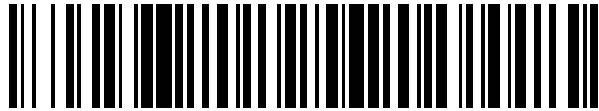


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 550 202**

51 Int. Cl.:

C12N 15/62 (2006.01)

C12N 5/10 (2006.01)

C12N 15/85 (2006.01)

A01K 67/027 (2006.01)

A61K 48/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **03.08.2010 E 10807022 (8)**

97 Fecha y número de publicación de la concesión europea: **15.07.2015 EP 2462230**

54 Título: **Métodos y composiciones para la modificación dirigida de genes**

30 Prioridad:

03.08.2009 US 230784 P

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

05.11.2015

73 Titular/es:

**RECOMBINETICS, INC. (100.0%)
1246 University Avenue West, Suite 301
Saint Paul, MN 55104, US**

72 Inventor/es:

**ESSNER, JEFFREY J.;
LIAO, HSIN-KAI y
FAHRENKRUG, SCOTT C.**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 550 202 T3

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

DESCRIPCIÓN

Métodos y composiciones para la modificación dirigida de genes

5 Declaración de los derechos de licencia del gobierno

El gobierno de Estados Unidos puede tener ciertos derechos sobre la invención a través de la financiación del Instituto Nacional de Sanidad a JJE para el desarrollo de modelos animales y materiales biológicos relacionados para la investigación R21, "Uso de RecA para potenciar la dirección de genes".

10

Campo técnico

El campo técnico se refiere a la modificación dirigida del genoma incluyendo, pero sin limitación, la inserción dirigida, la eliminación dirigida, la inactivación dirigida de genes y la mutagénesis dirigida.

15

Antecedentes

Un importante campo de interés en biología y medicina es la modificación dirigida de secuencias de nucleótidos genómicas. Dichas modificaciones incluyen la inserción, la eliminación y la sustitución de secuencias de ácido nucleico cromosómicas endógenas.

20

Sumario

Se desvelan, y también se ilustran con ejemplos de trabajo, composiciones y métodos que proporcionarán la modificación dirigida de secuencias genómicas. Se describen ciertas proteínas de fusión que administran de forma segura y eficaz secuencias exógenas al sitio diana para lograr el efecto deseado.

25

En el pasado, otros investigadores han intentado modificar secuencias genómicas en células cultivadas aprovechando el fenómeno natural de la recombinación homóloga. Véase, por ejemplo, Capecchi (1989) *Science* 244: 1288-1292; patentes de EE.UU. N° 6.528.313 y 6.528.314. Si un polinucleótido exógeno tiene suficiente homología con la región genómica que contiene la secuencia que se va a modificar, es posible que parte o toda la secuencia del polinucleótido exógeno reemplace la secuencia genómica por recombinación homóloga. Sin embargo, la frecuencia de recombinación homóloga en estas circunstancias es extremadamente baja. Por otra parte, la frecuencia de inserción del polinucleótido exógeno en las ubicaciones genómicas que carecen de homología de secuencia supera la frecuencia de recombinación homóloga en varios órdenes de magnitud.

30

35

Por lo tanto, los anteriores intentos por reemplazar determinadas secuencias han implicado poner en contacto una célula *ex vivo* con un polinucleótido exógeno (también denominado ADN donante) que comprende secuencias que portan homología con una región cromosómica diana), seguido de la selección de las células *ex vivo* en las que la molécula de ADN donante se había sometido a recombinación homóloga en el genoma. La tasa de éxito de estos métodos es baja, debido a la baja eficacia de la recombinación homóloga y a una alta frecuencia de inserción inespecífica del ADN donante en regiones del genoma distintas del sitio diana.

40

Debido a estos problemas conocidos con la eficacia y la especificidad de los métodos existentes para la recombinación dirigida, sigue existiendo la necesidad de métodos específicos, de alta eficacia, y de composiciones para la dirección de genes. Además de hacer la dirección de genes más accesible y práctica, dichos métodos y composiciones mejorados también reducirían los efectos secundarios resultantes de las inserciones no dirigidas. Véase, por ejemplo, Hacia-Bey-Abina *et al.* (2003), *Science* 302: 415-419.

45

La proteína RecA es el prototipo de una familia de proteínas procariotas y eucariotas que catalizan la recombinación genética (es decir, el intercambio de información de la secuencia de ADN entre dos moléculas de ADN). RecA y sus homólogos participan en la reparación de las roturas del ADN bicatenario mediante la catalización de la sinapsis de una molécula de ADN monocatenaria con secuencias homólogas en un ADN bicatenario para formar una molécula de heterodúplex. La migración de la ramificación en el heterodúplex puede producir la transferencia de información de la secuencia del ADN monocatenario a la molécula bicatenaria, como ocurre en los procesos de recombinación y conversión de genes.

50

55

Las actividades notables y diversas de RecA han llevado a los investigadores a investigar el uso de esta proteína, y sus homólogos, para estimular la recombinación homóloga y la orientación de genes en eucariotas. En el tabaco, la expresión de RecA bacteriana que contiene una señal de localización nuclear (NLS-RecA) aumentó la resistencia a la reticulación del ADN inducida por la mitomicina C, y también aumentó la recombinación intracromosómica somática (recombinación entre cromosomas homólogos) en diez veces. Reiss *et al.* (1996) *Proc. Natl. Acad. Sci. EE.UU.* 93: 3094-3098. En un estudio realizado por separado en el tabaco, se encontró que la expresión de NLS-RecA estimula el intercambio de cromátidas hermanas en 2,4 veces frente a los niveles de tipo silvestre. Reiss *et al.* (2000) *Proc. Natl. Acad. Sci. EE.UU.* 97: 3358-3363.

60

65

En las células de mamífero, se informó que la sobreexpresión de NLS-RecA estimula la dirección de genes a través de la recombinación homóloga en 10 veces. Shcherbakova *et al.* (2000) *Mutation Res.* 459:65-71. En las células humanas, la sobreexpresión del homólogo de RecA humano RAD51 fue capaz de estimular la recombinación solo de 2 a 3 veces frente a los niveles de tipo silvestre. Yanez *et al.* (1999) *Gene Ther.* 6:1282-1290. Otro estudio demostró que la inyección directa de filamentos preformados de nucleoproteína recubiertos con RecA en embriones de pez cebra podría corregir una forma mutante de la proteína verde fluorescente mejorada (eGFP), aunque a una frecuencia baja. Cui *et al.* (2003) *Marine Biotechnology.* 5:174-184. En experimentos similares de inyección realizados en embriones de pez cebra, otro grupo mostró que Rad52, un miembro del grupo de epistasias de Rad51, podría potenciar la hibridación de cadenas sencillas y la modificación de genes mediada por oligonucleótidos de bajo nivel. Takahashi *et al.* (2005) *Nucleic Acids Res.* 33:e120. Otras publicaciones relacionadas con RecA incluyen, por ejemplo, la patente de EE.UU. N° 7.229.767. Una divulgación reciente describió el uso de ataduras moleculares para la inserción dirigida de vectores de transposones, en el que la atadura comprende un dominio de unión al ADN que se une a un sitio diana en el vector, véase la publicación de patente de EE.UU. N° 2007/0031380. Otras publicaciones relacionadas con los transposones incluyen, por ejemplo, las patentes de EE.UU. N° 6.498.458, 7.160.682 y 7.527.966.

Las herramientas convencionales usadas para realizar la genética inversa y crear la modificación dirigida de genes específicos se limitan a unas cuantas especies, y requieren tecnologías sofisticadas y muy laboriosas que, por lo general, implican la clonación o la obtención mediante ingeniería genética de células madre embrionarias. Para hacer frente a estas limitaciones, se desarrollaron tecnologías innovadoras que se pueden usar para modificar regiones cromosómicas específicas mediante la inyección directa de complejos de proteína-ácido nucleico en cigotos fertilizados. Se describe una versión modificada de la proteína RecA bacteriana que es capaz de potenciar la recombinación homóloga o no homóloga, y la inserción de ADN exógeno en ubicaciones genómicas específicas para la modificación de genes. Esta versión modificada de RecA funciona a frecuencias de varios órdenes de magnitud superiores a las de los informes anteriores. Se trata de una forma muy activa de RecA que funciona en vertebrados, y que se espera que funcione en los animales y en las plantas en general. La modificación y el uso de RecA es un resultado inesperado y sorprendente que no se esperaba que funcionara así, es decir, que potenciara la recombinación homóloga.

Breve descripción de las figuras

La Figura 1 muestra la secuencia de aminoácidos de una proteína de fusión NLS/RecA/Gal4 (SEC ID N° 3).

La Figura 2 muestra la secuencia de nucleótidos que codifica una proteína de fusión NLS/RecA/Gal4 (SEC ID N° 4).

La Figura 3 muestra la secuencia de aminoácidos de una proteína de fusión NLS/RecA (SEC ID N° 5).

La Figura 4 muestra la secuencia de nucleótidos que codifica una proteína de fusión NLS/RecA (SEC ID N° 6).

La Figura 5 es una ilustración de tres tipos de proteínas de fusión RecA.

La Figura 6 representa los resultados que muestran que la inyección de filamento de ADNmc-NLS-RecA-Gal4 complementario conduce a la inserción específica del sitio. Como se muestra, este filamento provoca la pérdida de heterocigosidad (LOH) en embriones *gol* heterocigotos, generando la pigmentación ocular en mosaico. Panel A: Genotipo de los embriones usados para la inyección. Todos los embriones se inyectaron en la fase de una sola célula. Panel B: Vistas dorsales de los ojos en 3 dpf que muestran patrones de pigmentación de tipo silvestre en un embrión no inyectado (panel superior) y patrones mutantes (panel inferior) por los filamentos dirigidos inyectados de oligonucleótidos *gol*-NLS-RecA-Gal4. La inyección del filamento dirigido provoca la pérdida de heterocigosidad en el locus de *gol*. Panel C: Cuatro sondas de dirección de *gol* recubiertas por NLS-RecA-Gal4 diseñadas para su dirección bien a la región exón4/5 o exón6. Las sondas *gsg1* y *gsg2* portan la mutación del codón de terminación dentro de las sondas de ADN que se muestran con el círculo relleno. Las sondas *gbg1* y *gbg2* son de 60 nucleótidos de longitud y están sintetizadas a secuencias de ADN genómicas adyacentes. Estas dos sondas no contienen mutaciones.

La Figura 7 representa los resultados que muestran, para tres genes diferentes, que la expresión génica es un resultado de la integración específica del sitio. Como se muestra, la expresión de un gen indicador EGFP coincide con la integración específica del sitio en los locus de *gol*, *flh* y *prominin-1*. Los filamentos de NLS-RecA-Gal4 monocatenarios complementarios a los locus de (1) *gol*, (2) *flh* y (3) *prominin-1* se inyectaron conjuntamente con el casete del gen indicador EGFP. La expresión de EGFP coincidente con la expresión génica dirigida se observó en del 5 al 19 % de los embriones inyectados. Para el gen *gol*, se observó la expresión en el ojo, para el gen *flh*, en la notocorda, y para el gen *prominin-1*, en el diencéfalo dorsal.

La Figura 8 representa la inserción específica del sitio de ADN exógeno. Como se ilustra, los filamentos de ADNmc-NLS-RecA-Gal4 dirigieron la inserción específica del sitio de ADN de GFP. Panel A: Se amplificaron, desnaturalizaron y recubrieron con la proteína NLS-RecA-Gal4 dos regiones del gen *gol* para crear filamentos de ADNmc-NLS-RecA-Gal4. Se inyectaron los filamentos con ADN foráneo que contenía un aceptor de corte y

empalme (SA), seguido de la proteína verde fluorescente (GFP) en tres marcos de lectura y una señal de poliadenilación (pA). Panel B: Se obtuvo la amplificación por PCR de los fragmentos de unión entre el ADN foráneo y el locus de *gol* endógeno a partir de ADN aislado de embriones individuales. Las bandas marcadas con una estrella se verificaron secuencialmente como fragmentos de unión. Panel C: Mapa de fragmentos de unión de inserciones en el locus de *gol* que muestra la inserción de ADN exógeno cerca de los extremos de las regiones complementarias a los fragmentos de unión.

La Figura 9 representa la mutación diana para la creación de un animal transgénico. El locus de *gol* situado en la línea germinal de pez cebra se usó como la diana y se modificó. Un cruce de ensayo entre un fundador dirigido a *gol* y homocigotos de *gol*^{b1} produjo descendencia que no logró complementar el alelo *b1* (derecha) y su hermano con pigmento normal (izquierda).

La Figura 10 representa un modelo para la dirección de genes mediante ADNmc-NLSRecAGal4. Se inyectan conjuntamente filamentos de una sola cadena tanto directa como inversa de NLSRecAGal4 en embriones de pez cebra. La actividad de búsqueda de homología de RecA guía a los filamentos a la región diana. Los filamentos de ADNmc-NLSRecAGal4 sufren el emparejamiento homólogo y la invasión de la cadena, lo que provoca la formación de bucles D en el cromosoma diana. La estructura es estabilizada por los dominios de dimerización de Gal4 entre los filamentos complementarios. Existe la teoría de que dicha molécula unida de ADN bloquea la progresión de la horquilla de replicación, conduciendo a la rotura de la doble cadena (DSB).

Descripción detallada

Se observó la dirección de ácidos nucleicos recubiertos con ciertas proteínas de fusión a sitios diana específicos. Se incorporaron secuencias de ácido nucleico exógeno bicatenario en los sitios diana de la célula hospedadora, y fueron expresadas por las células. Los animales así transfectados expresaron de manera continua el gen y pasaron las alteraciones genéticas a la descendencia. Este método es muy sencillo y de gran alcance en comparación con las tecnologías convencionales, que se limitan a unas cuantas especies y requieren técnicas sofisticadas y muy laboriosas.

Se descubrió que la actividad de apareamiento de los filamentos de RecA-ADN se podría utilizar para dirigir las actividades bioquímicas a sitios cromosómicos específicos. Se ensayaron varios filamentos con proteínas RecA quiméricas con diversos genes. El modelo de animal vertebrado fue el de pez cebra. En este modelo, se demuestra la interrupción específica del sitio de un gen mediante la inducción de la pérdida de heterocigosidad (LOH) en el locus de *golden* del pez cebra tras la inyección en la fase de 1 célula. La LOH se observa por una pigmentación en mosaico, y se verificó mediante el análisis de ADN directo. Los resultados presentados en el presente documento demuestran que los filamentos de ADN, de diferentes tamaños, recubiertos con las proteínas de fusión son capaces de causar mutaciones dirigidas al sitio y eliminaciones cromosómicas diana en el pez cebra que se transmiten a las generaciones posteriores. Además, la inyección conjunta de ácidos nucleicos exógenos con la proteína de fusión potencia la inserción de los ácidos nucleicos exógenos en ubicaciones genómicas diana, probablemente a través de la vía de unión de extremos no homólogos.

Sin limitarse a una teoría en particular, se presenta un modelo mediante el cual el dominio de Gal4 de NLSRecAGal4 potencia la dimerización de filamentos monocatenarios complementarios (mcc) después de que los filamentos hayan encontrado su diana. El modelo proporciona un complejo que crea un bloque estérico a la replicación, dando lugar a la paralización de una horquilla de replicación y, bien a la reparación del locus o la rotura cromosómica.

Los datos del presente documento arrojan pruebas de que es posible dirigir específicamente moléculas de fusión proteicas a un sitio diana en un cromosoma del huésped y crear una rotura. Esta rotura se puede aprovechar para crear mutantes, descubrir la función de genes, insertar genes exógenos y otros fines. Los Ejemplos 1 y 2 describen detalladamente dicha molécula de fusión, específicamente, una molécula de dominio NLS-dominio RecA-dominio Gal4 o dominio NLS-dominio RecA, y construcciones de ADN para la creación la misma. Estas moléculas de fusión conservan la función de formación de filamentos nucleoproteicos (Ejemplo 3). En el Ejemplo 4, se usó el modelo de pez cebra para demostrar la interrupción de sitios de genes específicos. Se desnaturalizaron sondas de ADN bicatenario complementarias a diversos sitios en el locus de *gol* en cadenas sencillas, y se formaron filamentos con una molécula de proteína de fusión NLS/RecA/Gal4. Se usaron satisfactoriamente sondas tanto de 1.300 pares de bases (pb) como de 60 pb dirigidas a distintos sitios de *gol*. Las sondas no se integraron en el ADN de la célula hospedadora.

Otros datos mostraron que el ADN exógeno se pudo integrar luego en sitios específicos y que el método fue aplicable, en general, a genes y que no se limitó a *gol*. En el Ejemplo 5, se demostraron inserciones dirigidas mediante la expresión específica del tejido del gen de la proteína verde fluorescente mejorado (EGFP) tras la inyección conjunta con filamentos de ADNmc-NLS-RecA-Gal4 complementarios a los locus de *gol*, *prominin-1* y *floating head (flh)* (Figura 7). Se observó la expresión en ausencia de un promotor exógeno, es decir, el sitio de inserción se eligió de manera que se pudieran aprovechar los promotores nativos y la maquinaria celular. El análisis de los sitios de inserción mostró que estaban en el gen diana, y en aproximadamente 500 pb del sitio diana de la sonda. Por último, se demostró la transfección de la línea germinal y la progenie que expresaban los genes

exógenos (Ejemplo 6). Los mecanismos biológicos se detallan a continuación.

Se realizaron otros experimentos con una construcción de sustitución de *gol-mcherry-gol*, que era capaz de dirigirse al locus genómico de *gol*. Los exones del gen *gol* se designan E1 a E9. La construcción de dirección contenía el gen *mcherry* (M). La inyección de esta construcción de sustitución con NLSRecAGal4 produjo sectores rojos fluorescentes en el ojo. Además, se recuperó un fragmento de unión después de la amplificación por PCR anidada, lo que indicó de la recombinación homóloga entre el vector de sustitución de *gol* y el gen *gol* endógeno.

