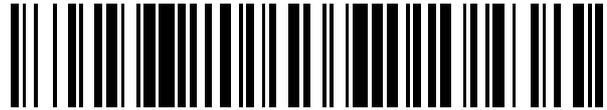


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 543 409**

51 Int. Cl.:

C07K 14/715 (2006.01)
C07K 14/435 (2006.01)
C12N 9/14 (2006.01)
C12N 15/85 (2006.01)
C12N 9/22 (2006.01)
C12N 15/62 (2006.01)
C07K 14/47 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **03.02.2005 E 05756438 (7)**

97 Fecha y número de publicación de la concesión europea: **06.05.2015 EP 1720995**

54 Título: **Métodos y composiciones para la escisión y recombinación dirigidas**

30 Prioridad:

05.02.2004 US 542780 P
26.03.2004 US 556831 P
01.06.2004 US 575919 P
06.08.2004 US 912932

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
19.08.2015

73 Titular/es:

SANGAMO BIOSCIENCES, INC. (100.0%)
POINT RICHMOND TECH CENTER, SUITE A100,
501 CANAL BOULEVARD
RICHMOND, CA 94804, US

72 Inventor/es:

MILLER, JEFFREY C. y
ZHANG, LEI

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 543 409 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).

Métodos y composiciones para la escisión y recombinación dirigidas

Descripción

5 CAMPO TÉCNICO

La presente divulgación está en los campos de la ingeniería del genoma, mutagénesis y recombinación homóloga dirigidas.

10 ANTECEDENTES

15 Un área importante de interés en la biología del genoma, especialmente en vista de la determinación de las secuencias de nucleótidos completas de varios genomas, es la alteración dirigida de secuencias del genoma. Pero para proporcionar un ejemplo, la anemia drepanocítica se produce por mutación de un único par de nucleótidos en el gen humano β -globina. Así, la capacidad para convertir la copia genómica endógena de este par de nucleótidos mutante en la secuencia no mutante en un forma estable y producir β -globina normal proporcionaría una cura para la anemia drepanocítica.

20 Se han hecho intentos por alterar secuencias genómicas en células cultivadas aprovechando el fenómeno natural de la recombinación homóloga. Véanse, por ejemplo, Capecchi (1989) Science 244:1288-1292; patentes de EE.UU. nº 6.528.313 y 6.528.314. Si un polinucleótido tiene homología suficiente con la región genómica que contiene la secuencia que va a alterarse, es posible que parte o toda la secuencia del polinucleótido sustituya la secuencia genómica por recombinación homóloga. Sin embargo, la frecuencia de recombinación homóloga bajo estas circunstancias es extremadamente baja. Además, la frecuencia de inserción del polinucleótido exógeno en localizaciones genómicas que carecen de homología de secuencias supera la frecuencia de recombinación homóloga en varios órdenes de magnitud.

25 Se ha mostrado que la introducción de una rotura bicatenaria en ADN genómico, en la región del genoma que lleva homología con un polinucleótido exógeno, estimula la recombinación homóloga en este sitio por varios cientos de veces en células cultivadas. Rouet et al., (1994) Mol. Cell. Biol. 14:8096-8106; Choulifa et al., (1995) Mol. Cell. Biol. 15:1968-1973; Donoho et al. (1998) Mol. Cell. Biol. 18:4070-4078. Véase también Johnson et al., (2001) Biochem. Soc. Trans. 29:196-201; y Yanez et al., (1998) Gene Therapy 5:149-159. En estos métodos, la escisión de ADN en la región genómica deseada se llevó a cabo insertando un sitio de reconocimiento para una meganucleasa (es decir, una endonucleasa cuya secuencia de reconocimiento es tan grande que no se produce, o se produce solo raramente, en el genoma de interés) en la región genómica deseada.

30 Sin embargo, la recombinación homóloga estimulada por la escisión de meganucleasas se basa en tanto la presencia fortuita de, como la inserción dirigida de, un sitio de reconocimiento de meganucleasa adecuado en la proximidad de la región genómica que va a alterarse. Como los sitios de reconocimiento de meganucleasa son raros (o inexistentes) en un genoma de mamífero típico, y la inserción de un sitio de reconocimiento de meganucleasa adecuado está plagado de las mismas dificultades que se han asociado a otras alteraciones genómicas, estos métodos no son ampliamente aplicables.

35 Bibikova et al., (Genetics 2002,161:1169-1175) desvelan la escisión cromosómica dirigida y mutagénesis en *Drosophila* usando nucleasas de dedos de cinc, en el que un dominio de escisión está localizado en el extremo C con respecto a un dominio de dedos de cinc.

40 Así, sigue existiendo la necesidad de composiciones y métodos para la alteración dirigida de secuencias en cualquier genoma.

50 RESUMEN

La presente invención se define por las reivindicaciones.

55 La presente divulgación proporciona composiciones y métodos para la escisión dirigida de cromatina celular en una región de interés y/o 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 el tratamiento en casos en los que las células y/o sus descendientes se devolverán al organismo después del tratamiento. Una región de interés en cromatina celular puede ser, por ejemplo, una secuencia genómica o porción de la misma. Las composiciones incluyen polipéptidos de fusión que comprenden un dominio de unión de dedos de cinc manipulado (por ejemplo, un dominio de unión de dedos de cinc que tiene una especificidad novedosa) y un dominio de escisión, y polipéptidos de fusión que comprenden un dominio de unión de dedos de cinc manipulado y un medio dominio de escisión. Los dominios de escisión y medios dominios de escisión pueden obtenerse, por ejemplo, de diversas endonucleasas de restricción y/o endonucleasas de asentamiento.

65 La cromatina celular puede estar presente en cualquier tipo de célula que incluye, pero no se limita a, células

procariotas y eucariotas, células fúngicas, células vegetales, células de animales, células de mamífero, células de primate y células humanas. La cromatina celular puede estar presente, por ejemplo, en cromosomas o en genomas intracelulares de bacterias o virus infectantes.

5 En un aspecto, se proporciona un método para la escisión de cromatina celular en una región de interés (por ejemplo, un método para la escisión dirigida de secuencias genómicas), comprendiendo el método (a) seleccionar la región de interés; (b) manipular un primer dominio de unión de dedos de cinc para unirse a una primera secuencia de nucleótidos en la región de interés; (c) proporcionar un segundo dominio de unión de dedos de cinc que se une a una segunda secuencia de nucleótidos en la región de interés, en el que la segunda secuencia se localiza entre 2 y
10 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 cinc 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 cinc y un segundo medio dominio de escisión; en el que la primera proteína de fusión se une a la primera secuencia de nucleótidos y la segunda proteína de fusión se une a la segunda secuencia de
15 nucleótidos, y adicionalmente en el que dicha unión posiciona los medios dominios de escisión de forma que la cromatina celular se escinda en la región de interés. La escisión puede producirse entre la primera y segunda secuencias de nucleótidos y, en ciertas realizaciones, el segundo dominio de unión de dedos de cinc se manipula para unirse a la segunda secuencia de nucleótidos.

20 Los medios dominios de escisión en la primera y segunda proteínas de fusión pueden ser de la misma endonucleasa o de endonucleasas diferentes. En ciertas realizaciones, la endonucleasa es una endonucleasa de restricción de tipo IIS. En realizaciones adicionales, la endonucleasa de restricción de tipo IIS es *Fok I*.

25 La polaridad de las proteínas de fusión puede ser de forma que el dominio de unión de dedos de cinc esté en el extremo N con respecto al medio dominio de escisión; alternatively, el medio dominio de escisión puede estar en el extremo N con respecto al dominio de unión de dedos de cinc. Cuando se usan dos proteínas de fusión de la misma polaridad, sus sitios de unión están en hebras opuestas del ADN en la región de interés.

30 En realizaciones adicionales, se usan dos proteínas de fusión de polaridad opuesta. En este caso, los sitios de unión para las dos proteínas están sobre la misma hebra de ADN.

El sitio de escisión está generalmente localizado entre los sitios de unión de las dos proteínas de fusión. Puede estar separado del borde próximo del uno o el otro de los sitios de unión por 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o
35 más nucleótidos.

Una proteína de fusión puede expresarse en una célula, por ejemplo, administrando la proteína de fusión a la célula o administrando un polinucleótido que codifica la proteína de fusión a una célula, en la que el polinucleótido, si es ADN, se transcribe, y una molécula de ARN administrada a la célula o un transcrito de una molécula de ADN administrado a la célula se traduce, para generar la proteína de fusión. Métodos para la administración de polinucleótidos y polipéptidos a células se conocen en la técnica y se presentan en cualquier parte en la presente divulgación.
40

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

50 En realizaciones adicionales, la escisión dirigida de cromatina celular en una región de interés se logra expresando dos proteínas de fusión en una célula, comprendiendo cada proteína de fusión un dominio de unión de dedos de cinc y un medio dominio de escisión. Uno o ambos de los dominios de unión de dedos de cinc de las proteínas de fusión pueden manipularse para unirse a una secuencia diana en la proximidad del sitio de escisión previsto. Si la expresión de las proteínas de fusión es por la administración de polinucleótidos, cada una de las dos proteínas de fusión puede codificarse por un polinucleótido separado, o un único polinucleótido puede codificar ambas proteínas de fusión.

55 Por consiguiente, un método para escindir cromatina celular en una región de interés puede comprender (a) seleccionar una primera secuencia en la región de interés; (b) manipular un primer dominio de unión de dedos de cinc 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 cinc 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 cinc y un segundo medio dominio de escisión, en el que la primera proteína de fusión se une a la primera secuencia, y la segunda proteína de fusión se une a una segunda secuencia localizada entre 2 y
60 50 nucleótidos de la primera secuencia, posicionándose así los medios dominios de escisión de forma que la cromatina celular se escinda en la región de interés.

65 En ciertas realizaciones, la unión de la primera y segunda proteínas de fusión posiciona los medios dominios de escisión de forma que se reconstituya un dominio de escisión funcional.

En ciertas realizaciones, el segundo dominio de unión de dedos de cinc se manipula para unirse a la segunda secuencia. En realizaciones adicionales, el primer y segundo medios dominios de escisión se derivan de la misma endonucleasa, que puede ser, por ejemplo, una endonucleasa de restricción (por ejemplo, una endonucleasa de restricción de tipo IIS tal como *Fok I*) o una endonucleasa de asentamiento.

5 En otras realizaciones, cualquiera de los métodos descritos en el presente documento puede comprender (a) seleccionar primera y segunda secuencias en una región de interés, en el que la primera y segunda secuencias están separadas entre 2 y 50 nucleótidos; (b) manipular un primer dominio de unión de dedos de cinc para unirse a la primera secuencia; (c) manipular un segundo dominio de unión de dedos de cinc para unirse a la segunda
10 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 cinc manipulado 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 cinc manipulado y un segundo medio dominio de escisión; en el que 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 el
15 primer y segundo medios dominios de escisión de forma que la cromatina celular se escinda en la región de interés.

En ciertas realizaciones, el primer y segundo medios dominios de escisión se derivan de la misma endonucleasa, por ejemplo, una endonucleasa de restricción de tipo IIS, por ejemplo, *Fok I*. En realizaciones adicionales, la cromatina celular se escinde en uno o más sitios entre la primera y segunda secuencias a las que se unen las
20 proteínas de fusión.

En realizaciones adicionales, un método para la escisión de cromatina celular en una región de interés comprende (a) seleccionar la región de interés; (b) manipular un primer dominio de unión de dedos de cinc para unirse a una primera secuencia en la región de interés; (c) proporcionar un segundo dominio de unión de dedos de cinc que se
25 une a una segunda secuencia en la región de interés, en el que la segunda secuencia se localiza 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 cinc 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 cinc y un segundo medio dominio de escisión; en el que 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
30 medios dominios de escisión de forma que la cromatina celular se escinda en la región de interés.

En cualquiera de los métodos descritos en el presente documento, el primer y segundo medios dominios de escisión pueden derivarse de la misma endonucleasa o de endonucleasas diferentes. En realizaciones adicionales, el
35 segundo dominio de unión de dedos de cinc se manipula para unirse a la segunda secuencia.

Si uno o más polinucleótidos que codifican las proteínas de fusión se introducen en la célula, un método a modo de ejemplo para la escisión dirigida de cromatina celular en una región de interés comprende (a) seleccionar la región de interés; (b) manipular un primer dominio de unión de dedos de cinc para unirse a una primera secuencia en la
40 región de interés; (c) proporcionar un segundo dominio de unión de dedos de cinc que se une a una segunda secuencia en la región de interés, en el que la segunda secuencia se localiza 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 cinc 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 cinc y un segundo medio dominio de escisión; en el que la primera y segunda proteínas de fusión 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, posicionando así los medios dominios de escisión de forma que la cromatina celular se escinda 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.
50

Para cualquiera de los métodos anteriormente mencionados, la cromatina celular puede estar en un cromosoma, episoma o genoma organular. Además, en cualquiera de los métodos descritos en el presente documento, al menos un dominio de unión de dedos de cinc se manipula, por ejemplo, por diseño o métodos de selección.

55 Similarmente, para cualquiera de los métodos anteriormente mencionados, el medio dominio de escisión puede derivarse de, por ejemplo, una endonucleasa de asentamiento o una endonucleasa de restricción, por ejemplo, una endonucleasa de restricción de tipo IIS. Una endonucleasa de restricción de tipo IIS a modo de ejemplo es *Fok I*.

60 Para cualquiera de los métodos de escisión dirigida, mutagénesis dirigida y/o recombinación dirigida desvelados en el presente documento que utilizan proteínas de fusión que comprenden un medio dominio de escisión, los bordes próximos de los sitios de unión de las proteínas de fusión pueden separarse por 5 ó 6 pares de bases.

En ciertas realizaciones, es posible obtener elevada especificidad de escisión utilizando proteínas de fusión en las que uno o ambos dominios de escisión contienen una alteración en la secuencia de aminoácidos de la interfase de dimerización.
65

La mutagénesis dirigida de una región de interés en cromatina celular puede producirse cuando un evento de escisión dirigida, como se ha descrito anteriormente, va seguido de unión de extremos no homólogos (NHEJ). Por consiguiente, se proporcionan métodos para la alteración de una primera secuencia de nucleótidos en una región de interés en cromatina celular, en los que los métodos comprenden las etapas de (a) manipular un primer dominio de unión de dedos de cinc para unirse a una segunda secuencia de nucleótidos en la región de interés, en el que la segunda secuencia comprende al menos 9 nucleótidos; (b) proporcionar un segundo dominio de unión de dedos de cinc para unirse a una tercera secuencia de nucleótidos, en el que la tercera secuencia comprende al menos 9 nucleótidos y se localiza 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 cinc 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 cinc y un segundo medio dominio de escisión; en el que la primera proteína de fusión se une a la segunda secuencia, y la segunda proteína de fusión se une a la tercera secuencia, posicionando así los medios dominios de escisión de forma que la cromatina celular se escinda en la región de interés y el sitio de escisión esté sometido a unión de extremos no homólogos.

Las mutaciones dirigidas resultantes del método anteriormente mencionado incluyen, pero no se limitan a, mutaciones puntuales (es decir, conversión de un único par de bases en un par de bases diferente), sustituciones (es decir, conversión de una pluralidad de pares de bases en una secuencia diferente de longitud idéntica), inserciones de uno o más pares de bases, deleciones de uno o más pares de bases y cualquier combinación de las alteraciones de secuencias anteriormente mencionadas.

También se proporcionan métodos para la recombinación dirigida (para, por ejemplo, alteración o sustitución de una secuencia en un cromosoma o una región de interés en cromatina celular). Por ejemplo, una secuencia genómica mutante puede sustituirse con una secuencia no mutante, por ejemplo, para el tratamiento de enfermedad genética o trastornos heredados. Además, una secuencia no mutante genómica puede sustituirse con una secuencia mutante, por ejemplo, para prevenir la función de un producto oncogénico o un producto de un gen que participa en una respuesta inflamatoria inapropiada. Además, un alelo de un gen puede sustituirse con un alelo diferente.

En tales métodos, una o más nucleasas dirigidas crean una rotura bicatenaria en cromatina celular en un sitio predeterminado, y un polinucleótido donante, que tiene homología con la secuencia de nucleótidos de la cromatina celular en la región de la rotura, se introduce en la célula. Los procesos de reparación de ADN celular se activan por la presencia de la rotura bicatenaria y el polinucleótido donante se usa como molde para la reparación de la rotura, produciendo la introducción de toda o parte de la secuencia de nucleótidos del donante en la cromatina celular. Así una primera secuencia en cromatina celular puede alterarse y, en ciertas realizaciones, puede convertirse en una secuencia presente en un polinucleótido donante.

En este contexto, el uso de los términos "sustituir" o "sustitución" pueden entenderse para representar la sustitución de una secuencia de nucleótidos por otra (es decir, sustitución de una secuencia en el sentido informativo), y no necesariamente requiere la sustitución física o química de un polinucleótido por otro.

Por consiguiente, en un aspecto, se proporciona un método para la sustitución de una región de interés en cromatina celular (por ejemplo, una secuencia genómica) con una primera secuencia de nucleótidos, comprendiendo el método: (a) manipular un dominio de unión de dedos de cinc 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 cinc y un dominio de escisión; y (c) poner en contacto la célula con un polinucleótido que comprende la primera secuencia de nucleótidos; en el que la proteína de fusión se une a la segunda secuencia de forma que la cromatina celular se escinda en la región de interés y una secuencia de nucleótidos en la región de interés se sustituye por la primera secuencia de nucleótidos. Generalmente, la cromatina celular se escinde en la región de interés en o adyacente a la segunda secuencia. En realizaciones adicionales, el dominio de escisión comprende dos medios dominios de escisión, que pueden derivarse de la misma nucleasa o de nucleasas diferentes.

Además, se proporciona un método para la sustitución de una región de interés en cromatina celular (por ejemplo, una secuencia genómica) con una primera secuencia de nucleótidos, comprendiendo el método: (a) manipular un primer dominio de unión de dedos de cinc para unirse a una segunda secuencia en la región de interés; (b) proporcionar un segundo dominio de unión de dedos de cinc 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 cinc 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 cinc 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 el que la primera proteína de fusión se une a la segunda secuencia y la segunda proteína de fusión se une a la tercera secuencia, posicionando así los medios dominios de escisión de forma que la cromatina celular se escinda en la región de interés y una secuencia de nucleótidos en la región de interés se sustituya por la primera secuencia de nucleótidos. Generalmente, la cromatina celular se escinde en la región de interés en un sitio entre la segunda y tercera secuencias.

Métodos adicionales para la sustitución de una región de interés en cromatina celular (por ejemplo, una secuencia

genómica) con una primera secuencia de nucleótidos comprenden: (a) seleccionar una segunda secuencia, en los que la segunda secuencia está en la región de interés y tiene una longitud de al menos 9 nucleótidos; (b) manipular un primer dominio de unión de dedos de cinc para unirse a la segunda secuencia; (c) seleccionar una tercera secuencia, en los que la tercera secuencia tiene una longitud de al menos 9 nucleótidos y se localiza entre 2 y 50 nucleótidos de la segunda secuencia; (d) proporcionar un segundo dominio de unión de dedos de cinc para unirse a la tercera secuencia; (e) expresar una primera proteína de fusión en una célula, comprendiendo la primera proteína de fusión el primer dominio de unión de dedos de cinc 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 cinc 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 los que la primera proteína de fusión se une a la segunda secuencia y la segunda proteína de fusión se une a la tercera secuencia, posicionando así los medios dominios de escisión de forma que la cromatina celular se escinda en la región de interés y una secuencia de nucleótidos en la región de interés se sustituya por la primera secuencia de nucleótidos. Generalmente, la cromatina celular se escinde en la región de interés en un sitio entre la segunda y tercera secuencias.

En otro aspecto, se proporcionan métodos para la recombinación dirigida en los que una primera secuencia de nucleótidos, localizada en una región de interés en cromatina celular, se sustituye con una segunda secuencia de nucleótidos. Los métodos comprenden (a) manipular un primer dominio de unión de dedos de cinc para unirse a una tercera secuencia en la región de interés; (b) proporcionar un segundo dominio de unión de dedos de cinc para unirse a una cuarta secuencia; (c) expresar una primera proteína de fusión en una célula, comprendiendo la proteína de fusión el primer dominio de unión de dedos de cinc 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 cinc 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 el que la primera proteína de fusión se une a la tercera secuencia y la segunda proteína de fusión se une a la cuarta secuencia, posicionando así los medios dominios de escisión de forma que la cromatina celular se escinda en la región de interés y la primera secuencia de nucleótidos se sustituya por la segunda secuencia de nucleótidos.

En realizaciones adicionales, se proporciona un método para la alteración de una primera secuencia de nucleótidos en una región de interés en cromatina celular, comprendiendo el método las etapas de (a) manipular un primer dominio de unión de dedos de cinc para unirse a una segunda secuencia de nucleótidos en la región de interés, en el que la segunda secuencia comprende al menos 9 nucleótidos; (b) proporcionar un segundo dominio de unión de dedos de cinc para unirse a una tercera secuencia de nucleótidos, en el que la tercera secuencia comprende al menos 9 nucleótidos y se localiza 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 cinc 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 cinc 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 el que la cuarta secuencia de nucleótidos es homóloga pero no idéntica a la primera secuencia de nucleótidos; en el que la primera proteína de fusión se une a la segunda secuencia, y la segunda proteína de fusión se une a la tercera secuencia, posicionando así los medios dominios de escisión de forma que la cromatina celular se escinda en la región de interés y la primera secuencia de nucleótidos se altere. En ciertas realizaciones, la primera secuencia de nucleótidos se convierte en la cuarta secuencia de nucleótidos. En realizaciones adicionales, la segunda y tercera secuencias de nucleótidos (es decir, los sitios de unión para las proteínas de fusión) están presentes en el polinucleótido que comprende la cuarta secuencia de nucleótidos (es decir, el polinucleótido donante) y el polinucleótido que comprende la cuarta secuencia de nucleótidos se escinde.

En los métodos anteriormente mencionados para la recombinación dirigida, los sitios de unión para las proteínas de fusión (es decir, la tercera y cuarta secuencias) pueden comprender cualquier número de nucleótidos. Preferentemente, tienen al menos nueve nucleótidos de longitud, pero también pueden ser mayores (por ejemplo, 10, 11, 12, 13, 14, 15, 16, 17, 18 y hasta 100 nucleótidos, que incluyen cualquier valor integral entre 9 y 100 nucleótidos); además la tercera y cuarta secuencias no necesitan ser de la misma longitud. La distancia entre los sitios de unión (es decir, la longitud de la secuencia de nucleótidos entre la tercera y cuarta secuencias) puede ser cualquier número entero de pares de nucleótidos entre 2 y 50 (por ejemplo, 5 ó 6 pares de bases) como se mide desde el extremo próximo de un sitio de unión hasta el extremo próximo del otro sitio de unión.

En los métodos anteriormente mencionados para la recombinación dirigida, la cromatina celular puede escindirse en un sitio localizado entre los sitios de unión de las dos proteínas de fusión. En ciertas realizaciones, los sitios de unión están sobre hebras de ADN opuestas; en realizaciones adicionales, los sitios de unión están sobre la misma hebra de ADN. Además, la expresión de las proteínas de fusión en la célula puede llevarse a cabo tanto por introducción de las proteínas en la célula como por 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, pueden introducirse dos polinucleótidos, cada uno que comprende secuencias que codifican una de las dos proteínas de fusión, en una célula. Alternativamente, un único polinucleótido que comprende secuencias que codifican ambas proteínas de fusión puede introducirse en la célula.

Así, en una realización, un método para la sustitución de una región de interés en cromatina celular (por ejemplo, una secuencia genómica) con una primera secuencia de nucleótidos comprende: (a) manipular un primer dominio de unión de dedos de cinc para unirse a una segunda secuencia en la región de interés; (b) proporcionar un segundo dominio de unión de dedos de cinc para unirse a una tercera secuencia; y (c) poner en contacto una célula con:

- (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 cinc 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 cinc y un segundo medio dominio de escisión;

en el que la primera y segunda proteínas de fusión 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, posicionando así los medios dominios de escisión de forma que la cromatina celular se escinda en la región de interés; y la región de interés se sustituya por la primera secuencia de nucleótidos.

En las realizaciones preferidas de métodos para la recombinación dirigida y/o sustitución y/o alteración de una secuencia en una región de interés en cromatina celular, una secuencia cromosómica se altera por recombinación homóloga con una secuencia de nucleótidos "donante" exógena. Tal recombinación homóloga se estimula por la presencia de una rotura bicatenaria en la cromatina celular, si las secuencias homólogas a la región de la rotura están presentes. Las roturas bicatenarias en cromatina celular también pueden estimular mecanismos celulares de unión de extremos no homólogos.

En cualquiera de los métodos descritos en el presente documento, la primera secuencia de nucleótidos (la "secuencia donante") puede contener secuencias que son homólogas, pero no idénticas, a secuencias genómicas en la región de interés, estimulando así la recombinación homóloga para insertar una secuencia no idéntica en la región de interés. Así, en ciertas realizaciones, porciones de la secuencia donante que son homólogas a secuencias en la región de interés presentan entre aproximadamente el 80 y el 99 % (o cualquier número entero entremedias) de identidad de secuencias con la secuencia genómica que se sustituye. En otras realizaciones, la homología entre la secuencia donante y genómica es superior al 99 %, por ejemplo, si solo 1 nucleótido se diferencia entre las secuencias donante y genómica de más de 100 pares de bases contiguos. En ciertos casos, una porción no homóloga de la secuencia donante puede contener secuencias no presentes en la región de interés, de forma que se introducen nuevas secuencias en la región de interés. En estos casos, la secuencia no homóloga está generalmente flanqueada por secuencias de 50-1.000 pares de bases (o cualquier valor entero entremedias) o cualquier número de pares de bases superior a 1.000, que son homólogas o idénticas a secuencias en la región de interés. En otras realizaciones, la secuencia donante es no homóloga a la primera secuencia, y se inserta en el genoma por mecanismos de recombinación no homóloga.

En métodos para la recombinación dirigida y/o sustitución y/o alteración de una secuencia de interés en cromatina celular, el primer y segundo medios dominios de escisión pueden derivarse de la misma endonucleasa o de endonucleasas diferentes. Las endonucleasas incluyen, pero no se limitan a, endonucleasas de asentamiento y endonucleasas de restricción. Endonucleasas de restricción a modo de ejemplo son endonucleasas de restricción de tipo IIS; una endonucleasa de restricción de tipo IIS a modo de ejemplo es *Fok I*.

La región de interés puede estar en un cromosoma, episoma o genoma organular. La región de interés puede comprender una mutación, que puede sustituirse con una secuencia no mutante (o con una secuencia mutante diferente), o la región de interés puede contener una secuencia no mutante que está sustituida con 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, transposiciones, inversiones y translocalizaciones. Las mutaciones pueden cambiar la secuencia codificante, introducir codón (codones) de terminación prematuro(s) y/o modificar la frecuencia de un motivo de secuencia repetitivo (por ejemplo, repetición de trinucleótidos) en un gen. Para aplicaciones en las que se usa la recombinación dirigida para sustituir una secuencia mutante, la cromatina celular se escinde generalmente en un sitio localizado dentro de 100 nucleótidos en cualquier lado de la mutación, aunque también pueden usarse sitios de escisión localizados hasta 6-10 kb del sitio de una mutación.

En cualquiera de los métodos descritos en el presente documento, el segundo dominio de unión de dedos de cinc puede manipularse, por ejemplo, diseñarse y/o seleccionarse.

Además, el polinucleótido donante puede ser ADN o ARN, puede ser lineal o circular, y puede ser monocatenario o bicatenario. Puede administrarse a la célula como ácido nucleico desnudo, como un complejo con uno o más agentes de administración (por ejemplo, liposomas, poloxámeros) o contenerse en un vehículo de administración viral, tal como, por ejemplo, un adenovirus, un virus adeno-asociado (AAV) o un vector de lentivirus no integrante. Las secuencias donantes pueden oscilar en longitud de 10 a 1.000 nucleótidos (o cualquier valor entero de nucleótidos entremedias) o más.

Similarmente, los polinucleótidos que codifican fusiones entre un dominio de unión de dedos de cinc y un dominio de escisión o medio dominio pueden ser ADN o ARN, pueden ser lineales o circulares, y pueden ser monocatenarios o bicatenarios. Pueden administrarse a la célula como ácido nucleico desnudo, como un complejo con uno o más agentes de administración (por ejemplo, liposomas, poloxámeros) o contenerse en un vehículo de administración viral, tal como, por ejemplo, un adenovirus, un virus adeno-asociado (AAV) o un vector de lentivirus no integrante. Un polinucleótido puede codificar una o más proteínas de fusión.

En los métodos para la recombinación dirigida, al igual que con los métodos para la escisión dirigida, un dominio de escisión o medio dominio puede derivarse de cualquier nucleasa, por ejemplo, una endonucleasa de asentamiento o una endonucleasa de restricción, en particular, una endonucleasa de restricción de tipo IIS. Los medios dominios de escisión pueden derivarse de las mismas endonucleasas o de endonucleasas diferentes. Una fuente a modo de ejemplo, de la que puede derivarse un medio dominio de escisión, es la endonucleasa de restricción de tipo IIS *Fok* I.

En ciertas realizaciones, la frecuencia de recombinación homóloga puede potenciarse deteniendo las células en la fase G2 del ciclo celular y/o activando la expresión de una o más moléculas (proteína, ARN) que participan en la recombinación homóloga y/o inhibiendo la expresión o actividad de proteínas que participan en la unión de extremos no homólogos.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La **Figura 1** muestra la secuencia de nucleótidos, en forma bicatenaria, de una porción del gen hSMC1L1 humano que codifica la porción del extremo amino de la proteína (SEC ID N°: 1) y la secuencia de aminoácidos codificada (SEC ID N°: 2). Las secuencias diana para las ZFP específicas de hSMC1 están subrayadas (una en cada hebra de ADN).

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

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 en la dirección 5' (izquierda de +1), el primer exón (entre +1 y el extremo derecho de la flecha marcada "secuencia codificante de SMC1") y una porción del primer intrón. También se proporcionan localizaciones de secuencias homólogas a los cebadores de amplificación iniciales y al cebador específico del cromosoma (véase la Tabla 3). La Figura 3C muestra la secuencia de nucleótidos del cromosoma X humano en la región del codón de iniciación de SMC1 (SEC ID N°: 3), la secuencia de aminoácidos codificada (SEC ID N°: 4) y los sitios diana para las proteínas de dedos de cinc específicas para SMC1. La Figura 3D muestra la secuencia de la región correspondiente de la molécula donante (SEC ID N°: 5), con diferencias entre secuencias donantes y cromosómicas subrayadas. Las secuencias contenidas en el cebador de amplificación específico de donante (Tabla 3) se indican por subrayado doble.

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

La **Figura 5** muestra el análisis por PCR de ADN de células HEK293 transfectadas. Desde la izquierda, los carriles muestran los 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 de ZFP-*Fok*I específicas de hSMC1 (ZFP solas), células transfectadas con dos concentraciones del plásmido donante de hSMC1 (donante solo) y células transfectadas con los dos plásmidos que codifican ZFP y el plásmido donante (ZFP + donante). Véase el Ejemplo 1 para detalles.

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

La **Figura 7** muestra la secuencia de nucleótidos de una porción del gen IL-2R γ humano que comprende el extremo 3' del segundo intrón y el extremo 5' del tercer exón (SEC ID N°: 7) y la secuencia de aminoácidos codificada por la porción expresada del tercer exón (SEC ID N°: 8). Las secuencias diana para el segundo par de ZFP específicas para IL-2R γ están subrayadas. Véase el Ejemplo 2 para detalles.

La **Figura 8** muestra un diagrama esquemático de un plásmido que codifica una fusión de ZFP-*Fok*I para la escisión dirigida del gen IL-2R γ .

La **Figura 9 A-D** muestra un diagrama esquemático del gen IL-2R γ . La Figura 9A muestra un esquema de una porción del cromosoma X humano que incluye el gen IL-2R γ . La Figura 9B muestra un esquema de una porción del gen IL-2R γ que incluye una porción del segundo intrón, el tercer exón y una porción del tercer intrón. También se proporcionan localizaciones de secuencias homólogas a los cebadores de amplificación iniciales y al cebador específico de cromosoma (véase la Tabla 5). La Figura 9C muestra la secuencia de nucleótidos del cromosoma X humano en la región del tercer exón del gen IL-2R γ (SEC ID N°: 9), la

secuencia de aminoácidos codificada (SEC ID N°: 10) y los sitios diana para el primer par de proteínas de dedos de cinc específicas para IL-2R γ . La Figura 9D muestra la secuencia de la región correspondiente de la molécula donante (SEC ID N°: 11), con diferencias entre las secuencias donantes y cromosómicas subrayadas. Las secuencias contenidas en el cebador de amplificación específico para donante (Tabla 5) se indican por sobrerayado doble.

La **Figura 10** muestra un diagrama esquemático de la construcción donante de IL-2R γ .

La **Figura 11** muestra el análisis por PCR de 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 de ZFP-*FokI* específicas de IL-2R γ (ZFP solas, carril 1), células transfectadas con dos concentraciones del plásmido donante de IL-2R γ (donante solo, 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 de ZFP-*FokI* específicas de IL-2R γ (identificadas como "par 1" y "par 2") y el uso de ambos pares produjo la producción del producto de amplificación de diagnóstico (marcado "producto quimérico esperado" en la figura). Véase el Ejemplo 2 para detalles.

La **Figura 12** muestra la secuencia de nucleótidos de un producto de amplificación derivado de un gen IL-2R γ mutado (SEC ID N°: 12) generado por recombinación homóloga dirigida. Las secuencias derivadas del vector en el que se clonó el producto de amplificación están subrayadas una vez, las secuencias cromosómicas no presentes en la molécula donante se indican por subrayado discontinuo (nucleótidos 460-552), las secuencias comunes al donante y al cromosoma no están subrayadas (nucleótidos 32-42 y 59-459) y un estiramiento de secuencia que contiene nucleótidos que distinguen secuencias donantes de secuencias cromosómicas está doblemente subrayado (nucleótidos 44-58). Las letras en minúscula representan nucleótidos cuya secuencia se diferencia entre el cromosoma y el donante.

La **Figura 13** muestra la secuencia de nucleótidos de una porción del gen de la beta-globina humana que codifica segmentos del promotor central, los dos primeros exones y el primer intrón (SEC ID N°: 13). Una mutación de aminoácido que cambia una A (en negrita y subrayada) en la posición 5212541 sobre el cromosoma 11 (BLAT, sitio de UCSC Genome Bioinformatics) a una T produce anemia drepanocítica. Se diseñó una primera proteína de fusión de dedos de cinc/*FokI* de forma que los contactos primarios estuvieran con la secuencia de 12 nucleótidos subrayada AAGGTGAACGTG (nucleótidos 305-316 de SEC ID N°: 13) y se diseñó una segunda proteína de fusión de dedos de cinc/*FokI* de forma que los primeros contactos estuvieran con el complemento de la secuencia de 12 nucleótidos subrayada CCGTTACTGCC (nucleótidos 325-336 de SEC ID N°: 13).

La **Figura 14** es un diagrama esquemático de un plásmido que codifica la fusión de ZFP-*FokI* 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 clonado que muestra la región en la dirección 5', primer y segundo exones, primer intrón y primeros sitios de unión.

La **Figura 16** es un diagrama esquemático de la construcción donante de beta-globina, pCR4-TOPO-HBBdonante.

La **Figura 17** muestra el análisis por PCR de ADN de células transfectadas con dos pares de ZFP-nucleasas específicas de β -globina y un plásmido donante de beta-globina. El panel de la izquierda es un control de carga en el que los cebadores amp 1 inicial y amp 2 inicial (Tabla 7) se usaron para la amplificación. En el experimento mostrado en el panel derecho, los cebadores "específicos de cromosoma" y "específicos de donante" (Tabla 7) se usaron para la amplificación. El carril más a la izquierda en cada panel contiene marcadores de peso molecular y el siguiente carril muestra productos de amplificación obtenidos de las células transfectadas con vector vacío. Los carriles restantes, de izquierda a derecha, muestran producto de amplificación de células transfectadas con: un plásmido que codifica GFP, 100 ng de cada plásmido que codifica ZFP/*FokI*, 200 ng de cada plásmido que codifica ZFP/*FokI*, 200 ng de plásmido donante, 600 ng de plásmido donante, 200 ng de plásmido donante + 100 ng de cada plásmido que codifica ZFP/*FokI* y 600 ng de plásmido donante + 200 ng de cada plásmido que codifica ZFP/*FokI*.

La **Figura 18** muestra la secuencia de nucleótidos de un producto de amplificación derivado de un gen de la beta-globina mutada (SEC ID N°: 14) generado por recombinación homóloga dirigida. Las secuencias cromosómicas no presentes en la molécula donante se indican por subrayado discontinuo (nucleótidos 1-72), secuencias comunes al donante y el cromosoma no están subrayadas (nucleótidos 73-376) y un estiramiento de secuencia que contiene nucleótidos que distinguen secuencias donantes de secuencias cromosómicas está doblemente subrayado (nucleótidos 377-408). Las letras en minúscula representan nucleótidos cuya secuencia se diferencia 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 γ) (SEC ID N°: 15). También se muestran (subrayadas) las secuencias diana para las proteínas de fusión de ZFP/*FokI* 5-8 y 5-10. Véase el Ejemplo 5 para detalles.

La **Figura 20** muestra la secuencia de aminoácidos de la fusión de ZFP/*FokI* 5-8 dirigida al exón 5 del gen IL-2R γ humano (SEC ID N°: 16). Los residuos de aminoácidos 1-17 contienen una secuencia de localización nuclear (NLS, subrayada); los residuos 18-130 contienen la porción de ZFP, con las regiones de reconocimiento de los dedos de cinc componentes mostrados en negrita; el conector de ZFP-*FokI* (conector de ZC, subrayada) se extiende desde los residuos 131 a 140 y el medio dominio de escisión de *FokI* empieza en el residuo 141 y se extiende hasta el extremo 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 de ZFP/*FokI* 5-10 dirigida al exón 5 del gen IL-2R γ humano (SEC ID N°: 17). Los residuos de aminoácidos 1-17 contienen una secuencia de localización nuclear (NLS, subrayada); los residuos 18-133 contienen la porción de ZFP, con las regiones de reconocimiento de dedos de cinc componentes mostradas en negrita; el conector de ZFP-*FokI* (conector de ZC, subrayado) se extiende desde los residuos 134 a 143 y el medio dominio de escisión de *FokI* empieza en el residuo 144 y se extiende hasta el extremo de la proteína en el residuo 339. El residuo que se alteró para generar la mutación E490K se muestra subrayado y en negrita.

La **Figura 22** muestra la secuencia de nucleótidos del gen de la proteína verde fluorescente potenciado (SEC ID N°: 18) derivado del gen GFP de *Aequorea victoria* (Tsien (1998) Ann. Rev. Biochem. 67:509-544). El codón de iniciación ATG, además de la región que se mutagenizó, están subrayados.

La **Figura 23** muestra la secuencia de nucleótidos de un gen eGFP defectuoso mutante (SEC ID N°: 19). Los sitios de unión para ZFP-nucleasas están subrayados y la región entre los sitios de unión se corresponde con la región que se modificó.

La **Figura 24** muestra las estructuras de plásmidos que codifican nucleasas de dedos de cinc dirigidas al gen eGFP.

La **Figura 25** muestra un autorradiograma de un gen de 10 % de acrilamida usado para analizar la escisión de ADN dirigida de un gen eGFP mutante por endonucleasas de dedos de cinc. Véase el Ejemplo 8 para detalles.

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

La **Figura 27** muestra niveles de ARNm de eGFPmut, normalizados a ARNm de GAPDH, en diversas líneas celulares obtenidas de la transfección de células HEK293 humanas. Las barras claras muestran niveles en células sin tratar; las barras oscuras muestran niveles en células que se habían tratado con 2 ng/ml de doxiciclina. Véase el Ejemplo 9 para detalles.

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

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

La **Figura 30** muestra un trazo de FACS de células T18 transfectadas con plásmidos que codifican dos ZFP-nucleasas y un plásmido que codifica una secuencia donante, que se detuvieron en la fase G2 del ciclo celular 24 horas después de la transfección con 100 ng/ml de nocodazol durante 48 horas. El medio se sustituyó y se dejó que las células se recuperaran durante 48 horas adicionales, y la corrección génica se midió por análisis de FACS. Véase el Ejemplo 11 para detalles.

La **Figura 31** muestra un trazo de FACS de células T18 transfectadas con plásmidos que codifican dos ZFP-nucleasas y un plásmido que codifica una secuencia donante, que se detuvieron en la fase G2 del ciclo celular 24 horas después de la transfección con vinblastina 0,2 μ M durante 48 horas. El medio se sustituyó y se dejó que las células se recuperaran durante 48 horas adicionales, y la corrección génica se midió por análisis de FACS. Véase el Ejemplo 11 para detalles.

La **Figura 32** muestra la secuencia de nucleótidos de un inserto de eGFP de 1.527 nucleótidos en pCR(R)4-TOPO (SEC ID N°: 21). La secuencia codifica una proteína verde fluorescente potenciada no modificada que carece de un codón de iniciación. Véase el Ejemplo 13 para detalles.

La **Figura 33** muestra un diagrama esquemático de un ensayo usado para medir la frecuencia de edición del gen IL-2R γ humano endógeno. Véase el Ejemplo 14 para detalles.

La **Figura 34** muestra autorradiogramas de geles de acrilamida usados en un ensayo para medir la frecuencia de edición de un gen celular endógeno por escisión y recombinación homóloga dirigidas. El carril marcado "GFP" muestra los resultados del ensayo de un control en el que las células se transfectaron con un vector que codifica eGFP; el carril marcado "ZFP solas" muestra los resultados de otro experimento de control en el que 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 "donante solo" muestran resultados de un experimento de control en el que células se transfectaron con 1 μ g de plásmido donante, pero no con el plásmido que codifica 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 μ g del plásmido donante, e 1D se refiere a células transfectadas con 1,0 μ g del plásmido donante. "+" se refiere a células que se expusieron a vinblastina 0,2 μ M; "-" se refieren a células que no se expusieron a vinblastina. "wt" se refiere al fragmento obtenido después de la digestión con *Bsr*BI de productos de amplificación obtenidos de cromosomas que contienen el gen IL-2R γ cromosómico no mutante; "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 de cromosomas que contienen secuencias del plásmido donante que se habían integrado por recombinación homóloga.

La **Figura 35** muestra una imagen autorradiográfica de una exposición de cuatro horas de un gel usado en un ensayo para medir la recombinación dirigida al sitio IL-2R γ humano en células K562. "wt" identifica una banda que es diagnóstica de ADN cromosómico que contiene la secuencia nativa de IL-2R γ de K562; "rflp"

identifica un diagnóstico doble de ADN cromosómico que contiene la secuencia alterada de IL-2R γ presente en la molécula de ADN donante. El símbolo "+" encima de un carril indica que las células se trataron con vinblastina 0,2 μ M; el símbolo "-" indica que las células no se trataron con vinblastina. Los números en los carriles "ZFP + donante" indican el porcentaje de ADN cromosómico total que contiene secuencia originalmente presente en la molécula de ADN donante, calculado usando la función "peak finder, automatic baseline" del software ImageQuant v. 5.1 de Molecular Dynamics como se describe en Cap. 8 del manual del fabricante (Guía del Usuario de ImageQuant de Molecular Dynamics; parte 218-415). "Untr" indica célula sin transfectar. Véase el Ejemplo 15 para detalles adicionales.

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

La **Figura 37** muestra un autorradiograma de una exposición de cuatro horas de una transferencia de ADN sondada con un fragmento específico para el gen IL-2R γ humano. La flecha a la derecha de la imagen indica la posición de una banda correspondiente a ADN genómico cuya secuencia ha sido alterada por recombinación homóloga. El símbolo "+" encima de un carril indica que las células se trataron con vinblastina 0,2 μ M; el símbolo "-" indica que las células no se trataron con vinblastina. Los números debajo de los carriles "ZFP + donante" indican el porcentaje de ADN cromosómico total que contiene secuencia originalmente presente en la molécula de ADN donante, calculado como se describe en el Ejemplo 35. Véase el Ejemplo 15 para detalles adicionales.

