dedos de zinc en un dominio de unión de dedos de zinc diseñado. Por ejemplo, para las ZFP en las que los motivos de los dedos no se unen a subsitios superpuestos, una secuencia diana de seis nucleótidos está unida por un dominio de unión de dos dedos; una secuencia diana de nueve nucleótidos está unida por un dominio de unión de tres dedos, *etc.* Como se indica en el presente documento, los sitios de unión para dedos de zinc individuales (es decir, subsitios) en un sitio diana no necesitan ser contiguos, sino que pueden estar separados por uno o varios nucleótidos, dependiendo de la longitud y la naturaleza de las secuencias de aminoácidos entre los dedos de zinc (es decir, los enlazadores entre dedos) en un dominio de unión de múltiples dedos.

**[0103]** En un dominio de unión de dedos de zinc de múltiples dedos, los dedos de zinc adyacentes se pueden separar por secuencias de enlazador de aminoácidos de aproximadamente 5 aminoácidos (los llamados enlazadores entre dedos "canónicos") o, alternativamente, por uno o más no canónicos enlazadores. Véanse, *p. ej.*, las patentes de los Estados Unidos de propiedad conjunta números 6,453,242 y 6,534,261. Para los dominios de unión de dedos de zinc diseñados que comprenden más de tres dedos, se puede preferir la inserción de enlazadores entre dedos más largos ("no canónicos") entre ciertos dedos de zinc, ya que puede aumentar la afinidad y/o especificidad de unión por la unión dominio. Véanse, *p. ej.*, la Patente de Estados Unidos Nº 6,479,626 y el documento WO 01/53480. Por consiguiente, los dominios de unión de dedos de zinc con múltiples dedos también se pueden caracterizar con respecto a la

presencia y ubicación de enlazadores entre dedos no canónicos. Por ejemplo, un dominio de unión de dedos de zinc de seis dedos que comprende tres dedos (unidos por dos enlazadores entre dedos canónicos), un enlazador largo y tres dedos adicionales (unidos por dos enlazadores entre dedos canónicos) se denomina configuración 2x3. De forma similar, un dominio de unión que comprende dos dedos (con un enlazador canónico entre ellos), un enlazador largo y dos dedos adicionales (unidos por un enlazador canónico) se denomina proteína 2x2. Una proteína que comprende tres unidades de dos dedos (en cada una de las cuales los dos dedos están unidos por un conector canónico), y en donde cada unidad de dos dedos está unida a la unidad adyacente de dos dedos por un conector largo, se conoce como Proteína 3x2.

**[0104]** La presencia de un enlace entre dedos largo o no canónico entre dos dedos de zinc adyacentes en un dominio de unión multi-dedo a menudo permite que los dos dedos se unan a los subsitios que no son inmediatamente contiguos en la secuencia diana. En consecuencia, puede haber huecos de uno o más nucleótidos entre subsitios en un sitio diana; es decir, un sitio diana puede contener uno o más nucleótidos que no son contactados por un dedo de zinc. Por ejemplo, un dominio de unión de dedos de zinc 2x2 puede unirse a dos secuencias de seis nucleótidos separadas por un nucleótido, es decir, se une a un sitio diana de 13 nucleótidos. Ver también Moore et al. (2001a) Proc. Natl. Acad. Sci. EE.UU. 98: 1432-1436; Moore y col. (2001b) Proc. Natl. Acad. Sci. EE.UU. 98: 1437-1441 y WO 01/53480.

**[0105]** Como se mencionó anteriormente, un subsitio diana es una secuencia de tres o cuatro nucleótidos que está obligado por un solo dedo de zinc. Para ciertos fines, una unidad de dos dedos se denomina módulo de enlace. Se puede obtener un módulo de unión, p. ej., seleccionando dos dedos adyacentes en el contexto de una proteína de múltiples dedos (generalmente tres dedos) que se unen a una secuencia diana particular de seis nucleótidos. Alternativamente, los módulos se pueden construir mediante el ensamblaje de dedos de zinc individuales. Véanse también los documentos WO 98/53057 y WO 01/53480.

## 25 Dominios de escisión

**[0106]** La porción del dominio de escisión de las proteínas de fusión descritas en el presente documento se puede obtener de cualquier endo- o exonucleasa. Los ejemplos de endonucleasas a partir de las cuales se puede derivar un dominio de escisión incluyen, pero no se limitan a, endonucleasas de restricción y endonucleasas de referencia. Ver, p. ej., Catálogo 2002-2003, New England Biolabs, Beverly, MA; y Belfort et al. (1997) Nucleic Acids Res. 25: 3379-3388. Se conocen enzimas adicionales que escinden el ADN (p. ej., Nucleasa S1; nucleasa de frijol mungo; DNasa pancreática I; nucleasa micrococcal; endonucleasa HO de levadura; véase también Linn et al. (Eds.) Nucleasas, Cold Spring Harbor Laboratory Press, 1993). Una o más de estas enzimas (o fragmentos funcionales de las mismas) pueden usarse como fuente de dominios de escisión y medios dominios de escisión.

**[0107]** De manera similar, un medio dominio de escisión (p. ej., proteínas de fusión que comprenden un dominio de unión de dedos de zinc y un medio dominio de escisión) puede derivarse de cualquier nucleasa o una porción de la misma, como se establece anteriormente, que requiere dimerización para la actividad de escisión. En general, se requieren dos proteínas de fusión para la escisión si las proteínas de fusión comprenden medios dominios de escisión. Los dos medios dominios de escisión pueden derivarse de la misma endonucleasa (o fragmentos funcionales de la misma), o cada medio dominio de escisión puede derivarse de una endonucleasa diferente (o fragmentos funcionales de la misma). Además, los sitios diana para las dos proteínas de fusión están dispuestos preferiblemente entre sí, de modo que la unión de las dos proteínas de fusión coloca los medios dominios de escisión en una orientación espacial entre sí que permite que los medios dominios de escisión formen un dominio de escisión funcional, p. ej., por dimerización. Por lo tanto, en ciertas realizaciones, los bordes cercanos de los sitios diana están separados por 5-8 nucleótidos o por 15-18 nucleótidos. Sin embargo, cualquier número integral de nucleótidos o pares de nucleótidos puede intervenir entre dos sitios diana (p. ej., de 2 a 50 nucleótidos o más). En general, el punto de escisión se encuentra entre los sitios diana.

**[0108]** En general, si se usan dos proteínas de fusión, cada una de las cuales comprende un medio dominio de escisión, la cadena de contacto primaria para la porción de dedo de zinc de cada proteína de fusión estará en una cadena de ADN diferente y en orientación opuesta. Es decir, para un par de fusiones de medio dominio de escisión/ZFP, las secuencias diana están en cadenas opuestas y las dos proteínas se unen en orientaciones opuestas.

**[0109]** Las endonucleasas de restricción (enzimas de restricción) están presentes en muchas especies y son capaces de unirse específicamente al ADN a la secuencia (en un sitio de reconocimiento) y escindir el ADN en o cerca del sitio de unión. Ciertas enzimas de restricción (p. ej., Tipo IIS) escinden el ADN en sitios retirados del sitio de reconocimiento y tienen dominios de unión y escisión separables. Por ejemplo, la enzima Tipo IIS *Fok I* cataliza la escisión de ADN de doble cadena, a 9 nucleótidos desde su sitio de reconocimiento en una cadena y 13 nucleótidos desde su sitio de reconocimiento en el otro. Véanse, p. ej., las patentes de los Estados Unidos 5,356,802; 5,436,150 y 5,487,994; así como Li et al. (1992) Proc. Natl. Acad. Sci. EE.UU. 89: 4275-4279; Li y col. (1993) Proc. Natl. Acad. Sci. EE.UU. 90: 2764-2768; Kim y col. (1994a) Proc. Natl. Acad. Sci. EE.UU. 91: 883-887; Kim y col. (1994b) J. Biol. Chem 269: 31,978-31,982. Por lo tanto, las proteínas de fusión pueden comprender el dominio de escisión (o medio dominio de escisión) de al menos una enzima de restricción de Tipo IIS y uno o más dominios de unión de dedos de zinc, que pueden o no estar diseñados.

[0110] Una enzima de restricción de Tipo IIS ejemplar, cuyo dominio de escisión es separable del dominio de unión, es *Fok I*. Esta enzima particular es activa como un dímero. Bitinaite y col. (1998) Proc. Natl. Acad. Sci. EE.UU. 95: 10.570-10.575. Por consiguiente, para los fines de la presente descripción, la porción de la enzima *Fok I* usada en las proteínas de fusión descritas se considera un medio dominio de escisión. Por lo tanto, para la escisión bicatenaria dirigida y/o la sustitución dirigida de secuencias celulares usando fusiones de dedo de zinc *Fok I*, se pueden usar dos proteínas de fusión, cada una de las cuales comprende un medio dominio de escisión *Fok I*, para reconstituir un dominio de escisión catalíticamente activo. Alternativamente, también se puede usar una sola molécula de polipéptido que contiene un dominio de unión de dedos de zinc y dos medios dominios de escisión *Fok I*. Los parámetros para la escisión dirigida y la alteración de la secuencia dirigida usando fusiones de dedo de zinc *Fok I* se proporcionan en otra parte de esta divulgación.

[0111] Enzimas de restricción de Tipo IIS ejemplares se enumeran en la Tabla 1. Enzimas de restricción adicionales también contienen dominios de unión y escisión separables, y estos se contemplan por la presente divulgación. Ver, p. ej., Roberts et al. (2003) Nucleic Acids Res. 31: 418-420.

**Tabla 1: Algunas enzimas de restricción de tipo IIS**

Aar I	BsrB I	SspD5 I
Ace III	BsrD I	Sth132 I
Aci I	BstF5 I	Sts I
Alo I	Btr I	TspDT I
Bae I	Bts I	TspGW I
Bbr7 I	Cdi I	Tth111 II
Bbv I	CjeP I	UbaP I
Bbv II	Drd II	Bsa I
BbvC I	Eci I	BsmB I
Bcc I	Eco31 I	
Bce83 I	Eco57 I	
BceA I	Eco57M I	
Bcef I	Esp3 I	
Bcg I	Fau I	
BciV I	Fin I	
Bfi I	Fok I	
Bin I	Gdi II	
Bmg I	Gsu I	
Bpu10 I	Hga I	
BsaX I	Hin4 II	
Bsb I	Hph I	
BscA I	Ksp632 I	
BscG I	Mbo II	
BseR I	Mly I	
BseY I	Mme I	
Bsi I	Mnl I	
Bsm I	Pfl1108 I	
BsmA I	Ple I	
BsmF I	Ppi I	
Bsp24 I	Psr I	
BspG I	RleA I	
BspM I	Sap I	
BspNC I	SfaN I	
Bsr I	Sim I	

**Fusiones del dominio de escisión del dominio del dedo de zinc**

[0112] Los expertos en la técnica conocen métodos para el diseño y la construcción de proteínas de fusión (y polinucleótidos que las codifican). Por ejemplo, los métodos para el diseño y la construcción de proteínas de fusión que comprenden proteínas de dedos de zinc (y polinucleótidos que los codifican) se describen en las patentes de los Estados Unidos 6,453,242 y 6,534,261 de propiedad conjunta. Se pueden construir polinucleótidos que codifican tales proteínas de fusión. Estos polinucleótidos pueden insertarse en un vector y el vector puede introducirse en una célula (véase más adelante para una descripción adicional con respecto a los vectores y métodos para introducir polinucleótidos en las células).

[0113] En ciertas realizaciones de los métodos descritos en este documento, una proteína de fusión comprende un dominio de unión de dedos de zinc y un medio dominio de escisión de la enzima de restricción *Fok I*, y dos de tales proteínas de fusión se expresan en una célula. La expresión de dos proteínas de fusión en una célula puede resultar

del suministro de las dos proteínas a la célula; suministro de una proteína y un ácido nucleico que codifica una de las proteínas a la célula; suministro de dos ácidos nucleicos, cada uno de los cuales codifica una de las proteínas, a la célula; o mediante el suministro de un único ácido nucleico, que codifica ambas proteínas, a la célula. Una proteína de fusión puede comprender una única cadena de polipéptidos que comprende dos medios dominios de escisión y un dominio de unión de dedos de zinc. En este caso, una proteína de fusión única se expresa en una célula y, sin desear estar unida por la teoría, se cree que escinde el ADN como resultado de la formación de un dímero intramolecular de los medios dominios de escisión.

[0114] En general, los componentes de las proteínas de fusión (*p. ej.*, fusiones ZFP-Fok I) están dispuestos de manera que el dominio del dedo de zinc está más cerca del extremo amino de la proteína de fusión, y el medio dominio de escisión está más cerca del término carboxi. Esto refleja la orientación relativa del dominio de escisión en los dominios de escisión dimerizantes de origen natural, como los derivados de la enzima Fok I, en donde el dominio de unión al ADN está más cerca del extremo amino y el medio dominio de escisión está más cerca del extremo carboxilo.

[0115] En las proteínas de fusión descritas, la secuencia de aminoácidos entre el dominio de unión de dedos de zinc (que está delimitado por la mayoría de los dos residuos de cisteína conservados en el extremo N y la mayoría de los dos residuos de histidina conservados en el extremo extremo C) y el dominio de escisión (o medio dominio) se denomina "enlazador ZC". El enlazador ZC debe distinguirse de los enlazadores entre dedos discutidos anteriormente. Por ejemplo, en una proteína de fusión ZFP-Fok I (en donde los componentes están dispuestos: dominio de unión de extremo N-dedos de zinc-Fok I medio dominio de escisión-extremo C), el enlazador ZC se encuentra entre el segundo residuo de histidina del dedo más zinc extremo C y el residuo de aminoácido más extremo N del medio dominio de escisión (que generalmente es glutamina (Q) en la secuencia QLV). El enlazador ZC puede ser cualquier secuencia de aminoácidos. Para obtener una escisión óptima, la longitud del enlazador y la **distancia** entre los sitios diana (sitios de unión) están relacionados entre sí. *Ver, p. ej.*, Smith et al. (2000) Nucleic Acids Res. 28:3361-3369; Bibikova y col. (2001) Mol. Cell. Biol. 21: 289-297, señalando que su notación para la longitud del enlazador difiere de la dada aquí. Por ejemplo, para las fusiones ZFP-Fok I que tienen una longitud de enlazador ZC de cuatro aminoácidos (como se define aquí), la escisión óptima ocurre cuando los sitios de unión para las proteínas de fusión están ubicados a 6 o 16 nucleótidos separados (medidos desde el borde cercano de cada sitio de unión).

### 30 Métodos para escisión dirigida

[0116] Los métodos y composiciones descritas pueden usarse para escindir ADN en una región de interés en cromatina celular (*p. ej.*, en un sitio deseado o predeterminado en un genoma, *p. ej.*, en un gen, mutante o tipo silvestre). Para tal escisión de ADN diana, un dominio de unión de dedos de zinc está diseñado para unirse a un sitio diana en o cerca del sitio de escisión predeterminado, y una proteína de fusión que comprende el dominio de unión de dedo de zinc diseñado y un dominio de escisión se expresa en una célula. Tras la unión de la porción de dedo de zinc de la proteína de fusión al sitio diana, el ADN se escinde cerca del sitio diana mediante el dominio de escisión. El sitio exacto de escisión puede depender de la longitud del enlazador ZC.

[0117] Alternativamente, dos proteínas de fusión, cada una de las cuales comprende un dominio de unión de dedos de zinc y un medio dominio de escisión, se expresan en una célula y se unen a sitios diana que se yuxtaponen de tal manera que se reconstituya un dominio de escisión funcional y ADN se escinde en las proximidades de los sitios diana. En una realización, la escisión se produce entre los sitios diana de los dos dominios de unión de dedos de zinc. Se pueden diseñar uno o ambos dominios de unión de dedos de zinc.

[0118] Para la escisión dirigida usando un polipéptido de fusión de dominio de escisión de dominio de unión de dedo de zinc, el sitio de unión puede abarcar el sitio de escisión, o el borde cercano del sitio de unión puede ser 1, 2, 3, 4, 5, 6, 10, 25, 50 o más nucleótidos (o cualquier valor integral entre 1 y 50 nucleótidos) del sitio de escisión. La ubicación exacta del sitio de unión, con respecto al sitio de escisión, dependerá del dominio de escisión particular, y la longitud del enlazador ZC. Para los métodos en los que se usan dos polipéptidos de fusión, cada uno de los cuales comprende un dominio de unión de dedos de zinc y un medio dominio de escisión, los sitios de unión generalmente se extienden a horcajadas sobre el sitio de escisión. Así, el borde cercano del primer sitio de unión puede ser 1, 2, 3, 4, 5, 6, 10, 25 o más nucleótidos (o cualquier valor integral entre 1 y 50 nucleótidos) en un lado del sitio de escisión, y el el borde cercano del segundo sitio de unión puede ser 1, 2, 3, 4, 5, 6, 10, 25 o más nucleótidos (o cualquier valor integral entre 1 y 50 nucleótidos) en el otro lado del sitio de escisión. Los expertos en la materia conocen métodos para mapear sitios de escisión *in vitro* e *in vivo*.

[0119] Por lo tanto, los métodos descritos en el presente documento pueden emplear un dominio de unión de dedos de zinc diseñado por ingeniería fusionada a un dominio de escisión. En estos casos, el dominio de unión está diseñado para unirse a una secuencia diana, en o cerca de la cual se desea la escisión. La proteína de fusión, o un polinucleótido que la codifica, se introduce en una célula. Una vez introducida o expresada en la célula, la proteína de fusión se une a la secuencia diana y se escinde en o cerca de la secuencia diana. El sitio exacto de escisión depende de la naturaleza del dominio de escisión y/o la presencia y/o naturaleza de las secuencias de enlace entre los dominios de unión y escisión. En los casos en que se usan dos proteínas de fusión, cada una de las cuales comprende un medio dominio de escisión, la distancia entre los bordes cercanos de los sitios de unión puede ser 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 25 o más nucleótidos (o cualquier valor integral entre 1 y 50 nucleótidos). Los niveles óptimos de escisión también pueden

depender tanto de la distancia entre los sitios de unión de las dos proteínas de fusión (véase, p. ej., Smith et al. (2000) Nucleic Acids Res. 28: 3361-3369; Bibikova et al. (2001) Mol. Cell. Biol. 21: 289-297) y la longitud del enlazador ZC en cada proteína de fusión.

5 **[0120]** Para las nucleasas de fusión ZFP-*Fok* I, la longitud del enlazador entre el ZFP y el medio dominio de escisión de *Fok* I (es decir, el ligador ZC) puede influir en la eficacia de escisión. En un sistema experimental que utiliza una fusión ZFP-*Fok* I con un enlazador ZC de 4 residuos de aminoácidos, se obtuvo una escisión óptima cuando los bordes cercanos de los sitios de unión para dos nucleasas ZFP-*Fok* I se separaron por 6 pares de bases. Esta nucleasa de fusión particular comprendía la siguiente secuencia de aminoácidos entre la porción de dedo de zinc y el semidominio de nucleasa:  
10 HQRTHQNKKQLV (SEQ ID NO:26) en donde las dos histidinas conservadas en la porción de extremo C del dedo de zinc y los tres primeros residuos en el medio dominio de escisión de *Fok* I están subrayados. Por consiguiente, la secuencia del enlazador en esta construcción es QNKK. Bibikova y col. (2001) Mol. Cell. Biol. 21: 289-297. Los presentes inventores han construido una serie de nucleasas de fusión ZFP-*Fok* I que tienen una variedad de longitudes y secuencias de enlazador ZC, y analizaron las eficiencias de escisión de estas nucleasas en una serie de sustratos que tienen diferentes distancias entre los sitios de unión de ZFP. Ver ejemplo 4.

15 **[0121]** En ciertas realizaciones, el dominio de escisión comprende dos medios dominios de escisión, los cuales son parte de un único polipéptido que comprende un dominio de unión, un primer medio dominio de escisión y un segundo medio dominio de escisión. Los medios dominios de escisión pueden tener la misma secuencia de aminoácidos o diferentes secuencias de aminoácidos, siempre que funcionen para escindir el ADN.

20 **[0122]** Los medios dominios de escisión también pueden proporcionarse en moléculas separadas. Por ejemplo, dos polipéptidos de fusión pueden introducirse en una célula, en donde cada polipéptido comprende un dominio de unión y un medio dominio de escisión. Los medios dominios de escisión pueden tener la misma secuencia de aminoácidos o diferentes secuencias de aminoácidos, siempre que funcionen para escindir el ADN. Además, los dominios de unión se unen a secuencias diana que típicamente están dispuestas de tal manera que, tras la unión de los polipéptidos de fusión, los dos medios dominios de escisión se presentan en una orientación espacial entre sí que permite la reconstitución de un dominio de escisión (p. ej., por dimerización de los semidominios), colocando así los semidominios entre sí para formar un dominio de escisión funcional, lo que da como resultado la escisión de la cromatina celular en una región de interés. Generalmente, la escisión por el dominio de escisión reconstituido ocurre en un sitio ubicado entre las dos secuencias diana. Una o ambas proteínas se pueden diseñar para unirse a su sitio diana.

25 **[0123]** Las dos proteínas de fusión pueden unirse en la región de interés en la misma u opuesta polaridad, y sus sitios de unión (es decir, sitios diana) pueden estar separados por cualquier número de nucleótidos, p. ej., de 0 a 200 nucleótidos o cualquier valor integral intermedio. En ciertas realizaciones, los sitios de unión para dos proteínas de fusión, que comprenden cada uno un dominio de unión de dedos de zinc y un medio dominio de escisión, se pueden ubicar entre 5 y 18 nucleótidos separados, p. ej., 5-8 nucleótidos separados, o 15-18 nucleótidos separados, o 6 nucleótidos separados, o 16 nucleótidos separados, medidos desde el borde de cada sitio de unión más cercano al otro sitio de unión, y la escisión ocurre entre los sitios de unión.

30 **[0124]** El sitio en donde se divide el ADN generalmente se encuentra entre los sitios de unión para las dos proteínas de fusión. La ruptura de doble cadena del ADN a menudo resulta de dos roturas de una sola cadena, o "muescas", compensadas por 1, 2, 3, 4, 5, 6 o más nucleótidos (p. ej., la escisión del ADN de doble cadena por *Fok* I nativo resulta de roturas monocatenarias compensadas por 4 nucleótidos). Por lo tanto, la escisión no ocurre necesariamente en sitios exactamente opuestos en cada cadena de ADN. Además, la estructura de las proteínas de fusión y la distancia entre los sitios diana pueden influir en si la escisión ocurre adyacente a un solo par de nucleótidos, o si la escisión ocurre en varios sitios. Sin embargo, para muchas aplicaciones, incluso la división por recombinación dirigida (ver *infra*) dentro de un rango de nucleótidos es generalmente suficiente, y no es necesaria la división entre pares de bases particulares.

35 **[0125]** Como se señaló anteriormente, la(s) proteína(s) de fusión se puede(n) introducir como polipéptidos y/o polinucleótidos. Por ejemplo, dos polinucleótidos, cada uno de los cuales comprende secuencias que codifican uno de los polipéptidos mencionados anteriormente, pueden introducirse en una célula, y cuando los polipéptidos se expresan y cada uno se une a su secuencia diana, la escisión ocurre en o cerca de la secuencia diana. Alternativamente, un solo polinucleótido que comprende secuencias que codifican ambos polipéptidos de fusión se introduce en una célula. Los polinucleótidos pueden ser ADN, ARN o cualquier forma modificada o análogos o ADN y/o ARN.

40 **[0126]** Para mejorar la especificidad de escisión, también se pueden emplear composiciones adicionales en los métodos descritos en este documento. Por ejemplo, los semi-dominios de escisión simple pueden exhibir una actividad de escisión bicatenaria limitada. En los métodos en los que se introducen en la célula dos proteínas de fusión, cada una de las cuales contiene un dominio de tres dedos de zinc y un medio dominio de escisión, cada proteína específica un sitio diana de aproximadamente 9 nucleótidos. Aunque la secuencia diana agregada de 18 nucleótidos es probable que sea única en un genoma de mamífero, cualquier sitio diana de 9 nucleótidos se produce, en promedio, aproximadamente 23.000 veces en el genoma humano. Por lo tanto, puede producirse una escisión no específica, debido a la unión específica del sitio de un solo medio dominio. Por consiguiente, los métodos descritos en la presente

contemplan el uso de un mutante dominante negativo de un medio dominio de escisión como *Fok I* (o un ácido nucleico que lo codifica) que se expresa en una célula junto con las dos proteínas de fusión. El mutante dominante negativo es capaz de dimerizarse, pero no puede escindir, y también bloquea la actividad de escisión de un medio dominio en donde se dimeriza. Al proporcionar el mutante dominante negativo en exceso molar a las proteínas de fusión, solo las regiones en las que se unen ambas proteínas de fusión tendrán una concentración local lo suficientemente alta de medios dominios de escisión funcional para que se produzca la dimerización y escisión. En los sitios donde solo se une una de las dos proteínas de fusión, su medio dominio de escisión forma un dímero con el medio dominio mutante negativo dominante, y no se produce una escisión no específica indeseable.

[0127] Se han identificado tres residuos de aminoácidos catalíticos en el medio dominio de escisión *Fok I*: Asp 450, Asp 467 y Lys 469. Bitinaite et al. (1998) Proc. Natl. Acad. Sci. EE.UU. 95: 10,570-10,575. Por lo tanto, una o más mutaciones en uno de estos residuos pueden usarse para generar una mutación negativa dominante. Además, muchos de los residuos de aminoácidos catalíticos de otras endonucleasas de tipo IIS son conocidos y/o pueden determinarse, p. ej., mediante la alineación con secuencias de *Fok I* y/o mediante la generación y prueba de mutantes para actividad catalítica.

#### Mutaciones del dominio de dimerización en el medio dominio de escisión

[0128] Los métodos para la escisión dirigida que implican el uso de fusiones entre una ZFP y un medio dominio de escisión (como, p. ej., una fusión ZFP/*Fok I*) requieren el uso de dos moléculas de fusión, cada una generalmente dirigida a una secuencia diana distinta. Las secuencias diana para las dos proteínas de fusión se pueden elegir de modo que la escisión dirigida se dirija a un sitio único en un genoma, como se discutió anteriormente. Una fuente potencial de especificidad de escisión reducida podría resultar de la homodimerización de una de las dos fusiones de medio dominio de escisión/ZFP. Esto podría ocurrir, p. ej., debido a la presencia, en un genoma, de repeticiones invertidas de las secuencias diana para una de las dos fusiones de medio dominio de escisión/ZFP, ubicadas para permitir que dos copias de la misma proteína de fusión se unan con una orientación y espaciado que permite la formación de un dímero funcional.

[0129] Un enfoque para reducir la probabilidad de este tipo de escisión aberrante en secuencias distintas del sitio diana previsto implica generar variantes del medio dominio de escisión que minimizan o evitan la homodimerización. Preferiblemente, se alteran uno o más aminoácidos en la región del medio dominio implicado en su dimerización. En la estructura de cristal del dímero de proteína *Fok I*, la estructura de los medios dominios de escisión se informa, es similar a la disposición de los medios dominios de escisión durante la escisión de ADN por *Fok I*. Wah y col. (1998) Proc. Natl. Acad. Sci. EE.UU. 95: 10564-10569. Esta estructura indica que los residuos de aminoácidos en las posiciones 483 y 487 juegan un papel clave en la dimerización de los medios dominios de escisión de *Fok I*. La estructura también indica que los residuos de aminoácidos en las posiciones 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537 y 538 están lo suficientemente cerca del interfaz de dimerización para influir en la dimerización. En consecuencia, las alteraciones de la secuencia de aminoácidos en una o más de las posiciones mencionadas anteriormente probablemente alterarán las propiedades de dimerización del medio dominio de escisión. Tales cambios pueden introducirse, p. ej., construyendo una biblioteca que contenga (o codifique) diferentes residuos de aminoácidos en estas posiciones y seleccionando variantes con las propiedades deseadas, o diseñando racionalmente mutantes individuales. Además de prevenir la homodimerización, también es posible que algunas de estas mutaciones puedan aumentar la eficiencia de escisión por encima de la obtenida con dos medios dominios de escisión de tipo silvestre.

[0130] En consecuencia, la alteración de un medio dominio de escisión de *Fok I* en cualquier residuo de aminoácido que afecte la dimerización puede usarse para prevenir uno de un par de fusiones ZFP/*Fok I* de someterse a homodimerización que puede conducir a la escisión en secuencias no deseadas. Por lo tanto, para la escisión dirigida usando un par de fusiones ZFP/*Fok I*, una o ambas proteínas de fusión pueden comprender una o más alteraciones de aminoácidos que inhiben la auto-dimerización, pero permiten que ocurra la heterodimerización de las dos proteínas de fusión de manera que se produzca la escisión en el sitio diana deseado. En ciertas realizaciones, las alteraciones están presentes en ambas proteínas de fusión, y las alteraciones tienen efectos aditivos; es decir, la homodimerización de cualquiera de las fusiones, que conduce a la escisión aberrante, se minimiza o se suprime, mientras que la heterodimerización de las dos proteínas de fusión se facilita en comparación con la obtenida con los medios dominios de escisión de tipo silvestre. Véase el Ejemplo 5.

#### Métodos para la alteración dirigida de secuencias genómicas y recombinación dirigida

[0131] También se describen en el presente documento métodos para reemplazar una secuencia genómica (p. ej., una región de interés en cromatina celular) con una secuencia homóloga no idéntica (es decir, recombinación dirigida). Los intentos anteriores para reemplazar secuencias particulares han implicado el contacto de una célula con un polinucleótido que comprende secuencias que tienen homología con una región cromosómica (es decir, un ADN donante), seguido de la selección de células en las que la molécula de ADN donante había sufrido una recombinación homóloga en el genoma. La tasa de éxito de estos métodos es baja, debido a la baja eficiencia de la recombinación homóloga y a la alta frecuencia de inserción no específica del ADN del donante en regiones del genoma distintas del sitio diana.

**[0132]** La presente divulgación proporciona métodos de alteración de la secuencia dirigida caracterizados por una eficiencia mayor de la recombinación específica y una menor frecuencia de los eventos de inserción no específicos. Los métodos implican la fabricación y el uso de dominios de unión de dedos de zinc diseñados por ingeniería fusionados para escindir dominios (o semidominios de escisión) para hacer una o más rupturas de doble cadena dirigidas en el ADN celular. Debido a que las rupturas bicatenarias en el ADN celular estimulan la recombinación homóloga varias veces en las proximidades del sitio de escisión, dicha escisión dirigida permite la alteración o el reemplazo (mediante recombinación homóloga) de secuencias en prácticamente cualquier sitio del genoma.

**[0133]** Además de las moléculas de fusión descritas en el presente documento, el reemplazo dirigido de una secuencia genómica seleccionada también requiere la introducción de la secuencia de reemplazo (o donante). La secuencia del donante puede introducirse en la célula antes de, simultáneamente con, o después de la expresión de la proteína o proteínas de fusión. El polinucleótido donante contiene suficiente homología con una secuencia genómica para soportar la recombinación homóloga entre él y la secuencia genómica con la que tiene homología. Aproximadamente 25, 50, 100 o 200 nucleótidos o más de homología de secuencia entre un donante y una secuencia genómica (o cualquier valor integral entre 10 y 200 nucleótidos, o más) apoyarán la recombinación homóloga entre ellos. Las secuencias donantes pueden variar en longitud de 10 a 5000 nucleótidos (o cualquier valor integral de nucleótidos entre ellas) o más. Será evidente que la secuencia del donante generalmente no es idéntica a la secuencia genómica que reemplaza. Por ejemplo, la secuencia del polinucleótido donante puede contener uno o más cambios, inserciones, deleciones, inversiones o reordenamientos de bases individuales con respecto a la secuencia genómica, siempre que haya suficiente homología para apoyar la recombinación homóloga. Alternativamente, una secuencia donante puede contener una secuencia no homóloga flanqueada por dos regiones de homología. Además, las secuencias del donante pueden comprender una molécula de vector que contiene secuencias que no son homólogas a la región de interés en la cromatina celular. Generalmente, la región o regiones homólogas de una secuencia donante tendrán al menos un 50% de identidad de secuencia con una secuencia genómica con la que se desea la recombinación. En ciertas realizaciones, está presente una identidad de secuencia del 60%, 70%, 80%, 90%, 95%, 98%, 99% o 99,9%. Puede estar presente cualquier valor entre 1% y 100% de identidad de secuencia, dependiendo de la longitud del polinucleótido donante.

**[0134]** Una molécula donante puede contener varias regiones discontinuas de homología con la cromatina celular. Por ejemplo, para la inserción dirigida de secuencias que normalmente no están presentes en una región de interés, dichas secuencias pueden estar presentes en una molécula de ácido nucleico donante y estar flanqueadas por regiones de homología para secuenciar en la región de interés.

**[0135]** Para simplificar los ensayos (*p. ej.*, hibridación, PCR, digestión con enzimas de restricción) para determinar la inserción exitosa de la secuencia donante, pueden existir ciertas diferencias de secuencia en la secuencia donante en comparación con la secuencia genómica. Preferiblemente, si se encuentra en una región de codificación, tales diferencias en la secuencia de nucleótidos no cambiarán la secuencia de aminoácidos, o harán cambios silenciosos de aminoácidos (es decir, cambios que no afectan la estructura o función de la proteína). El polinucleótido del donante puede contener opcionalmente cambios en las secuencias correspondientes a los sitios de unión al dominio del dedo de zinc en la región de interés, para evitar la escisión de las secuencias del donante que se han introducido en la cromatina celular por recombinación homóloga.

**[0136]** El polinucleótido donante puede ser ADN o ARN, monocatenario o bicatenario y puede introducirse en una célula en forma lineal o circular. Si se introduce en forma lineal, los extremos de la secuencia del donante pueden protegerse (*p. ej.*, de la degradación exonucleolítica) por métodos conocidos por los expertos en la materia. Por ejemplo, se añaden uno o más residuos de dideoxinucleótidos al extremo 3' de una molécula lineal y/o los oligonucleótidos autocomplementarios se unen a uno o ambos extremos. Ver, *p. ej.*, Chang et al. (1987) Proc. Natl. Acad. Sci. EE.UU. 84: 4959-4963; Nehls y col. (1996) Science 272: 886-889. Los métodos adicionales para proteger los polinucleótidos exógenos de la degradación incluyen, pero no se limitan a la adición de grupos amino terminales y el uso de enlaces internucleotídicos modificados tales como, *p. ej.*, fosforotioatos, fosforamidatos y residuos de O-metilo ribosa o desoxirribosa. Se puede introducir un polinucleótido en una célula como parte de una molécula de vector que tiene secuencias adicionales tales como, *p. ej.*, orígenes de replicación, promotores y genes que codifican resistencia a antibióticos. Además, los polinucleótidos donadores pueden introducirse como ácido nucleico desnudo, como ácido nucleico complejado con un agente como un liposoma o poloxámero, o pueden administrarse mediante virus (*p. ej.*, adenovirus, AAV).

**[0137]** Sin estar limitado por una teoría, parece que la presencia de una ruptura bicatenaria en una secuencia celular, junto con la presencia de una molécula de ADN exógeno que tiene homología con una región adyacente o que rodea la ruptura, activa mecanismos celulares que reparan la ruptura mediante la transferencia de información de secuencia de la molécula donante a la secuencia celular (*p. ej.*, genómica o cromosómica); es decir, por un proceso de recombinación homóloga. Los métodos de los solicitantes combinan ventajosamente las potentes capacidades de direccionamiento de las ZFP diseñadas con un dominio de escisión (o medio dominio de escisión) para dirigir específicamente una ruptura bicatenaria a la región del genoma en donde se desea la recombinación.

**[0138]** Para la alteración de una secuencia cromosómica, no es necesario copiar toda la secuencia del donante en el

cromosoma, siempre que se copie suficiente secuencia del donante para efectuar la alteración de secuencia deseada.

5 **[0139]** La eficacia de la inserción de secuencias de donantes por recombinación homóloga está inversamente relacionada con la distancia, en el ADN celular, entre la ruptura de doble cadena y el sitio en donde se desea la recombinación. En otras palabras, se observan mayores eficiencias de recombinación homólogas cuando la ruptura bicatenaria está más cerca del sitio en donde se desea la recombinación. En los casos en que un sitio preciso de recombinación no está predeterminado (*p. ej.*, el evento de recombinación deseado puede ocurrir durante un intervalo de secuencia genómica), la longitud y la secuencia del ácido nucleico donante, junto con el (los) sitio(s) de escisión, son seleccionados para obtener el evento de recombinación deseado. En los casos en que el evento deseado está diseñado para cambiar la secuencia de un solo par de nucleótidos en una secuencia genómica, la cromatina celular se divide dentro de 10.000 nucleótidos a cada lado de ese par de nucleótidos. En ciertas realizaciones, la escisión ocurre dentro de 500, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5 o 2 nucleótidos, o cualquier valor integral entre 2 y 1000 nucleótidos, a cada lado del par de nucleótidos cuya secuencia se va a cambiar.

15 **[0140]** Como se detalló anteriormente, los sitios de unión para dos proteínas de fusión, que comprenden cada uno un dominio de unión de dedos de zinc y un medio dominio de escisión, pueden ubicarse a 5-8 o 15-18 nucleótidos separados, medidos desde el borde de cada sitio de unión más cercano al otro sitio de unión, y se produce la escisión entre los sitios de unión. Si la escisión ocurre en un solo sitio o en múltiples sitios entre los sitios de unión es irrelevante, ya que las secuencias genómicas escindidas son reemplazadas por las secuencias del donante. Por lo tanto, para una alteración eficiente de la secuencia de un solo par de nucleótidos mediante recombinación dirigida, el punto medio de la región entre los sitios de unión está dentro de los 10.000 nucleótidos de ese par de nucleótidos, preferiblemente dentro de 1.000 nucleótidos, o 500 nucleótidos, o 200 nucleótidos, o 100 nucleótidos, o 50 nucleótidos, o 20 nucleótidos, o 10 nucleótidos, o 5 nucleótidos, o 2 nucleótidos, o un nucleótido, o en el par de nucleótidos de interés.

25 **[0141]** Un cromosoma homólogo puede servir como el polinucleótido donante. Así, *p. ej.*, la corrección de una mutación en un heterocigoto se puede lograr mediante el diseño de proteínas de fusión que se unen y escinden la secuencia mutante en un cromosoma, pero no escinden la secuencia de tipo silvestre en el cromosoma homólogo. La ruptura de doble cadena en el cromosoma portador de mutación estimula un proceso de "conversión de genes" basado en la homología en donde la secuencia de tipo silvestre del cromosoma homólogo se copia en el cromosoma escindido, restaurando así dos copias de la secuencia de tipo silvestre.

35 **[0142]** También se describen métodos y composiciones que pueden mejorar los niveles de recombinación dirigida incluyendo, pero no limitado al uso de fusiones de dominio funcional ZFP adicionales para activar la expresión de genes implicados en la recombinación homóloga, tales como, *p. ej.*, miembros del grupo de epistasis RAD52 (*p. ej.*, *Rad50*, *Rad51*, *Rad51B*, *Rad51C*, *Rad51D*, *Rad52*, *Rad54*, *Rad54B*, *Mre11*, *XRCC2*, *XRCC3*), genes cuyos productos interactúan con los productos genéticos mencionados anteriormente (*p. ej.*, *BRCA1*, *BRCA2*) y/o genes en el complejo NBS1. De manera similar, las fusiones de dominio funcional ZFP se pueden usar, en combinación con los métodos y composiciones descritas aquí, para reprimir la expresión de genes implicados en la unión de extremos no homólogos (*p. ej.*, *Ku70/80*, *XRCC4*, polimerasa de poli(ADP ribosa), ADN ligasa 4). *Ver, p. ej.*, Yáñez et al. (1998) *Gene Therapy* 5: 149-159; Hoeijmakers (2001) *Nature* 411: 366-374; Johnson y col. (2001) *Biochem. Soc. Trans.* 29: 196-201; Tsuchi y col. (2002) *Oncogene* 21: 8967-8980. Los métodos para la activación y represión de la expresión génica usando fusiones entre un dominio de unión de dedos de zinc y un dominio funcional se describen en la patente de EE.UU. N° 6,534,261. Los métodos de represión adicionales incluyen el uso de oligonucleótidos antisentido y/o ARN interferente pequeño (ARNip o ARNi) dirigido a la secuencia del gen a reprimir.

45 **[0143]** Como alternativa a, o además de activar la expresión de productos génicos implicados en la recombinación homóloga, pueden usarse fusiones de estas proteínas (o fragmentos funcionales de las mismas) con un dominio de unión de dedos de zinc dirigido a la región de interés, se puede utilizar para reclutar estas proteínas (proteínas de recombinación) a la región de interés, aumentando así su concentración local y estimulando aún más los procesos de recombinación homóloga. Alternativamente, un polipéptido implicado en la recombinación homóloga como se describió anteriormente (o un fragmento funcional del mismo) puede ser parte de una proteína de triple fusión que comprende un dominio de unión de dedos de zinc, un dominio de escisión (o medio dominio de escisión) y la proteína de recombinación (o fragmento funcional del mismo). Las proteínas adicionales implicadas en la conversión génica y la remodelación de la cromatina relacionada con la recombinación, que pueden usarse en los métodos y composiciones mencionadas anteriormente, incluyen acetiltransferasas de histona (*p. ej.*, *Esalp*, *Tip60*), metiltransferasas de histona (*p. ej.*, *Dot1p*), quinasas de histona y fosfatasa de histona.

60 **[0144]** Se ha informado que la proteína p53 desempeña un papel central en la represión de la recombinación homóloga (RH). *Ver, p. ej.*, Valerie et al., (2003) *Oncogene* 22: 5792-5812; Janz y col. (2002) *Oncogene* 21: 5929-5933. Por ejemplo, la tasa de RH en las líneas tumorales humanas deficientes en p53 es 10.000 veces mayor que en los fibroblastos humanos primarios, y hay un aumento de 100 veces en la RH en las células tumorales con un p53 no funcional en comparación con aquellos con p53 funcional. Mekeel y col. (1997) *Oncogene* 14: 1847-1857. Además, la sobreexpresión de mutantes negativos dominantes de p53 conduce a un aumento de 20 veces en la recombinación espontánea. Bertrand y col. (1997) *Oncogene* 14: 1117-1122. El análisis de diferentes mutaciones de p53 ha revelado que los roles de p53 en la transactivación transcripcional y el control del punto de control del ciclo celular G1 son separables de su participación en la RH. Saintigny y col. (1999) *Oncogene* 18: 3553-3563; Boehden y col. (2003)

Oncogene 22: 4111-4117. En consecuencia, la regulación negativa de la actividad de p53 puede servir para aumentar la eficiencia de la recombinación homóloga dirigida utilizando los métodos y composiciones descritas en este documento. Se puede usar cualquier método para la regulación negativa de la actividad de p53, que incluye, pero no se limita a cotransfección y sobreexpresión de un p53 dominante negativo mutante o represión dirigida de la expresión del gen p53 de acuerdo con los métodos divulgados, p. ej., en la Patente de los Estados Unidos de propiedad conjunta N° 6,534,261.

**[0145]** Otros incrementos en la eficiencia de la recombinación dirigida, en las células que comprenden una molécula de dedos de cinc/fusión de nucleasa y una molécula de ADN donadora, se logran mediante el bloqueo de las células en la fase G<sub>2</sub> del ciclo celular, cuando procesos de reparación conducidos por homología son máximamente activos. Tal arresto se puede lograr de varias maneras. Por ejemplo, las células pueden tratarse con, p. ej., fármacos, compuestos y/o moléculas pequeñas que influyen en la progresión del ciclo celular para detener las células en la fase G<sub>2</sub>. Las moléculas ejemplares de este tipo incluyen, pero no se limitan a, compuestos que afectan la polimerización de microtúbulos (p. ej., Vinblastina, nocodazol, taxol), compuestos que interactúan con el ADN (p. ej., dicloruro de diamina *cis*-platino (II), cisplatino, doxorubicina) y/o compuestos que afectan la síntesis de ADN (p. ej., timidina, hidroxiourea, L-mimosina, etopósido, 5-fluorouracilo). Se logran aumentos adicionales en la eficiencia de recombinación mediante el uso de inhibidores de desacetilasa de histona (HDAC) (p. ej., butirato de sodio, trichostatina A) que alteran la estructura de la cromatina para hacer que el ADN genómico sea más accesible para la maquinaria de recombinación celular.

**[0146]** Los métodos adicionales para la detención del ciclo celular incluyen la sobreexpresión de las proteínas que inhiben la actividad de las quinasas del ciclo celular CDK, p. ej., mediante la introducción de un ADNc que codifica la proteína en la célula o mediante la introducción en la célula de una ZFP de ingeniería que activa expresión del gen que codifica la proteína. La detención del ciclo celular también se logra inhibiendo la actividad de las ciclinas y las CDK, p. ej., utilizando métodos de ARNi (p. ej., la patente de EE.UU. N° 6,506,559) o introduciendo en la célula una ZFP diseñada que reprime la expresión de uno o más genes involucrados en la progresión del ciclo celular tal como, p. ej., ciclina y/o genes CDK. Véase, p. ej., la Patente de los Estados Unidos de copropiedad N° 6,534,261, para conocer los métodos para la síntesis de proteínas de dedos de zinc modificadas genéticamente para la regulación de la expresión génica.

**[0147]** Alternativamente, en ciertos casos, la escisión dirigida se realiza en ausencia de un polinucleótido donante (preferiblemente en fase S o G<sub>2</sub>), y se produce recombinación entre cromosomas homólogos.

#### Métodos para detectar factores celulares que facilitan la recombinación homóloga

**[0148]** Dado que la recombinación homóloga es un proceso de múltiples pasos que requiere la modificación de los extremos del ADN y el reclutamiento de varios factores celulares en un complejo proteico, la adición de uno o más factores exógenos, junto con con ADN donante y vectores que codifican fusiones de dominio de escisión con dedos de zinc, se pueden usar para facilitar la recombinación homóloga dirigida. Un método ejemplar para identificar dicho factor o factores emplea análisis de la expresión génica usando microarrays (p. ej., matrices Affymetrix Gene Chip®) para comparar los patrones de expresión de ARNm de diferentes células. Por ejemplo, las células que exhiben una mayor capacidad para estimular la recombinación homóloga impulsada por ruptura de doble cadena en presencia de ADN del donante y fusiones de dominio de escisión con dedos de zinc, sin ayuda o en condiciones que se sabe que aumentan el nivel de corrección génica, pueden analizarse para sus patrones de expresión génica en comparación con las células que carecen de dicha capacidad. Los genes que se regulan al alza o a la baja de una manera que se correlaciona directamente con el aumento de los niveles de recombinación homóloga se identifican de este modo y se pueden clonar en cualquiera de varios vectores de expresión. Estas construcciones de expresión pueden cotransfectarse junto con fusiones de dominio de escisión con dedos de zinc y construcciones de donantes para producir métodos mejorados para lograr una recombinación homóloga de alta eficiencia. Alternativamente, la expresión de tales genes puede regularse adecuadamente usando proteínas de dedos de zinc modificadas genéticamente que modulan la expresión (ya sea activación o represión) de uno o más de estos genes. Véase, p. ej., la Patente de los Estados Unidos de copropiedad N° 6,534,261, para conocer los métodos para la síntesis de proteínas de dedos de zinc modificadas genéticamente para regular la expresión génica.