Así pues, las realizaciones de la invención incluyen una molécula de fusión según lo definido en las reivindicaciones. La molécula de fusión puede incluir una señal de localización nuclear o transportarse de otro modo en la célula y en el núcleo. Los sistemas pueden incluir sondas y ADN exógeno para la inserción en una célula hospedadora. Estas características se detallan en el presente documento.

La práctica de la presente divulgación emplea, a menos que se indique lo contrario, los métodos y las técnicas convencionales en los campos de la biología celular, biología evolutiva, biología reproductiva, biología molecular, bioquímica, cultivo celular, ADN recombinante y campos relacionados como son los pertenecientes al alcance de la técnica. Dichas técnicas se describen en la bibliografía y, por lo tanto, se encuentran disponibles para los expertos en la materia. Véase, por ejemplo, Alberts, B. *et al.*, "Molecular Biology of the Cell", 5ª edición, Garland Science, New York, NY, 2008; Voet, D. *et al.* "Fundamentals of Biochemistry: Life at the Molecular Level", 3ª Edición, John Wiley & Sons, Hoboken, NJ, 2008; Sambrook, J. *et al.*, "Molecular Cloning: A Laboratory Manual", 3ª Edición, Cold Spring Harbor Laboratory Press, 2001; Ausubel, F. *et al.*, "Current Protocols in Molecular Biology", John Wiley & Sons, New York, 1987 y actualizaciones periódicas; Freshney, R. I., "Culture of Animal Cells: A Manual of Basic Technique", 4ª Edición, John Wiley & Sons, Somerset, NJ, 2000; y la serie "Methods in Enzymology", Academic Press, San Diego, CA.

Moléculas de fusión

Los métodos y las composiciones para la modificación dirigida del genoma que se desvelan en el presente documento implican, en ciertas realizaciones, el uso de moléculas de fusión. Para los fines de la presente divulgación, una molécula de fusión es una molécula artificial que contiene al menos dos dominios unidos entre sí dentro de una sola molécula, de manera que los dos dominios no se encuentran juntos en una molécula de origen natural. Los dominios pueden ser de origen natural o sintético. Los dominios pueden ser del mismo tipo químico de molécula, o pueden ser de diferentes tipos químicos de moléculas. La expresión "molécula de fusión proteica" se refiere a una molécula de fusión que tiene al menos dos dominios polipeptídicos. Dado que tiene dominios polipeptídicos, es "proteica", y puede comprender, además, características no proteicas, por ejemplo, enlazadores poliméricos. La expresión "dominio polipeptídico" se refiere a: un conjunto de péptidos unidos entre sí que, en conjunto y de forma independiente, cumplen una función biológica. Los ejemplos de dominios que satisfacen esta definición son los dedos de cinc, el dominio de mano EF de unión a calcio de la calmodulina, secuencias de péptidos que muestran unión específica a una diana predeterminada, NLS, secuencias de unión al ADN y partes de proteínas que realizan la función de la proteína de tipo silvestre (por ejemplo, un derivado de RecA). Por lo tanto, los dominios polipeptídicos, por ejemplo, se pueden mezclar y aparear mediante ingeniería genética entre una proteína y otra para crear proteínas quiméricas.

En ciertas realizaciones, una molécula de fusión proteica incluye al menos dos dominios seleccionados del grupo que consiste en un primer dominio que es un dominio de unión a ADN; por ejemplo, una proteína de unión a ADN o un fragmento funcional de una proteína de unión a ADN, un segundo dominio que comprende una secuencia polipeptídica que tiene actividad recombinasa y un tercer dominio que comprende una señal de localización nuclear. Estos dominios pueden estar en cualquier orden, por ejemplo, NLS-recombinasa-unión a ADN o unión a ADN-NLS-recombinasa. Los dominios pueden estar separados por enlazadores que sean peptídicos o estén hechos de otros materiales.

En ciertas realizaciones, cada uno de los dominios de molécula de fusión corresponde a una secuencia polipeptídica distinta; por ejemplo, un dominio de unión a ADN polipeptídico tal como Gal4 y secuencias polipeptídicas de RecA que tienen actividad recombinasa. Sin embargo, también es posible que las moléculas de fusión posean dominios no polipeptídicos. Por ejemplo, el dominio de unión a ADN puede comprender un polímero, un espaciador peptídico, un ácido nucleico formador de triplex, una poliamida, un enlazador al surco menor, un agente intercalante, un antibiótico y/o un ácido nucleico.

Las moléculas de fusión, incluyendo las proteínas de fusión y los ácidos nucleicos que las codifican, se construyen mediante métodos de clonación y conjugación bioquímica que son bien conocidos para los expertos en la materia. Las proteínas de fusión (y los ácidos nucleicos que las codifican) se pueden diseñar de manera que se conserve el marco de lectura de la traducción entre los componentes de la fusión.

Las fusiones entre secuencias polipeptídicas que poseen actividad recombinasa, por un lado, y un dominio de unión a ADN no proteico (por ejemplo, antibiótico, intercalador, enlazador al surco menor, de ácido nucleico), por el otro, se construyen mediante métodos de conjugación bioquímica conocidos por los expertos en la materia. Véase, por

ejemplo, el catálogo de Pierce Chemical Company (Rockford, I11). En otras realizaciones, se usa un enlazador químico para conectar dominios producidos de manera sintética o recombinante. Dichos enlazadores flexibles son conocidos por los expertos en la materia. Por ejemplo, hay enlazadores de poli(etilenglicol) disponibles en Shearwater Polymers, Inc. (Huntsville, Ala). Estos enlazadores tienen opcionalmente enlaces de amida, enlaces de sulfhidrilo o enlaces heterofuncionales.

Se han descrito métodos y composiciones para realizar fusiones entre un enlazador al surco menor y un polipéptido. Mapp *et al* (2000) *Proc. Natl. Acad. Sci. EE.UU.* 97: 3930-3935.

Con respecto a los polipéptidos de fusión, la expresión "unido operativamente" se refiere al hecho de que cada uno de los componentes realiza la misma función en la unión con el otro componente que haría si no estuviera enlazado.

Un fragmento funcional de una proteína, un polipéptido o un ácido nucleico es una proteína, un polipéptido o un ácido nucleico cuya secuencia no es idéntica a la proteína, al polipéptido o al ácido nucleico de longitud completa, aunque conserva la misma función que la proteína, el polipéptido o el ácido nucleico de longitud completa. Un fragmento funcional puede poseer más, menos o el mismo número de residuos que la molécula nativa correspondiente, y/o puede contener una o más sustituciones de aminoácidos o nucleótidos. Los métodos para determinar la función de un ácido nucleico (por ejemplo, la función de codificación, la capacidad para hibridarse con otro ácido nucleico) son bien conocidos en la técnica. Del mismo modo, también se conocen bien métodos para determinar la función de las proteínas. Por ejemplo, la función de unión al ADN de un polipéptido se puede determinar, por ejemplo, mediante ensayos de unión al filtro, de desplazamiento de la movilidad electroforética o de inmunoprecipitación. Véase Ausubel *et al.*, *supra*. La capacidad de una proteína para interactuar con otra proteína se puede determinar, por ejemplo, mediante coimmunoprecipitación, ensayos de dos híbridos o complementación, ya sea genética o 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. Por consiguiente, las realizaciones incluyen moléculas de fusión con un fragmento funcional de una o más de entre una recombinasa, RecA, NLS, Gal4 y dominios de unión a ADN de polipéptidos.

En ciertas realizaciones, una fusión entre un dominio de unión a ADN polipeptídico y secuencias polipeptídicas que poseen actividad recombinasa es codificada por un ácido nucleico de fusión. En dichos casos, el ácido nucleico se puede clonar en vectores intermedios para la transformación en células procariontas o eucariotas para la replicación y/o la expresión. Los vectores intermedios para el almacenamiento o la manipulación del ácido nucleico de fusión o la producción de proteína de fusión pueden ser vectores procariontas, (por ejemplo, plásmidos), vectores lanzadera, vectores de insectos o vectores virales, por ejemplo. También se puede clonar un ácido nucleico de fusión en un vector de expresión para la administración a una célula bacteriana, célula fúngica, célula de protozoo, célula vegetal o célula animal, por ejemplo, una célula de mamífero o una célula humana. Los vectores para la replicación, expresión, almacenamiento y/o manipulación de secuencias de ácidos nucleicos clonadas son bien conocidos en la técnica. Véase, por ejemplo, Sambrook, *supra*, y Ausubel, *supra*.

Por lo tanto, la expresión de una proteína de fusión en una célula puede producirse como consecuencia de la administración de la proteína de fusión a la célula o mediante la administración de un polinucleótido que codifica la proteína de fusión a una célula, en la que el polinucleótido se transcribe, y la transcripción se traduce para generar la proteína de fusión. El trans-corte y empalme, la escisión del polipéptido y la ligadura del polipéptido también pueden participar en la expresión de una proteína en una célula. Los métodos para la administración de polinucleótidos y polipéptidos a las células se presentan en otra parte en la presente divulgación.

Se pueden incluir dominios enlazadores entre los dominios de polipéptidos, por ejemplo, entre un dominio de unión a ADN y secuencias polipeptídicas que tienen actividad recombinasa. Dichos enlazadores pueden ser secuencias polipeptídicas tales como secuencias de poli-glicina de 1 a aproximadamente 200 aminoácidos. Los dominios enlazadores pueden comprender subsecuencias de aminoácidos flexibles que se sintetizan como parte de una proteína de fusión recombinante. Por ejemplo, un dominio enlazador puede comprender la secuencia de aminoácidos con una pluralidad de aminoácidos, por ejemplo, de 1 al 20. Los expertos en la materia apreciarán de inmediato que se contemplan todos los intervalos y valores de los intervalos indicados explícitamente, por ejemplo, 2 o 10, o de 3 a 11. Como alternativa, se pueden diseñar razonablemente enlazadores flexibles usando programas informáticos capaces de modelizar tanto los sitios de unión al ADN como los propios péptidos (Desjarlais y Berg (1993) *Proc. Natl. Acad. Sci. EE.UU.* 90: 2256-2260; Desjarlais *et al.* (1994) *Proc. Natl. Acad. Sci. EE.UU.* 91:11099-11103) o mediante métodos de presentación en fagos. También se han descrito métodos para obtener secuencias que median la unión no covalente entre dominios polipeptídicos. Wang *et al.* (1999) *Proc. Natl. Acad. Sci. EE.UU.* 96: 9.568-9.573.

Señales de localización nuclear

Las moléculas de fusión, como las desveladas en el presente documento, también comprenden opcionalmente señales de localización nuclear ("NLS"). Como se usa en el presente documento, la expresión "señal de localización nuclear" significa una secuencia de aminoácidos conocida, *in vivo*, por dirigir una proteína dispuesta en el citoplasma de una célula a través de la membrana nuclear y hacia el núcleo de la célula. Una señal de localización nuclear

también puede dirigirse a la superficie exterior de una célula. Por lo tanto, una señal de localización nuclear individual puede dirigir la entidad con la que está asociada al exterior de una célula y al núcleo de una célula. Dichas secuencias pueden ser de cualquier tamaño y composición, por ejemplo, de entre 4 y 400 aminoácidos. Los expertos en la materia apreciarán de inmediato que se contemplan todos los intervalos y valores de los intervalos indicados explícitamente, por ejemplo, más de 25, 25, 15, 12, 10, 8, 7, 6, 5 o 4 aminoácidos.

Las NLS son grupos peptídicos que señalan la importación de una proteína hacia el núcleo. Los ejemplos de NLS son el gran antígeno T de SV40, nucleoplasmina, el VIH-1 Rev y hnRNPA1 (M9), véase Escriou *et al.*, "NLS bioconjugates for targeting therapeutic genes to the nucleus, *Advanced Drug Delivery Reviews*", 55 (2003) 295-306. Se han derivado varios péptidos del antígeno T de SV40. Estos incluyen una NLS corta o NLS largas. Otros péptidos NLS se han derivado de la proteína M9, nucleoplasmina y c-myc.

Dominios de unión al ADN

En ciertas realizaciones, las composiciones y los métodos desvelados en el presente documento implican fusiones entre un dominio de unión al ADN y un dominio que tiene actividad recombinasa. Se puede usar cualquier dominio de unión al ADN conocido en la técnica como parte de una molécula de fusión. Un dominio de unión al ADN puede comprender cualquier entidad molecular capaz de unirse con especificidad de secuencia con el ADN cromosómico. La unión puede estar mediada por interacciones electrostáticas, interacciones hidrófobas o cualquier otro tipo de interacción química. Los ejemplos de restos que pueden comprender parte de un dominio de unión a ADN incluyen, pero sin limitación, enlazadores al surco menor, enlazadores al surco mayor, antibióticos, agentes intercalantes, péptidos, polipéptidos, ácidos nucleicos peptídicos, poliamidas, oligonucleótidos y polinucleótidos. Un ejemplo de un ácido nucleico de unión al ADN es un oligonucleótido formador de triplex.

Las realizaciones incluyen moléculas de fusión con un dominio de unión al ADN que se dirigen a las técnicas y los tratamientos para la conversión de genes, dirección de genes independientes de la homología, recombinación homóloga, mutagénesis dirigida, enfermedades genéticas, animales transgénicos, vectores de expresión y administración a plantas.

Los enlazadores al surco menor incluyen sustancias que, en virtud de sus propiedades estéricas y/o electrostáticas, interaccionan preferentemente con el surco menor de los ácidos nucleicos bicatenarios. Ciertos enlazadores al surco menor muestran una preferencia por determinadas composiciones de secuencias. Por ejemplo, la netropsina, la distamicina y CC-1065 son ejemplos de enlazadores al surco menor que se unen específicamente a secuencias ricas en AT, en particular, series de A o T. Documento WO 96/32496.

Por ejemplo, en la patente de Estados Unidos N° 6.555.692, se describen dominios de unión al ADN de poliamida. Por ejemplo, en las patentes de Estados Unidos N° 5.539.082; 5.773.571; 6.395.474; 6.451.968 y 7.378.485, se describen ácidos nucleicos peptídicos. Véase también Nielsen *et al.* (1991) *Science* 254: 1497-1500.

Se conocen numerosos antibióticos que ejercen sus efectos mediante la unión a ADN. La unión de antibióticos al ADN suele ser específica de la secuencia o mostrar preferencias por secuencias. La actinomicina, por ejemplo, es un agente de unión al ADN relativamente específico de GC.

Los dominios de unión a ADN polipeptídicos se encuentran, por ejemplo, en las proteínas implicadas en la replicación del ADN, la reparación del ADN, la recombinación y la transcripción. Se han observado regiones definidas dentro de la secuencia polipeptídica de diversos factores de transcripción que son responsables de la unión al ADN con especificidad de secuencia. Estas regiones incluyen, pero sin limitación, los motivos conocidos como cremalleras de leucina, dominios de hélice-bucle-hélice (HLH), dominios de hélice-giro-hélice, dedos de cinc, motivos de lámina β , motivos receptores de esteroides, homeodominios de dominios bZIP, ganchos de AT y otros. Se conocen las secuencias de aminoácidos de estos motivos y, en algunos casos, se han identificado los aminoácidos que son fundamentales para la especificidad de secuencia. Véase, por ejemplo, Pabo *et al.* (1992) *Ann. Rev. Biochem.* 61: 1053-1095 y las referencias citadas en el mismo. Los dominios de unión al ADN bien caracterizados ilustrativos incluyen aquellos para LexA, Gal4 y zif268. Webster *et al.* (1988) *Cell* 52: 169-178.

Las secuencias peptídicas implicadas en el reconocimiento específico del ADN, tales como las que se encuentran en los factores de transcripción, se pueden obtener a través de técnicas de clonación y expresión de ADN recombinante o mediante síntesis química, y se pueden unir a otros componentes de una molécula de fusión mediante métodos conocidos en la técnica.

Además de los dominios de unión al ADN de origen natural tales como los descritos anteriormente, también se pueden usar dominios de unión al ADN no naturales, obtenidos por ingeniería genética. A este respecto, el dominio de unión al ADN de dedo de cinc es útil, en tanto en cuanto permite diseñar proteínas de dedo de cinc para unirse a cualquier secuencia de ADN de elección. Un dominio de unión de dedo de cinc comprende una o más estructuras de dedo de cinc. Miller *et al.* (1985) *EMBO J* 4:1609-1614; Rhodes (1993) *Scientific American*, febrero: 56-65; patente de EE.UU. N° 6.453.242. Por lo general, un solo dedo de cinc es de aproximadamente 30 aminoácidos de longitud y contiene cuatro restos de aminoácidos que se coordinan con el cinc. Estudios estructurales han demostrado que el

motivo de dedo de cinc canónico (C_2H_2) contiene dos láminas β (sujetas en un giro β que, por lo general, contiene dos restos de cisteína que se coordinan con el cinc) empaquetadas contra una hélice α (que, en general, contiene dos restos de histidina que se coordinan con el cinc).

5 Los dedos de cinc incluyen tanto dedos de cinc C_2H_2 canónicos (es decir, aquellos en los que el ion de cinc es coordinado por dos restos de cisteína y dos restos de histidina) y dedos de cinc no canónicos tales como, por ejemplo, los dedos de cinc C_3H (aquellos en los que el ion de cinc es coordinado por tres restos de cisteína y un resto de histidina) y dedos de cinc C_4 (aquellos en los que el ion de cinc es coordinado por cuatro restos de cisteína). Los dedos de cinc no canónicos también pueden incluir aquellos en los que un aminoácido distinto de cisteína o histidina sustituye a uno de estos restos de coordinación del cinc. Véase, por ejemplo, el documento WO 02/057293 (25 de julio, 2002) y US 2003/0108880 (12 de junio, 2003).