La **Figura 38** muestra imágenes autorradiográficas de geles usados en un ensayo para medir recombinación dirigida en el locus IL-2R γ humano en células de médula ósea humana CD34⁺. El panel izquierdo muestra un patrón de referencia en el que el porcentaje establecido de ADN genómico humano normal (que contiene un sitio *Maell*) se añadió a ADN genómico de células Jurkat (que carecen de un sitio *Maell*), la mezcla se amplificó por PCR para generar un producto de amplificación radiomarcado y el producto de amplificación se digirió con *Maell*. "wt" identifica una banda que representa ADN sin digerir y "rflp" identifica una banda resultante de digestión con *Maell*.

El panel derecho muestra los resultados de un experimento en el que células CD34⁺ se transfectaron con ADN donante que contiene un sitio *BsrBI* y plásmidos que codifican endonucleasas de fusión de dedos de cinc-*FokI*. La región genómica relevante se amplificó a continuación y se marca, y el producto de amplificación marcado se digirió con *BsrBI*. "GFP" indica células de control que se transfectaron con un plásmido codificante de GFP; "Donante solo" indica células de control que se transfectaron solo con ADN donante, y "ZFP + Donante" indica células que se transfectaron con ADN donante y con plásmidos que codifican las nucleasas de dedos de cinc/*FokI*. "wt" identifica una banda que es diagnóstica de ADN cromosómico que contiene la secuencia nativa de IL-2R γ ; "rflp" identifica una banda que es diagnóstica de ADN cromosómico que contiene la secuencia alterada de IL-2R γ presente en la molécula de ADN donante. El carril más a la derecha contiene marcadores de tamaño de ADN. Véase el Ejemplo 16 para detalles adicionales.

La **Figura 39** muestra una imagen de una inmunotransferencia usada para probar 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 conjuntos de ARNip diferentes (véase el Ejemplo 18). Carril 1: 70 ng de conjunto D de ARNip; Carril 2: 140 ng de conjunto D de ARNip; Carril 3: 70 ng de conjunto E de ARNip; Carril 4: 140 ng de conjunto E de ARNip. "Ku70" indica la banda que representa la proteína Ku70; "TFIIB" indica una banda que representa el factor de transcripción TFIIB, usado como control.

La **Figura 40** muestra las secuencias de aminoácidos de cuatro dominios de dedos de cinc dirigidos al gen de β -globina humano: sca-29b (SEC ID N°: 22); sca-36a (SEC ID N°: 23); sca-36b (SEC ID N°: 24) y sca-36c (SEC ID N°: 25). El sitio diana para el dominio de sca-29b está sobre la hebra de ADN, y los sitios diana para los dominios de sca-36a, sca-36b y sca-36c están sobre la hebra opuesta. Véase el Ejemplo 20.

La **Figura 41** muestra los resultados de un ensayo *in vitro*, en el que diferentes combinaciones de nucleasas de fusión de dedos de cinc/*FokI* (ZFN) se probaron para la escisión de ADN específico de secuencia. El carril marcado "U" muestra una muestra del molde de ADN. Los cuatro siguientes carriles muestran resultados de incubación del molde de ADN con cada una de las cuatro ZFN dirigidas a β -globina (véase el Ejemplo 20 para la caracterización de estas ZFP). Los tres carriles más a la derecha muestran los resultados de la incubación del molde de ADN con la ZFN sca-29b y una de las ZFN sca-36a, sca-36b o sca-36c (todas las cuales están dirigidas a la hebra opuesta a la que se dirige sca-29b).

La **Figura 42** muestra 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 de eGFP, en células transfectadas con ADN donante y plásmidos que codifican nucleasas de dedos de cinc dirigidas a eGFP, en función de la concentración de doxiciclina.

La **Figura 43A-C** muestra diagramas esquemáticos de diferentes configuraciones de proteínas de fusión.

La Figura 43A muestra dos proteínas de fusión, en las que el dominio de dedos de cinc es el más próximo al extremo N y el medio dominio de escisión de *FokI* es el más próximo al extremo C, uniéndose a sitios diana de ADN sobre hebras opuestas cuyos extremos 5' son proximales entre sí. La Figura 43B muestra dos proteínas de fusión, en las que el medio dominio de escisión de *FokI* es el más próximo al extremo N y el dominio de dedos de cinc es el más próximo al extremo C, uniéndose a sitios diana de ADN sobre hebras opuestas cuyos extremos 3' son proximales entre sí. La Figura 43C muestra una primera proteína en la que el medio dominio de escisión de *FokI* es el más próximo al extremo N y el dominio de dedos de cinc es el más próximo al extremo C y una segunda proteína en la que el dominio de dedos de cinc es el más próximo al extremo N y el medio dominio de escisión de *FokI* es el más próximo al extremo C, uniéndose a sitios diana de ADN sobre la misma hebra, en el que el sitio diana para la primera proteína está en la dirección 5' (es decir, hacia el lado 5') del sitio de unión para la segunda proteína.

En todos los ejemplos se muestra que proteínas de tres dedos se unen a sitios diana de nueve nucleótidos. Se muestra la polaridad de 5' y 3' de las hebras de ADN, y se identifican los extremos N de las proteínas de fusión.

La **Figura 44** es un autorradiograma de un gel de acrilamida en el que se ensayó la escisión de un sustrato modelo por endonucleasas de dedos de cinc. El carril 1 muestra la migración de sustrato sin escindir. El carril 2 muestra sustrato después de la incubación con la proteína de fusión de dedos de cinc de IL2-1R/*FokI*. El carril 3 muestra el sustrato después de la incubación con la proteína de fusión de dedos de cinc de 5-9DR/*FokI*. El carril 4 muestra el sustrato después de la incubación con ambas proteínas. Tamaños aproximados (en pares de bases) del sustrato y sus productos de escisión se muestran a la derecha de la imagen. Debajo de la imagen se muestra la secuencia de nucleótidos (SEC ID N°: 211) de la porción del sustrato que contiene los sitios de unión para los dominios de unión de dedos de cinc de 5-9D y IL2-1. Los sitios de unión se identifican y se indican por subrayado.

La **Figura 45** es un autorradiograma de un gel de acrilamida en el que se ensayó la escisión de un sustrato modelo por endonucleasas de dedos de cinc. El carril 1 muestra la migración de sustrato sin escindir. El carril 2 muestra sustrato después de la incubación con la proteína de fusión de dedos de cinc de IL2-1C/*FokI*. El carril 3 muestra el sustrato después de la incubación con la proteína de fusión de dedos de cinc de IL2-1R/*FokI*. El carril 4 muestra el sustrato después de la incubación con la proteína de fusión de dedos de cinc de 5-9DR/*FokI*. El carril 5 muestra el sustrato después de la incubación con tanto las proteínas de fusión IL2-1R como 5-9DR. El carril 6 muestra el sustrato después de la incubación con ambas proteínas IL2-1C y 5-9DR. Tamaños aproximados (en pares de bases) del sustrato y sus productos de escisión se muestran a la derecha de la imagen. Debajo de la imagen se muestra la secuencia de nucleótidos (SEC ID N°: 212) de la porción del sustrato que contiene los sitios de unión para los dominios de unión de dedos de cinc de 5-9D y IL2-1. Los sitios de unión se identifican y se indican por subrayado.

DESCRIPCIÓN DETALLADA

La presente invención se define por las reivindicaciones.

En el presente documento se desvelan composiciones y métodos útiles para la escisión dirigida de cromatina celular y para la alteración dirigida de una secuencia de nucleótidos celular, por ejemplo, por escisión dirigida seguida de unión de extremos no homólogos o por escisión dirigida seguida de recombinación homóloga entre un polinucleótido exógeno (que comprende una o más regiones de homología con la secuencia de nucleótidos celular) y una secuencia genómica. Secuencias genómicas incluyen aquellas presentes en cromosomas, episomas, genomas organulares (por ejemplo, mitocondrias, cloroplastos), cromosomas artificiales y cualquier otro tipo de ácido nucleico presente en una célula tal como, por ejemplo, secuencias amplificadas, cromosomas dobles diminutos y los genomas de bacterias y virus endógenos o infectantes. Secuencias genómicas pueden ser normales (es decir, no mutantes) o mutantes; las secuencias mutantes pueden comprender, por ejemplo, inserciones, deleciones, translocalizaciones, transposiciones y/o mutaciones puntuales. Una secuencia genómica también puede comprender uno de varios alelos diferentes.

Composiciones útiles para la escisión y recombinación dirigidas incluyen proteínas de fusión que comprenden un dominio de escisión (o un medio dominio de escisión) y un dominio de unión de dedos de cinc, polinucleótidos que codifican estas proteínas y combinaciones de polipéptidos y polinucleótidos que codifican polipéptidos. Un dominio de unión de dedos de cinc puede comprender uno o más dedos de cinc (por ejemplo, 2, 3,4, 5, 6, 7, 8, 9 o más dedos de cinc), y puede manipularse para unirse a cualquier secuencia genómica. Así, identificando una región genómica diana de interés en la que se desea la escisión o recombinación, pueden, según los métodos desvelados en el presente documento, construirse una o más proteínas de fusión que comprenden un dominio de escisión (o medio dominio de escisión) y un dominio de dedos de cinc manipulado para reconocer una secuencia diana en dicha región genómica. La presencia de una proteína de fusión tal (o proteínas) en una célula producirá la unión de la(s) proteína(s) de fusión a su(s) sitio(s) de unión y la escisión dentro de o cerca de dicha región genómica. Además, si un polinucleótido exógeno homólogo a la región genómica también está presente en una célula tal, la recombinación homóloga se produce a una tasa mayor entre la región genómica y el polinucleótido exógeno.

General

La práctica de los métodos, además de la preparación y uso de las composiciones desveladas en el presente documento, emplea, a menos que se indique lo contrario, técnicas convencionales en biología molecular, bioquímica, estructura y análisis de cromatina, química computacional, cultivo celular, ADN recombinante y campos relacionados que están dentro de la experiencia de la materia. Estas técnicas se explican completamente en la bibliografía. Véase, por ejemplo, Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, segunda edición, Cold Spring Harbor Laboratory Press, 1989 y tercera edición, 2001; Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 y actualizaciones periódicas; la serie METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, tercera edición, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P.M. Wassarman y A. P. Wolffe, eds.), Academic Press, San Diego, 1999; y METHODS IN MOLECULAR BIOLOGY, vol. 119, "Chromatin Protocols" (P.B. Becker, ed.) Humana Press, Totowa, 1999.

Definiciones

Los términos "ácido nucleico", "polinucleótido" y "oligonucleótido" se usan indistintamente y se refieren a un desoxirribonucleótido o polímero de ribonucleótidos, en conformación lineal o circular, y en forma tanto mono- como 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 englobar análogos conocidos de nucleótidos naturales, además de nucleótidos que se modifican en los restos de base, azúcar y/o fosfato (por ejemplo, esqueletos de fosforotioato). En general, un análogo de un nucleótido particular tiene la misma especificidad por apareamiento de bases; es decir, un análogo de A se apareará con base con T.

Los términos "polipéptido", "péptido" y "proteína" se usan indistintamente para referirse a un polímero de residuos de aminoácidos. El término también se aplica a polímeros de aminoácidos en los que uno o más aminoácidos son análogos químicos o derivados modificados de un aminoácido que se produce naturalmente correspondiente.

"Unión" se refiere a una interacción no covalente específica de secuencias, entre macromoléculas (por ejemplo, entre una proteína y un ácido nucleico). No todos los componentes de una interacción de unión necesitan ser específicos de secuencia (por ejemplo, contactos con residuos de fosfato en un esqueleto de ADN), en tanto que la interacción en conjunto sea específica de secuencia. Tales interacciones se caracterizan generalmente por una constante de disociación (K_d) de 10^{-6} M⁻¹ o menor. "Afinidad" se refiere a la intensidad de la unión: correlacionándose elevada afinidad de unión con una menor K_d .

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

Una "proteína de unión a ADN de dedos de cinc" (o dominio de unión) es una proteína, o un dominio dentro de una proteína mayor, que se une a ADN en un modo específico de secuencia mediante uno o más dedos de cinc, que son regiones de secuencia de aminoácidos dentro del dominio de unión cuya estructura se estabiliza mediante la coordinación de un ión cinc. El término proteína de unión a ADN de dedos de cinc se abrevia frecuentemente proteína de dedos de cinc o ZFP.

Los dominios de unión de dedos de cinc pueden "manipularse" para unirse a una secuencia de nucleótidos predeterminada. Ejemplos no limitantes de métodos para manipular proteínas de dedos de cinc son diseño y selección. Una proteína de dedos de cinc diseñada es una proteína que no se produce en la naturaleza cuyo diseño/composición resulta principalmente de criterios racionales. Los criterios racionales para el diseño incluyen aplicación de reglas de sustitución y algoritmos computerizados para procesar la información en una base de datos que guarda información de diseños y datos de unión ZFP existentes. Véanse, por ejemplo, las patentes de EE.UU. 6.140.081; 6.453.242; y 6.534.261; véanse también los documentos WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 y WO 03/016496.

Una proteína de dedos de cinc "seleccionada" es una proteína que no se encuentra en la naturaleza cuya producción resulta principalmente de un proceso empírico tal como expresión en fago, trampa de interacción o selección de híbridos. Véanse, por ejemplo, los documentos US 5.789.538; US 5.925.523; US 6.007.988; US 6.013.453; US 6.200.759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970; WO 01/88197 y WO 02/099084.

El término "secuencia" se refiere a una secuencia de nucleótidos de cualquier longitud, que puede ser ADN o ARN; puede ser lineal, circular o ramificada y puede ser tanto monocatenaria como bicatenaria. El término "secuencia donante" se refiere a una secuencia de nucleótidos que se inserta en un genoma. Una secuencia donante puede ser de cualquier longitud, por ejemplo, entre 2 y 10.000 nucleótidos de longitud (o cualquier valor de número entero

entremedias o por encima), preferentemente entre aproximadamente 100 y 1.000 nucleótidos de longitud (o cualquier número entero entremedias), más preferentemente entre aproximadamente 200 y 500 nucleótidos de longitud.

5 Una "secuencia no idéntica homóloga" se refiere a una primera secuencia que comparte un grado de identidad de secuencias con una segunda secuencia, pero cuya secuencia no es idéntica a la de la segunda secuencia. Por ejemplo, un polinucleótido que comprende la secuencia no mutante 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 entremedias, utilizando mecanismos celulares normales. Dos secuencias no idénticas homólogas pueden ser de cualquier longitud y su grado de no homología puede ser tan pequeño como un único nucleótido (por ejemplo, para la corrección de una mutación puntual genómica por recombinación homóloga dirigida) o de hasta 10 o más kilobases (por ejemplo, para la inserción de un gen en un sitio ectópico predeterminado en un cromosoma). Dos polinucleótidos que comprenden las secuencias no idénticas homólogas no necesitan ser de la misma longitud. Por ejemplo, puede usarse un polinucleótido exógeno (es decir, polinucleótido donante) de entre 15 20 y 10.000 nucleótidos o pares de nucleótidos.

En la técnica también se conocen técnicas para determinar la identidad de secuencias de ácidos nucleicos y de aminoácidos. Normalmente, tales técnicas incluyen determinar la secuencia de nucleótidos del ARNm para un gen y/o determinar la secuencia de aminoácidos así codificada, y comparar estas secuencias con un segundo nucleótido o secuencia de aminoácidos. También pueden determinarse y compararse secuencias genómicas de esta forma. En general, identidad se refiere a una correspondencia exacta de nucleótido con nucleótido o aminoácido con aminoácido de dos polinucleótidos o secuencias de polipéptidos, respectivamente. Pueden compararse dos o más secuencias (polinucleótido o aminoácido) determinando su identidad en porcentaje. La identidad en porcentaje de dos secuencias, tanto de secuencias de ácidos nucleicos como de aminoácidos, es el número de apareamientos exactos entre dos secuencias alineadas dividido entre la longitud de las secuencias más cortas y multiplicado por 100. Un alineamiento aproximado para secuencias de ácidos nucleicos se proporciona por el algoritmo de homología local de Smith y Waterman, *Advances in Applied Mathematics* 2:482-489 (1981). Este algoritmo puede aplicarse a secuencias de aminoácidos usando la matriz de puntuación desarrollada por Dayhoff, *Atlas of Protein Sequences and Structure*, M.O. Dayhoff ed., 5 supl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., EE.UU., y normalizado por Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). Una implementación a modo de ejemplo de este algoritmo para determinar la identidad en porcentaje de una secuencia se proporciona por Genetics Computer Group (Madison, WI) en la aplicación informática "BestFit". Los parámetros por defecto para este método se describen en *Wisconsin Analysis Package Program Manual*, versión 8 (1995) (disponible de Genetics Computer Group, Madison, WI). Un método preferido de establecimiento de la identidad en porcentaje en el contexto de la presente divulgación es usar el paquete MPSRCH de programas con los derechos registrados por la Universidad de Edimburgo, desarrollado por John F. Collins y Shane S. Sturrok, y distribuido por IntelliGenetics, Inc. (Mountain View, CA). De este juego de paquetes puede emplearse el algoritmo de Smith-Waterman cuando se usan parámetros por defecto para la tabla de puntuación (por ejemplo, penalización por hueco abierto de 12, penalización por extensión de hueco de uno, y un hueco de seis). De los datos generados, el valor de "Apareamiento" refleja la identidad de secuencias. Otros programas adecuados para calcular la identidad en porcentaje o similitud entre secuencias son generalmente conocidos en la técnica, por ejemplo, otro programa de alineamiento es BLAST, usado con parámetros por defecto. Por ejemplo, pueden usarse BLASTN y BLASTP usando los siguientes parámetros por defecto: código genético = estándar; filtro = ninguno; cadena = ambas; corte = 60; esperado = 10; Matriz = BLOSUM62; Descripciones = 50 secuencias; clasificado por = PUNTUACIÓN ALTA; Bases de datos = no redundantes, traducciones de GenBank + EMBL + DDBJ + PDB + GenBank CDS + proteína Swiss + Spupdate + PIR. Detalles de estos programas pueden encontrarse 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 deseados de identidad de secuencias es aproximadamente del 80 % al 100 % y cualquier valor de número entero entremedias. Normalmente, las identidades en porcentaje entre secuencias son de al menos el 70-75 %, preferentemente el 80-82 %, más preferentemente el 85-90 %, incluso más preferentemente el 92 %, todavía más preferentemente el 95 %, y lo más preferentemente el 98 % de identidad de secuencias.

Alternativamente, el grado de similitud de secuencias entre polinucleótidos puede determinarse por hibridación de polinucleótidos en condiciones que permiten la formación de dúplex estables entre regiones homólogas, seguido de digestión con nucleasa(s) específicas monocatenarias, y determinación del tamaño de los fragmentos digeridos. Dos secuencias de ácidos nucleicos, o dos de polipéptidos, son sustancialmente homólogas entre sí cuando las secuencias presentan al menos aproximadamente el 70 %-75 %, preferentemente el 80 %-82 %, más preferentemente el 85 %-90 %, incluso más preferentemente el 92 %, todavía más preferentemente el 95 %, y lo más preferentemente el 98 % de identidad de secuencias sobre una longitud definida de las moléculas, como se determina usando los métodos anteriores. Como se usa en el presente documento, sustancialmente homólogo también se refiere a secuencias que muestran identidad completa con una secuencia de ADN o de polipéptidos especificada. Las secuencias de ADN que son sustancialmente homólogas pueden identificarse en un experimento de hibridación Southern bajo, por ejemplo, condiciones rigurosas, como se definen para ese sistema particular. La definición de condiciones de hibridación apropiadas está dentro de la experiencia de la materia. Véase, por ejemplo, Sambrook et al., arriba; *Nucleic Acid Hybridization: A Practical Approach*, editors B.D. Hames and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

La hibridación selectiva de dos fragmentos de ácido nucleico puede determinarse del siguiente modo. El grado de identidad de secuencias entre dos moléculas de ácidos nucleicos afecta la eficiencia e intensidad de los eventos de hibridación entre tales moléculas. Una secuencia de ácidos nucleicos parcialmente idéntica inhibirá al menos parcialmente la hibridación de una secuencia completamente idéntica con una molécula diana. La inhibición de la hibridación de la secuencia completamente idéntica puede evaluarse usando ensayos de hibridación que son muy conocidos en la técnica (por ejemplo, transferencia Southern (ADN), transferencia Northern (ARN), hibridación en disolución, o similares, véase Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, segunda edición, (1989) Cold Spring Harbor, N.Y.). Tales ensayos pueden realizarse usando grados variables de selectividad, por ejemplo, usando condiciones variables de baja a alta rigurosidad. Si se emplean condiciones de baja rigurosidad, la ausencia de unión no específica puede evaluarse usando una sonda secundaria que carece de incluso un grado parcial de identidad de secuencias (por ejemplo, una sonda que tiene menos de aproximadamente el 30 % de identidad de secuencias con la molécula diana), de forma que, en ausencia de eventos de unión no específicos, la sonda secundaria no se hibridará con la diana.

Si se utiliza un sistema de detección basado en hibridación, se elige una sonda de ácido nucleico que es complementaria a una secuencia de ácidos nucleicos de referencia, y luego por selección de condiciones apropiadas la sonda y la secuencia de referencia se hibridan selectivamente, o se unen, entre sí para formar una molécula de dúplex. Una molécula de ácido nucleico que puede hibridarse selectivamente con una secuencia de referencia bajo condiciones de hibridación de rigurosidad moderada normalmente se hibrida bajo condiciones que permitan la detección de una secuencia de ácidos nucleicos diana de al menos aproximadamente 10-14 nucleótidos de longitud que al menos aproximadamente el 70% de identidad de secuencias con la secuencia de la sonda de ácido nucleico seleccionada. Las condiciones de hibridación rigurosas permiten normalmente la detección de secuencias de ácidos nucleicos diana de al menos aproximadamente 10-14 nucleótidos de longitud que tienen una identidad de secuencias superior a aproximadamente el 90-95 % con la secuencia de la sonda de ácido nucleico seleccionada. Las condiciones de hibridación útiles para la hibridación de secuencias de la sonda/de referencia, en las que la secuencia de la sonda y de referencia tienen un grado específico de identidad de secuencias, pueden determinarse como se conoce en la técnica (véase, por ejemplo, *Nucleic Acid Hybridization: A Practical Approach*, editors B.D. Hames and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

Las condiciones para la hibridación son muy conocidas para aquellos expertos en la materia. Rigurosidad de la hibridación se refiere al grado al que las condiciones de hibridación desfavorecen la formación de híbridos que contienen nucleótidos desapareados, correlacionándose la mayor rigurosidad con una menor tolerancia por híbridos desapareados. Los factores que afectan la rigurosidad de hibridación son muy conocidos para aquellos expertos en la materia e incluyen, pero no se limitan a, temperatura, pH, fuerza iónica y concentración de disolventes orgánicos tales como, por ejemplo, formamida y sulfóxido de dimetilo. Como es conocido por aquellos expertos en la materia, la rigurosidad de hibridación aumenta por mayores temperaturas, menor fuerza iónica y menores concentraciones de disolvente.

Con respecto a las condiciones de rigurosidad para la hibridación, es muy conocido en la técnica que puedan emplearse numerosas condiciones equivalentes para establecer una rigurosidad particular variando, por ejemplo, los siguientes factores: la longitud y la naturaleza de las secuencias, composición básica de las diversas secuencias, concentraciones de sales y otros componentes de la disolución de hibridación, la presencia o ausencia de agentes de bloqueo en las disoluciones de hibridación (por ejemplo, sulfato de dextrano y polietilenglicol), parámetros de temperatura de reacción y de tiempo de hibridación, además de condiciones de lavado variables. La selección de un conjunto particular de condiciones de hibridación se selecciona siguiendo métodos estándar en la técnica (véanse, por ejemplo, Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, segunda edición, (1989) Cold Spring Harbor, N.Y.).

“Recombinación” se refiere a un proceso de intercambio de información genética entre dos polinucleótidos. Para los fines de la presente divulgación, “recombinación homóloga (HR)” se refiere a la forma especializada de tal intercambio que tiene lugar, por ejemplo, durante la reparación de roturas bicatenarias en células. Este proceso requiere homología de secuencias de nucleótidos, usa una molécula “donante” para la reparación del molde de una molécula “diana” (es decir, la que experimentó la rotura bicatenaria), y se conoce indistintamente como “conversión génica no de entrecruzamiento” o “conversión génica de extensión corta”, debido a que conduce a la transferencia de información genética del donante a la diana. Sin desear ceñirse a teoría alguna particular, tal transferencia puede implicar la corrección de desapareamientos del ADN de heterodúplex que se forma entre la diana rota y el donante, y/o la “hibridación de hebras dependiente de la síntesis”, en la que el donante se usa para resintetizar información genética que será parte de la diana, y/o procesos relacionados. Tal HR especializada frecuentemente produce una alteración de la secuencia de la molécula diana de forma que parte o toda la secuencia del polinucleótido donante se incorpora en el polinucleótido diana.

“Escisión” se refiere a la rotura del esqueleto covalente de una molécula de ADN. La escisión puede iniciarse mediante una variedad de métodos que incluyen, pero no se limitan a, hidrólisis enzimática o química de un enlace fosfodiéster. Tanto la escisión monocatenaria como la escisión bicatenaria son posibles, y la escisión bicatenaria puede producirse como resultado de dos eventos de escisión monocatenaria distintos. La escisión de ADN puede

producir la producción de tanto extremos romos como extremos en bisel. En ciertas realizaciones, los polipéptidos de fusión se usan para la escisión de ADN bicatenario dirigida.

5 Un “dominio de escisión” comprende una o más secuencias de polipéptidos que poseen actividad catalítica para la escisión de ADN. Un dominio de escisión puede estar contenido en una única cadena de polipéptidos o la actividad de escisión puede resultar de la asociación de dos (o más) polipéptidos.

10 Un “medio dominio de escisión” es una secuencia de polipéptidos que, conjuntamente con un segundo polipéptido (tanto idéntico como diferente), forma un complejo que tiene actividad de escisión (preferentemente actividad de escisión bicatenaria).

15 “Cromatina” es la estructura de nucleoproteína que comprende el genoma celular. La cromatina celular comprende ácido nucleico, principalmente ADN, y proteína, que incluye histonas y proteínas cromosómicas no histonas. La mayoría de la cromatina celular eucariota existe en forma de nucleosomas, en la que un núcleo de nucleosoma comprende aproximadamente 150 pares de bases de ADN asociados a un octámero que comprende dos de cada una de las histonas H2A, H2B, H3 y H4; y el ADN de conector (de longitud variable dependiendo del organismo) se extiende entre los núcleos de nucleosomas. Una molécula de histona H1 está generalmente asociada al ADN de conector. Para los fines de la presente divulgación, se indica que el término “cromatina” engloba todos los tipos de nucleoproteína celular, tanto procarionota como eucariota. La cromatina celular incluye tanto cromatina cromosómica como episómica.

20 Un “cromosoma” es un complejo de cromatina que comprende toda o una porción del genoma de una célula. El genoma de una célula se caracteriza frecuentemente por su cariotipo, que es el conjunto de todos los cromosomas que comprenden el genoma de la célula. El genoma de una célula puede comprender uno o más cromosomas.

25 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. Ejemplos de episomas incluyen plásmidos y ciertos genomas virales.

30 Una “región accesible” es un sitio en la cromatina celular en la que un sitio diana presente en el ácido nucleico puede unirse por una molécula exógena que reconoce el sitio diana. Sin desear ceñirse a teoría alguna particular, se cree que una región accesible es una que no está encapsidada en una estructura nucleosómica. La distinta estructura de una región accesible puede detectarse frecuentemente por su sensibilidad a sondas químicas y enzimáticas, por ejemplo, nucleasas.

35 Un “sitio diana” o “secuencia diana” es una secuencia de ácidos nucleicos que define una porción de un ácido nucleico con el que se unirá una molécula de unión, proporcionado condiciones suficientes para que exista la unión. Por ejemplo, la secuencia 5'-GAATTC-3' es un sitio diana para la endonucleasa de restricción Eco RI.

40 Una molécula “exógena” es una molécula que no está normalmente presente en una célula, pero puede introducirse en una célula por uno o más métodos genéticos, bioquímicos u otros. La “presencia normal en la célula” se determina con respecto a la etapa de desarrollo particular y las condiciones medioambientales de la célula. Así, por ejemplo, una molécula que está presente solo durante el desarrollo embrionario de músculo es una molécula exógena con respecto a una célula de músculo adulta. Similarmente, una molécula inducida por choque térmico es una molécula exógena con respecto a una célula no de choque térmico. Una molécula exógena puede comprender, por ejemplo, una versión con funcionalidad de una molécula endógena disfuncional o una versión disfuncional de una molécula endógena que funciona normalmente.

45 Una molécula exógena puede ser, entre otras cosas, una molécula pequeña, tal como se genera por un proceso de química combinatoria, o una macromolécula tal como una proteína, ácido nucleico, hidrato de carbono, lípido, glicoproteína, lipoproteína, polisacárido, cualquier derivado modificado de las moléculas anteriores, o cualquier complejo que comprende una o más de las moléculas anteriores. Los ácidos nucleicos incluyen ADN y ARN, pueden ser mono- o bicatenarios; pueden ser lineales, ramificados o circulares; y pueden ser de cualquier longitud. Los ácidos nucleicos incluyen aquellos que pueden formar dúplex, además de ácidos nucleicos que forman tríplex. Véanse, por ejemplo, las patentes de EE.UU. nº 5.176.996 y 5.422.251. Las proteínas incluyen, pero no se limitan a, proteínas de unión a ADN, factores de transcripción, factores de remodelación de cromatina, proteínas de unión a ADN metilado, polimerasas, metilasas, desmetilasas, acetilasas, desacetilasas, cinasas, fosfatasa, integrasas, recombinasas, ligasas, topoisomerasas, girasas y helicasas.

50 Una molécula exógena puede ser del mismo tipo de molécula que una molécula endógena, por ejemplo, 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. Métodos para la introducción de moléculas exógenas en células son conocidos para aquellos expertos en la materia e incluyen, pero no se limitan a, transferencia mediada por lípidos (es decir, liposomas, que incluyen lípidos neutros y catiónicos), electroporación, inyección directa, fusión de células, bombardeo de partículas, co-precipitación con fosfato de calcio, transferencia mediada por DEAE-dextrano y transferencia mediada por vector

viral.

5 Por el contrario, una molécula “endógena” es una que normalmente está presente en una célula particular en una etapa de desarrollo particular bajo condiciones medioambientales particulares. Por ejemplo, un ácido nucleico endógeno puede comprender un cromosoma, el genoma de una mitocondria, cloroplasto u otro orgánulo, o un ácido nucleico episómico que se produce naturalmente. Moléculas endógenas adicionales pueden incluir proteínas, por ejemplo, factores de transcripción y enzimas.

10 Una molécula de “fusión” es una molécula en la que dos o más moléculas de subunidad están asociadas, preferentemente covalentemente. Las moléculas de subunidad pueden ser el mismo tipo químico de molécula, o pueden ser tipos químicos diferentes de moléculas. Ejemplos del primer tipo de molécula de fusión incluyen, pero no se limitan a, proteínas de fusión (por ejemplo, una fusión entre un dominio de unión de ADN de ZFP y un dominio de escisión) y ácidos nucleicos de fusión (por ejemplo, un ácido nucleico que codifica la proteína de fusión descrita arriba). Ejemplos del segundo tipo de molécula de fusión incluyen, pero no se limitan a, una fusión entre un ácido nucleico formador de triplex y un polipéptido, y una fusión entre un ligando de unión al surco menor y un ácido nucleico.

15 La expresión de una proteína de fusión en una célula puede resultar de la administración de la proteína de fusión a la célula o por administración de un polinucleótido que codifica la proteína de fusión a una célula, en la que el polinucleótido se transcribe, y el transcrito se traduce, para generar la proteína de fusión. También pueden participar corte y empalme in trans, escisión de polipéptidos y ligación de polipéptidos en la expresión de una proteína en una célula. Métodos para la administración de polinucleótidos y polipéptidos a células se presentan en cualquier parte en la presente divulgación.

20 Un “gen”, para los fines de la presente divulgación, incluye una región de ADN que codifica un producto génico (véase más adelante), además de todas las ADN regiones que regulan la producción del producto génico, tanto si tales secuencias reguladoras son adyacentes o no a secuencias de codificación y/o 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 al ribosoma y sitios internos de entrada al ribosoma, potenciadores, silenciadores, aislantes, elementos de frontera, orígenes de replicación, sitios de unión a la matriz y regiones de control de locus.

25 “Expresión génica” se refiere a la conversión de la información, contenida en un gen, en un producto génico. Un producto génico puede ser el producto de la transcripción directa de un gen (por ejemplo, ARNm, ARNt, ARNr, ARN antisentido, ribozima, ARN estructural o cualquier otro tipo de ARN) o una proteína producida por la traducción de un ARNm. Productos génicos también incluyen ARN que se modifican, por procesos tales como polarización, poliadenilación, metilación y edición, y proteínas modificadas por, por ejemplo, metilación, acetilación, fosforilación, ubiquitinación, ribosilación de ADP, miristilación y glicosilación.

30 “Modulación” de la expresión génica se refiere a un cambio en la actividad de un gen. La modulación de la expresión puede incluir, pero no se limita a, activación génica y represión génica.

35 Células “eucariotas” incluyen, pero no se limitan a, células fúngicas (tales como levadura), células vegetales, células de animales, células de mamífero y células humanas.

40 Una “región de interés” es cualquier región de cromatina celular, tal como, por ejemplo, un gen o una secuencia no codificante dentro de o adyacente a un gen, en la que se desea unir una molécula exógena. La unión puede ser para los fines de escisión de ADN dirigida y/o recombinación dirigida. Una región de interés puede estar presente en, por ejemplo, un cromosoma, un episoma, un genoma organular (por ejemplo, mitocondria, cloroplasto), o un genoma viral infectante. Una región de interés puede estar dentro de la región codificante de un gen, dentro de regiones no codificantes transcritas tales como, por ejemplo, secuencias conductoras, secuencias de transporte o intrones, o dentro de regiones no transcritas, tanto en la dirección 5' como en la dirección 3' de la región codificante. Una región de interés puede ser tan pequeña como un único par de nucleótidos o hasta 2.000 pares de nucleótidos de longitud, o cualquier valor entero de pares de nucleótidos.

45 Los términos “enlace operativo” y “operativamente ligado” (u “operativamente ligado”) se usan indistintamente con referencia a una yuxtaposición de dos o más componentes (tales como elementos de secuencia), en los que los componentes están dispuestos de forma que ambos componentes funcionan normalmente y permiten la posibilidad de que al menos uno de los componentes pueda mediar en una función que es ejercida por al menos uno de los otros componentes. A modo de ilustración, una secuencia reguladora de la transcripción, tal como un promotor, está operativamente ligada 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 de la transcripción. Una secuencia reguladora de la transcripción está generalmente operativamente ligada en *cis* con una secuencia codificante, pero no necesita ser directamente adyacente a ella. Por ejemplo, un potenciador es una secuencia reguladora transcripcional que está operativamente ligada a una secuencia codificante, aún cuando no sean contiguas.

Con respecto a polipéptidos de fusión, el término “operativamente ligado” puede referirse al hecho de que cada uno de los componentes realice la misma función en el enlace con el otro componente que haría si no estuviera así ligado. Por ejemplo, con respecto a un polipéptido de fusión en el que un dominio de unión de ADN de ZFP está fusionado con un dominio de escisión, el dominio de unión de ADN de ZFP y el dominio de escisión están en enlace operativo si, en el polipéptido de fusión, la porción del dominio de unión de ADN de ZFP puede unirse a su sitio diana y/o su sitio de unión, mientras que el dominio de escisión puede escindir ADN en la proximidad del sitio diana.

Un “fragmento funcional” de una proteína, polipéptido o ácido nucleico es una proteína, polipéptido o ácido nucleico cuya secuencia no es idéntica a la proteína, polipéptido o ácido nucleico de longitud completa, pero retiene la misma función que la proteína, polipéptido o ácido nucleico de longitud completa. Un fragmento funcional puede poseer más, menos, o el mismo número de residuos que la molécula nativa correspondiente, y/o puede contener uno o más sustituciones de aminoácidos o de nucleótidos. Métodos para determinar la función de un ácido nucleico (por ejemplo, función codificante, capacidad para hibridarse con otro ácido nucleico) son muy conocidos en la técnica. Similarmente, métodos para determinar la función de proteínas son muy conocidos. Por ejemplo, la función de unión del ADN de un polipéptido puede determinarse, por ejemplo, por unión a filtro, desplazamiento de la movilidad electroforética o ensayos de inmunoprecipitación. La escisión de ADN puede ensayarse por electroforesis en gel. Véase Ausubel et al., arriba. La capacidad de una proteína para interactuar con otra proteína puede determinarse, por ejemplo, por co-inmunoprecipitación, ensayos de dos híbridos o complementación, tanto genética como bioquímica. Véase, por ejemplo, Fields et al., (1989) Nature 340:245-246; patente de EE.UU. n° 5.585.245 y documento PCT WO 98/44350.

Sitios diana

Los métodos y composiciones desvelados incluyen proteínas de fusión que comprenden un dominio de escisión (o un medio dominio de escisión) y un dominio de dedos de cinc, en los que el dominio de dedos de cinc, que se une a una secuencia en cromatina celular (por ejemplo, un sitio diana o un sitio de unión), dirige la actividad del dominio de escisión (o medio dominio de escisión) a la proximidad de la secuencia y, por tanto, induce la escisión en la proximidad de la secuencia diana. Como se expone en cualquier parte en la presente divulgación, un dominio de dedos de cinc puede manipularse para unirse a prácticamente cualquier secuencia deseada. Por consiguiente, después de identificar una región de interés que contiene una secuencia en la que se desea la escisión o recombinación, uno o más dominios de unión de dedos de cinc pueden manipularse para unirse 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 cinc y un dominio de escisión (o de dos proteínas de fusión, que comprende cada una un dominio de unión de dedos de cinc y un medio dominio de escisión), en una célula, afecta la escisión en la región de interés.

La selección de una secuencia en cromatina celular para la unión por un dominio de dedos de cinc (por ejemplo, un sitio diana) puede llevarse a cabo, por ejemplo, según los métodos desvelados en la patente de EE.UU. del mismo solicitante n° 6.453.242 (17 de septiembre de 2002), que también desvela métodos para diseñar ZFP para unirse a una secuencia seleccionada. Será evidente para aquellos expertos en la materia que la simple inspección visual de una secuencia de nucleótidos también puede usarse para la selección de un sitio diana. Por consiguiente, puede usarse cualquier medio para la selección de sitios diana en los métodos reivindicados.

Los sitios diana están generalmente compuestos de una pluralidad de subsitios diana adyacentes. Un subsitio diana se refiere a la secuencia (normalmente tanto un triplete de nucleótidos, como un cuadruplete de nucleótidos que pueden solaparse por un nucleótido con un cuadruplete adyacente) unida por un dedo de cinc individual. Véase, por ejemplo, el documento WO 02/077227. Si la hebra con la que una proteína de dedos de cinc hace la mayoría de los contactos se designa la hebra diana “hebra de reconocimiento primario”, o “hebra de contacto primario”, algunas proteínas de dedos de cinc se unen a un triplete de tres bases en la hebra diana y una cuarta base sobre la hebra no diana. Un sitio diana generalmente tiene una longitud de al menos 9 nucleótidos y, por consiguiente, está unido por un dominio de unión de dedos de cinc que comprende al menos tres dedos de cinc. Sin embargo, también es posible la unión de, por ejemplo, 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. Como será evidente, también es posible la unión de dominios de unión más grandes (por ejemplo, 7, 8, 9 dedos y más) a sitios diana más largos.

No es necesario que un sitio diana sea un múltiplo de tres nucleótidos. Por ejemplo, en casos en los que se producen interacciones de hebras cruzadas (véase, por ejemplo, la patente de EE.UU. 6.453.242 y el documento WO 02/077227), uno o más de los dedos de cinc individuales de un dominio de unión de múltiples dedos puede unirse a subsitios de cuadruplete que se solapan. Como resultado, una proteína de tres dedos puede unirse a una secuencia de 10 nucleótidos, en la que 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 la que el decimotercero nucleótido es parte de un cuadruplete unido por un dedo terminal, etc.

La longitud y naturaleza de secuencias conectoras de aminoácidos entre dedos de cinc individuales en un dominio de unión de múltiples dedos también afecta la unión a una secuencia diana. Por ejemplo, la presencia de un llamado

“conector no canónico”, “conector largo” o “conector estructurado” entre dedos de cinc adyacentes en un dominio de unión de múltiples dedos puede permitir que aquellos dedos se unan a subsitios que no están inmediatamente adyacentes. Ejemplos no limitantes de tales conectores se describen, por ejemplo, en la patente de EE.UU. n° 6.479.626 y el documento WO 01/53480. Por consiguiente, uno o más subsitios, en un sitio diana para un dominio de unión de dedos de cinc, pueden separarse entre sí por 1, 2, 3, 4, 5 o más nucleótidos. Pero para proporcionar un ejemplo, un dominio de unión de cuatro dedos puede unirse a un sitio diana de 13 nucleótidos que comprende, en secuencia, dos subsitios de 3 nucleótidos contiguos, un nucleótido intermedio y dos subsitios de triplete contiguos.

La distancia entre secuencias (por ejemplo, sitios diana) se refiere al número de nucleótidos o pares de nucleótidos intermedios entre dos secuencias, como se mide desde los bordes de las secuencias más próximas entre sí.

En ciertas realizaciones en las que la escisión depende de la unión de dos moléculas de fusión de dominio de dedos de cinc/medio dominio de escisión para separar sitios diana, los dos sitios diana pueden estar sobre hebras de ADN opuestas. En otras realizaciones, ambos sitios diana están sobre la misma hebra de ADN.

Dominios de unión de dedos de cinc

Un dominio de unión de dedos de cinc comprende uno o más dedos de cinc. Miller et al., (1985) EMBO J. 4:1609-1614; Rhodes (1993) Scientific American Feb.:56-65; patente de EE.UU. n° 6.453.242. Normalmente, un único dominio de dedos de cinc tiene aproximadamente 30 aminoácidos de longitud. Estudios estructurales han demostrado que cada dominio de dedos de cinc (motivo) contiene dos hojas beta (mantenidas 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 mediante coordinación de un átomo de cinc por las dos cisteínas y las dos histidinas.

Los dedos de cinc incluyen tanto dedos de cinc C_2H_2 canónicos (es decir, aquellos en los que el ión cinc está coordinado por dos residuos de cisteína y dos de histidina) y dedos de cinc no canónicos tales como, por ejemplo, dedos de cinc C_3H (aquellos en los que el ión cinc está coordinado por tres residuos de cisteína y un residuo de histidina) y dedos de cinc C_4 (aquellos en los que el ión cinc está coordinado por cuatro residuos de cisteína). Véase también el documento WO 02/057293.

Los dominios de unión de dedos de cinc pueden manipularse para unirse a una secuencia de elección. Véanse, por ejemplo, Beerli et al., (2002) Nature Biotechnol. 20:135-141; Pabo et al., (2001) Ann. Rev. Biochem. 70:313-340; Isalan et al., (2001) Nature Biotechnol. 19:656-660; Segal et al., (2001) Curr. Opin. Biotechnol. 12:632-637; Choo et al., (2000) Curr. Opin. Struct. Biol. 10:411-416. Un dominio de unión de dedos de cinc manipulado puede tener una especificidad de unión novedosa, en comparación con una proteína de dedos de cinc que se produce naturalmente. Los métodos de manipulación incluyen, pero no se limitan a, diseño racional y diversos tipos de selección. El diseño racional incluye, por ejemplo, usar bases de datos que comprenden secuencias de triplete (o cuadruplete) de nucleótidos y secuencias de aminoácidos de dedos de cinc individuales, en las que cada secuencia de nucleótidos de triplete o cuadruplete está asociada con una o más secuencias de aminoácidos de dedos de cinc que se unen a la secuencia de triplete o cuadruplete particular. Véanse, por ejemplo, las patentes de EE.UU. del mismo solicitante 6.453.242 y 6.534.261.

Métodos de selección a modo de ejemplo, que incluyen expresión en fago y sistemas de dos híbridos, se desvelan en las patentes de EE.UU. 5.789.538; 5.925.523; 6.007.988; 6.013.453; 6.410.248; 6.140.466; 6.200.759; y 6.242.568; además de los documentos WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 y GB 2.338.237.

La mejora de la especificidad de unión por dominios de unión de dedos de cinc se ha descrito, por ejemplo, en el documento del mismo solicitante WO 02/077227.