**[0149]** Como ejemplo, se observó que los diferentes clones obtenidos en los experimentos descritos en el Ejemplo 9 y la Figura 27 exhibieron una amplia gama de frecuencias de recombinación homólogas, cuando se transfectaron con ADN de donante y plásmidos que codifican fusiones de dominio de escisión con dedos de zinc. La expresión génica en clones que muestra una alta frecuencia de recombinación dirigida puede compararse con la de los clones que exhiben una baja frecuencia, y pueden identificarse patrones de expresión únicos para los clones anteriores.

**[0150]** Como un ejemplo adicional, los estudios que usan inhibidores del ciclo celular (p. ej., nocodazol o vinblastina, véanse, p. ej., los Ejemplos 11, 14 y 15) mostraron que las células detenidas en la fase G<sub>2</sub> del ciclo celular llevaron a cabo una recombinación homóloga a niveles más altos, lo que indica que los factores celulares responsables de la recombinación homóloga pueden expresarse preferentemente o ser activos en G<sub>2</sub>. Una forma de identificar estos factores es comparar los patrones de expresión de ARNm entre los clones de células HEK 293 transfectadas que llevan a cabo la corrección génica a niveles altos y bajos (p. ej., clon T18 vs. clon T7). Se hacen comparaciones

similares entre estas líneas celulares en respuesta a compuestos que detienen las células en la fase G2. Los genes candidatos que se expresan diferencialmente en células que llevan a cabo una recombinación homóloga a una velocidad mayor, sin ayuda o en respuesta a compuestos que detienen las células en G2, se identifican, clonan y reintroducen en las células para determinar si su expresión es suficiente para volver a capitular las tasas mejoradas. Alternativamente, la expresión de dichos genes candidatos se activa usando factores de transcripción de dedos de zinc diseñados como se describe, p. ej., en la Patente de Estados Unidos N° 6,534,261.

**Vectores de expresión**

[0151] Un ácido nucleico que codifica una o más ZFP o proteínas de fusión ZFP puede clonarse en un vector para transformación en células procariotas o eucariotas para replicación y/o expresión. Los vectores pueden ser vectores procariotas, p. ej., plásmidos o vectores lanzadera, vectores de insectos o vectores eucariotas. Un ácido nucleico que codifica una ZFP también puede clonarse en un vector de expresión, para administración a una célula vegetal, célula animal, preferiblemente una célula de mamífero o una célula humana, célula fúngica, célula bacteriana o célula protozoaria.

[0152] Para obtener la expresión de un gen clonado o ácido nucleico, las secuencias que codifican una proteína de fusión ZFP o ZFP se subclonan típicamente en un vector de expresión que contiene un promotor para dirigir la transcripción. Los promotores bacterianos y eucarióticos adecuados son bien conocidos en la técnica y se describen, p. ej., en Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2ª ed. 1989; 3ª ed., 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); y *Current Protocols in Molecular Biology* (Ausubel et al., supra). Los sistemas de expresión bacteriana para expresar la ZFP están disponibles, p. ej., en *E. coli*, *Bacillus sp.* y *Salmonella* (Palva et al., *Gene* 22: 229-235 (1983)). Los kits para tales sistemas de expresión están disponibles comercialmente. Los sistemas de expresión eucarióticos para células de mamíferos, levaduras y células de insectos son bien conocidos por los expertos en la materia y también están disponibles comercialmente.

[0153] El promotor utilizado para dirigir la expresión de un ácido nucleico que codifica ZFP depende de la aplicación particular. Por ejemplo, un promotor constitutivo fuerte se usa típicamente para la expresión y purificación de ZFP. Por el contrario, cuando se administra una ZFP *in vivo* para la regulación génica, ya sea constitutiva o inducible se usa un promotor, dependiendo del uso particular de la ZFP. Además, un promotor preferido para la administración de una ZFP puede ser un promotor débil, tal como HSV TK o un promotor que tiene una actividad similar. El promotor típicamente también incluye elementos que responden a la transactivación, p. ej., elementos de respuesta a hipoxia, elementos de respuesta Gal4, elemento de respuesta represor lac y sistemas de control de moléculas pequeñas como sistemas regulados por tet y el sistema RU-486 (ver, p. ej., Gossen & Bujard, *PNAS* 89: 5547 (1992); Oligino y col., *Gene Ther.* 5: 491-496 (1998); Wang y col., *Gene Ther.* 4: 432-441 (1997); Neering y col., *Blood* 88: 1147-1155 (1996); y Rendahl y col., *Nat. Biotechnol.* 16:757-761 (1998)). El promotor MNDU3 también se puede usar, y es preferentemente activo en células madre hematopoyéticas CD34\*.

[0154] Además del promotor, el vector de expresión contiene típicamente una unidad de transcripción o casete de expresión que contiene todos los elementos adicionales necesarios para la expresión del ácido nucleico en las células huésped, ya sea procariotas o eucariotas. Por lo tanto, un casete de expresión típico contiene un promotor unido operativamente, p. ej., a una secuencia de ácido nucleico que codifica la ZFP, y las señales requeridas, p. ej., para una poliadenilación eficiente del transcrito, terminación transcripcional, sitios de unión a ribosomas o terminación de traducción. Los elementos adicionales del casete pueden incluir, p. ej., potenciadores y señales de empalme heterólogo.

[0155] El vector de expresión particular usado para transportar la información genética a la célula se selecciona con respecto al uso previsto de la ZFP, p. ej., expresión en plantas, animales, bacterias, hongos, protozoos, etc. (ver los vectores de expresión descritos abajo). Los vectores de expresión bacterianos estándar incluyen plásmidos tales como plásmidos basados en pBR322, pSKF, pET23D y sistemas de expresión de fusión disponibles comercialmente como GST y LacZ. Una proteína de fusión ejemplar es la proteína de unión a maltosa, "MBP". Dichas proteínas de fusión se usan para la purificación de la ZFP. Las etiquetas de epítopo también se pueden agregar a proteínas recombinantes para proporcionar métodos convenientes de aislamiento, para controlar la expresión y para controlar la localización celular y subcelular, p. ej., c-myc o FLAG.

[0156] Los vectores de expresión que contienen elementos reguladores de virus eucariotas se usan a menudo en vectores de expresión eucariotas, p. ej., vectores SV40, vectores de virus del papiloma y vectores derivados del virus Epstein-Barr. Otros vectores eucarióticos ejemplares incluyen pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, y cualquier otro vector que permita la expresión de proteínas bajo la dirección del promotor temprano SV40, promotor tardío SV40, promotor tardío SV40, promotor de metalotioneína, virus de tumor mamario murino promotor, promotor del virus del sarcoma de Rous, promotor de polihedrina u otros promotores que se muestran efectivos para la expresión en células eucariotas.

[0157] Algunos sistemas de expresión tienen marcadores para la selección de líneas celulares transfectadas de forma estable tales como timidina quinasa, higromicina B fosfotransferasa y dihidrofolato reductasa. Los sistemas de expresión de alto rendimiento también son adecuados, como el uso de un vector de baculovirus en células de insecto,

con una secuencia que codifica ZFP bajo la dirección del promotor de polihedrina u otros promotores de baculovirus fuertes.

5 [0158] Los elementos que se incluyen típicamente en vectores de expresión también incluyen un replicón que funciona en *E. coli*, un gen que codifica resistencia a los antibióticos para permitir la selección de bacterias que albergan plásmidos recombinantes, y sitios de restricción únicos en regiones no esenciales del plásmido para permitir la inserción de secuencias recombinantes.

10 [0159] Métodos de transfección estándar se utilizan para producir líneas bacterianas de mamíferos, levaduras o células de insecto que expresan grandes cantidades de proteína, que se purifican luego utilizando técnicas estándar (véase, *p. ej.*, Colley et al., J. Biol. Chem. 264: 17619-17622 (1989); Guide to Protein Purification, in Methods in Enzymology, vol. 182 (Deutscher, ed., 1990)). La transformación de células eucariotas y procariontes se realiza de acuerdo con técnicas estándar (véase, *p. ej.*, Morrison, J. Bact. 132: 349-351 (1977); Clark-Curtiss & Curtiss, Methods in Enzymology 101: 347-362 (Wu et al., eds, 1983).

15 [0160] Se puede usar cualquiera de los procedimientos bien conocidos para introducir secuencias de nucleótidos extraños en las células huésped, que incluyen el uso de transfección con fosfato de calcio, polibreno, fusión de protoplastos, electroporación, métodos ultrasónicos (*p. ej.*, sonoporación), liposomas, microinyección, ADN desnudo, vectores plasmídicos, vectores virales, tanto episomales como integrativos, y cualquiera de los otros métodos bien conocidos para introducir ADN genómico clonado, ADNc, ADN sintético u otro material genético extraño en una célula huésped (*ver*, por ejemplo, Sambrook *et al.*, *supra*). Solo es necesario que el procedimiento de ingeniería genética particular utilizado sea capaz de introducir con éxito al menos un gen en la célula huésped capaz de expresar la proteína de elección.

#### 25 **Ácidos nucleicos que codifican proteínas de fusión y suministro a células**

[0161] Se pueden usar métodos de transferencia génica basados en virus y no virus convencionales para introducir ácidos nucleicos que codifican ZFP modificadas en células (*p. ej.*, células de mamífero) y tejidos diana. Tales métodos también pueden usarse para administrar ácidos nucleicos que codifican ZFP a células *in vitro*. Los ácidos nucleicos que codifican ZFP se pueden administrar para usos de terapia génica *in vivo* o *ex vivo*. Los sistemas de administración de vectores no virales incluyen plásmidos de ADN, ácido nucleico desnudo y ácido nucleico complejado con un vehículo de administración tal como un liposoma o un poloxámero. Los sistemas de entrega de vectores virales incluyen virus de ADN y ARN, que tienen genomas episomales o integrados después de la entrega a la célula. Para una revisión de los procedimientos de terapia génica, *ver* Anderson, Science 256: 808-813 (1992); Nabel y Felgner, TIBTECH 11: 211-217 (1993); Mitani y Caskey, TIBTECH 11: 162-166 (1993); Dillon, TIBTECH 11: 167-175 (1993); Miller, Nature 357: 455-460 (1992); Van Brunt, Biotechnology 6 (10): 1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8: 35-36 (1995); Kremer y Perricaudet, British Medical Bulletin 51 (1): 31-44 (1995); Haddada et al., En Current Topics in Microbiology and Immunology Doerfler y Böhm (eds) (1995); y Yu et al., Gene Therapy 1: 13-26 (1994).

40 [0162] Los métodos de suministro no viral de ácidos nucleicos que codifican ZFP modificadas incluyen electroporación, lipofección, microinyección, biolística, virosomas, liposomas, inmunoliposomas, polimerización o conjugados de lípidos:ácido nucleico, ADN desnudo, viriones artificiales y captación mejorada de agente ADN. La sonoporación que usa, *p. ej.*, el sistema Sonitron 2000 (RichMar) también se puede usar para la administración de ácidos nucleicos.

45 [0163] Los sistemas de entrega adicionales a modo de ejemplo de ácido nucleico incluyen los proporcionados por Amaxa Biosystems (Colonia, Alemania), MaxCyte, Inc. (Rockville, Maryland) y BTX Molecular Delivery Systems (Holliston, MA).

50 [0164] La lipofección se describe en, *p. ej.*, US 5,049,386, US 4,946,787; y US 4,897,355) y los reactivos de lipofección se venden comercialmente (*p. ej.*, Transfectam™ y Lipofectin™). Los lípidos catiónicos y neutros que son adecuados para la lipofección de polinucleótidos de reconocimiento de receptor eficiente incluyen los de Felgner, WO 91/17424, WO 91/16024. El suministro puede ser a células (administración *ex vivo*) o tejidos diana (administración *in vivo*).

55 [0165] La preparación de complejos de lípido:ácido nucleico, incluyendo liposomas dirigidos tales como complejos de inmunolípido, es bien conocido para un experto en la técnica (véase, *p. ej.*, Crystal, Science 270: 404-410 (1995); Blaese et al., Cancer Gene Ther. 2: 291-297 (1995); Behr et al., Bioconjugate Chem. 5: 382-389 (1994); Remy et al., Bioconjugate Chem. 5: 647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); Patentes de EE.UU. Nos 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028 y 4,946,787).

60 [0166] El uso de sistemas basados en virus de ARN o ADN para el suministro de ácidos nucleicos que codifican ZFP modificadas aprovecha los procesos altamente evolucionados para dirigir un virus a células específicas en el cuerpo y traficar la carga viral al núcleo. Los vectores virales pueden administrarse directamente a pacientes (*in vivo*) o pueden usarse para tratar células *in vitro* y las células modificadas se administran a pacientes (*ex vivo*). Los sistemas convencionales basados en virus para el suministro de ZFP incluyen, entre otros, vectores retrovirales, lentivirus,

adenovirales, adenoasociados, vaccinia y herpes simplex para la transferencia de genes. La integración en el genoma del huésped es posible con los métodos de transferencia génica de retrovirus, lentivirus y virus adenoasociados, lo que a menudo resulta en la expresión a largo plazo del transgen insertado. Además, se han observado altas eficiencias de transducción en muchos tipos diferentes de células y tejidos diana.

**[0167]** El tropismo de un retrovirus puede alterarse incorporando proteínas de envoltura extrañas, expandiendo la población diana potencial de células diana. Los vectores lentivirales son vectores retrovirales que son capaces de transducir o infectar células que no se dividen y típicamente producen altos títulos virales. La selección de un sistema de transferencia génica retroviral depende del tejido objetivo. Los vectores retrovirales están compuestos por repeticiones terminales largas que actúan en *cis* con capacidad de empaquetamiento de hasta 6-10 kb de secuencia extraña. Las LTR mínimas que actúan en *cis* son suficientes para la replicación y el empaquetamiento de los vectores, que luego se utilizan para integrar el gen terapéutico en la célula diana para proporcionar una expresión transgénica permanente. Los vectores retrovirales ampliamente utilizados incluyen aquellos basados en el virus de la leucemia murina (MuLV), el virus de la leucemia del mono gibón (GaLV), el virus de la inmunodeficiencia de Simian (SIV), el virus de la inmunodeficiencia humana (VIH) y sus combinaciones (véase, p. ej., Buchscher et al., J. Virol. 66: 2731-2739 (1992); Johann et al., J. Virol. 66: 1635-1640 (1992); Sommerfelt et al., Virol. 176: 58-59 (1990); Wilson y col., J. Virol. 63: 2374-2378 (1989); Miller y col., J. Virol. 65: 2220-2224 (1991); PCT/US94/05700).

**[0168]** En aplicaciones en las que se prefiere la expresión transitoria de una proteína de fusión ZFP, se pueden usar sistemas basados en adenovirus. Los vectores basados en adenovirus son capaces de una eficiencia de transducción muy alta en muchos tipos de células y no requieren división celular. Con tales vectores, se han obtenido un alto título y altos niveles de expresión. Este vector se puede producir en grandes cantidades en un sistema relativamente simple. Los vectores de virus adenoasociados ("AAV") también se usan para transducir células con ácidos nucleicos diana, p. ej., en la producción *in vitro* de ácidos nucleicos y péptidos, y para procedimientos de terapia génica *in vivo* y *ex vivo* (ver, p. ej., West et al., Virology 160: 38-47 (1987); Patente de Estados Unidos N° 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5: 793-801 (1994); Muzyczka, J. Clin. Invest. 94: 1351 (1994). La construcción de vectores de AAV recombinantes se describe en varias publicaciones, incluyendo la Patente de Estados Unidos N° 5,173,414; Tratschin y col., Mol. Cell. Biol. 5: 3251-3260 (1985); Tratschin y col., Mol. Cell. Biol. 4: 2072-2081 (1984); Hermonat y Muzyczka, PNAS 81: 6466-6470 (1984) y Samulski et al., J. Virol. 63: 03822-3828 (1989).

**[0169]** Al menos seis enfoques de vectores virales están actualmente disponibles para la transferencia de genes en ensayos clínicos, que utilizan enfoques que implican la complementación de vectores defectuosos por genes insertados en líneas celulares auxiliares para generar el agente transductor.

**[0170]** pLASN y MFG-S son ejemplos de vectores retrovirales que se han utilizado en ensayos clínicos (Dunbar et al., Blood 85: 3048-305 (1995); Kohn y col., Nat. Med. 1: 1017-102 (1995); Malech et al., PNAS 94:22 12133-12138 (1997)). PA317/pLASN fue el primer vector terapéutico utilizado en un ensayo de terapia génica. (Blaese y col., Science 270: 475-480 (1995)). Se han observado eficiencias de transducción de 50% o más para los vectores empaquetados de MFG-S. (Ellem y col., Immunol Immunother. 44 (1): 10-20 (1997); Dranoff y col., Hum. Gene Ther. 1: 111-2 (1997)).

**[0171]** Vectores de virus adenoasociados recombinantes (rAAV) son sistemas de suministro de genes alternativos un prometedor basado en el virus tipo 2 adenoasociado defectuoso y no patógeno de parvovirus. Todos los vectores se derivan de un plásmido que retiene sólo las repeticiones terminales invertidas de AAV 145 bp que flanquean el casete de expresión transgénica. Transferencia eficaz de genes y el suministro transgénico estable debido a la integración en los genomas de la célula transducida son características clave para este sistema de vectores (Wagner et al., Lancet 351: 9117 1702-3 (1998), Kearns et al., Gene Ther. 9: 748- 55 (1996)).

**[0172]** Los vectores adenovirales recombinantes deficientes en replicación (Ad) pueden producirse en un título alto e infectar fácilmente varios tipos de células diferentes. La mayoría de los vectores de adenovirus están diseñados de tal manera que un transgen reemplaza los genes Ad E1a, E1b, y/o E3; posteriormente, el vector defectuoso de replicación se propaga en células humanas 293 que suministran la función del gen eliminado en *trans*. Los vectores publicitarios pueden transducir múltiples tipos de tejidos *in vivo*, incluidas las células no divididas y diferenciadas, como las que se encuentran en el hígado, los riñones y los músculos. Los vectores publicitarios convencionales tienen una gran capacidad de carga. Un ejemplo del uso de un vector Ad en un ensayo clínico implicaba la terapia con polinucleótidos para la inmunización antitumoral con inyección intramuscular (Sternan et al., Hum. Gene Ther. 7: 1083-9 (1998)). Ejemplos adicionales del uso de vectores de adenovirus para la transferencia de genes en ensayos clínicos incluyen Rosenecker et al., Infection 24: 1 5-10 (1996); Sternan y col., Hum. Gene Ther. 9: 7 1083-1089 (1998); Welsh y col., Hum. Gene Ther. 2: 205-18 (1995); Alvarez y col., Hum. Gene Ther. 5: 597-613 (1997); Topf y col., Gene Ther. 5: 507-513 (1998); Sternan y col., Hum. Gene Ther. 7: 1083-1089 (1998).

**[0173]** Células de empaquetamiento se utilizan para formar partículas de virus que son capaces de infectar una célula huésped. Dichas células incluyen células 293, que empaquetan adenovirus, y células  $\psi$ 2 o células PA317, que empaquetan retrovirus. Los vectores virales utilizados en la terapia génica generalmente son generados por una línea celular productora que empaqueta un vector de ácido nucleico en una partícula viral. Los vectores típicamente contienen las secuencias virales mínimas requeridas para el empaquetamiento y la posterior integración en un huésped (si corresponde), siendo reemplazadas otras secuencias virales por un casete de expresión que codifica la

proteína a expresar. Las funciones virales faltantes son suministradas en *trans* por la línea celular de empaquetamiento. Por ejemplo, los vectores de AAV utilizados en terapia génica típicamente poseen secuencias de repetición terminal invertida (ITR) del genoma de AAV que se requieren para el empaquetamiento y la integración en el genoma del huésped. El ADN viral está empaquetado en una línea celular, que contiene un plásmido auxiliar que codifica los otros genes AAV, a saber, *rep* y *cap*, pero que carece de secuencias ITR. La línea celular también está infectada con adenovirus como ayudante. El virus auxiliar promueve la replicación del vector AAV y la expresión de genes AAV a partir del plásmido auxiliar. El plásmido auxiliar no está empaquetado en cantidades significativas debido a la falta de secuencias ITR. La contaminación con adenovirus puede reducirse mediante, p. ej., un tratamiento térmico al que el adenovirus es más sensible que el AAV.

**[0174]** En muchas aplicaciones de terapia génica, es deseable que el vector de terapia génica se administre con un alto grado de especificidad a un tipo de tejido particular. En consecuencia, un vector viral puede modificarse para tener especificidad para un tipo de célula dado al expresar un ligando como una proteína de fusión con una proteína de cubierta viral en la superficie externa del virus. El ligando se elige para tener afinidad por un receptor que se sabe que está presente en el tipo de célula de interés. Por ejemplo, Han et al., Proc. Natl. Acad. Sci. EE.UU. 92: 9747-9751 (1995), informó que el virus de la leucemia murina Moloney puede modificarse para expresar heregulina humana fusionada a gp70, y el virus recombinante infecta ciertas células de cáncer de mama humano que expresan el receptor del factor de crecimiento epidérmico humano. Este principio puede extenderse a otros pares de células diana-virus, en las que la célula diana expresa un receptor y el virus expresa una proteína de fusión que comprende un ligando para el receptor de la superficie celular. Por ejemplo, el fago filamentoso se puede diseñar para mostrar fragmentos de anticuerpos (p. ej., FAB o Fv) que tienen afinidad de unión específica para prácticamente cualquier receptor celular elegido. Aunque la descripción anterior se aplica principalmente a los vectores virales, los mismos principios se pueden aplicar a los vectores no virales. Dichos vectores se pueden diseñar para contener secuencias de captación específicas que favorecen la captación por células diana específicas.

**[0175]** Los vectores de terapia génica pueden administrarse *in vivo* mediante la administración a un paciente individual, típicamente por administración sistémica (p. ej., intravenosa, intraperitoneal, intramuscular, subdérmica, o infusión intracraneal) o aplicación tópica, como se describe a continuación. Alternativamente, los vectores pueden administrarse a las células *ex vivo*, como las células explantadas de un paciente individual (p. ej., linfocitos, aspirados de médula ósea, biopsia de tejido) o células madre hematopoyéticas de donantes universales, seguido de reimplantación de las células en un paciente, generalmente después selección para células que han incorporado el vector.

**[0176]** Transfección de células *ex vivo* para el diagnóstico, investigación, o para terapia génica (p. ej., a través de re-infusión de las transfectadas células en el organismo huésped) es bien conocido por los expertos en la técnica. Las células pueden aislarse del organismo sujeto, transfectarse con un ácido nucleico de ZFP (gen o ADNc) y reinfundirse nuevamente dentro del organismo sujeto (p. ej., paciente). Los expertos en la materia conocen bien varios tipos de células adecuadas para la transfección *ex vivo* (véase, p. ej., Freshney et al., Culture of Animal Cells, A Manual of Basic Technique (3ª ed. 1994)) y las referencias allí citadas para una discusión sobre cómo aislar y cultivar células de pacientes).

**[0177]** Las células madre pueden usarse en procedimientos *ex vivo* para la transfección celular y terapia génica. La ventaja de usar células madre es que pueden diferenciarse en otros tipos de células *in vitro*, o pueden introducirse en un mamífero (como el donante de las células) donde se injertarán en la médula ósea. Los métodos para diferenciar células CD34<sup>+</sup> *in vitro* en tipos de células inmunes clínicamente importantes usando citoquinas un GM-CSF tal, IFN- $\gamma$  y TNF- $\alpha$  son conocidos (véase Inaba et al., J. Exp. Med. 176:1693-1702 (1992)).

**[0178]** Las células madre se aíslan para transducción y diferenciación utilizando métodos conocidos. Por ejemplo, las células madre se aíslan de las células de la médula ósea mediante el barrido de las células de la médula ósea con anticuerpos que se unen a las células no deseadas, como CD4<sup>+</sup> y CD8<sup>+</sup> (células T), CD45<sup>+</sup> (células panB), GR-1 (granulocitos) y lad (células presentadoras de antígeno diferenciadas) (véase Inaba et al., J. Exp. Med. 176:1693-1702 (1992)).

**[0179]** Los vectores (p. ej., retrovirus, adenovirus, liposomas, etc.) que contienen ácidos nucleicos terapéuticos ZFP también se pueden administrar directamente a un organismo para la transducción de células *in vivo*. Alternativamente, se puede administrar ADN desnudo. La administración se realiza por cualquiera de las rutas que normalmente se utilizan para introducir una molécula en contacto final con las células sanguíneas o tisulares, incluidas, entre otras, inyección, infusión, aplicación tópica y electroporación. Los métodos adecuados para administrar tales ácidos nucleicos están disponibles y son bien conocidos por los expertos en la técnica, y, aunque se puede usar más de una ruta para administrar una composición particular, una ruta particular a menudo puede proporcionar una reacción más inmediata y más efectiva que Otra ruta.

**[0180]** Los métodos para la introducción de ADN en células madre hematopoyéticas se describen, p. ej., en la Patente de EE.UU. N° 5,928,638.

**[0181]** Los vehículos farmacéuticamente aceptables se determinan en parte por la composición particular que se

administra, así como por el método particular usado para administrar la composición. En consecuencia, existe una amplia variedad de formulaciones adecuadas de composiciones farmacéuticas disponibles, como se describe a continuación (véase, p. ej., Remington's Pharmaceutical Sciences, 17ª ed., 1989).

5 **[0182]** Las construcciones de ADN pueden introducirse en el genoma de una planta huésped deseada mediante una  
 variedad de técnicas convencionales. Para revisiones de tales técnicas, ver, p. ej., Weissbach & Weissbach Methods  
 for Plant Molecular Biology (1988, Academic Press, NY) Sección VIII, págs. 421-463; y Grierson & Corey, Plant  
 Molecular Biology (1988, 2ª ed.), Blackie, Londres, Cap. 7-9. Por ejemplo, la construcción de ADN puede introducirse  
 10 directamente en el ADN genómico de la célula vegetal utilizando técnicas como la electroporación y la microinyección  
 de protoplastos de células vegetales, o las construcciones de ADN pueden introducirse directamente en el tejido  
 vegetal utilizando métodos biolísticos, como el bombardeo de partículas de ADN. (véase, p. ej., Klein et al (1987)  
 Nature 327: 70-73). Alternativamente, las construcciones de ADN pueden combinarse con regiones flanqueantes de  
 ADN-T adecuadas e introducirse en un vector huésped de *Agrobacterium tumefaciens* convencional. Las técnicas de  
 15 transformación mediadas por *Agrobacterium tumefaciens*, incluido el desarme y el uso de vectores binarios, están bien  
 descritas en la literatura científica. Véanse, p. ej., Horsch et al (1984) Science 233: 496-498, y Fraley et al (1983) Proc.  
 Nat'l. Acad. Sci. EE.UU. 80: 4803. Las funciones de virulencia del huésped *Agrobacterium tumefaciens* dirigirán la  
 inserción de la construcción y el marcador adyacente en el ADN de la célula vegetal cuando la célula está infectada  
 por la bacteria usando el vector binario de ADN T (Bevan (1984) Nuc. Acid Res. 12: 8711-8721) o el procedimiento de  
 20 cocultivo (Horsch et al (1985) Science 227: 1229-1231). Generalmente, el sistema de transformación de *Agrobacterium*  
 se usa para diseñar plantas dicotiledóneas (Bevan et al (1982) Ann. Rev. Genet 16: 357-384; Rogers et al (1986)  
 Methods Enzymol. 118: 627-641). El sistema de transformación de *Agrobacterium* también puede usarse para  
 transformar, así como transferir, ADN a plantas monocotiledóneas y células vegetales. Ver Hernalsteen et al (1984)  
 EMBO J3: 3039-3041; Hooykass-Van Slogteren y otros (1984) Nature 311: 763-764; Grimsley y otros (1987) Nature  
 325: 1677-179; Boulton et al (1989) Plant Mol. Biol. 12: 31-40.; y Gould et al (1991) Plant Physiol. 95: 426-434.

25 **[0183]** Los métodos alternativos de transferencia y transformación génica incluyen, pero no se limitan a,  
 transformación de protoplastos a través de calcio, polietilenglicol (PEG) o absorción de ADN desnudo mediada por  
 electroporación (ver Paszkowski et al. (1984) EMBO J 3: 2717-2722, Potrykus et al. (1985) Molec. Gen. Genet. 199:  
 169-177; Fromm et al. (1985) Proc. Nat. Acad. Sci. EE.UU. 82: 5824-5828; y Shimamoto (1989) Nature 338: 274-276)  
 30 y electroporación de tejidos vegetales (D'Halluin et al. (1992) Plant Cell 4: 1495-1505). Los métodos adicionales para  
 la transformación de células vegetales incluyen microinyección, absorción de ADN mediada por carburo de silicio  
 (Kaeppeler et al. (1990) Plant Cell Reporter 9: 415-418) y bombardeo con microproyectiles (véase Klein et al. (1988)  
 Proc. Nat. Acad. Sci. EE.UU. 85: 4305-4309; y Gordon-Kamm et al. (1990) Plant Cell 2: 603-618).

35 **[0184]** Los métodos y composiciones descritas se pueden utilizar para insertar secuencias exógenas en una  
 predeterminada posición en un genoma de célula vegetal. Esto es útil en la medida en que la expresión de un transgen  
 introducido en el genoma de una planta depende de manera crítica de su sitio de integración. Por consiguiente, los  
 genes que codifican, p. ej., nutrientes, antibióticos o moléculas terapéuticas pueden insertarse, por recombinación  
 40 dirigida, en regiones del genoma de una planta favorables para su expresión.

**[0185]** Las células vegetales transformadas que se producen mediante cualquiera de las técnicas de transformación  
 anteriores pueden cultivarse para regenerar una planta completa que posee el genotipo transformado y, por lo tanto,  
 el fenotipo deseado. Dichas técnicas de regeneración se basan en la manipulación de ciertas fitohormonas en un  
 medio de crecimiento de cultivo de tejidos, que normalmente se basa en un marcador de biocida y/o herbicida que se  
 45 ha introducido junto con las secuencias de nucleótidos deseadas. La regeneración de plantas a partir de protoplastos  
 cultivados se describe en Evans, et al., "Protoplasts Isolation and Culture" en Handbook of Plant Cell Culture, págs.  
 124-176, Macmillan Publishing Company, Nueva York, 1983; y Binding, Regeneration of Plants, Plant Protoplasts, pp.  
 21-73, CRC Press, Boca Raton, 1985. La regeneración también se puede obtener a partir de callos de plantas,  
 explantes, órganos, pólenes, embriones o partes de los mismos. Dichas técnicas de regeneración se describen  
 50 generalmente en Klee et al (1987) Ann. Rev. de Plant Phys. 38: 467-486.

**[0186]** Ácidos nucleicos introducidos en una célula de la planta se pueden utilizar para conferir los rasgos deseados a  
 esencialmente cualquier planta. Se puede diseñar una amplia variedad de plantas y sistemas de células vegetales  
 para las características fisiológicas y agronómicas deseadas descritas en el presente documento usando las  
 55 construcciones de ácido nucleico de la presente descripción y los diversos métodos de transformación mencionados  
 anteriormente. Las plantas diana y las células vegetales para ingeniería incluyen, entre otras, las plantas  
 monocotiledóneas y dicotiledóneas, como los cultivos que incluyen cultivos de granos (p. ej., trigo, maíz, arroz, mijo,  
 cebada), cultivos de frutas (p. ej., tomate, manzana, pera, fresa, naranja), cultivos forrajeros (p. ej., alfalfa), cultivos de  
 hortalizas de raíz (p. ej., zanahoria, papa, remolacha azucarera, ñame), cultivos de hortalizas de hoja (p. ej., lechuga,  
 60 espinacas); plantas con flores (p. ej., petunia, rosa, crisantemo), coníferas y pinos (p. ej., abeto, píceas); plantas  
 utilizadas en fitorremediación (p. ej., plantas de acumulación de metales pesados); cultivos oleaginosos (p. ej., girasol,  
 colza) y plantas utilizadas con fines experimentales (p. ej., *Arabidopsis*). Por lo tanto, los métodos y composiciones  
 descritas tienen uso en una amplia gama de plantas, incluidas, entre otras, especies de los géneros *Asparagus*, *Avena*,  
 65 *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Cucurbita*, *Daucus*, *Glycine*, *Hordeum*, *Lactuca*, *Lycopersicon*, *Malus*, *Manihot*,  
*Nicotiana*, *Oryza*, *Persea*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Secale*, *Solanum*, *Sorghum*, *Triticum*, *Vitis*, *Vigna* y *Zea*.

[0187] Un experto en la materia reconocerá que después de que el casete de expresión se incorpore de manera estable en plantas transgénicas y se confirme que es operable, puede introducirse en otras plantas por cruce sexual. Se puede usar cualquiera de una serie de técnicas de reproducción estándar, dependiendo de la especie a cruzar.

5 [0188] Se puede identificar y aislar una célula vegetal transformada, un callo, un tejido o una planta seleccionando o cribando el material vegetal modificado para detectar rasgos codificados por los genes marcadores presentes en el ADN transformante. Por ejemplo, la selección puede realizarse haciendo crecer el material vegetal modificado en medios que contengan una cantidad inhibidora del antibiótico o herbicida al que la construcción del gen transformante confiere resistencia. Además, las plantas transformadas y las células vegetales también pueden identificarse mediante la detección de las actividades de cualquier gen marcador visible (*p. ej.*, los genes de  $\beta$ -glucuronidasa, luciferasa, B o C1) que pueden estar presentes en las construcciones de ácido nucleico recombinante. Dichas metodologías de selección y cribado son bien conocidas por los expertos en la materia.

15 [0189] Métodos físicos y bioquímicos también se pueden usar para identificar transformantes de plantas o células de plantas que contienen construcciones de genes insertados. Estos métodos incluyen, entre otros: 1) análisis Southern o amplificación por PCR para detectar y determinar la estructura del inserto de ADN recombinante; 2) transferencia Northern, protección S1 RNasa, extensión de cebador o transcriptasa inversa-amplificación por PCR para detectar y examinar transcripciones de ARN de las construcciones génicas; 3) ensayos enzimáticos para detectar la actividad enzimática o ribozima, donde dichos productos génicos están codificados por la construcción génica; 4) electroforesis en gel de proteínas, técnicas de transferencia Western, inmunoprecipitación o inmunoensayos ligados a enzimas, donde los productos de construcción génica son proteínas. También se pueden utilizar técnicas adicionales, tales como hibridación *in situ*, tinción enzimática e inmunotinción, para detectar la presencia o expresión de la construcción recombinante en órganos y tejidos vegetales específicos. Los métodos para realizar todos estos ensayos son bien conocidos por los expertos en la materia.

25 [0190] Efectos de la manipulación de genes utilizando los métodos descritos en este documento se pueden observar, *p. ej.*, transferencias northern de ARN (*p. ej.*, ARNm) aisladas de los tejidos de interés. Típicamente, si la cantidad de ARNm ha aumentado, se puede suponer que el gen endógeno correspondiente se está expresando a un ritmo mayor que antes. Se pueden usar otros métodos para medir la actividad del gen *y/o* CYP74B. Se pueden usar diferentes tipos de ensayos enzimáticos, dependiendo del sustrato utilizado y el método de detección del aumento o disminución de un producto o subproducto de reacción. Además, los niveles de *y/o* proteína CYP74B expresados se pueden medir inmunoquímicamente, es decir, ELISA, RIA, EIA y otros ensayos basados en anticuerpos bien conocidos por los expertos en la técnica, tales como ensayos de detección electroforética (ya sea con tinción o transferencia Western). El transgen puede expresarse selectivamente en algunos tejidos de la planta o en algunas etapas de desarrollo, o el transgen puede expresarse en sustancialmente todos los tejidos vegetales, sustancialmente a lo largo de todo su ciclo de vida. Sin embargo, cualquier modo de expresión combinatoria también es aplicable.

35 [0191] La presente descripción también abarca semillas de las plantas transgénicas descritas anteriormente en las que la semilla tiene el constructo transgénico o genético. La presente descripción abarca además la progenie, clones, líneas celulares o células de las plantas transgénicas descritas anteriormente en donde dicha progenie, clon, línea celular o célula tiene el transgen o construcción génica.

#### Vehículos de suministro

45 [0192] Un factor importante en la administración de compuestos de polipéptidos, tales como las proteínas de fusión ZFP, es asegurar que el polipéptido tenga la capacidad de atravesar la membrana plasmática de una célula, o la membrana de un compartimento intracelular como el núcleo. Las membranas celulares están compuestas por bicapas de lípidos y proteínas que son libremente permeables a compuestos lipofílicos no iónicos pequeños y son inherentemente impermeables a compuestos polares, macromoléculas y agentes terapéuticos o de diagnóstico. Sin embargo, se han descrito proteínas y otros compuestos como los liposomas, que tienen la capacidad de translocar polipéptidos como las ZFP a través de una membrana celular.

55 [0193] Por ejemplo, los "polipéptidos de translocación de membrana" tienen subsecuencias de aminoácidos anfífilos o hidrófobos que tienen la capacidad de actuar como portadores de translocación de membrana. En una realización, las proteínas de homeodominio tienen la capacidad de translocarse a través de las membranas celulares. Se encontró que el péptido internalizable más corto de una proteína de homeodominio, Antennapedia, es la tercera hélice de la proteína, desde la posición de aminoácidos 43 a 58 (*véase, p. ej.*, Prochiantz, *Current Opinion in Neurobiology* 6: 629-634 (1996)). Se encontró que otra subsecuencia, el dominio h (hidrófobo) de los péptidos señal, tenía características similares de translocación de la membrana celular (*véase, p. ej.*, Lin et al., *J. Biol. Chem.* 270: 14255-14258 (1995)).

60 [0194] Ejemplos de secuencias de péptidos que pueden ser vinculados a una proteína, para facilitar la absorción de la proteína en las células, incluyen, pero no se limitan a: un péptido de 11 aminoácidos de la proteína tat del VIH; una secuencia peptídica de 20 residuos que corresponde a los aminoácidos 84-103 de la proteína p16 (*véase* Fahraeus et al., *Current Biology* 6:84 (1996)); la tercera hélice del homeodominio largo de 60 aminoácidos de Antennapedia (Derossi et al., *J. Biol. Chem.* 269: 10444 (1994)); la región h de un péptido señal tal como la región h del factor de crecimiento de fibroblastos de Kaposi (K-FGF) (Lin et al., *supra*); o el dominio de translocación VP22 de HSV (Elliot y

O'Hare, Cell 88: 223-233 (1997)). Otros restos químicos adecuados que proporcionan una captación celular mejorada también pueden estar químicamente unidos a las ZFP. Los dominios de translocación de membrana (es decir, dominios de internalización) también se pueden seleccionar de bibliotecas de secuencias peptídicas aleatorizadas. Ver, p. ej., Yeh et al. (2003) Molecular Therapy 7 (5): S461, Resumen nº 1191.

**[0195]** Moléculas de toxina también tienen la capacidad de polipéptidos de transporte a través de membranas celulares. A menudo, tales moléculas (llamadas "toxinas binarias") se componen de al menos dos partes: un dominio o polipéptido de translocación/unión y un dominio o polipéptido de toxina separado. Típicamente, el dominio de translocación o polipéptido se une a un receptor celular, y luego la toxina se transporta a la célula. Se han utilizado varias toxinas bacterianas, incluidas la toxina iota *Clostridium perfringens*, la toxina diftérica (DT), la exotoxina *Pseudomonas A* (PE), la toxina pertussis (PT), la toxina *Bacillus anthracis* y la ciclasa de adenilato pertussis (CYA), para administrar péptidos a la célula citosol como fusiones internas o amino terminales (Arora et al., J. Biol. Chem., 268: 3334-3341 (1993); Perelle et al., Infect. Immun., 61: 5147-5156 (1993); Stenmark y col., J. Cell Biol. 113: 1025-1032 (1991); Donnelly y col., PNAS 90: 3530-3534 (1993); Carbonetti y col., Abstr. Annu. Meet. Am. Soc. Microbiol. 95: 295 (1995); Sebo y col., Infect. Immun. 63: 3851-3857 (1995); Klimpel y col., PNAS EE.UU. 89: 10277-10281 (1992); y Novak y col., J. Biol Chem. 267: 17186-17193 (1992)).

**[0196]** Dichas secuencias peptídicas pueden usarse para translocar ZFP a través de una membrana celular. Las ZFP pueden fusionarse convenientemente o derivatizarse con tales secuencias. Típicamente, la secuencia de translocación se proporciona como parte de una proteína de fusión. Opcionalmente, se puede usar un enlazador para vincular la ZFP y la secuencia de translocación. Se puede usar cualquier enlazador adecuado, p. ej., un enlazador peptídico.

**[0197]** La ZFP puede también ser introducida en una célula animal, preferiblemente una célula de mamífero, a través de liposomas y derivados de liposomas tales como inmunoliposomas. El término "liposoma" se refiere a vesículas compuestas de una o más bicapas lipídicas ordenadas concéntricamente, que encapsulan una fase acuosa. La fase acuosa típicamente contiene el compuesto que se administrará a la célula, es decir, una ZFP.

**[0198]** El liposoma se fusiona con la membrana plasmática, liberando así el fármaco en el citosol. Alternativamente, el liposoma es fagocitado o absorbido por la célula en una vesícula de transporte. Una vez en el endosoma o fagosoma, el liposoma se degrada o se fusiona con la membrana de la vesícula de transporte y libera su contenido.

**[0199]** En los métodos actuales de suministro de fármacos a través de liposomas, el liposoma finalmente se vuelve permeable y libera el compuesto encapsulado (en este caso, una ZFP) en el tejido o célula diana. Para el suministro sistémico o específico de tejido, esto se puede lograr, p. ej., de manera pasiva en donde la bicapa de liposomas se degrada con el tiempo a través de la acción de varios agentes en el cuerpo. Alternativamente, la liberación activa del fármaco implica el uso de un agente para inducir un cambio de permeabilidad en la vesícula del liposoma. Las membranas de los liposomas pueden construirse de modo que se desestabilicen cuando el ambiente se vuelva ácido cerca de la membrana de los liposomas (véase, p. ej., PNAS 84: 7851 (1987); Biochemistry 28: 908 (1989)). Cuando los liposomas son endocitados por una célula diana, p. ej., se desestabilizan y liberan su contenido. Esta desestabilización se denomina fusogénesis. La dioleoilfosfatidiletanolamina (DOPE) es la base de muchos sistemas "fusogénicos".

**[0200]** Tales liposomas comprenden típicamente una ZFP y un componente lipídico, p. ej., un lípido neutro y/o catiónico, que incluye opcionalmente una molécula de reconocimiento de receptor tal como un anticuerpo que se une a un receptor de superficie celular predeterminado o ligando (p. ej., un antígeno). Hay una variedad de métodos disponibles para preparar liposomas como se describe en, p. ej., Szoka et al., Ann. Rev. Biophys. Bioeng. 9:467 (1980), Pat. de EE.UU. N<sup>os</sup> 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,946,787, Publicación de PCT N<sup>o</sup> WO 9117424, Deamer & Bangham, Biochim. Biophys. Acta 443:629-634 (1976); Fraley, et al., PNAS 76:3348-3352 (1979); Hope et al., Biochim. Biophys. Acta 812:55-65 (1985); Mayer et al., Biochim. Biophys. Acta 858:161-168 (1986); Williams et al., PNAS 85:242-246 (1988); Liposomes (Ostro (ed.), 1983, Capítulo 1); Hope et al., Chem. Phys. Lip. 40:89 (1986); Gregoriadis, Liposome Technology (1984) y Lasic, Liposomes: from Physics to Applications (1993)). Los métodos adecuados incluyen, p. ej., sonicación, extrusión, alta presión/homogeneización, microfluidización, diálisis con detergente, fusión inducida por calcio de vesículas de liposomas pequeños y métodos de fusión con éter, todos los cuales son conocidos por los expertos en la técnica.

**[0201]** En ciertas realizaciones, es deseable apuntar a los liposomas usando restos dirigidos que son específicos para un tipo de célula particular, tejido y similares. Se ha descrito el direccionamiento de liposomas usando una variedad de restos de direccionamiento (p. ej., Ligandos, receptores y anticuerpos monoclonales). Véanse, p. ej., las Patentes de EE.UU. Números 4,957,773 y 4,603.044.

**[0202]** Los ejemplos de restos dirigidos incluyen anticuerpos monoclonales específicos para antígenos asociados con neoplasias, tales como antígeno específico de cáncer de próstata y MAGE. Los tumores también pueden diagnosticarse mediante la detección de productos genéticos resultantes de la activación o sobreexpresión de oncogenes, como ras o c-erbB2. Además, muchos tumores expresan antígenos normalmente expresados por el tejido fetal, como la alfafetoproteína (AFP) y el antígeno carcinoembrionario (CEA). Los sitios de infección viral se pueden

diagnosticar utilizando varios antígenos virales, como los antígenos de hepatitis B centrales y de superficie (HBVc, HBV), antígenos de hepatitis C, antígenos del virus de Epstein-Barr, virus de inmunodeficiencia humana tipo 1 (VIH1) y antígenos del virus del papiloma. La inflamación puede detectarse usando moléculas específicamente reconocidas por moléculas de superficie que se expresan en sitios de inflamación tales como integrinas (p. ej., VCAM-1), receptores de selectina (p. ej., ELAM-1) y similares.

[0203] Se pueden usar métodos estándar para acoplar agentes de direccionamiento a liposomas. Estos métodos generalmente implican la incorporación en liposomas de componentes lipídicos, p. ej., fosfatidiletanolamina, que puede activarse para la unión de agentes de direccionamiento o compuestos lipofílicos derivados, tales como bleomicina derivatizada con lípidos. Los liposomas dirigidos a anticuerpos pueden construirse usando, p. ej., liposomas que incorporan proteína A (véase Renneisen y col., J. Biol. Chem., 265: 16337-16342 (1990) y Leonetti y col., PNAS 87: 2448-2451 (1990).

### Dosis

[0204] Para aplicaciones terapéuticas, la dosis administrada a un paciente, o a una célula que se introducirá en un paciente, en el contexto de la presente divulgación, debería ser suficiente para efectuar una respuesta terapéutica beneficiosa en el paciente con el tiempo. Además, los regímenes de dosificación particulares pueden ser útiles para determinar los cambios fenotípicos en un entorno experimental, p. ej., en estudios de genómica funcional y en modelos celulares o animales. La dosis se determinará por la eficacia y  $K_d$  del ZFP empleado particular, el volumen nuclear de la célula diana y la condición del paciente, así como el peso corporal o el área de superficie del paciente a tratar. El tamaño de la dosis también estará determinado por la existencia, la naturaleza y extensión de cualesquiera efectos secundarios que acompañan la administración de un compuesto o vector particular en un paciente particular.