10 Los dominios de unión a dedos de cinc se pueden diseñar por ingeniería genética para que se unan a una secuencia de elección. Véase, 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. Por consiguiente, el dominio de unión de dedo de cinc se puede diseñar por ingeniería genética para que tenga una nueva especificidad de unión, en comparación con una proteína de dedo de cinc de origen natural. Los métodos de diseño por ingeniería genética incluyen, pero sin limitación, el diseño racional y varios tipos de métodos de selección empíricos. El diseño racional incluye, por ejemplo, el uso de bases de datos que comprenden secuencias de nucleótidos de triplete (o cuadruplete) y secuencias de aminoácidos de dedo 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 en particular. Véanse, por ejemplo, las patentes de EE.UU. N° 6.140.081; 6.453.242; 6.534.261; 6.610.512; 6.746.838; 6.866.997; 7.067.617; la publicación de solicitud de patente de EE.UU. N° 2002/0165356; 2004/0197892; 2007/0154989; 2007/0213269; y la publicación de solicitud de patente internacional N° WO 98/53059 y WO 2003/016496.

20 En las patentes de EE.UU. N° 5.789.538; 5.925.523; 6.007.988; 6.013.453; 6.140.466; 6.200.759; 6.242.568; 6.410.248; 6.733.970; 6.790.941; 7.029.847 y 7.297.491; así como en las publicaciones de solicitud de patente de EE.UU. N° 2007/0009948 y 2007/0009962; WO 98/37186; WO 01/60970 y GB 2.338.237, se desvelan métodos de selección ilustrativos, incluyendo presentación en fagos, trampa de interacción, selección de híbridos y sistemas de dos híbridos.

25 Maeder *et al.* (2008) *Mol Cell* 31:294-301 describen métodos adicionales para el diseño de dominios de unión al ADN de dedos de cinc específicos de secuencia.

30 La mejora de la especificidad de unión para los dominios de unión a los dedos de cinc se ha descrito, por ejemplo, en la patente de EE.UU. N° 6.794.136 (21 de septiembre de 2004). En la patente de EE.UU. N° 6.479.626 y en la publicación de solicitud de patente N° 2003/0119023, se desvelan aspectos adicionales del diseño por ingeniería genética de dedos de cinc, con respecto a secuencias de enlazadores entre dedos. Véase también Moore *et al.* (2001a) *Proc. Natl. Acad. Sci. EE.UU.* 98:1432-1436; Moore *et al.* (2001b) *Proc. Natl. Acad. Sci. EE.UU.* 98:1437-1441 y WO 01/53480.

35 Los dominios de unión al ADN de dedo de cinc, diseñados por ingeniería genética para unirse a una secuencia de ADN de elección, se encuentran disponibles en el mercado (CompoZr™, Sigma-Aldrich, St. Louis, MO). Akopian *et al.* (2003) *Proc. Natl. Acad. Sci. EE.UU.* 100: 8688-8691 han descrito fusiones entre un dominio de recombinasa y un dominio de unión al ADN de dedo de cinc.

40 Todas las referencias citadas en el presente apartado, titulado "dominios de unión al ADN", se incorporan en el presente documento por referencia en su totalidad para los fines de la divulgación de los dominios de unión al ADN y los métodos para el diseño, la selección y la obtención por ingeniería genética de dominios de unión al ADN de dedo de cinc reconocidos en la técnica.

45 En general, Gal4 se usa como un ejemplo de dominio de unión al ADN. Del mismo modo, RecA se usa como un ejemplo de recombinasa, siendo una proteína de fusión de ambos una proteína de fusión ilustrativa. Una de las utilidades de la proteína de fusión NLS-RecA-Gal4 es su capacidad como RecA para recubrir el ADN monocatenario y encontrar regiones homólogas en un genoma al filamento de RecA. Al hacer esto, la parte RecA de NLS-RecA-Gal4 aporta la actividad asociada con el dominio de unión al ADN Gal4 a la región cromosómica diana. El motivo de unión al ADN Gal4 contiene tanto un centro de coordinación de metal como un motivo de dimerización. Ambas actividades son como las necesarias para que el ADN recubierto de proteína de fusión NLS-RecA-Gal4 potencie las roturas cromosómicas mediante la posibilidad de crear horquillas de replicación atascadas. También se prevén otros mecanismos.

50 Con el uso de dicha capacidad de las proteínas de fusión RecA, se pueden llevar diferentes actividades a distintas ubicaciones cromosómicas mediante la sustitución del dominio Gal4 con otras proteínas o motivos. Por ejemplo, se podría sustituir una nucleasa con Gal4. En este caso, el filamento recubierto con RecA traería la nucleasa, a través

de su unión a RecA, a un sitio cromosómico específico. Varios informes recientes han puesto de relieve el uso de nucleasas de dedo de cinc para modificar regiones cromosómicas específicas mediante su capacidad para inducir roturas de cadenas dobles (Bibikova *et al.*, 2002; Porteus y Baltimore, 2003; Umov *et al.*, 2005; Wright *et al.*, 2005). La especificidad de esta técnica se basa en la observación de que la endonucleasa de restricción, *FokI*, solo está activa como un dímero. Por consiguiente, este sistema requiere que los dedos de cinc con *FokI* se unan a dos sitios distintos. Las roturas de cadenas dobles inducidas por *FokI* se pueden reparar a partir de un plásmido de ADN suministrado exógenamente que contenga una región de homología. Si el ADN suministrado exógenamente contiene un cambio con respecto a la diana cromosómica, este cambio se puede incorporar en el cromosoma reparado. Esta técnica parece funcionar bien en una variedad de sistemas, y se ha usado de manera eficaz para modificar cromosomas en *Drosophila* (Bibikova *et al.*, 2002), tabaco (Wright *et al.*, 2005) y células humanas (Porteus y Baltimore, 2003; Urnov *et al.*, 2005). Debido a que el diseño por ingeniería genética de los dedos de cinc puede requerir una importante selección, el uso generalizado de esta técnica puede ser limitado. La función de búsqueda de homología de *RecA* puede sustituir a los dedos de cinc en este sistema. Cualquiera de las proteínas quiméricas *RecA-FokI* o los filamentos recubiertos con *RecA* que se unen a *FokI* podría inducir roturas de cadena doble específicas en sitios cromosómicos específicos. También se podrían usar otras nucleasas tales como *I-sceI* y *EcoRI*.

Otras actividades que podrían sustituir al dominio Gal4 incluyen diferentes motivos de unión al ADN tales como los dedos de cinc y las proteínas hélice-giro-hélice. Esto potenciaría aún más actividades distintas y la posibilidad de modificar un cromosoma con especificidad del sitio.

Las técnicas incluyen la introducción de una proteína de fusión con un dominio de unión al ADN que no se une específicamente a un ADN. Las uniones triviales no son uniones específicas. La unión específica, como se usa comúnmente la expresión en biología, se refiere, en general, a una molécula que se une a una diana con una afinidad relativamente alta en comparación con los tejidos no diana y, por lo general, implica una pluralidad de interacciones no covalentes, tales como interacciones electrostáticas, interacciones de van der Waals, enlaces de hidrógeno, y similares. Las interacciones de unión específica caracterizan las interacciones de unión de anticuerpo-antígeno, de unión de enzima-sustrato y, específicamente, de unión de proteína-receptor. Aunque, de vez en cuando, dichas moléculas se pueden unir a tejidos además de sus dianas, se dice que dicha unión carece de especificidad y no es una unión específica. Así pues, un dominio de unión a ADN no presentará unión específica a un ácido nucleico a menos que haya una secuencia reconocida de manera específica por el dominio.

Recombinasas

El término recombinasa se refiere a una enzima de recombinación genética que cataliza enzimáticamente, en una célula, la unión de pedazos relativamente cortos de ADN entre dos cadenas de ADN relativamente más largas. Las recombinasas incluyen recombinasa Cre, recombinasa Hin, RecA, RAD51, Tre y FLP. La recombinasa Cre es una topoisomerasa de tipo I del bacteriófago P1 que cataliza la recombinación dirigida de sitio del DNA entre los sitios loxP. La recombinasa Hin es una proteína de 21 kD compuesta de 198 aminoácidos que se encuentra en la bacteria *Salmonella*. Hin pertenece a la familia de serina recombinasas de ADN invertidas que se basan en la serina del sitio activo para iniciar la división y la recombinación del ADN. RAD51 es un gen humano. La proteína codificada por este gen es un miembro de la familia de proteínas RAD51 que ayudan a la reparación de las roturas de cadenas dobles de ADN. Los miembros de la familia RAD51 son homólogos a la RecA bacteriana y a Rad51 de levadura. La recombinasa Tre es una enzima experimental que, en ensayos de laboratorio, ha eliminado con éxito el ADN insertado por el VIH de células infectadas. La enzima se deriva de la recombinasa Cre a través de la mutación selectiva con el fin de identificar marcadores del VIH que no están unidos por sitios loxP, deshabilitando, de ese modo, los intentos de recombinación de Cre-Lox. FLP se refiere a enzima de recombinación flipasa (FLP o Flp) derivada del plásmido 2 μ de levadura *Saccharomyces cerevisiae* del panadero.

RecA es conocida por su actividad recombinasa para catalizar el intercambio de cadenas durante la reparación de las roturas de las cadenas dobles por recombinación homóloga (McGrew y Knight, 2003) Radding, *et al.*, 1981; Seitz *et al.*, 1998). RecA también ha mostrado catalizar la proteólisis, por ejemplo, de las proteínas represoras LexA y λ , y poseen actividad ATPasa dependiente del ADN. Tras producirse una rotura de una cadena doble por radiación ionizante o algún otro daño, las exonucleasas vuelven a masticar los extremos 5' a 3' de ADN, exponiendo así una cadena de ADN (Cox, 1999; McGrew y Knight, 2003). El ADN monocatenario es estabilizado por la proteína de unión de una sola cadena (SSB). Después de la unión de SSB, RecA se une al ADN monocatenario (mc) y forma un filamento de nucleoproteína helicoidal (denominado filamento o filamento presináptico). Durante la reparación del ADN, las funciones de búsqueda de homología de RecA dirigen el filamento al ADN homólogo y catalizan el apareamiento de bases homólogas y el intercambio de cadenas. Esto da lugar a la formación de heterodúplex de ADN. Tras la invasión de la cadena, la ADN polimerasa alarga el ADNmc basándose en el molde de ADN homólogo para reparar la rotura del ADN, y se forman estructuras de cruce o uniones de Holliday. RecA también muestra una función motora que participa en la migración de las estructuras de cruce (Campbell y Davis, 1999).

La actividad recombinasa comprende una serie de diferentes funciones. Por ejemplo, las secuencias polipeptídicas que tienen actividad recombinasa son capaces de unirse de un modo no específico de la secuencia al ADN monocatenario para formar un filamento de nucleoproteína. Dichos filamentos de nucleoproteína unidos a la

recombinasa son capaces de interactuar de una manera no específica de la secuencia con una molécula de ADN bicatenaria, de buscar las secuencias de la molécula bicatenaria que son homólogas a secuencias del filamento y, cuando se encuentran dichas secuencias, desplazar una de las cadenas de la molécula bicatenaria para permitir el apareamiento de bases entre las secuencias del filamento y las secuencias complementarias de una de las cadenas de la molécula bicatenaria. Estas etapas se denominan colectivamente "sinapsis".

Se han examinado RecA y las proteínas de tipo RecA (denominadas Rad51 en especies no bacterianas) para estimular la dirección de genes y la recombinación homóloga en una variedad de sistemas eucariotas. En células de tabaco, la expresión de RecA bacteriana que contiene una señal de localización nuclear (NLS) aumenta la reparación de los daños en el ADN inducidos por la mitomicina C mediante recombinación homóloga y recombinación intracromosómica somática (recombinación entre cromosomas homólogos) de tres a diez veces (Reiss *et al.*, 1996). La expresión de NLSRecA en el tabaco también puede estimular el intercambio de cromátidas hermanas 2,4 veces frente a los niveles de tipo silvestre (Reiss *et al.*, 2000). En células somáticas de mamífero, la sobreexpresión de NLSRecA multiplica por 10 la dirección génica mediante recombinación homóloga (Shcherbakova *et al.*, 2000). Sin embargo, en las células humanas, la sobreexpresión de un homólogo humano de RecA, hRAD51, solo estimula la recombinación de 2 a 3 veces frente a los niveles de tipo silvestre en virtud de la selección de antibióticos (Yanez y Porter, 1999). En el pez cebra, se corrigió una forma mutante de la proteína verde fluorescente mejorada (EGFP) a baja frecuencia mediante la inyección de filamentos RecA-ADNmc directamente (Cui *et al.*, 2003). Rad52, un miembro del grupo de epistasis Rad51, también potencia la hibridación de una sola cadena y la alteración de genes de bajo nivel en el pez cebra usando oligonucleótidos mutados (Takahashi y Dawid, 2005). Tomados en conjunto, estos estudios indican que la expresión ectópica de RecA o Rad51 genera una estimulación modesta de la recombinación homóloga, pero no aumenta los niveles lo suficiente como para que sea útil para la dirección génica.

Así pues, las actividades recombinasa incluyen, pero sin limitación, la unión a ADN monocatenario, la sinapsis, la búsqueda de homología, la invasión de dúplex por ADN monocatenario, la formación de heterodúplex, la hidrólisis de ATP y la proteólisis. La recombinasa prototípica es la proteína RecA de *E. coli*. Véase, por ejemplo, la patente de EE.UU. N° 4.888.274. También se han descrito proteínas de tipo RecA procariontes en especies de *Salmonella*, *Bacillus* y *Proteus*. Una proteína RecA termoestable, de *Thermus aquaticus*, se ha descrito en la patente de EE.UU. N° 5.510.473. Se ha descrito un homólogo de RecA de bacteriófago T4, la proteína UvsX. Se han descrito mutantes de RecA, que tienen actividades de recombinasa modificadas, por ejemplo, en las patentes de EE.UU. N° 6.774.213.; 7.176.007 y 7.294.494. Los homólogos de RecA vegetales se describen, por ejemplo, en las patentes de EE.UU. N° 5.674.992.; 6.388.169 y 6.809.183. Se han descrito fragmentos de RecA que contienen actividad recombinasa, por ejemplo, en la patente de EE.UU. N° 5.731.411. Se han descrito proteínas RecA mutantes que tienen una mayor actividad recombinasa tales como, por ejemplo, RecA803. Véase, por ejemplo, Madiraju *et al.* (1988) *Proc. Natl. Acad. Sci.* EE.UU. 85: 6592-6596.

Un homólogo eucariota de RecA, que también posee actividad recombinasa, es la proteína Rad51, que se identificó por primera vez en la levadura *Saccharomyces cerevisiae*. Véase Bishop *et al.*, (1992) *Cell* 69: 439-56 y Shinohara *et al.*, (1992) *Cell*: 457-70 Aboussekhra, *et al.*, (1992) *Mol. Cell. Biol.* 72:3224-3234. Basile *et al.*, (1992) *Mol. Cell. Biol.* 12, 3235-3246. En las patentes de EE.UU. N° 6.541.684; 6.720.478; 6.905.857 y 7.034.117, se describen secuencias de Rad51 vegetal. Otra proteína de levadura que es homóloga a RecA es la proteína Dmcl. Se han descrito homólogos de RecA/Rad51 en organismos distintos de *E. coli* y *S. cerevisiae*. Morita *et al.* (1993) *Proc. Natl. Acad. Sci.* EE.UU. 90:6577-6580; Shinohara *et al.* (1993) *Nature Genet.* 4:239-243; Heyer (1994) *Experientia* 50:223-233; Maeshima *et al.* (1995) *Gene* 160:195- 200; patentes de EE.UU. N° 6.541.684 y 6.905.857.

En el presente documento, "RecA" o "proteína RecA" se refiere a una familia de proteínas de recombinación de tipo RecA que tienen prácticamente todas o la mayor parte de las mismas funciones, en particular: (i) la capacidad de colocar correctamente oligonucleótidos o polinucleótidos en sus dianas homólogas para la posterior extensión por parte de ADN polimerasas; (ii) la capacidad topológica de preparar ácido nucleico en dúplex para la síntesis de ADN; y (iii) la capacidad de los complejos de RecA/oligonucleótido o de RecA/polinucleótido para encontrar de manera eficaz y unirse a secuencias complementarias. La proteína RecA mejor caracterizada es de *E. coli*. Además de la forma alélica original de la proteína, se ha identificado una serie de proteínas de tipo RecA mutantes, por ejemplo, RecA803. Además, muchos organismos tienen proteínas de tipo RecA de transferencia de cadena, incluyendo, por ejemplo, levadura, *Drosophila*, mamíferos incluyendo seres humanos y plantas. Estas proteínas incluyen, por ejemplo, Rec1, Rec2, Rad51, Rad51B, Rad51C, Rad51D, Rad51E, XRCC2 y DMC1. Una realización de la proteína de recombinación es la proteína RecA de *E. coli*. Como alternativa, la proteína RecA puede ser la proteína mutante RecA-803 de *E. coli*, una proteína RecA de otra fuente bacteriana o una proteína de recombinación homóloga de otro organismo.

Por ejemplo, en Fugisawa *et al.* (1985) *Nucl. Acids Res.* 13:7473; Hsieh *et al.* (1986) *Cell* 44:885; Hsieh *et al.* (1989) *J. Biol. Chem.* 264:5089; Fishel *et al.* (1988) *Proc. Natl. Acad. Sci.* EE.UU. 85:3683; Cassuto *et al.*, (1987) *Mol. Gen. Genet.* 208:10; Ganea *et al.* (1987) *Mol. Cell Biol.* 7:3124; Moor *et al.* (1990) *J. Biol. Chem.* 11108; Keene *et al.* (1984) *Nucl. Acids Res.* 12:3057; Kimiec (1984) *Cold Spring Harbor Symp.* 48:675; Kimeic (1986) *Cell* 44:545; Kolodner *et al.* (1987) *Proc. Natl. Acad. Sci.* EE.UU. 84:5560; Sugino *et al.* (1985) *Proc. Natl. Acad. Sci.* EE.UU. 85: 3683; Halbrook *et al.* (1989) *J. Biol. Chem.* 264:21403; Eisen *et al.* (1988) *Proc. Natl. Acad. Sci.* EE.UU. 85:7481;

McCarthy *et al.* (1988) *Proc. Natl Acad. Sci.* EE.UU. 85:5854; y Lowenhaupt *et al.* (1989) *J. Biol Chem.* 264:20568, que se incorporan en el presente documento por referencia, se pueden encontrar otras descripciones adicionales de proteínas que tienen actividad recombinasa. Véase también Brendel *et al.* (1997) *J. Mol. Evol.* 44:528 541.