Como un dedo de cinc 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 cinc adyacente), la longitud de una secuencia con la que se manipula un dominio de unión de dedos de cinc para unirse (por ejemplo, una secuencia diana) determinará el número de dedos de cinc en un dominio de unión de dedos de cinc manipulado. Por ejemplo, para ZFP en las que los motivos de dedos no se unen a subsitios de solapamiento, una secuencia diana de seis nucleótidos se une por un dominio de unión de dos dedos; una secuencia diana de nueve nucleótidos se une por un dominio de unión de tres dedos, etc. Como se observa en el presente documento, los sitios de unión para dedos de cinc 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 naturaleza de las secuencias de aminoácidos entre los dedos de cinc (es decir, los conectores entre dedos) en un dominio de unión de múltiples dedos.

En un dominio de unión de dedos de cinc de múltiples dedos, dedos de cinc adyacentes pueden estar separados por secuencias conectoras de aminoácidos de aproximadamente 5 aminoácidos (llamados conectores entre dedos “canónicos”) o, alternativamente, por uno o más conectores no canónicos. Véanse, por ejemplo, las patentes de EE.UU. n° del mismo solicitante 6.453.242 y 6.534.261. Para dominios de unión de dedos de cinc manipulados que

comprenden más tres dedos, puede preferirse la inserción de conectores entre dedos más largos (“no canónicos”) entre ciertos de los dedos de cinc ya que puede aumentar la afinidad y/o especificidad de unión por el dominio de unión. Véase, por ejemplo, la patente de EE.UU. n° 6.479.626 y documento WO 01/53480. Por consiguiente, también pueden caracterizarse dominios de unión de dedos de cinc de múltiples dedos con respecto a la presencia y localización de conectores entre dedos no canónicos. Por ejemplo, un dominio de unión de dedos de cinc de seis dedos que comprende tres dedos (unidos por dos conectores entre dedos canónicos), un conector largo y tres dedos adicionales (unidos por dos conectores entre dedos canónicos) se indica una configuración 2x3. Similarmente, un dominio de unión que comprende dos dedos (con un conector canónico entremedias), un conector largo y dos dedos adicionales (unidos por un conector canónico) se induce una proteína 2x2. Una proteína que comprende tres unidades de dos dedos (en cada una de las cuales los dos dedos se unen por un conector canónico), y en la que cada unidad de dos dedos está unida a la unidad de dos dedos adyacentes por un conector largo, se denomina una proteína 3x2.

La presencia de un conector largo o entre dedos no canónico entre dos dedos de cinc adyacentes en un dominio de unión de múltiples dedos frecuentemente permite que los dos dedos se unan a subsitios que no son inmediatamente contiguos en la secuencia diana. Por consiguiente, 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 están en contacto por un dedo de cinc. Por ejemplo, un dominio de unión de dedos de cinc 2x2 puede unirse a dos secuencias de seis nucleótidos separadas por un nucleótido, es decir, se une a un sitio diana de 13 nucleótidos. Véase también Moore et al., (2001a) Proc. Natl. Acad. Sci. USA 98:1432-1436; Moore et al., (2001b) Proc. Natl. Acad. Sci. USA 98:1437-1441 y el documento WO 01/53480.

Como se ha mencionado previamente, un subsitio diana es una secuencia de tres o cuatro nucleótidos que se une por un único dedo de cinc. Para ciertos fines, una unidad de dos dedos se indica un módulo de unión. Un módulo de unión puede obtenerse, por ejemplo, seleccionando dos dedos adyacentes en el contexto de una proteína de múltiples dedos (generalmente tres dedos) que se unen a una secuencia diana de seis nucleótidos particular. Alternativamente, pueden construirse módulos por ensamblaje de dedos de cinc individuales. Véanse también los documentos WO 98/53057 y WO 01/53480.

30 **Dominios de escisión**

La porción del dominio de escisión de las proteínas de fusión desveladas en el presente documento puede obtenerse de cualquier endo- o exonucleasa. Endonucleasas a modo de ejemplo de las que puede derivarse un dominio de escisión incluyen, pero no se limitan a, endonucleasas de restricción y endonucleasas de asentamiento. Véase, por ejemplo, 2002-2003 Catalogue, New England Biolabs, Beverly, MA; y Belfort et al., (1997) Nucleic Acids Res. 25:3379-3388. Se conocen enzimas adicionales que escinden ADN (por ejemplo, nucleasas S1; nucleasa de judías mungo; DNasa I pancreática; nucleasa microcócica; endonucleasa HO de levadura; véase también Linn et al., (eds.) Nucleases, Cold Spring Harbor Laboratory Press, 1993). Una o más de estas enzimas (o fragmentos funcionales de las mismas) pueden usarse como fuente de dominios de escisión y medios dominios de escisión.

Similarmente, un medio dominio de escisión (por ejemplo, proteínas de fusión que comprenden un dominio de unión de dedos de cinc y un medio dominio de escisión) pueden derivarse de cualquier nucleasa o porción de la misma, como se expone 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. Alternativamente, puede usarse una única proteína que comprende dos 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 preferentemente dispuestos, el uno con respecto al otro, tal que la unión de las dos proteínas de fusión ponga los medios dominios de escisión en una orientación espacial entre sí que permita que los medios dominios de escisión formen un dominio de escisión funcional, por ejemplo, por dimerización. Así, en ciertas realizaciones, los bordes próximos de los sitios diana están separados por 5-8 nucleótidos o por 15-18 nucleótidos. Sin embargo, puede interponerse cualquier número entero de nucleótidos o pares de nucleótidos entre dos sitios diana (por ejemplo, de 2 a 50 nucleótidos o más). En general, el punto de escisión se encuentra entre los sitios diana.

Las endonucleasas de restricción (enzimas de restricción) están presentes en muchas especies y son capaces de unión específica de secuencia a ADN (en un sitio de reconocimiento), y escindir ADN en o cerca del sitio de unión. Ciertas enzimas de restricción (por ejemplo, tipo IIS) escinden ADN en sitios eliminados del sitio de reconocimiento y tienen dominios de unión y de escisión separables. Por ejemplo, la enzima de tipo IIS Fok I cataliza la escisión de ADN bicatenario, en 9 nucleótidos desde su sitio de reconocimiento sobre una cadena y 13 nucleótidos desde su sitio de reconocimiento sobre la otra. Véanse, por ejemplo, las patentes de EE.UU. 5.356.802; 5.436.150 y 5.487.994; además de Li et al., (1992) Proc. Natl. Acad. Sci. USA 89:4275-4279; Li et al., (1993) Proc. Natl. Acad. Sci. USA 90:2764-2768; Kim et al., (1994a) Proc. Natl. Acad. Sci. USA 91:883-887; Kim et al., (1994b) J. Biol. Chem. 269:31,978-31,982. Así, en una realización, las proteínas de fusión comprenden el dominio de escisión (o medio dominio de escisión) desde al menos una enzima de restricción de tipo IIS y uno o más dominios de unión de dedos de cinc, que pueden o pueden no estar manipulados.

Una Enzima de restricción de tipo IIS a modo de ejemplo, cuyo dominio de escisión es separable del dominio de unión, es *Fok I*. Esta enzima particular es activa como dímero. Bitinaite et al., (1998) Proc. Natl. Acad. Sci. USA 95: 10.570-10.575. Por consiguiente, para los fines de la presente divulgación, la porción de la enzima *Fok I* usada en las proteínas de fusión desveladas se considera un medio dominio de escisión. Así, para la escisión bicatenaria dirigida y/o sustitución dirigida de secuencias celulares usando fusiones de dedo de cinc-*Fok I*, pueden usarse dos proteínas de fusión, cada una que comprende un medio dominio de escisión de *FokI*, para reconstituir un dominio de escisión catalíticamente activo. Alternativamente, también pueden usarse una única molécula de polipéptido que contiene un dominio de unión de dedos de cinc y dos medios dominios de escisión de *Fok I*. Parámetros para la escisión dirigida y alteración de secuencias dirigida usando fusiones de dedo de cinc-*Fok I* se proporcionan en cualquier parte en la presente divulgación.

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

Enzimas de restricción de tipo IIS a modo de ejemplo se enumeran en la Tabla 1. Enzimas de restricción adicionales también contienen dominios de unión y de escisión separables, y éstos se contemplan por la presente divulgación. Véase, por ejemplo, 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
25	Aci I	BstF5 I	Sts I
	Alo I	Btr I	TspDT I
	Bae I	Bts I	TspGW I
	Bbr7 I	Cdi I	Tth111 II
30	Bbv I	CjeP I	UbaP I
	Bbv II	Drd II	Bsa I
	BbvC I	Eci I	BsmB I
	Bcc I	Eco31 I	
35	Bce83 I	Eco57 I	
	BceA I	Eco57M I	
	Bcef I	Esp3 I	
	Bcg I	Fau I	
40	BciV I	Fin I	
	Bfi I	Fok I	
	Bin I	Gdi II	
	Bmg I	Gsu I	
45	Bpu10 I	Hga I	
	BsaX I	Hin4 II	
	Bsb I	Hph I	
	BscA I	Ksp632 I	
50	BscG I	Mbo II	
	BseR I	Mly I	
	BseY I	Mme I	
	Bsi I		Mn1 I
55	Bsm I		Pfl1108 I
	BsmA I		Ple I
	BsmF I		Ppi I
	Bsp24 I		Psr I
60	BspG I		RleA I
	BspM I		Sap I
	BspNC I		SfaN I
65	Bsr I		Sim I

Fusiones de dominio de dedos de cinc-dominio de escisión

Métodos para el diseño y construcción de proteínas de fusión (y polinucleótidos que codifican las mismas) son conocidos para aquellos expertos en la materia. Por ejemplo, métodos para el diseño y construcción de proteína de fusión que comprende proteínas de dedos de cinc (y polinucleótidos que codifican las mismas) se describen en las patentes de EE.UU. del mismo solicitante 6.453.242 y 6.534.261. En ciertas realizaciones, se construyen 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 divulgación adicional referente a vectores y métodos para introducir polinucleótidos en células).

En ciertas realizaciones de los métodos descritos en el presente documento, una proteína de fusión comprende un dominio de unión de dedos de cinc y un medio dominio de escisión de la enzima de restricción *Fok I*, y dos proteínas de fusión tales se expresan en una célula. La expresión de dos proteínas de fusión en una célula puede resultar de la administración de las dos proteínas a la célula; administración de una proteína y un ácido nucleico que codifica una de las proteínas a la célula; administración de dos ácidos nucleicos, codificando cada una una de las proteínas, a la célula; o por administración de un único ácido nucleico, que codifica ambas proteínas, a la célula. En realizaciones adicionales, una proteína de fusión comprende una única cadena de polipéptidos que comprende dos medios dominios de escisión y un dominio de unión de dedos de cinc. En este caso, una única proteína de fusión se expresa en una célula y, sin desear ceñirse a teoría alguna, se cree que escindir ADN como resultado de la formación de un dímero intramolecular de los medios dominios de escisión.

En ciertas realizaciones, los componentes de las proteínas de fusión (por ejemplo, fusiones de ZFP-*Fok I*) están dispuestos de forma que el dominio de dedos de cinc sea el más próximo al extremo amino de la proteína de fusión, y el medio dominio de escisión sea el más próximo al extremo carboxi. Esto refleja la orientación relativa del dominio de escisión en dominios de escisión de dimerización que se producen naturalmente tales como aquellos derivados de la enzima *Fok I*, en la que el dominio de unión a ADN es el más próximo al extremo amino y el medio dominio de escisión es el más próximo al extremo carboxi. En estas realizaciones, la dimerización de los medios dominios de escisión para formar una nucleasa funcional se provoca uniendo las proteínas de fusión a sitios sobre hebras de ADN opuestas, siendo los extremos 5' de los sitios de unión proximales entre sí. Véase la Figura 43A.

En realizaciones adicionales, los componentes de las proteínas de fusión (por ejemplo, fusiones de ZFP-*Fok I*) se disponen de forma que el medio dominio de escisión sea el más próximo al extremo amino de la proteína de fusión, y el dominio de dedos de cinc sea el más próximo al extremo carboxi. En estas realizaciones, la dimerización de los medios dominios de escisión para formar una nucleasa funcional se provoca uniendo las proteínas de fusión a sitios sobre hebras de ADN opuestas, siendo los extremos 3' de los sitios de unión proximales entre sí. Véase la Figura 43B.

En otras realizaciones adicionales, una primera proteína de fusión contiene el medio dominio de escisión más próximo al extremo amino de la proteína de fusión, y el dominio de dedos de cinc más próximo al extremo carboxi, y una segunda proteína de fusión está dispuesta de forma que el dominio de dedos de cinc sea el más próximo al extremo amino de la proteína de fusión, y el medio dominio de escisión sea el más próximo al extremo carboxi. En estas realizaciones, ambas proteínas de fusión se unen a la misma hebra de ADN, conteniendo el sitio de unión de la primera proteína de fusión el dominio de dedos de cinc más próximo al extremo carboxi localizado con respecto al lado 5' del sitio de unión de la segunda proteína de fusión que contiene el dominio de dedos de cinc más próximo al extremo amino. Véase la Figura 43C.

En las proteínas de fusión desveladas, la secuencia de aminoácidos entre el dominio de dedos de cinc y el dominio de escisión (o medio dominio de escisión) se indica el "conector de ZC". El conector de ZC debe distinguirse de los conectores entre dedos tratados anteriormente. Para los fines de determinar la longitud de un conector de ZC, se usa la estructura de dedo de cinc descrita por Pabo et al., (2001) Ann. Rev. Biochem. 70:313-340:

X-X-C-X_{2,4}-C-X₁₂-H-X_{3,5}-H (SEC ID N°: 201)

En esta estructura, el primer residuo de un dedo de cinc es el aminoácido localizado dos residuos del extremo amino con respecto al primer residuo de cisteína conservado. En la mayoría de las proteínas de dedos de cinc que se producen naturalmente, esta posición está ocupada por un aminoácido hidrófobo (normalmente tanto fenilalanina como tirosina). En las proteínas de fusión desveladas, el primer residuo de un dedo de cinc será así frecuentemente un residuo hidrófobo, pero puede ser cualquier aminoácido. El residuo de aminoácido final de un dedo de cinc, como se muestra anteriormente, es el segundo residuo de histidina conservado.

Así, en las proteínas de fusión desveladas que tienen una polaridad en la que el dominio de unión de dedos de cinc está en el extremo amino con respecto al dominio de escisión (o medio dominio de escisión), el conector de ZC es la secuencia de aminoácidos entre el segundo residuo de histidina conservado del dedo de cinc más terminal del extremo C y el aminoácido más terminal del extremo N del dominio de escisión (o medio dominio de escisión). Por ejemplo, en ciertas proteínas de fusión cuya construcción se ejemplifica en los ejemplos sección, el aminoácido más terminal del extremo N de un medio dominio de escisión es un residuo de glutamina (Q) correspondiente al número

de aminoácido 384 en la secuencia de *FokI* de Looney et al., (1989) Gene 80:193-208.

Para proteínas de fusión que tienen una polaridad en la que el dominio de escisión (o medio dominio de escisión) está en el extremo amino con respecto al dominio de unión de dedos de cinc, el conector de ZC es la secuencia de aminoácidos entre el residuo de aminoácido más terminal del extremo C del dominio de escisión (o medio dominio) y el primer residuo del dedo de cinc más terminal del extremo N del dominio de unión de dedos de cinc (es decir, el residuo localizado dos residuos en la dirección 5' del primer residuo de cisteína conservado). En ciertas proteínas de fusión a modo de ejemplo, el aminoácido más terminal del extremo C de un medio dominio de escisión es un residuo de fenilalanina (F) correspondiente al número de aminoácido 579 en la secuencia de *FokI* de Looney et al., (1989) Gene 80:193-208.

El conector de ZC puede ser cualquier secuencia de aminoácidos. Para obtener escisión óptima, la longitud del conector de ZC y la distancia entre los sitios diana (sitios de unión) están relacionadas entre sí. Véase, por ejemplo, Smith et al., (2000) Nucleic Acids Res. 28:3361-3369; Bibikova et al., (2001) Mol. Cell. Biol. 21:289-297, señalando que su notación para la longitud de conectores se diferencia de la dada aquí. Por ejemplo, para fusiones de ZFP-*FokI* en las que el dominio de unión de dedos de cinc está en el extremo amino con respecto al medio dominio de escisión, y que tiene una longitud de conector de ZC de cuatro aminoácidos como se define en el presente documento (y se indica L0 por otros), la escisión óptima se produce cuando los sitios de unión para las proteínas de fusión se localizan 6 ó 16 nucleótidos separados (como se mide desde el borde próximo de cada sitio de unión). Véase el Ejemplo 4.

Métodos de escisión dirigida

Los métodos y composiciones desvelados pueden usarse para escindir ADN en una región de interés en cromatina celular (por ejemplo, en un sitio deseado o predeterminado en un genoma, por ejemplo, en un gen, tanto mutante como no mutante). Para tal escisión de ADN dirigida, un dominio de unión de dedos de cinc se manipula para unirse a un sitio diana en o cerca del sitio predeterminado de escisión, y una proteína de fusión que comprende el dominio de unión de dedos de cinc manipulado y un dominio de escisión se expresa en una célula. Tras la unión de la porción de dedo de cinc de la proteína de fusión al sitio diana, el ADN se escinde próximo al sitio diana por el dominio de escisión. El sitio de escisión exacto puede depender de la longitud del conector de ZC.

Alternativamente, dos proteínas de fusión, que comprende cada una un dominio de unión de dedos de cinc y un medio dominio de escisión, se expresan en una célula, y se unen a sitios diana que están yuxtapuestos de tal forma que se reconstituye un dominio de escisión funcional y se escinde ADN en la proximidad de los sitios diana. En una realización, la escisión se produce entre los sitios diana de los dos dominios de unión de dedos de cinc. Pueden manipularse uno o ambos de los dominios de unión de dedos de cinc.

Para la escisión dirigida usando un polipéptido de fusión de dominio de unión de dedos de cinc-dominio de escisión, el sitio de unión puede englobar el sitio de escisión, o el borde próximo del sitio de unión puede tener 1, 2, 3, 4, 5, 6, 10, 25, 50 o más nucleótidos (o cualquier valor entero entre 1 y 50 nucleótidos) desde el sitio de escisión. La localización exacta del sitio de unión, con respecto al sitio de escisión, dependerá del dominio de escisión particular, y la longitud del conector de ZC. Para métodos en los que se usan dos polipéptidos de fusión, que comprenden cada uno un dominio de unión de dedos de cinc y un medio dominio de escisión, los sitios de unión generalmente cubren el sitio de escisión. Así, el borde próximo del primer sitio de unión puede tener 1, 2, 3, 4, 5, 6, 10, 25 o más nucleótidos (o cualquier valor entero entre 1 y 50 nucleótidos) sobre un lado del sitio de escisión, y el borde próximo del segundo sitio de unión puede tener 1, 2, 3, 4, 5, 6, 10, 25 o más nucleótidos (o cualquier valor entero entre 1 y 50 nucleótidos) sobre el otro lado del sitio de escisión. Métodos para mapear sitios de escisión *in vitro* e *in vivo* son conocidos para aquellos expertos en la materia.

Así, los métodos descritos en el presente documento pueden emplear un dominio de unión de dedos de cinc manipulado fusionado con un dominio de escisión. En estos casos, el dominio de unión se manipula para unirse a una secuencia diana, en o cerca de la cual se desea la escisión. La proteína de fusión, o un polinucleótido que codifica la misma, se introduce en una célula. Una vez introducida en, 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 de escisión exacto depende de la naturaleza del dominio de escisión y/o la presencia y/o naturaleza de secuencias conectoras entre los dominios de unión y de escisión. En casos en los que se usan dos proteínas de fusión, que comprenden cada una un medio dominio de escisión, la distancia entre los bordes próximos de los sitios de unión pueden ser 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 25 o más nucleótidos (o cualquier valor entero entre 1 y 50 nucleótidos). Niveles de escisión óptimos también pueden depender de tanto la distancia entre los sitios de unión de las dos proteínas de fusión (véase, por ejemplo, Smith et al., (2000) Nucleic Acids Res. 28:3361-3369; Bibikova et al., (2001) Mol. Cell. Biol. 21:289-297) y la longitud del conector de ZC en cada proteína de fusión.

Para nucleasas de fusión de ZFP-*FokI*, la longitud del conector entre la ZFP y el medio dominio de escisión de *FokI* (es decir, el conector de ZC) puede influir en la eficiencia de escisión. En un sistema experimental que utiliza una fusión de ZFP-*FokI* con un conector de ZC de 4 residuos de aminoácidos se obtuvo escisión óptima cuando los bordes próximos de los sitios de unión para dos nucleasas ZFP-*FokI* se separaron por 6 pares de bases. Esta

nucleasa de fusión particular comprendió la siguiente secuencia de aminoácidos entre la porción de dedo de cinc y el medio dominio de nucleasa:

HQRTHQNKKQLV (SEC ID N°: 26)

5 en la que están subrayadas las dos histidinas conservadas en la porción del extremo C del dedo de cinc y los tres primeros residuos en el medio dominio de escisión de *FokI*. Por consiguiente, la secuencia del conector de ZC de esta construcción es QNKK. Bibikova et al., (2001) Mol. Cell. Biol. 21:289-297. Los presentes inventores han construido varias nucleasas de fusión de ZFP-*FokI* que tienen una variedad de longitudes y secuencias de conectores de ZC, y analizado las eficiencias de escisión de estas nucleasas sobre una serie de sustratos que tienen diferentes distancias entre los sitios de unión de ZFP. Véase el Ejemplo 4.

15 En ciertas realizaciones, el dominio de escisión comprende dos medios dominios de escisión, ambos de 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, mientras que sirven para escindir el ADN.

20 También pueden proporcionarse medios dominios de escisión en moléculas separadas. Por ejemplo, pueden introducirse dos polipéptidos de fusión en una célula, en la que 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, mientras que sirven para escindir el ADN. Además, los dominios de unión se unen a secuencias diana que normalmente están dispuestas de tal forma que, tras la unión de los polipéptidos de fusión, los dos medios dominios de escisión se presentan en una orientación espacial entre sí que permita la reconstitución de un dominio de escisión (por ejemplo, por dimerización de los medios dominios), posicionándose así los medios dominios el uno con respecto al otro para formar un dominio de escisión funcional, produciendo la escisión de cromatina celular en una región de interés. Generalmente, la escisión por el dominio de escisión reconstituido se produce en un sitio localizado entre las dos secuencias diana. Una o ambas de las proteínas puede manipularse para unirse a su sitio diana.

30 Las dos proteínas de fusión pueden unirse en la región de interés en la misma polaridad o polaridad opuesta, y sus sitios de unión (es decir, sitios diana) pueden separarse por cualquier número de nucleótidos, por ejemplo, de 0 a 200 nucleótidos o cualquier valor entero entremedias. En ciertas realizaciones, los sitios de unión para dos proteínas de fusión, que comprenden cada una un dominio de unión de dedos de cinc y un medio dominio de escisión, pueden localizarse separados entre 5 y 18 nucleótidos, por ejemplo, separados 5-8 nucleótidos, o separados 15-18 nucleótidos, o separados 6 nucleótidos, o separados 16 nucleótidos, como se mide desde el borde de cada sitio de unión más próximo al otro sitio de unión, y la escisión se produce entre los sitios de unión.

40 El sitio en el que el ADN se escinde generalmente se encuentra entre los sitios de unión para las dos proteínas de fusión. La rotura bicatenaria de ADN frecuentemente resulta de dos roturas monocatenarias, o "cortes", compensados por 1, 2, 3, 4, 5, 6 o más nucleótidos (por ejemplo, la escisión de ADN bicatenario por *FokI* nativa resulta de roturas monocatenarias compensadas por 4 nucleótidos). Así, la escisión no se produce necesariamente en sitios exactamente opuestos sobre cada hebra de ADN. Además, la estructura de las proteínas de fusión y la distancia entre los sitios diana puede influir en si la escisión se produce adyacente a un único par de nucleótidos, o si la escisión se produce en varios sitios. Sin embargo, para muchas aplicaciones, que incluyen recombinación dirigida y mutagénesis dirigida (véase más adelante), la escisión dentro de un intervalo de nucleótidos es generalmente suficiente, y no se requiere la escisión entre pares de bases particulares.

50 Como se observa anteriormente, la(s) proteína(s) de fusión puede(n) introducirse como polipéptidos y/o polinucleótidos. Por ejemplo, pueden introducirse dos polinucleótidos, que comprenden cada uno secuencias que codifican uno de los polipéptidos anteriormente mencionados, en una célula, y cuando los polipéptidos se expresan y cada uno se une a su secuencia diana, la escisión se produce en o cerca de la secuencia diana. Alternativamente, un único 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.

55 Para potenciar la especificidad por escisión, también pueden emplearse composiciones adicionales en los métodos descritos en el presente documento. Por ejemplo, medios dominios de escisión individuales pueden presentar actividad de escisión bicatenaria limitada. En métodos en los que se introducen dos proteínas de fusión, que contiene cada una un dominio de dedos de cinc de tres dedos y un medio dominio de escisión, en la célula, cualquier proteína específica un sitio diana de aproximadamente 9 nucleótidos. Aunque es probable que la secuencia diana total de 18 nucleótidos sea única en un genoma de mamífero, cualquier sitio diana de 9 nucleótidos dado se produce, en promedio, aproximadamente 23.000 veces en el genoma humano. Así, no puede producirse escisión específica, debido a la unión específica al sitio de un único medio dominio. Por consiguiente, los métodos descritos en el presente documento contemplan el uso de un mutante negativo dominante de un medio dominio de escisión tal como *FokI* (o un ácido nucleico que codifica el mismo) que se expresa en una célula junto con las dos proteínas de fusión. El mutante negativo dominante puede dimerizarse, pero es incapaz de escindir, y también bloquea la actividad de escisión de un medio dominio con el que se dimeriza. Proporcionando el mutante negativo dominante

en exceso molar a las proteínas de fusión, solo las regiones en las que ambas proteínas de fusión están unidas tendrán una concentración local suficientemente alta de medios dominios de escisión funcionales para que se produzca la dimerización y escisión. En sitios en los que solo se une una de las dos proteínas de fusión, su medio dominio de escisión forma un dímero con el medio dominio del mutante negativo dominante, y no se produce escisión no específica no deseable.

Se han identificado tres residuos de aminoácidos catalíticos en el medio dominio de escisión de *Fok I*: Asp 450, Asp 467 y Lys 469. Bitinaite et al., (1998) Proc. Natl. Acad. Sci. USA 95: 10,570-10,575. Así, puede usarse una o más mutaciones en uno de estos residuos para generar una mutación negativa dominante. Además, se conocen muchos de los residuos de aminoácidos catalíticos de otras endonucleasas de tipo IIS y/o puede determinarse, por ejemplo, por alineamiento con secuencias de *Fok I* y/o por generación y prueba de mutantes para actividad catalítica.

Mutaciones de dominios de dimerización en el medio dominio de escisión

Métodos para la escisión dirigida que implican el uso de fusiones entre una ZFP y un medio dominio de escisión (tales como, por ejemplo, una fusión de ZFP/*FokI*) requieren el uso de dos moléculas de fusión tales, cada una generalmente dirigida a una secuencia diana distinta. Pueden elegirse secuencias diana para las dos proteínas de fusión de manera que la escisión dirigida se dirija a un único sitio en un genoma, como se trata anteriormente. Una posible fuente de especificidad de escisión reducida podría resultar de la homodimerización de uno de las dos fusiones de ZFP/medio dominio de escisión. Esto podría producirse, por ejemplo, debido a la presencia, en un genoma, de repeticiones invertidas de las secuencias diana para una de las dos fusiones de ZFP/medio dominio de escisión, localizadas de manera que se permita que dos copias de la misma proteína de fusión se unan con una orientación y separación que permita la formación de un dímero funcional.

Un enfoque para reducir la probabilidad de este tipo de escisión anómala en secuencias distintas del sitio diana previsto implica generar variantes del medio dominio de escisión que minimizan o previenen la homodimerización. Preferentemente, uno o más aminoácidos en la región del medio dominio que participan en su dimerización están alterados. En la estructura cristalina del dímero de la proteína *FokI*, se informa que la estructura de los medios dominios de escisión es similar a la disposición de los medios dominios de escisión durante la escisión de ADN por *FokI*. Wah et al., (1998) Proc. Natl. Acad. Sci. USA 95:10564-10569. Esta estructura indica que los residuos de aminoácidos en las posiciones 483 y 487 desempeñan una función clave en la dimerización de los medios dominios de escisión de *FokI*. 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 todos suficientemente próximos a la interfase de dimerización para influir en la dimerización. Por consiguiente, alteraciones de la secuencia de aminoácidos en una o más de las posiciones anteriormente mencionadas alterarán probablemente las propiedades de dimerización del medio dominio de escisión. Tales cambios pueden introducirse, por ejemplo, construyendo una biblioteca que contiene (o que codifica) 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 no mutantes.

Por consiguiente, la alteración de un medio dominio de escisión de *FokI* en cualquier residuo de aminoácido que afecte la dimerización puede usarse para prevenir que una de un par de fusiones de ZFP/*FokI* experimente homodimerización que puede conducir a la escisión en secuencias no deseadas. Así, para la escisión dirigida usando un par de fusiones de ZFP/*FokI*, una o ambas de las proteínas de fusión pueden comprender una o más alteraciones de aminoácidos que inhiben la auto-dimerización, pero permiten que se produzca la heterodimerización de las dos proteínas de fusión de forma que la escisión se produzca 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 cualquier fusión, que conduce a la escisión anómala, se minimiza o suprime, mientras que la heterodimerización de las dos proteínas de fusión se facilita en comparación con la obtenida con medios dominios de escisión no mutantes. Véase el Ejemplo 5.

Métodos para la alteración dirigida de secuencias genómicas y recombinación dirigida

En el presente documento también se describen métodos de sustitución de una secuencia genómica (por ejemplo, una región de interés en cromatina celular) con una secuencia no idéntica homóloga (es decir, recombinación dirigida). Intentos previos por sustituir secuencias particulares han implicado poner en contacto una célula con un polinucleótido que comprende secuencias que llevan homología con una región cromosómica (es decir, un ADN donante), seguido de selección de células en las que la molécula de ADN donante ha experimentado recombinación homóloga en el genoma. La tasa de éxito de estos métodos es baja, debido a la mala eficiencia de recombinación homóloga y una alta frecuencia de inserción no específica del ADN donante en regiones del genoma distintas del sitio diana.

La presente divulgación proporciona métodos de alteración de secuencias dirigida caracterizados por una mayor eficiencia de recombinación dirigida y una menor frecuencia de eventos de inserción no específicos. Los métodos implican hacer y usar dominios de unión de dedos de cinc manipulados fusionados con dominios de escisión (o

medios dominios de escisión) para hacer una o más roturas bicatenarias dirigidas en ADN celular. Debido a que las roturas bicatenarias en ADN celular estimulan la recombinación homóloga varios miles de veces en la proximidad del sitio de escisión, tal escisión dirigida permite la alteración o sustitución (mediante recombinación homóloga) de secuencias en prácticamente cualquier sitio en el genoma.

Además de las moléculas de fusión descritas en el presente documento, la sustitución dirigida de una secuencia genómica seleccionada también requiere la introducción de la secuencia de sustitución (o donante). La secuencia donante puede introducirse en la célula antes de, simultáneamente con, o posterior a, la expresión de la(s) proteína(s) de fusión. El polinucleótido donante contiene homología suficiente con una secuencia genómica para soportar la recombinación homóloga entre ella y la secuencia genómica con la que tiene homología. Aproximadamente 25, 50 100 ó 200 nucleótidos o más de homología de secuencias entre una secuencia donante y una genómica (o cualquier valor entero entre 10 y 200 nucleótidos, o más) soportarán la recombinación homóloga entremedias. Las secuencias donantes puede oscilar en longitud de 10 a 5.000 nucleótidos (o cualquier valor entero de nucleótidos entremedias) o más. Será rápidamente evidente que la secuencia donante normalmente no es idéntica a la secuencia genómica que sustituye. Por ejemplo, la secuencia del polinucleótido donante puede contener uno o más cambios, inserciones, deleciones, inversiones o transposiciones de una sola base con respecto a la secuencia genómica, mientras que está presente suficiente homología para soportar la recombinación homóloga. Alternativamente, una secuencia donante puede contener una secuencia no homóloga flanqueada por dos regiones de homología. Adicionalmente, las secuencias donantes pueden comprender una molécula de vector que contiene secuencias que no son homólogas a la región de interés en cromatina celular. Generalmente, la(s) región (regiones) homóloga(s) de una secuencia donante tendrán al menos el 50 % de identidad de secuencias con una secuencia genómica con la que se desea recombinación. En ciertas realizaciones, está presente el 60 %, 70 %, 80 %, 90 %, 95 %, 98 %, 99 % o el 99,9 % de identidad de secuencias. Puede estar presente cualquier valor entre el 1 % y el 100 % de identidad de secuencias, dependiendo de la longitud del polinucleótido donante.

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

Para simplificar los ensayos (por ejemplo, hibridación, PCR, digestión con enzimas de restricción) para determinar la inserción satisfactoria de la secuencia donante, pueden estar presentes ciertas diferencias de secuencias en la secuencia donante en comparación con la secuencia genómica. Preferentemente, si está localizada en una región codificante, tales diferencias de la secuencia de nucleótidos no cambiarán la secuencia de aminoácidos, o harán cambios de aminoácidos silenciosos (es decir, cambios que no afectan a la estructura o función de la proteína). El polinucleótido donante puede contener opcionalmente cambios en secuencias correspondientes a sitios de unión del dominio de dedos de cinc en la región de interés, para prevenir la escisión de secuencias donantes que se han introducido en cromatina celular por recombinación homóloga.

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 donante pueden protegerse (por ejemplo, de la degradación exonucleolítica) por métodos conocidos para aquellos expertos en la materia. Por ejemplo, se añaden uno o más residuos de dideoxinucleótidos al extremo 3' de una molécula lineal y/o se ligan oligonucleótidos auto-complementarios a uno o ambos extremos. Véase, por ejemplo, Chang et al., (1987) Proc. Natl. Acad. Sci. USA 84:4959-4963; Nehls et al., (1996) Science 272:886-889. Métodos adicionales para proteger polinucleótidos exógenos de la degradación incluyen, pero no se limitan a, adición de grupo(s) amino terminal(es) y el uso de enlaces internucleotídicos modificados tales como, por ejemplo, fosforotioatos, fosforamidatos y residuos de O-metilribosa o desoxirribosa. Puede introducirse un polinucleótido en una célula como parte de una molécula de vector que tiene secuencias adicionales tales como, por ejemplo, orígenes de replicación, promotores y genes que codifican resistencia a antibióticos. Además, los polinucleótidos donantes pueden introducirse como ácido nucleico desnudo, como ácido nucleico complejo con un agente tal como un liposoma o poloxámero, o pueden administrarse por virus (por ejemplo, adenovirus, AAV).

Sin desear quedar ligado a teoría, parece que la presencia de una rotura bicatenaria en una secuencia celular, acoplada a la presencia de una molécula de ADN exógeno que tiene homología con una región adyacente a o que rodea la rotura, activa mecanismos celulares que reparan la rotura por transferencia de información de secuencias de la molécula donante en la secuencia celular (por ejemplo, genómica o cromosómica); es decir, por un proceso de recombinación homóloga. Los métodos de los solicitantes combinan ventajosamente las poderosas capacidades de direccionamiento de las ZFP manipuladas con un dominio de escisión (o medio dominio de escisión) para dirigir específicamente una rotura bicatenaria a la región del genoma en la que se desea recombinación.

Para la alteración de una secuencia cromosómica es necesario que la secuencia entera del donante sea copiada en el cromosoma, en tanto que sea copiado suficiente de la secuencia donante para efectuar la alteración de la secuencia deseada.

La eficiencia de inserción de secuencias donantes por recombinación homóloga está inversamente relacionada con

la distancia, en el ADN celular, entre la rotura bicatenaria y el sitio en el que se desea la recombinación. En otras palabras, se observan mayores eficiencias de recombinación homóloga cuando la rotura bicatenaria está más próxima al sitio en el que se desea la recombinación. En los casos en los que no se ha predeterminado un sitio de recombinación preciso (por ejemplo, el evento de recombinación deseado puede producirse sobre un intervalo de secuencia genómica), se seleccionan la longitud y secuencia del ácido nucleico donante, junto con el (los) sitio(s) de escisión, para obtener el evento de recombinación deseados. En los casos en los que el evento deseado se diseña para cambiar la secuencia de un único par de nucleótidos en una secuencia genómica, la cromatina celular se escinde dentro de 10.000 nucleótidos en cualquier lado de ese par de nucleótidos. En ciertas realizaciones, la escisión se produce dentro de 500, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5 ó 2 nucleótidos, o cualquier valor entero entre 2 y 1.000 nucleótidos, en cualquier lado del par de nucleótidos cuya secuencia va a cambiarse.

Como se ha detallado anteriormente, los sitios de unión para dos proteínas de fusión, que comprenden cada uno un dominio de unión de dedos de cinc y un medio dominio de escisión, pueden localizarse separados 5-8 ó 15-18 nucleótidos, como se mide desde el borde de cada sitio de unión más próximo al otro sitio de unión, y la escisión se produce entre los sitios de unión. Si la escisión se produce o no en un único sitio o en múltiples sitios entre los sitios de unión es irrelevante, ya que las secuencias genómicas escindidas están sustituidas con las secuencias donantes. Así, para la eficaz alteración de la secuencia de un único par de nucleótidos por recombinación dirigida, el punto medio de la región entre los sitios de unión está dentro de 10.000 nucleótidos de ese par de nucleótidos, preferentemente 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.

En ciertas realizaciones, un cromosoma homólogo puede servir de polinucleótido donante. Así, por ejemplo, la corrección de una mutación en un heterocigoto puede lograrse manipulando proteínas de fusión que se unen a y escinden la secuencia mutante en un cromosoma, pero no escinden la secuencia no mutante en el cromosoma homólogo. La rotura bicatenaria en el cromosoma que tiene la mutación estimula un proceso de "conversión génica" basado en homología en el que la secuencia no mutante del cromosoma homólogo se copia en el cromosoma escindido, restaurando así dos copias de la secuencia no mutante.

También se proporcionan métodos y composiciones que pueden potenciar los niveles de recombinación dirigida que incluyen, pero no se limitan a, el uso de fusiones de ZFP-dominio funcional adicionales para activar la expresión de genes que participan en recombinación homóloga, tales como, por ejemplo, miembros del grupo de epistasia RAD52 (por ejemplo, *Rad50*, *Rad51*, *Rad51B*, *Rad51C*, *Rad51D*, *Rad52*, *Rad54*, *Rad54B*, *Mre11*, *XRCC2*, *XRCC3*), genes cuyos productos interactúan con los productos génicos anteriormente mencionados (por ejemplo, BRCA1, BRCA2) y/o genes en el complejo de NBS1. Similarmente, pueden usarse fusiones de ZFP-dominio funcional, en combinación con los métodos y composiciones desvelados en el presente documento, para reprimir la expresión de genes que participan en la unión de extremos no homólogos (por ejemplo, Ku70/80, XRCC4, polimerasa poli(ADP ribosa), ADN ligasa 4). Véanse, por ejemplo, Yanez et al., (1998) *Gene Therapy* 5:149-159; Hoeijmakers (2001) *Nature* 411:366-374; Johnson et al., (2001) *Biochem. Soc. Trans.* 29:196-201; Tauchi et al., (2002) *Oncogene* 21:8967-8980. Métodos para la activación y represión de la expresión génica usando fusiones entre un dominio de unión de dedos de cinc y un dominio funcional se desvelan en la patente de EE.UU. del mismo solicitante nº 6.534.261. Métodos de represión adicionales incluyen el uso de oligonucleótidos antisentido y/o ARN interferente pequeño (ARNip o iARN) dirigido a la secuencia del gen que va a reprimirse.

Como una alternativa a o, además de, puede usarse la activación de la expresión de productos génicos que participa en la recombinación homóloga, fusiones de estas proteínas (o fragmentos funcionales de las mismas) con un dominio de unión de dedos de cinc dirigido a la región de interés, para reclutar estas proteínas (proteínas de recombinación) a la región de interés, aumentando así su concentración local y estimulando adicionalmente los procesos de recombinación homóloga. Alternativamente, un polipéptido que participa en la recombinación homóloga como se ha descrito anteriormente (o un fragmento funcional del mismo) puede ser parte de una proteína de fusión triple que comprende un dominio de unión de dedos de cinc, un dominio de escisión (o medio dominio de escisión) y la proteína de recombinación (o fragmento funcional de la misma). Proteínas adicionales que participan en la conversión génica y remodelación de cromatina relacionada con la recombinación, que pueden usarse en los métodos y composiciones anteriormente mencionados, incluyen histona acetiltransferasas (por ejemplo, Esalp, Tip60), histona metiltransferasas (por ejemplo, Dot1p), histona cinasas e histona fosfatasa.

Se ha informado que la proteína p53 desempeña una función fundamental en reprimir la recombinación homóloga (HR). Véanse, por ejemplo, Valerie et al., (2003) *Oncogene* 22:5792-5812; Janz, et al., (2002) *Oncogene* 21:5929-5933. Por ejemplo, la tasa de HR en líneas tumorales humanas deficientes en p53 es 10,000 veces superior a en fibroblastos humanos primarios, y hay un aumento de 100 veces en HR en células tumorales con una p53 no funcional en comparación con aquellas con p53 funcional. Mekeel et al., (1997) *Oncogene* 14:1847-1857. Además, la expresión en exceso de mutantes negativos dominantes de p53 conduce a un aumento de 20 veces en la recombinación espontánea. Bertrand et al., (1997) *Oncogene* 14:1117-1122. El análisis de diferentes mutaciones de p53 ha revelado que las funciones de p53 en la transactivación transcripcional y el control del punto de regulación del ciclo celular G1 son separables de su participación en HR. Saintigny et al., (1999) *Oncogene* 18:3553-3563; Boehden et al., (2003) *Oncogene* 22:4111-4117. Por consiguiente, la regulación por disminución de la actividad de

p53 puede servir para aumentar la eficiencia de recombinación homóloga dirigida usando los métodos y composiciones desvelados en el presente documento. Puede usarse cualquier método para la regulación por disminución de la actividad de p53, que incluye, pero no se limita a, cotransfección y expresión en exceso de un mutante negativo dominante de p53 o represión dirigida de la expresión génica de p53 según métodos desvelados, por ejemplo, en la patente de EE.UU. del mismo solicitante nº 6.534.261.

Se consiguen aumentos adicionales en la eficiencia de recombinación dirigida, en células que comprenden una molécula de fusión de dedos de cinc/nucleasa y una molécula de ADN donante, bloqueando las células en la fase G₂ del ciclo celular, cuando los procesos de reparación conducidos por homología son máximamente activos. Tal parada puede lograrse de varias formas. Por ejemplo, pueden tratarse células con, por ejemplo, fármacos, compuestos y/o moléculas pequeñas que influyen en la progresión del ciclo celular de manera que se detengan las células en la fase G₂. Moléculas a modo de ejemplo de este tipo incluyen, pero no se limitan a, compuestos que afectan la polimerización microtúbulos (por ejemplo, vinblastina, nocodazol, Taxol), compuestos que interactúan con ADN (por ejemplo, dicloruro de cis-platino(II)-diamina, cisplatino, doxorubicina) y/o compuestos que afectan la síntesis de ADN (por ejemplo, timidina, hidroximetiluracilo, L-mimosina, etopósido, 5-fluorouracilo). Aumentos adicionales en la eficiencia de recombinación se logran usando inhibidores de la histona desacetilasa (HDAC) (por ejemplo, butirato de sodio, tricostatina A) que alteran la estructura de la cromatina para hacer el ADN genómico más accesible a la maquinaria de recombinación celular.

Métodos adicionales para la parada del ciclo celular incluyen la expresión en exceso de proteínas que inhiben la actividad de las cinasas del ciclo celular CDK, por ejemplo, introduciendo un ADNc que codifica la proteína en la célula o introduciendo en la célula una ZFP manipulada que activa la expresión del gen que codifica la proteína. La parada del ciclo celular también se logra inhibiendo la actividad de ciclinas y CDK, por ejemplo, usando métodos de iARN (por ejemplo, la patente de EE.UU. nº 6.506.559) o introduciendo en la célula una ZFP manipulada que reprime la expresión de uno o más genes que participan en la progresión del ciclo celular tal como, por ejemplo, genes de ciclina y/o CDK. Véase, por ejemplo, la patente de EE.UU. del mismo solicitante nº 6.534.261 para métodos para la síntesis de proteínas de dedos de cinc manipuladas para la regulación de la expresión génica.