[0205] La dosificación terapéuticamente eficaz máxima de ZFP para aproximadamente el 99% de unión a sitios diana se calcula en el intervalo de menos de aproximadamente  $1,5 \times 10^5$  a  $1,5 \times 10^6$  copias de la molécula específica ZFP por célula. El número de ZFP por célula para este nivel de unión se calcula de la siguiente manera, utilizando el volumen de un núcleo de células HeLa (aproximadamente  $1000 \mu\text{m}^3$  o  $10^{-12}$  L; Cell Biology, (Altman y Katz, eds. (1976)). Al ser el núcleo de HeLa relativamente grande, este número de dosis se recalcula según sea necesario utilizando el volumen del núcleo de la célula diana. Este cálculo tampoco tiene en cuenta la competencia por la unión de ZFP por otros sitios. Este cálculo también supone que esencialmente todos ZFP se localiza en el núcleo. Se utiliza un valor de  $100 \times K_d$  para calcular aproximadamente el 99% de unión al sitio diana, y un valor de  $10 \times K_d$  se usa para calcular aproximadamente el 90% de unión al sitio diana. ejemplo,  $K_d = 25$  nM

ZFP + sitio diana  $\leftrightarrow$  complejo,  
es decir, ADN + proteína  $\leftrightarrow$  complejo ADN:proteína

$$K_d = \frac{[\text{ADN}] [\text{proteína}]}{[\text{complejoADN:proteína}]}$$

Cuando se une el 50% de ZFP,  $K_d = [\text{proteína}]$   
Entonces cuando  $[\text{proteína}] = 25$  nM y el volumen del núcleo es  $10^{-12}$  L  $[\text{proteína}] = (25 \times 10^{-9} \text{ moles/L}) (10^{-12} \text{ L/núcleo}) (6 \times 10^{23} \text{ moléculas/mol}) = 15.000 \text{ moléculas/núcleo para } 50\% \text{ de unión}$   
Cuando 99% diana está vinculado;  $100 \times K_d = [\text{proteína}]$   
 $100 \times K_d = [\text{proteína}] = 2,5 \mu\text{m}$   
 $(2,5 \times 10^{-6} \text{ moles/L}) (10^{-12} \text{ L/núcleo}) (6 \times 10^{23} \text{ moléculas/mol})$   
= aproximadamente 1,500.000 moléculas por núcleo para un 99% de unión del sitio diana.

[0206] La dosis apropiada de un vector de expresión que codifica una ZFP también se puede calcular teniendo en cuenta la tasa promedio de expresión de ZFP del promotor y la tasa promedio de degradación de ZFP en la célula. En ciertas realizaciones, se usa un promotor débil tal como un promotor HSK TK de tipo silvestre o mutante, como se describió anteriormente. La dosis de ZFP en microgramos se calcula teniendo en cuenta el peso molecular de la ZFP particular que se está empleando.

[0207] Al determinar la cantidad efectiva de la ZFP que se administrará en el tratamiento o la profilaxis de la enfermedad, el médico evalúa los niveles plasmáticos circulantes de la ZFP o el ácido nucleico que codifica la ZFP, las posibles toxicidades de la ZFP, la progresión de la enfermedad y la producción de anticuerpos anti-ZFP. La administración se puede lograr a través de dosis únicas o divididas.

### Composiciones farmacéuticas y administración

[0208] Los ZFPs y vectores de expresión que codifican ZFPs pueden administrarse directamente al paciente para la escisión específica y/o la recombinación, y para aplicaciones terapéuticas o profilácticas, p. ej., cáncer, isquemia, retinopatía diabética, degeneración macular, artritis reumatoide, psoriasis, infección por VIH, anemia falciforme, enfermedad de Alzheimer, distrofia muscular, enfermedades neurodegenerativas, enfermedad vascular, fibrosis quística, accidente cerebrovascular y similares. Los ejemplos de microorganismos que pueden ser inhibidos por la

terapia génica ZFP incluyen bacterias patógenas, p. ej., clamidia, bacterias rickettsiales, micobacterias, estafilococos, estreptococos, neumococos, meningococos y conococos, clebsiella, proteus, serratia, pseudomonas, legionella, difteria, salmonella, bacilos, cólera, tétanos, botulismo, ántrax, peste, leptospirosis y bacterias de la enfermedad de Lyme; hongos infecciosos, p. ej., especies de *Aspergillus*, *Candida*; protozoos como los esporozoos (p. ej., *Plasmodia*), rizopodos (p. ej., *Entamoeba*) y flagelados (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); enfermedades virales, p. ej., hepatitis (A, B o C), virus del herpes (p. ej., VZV, HSV-1, HSV-6, HSV-II, CMV y EBV), VIH, Ébola, adenovirus, virus de la influenza, flavivirus, echovirus, rinovirus, virus coxsackie, coronavirus, virus sincitial respiratorio, virus de paperas, rotavirus, virus del sarampión, virus de la rubéola, parvovirus, virus vaccinia, virus HTLV, virus del dengue, virus del papiloma, poliovirus, virus de la rabia, y el virus de la encefalitis por arbovirus, etc.

**[0209]** La administración de cantidades terapéuticamente eficaces es por cualquiera de las rutas utilizadas normalmente para introducir ZFP en contacto último con el tejido a tratar. Las ZFP se administran de cualquier manera adecuada, preferiblemente con vehículos farmacéuticamente aceptables. Los métodos adecuados para administrar tales moduladores están disponibles y son bien conocidos por los expertos en la técnica, y, aunque se puede usar más de una ruta para administrar una composición particular, una ruta particular a menudo puede proporcionar una reacción más inmediata y más efectiva que otra ruta.

**[0210]** Los vehículos farmacéuticamente aceptables se determinan en parte por la composición particular que se administra, así como por el método particular usado para administrar la composición. En consecuencia, existe una amplia variedad de formulaciones adecuadas de composiciones farmacéuticas que están disponibles (véase, p. ej., Remington's Pharmaceutical Sciences, 17ª edición, 1985)).

**[0211]** Las ZFPs, solas o en combinación con otros componentes adecuados, se pueden convertir en aerosoles las formulaciones (es decir, que pueden ser "nebulizadas") para administrarse por inhalación. Las formulaciones de aerosol se pueden colocar en propulsores aceptables a presión, tales como diclorodifluorometano, propano, nitrógeno y similares.

**[0212]** Las formulaciones adecuadas para administración parenteral, tales como, p. ej., por vía intravenosa, intramuscular, intradérmica y subcutánea, incluyen soluciones de inyección estériles isotónicas acuosas y no acuosas, que pueden contener antioxidantes, tampones, bacteriostatos y solutos que hace que la formulación sea isotónica con la sangre del receptor previsto y suspensiones estériles acuosas y no acuosas que pueden incluir agentes de suspensión, solubilizantes, agentes espesantes, estabilizadores y conservantes. Las composiciones descritas se pueden administrar, p. ej., por infusión intravenosa, por vía oral, tópica, intraperitoneal, intravesical o intratecal. Las formulaciones de compuestos pueden presentarse en envases sellados de dosis unitarias o multidosis, como ampollas y viales. Las soluciones y suspensiones de inyección pueden prepararse a partir de polvos, gránulos y tabletas estériles del tipo descrito anteriormente.

### Aplicaciones

**[0213]** Los métodos y composiciones descritas para la escisión dirigida pueden usarse para inducir mutaciones en una secuencia genómica, p. ej., escindiendo en dos sitios y eliminando secuencias en el medio, por escisión en un sitio único seguido de unión de extremo no homólogo, y/o escindiendo en un sitio para eliminar uno o dos o unos pocos nucleótidos. La escisión dirigida también se puede usar para crear genes inactivados (p. ej., para genómica funcional o validación de dianas) y para facilitar la inserción selectiva de una secuencia en un genoma (es decir, genes inactivados); por ejemplo, para fines de ingeniería celular o sobreexpresión de proteínas. La inserción puede realizarse mediante reemplazos de secuencias cromosómicas a través de recompilación homóloga o mediante integración dirigida, en donde una nueva secuencia (es decir, una secuencia no presente en la región de interés), flanqueada por secuencias homólogas a la región de interés en el cromosoma, se inserta en un sitio diana predeterminado.

**[0214]** Los mismos métodos también se pueden usar para reemplazar una secuencia de tipo silvestre con una secuencia mutante, o para convertir un alelo en un alelo diferente.

**[0215]** La escisión dirigida de genomas virales infectados o integrados puede usarse para tratar infecciones virales en un huésped. Además, la escisión dirigida de genes que codifican receptores para virus puede usarse para bloquear la expresión de dichos receptores, evitando así la infección viral y/o la propagación viral en un organismo huésped. La mutagénesis dirigida de genes que codifican receptores virales (p. ej., los receptores CCR5 y CXCR4 para el VIH) puede usarse para hacer que los receptores no puedan unirse al virus, evitando así nuevas infecciones y bloqueando la propagación de infecciones existentes. Entre los ejemplos no limitantes de virus o receptores virales que pueden atacarse se incluyen el virus del herpes simple (HSV), como HSV-1 y HSV-2, el virus varicela zoster (VZV), el virus de Epstein-Barr (EBV) y el citomegalovirus (CMV), HHV6 y HHV7. La familia de virus de la hepatitis incluye el virus de la hepatitis A (VHA), el virus de la hepatitis B (VHB), el virus de la hepatitis C (VHC), el virus de la hepatitis delta (HDV), el virus de la hepatitis E (HEV) y el virus de la hepatitis G (HGV). Otros virus o sus receptores pueden estar dirigidos, incluidos, entre otros, Picornaviridae (p. ej., Poliovirus, etc.); Caliciviridae; Togaviridae (p. ej., virus de la rubéola, virus del dengue, etc.); Flaviviridae; Coronaviridae; Reoviridae; Birnaviridae; Rhabdoviridae (p. ej., virus de la rabia, etc.); Filoviridae; Paramyxoviridae (p. ej., virus de las paperas, virus del sarampión, virus sincitial respiratorio, etc.); Orthomyxoviridae (p. ej., virus de influenza tipos A, B y C, etc.); Bunyaviridae; Arenaviridae; Retroviridae; lentivirus

(*p. ej.*, HTLV-I; HTLV-II; VIH-1 (también conocido como HTLV-III, LAV, ARV, hTLR, *etc.*) VIH-II); virus de inmunodeficiencia de simio (SIV), virus del papiloma humano (VPH), virus de la gripe y los virus de la encefalitis transmitida por garrapatas. Véase, *p. ej.*, Virology, 3ª Edición (WK Joklik ed. 1988); Fundamental Virology, 2ª edición (BN Fields y DM Knipe, eds. 1991), para una descripción de estos y otros virus. Los receptores para el VIH, *p. ej.*, incluyen CCR-5 y CXCR-4.

**[0216]** De manera similar, el genoma de una bacteria infectante puede mutagenizarse mediante escisión de ADN dirigida seguido de unión final no homóloga, para bloquear o mejorar infecciones bacterianas.

**[0217]** Los métodos descritos para la recombinación dirigida pueden usarse para reemplazar cualquier secuencia genómica con una secuencia homóloga, no idéntica. Por ejemplo, una secuencia genómica mutante puede ser reemplazada por su contraparte de tipo silvestre, proporcionando así métodos para el tratamiento de, *p. ej.*, enfermedades genéticas, trastornos hereditarios, cáncer y enfermedades autoinmunes. De manera similar, un alelo de un gen puede ser reemplazado por un alelo diferente usando los métodos de recombinación dirigida descritos aquí.

**[0218]** Las enfermedades genéticas ejemplares incluyen, pero no se limitan a, acondroplasia, acromatopsia, deficiencia de maltasa ácida, deficiencia de adenosina desaminasa (OMIM N° 102700), adrenoleucodistrofia, síndrome de aicardi, deficiencia de antitripsina alfa-1, alfa-talasemia y síndrome de insensibilidad a los andrógenos, síndrome de apertura, ventrículo derecho arritmogénico, displasia, ataxia telangictasia, síndrome de Barth, beta-talasemia, síndrome de nevo azul de la ampolla de goma, enfermedad de Canavan, enfermedades granulomatosas crónicas (CGD), síndrome de cri du chat, fibrosis quística, enfermedad de dermis, displasia ectodérmica, anemia de Fanconi, fibrodisplasia osificante progresiva, síndrome de X frágil, galactosemia, enfermedad de Gaucher, gangliosidosis generalizadas (*p. ej.*, GM1), la hemocromatosis, la mutación de la hemoglobina C en el 6º codón de beta-globina (HbC), la hemofilia, la enfermedad de Huntington, síndrome de Hurler, hipofosfatasa, síndrome de Klinefelter, enfermedad de Krabbes, síndrome de Langer-Giedion, deficiencia de adhesión leucocitaria (LAD, OMIM N° 116920), leucodistrofia, síndrome de QT largo, síndrome de Marfan, síndrome de Moebius, mucopolisacaridosis (MPS), síndrome de la rótula ungueal, diabetes insípida nefrogénica, neurofibromatosis, enfermedad de Neimann-Pick, osteogénesis imperfecta, porfiria, síndrome de Prader-Willi, progeria, síndrome de Proteus, retinoblastoma, síndrome de retinoblastoma, síndrome de Rubinstein-Taybi, síndrome de Sanfilippo, inmunodeficiencia combinada severa (SCID), síndrome de Shwachman, enfermedad de células falciformes (anemia de células falciformes), síndrome de Smith-Magenis, síndrome de Stickler, enfermedad de Tay-Sachs, síndrome de trombocitopenia de radio ausente (TAR), Síndrome de Treacher Collins, trisomía, esclerosis tuberosa, síndrome de Turner, trastorno del ciclo de la urea, enfermedad de von Hippel-Landau, síndrome de Waardenburg, síndrome de Williams, enfermedad de Wilson, síndrome de Wiskott-Aldrich, síndrome linfoproliferativo ligado al cromosoma X (XLP, OMIM N° 308240).

**[0219]** Las enfermedades ejemplares adicionales que pueden tratarse mediante escisión de ADN dirigida y/o recombinación homóloga incluyen inmunodeficiencias adquiridas, enfermedades de almacenamiento lisosómico (*p. ej.*, enfermedad de Gaucher, GM1, enfermedad de Fabry y enfermedad de Tay-Sachs), mucopolisacahidosis (*p. ej.*, enfermedad de Hunter, enfermedad de Hurler), hemoglobinopatías (*p. ej.*, enfermedades de células falciformes, HbC,  $\alpha$ -talasemia,  $\beta$ -talasemia) y hemofilias.

**[0220]** En ciertos casos, se desea la alteración de una secuencia genómica en una célula pluripotente (*p. ej.*, una célula madre hematopoyética). Los métodos para la movilización, enriquecimiento y cultivo de células madre hematopoyéticas son conocidos en la técnica. Véanse, *p. ej.*, las patentes de los Estados Unidos 5,061,620; 5,681,559; 6,335,195; 6,645,489 y 6,667,064. Las células madre tratadas pueden devolverse a un paciente para el tratamiento de diversas enfermedades, incluidas, entre otras, SCID y anemia falciforme.

**[0221]** En muchos de estos casos, una región de interés comprende una mutación, y el polinucleótido donante comprende la secuencia de tipo silvestre correspondiente. De manera similar, una secuencia genómica de tipo silvestre puede ser reemplazada por una secuencia mutante, si es deseable. Por ejemplo, la sobreexpresión de un oncogén puede revertirse mutando el gen o reemplazando sus secuencias de control con secuencias que soportan un nivel de expresión más bajo y no patológico. Como otro ejemplo, el alelo de tipo silvestre del gen ApoA1 puede reemplazarse por el alelo ApoA1 Milano, para tratar la aterosclerosis. De hecho, cualquier patología dependiente de una secuencia genómica particular, de cualquier manera, puede corregirse o aliviarse usando los métodos y composiciones descritas aquí.

**[0222]** La escisión dirigida y la recombinación dirigida también se pueden usar para alterar secuencias no codificantes (*p. ej.*, secuencias reguladoras tales como promotores, potenciadores, iniciadores, terminadores, sitios de empalme) para alterar los niveles de expresión de un producto génico. Dichos métodos pueden usarse, *p. ej.*, con fines terapéuticos, genómica funcional y/o estudios de validación de dianas.

**[0223]** Las composiciones y métodos descritos en el presente documento también permiten enfoques y sistemas novedosos para abordar las reacciones inmunes de un huésped a los injertos alogénicos. En particular, un problema importante que se enfrenta cuando las células madre alogénicas (o cualquier tipo de célula alogénica) se injertan en un receptor huésped es el alto riesgo de rechazo por parte del sistema inmunitario del huésped, mediado principalmente a través del reconocimiento del complejo mayor de histocompatibilidad (MHC) en la superficie de las

células injertadas. El MHC comprende la(s) proteína(s) HLA clase I que funcionan como heterodímeros que están compuestos por una subunidad  $\beta$  común y subunidades  $\alpha$  variables. Se ha demostrado que los injertos de tejido derivados de células madre que carecen de HLA escapan de la respuesta inmune del huésped. Ver, p. ej., Coffman et al. J Immunol 151, 425-35. (1993); Markmann y col. Transplantation 54, 1085-9. (1992); Koller y col. Science 248, 1227-30. (1990) Usando las composiciones y métodos descritos en este documento, los genes que codifican las proteínas HLA involucradas en el rechazo del injerto se pueden escindir, mutagenizar o alterar por recombinación, ya sea en sus secuencias codificantes o reguladoras, de modo que su expresión se bloquee o expresen un producto no funcional. Por ejemplo, al inactivar el gen que codifica el gen de la subunidad  $\beta$  común (microglobulina  $\beta 2$ ) utilizando proteínas de fusión ZFP como se describe en este documento, el HLA clase I se puede eliminar de las células para generar rápida y confiablemente células madre *nulas* HLA clase I de cualquier donante, reduciendo así la necesidad de haplotipos MHC de donantes/receptores estrechamente emparejados durante el injerto de células madre.

**[0224]** La inactivación de cualquier gen (p. ej., el gen de la microglobulina  $\beta 2$ ) se puede lograr, p. ej., mediante un solo evento de escisión, por escisión seguida de unión final no homóloga, por escisión en dos sitios seguidos de la unión para eliminar la secuencia entre los dos sitios de escisión, por recombinación dirigida de un codón sin sentido o sin sentido en la región de codificación, o por recombinación dirigida de una secuencia irrelevante (es decir, una secuencia de "relleno") en el gen o su región reguladora, de modo que para interrumpir el gen o la región reguladora.

**[0225]** La modificación dirigida de la estructura de la cromatina, como se describe en el documento WO 01/83793, copropietario, puede usarse para facilitar la unión de proteínas de fusión a la cromatina celular.

**[0226]** En realizaciones adicionales, se pueden usar una o más fusiones entre un dominio de unión de dedos de zinc y una recombinasa (o fragmento funcional del mismo), además de o en lugar de las fusiones de dominio de escisión de dedos de zinc descritas aquí, para facilitar la recombinación dirigida. Véanse, p. ej., la patente de los Estados Unidos de propiedad conjunta N° 6,534,261 y Akopian et al. (2003) Proc. Natl. Acad. Sci. EE.UU. 100: 8688-8691.

**[0227]** En realizaciones adicionales de la descripción, los métodos y composiciones descritas se usan para proporcionar fusiones de dominios de unión a ZFP con dominios de activación o represión transcripcionales que requieren dimerización (ya sea homodimerización o heterodimerización) para su actividad. En estos casos, un polipéptido de fusión comprende un dominio de unión de dedos de zinc y un monómero de dominio funcional (p. ej., un monómero de un dominio de activación transcripcional o represión dimérico). La unión de dos de tales polipéptidos de fusión a sitios diana adecuadamente ubicados permite la dimerización para reconstituir un dominio de activación o represión de la transcripción funcional.

## EJEMPLOS

### **Ejemplo 1: Edición de un gen hSMC1L1 cromosómico mediante recombinación dirigida**

**[0228]** El gen hSMC1L1 es el ortólogo humano del *mantenimiento estructural de los cromosomas 1* del gen de levadura incipiente. Una región de este gen que codifica una porción amino terminal de la proteína que incluye el dominio Walker ATPasa fue mutagenizado por escisión dirigida y recombinación. La escisión se dirigió a la región del codón de iniciación de la metionina (nucleótidos 24-26, Figura 1), mediante el diseño de nucleasas quiméricas, que comprenden un dominio de unión al ADN de dedo de zinc y un medio dominio de escisión *Fok I*, que se unen en la vecindad del codon. Por lo tanto, se diseñaron dos dominios de unión de dedos de zinc, uno de los cuales reconoce los nucleótidos 23-34 (contactos primarios a lo largo de la cadena superior como se muestra en la Figura 1), y el otro reconoce los nucleótidos 5-16 (contactos primarios a lo largo de la cadena inferior). Las proteínas de los dedos de zinc se diseñaron como se describe en las patentes estadounidenses de propiedad conjunta 6,453,242 y 6,534,261. Ver la Tabla 2 para las secuencias de aminoácidos de las regiones de reconocimiento de las proteínas de dedo de zinc.

**[0229]** Las secuencias que codifican cada uno de estos dos dominios de unión de ZFP se fusionaron a las secuencias que codifican un medio dominio de escisión *Fok I* (aminoácidos 384-579 de la secuencia nativa *Fok I*; Kita et al (1989) J. Biol. Chem. 264: 5751-5756), de modo que la proteína codificada contenía secuencias *Fok I* en el terminal carboxi y secuencias ZFP en el terminal amino. Cada una de estas secuencias de fusión se clonó luego en un vector de expresión de mamífero modificado pcDNA3 (Figura 2).

Tabla 2: Diseños de dedos de zinc para el gen hSMC1L1

Secuencia diana	F1	F2	F3	F4
CATGGGGTTCCT (SEQ ID NO:27)	RSHDLIE (SEQ ID NO:28)	TSSSLSR (SEQ ID NO:29)	RSDHLST (SEQ ID NO:30)	TNSNRIT (SEQ ID NO:31)
GCGGCGCCGGCG (SEQ ID NO:32)	RSDDLRS (SEQ ID NO:33)	RSDDRKT (SEQ ID NO:34)	RSEDLIR (SEQ ID NO:35)	RSDTLSR (SEQ ID NO:36)

Nota: las secuencias de aminoácidos del dedo de zinc que se muestran arriba (en el código de una letra representan los residuos -1 a +6, con respecto al inicio de la porción alfa-helicoidal de cada dedo de zinc. El dedo F1 está más cerca del extremo amino del proteína, y el dedo F4 está más cerca del terminal carboxi.

**[0230]** Se obtuvo una molécula de ADN donante como sigue. Primero, se amplificó un fragmento de 700 pares de bases de ADN genómico humano que representa los nucleótidos 52415936-52416635 de la cadena "-" del cromosoma X (liberación del genoma humano UCSC en julio de 2003), que incluye el primer exón del gen hSMC1L1 humano. usando ADN genómico de células HEK293 como plantilla. Las secuencias de cebadores utilizados para la amplificación se muestran en la Tabla 3 ("Amplificador inicial 1" y "Amplificador inicial 2"). El producto de PCR se alteró luego, utilizando la metodología de PCR de extensión de solapamiento estándar (ver, p. ej., Ho, et al. (1989) Gene 77: 51-59), lo que resultó en el reemplazo de la secuencia ATGGGG (nucleótidos 24-29 en la Figura 1) a ATAAGAAGC. Este cambio resultó en la conversión del codón ATG (metionina) en un codón ATA (isoleucina) y el reemplazo de GGG (nucleótidos 27-29 en la Figura 1) por la secuencia AGAAGC, permitiendo la discriminación entre secuencias derivadas de donantes y secuencias cromosómicas endógenas después de la recombinación. En la Figura 3 se muestra un diagrama esquemático del gen hSMC1, que incluye secuencias del ADN cromosómico en la región del codón de iniciación y secuencias en el ADN del donante que difieren de la secuencia cromosómica. El fragmento del donante de 700 pares de bases resultante se clonó en pCR4BluntTopo, que no contiene ninguna secuencia homóloga al genoma humano. Ver Figura 4.

**[0231]** Para la mutación dirigida del gen cromosómico hSMC1L1, los dos plásmidos que codifican las fusiones *ZFP-Fok I* y el plásmido donante se introdujeron en  $1 \times 10^6$  células HEK293 mediante transfección usando Lipofectamine 2000® (Invitrogen). Los controles incluyeron células transfectadas solo con los dos plásmidos que codifican las fusiones de *ZFP-Fok I*, células transfectadas solo con el plásmido donante y células transfectadas con un plásmido de control (pEGFP-N1, Clontech). Las células fueron cultivadas en 5% de CO<sub>2</sub> a 37°C. A las 48 horas después de la transfección, se aisló el ADN genómico de las células, y se usaron 200 ng como plantilla para la amplificación por PCR, usando un cebador complementario a una región del gen fuera de su región de homología con las secuencias del donante (nucleótidos 52416677-52416701 en la cadena "-" del cromosoma X; UCSC julio de 2003), y un segundo cebador complementario a una región de la molécula donante en donde se introdujeron mutaciones distintivas. Usando estos dos cebadores, se obtendrá un producto de amplificación de 400 pares de bases del ADN genómico si se ha producido un evento de recombinación dirigido. Las secuencias de estos cebadores se dan en la Tabla 3 (etiquetadas como "específicas de cromosoma" y "específicas de donante", respectivamente). Las condiciones para la amplificación fueron: 94°C, 2 min, seguido de 40 ciclos de 94°C, 30 segundos, 60°C, 1 min, 72°C, 1 min; y un paso final de 72°C, 7 min.

**[0232]** Los resultados de este análisis (Figura 5) indican que se obtuvo un producto de amplificación de 400 pares de bases (marcado como "ADN quimérico" en la Figura) solo con ADN extraído de células que habían sido transfectadas con el plásmido donante y ambos plásmidos *ZFP-FokI*.

Tabla 3: Cebadores de amplificación para el gen hSMC1L1

<b>Amplificador inicial 1</b>	AGCAACAACCTCCTCCGGGGATC (SEQ ID NO:37)
<b>Amplificador inicial 2</b>	TTCCAGACGCGACTCTTTGGC (SEQ ID NO:38)
<b>Cromosoma específico</b>	CTCAGCAAGCGTGAGCTCAGGTCTC (SEQ ID NO:39)
<b>Específico del donante</b>	CAATCAGTTTCAGGAAGCTTCTT (SEQ ID NO:40)
<b>Exterior 1</b>	CTCAGCAAGCGTGAGCTCAGGTCTC (SEQ ID NO:41)
<b>Exterior 2</b>	GGGGTCAAGTAAGGCTGGGAAGC (SEQ ID NO:42)

**[0233]** Para confirmar este resultado, se realizaron dos experimentos adicionales. Primero, el producto de amplificación se clonó en pCR4Blunt-Topo (Invitrogen) y se determinó su secuencia de nucleótidos. Como se muestra en la Figura 6 (SEQ ID NO:6), la secuencia amplificada obtenida del ADN cromosómico de las células transfectadas con los dos plásmidos que codifican *ZFP-FokI* y el plásmido donante contiene la secuencia AAGAAGC que es única para el donante (nucleótidos 395-401 de la secuencia presentada en la Figura 6) unido covalentemente a secuencias cromosómicas que no están presentes en la molécula donante (nucleótidos 32-97 de la Figura 6), lo que indica que las secuencias de los donantes se han recombinado en el cromosoma. En particular, la mutación G → A que convierte

el codón de iniciación en un codon de isoleucina se observa en la posición 395 en la secuencia.

[0234] En un segundo experimento, ADN cromosómico de células transfectadas solo con plásmido donante, células transfectadas con plásmidos de fusión *ZFP-Fok I*, células transfectadas con el plásmido donante y plásmidos de fusión *ZFP-Fok I* o células transfectadas con el plásmido de control EGFP se usó como plantilla para la amplificación, usando cebadores complementarios a secuencias fuera de la región de homología de 700 nucleótidos entre las secuencias de donantes y cromosómicas (identificadas como "Exterior 1" y "Exterior 2" en la Tabla 3). El producto de amplificación resultante se purificó y se usó como plantilla para una segunda reacción de amplificación usando los cebadores específicos de donantes y cromosomas descritos anteriormente (Tabla 3). Esta amplificación produjo un producto de 400 nucleótidos solo de células transfectadas con la construcción del donante y ambas construcciones de fusión *ZFP-Fok I*, un resultado consistente con el reemplazo de secuencias genómicas por recombinación dirigida en estas células.

**Ejemplo 2: Edición de un gen cromosómico de IL2R $\gamma$  por recombinación dirigida**

[0235] El gen IL<sub>2</sub>R $\gamma$  codifica una proteína, conocida como la "cadena gamma del receptor de citocina común", que funciona como una subunidad de varios receptores de interleucina (incluida IL-2R, IL-4R, IL-7R, IL-9R, IL-15R e IL-21R). Las mutaciones en este gen, incluidas las que rodean el extremo 5' del tercer exón (*p. ej.*, el codón de tirosina 91), pueden causar inmunodeficiencia combinada severa ligada al cromosoma X (SCID). Ver, *p. ej.*, Puck et al. (1997) Blood 89: 1968-1977. Se introdujo una mutación en el codón de tirosina 91 (nucleótidos 23-25 de SEQ ID NO:7; Figura 7) en el gen IL2R $\gamma$  mediante escisión y recombinación dirigidas. La escisión se dirigió a esta región mediante el diseño de dos pares de proteínas de dedos de zinc. El primer par (primeras dos filas de la Tabla 4) comprende una proteína de dedo de zinc diseñada para unirse a los nucleótidos 29-40 (contactos primarios a lo largo de la cadena superior como se muestra en la Figura 7) y una proteína de dedo de zinc diseñada para unirse a los nucleótidos 8-20 (contactos primarios a lo largo del hilo inferior). El segundo par (filas tercera y cuarta de la Tabla 4) comprende dos proteínas de dedos de zinc, la primera de las cuales reconoce los nucleótidos 23-34 (contactos primarios a lo largo de la cadena superior como se muestra en la Figura 7) y la segunda reconoce los nucleótidos 8-16 (contactos primarios a lo largo del hilo inferior). Las proteínas de los dedos de zinc se diseñaron como se describe en las patentes estadounidenses de propiedad conjunta 6,453,242 y 6,534,261. Ver la Tabla 4 para las secuencias de aminoácidos de las regiones de reconocimiento de las proteínas de dedo de zinc.

[0236] Las secuencias que codifican los dominios de unión a ZFP se fusionaron con secuencias que codifican un medio dominio de escisión de *Fok I* (aminoácidos 384-579 de la secuencia nativa de *Fok I*, Kita et al., *Supra*), de modo que la proteína codificada contenía secuencias de *Fok I* en las secuencias de carboxilo terminal y ZFP en el terminal amino. Cada una de estas secuencias de fusión se clonó luego en un vector de expresión de mamífero modificado pcDNA3. Vea la Figura 8 para un diagrama esquemático de las construcciones.

**Tabla 4: Diseños de dedos de zinc para el gen IL2R $\gamma$**

Secuencia diana	F1	F2	F3	F4
AACTCGGATAAT (SEQ ID NO:43)	DRSTLIE (SEQ ID NO: 44)	SSSNLSR (SEQ ID NO:45)	RSDDLK (SEQ ID NO:46)	DNSNRIK (SEQ ID NO:47)
TAGAGGaGAAAGG (SEQ ID NO:48)	RSDNLSN (SEQ ID NO:49)	TSSSRIN (SEQ ID NO:50)	RSDHLSQ (SEQ ID NO:51)	RNADRKT (SEQ ID NO:52)
TACAAGAACTCG (SEQ ID NO:53)	RSDDLK (SEQ ID NO:54)	DNSNRIK (SEQ ID NO:55)	RSDALSV (SEQ ID NO:56)	DNANRTK (SEQ ID NO:57)
GGAGAAAGG (SEQ ID NO:58)	RSDHLTQ (SEQ ID NO:59)	QSGNLAR (SEQ ID NO:60)	RSDHLSR (SEQ ID NO:61)	
Nota: Las secuencias de aminoácidos del dedo de zinc que se muestran arriba (en código de una letra) representan los residuos -1 a +6, con respecto al comienzo de la porción alfa-helicoidal de cada dedo de zinc. El dedo F1 está más cerca del extremo amino de la proteína.				

[0237] Se obtuvo una molécula de ADN donante como sigue. Primero, se amplificó un fragmento de 700 pares de bases de ADN humano correspondiente a las posiciones 69196910-69197609 en la cadena "-" del cromosoma X (UCSC, julio de 2003), que incluye el exón 3 del gen IL2R $\gamma$ , utilizando ADN genómica de células K562 como plantilla. Consulte la Figura 9. Las secuencias de los cebadores utilizados para la amplificación se muestran en la Tabla 5 (etiquetada amp inicial 1 y amp inicial 2). El producto de PCR se alteró luego mediante la metodología estándar de PCR de extensión de superposición (Ho, et al., *Supra*) para reemplazar la secuencia TACAAGAACTCGGATAAT (SEQ ID NO:62) con la secuencia TAAAAGAATTCCGACAAC (SEQ ID NO:63). Este reemplazo da como resultado la introducción de una mutación puntual en el nucleótido 25 (Figura 7), que convierte el codón TAC de tirosina 91 en un codón de terminación TAA y permite la discriminación entre las secuencias cromosómicas endógenas y derivadas del donante después de la recombinación, debido a diferencias en las secuencias aguas abajo del codón 91. El fragmento resultante de 700 pares de bases se clonó en pCR4BluntTopo que no contiene ninguna secuencia homóloga al

genoma humano. Ver Figura 10.

**[0238]** Para la mutación dirigida del gen cromosómico IL2R $\gamma$ , el plásmido donante, junto con dos plásmidos que codifican cada uno de un par de fusiones ZFP/*Fok* I, se introdujeron en  $2 \times 10^6$  células K652 usando lipofección/electroporación mixta (Amaxa). Cada uno de los pares ZFP/*Fok* I (ver Tabla 4) se probó en experimentos separados. Los controles incluyeron células transfectadas solo con dos plásmidos que codifican fusiones ZFP-*Fok* I, y células transfectadas solo con el plásmido donante. Las células fueron cultivadas en 5% de CO<sub>2</sub> a 37°C. A las 48 horas después de la transfección, se aisló el ADN genómico de las células y se usaron 200 ng como plantilla para la amplificación por PCR, usando un cebador complementario a una región del gen fuera de su región de homología con las secuencias del donante (nucleótidos 69196839-69196863 en la cadena "+" del cromosoma X; UCSC, julio de 2003), y un segundo cebador complementario a una región de la molécula donante en donde se introdujeron mutaciones distintivas (ver arriba) y cuya secuencia, por lo tanto, difiere de la del ADN cromosómico. Ver la Tabla 5 para las secuencias de cebadores, etiquetadas como "específicas de cromosoma" y "específicas de donante", respectivamente. Usando estos dos cebadores, se obtiene un producto de amplificación de 500 pb del ADN genómico en donde se ha producido un evento de recombinación dirigido. Las condiciones para la amplificación fueron: 94°C, 2 min, seguido de 35 ciclos de 94°C, 30 segundos, 62°C, 1 minuto, 72°C, 45 segundos; y un paso final de 72°C, 7 min.

**[0239]** Los resultados de este análisis (Figura 11) indican que se obtiene un producto de amplificación del tamaño esperado (500 pares de bases) con ADN extraído de células que habían sido transfectadas con el plásmido donante y cualquiera de los pares de plásmidos de codificación ZFP-*Fok* I. El ADN de las células transfectadas con plásmidos que codifican solo un par de ZFP (sin plásmido donante) no dio como resultado la generación del producto de 500 pb, ni el ADN de las células transfectadas solo con el plásmido donante.

**Tabla 5: Cebadores de amplificación para IL2R $\gamma$  génica**

<b>Amplificador inicial 1</b>	TGTCGAGTACATGAATTGCACTTGG (SEQ ID NO: 64)
<b>Amplificador inicial 2</b>	TTAGGTTCTCTGGAGCCCAGGG (SEQ ID NO:65)
<b>Específico al cromosoma</b>	CTCCAAACAGTGGTTCAAGAATCTG (SEQ ID NO:66)
<b>Específico al donante</b>	TCCTCTAGGTAAGAATTCCGACAAC (SEQ ID NO:67)

**[0240]** Para confirmar este resultado, el producto de amplificación obtenido del experimento usando el segundo par de fusiones ZFP/*Fok* I se clonó en pCR4Blunt-Topo (Invitrogen) y se determinó su secuencia de nucleótidos. Como se muestra en la Figura 12 (SEQ ID NO:12), la secuencia consiste en una fusión entre secuencias cromosómicas y secuencias del plásmido donante. En particular, la mutación G a A que convierte la tirosina 91 en un codón de parada se observa en la posición 43 en la secuencia. Las posiciones 43-58 contienen nucleótidos únicos para el donante; los nucleótidos 32-42 y 59-459 son secuencias comunes para el donante y el cromosoma, y los nucleótidos 460-552 son únicos para el cromosoma. La presencia de secuencias únicas del donante unidas covalentemente a secuencias presentes en el cromosoma pero no en el donante indica que el ADN del plásmido del donante se introdujo en el cromosoma mediante recombinación homóloga.

**Ejemplo 3: Edición de un gen de la  $\beta$ -globina cromosómica mediante recombinación dirigida**

**[0241]** El gen de la beta globina humana es uno de los dos productos génicos responsables de la estructura y función de la hemoglobina en los eritrocitos humanos adultos. Las mutaciones en el gen de la beta-globina pueden provocar anemia falciforme. Se diseñaron dos proteínas de dedo de zinc para unirse dentro de esta secuencia, cerca de la ubicación de un nucleótido que, cuando muta, causa anemia de células falciformes. La Figura 13 muestra la secuencia de nucleótidos de una porción del gen de la beta-globina humana, y los sitios diana para las dos proteínas de dedo de zinc están subrayados en la secuencia presentada en la Figura 13. Secuencias de aminoácidos de las regiones de reconocimiento de las dos proteínas de dedo de zinc se muestran en la Tabla 6. Las secuencias que codifican cada uno de estos dos dominios de unión a ZFP se fusionaron con secuencias que codifican un medio dominio de escisión de *Fok* I, como se describió anteriormente, para crear nucleasas de ZFP diseñadas que se dirigieron al gen de la beta globina endógena. Cada una de estas secuencias de fusión se clonó luego en el vector de expresión de mamífero pcDNA3,1 (Figura 14).

**Tabla 6: Diseños de dedos de zinc para el gen beta-globina**

Secuencia diana	F1	F2	F3	F4
GGGCAGTAACGG (SEQ ID NO:68)	RSDHLS (SEQ ID NO:69)	QSANRTR (SEQ ID NO:70)	RSDNLSA (SEQ ID NO:71)	RSQNRTR (SEQ ID NO:72)
AAGGTGAACGTG (SEQ ID NO:73)	RSDLSR (SEQ ID NO:74)	DSSNRKT (SEQ ID NO:75)	RSDLSA (SEQ ID NO:76)	RNDNRKT (SEQ ID NO:77)
Nota: Las secuencias de aminoácidos del dedo de zinc que se muestran arriba (en código de una letra) representan los residuos -1 a +6, con respecto al comienzo de la porción alfa-helicoidal de cada dedo de zinc. El dedo F1 está más cerca del extremo amino de la proteína, y el dedo F4 está más cerca del extremo carboxilo.				

[0242] Se obtuvo una molécula de ADN donante como sigue. Primero, un fragmento de 700 pares de bases de ADN genómico humano correspondiente a los nucleótidos 5212134 - 5212833 en la cadena "-" del cromosoma 11 (BLAT, sitio del genoma humano UCSC) se amplificó por PCR, usando ADN genómico de células K562 como plantilla. Las secuencias de los cebadores utilizados para la amplificación se muestran en la Tabla 7 (etiquetada amp inicial 1 y amp inicial 2). El fragmento amplificado resultante contiene secuencias correspondientes al promotor, los dos primeros exones y el primer intrón del gen de la beta globina humana. Consulte la Figura 15 para ver un esquema que ilustra las ubicaciones de los exones 1 y 2, el primer intrón y los sitios de unión del cebador en la secuencia de beta globina. El producto clonado se modificó luego por PCR para introducir un conjunto de cambios de secuencia entre los nucleótidos 305-336 (como se muestra en la Figura 13), que reemplazó la secuencia CCGTTACTGCCCTGTGGGGCAAGGTGAACGTG (SEQ ID NO:78) con **gCGTTAgTGCCCGAATTCGAtcGTcAACcac** (SEQ ID NO:79) (cambios en negrita). Algunos de estos cambios (mostrados en minúsculas) fueron diseñados específicamente para evitar que las proteínas de fusión ZFP/Fok I se unan y escinden la secuencia del donante, una vez integradas en el cromosoma. Además, todos los cambios de secuencia permiten la discriminación entre las secuencias cromosómicas endógenas y de donantes después de la recombinación. El fragmento de 700 pares de bases resultante se clonó en pCR4-TOPO, que no contiene ninguna secuencia homóloga al genoma humano (Figura 16).

[0243] Para la mutación dirigida del gen de la globina beta cromosómica, los dos plásmidos que codifican las fusiones de ZFP-Fok I y el plásmido donador (pCR4-TOPO-HBBdonante) se introdujeron en células 1 X 10<sup>6</sup> K562 mediante transfección usando solución Nucleofector™ (Amaxa Biosystems). Los controles incluyeron células transfectadas solo con 100 ng (bajo) o 200 ng (alto) de los dos plásmidos que codifican las fusiones ZFP-Fok I, células transfectadas solo con 200 ng (bajo) o 600 ng (alto) del plásmido donante, células transfectadas con un plásmido que codifica GFP y simulando células transfectadas. Las células se cultivaron en medio RPMI 1640 (Invitrogen), suplementado con suero bovino fetal al 10% (FBS) (Hyclone) y L-glutamina 2 mM. Las células se mantuvieron a 37°C en una atmósfera de 5% de CO<sub>2</sub>. A las 72 horas después de la transfección, se aisló el ADN genómico de las células y se usaron 200 ng como plantilla para la amplificación por PCR, usando un cebador complementario a una región del gen fuera de su región de homología con las secuencias del donante (nucleótidos 5212883-5212905 en la cadena "-" del cromosoma 11), y un segundo cebador complementario a una región de la molécula donante en donde se introdujeron mutaciones distintivas en la secuencia donante (ver arriba). Las secuencias de estos cebadores se dan en la Tabla 7 (etiquetadas como "específicas de cromosomas" y "específicas al donante", respectivamente). Usando estos dos cebadores, se obtendrá un producto de amplificación de 415 pares de bases del ADN genómico si se ha producido un evento de recombinación dirigido. Como control para la carga de ADN, las reacciones de PCR también se llevaron a cabo utilizando los cebadores Initial amp 1 e Initial amp 2 para garantizar que se agregaran niveles similares de ADN genómico a cada reacción de PCR. Las condiciones para la amplificación fueron: 95°C, 2 min, seguido de 40 ciclos de 95°C, 30 segundos, 60°C, 45 segundos, 68°C, 2 minutos; y un paso final de 68°C, 10 min.

[0244] Los resultados de este análisis (Figura 17) indican que un producto de amplificación de par 415 de base se obtuvo solamente con el ADN extraído a partir de células que habían sido transfectadas con la concentración "alta" de plásmido donante y los dos plásmidos ZFP-Fok I, en concordancia con recombinación dirigida de secuencias de donantes en el locus cromosómico de beta-globina.

**Tabla 7: Cebadores de amplificación para el gen de la beta globina humana**

<b>Amplificador inicial 1</b>	TACTGATGGTATGGGGCCAAGAG (SEQ ID NO:80)
<b>Amplificador inicial 2</b>	CACGTGCAGCTTGTACAGTGC (SEQ ID NO:81)
<b>Específico al cromosoma</b>	TGCTTACCAAGCTGTGATTCCA (SEQ ID NO:82)
<b>Específico al donante</b>	GGTTGACGATCGGAATTC (SEQ ID NO:83)

[0245] Para confirmar este resultado, el producto de amplificación se clonó en pCR4-TOPO (Invitrogen) y se determinó su secuencia de nucleótidos. Como se muestra en la Figura 18 (SEQ ID NO:14), la secuencia consiste en una fusión entre secuencias cromosómicas que no están presentes en el plásmido donador y secuencias únicas del plásmido donador. Por ejemplo, se observan dos mutaciones C → G que interrumpen la unión de ZFP en las posiciones 377 y 383 en la secuencia. Los nucleótidos 377-408 representan la secuencia obtenida del plásmido donante que contiene los cambios de secuencia descritos anteriormente; nucleótidos 73-376 son secuencias comunes al donante y al cromosoma, y los nucleótidos 1-72 son únicos para el cromosoma. El enlace covalente de las secuencias específicas al donante y específicas al cromosoma en el genoma confirma la recombinación exitosa de la secuencia del donante en el locus correcto dentro del genoma de las células K562.

**Ejemplo 4: Optimización del enlazador ZFP-FokI (enlazador ZC)**

[0246] Para probar el efecto de la longitud del enlazador ZC sobre la eficacia de escisión, se fusionó un dominio de unión de ZFP de cuatro dedos a un medio dominio de escisión Fok I, usando enlazadores ZC de varias longitudes. El sitio diana para la ZFP es 5'-AACTCGGATAAT-3' (SEQ ID NO:84) y las secuencias de aminoácidos de las regiones de reconocimiento (posiciones -1 a +6 con respecto al comienzo de la hélice alfa) de cada uno de los dedos de zinc

fueron los siguientes (en donde F1 es más N y F4 es el dedo de zinc más C):

F1: DRSTLIE (SEQ ID NO:85)

5 F2: SSSNLSR (SEQ ID NO:86)

F3: RSDDLK (SEQ ID NO:87)

10 F4: DNSNRIK (SEQ ID NO:88)

**[0247]** Fusiones *ZFP-Fok I*, en donde el mencionado dominio de unión *ZFP* y un medio dominio de escisión *Fok I* fueron separados por 2, 3, 4, 5, 6 o 10 residuos de aminoácidos, se construyeron. Cada una de estas proteínas se probó para la escisión de sustratos que tienen una repetición invertida del sitio diana de *ZFP*, con repeticiones separadas por 4, 5, 6, 7, 8, 9, 12, 15, 16, 17, 22 o 26 pares de bases.