5 Los ejemplos de proteínas que tienen actividad recombinasa incluyen recA, recA803, uvsX, y otras recA mutantes y recombinasas de tipo recA (Roca (1990) *Crit. Rev. Biochem. Molec. Biol.* 25:415), sepl (Kolodner *et al.* (1987) *Proc. Natl. Acad. Sci.* EE.UU. 84:5560; Tishkoff *et al.* (1991) *Molec. Cell Biol* 11:2593), RuvC (Dunderdale *et al.* (1991) *Nature* 354:506), DST2, KEM1 y XRN1 (Dykstra *et al.* (1991) *Molec. Cell. Biol* 11:2583), STPa/DSTI (Clark *et al.* (1991) *Molec. Cell Biol* 11:2576), HPP-1 (Moore *et al.* (1991) *Proc. Natl. Acad. Sci.* EE.UU. 88:9067), otras recombinasas eucariotas (Bishop *et al.* (1992) *Cell* 69:439; y Shinohara *et al.* (1992) *Cell* 69:457); incorporados en el presente documento por referencia.

15 En la patente de EE.UU. N° 6.686.515, se han descrito proteínas evolucionadas *in vitro* que tienen actividad recombinasa. Otras publicaciones relacionadas con las recombinasas incluyen, por ejemplo, las patentes de EE.UU. N° 7.732.585, 7.361.641, 7.144.734. Para una revisión de las recombinasas, véase Cox (2001) *Proc. Natl. Acad. Sci.* EE.UU. 98: 8173-8180.

Métodos de formación de filamentos de nucleoproteína

20 En ciertas realizaciones, se pone en contacto una molécula de fusión como la desvelada en el presente documento con un ácido nucleico para formar un filamento de nucleoproteína o "filamento". El término filamento, en el contexto de la formación de una estructura con una recombinasa, es un término conocido por los expertos en estos campos. A continuación, el filamento de nucleoproteína así formado se puede poner en contacto, por ejemplo, con otro ácido nucleico o introducirse en una célula. Los métodos de formación de filamentos de nucleoproteína, en los que los filamentos comprenden secuencias polipeptídicas que tienen actividad recombinasa y un ácido nucleico, son bien conocidos en la técnica. Véase, por ejemplo, Cui *et al.* (2003) *Marine Biotechnol.* 5:174-184 y las patentes de EE.UU. N° 4.888.274; 5.763.240; 5.948.653 y 7.199.281, cuyas divulgaciones se incorporan por referencia con fines divulgativos de las técnicas ilustrativas para la unión de las recombinasas a los ácidos nucleicos para formar filamentos de nucleoproteína.

30 En general, una molécula que tiene actividad recombinasa se pone en contacto con un ácido nucleico monocatenario lineal. El ácido nucleico monocatenario lineal puede ser una sonda. La preparación de dichos ácidos nucleicos monocatenarios es conocida. Por lo general, la mezcla de reacción contiene un ion de magnesio. Opcionalmente, la mezcla de reacción está tamponada y, opcionalmente, también contiene ATP, dATP o un análogo de ATP no hidrolizable tal como, por ejemplo, γ -tio-ATP (ATP- γ -S) o γ -tio-GTP (GTP- γ -S). Las mezclas de reacción también pueden contener opcionalmente un sistema de generación de ATP. Las moléculas de ADN bicatenarias se pueden desnaturalizar (por ejemplo, por calor o álcali) ya sea antes de, o durante, la formación de los filamentos. La optimización de la proporción molar de recombinasa con respecto a un ácido nucleico pertenece al alcance de la técnica. Por ejemplo, se puede añadir una serie de diferentes concentraciones de recombinasa a una cantidad constante de ácido nucleico, y ensayarse la formación de filamentos mediante la movilidad en un gel de agarosa o acrilamida. Debido a que la proteína unida retarda la movilidad electroforética de un polinucleótido, la formación de filamentos se pone de manifiesto por la movilidad retardada del ácido nucleico. Se puede usar cualquier grado máximo de retraso o cantidad máxima de ácido nucleico que migre con una movilidad retardada para indicar proporciones óptimas de recombinasa:ácido nucleico. La asociación entre el ADN y la proteína también se puede cuantificar midiendo la capacidad de un polinucleótido para unirse a nitrocelulosa.

Secuencias exógenas

50 Los métodos y las composiciones expuestos en el presente documento se pueden usar para la integración dirigida de secuencias exógenas (también denominadas en el presente documento secuencias donantes) a una región de interés del genoma de una célula. La integración dirigida de una secuencia exógena se puede producir mediante mecanismos tanto dependientes de la homología como independientes de la homología. Los datos proporcionados en el presente documento muestran la amplia aplicabilidad de estas técnicas a través de las especies y la amplia incorporación de los ADN en general. Por consiguiente, las realizaciones incluyen la inserción de ADN para el tratamiento de las diversas afecciones descritas en el presente documento, así como terapias para insertar ADN de tipo silvestre no defectuosos en las células para reemplazar secuencias de ácidos nucleicos defectuosos. Por lo tanto, las realizaciones incluyen ácidos nucleicos exógenos dirigidos a técnicas y tratamientos para la conversión de genes, dirección génica independiente de la homología, recombinación homóloga, mutagénesis dirigida, enfermedades genéticas, animales transgénicos, vectores de expresión y administración a plantas.

60 Las secuencias exógenas ilustrativas incluyen, pero sin limitación, ADNc, secuencias promotoras, secuencias potenciadoras, marcadores de epítomos, genes marcadores, sitios de reconocimiento de enzima de escisión y diversos tipos de construcciones de expresión. Los genes marcadores incluyen, pero sin limitación, secuencias que codifican proteínas que median la resistencia a los antibióticos (por ejemplo, resistencia a la ampicilina, resistencia a la neomicina, resistencia a G418, resistencia a la puromicina), secuencias que codifican proteínas coloreadas o fluorescentes o luminiscentes (por ejemplo, proteína verde fluorescente, proteína verde fluorescente mejorada,

65

proteína roja fluorescente, luciferasa) y las proteínas que median el crecimiento celular mejorado y/o la amplificación de genes (por ejemplo, la dihidrofolato reductasa). Los marcadores de epítomos incluyen, por ejemplo, una o más copias de FLAG, His, myc, Tap, HA o cualquier secuencia de aminoácidos detectable.

5 Opcionalmente, las construcciones de expresión de proteínas incluyen, por ejemplo, ADNc y secuencias de control de la transcripción en unión operativa con secuencias de ADNc. Las secuencias de control de la transcripción incluyen promotores, potenciadores y aislantes. Las secuencias adicionales reguladoras de la transcripción y de la traducción que se pueden incluir en las construcciones de expresión incluyen, por ejemplo, sitios de entrada al ribosoma internos, secuencias que codifican péptidos 2A y señales de poliadenilación. Para la expresión óptima de una o más proteínas codificadas por secuencias exógenas integradas en un genoma, el sitio de integración cromosómica debe ser compatible con la transcripción de alto nivel de las secuencias integradas, preferentemente en una amplia selección de tipos de células y estados de desarrollo. Sin embargo, se ha observado que la transcripción de las secuencias integradas varía dependiendo del sitio de integración debido, entre otras cosas, a la estructura de la cromatina del genoma en el sitio de integración. Por consiguiente, son deseables los sitios diana genómicos que apoyan la transcripción de alto nivel de las secuencias integradas. Los ejemplos no limitantes de regiones cromosómicas que no codifican un gen esencial y apoyan la transcripción de alto nivel de secuencias integradas en las mismas incluyen el locus Rosa26 murino (y su homólogo humano), el locus CCR5 humano y el sitio de integración de AAV P1 en el cromosoma humano 19. Los sitios diana genómicos adicionales que apoyan la transcripción de alto nivel de las secuencias integradas se pueden identificar como regiones de la cromatina abierta como se describe, por ejemplo, en las publicaciones de solicitud de patente de EE.UU. n° 2002/0064802 (30 de mayo de 2002) y 2002/0081603 (Jun. 27, 2002).

Los sitios de reconocimiento de las enzimas de escisión incluyen, por ejemplo, las secuencias reconocidas por las endonucleasas de restricción, las endonucleasas buscadoras y/o las meganucleasas. La integración dirigida de un sitio de reconocimiento de enzimas de escisión (ya sea mediante mecanismos dependientes de la homología o independientes de la homología) es útil para la generación de células cuyo genoma solo contiene un único sitio que puede ser escindido por una determinada enzima. El contacto de dichas células con una enzima que realice el reconocimiento y la escisión en el sitio único facilita la posterior integración dirigida de secuencias exógenas (ya sea mediante mecanismos dependientes de la homología o independientes de la homología) y/o la mutagénesis dirigida en el sitio que es escindido.

Para ciertas realizaciones, es deseable que no haya un sitio de integración en un gen esencial (por ejemplo, un gen esencial para la viabilidad celular), de modo que la inactivación de dicho gen esencial no produzca la integración de las secuencias exógenas. Por otro lado, si la intención es la de desactivar la función del gen (es decir, crear un gen "desactivado") la integración dirigida de una secuencia exógena para interrumpir un gen endógeno es un método eficaz. En estos casos, la secuencia exógena puede ser cualquier secuencia capaz de bloquear la transcripción del gen endógeno o de generar un producto de traducción no funcional, por ejemplo, un tramo corto de secuencia de aminoácidos, que sea opcionalmente detectable (véase la información anterior). En ciertas realizaciones, las secuencias exógenas pueden comprender un gen marcador (descrito anteriormente), lo que permite la selección de células que hayan sufrido la integración dirigida. En ciertas realizaciones, también será deseable que la integración de las secuencias exógenas no produzca la activación ectópica de uno o más genes celulares (por ejemplo, oncogenes). Por otro lado, en el caso de la integración de secuencias promotoras y/o potenciadoras, puede ser deseable la expresión ectópica.

45 En ciertas realizaciones, la integración dirigida se usa para insertar una construcción de expresión de ARN, por ejemplo, las secuencias responsables de la expresión regulada de ARN micro, ARNip o ARNhp. Los promotores, los potenciadores y las secuencias reguladoras de la transcripción adicionales, según lo descrito anteriormente, también se pueden incorporar en una construcción de expresión de ARN.

50 **Sondas**

Los datos presentados en el presente documento muestran que es posible asociar una sonda con un filamento de nucleoproteína para dirigir el filamento con especificidad a una diana de un cromosoma de una célula hospedadora. Una diana se refiere a una molécula, un tejido o un lugar predeterminado con el que el usuario tiene la intención de unir la sonda. Una sonda, en el contexto de un filamento de nucleoproteína, se refiere a un ácido nucleico con complementariedad a una secuencia de ácido nucleico diana. Los expertos en la materia están familiarizados con los métodos para identificar los sitios de interés y desarrollar sondas. Las sondas se pueden seleccionar como aquellas adecuadas para la recombinasa seleccionada. El tamaño de la sonda se puede seleccionar correspondientemente. Los ejemplos incluyen sondas con 60 pb o 1.300 pb, o con una longitud en el intervalo de aproximadamente 10 y aproximadamente 10.000 restos. Los expertos en la materia apreciarán de inmediato que se contemplan todos los intervalos y valores de los intervalos indicados explícitamente, por ejemplo, de aproximadamente 40 a aproximadamente 5000, de aproximadamente 60 a aproximadamente 1.300, de aproximadamente 10 a aproximadamente 3.000, al menos 10, al menos aproximadamente 30.

65 Las sondas se pueden seleccionar con el grado de especificidad deseado. Como se demuestra en el presente documento, las secuencias exógenas se pueden colocar con un alto grado de exactitud reproducible. Las

secuencias se pueden colocar en el gen diana. La especificidad de la colocación se puede medir en pares de bases (pb) mediante la comparación de la punta de la sonda que se encuentra más secuencia arriba hasta el punto de la secuencia exógena insertada que se encuentra más secuencia arriba, siendo la diferencia entre estos dos puntos la distancia desde la sonda hasta el lugar de inserción. Por consiguiente, los ácidos nucleicos exógenos se pueden colocar, y las sondas se pueden diseñar para la colocación, con una especificidad predeterminada; por ejemplo, de aproximadamente 200 a aproximadamente 2.000 pb. Los expertos en la materia apreciarán de inmediato que se contemplan todos los intervalos y valores de los intervalos indicados explícitamente, por ejemplo, menos de aproximadamente 500 pb, menos de aproximadamente 1.000 pb, menos de aproximadamente 5.000 pb, o de aproximadamente 500 a menos de aproximadamente 5.000 pb. Se puede medir una especificidad predeterminada *in vitro* usando el modelo animal de pez cebra y siguiendo los procedimientos de los Ejemplos, con embriones de pez cebra, inyectados directamente, que tienen incorporado un ADN exógeno dentro del intervalo indicado con un 90 % de exactitud, medida para los embriones que están transfectados con éxito.

Las realizaciones incluyen sondas dirigidas a las técnicas y a los tratamientos para la conversión de genes, dirección génica independiente de la homología, recombinación homóloga, mutagénesis dirigida, enfermedades genéticas, animales transgénicos, vectores de expresión y administración a plantas.

Aplicaciones

Dado que las recombinasas están muy conservadas, tanto entre eucariotas como entre procariotas, y que la actividad de potenciación de la recombinación de las proteínas de fusión desveladas en el presente documento no depende de la presencia de un sitio de unión para el dominio de unión al ADN específico de secuencia presente en la proteína de fusión, los métodos y las composiciones desvelados serán ampliamente aplicables en muchas especies. Estas incluyen, pero sin limitación, procariotas, eucariotas, plantas, metazoos, vertebrados, mamíferos y seres humanos. Las plantas incluyen especies monocotiledóneas y dicotiledóneas. Los ejemplos de plantas incluyen *Arabidopsis*. Los ejemplos de metazoos incluyen moscas de la fruta (*Drosophila*) y ascárids (*Caenorhabditis*). Los ejemplos de vertebrados incluyen ranas (por ejemplo, *Xenopus*) y peces (por ejemplo, *Danio*). Los ejemplos de mamíferos incluyen bovinos, porcinos, ovinos, caprinos, equinos, felinos, caninos, murinos y seres humanos.

Los métodos y las composiciones desvelados en el presente documento encontrarán uso tanto en la investigación como en aplicaciones terapéuticas, como se describirá ahora.

Conversión de genes

En ciertas realizaciones, la introducción, en una célula, de una molécula de fusión que comprende un dominio de recombinasa y un dominio de unión al ADN específico de secuencia conduce a un aumento generalizado, en todo el genoma, de los acontecimientos de recombinación, que se puede manifestar como la conversión génica o la pérdida de heterocigosidad. La selección de un acontecimiento de recombinación de interés permite el aislamiento de nuevas secuencias, incluyendo, por ejemplo, diferentes alelos o haplotipos de secuencias genómicas, secuencias mutantes, secuencias de tipo silvestre, inserciones, eliminaciones o reordenamientos.

La actividad de escisión del ADN de un dominio de recombinasa puede ser dirigida por la formación de un filamento de nucleoproteína que contenga una molécula de fusión que comprende el dominio de recombinasa, como se desvela en el presente documento, y una secuencia homóloga a una secuencia genómica de interés. Dichas moléculas de fusión, cuando se introducen en una célula, pueden facilitar la mutagénesis dirigida en una región genómica de interés como resultado de la escisión en la región de interés seguida de la unión de extremos no homólogos. Dichas mutaciones pueden dar lugar, por ejemplo, a genes desactivados, por ejemplo, para la genómica funcional o la validación de dianas.

La escisión dirigida del ADN, mediada ya sea por una molécula de fusión como la desvelada en el presente documento o por un filamento de nucleoproteína como el desvelado en el presente documento, realizada en ausencia de un polinucleótido exógeno (preferentemente en fase S o G₂), también puede estimular la recombinación entre cromosomas homólogos.

Dirección génica independiente de la homología

La integración de secuencias exógenas en una región de interés de un genoma, cuando las secuencias exógenas carecen de homología con la región de interés, se ve facilitada por la introducción en una célula, junto con las secuencias exógenas, de un filamento de nucleoproteína que se compone de secuencias homólogas a la región de interés recubierta con moléculas de fusión que comprenden un dominio de recombinasa y un dominio de unión al ADN específico de secuencia. La inclusión del dominio de unión al ADN específico de secuencia en la proteína de fusión aumenta la frecuencia de recombinación observada en comparación con los casos en los que se forma un filamento de nucleoproteína usando solo una recombinasa. No es necesario que haya una secuencia diana, o un sitio de unión, para el dominio de unión al ADN específico de secuencia en la secuencia exógena o la región genómica de interés.