Alternativamente, en ciertos casos, la escisión dirigida se realiza en ausencia de un polinucleótido donante (preferentemente en fase S o G₂), y la recombinación se produce entre cromosomas homólogos.

Métodos para cribar factores celulares que facilitan la recombinación homóloga

Como la recombinación homóloga es un proceso de múltiples etapas que requieren la modificación de extremos de ADN y el reclutamiento de varios factores celulares en un complejo de proteína, la adición de uno o más factores exógenos, junto con ADN donante y vectores que codifican fusiones de dedo de cinc-dominio de escisión, puede usarse para facilitar la recombinación homóloga dirigida. Un método a modo de ejemplo para identificar un factor tal o factores emplea análisis de expresión génica usando micromatrices (por ejemplo, matrices Gene Chip[®] de Affymetrix) para comparar los patrones de expresión de ARNm de diferentes células. Por ejemplo, pueden analizarse células que presentan una mayor capacidad para estimular la recombinación homóloga conducida por la rotura bicatenaria en presencia de ADN donante y fusiones de dedo de cinc-dominio de escisión, tanto sin ayuda como en condiciones conocidas por aumentar el nivel de corrección génica, para sus patrones de expresión génica en comparación con células que carecen de tal capacidad. Los genes que están regulados por incremento o regulados por disminución de un modo que se correlaciona directamente con elevados niveles de recombinación homóloga son así identificados y pueden clonarse en uno cualquiera de varios vectores de expresión. Estas construcciones de expresión pueden co-transfectarse junto con fusiones de dedo de cinc-dominio de escisión y construcciones donantes para dar métodos mejorados para lograr la recombinación homóloga de alta eficiencia. Alternativamente, la expresión de tales genes puede regularse apropiadamente usando proteínas de dedo de cinc manipuladas que modulan la expresión (tanto activación como represión) de uno o más de estos genes. Véase, por ejemplo, la patente de EE.UU. del mismo solicitante nº 6.534.261 para métodos para la síntesis de proteínas de dedos de cinc manipuladas para la regulación de expresión génica.

Como un ejemplo, se observó que los diferentes clones obtenidos en los experimentos descritos en el Ejemplo 9 y la Figura 27 presentaron un amplio intervalo de frecuencias de recombinación homóloga cuando se transfectaron con ADN donante y plásmidos que codifican fusiones de dedo de cinc-dominio de escisión. La expresión génica en clones que muestran una alta frecuencia de recombinación dirigida puede así compararse con aquella en clones que presentan una baja frecuencia, y pueden identificarse patrones de expresión únicos para los primeros clones.

Como un ejemplo adicional, estudios usando inhibidores del ciclo celular (por ejemplo, nocodazol o vinblastina, véanse, por ejemplo, los Ejemplos 11, 14 y 15) mostraron que las células paradas en la fase G₂ fase del ciclo celular llevaron a cabo la recombinación homóloga a mayores tasas, que indica que los factores celulares responsables de la recombinación homóloga pueden expresarse preferencialmente o ser activos en G₂. Una forma para identificar estos factores es comparar los patrones de expresión de ARNm entre los clones de células HEK 293 establemente transfectadas que lleva a cabo la corrección génica a altos y bajos niveles (por ejemplo, el clon T18 frente al clon T7). Se hacen comparaciones similares entre estas líneas celulares en respuesta a compuestos que paran las células en la fase G₂. Genes candidatos que se expresan diferencialmente en células que llevan a cabo la

recombinación homóloga a una mayor tasa, tanto sin ayuda como en respuesta a compuestos que paran las células en G2, se identifican, clonan y re-introducen en las células para determinar si su expresión es suficiente para dar lugar a las tasas mejoradas. Alternativamente, la expresión de dichos genes candidatos se activa usando factores de transcripción de dedos de cinc manipulados como se describen, por ejemplo, en la patente de EE.UU. del mismo solicitante n° 6.534.261.

Vectores de expresión

Un ácido nucleico que codifica una o más ZFP o proteínas de fusión de ZFP puede clonarse en un vector para la transformación en células procariotas o eucariotas para la replicación y/o expresión. Los vectores pueden ser vectores procariotas, por ejemplo, plásmidos, o vectores lanzadera, vectores de insecto o vectores eucariotas. Un ácido nucleico que codifica una ZFP también puede clonarse en un vector de expresión, para administración a una célula de planta, célula de animal, preferentemente una célula de mamífero o una célula humana, célula fúngica, célula bacteriana, o célula protozoica.

Para obtener la expresión de un gen o ácido nucleico clonado, las secuencias que codifica una ZFP o proteína de fusión de ZFP se subclonan normalmente en un vector de expresión que contiene un promotor para dirigir la transcripción. Promotores bacterianos y eucariotas adecuados son muy conocidos en la técnica y se describen, por ejemplo, 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., arriba). Sistemas de expresión bacterianos para expresar las ZFP están disponibles en, por ejemplo, *E. coli*, *Bacillus sp.*, y *Salmonella* (Palva et al., *Gene* 22:229-235 (1983)). Están comercialmente disponibles kits para tales sistemas de expresión. Sistemas de expresión eucariotas para células de mamífero, levadura y células de insecto son muy conocidos por aquellos expertos en la materia y también están comercialmente disponibles.

Un promotor usado para dirigir la expresión de un ácido nucleico que codifica ZFP depende de la aplicación particular. Por ejemplo, normalmente se usa un promotor constitutivo fuerte para la expresión y purificación de ZFP. A diferencia, cuando ZFP se administra *in vivo* para la regulación génica, se usa tanto un promotor constitutivo como uno inducible, dependiendo del uso particular de la ZFP. Además, un promotor preferido para la administración de una ZFP puede ser un promotor débil, tal como TK del VHS o un promotor que tiene actividad similar. Un promotor normalmente también puede incluir elementos que son sensibles a la transactivación, por ejemplo, elementos de respuesta a hipoxia, elementos de respuesta a Gal4, elemento de respuesta al represor lac y sistemas de control de molécula pequeña tales como sistemas regulados por tet y el sistema RU-486 (véase, por ejemplo, Gossen & Bujard, *PNAS* 89:5547 (1992); Oligino et al., *Gene Ther.* 5:491-496 (1998); Wang et al., *Gene Ther.* 4:432-441 (1997); Neering et al., *Blood* 88:1147-1155 (1996); y Rendahl et al., *Nat. Biotechnol.* 16:757-761 (1998)). También puede usarse el promotor MNDU3, y es preferencialmente activo en citoblastos hematopoyéticos CD34⁺.

Además de un promotor, el vector de expresión normalmente contiene una unidad de transcripción o casete de expresión que contiene todos los elementos adicionales requeridos para la expresión del ácido nucleico en células huésped, tanto procariotas como eucariotas. Así, un casete de expresión típico contiene un promotor operativamente ligado, por ejemplo, a una secuencia de ácidos nucleicos que codifica la ZFP, y señales requeridas, por ejemplo, para la eficaz poliadenilación del transcrito, terminación transcripcional, sitios de unión al ribosoma, o terminación de la traducción. Elementos adicionales del casete pueden incluir, por ejemplo, potenciadores y señales de corte y empalme heterólogas.

El vector de expresión particular usado para el transporte de la información genética a la célula se selecciona con respecto al uso previsto de la ZFP, por ejemplo, expresión en plantas, animales, bacterias, hongos, protozoos, etc. (véanse los vectores de expresión descritos más adelante). Vectores de expresión bacterianos estándar incluyen plásmidos tales como plásmidos basados en pBR322, pSKF, pET23D y comercialmente sistemas de expresión de fusión disponibles tales como GST y LacZ. Una proteína de fusión a modo de ejemplo es la proteína de unión a maltosa, "MBP". Tales proteínas de fusión se usan para la purificación de la ZFP. También pueden añadirse marcas de epítopes a proteínas recombinantes para proporcionar métodos convenientes de aislamiento, para monitorizar la expresión y para monitorizar la localización celular y subcelular, por ejemplo, c-myc o FLAG.

Vectores de expresión que contienen elementos reguladores de virus eucariotas se usan frecuentemente en vectores de expresión eucariotas, por ejemplo, vectores del SV40, vectores del virus del papiloma y vectores derivados del virus de Epstein-Barr. Otros vectores eucariotas a modo de ejemplo 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 del SV40, promotor tardío del SV40, promotor de metalotioneína, promotor del virus del tumor mamario murino, promotor del virus del sarcoma de Rous, promotor de la poliedrina, u otros promotores que mostró que eran eficaces para la expresión en células eucariotas.

Algunos sistemas de expresión tienen marcadores para la selección de líneas celulares establemente transfectadas tales como timidina cinasa, higromicina B fosfotransferasa y dihidrofolato reductasa. También son adecuados sistemas de expresión de alto rendimiento, tales como usando un vector de baculovirus en células de insecto, con una secuencia codificante de ZFP bajo la dirección del promotor de la poliedrina u otros promotores de baculovirus

fuertes.

5 Los elementos que normalmente están incluidos en vectores de expresión también incluyen un replicón que funciona en *E. coli*, un gen que codifica la resistencia a antibióticos para permitir la selección de bacterias que alojan plásmidos recombinantes, y sitios de restricción únicos en regiones no esenciales del plásmido para permitir la inserción de secuencias recombinantes.

10 Se usan métodos de transfección estándar para producir líneas celulares bacterianas, de mamífero, de levadura o insecto que expresan grandes cantidades de proteína, que luego se purifica usando técnicas convencionales (véase, por ejemplo, Colley et al., J. Biol. Chem. 264:17619-17622 (1989); Guide to Protein Purification, en Methods in Enzymology, vol. 182 (Deutscher, ed., 1990)). La transformación de células eucariotas y procariontes se realiza según técnicas convencionales (véase, por ejemplo, Morrison, J. Bact. 132:349-351 (1977); Clark-Curtiss & Curtiss, Methods in Enzymology 101:347-362 (Wu et al., eds, 1983).

15 Puede usarse cualquiera de los procedimientos muy conocidos para introducir secuencias de nucleótidos extrañas en células huésped. Éstos incluyen el uso de transfección con fosfato de calcio, polibreno, fusión de protoplastos, electroporación, métodos ultrasónicos (por ejemplo, sonoporación), liposomas, microinyección, ADN desnudo, vectores plasmídicos, vectores virales, tanto episómicos como integrativos, y cualquiera de los otros métodos muy conocidos para introducir ADN genómico clonado, ADNc, ADN sintético u otro material genético extraño en una célula huésped (véase, por ejemplo, Sambrook et al., arriba). Solo es necesario que el procedimiento de ingeniería genética particular usado sea capaz de introducir satisfactoriamente al menos un gen en la célula huésped que puede expresar la proteína de elección.

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

Pueden usarse métodos de transferencia génica basados en virus y no virales convencionales para introducir ácidos nucleicos que codifican ZFP manipuladas en células (por ejemplo, células de mamífero) y tejidos diana. Tales métodos también pueden usarse para administrar ácidos nucleicos que codifican ZFP a células *in vitro*. En ciertas realizaciones, los ácidos nucleicos que codifican ZFP se administran para usos de terapia génica *in vivo* o *ex vivo*.
30 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 poloxámero. Sistemas de administración de vectores virales incluyen virus de ADN y de ARN, que tienen tanto genomas episómicos como integrados después de la administración a la célula. Para una revisión de los procedimientos de terapia génica véase Anderson, Science 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 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 Métodos de administración no viral de ácidos nucleicos que codifican ZFP manipuladas incluyen electroporación, lipofección, microinyección, biolística, virosomas, liposomas, inmunoliposomas, conjugados de polimerización o lípido:ácido nucleico, ADN desnudo, viriones artificiales y captación de ADN potenciada por agente. También puede usarse sonoporación usando, por ejemplo, el sistema Sonitron 2000 (Rich-Mar) para la administración de ácidos nucleicos.

45 Sistemas de administración de ácidos nucleicos a modo de ejemplo adicionales incluyen aquellos proporcionados por Amaxa Biosystems (Colonia, Alemania), Maxcyte, Inc. (Rockville, Maryland) y BTX Molecular Delivery Systems (Holliston, MA).

50 La lipofección se describe en, por ejemplo, los documentos US 5.049.386, US 4.946.787; y US 4.897.355) y los reactivos de lipofección se venden comercialmente (por ejemplo, Transfectam™ y Lipofectin™). Lípidos catiónicos y neutros que son adecuados para la eficaz lipofección de reconocimiento de receptores de polinucleótidos incluyen aquellos de Felgner, documentos WO 91/17424, WO 91/16024. La administración puede ser a células (administración *ex vivo*) o tejidos diana (administración *in vivo*).

55 La preparación de complejos de lípido:ácido nucleico, que incluye liposomas dirigidos tales como complejos de inmunolípidos, es muy conocido para un experto en la materia (véase, por ejemplo, Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); las patentes de EE.UU. nº 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).

65 El uso de sistemas basados en virus de ARN o ADN para la administración de ácidos nucleicos que codifican ZFP manipuladas se aprovechan de procesos altamente desarrollados para dirigir un virus a células específicas en el cuerpo y transportar 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*).

Sistemas basados en virus convencionales para la administración de ZFP incluyen, pero no se limitan a, vectores retrovirales, de lentivirus, adenovirales, adeno-asociados, de la variolovacuna y del virus del herpes simple para la transferencia génica. Es posible la integración en el genoma del huésped con los métodos de transferencia de genes de retrovirus, lentivirus y virus adeno-asociados, produciendo frecuentemente expresión a largo plazo del transgén insertado. Adicionalmente, se han observado altas eficiencias de transducción en muchos tipos diferentes de células y tejidos diana.

El tropismo de un retrovirus puede alterarse incorporando proteínas de la envuelta extrañas, que amplían la posible población diana de células diana. Vectores lentivirales son vectores retrovirales que pueden transducir o infectar células no en división y normalmente producen títulos virales altos. La selección de un sistema de transferencia génica retroviral depende del tejido diana. Los vectores retrovirales comprenden repeticiones terminales largas que actúan en *cis* con capacidad de encapsidación para hasta 6-10 kb de secuencia extraña. Las LTR que actúan en *cis* mínimas son suficientes para la replicación y encapsidación de los vectores, que luego se usan para integrar el gen terapéutico en la célula diana para proporcionar expresión transgénica permanente. Vectores retrovirales ampliamente usados incluyen aquellos basados en el virus de la leucemia murina (MuLV), virus de la leucemia de mono gibón (GaLV), virus de la inmunodeficiencia simia (VIS), virus de la inmunodeficiencia humana (VIH) y combinaciones de los mismos (véanse, por ejemplo, Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommerfelt et al., Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); documento PCT/US94/05700).

En aplicaciones en las que se prefiere la expresión transitoria de una proteína de fusión de ZFP, pueden usarse sistemas basados en adenovirus. Los vectores basados en adenovirus son capaces de eficiencia de transducción muy alta en muchos tipos de células y no requieren división celular. Con tales vectores se han obtenido alto título y altos niveles de expresión. Este vector puede producirse en grandes cantidades en un sistema relativamente simple. También se usan vectores de virus adeno-asociados ("AAV") para transducir células con ácidos nucleicos diana, por ejemplo, en la producción *in vitro* de ácidos nucleicos y péptidos, y para los procedimientos de terapia génica *in vivo* y *ex vivo* (véanse, por ejemplo, West et al., Virology 160:38-47 (1987); patente de EE.UU. n° 4.797.368; documento WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). La construcción de vectores de AAV recombinantes se describen en varias publicaciones, que incluyen la patente de EE.UU. n° 5.173.414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); y Samulski et al., J. Virol. 63:03822-3828 (1989).

Al menos seis enfoques de vectores virales están actualmente disponibles para transferencia génica en ensayos clínicos, que utilizan enfoques que implican complementación de vectores defectuosos por genes insertados en líneas celulares auxiliares para generar el agente de transducción.

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

Los vectores de virus adeno-asociados recombinantes (rAAV) son un sistema de administración de genes alternativo prometedor basado en el virus de tipo 2 adeno-asociado de parvovirus defectuoso y no patógeno. Todos los vectores se derivan de un plásmido que retiene solo las repeticiones terminales invertidas de 145 pb de AAV que flanquean el casete de expresión del transgén. La eficaz transferencia de genes y la administración de transgenes estable debido a la integración en los genomas de la célula transducida son características clave para este sistema de vector (Wagner et al., Lancet 351:9117 1702-3 (1998), Kearns et al., Gene Ther. 9:748-55 (1996)).

Pueden producirse vectores adenovirales (Ad) recombinantes deficientes en la replicación a alto título e infectar fácilmente varios tipos diferentes de células. La mayoría de los vectores de adenovirus se manipulan de forma que un transgén sustituya los genes E1a, E1b y/o E3 de Ad; posteriormente, el vector defectuoso en la replicación se propaga en células 293 humanas que suministran la función génica delecionada en *trans*. Los vectores Ad pueden transducir múltiples tipos de tejidos *in vivo*, que incluyen células diferenciadas que no se dividen, tales como aquellas encontradas en hígado, riñón y músculo. Los vectores Ad convencionales Ad tienen una gran capacidad portadora. Un ejemplo del uso de un vector Ad en un ensayo clínico implicó terapia con polinucleótidos para inmunización antitumoral con inyección intramuscular (Serman et al., Hum. Gene Ther. 7:1083-9 (1998)). Ejemplos adicionales del uso de vectores de adenovirus para transferencia génica en ensayos clínicos incluyen Rosenecker et al., Infection 24:15-10 (1996); Serman et al., Hum. Gene Ther. 9:7 1083-1089 (1998); Welsh et al., Hum. Gene Ther. 2:205-18 (1995); Alvarez et al., Hum. Gene Ther. 5:597-613 (1997); Topf et al., Gene Ther. 5:507-513 (1998); Serman et al., Hum. Gene Ther. 7:1083-1089 (1998).

Se usan células encapsidadas para formar partículas de virus que pueden infectar una célula huésped. Tales células incluyen células 293, que encapsidan adenovirus, y células Ψ2 o células PA317, que encapsidan retrovirus. Los vectores virales usados en terapia génica se generan normalmente por una línea celular productora que encapsida

un vector de ácido nucleico en una partícula viral. Los vectores normalmente contienen las secuencias virales mínimas requeridas para la encapsidación y posterior integración en un huésped (si es aplicable), siendo otras secuencias virales sustituidas con un casete de expresión que codifica la proteína que va a expresarse. Las funciones virales ausentes se suministran en *trans* por la línea celular de encapsidación. Por ejemplo, vectores de AAV usados en terapia génica normalmente solo poseen secuencias de repeticiones terminales invertidas (ITR) del genoma de AAV que se requieren para la encapsidación e integración en el genoma del huésped. El ADN viral se encapsida en una línea celular, que contiene un plásmido auxiliar que codifica los otros genes de AAV, concretamente *rep* y *cap*, pero que carece de secuencias de ITR. La línea celular también se infecta con adenovirus como auxiliar. El virus auxiliar promueve la replicación del vector de AAV y la expresión de genes de AAV del plásmido auxiliar. El plásmido auxiliar no está encapsidado en cantidades significativas debido a que carece de secuencias de ITR. La contaminación con adenovirus puede reducirse por, por ejemplo, tratamiento térmico al cual el adenovirus es más sensible que AAV.

En muchas aplicaciones de terapia génica se desea administrar el vector de terapia génica con un alto grado de especificidad a un tipo de tejido particular. Por consiguiente, puede modificarse un vector viral para que tenga especificidad por un tipo de célula dado que expresa un ligando como una proteína de fusión con una proteína de la envuelta viral sobre la superficie externa del virus. Se elige el ligando por tener afinidad por un receptor conocido por estar presente sobre el tipo de célula de interés. Por ejemplo, Han et al., Proc. Natl. Acad. Sci. USA 92:9747-9751 (1995), informaron de que puede modificarse el virus de la leucemia murina de Moloney para expresar heregulina humana fusionada con gp70, y el virus recombinante infecta ciertas células humanas de cáncer de mama que expresan el receptor del factor de crecimiento epidérmico humano. Este principio puede extenderse a otros pares de virus-célula diana, en los 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, puede manipularse fago filamentoso para mostrar fragmentos de anticuerpos (por ejemplo, FAB o Fv) que tienen afinidad de unión específica por prácticamente cualquier receptor celular elegido. Aunque la descripción anterior se aplica principalmente a vectores virales, los mismos principios pueden aplicarse a vectores no virales. Tales vectores pueden manipularse para contener secuencias de captación específicas que favorecen la captación por células diana específicas.

Pueden administrarse vectores de terapia génica *in vivo* por administración a un paciente individual, normalmente por administración sistémica (por ejemplo, infusión intravenosa, intraperitoneal, intramuscular, subdérmica o intracraneal) o administración tópica, como se describe más adelante. Alternativamente, los vectores pueden administrarse a células *ex vivo*, tales como células explantadas de un paciente individual (por ejemplo, linfocitos, aspirados de la médula ósea, biopsia de tejido) o citoblastos hematopoyéticos de donante universal, seguido de reimplantación de las células en un paciente, normalmente después de la selección de células que han incorporado el vector.

La transfección de células *ex vivo* para diagnóstico, investigación o para terapia génica (por ejemplo, mediante re-infusión de las células transfectadas en el organismo huésped) es muy conocida para aquellos expertos en la materia. En una realización preferida, se aíslan células del organismo del sujeto, se transfectan con un ácido nucleico de ZFP (gen o ADNc) y se vuelven a reinfundir en el organismo del sujeto (por ejemplo, paciente). Diversos tipos de células adecuadas para transfección *ex vivo* son muy conocidas para aquellos expertos en la materia (véase, por ejemplo, Freshney et al., Culture of Animal Cells, A Manual of Basic Technique (3ª ed. 1994)) y las referencias citadas en su interior para una discusión de cómo aislar y cultivar células de pacientes).

En una realización, se usan citoblastos en procedimientos *ex vivo* para la transfección de células y terapia génica. La ventaja de usar citoblastos es que pueden diferenciarse en otros tipos de células *in vitro*, o pueden introducirse en un mamífero (tal como el donante de las células) en el que se implantarán en la médula ósea. Se conocen métodos para diferenciar células CD34+ *in vitro* en tipos inmunitarios clínicamente importantes de células usando citocinas tales como GM-CSF, IFN- γ y TNF- α (véase Inaba et al., J. Exp. Med. 176:1693-1702 (1992)).

Se aíslan citoblastos para la transducción y diferenciación usando métodos conocidos. Por ejemplo, los citoblastos se aíslan de células de la médula ósea inmunopurificando las células de la médula ósea con anticuerpos que se unen a células no deseadas, tales como CD4+ y CD8+ (linfocitos T), CD45+ (células panB), GR-1 (granulocitos) e lad (células presentadoras de antígeno diferenciadas) (véase Inaba et al., J. Exp. Med. 176:1693-1702 (1992)).

También pueden administrarse directamente vectores (por ejemplo, retrovirus, adenovirus, liposomas, etc.) que contienen ácidos nucleicos de ZFP terapéuticos a un organismo para la transducción de células *in vivo*. Alternativamente, puede administrarse ADN desnudo. La administración es por cualquiera de las vías usadas normalmente para introducir una molécula en contacto máximo con sangre o células de tejido que incluyen, pero no se limitan a, inyección, infusión, administración tópica y electroporación. Métodos adecuados de administrar tales ácidos nucleicos están disponibles y son muy conocidos para aquellos expertos en la materia, y, aunque puede usarse más de una vía para administrar una composición particular, una vía particular puede proporcionar frecuentemente una reacción más inmediata y más eficaz que otra vía.

Métodos para la introducción de ADN en citoblastos hematopoyéticos se desvelan, por ejemplo, en la patente de EE.UU. n° 5.928.638. Vectores útiles para la introducción de transgenes en citoblastos hematopoyéticos, por

ejemplo, células CD34⁺, incluyen adenovirus tipo 35.

Vectores adecuados para la introducción de transgenes en células inmunitarias (por ejemplo, linfocitos T) incluyen vectores de lentivirus no integrantes. Véanse, por ejemplo, Ory et al., (1996) Proc. Natl. Acad. Sci. USA 93:11382-11388; Dull et al., (1998) J. Virol. 72:8463-8471; Zuffery et al., (1998) J. Virol. 72:9873-9880; Follenzi et al., (2000) Nature Genetics 25:217-222.

Los vehículos farmacéuticamente aceptables se determinan en parte por la composición particular que se administra, además de por el método particular usado para administrar la composición. Por consiguiente, hay una amplia variedad de formulaciones adecuadas de composiciones farmacéuticas disponibles, como se describe más adelante (véase, por ejemplo, Remington's Pharmaceutical Sciences, 17^a ed., 1989).

Pueden introducirse construcciones de ADN en el genoma de un huésped de planta deseado mediante una variedad de técnicas convencionales. Para revisiones de tales técnicas véanse, por ejemplo, Weissbach & Weissbach Methods for Plant Molecular Biology (1988, Academic Press, N.Y.) Sección VIII, pág. 421-463; y Grierson & Corey, Plant Molecular Biology (1988, 2^a ed.), Blackie, London, Cap. 7-9. Por ejemplo, la construcción de ADN puede introducirse directamente en el ADN genómico de la célula de planta usando técnicas tales como electroporación y microinyección de protoplastos de célula de planta, o las construcciones de ADN pueden introducirse directamente en tejido de planta usando métodos biolísticos, tales como bombardeo de partículas de ADN (véase, por ejemplo, Klein et al., (1987) Nature 327:70-73). Alternativamente, las construcciones de ADN pueden combinarse con regiones flanqueantes de T-ADN adecuadas e introducirse en un vector huésped de *Agrobacterium tumefaciens* convencional. Las técnicas de transformación mediadas por *Agrobacterium tumefaciens*, que incluyen el desarme y uso de vectores binarios, se han descrito bien en la bibliografía científica. Véanse, por ejemplo, Horsch et al., (1984) Science 233:496-498, y Fraley et al., (1983) Proc. Nat'l. Acad. Sci. USA 80:4803. Las funciones de virulencia del huésped de *Agrobacterium tumefaciens* dirigirán la inserción de la construcción y marcador adyacente en el ADN de la célula de planta cuando la célula se infecta por las bacterias usando el vector T-ADN binario (Bevan (1984) Nuc. Acid Res. 12:8711-8721) o el procedimiento de co-cultivo (Horsch et al., (1985) Science 227:1229-1231). Generalmente, el sistema de transformación con *Agrobacterium* se usa para manipular 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 con *Agrobacterium* también puede usarse para transformar, además de transferir, ADN a plantas monocotiledóneas y células vegetales. Véase Hernalsteen et al., (1984) EMBO J 3:3039-3041; Hooykass-Van Slogteren et al., (1984) Nature 311:763-764; Grimsley et al., (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.

Métodos de transferencia y de transformación génica alternativos incluyen, pero no se limitan a, transformación de protoplastos mediante captación mediada por calcio, polietilenglicol (PEG) o electroporación de ADN desnudo (véanse 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. USA 82:5824-5828; y Shimamoto (1989) Nature 338:274-276) y electroporación de tejidos de planta (D'Halluin et al., (1992) Plant Cell 4:1495-1505). Métodos adicionales para la transformación de células de planta incluyen microinyección, captación de ADN mediada por carburo de silicio (Kaeppeler et al., (1990) Plant Cell Reporter 9:415-418) y bombardeo con microproyectiles (véanse Klein et al., (1988) Proc. Nat. Acad. Sci. USA 85:4305-4309; y Gordon-Kamm et al., (1990) Plant Cell 2:603-618).

Los métodos y composiciones desvelados pueden usarse para insertar secuencias exógenas en una localización predeterminada en un genoma de célula de planta. Esto es útil en la medida en que la expresión de un transgén introducido en un genoma de planta depende críticamente de su sitio de integración. Por consiguiente, genes que codifican, por ejemplo, nutrientes, antibióticos o moléculas terapéuticas pueden insertarse, por recombinación dirigida, en regiones de un genoma de planta favorable para su expresión.

Células vegetales transformadas que se producen por cualquiera de las técnicas de transformación anteriores pueden cultivarse para regenerar una planta completa que posee el genotipo transformado y así el fenotipo deseado. Tales técnicas de regeneración se basan en la manipulación de ciertas fitohormonas en un medio de crecimiento de cultivo de tejido, que normalmente se basa en un marcador biocida y/o herbicida que se ha introducido junto con las secuencias de nucleótidos deseadas. La regeneración de plantas de protoplastos cultivados se describe en Evans, et al., "Protoplasts Isolation and Culture" en Handbook of Plant Cell Culture, pág. 124-176, Macmillan Publishing Company, New York, 1983; y Binding, Regeneration of Plants, Plant Protoplasts, pág. 21-73, CRC Press, Boca Raton, 1985. La regeneración también puede obtenerse de callos de planta, explantes, órganos, polen, embriones o partes de los mismos. Tales técnicas de regeneración se describen generalmente en Klee et al., (1987) Ann. Rev. of Plant Phys. 38:467-486.

Los ácidos nucleicos introducidos en una célula de planta pueden usarse para conferir rasgos deseados a esencialmente cualquier planta. Puede manipularse una amplia variedad de plantas y sistemas de células de planta para las características fisiológicas y agronómicas deseadas como se ha descrito en el presente documento usando las construcciones de ácidos nucleicos de la presente divulgación y los diversos métodos de transformación mencionados anteriormente. En realizaciones preferidas, plantas y células vegetales diana para manipular incluyen, pero no se limitan a, aquellas plantas monocotiledóneas y dicotiledóneas, tales como cultivos que incluyen cultivos

de grano (por ejemplo, trigo, maíz, arroz, mijo, cebada), cultivos de frutos (por ejemplo, tomate, manzana, pera, fresa, naranja), cultivos de forraje (por ejemplo, alfalfa), cultivos de verduras de raíz (por ejemplo, zanahoria, patata, remolachas azucareras, boniato), cultivos de verduras de hoja (por ejemplo, lechuga, espinaca); plantas de flor (por ejemplo, petunia, rosa, crisantemo), coníferas y pinos (por ejemplo, pino abeto, píceas); plantas usadas en fitorremediación (por ejemplo, plantas que acumulan metales pesados); cultivos de aceite (por ejemplo, girasol, semilla de colza) y plantas usadas para fines experimentales (por ejemplo, *Arabidopsis*). Así, los métodos y composiciones desvelados tienen uso en una amplia variedad de plantas, que incluyen, pero no se limitan a, especies de los géneros *Asparagus*, *Avena*, *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*.

Un experto en la materia reconocerá que después de incorporar establemente el casete de expresión en plantas transgénicas y de confirmarse que es operativo, puede introducirse en otras plantas por cruce sexual. Puede usarse cualquiera de varias técnicas de cultivo estándar, dependiendo de las especies que vayan a cruzarse.

Una célula de planta transformada, callo, tejido o planta puede identificarse y aislarse seleccionando o cribando el material de planta manipulado para rasgos codificados por los genes marcadores presentes en el ADN transformante. Por ejemplo, puede realizarse selección cultivando el material de planta manipulado en medios que contienen una cantidad inhibitoria del antibiótico o herbicida al que la construcción de gen transformante confiere resistencia. Además, también pueden identificarse plantas y células vegetales transformadas cribando las actividades de cualquier marcador de gen visible (por ejemplo, los genes de la β -glucuronidasa, luciferasa, B o C1) que pueden estar presentes en las construcciones de ácidos nucleicos recombinantes. Tales metodologías de selección y cribado son muy conocidas para aquellos expertos en la materia.

También pueden usarse métodos físicos y bioquímicos para identificar transformantes de plantas o células de planta que contienen construcciones de genes insertadas. Estos métodos incluyen, pero no se limitan a: 1) análisis de Southern o amplificación por PCR para detectar y determinar la estructura del inserto de ADN recombinante; 2) transferencia Northern, protección de RNasa S1, amplificación por PCR por extensión de cebadores o con transcriptasa inversa para detectar y examinar transcritos de ARN de las construcciones de gen; 3) ensayos enzimáticos para detectar actividad de enzima o ribozima, en la que tales productos génicos están codificados por la construcción de gen; 4) electroforesis en gel de proteína, técnicas de transferencia Western, inmunoprecipitación o enzimoanálisis, en los que los productos de la construcción de gen son proteínas. También pueden usarse técnicas adicionales, tales como hibridación *in situ*, tinción de enzimas e inmunotinción, para detectar la presencia o expresión de la construcción recombinante en órganos y tejidos de planta específicos. Los métodos para hacer todos estos ensayos son muy conocidos para aquellos expertos en la materia.

Los efectos de la manipulación génica usando los métodos desvelados en el presente documento pueden observarse, por ejemplo, por transferencias Northern del ARN (por ejemplo, ARNm) aislado de los tejidos de interés. Normalmente, si la cantidad de ARNm ha aumentado, puede asumirse que el gen endógeno correspondiente está siendo expresado a una mayor tasa que antes. Pueden usarse otros métodos de medición de la actividad de genes y/o de CYP74B. Pueden usarse diferentes tipos de ensayos enzimáticos, dependiendo del sustrato usado y el método de detección del aumento o disminución de un producto de reacción o subproducto. Además, los niveles de y/o proteína CYP74B expresada pueden medirse inmunológicamente, es decir, ELISA, RIA, EIA y otros ensayos basados en anticuerpos muy conocidos para aquellos expertos en la materia, tales como por ensayos de detección electroforética (tanto con tinción como transferencia Western). El transgén puede expresarse selectivamente en algunos tejidos de la planta o en algunos estadios de desarrollo, o el transgén puede expresarse en sustancialmente todos los tejidos de la planta, sustancialmente a lo largo de su ciclo de vida entero. Sin embargo, también es aplicable cualquier modo de expresión combinatoria.

La presente divulgación también engloba semillas de las plantas transgénicas descritas anteriormente en las que la semilla tiene el transgén o construcción de gen. La presente divulgación engloba adicionalmente la progenie, clones, líneas celulares o células de las plantas transgénicas descritas anteriormente en las que dicha progenie, clon, línea celular o célula tiene el transgén o construcción de gen.

55 Vehículos de administración

Un factor importante en la administración de compuestos de polipéptido, tales como proteínas de fusión de ZFP, es asegurar que el polipéptido tiene la capacidad de atravesar la membrana plasmática de una célula, o la membrana de un compartimento intracelular tal como el núcleo. Las membranas celulares están compuestas de bicapas de lípido-proteína que son libremente permeables a compuestos lipófilos 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 tales como liposomas, que tienen la capacidad de translocar polipéptidos tales como ZFP a través de una membrana celular.

Por ejemplo, los "polipéptidos de translocalización hacia la membrana" tienen subsecuencias de aminoácidos anfífilos o hidrófobos que tienen la capacidad de actuar de vehículos de translocación hacia la membrana. En una

realización, las proteínas de homeodominio tienen la capacidad de translocar a través de membranas celulares. Se encontró que el péptido internalizable más corto de una proteína de homeodominio, Antennapedia, era la tercera hélice de la proteína, desde la posición de aminoácido 43 a 58 (véase, por ejemplo, Prochiantz, Current Opinion in Neurobiology 6:629-634 (1996)). Se encontró que otra subsecuencia, el dominio h (hidrófobo) de péptidos señal, tenía características de translocalización hacia la membrana celular similares (véase, por ejemplo, Lin et al., J. Biol. Chem. 270:1 4255-14258 (1995)).

Ejemplos de secuencias de péptidos que pueden ligarse a una proteína, para facilitar la captación de la proteína en células, incluyen, pero no se limitan a: un péptido de 11 aminoácidos de la proteína tat del VIH; una secuencia de péptidos de 20 residuos que se corresponde con los aminoácidos 84-103 de la proteína p16 (véase Fahraeus et al., Current Biology 6:84 (1996)); la tercera hélice del homeodominio de 60 aminoácidos de longitud 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., arriba); o el dominio de translocalización de VP22 del VHS (Elliot & O'Hare, Cell 88:223-233 (1997)). Otros restos químicos adecuados que proporcionan captación celular potenciada también pueden estar químicamente ligados a ZFP. También pueden seleccionarse dominios de translocalización de membrana (es decir, dominios de internalización) de bibliotecas de secuencias de péptidos aleatorizadas. Véase, por ejemplo, Yeh et al., (2003) Molecular Therapy 7(5):S461, Resumen n° 1191.

Las moléculas de toxina también tienen la capacidad de transportar polipéptidos a través de membranas celulares. Frecuentemente, tales moléculas (llamadas "toxinas binarias") están compuestas de al menos dos partes: un dominio de translocalización/unión o polipéptido y un dominio de toxina o polipéptido separado. Normalmente, el dominio de translocalización o polipéptido se une a un receptor celular, y a continuación la toxina se transporta en la célula. Se han usado varias toxinas bacterianas, que incluyen toxina iota de *Clostridium perfringens*, toxina diftérica (DT), exotoxina A de *Pseudomonas* (PE), toxina pertussis (PT), toxina de *Bacillus anthracis* y adenilato ciclasa de pertussis (CYA) para administrar péptidos al citosol de la célula como fusiones internas o del extremo amino (Arora et al., J. Biol. Chem., 268:3334-3341 (1993); Perelle et al., Infect. Immun., 61:5147-5156 (1993); Stenmark et al., J. Cell Biol. 113:1025-1032 (1991); Donnelly et al., PNAS 90:3530-3534 (1993); Carbonetti et al., Abstr. Annu. Meet. Am. Soc. Microbiol. 95:295 (1995); Sebo et al., Infect. Immun. 63:3851-3857 (1995); Klimpel et al., PNAS U.S.A. 89:10277-10281 (1992); y Novak et al., J. Biol. Chem. 267:17186-17193 (1992)).

Tales secuencias de péptidos pueden usarse para translocar ZFP a través de una membrana celular. Las ZFP pueden fusionarse convenientemente con o derivatizarse con tales secuencias. Normalmente, la secuencia de translocalización se proporciona como parte de una proteína de fusión. Opcionalmente, puede usarse un conector para enlazar la ZFP y la secuencia de translocalización. Puede usarse cualquier conector adecuado, por ejemplo, un conector peptídico.

La ZFP también puede introducirse en una célula de animal, preferentemente una célula de mamífero, mediante liposomas y derivados de liposomas tales como inmunoliposomas. El término "liposoma" se refiere a vesículas que comprenden una o más bicapas de lípidos concéntricamente ordenadas, que encapsulan una fase acuosa. La fase acuosa normalmente contiene el compuesto que va a administrarse a la célula, es decir, un ZFP.

El liposoma se fusiona con la membrana plasmática, liberando así el fármaco en el citosol. Alternativamente, el liposoma es fagocitado o captado por la célula en una vesícula de transporte. Una vez en el endosoma o fagosoma, el liposoma tanto es degradado como se fusiona con la membrana de la vesícula de transporte y libera su contenido.

En los actuales métodos de administración de fármacos mediante liposomas, el liposoma se vuelve por último lugar permeable y libera el compuesto encapsulado (en este caso, una ZFP) en el tejido diana o célula. Para administración sistémica o específica de tejido, esto puede llevarse a cabo, por ejemplo, en un modo pasivo en el que la bicapa de liposoma se degrada con el tiempo mediante la acción de diversos agentes en el cuerpo. Alternativamente, la liberación de fármaco activo implica usar un agente para inducir un cambio de permeabilidad en la vesícula del liposoma. Las membranas de liposomas pueden construirse de manera que se desestabilicen cuando el entorno se vuelve ácido cerca de la membrana del liposoma (véase, por ejemplo, PNAS 84:7851 (1987); Biochemistry 28:908 (1989)). Cuando los liposomas son endocitados por una célula diana, por ejemplo, se desestabilizan y liberan su contenido. Esta desestabilización se llama fusogénesis.

La dioleoilfosfatidiletanolamina (DOPE) es la base de muchos sistemas "fusogénicos".

Tales liposomas normalmente comprenden una ZFP y un componente de lípido, por ejemplo, un lípido neutro y/o catiónico, que opcionalmente incluye una molécula de reconocimiento de receptor tal como un anticuerpo que se une a un receptor de la superficie celular o ligando predeterminado (por ejemplo, un antígeno). Están disponibles una variedad de métodos para preparar liposomas como se describe en, por ejemplo, Szoka et al., Ann. Rev. Biophys. Bioeng. 9:467 (1980), las patentes de EE.UU. n° 4.186.183, 4.217.344, 4.235.871, 4.261.975, 4.485.054, 4.501.728, 4.774.085, 4.837.028, 4.235.871, 4.261.975, 4.485.054, 4.501.728, 4.774.085, 4.837.028, 4.946.787, publicación PCT n° WO 91/17424, 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)). Métodos adecuados incluyen, por ejemplo, sonicación, extrusión, alta presión/homogeneización, microfluidización, diálisis con detergentes, fusión inducida por calcio de vesículas de liposomas pequeñas y métodos de fusión en éter, todos los cuales son conocidos para aquellos expertos en la materia.

En ciertas realizaciones, se desea dirigir liposomas usando restos de direccionamiento 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 (por ejemplo, ligandos, receptores y anticuerpos monoclonales). Véanse, por ejemplo, las patentes de EE.UU. nº 4.957.773 y 4.603.044.

Ejemplos de restos de direccionamiento incluyen anticuerpos monoclonales específicos para antígenos asociados a neoplasias, tales como antígeno específico del cáncer de próstata y MAGE. También pueden diagnosticarse tumores detectando productos génicos resultantes de la activación o expresión en exceso de oncogenes, tales como ras o c-erbB2. Además, muchos tumores expresan antígenos normalmente expresados por tejido fetal, tales como la alfa-fetoproteína (AFP) y antígeno carcinoembrionario (CEA). Pueden diagnosticarse sitios de infección viral usando diversos antígenos virales tales como antígenos del núcleo y de la superficie de la hepatitis B (VHBc, VHBs) antígenos de la hepatitis C, antígenos del virus de Epstein-Barr, antígenos del virus de la inmunodeficiencia humana tipo-1 (VHI1) y 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 (por ejemplo, VCAM-1), receptores de selectina (por ejemplo, ELAM-1) y similares.

Pueden usarse métodos convencionales para acoplar agentes de direccionamiento a liposomas. Estos métodos generalmente implican la incorporación en liposomas de componentes de lípido, por ejemplo, fosfatidiletanolamina, que pueden activarse para la unión de agentes de direccionamiento, o compuestos lipófilos derivatizados tales como bleomicina derivatizada de lípidos. Pueden construirse liposomas dirigidos a anticuerpo usando, por ejemplo, liposomas que incorporan proteína A (véase Renneisen et al., J. Biol. Chem., 265:16337-16342 (1990) y Leonetti et al., PNAS 87:2448-2451 (1990).

Dosificaciones

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, debe ser suficiente para efectuar una respuesta terapéutica beneficiosa en el paciente con el tiempo. Además, pautas de dosificación particulares pueden ser útiles para determinar cambios fenotípicos en un entorno experimental, por ejemplo, en estudios de genómica funcionales, y en modelos de célula o animales. La dosis se determinará por la eficacia y K_d de la ZFP particular empleada, el volumen nuclear de la célula diana y la afección del paciente, además del peso corporal o área superficial del paciente que va a tratarse. El tamaño de la dosis también se determinará por la existencia, naturaleza y grado de cualquier efecto secundario adverso que acompaña a la administración de un compuesto particular o vector en un paciente particular.

La máxima dosificación terapéuticamente eficaz de ZFP para aproximadamente el 99 % de la unión a sitios diana se calcula que está en el intervalo de menos de aproximadamente $1,5 \times 10^5$ a $1,5 \times 10^6$ copias de la molécula de ZFP específica por célula. El número de ZFP por célula para este nivel de unión se calcula del siguiente modo, usando el volumen de un núcleo de células HeLa (aproximadamente $1000 \mu\text{m}^3$ o 10^{-12} l; Cell Biology (Altman & Katz, eds. (1976)). Como el núcleo de HeLa es relativamente grande, este número de dosificación se recalcula según se necesite usando el volumen del núcleo de la célula diana. Este cálculo no tiene en cuenta la competición de la unión de ZFP por otros sitios. Este cálculo también supone que esencialmente toda la ZFP está localizada en el núcleo. Se usa un valor de $100 \times K_d$ para calcular aproximadamente el 99 % de unión al sitio diana, y se usa un valor de $10 \times K_d$ para calcular aproximadamente el 90 % de unión al sitio diana. Para este ejemplo, $K_d = 25$ nM

ZFP + sitio diana \leftrightarrow complejo $K_d = \frac{[\text{ADN}] [\text{proteína}]}{[\text{ADN:complejo proteína}]}$ Cuando está
 es decir, ADN + proteína \leftrightarrow complejo de ADN:proteína
 unido el 50 % de ZFP, $K_d = [\text{proteína}]$ Así, cuando $[\text{proteína}] = 25$ nM y el volumen de núcleos 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 el 50 % de la unión Cuando está unido el 99 % de diana; $100 \times K_d =$
 $[\text{proteína}] 100 \times K_d = [\text{proteína}] = 2,5 \mu\text{M} (2,5 \times 10^{-6} \text{ moles/l}) (10^{-12} \text{ l/núcleo}) (6 \times 10^{23} \text{ moléculas/mol}) =$
 aproximadamente 1.500.000 moléculas por núcleo para el 99 % de unión del sitio diana.