15 **[0248]** Las secuencias de aminoácidos de las construcciones de fusión, en la región de la unión *ZFP-Fok I* (con la secuencia del enlazador *ZC* subrayada), son las siguientes:

20

Enlazador de 10 residuos HTKIHLRQKDAARGSQLV (SEQ ID NO:89)

Enlazador de 6 residuos HTKIHLRQKGSSQLV (SEQ ID NO:90)

25 Enlazador de 5 residuos HTKIHLRQGSSQLV (SEQ ID NO:91)

Enlazador de 4 residuos HTKIHLRGSSQLV (SEQ ID NO:92)

Enlazador de 3 residuos HTKIHLGSQLV (SEQ ID NO: 93)

Enlazador de 2 residuos HTKIHGSQLV (SEQ ID NO:90)

30

**[0249]** Las secuencias de los diversos sustratos de escisión, con el objetivo ZFP Los sitios subrayados son los siguientes:

5	Separación 4bp	CTAGCATTATCCGAGTTACACA <u>ACTCGGATAATGCTAG</u> GATCGTAATAGGCTCAATGTGTTGAGCCTATTACGATC (SEQ ID NO:95)
10	Separación 5bp	CTAGCATTATCCGAGTTACACA <u>ACTCGGATAATGCTAG</u> GATCGTAATAGGCTCAAGTGTGTTGAGCCTATTACGATC (SEQ ID NO:96)
15	Separación 6bp	CTAGCATTATCCGAGTTACCACA <u>ACTCGGATAATGACTAG</u> GATCCGTAATAGGCTCAAGTGGTGTGAGCCTATTACTGATC (SEQ ID NO:97)
20	Separación 7bp	CTAGCATTATCCGAGTTACACACA <u>ACTCGGATAATGCTAG</u> GATCGTAATAGGCTCAAGTGTGTTGAGCCTATTACGATC (SEQ ID NO:98)
25	Separación 8bp	CTAGCATTATCCGAGTTACCACACA <u>ACTCGGATAATGCTAG</u> GATCGTAATAGGCTCAAGTGGTGTGTTGAGCCTATTACGATC (SEQ ID NO:99)
30	Separación 9bp	CTAGCATTATCCGAGTTACACACACA <u>ACTCGGATAATGCTAG</u> GATCGTAATAGGCTCAAGTGTGTTGAGCCTATTACGATC (SEQ ID NO:100)
35	Separación 12bp	CTAGCATTATCCGAGTTACCACCAACACA <u>ACTCGGATAATGCTAG</u> GATCGTAATAGGCTCAAGTGGTGGTGTGTTGAGCCTATTACGATC (SEQ ID NO:101)
40	Separación 15bp	CTAGCATTATCCGAGTTACCACCAACCACACA <u>ACTCGGATAATGCTAG</u> GATCGTAATAGGCTCAAGTGGTGGTGTGTTGAGCCTATTACGATC (SEQ ID NO:102)
	Separación 16bp	CTAGCATTATCCGAGTTACCACCAACCACCA <u>ACTCGGATAATGCTAG</u> GATCGTAATAGGCTCAAGTGGTGGTGTGTTGAGCCTATTACGATC (SEQ ID NO:103)
	Separación 17bp	CTAGCATTATCCGAGTTCAACCACCAACCACCA <u>ACTCGGATAATGCTAG</u> GATCGTAATAGGCTCAAGTGGTGGTGTGTTGAGCCTATTACGATC (SEQ ID NO:104)
	Separación 22bp	CTAGCATTATCCGAGTTCAACCACCAACCACCAACA <u>ACTCGGATAATGCTAG</u> GATCGTAATAGGCTCAAGTGGTGGTGTGTTGAGCCTATTACGATC (SEQ ID NO:105)
	Separación 26bp	CTAGCATTATCCGAGTTCAACCACCAACCACCAACCACCA <u>ACTCGGATAATGCTAG</u> GATCGTAATAGGCTCAAGTGGTGGTGTGTTGAGCCTATTACGATC (SEQ ID NO:106)

45 **[0250]** Los plásmidos que codifican las diferentes proteínas de fusión ZFP-*Fok I* (véase más arriba) fueron construidos por técnicas de biología molecular estandar y un sistema de transcripción/traducción acoplado *in vitro* se utilizaron para expresar las proteínas codificadas. Para cada construcción, se incubaron 200 ng de ADN plasmídico linealizado en 20 µl de mezcla de TnT y se incubaron a 30°C durante 1 hora y 45 minutos. La mezcla de TnT contiene 100 µl de lisado TnT (Promega, Madison, WI) con 4 µl de polimerasa ARN T7 (Promega) + 2 µl de metionina (1 mM) + 2,5 µl de ZnCl<sub>2</sub> (20 mM).

50 **[0251]** Para el análisis de la escisión del ADN por las diferentes fusiones de ZFP-*Fok I*, se combinó 1 ul de la mezcla de reacción de transcripción/traducción acoplada con aproximadamente 1 ng de sustrato de ADN (marcado en el extremo con <sup>32</sup>P usando polinucleótido quinasa T4), y la mezcla se diluyó a un volumen final de 19 µl con tampón de escisión *Fok I*. El tampón de escisión *Fok I* contiene Tris-HCl 20 mM, pH 8,5, NaCl 75 mM, ZnCl<sub>2</sub> 10 µM, DTT 1 mM, glicerol al 5%, BSA 500 µg/ml. La mezcla se incubó durante 1 hora a 37°C. Luego se añadieron 6,5 µl de tampón *Fok I*, que también contenía MgCl<sub>2</sub> 8 mM, y se continuó la incubación durante una hora a 37°C. La proteína se extrajo añadiendo 10 µl de solución de fenol-cloroformo a cada reacción, mezcla y centrifugación para separar las fases. Diez microlitros de la fase acuosa de cada reacción se analizaron por electroforesis en un gel de poliacrilamida al 10%.

60 **[0252]** El gel se sometió a autorradiografía, y la eficiencia de escisión para cada par de fusión/sustrato ZFP-*Fok I* se calculó cuantificando la radioactividad en las bandas correspondientes al sustrato escindido y sin escindir, sumando para obtener la radiactividad total y determinando el porcentaje de la radiactividad total presente en las bandas que representan los productos de escisión.

65 **[0253]** Los resultados de este experimento se muestran en la Tabla 8. Estos datos permiten la selección de un

enlazador ZC que proporciona una eficacia de escisión óptima para una separación de sitio diana dada. Estos datos también permiten la selección de longitudes de enlace que permiten la escisión en un par seleccionado de sitios diana, pero discriminan contra la escisión en los mismos sitios diana ZFP o similares que tienen una separación que es diferente de la del sitio de escisión previsto.

**Tabla 8: Eficacia de escisión de ADN para diversas longitudes de enlazador ZC y diversas separaciones sitio de unión\***

		2 residuos	3 residuos	4 residuos	5 residuos	6 residuos	10 residuos
10	4 pb	74%	81%	74%	12%	6%	4%
	5 pb	61%	89%	92%	80%	53%	40%
	6 pb	78%	89%	95%	91%	93%	76%
	7 pb	15%	55%	80%	80%	70%	80%
	8 pb	0%	0%	8%	11%	22%	63%
15	9 pb	2%	6%	23%	9%	13%	51%
	12 pb	8%	12%	22%	40%	69%	84%
	15 pb	73%	78%	97%	92%	95%	88%
	16 pb	59%	89%	100%	97%	90%	86%
	17 pb	5%	22%	77%	71%	85%	82%
20	22 pb	1%	3%	5%	8%	18%	58%
	26 pb	1%	2%	35%	36%	84%	78%

\*Las columnas representan diferentes construcciones de fusión ZFP-Fok I con el número indicado de residuos que separan el medio dominio de escisión ZFP y Fok I. Las filas representan diferentes sustratos de ADN con el número indicado de pares de bases que separan las repeticiones invertidas del sitio diana de ZFP.

[0254] Para las fusiones ZFP-Fok I con cuatro enlazadores de residuos, también se varió la secuencia de aminoácidos del enlazador. En construcciones separadas, la secuencia de enlace LRGS original (SEQ ID NO:107) se cambió a LGGS (SEQ ID NO:108), TGGS (SEQ ID NO:109), GGGS (SEQ ID NO:110), LPGS (SEQ ID NO:111), LRKS (SEQ ID NO:112) y LRWS (SEQ ID NO:113); y las fusiones resultantes se probaron en sustratos que tienen una separación de seis pares de bases entre los sitios de unión. Se observó que las fusiones que contenían la secuencia enlazadora LGGS (SEQ ID NO:108) se escindieron de manera más eficiente que las que contenían la secuencia LRGS original (SEQ ID NO:107). las fusiones que contienen las secuencias LRKS (SEQ ID NO:112) y LRWS (SEQ ID NO:113) se cortaron con menos eficiencia que la secuencia LRGS (SEQ ID NO:107), mientras que las eficiencias de escisión de las fusiones restantes fueron similares a las de la fusión comprende la secuencia LRGS original (SEQ ID NO:107).

**Ejemplo 5: Aumento de la especificidad de escisión resultante de la alteración del medio dominio de escisión de Fok I en la interfaz de dimerización**

[0255] Se diseñó un par de proteínas de fusión ZFP/Fok I (indicadas 5-8 y 5-10) para unirse a los sitios diana en el quinto exón del gen IL-2R $\gamma$ , para promover la escisión en la región entre los sitios diana. La región relevante del gen, incluidas las secuencias diana de las dos proteínas de fusión, se muestra en la Figura 19. La secuencia de aminoácidos de la proteína 5-8 se muestra en la Figura 20, y la secuencia de aminoácidos de la proteína 5-10 se muestra en la Figura 21. Ambas proteínas contienen un enlazador ZC de 10 aminoácidos. Con respecto a la porción de dedo de zinc de estas proteínas, las secuencias diana de ADN, así como las secuencias de aminoácidos de las regiones de reconocimiento en los dedos de zinc, se dan en la Tabla 9.

**Tabla 9: Diseños de dedos de zinc para IL2R $\gamma$  génica**

Fusión	Secuencia diana	F1	F2	F3	F4
5-8	ACTCTGTGGAAG (SEQ ID NO:114)	RSDNLSE (SEQ ID NO:115)	ARNHRIN (SEQ ID NO:116)	RSDDLSE (SEQ ID NO:117)	ARSTRIT (SEQ ID NO:118)
5-10	AACACGaAACGTG (SEQ ID NO:119)	RSDLSLR (SEQ ID NO:120)	DSSNRKT (SEQ ID NO:121)	RSDLSLV (SEQ ID NO:122)	DRSNRIT (SEQ ID NO:123)

Nota: Las secuencias de aminoácidos del dedo de zinc que se muestran arriba (en código de una letra) representan los residuos -1 a +6, con respecto al comienzo de la porción alfa-helicoidal de cada dedo de zinc. El dedo F1 está más cerca del extremo amino de la proteína.

[0256] La capacidad de este par de proteínas de fusión para catalizar la escisión específica de ADN entre sus secuencias diana (ver Figura 19) se probó *in vitro* usando una plantilla de ADN marcada que contiene la secuencia diana y analizando la presencia de productos de diagnóstico de digestión. La escisión específica se obtuvo cuando se usaron ambas proteínas (Tabla 10, primera fila). Sin embargo, la proteína de fusión 5-10 (que comprende un medio dominio de escisión Fok I de tipo silvestre) también fue capaz de escisión aberrante en un sitio no diana en ausencia de la proteína 5-8 (Tabla 10, segunda fila), posiblemente debido a la autodimerización.

[0257] En consecuencia, 5-10 se modificó en su medio dominio de escisión *Fok I* convirtiendo el residuo de aminoácido 490 de ácido glutámico (E) en lisina (K). (La numeración de los residuos de aminoácidos en la proteína *Fok I* está de acuerdo con Wah *et al.*, *Supra*). Esta modificación fue diseñada para prevenir la homodimerización al alterar un residuo de aminoácido en la interfaz de dimerización. El mutante 5-10 (E490K), a diferencia de la proteína 5-10 parental, no pudo escindirse en sitios aberrantes en ausencia de la proteína de fusión 5-8 (Tabla 10, Fila 3). Sin embargo, el mutante 5-10 (E490K), junto con la proteína 5-8, catalizó la escisión específica del sustrato (Tabla 10, Fila 4). Por lo tanto, la alteración de un residuo en el medio dominio de escisión de 5-10, que está implicado en la dimerización, evitó la escisión aberrante por esta proteína de fusión debido a la auto-dimerización. Un mutante E490R también exhibe niveles más bajos de homodimerización que la proteína original.

[0258] Además, la proteína 5-8 se modificó en su interfaz de dimerización reemplazando el residuo de glutamina (Q) en la posición 486 con ácido glutámico (E). Este mutante 5-8 (Q486E) se probó por su capacidad para catalizar la escisión dirigida en presencia de la proteína 5-10 de tipo silvestre o del mutante 5-10 (E490K). La escisión del ADN no se observó cuando el sustrato marcado se incubó en presencia de 5-8 (Q486E) y de tipo silvestre 5-10 (Tabla 10, Fila 5). Sin embargo, la escisión se obtuvo cuando los mutantes 5-8 (Q486E) y 5-10 (E490K) se usaron en combinación (Tabla 10, Fila 6).

[0259] Estos resultados indican que la escisión del ADN por un par de proteínas de fusión ZFP/*Fok I*, en regiones distintas de la definida por las secuencias diana de las dos proteínas de fusión, puede minimizarse o abolirse alterando la secuencia de aminoácidos del medio dominio de la escisión en una o ambas proteínas de fusión.

**Tabla 10: Escisión de ADN por pares de proteínas de fusión ZFP/*Fok I* que contienen medios dominios de escisión de tipo silvestre y mutante**

	<b>Dominio de unión de ZFP 5-8</b>	<b>Dominio de unión de ZFP 5-10</b>	<b>Escisión de ADN</b>
1	<i>Fok I</i> de tipo silvestre	<i>Fok I</i> de tipo silvestre	Específica
2	No presente	<i>Fok I</i> de tipo silvestre	No específica
3	No presente	<i>Fok I</i> E490K	Ninguna
4	<i>Fok I</i> de tipo silvestre	<i>Fok I</i> E490K	Específica
5	<i>Fok I</i> Q486E	<i>Fok I</i> de tipo silvestre	Ninguna
6	<i>Fok I</i> Q486E	<i>Fok I</i> E490K	Específica

Nota: Cada fila de la tabla presenta los resultados de un experimento separado en donde se analizaron las proteínas de fusión ZFP/*Fok I* para determinar la escisión de un sustrato de ADN marcado. Una de las proteínas de fusión contenía el dominio de unión al ADN 5-8, y la otra proteína de fusión contenía el dominio de unión al ADN 5-10 (ver Tabla 9 y Figura 19). La porción de medio dominio de escisión de las proteínas de fusión fue como se indica en la Tabla. Por lo tanto, las entradas en la columna ZFP 5-8 indican el tipo de dominio de escisión *Fok I* fusionado a ZFP 5-8; y las entradas en la columna ZFP 5-10 indican el tipo de dominio de escisión *Fok I* fusionado a ZFP 5-10. Para los mutantes de medio dominio de escisión de *Fok I*, el número se refiere al residuo de aminoácido en la proteína *Fok I*; la letra que precede al número se refiere al aminoácido presente en la proteína de tipo silvestre y la letra que sigue al número indica el aminoácido al que se cambió el residuo de tipo silvestre al generar la proteína modificada. "No presente" indica que la proteína de fusión ZFP/*Fok I* completa se omitió de ese experimento en particular. El sustrato de ADN utilizado en este experimento fue un producto de PCR de aproximadamente 400 pb que contenía los sitios diana tanto para ZFP 5-8 como para ZFP 5-10. Consulte la Figura 19 para ver las secuencias y la orientación relativa de los dos sitios diana.

**Ejemplo 6: Generación de un gen defectuoso de la proteína verde fluorescente mejorada (eGFP)**

[0260] La proteína verde fluorescente mejorada (eGFP) es una forma modificada de la proteína verde fluorescente (GFP); véase, *p. ej.*, Tsien (1998) *Ann. Rev. Biochem.* 67: 509-544) que contiene cambios en el aminoácido 64 (phe a leu) y 65 (ser a thr). Heim y col. (1995) *Nature* 373: 663-664; Cormack y col. (1996) *Gene* 173: 33-38. Se construyó un sistema informador basado en eGFP generando una forma defectuosa del gen eGFP, que contenía un codón de parada y una mutación de desplazamiento de marco de 2 pb. La secuencia del gen eGFP se muestra en la Figura 22. Las mutaciones se insertaron superponiendo la mutagénesis por PCR, utilizando el kit Platinum® *Taq* DNA Polymerase High Fidelity (Invitrogen) y los oligonucleótidos GFPBam, GFP-Xba, parada sentido2 y parada anti2 como cebadores (las secuencias de oligonucleótidos se enumeran a continuación en la Tabla 11). GFPBam y GFP-Xba sirvieron como cebadores externos, mientras que los cebadores parada sentido2 y parada anti2 como cebadores internos que codifican los cambios de nucleótidos. El vector peGFP-NI (BD Biosciences), que codifica un gen eGFP de longitud completa, se usó como plantilla de ADN en dos reacciones de amplificación separadas, la primera utilizando los oligonucleótidos GFP-Bam y parada anti2 como cebadores y la segunda utilizando GFP-Xba y oligonucleótidos parada sentido2 como cebadores. Esto generó dos productos de amplificación cuyas secuencias se superponen. Estos productos se combinaron y usaron como plantilla en una tercera reacción de amplificación, usando los oligonucleótidos externos GFP-Bam y GFP-Xba como cebadores, para regenerar un gen eGFP modificado en donde la secuencia GACCACAT (SEQ ID NO:124) en los nucleótidos 280-287 se reemplazó con la secuencia

TAACAC (SEQ ID NO:125). Las condiciones de PCR para todas las reacciones de amplificación fueron las siguientes: la plantilla se desnaturizó inicialmente durante 2 minutos a 94 grados y seguido por 25 ciclos de amplificación incubando la reacción durante 30 seg. a 94 grados C, 45 seg. a 46 grados C y 60 seg. a 68 grados C. Se realizó una ronda final de extensión a 68 grados C durante 10 minutos. La secuencia del producto de amplificación final se muestra en la Figura 23. Este fragmento de 795 pb se clonó en el vector pCR(R)4-TOPO utilizando el kit de clonación TOPO-TA (Invitrogen) para generar la construcción pCR(R)4-TOPO-GFPmut.

**Tabla 11: Secuencias de oligonucleótidos para la secuencia de GFP**

Oligo	Secuencia 5'-3'
GFP-Bam	CGAATTCTGCAGTCGAC (SEQ ID NO:126)
GFP-Xba	GATTATGATCTAGAGTCG (SEQ ID NO:127)
Parada sentido2	AGCCGCTACCCCTAACACGAAGCAG (SEQ ID NO:128)
Parada anti2	CTGCTTCGTGTTAGGGGTAGCGGCT (SEQ ID NO:129)

### Ejemplo 7: Diseño y ensamblaje de Nucleasas de Dedo de Zinc dirigidas a eGFP

**[0261]** Se diseñaron dos ZFP de tres dedos para unir una región del gen GFP mutado (Ejemplo 6) correspondiente a los nucleótidos 271-294 (numeración según la Figura 23). Los sitios de unión para estas proteínas ocurren en orientación opuesta con 6 pares de bases que separan los dos sitios de unión. Ver Figura 23. ZFP 287A une los nucleótidos 271-279 en la cadena no codificante, mientras que ZFP 296 une los nucleótidos 286-294 en la cadena codificante. La diana de ADN y la secuencia de aminoácidos para las regiones de reconocimiento de las ZFP se enumeran a continuación, y en la Tabla 12:

287A:

F1 (GCGg) RSDDLTR (SEQ ID NO:130)  
 F2 (GTA) QSGALAR (SEQ ID NO:131)  
 F3 (GGG) RSDHLR (SEQ ID NO:132)

296S:

F1 (GCA) QSGSLTR (SEQ ID NO:133)  
 F2 (GCA) QSGDLTR (SEQ ID NO:134)  
 F3 (GAA) QSGNLAR (SEQ ID NO:135)

**Tabla 12: Diseños de dedos de zinc para el gen GFP**

Proteína	Secuencia diana	F1	F2	F3
287A	GGGGTAGCGg (SEQ ID NO:136)	RSDDLTR (SEQ ID NO:137)	QSGALAR (SEQ ID NO:138)	RSDHLR (SEQ ID NO:139)
296S	GAAGCAGCA (SEQ ID NO:140)	QSGSLTR (SEQ ID NO:141)	QSGDLTR (SEQ ID NO:142)	QSGNLAR (SEQ ID NO:143)
Nota: Las secuencias de aminoácidos del dedo de zinc que se muestran arriba (en código de una letra) representan los residuos -1 a +6, con respecto al comienzo de la porción alfa-helicoidal de cada dedo de zinc. El dedo F1 está más cerca del extremo amino de la proteína, y el dedo F3 está más cerca del extremo carboxilo.				

**[0262]** Las secuencias que codifican estas proteínas se generaron mediante ensamblaje de PCR (*p. ej.*, Patente de Estados Unidos N° 6,534,261), clonado entre los sitios *Kph* I y *Bam* HI del vector pcDNA3,1 (Invitrogen), y fusionado en marco con el dominio catalítico de la endonucleasa *Fok* I (aminoácidos 384-579 de la secuencia de Looney et al. (1989) Gene 80: 193-208). Las construcciones resultantes se denominaron pcDNA3,1-GFP287-FokI y pcDNA3,1-GFP296-FokI (Figura 24).

### Ejemplo 8: Escisión de ADN *in vitro* dirigida por nucleasas de dedo de zinc diseñadas

**[0263]** La construcción pCR(R)4-TOPO-GFPmut (Ejemplo 6) se usó para proporcionar una plantilla para probar la capacidad de las proteínas de dedo de zinc 287 y 296 para reconocer específicamente sus sitios diana y escindir esta forma modificada de eGFP *in vitro*.

**[0264]** Se obtuvo un fragmento de ADN que contenía el inserto codificador de eGFP defectuoso por amplificación por PCR, usando los cebadores universales T7 y T3 y pCR(R)4-TOPO-GFPmut como plantilla. Este fragmento se marcó en el extremo usando polinucleótido quinasa  $\gamma$ -<sup>32</sup>P-ATP y T4. El nucleótido no incorporado se eliminó usando una

columna microspin G-50 (Amersham).

**[0265]** Se usó un sistema de transcripción/traducción acoplado *in vitro* para expresar las nucleasas de dedos de zinc 287 y 296 descritas en el Ejemplo 7. Para cada construcción, se incubaron 200 ng de ADN plasmídico linealizado en una mezcla de TnT de 20 ml y se incubaron a 30°C. por 1 hora y 45 minutos. La mezcla de TnT contiene 100 µl de lisado TnT (que incluye polimerasa ARN de T7, Promega, Madison, WI), suplementado con 2 µl de metionina (1 mM) y 2,5 µl de ZnCl<sub>2</sub> (20 mM).

**[0266]** Para el análisis de la escisión del ADN, se combinaron alícuotas de cada una de las mezclas de reacción de transcripción/traducción acopladas 287 y 296, luego se diluyeron en serie con tampón de escisión. La escisión tampón contiene 20 mM Tris-HCl pH 8,5, 75 mM NaCl, 10 mM MgCl<sub>2</sub>, 10 µM ZnCl<sub>2</sub>, DTT 1 mM, glicerol al 5%, 500 µg/ml de BSA. Se combinaron 5 ml de cada dilución con aproximadamente 1 ng de sustrato de ADN (marcado en el extremo con <sup>32</sup>P usando polinucleótido quinasa T4 como se describe anteriormente), y cada mezcla se diluyó adicionalmente para generar una reacción de escisión de 20 ml con la siguiente composición: 20 mM TrisHCl pH 8,5, 75 mM NaCl, 10 mM MgCl<sub>2</sub>, 10 µM ZnCl<sub>2</sub>, 1 mM DTT, glicerol al 5%, 500 µg/ml de BSA. Las reacciones de escisión se incubaron durante 1 hora a 37°C. La proteína se extrajo agregando 10 µl de solución de fenol-cloroformo a cada reacción, mezclando y centrifugando para separar las fases. Diez microlitros de la fase acuosa de cada reacción se analizaron por electroforesis en un gel de poliacrilamida al 10%.

**[0267]** El gel se sometió a autorradiografía, y los resultados de este experimento se muestran en la Figura 25. Los cuatro carriles de la izquierda muestran los resultados de las reacciones en las que la dilución final de cada mezcla de reacción de transcripción/traducción acoplada (en la escisión reacción) fue 1/156,25, 1/31,25, 1/12,5 y 1/5, respectivamente, resultando en volúmenes efectivos de 0,032, 0,16, 0,4, y 1 ul, respectivamente, de cada reacción de transcripción/traducción acoplada. La aparición de dos fragmentos de ADN que tienen pesos moleculares más bajos que el fragmento inicial (carril marcado como "control sin cortar" en la Figura 25) se correlaciona con cantidades crecientes de las endonucleasas de dedo de zinc 287 y 296 en la mezcla de reacción, lo que muestra que se obtuvo la escisión del ADN en el sitio diana esperado.

#### **Ejemplo 9: Generación de líneas celulares estables que contienen un gen eGFP defectuoso integrado**

**[0268]** Un fragmento de ADN que codifica el eGFP mutado, eGFPmut, se escindió del vector pCR(R)4-TOPO-GFPmut (Ejemplo 6) y se clonó en los sitios *Hind* III y *Not* I de pcDNA4/TO, colocando así este gen bajo control de un promotor CMV inducible por tetraciclina. El plásmido resultante se denominó pcDNA4/TO/GFPmut (Figura 26). Se cultivaron células T-Rex 293 (Invitrogen) en medio Eagle modificado por Dulbecco (DMEM) (Invitrogen) suplementado con suero bovino fetal (FBS) 10% Tetfree (HyClone). Las células se colocaron en una placa de 6 pocillos con una confluencia del 50%, y cada uno de los dos pocillos se transfeció con pcDNA4/TO/GFPmut. Se permitió que las células se recuperaran durante 48 horas, luego se combinaron las células de ambos pocillos y se dividieron en 10 X 15 cm<sup>2</sup> placas en medio selectivo, es decir, medio suplementado con 400 µg/ml de Zeocin (Invitrogen). El medio se cambió cada 3 días, y después de 10 días se aislaron colonias individuales y se expandieron aún más. Cada línea clonal se probó individualmente para la expresión inducible por doxiciclina (dox) del gen eGFPmut por RT-PCR cuantitativa (TaqMan®).

**[0269]** Para el análisis cuantitativo de RT-PCR, se aisló el ARN total de las células tratadas y no tratadas con dox usando el Kit de aislamiento de alta pureza (Roche Molecular Biochemicals), y 25 ng de ARN total de cada muestra se sometió a RT cuantitativa en tiempo real. PCR para analizar la expresión génica endógena, utilizando ensayos TaqMan®. Las secuencias de sonda y cebador se muestran en la Tabla 13. Las reacciones se llevaron a cabo en una máquina ABI 7700 SDS (PerkinElmer Life Sciences) en las siguientes condiciones. La reacción de transcripción inversa se realizó a 48°C durante 30 minutos con transcriptasa inversa MultiScribe (PerkinElmer Life Sciences), seguido de un paso de desnaturalización de 10 minutos a 95°C. La reacción en cadena de la polimerasa (PCR) se realizó con la ADN polimerasa AmpliGold (PerkinElmer Life Sciences) durante 40 ciclos a 95°C durante 15 segundos y 60°C durante 1 minuto. Los resultados se analizaron utilizando el software SDS versión 1,7 y se muestran en la Figura 27, con la expresión del gen eGFPmut normalizada a la expresión del gen GAPDH humano. Varias líneas celulares exhibieron expresión dependiente de doxiciclina de eGFP; la línea 18 (T18) se eligió como una línea celular modelo para estudios posteriores.

Tabla 13: Oligonucleótidos para el análisis de ARNm

Oligonucleótido	Secuencia
Cebador eGFP 1 (5T)	CTGCTGCCCGACAACCA (SEQ ID NO:144)
Cebador eGFP 2 (3T)	CCATGTGATCGCGCTTCTC (SEQ ID NO:145)
Sonda eGFP	CCCAGTCCGCCCTGAGCAAAGA (SEQ ID NO:146)
Cebador GAPDH 1	CCATGTTTCGTCATGGGTGTGA (SEQ ID NO:147)
Cebador GAPDH 2	CATGGACTGTGGTCATGAGT (SEQ ID NO:148)
Sonda GAPDH	TCCTGCACCACCAACTGCTTAGCA (SEQ ID NO:149)

#### 10 Ejemplo 10: Generación de una secuencia de donante para la corrección de un gen eGFP cromosómico defectuoso

15 [0270] Se construyó una construcción de donante que contiene la información genética para corregir el gen eGFPmut defectuoso por PCR. La reacción de PCR se llevó a cabo como se describió anteriormente, usando el vector peGFP-NI como plantilla. Para prevenir la expresión de fondo del constructo donante en experimentos de recombinación dirigida, los primeros 12 pb y el codón de inicio se eliminaron del donante por PCR usando los cebadores GFPnstart y GFP-Xba (secuencias proporcionadas en la Tabla 14). El fragmento de PCR resultante (734 pb) se clonó en el vector pCR(R)4-TOPO, que no contiene un promotor de células de mamífero, mediante clonación TOPO-TA para crear pCR(R)4-TOPO-GFPdonante5 (Figura 28). La secuencia del inserto eGFP de esta construcción (correspondiente a los nucleótidos 64-797 de la secuencia que se muestra en la Figura 22) se muestra en la Figura 29 (SEQ ID NO:20).

Tabla 14: oligonucleótidos para la construcción de la molécula donante

Oligonucleótido	Secuencia 5'-3'
GFPnstart	GGCGAGGAGCTGTTAC (SEQ ID NO:150)
GFP-Xba	GATTATGATCTAGAGTCG (SEQ ID NO:151)

#### 30 Ejemplo 11: Corrección de una mutación en un gen eGFP cromosómico integrado por escisión y recombinación dirigida

35 [0271] La línea celular estable T18 (Ejemplo 9) se transfectó con uno o ambos del plásmido de expresión ZFP-Fok I (pcDNA3,1-GFP287 -FokI y pcDNA3,1-GFP296-FokI, Ejemplo 7) y 300 ng del plásmido donante pCR(R)4-TOPOGFPdonante5 (Ejemplo 10) usando el Reactivo LipofectAMINE 2000 (Invitrogen) en medio sérico reducido Opti-MEM I, según el protocolo del fabricante. La expresión del gen eGFP cromosómico defectuoso se indujo 5-6 horas después de la transfección mediante la adición de 2 ng/ml de doxiciclina al medio de cultivo. Las células fueron detenidas en la fase G2 del ciclo celular mediante la adición, a las 24 horas después de la transfección, de 100 ng/ml de Nocodazol (Figura 30) o Vinblastina 0,2  $\mu$ M (Figura 31). Se permitió que el arresto G2 continuara durante 24-48 horas, y luego se liberó mediante la eliminación del medio. Las células se lavaron con PBS y el medio se reemplazó con DMEM que contenía FBS libre de tetraciclina y doxiciclina 2 ng/ml. Se permitió que las células se recuperaran durante 24-48 horas, y se midió la eficacia de corrección génica controlando el número de células que exhibían fluorescencia de eGFP, mediante análisis de clasificación de células activadas por fluorescencia (FACS). El análisis FACS se realizó utilizando un instrumento Beckman-Coulter EPICS XL-MCL y el software System II Data Acquisition and Display, versión 2.0. La fluorescencia de eGFP se detectó por excitación a 488 nm con un láser de argón y monitoreo de emisiones a 525 nm (eje x). El fondo o la autofluorescencia se midieron monitoreando las emisiones a 570 nm (eje y). Las células que exhiben alta emisión fluorescente a 525 nm y baja emisión a 570 nm (región E) se puntuaron como positivas para la corrección génica.

50 [0272] Los resultados se resumen en la Tabla 15 y las Figuras 30 y 31. Las Figuras 30 y 31 muestran los resultados en donde T18 se transfectaron células con el pcDNA3,1-GFP287-FokI y plásmidos pcDNA3,1-GFP296-FokI que codifica nucleasas de ZFP y el plásmido pCR(R)4-TOPO-GFPdonante5, la expresión de eGFP se indujo con doxiciclina y las células se detuvieron en G2 con nocodazol (Figura 30) o vinblastina (Figura 31). Ambas figuras muestran trazas de FACS, en las que las células que exhiben fluorescencia de eGFP están representadas en la parte inferior derecha de la traza (identificada como Región E, que es la parte de la Tabla 4 debajo de la curva). Para las células transfectadas que habían sido tratadas con nocodazol, el 5,35% de las células exhibían fluorescencia GFP, indicativa de corrección del gen cromosómico mutante eGFP (Figura 30), mientras que el 6,7% de las células tratadas con vinblastina se sometieron a corrección del gen eGFP (Figura 31). Estos resultados se resumen, junto con experimentos de control adicionales, en las filas 1-8 de la tabla 15.

60 [0273] En resumen, estos experimentos muestran que, en presencia de dos nucleasas ZFP y una secuencia donante, aproximadamente 1% de las células tratadas se sometieron a la corrección de genes, y que este nivel de corrección se aumentó 4-5 veces por la detención de las células tratadas en la fase G2 del ciclo celular.

65

Tabla 15: Corrección de un gen eGFP cromosómico defectuoso

Expte.	Tratamiento <sup>1</sup>	Porcentaje de células con gen eGFP corregido <sup>2</sup>
1	300 ng solo donante	0,01
2	100 ng ZFP 287 + 300 ng donante	0,16
3	100 ng ZFP 296 + 300 ng donante	0,6
4	50 ng ZFP 287 + 50 ng ZFP 296 + 300 ng donante	1,2
5	como 4 + 100 ng/ml de nocodazol	5,35
6	como 4 + 0,2 uM de vinblastina	6,7
7	sin donante, sin ZFP, 100 ng/ml de nocodazol	0,01
8	sin donante, sin ZFP, 0,2 uM de vinblastina	0.0
9	100 ng ZFP287/Q486E + 300 ng donante	0.0
10	100 ng ZFP296/E490K + 300 ng donante	0,01
11	50 ng 287/Q486E + 50 ng 296/E490K + 300 ng donante	0,62
12	como 11 + 100 ng/ml de nocodazol	2,37
13	como 11 + 0,2 uM de vinblastina	2,56
<b>Notas:</b>		
1: Las células T18, que contienen un gen eGFP cromosómico defectuoso, se transfectaron con plásmidos que codifican una o dos nucleasas ZFP y/o un plásmido donante que codifica una secuencia de eGFP no defectuosa, y se indujo la expresión del gen eGFP cromosómico con doxiciclina. Las células fueron detenidas opcionalmente en la fase G2 del ciclo celular después de la inducción de eGFP. El análisis FACS se realizó 5 días después de la transfección.		
2: El número es el porcentaje de fluorescencia total que exhibe alta emisión a 525 nm y baja emisión a 570 nm (región E de la traza FACS).		

#### Ejemplo 12: Corrección de un gen cromosómico defectuoso utilizando nucleasas de dedo de zinc con alteraciones de secuencia en la interfaz de dimerización

**[0274]** Nucleasas de dedo de zinc cuyas secuencias se habían alterado en la interfaz de dimerización se probó por su capacidad para catalizar la corrección de un gen eGFP cromosómico defectuoso. Se usó el protocolo descrito en el Ejemplo 11, excepto que la porción de nucleasa de las nucleasas ZFP (es decir, los medios dominios de escisión de *Fok I*) se alteraron como se describe en el Ejemplo 5. Por lo tanto, un medio dominio de escisión E490K se fusionó con el dominio GFP296 ZFP (Tabla 12) y un medio dominio de escisión Q486E se fusionaron con el ZFP GFP287 (Tabla 12).

**[0275]** Los resultados se muestran en las filas 9-11 de la tabla 15 e indican que se obtuvo un aumento significativo en la frecuencia de corrección génica en presencia de dos nucleasas de ZFP que tienen alteraciones en sus interfaces de dimerización, en comparación con el obtenido en presencia de cualquiera de las nucleasas solas. Experimentos adicionales, en los que las células T18 se transfectaron con plásmido donante y plásmidos que codifican las nucleasas de dedo de zinc 287/Q486E y 296/E490K, luego se detuvieron en G2 con nocodazol o vinblastina, mostraron un aumento adicional en la frecuencia de corrección génica, con más del 2% de células que exhiben fluorescencia de eGFP, indicativo de un gen eGFP cromosómico corregido (Tabla 15, Filas 12 y 13).

#### Ejemplo 13: Efecto de la longitud del donante sobre la frecuencia de corrección génica

**[0276]** En un experimento similar a los descritos en el Ejemplo 11, se probó el efecto de la longitud de la secuencia del donante sobre la frecuencia de la recombinación dirigida. Las células T18 se transfectaron con las dos nucleasas ZFP, y la expresión de eGFP se indujo con doxiciclina, como en el Ejemplo 11. Las células también se transfectaron con el plásmido pCR(R)4-TOPOGFPdonante5 (Figura 28) que contiene un inserto eGFP de 734 pb (Figura 29) como en el Ejemplo 11, o un plásmido similar que contiene un inserto de secuencia de 1527 pb (Figura 32) homólogo al gen eGFP cromosómico mutado. Además, se evaluó el efecto de la detención de G2 con nocodazol en la frecuencia de recombinación.

**[0277]** En un segundo experimento, se compararon longitudes de donante de 0,7, 1,08 y 1,5 kbp. Las células T18 se transfectaron con 50 ng de los plásmidos de expresión 287-FokI y 296-FokI (Ejemplo 7, Tabla 12) y 500 ng de donantes de 0,7 kbp, 1,08 kbp o 1,5 kbp, como se describe en el Ejemplo 11. Cuatro días después de la transfección, las células fueron analizadas para la corrección del gen defectuoso eGFP por FACS, monitoreando la fluorescencia de GFP.

**[0278]** Los resultados de estos dos experimentos, mostrados en la Tabla 16, muestran que la secuencia donante ya aumenta la frecuencia de recombinación dirigida (y, por lo tanto, de la corrección de genes) y confirmar que la detención de las células en la fase G2 del ciclo celular también aumenta la frecuencia de recombinación dirigida.

**Tabla 16: Efecto de la duración del donante y la detención del ciclo celular sobre la frecuencia de recombinación dirigida**

Longitud del donante (kb)	Experimento 1 Concentración de nocodazol:		Experimento 2 -
	0 ng/ml	100 ng/ml	
0,7	1,41	5,84	1,2
1,08	no hecho	no hecho	2,2
1,5	2,16	8,38	2,3

**Nota:** Los números representan el porcentaje de fluorescencia total en la Región E de la traza FACS (ver Ejemplo 11), que es una indicación de la fracción de células que se han sometido a recombinación dirigida para corregir el gen eGFP cromosómico defectuoso.

**Ejemplo 14: Edición del gen endógeno humano IL-2R $\gamma$  mediante escisión dirigida y recombinación usando nucleasas de dedos de zinc**

**[0279]** Dos vectores de expresión, cada uno de los cuales codifica una ZFP-nucleasa dirigida al gen humano IL-2R $\gamma$ , se construyeron. Cada nucleasa de ZFP contenía un dominio de unión a ADN basado en proteína de dedo de zinc (ver Tabla 17) fusionado al dominio de nucleasa de la enzima de restricción tipo IIS *Fok I* (aminoácidos 384-579 de la secuencia de Wah et al. (1998) Proc Natl. Acad. Sci. EE.UU. 95: 10564-10569) a través de un enlazador ZC de cuatro aminoácidos (véase el Ejemplo 4). Las nucleasas se diseñaron para unirse a las posiciones en el exón 5 del gen cromosómico IL-2R $\gamma$  que rodea los codones 228 y 229 (un punto de acceso mutacional en el gen) e introducir una ruptura de doble cadena en el ADN entre sus sitios de unión.

**Tabla 17: Diseños de dedo de zinc para el exón 5 de la IL2R $\gamma$  génica**

Secuencia diana	F1	F2	F3	F4
ACTCTGTGGAAG (SEQ ID NO:152) 5-8G	RSDNLSV (SEQ ID NO:153)	ARNHRIN (SEQ ID NO:154)	RSDLTSE (SEQ ID NO:155)	ARSTRTN (SEQ ID NO:156)
AAAGCGGCTCCG (SEQ ID NO:157) 5-9D	RSDLTSE (SEQ ID NO:158)	ARSTRTT (SEQ ID NO:159)	RSDLSK (SEQ ID NO:160)	QRSNLKV (SEQ ID NO:161)

Nota: Las secuencias de aminoácidos del dedo de zinc que se muestran arriba (en código de una letra) representan los residuos -1 a +6, con respecto al comienzo de la porción alfa-helicoidal de cada dedo de zinc. El dedo F1 está más cerca del extremo amino de la proteína.

**[0280]** La porción completa de unión al ADN de cada una de las endonucleasas químicas fue la siguiente:

nucleasa dirigida a ACTCTGTGGAAG (SEQ ID NO:152)

MAERPFQCRICMRNFSRSDNLSVHIRTHTGKPFACDICGRKFARNAHRIN  
HTKIHTGSQKPFQCRICMRNFSRSDLSEHIRTHTGKPFACDICGRKFAARSTR  
NHTKIHLRGS (SEQ ID NO:162)

nucleasa dirigida a AAAGCGGCTCCG (SEQ ID NO:157)

MAERPFQCRICMRNFSRSDLSEHIRTHTGKPFACDICGRKFAARSTRTHHTKIH  
TGSQKPFQCRICMRNFSRSDLSKHIRTHTGKPFACDICGRKFAQRSNLKVHTKI  
HLRGS (SEQ ID NO:163)

**[0281]** Las células 293 de riñón embrionario humano fueron transfectadas (Lipofectamine 2000; Invitrogen) con dos construcciones de expresión, cada una de las cuales codifica una de las nucleasas ZFP descritas en el párrafo anterior. Las células también se transfectaron con una construcción de donante que llevaba como inserto un fragmento de 1.543 pb del locus IL2R $\gamma$  correspondiente a las posiciones 69195166-69196708 de la cadena "menos" del cromosoma X (liberación del genoma humano UCSC julio de 2003), en el vector pCR4Blunt Topo (Invitrogen). La secuencia de inserción de IL-2R $\gamma$  contenía las siguientes dos mutaciones puntuales en la secuencia del exón 5 (subrayado):

**FRVRSRFNPLCGS** (SEQ ID NO:164)  
TTTCGTGTTCCGAGCCGGTTTAACCCGCTCTGTGGAAGT (SEQ ID NO:165)

**[0282]** La primera mutación (CGC  $\rightarrow$  CGG) no cambia la secuencia de aminoácidos (línea superior) y sirve para afectar negativamente la capacidad de la nucleasa de ZFP para unirse al ADN del donante y al ADN cromosómico después de la recombinación. La segunda mutación (CCA  $\rightarrow$  CCG) no cambia la secuencia de aminoácidos y crea un sitio de

reconocimiento para la enzima de restricción *BsrBI*.

[0283] Se usaron 50 o 100 nanogramos de cada construcción de expresión de nucleasa de ZFP y 0,5 o 1 microgramo de la construcción de donante en transfecciones duplicadas. También se realizaron los siguientes experimentos de control: transfección con un plásmido de expresión que codifica la proteína eGFP; transfección con construcción de donante solamente; y transfección con plásmidos que expresan solo las nucleasas de ZFP. Veinticuatro horas después de la transfección, se añadió vinblastina (Sigma) a una concentración final de 0,2  $\mu\text{m}$  a una muestra en cada conjunto de duplicados, mientras que la otra permaneció sin tratamiento. La vinblastina afecta la capacidad de la célula para ensamblar el huso mitótico y, por lo tanto, actúa como un potente agente de detención de  $G_2$ . Este tratamiento se realizó para mejorar la frecuencia de la focalización porque la vía de reparación de rotura bicatenaria dirigida por homología es más activa que la unión final no homóloga en la fase  $G_2$  del ciclo celular. Después de un período de tratamiento de 48 h con vinblastina 0,2  $\mu\text{m}$ , se reemplazó el medio de crecimiento y se permitió que las células se recuperaran del tratamiento con vinblastina durante 24 horas adicionales. Luego se aisló el ADN genómico de todas las muestras de células usando el kit de tejido DNeasy (Qiagen). Cinco cien nanogramos de ADN genómico de cada muestra se ensayaron entonces para la frecuencia de modificación génica dirigida, mediante el ensayo para la presencia de un nuevo sitio *BsrBI* en el locus cromosómico IL-2R $\gamma$ , utilizando el ensayo descrito esquemáticamente en la Figura 33.

[0284] En resumen, 20 ciclos de PCR se realizaron usando los cebadores mostrados en la Tabla 18, cada uno de los cuales hibrida con el locus cromosómico IL-2R $\gamma$  inmediatamente fuera de la región homóloga a la secuencia donante de 1,5 kb. Se incluyeron veinte microcuries de  $\alpha$ - $^{32}\text{P}$ -dCTP y  $\alpha$ - $^{32}\text{P}$ -dATP en cada reacción de PCR para permitir la detección de productos de PCR. Las reacciones de PCR se desalaron en una columna G-50 (Amersham) y se digirieron durante 1 hora con 10 unidades de *BsrBI* (New England Biolabs). Los productos de digestión se resolvieron en un gel de poliacrilamida no desnaturizante al 10% (BioRad), y el gel se secó y autorradiografió (Figura 34). Además del producto de PCR principal, correspondiente al fragmento amplificado de 1,55 kb del locus IL2R $\gamma$  ("wt" en la Figura 34), se observó una banda adicional ("rflp" en la Figura 34) en los carriles correspondientes a muestras de células que se transfectaron con la construcción de ADN donante y ambas construcciones de nucleasa de ZFP. Esta banda adicional no apareció en ninguno de los carriles de control, lo que indica que en este experimento se produjo la recombinación facilitada por nucleasa de ZFP de la secuencia del donante que contiene *BsrBI* RFLP en el cromosoma.

[0285] Experimentos adicionales, en los que se añadieron trazas de una secuencia de ADN de IL-2R $\gamma$  que contiene RFLP al ADN genómico humano (que contiene el gen de IL-2R $\gamma$  natural), y la mezcla resultante se amplificó y se sometió a digestión con una enzima de restricción que se escinde en la RFLP, han indicado que se puede detectar cuantitativamente tan solo 0,5% de secuencia que contiene RFLP usando este ensayo.

**Tabla 18: Oligonucleótidos para el análisis del gen humano IL-2R $\gamma$**

Oligonucleótido	Secuencia
Ex5_1,5detF1	GATTCAACCAGACAGATAGAAGG (SEQ ID NO:166)
Ex5_1,5detR1	TTACTGTCTCATCCTTTACTCC (SEQ ID NO:167)

#### **Ejemplo 15: Recombinación dirigida en el locus de IL-2R $\gamma$ en células K562**

[0286] K562 es una línea celular derivada de una leucemia mielógena crónica humana. Las proteínas utilizadas para la escisión dirigida fueron fusiones *Fok I* a los dominios de unión a ADN de dedos de zinc 5-8G y 5-9D (Ejemplo 14, Tabla 17). La secuencia donante fue el fragmento de 1,5 kbp del gen de IL-2R $\gamma$  humano que contiene un sitio *BsrBI* introducido por mutación, descrito en el Ejemplo 14.

[0287] Las células K562 fueron cultivadas en medio RPMI 1640 (Invitrogen), complementado con 10% de suero bovino fetal (FBS) (Hyclone) y 2 mM L-glutamina. Todas las células se mantuvieron a 37°C en una atmósfera de 5% de CO<sub>2</sub>. Estas células fueron transfectadas por Nucleofection™ (Solución V, Programa T16) (Amaxa Biosystems), de acuerdo con el protocolo del fabricante, transfectando 2 millones de células por muestra. Los ADN para la transfección, utilizados en diversas combinaciones como se describe a continuación, fueron un plásmido que codifica la endonucleasa de fusión 5-8G ZFP-*Fok I*, un plásmido que codifica la endonucleasa de fusión 5-9D ZFP-*Fok I*, un plásmido que contiene la secuencia donante (descrita anteriormente y en el Ejemplo 14) y el vector peGFP-N1 (BD Biosciences) utilizado como control.