Sin el deseo de quedar ligados a teoría alguna, una posible explicación de la capacidad de los filamentos de nucleoproteína que comprenden las proteínas de fusión desveladas en el presente documento para estimular la dirección de genes es que la porción recombinasa de la proteína de fusión cataliza las roturas de las cadenas dobles de ADN genómico homólogo a la secuencia de nucleótidos del componente de ADN del filamento. Es bien sabido que las roturas de las cadenas dobles del ADN celular estimulan los mecanismos de reparación celular, en varios miles de veces, en las proximidades del sitio de escisión, facilitando tanto la integración dependiente de la homología (véase más adelante) como la integración independiente de la homología de secuencias exógenas. Véase, por ejemplo, Rouet *et al.* (1994) *Mol Cell Biol.* 14: 8096-8106; Choulika *et al.* (1995) *Mol. Cell Biol.* 15:1968-1973; Donoho *et al.* (1998) *Mol Cell. Biol* 18: 4070-4078; Johnson *et al.* (2001) *Biochem. Soc. Trans.* 29: 196-201; y Yanez *et al.*, (1998) *Gene Therapy* 5: 149-159.

La integración no dependiente de la homología dirigida, como se describe anteriormente, se puede usar, por ejemplo, para los fines del diseño por ingeniería genética de células y/o la sobreexpresión de proteínas. Las realizaciones incluyen secuencias donantes que carecen de homología con el ADN del hospedador y/o que carecen de homología con el sitio previsto de inserción. Por ejemplo, el ácido nucleico donante se puede diseñar o seleccionar para que carezca de homología con los ácidos nucleicos que se encuentran en o cerca del sitio diana mediante la sonda, por ejemplo, de 0 a 500.000 pb de sonda. Los expertos en la materia apreciarán de inmediato que se contemplan todos los intervalos y valores de los intervalos establecidos expresamente, por ejemplo, en aproximadamente 100.000 pb. La falta de homología puede ser, por ejemplo, que tenga no más del 50 % de identidad de secuencia y/o que carezca de hibridación específica a baja rigurosidad. La falta de homología puede incluir además el criterio de que tenga identidad no superior a 9 pb. Otros criterios para la falta de homología se pueden inferir de la siguiente descripción de la recombinación homóloga. Las realizaciones incluyen células *in vitro*, células tratadas *ex vivo* para la reincorporación en el animal hospedador (por ejemplo, ser humano), animales no humanos, animales transgénicos no humanos y ADN sintético modificado con ADN donante no homólogo, así como sistemas y métodos de producción de los mismos como se desvela en el presente documento.

Recombinación homóloga

También se describen en el presente documento métodos para facilitar la recombinación homóloga entre un locus cromosómico y un ácido nucleico exógeno que porta secuencias que son homólogas al locus cromosómico (por ejemplo, dirección génica). Dichos mecanismos pueden dar lugar bien a la sustitución de una secuencia genómica (por ejemplo, una región de interés en un genoma celular) con una secuencia no idéntica homóloga o a la inserción, en un genoma, de secuencias exógenas normalmente no presentes en ese genoma (siempre que las secuencias que normalmente no se encuentran presentes en el genoma estén ligadas, en el ácido nucleico exógeno, con secuencias que sean homólogas a una región de interés del genoma). Las realizaciones incluyen células *in vitro*, células tratadas *ex vivo* para la reincorporación en el animal hospedador (por ejemplo, ser humano), animales no humanos, animales transgénicos no humanos y ADN sintético modificado con ADN donante no homólogo, así como sistemas y métodos de producción de los mismos como se desvela en el presente documento.

Los métodos desvelados para la recombinación dirigida implican la introducción, en una célula, de un ácido nucleico exógeno que comprende secuencias homólogas a la región de interés, junto con una molécula de fusión que comprende un dominio de recombinasa y un dominio de unión al ADN específico de secuencia. Las moléculas de fusión se han descrito anteriormente, y comprenden opcionalmente una señal de localización nuclear. La secuencia de ácido nucleico exógena, también denominada en el presente documento "secuencia donante", se puede introducir en la célula antes de, simultáneamente con, o después de, la introducción de la molécula de fusión.

Una "secuencia no idéntica homóloga" se refiere a una primera secuencia que comparte un grado de identidad de secuencia con una segunda secuencia, pero cuya secuencia no es idéntica a la segunda secuencia. Por ejemplo, un polinucleótido que comprende la secuencia de tipo silvestre de un gen mutante es homólogo y no idéntico a la secuencia del gen mutante. Del mismo modo, dos alelos de un gen son secuencias homólogas, no idénticas, al igual que dos haplotipos de un determinado locus genómico.

Las realizaciones incluyen un ácido nucleico exógeno, o donante, que contiene homología considerable, que es una homología suficiente con una secuencia genómica para apoyar la recombinación homóloga (o la reparación dirigida por homología) entre ésta y la secuencia genómica con la que guarda homología: aproximadamente 25, 50, 100, 200, 500, 750, 1.000, 1.500, 2.000 nucleótidos o más de homología de secuencia entre un donante y una secuencia genómica (o cualquier valor integral entre 10 y 2.000 nucleótidos, o más), por lo general, soportará la recombinación homóloga entre las mismas. Las secuencias donantes pueden variar en longitud, por ejemplo, de 10 a 10.000 nucleótidos (o cualquier valor integral de nucleótidos entre los mismos) o más. Será muy evidente que la secuencia donante, normalmente, no sea idéntica a la secuencia genómica que reemplaza. Por ejemplo, la secuencia del polinucleótido donante puede contener uno o más cambios, inserciones, eliminaciones, inversiones o reordenamientos de bases individuales con respecto a la secuencia genómica, siempre que haya suficiente homología con las secuencias cromosómicas. Como alternativa, una secuencia donante puede contener una secuencia no homóloga flanqueada por dos regiones de homología, o una secuencia homóloga flanqueada por dos regiones de no homología. Además, las secuencias donantes pueden comprender una molécula de vector que contenga secuencias que no sean homólogas a la región de interés en la cromatina celular. Una molécula donante

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

En general, la o las regiones homólogas de una secuencia donante tendrán al menos un 50 % de identidad de secuencia con una secuencia genómica con la que se desee la recombinación. En ciertas realizaciones, hay un 60 %, 70 %, 80 %, 90 %, 95 %, 98 %, 99 % o 99,9 % de identidad de secuencia. Los expertos en la materia apreciarán de inmediato que se contemplan todos los intervalos y valores pertenecientes a los valores indicados expresamente.

Por lo tanto, en ciertas realizaciones, el grado de homología entre dos secuencias (por ejemplo, un locus genómico y un ácido nucleico exógeno) es considerable, para permitir la recombinación homóloga entre las mismas. Dos secuencias homólogas no idénticas pueden tener cualquier longitud, y su grado de no homología puede ser tan bajo como de un solo nucleótido (por ejemplo, para la corrección de una mutación genómica puntual mediante recombinación homóloga dirigida) o tan alto como de 10 o más kilobases (por ejemplo, para la inserción de un gen en un sitio ectópico predeterminado de un cromosoma). Dos polinucleótidos que comprenden las secuencias homólogas no idénticas no tienen que ser de la misma longitud. Por ejemplo, se puede usar un polinucleótido exógeno (es decir, un polinucleótido donante) de entre 20 y 10.000 nucleótidos o pares de nucleótidos.

Las técnicas para determinar la identidad y la homología entre las secuencias de ácido nucleico y aminoácidos se conocen en la materia. Por lo general, dichas técnicas incluyen la determinación de la secuencia de nucleótidos del ARNm para un gen y/o la determinación de la secuencia de aminoácidos codificada por el mismo, y la comparación de estas secuencias con una segunda secuencia de nucleótidos o de aminoácidos. Las secuencias genómicas también se pueden determinar y comparar de esta manera. En general, la identidad se refiere a una correspondencia exacta de nucleótido a nucleótido y de aminoácido a aminoácido de dos polinucleótidos o secuencias polipeptídicas, respectivamente. Dos o más secuencias (polinucleótido o aminoácido) se pueden comparar determinando su porcentaje de identidad. El porcentaje de identidad de dos secuencias, ya sean secuencias de ácido nucleico o de aminoácidos, es el número de coincidencias exactas 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 es proporcionado por el algoritmo de homología local de Smith y Waterman, "Advances in Applied Mathematics" 2: 482-489 (1981). Este algoritmo se puede aplicar 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., normalizada por Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). El Genetics Computer Group (Madison, WI) proporciona una implementación ilustrativa de este algoritmo para determinar el porcentaje de identidad de una secuencia en la herramienta "BestFit". Los parámetros por defecto de este método se describen en el Manual Wisconsin Sequence Analysis Package Program Manual, Versión 8 (1995) (disponible en Genetics Computer Group, Madison, WI).

Otro método de establecimiento del porcentaje de identidad es el uso del paquete de programas MPSRCH registrado por la Universidad de Edimburgo, desarrollado por John F. Collins y Shane S. Sturrok, y distribuido por IntelliGenetics, Inc. (Mountain View, CA). A partir de este conjunto de paquetes, se puede emplear el algoritmo de Smith-Waterman usando los parámetros por defecto para la tabla de puntuación (por ejemplo, penalización de hueco abierto de 12, penalización de extensión de hueco de uno, y un hueco de seis). A partir de los datos generados, el valor "Coincidencia" refleja la identidad de secuencia.

En la técnica, se conocen en general otros programas adecuados para calcular el porcentaje de identidad o similitud entre secuencias, por ejemplo, otro programa de alineamiento es el BLAST, usado con los parámetros por defecto. Por ejemplo, BLASTN y BLASTP se pueden emplear usando los siguientes parámetros por defecto: código genético = estándar; filtro = ninguno; cadena = ambas; punto de corte = 60; esperado = 10; Matriz = BLOSUM62; Descripción = 50 secuencias; ordenado por = HIGH SCORE; Bases de datos = no redundante, GenBank + EMBL + DDBJ + AP + GenBank CDS translations + Swiss protein + Spupdate + PIR. Los detalles de estos programas se pueden encontrar en la siguiente dirección de internet: <http://www.ncbi.nlm.gov/cgi-bin/-BLAST>. Con respecto a las secuencias descritas en el presente documento, el intervalo de grados deseados de identidad de secuencia es del aproximadamente 80 % al 100 % y de cualquier valor entero que se encuentre entre los mismos. Por lo general, los porcentajes de identidad entre las secuencias son de al menos el 70-75 %, como alternativa, 80-82 %, como alternativa 85-90 %, 92 % o más, 95 % o más, 98 % o más, o 99 % o más.

Como alternativa, el grado de similitud de secuencia entre los polinucleótidos se puede determinar mediante hibridación de los polinucleótidos en condiciones que permitan la formación de dúplex estables entre regiones homólogas, seguida del ensayo para el ácido nucleico bicatenario (por ejemplo, hiperchromicidad, unión a hidroxipatita o digestión con una o varias nucleasas específicas de una sola cadena y la determinación del tamaño de los fragmentos digeridos). Dos secuencias de ácido nucleico o dos secuencias polipeptídicas son considerablemente homólogas entre sí cuando las secuencias presentan al menos aproximadamente 70 %-75 %, como alternativa 80-82 %, como alternativa 85 %-90 %, 92 % o más, 95 % o más, 98 % o más, o 99 % o más de

identidad de secuencia sobre una longitud definida de las moléculas, determinada usando los métodos anteriores. Como se usa en el presente documento, considerablemente homólogas también se refiere a secuencias que muestran identidad completa con una secuencia de ADN o polipeptídica especificada. Las secuencias de ADN que son considerablemente homólogas se pueden identificar en un experimento de hibridación Southern, por ejemplo, en condiciones rigurosas tales como las definidas para ese sistema en particular. La definición de las condiciones de hibridación apropiadas está dentro de la experiencia de la técnica. Véase, por ejemplo, Sambrook *et al, supra*; "Nucleic Acid Hybridization: A Practical Approach", editores B. D. Hames y S. J. Higgins, (1985) Oxford; Washington, D. C.; IRL Press.

La hibridación selectiva de dos fragmentos de ácido nucleico se puede determinar de la siguiente manera. El grado de identidad de secuencia entre dos moléculas de ácido nucleico afecta a la eficacia y a la resistencia de las hibridaciones entre dichas moléculas. Una secuencia de ácido nucleico parcialmente idéntica inhibirá al menos parcialmente la hibridación de una secuencia completamente idéntica a una molécula diana homóloga o idéntica. La inhibición de la hibridación de la secuencia completamente idéntica se puede evaluar usando ensayos de hibridación que son bien conocidos en la técnica (por ejemplo, transferencia Southern (ADN), transferencia Northern (ARN), hibridación en solución, o similares, véase Sambrook, *et al, supra*). Dichos ensayos se pueden realizar usando grados variables de selectividad, por ejemplo, usando condiciones que varían de baja a alta rigurosidad. Si se emplean condiciones de baja rigurosidad, se puede evaluar la ausencia de unión inespecífica usando una sonda secundaria que carezca incluso de un grado parcial de identidad de secuencia (por ejemplo, una sonda que tenga menos del aproximadamente 30 % de identidad de secuencia con la molécula diana), de manera que, en ausencia de uniones inespecíficas, la sonda secundaria no se hibridará a la diana.

Cuando se utiliza un sistema de detección basado en la hibridación, se selecciona una sonda de ácido nucleico que es complementaria a una secuencia de ácido nucleico de referencia y, a continuación, mediante la selección de las condiciones apropiadas, la sonda y la secuencia de referencia se hibridan selectivamente, o se aparean, entre sí para formar una molécula dúplex. Una molécula de ácido nucleico que es capaz de hibridarse selectivamente a una secuencia de referencia en condiciones de hibridación moderadamente rigurosas, normalmente, se hibrida en condiciones que permiten la detección de una secuencia de ácido nucleico diana de al menos aproximadamente 10-14 nucleótidos de longitud que tenga al menos aproximadamente un 70 % de identidad de secuencia con la secuencia de la sonda de ácido nucleico seleccionada. Las condiciones de hibridación rigurosas normalmente permiten la detección de secuencias de ácidos nucleicos diana de al menos aproximadamente 10-14 nucleótidos de longitud que tienen una identidad de secuencia superior al aproximadamente 90-95 % con la secuencia de la sonda de ácido nucleico seleccionada. Las condiciones de hibridación útiles para la hibridación de la sonda/secuencia de referencia, donde la sonda y la secuencia de referencia tienen un grado específico de identidad de secuencia, se pueden determinar como se conoce en la técnica (véase, por ejemplo, "Nucleic Acid Hybridization: A Practical Approach", editores B. D. Hames y Higgins, S. J. (1985) Oxford, Washington, DC; IRL Press).

Las condiciones para la hibridación son bien conocidas por los expertos en la materia. La rigurosidad de la hibridación se refiere al grado en que las condiciones de hibridación desfavorecen la formación de híbridos que contienen nucleótidos no coincidentes, estando la mayor rigurosidad correlacionada con una menor tolerancia para los híbridos no coincidentes. Los factores que afectan a la rigurosidad de la hibridación son bien conocidos por los expertos en la materia, e incluyen, pero sin limitación, temperatura, pH, fuerza iónica y concentración de disolventes orgánicos tales como, por ejemplo, formamida y dimetilsulfóxido. Como es sabido por los expertos en la materia, la rigurosidad de la hibridación aumenta al elevarse las temperaturas, reducirse la fuerza iónica y reducirse las concentraciones de disolventes.

Con respecto a las condiciones de rigurosidad para la hibridación, es bien sabido en la técnica que se pueden emplear numerosas condiciones equivalentes para establecer una determinada rigurosidad variando, por ejemplo, los siguientes factores: la longitud y la naturaleza de las secuencias, la composición de las bases de las diversas secuencias, las concentraciones de sales y otros componentes de la solución de hibridación, la presencia o ausencia de agentes bloqueantes en las soluciones de hibridación (por ejemplo, sulfato de dextrano, polietilenglicol), la temperatura de la reacción de hibridación, y los parámetros de tiempo y las condiciones de lavado.

El polinucleótido donante, exógeno, puede ser ADN o ARN, monocatenario o bicatenario, y se puede introducir en una célula en forma lineal o circular. Si se introduce en forma lineal, los extremos de la secuencia donante se pueden proteger (por ejemplo, de la degradación exonucleolítica) mediante métodos conocidos por los expertos en la materia. Por ejemplo, se añaden uno o más restos de dideoxinucleótidos al extremo 3' de una molécula lineal y/o se ligan oligonucleótidos autocomplementarios a uno o ambos extremos. Véase, por ejemplo, Chang *et al.* (1987) *Proc. Natl. Acad. Sci. EE.UU.* 84: 4.959-4.963; Nehls *et al.* (1996) *Science* 272: 886-889. Otros métodos adicionales para la protección de polinucleótidos exógenos de la degradación incluyen, pero sin limitación, la adición de grupo/s terminal/es amino y el uso de enlaces internucleótidos modificados tales como, por ejemplo, fosforotioatos, fosforamidatos y restos de O-metilribosa o desoxirribosa. Un polinucleótido se puede introducir en una célula como parte de una molécula de vector que tenga secuencias adicionales, tales como, por ejemplo, orígenes de replicación, promotores y genes que codifiquen la resistencia a los antibióticos. Por otra parte, los polinucleótidos donantes se pueden introducir como ácido nucleico desnudo, como ácido nucleico complejo con un agente tal como un liposoma o poloxámero, o pueden ser administrados por virus (por ejemplo, adenovirus, AAV, herpesvirus, retrovirus,

lentivirus).

En realizaciones adicionales, se pueden modificar los extremos de una molécula de ácido nucleico donante exógena de manera que se cree un sustrato más adecuado para la recombinación. Por ejemplo, una molécula de ácido nucleico exógena para su integración en un genoma, mediante un mecanismo bien dependiente o independiente de la homología, puede contener extremos monocatenarios que sobresalgan en 3' ("salientes 3'"). Los métodos de generación de dichos extremos (por ejemplo, el tratamiento de ADN bicatenario lineal con exonucleasas específicas de 5', tales como exonucleasa λ o exonucleasa T7) son conocidos en la técnica.