La dosis apropiada de un vector de expresión que codifica una ZFP también puede calcularse teniendo en cuenta la tasa promedio de expresión de ZFP de un promotor y la tasa promedio de degradación de ZFP en la célula. En ciertas realizaciones se usa un promotor débil tal como un promotor TK del VHS no mutante o mutante, como se ha descrito anteriormente. La dosis de ZFP en microgramos se calcula teniendo en cuenta el peso molecular de la ZFP particular que se emplea.

En la determinación de la cantidad eficaz de la ZFP que va a administrarse en el tratamiento o profilaxis de la enfermedad, el médico evalúa los niveles en plasma circulantes de la ZFP o ácido nucleico que codifica la ZFP,

posibles toxicidades de ZFP, progresión de la enfermedad y la producción de anticuerpos anti-ZFP. La administración puede llevarse a cabo mediante dosis únicas o divididas.

Composiciones farmacéuticas y administración

5 Pueden administrarse ZFP y vectores de expresión que codifican ZFP directamente al paciente para la escisión y/o recombinación dirigidas, y para aplicaciones terapéuticas o profilácticas, por ejemplo, cáncer, isquemia, retinopatía diabética, degeneración macular, artritis reumatoide, psoriasis, infección por el VIH, anemia drepanocítica, enfermedad de Alzheimer, distrofia muscular, enfermedades neurodegenerativas, enfermedad vascular, fibrosis quística, accidente cerebrovascular y similares. Ejemplos de microorganismos que pueden inhibirse por terapia génica con ZFP incluyen bacterias patógenas, por ejemplo, clamidias, bacterias rickettsias, micobacterias, estafilococos, estreptococos, neumococos, meningococos y conococos, klebsiella, proteus, serratia, pseudomonas, legionella, difteria, salmonella, bacilos, cólera, tétanos, botulismo, ántrax, peste, leptospirosis y bacterias de la enfermedad de Lyme; hongo infeccioso, por ejemplo, *Aspergillus*, especies de *Candida*; protozoos tales como esporozoos (por ejemplo, *Plasmodia*), rizópodos (por ejemplo, *Entamoeba*) y flagelados (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); enfermedades virales, por ejemplo, hepatitis (A, B o C), virus del herpes (por ejemplo, VVZ, VHS-1, VHS-6, VHS-II, CMV y VEB), VIH, Ébola, adenovirus, virus de la gripe, flavivirus, ecovirus, rinovirus, virus de Coxsackie, coronavirus, virus respiratorio sincitial, virus de las paperas, rotavirus, virus del sarampión, virus de la rubeola, parvovirus, virus de la variolovacuna, virus HTLV, virus del dengue, virus del papiloma, virus de la poliomielitis, virus de la rabia y virus de la encefalitis arboviral, etc.

La administración de cantidades terapéuticamente eficaces es por cualquiera de las vías normalmente usadas para introducir ZFP en contacto máximo con el tejido que va a tratarse. Las ZFP se administran de cualquier manera adecuada, preferentemente con vehículos farmacéuticamente aceptables. Están disponibles métodos adecuados de administración de tales moduladores y son muy conocidos para aquellos expertos en la materia, y, aunque puede usarse más de una vía para administrar una composición particular, una vía particular puede frecuentemente proporcionar una reacción más inmediata y más eficaz que otra vía.

Vehículos farmacéuticamente aceptables se determinan en parte por la composición particular que se administra, además de por el método particular usado para administrar la composición. Por consiguiente, hay una amplia variedad de formulaciones adecuadas de composiciones farmacéuticas que están disponibles (véase, por ejemplo, Remington's Pharmaceutical Sciences, 17^a ed. 1985)).

Las ZFP, solas o en combinación con otros componentes adecuados, pueden prepararse en formulaciones en aerosol (es decir, pueden "nebulizarse") para administrarse mediante inhalación. Las formulaciones en aerosol pueden colocarse en propulsores aceptables presurizados, tales como diclorodifluorometano, propano, nitrógeno y similares.

Formulaciones adecuadas para administración parenteral, tales como, por ejemplo, por vías intravenosa, intramuscular, intradérmica y subcutánea, incluyen disoluciones acuosas y no acuosas para inyección estéril isotónica, que pueden contener antioxidantes, tampones, bacteriostáticos y solutos que convierten la formulación en isotónica con la sangre del receptor previsto, y suspensiones estériles acuosas y no acuosas que pueden incluir agentes de suspensión, solubilizantes, espesantes, estabilizadores y conservantes. Las composiciones desveladas pueden administrarse, por ejemplo, por infusión intravenosa, por vía oral, tópicamente, intraperitonealmente, intravesicalmente o intratecalmente. Las formulaciones de compuestos pueden presentarse en envases sellados de dosis unitaria o multi-dosis, tales como ampollas y viales. Pueden prepararse disoluciones y suspensiones para inyección a partir de polvos estériles, gránulos y comprimidos del tipo previamente descrito.

Aplicaciones

50 Los métodos y composiciones desvelados para la escisión dirigida pueden usarse para inducir mutaciones en una secuencia genómica, por ejemplo, escindiendo en dos sitios y delecionando secuencias entremedias, escindiendo en un único sitio seguido de unión de extremos no homólogos y/o escindiendo en un sitio de manera que se eliminen uno o dos o algunos nucleótidos. La escisión dirigida también puede usarse para crear inactivaciones de genes (por ejemplo, para genómica funcional o la validación de dianas) y para facilitar la inserción dirigida de una secuencia en un genoma (es decir, activaciones de genes); por ejemplo, para fines de ingeniería de células o expresión en exceso de proteínas. La inserción puede ser por medio de sustituciones de secuencias cromosómicas mediante recombinación homóloga o por integración dirigida, en la que 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.

También pueden usarse los mismos métodos para sustituir una secuencia no mutante con una secuencia mutante, o para convertir un alelo en un alelo diferente.

65 La escisión dirigida de genomas virales infectantes o integrados puede usarse para tratar infecciones virales en un huésped. Adicionalmente, la escisión dirigida de genes que codifican receptores para virus puede usarse para

bloquear la expresión de tales receptores, previniendo así la infección viral y/o proliferación viral en un organismo huésped. Puede usarse mutagénesis dirigida de genes que codifican receptores virales (por ejemplo, los receptores CCR5 y CXCR4 para el VIH) para hacer que los receptores sean incapaces de unirse a virus, previniendo así la nueva infección y bloqueando la propagación de infecciones existentes. Ejemplos no limitantes de virus o receptores virales que pueden ser elegidos como diana incluyen virus del herpes simple (VHS), tales como VHS-1 y VHS-2, virus de la varicela zóster (VVZ), virus de Epstein-Barr (VEB) y citomegalovirus (CMV), VHH6 y VHH7. Los familia de virus de la hepatitis incluye virus de la hepatitis A (VHA), virus de la hepatitis B (VHB), virus de la hepatitis C (VHC), el virus de la hepatitis delta (VHD), virus de la hepatitis E (VHE) y virus de la hepatitis G (VHG). Otros virus o sus receptores pueden elegidos como diana, que incluyen, pero no se limitan a, Picornaviridae (por ejemplo, virus de la poliomielitis, etc.); Caliciviridae; Togaviridae (por ejemplo, virus de la rubeola, virus del dengue, etc.); Flaviviridae; Coronaviridae; Reoviridae; Birnaviridae; Rhabdoviridae (por ejemplo, virus de la rabia, etc.); Filoviridae; Paramyxoviridae (por ejemplo, virus de las paperas, virus del sarampión, virus respiratorio sincitial, etc.); Orthomyxoviridae (por ejemplo, virus de la gripe tipos A, B y C, etc.); Bunyaviridae; Arenaviridae; Retroviridae; lentivirus (por ejemplo, HTLV-I; HTLV-II; VIH-1 (también conocidos como HTLV-III, LAV, ARV, hTLR, etc.) VIH-II); virus de la inmunodeficiencia simia (VIS), virus del papiloma humano (VPH), virus de la gripe y los virus de la encefalitis transmitida por garrapatas. Véase, por ejemplo, *Virology*, 3ª edición (W. K. Joklik ed. 1988); *Fundamental Virology*, 2ª edición (B. N. Fields y D. M. Knipe, eds. 1991), para una descripción de estos y otros virus. Los receptores para el VIH, por ejemplo, incluyen CCR-5 y CXCR-4.

De manera similar, el genoma de una bacteria infectante puede mutagenizarse por escisión de ADN dirigida seguido de unión de extremos no homólogos, para bloquear o mejorar las infecciones bacterianas.

Los métodos desvelados para la recombinación dirigida pueden usarse para sustituir cualquier secuencia genómica con una secuencia no idéntica homóloga. Por ejemplo, puede sustituirse una secuencia genómica mutante con su homólogo no mutante, proporcionando así métodos para el tratamiento de, por ejemplo, enfermedad genética, trastornos heredados, cáncer y enfermedad autoinmunitaria. De manera similar, un alelo de un gen puede sustituirse con un alelo diferente usando los métodos de recombinación dirigida desvelados en el presente documento.

Enfermedades genéticas a modo de ejemplo 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 alfa-1 antitripsina, alfa-talasemia, síndrome de insensibilidad a andrógenos, síndrome de Apert, displasia ventricular derecha arritmogénica, ataxia telangiectasia, síndrome de Bart, beta-talasemia, síndrome de Bean, enfermedad de Canavan, enfermedades granulomatosas crónicas (EGC), síndrome del maullido, fibrosis quística, enfermedad de Dercum, displasia ectodérmica, anemia de Fanconi, fibrodisplasia osificante progresiva, síndrome del X frágil, galactosemia, enfermedad de Gaucher, gangliosidosis generalizada (por ejemplo, GM1), hemocromatosis, la mutación de hemoglobina C en el 6º codón de beta-globina (HbC), hemofilia, enfermedad de Huntington, síndrome de Hurler, hipofosfatasa, síndrome de Klinefelter, enfermedad de Krabbes, síndrome de Langer-Giedion, deficiencia de adhesión de leucocitos (LAD, OMIM No. 116920), leucodistrofia, síndrome de QT largo, síndrome de Marfan, síndrome de Moebius, mucopolisacaridosis (MPS), síndrome oncorrotuliano, diabetes insípida nefrogénica, neurofibromatosis, enfermedad de Neimann-Pick, osteogénesis imperfecta, porfiria, síndrome de Prader-Willi, progeria, síndrome de Proteus, retinoblastoma, síndrome de Rett, síndrome de Rubinstein-Taybi, síndrome de Sanfilippo, inmunodeficiencia combinada grave (IDCG), síndrome de Shwachman, drepanocitosis (anemia drepanocítica), síndrome de Smith-Magenis, síndrome de Stickler, enfermedad de Tay-Sachs, síndrome de trombocitopenia con radios ausente (TAR), síndrome de Treacher-Collins, trisomía, esclerosis tuberculosa, 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 asociado a X (XLP, OMIM No. 308240).

Enfermedades a modo de ejemplo adicionales que pueden tratarse por escisión de ADN y/o recombinación homóloga dirigidas incluyen inmunodeficiencias adquiridas, enfermedades de almacenamiento lisosómico (por ejemplo, enfermedad de Gaucher, GM1, enfermedad de Fabry y enfermedad de Tay-Sachs), mucopolisacaridosis (por ejemplo, enfermedad de Hunter, enfermedad de Hurler), hemoglobinopatías (por ejemplo, enfermedades drepanocíticas, HbC, α -talasemia, β -talasemia) y hemofilias.

En ciertos casos, se desea la alteración de una secuencia genómica en una célula pluripotente (por ejemplo, un citoblasto hematopoyético). Se conocen en la técnica métodos para la movilización, enriquecimiento y cultivo de citoblastos hematopoyéticos. Véanse, por ejemplo, las patentes de EE.UU. 5.061.620; 5.681.559; 6.335.195; 6.645.489 y 6.667.064. Los citoblastos tratados pueden devolverse a un paciente para el tratamiento de diversas enfermedades que incluyen, pero no se limitan a, IDCG y anemia drepanocítica.

En muchos de estos casos, una región de interés comprende una mutación, y el polinucleótido donante comprende la secuencia no mutante correspondiente. Similarmente, una secuencia no mutante genómica puede sustituirse por una secuencia mutante, si se desea. Por ejemplo, la expresión en exceso de un oncogén puede invertirse tanto mutando el gen como reemplazando sus secuencias de control con secuencias que soporten un menor nivel de expresión no patológico. Como otro ejemplo, el alelo no mutante del gen ApoA1 puede sustituirse con el alelo Milano de ApoA1, para tratar aterosclerosis. De hecho, cualquier patología dependiente de una secuencia genómica

particular, en cualquier modo, puede corregirse o aliviarse usando los métodos y composiciones desvelados en el presente documento.

5 También pueden usarse escisión dirigida y recombinación dirigida para alterar secuencias no codificantes (por ejemplo, secuencias reguladoras tales como promotores, potenciadores, iniciadores, terminadores, sitios de corte y empalme) para alterar los niveles de expresión de un producto génico. Tales métodos pueden usarse, por ejemplo, para fines terapéuticos, genómica funcional y/o estudios de validación de dianas.

10 Las composiciones y métodos descritos en el presente documento también permiten novedosos enfoques y sistemas para dirigir reacciones inmunitarias de un huésped a injertos alógenos. En particular, un problema importante que se afronta cuando los citoblastos alógenos (o cualquier tipo de célula alógena) se injertan en un receptor huésped es el alto riesgo de rechazo por el sistema inmunitario del huésped, principalmente mediado por el reconocimiento del complejo mayor de histocompatibilidad (MHC) sobre la superficie de las células implantadas. El MHC comprende la(s) proteína(s) de clase I de HLA que sirven de heterodímeros que comprenden una subunidad β común y subunidades α variables. Se ha demostrado que los injertos de tejido derivados de citoblastos que carecen de HLA escapan de la respuesta inmunitaria del huésped. Véanse, por ejemplo, Coffman et al., *J Immunol* 151, 425-35. (1993); Markmann et al., *Transplantation* 54, 1085-9. (1992); Koller et al., *Science* 248, 1227-30. (1990). Usando las composiciones y métodos descritos en el presente documento, los genes que codifican proteínas HLA que participaron en el rechazo del injerto pueden escindirse, mutagenizarse o alterarse por recombinación, en tanto sus secuencias codificantes como reguladoras, de manera que su expresión se bloquee o expresen un producto no funcional. Por ejemplo, inactivando el gen que codifica el gen de subunidad β común (β 2-microglobulina) usando proteínas de fusión de ZFP como se describen en el presente documento, la clase I de HLA puede eliminarse de las células para generar rápida y de forma fiable citoblastos *nulos* de la clase I de HLA de cualquier donante, reduciendo así la necesidad de hacer coincidir estrechamente los haplotipos del MHC del donante/receptor durante el injerto de citoblastos.

25 La inactivación de cualquier gen (por ejemplo, el gen de β 2-microglobulina) puede lograrse, por ejemplo, por un evento de escisión único, por escisión seguida de unión de extremos no homólogos, por escisión en dos sitios seguido de unión de manera que se deleccione la secuencia entre los dos sitios de escisión, por recombinación dirigida de un codón de sentido alterado o finalizador en la región codificante, o por recombinación dirigida de una secuencia irrelevante (es decir, una secuencia de "relleno") en el gen o su región reguladora, de manera que rompan el gen o región reguladora.

30 La modificación dirigida de la estructura de cromatina, como se ha desvelado en el documento del mismo solicitante WO 01/83793, puede usarse para facilitar la unión de proteínas de fusión a cromatina celular.

35 En realizaciones adicionales pueden usarse una o más fusiones entre un dominio de unión de dedos de cinc y una recombinasa (o fragmento funcional de la misma), además de o en lugar de las fusiones de dedos de cinc-dominio de escisión desveladas en el presente documento, para facilitar la recombinación dirigida. Véanse, por ejemplo, la patente de EE.UU. del mismo solicitante nº 6.534.261 y Akopian et al., (2003) *Proc. Natl. Acad. Sci. USA* 100:8688-8691.

40 En realizaciones adicionales, los métodos y composiciones desvelados se usan para proporcionar fusiones de dominios de unión de ZFP con activación transcripcional o dominios de represión que requieren dimerización (tanto homodimerización como heterodimerización) para su actividad. En estos casos, un polipéptido de fusión comprende un dominio de unión de dedos de cinc y un monómero de dominio funcional (por ejemplo, un monómero de una activación transcripcional dímera o dominio de represión). La unión de dos de tales polipéptidos de fusión a sitios diana apropiadamente situados permite la dimerización de manera que se reconstituya una activación de la transcripción funcional o dominio de represión.

50 EJEMPLOS

Ejemplo 1: Edición de un gen hSMC1L1 cromosómico por recombinación dirigida

55 El gen hSMC1L1 es el ortólogo humano del gen de la levadura en gemación *mantenimiento estructural de cromosomas 1*. Una región de este gen que codifica una porción del extremo amino de la proteína que incluye el dominio de ATPasa Walker se mutagenizó por escisión y recombinación dirigidas. La escisión se dirigió a la región del codón de iniciación de metionina (nucleótidos 24-26, Figura 1), diseñando nucleasas químicas, que comprenden un dominio de unión de ADN de dedos de cinc y un medio dominio de escisión de *FokI*, que se unen en la proximidad del codón. Así, se diseñaron dos dominios de unión de dedos de cinc, uno de los cuales reconoce los nucleótidos 23-34 (contactos primarios a lo largo de la hebra superior como se muestra en la Figura 1) y el otro de los cuales reconoce los nucleótidos 5-16 (contactos primarios a lo largo de la hebra inferior). Se diseñaron proteínas de dedos de cinc como se ha descrito en las patentes de EE.UU. del mismo solicitante 6.453.242 y 6.534.261. Véase la Tabla 2 para las secuencias de aminoácidos de las regiones de reconocimiento de las proteínas de dedos de cinc.

Las secuencias que codifican cada uno de estos dos dominios de unión de ZFP se fusionaron con secuencias que codifican un medio dominio de escisión de *FokI* (aminoácidos 384-579 de la secuencia nativa de *FokI*; Kita et al., (1989) J. Biol. Chem. 264:5751-5756), de forma que la proteína codificada contuviera secuencias de *FokI* en el extremo carboxi y secuencias de ZFP en el extremo amino. Cada una de estas secuencias de fusión se clonó entonces en un vector de expresión de mamífero modificado pcDNA3 (Figura 2).

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

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

Nota: Las secuencias de aminoácidos de dedos de cinc mostradas anteriormente (en el código de una letra) representan los residuos -1 a +6, con respecto al inicio de la porción de hélice alfa de cada dedo de cinc. El dedo F1 es el más próximo al extremo amino de la proteína y el dedo 4 es el más próximo al extremo carboxi.

Se obtuvo una molécula de ADN donante del siguiente modo. Primero, se amplificó un fragmento de 700 pares de bases de ADN genómico humano que representa los nucleótidos 52415936-52416635 de la hebra “-” del cromosoma X (publicación del genoma humano de UCSC de julio de 2003), que incluye el primer exón del gen hSMC1L1 humano, usando ADN genómico de células HEK293 como molde. Las secuencias de cebadores usadas para la amplificación se muestran en la Tabla 3 (“Amp inicial 1” y “Amp inicial 2”). A continuación se alteró el producto de PCR usando metodología de PCR de extensión por solapamiento estándar (véase, por ejemplo, Ho, et al., (1989) Gene 77:51-59), produciendo la sustitución de la secuencia ATGGGG (nucleótidos 24-29 en la Figura 1) a ATAAGAAGC. Este cambio produjo la conversión del codón ATG (metionina) en un codón ATA (isoleucina) y la sustitución de GGG (nucleótidos 27-29 en la Figura 1) por la secuencia AGAAGC, permitiendo la discriminación entre secuencias derivadas de donante y secuencias cromosómicas endógenas tras la recombinación. Un diagrama esquemático del gen hSMC1, que incluye secuencias del ADN cromosómico en la región del codón de iniciación, y secuencias en el ADN donante que se diferencian de la secuencia cromosómica, se facilita en la Figura 3. El fragmento donante de 700 pares de bases resultante se clonó en pCR4BluntTopo, que no contiene ninguna secuencia homóloga al genoma humano. Véase la Figura 4.

Para la mutación dirigida del gen hSMC1L1 cromosómico, los dos plásmidos que codifican fusiones de ZFP-*FokI* y el plásmido donante se introdujeron en 1×10^6 células HEK293 por transfección usando Lipofectamine 2000® (Invitrogen). Los controles incluyeron células transfectadas solo con los dos plásmidos que codifican las fusiones de ZFP-*FokI*, células transfectadas solo con el plásmido donante y células transfectadas con un plásmido de control (pEGFP-N1, Clontech). Las células se cultivaron en 5 % de CO₂ a 37 °C. 48 horas después de la transfección, se aisló ADN genómico de las células, y se usaron 200 ng como molde para la amplificación por PCR, usando un cebador complementario a una región del gen fuera de su región de homología con las secuencias donantes (nucleótidos 52416677-52416701 sobre la HEBRA “-” del cromosoma X; UCSC Julio de 2003), y un segundo cebador complementario a una región de la molécula donante en la que se introdujeron mutaciones distintivas. Usando estos dos cebadores, se obtendrá un producto de amplificación de 400 pares de bases a partir del ADN genómico si se ha producido un evento de recombinación dirigida. Las secuencias de estos cebadores se facilitan en la Tabla 3 (marcado “específica de cromosoma” y “específica de donante”, respectivamente). Las condiciones para la amplificación fueron: 94 °C, 2 min, seguido de 40 ciclos de 94 °C, 30 s, 60 °C, 1 min, 72 °C, 1 min; y una etapa final de 72 °C, 7 min.

Los resultados de este análisis (Figura 5) indican que un producto de amplificación de 400 pares de bases (marcado “ADN quimérico” en la figura) se obtuvo solo con ADN extraído de células que se había transfectado con el plásmido donante y ambos plásmidos de ZFP-*FokI*.

Tabla 3: Cebadores de amplificación para el gen hSMC1L1

Amp. inicial 1	AGCAACAACCTCCTCCGGGGATC (SEC ID NO: 37)
Amp. inicial 2	TTCCAGACGCGACTCTTTGGC (SEC ID NO: 38)
Cromosoma específico	CTCAGCAAGCGTGAGCTCAGGTCTC (SEC ID NO: 39)
Donante específico	CAATCAGTTTCAGGAAGCTTCTT (SEC ID NO: 40)
Exterior 1	CTCAGCAAGCGTGAGCTCAGGTCTC (SEC ID NO: 41)
Exterior 2	GGGGTCAAGTAAGGCTGGGAAGC (SEC ID NO: 42)

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 (SEC ID N°: 6), la secuencia amplificada se obtuvo a partir del ADN cromosómico de 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) ligada covalentemente a secuencias cromosómicas no presentes en la molécula donante (nucleótidos 32-97 de la Figura 6), que indica que las secuencias donantes se han recombinado en el cromosoma. En particular, la mutación de G→A que convierte el codón de iniciación en un codón de isoleucina se observa en la posición 395 en la secuencia.

En un segundo experimento, ADN cromosómico de células transfectadas solo con plásmido donante, células transfectadas con ambos plásmidos de fusión de ZFP-*FokI*, células transfectadas con el plásmido donante y ambos plásmidos de fusión de ZFP-*FokI* o células transfectadas con el plásmido de control EGFP se usaron como molde para la amplificación, usando cebadores complementarios a secuencias fuera de la región de homología de 700 nucleótidos entre las secuencias donantes y cromosómicas (identificado como “Fuera de 1” y “Fuera de 2” en la Tabla 3). El producto de amplificación resultante se purificó y se usó como molde para una segunda reacción de amplificación usando los cebadores específicos de donante y específicos de cromosoma descritos anteriormente (Tabla 3). Esta amplificación dio un producto de 400 nucleótidos solo de células transfectadas con la construcción donante y ambas construcciones de fusión de ZFP-*FokI*, un resultado de acuerdo con la sustitución de secuencias genómicas por recombinación dirigida en estas células.

Ejemplo 2: Edición de un gen IL-2R γ cromosómico por recombinación dirigida

El gen IL-2R γ codifica una proteína, conocida como la “cadena gamma del receptor de citocinas común”, que sirve de subunidad de varios receptores de interleucina (que incluyen IL-2R, IL-4R, IL-7R, IL-9R, IL-15R y IL-21R). Las mutaciones en este gen, que incluyen aquellas que rodean el extremo 5' del tercer exón (por ejemplo, el codón de tirosina 91), pueden producir inmunodeficiencia combinada grave (IDCG) ligada al X. Véase, por ejemplo, Puck et al., (1997) Blood 89:1968-1977. Se introdujo una mutación en el codón de tirosina 91 (nucleótidos 23-25 de SEC ID N°: 7; Figura 7) en el gen IL-2R γ por escisión y recombinación dirigidas. La escisión se dirigió a esta región diseñando dos pares de proteínas de dedos de cinc. El primer par (dos primeras filas de la Tabla 4) comprende una proteína de dedos de cinc diseñada para unirse a los nucleótidos 29-40 (contactos primarios a lo largo de la hebra superior como se muestra en la Figura 7) y una proteína de dedos de cinc diseñada para unirse a los nucleótidos 8-20 (contactos primarios a lo largo de la hebra inferior). El segundo par (tercera y cuarta filas de la Tabla 4) comprende dos proteínas de dedos de cinc, reconociendo la primera los nucleótidos 23-34 (contactos primarios a lo largo de la hebra superior como se muestra en la Figura 7) y reconociendo la segunda los nucleótidos 8-16 (contactos primarios a lo largo de la hebra inferior). Se diseñaron proteína de dedos de cinc como se describe en las patentes de EE.UU. del mismo solicitante 6.453.242 y 6.534.261. Véase la Tabla 4 para las secuencias de aminoácidos de las regiones de reconocimiento de las proteínas de dedos de cinc.

Las secuencias que codifican los dominios de unión de ZFP se fusionaron con las secuencias que codifican un medio dominio de escisión de *FokI* (aminoácidos 384-579 de la secuencia nativa de *FokI*, Kita et al., arriba), de forma que la proteína codificada contuviera las secuencias de *FokI* en el extremo carboxi y las secuencias de ZFP en el extremo amino. Cada una de estas secuencias de fusión se clonó entonces en un vector de expresión de mamífero modificado pcDNA3. Véase la Figura 8 para un diagrama esquemático de las construcciones.

Tabla 4: Diseños de dedos de cinc para el gen IL-2R γ

Secuencia objetivo	F1	F2	F3	F4
AACTCGGATAAT (SEC ID NO: 43)	DRSTLIE (SEC ID NO:44)	SSSNLSR (SEC ID NO:45)	RSDDLK (SEC ID NO:46)	DNSNRIK (SEC ID NO:47)
TAGAGGaGAAAGG (SEC ID NO:48)	RSDNLSN (SEC ID NO:49)	TSSSRIN (SEC ID NO:50)	RSDHLSQ (SEC ID NO:51)	RNADRKT (SEC ID NO:52)
TACAAGAACTCG (SEC ID NO:53)	RSDDLK (SEC ID NO:54)	DNSNRIK (SEC ID NO:55)	RSDALSV (SEC ID NO:56)	DNANRTK (SEC ID NO:57)
GGAGAAAGG (SEC ID NO:58)	RSDHLTQ (SEC ID NO:59)	QSGNLAR (SEC ID NO:60)	RSDHLSR (SEC ID NO:61)	

Nota: Las secuencias de aminoácidos de dedos de cinc mostradas anteriormente (en el código de una letra) representan los residuos -1 a +6, con respecto al inicio de la porción de hélice alfa de cada dedo de cinc. El dedo F1 es el más próximo al extremo amino de la proteína.

Se obtuvo una molécula de ADN donante del siguiente modo. Primero, se amplificó un fragmento de 700 pares de bases de ADN humano correspondiente a las posiciones 69196910-69197609 sobre la hebra “-” del cromosoma X (UCSC, Julio de 2003), que incluye el exón 3 del gen IL-2R γ , usando ADN genómico de células K562 como molde. Véase la Figura 9. Las secuencias de cebadores usadas para la amplificación se muestran en la Tabla 5 (marcada Amp inicial 1 y Amp inicial 2). A continuación se alteró el producto de PCR mediante metodología de PCR de extensión por solapamiento estándar (Ho, et al., arriba) para sustituir la secuencia TACAAGAACTCGGATAAT (SEC ID N°: 62) con la secuencia TAAAAGAATTCCGACAAC (SEC ID N°: 63). Esta sustitución produce la introducción de una mutación puntual en el nucleótido 25 (Figura 7), que convierte el codón de tirosina 91 TAC en un codón de terminación TAA y permite la discriminación entre secuencias cromosómicas derivadas de donante y endógenas tras la recombinación, debido a las diferencias en las secuencias en la dirección 3' del codón 91. El fragmento de 700 pares de bases resultante se clonó en pCR4BluntTopo que no contiene ninguna secuencia homóloga al genoma humano. Véase la Figura 10.

Para la mutación dirigida del gen IL-2R γ cromosómico, el plásmido donante, junto con dos plásmidos que codifican cada uno uno de un par de fusiones de ZFP-FokI, se introdujeron en 2×10^6 células K652 usando lipofección/electroporación mixta (Amaxa). Cada uno de los pares de ZFP/FokI (véase la Tabla 4) se probó en experimentos separados. Los controles incluyeron células transfectadas solo con dos plásmidos que codifican fusiones de ZFP-FokI, y células transfectadas solo con el plásmido donante. Las células se cultivaron en 5 % de CO₂ a 37 °C. 48 horas después de la transfección, se aisló ADN genómico de las células y se usaron 200 ng como molde para la amplificación por PCR, usando un cebador complementario a una región del gen fuera de su región de homología con las secuencias donantes (nucleótidos 69196839-69196863 sobre la hebra “+” del cromosoma X; UCSC, Julio de 2003) y un segundo cebador complementario a una región de la molécula donante en la que se introdujeron mutaciones distintivas (véase anteriormente) y cuya secuencia, por tanto, diverge de aquella del ADN cromosómico. Véase la Tabla 5 para las secuencias de cebadores, marcadas “específica de cromosoma” y “específica de donante”, respectivamente. Usando estos dos cebadores, se obtiene un producto de amplificación de 500 pb a partir del ADN genómico en el que se ha producido un evento de recombinación dirigida. Las condiciones para la amplificación fueron: 94 °C, 2 min, seguido de 35 ciclos de 94 °C, 30 s, 62 °C, 1 min, 72 °C, 45 s; y una etapa final de 72 °C, 7 min.

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 se habían transfectado con el plásmido donante y cualquiera de los pares de ZFP-FokI que codifican plásmidos. El ADN de células transfectadas con plásmidos que codifican un par de ZFP solas (sin plásmido donante) no produjo generación del producto de 500 pb, ni produjo ADN de células transfectadas solo con el plásmido donante.

Tabla 5: Cebadores de amplificación para el gen IL-2R γ

Amp. Inicial 1	TGTCGAGTACATGAATTGCACTTGG (SEC ID NO:64)
Amp. Inicial 2	TTAGGTTCTCTGGAGCCCAGGG (SEC ID NO:65)
Cromosoma específico	CTCCAAACAGTGGTTCAAGAATCTG (SEC ID NO:66)
Donante específico	TCCTCTAGGTAAAGAATTCCGACAAC (SEC ID NO:67)

Para confirmar este resultado, el producto de amplificación obtenido a partir del experimento usando el segundo par de fusiones de ZFP/FokI se clonó en pCR4Blunt-Topo (Invitrogen) y se determinó su secuencia de nucleótidos. Como se muestra en la Figura 12 (SEC ID N°: 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 terminación 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 al donante y el cromosoma, y los nucleótidos 460-552 son únicos para el cromosoma. La presencia de secuencias únicas del donante ligadas covalentemente a secuencias presentes en el cromosoma, pero no en el donante, indica que el ADN del plásmido donante se introdujo en el cromosoma por recombinación homóloga.

Ejemplo 3: Edición de un gen de β -globina cromosómico por recombinación dirigida

El gen de la beta-globina humana es uno de los dos productos génicos responsables de la estructura y función de la hemoglobina en eritrocitos humanos adultos. Mutaciones en el gen de la beta-globina pueden producir anemia drepanocítica. Se diseñaron dos proteínas de dedos de cinc para unirse dentro de esta secuencia, cerca de la localización de un nucleótido que, cuando se muta, produce anemia drepanocítica. 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 dedos de cinc están subrayados en la secuencia presentada en la Figura 13. Las secuencias de aminoácidos de las regiones de reconocimiento de las dos proteínas de dedos de cinc se muestran en la Tabla 6. Las secuencias que codifican cada uno de estos dos dominios de unión de ZFP se fusionaron con secuencias que codifican un medio dominio de escisión de FokI, como se ha descrito anteriormente, para crear ZFP-nucleasas manipuladas que

digirieron el gen de la beta-globina endógena. Cada una de estas secuencias de fusión se clonó entonces en el vector de expresión de mamífero pcDNA3.1 (Figura 14).

Tabla 6: Diseños de dedos de cinc para el gen de la beta-globina

Secuencia Objetivo	F1	F2	F3	F4
GGGCAGTAACGG (SEC ID NO: 68)	RSDHLS (SEC ID NO: 69)	QSANRTK (SEC ID NO: 70)	RSDNLSA (SEC ID NO: 71)	RSQNRTR (SEC ID NO: 72)
AAGGTGAACGTG (SEC ID NO: 73)	RSDSLSR (SEC ID NO: 74)	DSSNRKT (SEC ID NO: 75)	RSDLSLA (SEC ID NO: 76)	RNDNRKT (SEC ID NO: 77)

Nota: Las secuencias de aminoácidos de dedos de cinc mostradas anteriormente (en el código de una letra) representan los residuos -1 a +6, con respecto al inicio de la porción de hélice alfa de cada dedo de cinc. El dedo F1 es el más próximo al extremo amino de la proteína y el dedo 4 es el más próximo al extremo carboxi.

Se obtuvo una molécula de ADN donante del siguiente modo. Primero, se amplificó un fragmento de 700 pares de bases de ADN humano correspondiente a los nucleótidos 5212134 - 5212833 sobre la hebra “-” del cromosoma 11 (BLAT, sitio del genoma humano UCSC) por PCR usando ADN genómico de células K562 como molde. Las secuencias de cebadores usadas para la amplificación se muestran en la Tabla 7 (marcada Amp inicial 1 y Amp inicial 2). El fragmento amplificado resultante contiene secuencias correspondientes a un promotor, los dos primeros exones y el primer intrón del gen de la beta-globina humana. Véase la Figura 15 para un esquema que ilustra las localizaciones de exones 1 y 2, el primer intrón y los primeros sitios de unión en la secuencia de la beta-globina. Entonces, el producto clonado se modificó adicionalmente por PCR para introducir un conjunto de cambios de secuencias entre los nucleótidos 305-336 (como se muestra en la Figura 13), que sustituyó la secuencia CCGTACTGCCCTGTGGGGCAAGGTGAACGTG (SEC ID N°: 78) con **gCGTTAgTGCCCGAATTCGAtcGTcAACcac** (SEC ID N°: 79) (cambios en negrita). Ciertos de estos cambios (mostrados en minúscula) se manipularon específicamente para prevenir que las proteínas de fusión de ZFP/*FokI* se unieran a y escindieran la secuencia donante, una vez integradas en el cromosoma. Además, todos los cambios de secuencia permiten la discriminación entre secuencias cromosómicas donantes y endógenas tras 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).

Para la mutación dirigida del gen de la beta-globina cromosómico, los dos plásmidos que codifican fusiones de ZFP-*FokI* y el plásmido donante (pCR4-TOPO-HBBdonante) se introdujeron en 1×10^6 células K562 por transfección usando disolución Nucleofector™ (Amaya Biosystems). Los controles incluyeron células transfectadas solo con 100 ng (bajo) o 200 ng (alto) de los dos plásmidos que codifican las fusiones de ZFP-*FokI*, células transfectadas solo con 200 ng (bajo) o 600 ng (alto) del plásmido donante, células transfectadas con un plásmido que codifica GFP y células transfectadas con vector vacío. Las células se cultivaron en medio RPMI 1640 (Invitrogen) complementado con 10 % de suero bovino fetal (FBS) (Hyclone) y L-glutamina 2 mM. Las células se mantuvieron a 37 °C en una atmósfera de 5 % de CO₂. 72 horas después de la transfección, se aisló ADN genómico de las células y se usaron 200 ng como molde para la amplificación por PCR, usando un cebador complementario a una región del gen fuera de su región de homología con las secuencias donantes (nucleótidos 5212883-5212905 sobre la hebra “-” del cromosoma 11) y un segundo cebador complementario a una región de la molécula donante en la que se introdujeron mutaciones distintivas en la secuencia donante (véase arriba). Las secuencias de estos cebadores se facilitan en la Tabla 7 (marcada “específica de cromosoma” y “específica de donante”, respectivamente). Usando estos dos cebadores, se obtendrá un producto de amplificación de 415 pares de bases a partir del ADN genómico si se ha producido un evento de recombinación dirigida. Como control para la carga de ADN, también se llevaron a cabo reacciones de PCR usando los cebadores Amp inicial 1 y Amp inicial 2 para garantizar que niveles de ADN genómico similares se añadieron a cada reacción de PCR. Las condiciones para la amplificación fueron: 95 °C, 2 min, seguido de 40 ciclos de 95 °C, 30 s, 60 °C, 45 s, 68 °C, 2 min; y una etapa final de 68 °C, 10 min.

Los resultados de este análisis (Figura 17) indican que se obtuvo un producto de amplificación de 415 pares de bases solo con ADN extraído de células que se habían transfectado con la concentración “alta” de plásmido donante y ambos plásmidos de ZFP-*FokI*, de acuerdo con la recombinación dirigida de secuencias donantes en el sitio de beta-globina cromosómico.

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

Amp. inicial 1	TACTGATGGTATGGGGCCAAGAG (SEC ID NO:80)
Amp. inicial 2	CACGTGCAGCTTGTACAGTGC (SEC ID NO:81)
Cromosoma específico	TGCTTACCAAGCTGTGATTCCA (SEC ID NO:82)
Donante específico	GGTTGACGATCGGAATTC (SEC ID NO:83)

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 (SEC ID N°: 14), la secuencia consiste en una fusión entre las secuencias cromosómicas no presentes en el plásmido donante y las secuencias únicas para el plásmido donante. Por ejemplo, dos mutaciones C→G que rompen la unión de ZFP se observan en las posiciones 377 y 383 en la secuencia. Los nucleótidos 377-408 representan secuencia obtenida a partir del plásmido donante que contiene los cambios de secuencia descritos anteriormente; los nucleótidos 73-376 son secuencias comunes al donante y el cromosoma y los nucleótidos 1-72 son únicos para el cromosoma. El enlace covalente de secuencias específicas de donante y específicas de cromosoma en el genoma confirma la satisfactoria recombinación de la secuencia donante en el locus correcto dentro del genoma de células K562.

Ejemplo 4: Optimización del conector de ZFP-FokI (conector de ZC)

Con el fin de probar el efecto de la longitud del conector de ZC sobre la eficiencia de escisión, un dominio de unión de ZFP de cuatro dedos se fusionó con un medio dominio de escisión de FokI, usando conectores de ZC de diversas longitudes. El sitio diana para la ZFP es 5'-AACTCGGATAAT-3' (SEC ID N°: 84) y las secuencias de aminoácidos de las regiones de reconocimiento (posiciones - 1 a +6 con respecto al inicio de la hélice alfa) de cada uno de los dedos de cinc fueron del siguiente modo (en las que F1 es el dedo de cinc más terminal del extremo N y F4 es el más terminal del extremo C):

F1: DRSTLIE (SEC ID N°: 85)
 F2: SSSNLSR (SEC ID N°: 86)
 F3: RSDDLK (SEC ID N°: 87)
 F4: DNSNRIK (SEC ID N°: 88)

Se construyeron fusiones de ZFP-FokI, en las que el dominio de unión de ZFP anteriormente mencionado y un medio dominio de escisión de FokI se separaron por 2, 3, 4, 5, 6 ó 10 residuos de aminoácidos. 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 ó 26 pares de bases.

Las secuencias de aminoácidos de las construcciones de fusión, en la región del empalme de ZFP-FokI (con las secuencias de los diversos sustratos de escisión, con los sitios diana de ZFP subrayados, son del siguiente modo:

Conector de 10 residuos	HTKIH <u>LRQKDA</u> ARGSQLV (SEC ID N°: 89)
Conector de 6 residuos	HTKIH <u>LRQKGS</u> QLV (SEC ID N°: 90)
Conector de 5 residuos	HTKIH <u>LR</u> OGSQLV (SEC ID N°: 91)
Conector de 4 residuos	HTKIH <u>LRGS</u> QLV (SEC ID N°: 92)
Conector de 3 residuos	HTKIH <u>LGS</u> QLV (SEC ID N°: 93)
Conector de 2 residuos	HTKIH <u>GS</u> QLV (SEC ID N°: 94)

Las secuencias de los diversos sustratos de escisión, con los sitios diana de ZFP subrayados, son del siguiente modo:

separación de 4 pb	CTAGCATTATCCGAGTTACACA <u>ACTCGGATAAT</u> GCTAG GATCGT <u>AAATAGGCTCAAT</u> GTGTTGAGCCTATTACGATC (SEQ ID NO:95)
separación de 5 pb	CTAGCATTATCCGAGTTCACACA <u>ACTCGGATAAT</u> GCTAG GATCGT <u>AAATAGGCTCAAGT</u> GTGTTGAGCCTATTACGATC (SEQ ID NO:96)
separación de 6 pb	CTAGGCATTATCCGAGTTCACCACA <u>ACTCGGATAAT</u> GACTAG GATCCGT <u>AAATAGGCTCAAGT</u> GGTGTGTTGAGCCTATTACTGATC (SEQ ID NO:97)
separación de 7 pb	CTAGCATTATCCGAGTTCACACACA <u>ACTCGGATAAT</u> GCTAG GATCGT <u>AAATAGGCTCAAGT</u> GTGTTGAGCCTATTACGATC (SEQ ID NO:98)
separación de 8 pb	CTAGCATTATCCGAGTTCACCACACA <u>ACTCGGATAAT</u> GCTAG GATCGT <u>AAATAGGCTCAAGT</u> GGTGTGTTGAGCCTATTACGATC (SEQ ID NO:99)
separación de 9 pb	CTAGCATTATCCGAGTTCACACACA <u>ACTCGGATAAT</u> GCTAG GATCGT <u>AAATAGGCTCAAGT</u> GTGTTGAGCCTATTACGATC (SEQ ID NO:100)
separación de 12 pb	CTAGCATTATCCGAGTTCACCACCAACACA <u>ACTCGGATAAT</u> GCTAG GATCGT <u>AAATAGGCTCAAGT</u> GGTGTGTTGAGCCTATTACGATC (SEQ ID NO:101)
separación de 15 pb	CTAGCATTATCCGAGTTCACCACCAACCACACA <u>ACTCGGATAAT</u> GCTAG GATCGT <u>AAATAGGCTCAAGT</u> GGTGTGTTGAGCCTATTACGATC (SEQ ID NO:102)

Para fusiones de ZFP-*FokI* con conectores de cuatro residuos, la secuencia de aminoácidos del conector también se varió. En construcciones separadas, la secuencia conectora de LRGS original (SEC ID N°: 107) se cambió a LGGS (SEC ID N°: 108), TGGG (SEC ID N°: 109), GGGG (SEC ID N°: 110), LPGS (SEC ID N°: 111), LRKS (SEC ID N°: 112) y LRWS (SEC ID N°: 113); y las fusiones resultantes se probaron sobre sustratos que tenían una separación de seis pares de bases entre sitios de unión. Se observaron fusiones que contenían la secuencia conectora LGGS (SEC ID N°: 108) para escindir más eficazmente que aquellas que contenían la secuencia LRGS original (SEC ID N°: 107). Las fusiones que contienen las secuencias LRKS (SEC ID N°: 112) y LRWS (SEC ID N°: 113) se escindieron con menos eficiencia que la secuencia LRGS (SEC ID N°: 107), mientras que las eficiencias de escisión de las restantes fusiones fueron similares a las de la fusión que comprende la secuencia LRGS original (SEC ID N°: 107).