[0288] En el primer experimento, las células se transfectaron con diversos plásmidos o combinaciones de plásmidos como se muestra en la Tabla 19.

Tabla 19

Muestra N°	p-eGFP-N1	p5-8G	p5-9D	donante	vinblastina
1	5 µg	-	-	-	-
2	-	-	-	50 µg	-
3	-	-	-	50 µg	si
4	-	10 µg	10 µg	-	-
5	-	5 µg	5 µg	25 µg	-
6	-	5 µg	5 µg	25 µg	si
7	-	7,5 µg	7,5 µg	25 µg	-
8	-	7,5 µg	7,5 µg	25 µg	si
9	-	7,5 µg	7,5 µg	50 µg	-
10	-	7,5 µg	7,5 µg	50 µg	si

[0289] Las células tratadas con vinblastina se expusieron a vinblastina 0,2 µM a las 24 horas después de la transfección durante 30 horas. Las células se recogieron, se lavaron dos veces con PBS y se volvieron a sembrar en medio de crecimiento. Las células se cosecharon 4 días después de la transfección para el análisis del ADN genómico.

[0290] Se extrajo ADN genómico de las células usando el kit DNEasy (Qiagen). Se usaron cien nanogramos de ADN genómico de cada muestra en una reacción de PCR con los siguientes cebadores:

Exón 5 delantero: GCTAAGGCCAAGAAAGTAGGGCTAAAG (SEQ ID NO:168)

Exón 5 inverso: TTCCTTCATCACCAAACCTCTTG (SEQ ID NO:169)

[0291] Estos cebadores amplifican un fragmento de 1.669 pb del cromosoma X correspondiente a las posiciones 69195100-69196768 en la cadena "-" (liberación del genoma humano UCSC, julio de 2003) que contiene el exón 5 del gen IL2Ry. La amplificación del ADN genómico que ha sufrido una recombinación homóloga con el ADN del donante produce un producto que contiene un sitio de *Bsr* BI; mientras que el producto de amplificación de ADN genómico que no ha sufrido recombinación homóloga con ADN de donante no contendrá este sitio de restricción.

[0292] Diez microcuries cada uno de  $\alpha$ -<sup>32</sup>PdCTP y  $\alpha$ -<sup>32</sup>PdATP se incluyeron en cada reacción de amplificación para permitir la visualización de los productos de reacción. Después de 20 ciclos de PCR, la reacción se desaló en una columna Sephadex G-50 (Farmacia) y se digirió con 10 unidades de *Bsr* BI (New England Biolabs) durante 1 hora a 37°C. La reacción se resolvió luego en una PAGE no desnaturalizante al 10%, se secó y se expuso a una pantalla PhosphorImager.

[0293] Los resultados de este experimento se muestran en la Figura 35. Cuando las células se transfectaron con el plásmido de GFP control, donante plásmido solo o los dos plásmidos que codifican ZFP en ausencia de donante, ningún sitio *Bsr*BI estaba presente en el producto de amplificación, como se indica por la ausencia de la banda marcada "rflp" en los carriles correspondientes a estas muestras en la Figura 35. Sin embargo, el ADN genómico de las células que se transfectaron con el plásmido donante y ambos plásmidos que codifican ZFP contenían el sitio de BI de *Bsr* introducido por recombinación homóloga con el ADN del donante (banda etiquetada "rflp"). La cuantificación del porcentaje de señal representada por el ADN que contiene RFLP, que se muestra en la Figura 35, indicó que, en condiciones óptimas, hasta el 18% de todos los genes de IL-2Ry en la población celular transfectada se alteraron por recombinación homóloga.

[0294] Un segundo experimento se llevó a cabo según el protocolo que se acaba de describir, excepto que se expandieron las células durante 10 días después de la transfección. Los ADN utilizados para la transfección se muestran en la Tabla 20.

Tabla 20

Muestra N°	p-eGFP-N1	p5-8G	p5-9D	donante	vinblastina
1	50 µg	-	-	-	-
2	-	-	-	50 µg	-
3	-	-	-	50 µg	si
4	-	7,5 µg	7,5 µg	-	-
5	-	5 µg	5 µg	25 µg	-
6	-	5 µg	5 µg	25 µg	si
7	-	7,5 µg	7,5 µg	50 µg	-
8	-	7,5 µg	7,5 µg	50 µg	si

[0295] El análisis de la digestión con *Bsr* BI del ADN amplificado, que se muestra en la Figura 36, demostró nuevamente que hasta el 18% de los genes de IL-2R $\gamma$  habían sufrido alteración de secuencia por recombinación homóloga, después de múltiples rondas de división celular. Por lo tanto, los eventos de recombinación dirigidos son estables.

[0296] Además, el ADN de las células transfectadas en este segundo experimento se analizó mediante transferencia Southern. Para este análisis, doce microgramos de ADN genómico de cada muestra se digirieron con 100 unidades de *Eco* RI, 50 unidades de *Bsr* BI, y 40 unidades de *Dpn* I (todos de New England Biolabs) durante 12 horas a 37°C. Esta digestión genera un fragmento *Eco* RI de 7,7 kbp del gen IL-2R $\gamma$  nativo (que carece de un sitio *Bsr*BI) y fragmentos de 6,7 y 1,0 kbp de un gen IL-2R $\gamma$  cromosómico cuya secuencia ha sido alterada, por recombinación homóloga, para incluir el sitio de *Bsr*BI. *Dpn* I, una enzima de restricción dependiente de la metilación, se incluyó para destruir el ADN del donante metilado en la presa. ADN genómico de la célula no metilada K562 es resistente a la digestión *Dpn* I.

[0297] Después de la digestión, el ADN genómico se purificó por extracción con fenol-cloroformo y precipitación con etanol, se resuspendió en tampón TE y se resolvió en un gel de agarosa al 0,8% junto con una muestra de ADN genómico digerido con *Eco*RI y *Sph*I para generar un marcador de tamaño. El gel se procesó para transferencia alcalina siguiendo el procedimiento estándar y el ADN se transfirió a una membrana de nylon (Schleicher y Schuell). Luego se realizó la hibridación a la transferencia usando un fragmento radiomarcado del locus de IL-2R $\gamma$  correspondiente a las posiciones 69198428-69198769 de la cadena "-" del cromosoma X (liberación de julio de 2003 del genoma humano UCSC). Esta región del gen está fuera de la región homóloga al ADN del donante. Después de la hibridación, la membrana se expuso a una placa PhosphorImager y los datos se cuantificaron utilizando el software Molecular Dynamics. La alteración de la secuencia cromosómica de IL-2R $\gamma$  se midió analizando la intensidad de la banda correspondiente al fragmento *Eco* RI-*Bsr* BI (flecha al lado de la autorradiografía; sitio *Bsr*BI indicado por el triángulo lleno en el mapa sobre la autorradiografía).

[0298] Los resultados, mostrados en la Figura 37, indican hasta un 15% de secuencias cromosómicas IL-2R $\gamma$  fueron alteradas por recombinación homóloga, lo que confirma los resultados obtenidos por análisis PCR de que el evento de recombinación específica era estable a través de múltiples rondas de división celular. Los resultados de la transferencia Southern también indican que los resultados que se muestran en la Figura 36 no son el resultado de un artefacto de amplificación.

#### **Ejemplo 16: Recombinación dirigida en el locus de IL-2R $\gamma$ en células madre hematopoyéticas positivas para CD34**

[0299] Las enfermedades genéticas (*p. ej.*, Inmunodeficiencia combinada severa (SCID) y anemia de células falciformes) pueden tratarse mediante corrección homóloga mediada por recombinación de la alteración específica de la secuencia de ADN responsable de la enfermedad. En ciertos casos, la eficacia y la estabilidad máximas del tratamiento resultarían de la corrección del defecto genético en una célula pluripotente. Con este fin, este ejemplo demuestra la alteración de la secuencia del gen IL-2R $\gamma$  en células de médula ósea positivas para CD34 humanas. Las células CD34<sup>+</sup> son células madre hematopoyéticas pluripotenciales que dan origen a los linajes eritroides, mieloides y linfoides.

[0300] Las células CD34 humanas derivadas de médula ósea se compraron de AllCells, LLC y se enviaron como reservas congeladas. Estas células se descongelaron y se dejaron reposar durante 2 horas a 37°C en una atmósfera de 5% de CO<sub>2</sub> en Medio RPMI 1640 (Invitrogen), complementado con 10% de suero bovino fetal (FBS) (Hyclone) y 2 mM L-glutamina. Las muestras celulares (1x10<sup>6</sup> o 2x10<sup>6</sup> células) se transfectaron por Nucleofection™ (Amaxa Biosystems) utilizando la célula CD34 humano Nucleofector™ Kit, según el protocolo de los fabricantes. Después de la transfección, las células se cultivaron en medio RPMI 1640 (Invitrogen), suplementado con FBS al 10%, L-glutamina 2 mM, factor estimulante de colonias de granulocitos (G-CSF) 100 ng/ml, factor de células madre (SCF) 100 ng/ml, 100 ng/ml de trombopoyetina (TPO), 50 ng/ml de ligando Flt3 y 20 ng/ml de interleucina-6 (IL-6). El inhibidor de caspasa zVAD-FMK (Sigma-Aldrich) se añadió a una concentración final de 40  $\mu$ M en el medio de crecimiento inmediatamente después de la transfección para bloquear la apoptosis. Se añadió inhibidor de caspasa adicional 48 horas más tarde a una concentración final de 20  $\mu$ M para prevenir adicionalmente la apoptosis. Estas células se mantuvieron a 37°C en una atmósfera de 5% de CO<sub>2</sub> y se cosecharon 3 días después de la transfección.

[0301] Los números de células y los ADN utilizados para la transfección se muestran en la Tabla 21.

Tabla 21

Muestra	Nº células	p-eGFP-N1 1	Donante 2	p5-8G 3	p5-9D 3
1	1x10 <sup>6</sup>	5 ug	-	-	-
2	2x10 <sup>6</sup>	-	50 ug	-	-
3	2x10 <sup>6</sup>	-	50 ug	7,5 ug	7,5 ug

1. Este es un plásmido de control que codifica una proteína fluorescente verde mejorada.  
2. El ADN donante es un fragmento de 1,5 kbp que contiene secuencias del exón 5 del gen IL-2R $\gamma$  con un sitio de *Bsr* BI introducido (véase el Ejemplo 14). 3. Estos son plásmidos que codifican las fusiones de *Fok* I con los dominios de unión a ADN de zinc con dedos 5-8 G y 5-9D (ver Tabla 17).

**[0302]** Se extrajo ADN genómico de las células usando el kit de purificación de ADN MasterPure (epicentro). Debido a la presencia de glucógeno en el precipitado, la cuantificación precisa de este ADN utilizado como entrada en la reacción de PCR es imposible; las estimaciones que utilizan análisis de geles de agarosa teñidos con bromuro de etidio indican que se usó ca. 50 ng de ADN genómico en cada muestra. Treinta ciclos de PCR fueron luego realizados utilizando los siguientes cebadores, cada uno de los cuales hibrida con el locus cromosómico IL-2R $\gamma$  inmediatamente fuera de la región homóloga al donante 1,5 kb:

ex5\_1,5detF3      GCTAAGGCCAAGAAAGTAGGGCTAAAG (SEQ ID NO:170)  
ex5\_1,5detR3      TTCCTTCCATCACCAAACCCTCTTG (SEQ ID NO:171)

**[0303]** Veinte microcuries cada uno de  $\alpha$ -<sup>32</sup>PdCTP y  $\alpha$ -<sup>32</sup>PdATP se incluyeron en cada reacción de PCR para permitir la detección de productos de PCR. Para proporcionar una referencia de cuantificación en gel, se explotó la existencia de un SNP espontáneo en el exón 5 del gen IL-2R $\gamma$  en las células Jurkat: este SNP crea un RFLP al destruir un sitio MaeII que está presente en el ADN humano normal. Por lo tanto, se creó un estándar de referencia agregando 1 o 10 nanogramos de ADN genómico humano normal (obtenido de Clontech, Palo Alto, CA) a 100 o 90 ng de ADN genómico Jurkat, respectivamente, y realizando la PCR como se describió anteriormente. Las reacciones de PCR se desalaron en una columna G-50 (Amersham) y se digirieron durante 1 hora con enzima de restricción: las muestras experimentales se digirieron con 10 unidades de *Bsr* BI (New England Biolabs); Las reacciones del "patrón de referencia" se digirieron con *Mae* II. Los productos de digestión se resolvieron en una PAGE al 10% no desnaturizante (BioRad), el gel se secó y se analizó por exposición a una placa PhosphorImager (Molecular Dynamics).

**[0304]** Los resultados se muestran en la Figura 38. Además del producto de PCR principal, correspondiente al fragmento de 1,6 kb del locus IL2R $\gamma$  ("wt" en el panel de la derecha de la Figura 38), una banda adicional (etiquetada "rflp") se observó en los carriles correspondientes a muestras de células que se transflectaron con plásmidos que codifican tanto las nucleasas ZFP como la construcción de ADN del donante. Esta banda adicional no apareció en los carriles de control, lo que concuerda con la idea de que la orientación del gen asistido por nucleasa de ZFP del exón 5 del gen de la cadena gamma común ocurrió en este experimento.

**[0305]** Aunque la cuantificación precisa de la velocidad de focalización se complica por la proximidad de la banda RFLP a la banda de tipo silvestre; La frecuencia de focalización se estimó, en comparación con el estándar de referencia (panel izquierdo), entre 1-5%.

#### 45 Ejemplo 17: Efectos de la homología donante-diana

**[0306]** El efecto, sobre la frecuencia de la recombinación homóloga, del grado de homología entre el ADN del donante y la secuencia cromosómica con la que se recombina se examinó en la línea celular T18, descrita en el Ejemplo 9. Esta línea contiene un gen eGFP defectuoso integrado cromosómicamente, y el ADN del donante contiene cambios de secuencia, con respecto al gen cromosómico, que corrige el defecto.

**[0307]** En consecuencia, la secuencia del donante descrita en el Ejemplo 10 se modificó, mediante mutagénesis por PCR, para generar una serie de construcciones de donantes de ~700 pb con diferentes grados de no homología con la diana. Todos los donantes modificados contenían cambios de secuencia que corrigieron el defecto en el gen eGFP cromosómico y contenían mutaciones silenciosas adicionales (mutaciones de ADN que no cambian la secuencia de la proteína codificada) insertadas en la región de codificación que rodea el sitio de escisión. Estas mutaciones silenciosas pretendían evitar la unión y la escisión de la secuencia del donante por las fusiones del dominio de escisión con dedos de zinc, reduciendo así la competencia entre la diana cromosómica deseada y el plásmido donante para la unión por las nucleasas quiméricas. Además, después de la recombinación homóloga, la capacidad de las nucleasas quiméricas para unirse y volver a escindir las secuencias cromosómicas recién insertadas (y posiblemente estimular otra ronda de recombinación, o causar una unión final no homóloga u otras alteraciones impulsadas por rotura de doble cadena del genoma) se minimizaría.

**[0308]** Se probaron cuatro secuencias de donantes diferentes. El donante 1 contiene 8 desajustes con respecto a la secuencia diana de eGFP defectuosa cromosómica, el donante 2 tiene 10 desajustes, el donante 3 tiene 6 desajustes y el donante 5 tiene 4 desajustes. Tenga en cuenta que la secuencia del donante 5 es idéntica a la secuencia de eGFP

de tipo silvestre, pero contiene 4 desajustes con respecto a la secuencia de eGFP cromosómica defectuosa en la línea celular T18. La Tabla 22 proporciona la secuencia de cada donante entre los nucleótidos 201-242. Los nucleótidos que son divergentes de la secuencia del gen eGFP defectuoso integrado en el genoma de la línea celular T18 se muestran en negrita y subrayados. También se muestran las secuencias correspondientes del gen eGFP cromosómico defectuoso (GFP mut) y el gen eGFP normal (GFP wt).

Tabla 22

Donante	Secuencia	SEQ ID NO.
Donante 1	CTTCAGCCGCTATCC <b>AGACCACATGAAACAAC</b> ACGACTTCTT	172
Donante 2	CTTCAGCCG <b>G</b> TATCC <b>AGACCACATGAAACAAC</b> ATGACTTCTT	173
Donante 3	CTTCAGCCGCTACCC <b>AGACCACATGAAACA</b> CAGCAGCAGACTTCTT	174
Donante 5	CTTCAGCCGCTACCC <b>CGACCACATGAAGCAGCAGCAG</b> ACTTCTT	175
GFP mut	CTTCAGCCGCTACCCCTAACAC--GAAGCAGCAGCAGACTTCTT	176
GFP wt	CTTCAGCCGCTACCC <b>CGACCACATGAAGCAGCAGCAG</b> ACTTCTT	177

**[0309]** La línea celular T18 fue transfectada, como se describe en el Ejemplo 11, con 50 ng de las construcciones de expresión 287-Fok I y 296-Fok I (Ejemplo 7 y Tabla 12) y 500 ng de cada constructo de donantes. El análisis FACS se realizó como se describe en el Ejemplo 11.

**[0310]** Los resultados, mostrados en la Tabla 23, indican que un grado decreciente de desajuste entre la secuencia diana cromosómica y del donante (es decir, una mayor homología) da como resultado una frecuencia aumentada de recombinación homóloga según se evalúa mediante la restauración de la función GFP.

Tabla 23<sup>1</sup>

Donante	Nº de discrepancias	Porcentaje de células con eGFP corregido gen 2
Donante 2	10	0,45%
Donante 1	8	0,53%
Donante 3	6	0,89%
Donante 5	4	1,56%
Donante	Nº de discrepancias	Porcentaje de células con eGFP corregido gen 2
1: Las células T18, que contienen un gen eGFP cromosómico defectuoso, se transfectaron con plásmidos que codifican dos nucleasas ZFP y con plásmidos donadores que codifican una secuencia de eGFP no defectuosa que tiene diferentes números de desajustes de secuencia con la secuencia diana cromosómica. La expresión del gen cromosómico eGFP se indujo con doxiciclina y el análisis FACS se realizó 5 días después de la transfección.		
2: El número es el porcentaje de fluorescencia total que exhibe alta emisión a 525 nm y baja emisión a 570 nm (región E de la traza FACS).		

**[0311]** Los resultados anteriores muestran que los niveles de recombinación homóloga aumentan disminuyendo el grado de divergencia de la secuencia del donante diana. Sin desear limitarse a ninguna teoría en particular o proponer un mecanismo en particular, se observa que una mayor homología entre el donante y la diana podría facilitar la recombinación homóloga al aumentar la eficiencia mediante la cual la maquinaria de recombinación homóloga celular reconoce la molécula donante como una plantilla adecuada. Alternativamente, un aumento en la homología del donante con la diana también podría conducir a la escisión del donante por las nucleasas quiméricas de ZFP. Un donante escindido podría ayudar a facilitar la recombinación homóloga al aumentar la tasa de invasión de cadena o podría ayudar en el reconocimiento del extremo del donante escindido como un tramo homólogo de ADN durante la búsqueda de homología por la maquinaria de recombinación homóloga. Además, estas posibilidades no son mutuamente excluyentes.

#### Ejemplo 18: Preparación de ARNip

**[0312]** Para probar si la disminución de los niveles celulares de proteínas involucradas en la unión final no homóloga (NHEJ) facilita la recombinación homóloga dirigida, se realizó un experimento en donde los niveles de la proteína Ku70 se redujeron mediante la inhibición de ARNip. Las moléculas de ARNip dirigidas al gen Ku70 se generaron mediante la transcripción de ADNc de Ku70 seguido de escisión de la transcripción bicatenaria con enzima Dicer.

**[0313]** Brevemente, se usó un grupo de ADNc generado a partir de células 293 y U2OS en cinco reacciones de amplificación separadas, cada una usando un conjunto diferente de cebadores de amplificación específicos para el gen Ku70, para generar cinco grupos de fragmentos de ADNc (grupos AE), que varían en tamaño de 500-750 pb. Los fragmentos en cada uno de estos cinco grupos se volvieron a amplificar usando cebadores que contienen el elemento promotor de la polimerasa de ARN del bacteriófago T7, nuevamente usando un conjunto diferente de cebadores para cada grupo de ADNc. La generación de ADNc y las reacciones de PCR se realizaron utilizando el sistema de ADNc

Superscript Choice y Platinum Taq High Fidelity Polymerase (ambas de Invitrogen, Carlsbad, CA), de acuerdo con los protocolos y recomendaciones de los fabricantes.

[0314] Cada uno de los grupos de ADN amplificado se transcribió luego *in vitro* con la polimerasa de ARN del bacteriófago T7 para generar cinco grupos (AE) de ARN de doble cadena (ARNds), usando el kit de transcripción *in vitro* RNAMAXX (Stratagene, San Diego, CA) de acuerdo con las instrucciones del fabricante. Después de la precipitación con etanol, el ARN en cada uno de los grupos se resuspendió y se escindió *in vitro* usando la enzima Dicer recombinante (Stratagene, San Diego, CA) de acuerdo con las instrucciones del fabricante. Los productos de ARNip de 21-23 pb en cada uno de los cinco grupos se purificaron mediante un método de dos pasos, primero usando una columna Microspin G-25 (Amershan), seguido de una columna Microcon YM-100 (Amicon). Cada grupo de productos de ARNip se transfectó transitoriamente en la línea celular T7 usando Lipofectamone 2000®.

[0315] Se realizaron transferencias Western para analizar la efectividad relativa de los grupos de ARNip en la supresión de la expresión de Ku70 aproximadamente 3 días después de la transfección. Brevemente, las células se lisaron e interrumpieron usando tampón RIPA (Santa Cruz Biotechnology), y se homogeneizaron pasando los lisados a través de un triturador QIA (Qiagen, Valencia, CA). Los lisados clarificados se trataron luego con tampón de muestra SDS PAGE (con  $\beta$  mercaptoetanol utilizado como agente reductor) y se hirvieron durante 5 minutos. Las muestras se resolvieron luego en un gel NUPAGE con gradiente de 4-12% y se transfirieron a una membrana de PVDF. La parte superior de la transferencia se expuso a un anticuerpo anti-Ku70 (Santa Cruz sc-5309) y la parte inferior se expuso a un anticuerpo anti-TF IIB (Santa Cruz sc-225, usado como control de entrada). La mancha se expuso luego al anticuerpo secundario anti-ratón de cabra conjugado con peroxidasa de rábano picante y se procesó para la detección electroquimioluminiscente (ECL) usando un kit de Pierce Chemical Co. de acuerdo con las instrucciones del fabricante.

[0316] La Figura 39 muestra resultados representativos después de la transfección de dos de los grupos de ARNip (grupos D y E) en células T7. La transfección con 70 ng de ARNip E da como resultado una disminución significativa en los niveles de proteína Ku70 (Figura 39, carril 3).

**Ejemplo 19: Aumento de la frecuencia de recombinación homóloga por inhibición de la expresión de una proteína implicada en la unión de extremos no homólogos**

[0317] La reparación de una ruptura bicatenaria en el ADN genómico puede proceder a lo largo de dos rutas celulares diferentes; recombinación homóloga (HR) o unión final no homóloga (NHEJ). Ku70 es una proteína involucrada en NHEJ, que se une a los extremos libres del ADN como resultado de una ruptura bicatenaria en el ADN genómico. Para probar si la disminución de la concentración intracelular de una proteína involucrada en NHEJ aumenta la frecuencia de la RH, se usaron pequeños ARN interferentes (ARNip), preparados como se describe en el Ejemplo 18, para inhibir la expresión del ARNm de Ku70, disminuyendo así los niveles de proteína Ku70, en células cotransfectadas con ADN donante y con plásmidos que codifican nucleasas quiméricas.

[0318] Para estos experimentos, se usó la línea celular T7 (ver Ejemplo 9 y Figura 27). Estas células contienen un gen eGFP defectuoso integrado cromosómicamente, pero se ha observado que exhiben niveles más bajos de recombinación homóloga dirigida que la línea celular T18 usada en los Ejemplos 11-13.

[0319] Las células T7 se transfectaron, como se describe en el Ejemplo 11, con 70 o 140 ng de uno de los dos grupos de producto de corte dirigido a Ku70 (véase el Ejemplo 18). El análisis de transferencia de proteínas se realizó en extractos derivados de las células transfectadas para determinar si el tratamiento de las células con ARNip resultó en una disminución en los niveles de la proteína Ku70 (véase el ejemplo anterior). La Figura 39 muestra que los niveles de la proteína Ku70 se reducen en las células que habían sido tratadas con 70 ng de ARNip de la reserva E.

[0320] Se cotransfectaron muestras de células separadas en el mismo experimento con 70 o 140 ng de ARNip (grupo D o grupo E) junto con 50 ng de cada una de las construcciones de expresión 287-FokI y 296-FokI (Ejemplo 7 y Tabla 12) y 500 ng del donante GFP de 1,5 kbp (Ejemplo 13), para determinar si la disminución de los niveles de Ku70 aumentaba la frecuencia de la recombinación homóloga. El protocolo experimental se describe en la Tabla 24. La restauración de la actividad de eGFP, debido a la recombinación homóloga, se ensayó mediante análisis FACS como se describe en el Ejemplo 11.

**Tabla 24**

Expte. N°	Donante <sup>1</sup>	ZFNs <sup>2</sup>	ARNip <sup>3</sup>	% corrección <sup>4</sup>
1	500 ng	-	-	0,05
2	-	50 ng cada uno	-	0,01
3	500 ng	50 ng cada uno	-	0,79
4	500 ng	50 ng cada uno	70 ng piscina D	0,68
5	500 ng	50 ng cada uno	140 ng piscina D	0,59
6	500 ng	50 ng cada uno	70 ng grupo E	1,25

(Continuación)

Expte. N°	Donante <sup>1</sup>	ZFNs <sup>2</sup>	ARNip <sup>3</sup>	% corrección <sup>4</sup>
1	500 ng	-	-	0,05
2	-	50 ng cada uno	-	0,01
3	500 ng	50 ng cada uno	-	0,79
4	500 ng	50 ng cada uno	70 ng piscina D	0,68
5	500 ng	50 ng cada uno	140 ng piscina D	0,59
6	500 ng	50 ng cada uno	70 ng grupo E	1,25
7	500 ng	50 ng cada uno	140 ng grupo E	0,92

1. Un plásmido que contiene una secuencia de 1,5 kbp que codifica una proteína eGFP funcional que es homóloga al gen eGFP defectuoso integrado en el cromosoma  
2. Los plásmidos que codifican la proteína de dedo de zinc 287 y 296 dirigida a eGFP/endonucleasas de fusión *Fok I*  
3. Ver Ejemplo 18  
4. Porcentaje de fluorescencia total que exhibe alta emisión a 525 nm y emisión baja a 570 nm (región e de la FACS traza, véase el Ejemplo 11).

[0321] El porcentaje de corrección del gen eGFP defectuoso en las células T7 transfectadas (indicativas de la frecuencia de recombinación homóloga dirigida) se muestra en la columna más a la derecha de Tabla 24. La frecuencia más alta de recombinación dirigida se observa en el Experimento 6, en donde las células se transfectaron con ADN del donante, los plásmidos que codifican las dos nucleasas de fusión dirigidas a eGFP y 70 ng de ARNip reserva E. La referencia al Ejemplo 18 y la Figura 39 indica que 70 ng de ARNip reserva E deprimieron significativamente los niveles de proteína Ku70. Por lo tanto, los métodos que reducen los niveles celulares de proteínas involucradas en NHEJ pueden usarse como un medio para facilitar la recombinación homóloga.

#### Ejemplo 20: Nucleasas de fusión *Fok I* de dedo de zinc dirigidas al gen de la $\beta$ -globina humana

[0322] Se diseñaron varios dominios de unión al ADN del dedo de cinc de cuatro dedos, dirigidos al gen de la  $\beta$ -globina humana, y plásmidos que codifican cada dominio de dedo de zinc, fusionado a un medio dominio de escisión *Fok I*, se construyeron. Cada dominio de dedos de zinc contenía cuatro dedos de zinc y reconocía un sitio diana de 12 pb en la región del gen de la  $\beta$ -globina humana que codifica la mutación responsable de la anemia de células falciformes. Se evaluó la afinidad de unión de cada una de estas proteínas para su secuencia diana, y cuatro proteínas que exhiben fuerte unión (sca-r29b, sca-36a, sca-36b, y sca-36c) se utiliza para la construcción de endonucleasas de fusión *Fok I*.

[0323] Los sitios diana de los dominios de unión a ADN de ZFP, alineados con la secuencia del gen de globina  $\beta$  humana, se muestran a continuación. El codón de inicio de la traducción (ATG) está en negrita y subrayado, al igual que la sustitución de AT que causa anemia de células falciformes.

sca-36a	GAAGTCTGCCGT	(SEQ ID NO:178)
sca-36b	GAAGTctGCCGT	(SEQ ID NO:179)
sca-36c	GAAGTctGCCGT	(SEQ ID NO:180)
	CAAACAGACACC <b><u>ATG</u></b> GTGCATCTGACTCCTG <b><u>T</u></b> GGAGAAGTCTGCCGTTACTG	(SEQ ID NO:181)
	GTTTGTCTGTGGTACCACGTAGACTGAGGAC <b><u>A</u></b> CTCTTCAGACGGCAATGAC	
sca-r29b	ACGTAGaCTGAGG	(SEQ ID NO:182)

[0324] Las secuencias de aminoácidos de las regiones de reconocimiento de los dedos de zinc en estas cuatro proteínas se muestran en la Tabla 25. Las secuencias de aminoácidos completas de estos dominios de dedos de zinc se muestran en la Figura 40. El dominio sca-36a reconoce un sitio diana que tiene 12 nucleótidos contiguos (mostrados en mayúscula arriba), mientras que los otros tres dominios reconocen una secuencia de trece nucleótidos que consta de dos sitios diana de seis nucleótidos (mostrados en mayúscula) separados por un solo nucleótido (mostrado en minúsculas). Por consiguiente, los dominios sca-r29b, sca-36b y sca-36c contienen un enlazador entre dedos no canónico que tiene la secuencia de aminoácidos TGGGGSQKP (SEQ ID NO:183) entre el segundo y el tercero de sus cuatro dedos.

Tabla 25

ZFP	F1	F2	F3	F4
sca- r29b	QSGDLTR (SEQ ID NO:184)	TSANLSR (SEQ ID NO:185)	DRSALSR (SEQ ID NO:186)	QSGHLSR (SEQ ID NO:187)
sca- 36a	RSQTRKT (SEQ ID NO:188)	QKRNRTK (SEQ ID NO:189)	DRSALSR (SEQ ID NO:190)	QSGNLR (SEQ ID NO:191)
sca- 36b	TSGLSR (SEQ ID NO:192)	DRSDLSR (SEQ ID NO:193)	DRSALSR (SEQ ID NO:194)	QSGNLR (SEQ ID NO:195)
sca- 36c	TSSLSR (SEQ ID NO:196)	DRSDLSR (SEQ ID NO:197)	DRSALSR (SEQ ID NO:198)	QSGNLR (SEQ ID NO:199)

**Ejemplo 21: Escisión *in vitro* de una secuencia diana de ADN por endonucleasas de fusión ZFP/*Fok* I dirigidas a  $\beta$ -globina**

[0325] Proteínas de fusión que contienen un medio dominio de escisión *Fok* I y uno de los cuatro dominios de unión a ADN de ZFP descritos en el ejemplo anterior se probaron para determinar su capacidad para escindir ADN *in vitro* con la especificidad de secuencia predicha. Estos dominios ZFP se clonaron en el vector de expresión pcDNA3,1 a través de los sitios KpnI y BamHI y se fusionaron en el inframundo al dominio de escisión *Fok* I a través de un enlazador ZC de 4 aminoácidos, como se describió anteriormente. Se clonó un fragmento de ADN que contenía 700 pb del gen de la  $\beta$ -globina humana a partir del ADN genómico obtenido de las células K562. El aislamiento y la secuencia de este fragmento se describieron en el Ejemplo 3, *supra*.

[0326] Para producir endonucleasas de fusión (ZFN) para el ensayo *in vitro*, se incubaron plásmidos circulares que codifican fusiones *Fok* I a las proteínas scar29b, sca-36a, sca-36b y sca-36c en un sistema de transcripción/traducción *in vitro*. Véase el Ejemplo 4. Se agregaron un total de 2  $\mu$ l de la reacción TNT (2  $\mu$ l de una reacción única cuando se estaba analizando una proteína única o 1  $\mu$ l de cada reacción cuando se estaba analizando un par de proteínas) a 13  $\mu$ l de la mezcla de tampón de escisión y 3  $\mu$ l de sonda marcada (~1 ng/ $\mu$ l). La sonda se marcó en el extremo con  $^{32}$ P usando polinucleótido quinasa. Esta reacción se incubó durante 1 hora a temperatura ambiente para permitir la unión de los ZFN. La escisión se estimuló mediante la adición de 8  $\mu$ l de MgCl<sub>2</sub> 8 mM, diluido en tampón de escisión, a una concentración final de aproximadamente 2,5 mM. La reacción de escisión se incubó durante 1 hora a 37°C y se detuvo mediante la adición de 11  $\mu$ l de fenol/cloroformo. El ADN se aisló por extracción con fenol/cloroformo y se analizó por electroforesis en gel, como se describe en el Ejemplo 4. Como control, se analizaron 3  $\mu$ l de sonda en el gel para marcar la migración de ADN sin cortar (marcado con "U" en la figura 41).

[0327] Los resultados se muestran en la Figura 41. La incubación del ADN objetivo con cualquier fusión de dedo de zinc/*Fok* I no dio como resultado ningún cambio en el tamaño del ADN de plantilla. Sin embargo, la combinación de la nucleasa sca-r29b con cualquiera de las nucleasas sca-36b o sca-36c dio como resultado la escisión del ADN diana, como lo demuestra la presencia de dos fragmentos de ADN más cortos (los dos carriles más a la derecha de la Figura 41).

**Ejemplo 22: Endonucleasas de fusión ZFP/*Fok* I, dirigidas al gen de la  $\beta$ -globina, analizadas en un sistema indicador de GFP cromosómico**

[0328] Se sintetizó un fragmento de ADN que contenía la secuencia del gen de  $\beta$ -globina humana dirigida por los ZFN descritos en el Ejemplo 20 y se clonó en un sitio *Spe* I en un gen informador de eGFP, interrumpiendo así la expresión de eGFP. El fragmento contenía la siguiente secuencia, en donde el nucleótido responsable de la mutación de células falciformes está en negrita y subrayado):

CTAGACACCATGGTGCATCTGACTCCTG**T**GGAGAAGTCTGCCGTTACT  
GCCCTAG (SEQ ID NO:200)

[0329] Este gen eGFP alterado que contiene secuencias de  $\beta$ -globina insertadas se clonó en pcDNA4/TO (Invitrogen, Carlsbad, CA) usando los sitios *Hind* III y *Not* I, y el vector resultante se transfeció en células HEK293 TRex (Invitrogen). Los clones estables individuales se aislaron y crecieron, y los clones se ensayaron para determinar la recombinación homóloga dirigida mediante la transfección de cada una de las proteínas sca-36 (sca-36a, sca-36b, sca-36c) emparejadas con sca-29b (ver Ejemplo 20 y Tabla 25 para secuencias y sitios de unión de estas nucleasas quiméricas). Las células se transfecieron con 50 ng de plásmido que codifica cada uno de los ZFN y con 500 ng del donante de GFP de 1,5 kb (Ejemplo 13). Cinco días después de la transfección, las células se analizaron para determinar la recombinación homóloga en el locus eGFP defectuoso insertado. Inicialmente, las células fueron examinadas por microscopía de fluorescencia para la función eGFP. Las células que exhiben fluorescencia se analizaron cuantitativamente usando un ensayo FACS para la fluorescencia de eGFP, como se describe en el Ejemplo 11.

[0330] Los resultados mostraron que todas las líneas celulares transfectadas con sca-29b y sca-36a fueron negativas para la función eGFP, cuando se analizaron por microscopía de fluorescencia. Algunas de las líneas transfectadas con sca-29b emparejadas con sca-36b o sca-36c fueron positivas para la expresión de eGFP, cuando se analizaron por microscopía de fluorescencia, y por lo tanto se analizaron adicionalmente mediante análisis FACS. Los resultados del análisis FACS de dos de estas líneas se muestran en la Tabla 26, e indican que las nucleasas de dedo de zinc dirigidas a secuencias de  $\beta$ -globina son capaces de catalizar la escisión de ADN de doble cadena específica de secuencia para facilitar la recombinación homóloga en células vivas.

Tabla 26

Línea celular	ADN transfectado:				% corr. <sup>1</sup>
	sca-29b	sca-36a	sca-36b	sca-36c	
N° 20	+	+			0
	+		+		0,08
	+			+	0,07
N° 40	+	+			0
	+		+		0,18
	+			+	0,12

1. Porcentaje de fluorescencia total que exhibe alta emisión a 525 nm y baja emisión a 570 nm (región E de la traza FACS, ver Ejemplo 11).

#### Ejemplo 23: Efecto del nivel de transcripción en la recombinación homóloga dirigida

[0331] Dado que la transcripción de una secuencia de ADN cromosómico implica alteraciones en su estructura de cromatina (generalmente para hacer que las secuencias transcritas sean más accesibles), es posible que un gen transcrito activamente sea más sustrato favorable para la recombinación homóloga dirigida. Esta idea se probó utilizando la línea celular T18 (Ejemplo 9) que contiene secuencias cromosómicas que codifican un gen eGFP defectuoso cuya transcripción está bajo el control de un promotor inducible por doxiciclina.

[0332] Se transfectaron muestras separadas de células T18 con plásmidos que codifican las proteínas de fusión 287 y 296 de dedo de zinc/*Fok I* dirigidas a eGFP (Ejemplo 7) y una molécula de ADN donante de 1,5 kbp que contiene secuencias que corrigen el defecto en el gen eGFP cromosómico (Ejemplo 9). Cinco horas después de la transfección, las células transfectadas se trataron con diferentes concentraciones de doxiciclina, luego se midieron los niveles de ARNm de eGFP 48 horas después de la adición de doxiciclina. La fluorescencia de eGFP a 520 nm (indicativo de recombinación dirigida de la secuencia del donante en el cromosoma para reemplazar las secuencias de  $\beta$ -globina insertadas) se midió mediante FACS a los 4 días después de la transfección.

[0333] Los resultados se muestran en la Figura 42. Las barras indican niveles crecientes de ARNm de eGFP normalizados a ARNm de GAPDH (equivalente, a una primera aproximación, a la velocidad de transcripción del gen eGFP cromosómico defectuoso). El número sobre cada barra indica el porcentaje de células que exhiben fluorescencia de eGFP. Los resultados muestran que el aumento de la tasa de transcripción del gen diana se acompaña de frecuencias más altas de recombinación dirigida. Esto sugiere que la activación dirigida de la transcripción (como se describe, p. ej., en las Patentes de los Estados Unidos de propiedad conjunta 6,534,261 y 6,607,882) se puede usar, junto con la escisión de ADN dirigida, para estimular la recombinación homóloga dirigida en las células.

#### Ejemplo 24: Generación de una línea celular que contiene una mutación en el gen IL-2R $\gamma$

[0334] Las células K562 se transfectaron con plásmidos que codifican las nucleasas de dedos de zinc 5-8GL0 y 5-9DL0 (ZFNs) (véase el Ejemplo 14; Tabla 17) y con una construcción de donante *Dra I* de 1,5 kbp. El donante *Dra I* se compone de una secuencia con homología a la región que codifica el 5° exón del gen IL2R $\gamma$ , pero inserta una base extra entre las ZFN de unión a los sitios para crear un marco de lectura y generar un sitio *Dra I*.

[0335] 24 horas después de la transfección, las células se trataron con vinblastina 0,2  $\mu$ M (concentración final) durante 30 horas. Las células se lavaron tres veces con PBS y se volvieron a sembrar en medio. Se permitió que las células se recuperaran durante 3 días y se retiró una alícuota de células para realizar un ensayo RFLP basado en PCR, similar al descrito en el Ejemplo 14, probando la presencia de un sitio *Dra I*. Se determinó que la frecuencia de corrección genética dentro de la población era aproximadamente del 4%.

[0336] Se permitió que las células se recuperaran durante 2 días adicionales y 1600 células individuales se colocaron en 40x placas de 96 pocillos en 100  $\mu$ l de medio.

[0337] Las células se cultivan durante aproximadamente 3 semanas, y las células homocigotas para el fenotipo mutante *Dra I* se aíslan. Las células se prueban para la modificación del genoma (mediante la prueba de la presencia de un sitio *Dra I* en el exón 5 del gen IL-2R $\gamma$ ) y para los niveles de ARNm de IL-2R $\gamma$  (por PCR en tiempo real) y proteína

(por transferencia Western) para determinar el efecto de la mutación en la expresión génica. Las células se prueban para su función mediante análisis FACS.

5 **[0338]** Las células que contienen la mutación de desplazamiento de marco *Dra* I en el gen de IL-2R $\gamma$  se transfectan con plásmidos que codifican las proteínas de fusión 5-8GL0 y 5-9DL0 y una construcción de donante de *Bsr* BI de 1,5 kb (Ejemplo 14) para reemplazar la mutación de desplazamiento de marco *Dra* I con una secuencia que codifica una proteína funcional. Se obtienen niveles de recombinación homóloga superiores al 1% en estas células, tal como se mide mediante el ensayo de la presencia de un sitio de *Bsr* BI como se describe en el Ejemplo 14. La recuperación de la función génica se demuestra midiendo los niveles de ARNm y proteína y mediante análisis FACS.