10 Métodos auxiliares de mejora de la frecuencia de recombinación

También se proporcionan métodos y composiciones que mejoran los niveles de recombinación dirigida, incluyendo, pero sin limitación, el uso de ADNc y/o factores de transcripción obtenidos por ingeniería genética para aumentar la expresión de genes implicados en la 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 del complejo NBS1. Cuando se desea la recombinación homóloga, se pueden usar métodos similares, en combinación con los métodos y las composiciones desvelados en el presente documento, para reprimir la expresión de genes implicados en la unión de extremos no homólogos (por ejemplo, Ku70/80, XRCC4, poli(ADP ribosa) polimerasa, ADN ligasa 4). Véase, 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. Los métodos de activación y represión de la expresión génica usando fusiones entre un dominio de unión de dedo de cinc y un dominio funcional se desvelan, por ejemplo, en las patentes de EE.UU. N° 6.534.261; 6.824.978 y 6.933.113. Los métodos adicionales de represión incluyen el uso de oligonucleótidos antisentido y/o ARN pequeño de interferencia (ARNip o ARNi) dirigidos a la secuencia del gen que se va a reprimir.

Otras proteínas adicionales que participan en la conversión génica y la remodelación de la cromatina relacionada con la recombinación, que se pueden usar en los métodos y en las composiciones mencionados anteriormente, incluyen histona acetiltransferasas (por ejemplo, Esa1p, Tip60), histona metiltransferasas (por ejemplo, Dot1p), histona quinasas e histona fosfatasa.

Se ha publicado que la proteína p53 desempeña un papel central en la represión de la recombinación homóloga. Véase, por ejemplo, Valerie *et al.* (2003) *Oncogene* 22: 5792-5812; Janz, *et al.* (2002) *Oncogene* 21: 5929-5933. Por ejemplo, la tasa de recombinación homóloga en las líneas tumorales humanas deficientes en p53 es 10.000 veces superior que en los fibroblastos humanos primarios, y hay un aumento de 100 veces en la recombinación homóloga 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 sobreexpresión 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 de la transcripción y el control de punto de control del ciclo celular G1 son independientes de su participación en la recombinación homóloga. Saintigny *et al.* (1999) *Oncogene* 18:3553-3563; Boehden *et al.* (2003) *Oncogene* 22:4111-4117. Por consiguiente, la regulación negativa de la actividad de p53 puede servir para aumentar la eficacia de la recombinación homóloga dirigida usando los métodos y las composiciones desvelados en el presente documento. Se puede usar cualquier método para la regulación negativa de la actividad de p53, incluyendo, pero sin limitación, la cotransfección y la sobreexpresión de una p53 mutante dominante negativa o la represión dirigida de la expresión génica de p53 de acuerdo con métodos desvelados, por ejemplo, en la patente de EE.UU. N° 6.534.261.

Se pueden lograr otros aumentos de la eficacia de la recombinación dirigida mediante el bloqueo de las células en la fase G₂ del ciclo celular, cuando los procesos de reparación impulsados por la homología están activados al máximo. Dicha detención se puede lograr de una serie de maneras. Por ejemplo, se pueden tratar las células con, por ejemplo, fármacos, compuestos y/o pequeñas moléculas que influyan en la progresión del ciclo celular con el fin de detener las células que se encuentran en la fase G₂. Las moléculas ilustrativas de este tipo incluyen, pero sin limitación, compuestos que afectan a la polimerización de microtúbulos (por ejemplo, vinblastina, nocodazol, Taxol), compuestos que interactúan con el ADN (por ejemplo, dicloruro de diamina de cis-platino (II), cisplatino, doxorubicina) y/o compuestos que afectan a la síntesis de ADN (por ejemplo, timidina, hidroxiurea, L-mimosina, etopósido, 5-fluorouracilo). Se consiguen otros aumentos adicionales en la eficacia de la recombinación mediante el uso de inhibidores de la histona desacetilasa (HDAC) (por ejemplo, butirato de sodio, tricostatina A) que modifican la estructura de la cromatina para hacer el ADN genómico más accesible a la maquinaria de recombinación celular.

Otros métodos adicionales de detención del ciclo celular incluyen la sobreexpresión de las proteínas que inhiben la actividad de las quinasas del ciclo de células CDK, por ejemplo, mediante la introducción de un ADNc que codifica dicha proteína en la célula o mediante la activación de la expresión del gen que codifica la proteína en la célula. La detención del ciclo celular también se logra mediante la inhibición de la actividad de las ciclinas y las CDK, por ejemplo, usando métodos de ARNi (por ejemplo, patente de EE.UU. N° 6.506.559) o mediante la inhibición de la expresión de uno o más genes implicados en la progresión del ciclo celular tales como, por ejemplo, ciclina y/o CDK.

La integración dirigida dependiente de la homología, como se ha descrito anteriormente, se puede usar, por ejemplo, con fines de diseño por ingeniería genética de células y/o sobreexpresión de la proteína o para sustituir una secuencia de tipo silvestre con una secuencia mutante (o viceversa).

5 Mutagénesis dirigida

Para la mutagénesis dirigida, se puede usar cualquiera de los métodos desvelados en el presente documento mediante, por ejemplo, la inserción de una secuencia en un gen con el fin de interrumpir el gen, la introducción de una eliminación, la introducción de una mutación puntual o la sustitución de un gen con un alelo no funcional. Dicha mutagénesis dirigida se puede usar para una serie de fines. Por ejemplo, la mutagénesis dirigida de genes que codifican receptores virales (por ejemplo, los receptores CCR5 y CXCR4 para el VIH) se puede usar para volver a los receptores incapaces de unirse a virus, evitando de este modo una nueva infección y bloqueando la propagación de infecciones existentes. Los ejemplos no limitantes de virus o receptores virales que pueden servir de diana incluyen virus del herpes simple (VHS) tales como VHS-1 y VHS-2, virus de la varicela zoster (VVZ), virus de Epstein-Barr (VEB) y citomegalovirus (CMV), HHV6 y HHV7. La familia de los virus de la hepatitis incluye virus de la hepatitis A (VHA), virus de la hepatitis B (VHB), virus de la hepatitis C (VHC), virus de la hepatitis delta (VHD), virus de la hepatitis E (VHE) y virus de la hepatitis G (VHG). Hay otros virus o sus receptores que pueden servir de diana, incluyendo, pero sin limitación, *Picornaviridae* (por ejemplo, poliovirus, etc.); *Caliciviridae*; *Togaviridae* (por ejemplo, virus de la rubéola, virus del dengue, etc.); *Flaviviridae*; *Coronaviridae*; *Reoviridae*; *Bimaviridae*; *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 gripales de tipo A, B y C, etc.); *Bunyaviridae*; *Arenaviridae*; *Retroviridae*; lentivirus (por ejemplo, HTLV-I; HTLV-II; VIH-1 (también conocido como HTLV-III, LAV, ARV, hTLR, etc.) HFV-II); virus de la inmunodeficiencia simia (VIS), virus del papiloma humano (VPH) 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.

De manera similar, se puede someter a mutagénesis el genoma de una bacteria infecciosa mediante uno o más de los métodos desvelados en el presente documento, para bloquear o mejorar las infecciones bacterianas.

La escisión dirigida y la recombinación dirigida del ADN, como se desvela en el presente documento, se pueden usar para modificar secuencias no codificantes (por ejemplo, secuencias reguladoras tales como promotores, potenciadores, iniciadores, terminadores, sitios de corte y empalme) con el fin de modificar los niveles de expresión de un producto génico. Dichos métodos se pueden usar, por ejemplo, con fines terapéuticos, de genómica funcional y/o estudios de validación de dianas.

En realizaciones adicionales que utilizan las composiciones y los métodos descritos en el presente documento, las proteínas HLA de codificación de genes implicadas en el rechazo de los injertos se pueden escindir, someter a mutagénesis o modificar mediante recombinación, bien en las secuencias de codificación o regulación, de modo que se bloquee su expresión o que expresen un producto no funcional. Por ejemplo, mediante la inactivación del gen que codifica el gen de la subunidad β común (β 2-microglobulina), se pueden generar células madre nulas en HLA de clase I a partir de cualquier donante, reduciendo así la necesidad de aparear estrechamente haplotipos MHC de donante/receptor durante el injerto de células madre.

45 Enfermedades genéticas

Los métodos desvelados para la recombinación dirigida (tanto dependiente como independiente de la homología) se pueden usar para reemplazar cualquier secuencia genómica por una secuencia homóloga no idéntica. Por ejemplo, una secuencia genómica mutante se puede reemplazar por su homólogo de tipo silvestre, proporcionando así métodos para el tratamiento de, por ejemplo, enfermedades genéticas, trastornos hereditarios, cáncer y enfermedades autoinmunes. De manera similar, se puede reemplazar un alelo de un gen por un alelo diferente, usando los métodos de recombinación dirigida desvelados en el presente documento.

Las enfermedades genéticas ilustrativas incluyen, pero sin limitación, acondroplasia, acromatopsia, deficiencia de maltasa ácida, inmunodeficiencias adquiridas, deficiencia de adenosina desaminasa (OMIM N° 102700), adrenoleucodistrofia, síndrome de Aicardi, deficiencia de antitripsina α -I, α -talasemia, síndrome de insensibilidad andrógena, síndrome de Apert, ventrículo derecho arritmogénico, displasia, telangiectasia ataxia, síndrome de Barth, β -talasemia, Síndrome de Bean, enfermedad de Canavan, enfermedades granulomatosas crónicas (CGD), síndrome del maullido de gato, fibrosis quística, enfermedad de Dercum, displasia ectodérmica, anemia de Fanconi, fibrodisplasia osificante progresiva, síndrome X frágil, galactosemia, enfermedad de Gaucher, gangliosidosis generalizadas (por ejemplo, GM1), hemocromatosis, hemoglobinopatías (por ejemplo, anemia de células falciformes, la mutación de la hemoglobina C en el 6º codón de β -globina, α -talasemia, β -talasemia), hemofilia, enfermedad de Huntington, síndrome de Hurler, hipofosfatasa, síndrome Klinefelter, enfermedad de Krabbes, síndrome de Langer-Giedion, deficiencia de adhesión de leucocitos (LAD, OMIM N° 116920), leucodistrofia, síndrome de QT largo, enfermedades de almacenamiento lisosomal (por ejemplo, enfermedad de Gaucher, GM1, enfermedad de Fabry y enfermedad de Tay-Sachs), síndrome de Marfan, síndrome de Moebius, mucopolisacaridosis (por ejemplo,

enfermedad de Hunter, enfermedad de Hurler), síndrome de ña-rótula, 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 San Filippo, inmunodeficiencia combinada severa (SCID), síndrome de Shwachman, enfermedad de células falciformes (anemia de células falciformes), síndrome de Smith-Magenis, síndrome de Stickler, enfermedad de Tay-Sachs, síndrome de trombocitopenia con radios ausentes (TAR), síndrome de Treacher Collins, trisomía, esclerosis tuberosa, síndrome de Turner, trastorno del ciclo de la urea, enfermedad de von Hippel-Landau, síndrome de Waardenburg, síndrome de Williams, enfermedad de Wilson, síndrome de Wiskott-Aldrich, síndrome linfoproliferativo ligado al cromosoma X (XLP, OMIM N° 308240).

En muchos de estos casos, una región de interés comprende una mutación, y el ácido nucleico exógeno o donante comprende la secuencia de tipo silvestre correspondiente. Del mismo modo, una secuencia genómica de tipo silvestre puede reemplazarse por una secuencia mutante, si así se desea. Por ejemplo, la sobreexpresión de un oncogén se puede invertir, bien mediante la mutación del gen o mediante la sustitución de sus secuencias de control con secuencias que soportan un nivel no patológico inferior de expresión. Como otro ejemplo, el alelo de tipo silvestre del gen ApoA1 se puede reemplazar por el alelo ApoA1 Milano, para tratar la aterosclerosis. De hecho, se puede corregir o aliviar cualquier patología dependiente de una determinada secuencia genómica, de cualquier manera, usando los métodos y las composiciones desvelados en el presente documento.

En ciertos casos, se desea la modificación de una secuencia genómica de una célula pluripotente (por ejemplo, una célula madre hematopoyética). Los métodos para la movilización, el enriquecimiento y el cultivo de células madre hematopoyéticas son conocidos en la técnica. Véanse, por ejemplo, las patentes de EE.UU. N° 5.061.620; 5.681.559; 6.335.195; 6.645.489 y 6.667.064. Las células madre tratadas se pueden devolver a un paciente para el tratamiento de diversas enfermedades incluyendo, pero sin limitación, SCID y anemia de células falciformes.

Del mismo modo, los genomas de células madre pluripotentes inducidas (células iPS) también se pueden modificar de acuerdo con los métodos y las composiciones desvelados. Las células madre pluripotentes inducidas se describen, por ejemplo, en Yu *et al.* (2007) *Science* 318: 1917-1920 y Dimos *et al.* (2008) *Science* 321:1218-1221.

Por consiguiente, las realizaciones de la divulgación incluyen la dirección de una molécula de fusión proteica, una sonda y un ácido nucleico exógeno a un animal (incluyendo mamíferos humanos o no humanos, y vertebrados) para tratar al animal. El ácido nucleico exógeno expresa una proteína que proporciona un efecto terapéutico. En otros casos, se elimina la base genética infractora para la enfermedad. El método se puede realizar sin un vector, por ejemplo, sin un virus y sin un transposón.

Animales transgénicos

Los métodos y las composiciones desvelados se pueden usar para la generación de ganado y de grandes mamíferos transgénicos, como se desvela, por ejemplo, en la patente de EE.UU. N° 7.199.218 y el documento US N° de serie 12/504.364, presentado el 16 de julio de 2009 (publicación de EE.UU. N° 2010/0146655, "Methods And Materials For Producing Transgenic Animals").

Se pueden crear artiodáctilos transgénicos (por ejemplo, cerdos, ovejas, cabras y vacas). Las células nucleadas de los artiodáctilos transgénicos proporcionados en el presente documento contienen una construcción de ácido nucleico descrita en el presente documento. Como se usa en el presente documento, "artiodáctilo transgénico" incluye artiodáctilos transgénicos fundadores, así como la progenie de los fundadores, la progenie de la progenie, y así sucesivamente, siempre que la progenie conserve la construcción de ácido nucleico. Por ejemplo, se puede usar un animal transgénico fundador para criar animales adicionales que contengan la construcción de ácido nucleico.

Los tejidos obtenidos a partir de los artiodáctilos transgénicos (por ejemplo, cerdos transgénicos) y las células derivadas de los artiodáctilos transgénicos (por ejemplo, cerdos transgénicos) también se proporcionan en el presente documento. Como se usa en el presente documento, "derivado de" indica que las células se pueden aislar directamente del animal o pueden ser progenie de dichas células. Por ejemplo, de un artiodáctilo transgénico, se pueden obtener cerebro, pulmón, hígado, páncreas, corazón y válvulas cardíacas músculo, riñón, tiroides, córnea, piel, vasos sanguíneos u otro tejido conjuntivo (por ejemplo, de cerdo transgénico). También se pueden obtener de artiodáctilos transgénicos, por ejemplo, sangre y células hematopoyéticas, islotes de Langerhans, células β , células cerebrales, hepatocitos, células renales y células de otros órganos y fluidos corporales (por ejemplo, de cerdos transgénicos). Los órganos y las células de cerdos transgénicos se pueden trasplantar a un paciente humano. Por ejemplo, los islotes de cerdos transgénicos se pueden trasplantar a pacientes diabéticos humanos.

Se pueden usar diversas técnicas conocidas en la materia para introducir construcciones de ácidos nucleicos en animales no humanos para producir líneas fundadoras, en las que la construcción de ácido nucleico esté integrada en el genoma. Dichas técnicas incluyen, sin limitación, microinyección pronuclear (patente de EE.UU. N° 4.873.191), transferencia génica mediada por retrovirus en líneas germinales (Van der Putten *et al.* (1985) *Proc. Natl. Acad. Sci.* EE.UU. 82, 6148-1652), dirección génica hacia células madre embrionarias (Thompson *et al.* (1989) *Cell* 56, 313-321), electroporación de embriones (Lo (1983). *Mol Cell. Biol.* 3, 1803-1814), transferencia de genes mediada por

espermatozoides (Lavitrano *et al.* (2002) *Proc. Natl Acad. Sci EE.UU.* 99, 14230-14235; Lavitrano *et al.* (2006) *Reprod. Fert. Develop.* 18, 19-23) y transformación *in vitro* de células somáticas, tales como cúmulo o células mamarias, o células madre embrionarias, de adulto o fetales, seguida del trasplante nuclear (Wilmot *et al.* (1997) *Nature* 385, 810-813; y Wakayama *et al.* (1998) *Nature* 394, 369-374). La microinyección pronuclear, la transferencia de genes mediada por espermatozoides y la transferencia nuclear de células somáticas son técnicas especialmente útiles.

Por lo general, en la microinyección pronuclear, se introduce una construcción de ácido nucleico descrita en el presente documento en un óvulo fertilizado; se usan 1 o 2 óvulos fertilizados con células cuando los pronúcleos que contienen el material genético de la cabeza del espermatozoide y el óvulo son visibles dentro del protoplasma. Los óvulos fertilizados en fase pronuclear se pueden obtener *in vitro* o *in vivo* (es decir, extraerse quirúrgicamente del oviducto de los animales donantes). Los óvulos fertilizados *in vitro* se pueden producir de la siguiente manera. Por ejemplo, se pueden recoger ovarios de cerdo de un matadero y mantenerlos a 22-28 °C durante el transporte. Los ovarios se pueden lavar y aislar para la aspiración folicular, y se pueden aspirar los folículos que varían de 4 a 8 mm en tubos de centrifugación cónicos de 50 ml usando agujas de calibre 18 y al vacío. Se pueden aclarar el líquido folicular y los ovocitos aspirados a través de prefiltros con TL-HEPES comercial (Minitube, Verona, WI). Se pueden seleccionar los ovocitos rodeados por una masa de cúmulo compacta y colocarlos en medio de maduración de ovocitos TCM-199 (Minitube, Verona, WI) suplementado con 0,1 mg/ml de cisteína, 10 ng/ml de factor de crecimiento epidérmico, fluido folicular porcino al 10 %, 2-mercaptoetanol 50 µM, 0,5 mg/ml de AMPc, 10 UI/ml de gonadotropina de suero de yegua preñada (PMSG) y gonadotropina coriónica humana (hCG) durante aproximadamente 22 horas en aire humidificado a 38,7 °C y CO₂ al 5 %. Posteriormente, los ovocitos se pueden traspasar a medio de maduración TCM-199 recién preparado que no contendrá AMPc, PMSG ni hCG, e incubarse durante un período adicional de 22 horas. Los ovocitos madurados se pueden separar de sus células del cúmulo por agitación en hialuronidasa al 0,1 % durante 1 minuto.