Ejemplo 5: Especificidad de escisión elevada resultante de la alteración del medio dominio de escisión de *FokI* en la interfase de dimerización

Se diseñaron un par de proteínas de fusión de ZFP/*FokI* (indicadas 5-8 y 5-10) para unirse a sitios diana en el quinto exón del gen IL-2R γ , para promover la escisión en la región entre los sitios diana. La región relevante del gen, que incluye 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 conector de ZC de 10 aminoácidos. Con respecto a la porción de dedos de cinc de estas proteínas, las secuencias diana de ADN, además de las secuencias de aminoácidos de las regiones de reconocimiento en los dedos de cinc, se facilitan en la Tabla 9.

Tabla 9: Diseños de dedos de cinc para el gen IL-2R γ

Fusión	Secuencia Objetivo	F1	F2	F3	F4
5-8	ACTCTGTGGAAG (SEC ID NO: 114)	RSDNLSE (SEC ID NO:115)	RNAHRIN (SEC ID NO:116)	RSDTLSE (SEC ID NO:117)	ARSTRIT (SEC ID NO:118)
5-10	AACACG _a AACGTG (SEC ID NO: 119)	RSDLSLR (SEC ID NO:120)	DSSNRKT (SEC ID NO:121)	RSDLSLV (SEC ID NO:122)	DRSNRIT (SEC ID NO:123)

Nota: Las secuencias de aminoácidos de dedos de cinc mostradas anteriormente (en el código de una letra) representan los residuos -1 a +6, con respecto al inicio de la porción de hélice alfa de cada dedo de cinc. El dedo F1 es el más próximo al extremo amino de la proteína.

La capacidad de este par de proteínas de fusión para catalizar la escisión de ADN específica entre sus secuencias diana (véase la Figura 19) se probó *in vitro* usando un molde de ADN marcado que contenía la secuencia diana y ensayando la presencia de productos de digestión de diagnóstico. Se obtuvo escisión específica cuando se usaron ambas proteínas (Tabla 10, primera fila). Sin embargo, la proteína de fusión 5-10 (que comprende un medio dominio de escisión de *FokI* no mutante) también fue capaz de escisión anómala en un sitio no diana en ausencia de la proteína 5-8 (Tabla 10, segunda fila), posiblemente debido a auto-dimerización.

Por consiguiente, 5-10 se modificó en su medio dominio de escisión de *FokI* convirtiendo el residuo de aminoácido 490 de ácido glutámico (E) en lisina (K) (la numeración de residuos de aminoácidos en la proteína *FokI* es según Wah et al., arriba). Esta modificación se diseñó para prevenir la homodimerización alterando un residuo de aminoácido en la interfase de dimerización. La proteína 5-10 mutante (E490K), a diferencia de la parental 5-10, fue incapaz de escindir en sitios anómalos en ausencia de la proteína de fusión 5-8 (Tabla 10, fila 3). Sin embargo, la proteína 5-10 (E490K) mutante, junto con la 5-8, catalizaron la escisión específica del sustrato (Tabla 10, fila 4). Así, la alteración de un residuo en el medio dominio de escisión de 5-10, que participa en la dimerización, previno la escisión anómala por esta proteína de fusión debido a auto-dimerización. Un mutante E490R también presenta menores niveles de homodimerización que la proteína parental.

Además, la proteína 5-8 se modificó en su interfase de dimerización reemplazando el residuo de glutamina (Q) en la posición 486 con ácido glutámico (E). Este mutante 5-8 (Q486E) se probó para su capacidad para catalizar la escisión dirigida en presencia de cualquiera de la proteína 5-10 no mutante o la mutante 5-10 (E490K). No se observó escisión de ADN cuando el sustrato marcado se incubó en presencia de tanto 5-8 (Q486E) como 5-10 no mutante (Tabla 10, fila 5). Sin embargo, se obtuvo escisión cuando los mutantes 5-8 (Q486E) y 5-10 (E490K) se usaron en combinación (Tabla 10, fila 6).

Estos resultados indican que la escisión de ADN por un par de proteínas de fusión de ZFP/*FokI*, en regiones distintas de aquellas definidas por las secuencias diana de las dos proteínas de fusión, puede minimizarse o suprimirse alterando la secuencia de aminoácidos del medio dominio de escisión en una o ambas de las proteínas de fusión.

Tabla 10: Escisión de ADN por los pares de proteínas de fusión de ZFP/FokI que contienen medios dominios de escisión no mutantes y mutantes

	ZFP 5-8 dominio de unión	ZFP 5-10 dominio de unión	Escisión ADN	
5	1	FokI tipo salvaje	FokI tipo salvaje	Específico
	2	No presente	FokI tipo salvaje	No Específico
10	3	No presente	FokI E490K	Ninguno
	4	FokI tipo salvaje	FokI E490K	Específico
	5	FokI Q486E	FokI tipo salvaje	Ninguno
15	6	FokI Q486E	FokI E490K	Específico
20	<p>Nota: Cada fila de la tabla presenta los resultados de un experimento separado en el que las proteínas de fusión de ZFP/FokI se probaron para la escisión de un sustrato de ADN marcado. Una de las proteínas de fusión contuvo el dominio de unión de ADN de 5-8 y la otra proteína de fusión contuvo el dominio de unión de ADN de 5-10 (véanse la Tabla 9 y la Figura 19). La porción de medio dominio de escisión de las proteínas de fusión fue como se indica en la tabla. Así, las entradas en la columna de ZFP 5-8 indican el tipo de dominio de escisión de FokI fusionado con ZFP 5-8; y las entradas en la columna de ZFP 5-10 indican el tipo de dominio de escisión de FokI fusionado con ZFP 5-10. Para los mutantes de medio dominio de escisión de FokI, el número se refiere al residuo de aminoácidos en la proteína FokI; la letra que precede al número se refiere al aminoácido presente en la proteína no mutada y la letra que sigue al número indica el aminoácido al que el residuo no mutante se cambió en la generación de la proteína modificada.</p> <p>'No presente' indica que la proteína de fusión de ZFP/FokI entera se omitió de ese experimento particular. El sustrato de ADN usado en este experimento fue un producto de PCR de aproximadamente 400 pb que contiene los sitios diana para tanto ZFP 5-8 como ZFP 5-10. Véase la Figura 19 para las secuencias y orientación relativa de los dos sitios diana.</p>			
25				
30				

Ejemplo 6: Generación de un gen de proteína verde fluorescente potenciada (eGFP) defectuoso

La proteína verde fluorescente potenciada (eGFP) es una forma modificada de la proteína verde fluorescente (GFP; véase, por ejemplo, Tsien (1998) Ann. Rev. Biochem. 67:509-544) que contiene cambios en el aminoácido 64 (phe a leu) y 65 (ser a thr). Heim et al., (1995) Nature 373:663-664; Cormack et al., (1996) Gene 173:33-38. Se construyó un sistema indicador basado en eGFP generando una forma defectuosa del gen eGFP, que contenía un codón de terminación y una mutación del marco de lectura de 2 pb. La secuencia del gen eGFP se muestra en la Figura 22. Las mutaciones se insertaron por mutagénesis de PCR por solapamiento, usando el kit de alta fidelidad de Taq ADN polimerasa Platinum® (Invitrogen) y los oligonucleótidos GFP-Bam, GFP-Xba, stop sense2 y stop anti2 como cebadores (las secuencias de oligonucleótidos se enumeran debajo en la Tabla 11). GFP-Bam y GFP-Xba sirvieron de cebadores externos, mientras que los cebadores stop sense2 y stop anti2 sirvieron de cebadores internos que codifican los cambios de nucleótidos. El vector pGFP-NI (BD Biosciences), que codifica un gen eGFP de longitud completa, se usó como el molde de ADN en dos reacciones de amplificación separadas, la primera que utiliza los oligonucleótidos GFP-Bam y stop anti2 como cebadores y la segunda que usa los oligonucleótidos GFP-Xba y stop sense2 como cebadores. Esto generó dos productos de amplificación cuyas secuencias se solaparon. Estos productos se combinaron y se usaron como molde en una tercera reacción de amplificación, usando los oligonucleótidos GFP-Bam y GFP-Xba externos como cebadores, para regenerar un gen eGFP modificado en el que la secuencia GACCACAT (SEC ID N°: 124) en los nucleótidos 280-287 se sustituyó con la secuencia TAACAC (SEC ID N°: 125). Las condiciones de PCR para todas las reacciones de amplificación fueron del siguiente modo: el molde se desnaturalizó inicialmente durante 2 minutos a 94 grados y seguido de 25 ciclos de amplificación incubando la reacción durante 30 s a 94 °C, 45 s a 46 °C y 60 s a 68 °C. Se llevó a cabo una ronda de extensión final a 68 °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 usando el kit de clonación TOPO-TA (Invitrogen) para generar la construcción pCR(R)4-TOPO-GFPmut.

Tabla 11: Secuencias de oligonucleótidos para GFP

Oligo	secuencia 5'-3'
GFP-Bam	CGAATTCTGCAGTCGAC (SEC ID NO:126)
GFP-Xba	GATTATGATCTAGAGTCG (SEC ID NO:127)
stop sense2	AGCCGCTACCCCTAACACGAAGCAG (SEC ID NO:128)
stop anti2	CTGCTTCGTGTTAGGGGTAGCGGCT (SEC ID NO:129)

Ejemplo 7: Diseño y ensamblaje de nucleasas de dedos de cinc dirigidas a eGFP

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 se producen en orientación opuesta separando 6 pares de bases los dos sitios de unión. Véase la Figura 23. ZFP 287A se une a los nucleótidos 271-279 sobre la hebra no codificante, mientras que ZFP 296 se une a los nucleótidos 286-294 sobre la hebra codificante. La secuencia diana de ADN y de aminoácidos para las regiones de reconocimiento de las ZFP se enumeran a continuación, y en la Tabla 12:

287A:

F1 (GCGg) RSDDLTR (SEC ID N°: 130)
 F2 (GTA) QSGALAR (SEC ID N°: 131)
 F3 (GGG) RSDHLSR (SEC ID N°: 132)

296S:

F1 (GCA) QSGSLTR (SEC ID N°: 133)
 F2 (GCA) QSGDLTR (SEC ID N°: 134)
 F3 (GAA) QSGNLAR (SEC ID N°: 135)

Tabla 12: Diseños de dedos de cinc para el gen GFP

Proteína	Secuencia Objetivo	F1	F2	F3
287A	GGGGTAGCGg (SEC ID NO:136)	RSDDLTR (SEC ID NO:137)	QSGALAR (SEC ID NO:138)	RSDHLSR (SEC ID NO:139)
296S	GAAGCAGCA (SEC ID NO:140)	QSGSLTR (SEC ID NO:141)	QSGDLTR (SEC ID NO:142)	QSGNLAR (SEC ID NO:143)

Nota: Las secuencias de aminoácidos de dedos de cinc mostradas anteriormente (en el código de una letra) representan los residuos -1 a +6, con respecto al inicio de la porción de hélice alfa de cada dedo de cinc. El dedo F1 es el más próximo al extremo amino de la proteína y el dedo 3 es el más próximo al extremo carboxi.

Se generaron secuencias que codifican estas proteínas por ensamblaje por PCR (por ejemplo, la patente de EE.UU. n° 6.534.261), se clonaron entre los sitios *KpnI* y *BamHI* del vector pcDNA3.1 (Invitrogen) y se fusionaron en marco con el dominio catalítico de la endonucleasa *FokI* (aminoácidos 384-579 de la secuencia de Looney et al., (1989) Gene 80:193-208). Las construcciones resultantes se llamaron pcDNA3.1-GFP287-FokI y pcDNA3.1-GFP296-FokI (Figura 24).

Ejemplo 8: Escisión de ADN *in vitro* dirigida por nucleasas de dedos de cinc diseñadas

Se usó la construcción pCR(R)4-TOPO-GFPmut (Ejemplo 6) para proporcionar un molde para probar la capacidad de las proteínas de dedos de cinc 287 y 296 para reconocer específicamente sus sitios diana y escindir la forma modificada de eGFP *in vitro*.

Se obtuvo un fragmento de ADN que contiene el inserto que codifica eGFP defectuoso por amplificación por PCR, usando los cebadores universales T7 y T3 y pCR(R)4-TOPO-GFPmut como molde. Este fragmento se marcó en los extremos usando γ -³²P-ATP y T4 polinucleótido cinasa. El nucleótido sin incorporar se eliminó usando una columna MicroSpin G-50 (Amersham).

Se usó un sistema de transcripción/traducción acoplado *in vitro* para expresar las nucleasas de dedos de cinc 287 y 296 descritas en el Ejemplo 7. Para cada construcción, se incubaron 200 ng de ADN de plásmido linealizados en 20 μ l de mezcla de TnT y se incubaron a 30 °C durante 1 hora y 45 minutos. La mezcla de TnT contiene 100 μ l de lisado de TnT (que incluye ARN polimerasa T7, Promega, Madison, WI) complementado con 2 μ l de metionina (1 mM) y 2,5 μ l de ZnCl₂ (20 mM).

Para el análisis de la escisión de ADN, se combinaron alícuotas de cada una de las mezclas de reacción de transcripción/traducción acopladas 287 y 296, a continuación se diluyeron en serie con tampón de escisión. El tampón de escisión contiene Tris-HCl 20 mM a pH 8,5, NaCl 75 mM, MgCl₂ 10 mM, ZnCl₂ 10 μ M, DTT 1 mM, 5 % de glicerol, 500 μ g/ml de BSA. Se combinaron 5 μ l de cada dilución con aproximadamente 1 ng de sustrato de ADN (marcado en los extremos con ³²P usando T4 polinucleótido cinasa como se ha descrito anteriormente), y cada mezcla se diluyó adicionalmente para generar una reacción de escisión de 20 μ l que tenía la siguiente composición: Tris-HCl 20 mM a pH 8,5, NaCl 75 mM, MgCl₂ 10 mM, ZnCl₂ 10 μ M, DTT 1 mM, 5 % de glicerol, 500 μ g/ml de BSA. Las reacciones de escisión se incubaron durante 1 hora a 37 °C. La proteína se extrajo añadiendo 10 μ l de

disolución de fenol-cloroformo a cada reacción, mezclando y centrifugando para separar las fases. Se analizaron diez microlitros de la fase acuosa de cada reacción por electroforesis en un gel de 10 % de poliacrilamida.

5 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 reacciones en las que la dilución final de cada mezcla de reacción de transcripción/traducción acoplada (en la reacción de escisión) fue 1/156,25, 1/31,25, 1/12,5 y 1/5, respectivamente, produciendo volúmenes eficaces 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 menores que el fragmento de partida (carril marcado “control sin cortar” en la Figura 25) se correlaciona con cantidades crecientes de las endonucleasas de dedos de cinc 287 y 296 en la mezcla de reacción, 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

15 Se escindió un fragmento de ADN que codifica la eGFP mutada, eGFPmut, del vector pCR(R)4-TOPO-GFPmut (Ejemplo 6) y se clonó en los sitios *HindIII* y *NotI* de pcDNA4/TO, poniendo así este gen bajo el control de un promotor del CMV inducible por tetraciclina. El plásmido resultante se llamó pcDNA4/TO/GFPmut (Figura 26). Se cultivaron células T-Rex 293 (Invitrogen) en medio Eagle modificado con Dulbecco (DMEM) (Invitrogen) complementado con 10 % de suero bovino fetal (FBS) libre de Tet (HyClone). Las células se sembraron en placa de 6 pocillos al 50 % de confluencia, y dos pocillos se transfectaron cada uno con pcDNA4/TO/GFPmut. Las células se dejaron recuperar durante 48 horas, luego se combinaron las células de ambos pocillos y se fraccionaron en placas de 10 x 15 cm² en medio selectivo, es decir, medio complementado con 400 ug/ml de zeocina (Invitrogen). El medio se cambió cada 3 días, y después de 10 días se aislaron colonias individuales y se expandieron adicionalmente. Cada línea clónica se probó individualmente para la expresión inducible por doxiciclina (dox) del gen eGFPmut por RT-PCR cuantitativa (TaqMan®).

Para el análisis por RT-PCR cuantitativa, se aisló ARN total de células tratadas con dox y sin tratar usando el kit High Pure Isolation (Roche Molecular Biochemicals) y 25 ng de ARN total de cada muestra se sometieron a RT-PCR cuantitativa en tiempo real para analizar la expresión del gen endógeno, usando ensayos TaqMan®. Secuencias de la sonda y el cebador se muestran en la Tabla 13. Las reacciones se llevaron a cabo en una máquina ABI 7700 SDS (PerkinElmer Life Sciences) bajo 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 una etapa de desnaturalización de 10 minutos a 95 °C. La reacción en cadena de la polimerasa (PCR) se llevó a cabo con 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 usando el software SDS versión 1.7 y se muestran en la Figura 27, con expresión del gen eGFPmut normalizado a la expresión del gen GAPDH humano. Varias líneas celulares presentaron expresión dependiente de doxiciclina de eGFP; la línea 18 (T18) se eligió como una línea celular modelo para estudios adicionales.

Tabla 13: Oligonucleótidos para análisis de ARNm

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

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

Se construyó por PCR una construcción donante que contenía la información genética para corregir el gen eGFPmut defectuoso. La reacción de PCR se llevó a cabo como se ha descrito anteriormente, usando el vector peGFP-NI como molde. Para prevenir la expresión de fondo de la construcción donante en experimentos de recombinación dirigida, se eliminaron el primer codón de 12 pb y de iniciación 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 célula de mamífero, clonando TOPO-TA para crear pCR(R)4-TOPO-GFPdonante5 (Figura 28). La secuencia del inserto de eGFP de esta construcción (correspondiente a los nucleótidos 64-797 de la secuencia mostrados en la Figura 22) se muestra en la Figura 29

(SEC ID N°: 20).

Tabla 14: Oligonucleótidos para construcción de molécula donante

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

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

La línea celular estable T18 (Ejemplo 9) se transfectó con uno o ambos del plásmido de expresión de ZFP-FokI (pcDNA3.1-GFP287-FokI y pcDNA3.1-GFP296-FokI, Ejemplo 7) y 300 ng del plásmido donante pCR(R)4-TOPO-GFPdonante5 (Ejemplo 10) usando el reactivo LipofectAMINE 2000 (Invitrogen) en medio de suero 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 se pararon en la fase G2 del ciclo celular mediante la adición, 24 horas después de la transfección, de 100 ng/ml de Nocodazol (Figura 30) o vinblastina 0,2 uM (Figura 31). Se dejó que la parada de G2 continuara durante 24-48 horas, y luego se liberó por la eliminación del medio. Las células se lavaron con PBS y el medio se sustituyó con DMEM que contenía FBS libre de tetraciclina y 2 ng/ml de doxiciclina. Se dejó que las células se recuperaran durante 24-48 horas, y la eficiencia de corrección del gen se midió monitorizando el número de células que presentaron fluorescencia de eGFP, por análisis de citometría de flujo activada por fluorescencia (FACS). Se llevó a cabo análisis de FACS usando 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 monitorizando las emisiones a 525 nm (eje x). El fondo o la autofluorescencia se midieron monitorizando emisiones a 570 nm (eje y). Las células que presentaron emisión de alta fluorescencia a 525 nm y baja emisión a 570 nm (región E) se puntuaron positivas para la corrección génica.

Los resultados se resumen en la Tabla 15 y Figuras 30 y 31. Las Figuras 30 y 31 muestran resultados en los que las células T18 se transfectaron con los plásmidos pcDNA3.1-GFP287-FokI y pcDNA3.1-GFP296-FokI que codifican ZFP-nucleasas y el plásmido pCR(R)4-TOPO-GFPdonante5, la expresión de eGFP se indujo con doxiciclina y las células se pararon en G2 con tanto nocodazol (Figura 30) como vinblastina (Figura 31). Ambas figuras muestran trazos de FACS, en los que las células que presentan fluorescencia de eGFP se representan en la porción derecha inferior del trazo (identificada como la región E, que es la porción del cuadrante 4 debajo de la curva). Para células transfectadas que se habían tratado con nocodazol, el 5,35 % de las células presentaron fluorescencia de GFP, indicativa de la corrección del gen eGFP cromosómico mutante (Figura 30), mientras que el 6,7 % de las células tratadas con vinblastina experimentaron 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.

En resumen, estos experimentos muestran que, en presencia de dos ZFP-nucleasas y una secuencia donante, aproximadamente el 1 % de las células tratadas experimentaron corrección génica, y que este nivel de corrección aumentó 4-5 veces parando las células tratadas en la fase G2 del ciclo celular.

Tabla 15: Corrección de un gen eGFP cromosómico defectuoso

Expt.	Tratamiento ¹	Porcentaje de células con gen eGFP corregido ²
1	300 ng donante solo	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 nocodazol	5.35
6	como 4 + 0.2 uM vinblastina	6.7
7	no donante, no ZFP, 100 ng/ml nocodazol	0.01
8	no donante, no ZFP, 0.2 uM vinblastina	0.0
9	100 ng ZFP287/Q486E + 300 ng donante	0.0

Expt.	Tratamiento ¹	Porcentaje de células con gen eGFP corregido ²
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 nocodazol	2.37
13	como 11 + 0.2 uM vinblastina	2.56

Notas:

1: Células T18, que contienen un gen eGFP cromosómico defectuoso, se transfectoron con plásmidos que codifican una o dos ZFP-nucleasas y/o un plásmido donante que codifica una secuencia de eGFP no defectuosa, y la expresión del gen eGFP cromosómico se indujo con doxiciclina. Las células se pararon opcionalmente en la fase G2 del ciclo celular después de una inducción de eGFP. El análisis de FACS se realizó 5 días después de la transfección. 2: El número es el porcentaje de fluorescencia total que presenta alta emisión a 525 nm y baja emisión a 570 nm (región E del trazo de FACS).

Ejemplo 12: Corrección de un gen cromosómico defectuoso usando nucleasas de dedos de cinc con alteraciones de secuencia en la interfase de dimerización

Se probaron nucleasas de dedos de cinc cuyas secuencias se habían alterado en la interfase de dimerización para 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 ZFP-nucleasas (es decir, los medios dominios de escisión de *FokI*) se alteraron como se ha descrito en el Ejemplo 5. Así, un medio dominio de escisión E490K se fusionó con el dominio de ZFP GFP296 (Tabla 12) y un medio dominio de escisión Q486E se fusionó con la ZFP GFP287 (Tabla 12).

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 ZFP-nucleasas que tienen alteraciones en sus interfases de dimerización, en comparación con aquel obtenido en presencia de cualquiera de las nucleasas solas. Experimentos adicionales, en los que células T18 se transfectoron con plásmido donante y plásmidos que codifican las nucleasas de dedos de cinc 287/Q486E y 296/E490K, se pararon entonces en G2 con nocodazol o vinblastina, mostraron un aumento adicional en la frecuencia de corrección génica, con más del 2 % de las células mostrando fluorescencia de eGFP, indicativa 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

En un experimento similar a aquellos descritos en el Ejemplo 11, se probó el efecto de la longitud de la secuencia donante sobre la frecuencia de recombinación dirigida. Se transfectoron células T18 con las dos ZFP-nucleasas, y se indujo la expresión de eGFP con doxiciclina, como en el Ejemplo 11. Las células también se transfectoron con tanto el plásmido pCR(R)4-TOPO-GFPdonante5 (Figura 28) que contiene un inserto de eGFP de 734 pb (Figura 29) como en el Ejemplo 11, como un plásmido similar que contiene un inserto de secuencia de 1527 pb (Figura 32) homólogo al gen eGFP cromosómico mutado. Adicionalmente, se evaluó el efecto de la parada de G2 con nocodazol sobre la frecuencia de recombinación.

En un segundo experimento, se compararon longitudes del donante de 0,7, 1,08 y 1,5 kpb. Se transfectoron células T18 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 kpb, 1,08 kpb o 1,5 kpb, como se describe en el Ejemplo 11. Cuatro días después de la transfección, las células se ensayaron para la corrección del gen eGFP defectuoso por FACS, monitorizando la fluorescencia de GFP.

Los resultados de estos dos experimentos, mostrados en la Tabla 16, muestran que secuencia donante más larga aumenta la frecuencia de recombinación dirigida (y, por tanto, de corrección génica) y confirman que la parada de las células en la fase G2 del ciclo celular también aumenta la frecuencia de recombinación dirigida.

Tabla 16: Efecto de la longitud del donante y parada del ciclo celular sobre la frecuencia de recombinación dirigida

Longitud donante (kb)	<u>Experimento 1 Concentración Nocodazol:</u>		Experimento 2
	0 ng/ml	100 ng/ml	
0.7	1.41	5.84	1.2
1.08	No dona	No dona	2.2
1.5	2.16	8.38	2.3

Nota: Los números representan el porcentaje de fluorescencia total en la región E del trazo de FACS (véase el Ejemplo 11) que es una indicación de la fracción de células que han experimentado recombinación dirigida para corregir el gen eGFP cromosómico defectuoso.

Ejemplo 14: Edición del gen IL-2R γ humano endógeno por escisión dirigida y recombinación usando nucleasas de dedos de cinc

Se construyeron dos vectores de expresión, codificando cada uno una ZFP-nucleasa dirigida al gen IL-2R γ humano. Cada ZFP-nucleasa contuvo un dominio de unión de ADN basado en proteínas de dedos de cinc (véase la Tabla 17) fusionado con el dominio de nucleasa de la enzima de restricción de tipo IIS *FokI* (aminoácidos 384-579 de la secuencia de Looney et al., (1989) Gene 80:193-208) mediante un conector de ZC de cuatro aminoácidos (véase el Ejemplo 4). Las nucleasas se diseñaron para unirse a posiciones en el exón 5 del gen IL-2R γ cromosómico que rodea los codones 228 y 229 (un punto caliente mutacional en el gen) y para introducir una rotura bicatenaria en el ADN entre sus sitios de unión.

Tabla 17: Diseños de dedos de cinc para el exón 5 del gen IL-2R γ

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

Nota: Las secuencias de aminoácidos de dedos de cinc mostradas anteriormente (en el código de una letra) representan los residuos -1 a +6, con respecto al inicio de la porción de hélice alfa de cada dedo de cinc. El dedo 11 es el más próximo al extremo amino de la proteína.

La porción de unión a ADN completa de cada una de las endonucleasas quiméricas fue del siguiente modo:

Nucleasa dirigida a ACTCTGTGGAAG (SEC ID N°: 152)

**MAERPFQCRICMRNFSRSDNLSVHIRTHTGKPFACDICGRKFARNAH
RINHTKIHTGSQKPFQCRICMRNFSRSDLTSEHIRTHTGKPFACDICGRKFAA
RSTRTNHTKIHLRGS (SEQ ID NO:162)**

Nucleasa dirigida a AAAGCGGCTCCG (SEC ID N°: 157)

**MAERPFQCRICMRNFSRSDLTSEHIRTHTGKPFACDICGRKFAARSTRTHTK
IHTGSQKPFQCRICMRNFSRSDSLSKHIRTHTGKPFACDICGRKFAQRSNLKV
HTKIHLRGS (SEQ ID NO:163)**

Se transfectaron células 293 de riñón embrionario humano (Lipofectamine 2000; Invitrogen) con dos construcciones de expresión, codificando cada una de las ZFP-nucleasas descritas en el párrafo precedente. Las células también se transfectaron con una construcción donante que lleva como inserto un fragmento de 1.543 pb del locus de IL-2R γ correspondiente a las posiciones 69195166-69196708 de la hebra "menos" del cromosoma X (publicación del genoma humano de UCSC de julio de 2003), en el vector pCR4Blunt Topo (Invitrogen). La secuencia de inserción de IL-2R γ contuvo las dos siguientes mutaciones puntuales en la secuencia del exón 5 (subrayada):

F R V R S R F N P L C G S (SEC ID N°: 164)

TTTCGTGTTCCGGAGCCGGTTTAACCCGCTCTGTGGAAGT (SEC ID N°: 165)

La primera mutación (CGC→CGG) no cambia la secuencia de aminoácidos (línea superior) y sirve para afectar adversamente la capacidad de la ZFP-nucleasa para unirse al ADN donante, y a ADN cromosómico tras la recombinación. La segunda mutación (CCA→CCG) no cambia la secuencia de aminoácidos y crea un sitio de reconocimiento para la enzima de restricción *BsrBI*.

Se usaron tanto 50 como 100 nanogramos de construcción de expresión de ZFP-nucleasa y 0,5 ó 1 microgramo de la construcción donante por transfecciones duplicadas. También se realizaron los siguientes experimentos de control: transfección con un plásmido de expresión que codifica la proteína eGFP; transfección con construcción donante solo; y transfección con plásmidos que expresan las ZFP-nucleasas solo. Veinticuatro horas después de la transfección, se añadió vinblastina (Sigma) a concentración final 0,2 μM a una muestra en cada conjunto de duplicados, mientras que la otra siguió sin tratar. La vinblastina afecta la capacidad de las células para ensamblar el huso mitótico y, por tanto, actúa de un potente agente de parada de G_2 . Este tratamiento se realizó para potenciar la frecuencia de direccionamiento debido a que la vía de reparación de la rotura bicatenaria dirigida por homología es más activa que la unión de extremos no homólogos en la fase G_2 del ciclo celular. Tras un periodo de tratamiento de 48 h con vinblastina 0,2 μM , el medio de crecimiento se sustituyó, y las células se dejaron recuperar del tratamiento con vinblastina durante 24 horas adicionales. Entonces se aisló ADN genómico de todas las muestras de células usando el kit DNEasy Tissue (Qiagen). Entonces se ensayaron quinientos nanogramos de ADN genómico de cada muestra para la frecuencia de direccionamiento génico, probando la presencia de un nuevo sitio *BsrBI* en el locus cromosómico de IL-2R γ , usando el ensayo descrito esquemáticamente en la Figura 33.

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

Experimentos adicionales, en los que cantidades traza de una secuencia de ADN de IL-2R γ que contiene RFLP se añadieron a ADN genómico humano (que contiene el gen IL-2R γ no mutante), y la mezcla resultante se amplificó y sometió a digestión con una enzima de restricción que escinde en el RFLP, han indicado que tan solo el 0,5 % de la secuencia que contiene RFLP puede detectarse cuantitativamente usando este ensayo.

Tabla 18: Oligonucleótidos para el análisis del gen IL-2R γ humano

Oligonucleótido	Secuencia
Ex5_1.5detF1	GATTCAACCAGACAGATAGAAGG (SEC ID NO:166)
Ex5_1.5detR1	TTACTGTCTCATCCTTTACTCC (SEC ID NO:167)

Ejemplo 15: Recombinación dirigida en el locus IL-2R γ en células K562

K562 es una línea celular derivada de leucemia mielógena crónica humana. Las proteínas usadas para la escisión dirigida fueron fusiones de *FokI* con los dominios de unión de ADN de dedos de cinc 5-8G y 5-9D (Ejemplo 14, Tabla 17). La secuencia donante fue el fragmento de 1,5 kpb del gen IL-2R γ humano que contiene un sitio *BsrBI* introducido por mutación, descrito en el Ejemplo 14.

Se cultivaron células K562 en medio RPMI 1640 (Invitrogen) complementado con 10 % de suero bovino fetal (FBS) (Hyclone) y L-glutamina 2 mM. Todas las células se mantuvieron a 37 °C en una atmósfera de 5 % de CO₂. Estas células se transfectaron por Nucleofection™ (disolución V, Programa T16) (Amaxa Biosystems), según el protocolo del fabricante, transfectando 2 millones de células por muestra. Los ADN para la transfección, usados en diversas combinaciones como se describen más adelante, fueron un plásmido que codifica la endonucleasa de fusión de ZFP 5-8G-*FokI*, un plásmido que codifica la endonucleasa de fusión de ZFP 5-9D-*FokI*, un plásmido que contiene la secuencia donante (descrito anteriormente y en el Ejemplo 14) y el vector peGFP-N1 (BD Biosciences) usado como control.

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 # p-eGFP-N1 p5-8G p5-9D Donante vinblastina

Muestra #	p-eGFP-N1	p5-8G	p5-9D	Donante	vinblastina
1	5 mg	-	-	-	-
2	-	-	-	50 mg	-
3	-	-	-	50 mg	si
4	-	10 mg	10 mg	-	-
5	-	5 mg	5 mg	25 mg	-
6	-	5 mg	5 mg	25 mg	si
7	-	7.5 mg	7.5 mg	25 mg	-
8	-	7.5 mg	7.5 mg	25 mg	si

Se expusieron células tratadas con vinblastina a vinblastina 0,2 μ M 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 recogieron 4 días después de la transfección para el análisis de ADN genómico.

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 siguiente cebadores:

Exón 5 directo: GCTAAGGCCAAGAAAAGTAGGGCTAAAG (SEC ID N°: 168)
 Exón 5 inverso: TTCCTTCCATCACCAAACCTCTTG (SEC ID N°: 169)

Estos cebadores amplifican un fragmento de 1.669 pb del cromosoma X correspondiente a las posiciones 69195100-69196768 sobre la hebra “-” (publicación del genoma humano de UCSC de julio de 2003) que contiene el exón 5 del gen IL-2R γ . La amplificación de ADN genómico que ha experimentado recombinación homóloga con el ADN donante da un producto que contiene un sitio *Bsr*BI; mientras que el producto de amplificación de ADN genómico que no ha experimentado recombinación homóloga con ADN donante no contendrá este sitio de restricción.

Se incluyeron diez microcuries de cada uno de α -³²PdCTP y α -³²PdATP en cada reacción de amplificación para permitir la visualización de productos de reacción. Tras 20 ciclos de PCR, la reacción se desaló sobre una columna Sephadex G-50 (Farmacia) y se digirió con 10 unidades de *Bsr*BI (New England Biolabs) durante 1 hora a 37 °C. Entonces, la reacción se resolvió sobre 10 % de PAGE no desnaturizante, se secó y se expuso a una pantalla de PhosphorImager.

Los resultados de este experimento se muestran en la Figura 35. Cuando las células se transfectaron con el plásmido GFP de control, plásmido donante solo, o los dos plásmidos que codifican ZFP en ausencia de donante, no estuvo presente sitio *Bsr*BI 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 células que se transfectaron con el plásmido donante y ambos plásmidos que codifican ZFP contuvieron el sitio *Bsr*BI introducido por recombinación homóloga con el ADN donante (banda marcada “rflp”). La cuantificación del porcentaje de señal representada por el ADN que contiene RFLP, mostrada en la Figura 35, indicó que, bajo condiciones óptimas, hasta el 18 % de todos los genes IL-2R γ en la población de células transfectadas se alteraron por recombinación homóloga.

Se realizó un segundo experimento según el protocolo que se acaba de describir, excepto que las células se expandieron durante 10 días después de la transfección. Los ADN usados para la transfección se muestran en la Tabla 20.

Tabla 20

Muestra #	p-eGFP-N1	p5-8G	p5-9D	Donante	vinblastina
1	50 mg	-	-	-	-
2	-	-	-	50 mg	-

Muestra #	p-eGFP-N1	p5-8G	p5-9D	Donante	vinblastina
3	-	-	-	50 mg	si
4	-	7.5 mg	7.5 mg	-	-
5	-	5 mg	5 mg	25 mg	-
6	-	5 mg	5 mg	25 mg	si
7	-	7.5 mg	7.5 mg	50 mg	-
8	-	7.5 mg	7.5 mg	50 mg	si

15 El análisis de la digestión con *BsrBI* de ADN amplificado, mostrado en la Figura 36, demostró de nuevo que hasta el 18 % del los genes IL-2R γ habían experimentado alteración de secuencia mediante recombinación homóloga, después de múltiples rondas de división celular. Así, los eventos de recombinación dirigida son estables.

20 Además, se analizó ADN de células transfectadas en este segundo experimento por transferencia Southern. Para este análisis, se digirieron doce microgramos de ADN genómico de cada muestra con 100 unidades de *EcoRI*, 50 unidades de *BsrBI* y 40 unidades de *DpnI* (todos de New England Biolabs) durante 12 horas a 37 °C. Esta digestión genera un fragmento de *EcoRI* de 7,7 kpb del gen nativo IL-2R γ (que carece de un sitio *BsrBI*) y fragmentos de 6,7 y 1,0 kpb de un gen IL-2R γ cromosómico cuya secuencia ha sido alterada, por recombinación homóloga, para incluir el sitio *BsrBI*. Se incluyó *DpnI*, una enzima de restricción dependiente de la metilación, para destruir el ADN donante metilado en *dam*. El ADN genómico de células K562 sin metilar es resistente a digestión con *DpnI*.

30 Tras la digestión, el ADN genómico se purificó por extracción con fenol-cloroformo y precipitación con etanol, se resuspendió en tampón TE y se resolvió sobre un gel al 0,8 % de agarosa junto con una muestra de ADN genómico digerido con *EcoRI* y *SphI* para generar un marcador de tamaño. El gel se procesó para la transferencia alcalina tras el procedimiento convencional y se transfirió ADN a una membrana de nailon (Schleicher & Schuell). Entonces se realizó la hibridación a la transferencia usando un fragmento radiomarcado del locus IL-2R γ correspondiente a las posiciones 69198428-69198769 de la hebra “-” del cromosoma X (publicación del genoma humano de UCSC de julio de 2003). Esta región del gen está fuera de la región homóloga al ADN donante. Después de la hibridación, la membrana se expuso a una placa PhosphorImager y los datos se cuantificaron usando software de Molecular Dynamics. La alteración de la secuencia de IL-2R γ cromosómica se midió analizando la intensidad de la banda correspondiente al fragmento *EcoRI-BsrBI* (flecha próxima a la autorradiografía; sitio *BsrBI* indicado por triángulo relleno en el mapa encima de la autorradiografía).

40 Los resultados, mostrados en la Figura 37, indican que hasta el 15 % de las secuencias de IL-2R γ cromosómicas se alteraron por recombinación homóloga, confirmando así los resultados obtenidos por análisis por PCR de que el evento de recombinación dirigido fue estable a través de múltiples rondas de división celular. Los resultados de transferencia Southern también indican que los resultados mostrados en la Figura 36 no produjeron un artefacto de amplificación.

45 **Ejemplo 16: Recombinación dirigida en el locus IL-2R γ en citoblastos hematopoyéticos positivos para CD34**

Las enfermedades genéticas (por ejemplo, inmunodeficiencia combinada grave (IDCG) y anemia drepanocítica) pueden tratarse por corrección mediada por recombinación homóloga de la alteración de la secuencia de ADN específica responsable de la enfermedad. En ciertos casos, la eficiencia máxima y la estabilidad del tratamiento resultarían de la corrección del defecto genético en una célula pluripotente. Para este fin, este ejemplo demuestra la alteración de la secuencia del gen IL-2R γ en células de médula ósea positivas para CD34 humanas. Las células CD34⁺ son citoblastos hematopoyéticos pluripotenciales que dan lugar a los linajes eritroide, mieloide y linfoide.

55 Se compraron células CD34 humanas derivadas de la médula ósea de AllCells, LLC y se transportaron como disoluciones madre congeladas. Estas células se descongelaron y se dejó que reposaran durante 2 horas a 37 °C en una atmósfera de 5 % de CO₂ en medio RPMI 1640 (Invitrogen) complementado con 10 % de suero bovino fetal (FBS) (Hyclone) y L-glutamina 2 mM. Se transfectaron muestras de células (1x10⁶ o 2x10⁶ células) por Nucleofection™ (Amaxa Biosystems) usando el kit Human CD34 Cell Nucleofector™, según el protocolo del fabricante. Después de la transfección, las células se cultivaron en medio RPMI 1640 (Invitrogen) complementado con 10 % de FBS, L-glutamina 2 mM, 100 ng/ml de factor estimulante de colonias de granulocitos (G-CSF), 100 ng/ml de factor de citoblastos (SCF), 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 uM en el medio de crecimiento inmediatamente después de la transfección para bloquear la apoptosis. Se añadió inhibidor de caspasa adicional 48 horas después a una concentración final de 20 uM para prevenir adicionalmente la apoptosis. Estas células se mantuvieron a 37 °C en una atmósfera de 5 % de CO₂ y se recogieron 3 días después

de la transfección.

Los números de células y ADN usados para la transfección se muestran en la Tabla 21.

5

Tabla 21

Muestra	# células	p-eGFP-N1 ¹	Donante ²	p5-8G ³	p5-9D ³
1	1x10 ⁶	5 ug	-	-	-
2	2x10 ⁶	-	50 ug	-	-
3	2x10 ⁶	-	50 ug	7.5 ug	7.5 ug

10

15

1. Esto es un plásmido de control que codifica una proteína verde fluorescente potenciada.
 2. El ADN donante es un fragmento de 1,5 kpb que contiene secuencias del exón 5 del gen IL-2R γ con un sitio *Bsr*BI introducido (véase el Ejemplo 14).
 3. Éstos son plásmidos que codifican fusiones de *Fok*I con los dominios de unión de ADN de dedos de cinc 5-8G y 5-9D (véase la Tabla 17).

20

Se extrajo ADN genómico de las células usando el kit de purificación de ADN MasterPure (Epicentre). Debido a la presencia de glicógeno en el precipitado, la cuantificación precisa de este ADN usado como entrada en la reacción de PCR es imposible; los cálculos estimados usando análisis de geles de agarosa teñidos con bromuro de etidio indican que se usaron aprox. 50 ng de ADN en cada muestra. Entonces se realizaron treinta ciclos de PCR usando los siguientes cebadores, cada uno de los cuales se hibrida con el locus IL-2R γ cromosómico inmediatamente fuera de la región homóloga al donante de 1,5 kb: ex5_1,5detF3 GCTAAGGCCAAGAAAGTAGGGCTAAAG (SEC ID N°: 170) ex5_1,5detR3 TTCCTTCCATCACCAACCCTCTTG (SEC ID N°: 171)

25

30

Se incluyeron veinte microcuries de cada uno de α -³²P-dCTP y α -³²P-dATP 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 que se produce espontáneamente en el exón 5 del gen IL-2R γ en células Jurkat: este SNP crea un RFLP que destruye un sitio *Maell* que está presente en ADN humano normal. Por tanto, se creó un patrón de referencia añadiendo 1 ó 10 nanogramos de ADN genómico humano normal (obtenido a partir de Clontech, Palo Alto, CA) a 100 ó 90 ng de ADN genómico de Jurkat, respectivamente, y realizando la PCR como se ha descrito anteriormente. Las reacciones de PCR se desalaron sobre una columna G-50 (Amersham) y se digirieron durante 1 hora con enzima de restricción: se digirieron muestras experimentales con 10 unidades de *Bsr*BI (New England Biolabs); las reacciones "patrón de referencia" se digirieron con *Maell*. Los productos de digestión se resolvieron en PAGE al 10 % no desnaturizante (BioRad), el gel se secó y se analizó por exposición a una placa de PhosphorImager (Molecular Dynamics).

35

40

Los resultados se muestran en la Figura 38. Además del producto de PCR principal, correspondiente al fragmento de 1,6 kb del locus IL-2R γ ("wt" en el panel derecho de la Figura 38), se observó una banda adicional (marcada "rflp") en carriles correspondientes a muestras de células que se transfectaron con plásmidos que codifican tanto ZFP-nucleasas como la construcción de ADN donante. Esta banda adicional no apareció en los carriles de control, de acuerdo con la idea de que la ZFP-nucleasa ayudó en el direccionamiento génico del exón 5 del gen de la cadena gamma común producido en este experimento.

45

50

Aunque la cuantificación precisa de la tasa de direccionamiento se complica por la proximidad de la banda de RFLP a la banda no mutante; se estimó que la frecuencia dirigida, por comparación con el patrón de referencia (panel izquierdo), estaba entre 1-5 %.