10 LISTA DE SECUENCIAS

**[0339]**

15 <110> SANGAMO BIOSCIENCES, INC.  
 <120> MÉTODOS Y COMPOSICIONES PARA LA LIMPIEZA Y RECOMBINACIÓN DIRIGIDA  
 <130> 8325-0036,40  
 20 <140> PCT/US04/25407  
 <141> 2004-08-06  
 <150> 60/493,931  
 25 <151> 2003-08-08  
 <150> 60/518,253  
 <151> 2003-11-07  
 30 <150> 60/530,541  
 <151> 2003-12-18  
 <150> 60/542,780  
 35 <151> 2004-02 -05  
 <150> 60/556,831  
 <151> 2004-03-26  
 40 <150> 60/575,919  
 <151> 2004-06-01  
 <160> 200  
 <170> Patente versión 3,3  
 45 <210> 1  
 <211> 44  
 <212> ADN  
 <213> Artificial  
 50 <220>  
 <223> gen hSMCILI humano  
 <220>  
 55 <221> misc\_feature  
 <222> (5)..(16)  
 <223> Secuencia diana para la ZFP específica de hSMCI  
 <220>  
 60 <221> misc\_feature  
 <222> (23)..(34)  
 <223> Secuencia diana para la ZFP específica de hSMCI  
 <400> 1  
 65 ctgccgccg cgccgccg gtcattgggt tcctgaaact gatt 44



tagtctctgca ggtttaaacg aattcgccct tctcagcaag cgtgagctca ggtctcccc 60  
 gcctccttga acctcaagaa ctgctctgac tccgccagc aacaactcct cgggggatct 120  
 5 ggtccgcagg agcaagtgtt tgttggtgcc atgcaacaag aaaagggggc ggaggcacca 180  
 cgccagtcgt cagctcgctc ctcgatatag caacatcagt ccccgcccct ggtcccactc 240  
 10 ctgccggaag gcgaagatcc cgtaggcct ggacgtattc tcgcgacatt tgccggtcgc 300  
 ccggcttgca ctgcggcgtt tcccgcgcg gctacctcag ttctcggggc tacggcgcgg 360  
 15 cctgtcctac tgctgccggc gccgcggccg tcataagaag cttcctgaaa ctgattgaag 420  
 ggogaattcg cggccgctaa attcaattcg ccctatagtg agt 463

20 <210> 7  
 <211> 50  
 <212> ADN  
 <213> Artificial

25 <220>  
 <223> gen IL2Rgamma humano

30 <220>  
 <221> misc\_feature  
 <222> (8). (16)  
 <223> Secuencia diana para el segundo par de IL2Rgamma ZFP específico

35 <220>  
 <221> misc\_feature  
 <222> (23). (34)  
 <223> Secuencia diana para el segundo par de IL2Rgamma ZFP específico

40 <400> 7  
 ctccaacct ttctctcta ggtacaagaa ctggataat gataaagtc 50

45 <210> 8  
 <211> 9  
 <212> PRT  
 <213> Artificial

50 <220>  
 <223> gen IL2Rgamma humano

<400> 8

50 Tyr Lys Asn Ser Asp Asn Asp Lys Val  
 1 5

55 <210> 9  
 <211> 59  
 <212> ADN  
 <213> Artificial

60 <220>  
 <223> Gen IL2Rgamma

<400> 9  
 gttcctctc ctccaacct ttctctcta ggtacaagaa ctggataat gataaagtc 59

65 <210> 10  
 <211> 9

<212> PRT  
 <213> Artificial  
  
 <220>  
 5 <223> gen IL2Rgamma  
  
 <400> 10  
  
 <210> 11  
 10 <211> 59  
 <212> ADN  
 <213> Artificial  
  
 <220>  
 15 <223> oligonucleótido donador artificial  
  
 <400> 11  
 gttcctctc ctccaacct ttctctcta ggtaaagaa ttccgacaac gataaagtc 59  
  
 <210> 12  
 20 <211> 624  
 <212> ADN  
 <213> Artificial  
  
 <220>  
 25 <2220> Gen de IL2Rgamma  
  
 <400> 12  
  
 30 tagtcctgca ggtttaaacg aattgocctt ttctcttagg taaaagaatt ccgacaacga 60  
 taaagtccag aagtgcagcc actatctatt ccctgaagaa atcacttctg gctgtcagtt 120  
 35 gcaaaaaaag gagatccacc tctaccaaac atttgttggt cagctccagg acccacggga 180  
 acccaggaga caggccacac agatgctaaa actgcagaat ctgggtaatt tggaaagaaa 240  
 40 gggcaagag accagggata ctgtgggaca ttggagtcta cagagtagtg ttcttttatac 300  
 ataagggtac atgggcagaa aagaggaggt aggggatcat gatgggaagg gaggaggtat 360  
 taggggcact accttcagga tcctgacttg tctaggccag gggaatgacc acatatgcac 420  
 45 acatatctcc agtgatcccc tgggctccag agaacctaac acttcacaaa ctgagtgaat 480  
 cccagctaga actgaactgg aacaacagat tcttgaacca ctgtttgag cacttggtgc 540  
 50 agtaccggac taagggcgaa ttcgoggccg ctaaattcaa ttcgccctat agtgagtcgt 600  
 attacaattc actggccgtc gttt 624  
  
 <210> 13  
 55 <211> 700  
 <212> ADN  
 <213> Artificial  
  
 <220>  
 60 <223> gen de beta-globina humana  
  
 <400> 13

ES 2 808 687 T3

tactgatggt atggggccaa gagatatac ttagagggag ggctgagggg ttgaagtcca 60  
 5 actcctaagc cagtgccaga agagccaagg acaggtacgg ctgtcatcac ttagacctca 120  
 ccctgtggag ccacacccta gggttggcca atctactccc aggagcaggg agggcaggag 180  
 ccagggctgg gcataaaagt cagggcagag ccatctattg cttacatttg cttctgacac 240  
 10 aactgtgttc actagcaacc tcaaacagac accatggtgc atctgactcc tgaggagaag 300  
 tctgccgta ctgccctgtg gggcaagggt aacgtggatg aagttggtgg tgaggccctg 360  
 15 ggagagagag aagactcttg ggtttctgat aggactgac tctctctgcc tatttgtcta 480  
 ggagacagag aagactcttg ggtttctgat aggactgac tctctctgcc tatttgtcta 480  
 ttttcccacc cttaggctgc tgggtgtcta ccottggacc cagaggttct ttgagtcctt 540  
 20 tggggatctg tccactcctg atgctgttat gggcaaccct aaggtgaagg ctcattggcaa 600  
 gaaagtgtc ggtgccttta gtgatggcct ggctcacctg gacaacctca agggcacctt 660  
 25 tgccacactg agtgagctgc actgtgacaa gctgcacgtg 700

<210> 14  
 <211> 408  
 <212> ADN  
 <213> Artificial

<220>  
 <223> un producto de amplificación derivado de un mutado gen de la beta-globina

<400> 14

tgcttaccaa gctgtgattc caaatattac gtaaatacac ttgcaaagga ggatgttttt 60  
 40 agtagcaatt tgtactgatg gtatggggcc aagagatata tcttagaggg agggctgagg 120  
 gtttgaagtc caactcctaa gccagtgcca gaagagccaa ggacaggtac ggctgtcatc 180  
 45 acttagacct caccctgtgg agccacaccc tagggttggc caatctactc ccaggagcag 240  
 ggagggcagg agccagggct gggcataaaa gtcagggcag agccatctat tgcttacatt 300  
 tgcttctgac acaactgtgt tcaactagcaa cctcaaacag acaccatggt gcatctgact 360  
 50 cctgaggaga agtctggcgt tagtgcccga attccgatcg tcaaccac 408

<210> 15  
 <211> 42  
 <212> ADN  
 <213> Artificial

<220>  
 <223> gen IL-2Rgamma

<220>  
 <221> misc\_feature  
 <222> (1). (13)  
 <223> secuencias diana para la proteína de fusión 5-10 ZFP/FokI

<220>

ES 2 808 687 T3

<221> misc\_feature  
 <222> (31)..(42)  
 <223> secuencias diana para la proteína de fusión 5-8 ZFP/FokI

5 <400> 15  
 cacgtttcgt gttoggagcc gctttaacct actctgtgga ag 42

10 <210> 16  
 <211> 336  
 <212> PRT  
 <213> Artificial

15 <220>  
 <223> 5-8 ZFP/FokI

20 <220>  
 <221> MISC\_FEATURE  
 <222> (1)..(17)  
 <223> NLS

25 <220>  
 <221> MISC\_FEATURE  
 <222> (18)..(130)  
 <223> ZFP

30 <220>  
 <221> MISC\_FEATURE  
 <222> (131)..(140)  
 <223> enlazador ZC

35 <220>  
 <221> MISC\_FEATURE  
 <222> (141)..(336)  
 <223> Medio dominio de escisión FokI

35 <400> 16

40 **Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala**  
 1 5 10 15

45 **Ala Met Ala Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe**  
 20 25 30

50 **Ser Arg Ser Asp Asn Leu Ser Glu His Ile Arg Thr His Thr Gly Glu**  
 35 40 45

55 **Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Asn Ala**  
 50 55 60

60 **His Arg Ile Asn His Thr Lys Ile His Thr Gly Ser Gln Lys Pro Phe**  
 65 70 75 80

65 **Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp Thr Leu Ser**  
 85 90 95

65 **Glu His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala Cys Asp Ile**

ES 2 808 687 T3

5																	
	Cys	Gly	Arg	Lys	Phe	Ala	Ala	Arg	Ser	Thr	Arg	Thr	Thr	His	Thr	Lys	
			115					120					125				
10																	
	Ile	His	Leu	Arg	Gln	Lys	Asp	Ala	Ala	Arg	Gly	Ser	Gln	Leu	Val	Lys	
		130					135					140					
15																	
	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	Tyr	
	145					150					155					160	
20																	
	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	Thr	
					165					170					175		
25																	
	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val	Met	Glu	Phe	Phe	Met	Lys	Val	
				180					185					190			
30																	
	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	Gly	
			195					200					205				
35																	
	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	Asp	
		210					215					220					
40																	
	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	Asp	
	225					230					235					240	
45																	
	Glu	Met	Gln	Arg	Tyr	Val	Glu	Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	Ile	
					245					250					255		
50																	
	Asn	Pro	Asn	Glu	Trp	Trp	Lys	Val	Tyr	Pro	Ser	Ser	Val	Thr	Glu	Phe	
				260					265					270			
55																	
	Lys	Phe	Leu	Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala	Gln	
			275					280					285				
60																	
	Leu	Thr	Arg	Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu	Ser	
		290					295					300					
65																	
	Val	Glu	Glu	Leu	Leu	Ile	Gly	Gly	Glu	Met	Ile	Lys	Ala	Gly	Thr	Leu	
	305					310					315					320	
70																	
	Thr	Leu	Glu	Glu	Val	Arg	Arg	Lys	Phe	Asn	Asn	Gly	Glu	Ile	Asn	Phe	
					325					330					335		

5  
<210> 17  
<211> 339  
<212> PRT  
<213> Artificial

<220>  
<223> fusión 5-10 ZFP/FokI

10  
<220>  
<221> MISC\_FEATURE  
<222> (1)..(17)  
<223> NLS

15  
<220> <221> MISC\_FEATURE  
<222> (18)..(133)  
<223> ZFP

20  
<220>  
<221> MISC\_FEATURE  
<222> (134)..(143)  
<223> enlazador ZC

25  
<220>  
<221> MISC\_FEATURE  
<222> (144)..(339)  
<223> Medio dominio de escisión FokI

<400> 17

5 Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys  
 145 150 155 160  
 Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg  
 165 170 175  
 10 Asn Ser Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe  
 180 185 190  
 15 Met Lys Val Tyr Gly Tyr Arg Gly Lys His Leu Gly Gly Ser Arg Lys  
 195 200 205  
 20 Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val  
 210 215 220  
 25 Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly  
 225 230 235 240  
 30 Gln Ala Asp Glu Met Gln Arg Tyr Val Glu Glu Asn Gln Thr Arg Asn  
 245 250 255  
 35 Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val  
 260 265 270  
 40 Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr  
 275 280 285  
 45 Lys Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala  
 290 295 300  
 50 Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala  
 305 310 315 320  
 55 Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu  
 325 330 335  
 Ile Asn Phe

60 <21 0> 18  
 <211> 797  
 <212> ADN  
 <213> Artificial

65 <220>  
 <223> gen de proteína verde fluorescente mejorado

ES 2 808 687 T3

<400> 18

5	cgaattctgc agtcgacggt accgcgggcc cgggatccac cggtcgccac catggtgagc	60
	aagggcgagg agctgttcac cggggtggtg cccatcctgg tcgagctgga cggcgacgta	120
10	aacggccaca agttcagcgt gtccggcgag ggcgagggcg atgccaccta cggcaagctg	180
	accctgaagt tcatctgcac caccggcaag ctgccctgac cctggccac cctcgtgacc	240
	accctgacct acggcgtgca gtgcttcagc cgctaccccg accacatgaa gcagcacgac	300
15	ttcttcaagt ccgccatgcc cgaaggctac gtccaggagc gcacatctt cttcaaggac	360
	gacggcaact acaagaccgg cgccgaggtg aagttcgagg gcgacaccct ggtgaaccgc	420
20	atcgagctga agggcatcga cttcaaggag gacggcaaca tcctggggca caagctggag	480
	tacaactaca acagccacaa cgtctatata atggccgaca agcagaagaa cggcatcaag	540
	gtgaacttca agatccgcca caacatcgag gacggcagcg tgcagctcgc cgaccactac	600
25	cagcagaaca ccccatcgg cgacggcccc gtgctgctgc ccgacaacca ctacctgagc	660
	accagtccg ccctgagcaa agaccccaac gagaagcgcg atcacatggt cctgctggag	720
30	ttcgtgaccg ccgccgggat cactctcggc atggacgagc tgtacaagta aagcggccgc	780
	gactctagat cataatc	797

35 <210> 19  
 <211> 795  
 <212> ADN  
 <213> Artificial

40 <220>  
 <223> gen eGFP de mutante defectuoso

<400> 19

ES 2 808 687 T3

5 cgaattctgc agtcgacggt accgcgggcc cgggatccac cggtcgccac catggtgagc 60  
 aagggcgagg agctgttcac cggggtggtg cccatcctgg tcgagctgga cggcgacgta 120  
 aacggccaca agttcagcgt gtccggcgag ggcgagggcg atgccaccta cggcaagctg 180  
 10 accctgaagt tcatctgcac caccggcaag ctgcccgtgc cctggcccac cctcgtgacc 240  
 accctgacct acggcgtgca gtgcttcagc cgctaccctt aacacgaagc agcacgactt 300  
 cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc accatcttct tcaaggacga 360  
 15 cggcaactac aagacccgcg ccgaggtgaa gttcgagggc gacaccctgg tgaaccgcat 420  
 cgagctgaag ggcatcgact tcaaggagga cggcaacatc ctggggcaca agctggagta 480  
 caactacaac agccacaacg tctatatcat ggccgacaag cagaagaacg gcatcaaggt 540  
 20 gaacttcaag atccgccaca acatcgagga cggcagcgtg cagctcgccg accactacca 600  
 gcagaacacc cccatcggcg acggccccgt gctgctgccc gacaaccact acctgagcac 660  
 25 ccagtccgcc ctgagcaaag accccaacga gaagcgcgat cacatggtcc tgctggagtt 720  
 cgtgaccgcc gccgggatca ctctcggcat ggacgagctg tacaagtaa gcgcccgca 780  
 ctctagatca taatc 795

30 <210> 20  
 <211> 734  
 <212> ADN  
 <213> Artificial

35 <220>  
 <223> inserto eGFP en pCR(R)4-TOPO-GFPdonante5

40 <400> 20

45 ggcgaggagc tgttcaccgg ggtggtgccc atcctggtcg agctggacgg cgacgtaaac 60  
 ggccacaagt tcagcgtgtc cggcgagggc gaggcgcgatg ccacctacgg caagctgacc 120  
 ctgaagttca tctgcaccac cggcaagctg cccgtgcctt ggcccaccct cgtgaccacc 180  
 ctgacctacg gcgtgcagtg cttcagccgc taccocgacc acatgaagca gcacgacttc 240  
 50 ttcaagtccg ccatgcccga aggtacgtc caggagcgc ccatcttctt caaggacgac 300  
 ggcaactaca agacccgcg cagaggtgaag ttcgagggcg acaccctggt gaaccgcatc 360  
 gagctgaagg gcatcgactt caaggaggac ggcaacatcc tggggcaca gctggagtac 420  
 55 aactacaaca gccacaacgt ctatatcatg gccgacaagc agaagaacgg catcaaggtg 480  
 aacttcaaga tccgccaca catcgaggac ggcagcgtgc agctcgccga cactaccag 540  
 60 cagaacaccc ccatcggcga cggccccgtg ctgctgccc acaaccacta cctgagcacc 600  
 cagtccgcc tgagcaaaga cccaacgag aagcgcgatc acatggtcct gctggagttc 660  
 gtgaccgcc cgggatcac tctcggcatg gacgagctgt acaagtaaag cggcccgac 720  
 65 tctagatcat aatc 734

ES 2 808 687 T3

<210> 21  
 <211> 1527  
 <212> ADN  
 <213> Artificial

5

<220>  
 <223> inserto eGFP en pCR(R)4-TOPO

10

<400> 21

	ggcgaggagc tgttcaccgg ggtggtgccc atcctggtcg agctggacgg cgacgtaaac	60
15	ggccacaagt tcagcgtgtc cggcgagggc gagggcgatg ccacctacgg caagctgacc	120
	ctgaagttca tctgcaccac cggcaagctg cccgtgacct ggcccacct cgtgaccacc	180
	ctgacctacg gcgtgcagtg cttcagccgc taccctgacc acatgaagca gcacgacttc	240
20	ttcaagtccg ccatgcccga aggctacgtc caggagcgca ccatcttctt caaggacgac	300
	ggcaactaca agaccgcgc cgaggtgaag ttogagggcg acaccctggt gaaccgcatc	360
25	gagctgaagg gcatcgactt caaggaggac ggcaacatcc tggggcaca gctggagtac	420
	aactacaaca gccacaacgt ctatatcatg gccgacaagc agaagaacgg catcaaggtg	480
	aacttcaaga tccgccacaa catcgaggac ggcagcgtgc agctcgccga cactaccag	540
30	cagaacaccc ccatcggcga cggccccgtg ctgctgcccg acaaccacta cctgagcacc	600
	cagtccgcc tgagcaaaga cccaacgag aagcgcgatc acatggtcct gctggagttc	660
35	gtgaccgccg ccgggatcac tctcggcatg gacgagctgt acaagtaaag cggccgctcg	720
	agtctagagg gcccgttta acccgctgat cagcctcgac tgtgccttct agttgccagc	780
	catctgttgt ttgcccctcc cccgtgcctt ccttgacctt ggaaggtgcc actcccactg	840
40	tcctttccta ataaaatgag gaaattgcat cgcattgtct gagtaggtgt cattctattc	900
	tggggggtgg ggtggggcag gacagcaagg gggaggattg ggaagacaat agcaggcatg	960
45	ctggggatgc ggtgggctct atggcttctg aggcggaaag aaccagctgg ggctctaggg	1020
	ggtatcccca cgcgccctgt agcggcgcac taagcgcggc ggggtgtggtg gttacgcgca	1080
	gcgtgaccgc tacacttgcc agcgccttag cgcctgctcc ttcgctttc ttccttctct	1140
50	ttctcgccac gttcgccggc tttccccgtc aagctctaaa tcgggggctc ccttaggggt	1200
	tccgatttag tgctttacgg cacctcgacc ccaaaaaact tgattaggggt gatggttcac	1260
55	gtagtgggoc atcgccctga tagacggttt ttgcctctt gacgttggag tccacgttct	1320
	ttaatagtggt actcttgctt caaactggaa caaactcaa ccctatctcg gtctattctt	1380
	ttgatttata agggattttg ccgatttcgg cctattgggt aaaaaatgag ctgatttaac	1440
60	aaaaatttaa cgcgaattaa ttctgtggaa tgtgtgtcag ttaggggtgtg gaaagtcccc	1500
	aggctcccca gcaggcagaa gtatgca	1527

65

ES 2 808 687 T3

<210> 22  
 <211> 116  
 <212> PRT  
 <213> Artificial

5

<220>  
 <223> sca-29b

10

<400> 22

Met Ala Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser  
 1 5 10 15

15

Gln Ser Gly Asp Leu Thr Arg His Ile Arg Thr His Thr Gly Glu Lys  
 20 25 30

20

Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Thr Ser Ala Asn  
 35 40 45

25

Leu Ser Arg His Thr Lys Ile His Thr Gly Gly Gly Ser Gln Lys  
 50 55 60

30

Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Arg Ser Ala  
 65 70 75 80

35

Leu Ser Arg His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala Cys  
 85 90 95

40

Asp Ile Cys Gly Arg Lys Phe Ala Gln Ser Gly His Leu Ser Arg His  
 100 105 110

45

Thr Lys Ile His  
 115

<210> 23  
 <211> 113  
 <212> PRT  
 <213> Artificial

50

<220>  
 <223> sca-36a

55

<400> 23

ES 2 808 687 T3

5 Met Ala Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser  
 1 5 10 15

10 Arg Ser Gln Thr Arg Lys Thr His Ile Arg Thr His Thr Gly Glu Lys  
 20 25 30

15 Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Gln Lys Arg Asn  
 35 40 45

20 Arg Thr Lys His Thr Lys Ile His Thr Gly Ser Gln Lys Pro Phe Gln  
 50 55 60

25 Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Arg Ser Ala Leu Ser Arg  
 65 70 75 80

30 His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala Cys Asp Ile Cys  
 85 90 95

35 Gly Arg Lys Phe Ala Gln Ser Gly Asn Leu Ala Arg His Thr Lys Ile  
 100 105 110

35 His

40 <210> 24  
 <211> 116  
 <212> PRT  
 <213> Artificial

45 <220>  
 <223> sca-36b

<400> 24

ES 2 808 687 T3

1 Met Ala Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser  
 5 Thr Ser Gly Ser Leu Ser Arg His Ile Arg Thr His Thr Gly Glu Lys  
 10 Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Asp Arg Ser Asp  
 15 Leu Ser Arg His Thr Lys Ile His Thr Gly Gly Gly Gly Ser Gln Lys  
 20 Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Arg Ser Ala  
 25 Leu Ser Arg His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala Cys  
 30 Asp Ile Cys Gly Arg Lys Phe Ala Gln Ser Gly Asn Leu Ala Arg His  
 35 Thr Lys Ile His  
 <210> 25  
 <211> 116  
 <212> PRT  
 <213> Artificial  
 <220>  
 <223> sca-36c  
 45 <400> 25

ES 2 808 687 T3

1 Met Ala Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser  
 5 Thr Ser Ser Ser Leu Ser Arg His Ile Arg Thr His Thr Gly Glu Lys  
 10 Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Asp Arg Ser Asp  
 15 Leu Ser Arg His Thr Lys Ile His Thr Gly Gly Gly Gly Ser Gln Lys  
 20 Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Arg Ser Ala  
 25 Leu Ser Arg His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala Cys  
 30 Asp Ile Cys Gly Arg Lys Phe Ala Gln Ser Gly Asn Leu Ala Arg His  
 35 Thr Lys Ile His  
 115

<210> 26  
 <211> 12  
 <212> PRT  
 <213> Artificial

<220>  
 <223> nucleasa de fusión

<400> 26

His Gln Arg Thr His Gln Asn Lys Lys Gln Leu Val  
 1 5 10

<210> 27  
 <211> 12  
 <212> ADN  
 <213> Artificial

<220>  
 <223> Secuencia diana del gen hSMCILI

<400> 27  
 catggggttc ct 12

<210> 28  
 <211> 7  
 <212> PRT  
 <213> Artificial

ES 2 808 687 T3

<220>  
<223> Gen F1 hSMCILI

5 <400> 28

Arg Ser His Asp Leu Ile Glu  
1 5

10

<210> 29  
<211> 7  
<212> PRT  
<213> Artificial

15

<220>  
<223> Gen F2 hSMCILI

20 <400> 29

Thr Ser Ser Ser Leu Ser Arg  
1 5

25

<210> 30  
<211> 7  
<212> PRT  
<213> Artificial

30

<220>  
<223> Gen F3 hSMCILI

35 <400> 30

Arg Ser Asp His Leu Ser Thr  
1 5

40

<210> 31  
<211> 7  
<212> PRT  
<213> Artificial

45

<220>  
<223> Gen F4 hSMCILI

50 <400> 31

Thr Asn Ser Asn Arg Ile Thr  
1 5

55

<210> 32  
<211> 12  
<212> ADN  
<213> Artificial

60

<220>  
<223> Secuencia diana de gen hSMCILI

65 <400> 32

ES 2 808 687 T3

gcggcgccgg cg 12

5 <210> 33  
<211> 7  
<212> PRT  
<213> Artificial  
  
<220>  
10 <223> Gen F1 hSMCILI  
  
<400> 33

15 Arg Ser Asp Asp Leu Ser Arg  
1 5

20 <210> 34  
<211> 7  
<212> PRT  
<213> Artificial  
  
<220>  
25 <223> Gen F2 hSMC1L1  
  
<400> 34

30 Arg Ser Asp Asp Arg Lys Thr  
1 5

35 <210> 35  
<211> 7  
<212> PRT  
<213> Artificial  
  
<220>  
40 <223> Gen F3 hSMC1L1  
  
<400> 35

45 Arg Ser Glu Asp Leu Ile Arg  
1 5

50 <210> 36  
<211> 7  
<212> PRT  
<213> Artificial  
  
<220>  
55 <223> Gen F4 hSMC1L1  
  
<400> 36

60 Arg Ser Asp Thr Leu Ser Arg  
1 5

65

5 <210> 37  
 <211> 22  
 <212> ADN  
 <213> Artificial

<220>  
 <223> Gen hSMC1L1 Amp inicial 1

10 <400> 37  
 agcaacaact cctccgggga tc 22

15 <210> 38  
 <211> 21  
 <212> ADN  
 <213> Artificial

<220>  
 <223> Gen hSMC1L1 Amp inicial 2

20 <400> 38  
 ttccagacgc gactcttgg c 21

25 <210> 39  
 <211> 25  
 <212> ADN  
 <213> Artificial

30 <220>  
 <223> Gen hSMC1L1 Cebador específico de cromosomas

<400> 39  
 ctcaagaagc gtagctcag gtctc 25

35 <210> 40  
 <211> 23  
 <212> ADN  
 <213> Artificial

40 <220>  
 <223> hSMC1L1 Cebador específico de donante de genes

<400> 40  
 caatcagttt caggaagctt ctt 23

45 <210> 41  
 <211> 25  
 <212> ADN  
 <213> Artificial

50 <220>  
 <223> hSMC1L1 Gen fuera 1

<400> 41  
 ctcaagaagc gtagctcag gtctc 25

55 <210> 42  
 <211> 23  
 <212> ADN  
 <213> Artificial

60 <220>  
 <223> hSMC1L1 Gen fuera 2

65 <400> 42  
 ggggtcaagt aaggctggga agc 23

ES 2 808 687 T3

5 <210> 43  
<211> 12  
<212> ADN  
<213> Artificial  
  
<220>  
<223> secuencia de diana de gen IL2Rgamma  
  
10 <400> 43  
aactcggata en 12  
  
15 <210> 44  
<211> 7  
<212> PRT  
<213> Artificial  
  
20 <220>  
<223> gen F1 IL2Rgammae  
  
<400> 44  
  
25 **Asp Arg Ser Thr Leu Ile Glu**  
**1 5**  
  
30 <210> 45  
<211> 7  
<212> PRT  
<213> Artificial  
  
35 <220>  
<223> gen F2 IL2Rgamma  
  
<400> 45  
  
40 **Ser Ser Ser Asn Leu Ser Arg**  
**1 5**  
  
45 <210> 46  
<211> 7  
<212> PRT  
<213> Artificial  
  
50 <220>  
<223> gen F3 IL2Rgammae  
  
<400> 46  
  
55 **Arg Ser Asp Asp Leu Ser Lys**  
**1 5**  
  
60 <210> 47  
<211> 7  
<212> PRT  
<213> Artificial  
  
65 <220>  
<223> gen F4 IL2Rgamma

<400> 47

5  
Asp Asn Ser Asn Arg Ile Lys  
1 5

10  
<210> 48  
<211> 13  
<212> ADN  
<213> Artificial

15  
<220>  
<223> Secuencia diana del gen IL2Rgamma

20  
<400> 48  
tagaggagaa agg 13

<210> 49  
<211> 7  
<212> PRT  
<213> Artificial

25  
<220>  
<223> gen F1 IL2Rgamma

30  
<400> 49

35  
Arg Ser Asp Asn Leu Ser Asn  
1 5

<210> 50  
<211> 7  
<212> PRT  
<213> Artificial

40  
<220>  
<223> gen F2 IL2Rgamma

45  
<400> 50

50  
Thr Ser Ser Ser Arg Ile Asn  
1 5

55  
<210> 51  
<211> 7  
<212> PRT  
<213> Artificial

60  
<220>  
<223> gen F3 IL2Rgamma

<400> 51

65  
Arg Ser Asp His Leu Ser Gln  
1 5

ES 2 808 687 T3

5 <210> 52  
<211> 7  
<212> PRT  
<213> Artificial

10 <220>  
<223> gen F4 IL2Rgamma  
  
<400> 52

15 **Arg Asn Ala Asp Arg Lys Thr**  
**1 5**

20 <210> 53  
<211> 12  
<212> ADN  
<213> Artificial

25 <220>  
<223> secuencia diana de gen IL2Rgamma

30 <400> 53  
tacaagaact cg 12

35 <210> 54  
<211> 7  
<212> PRT  
<213> Artificial

40 <220>  
<223> gen F1 IL2Rgamma  
  
<400> 54

45 **Arg Ser Asp Asp Leu Ser Lys**  
**1 5**

50 <210> 55  
<211> 7  
<212> PRT  
<213> Artificial

55 <220>  
<223> gen F2 IL2Rgamma  
  
<400> 55

60 **Asp Asn Ser Asn Arg Ile Lys**  
**1 5**

65 <210> 56  
<211> 7  
<212> PRT  
<213> Artificial  
  
<220>

ES 2 808 687 T3

<223> gen F3 IL2Rgamma

<400> 56

5

Arg Ser Asp Ala Leu Ser Val  
1 5

10

<210> 57  
<211> 7  
<212> PRT  
<213> Artificial

15

<220>  
<223> gen F4 IL2Rgamma

<400> 57

20

Asp Asn Ala Asn Arg Thr Lys  
1 5

25

<210> 58  
<211> 9  
<212> ADN  
<213> Artificial

30

<220>  
<223> secuencia diana del gen IL2Rgamma

<400> 58  
ggagaaagg 9

35

<210> 59  
<211> 7  
<212> PRT  
<213> Artificial

40

<220>  
<223> gen F1 IL2Rgamma

<400> 59

45

Arg Ser Asp His Leu Thr Gln  
1 5

50

<210> 60  
<211> 7  
<212> PRT  
<213> Artificial

55

<220>  
<223> gen F2 IL2Rgamma

60

<400> 60

Gln Ser Gly Asn Leu Ala Arg  
1 5

65

ES 2 808 687 T3

5 <210> 61  
 <211> 7  
 <212> PRT  
 <213> artificial

10 <220>  
 <223> gen F3 IL2Rgamma  
 <400> 61

15 **Arg Ser Asp His Leu Ser Arg**  
 1 5

20 <210> 62  
 <211> 18  
 <212> ADN  
 <213> Artificial

25 <220>  
 <223> secuencia reemplazada

30 <400> 62  
 tacaagaact cggataat 18

35 <210> 63  
 <211> 18  
 <212> ADN  
 <213> Artificial

40 <220>  
 <223> secuencia de reemplazo

45 <400> 63  
 taaaagaatt ccgacaac 18

50 <210> 64  
 <211> 25  
 <212> ADN  
 <213> Artificial

55 <220>  
 <223> gen IL2Rgamma Amp inicial 1

60 <400> 64  
 tgtcgagtac atgaattgca cttgg 25

65 <210> 65  
 <211> 22  
 <212> ADN  
 <213> Artificial

<220>  
 <223> gen IL2Rgamma Amp inicial 2

60 <400> 65  
 ttaggttctc tggagcccag gg 22

65 <210> 66  
 <211> 25  
 <212> ADN  
 <213> Artificial

ES 2 808 687 T3

<220>  
<223> Cebador específico del cromosoma de gen IL2Rgamma

5 <400> 66  
ctccaaacag tgggtcaaga atctg 25

<210> 67  
<211> 26  
10 <212> ADN  
<213> Artificial

<220>  
<223> Cebador específico del donante de gen IL2Rgamma

15 <400> 67  
tcctctaggt aaagaattcc gacaac 26

<210> 68  
<211> 12  
20 <212> ADN  
<213> Artificial

<220>  
<223> Secuencia diana del gen de beta-globina

25 <400> 68  
gggcagtaac gg 12

<210> 69  
<211> 7  
30 <212> PRT  
<213> Artificial

<220>  
<223> Gen F1 de beta-globina

35 <400> 69

40 Arg Ser Asp His Leu Ser Glu  
1 5

<210> 70  
<211> 7  
45 <212> PRT  
<213> Artificial

50 <220>  
<223> Gen F2 de beta-globina

<400> 70

55 Gln Ser Ala Asn Arg Thr Lys  
1 5

<210> 71  
<211> 7  
60 <212> PRT  
<213> Artificial

65

ES 2 808 687 T3

<220>  
<223> Gen F3 de beta-globina

5 <400> 71

Arg Ser Asp Asn Leu Ser Ala  
1 5

10

<210> 72  
<211> 7  
<212> PRT  
<213> Artificial

15

<220>  
<223> Gen F4 de beta-globina

20 <400> 72

Arg Ser Gln Asn Arg Thr Arg  
1 5

25

<210> 73  
<211> 12  
<212> ADN  
<213> Artificial

30

<220>  
<223> Secuencia diana del gen de beta-globina

35 <400> 73

aaggtgaacg tg 12

<210> 74  
<211> 7  
<212> PRT  
<213> Artificial

40

<220>  
<223> Gen F1 de beta-globina

45 <400> 74

Arg Ser Asp Ser Leu Ser Arg  
1 5

50

<210> 75  
<211> 7  
<212> PRT  
<213> Artificial

55

<220>  
<223> Gen F2 de beta-globina

60 <400> 75

Asp Ser Ser Asn Arg Lys Thr  
1 5

65

5 <210> 76  
 <211> 7  
 <212> PRT  
 <213> Artificial  
  
 <220>  
 <223> Gen F3 de beta-globina  
 10 <400> 76  
  
 Arg Ser Asp Ser Leu Ser Ala  
 1 5  
 15  
  
 <210> 77  
 <211> 7  
 <212> PRT  
 <213> Artificial  
 20  
  
 <220>  
 <223> Gen F4 de beta-globina  
 25 <400> 77  
  
 Arg Asn Asp Asn Arg Lys Thr  
 1 5  
 30  
  
 <210> 78  
 <211> 32  
 <212> ADN  
 <213> Artificial  
 35  
  
 <220>  
 <223> secuencia reemplazada  
  
 <400> 78  
 40 ccgttactgc cctgtggggc aaggagaacg tg 32  
  
 <210> 79  
 <211> 32  
 <212> ADN  
 <213> Artificial  
 45  
  
 <220>  
 <223> secuencia de reemplazo  
  
 <400> 79  
 50 gcgtagtagc ccgaattccg atcgtaacc ac 32  
  
 <210> 80  
 <211> 23  
 <212> ADN  
 <213> Artificial  
 55  
  
 <220>  
 <223> gen de beta-globina humana amp inicial 1  
 60  
  
 <400> 80  
 tactgatggt atggggccaa gag 23  
  
 <210> 81  
 <211> 22  
 <212> ADN  
 65

ES 2 808 687 T3

<213> Artificial  
<220>  
<223> gen de beta-globina humana amp inicial 2  
5  
<400> 81  
cacgtgcagc ttgtcacagt gc 22  
10  
<210> 82  
<211> 22  
<212> ADN  
<213> Artificial  
15  
<220>  
<223> Cebador específico de cromosoma de gen de beta-globina humana  
<400> 82  
tgcttaccaa gctgtgattc ca 22  
20  
<210> 83  
<211> 18  
<212> ADN  
<213> Artificial  
25  
<220>  
<223> Cebador específico del donante de gen humano de beta-globina  
<400> 83  
ggttgacgat cggaattc 18  
30  
<210> 84  
<211> 12  
<212> ADN  
<213> Artificial  
35  
<220>  
<223> sitio diana para la ZFP  
<400> 84  
aactcggata en 12  
40  
<210> 85  
<211> 7  
<212> PRT  
<213> Artificial  
45  
<220>  
<223> F1  
50  
<400> 85  
  
**Asp Arg Ser Thr Leu Ile Glu**  
**1 5**  
55  
<210> 86  
<211> 7  
<212> PRT  
<213> Artificial  
60  
<220>  
<223> F2  
65  
<400> 86

ES 2 808 687 T3

Ser Ser Ser Asn Leu Ser Arg  
1 5

5

10 <210> 87  
<211> 7  
<212> PRT  
<213> Artificial

15 <220>  
<223> F3  
<400> 87

20 Arg Ser Asp Asp Leu Ser Lys  
1 5

25 <210> 88  
<211> 7  
<212> PRT  
<213> Artificial

30 <220>  
<223> F4  
<400> 88

35

Asp Asn Ser Asn Arg Ile Lys  
1 5

40

45 <210> 89  
<211> 18  
<212> PRT  
<213> Artificial

<220>  
<223> Conector de 10 residuos

50 <220>  
<221> MISC\_FEATURE  
<222> (6)..(15)  
<223> enlazador ZC

55 <400> 89

His Thr Lys Ile His Leu Arg Gln Lys Asp Ala Ala Arg Gly Ser Gln  
1 5 10 15

60

Leu Val

65 <210> 90  
<211> 14

```

<212> PRT
<213> Artificial

5
<220>
<223> enlazador de 6 residuos

<220>
<221> MISC_FEATURE
<222> (6)..(11)
10 <223> enlazador ZC

<400> 90

15
      His Thr Lys Ile His Leu Arg Gln Lys Gly Ser Gln Leu Val
      1           5           10

20
<210> 91
<211> 13
<212> PRT
<213> Artificial

25
<220>
<223> enlazador de 5 residuos

<220>
<221> MISC_FEATURE
30 <222> (6)..(10)
<223> enlazador ZC

<400> 91

35
      His Thr Lys Ile His Leu Arg Gln Gly Ser Gln Leu Val
      1           5           10

40
<210> 92
<211> 12
<212> PRT
<213> Artificial

45
<220>
<223> enlazador de 4 residuos

<220>
<221> MISC_FEATURE
50 <222> (6)..(9)
<223> enlazador ZC

<400> 92

55
      His Thr Lys Ile His Leu Arg Gly Ser Gln Leu Val
      1           5           10

60
<210> 93
<211> 11
<212> PRT
<213> Artificial

65
<220>

```

# ES 2 808 687 T3

<223> conector de 3 residuos  
 <220>  
 <221> MISC\_FEATURE  
 5 <222> (6)..(8)  
 <223> Conector ZC  
 <400> 93  
 10  
 His Thr Lys Ile His Leu Gly Ser Gln Leu Val  
 15                   1                   5                   10  
 <210> 94  
 <211> 10  
 <212> PRT  
 <213> Artificial  
 25 <220>  
 <223> enlazador de 2 residuos  
 <220>  
 <221> MISC\_FEATURE  
 30 <222> (6)..(7)  
 <223> enlazador ZC  
 <400> 94  
 35                   His Thr Lys Ile His Gly Ser Gln Leu Val  
                   1                   5                   10  
 <210> 95  
 <211> 38  
 <212> ADN  
 <213> Artificial  
 45 <220>  
 <223> Separación de 4 pb  
 <220>  
 <221> misc\_feature  
 50 <222> (6)..(17)  
 <223> Sitio diana de ZFP  
 <220>  
 <221> misc\_feature  
 55 <222> (22)..(33)  
 <223> Sitio diana de ZFP  
 <400> 95  
 60 ctagcattat ccgagttaca caactcggat aatgctag 38  
 <210> 96  
 <211> 39  
 <212> ADN  
 65 <213> Artificial

<220>  
 <223> Separación de 5 pb

5

<220>  
 <221> misc\_feature  
 <222> (6)..(17)  
 <223> Sitio diana de ZFP

10

<220>  
 <221> misc\_feature  
 <222> (23)..(34)  
 <223> Sitio diana de ZFP

15

<400> 96  
 ctgacattat ccgagttcac acaactcgga taatgctag 39

20

<210> 97  
 <211> 42  
 <212> ADN  
 <213> Artificial

25

<220>  
 <223> Separación 6bp

30

<220>  
 <221> misc\_feature  
 <222> (7)..(18)  
 <223> Sitio diana de ZFP

35

<220>  
 <221> misc\_feature  
 <222> (25)..(36)  
 <223> Sitio diana de ZFP

40

<400> 97  
 ctaggatta tccgagtca ccacaactcg gataatgact ag 42

45

<210> 98  
 <211> 41  
 <212> ADN  
 <213> Artificial

50

<220>  
 <223> Separación de 7 pb

55

<220>  
 <221> misc\_feature  
 <222> (6)..(17)  
 <223> Secuencia diana de ZFP

60

<220>  
 <221> misc\_feature  
 <222> (25)..(36)  
 <223> Secuencia diana de ZFP

65

<400> 98  
 ctgacattat ccgagttcac acacaactcg gataatgcta g 41

<210> 99  
 <211> 42  
 <212> ADN  
 <213> Artificial

<220>  
 <223> Separación de 8 pb

<220>  
 <221> misc\_feature  
 <222> (6)..(17)  
 5 <223> Sitio diana de ZPF

<220>  
 <221> misc\_feature  
 <222> (26)..(37)  
 10 <223> Sitio diana de ZPF

<400> 99  
 ctgacattat ccgagttcac cacacaactc ggataatgct ag 42

15 <210> 100  
 <211> 43  
 <212> ADN  
 <213> Artificial

20 <220>  
 <223> Separación de 9 pb

25 <220>  
 <221> misc\_feature  
 <222> (6)..(17)  
 <223> Sitio diana de ZPF

30 <220>  
 <221> misc\_feature  
 <222> (27)..(38)  
 <223> Sitio diana de ZPF

35 <400> 100  
 ctgacattat ccgagttcac acacacaact cggataatgc tag 43

40 <210> 101  
 <211> 46  
 <212> ADN  
 <213> Artificial

<220>  
 <223> Separación de 12 pb

45 <220>  
 <221> misc\_feature  
 <222> (6)..(17)  
 <223> Sitio diana de ZPF

50 <220>  
 <221> misc\_feature  
 <222> (30)..(41)  
 <223> Sitio diana de ZPF

55 <400> 101  
 ctgacattat ccgagttcac caccaacaca actcggataa tgctag 46

60 <210> 102  
 <211> 49  
 <212> ADN  
 <213> Artificial

<220>  
 <223> Separación de 15bp

65 <220>  
 <221> misc\_feature

<222> (6)..(17)  
 <223> Sitio diana de ZFP  
  
 <220>  
 5 <221> misc\_feature  
 <222> (33)..(44)  
 <223> Sitio diana de ZFP  
  
 <400> 102  
 10 ctgacattat ccgagttcac caccaaccac acaactcgga taatgctag 49  
  
 <210> 103  
 <211> 50  
 <212> ADN  
 15 <213> Artificial  
  
 <220>  
 <223> Separación de 16 pb  
  
 <220>  
 20 <221> misc\_feature  
 <222> (6)..(17)  
 <223> Sitio diana de ZFP  
  
 <220> <221> misc\_feature  
 25 <222> (34)..(45)  
 <223> Sitio diana de ZFP  
  
 <400> 103  
 30 ctgacattat ccgagttcac caccaaccac accaactcgg ataatgctag 50  
  
 <210> 104  
 <211> 51  
 <212> ADN  
 35 <213> Artificial  
  
 <220>  
 <223> Separación de 17 pb  
  
 <220>  
 40 <221> misc\_feature  
 <222> (6)..(17)  
 <223> Sitio diana de ZFP  
  
 <220>  
 45 <221> misc\_feature  
 <222> (35)..(46)  
 <223> Sitio diana de ZFP  
  
 <400> 104  
 50 ctgacattat ccgagttcaa ccaccaacca caccaactcg gataatgcta g 51  
  
 <210> 105  
 <211> 56  
 55 <212> ADN  
 <213> Artificial  
  
 <220>  
 <223> Separación de 22 pb  
  
 <220>  
 60 <221> misc\_feature  
 <222> (6)..(17)  
 <223> Sitio diana de ZFP  
  
 <220>  
 65

ES 2 808 687 T3

<221> misc\_feature  
<222> (40)..(51)  
<223> Sitio diana de ZFP

5 <400> 105  
ctagcattat cggagttcaa ccaccaacca c accaacaca actcggataa tgctag 56

<210> 106  
<211> 60  
10 <212> ADN  
<213> Artificial

<220>  
<223> Separación de 26 pb

15 <220>  
<221> misc\_feature  
<222> (6)..(17)  
<223> Sitio diana de ZFP

20 <220>  
<221> misc\_feature  
<222> (44)..(55)  
<223> Sitio diana de ZFP

25 <400> 106  
ctagcattat cggagttcaa ccaccaacca caccaacacc accaactcgg ataatgctag 60

<210> 107  
<211> 4  
30 <212> PRT  
<213> Artificial

<220>  
<223> enlazador

35 <400> 107

40 **Leu Arg Gly Ser**  
**1**

45 <210> 108  
<211> 4  
<212>  
<213> Artificial

50 <220>  
<223> enlazador

55 <400> 108

**Leu Gly Gly Ser**  
**1**

60 <210> 109  
<211> 4  
<212> PRT  
65 <213> Artificial

<220>  
<223> enlazador

5 <400> 109

Thr Gly Gly Ser  
1

10

15 <210> 110  
<211> 4  
<212>  
<213> Artificial

20 <220>  
<223> enlazador  
<400> 110

25 Gly Gly Gly Ser  
1

30 <210> 111  
<211> 4  
<212> PRT  
<213> Artificial

35 <220>  
<223> enlazador  
<400> 111

40 Leu Pro Gly Ser  
1

45 <210> 112  
<211> 4  
<212> **PRT**  
<213> Artificial

50 <220>  
<223> enlazador  
<400> 112

55 Leu Arg Lys Ser  
1

60

65 <210> 113  
<211> 4  
<212> **PRT**  
<213> Artificial

ES 2 808 687 T3

<220>  
<223> enlazador

5 <400> 113

**Leu Arg Trp Ser**  
**1**

10

<210> 114  
<211> 12  
<212> ADN  
<213> Artificial

15

<220>  
<223> fusión de gen IL2Rgamma 5-8 secuencia diana

20

<400> 114  
actctgtgga ag 12

25

<210> 115  
<211> 7  
<212> PRT  
<213> Artificial

30

<220>  
<223> fusión de gen IL2Rgamma 5-8 F1

<400> 115

35

**Arg Ser Asp Asn Leu Ser Glu**  
**1 5**

40

<210> 116  
<211> 7  
<212> PRT  
<213> Artificial

45

<220>  
<223> fusión de gen IL2Rgamma 5-8 F2

<400> 116

50

**Arg Asn Ala His Arg Ile Asn**  
**1 5**

55

<210> 117  
<211> 7  
<212>  
<213> Artificial

60

<220>  
<223> fusión de gen IL2Rgamma 5-8 F3

<400> 117

ES 2 808 687 T3

Arg Ser Asp Thr Leu Ser Glu  
1 5

5

<210> 118  
<211> 7  
<212> PRT  
<213> Artificial

10

<220>  
<223> fusión de gen IL2Rgamma 5-8 F4

15

<400> 118

Ala Arg Ser Thr Arg Thr Thr  
1 5

20

<210> 119  
<211> 13  
<212> ADN  
<213> Artificial

25

<220>  
<223> fusión de gen IL2Rgamma 5-10 secuencia diana

30

<400> 119  
aacacgaaac gtg 13

35

<210> 120  
<211> 7  
<212> PRT  
<213> Artificial

40

<220>  
<223> fusión de gen IL2Rgamma 5-10 F1

<400> 120

45

Arg Ser Asp Ser Leu Ser Arg  
1 5

50

<210> 121  
<211> 7  
<212> PRT  
<213> Artificial

55

<220>  
<223> fusión de gen IL2Rgamma 5-10 F2

<400> 121

60

Asp Ser Ser Asn Arg Lys Thr  
1 5

65

<210> 122  
<211> 7

ES 2 808 687 T3

<212> PRT  
<213> Artificial  
<220> **PRT**  
<223> fusión de gen IL2Rgamma 5-10 F3  
5  
<400> 122  
  
Arg Ser Asp Ser Leu Ser Val  
1 5  
10  
  
<210> 123  
<211> 7  
<212> PRT  
<213> Artificial  
15  
  
<220>  
<223> fusión de gen IL2Rgamma 5-10 F4  
20  
<400> 123  
  
Asp Arg Ser Asn Arg Ile Thr  
1 5  
25  
  
<210> 124  
<211> 8  
<212> ADN  
<213> Artificial  
30  
  
<220>  
<223> secuencia reemplazada  
35  
<400> 124  
gaccacat 8  
  
<210> 125  
<211> 6  
<212> ADN  
<213> Artificial  
40  
  
<220>  
<223> secuencia de reemplazo  
45  
<400> 125  
taacac 6  
  
<210> 126  
<211> 17  
<212> ADN  
<213> Artificial  
50  
  
<220>  
<223> GFP-Bam  
55  
<400> 126  
cgaattctgc agtcgac 17  
60  
  
<210> 127  
<211> 18  
<212> ADN  
<213> Artificial  
65  
<220>

ES 2 808 687 T3

<223> GFP-Xba  
<400> 127  
gattatgac tagagtcg 18  
5  
<210> 128  
<211> 25  
<212> ADN  
<213> Artificial  
10  
<220>  
<223> parada sentido2  
<400> 128  
15 agccgctacc cctaacacga agcag 25  
<210> 129  
<211> 25  
<212> ADN  
20 <213> Artificial  
<220>  
<223> parada anti2  
25 <400> 129  
ctgcttcgtg ttaggggtag cggct 25  
<210> 130  
<211> 7  
30 <212> PRT  
<213> Artificial  
<220>  
<223> 287A F1  
35 <400> 130  
  
40 **Arg Ser Asp Asp Leu Thr Arg**  
**1 5**  
  
<210> 131  
<211> 7  
45 <212> PRT  
<213> Artificial  
<220>  
<223> 287A F2  
50 <400> 131  
  
55 **Gln Ser Gly Ala Leu Ala Arg**  
**1 5**  
  
60 <210> 132  
<211> 7  
<212> PRT  
<213> Artificial  
65 <220>  
<223> 287A F3

ES 2 808 687 T3

<400> 132

5

Arg Ser Asp His Leu Ser Arg  
1 5

10

<2 10> 133  
<211> 7  
<212> PRT  
<213> Artificial

15

<220>  
<223> 296S F1

<400> 133

20

Gln Ser Gly Ser Leu Thr Arg  
1 5

25

<210> 134  
<211> 7  
<212> PRT  
<213> Artificial

30

<220>  
<223> 296S F2

<400> 134

35

Gln Ser Gly Asp Leu Thr Arg  
1 5

40

<210> 135  
<211> 7  
<212> PRT  
<213> Artificial

45

<220>  
<223> 296S F3

<400> 135

50

Gln Ser Gly Asn Leu Ala Arg  
1 5

55

<210> 136  
<211> 10  
<212> **ADN**  
<213> Artificial

60

<220>  
<223> 287A secuencia diana

65

<400> 136  
gggtagcgg 10

ES 2 808 687 T3

5 <210> 137  
<211> 7  
<212> PRT  
<213> Artificial

<220>  
<223> 287A F1

10 <400> 137

Arg Ser Asp Asp Leu Thr Arg  
1 5

15

20 <210> 138  
<211> 7  
<212> PRT  
<213> Artificial

<220>  
<223> 287A F2

25 <400> 138

Gln Ser Gly Ala Leu Ala Arg  
1 5

30

35 <210> 139  
<211> 7  
<212> PRT  
<213> Artificial

<220>  
<223> 287A F3

40 <400> 139

Arg Ser Asp His Leu Ser Arg  
1 5

45

50 <210> 140  
<211> 9  
<212> ADN  
<213> Artificial

<220>  
<223> Secuencia diana 296S

<400> 140  
gaagcagca 9

60 <210> 141  
<211> 7  
<212> PRT  
<213> Artificial

65 <220>  
<223> 296S F1

ES 2 808 687 T3

<400> 141

5

Gln Ser Gly Ser Leu Thr Arg  
1 5

10

<210> 142  
<211> 7  
<212> PRT  
<213> Artificial

15

<220>  
<223> 296S F2  
<400> 142

20

Gln Ser Gly Asp Leu Thr Arg  
1 5

25

<210> 143  
<211> 7  
<212> PRT  
<213> Artificial

30

<220>  
<223> 296S F3  
<400> 143

35

Gln Ser Gly Asn Leu Ala Arg  
1 5

40

<210> 144  
<211> 17  
<212> ADN  
<213> Artificial

45

<220>  
<223> eGFP cebador 1 (5T)