Los ovocitos maduros se pueden fertilizar en 500 µl del sistema de medio MINITUBE PORCPRO IVF (Minitube, Verona, WI) en placas de fertilización de 5 pocillos Minitube. En la preparación de la fertilización *in vitro* (IVF), se puede lavar semen porcino recién recogido o congelado y volverlo a suspender en medio PORCPRO IVF a 4 x 10⁵ espermatozoides. Las concentraciones de espermatozoides se pueden analizar mediante análisis del semen asistido por ordenador (SPERMVISION, Minitube, Verona, WI). La inseminación *in vitro* final se puede realizar en un volumen de 10 µl a una concentración final de aproximadamente 40 espermatozoides móviles/ovocito, dependiendo del cerdo. Se incuban todos los ovocitos fertilizantes a 38,7 °C en atmósfera de CO₂ al 5,0 % durante 6 horas. Seis horas después de la inseminación, los supuestos cigotos se pueden lavar dos veces en NCSU-23 y trasladar a 0,5 ml del mismo medio. Este sistema puede producir del 20 al 30 % de blastocitos de forma rutinaria en la mayoría de los cerdos con una tasa de inseminación poliespérmica del 10-30 %.

Se pueden inyectar construcciones de ácido nucleico linealizadas en uno de los pronúcleos, entonces los óvulos inyectados se pueden transferir a una hembra receptora (por ejemplo, en los oviductos de una hembra receptora) y permitir el desarrollo en la hembra receptora para producir los animales transgénicos. En particular, los embriones fertilizados *in vitro* se pueden centrifugar a 15.000 X g durante 5 minutos para sedimentar los lípidos permitiendo la visualización del pronúcleo. Los embriones se pueden inyectar con aproximadamente 5 picolitros del cóctel de transposón/transposasa usando un inyector Eppendorf FEMTOJET, y se pueden cultivar hasta la formación de blastocistos (~144 horas) en medio NCSU 23 (véase, por ejemplo, el documento WO/2006/036975). Se pueden registrar las tasas de escisión de embriones, y la formación y la calidad de los blastocistos.

Los embriones se pueden transferir quirúrgicamente a úteros de receptoras asíncronas. Para la transferencia quirúrgica de los embriones, se puede inducir la anestesia con una combinación de lo siguiente: ketamina (2 mg/kg); tiletamina/zolazepam (0,25 mg/kg); xilazina (1 mg/kg); y atropina (0,03 mg/kg) (todas de Columbus Serum). En decúbito dorsal, se pueden preparar las receptoras asépticamente para la cirugía y realizar una incisión ventral caudal para dejar al descubierto y examinar el tracto reproductivo. Por lo general, se pueden depositar 100-200 (por ejemplo, 150-200) embriones en la unión entre la ampolla y el istmo del oviducto usando un catéter TOMCAT[®] de aproximadamente 14 cm (5,5 pulgadas). Tras la cirugía, se puede realizar una ecografía en tiempo real del embarazo usando un escáner de ultrasonidos ALOKA 900 (Aloka Co. Ltd, Wallingford, CT) con una sonda transabdominal de 3,5 MHz conectada. La monitorización del inicio del embarazo puede comenzar a los 23 días de la fusión y se puede repetir semanalmente durante el embarazo. La cría de las receptoras se puede mantener como cerdas gestantes normales.

En la transferencia nuclear de células somáticas, se puede introducir una célula de artiodáctilo transgénico (por ejemplo, una célula de cerdo transgénico) tal como un blastómero embrionario, fibroblasto fetal, fibroblasto de oreja de adulto o célula de la granulosa que incluya una construcción de ácido nucleico descrita anteriormente en un ovocito enucleado para establecer una célula combinada. Los ovocitos se pueden enucleo mediante la disección de la zona parcial cercana al cuerpo polar y luego presionando el citoplasma en la zona de disección. Por lo general, se usa una pipeta de inyección con una punta afilada y biselada para inyectar la célula transgénica en un ovocito enucleado detenido en la meiosis 2. En algunos convenios, los ovocitos detenidos en la meiosis 2 se denominan "óvulos". Después de producir un embrión porcino (por ejemplo, mediante la fusión y la activación del ovocito), el

embrión porcino se transfiere a los oviductos de una hembra receptora, aproximadamente 20 a 24 horas después de la activación. Véase, por ejemplo, Cibelli *et al.* (1998) *Science* 280, 1256-1258 y patente de EE.UU. Nº 6.548.741. Para los cerdos, se puede comprobar el embarazo de las hembras receptoras aproximadamente de 20 a 21 días después de la transferencia de los embriones.

5 Se pueden usar técnicas de reproducción convencionales para crear animales que sean homocigóticos para el ácido nucleico diana de los animales fundadores iniciales heterocigotos. No obstante, la homocigosidad puede no ser necesaria. Los cerdos transgénicos descritos en el presente documento se pueden criar con otros cerdos de interés.

10 Una vez generados los animales transgénicos, se puede evaluar la expresión de un ácido nucleico diana usando técnicas convencionales. La selección inicial se puede realizar mediante análisis de transferencia Southern para determinar si ha tenido lugar o no la integración de la construcción. Para consultar una descripción del análisis Southern, véanse los apartados 9.37-9.52 de Sambrook *et al.*, 1989, "Molecular Cloning, A Laboratory Manual", segunda edición, Cold Spring Harbor Press, Plainview, NY. También se pueden usar técnicas de reacción en cadena de la polimerasa (PCR) en la selección inicial. PCR se refiere a un procedimiento o una técnica en la que
15 ácidos nucleicos diana se amplifican. En general, la información de las secuencias de los extremos de la región de interés o de más allá se emplea para diseñar cebadores de oligonucleótidos que sean idénticos o similares en secuencia a cadenas opuestas del molde por amplificar. La PCR se puede usar para amplificar secuencias específicas de ADN, así como de ARN, incluyendo secuencias de ADN genómico total o ARN celular total. Los
20 cebadores normalmente son de 14 a 40 nucleótidos de longitud, pero pueden variar de 10 nucleótidos a cientos de nucleótidos de longitud. La PCR se describe, por ejemplo, en "PCR Primer: A Laboratory Manual", ed. Dieffenbach y Dveksler, Cold Spring Harbor Laboratory Press, 1995. Los ácidos nucleicos también se pueden amplificar mediante la reacción en cadena de la ligasa, la amplificación por desplazamiento de cadena, la replicación de secuencia auto-sostenida o la amplificación basada en secuencias de ácido nucleico. Véase, por ejemplo, Lewis (1992) *Genetic
25 Engineering News* 12,1; Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci EE.UU.* 87, 1874-1878; y Weiss (1991) *Science* 254, 1292-1293. En la etapa de blastocisto, los embriones se pueden procesar individualmente para el análisis por PCR, la hibridación Southern y la PCR splinkerette (véase, por ejemplo, Dupuy *et al Proc Natl Acad Sci EE.UU.* (2002) 99 (7):4495-4499).

30 La expresión de una secuencia de ácido nucleico que codifica un polipéptido en los tejidos de cerdos transgénicos se puede evaluar usando técnicas que incluyen, sin limitación, análisis de transferencia Northern de muestras de tejido obtenidas del animal, análisis de hibridación *in situ*, análisis Western, inmunoensayos tales como ensayos de inmunoabsorción ligados a una enzima y PCR de transcriptasa inversa (RT-PCR).

35 **Administración**

Las moléculas de fusión, las nucleoproteínas y/o los ácidos nucleicos desvelados en el presente documento se pueden administrar directamente a un sujeto para aplicaciones terapéuticas o profilácticas tales como las descritas en el presente documento. Los sujetos pueden ser animales o plantas. En particular, las plantas pueden ser
40 monocotiledóneas o dicotiledóneas. Los animales pueden ser vertebrados, en particular, mamíferos (por ejemplo, ganado, animales domésticos), en particular, primates, en particular, seres humanos.

En general, y en vista de la descripción del presente documento, la referencia a la introducción de una proteína de fusión en un sujeto puede significar bien que se introduce la propia proteína de fusión o que se introduce un ácido
45 nucleico que codifica una proteína de fusión en una forma que se pueda expresar en el sujeto.

Con respecto a la introducción de ácidos nucleicos y filamentos de nucleoproteína en las células, también se puede usar cualquiera de los procedimientos bien conocidos para introducir ácidos nucleicos en células para la introducción de filamentos de nucleoproteína. Por ejemplo, los métodos de administración no viral de ácidos nucleicos y
50 filamentos de nucleoproteína incluyen, pero sin limitación, electroporación, lipofección, microinyección, biolística, virosomas, liposomas, inmunoliposomas, conjugados policatiónicos o conjugados de lípido:ácido nucleico, polibreno, fusión de protoplastos, transfección mediada por fosfato de calcio, transfección mediada por dextrano de DEAE, ADN desnudo, viriones artificiales, y absorción de ADN potenciada por un agente. La lipofección se describe en, por ejemplo, la patente de EE.UU. Nº 5.049.386, 4.946.787; y 4.897.355), habiendo reactivos de lipofección disponibles en el mercado (por ejemplo, Transfectam™, Lipofectamine® y lipofectina™). Los lípidos catiónicos y neutros que son
55 adecuados para la lipofección eficaz de reconocimiento de receptores de polinucleótidos incluyen los de Felgner, y de los documentos WO 91/17424 y WO 91/16024. La administración puede ser a células (administración *in vitro* o *ex vivo*) o a tejidos diana (administración *in vivo*). Véase también Sambrook *et al.*, *supra* y Ausubel *et al.*, *supra*.

60 En el caso de la introducción de una proteína de fusión y de un ácido nucleico exógeno, el ácido nucleico exógeno puede estar presente en exceso molar, medido por los moles de proteína de fusión y los moles de cadenas de ADN de ácido nucleico exógeno. Por ejemplo, el fragmento de ADN exógeno puede estar presente en una concentración molar que supere la concentración molar de la proteína de fusión, siendo el exceso opcionalmente de al menos el doble o de entre aproximadamente el doble y 500 veces. Los expertos en la materia apreciarán de inmediato que se
65 contemplan todos los intervalos y los valores de entre los valores indicados explícitamente, por ejemplo, 10 veces o de aproximadamente 5 veces a aproximadamente 50 veces.

Vectores de expresión

Los ácidos nucleicos se pueden clonar en varios tipos de vectores para la transformación en células procariotas o eucariotas para la replicación y/o la expresión, como se conoce en la técnica.

5 Para obtener la expresión de, por ejemplo, una proteína de fusión, se pueden insertar un ácido nucleico que codifique la proteína de fusión en un vector de expresión que contenga un promotor para dirigir la transcripción. Los 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 Ausubel *et al.*, *supra*. Hay sistemas de expresión bacterianos disponibles, por ejemplo, en *E. coli*, *Bacillus* sp., y *Salmonella* (Palva *et al.* (1983) *Gene* 22:229-235). Hay kits para dichos sistemas de expresión disponibles en el mercado. Los sistemas de expresión eucariotas para células de mamíferos, levaduras y células de insecto son muy conocidos en la técnica, y también se encuentran disponibles en el mercado.

15 Los ácidos nucleicos se pueden incorporar en vectores. Lo más frecuente es que los vectores contengan uno o más casetes de expresión que comprendan una o más secuencias de control de la expresión, en los que una secuencia de control de la expresión es una secuencia de ADN que controla y regula la transcripción y/o la traducción de otra secuencia de ADN o ARNm, respectivamente. Las secuencias de control de la expresión incluyen, por ejemplo, secuencias promotoras, elementos potenciadores de la transcripción, codones de inicio, codones de terminación, y cualquier otro elemento de ácido nucleico necesario para la unión, el inicio o la terminación de la transcripción por ARN polimerasas. En la técnica, se conoce bien una amplia selección de secuencias de control de la expresión, y se encuentra disponible en el mercado. Por lo tanto, una unidad transcripcional de un vector puede comprender una secuencia de control de la expresión unida operativamente a una secuencia de ácido nucleico exógena. Por ejemplo, una secuencia de ADN está unida operativamente a una secuencia de control de la expresión, tal como un promotor, cuando la secuencia de control de la expresión controla y regula la transcripción y la traducción de esa secuencia de ADN. Los ejemplos de vectores incluyen: plásmidos (que también pueden ser un vehículo de otro tipo de vector), adenovirus, virus adeno-asociados (AAV), lentivirus (por ejemplo, VIH-1, VIS o VIF modificados), retrovirus (por ejemplo, ASV, ALV o MoMLV) y transposones (por ejemplo, Sleeping Beauty, P-elementos, Tol-2, Frog Prince, piggyBac).

Administración, vehículos, composiciones farmacéuticas

35 La administración de cantidades terapéuticamente eficaces de las composiciones desveladas en el presente documento se realiza por cualquiera de las vías usadas normalmente para la introducción de macromoléculas en contacto final con el tejido que se vaya a tratar. Las moléculas de fusión, o sus ácidos nucleicos codificantes, o filamentos de nucleoproteína, se administran de cualquier manera adecuada, opcionalmente, en formulación con vehículos farmacéuticamente aceptables. Se pueden administrar múltiples composiciones simultáneamente o por separado por la misma vía o por diferentes vías. Los métodos adecuados para la administración de dichas composiciones están disponibles, y son bien conocidos por los expertos en la materia y, aunque se puede usar más de una vía para administrar una determinada composición, una vía en particular a menudo puede proporcionar una reacción más inmediata y más eficaz que otra vía.

45 Las composiciones farmacéuticas se determinan, en parte, por la sustancia que se administra en particular, así como por el método usado en particular para administrar la sustancia. Por consiguiente, hay una amplia variedad de formulaciones adecuadas de composiciones farmacéuticas. Véase, por ejemplo, "Remington's Pharmaceutical Sciences", 17ª ed. 1985; Brunton *et al.*, "Goodman and Gilman's The Pharmacological Basis of Therapeutics", McGraw-Hill, 2005; University of the Sciences in Philadelphia (eds.), "Remington: The Science and Practice of Pharmacy", Lippincott Williams & Wilkins, 2005; y University of the Sciences in Philadelphia (eds.), "Remington: The Principles of Pharmacy Practice", Lippincott Williams & Wilkins, 2008.

55 Las composiciones farmacéuticas de la presente divulgación se pueden preparar en formulaciones en aerosol (es decir, se pueden "nebulizar") para su administración por inhalación. Las formulaciones en aerosol se pueden disponer en propulsores presurizados aceptables tales como diclorodifluorometano, propano, nitrógeno, y similares.

60 Las formulaciones adecuadas para la administración parenteral, tales como, por ejemplo, las vías intravenosa, intramuscular, intradérmica y subcutánea, incluyen soluciones inyectables estériles isotónicas, , acuosas y no acuosas, que pueden contener antioxidantes, tampones, bacteriostáticos y solutos que vuelven la formulación isotónica con la sangre del receptor final, y suspensiones estériles, acuosas y no acuosas, que pueden incluir agentes de suspensión, solubilizantes, agentes espesantes, estabilizantes y conservantes. En la práctica de los métodos desvelados, las composiciones se pueden administrar, por ejemplo, por infusión intravenosa, u oral, tópica, intraperitoneal, intravesical o intratecalmente. Las formulaciones se pueden presentar en recipientes monodosis o multidosis sellados, tales como ampollas y viales. La administración se puede realizar a través de dosis únicas o divididas. Las suspensiones y soluciones inyectables se pueden preparar a partir liofilizados estériles, polvos, gránulos y comprimidos.

65

Las dosis apropiadas dependerán del efecto deseado, del tamaño y/o del peso del sujeto, y del estado de salud general del sujeto, y se pueden determinar mediante el aumento de la dosis en varias sesiones de tratamiento. El tamaño de la dosis también puede verse influido por la existencia, la naturaleza y el grado de cualquiera de los efectos secundarios adversos que acompañen a la administración.

5 Las composiciones terapéuticas desveladas pueden incluir materiales, composiciones o vehículos farmacéuticamente aceptables, tales como una carga líquida o sólida, diluyente, excipiente, disolvente o material de encapsulación, es decir, vehículos. Estos vehículos participan en el transporte del producto químico en cuestión desde un órgano, o una zona del cuerpo, a otro órgano u otra zona del cuerpo. Cada vehículo debería ser
10 "aceptable" en el sentido de ser compatible con el resto de ingredientes de la formulación y no perjudicial para el paciente. Algunos ejemplos de materiales que pueden servir como vehículos farmacéuticamente aceptables incluyen: azúcares, tales como lactosa, glucosa y sacarosa; almidones, tales como almidón de maíz y almidón de patata; celulosa y sus derivados, tales como carboximetilcelulosa sódica, etilcelulosa y acetato de celulosa; tragacanto en polvo; malta; gelatina; talco; excipientes tales como manteca de cacao y ceras para supositorios;
15 aceites, tales como aceite de cacahuete, aceite de semilla de algodón, aceite de cártamo, aceite de sésamo, aceite de oliva, aceite de maíz y aceite de soja; glicoles, tales como propilenglicol; polioles, tales como glicerina, sorbitol, manitol y polietilenglicol; ésteres, tales como oleato de etilo y laurato de etilo; agar; agentes de tamponamiento, tales como hidróxido de magnesio e hidróxido de aluminio; ácido alginico; agua libre de pirógenos; solución salina isotónica; solución de Ringer; alcohol etílico; soluciones tampón de fosfato; y otras sustancias compatibles no tóxicas
20 empleadas en formulaciones farmacéuticas. Los agentes humectantes, los emulsionantes y los lubricantes, tales como laurilsulfato sódico y estearato de magnesio, así como los agentes colorantes, agentes de liberación, agentes de recubrimiento, edulcorantes, aromatizantes y perfumantes, conservantes y antioxidantes también pueden estar presentes en las composiciones terapéuticas.