Ejemplo 17: Efectos de la homología donante-diana

55

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

60

Por consiguiente, la secuencia donante descrita en el Ejemplo 10 se modificó, por mutagénesis por PCR, para generar una serie de construcciones donantes de ~700 pb con diferentes grados de no homología con la diana. Todos los donantes modificados contuvieron cambios de secuencia que corrigieron el defecto en el gen eGFP cromosómico y contuvieron mutaciones silenciosas adicionales (mutaciones de ADN que no cambian la secuencia de la proteína codificada) insertadas en la región codificante que rodea el sitio de escisión. Estas mutaciones silenciosas tuvieron la intención de prevenir la unión a, y escisión de, la secuencia donante por las fusiones de dedo de cinc-dominio de escisión, reduciendo así la competición entre la diana cromosómica prevista y el plásmido donante por la unión por las nucleasas quiméricas. Además, tras la recombinación homóloga, se minimizaría la

65

capacidad de las nucleasas quiméricas para unirse y re-escindir las secuencias cromosómicas recientemente insertadas (y que posiblemente estimulan otra ronda de recombinación, o que producen la unión de extremos no homólogos u otras alteraciones conducidas por la rotura bicatenaria del genoma).

5 Se probaron cuatro secuencias donantes diferentes. El donante 1 contiene 8 desapareamientos con respecto a la secuencia diana de eGFP defectuosa cromosómica, el donante 2 tiene 10 desapareamientos, el donante 3 tiene 6 desapareamientos y el donante 5 tiene 4 desapareamientos. Obsérvese que la secuencia donante 5 es idéntica a la secuencia de eGFP no mutante, pero contiene 4 desapareamientos 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	SEC ID NO.
Donante1	CTTCAGCCGCTATCCAGACCACATGAAACAACACGACTTCTT	172
Donante2	CTTCAGCCGGTATCCAGACCACATGAAACAACATGACTTCTT	173
Donante3	CTTCAGCCGCTACCCAGACCACATGAAACAGCAGACTTCTT	174
Donante5	CTTCAGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTT	175
GFP mut	CTTCAGCCGCTACCCCTAACAC---GAAGCAGCAGACTTCTT	176
GFP wt	CTTCAGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTT	177

Se transfectó la línea celular T18, como se ha descrito en el Ejemplo 11, con 50 ng de las construcciones de expresión 287-FokI y 296-FokI (Ejemplo 7 y Tabla 12) y 500 ng de cada construcción donante. Se realizó análisis de FACS como se ha descrito en el Ejemplo 11.

Los resultados, mostrados en la Tabla 23, indican que una disminución del grado de desapareamiento entre la secuencia donante y cromosómica diana (es decir, elevada homología) produce una elevada frecuencia de recombinación homóloga como se evalúa por restauración de la función de GFP.

Tabla 23¹

Donante	# Diferencias	Porcentaje de células de gen con eGFP corregido ²
Donante 2	10	0.45%
Donante 1	8	0.53%
Donante 3	6	0.89%
Donante 5	4	1.56%

1: Células T18, que contienen un gen eGFP cromosómico defectuoso, se transfectaron con plásmidos que codificaban dos ZFP-nucleasas y con plásmidos donantes que codifican una secuencia de eGFP no defectuosa que tiene diferentes números de desapareamientos de secuencia con la secuencia diana cromosómica. La expresión del gen eGFP cromosómico se indujo con doxiciplina y el análisis de FACS se realizó 5 días después de la transfección.

2: El número es el porcentaje de fluorescencia total que presenta alta emisión a 525 nm y baja emisión a 570 nm (región E del trazo de FACS).

Los resultados anteriores muestran que los niveles de recombinación homóloga aumentan disminuyendo el grado de divergencia de secuencias diana-donante. Sin desear ceñirse a teoría particular alguna o proponer un mecanismo particular, se observa que la mayor homología entre el donante y la diana facilitaría la recombinación homóloga aumentando la eficiencia por la que la maquinaria de recombinación homóloga celular reconoce la molécula donante como molde adecuado. Alternativamente, un aumento en la homología del donante con la diana podría también conducir a escisión del donante por las ZFP-nucleasas quiméricas. Un donante escindido podría ayudar a facilitar la recombinación homóloga aumentando la tasa de invasión de hebras o podría ayudar en el reconocimiento del extremo del donante escindido como estiramiento homólogo de ADN durante la búsqueda de homología por la maquinaria de recombinación homóloga. Además, estas posibilidades no son mutuamente exclusivas.

Ejemplo 18: Preparación de ARNip

Para probar si la disminución de los niveles celulares de proteínas que participaron en la unión de extremos no homólogos (NHEJ) facilita la recombinación homóloga dirigida, se realizó un experimento en el que los niveles de la proteína Ku70 se disminuyeron mediante inhibición de ARNip. Se generaron moléculas de ARNip dirigidas al gen Ku70 por transcripción de ADNc de Ku70, seguido de escisión del transcrito bicatenario con la enzima Dicer.

Brevemente, se usó un conjunto 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 conjuntos de fragmentos de ADNc (conjuntos A-E), que oscilan en tamaño de 500-750 pb. Los fragmentos en cada uno de estos cinco conjuntos se reamplificaron entonces usando cebadores que contienen el elemento promotor de la ARN polimerasa de bacteriófago T7, usando de nuevo un conjunto diferente de cebadores para cada conjunto de ADNc. La generación de ADNc y las reacciones de PCR se realizaron usando el sistema de ADNc Superscript Choice y polimerasa de alta fidelidad Platinum Taq (ambos de Invitrogen, Carlsbad, CA), según los protocolos y recomendaciones del fabricante.

Cada uno de los conjuntos de ADN amplificado se transcribió entonces *in vitro* con ARN polimerasa del bacteriófago T7 para generar cinco conjuntos (A-E) de ARN bicatenario (ARNbc), usando el kit de transcripción *in vitro* RNAMAXX (Stratagene, San Diego, CA) según las instrucciones del fabricante. Después de la precipitación con etanol, el ARN en cada uno de los conjuntos se resuspendió y se escindió *in vitro* usando la enzima Dicer recombinante (Stratagene, San Diego, CA) según las instrucciones del fabricante. Se purificaron productos de ARNip de 21-23 pb en cada uno de los cinco conjuntos por un método de dos etapas, usando primero una columna Microspin G-25 (Amershan), seguido de una columna Microcon YM-100 (Amicon). Cada conjunto de productos de ARNip se transfectó transitoriamente en la línea celular T7 usando Lipofectamine 2000®.

Se realizaron transferencias Western para ensayar la eficacia relativa de los conjuntos de ARNip en la supresión de la expresión de Ku70 aproximadamente 3 días después de la transfección. Brevemente, se lisaron células y se rompieron usando tampón RIPA (Santa Cruz Biotechnology), y se homogeneizaron pasando los lisados a través de un QIAshredder (Qiagen, Valencia, CA). Entonces, los lisados clarificados se trataron con tampón de muestra de SDS-PAGE (con β -mercaptoetanol usado como agente reductor) y se hirvieron durante 5 minutos. Entonces se resolvieron muestras sobre un gel NUPAGE al 4-12 % de gradiente y se transfirieron sobre una membrana de PVDF. La porción superior de la transferencia se expuso a un anticuerpo anti-Ku70 (Santa Cruz sc-5309) y la porción inferior se expuso a un anticuerpo anti-TF IIB (Santa Cruz sc-225, usado como control de entrada). Entonces, la transferencia se expuso a anticuerpo secundario de cabra anti-ratón conjugado con peroxidasa de rábano picante y se procesó para detección electroquimioluminiscente (ECL) usando un kit de Pierce Chemical Co. según las instrucciones del fabricante.

La Figura 39 muestra resultados representativos tras la transfección de dos de los conjuntos de ARNip (conjuntos D y E) en células T7. La transfección con 70 ng de ARNip E produce 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 que participa en la unión de extremos no homólogos

La reparación de una rotura bicatenaria en ADN genómico puede avanzar a lo largo de dos vías celulares diferentes; recombinación homóloga (HR) o unión de extremos no homólogos (NHEJ). Ku70 es una proteína que participó en NHEJ, que se une a los extremos de ADN libres resultantes de una rotura bicatenaria en el ADN genómico. Para probar si la reducción de la concentración intracelular de una proteína que participa en NHEJ aumenta la frecuencia de HR, se usaron ARN interferentes pequeños (ARNip), preparados como se describe en Ejemplo 18, para inhibir la expresión del ARNm de Ku70, reduciendo así los niveles de proteína Ku70, en células co-transfectadas con ADN donante y con plásmidos que codifican nucleasas quiméricas.

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

Se transfectaron células T7, como se ha descrito en el Ejemplo 11, con tanto 70 como 140 ng de uno de dos conjuntos de producto Dicer que se dirige a Ku70 (véase el Ejemplo 18). Se realizaron análisis de transferencia de proteína sobre extractos derivados de las células transfectadas para determinar si el tratamiento de células con ARNip produjo una disminución en los niveles de la proteína Ku70 (véase el ejemplo previo). La Figura 39 muestra que los niveles de la proteína Ku70 se redujeron en células que se habían tratado con 70 ng de ARNip del conjunto E.

Se co-transfectaron muestras celulares separadas en el mismo experimento con 70 ó 140 ng de ARNip (conjunto D o conjunto 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 de GFP de 1,5 kpb (Ejemplo 13), para determinar si la reducción de los niveles de Ku70 aumentó la frecuencia de recombinación homóloga. El protocolo experimental se describe en la Tabla 24. La

restauración de la actividad de eGFP, debido a recombinación homóloga, se ensayó por análisis de FACS como se describe en el Ejemplo 11.

Tabla 24

Expt. #	Donante ¹	ZFNs ²	SiRNA ³	% corrección ⁴
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 conjunto D	0.68
5	500 ng	50 ng cada uno	140 ng conjunto D	0.59
6	500 ng	50 ng cada uno	70 ng conjunto E	1.25
7	500 ng	50 ng cada uno	140 ng conjunto E	0.92

1. Un plásmido que contiene una secuencia de 1,5 kpb que codifica un proteína eGFP funcional que es homóloga al gen eGFP defectuoso cromosómicamente integrado
 2. Plásmidos que codifican las endonucleasas de fusión de proteína de dedo de cinc /*FokI* 287 y 296 dirigidas a eGFP
 3. Véase el Ejemplo 18
 4. Porcentaje de la fluorescencia total que presenta alta emisión a 525 nm y baja emisión a 570 nm (región E del trazo de FACS, véase el Ejemplo 11).

El porcentaje de corrección del gen eGFP defectuoso en las células T7 transfectadas (indicativo de la frecuencia de recombinación homóloga dirigida) se muestra en la columna derecha de la Tabla 24. La mayor frecuencia de recombinación dirigida se observa en el Experimento 6, en el que se transfectaron células con ADN donante, plásmidos que codifican las dos nucleasas de fusión dirigidas a eGFP y 70 ng del conjunto E de ARNip. Referencia al Ejemplo 18 y Figura 39 indica que 70 ng de ARNip del conjunto E redujeron significativamente los niveles de proteína Ku70. Así, pueden usarse métodos que reducen los niveles celulares de proteínas que participaron en NHEJ como un medio para facilitar la recombinación homóloga.

Ejemplo 20: Nucleasas de fusión de dedos de cinc-*FokI* dirigidas al gen de la β-globina humana

Se diseñaron varios dominios de unión a ADN de dedos de cinc de cuatro dedos, dirigidos al gen de la β-globina humana, y se construyeron plásmidos que codificaban cada dominio de dedos de cinc, fusionados con un medio dominio de escisión de *FokI*. Cada dominio de dedos de cinc contuvo cuatro dedos de cinc y reconoció un sitio diana de 12 pb en la región del gen de la β-globina humana que codifica la mutación responsable de la anemia drepanocítica. Se evaluó la afinidad de unión de cada una de estas proteínas por su secuencia diana, y se usaron cuatro proteínas que presentan fuerte unión (*sca-r29b*, *sca-36a*, *sca-36b* y *sca-36c*) para la construcción de endonucleasas de fusión de *FokI*.

Los sitios diana de los dominios de unión de ADN de ZFP, alineados con la secuencia del gen de la β-globina humana, se muestran a continuación. El codón de inicio de la traducción (ATG) está en negrita y subrayado, ya que es la sustitución A-T que causa la anemia drepanocítica.

sca-36a GAAGTCTGCCGT (SEQ ID NO:178)
sca-36b GAAGTctGCCGTT (SEQ ID NO:179)
sca-36c GAAGTctGCCGTT (SEQ ID NO:180)
 CAAACAGACACC**ATGGTGCATCTGACTCCTGTGGAGAAGTCTGCCGTTACTG**
 GTTTGTCTGTGGTACCACGTAGACTGAGGAC**ACCTCTTCAGACGGCAATGAC** (SEQ ID NO:181)
sca-r29b ACGTAGaCTGAGG (SEQ ID NO:182)

Se muestran secuencias de aminoácidos de las regiones de reconocimiento de los dedos de cinc en estas cuatro proteínas en la Tabla 25. Las secuencias de aminoácidos completas de estos dominios de dedos de cinc se muestran en la Figura 40. El dominio *sca-36a* reconoce un sitio diana que tiene 12 nucleótidos contiguos (mostrados en mayúscula anteriormente), mientras que los otros tres dominios reconocen una secuencia de trece nucleótidos que consiste en dos sitios diana de seis nucleótidos (mostrados en mayúscula) separados por un único nucleótido (mostrado en minúscula). Por consiguiente, los dominios *sca-r29b*, *sca-36b* y *sca-36c* contienen un conector entre

dedos no canónico que tiene la secuencia de aminoácidos TGGGGSQKP (SEC ID N°: 183) entre el segundo y el tercero de sus cuatro dedos.

Tabla 25

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

Ejemplo 21: Escisión *in vitro* de una secuencia diana de ADN por endonucleasas de fusión de ZFP/FokI dirigidas a β -globina

Se probaron proteínas de fusión que contenían un medio dominio de escisión de *FokI* y uno de los cuatro dominios de unión de ADN de ZFP descritas en el ejemplo previo para su capacidad para escindir ADN *in vitro* con la especificidad de secuencia predicha. Estos dominios de ZFP se clonaron en el vector de expresión pcDNA3.1 mediante los sitios KpnI y BamHI y se fusionaron en marco con el dominio de escisión de *FokI* mediante un conector de ZC de 4 aminoácidos, como se ha descrito anteriormente. Se clonó un fragmento de ADN que contenía 700 pb del gen de la β -globina humana a partir del ADN genómico obtenido a partir de células K562. El aislamiento y secuencia de este fragmento fue como se ha descrito en el Ejemplo 3, arriba.

Para producir endonucleasas de fusión (ZFN) para el ensayo *in vitro*, plásmidos circulares que codifican fusiones de *FokI* con proteína sca-r29b, sca-36a, sca-36b y sca-36c se incubaron en un sistema de transcripción/traducción *in vitro*. Véase el Ejemplo 4. Se añadió un total de 2 ul de la reacción de TNT (2 ul de una reacción única cuando una proteína única estaba siendo ensayada o 1 ul de cada reacción cuando un par de proteínas estaban siendo ensayadas) a 13 ul de la mezcla de tampón de escisión y 3 ul de sonda marcada (~1 ng/ul). La sonda se marcó en los extremos con ^{32}P usando polinucleótido cinasa. Esta reacción se incubó durante 1 hora a temperatura ambiente para permitir la unión de las ZFN. La escisión se estimuló mediante la adición de 8 ul de MgCl_2 8 mM, se diluyó 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 ul 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 ul de sonda sobre el gel para marcar la migración de ADN sin cortar (marcado "U" en la Figura 41).

Los resultados se muestran en la Figura 41. La incubación del ADN diana con cualquier fusión de un solo dedo de cinc/*FokI* no produjo cambio en el tamaño del molde de ADN. Sin embargo, la combinación de la nucleasa sca-r29b con cualquiera de las nucleasas sca-36b o sca-36c produjo la escisión del ADN diana, como se demuestra por la presencia de dos fragmentos de ADN más cortos (más a la derecha de los dos carriles de la Figura 41).

Ejemplo 22: Endonucleasas de fusión de ZFP/*FokI*, dirigidas al gen de la β -globina, probadas en un sistema indicador de GFP cromosómico

Se sintetizó un fragmento de ADN que contiene la secuencia del gen de la β -globina humana dirigida por las ZFN descritas en el ejemplo 20 y se clonó en un sitio *SpeI* en un gen indicador de eGFP, alterándose así la expresión de eGFP. El fragmento contenido tuvo la siguiente secuencia, en la que el nucleótido responsable de la mutación del drepanocito está en negrita y subrayado):

CTAGACACCATGGTGCATCTGACTCCTG**I**GGAGAAGTCTGCCGTTA

CTGCCCTAG (SEQ ID NO:200)

Este gen eGFP alterado que contiene secuencias de β -globina insertadas se clonó en pcDNA4/TO (Invitrogen, Carlsbad, CA) usando los sitios *HindIII* y *NotI*, y el vector resultante se transfeció en células HEK293 TRex (Invitrogen). Se aislaron clones estables individuales y se cultivaron, y los clones se probaron para la recombinación homóloga dirigida transfeciendo cada una de las proteínas sca-36 (sca-36a, sca-36b, sca-36c) emparejadas con sca-29b (véase el Ejemplo 20 y la Tabla 25 para secuencias y sitios de unión de estas nucleasas quiméricas). Las células se transfecieron con 50 ng de plásmido que codifica cada una de las ZFN y con 500 ng del donante de GFP de 1,5 kb (Ejemplo 13). Cinco días después de la transfección, las células se probaron para la recombinación homóloga en el locus de eGFP defectuoso insertado. Inicialmente, las células se examinaron por microscopía de

fluorescencia para la función de eGFP. Las células que presentan la fluorescencia se analizaron entonces cuantitativamente usando un ensayo de FACS para fluorescencia de eGFP, como se describe en el Ejemplo 11.

Los resultados mostraron que todas las líneas celulares transfectadas con sca-29b y sca-36a fueron negativas para la función de eGFP, cuando se ensayaron por microscopía de fluorescencia. Algunas de las líneas transfectadas con sca-29b emparejadas con tanto sca-36b como sca-36c fueron positivas para la expresión de eGFP, cuando se ensayaron por microscopía de fluorescencia, y, por tanto, se analizaron adicionalmente por análisis de FACS. Los resultados del análisis de FACS de dos de estas líneas se muestran en la Tabla 26, e indican que las nucleasas de dedos de cinc dirigidas a secuencias de β -globina pueden catalizar la escisión de ADN bicatenario específica de secuencia para facilitar la recombinación homóloga en células vivas.

Tabla 26

Linea de célula	ADN transfectado				% corr. ¹
	sca-29b	sca-36a	sca-36b	sca-36c	
#20	+	+			0
	+		+		0.08
	+			+	0.07
#40	+	+			0
	+		+		0.18
	+			+	0.12

1. Porcentaje de fluorescencia total que presenta alta emisión a 525 nm y baja emisión a 570 nm (región E del trazo de FACS, véase el Ejemplo 11).

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

Como la transcripción de una secuencia cromosómica de ADN implica alteraciones en su estructura de cromatina (generalmente para hacer las secuencias transcritas más accesibles), es posible que un gen activamente transcrito pudiera ser un sustrato más favorable para la recombinación homóloga dirigida. Esta idea se probó usando 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 de doxiciclina.

Se transfectaron muestras separadas de células T18 con plásmidos que codifican las proteínas de fusión de dedos de cinc/*FokI* 287 y 296 dirigidas a eGFP (Ejemplo 7) y una molécula de ADN donante de 1,5 kpb 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 niveles de ARNm de eGFP 48 horas después de la adición, de doxiciclina. Se midió la fluorescencia de eGFP a 520 nm (indicativa de recombinación dirigida de la secuencia donante en el cromosoma para sustituir las secuencias de β -globina insertadas) por FACS 4 días después de transfección.

Los resultados se muestran en la Figura 42. Niveles en estado estacionario crecientes de ARNm de eGFP normalizados a ARNm de GAPDH (equivalente, en una primera aproximación, a la tasa de transcripción del gen eGFP cromosómico defectuoso) se indican por las barras. El número encima de cada barra indica el porcentaje de células que presenta fluorescencia de eGFP. Los resultados muestran que el aumento de la tasa de transcripción del gen diana va acompañado de mayores frecuencias de recombinación dirigida. Esto sugiere que la activación dirigida de la transcripción (como se ha desvelado, por ejemplo, en las patentes de EE.UU. del mismo solicitante 6.534.261 y 6.607.882) puede usarse, conjuntamente con escisión de ADN dirigida, para estimular la recombinación homóloga dirigida en células.

Ejemplo 24: Generación de una línea celular que contiene una mutación en el gen IL-2R γ

Se transfectaron células K562 con plásmidos que codifican las nucleasas de dedos de cinc 5-8GL0 y 5-9DL0 (ZFN) (véase Ejemplo 14; Tabla 17) y con una construcción donante de *Dral* de 1,5 kpb. El donante *Dral* comprende una secuencia con homología con la región que codifica el 5^o exón del gen IL-2R γ , pero inserta una base adicional entre los sitios de unión de ZFN para crear un desplazamiento del marco y generar un sitio *Dral*.

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

Se dejó que las células se recuperaran durante 2 días adicionales y se sembraron 1600 células individuales en 40x

placas de 96 pocillos en 100 ul de medio.

Las células se cultivan durante aproximadamente 3 semanas, y se aíslan las células homocigóticas para el fenotipo mutante de *Dral*. Las células se prueban para modificación del genoma (probando la presencia de un sitio *Dral* en el exón 5 del gen IL-2R γ) y para niveles de ARNm de IL-2R γ (por PCR en tiempo real) y proteína (por transferencia Western) para determinar el efecto de la mutación sobre la expresión génica. Las células se prueban para la función por análisis de FACS.

Se transfectan células que contienen la mutación del marco de lectura *Dral* en el gen IL-2R γ con plásmidos que codifican las proteínas de fusión 5-8GL0 y 5-9DL0 y una construcción donante de *BsrBI* de 1,5 kb (Ejemplo 14) para sustituir la mutación del marco de lectura *Dral* con una secuencia que codifica una proteína funcional. Se obtienen niveles de recombinación homóloga superiores al 1 % en estas células, como se mide ensayando la presencia de un sitio *BsrBI* como se describe en el Ejemplo 14. La recuperación de la función génica se demuestra midiendo los niveles de ARNm y de proteína y por análisis de FACS.

Ejemplo 25: Endonucleasas de fusión de ZFP/*FokI* con diferentes polaridades

Se construyó un vector que codifica una fusión de ZFP/*FokI*, en la que el dominio de ZFP estaba en el extremo N con respecto al dominio de *FokI*. El dominio de ZFP, indicado IL2-1, contuvo cuatro dedos de cinc, y se dirigió a la secuencia AACTCGGATAAT (SEC ID N°: 202), localizada en el tercer exón del gen IL-2R γ . Las secuencias de aminoácidos de las regiones de reconocimiento de los dedos de cinc se facilitan en la Tabla 27.

Tabla 27: Diseño de dedos de cinc del dominio de unión de IL2-1

Secuencia Objetivo	F1 (AAT)	F2 (GAT)	F3 (TCG)	F4 (AAC)
AACTCGGATAAT (SEC ID NO: 203)	DRSTLIE (SEQ ID NO:204)	SSNSLR (SEQ ID NO:205)	RSDDLK (SEQ ID NO:206)	DNSNRK (SEQ ID NO:207)
Nota: La secuencia diana de ADN se muestra en la columna más a la izquierda. Las restantes columnas muestran las secuencias de aminoácidos (en el código de una letra) de residuos -1 a +6 de cada uno de los cuatro dedos de cinc, con respecto al inicio de la porción de hélice alfa de cada dedo de cinc. El dedo F1 es el más próximo al extremo amino de la proteína. El subsitio de tres nucleótidos unidos por cada dedo se muestra en la fila superior adyacente a la designación del dedo.				

Se unieron secuencias que codifican este dominio de dedos de cinc a secuencias que codifican el medio dominio de escisión de la endonucleasa de restricción *FokI* (aminoácidos 384-579 según Looney et al., (1989) Gene 80:193-208) de forma que un conector de cuatro aminoácidos estuvo presente entre el dominio de ZFP y el medio dominio de escisión (es decir, un conector de ZC de cuatro aminoácidos). El medio dominio de escisión de *FokI* se obtuvo por amplificación por PCR de ADN genómico aislado de la cepa bacteriana *Planomicrobium okeanoites* (ATCC 33414) usando los siguiente cebadores:

5'-GGATCCCAACTAGTCAAAAGTGAAC (SEC ID N°: 208)

5'-CTCGAGTAAAAGTTTATCTCGCCG (SEC ID N°: 209).

El producto de PCR se digirió con *BamHI* y *XhoI* (sitios subrayados en las secuencias mostradas anteriormente) y a continuación se ligó con un fragmento de vector preparado a partir del plásmido pcDNA-nls-ZFP1656-VP16-flag después de la digestión con *BamHI* y *XhoI*. La construcción resultante, pcDNA-nls-ZFP1656-*FokI*, codifica una proteína de fusión que contiene, del extremo N al extremo C, una señal de localización nuclear derivada del antígeno T grande del SV40 (NLS, Kalderon et al., (1984) Cell 39:499-509), ZFP1656, y un medio dominio de escisión de *FokI*, en un esqueleto de vector pcDNA3.1 (Invitrogen, Carlsbad, CA). Esta construcción se digirió con *KpnI* y *BamHI* para liberar las secuencias codificantes de ZFP1656, y un fragmento *KpnI/BamHI* que codificaba el dominio de unión de dedos de cinc de IL2-1 se insertó por ligación. La construcción resultante (pIL2-1C) codifica una proteína de fusión que comprende, del extremo N al C, una señal de localización nuclear, el dominio de unión de dedos de cinc de IL2-1 de cuatro dedos y un medio dominio de escisión de *FokI*, con un conector de ZC de cuatro aminoácidos.

También se construyó un vector que codifica una proteína de fusión de ZFP/*FokI*, en la que las secuencias de *FokI* estuvieron en el extremo N con respecto a las secuencias de ZFP. Se insertó el dominio de dedos de cinc de cuatro dedos de IL2-1, como un fragmento *KpnI/BamHI*, en un vector que codifica una proteína de fusión que contiene una NLS, el dominio de represión KOX-1, EGFP y una marca de epítipo FLAG, que se había digerido con *KpnI* y *BamHI* para liberar las secuencias codificantes de EGFP. Esto generó un vector que contiene secuencias que codifican, del extremo N al extremo C, una NLS (del antígeno T grande del SV40), un dominio de represión KOX, el dominio de dedos de cinc de IL2-1 y una marca FLAG de epítipo. Esta construcción se digirió entonces con *EcoRI* y *KpnI* para liberar las secuencias codificantes de NLS y KOX, y se insertó un fragmento *EcoRI/KpnI* (generado por PCR usando, como molde, un vector que codifica *FokI*) que codifica los aminoácidos 384-579 de la enzima de restricción de *FokI* y una NLS. La construcción resultante, pIL2-1R, codifica una proteína de fusión que contiene, del extremo N al

extremo C, un medio dominio de escisión de *FokI*, una NLS y el dominio de unión de IL2-1-ZFP de cuatro dedos. El conector de ZC en esta construcción tiene 21 aminoácidos de longitud e incluye la secuencia de siete aminoácidos de localización nuclear (PKKKRKV; SEC ID N°: 210).

5 El dominio de dedos de cinc 5-9D se une a la secuencia diana de 12 nucleótidos AAAGCGGCTCCG (SEC ID N°: 157) localizada en el quinto exón del gen IL-2R γ . Véase el Ejemplo 14 (Tabla 17). Las secuencias que codifican el dominio de dedos de cinc 5-9D se insertaron en un vector para generar una fusión de *FokI*/ZFP, en la que las secuencias de *FokI* estuvieron en el extremo N con respecto a las secuencias de ZFP. Para preparar esta construcción, el plásmido pIL2-1R descrito en el párrafo previo se digirió con *KpnI* y *BamHI* para liberar un fragmento que contiene secuencias que codifican el dominio de unión de dedos de cinc de IL2-1, y un fragmento *KpnI*/*BamHI* que codifica el dominio de unión de dedos de cinc 5-9D se insertó en su sitio. La construcción resultante, p5-9DR, codifica una proteína de fusión que contiene, del extremo N al extremo C, un medio dominio de escisión de *FokI*, una NLS y el dominio de unión de dedos de cinc 5-9D de cuatro dedos. El conector de ZC en esta construcción tiene 22 aminoácidos de longitud e incluye la secuencia de localización nuclear de siete aminoácidos (PKKKRKV; SEC ID N°: 210).

Véanse las patentes de EE.UU. 6.453.242 y 6.534.261 del mismo solicitante para detalles adicionales de la construcción de vectores.

20 **Ejemplo 26: Construcción de sustratos sintéticos para la escisión de ADN**

Las secuencias diana unidas por las proteínas de fusión de IL2-1 y 5-9D descritas anteriormente se introdujeron en fragmentos de ADN bicatenario en una variedad de orientaciones, para probar la capacidad de escisión de proteínas de fusión de dedos de cinc/*FokI* que tienen una polaridad alterada en la que el dominio de *FokI* está en el extremo N con respecto al dominio de ZFP. En el molde 1, el sitio diana de 5-9D está presente en una hebra y el sitio diana de IL2-1 está presente sobre la hebra complementaria, estando los extremos 3' de los sitios de unión proximales entre sí y separados por seis pares de nucleótidos intermedios. En el molde 2, los sitios diana de 5-9D y IL2-1 están presentes sobre la misma hebra de ADN, con el extremo 3' del sitio de unión de 5-9D separado por seis pares de nucleótidos desde el extremo 5' del sitio de unión de IL2-1.

Se obtuvieron fragmentos de ADN de aproximadamente 442 pares de bases, que contienen las secuencias descritas anteriormente, como productos de amplificación de plásmidos en los que los moldes se habían clonado. Los sitios diana de IL2-1 y 5-9D se localizaron dentro de estos fragmentos de forma que la escisión del ADN bicatenario entre los dos sitios diana generara fragmentos de ADN de aproximadamente 278 y 164 pares de bases. Los productos de amplificación se marcaron radiactivamente por transferencia de ortofosfato de γ -32P-ATP usando T4 polinucleótido cinasa.

Ejemplo 27: Escisión de ADN dirigida con fusiones de dedos de cinc/*FokI* que tienen polaridad alterada

40 Las proteínas de fusión de IL2-1C, IL2-1R y 5-9DR se obtuvieron incubando plásmidos que codifican estas proteínas en un lisado de reticulocitos acoplado a TNT (Promega, Madison, WI). Las reacciones de escisión se realizaron en 23 μ l de una mezcla que contenía 1 μ l de reacción de TNT para cada proteína de fusión, 1 μ l de sustrato de digestión marcado y 20 μ l de tampón de escisión. El tampón de escisión se preparó añadiendo 1 μ l de ditiotreitolo 1 M y 50 μ l de albúmina de suero bovino (10 mg/ml) a 1 ml de Tris-Cl 20 mM, pH 8,5, NaCl 75 mM, ZnCl₂ 10 μ M, 5 % de (v/v) glicerol. Las reacciones de escisión se incubaron a 37 °C durante 2 horas, luego se agitaron con 13 μ l de fenol/cloroformo/alcohol isoamílico (25:24:1). Después de la centrifugación, se analizaron 10 μ l de la fase acuosa sobre un gel al 10 % de poliacrilamida. La radiactividad en el gel se detectó usando un PhosphorImager (Molecular Dynamics) y se cuantificó usando el software ImageQuant (Molecular Dynamics).

50 La Figura 44 muestra los resultados obtenidos usando dos nucleasas químicas que tienen una polaridad de NH₂-dominio de *FokI*-dominio de dedos de cinc-COOH para escindir un sustrato en el que los sitios de unión para las dos nucleasas químicas se localizan sobre hebras opuestas y los extremos 3' de los sitios de unión son proximales entre sí y están separados por seis pares de nucleótidos. La incubación del sustrato con cualquiera de las nucleasas IL2-1R o 5-9DR sola no produce la escisión del sustrato (compárense los carriles 2 y 3 con el carril 1), mientras que la incubación de ambas nucleasas produce escisión casi completa del sustrato de ADN en el sitio diana previsto (carril 4).

60 La Figura 45 muestra la capacidad de una primera nucleasa química que tiene una polaridad de NH₂-dominio de dedos de cinc-COOH y una segunda nucleasa química que tiene una polaridad de NH₂-dominio de *FokI*-dominio de dedos de cinc-COOH para escindir un sustrato en el que los sitios de unión para las dos nucleasas químicas se localizan sobre la misma cadena, y el extremo 3' del primer sitio de unión es proximal al extremo 5' del segundo sitio de unión y separado de él por seis pares de nucleótidos. Solo la combinación de las nucleasas 5-9DR y IL2-1C (es decir, cada nucleasa que tiene una polaridad diferente) fue satisfactoria en escindir el sustrato que tiene ambos sitios diana sobre la misma hebra (compárense el carril 6 con los carriles 1-5).

65 **Ejemplo 28: Nucleasas químicas con diferentes longitudes de conector de ZC**

Se diseñaron dos conjuntos de proteínas de fusión con diferentes longitudes de conector de ZC, en el que el dominio de *FokI* está en el extremo amino con respecto al dominio de ZFP. El dominio de *FokI* tiene los aminoácidos 384-579 según Looney et al., (1989) Gene 80:193-208. El dominio de ZFP se seleccionó de los dominios de IL1-2 (Tabla 27), 5-8G (Tabla 17) y 5-9D (Tabla 17). El primer conjunto tuvo la estructura NH₂-NLS-*FokI*-ZFP-Flag-COOH. En este conjunto, se diseñaron proteínas que tienen longitudes de conector de ZC de 13, 14, 18, 19, 28 y 29 aminoácidos. El segundo conjunto tuvo la estructura NH₂-*FokI*-NLS-ZFP-Flag-COOH y se diseñaron proteínas con conectores de ZC de 21, 22, 23, 24, 28, 29, 38 y 39 aminoácidos. Obsérvese que, en el segundo conjunto, la NLS es parte del conector de ZC. También se construyen plásmidos que codifican estas proteínas de fusión.

Se diseñaron secuencias de ADN modelo para probar la actividad de escisión de estas proteínas de fusión y para determinar las longitudes de conector de ZC óptimas en función de la distancia entre los sitios diana para las dos proteínas de fusión. Se diseñaron las siguientes secuencias:

1. Sitio diana de 5-9D y sitio diana de IL2-1 sobre hebras opuestas
2. Sitio diana de 5-9D y sitio diana de IL2-1 sobre la misma hebra
3. Sitio diana de 5-9D y sitio diana de 5-8G sobre hebras opuestas
4. Sitio diana de 5-9D y sitio diana de 5-8G sobre la misma hebra

Para cada uno de estos cuatro pares de sitios diana, se construyen secuencias en las que la separación entre los dos sitios diana es 4, 5, 6 ó 7 pares de bases.

Estas secuencias se introducen en sustratos marcados como se describe en el Ejemplo 26 y se usan para probar las diversas proteínas de fusión descritas en este ejemplo para su capacidad para escindir ADN, según los métodos descritos en el Ejemplo 27.

Reivindicaciones

1. Un método para escindir cromatina celular en una región de interés, comprendiendo el método:
 - 5 (a) seleccionar la región de interés;
 - (b) manipular un primer dominio de unión de dedos de cinc para unirse a una primera secuencia de nucleótidos en la región de interés;
 - (c) proporcionar un segundo dominio de unión de dedos de cinc que se une a una segunda secuencia de nucleótidos en la región de interés, en el que la segunda secuencia se localiza entre 2 y 50 nucleótidos de la primera secuencia;
 - 10 (d) expresar una primera proteína de fusión en la célula, comprendiendo la primera proteína de fusión el primer dominio de unión de dedos de cinc 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 cinc y un segundo medio dominio de escisión;
 - 15 en el que
 - (i) la primera proteína de fusión se une a la primera secuencia de nucleótidos,
 - (ii) la segunda proteína de fusión se une a la segunda secuencia de nucleótidos,
 - 20 (iii) dicha unión de la primera y segunda proteínas de fusión posiciona los medios dominios de escisión de forma que la cromatina celular se escinda en la región de interés, y
 - iv) en al menos una de la primera o segunda proteínas de fusión, el medio dominio de escisión está más próximo al extremo N y el dominio de unión de dedos de cinc está más próximo al extremo C.
2. El método de la reivindicación 1, en el que la escisión se produce entre la primera y segunda secuencias de nucleótidos.
3. El método de la reivindicación 1, en el que el segundo dominio de unión de dedos de cinc se manipula para unirse a la segunda secuencia de nucleótidos.
4. El método de la reivindicación 1, en el que el primer y segundo medios dominios de escisión son de la misma endonucleasa.
5. El método de la reivindicación 4, en el que la endonucleasa es una endonucleasa de restricción de tipo IIS.
6. El método de la reivindicación 5, en el que la endonucleasa de restricción de tipo IIS es *Fok I*.
7. El método de la reivindicación 1, en el que la cromatina celular está en un cromosoma.
8. El método de la reivindicación 1, en el que el primer medio dominio de escisión es de una endonucleasa de restricción de tipo IIS.
9. El método de la reivindicación 1, en el que el segundo medio dominio de escisión es de una endonucleasa de restricción de tipo IIS.
10. El método de la reivindicación 1, en el que la primera y segunda secuencias de nucleótidos están sobre hebras opuestas de ADN.
11. El método de la reivindicación 10, en el que, en la primera y segunda proteínas de fusión, los medios dominios de escisión están más próximos a los extremos N y los dominios de unión de dedos de cinc están más próximos a los extremos C.
12. El método de la reivindicación 1, en el que la primera y segunda secuencias de nucleótidos están sobre la misma hebra de ADN.
13. El método de la reivindicación 12, en el que, en la primera proteína de fusión, el medio dominio de escisión está más próximo al extremo N y el dominio de unión de dedos de cinc está más próximo al extremo C y, en la segunda proteína de fusión, el dominio de unión de dedos de cinc está más próximo al extremo N y el medio dominio de escisión está más próximo al extremo C.
14. El método de la reivindicación 12, en el que, en la primera proteína de fusión, el dominio de unión de dedos de cinc está más próximo al extremo N y el medio dominio de escisión está más próximo al extremo C y, en la segunda proteína de fusión, el medio dominio de escisión está más próximo al extremo N y el dominio de unión de dedos de cinc está más próximo al extremo C.