<400> 144  
ctgctgcccg acaacca 17

50

<210> 145  
<211> 19  
<212> ADN  
<213> Artificial

55

<220>  
<223> eGFP cebador 2 (3T)

<400> 145  
ccatgtgatac ggccttctc 19

60

<210> 146  
<211> 22  
<212> ADN  
<213> Artificial

65

<220>

<223> Sonda eGFP  
 <400> 146  
 cccagtccgc cctgagcaaa ga 22  
 5  
 <210> 147  
 <211> 21  
 <212> ADN  
 <213> Artificial  
 10  
 <220>  
 <223> GAPDH cebador 1  
 <400> 147  
 ccatgttcgt catgggtgtg a 21  
 15  
 <210> 148  
 <211> 20  
 <212> ADN  
 <213> Artificial  
 20  
 <220>  
 <223> GAPDH cebador 2  
 <400> 148  
 catggactgt ggcatgagt 20  
 25  
 <210> 149  
 <211> 24  
 <212> ADN  
 <213> Artificial  
 30  
 <220>  
 <223> Sonda GAPDH  
 <400> 149  
 tcctgcacca ccaactgctt agca 24  
 35  
 <210> 150  
 <211> 17  
 <212> ADN  
 <213> Artificial  
 40  
 <220>  
 <223> GFPnostart  
 <400> 150  
 ggcgaggagc tgttcac 17  
 45  
 <210> 151  
 <211> 18  
 <212> ADN  
 <213> Artificial  
 50  
 <220>  
 <223> GFP-Xba  
 <400> 151  
 gattatgac tagagtcg 18  
 55  
 <210> 152  
 <211> 12  
 <212> ADN  
 <213> Artificial  
 60  
 <220>  
 65

ES 2 808 687 T3

<223> Secuencia diana del gen IL2Rgamma

<400> 152  
actctgtgga ag 12

5

<210> 153  
<211> 7  
<212> PRT  
<213> Artificial

10

<220>  
<223> gen F1 IL2Rgamma

15

<400> 153

Arg Ser Asp Asn Leu Ser Val  
1 5

20

<210> 154  
<211> 7  
<212> **PRT**  
<213> Artificial

25

<220>  
<223> gen F2 IL2Rgamma

30

<400> **154**

Arg Asn Ala His Arg Ile Asn  
1 5

35

<210> 155  
<211> 7  
<212> PRT  
<213> Artificial

40

<220>  
<223> gen F3 IL2Rgamma

45

<400> 155

Arg Ser Asp Thr Leu Ser Glu  
1 5

50

<210> 156  
<211> 7  
<212> **PRT**  
<213> Artificial

55

<220>  
<223> gen F4 IL2Rgamma

60

<400> 156

Ala Arg Ser Thr Arg Thr Asn  
1 5

65

ES 2 808 687 T3

5 <210> 157  
<211> 12  
<212> ADN  
<213> Artificial

<220>  
<223> Secuencia diana del gen IL2Rgamma

10 <400> 157  
aaagcggctc cg 12

15 <210> 158  
<211> 7  
<212> PRT  
<213> Artificial

<220>  
<223> gen F1 IL2Rgamma

20 <400> 158

25 Arg Ser Asp Thr Leu Ser Glu  
1 5

30 <210> 159  
<211> 7  
<212> **PRT**  
<213> Artificial

<220>  
<223> gen F2 IL2Rgamma

35 <400> 159

40 Ala Arg Ser Thr Arg Thr Thr  
1 5

45 <210> 160  
<211> 7  
<212> PRT  
<213> Artificial

<220>  
<223> gen F3 IL2Rgamma

50 <400> 160

55 Arg Ser Asp Ser Leu Ser Lys  
1 5

60 <210> 161  
<211> 7  
<212> **PRT**  
<213> Artificial

<220>  
<223> gen F4 IL2Rgamma

65 <400> 161

Gln Arg Ser Asn Leu Lys Val  
 1 5

5 <210> 162  
 <211> 117  
 <212> PRT  
 <213> Artificial

10 <220>  
 <223> nucleasa  
 <400> 162

15 Met Ala Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser  
 1 5 10 15

20 Arg Ser Asp Asn Leu Ser Val His Ile Arg Thr His Thr Gly Glu Lys  
 20 25 30

25 Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Asn Ala His  
 35 40 45

30 Arg Ile Asn His Thr Lys Ile His Thr Gly Ser Gln Lys Pro Phe Gln  
 50 55 60

35 Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp Thr Leu Ser Glu  
 65 70 75 80

40 His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala Cys Asp Ile Cys  
 85 90 95

45 Gly Arg Lys Phe Ala Ala Arg Ser Thr Arg Thr Asn His Thr Lys Ile  
 100 105 110

50 His Leu Arg Gly Ser  
 115

50 <210> 163  
 <211> 117  
 <212> PRT  
 <213> Artificial

55 <220>  
 <223> nucleasa  
 <400> 163

ES 2 808 687 T3

5  
 10  
 15  
 20  
 25  
 30  
 35

Met Ala Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser  
 1 5 10 15

Arg Ser Asp Thr Leu Ser Glu His Ile Arg Thr His Thr Gly Glu Lys  
 20 25 30

Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Ala Arg Ser Thr  
 35 40 45

Arg Thr Thr His Thr Lys Ile His Thr Gly Ser Gln Lys Pro Phe Gln  
 50 55 60

Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp Ser Leu Ser Lys  
 65 70 75 80

His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala Cys Asp Ile Cys  
 85 90 95

Gly Arg Lys Phe Ala Gln Arg Ser Asn Leu Lys Val His Thr Lys Ile  
 100 105 110

His Leu Arg Gly Ser  
 115

40  
 45

<210> 164  
 <211> 13  
 <212> PRT  
 <213> Artificial

<220>  
 <223> Secuencia de inserción de IL-2Rgamma

<400> 164

50

Phe Arg Val Arg Ser Arg Phe Asn Pro Leu Cys Gly Ser  
 1 5 10

55  
 60  
 65

<210> 165  
 <211> 39  
 <212> **ADN**  
 <213> Artificial

<220>  
 <223> Secuencia de inserción de IL-2Rgamma

<400> 165  
 ttctgttc ggagccggtt taaccgctc tgtggaagt 39

<210> 166  
 <211> 23  
 <212> ADN  
 <213> Artificial

<220>  
 <223> Ex5\_1,5detF1  
 5 <400> 166  
 gattcaacca gacagataga agg 23  
 <210> 167  
 <211> 22  
 10 <212> ADN  
 <213> Artificial  
 <220>  
 <223> Ex5\_1,5detR1  
 15 <400> 167  
 ttactgtctc atcctttact cc 22  
 <210> 168  
 <211> 27  
 20 <212> ADN  
 <213> Artificial  
 <220>  
 25 <223> Exón 5 cebador directo  
 <400> 168  
 gctaaggcca agaaagtagg gctaaag 27  
 30 <210> 169  
 <211> 25  
 <212> ADN  
 <213> Artificial  
 35 <220>  
 <223> Cebador inverso de exón 5  
 <400> 169  
 40 ttcttccat caccaaacc tcttg 25  
 <210> 170  
 <211> 27  
 <212> ADN  
 <213> Artificial  
 45 <220>  
 <223> ex5\_1,5detF3  
 <400> 170  
 50 gctaaggcca agaaagtagg gctaaag 27  
 <210> 171  
 <211> 25  
 <212> ADN  
 55 <213> Artificial  
 <220>  
 <223> ex5\_1,5detR3  
 60 <400> 171  
 ttcttccat caccaaacc tcttg 25  
 <210> 172  
 <211> 42  
 65 <212> ADN  
 <213> Artificial

<220>  
 <223> Donante1  
 5 <400> 172  
 cttcagccgc tatccagacc acatgaaaca acacgacttc tt 42  
 <210> 173  
 <211> 42  
 10 <212> ADN  
 <213> Artificial  
 <220>  
 <223> Donante2  
 15 <400> 173  
 cttcagccgg tatccagacc acatgaaaca acatgacttc tt 42  
 <210> 174  
 <211> 42  
 20 <212> ADN  
 <213> Artificial  
 <220>  
 <223> Donante3  
 25 <400> 174  
 cttcagccgc taccagacc acatgaaaca gcacgacttc tt 42  
 <210> 175  
 <211> 42  
 30 <212> ADN  
 <213> Artificial  
 <220>  
 <223> Donante 5  
 35 <400> 175  
 cttcagccgc taccgacc gtccac 42  
 <210> 176  
 <211> 40  
 40 <212> ADN  
 <213> Artificial  
 45 <220>  
 <223> GFP mut  
 <400> 176  
 50 cttcagccgc taccctaac acgaagcagc acgacttctt 40  
 <210> 177  
 <211> 42  
 55 <212> ADN  
 <213> Artificial  
 <220>  
 <223> GFP wt  
 60 <400> 177  
 cttca gccgc taccgacc acatgaagca gcacgacttc tt 42  
 <210> 178  
 <211> 12  
 65 <212> ADN  
 <213> Artificial

<220>  
 <223> sca-36a  
 5 <400> 178  
 gaagtctgcc gt 12  
 <210> 179  
 <211> 13  
 10 <212> ADN  
 <213> Artificial  
 <220>  
 <223> sca-36b  
 15 <400> 179  
 gaagtctgcc gtt 13  
 <210> 180  
 <211> 13  
 20 <212> ADN  
 <213> Artificial  
 <220>  
 <223> sca-36c  
 25 <400> 180  
 gaagtctgcc gtt 13  
 30 <210> 181  
 <211> 52  
 <212> ADN  
 <213> Artificial  
 35 <220>  
 <223> gen de beta-globina humana  
 <400> 181  
 caaacagaca ccatggtgca tctgactcct ggggagaagt ctgccgttac tg 52  
 40 <210> 182  
 <211> 13  
 <212> ADN  
 <213> Artificial  
 45 <220>  
 <223> sca-r29b  
 <400> 182  
 50 acgtagactg agg 13  
 <210> 183  
 <211> 9  
 <212> PRT  
 55 <213> Artificial  
 <220>  
 <223> enlazador no canónico entre dedos  
 60 <400> 183  
 Thr Gly Gly Gly Gly Ser Gln Lys Pro  
 1 5  
 65

ES 2 808 687 T3

5  
<210> 184  
<211> 7  
<212> PRT  
<213> Artificial  
  
<220>  
<223> sca-r29b F1  
  
10  
<400> 184  
  
Gln Ser Gly Asp Leu Thr Arg  
1 5  
  
15  
<210> 185  
<211> 7  
<212> PRT  
<213> Artificial  
  
20  
<220>  
<223> sca-r29b F2  
  
25  
<400> 185  
  
Thr Ser Ala Asn Leu Ser Arg  
1 5  
  
30  
<210> 186  
<211> 7  
<212> PRT  
<213> Artificial  
  
35  
<220>  
<223> sca-r29b F3  
  
<400> 186  
  
40  
  
Asp Arg Ser Ala Leu Ser Arg  
1 5  
  
45  
<210> 187  
<211> 7  
<212> PRT  
<213> Artificial  
  
50  
<220>  
<223> sca-r29b F4  
  
<400> 187  
  
55  
  
Gln Ser Gly His Leu Ser Arg  
1 5  
  
60  
<210> 188  
<211> 7  
<212> PRT  
<213> Artificial  
  
65  
<220>

ES 2 808 687 T3

<223> sca-36a F1

<400> 188

5

Arg Ser Gln Thr Arg Lys Thr  
1 5

10

<210> 189  
<211> 7  
<212> PRT  
<213> Artificial

15

<220>  
<223> sca-36a F2

20

<400> 189

Gln Lys Arg Asn Arg Thr Lys  
1 5

25

<210> 190  
<211> 7  
<212> PRT  
<213> Artificial

30

<220>  
<223> sca-36a F3

35

<400> 190

Asp Arg Ser Ala Leu Ser Arg  
1 5

40

<210> 191  
<211> 7  
<212> PRT  
<213> Artificial

45

<220>  
<223> sca-36a F4

50

<400> 191

Gln Ser Gly Asn Leu Ala Arg  
1 5

55

<210> 192  
<211> 7  
<212> PRT  
<213> Artificial

60

<220>  
<223> sca-36b F1

65

<400> 192

Thr Ser Gly Ser Leu Ser Arg  
1 5

ES 2 808 687 T3

5 <210> 193  
<211> 7  
<212> PRT  
<213> Artificial

10 <220>  
<223> sca-36b F2  
  
<400> 193

15 **Asp Arg Ser Asp Leu Ser Arg**  
**1 5**

20 <210> 194  
<211> 7  
<212> PRT  
<213> Artificial

25 <220>  
<223> sca-36b F3  
  
<400> 194

30 **Asp Arg Ser Ala Leu Ser Arg**  
**1 5**

35 <210> 195  
<211> 7  
<212> PRT  
<213> Artificial

40 <220>  
<223> sca-36b F4  
  
<400> 195

45 **Gln Ser Gly Asn Leu Ala Arg**  
**1 5**

50 <210> 196  
<211> 7  
<212> PRT  
<213> Artificial

55 <220>  
<223> sca-36c F1  
  
<400> 196

60 **Thr Ser Ser Ser Leu Ser Arg**  
**1 5**

65

5 <210> 197  
 <211> 7  
 <212> PRT  
 <213> Artificial

10 <220>  
 <223> sca-36c F2  
 <400> 197

15 **Asp Arg Ser Asp Leu Ser Arg**  
 1 5

20 <210> 198  
 <211> 7  
 <212> PRT  
 <213> Artificial

25 <220>  
 <223> sca-36c F3  
 <400> 198

30 **Asp Arg Ser Ala Leu Ser Arg**  
 1 5

35 <210> 199  
 <211> 7  
 <212> PRT  
 <213> Artificial

40 <220>  
 <223> sca-36c F4  
 <400> 199

45 **Gln Ser Gly Asn Leu Ala Arg**  
 1 5

50 <210> 200  
 <211> 55  
 <212> ADN  
 <213> Secuencia artificial

55 <220>  
 <223> gen de la beta-globina humana dirigida por ZFNs  
 <400>200  
 ctgacacca tgggtcatct gactcctgtg gagaagtctg ccgttactgc cctag 55

60

65

**REIVINDICACIONES**

5 **1.** Un método in vitro para la escisión selectiva de un gen HLA clase I, un gen HLA que codifica una proteína de clase 1 del Complejo de Histocompatibilidad Mayor (MHC) o un gen HLA que codifica un gen de la subunidad MHC  $\beta$  (microglobulina  $\beta 2$ ), en una célula madre hematopoyética, comprendiendo el método el uso de una o más nucleasas dirigidas para crear una ruptura bicatenaria en el gen HLA de modo que el gen HLA esté inactivado, en donde la ruptura bicatenaria en el gen HLA es seguida por una unión de extremo no homólogo (NHEJ), y en donde la una o más nucleasas dirigidas es una proteína de fusión que comprende un dominio de unión de dedos de zinc diseñado y un dominio de escisión o medio dominio de escisión.

10 **2.** Una célula madre hematopoyética aislada producida por el método in vitro de la reivindicación 1, en donde el gen HLA está inactivo.

15

20

25

30

35

40

45

50

55

60

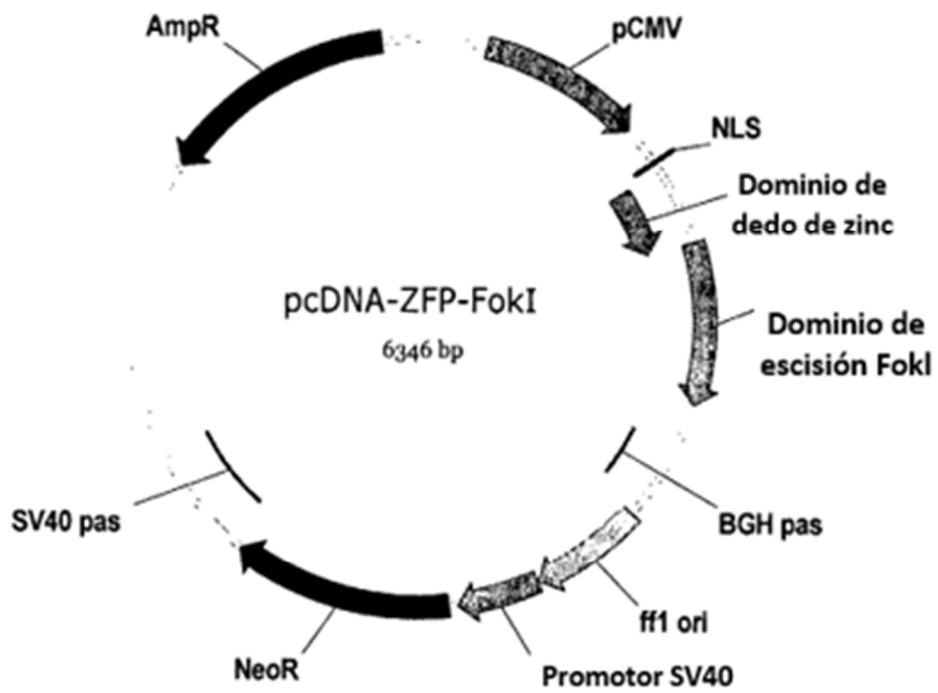
65

# FIGURA 1

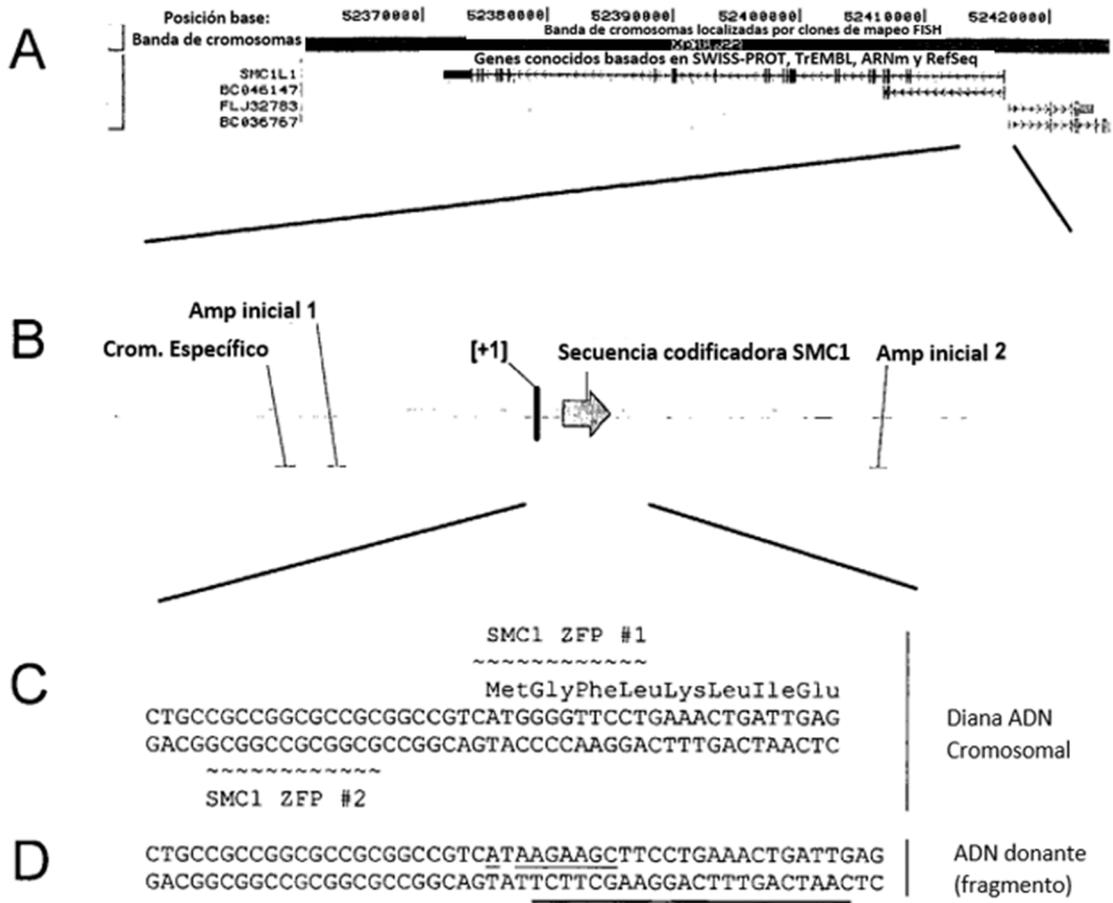
MetGlyPheLeuLysLeuIle (SEQ ID NO:2)  
CTGCCGCCGGCGCCGCGGCCGTCATGGGGTTCCTGAAACTGATT  
GACGGCGGCCGCGGCCGCGGCAGTACCCCAAGGACTTTGACTAA (SEQ ID NO:1)

5 10 15 20 25 30 35 40

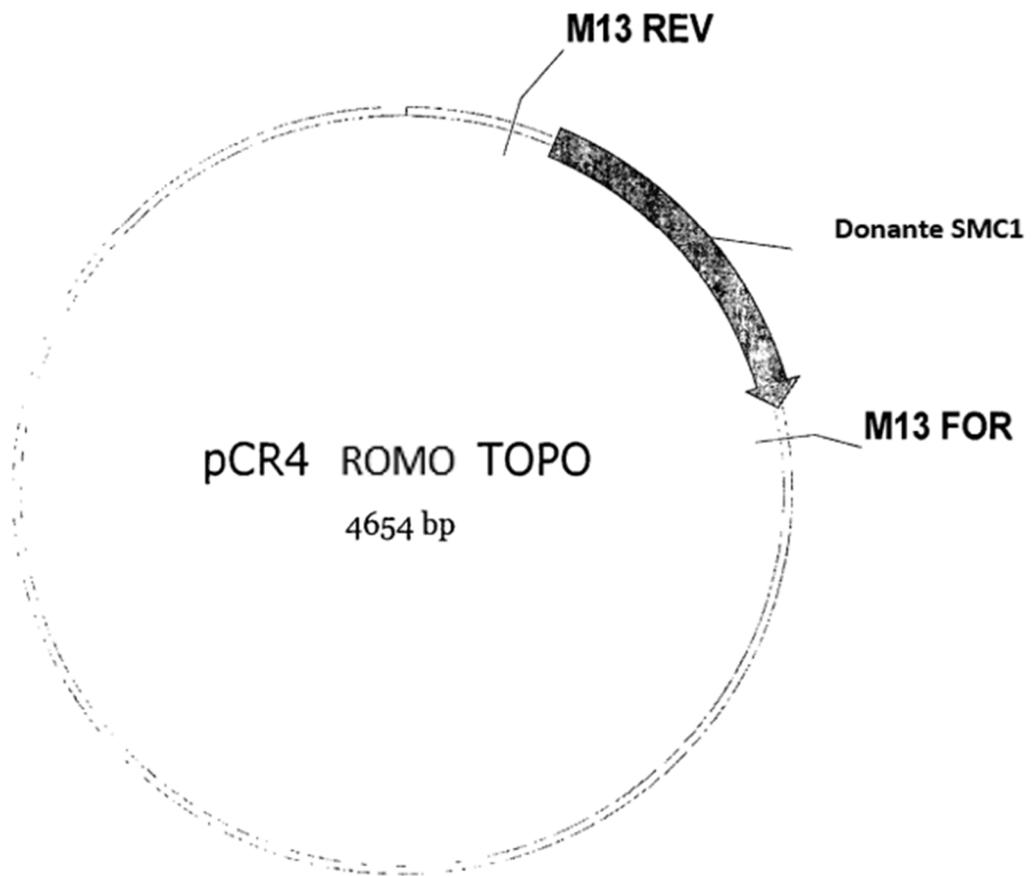
## FIGURA 2



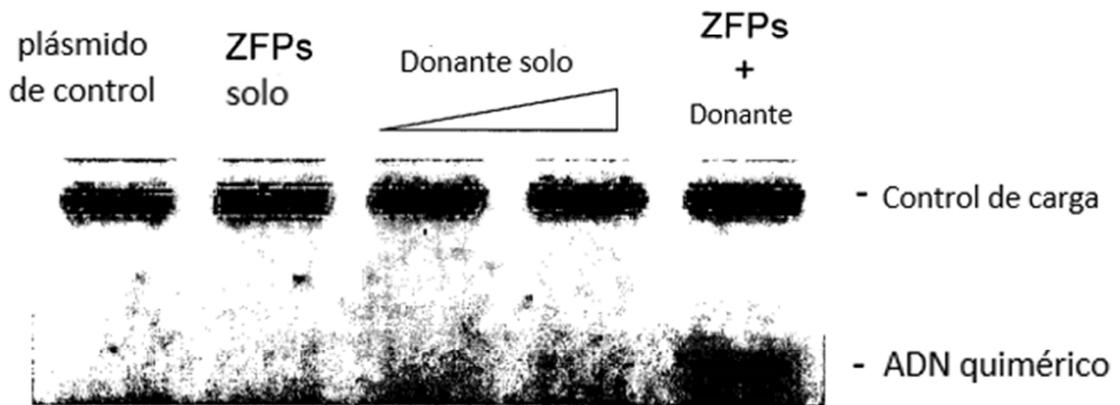
# FIGURA 3



# FIGURA 4



# FIGURA 5



## FIGURA 6

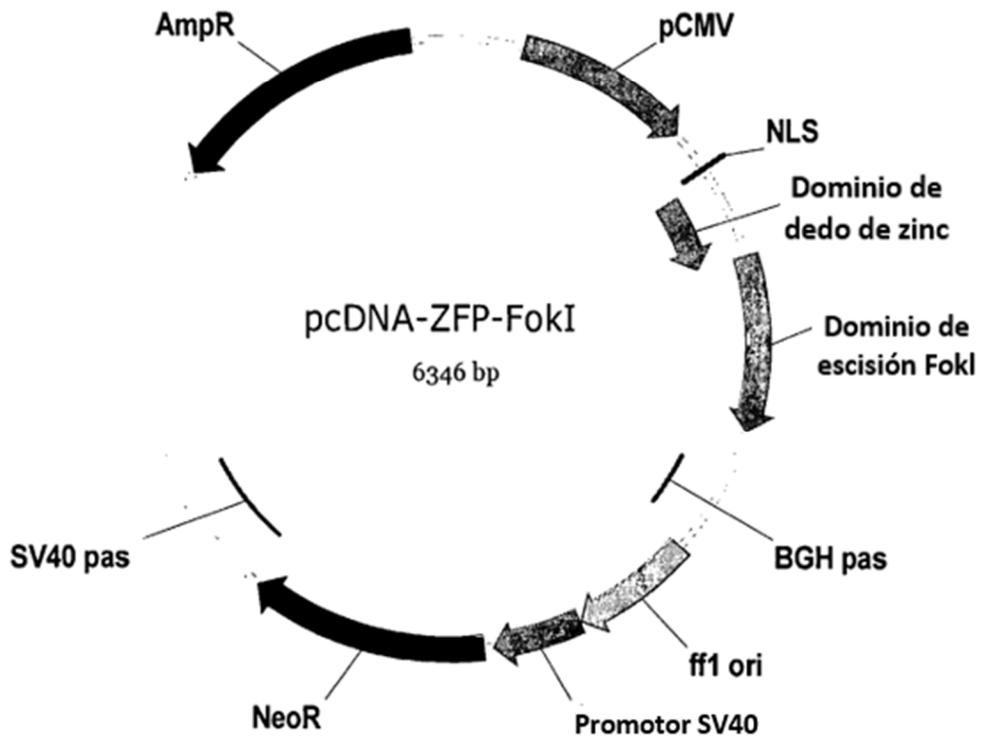
<u>TAGTCCTGCAGGTTTAAACGAATTCGCCCTTCTCAGCAAGCGTGAGCTCA</u>	50
<u>GGTCTCCCCCGCCTCCTTGAACCTCAAGAACTGCTCTGACTCCGCCCAGC</u>	100
<u>AACAACCTCCTCCGGGGATCTGGTCCGCAGGAGCAAGTGTTTGTTGTTGCC</u>	150
<u>ATGCAACAAGAAAAGGGGGCGGAGGCACCACGCCAGTCGTCAGCTCGCTC</u>	200
<u>CTCGTATACGCAACATCAGTCCCCGCCCTGGTCCCACCTCCTGCCGGAAG</u>	250
<u>GCGAAGATCCCGTTAGGCCTGGACGTATTCTCGCGACATTTGCCGGTCGC</u>	300
<u>CCGGCTTGCACTGCGGCGTTTCCCGCGCGGGCTACCTCAGTTCTCGGGCG</u>	350
<u>TACGGCGCGGCCTGTCCCTACTGCTGCCGGCGCCGCGGCCGTCAT<u>aagaag</u></u>	400
<u><u>c</u>TTCCTGAAACTGATTGAAGGGCGAATTCGCGGCCGCTAAATTCAATTCG</u>	450
<u>CCCTATAGTGAGT</u>	

(SEQ ID NO: 6)

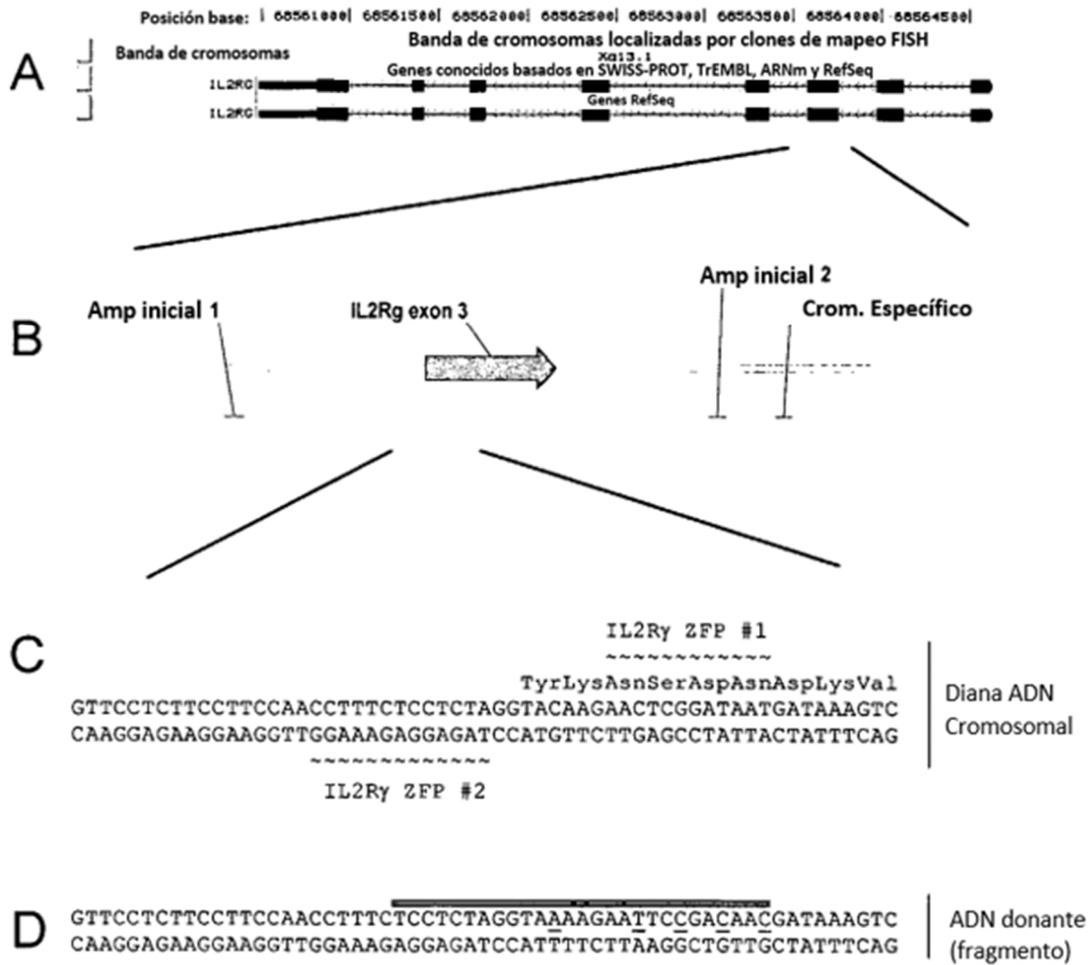
# FIGURA 7

NO: 8) TyrLysAsnSerAspAsnAspLysVal (SEQ ID  
CTTCCAACCTTTCTCCTCTAGGTACAAGAACTCGGATAATGATAAAAGTCC  
GAAGGTTGGAAAGAGGAGATCCATGTTCTTGAGCCTATTACTATTTTCAGG (SEQ ID  
NO: 7)  
| | | | | | | | | |  
5 10 15 20 25 30 35 40 45 50

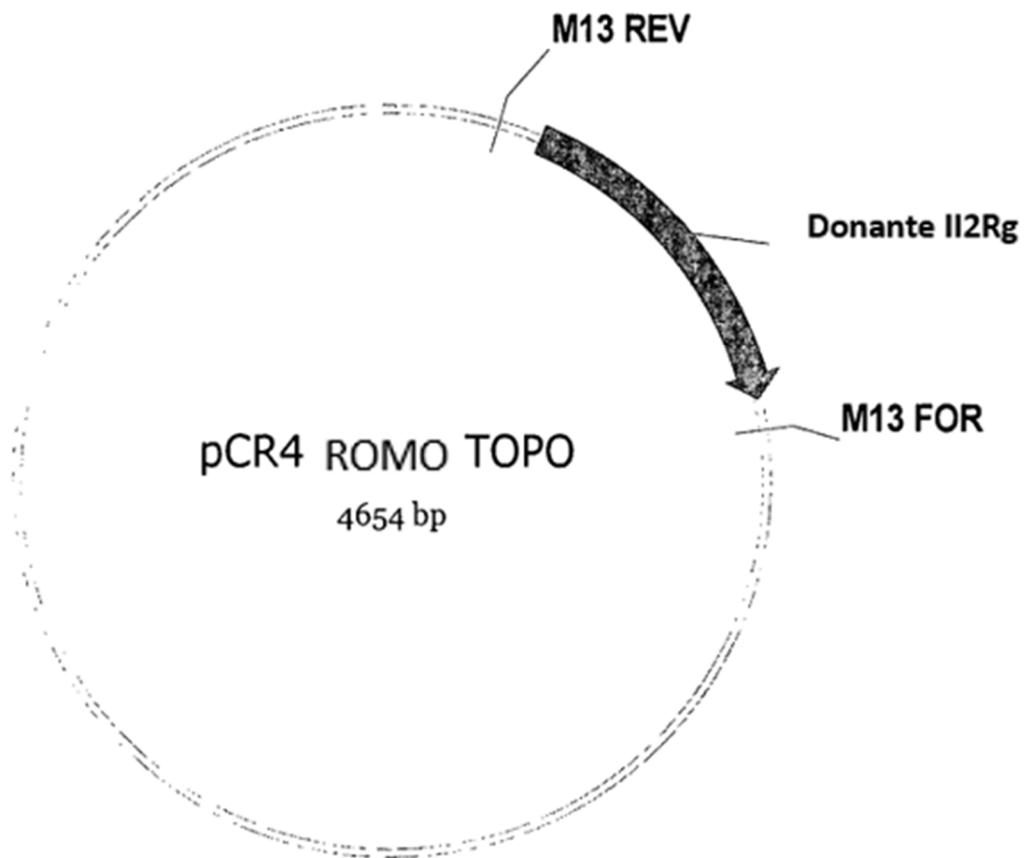
# FIGURA 8



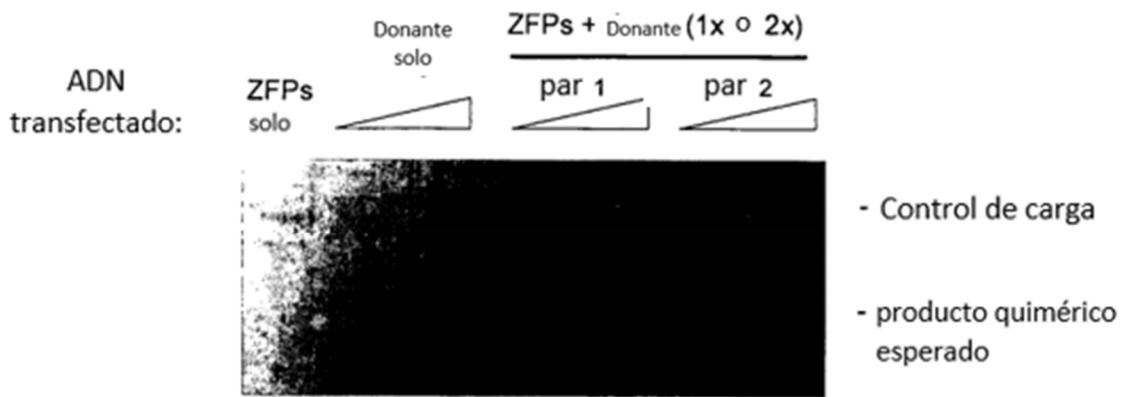
# FIGURA 9



# FIGURA 10



# FIGURA 11



**FIGURA 12**

<u>TAGTCCTGCAGGTTTAAACGAATTCGCCCTTTCCTCTAGGTAaAAGAAtT</u>	50
<u>CcGAcAAcGATAAAGTCCAGAAGTGCAGCCACTATCTATTCCCTGAAGAA</u>	100
ATCACTTCTGGCTGTCAGTTGCAAAAAAAGGAGATCCACCTCTACCAAAC	150
ATTTGTTGTTCACTCCAGGACCCACGGGAACCCAGGAGACAGGCCACAC	200
AGATGCTAAAACCTGCAGAATCTGGGTAATTTGGAAAGAAAGGGTCAAGAG	250
ACCAGGGATACTGTGGGACATTGGAGTCTACAGAGTAGTGTTCTTTTATC	300
ATAAGGGTACATGGGCAGAAAAGAGGAGGTAGGGGATCATGATGGGAAGG	350
GAGGAGGTATTAGGGGCACTACCTTCAGGATCCTGACTTGTCTAGGCCAG	400
GGGAATGACCACATATGCACACATATCTCCAGTGATCCCCTGGGCTCCAG	450
AGAACCTAACACTTCACAAACTGAGTGAATCCCAGCTAGAACTGAAGTGG	500
AACAACAGATTCTTGAACCACTGTTTGGAGCACTTGGTGCAGTACCGGAC	550
TAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGT	600
<u>ATTACAATTCACTGGCCGTCGTTT</u>	

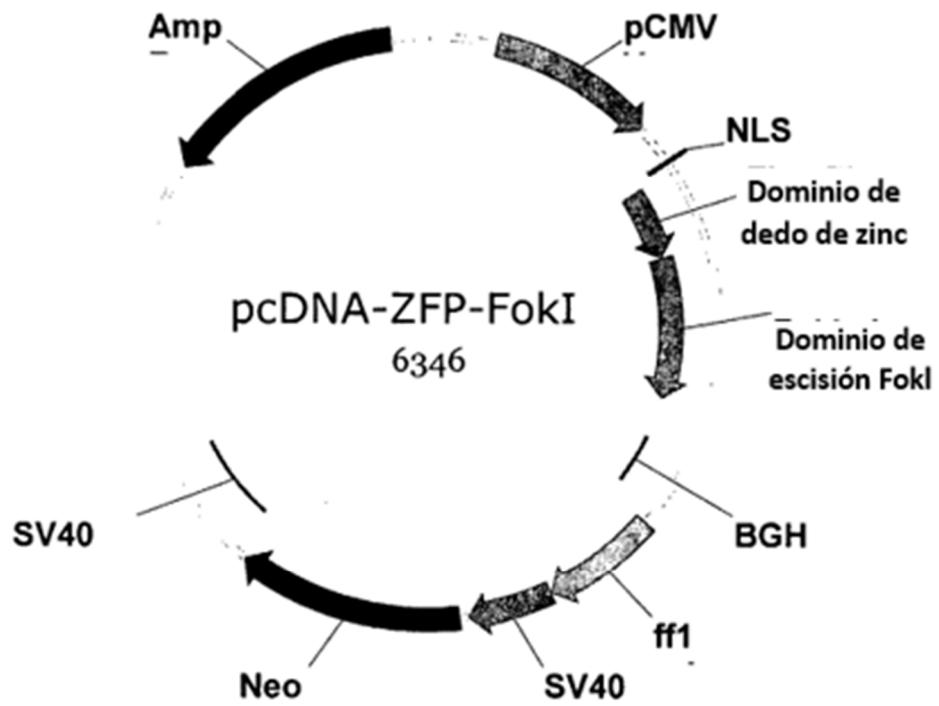
(SEQ ID NO:12)

## FIGURA 13

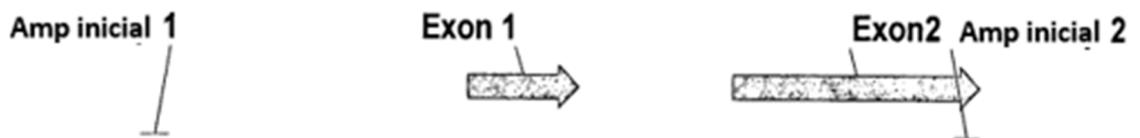
TACTGATGGTATGGGGCCAAGAGATATATCTTAGAGGGAGGGCTGAGGGT	50
TTGAAGTCCAACCTCCTAAGCCAGTGCCAGAAGAGCCAAGGACAGGTACGG	100
CTGTCATCACTTAGACCTCACCCCTGTGGAGCCACACCCTAGGGTTGGCCA	150
ATCTACTCCCAGGAGCAGGGAGGGCAGGAGCCAGGGCTGGGCATAAAAGT	200
CAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAACTGTGTTC	250
ACTAGCAACCTCAAACAGACACCCATGGTGCATCTGACTCCTG <u>AGG</u> GAGAAG	300
TCTG <u>CCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGG</u>	350
TGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGGTTTAAGGAGA	400
CCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTGGGGTTTCTGAT	450
AGGCACTGACTCTCTCTGCCTATTGGTCTATTTCCACCCCTTAGGCTGC	500
TGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTG	550
TCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAA	600
GAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCA	650
AGGGCACCTTTGCCCACTGAGTGAGCTGCACTGTGACAAGCTGCACGTG	700

(SEQ ID NO:13)

# FIGURA 14

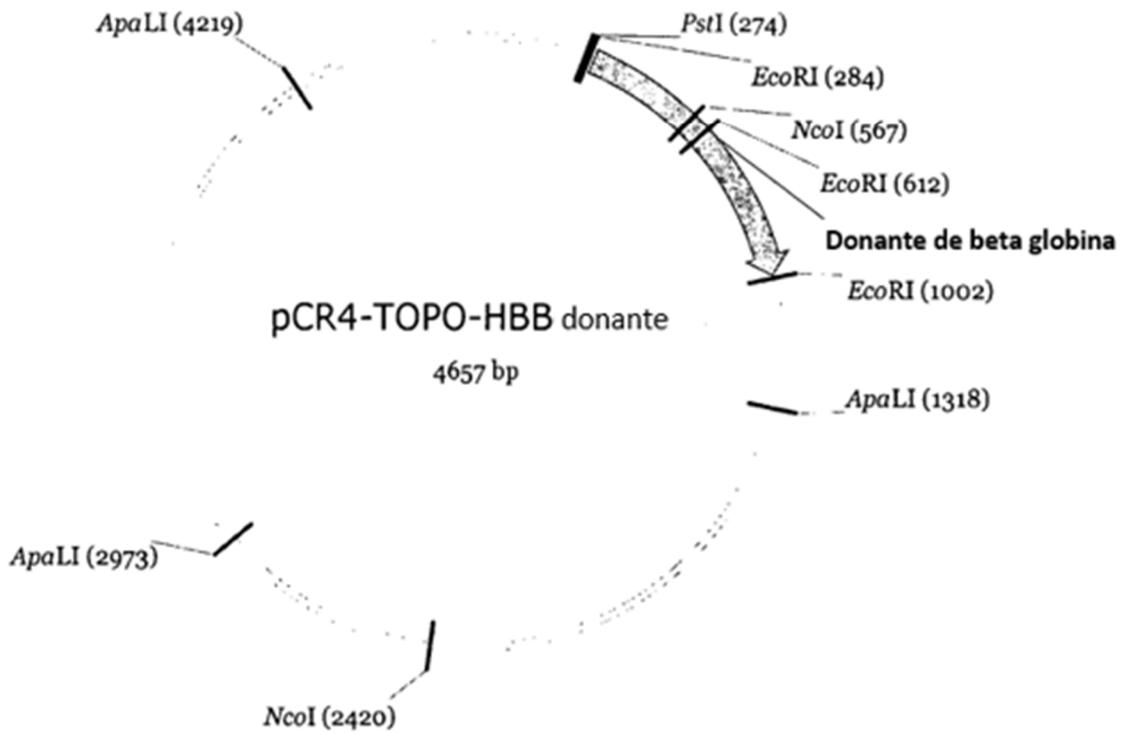


## FIGURA 15

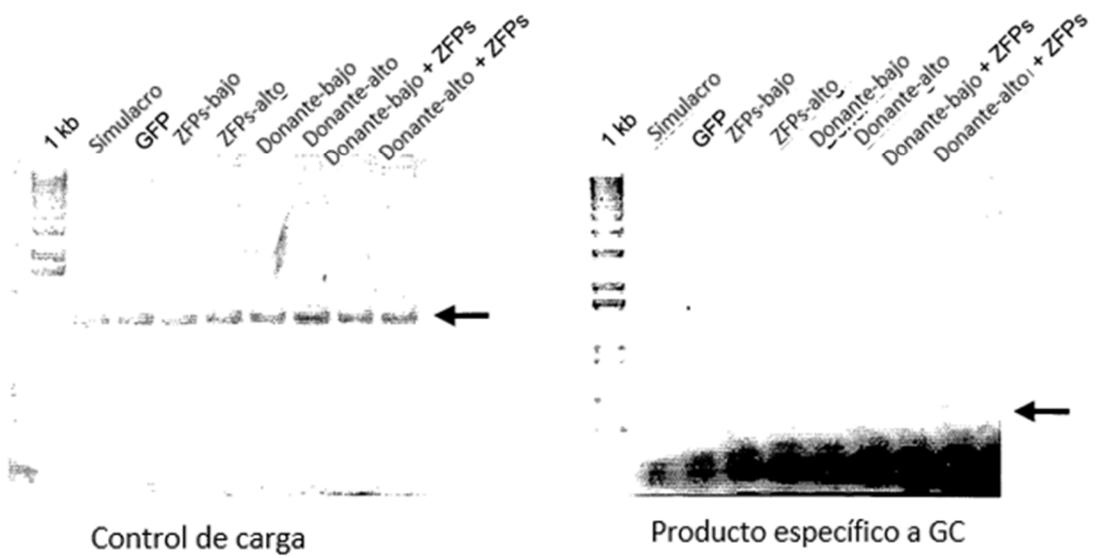


DIANA genómica humana de beta globina  
700 bp

# FIGURA 16



# FIGURA 17



## FIGURA 18

<u>TGCTTACCAAGCTGTGATTCCAAATATTACGTAAATACACTTGCAAAGGA</u>	50
<u>GGATGTTTTTAGTAGCAATTTGTA</u> CTGATGGTATGGGGCCAAGAGATATA	100
TCTTAGAGGGAGGGCTGAGGGTTTGAAGTCCAACCTCCTAAGCCAGTGCCA	150
GAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTCACCTGTGG	200
AGCCACACCCTAGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGG	250
AGCCAGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATT	300
TGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGT	350
GCATCTGACTCCTGAGGAGAAGTCTG <u>CGTTAgTGCCCGaattccgAtcG</u>	400
<u>TcAACcac</u>	