FIGURA 1

MetGlyPheLeuLysLeuIle (SEC ID NO:2)
CTGCCGCCGGCGCCGGCCGTCATGGGGTTCCCTGAAACTGATT
GACGGCGGCCGGCGCCGGCAGTACCCCAAGGACTTTGACTAA (SEC ID NO:1)

5 10 15 20 25 30 35 40

FIGURA 2

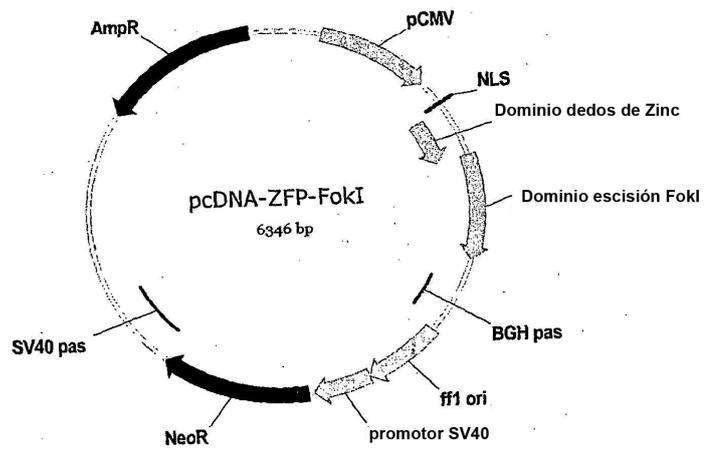


FIGURA 3

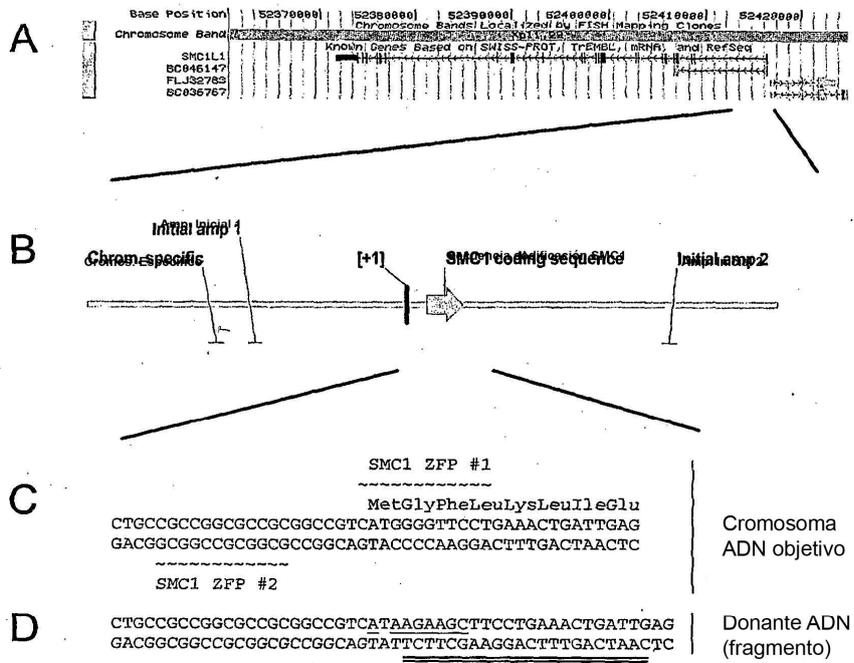


FIGURA 4

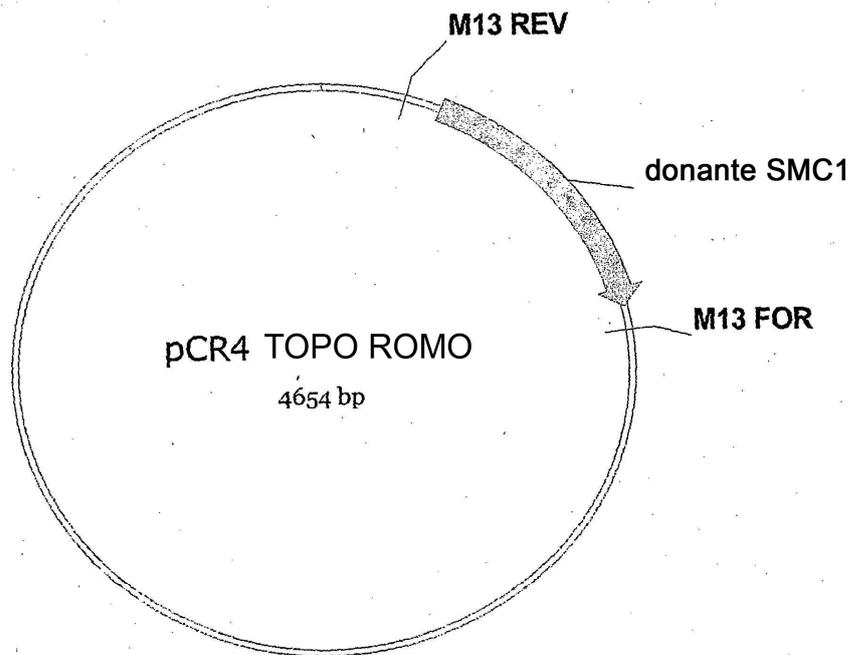


FIGURA 5

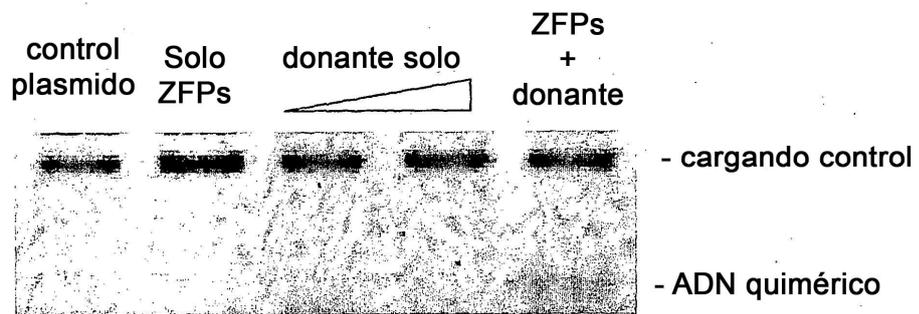


FIGURA 6

TAGTCCTGCAGGTTTAAACGAATTCGCCCTTCTCAGCAAGCGTGAGCTCA 50
GGTCTCCCCGCCTCCTTGAACTCAAGAACTGCTCTGACTCCGCCAGC 100
AACAACCTCTCCGGGATCTGGTCCGCAGGAGCAAGTGTGTTGTTGTTGCC 150
ATGCAACAAGAAAAGGGGGCGGAGGCACCACGCCAGTCGTCAGCTCGCTC 200
CTCGTATACGCAACATCAGTCCCCGCCCTGGTCCCCTCCTGCCGGAAG 250
GCGAAGATCCCGTTAGGCCTGGACGTATTCFCGCGACATTTGCCGGTCCG 300
CCGGCTTGCACTGCGGCGTTTCCCGCGGGGTACCTCAGTTCFCGGGCG 350
TACGGCGCGGCTGTCCTACTGCTGCCGGCGCCGCGGCCGTCATaagaag 400
cTTCCTGAACTGATTTGAAGGGCGAATTCGCCGCCGCTAAATTC AATTCG 450
CCCTATAGTGAGT

(SEC ID NO: 6)

FIGURA 7

TyrLysAsnSerAspAsnAspLysVal (SEC ID
NO:8)
CTTCCAACCTTTCTCCTCTAGGTACAAGAACTCGGATAATGATAAAGTCC
GAAGGTTGGAAAGAGGAGATCCATGTTCTTGAGCCTATTACTATTTTCAGG (SEC ID
NO:7)

5 10 15 20 25 30 35 40 45 50

FIGURA 8

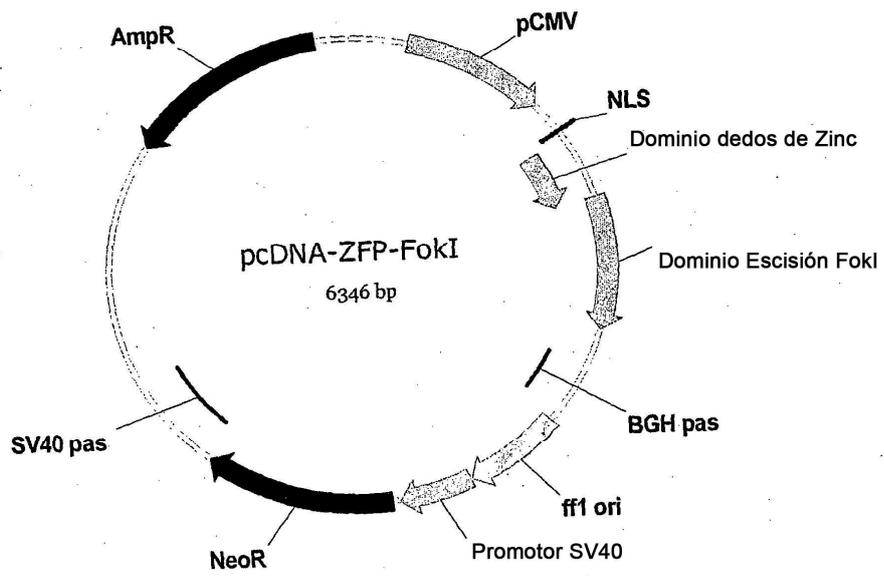


FIGURA 9

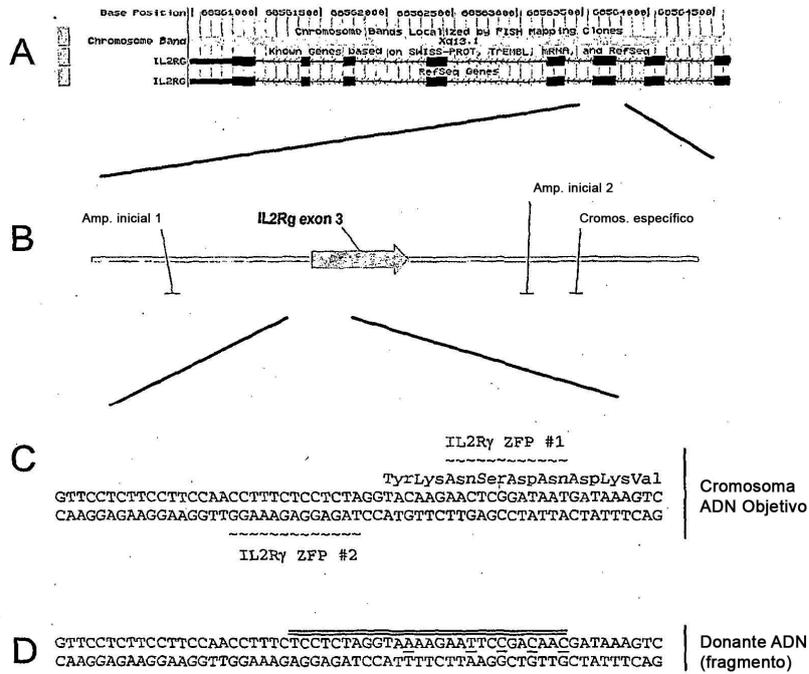


FIGURA 10

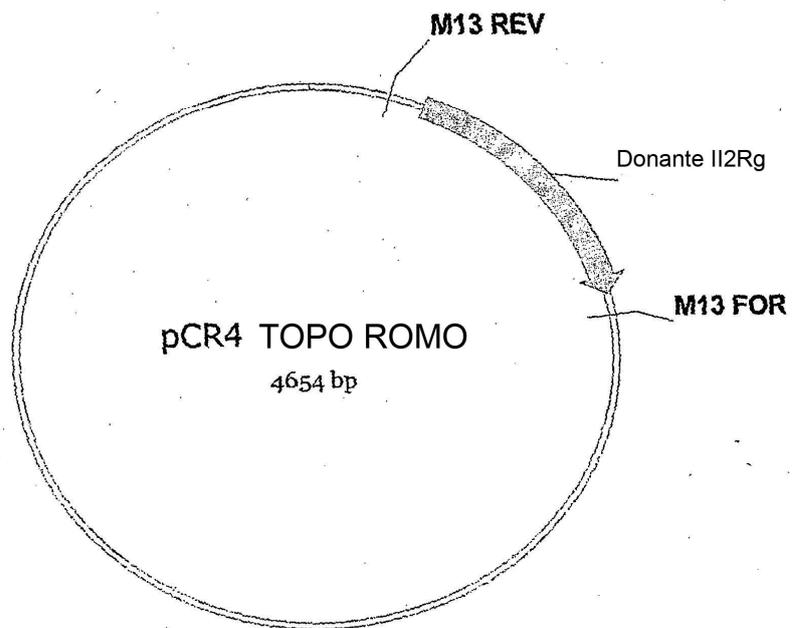


FIGURA 11

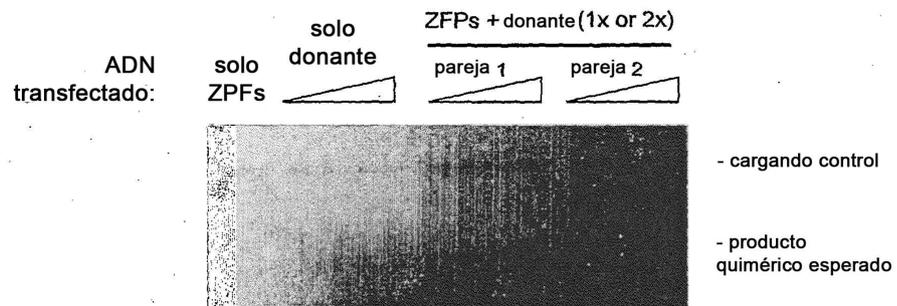


FIGURA 12

<u>TAGTCCTGCAGGTTTAAACGAATTCGCCCTTTCCTCTAGGTAaAAGAAAT</u>	50
<u>CcGAcAAcGATAAAAGTCCAGAAGTGCAGCCACTATCTATTC CCTGAAGAA</u>	100
ATCACTTCTGGCTGTCAGTTGCAAAAAAAGGAGATCCACCTCTACCAAAC	150
ATTTGTTGTTTCAGCTCCAGGACCCACGGGAACCCAGGAGACAGGCCACAC	200
AGATGCTAAAAGTGCAGAATCTGGGTAATTTGGAAAGAAAGGGTCAAGAG	250
ACCAGGGATACTGTGGGACATTGGAGTCTACAGAGTAGTGTTC TTTTATC	300
ATAAGGGTACATGGGCAGAAAAGAGGAGGTAGGGGATCATGATGGGAAGG	350
GAGGAGGTATTAGGGGCACTACCTTCAGGATCCTGACTTGTCTAGGCCAG	400
GGGAATGACCACATATGCACACATATCTCCAGTGATCCCC TGGGCTCCAG	450
AGAACCTAACACTTCACAAACTGAGTGAATCCCAGCTAGAACTGAACTGG	500
<u>AACAACAGATTCTTGAACCACTGTTTGGAGCACTTGGTGCAGTACCGGAC</u>	550
<u>TAAGGGCGAATTCGCGGCCGCTAAATTC AATTCGCCCTATAGTGAGTCGT</u>	600
<u>ATTACAATTCACTGGCCGTCGTTT</u>	

(SEC ID NO:12)

FIGURA 13

TACTGATGGTATGGGGCCAAGAGATATATCTTAGAGGGAGGGCTGAGGGT	50
TTGAAGTCCAACCTCCTAAGCCAGTGCCAGAAGAGCCAAGGACAGGTACGG	100
CTGTCATCACTTAGACCTCACCCCTGTGGAGCCACACCCCTAGGGTTGGCCA	150
ATCTACTCCCAGGAGCAGGGAGGGCAGGAGCCAGGGCTGGGCATAAAAGT	200
CAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAACCTGTGTTC	250
ACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGAGGAGAAG	300
TCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGG	350
TGAGGCCCTGGGCAGGTGGTATCAAGGTTACAAGACAGGTTTAAGGAGA	400
CCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTGAT	450
AGGCACTGACTCTCTGCCCTATTGGTCTATTTCCACCCCTTAGGCTGC	500
TGGTGGTCTACCCCTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTG	550
TCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAA	600
GAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCA	650
AGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTG	700

(SEC ID NO:13)

FIGURA 14

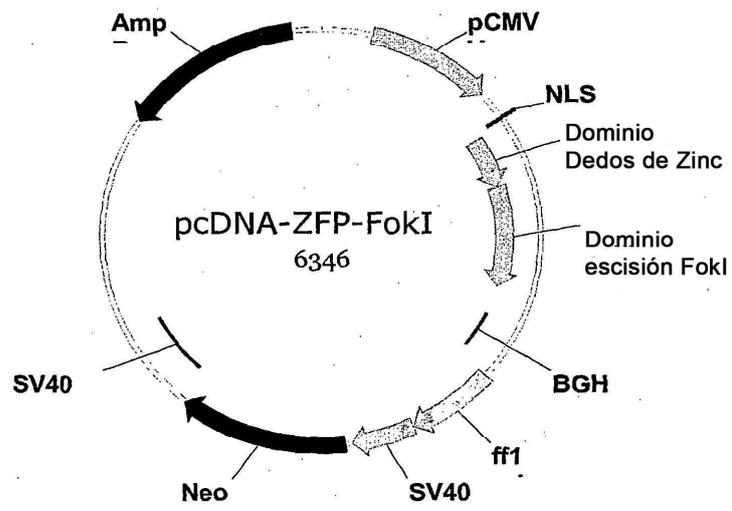
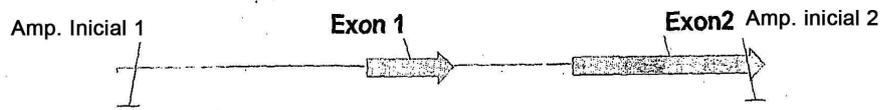


FIGURA 15



OBJETIVO genómico globina humana Beta
700 bp

FIGURA 16

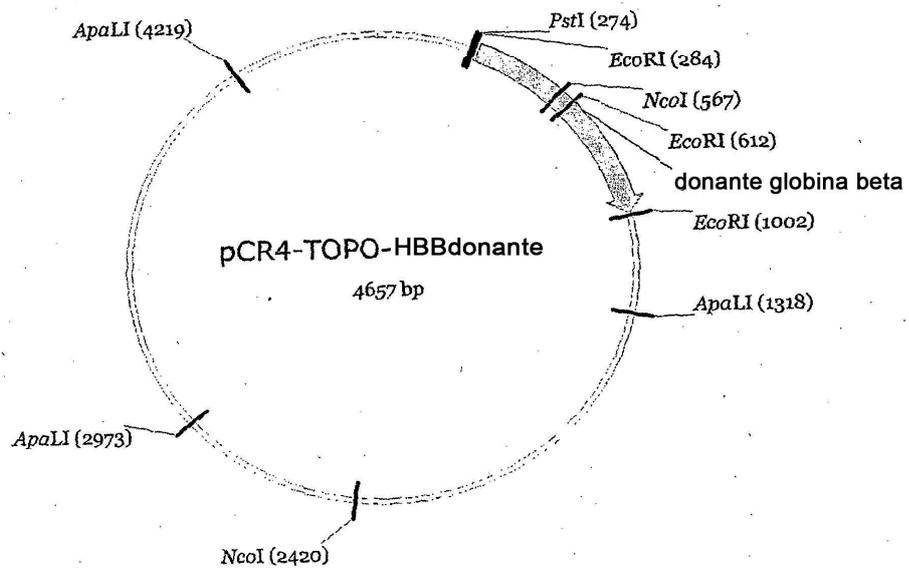


FIGURA 17

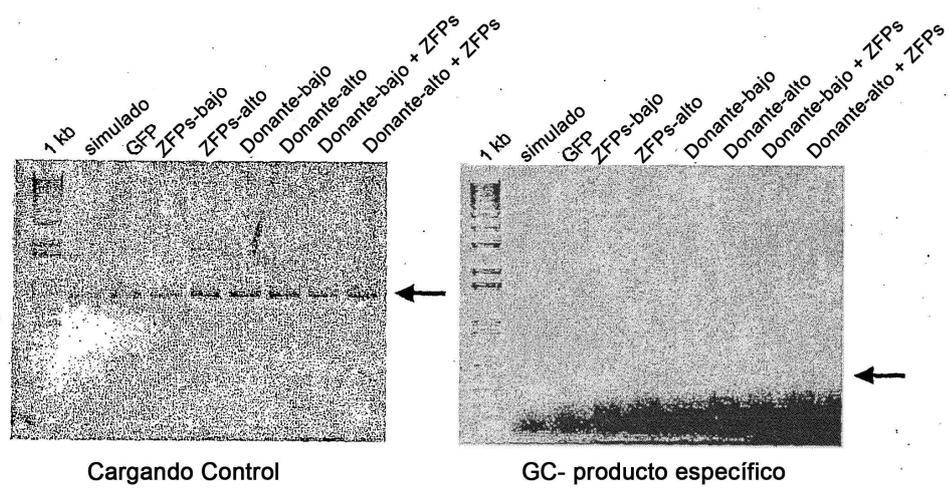


FIGURA 18

<u>TGCTTACCAAGCTGTGATTCCAAATATTACGTAAATACACTTGCAAAGGA</u>	50
<u>GGATGTTTTTAGTAGCAATTTGTACTGATGGTATGGGCCAAGAGATATA</u>	100
<u>TCTTAGAGGGAGGGCTGAGGGTTGAAGTCCAACCTCCTAAGCCAGTGCCA</u>	150
<u>GAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTCACCTGTGG</u>	200
<u>AGCCACACCCTAGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGG</u>	250
<u>AGCCAGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATT</u>	300
<u>TGCTTCTGACACAACGTGTTCCTAGCAACCTCAAACAGACACCATGGT</u>	350
<u>GCATCTGACTCCTGAGGAGAAGTCTG<u>CGTTAgTGCC</u><u>CgaattccgAtcG</u></u>	400
<u>TcAACcac</u>	

(SEC ID NO:14)

FIGURA 19

5-8
CACGTTTCGTGTTTCGGAGCCGCTTTAACCCACTCTGTGGAAG
GTGCAAAGCACAGCCTCGGCGAAATTGGGTGAGACACCTTC
5-10

(SEC ID NO:15)

FIGURA 20

1 MAPKKRKVQ IHGVPAAAE RPFQCRICMR NFSRSDNLSE HIRHTGKEP FACDICGRKF
61 ARNARRINHT KIHTGSQKPF QCRICMRNFS RSDTLSEHIR THTGKPFAC DICGRKFAAR
121 STRTHHTKIH LRQKDAARG QLVKSELEEK KSELRHKLKY VPHEYIELIE IARNSTQDRI
181 LEMKVMEFFM KVYGYRGKHL GGSRKPDGAI YTVGSPIDYG VIVDTKAYSG GYNLPIGQAD
241 EMQRYVEENQ TRNKHINPNE WWKVYPSSVT EPKFLFVSGH FKGNYKAQLT RLNHITNCNG
301 AVLSVEELLI GGEMIKAGTL TLEEVRRKFN NGEINF

(SEC ID NO:16)

FIGURA 21

1 MAPKKRKVG IHGVAAMAE RPFQCRICMR NFSRSDLSR HIRHTGKEP FACDICGRKF
61 ADSSNRKTH KIHTGGGSQ KPFQCRICMR NFSRSDLSV HIRHTGKEP FACDICGRKF
121 ADRSNRITH KIHLRQKDA RGSQLVKSEL EEKSELRHK LKYVPHEYIE LIEIARNSTQ
181 DRILEMKVME FFMKVYGYRG KHLGSRKPD GAIYTVGSPY DYGVIVDTKA YSGGYNLPFG
241 QADEMQRYVE ENQTRNKHIN FNEWVKVYPS SVTEPKFLFV SGHPKGNKYA QLTRLNHTTN
301 CNGAVLSVEE LLIGGEMIKA GTLTLEEVRR KFNNGEINF

(SEC ID NO:17)

FIGURA 22

CGAATTCTGCACTCGACGGTACCGCGGGCCCGGGATCCACCGGTGCGCCACCATGGTGAGC
AAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACC
ACCCTGACCTACGGCGTGCAGTGTTCAGCCGCTACCCCGACCACATGAAGCAGCACGAC
TTCTTCAAGTCCGCCATGCCC GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCC TGGTGAACCGC
ATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAG
TACAACTACAACAGCCACAACGTC TATATCATGGCCGACAAGCAGAAGAACGGCATCAAG
GTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTAC
CAGCAGAACACCCCATCGGGCAGCGCCCGTGCTGCTGCCC GACAACCCTACCTGAGC
ACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAG
TTCGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGC
GACTCTAGATCATAATC

(SEC ID NO:18)

FIGURA 23

CGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGTCGCCACCATGGTGAGC
AAGGGCGAGGAGCTGTTTACCGGGGTGGTGCCCATCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACC
ACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCTAACACGAAGCAGCACGACTT
CTTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGA
CGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCAT
CGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTA
CAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGT
GAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACCA
GCAGAACACCCCATCGGCGACGGCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTT
CGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGA
CTCTAGATCATAATC

(SEC ID NO:19)

FIGURA 24

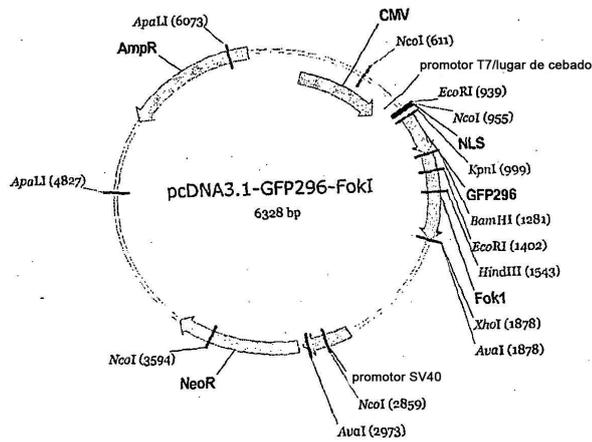
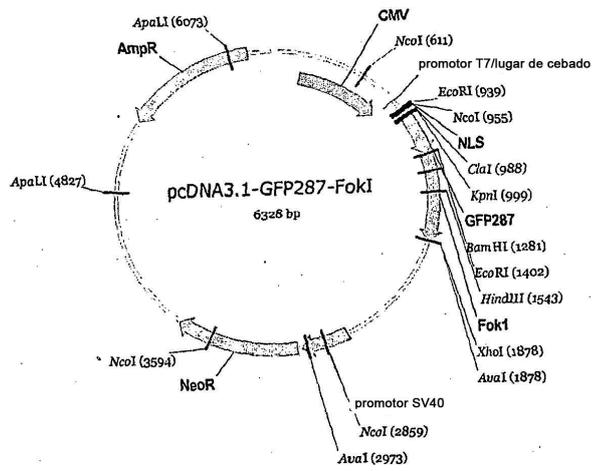


FIGURA 25

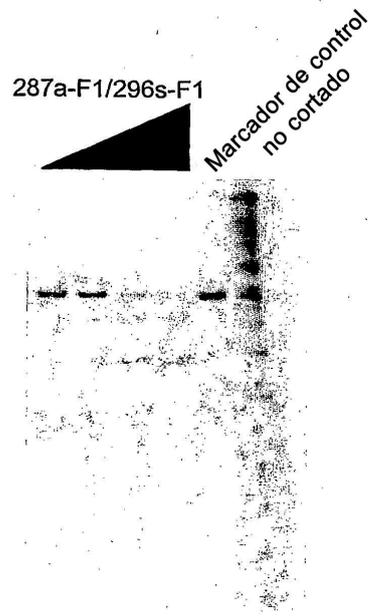


FIGURA 26

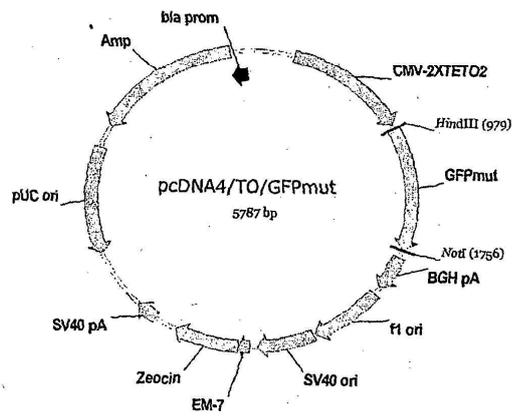


FIGURA 27

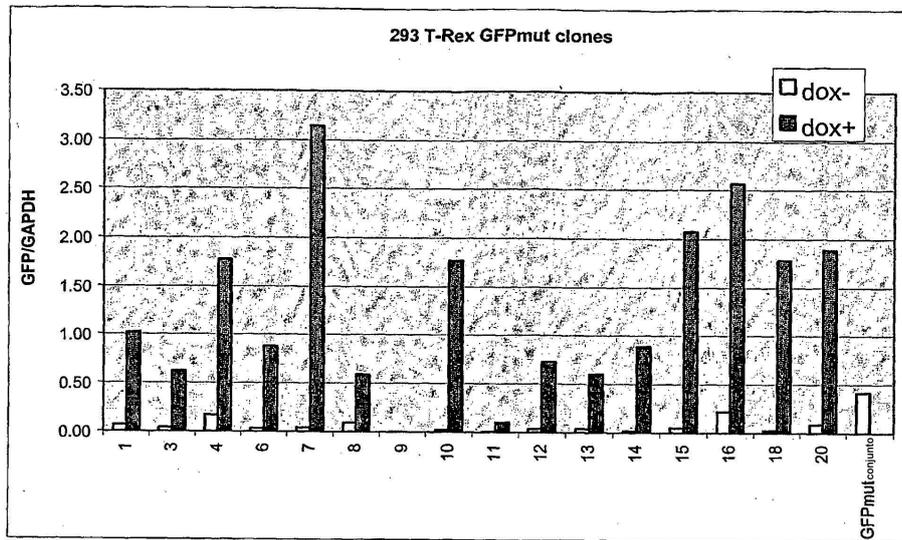


FIGURA 28

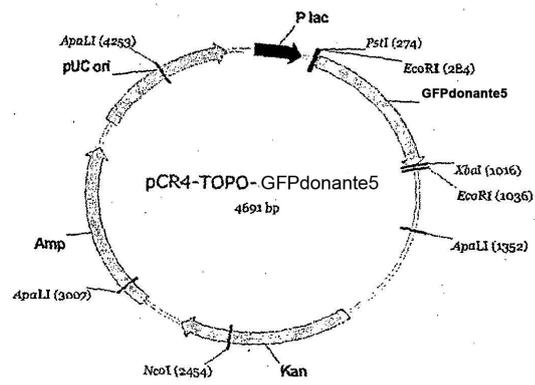


FIGURA 29

GGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAAC
GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACC
CTGAAGTTCATCTGCACCACCGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACC
CTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTC
TTC AAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGAC
GGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCCTGGTGAACCGCATC
GAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC
AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTG
AACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG
CAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACC
CAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTC
GTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGAC
TCTAGATCATAATC

(SEC ID NO:20)

FIGURA 30

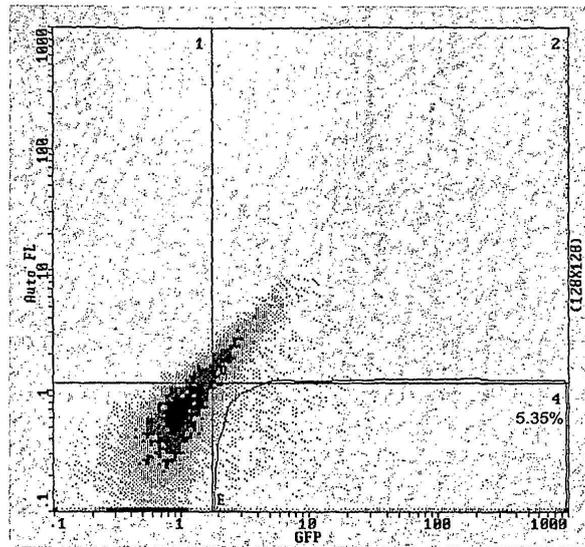


FIGURA 31

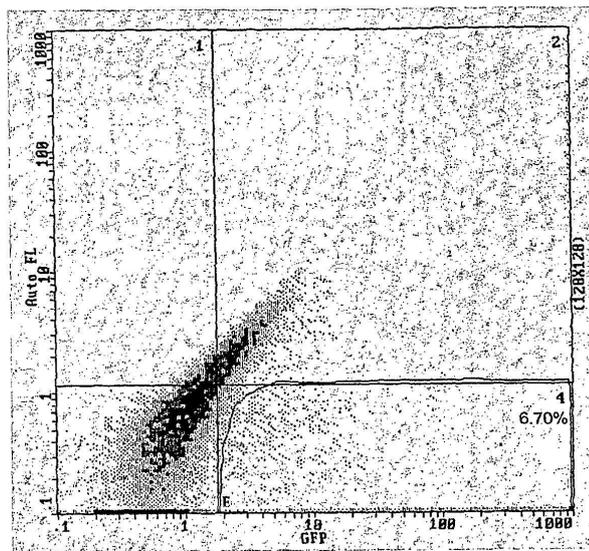


FIGURA 32

GGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACGTAAAC
GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACC
CTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACCACC
CTGACCTACGGCGTGCAGTGTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTTC
PTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGAC
GGCAACTACAAGACCCGCGCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATC
GAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC
AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTG
AACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACCAG
CAGAACACCCCATCGGCGACGGCCCGTGTGTGCTGCCCGACAACCACTACCTGAGCACC
CAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGTGGAGTTC
GTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCTCG
AGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGC
CATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTG
TCCTTTCCTAATAAAAATGAGGAAATTCATCGCATTGTCTGAGTAGGTGTCAATTCATTC
TGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG
CTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACAGCTGGGGCTCTAGGG
GGTATCCCCACGCGCCCTGTAGCGGCGCATTAAAGCGCGGGCGGGTGTGGTGGTTACGCGCA
GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCT
TTCTCGCCACGTTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGT
TCCGATTTAGTGCTTTACGGCACCTCGACCCAAAAAACTTGATTAGGGTGTGGTTTAC
GTAGTGGGCCATCGCCCTGATAGACGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCT
TTAATAGTGGACTCTTGTTCAAAACGGAACAACACTCAACCCTATCTCGGTCTATTCTT
TTGATTTATAAGGGATTTTGCCGATTTCCGCCCTATTGGTTAAAAAATGAGCTGATTTAAC
AAAAATTTAACGCGAATTAATCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCC
AGGCTCCCCAGCAGGCAGAAGTATGCA

(SEC ID NO:21)

FIGURA 33

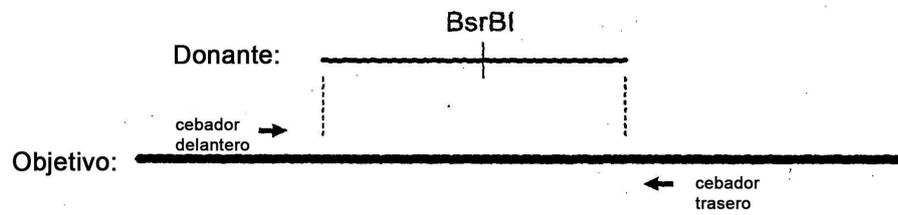


FIGURA 34

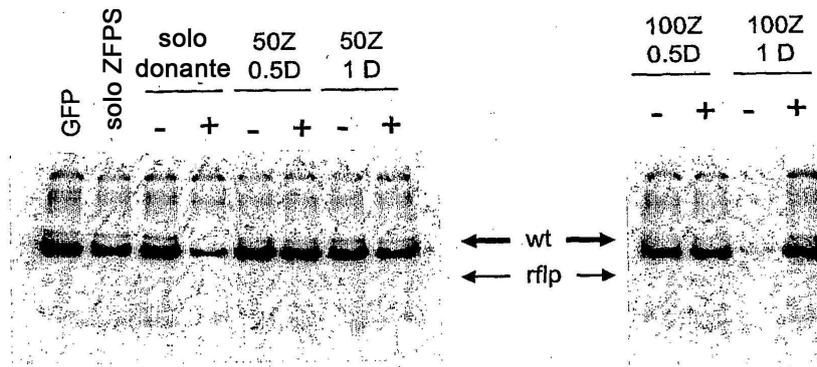


FIGURA 35

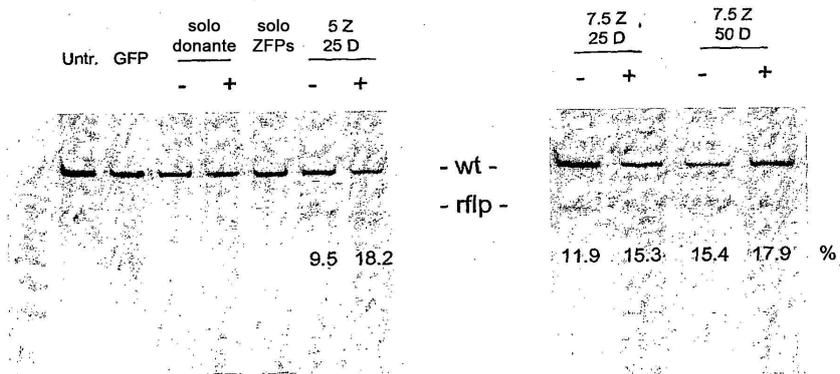


FIGURA 36

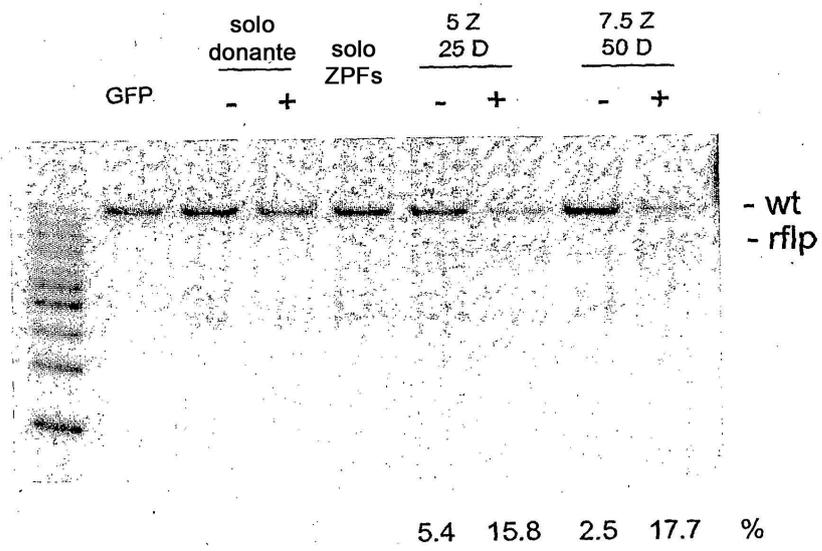


FIGURA 37

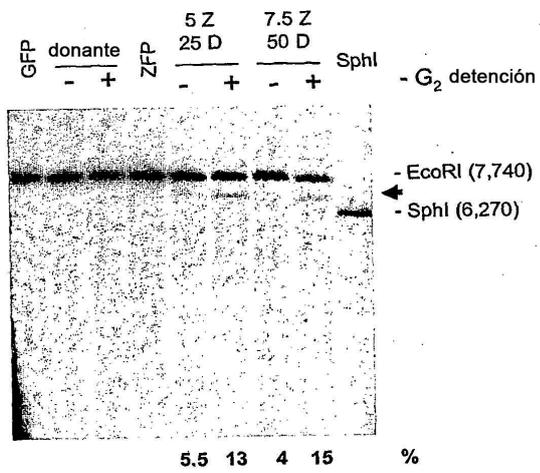
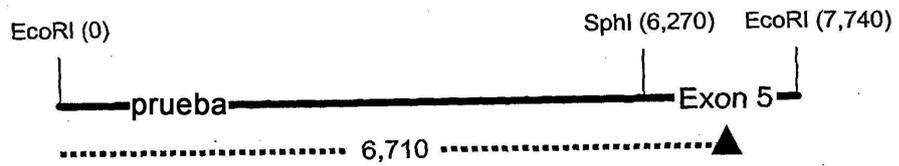


FIGURA 38

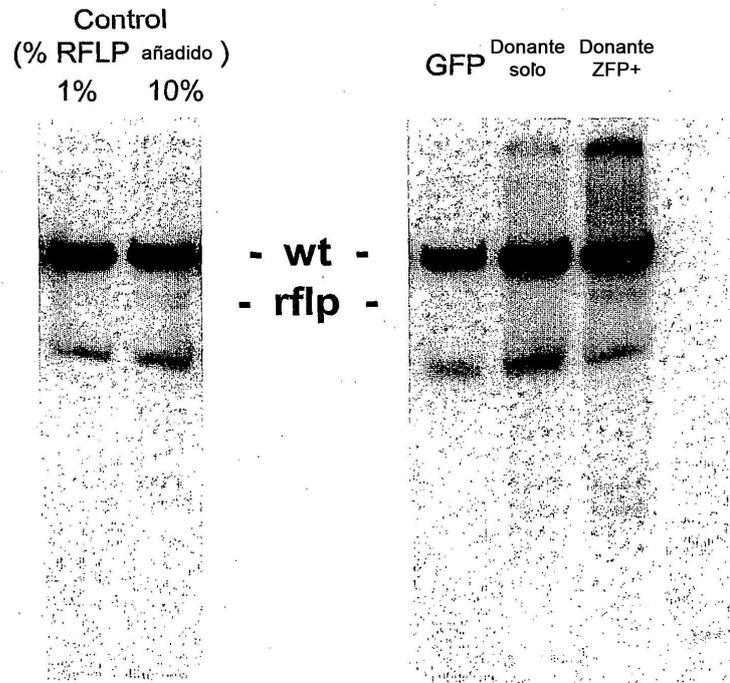


FIGURA 39

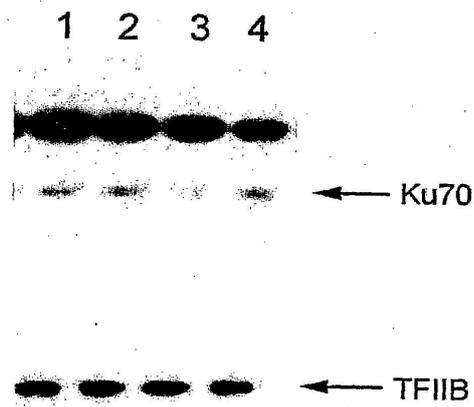


FIGURA 40

sca-29b:

MAERPFQCRICMRNFSQSGDLTRHIRTHTGEKPFACDICGRKFATSANLSRHTK
IHTGGGGSQKPFQCRICMRNFSDRSALSRHIRTHTGEKPFACDICGRKFAQSG
HLSRHTKIH (SEQ ID NO:22)

sca-36a:

MAERPFQCRICMRNFSRSQTRKTHIRTHTGEKPFACDICGRKFAQKRNRTKHT
KIHTGSQKPFQCRICMRNFSDRSALSRHIRTHTGEKPFACDICGRKFAQSGNLA
RHTKIH (SEQ ID NO:23)

sca-36b:

MAERPFQCRICMRNFSTSGSLSRHIRTHTGEKPFACDICGRKFADRSDLRHTK
IHTGGGGSQKPFQCRICMRNFSDRSALSRHIRTHTGEKPFACDICGRKFAQSG
NLARHTKIH (SEQ ID NO:24)

sca-36c:

MAERPFQCRICMRNFSTSSLSRHTHTGEKPFACDICGRKFADRSDLRHTK
IHTGGGGSQKPFQCRICMRNFSDRSALSRHIRTHTGEKPFACDICGRKFAQSG
NLARHTKIH (SEQ ID NO:25)

FIGURA 41

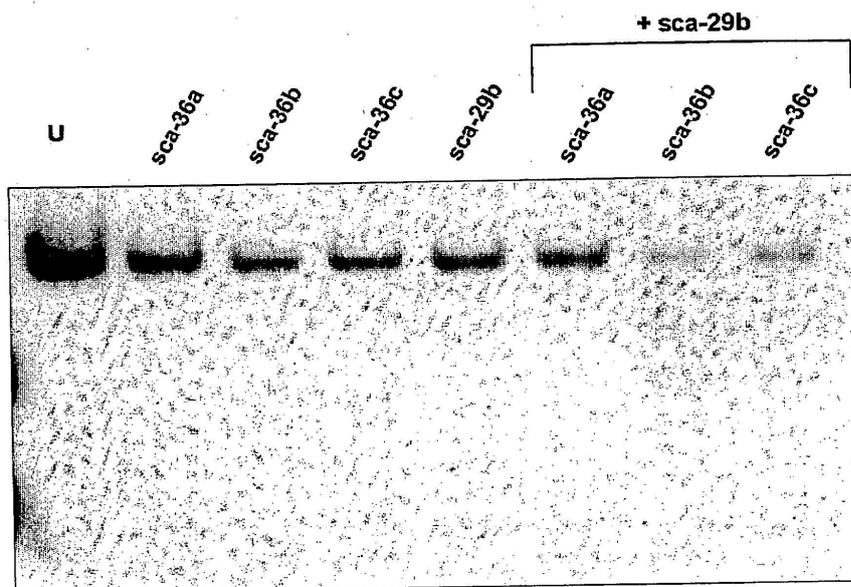


FIGURA 42

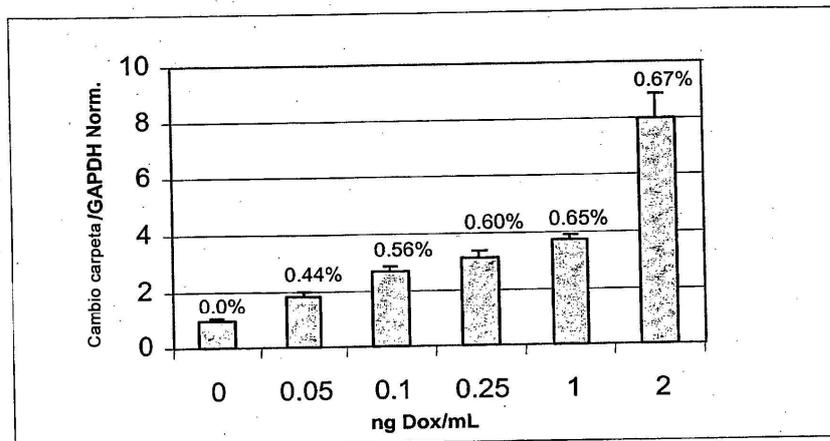


FIGURA 43

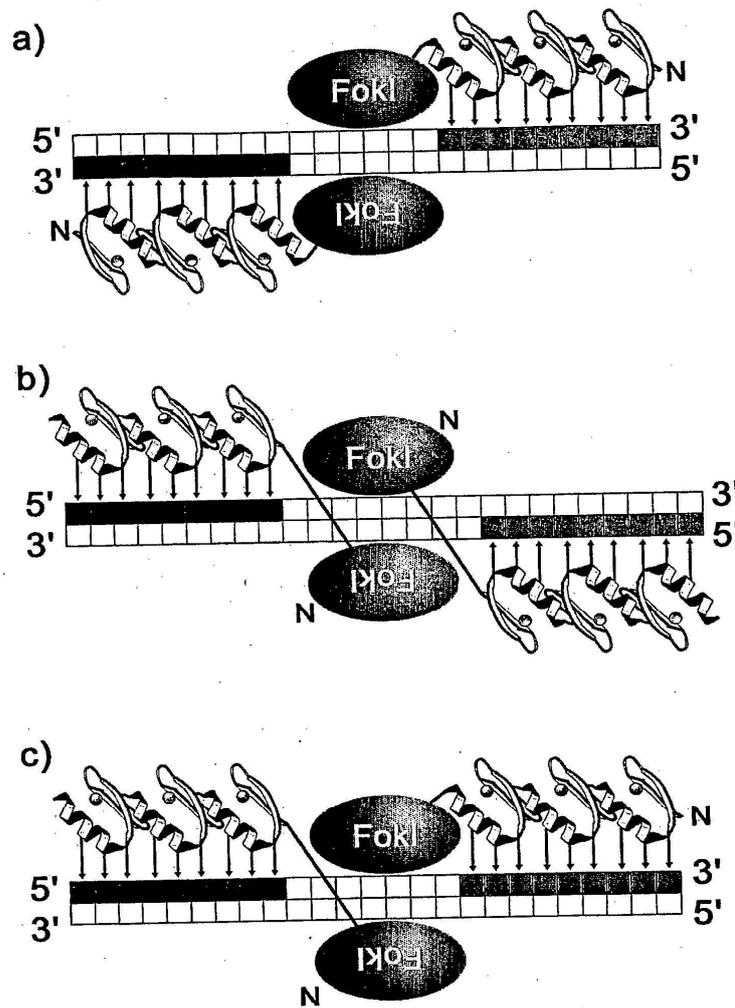
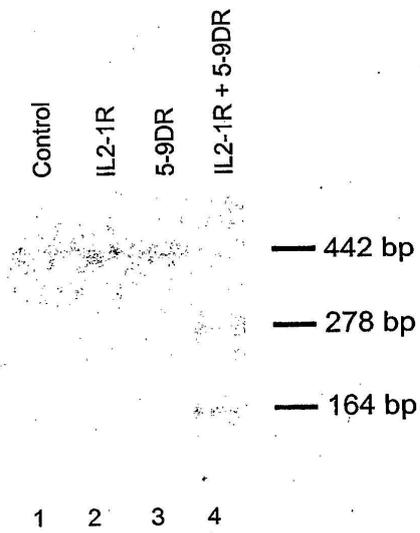


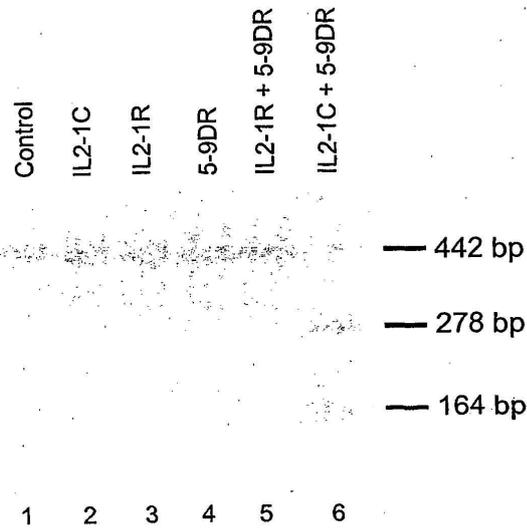
FIGURA 44



59D
-tAAAGCGGCTCCGaaccacATTATCCGAGTTc-
-aTTTCGCCGAGGcctggtgTAATAGGCTCAAg-
IL2-1

(SEC ID NO: 211)

FIGURA 45



59D IL2-1
-tAAAGCGGCTCCGaaccacAACTCGGATAATg-
-aTTTCGCCGAGGCTtggTgTTGAGCCTATTAc-

(SEC ID NO: 212)