(SEQ ID NO:14)

# FIGURA 19

5-8  
CACGTTTCGTGTTTCGGAGCCGCTTTAACCCACTCTGTGGAAG  
GTGCAAAGCACAAGCCTCGGCGAAATTGGGTGAGACACCTTC  
5-10

(SEQ ID NO:15)

## FIGURA 20

1 MAPKKRKVGV IHGVPAAMAE RPFQCRICMR NFSRSDNLSE HIRTHTGEKP FACDICGRKF  
61 ARNAHRINHT KIHTGSQKPF QCRICMRNFS **RSDTLSE**HIR THTGEKPFAC DICGRKFAAR  
121 **STR**THTKIH LRQKDAARGS QLVKSELEEK KSELRHKLKY VPHEYIELIE IARNSTQDRI  
181 LEMKVMEFFM KVYGYRGKHL GGSRKPDGAI YTVGSPIDYG VIVDTKAYSG GYNLPIGQAD  
241 EMQRYVEENQ TRNKHINPNE WWKVYPSSVT EFKFLFVSGH FKGNYKAQLT RLNHITNCNG  
301 AVLSVEELLI GGEMIKAGTL TLEEVRRKFN NGEINF

(SEQ ID NO:16)

## FIGURA 21

```
1  MAPKKRKRKVG  IHGVPAAMAE  RPFQCRICMR  NFSRSDSLSR  HIRTHTGEKP  FACDICGRKF
61  ADSSNRKTHT  KIHTGGGGSQ  KPFQCRICMR  NFSRSDSLSV  HIRTHTGEKP  FACDICGRKF
121 ADRSNRITHT  KIHLRQKDAA  RGSQLVKSEL  EEKSELRHK  LKYVPHEYIE  LIETARNSTQ
181 DRILEMKVME  FFMKVYGYRG  KHLGGSRKPD  GAIYTVGSPI  DYGVIVDTKA  YSGGYNLPIG
241 QADEMQRYVE  ENQTRNKHIN  PNEWKVYPS  SVTEFKFLV  SGHFKNYKA  QLTRLNHITN
301 CNGAVLSVEE  LLIGGEMIKA  GTLTLEEVRR  KFNNGEINF
```

(SEQ ID NO:17)

## FIGURA 22

CGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGTCGCCACCATGGTGAGC  
AAGGGCGAGGAGCTGTTACCGGGGTGGTGCCATCCTGGTCGAGCTGGACGGCGACGTA  
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG  
ACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCTGCCCTGGCCCACCCTCGTGACC  
ACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGAC  
TTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGAC  
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC  
ATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAG  
TACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAG  
GTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTAC  
CAGCAGAACACCCCCATCGGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGC  
ACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAG  
TTCGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGC  
GACTCTAGATCATAATC

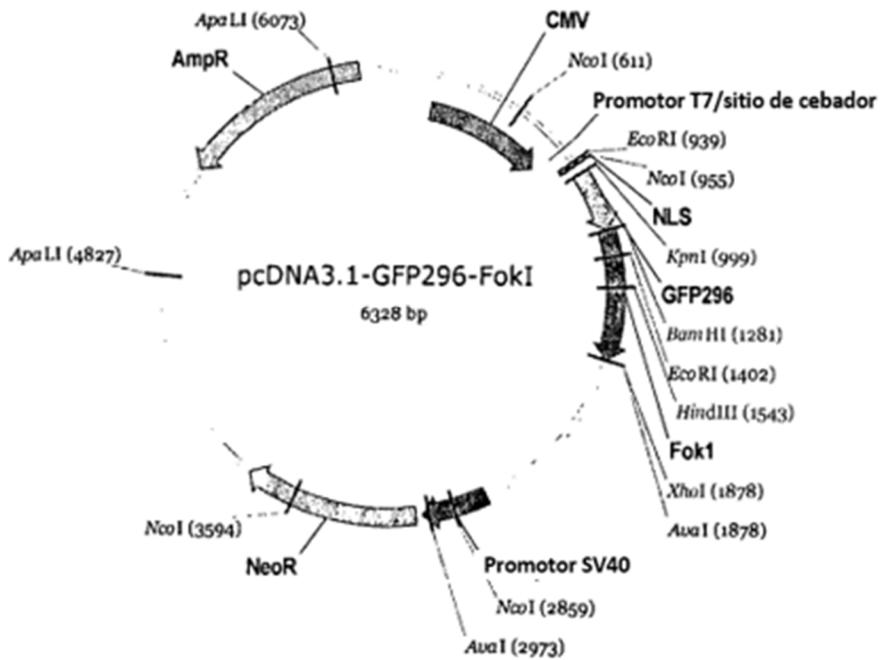
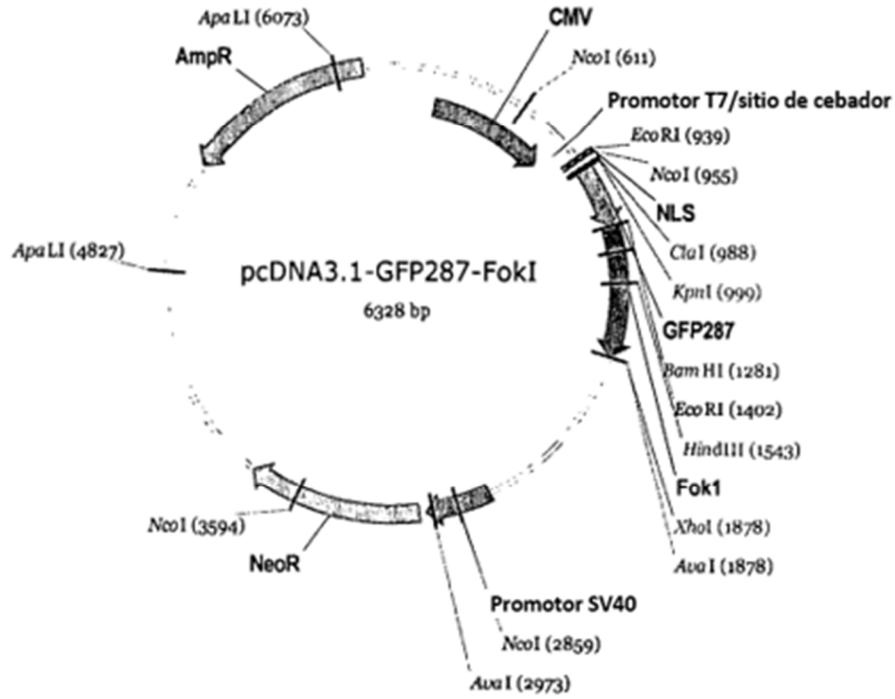
(SEQ ID NO:18)

## FIGURA 23

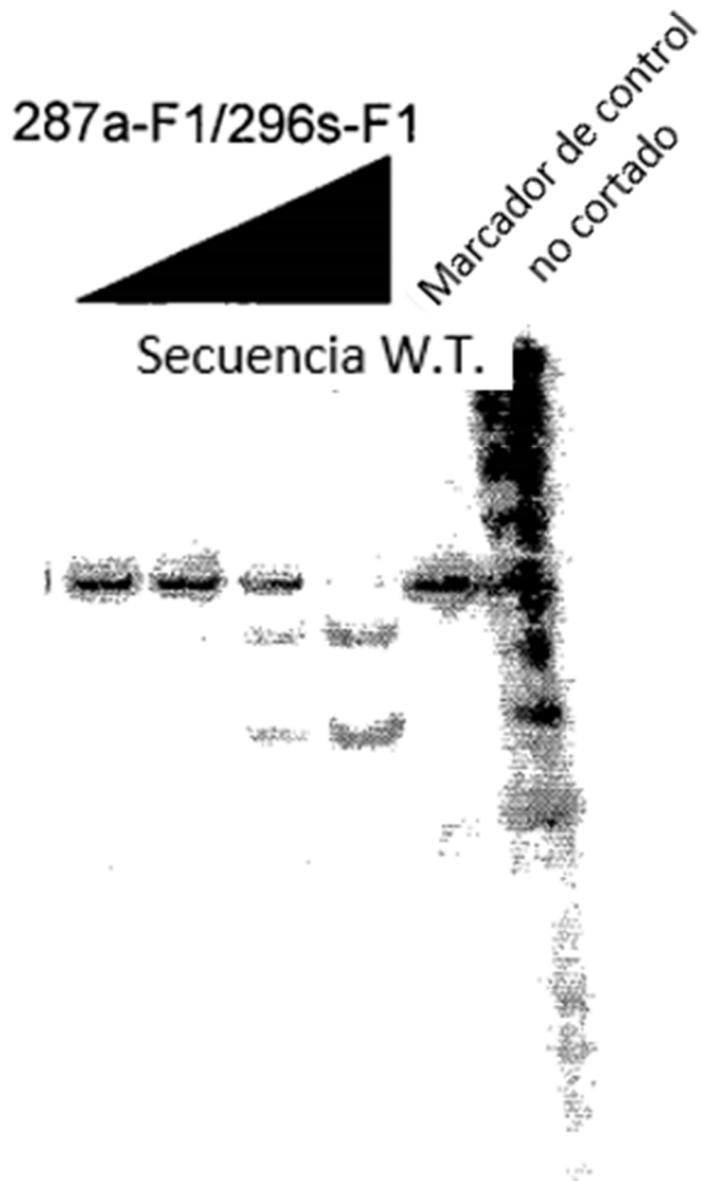
CGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGTCGCCACCATGGTGAGC  
AAGGGCGAGGAGCTGTTCCACCGGGTGGTGCCATCCTGGTCGAGCTGGACGGCGACGTA  
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG  
ACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACC  
ACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCTAACACGAAGCAGCAGCACTT  
CTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGA  
CGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCAT  
CGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTA  
CAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGT  
GAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCA  
GCAGAACACCCCATCGGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC  
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTT  
CGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGA  
CTCTAGATCATAATC

(SEQ ID NO:19)

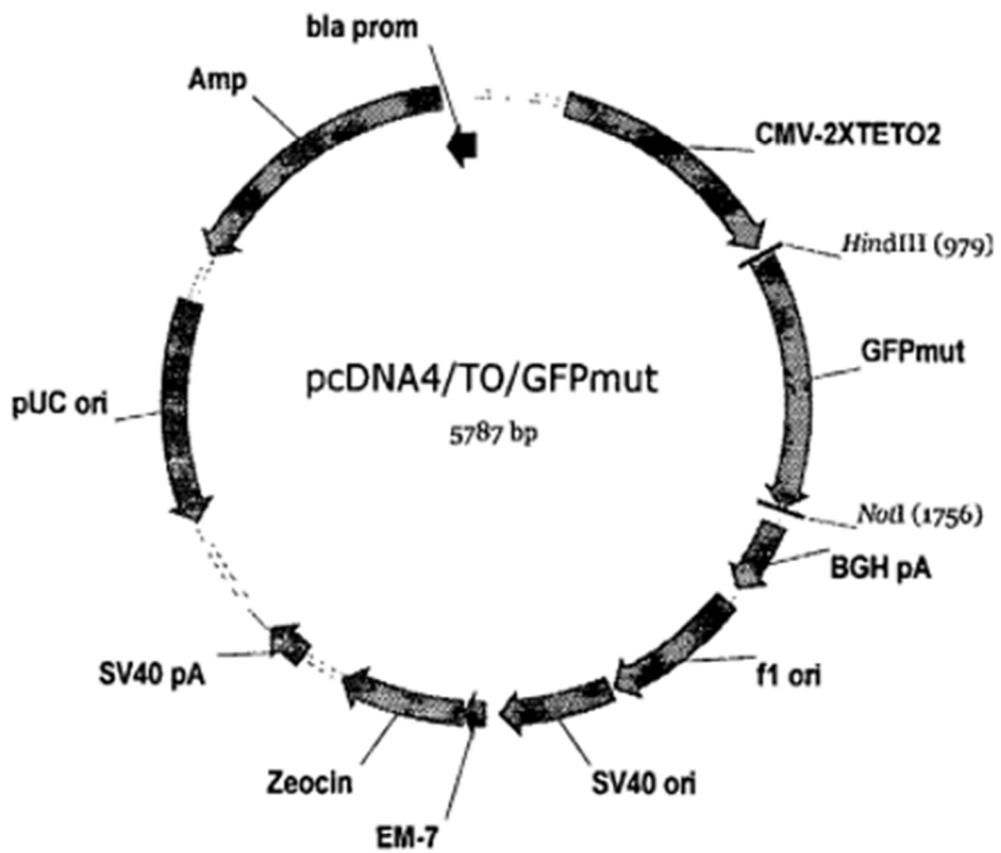
# FIGURA 24



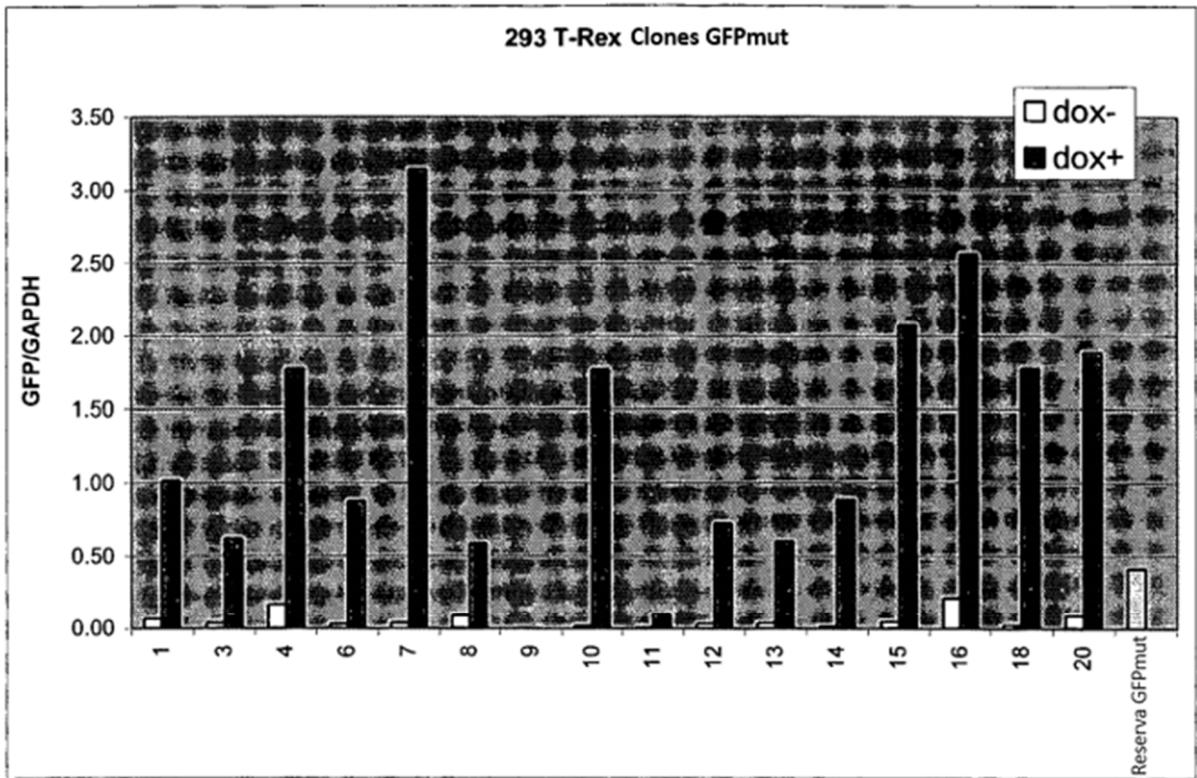
# FIGURA 25



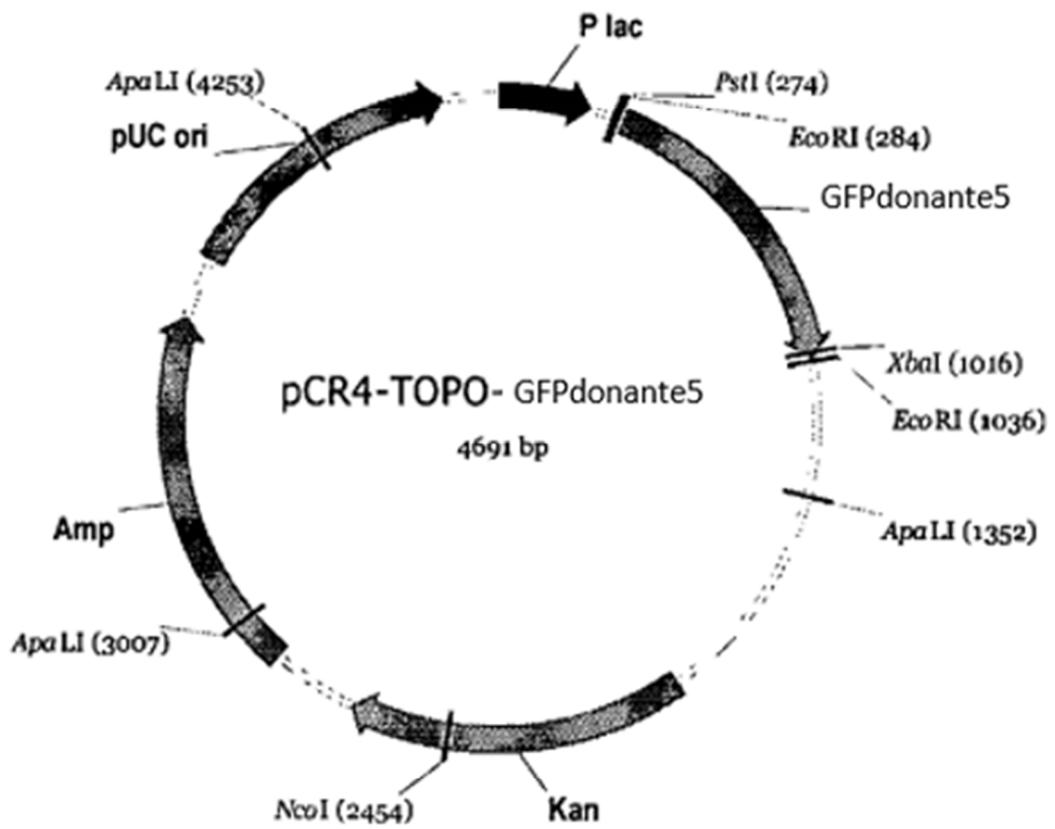
# FIGURA 26



**FIGURA 27**



# FIGURA 28

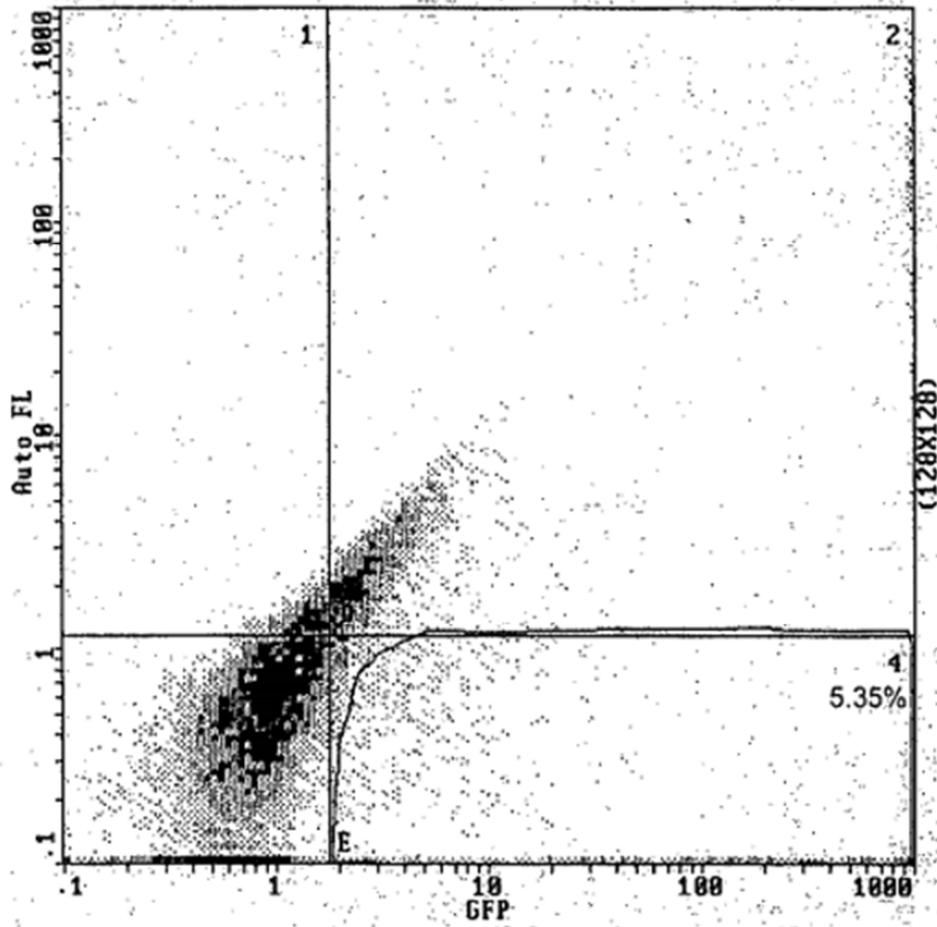


## FIGURA 29

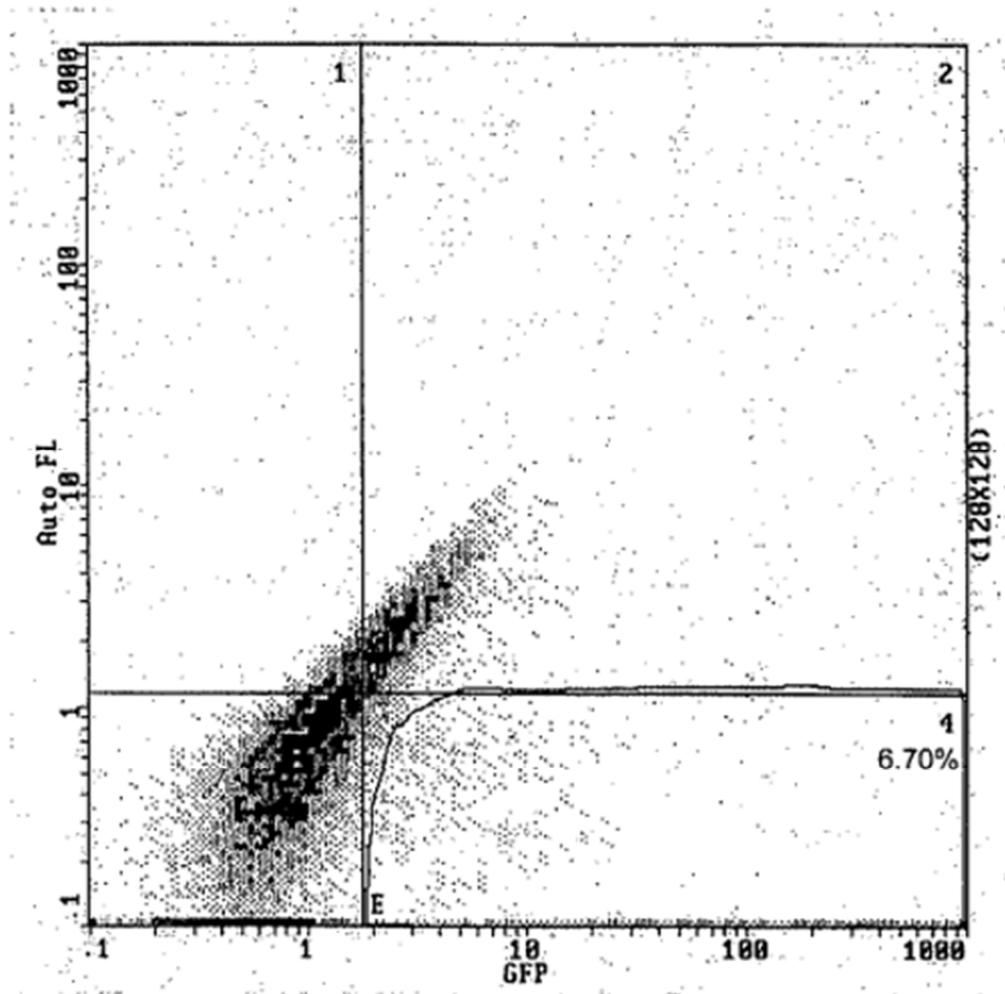
GGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGGCGACGTAAAC  
GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACC  
CTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCACCCTCGTGACCACC  
CTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTC  
TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGAC  
GGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATC  
GAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC  
AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTG  
AACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG  
CAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACC  
CAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTC  
GTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGAC  
TCTAGATCATAATC

(SEQ ID NO:20)

# FIGURA 30



# FIGURA 31

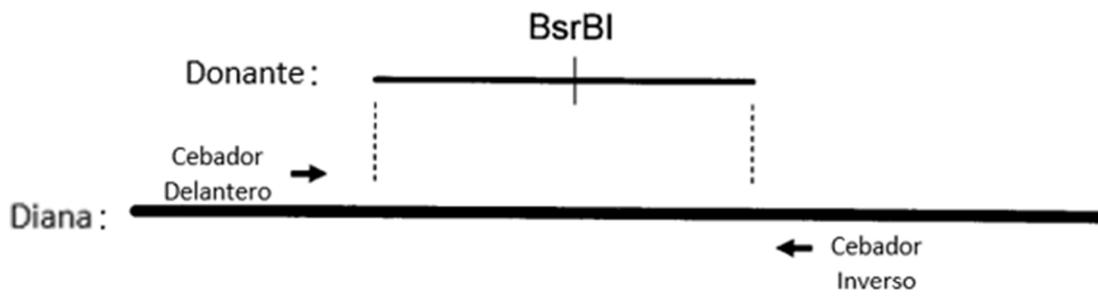


## FIGURA 32

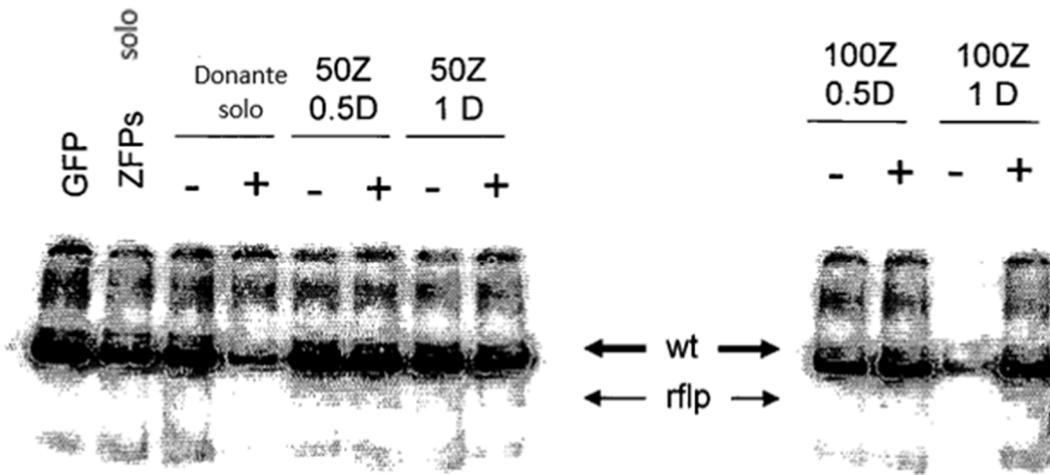
GGCGAGGAGCTGTTACACGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAAC  
GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACC  
CTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACCACC  
CTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTC  
TTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGAC  
GGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATC  
GAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC  
AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTG  
AACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACCAG  
CAGAACACCCCATCGGGCAGGGCCCGTGTCTGCTGCCCGACAACCACTACCTGAGCACC  
CAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTC  
GTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCTCG  
AGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGC  
CATCTGTTGTTTGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTG  
TCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTC  
TGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG  
CTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGG  
GGTATCCCCACGCGCCCTGTAGCGGCGCATTAAAGCGCGGGCGGGTGTGGTGGTTACGCGCA  
GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCGCTCCTTTCGCTTCTTCCCTTCCT  
TTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTAGGGT  
TCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGTGTTTAC  
GTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCT  
TTAATAGTGGACTCTTGTTCCAAACCTGGAACAACACTCAACCCTATCTCGGTCTATTCTT  
TTGATTTATAAGGGATTTTGGCGATTTGGCCTATTGGTTAAAAAATGAGCTGATTTAAC  
AAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTGAGTTAGGGTGTGGAAAGTCCCC  
AGGCTCCCCAGCAGGCAGAAGTATGCA

(SEQ ID NO:21)

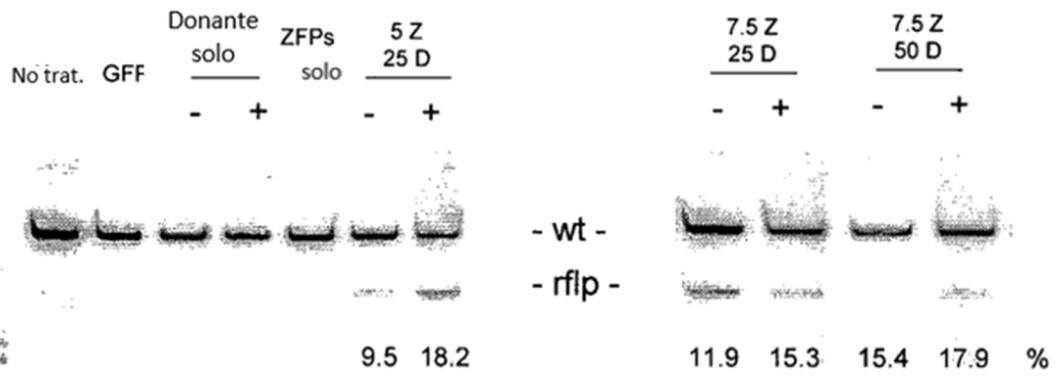
# FIGURA 33



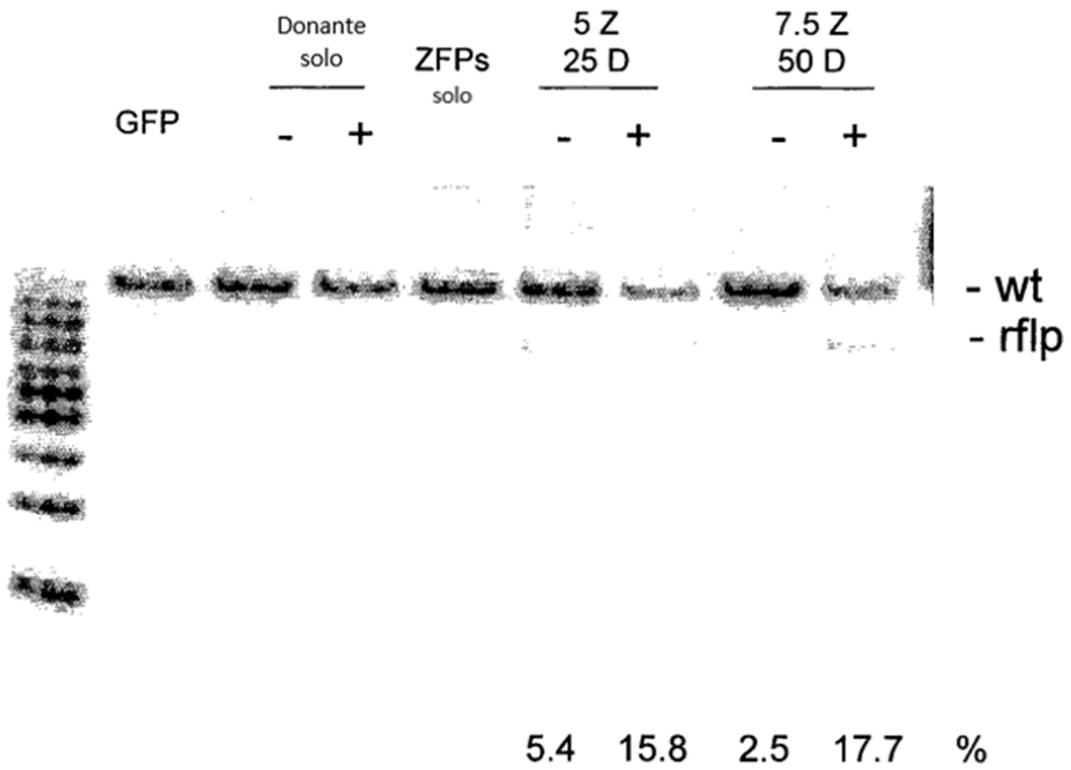
# FIGURA 34



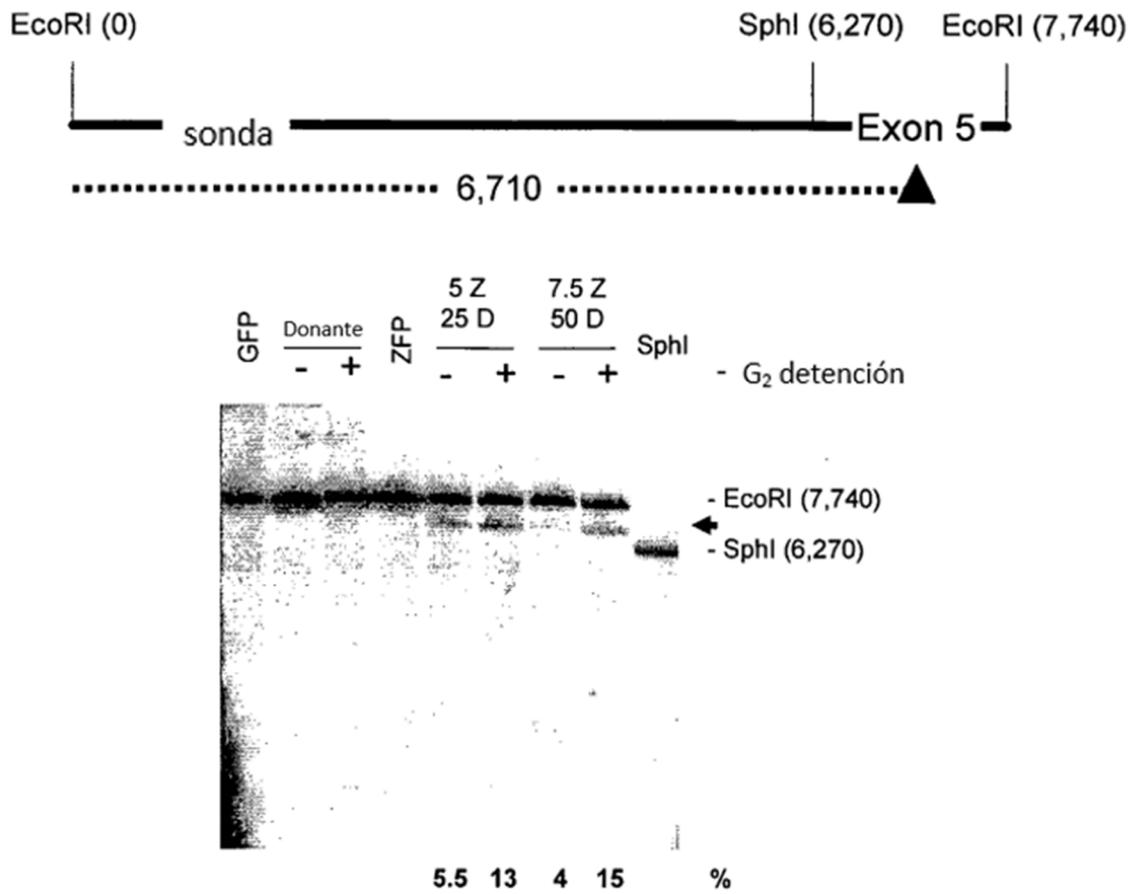
# FIGURA 35



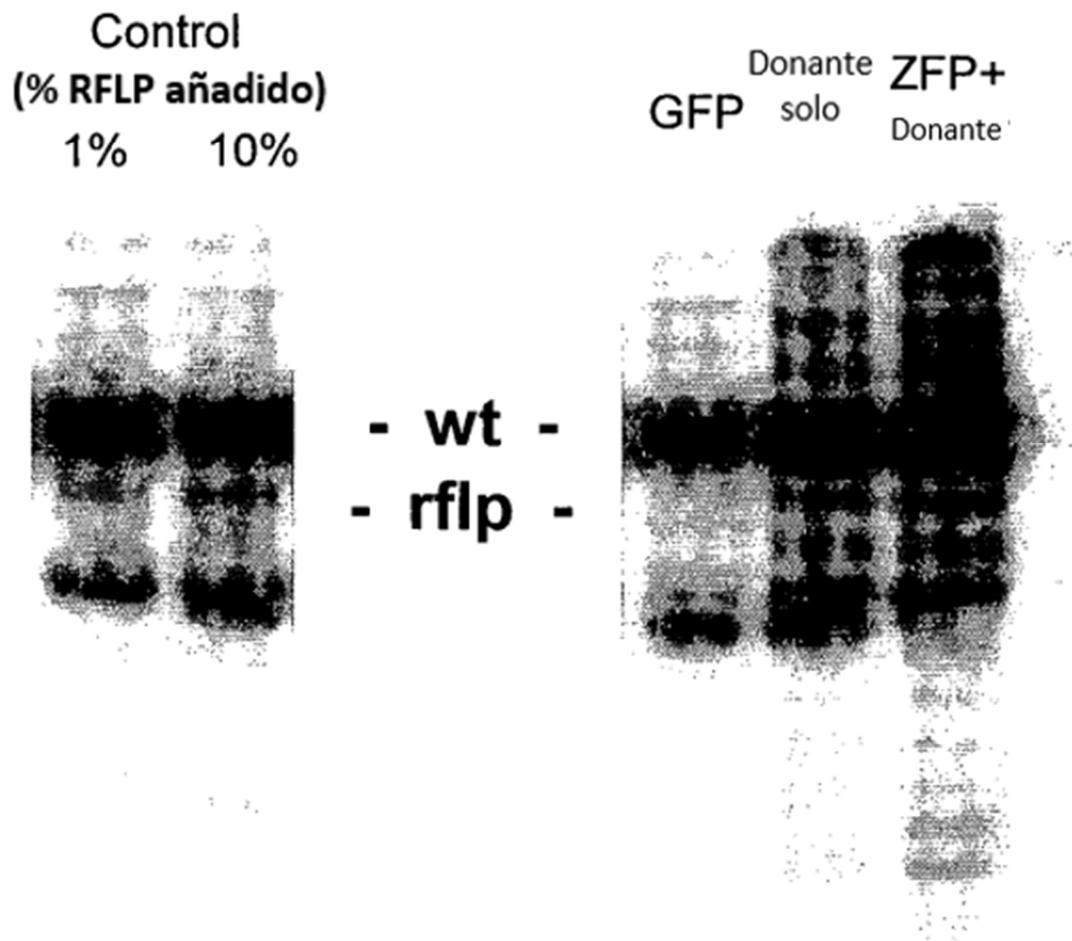
# FIGURA 36



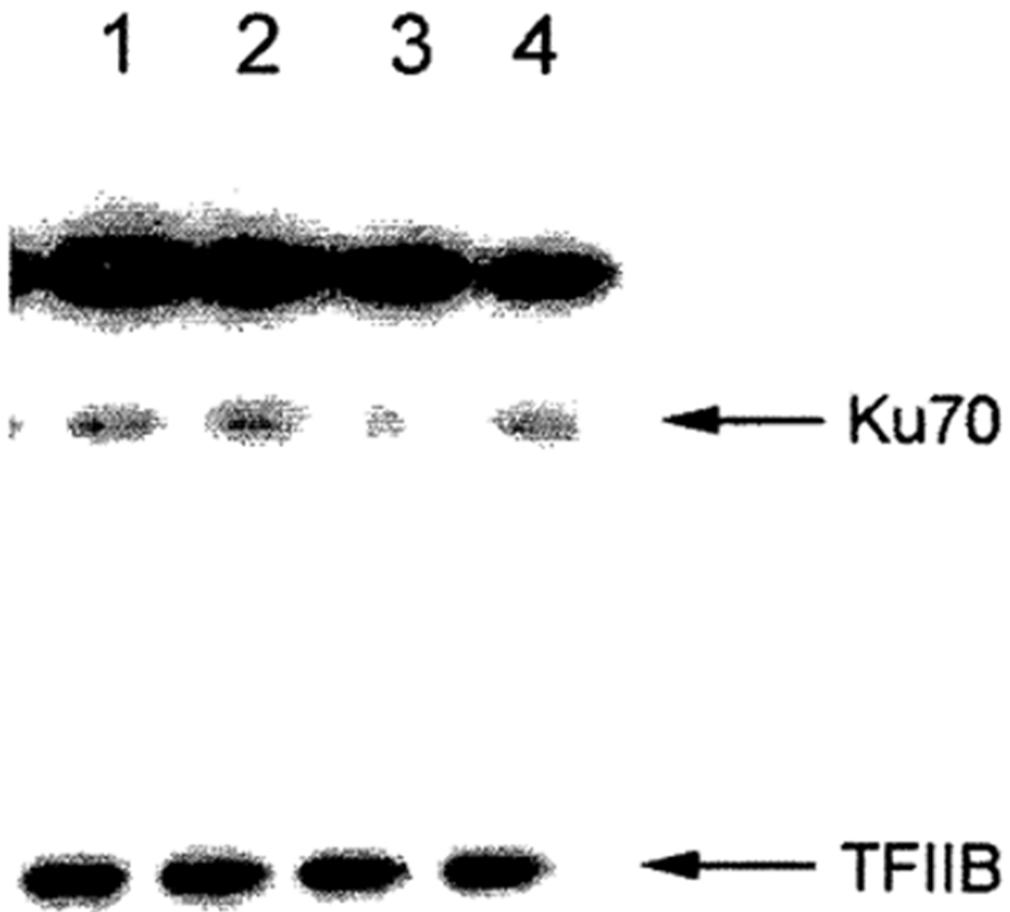
# FIGURA 37



# FIGURA 38



# FIGURA 39



## FIGURA 40

**sca-29b:**

MAERPFQCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFATSANLSRHTK  
IHTGGGGSQKPFQCRICMRNFSDRSALSRHIRTHTGEKPFACDICGRKFAQSG  
HLSRHTKIH (SEQ ID NO:22)

**sca-36a:**

MAERPFQCRICMRNFSRSQTRKTHIRTHTGEKPFACDICGRKFAQKRNRKHT  
KIHTGSQKPFQCRICMRNFSDRSALSRHIRTHTGEKPFACDICGRKFAQSGNLA  
RHTKIH (SEQ ID NO:23)

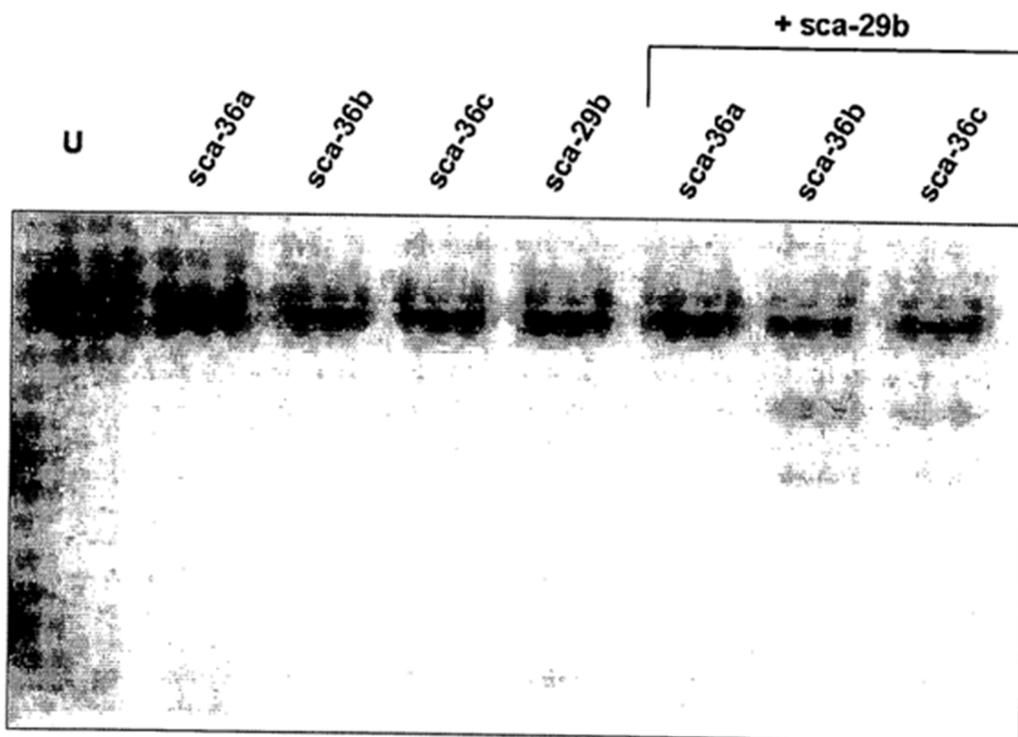
**sca-36b:**

MAERPFQCRICMRNFSTSGLSRHIRTHTGEKPFACDICGRKFADRSDLSRHTK  
IHTGGGGSQKPFQCRICMRNFSDRSALSRHIRTHTGEKPFACDICGRKFAQSG  
NLARHTKIH (SEQ ID NO:24)

**sca-36c:**

MAERPFQCRICMRNFSTSSSLSRHIRTHTGEKPFACDICGRKFADRSDLSRHTK  
IHTGGGGSQKPFQCRICMRNFSDRSALSRHIRTHTGEKPFACDICGRKFAQSG  
NLARHTKIH (SEQ ID NO:25)

# FIGURA 41



## FIGURA 42