25 Administración a plantas

Para la administración de polinucleótidos y nucleoproteínas a células vegetales, los ácidos nucleicos se pueden clonar en vectores intermedios para la transformación en células procariotas o eucariotas (por ejemplo, plantas) para la replicación y/o la expresión. Los vectores intermedios para el almacenamiento o la manipulación del ácido
30 nucleico o la producción de la proteína pueden ser vectores procariotas, (por ejemplo, plásmidos), vectores lanzadera, vectores de insectos o vectores virales, por ejemplo. Los ácidos nucleicos también se pueden clonar en un vector de expresión, para la administración a una célula bacteriana, célula fúngica, célula de protozoo o célula vegetal.

35 Los vectores de expresión vegetales y los genes indicadores se conocen en la técnica en general. Véase, por ejemplo, Gruber *et al.* (1993) en "Methods of Plant Molecular Biology and Biotechnology", Bernard R. Glick y John E. Thompson, eds., CRC Press, Boca Ratón, FL. Dichos sistemas incluyen técnicas de ADN recombinante *in vitro* e *in vivo*, y pueden utilizar cualquier otro método de recombinación sintético o natural. Véase, por ejemplo, "Transgenic Plants: A Production System for Industrial and Pharmaceutical Proteins", Owen y Pen eds., John Wiley & Sons,
40 1996; "Transgenic Plants", Galun and Breiman eds., Imperial College Press, 1997; y "Applied Plant Biotechnology", Chopra, Malik, y Bhat eds., Science Publishers, Inc., 1999.

El promotor usado para dirigir la expresión del ácido nucleico de elección depende de la aplicación en particular. Por ejemplo, normalmente se usa un potente promotor constitutivo para la expresión y la purificación. Por el contrario,
45 cuando una proteína se va a usar *in vivo*, se puede usar un promotor constitutivo o un promotor inducible, dependiendo de la función concreta de la proteína codificada. Además, se puede usar un promotor débil cuando se requieran niveles de proteína bajos, pero sostenidos. Por lo general, el promotor también puede incluir elementos que sean sensibles a la transactivación, por ejemplo, elementos de respuesta a la hipoxia y sistemas de control de moléculas pequeñas tales como sistemas regulados por tet y el sistema RU-486. Véase, por ejemplo, Gossen *et al.* (1992) *Proc. Natl. Acad. Sci. EE.UU.* 89:5547-5551; Oligino *et al.* (1998) *Gene Ther.* 5:491-496; Wang *et al.* (1997) *Gene Ther.* 4:432-441; Neering *et al.* (1996) *Blood* 88:1147-1155; y Rendahl *et al.* (1998) *Nature Biotechnol* 16:757-761.

Los promotores adecuados para su uso en sistemas de expresión de plantas incluyen, pero sin limitación,
55 promotores virales tales como promotores de ARN 35S y promotores de ARN 19S del virus del mosaico de la coliflor (CaMV) (Brisson *et al.*, (1984) *Nature* 310:511-514) y el promotor de la proteína de la cubierta del virus del mosaico del tabaco (TMV) (Takamatsu *et al.* (1987) *EMBO J.* 6:307-311); promotores vegetales tales como el promotor para el gen que codifica la subunidad pequeña de la ribulosa-1,5-bis-fosfato carboxilasa (RUBISCO) (Coruzzi *et al.* (1984) *EMBO J.* 3:1671-1680; Broglie *et al.* (1984) *Science* 224:838-843; y promotores de choque térmico vegetales, por ejemplo, hsp17.5-E o hsp17.3-B de soja (Gurley *et al.*, (1986) *Cell Biol* 6:559-565). Otros ejemplos de promotores
60 que se pueden usar para la expresión en células vegetales incluyen promotores de plásmidos inductores de tumores de *Agrobacterium tumefaciens*, tales como promotores de nopalina sintasa (NOS) y octopina sintasa; promotores de ADN-T bacterianos tales como promotores mas y ocs; o el promotor 35S del virus del mosaico de la escrofularia.

65 En ciertas realizaciones, se usa el promotor 35S del virus del mosaico de la coliflor (CaMV). La familia de los caulimovirus ha proporcionado una serie de promotores ilustrativos para la expresión transgénica en las plantas, en

particular, el promotor 35S (CaMV). Véase, por ejemplo, Kay *et al.* (1987) *Science* 236:1299. En la técnica, se han descrito promotores adicionales de esta familia, tales como el promotor del virus del mosaico de la escrofularia, el promotor del virus del moteado amarillo de la comelina, y el promotor del virus baciliforme del arroz tungro, y también se pueden usar en los métodos y en las composiciones desvelados en el presente documento. Véase, por ejemplo, Sanger *et al.* (1990) *Plant Mol. Biol.* 14:433-443; Medberry *et al.* (1992) *Plant Cell* 4:195-192; Yin *et al.* (1995) *Plant J.* 7:969-980.

Si se desea, los promotores vegetales se pueden modificar para afectar a su capacidad de respuesta reguladora. Por ejemplo, el promotor 35S del CaMV se puede unir a la porción del gen RUBISCO que reprime la expresión de RUBISCO en ausencia de luz, para crear un promotor que sea activo en las hojas, pero no en las raíces. También se pueden usar los promotores vegetales constitutivos tales como la actina y la ubiquitina, que tienen propiedades de expresión generales conocidas en la técnica. Véase, por ejemplo, McElroy *et al.* (1990) *Plant Cell* 2:163-171; Christensen *et al.* (1992) *Plant Mol. Biol.* 18:675-689.

Además, dependiendo del tejido deseado, la expresión se puede dirigir al endospermo, a la capa de aleurona, al embrión (o sus partes, tales como el escutelo y los cotiledones), al pericarpio, al tallo, a las hojas, a los tubérculos, a las raíces, etc. Los ejemplos de promotores específicos de tejidos conocidos incluyen el promotor de la patatina de clase I dirigido al tubérculo, los promotores asociados con genes ADPGPP del tubérculo de la patata, el promotor de soja de β -conglucina (proteína 7S) que dirige la transcripción dirigida a las semillas, y los promotores dirigidos a las semillas de los genes de zeína del endospermo del maíz. Véase, por ejemplo, Bevan *et al.* (1986) *Nucleic Acids Res.* 14:4625-4638; Muller *et al.* (1990) *Mol. Gen. Genet.* 224:136-146; Bray (1987) *Planta* 172:364-370; y Pedersen *et al.* (1982) *Cell* 29:1015-1026. Otros promotores adicionales específicos de las semillas incluyen los promotores de faseolina y napina.

Las construcciones recombinantes también pueden incluir genes marcadores seleccionables o rastreables expresables en plantas para aislar, identificar o rastrear células vegetales transformadas por estas construcciones. Los marcadores seleccionables incluyen, pero sin limitación, genes que confieren resistencia a los antibióticos (por ejemplo, resistencia a la kanamicina o a la higromicina) o resistencia a herbicidas (por ejemplo, resistencia a la sulfonilurea, fosfotricina o glifosato). Los marcadores detectables incluyen, pero sin limitación, los genes que codifican la β -glucuronidasa (Jefferson (1987) *Plant Molec Biol. Rep.* 5:387-405), luciferasa (Ow *et al.*, (1986) *Science* 234: 856-859) y productos génicos B y C1 que regulan la producción de pigmento de antocianina (Goff *et al.* (1990) *EMBO J.* 9: 2517- 2522).

Otros elementos opcionalmente presentes en los vectores de expresión incluyen un replicón que funciona en *E. coli* (o en otra célula hospedadora vegetal o de insecto, procarionta), un marcador selectivo que funciona en un hospedador procarionta, por ejemplo, un gen que codifica la resistencia a los antibióticos, para permitir la selección de bacterias que alberguen plásmidos recombinantes, y sitios de restricción únicos en regiones no esenciales del vector con el fin de permitir la inserción de secuencias recombinantes.

Los sistemas de transformación para plantas se conocen en la técnica. Véase, por ejemplo, Weissbach y Weissbach, "Methods for Plant Molecular Biology", Academic Press, NY, Apartado VIII, pág. 421-463 (1988); y Grierson y Corey, "Plant Molecular Biology", 2ª Ed., Blackie, Londres, Cap. 7-9 (1988). Por ejemplo, a menudo, se emplea satisfactoriamente *Agrobacterium* para introducir ácidos nucleicos en plantas. Dicha transformación usa preferentemente vectores de ADN-T binarios de *Agrobacterium* que se pueden usar para transformar plantas dicotiledóneas, plantas monocotiledóneas y células vegetales. Bevan (1984) *Nuc. Acid Res.* 12:8711-8721; Horsch *et al.* (1985) *Science* 227:1229-1231; Bevan *et al.* (1982) *Ann. Rev. Genet.* 16:357-384; Rogers *et al.* (1986) *Methods Enzymol.* 118:627-641; y Hernalsteen *et al.* (1984) *EMBO J.* 3:3039-3041. En realizaciones que utilizan el sistema de *Agrobacterium* para la transformación de plantas, las construcciones de ADN recombinante normalmente comprenden al menos la secuencia del borde derecho de ADN-T que flanquea las secuencias de ADN que se van a transformar en la célula vegetal. En realizaciones preferidas, las secuencias que se van a transferir están flanqueadas por las secuencias de los bordes derecho e izquierdo de ADN-T. El diseño y la construcción de dichos vectores de transformación basados en ADN-T son bien conocidos para los expertos en la materia.

Otros métodos de transferencia y transformación génica incluyen, pero sin limitación, la transformación de protoplastos a través de la absorción mediada por calcio, polietilenglicol (PEG) o electroporación del ADN desnudo. (véase, por ejemplo, Paszlcowski *et al.* (1984) *EMBO J.* 3:2717-2722; Potrykus *et al.* (1985) *Molec. Gen. Genet.* 199:169-177; Fromm *et al.* (1985) *Proc. Nat. Acad. Sci. EE.UU.* 82:5824-5828; y Shimamoto (1989) *Nature* 338:274-276); electroporación de tejidos vegetales (por ejemplo, D'Halluin *et al.* (1992) *Plant Cell* 4:1495-1505); microinyección, absorción de ADN mediada por carburo de silicio (por ejemplo, Kaeppeler *et al.* (1990) *Plant Cell Reporter* 9: 415-418), bombardeo de microproyectiles (por ejemplo, Klein *et al.* (1988) *Proc. Nat. Acad. Sci. EE.UU.* 85:4305-4309; y Gordon-Kamm *et al.* (1990) *Plant Cell* 2: 603-618); transferencia directa de genes, transformación de protoplastos *in vitro*, transformación de plantas mediada por virus, transformación mediada por liposomas y aceleración de partículas balísticas (por ejemplo, Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722; patentes de EE.UU. N° 4.684.611; 4.407.956; 4.536.475; Crossway *et al.* (1986) *Biotechniques* 4:320-334; Riggs *et al.* (1986) *Proc. Nati. Acad. Sci. EE.UU.* 83:5602-5606; Hinchee *et al.* (1988) *Biotechnology* 6:915-921; y patente de EE.UU. N° 4.945.050).

Se puede usar una amplia variedad de células hospedadoras, plantas y sistemas de células vegetales, incluyendo, pero sin limitación, aquellas plantas monocotiledóneas y dicotiledóneas, tales como cultivos, incluyendo cultivos de cereales (por ejemplo, trigo, maíz, arroz, mijo, cebada), cultivos frutales (por ejemplo, tomate, manzana, pera, fresa, naranja), cultivos forrajeros (por ejemplo, alfalfa), cultivos de hortalizas de raíz (por ejemplo, zanahoria, patata, remolacha azucarera, ñame), cultivos de hortalizas de hoja verde (por ejemplo, lechuga, espinaca); plantas con flores (por ejemplo, petunia, rosa, crisantemo), coníferas y pinos (por ejemplo, pino, abeto, abeto falso); plantas usadas en la fitorremediación (por ejemplo, plantas que acumulan metales pesados); cultivos oleaginosos (por ejemplo, girasol, semillas de colza) y plantas usadas con fines experimentales (por ejemplo, *Arabidopsis*).

Las secuencias exógenas también se pueden expresar en las semillas (por ejemplo, semilla de canola, de maíz, de soja, de arroz y de cebada), usando técnicas de producción a base de semillas, y, si se desea, los productos de expresión se pueden recoger durante la germinación de las semillas. Véanse, por ejemplo, las publicaciones PCT N° WO 99/40210; WO 99/16890; WO 99/07206; la patente de EE.UU. N° 5.866.121; y la patente de EE.UU. N° 5.792.933; y todas las referencias citadas en las mismas.

En realizaciones adicionales, las moléculas de fusión (por ejemplo, las proteínas de fusión) se administran directamente a células vegetales diana (en lugar de la introducción de un ácido nucleico que codifique una proteína de fusión). En ciertas situaciones *in vitro*, las células diana se cultivan en un medio que contiene una molécula de fusión como la desvelada en el presente documento. Un factor importante en la administración de compuestos polipeptídicos a plantas es garantizar que el polipéptido tenga la capacidad de atravesar una pared celular. No obstante, las proteínas, los virus, las toxinas, los métodos balísticos y similares tienen la capacidad de trasladar polipéptidos a través de una pared celular vegetal.

Por ejemplo, "plasmodesmos" es el término dado para explicar el transporte de célula a célula de proteínas endógenas y virales, y complejos ribonucleoproteicos (RNPC) en plantas. Los ejemplos de virus que pueden unirse a una molécula de fusión para facilitar su absorción en las células vegetales incluyen, el virus del mosaico del tabaco (Oparka *et al.* (1997) *Plant J.* 12: 781-789); tiorredoxina del floema del arroz (Ishiwatari *et al.* (1998) *Planta* 205: 12-22); y el virus X de la patata (Cruz *et al.* (1998) *Plant Cell* 10: 495 a 510). También se pueden unir otros restos químicos adecuados que proporcionen una mayor absorción celular, ya sea de forma covalente o no covalente, a moléculas de fusión para facilitar la penetración de una célula vegetal. Las moléculas de toxina también tienen la capacidad de transportar polipéptidos a través de las paredes celulares.

También se pueden usar técnicas de administración mediada por partículas (por ejemplo, inyección balística) como se describe anteriormente con respecto a los ácidos nucleicos para introducir polipéptidos en una célula vegetal.

Ácidos nucleicos

Ciertas realizaciones se dirigen a los ácidos nucleicos. Como se usa en el presente documento, el término "ácido nucleico" se refiere tanto a ARN como a ADN, incluyendo ARNip, ARNhp, ARNim, ADNc, ADN genómico, ADN sintético (por ejemplo, sintetizado químicamente), así como de origen natural y ácidos nucleicos modificados químicamente, por ejemplo, bases sintéticas o cadenas principales alternativas. Una molécula de ácido nucleico puede ser bicatenaria o monocatenaria (es decir, una sola cadena sentido o antisentido). Un ácido nucleico aislado se refiere a un ácido nucleico que está separado de otras bases de ácido nucleico que están presentes en un genoma, incluyendo los ácidos nucleicos que normalmente flanquean uno o ambos lados de una secuencia de ácido nucleico de un genoma de vertebrado (por ejemplo, ácidos nucleicos que flanquean un gen). Un ácido nucleico sustituido de forma conservadora se refiere a la sustitución de un codón con otro codón que codifica el mismo aminoácido, y también se refiere a ácidos nucleicos que codifican aminoácidos sustituidos de forma conservadora, como se describe en el presente documento con respecto a los polipéptidos. Significativamente, la combinación de posibles codones para un polipéptido de solo aproximadamente seis restos es manejablemente pequeña.

Las secuencias de ácido nucleico expuestas en el presente documento pretenden representar secuencias tanto de ADN como de ARN, de acuerdo con la práctica convencional de permitir que la abreviatura "T" signifique "T" o "U", como puede ser el caso para el ADN o el ARN. Los polinucleótidos son moléculas de ácido nucleico de al menos tres subunidades de nucleótidos. Los análogos de polinucleótidos o ácidos polinucleicos son polinucleótidos o ácidos polinucleicos modificados químicamente. En algunas realizaciones, los análogos de polinucleótidos se pueden generar mediante la sustitución de partes de la cadena principal de azúcar-fosfato de un polinucleótido con grupos funcionales alternativos. Los polinucleótidos modificados con morfolino, denominados en el presente documento "morfolinos" son análogos de polinucleótidos en los que las bases están unidas por una cadena principal de morfolino-fosforodiamidato (véanse, por ejemplo, las patentes de EE.UU. N° 5.142.047 y 5.185.444). Además de los morfolinos, otros ejemplos de análogos de polinucleótidos incluyen análogos en los que las bases están unidas por una cadena principal de polivinilo, ácidos nucleicos peptídicos (PNA) en los que las bases están unidas por enlaces de amida formados por grupos 2-aminoetil-glicina pseudopeptídicos, análogos en los que las subunidades de nucleósidos están unidas por grupos metilfosfonato, análogos en los que los restos de fosfato que unen las subunidades de nucleósidos están reemplazados por grupos fosfoamidato y ADN fosfortiolados, análogos que contienen restos de azúcar que tienen grupo O-metilo 2'). Los polinucleótidos de la invención se pueden producir a través de la técnica bien conocida y usada habitualmente de síntesis en fase sólida. Por otra parte, se pueden usar

otros métodos adecuados para dicha síntesis (por ejemplo, la clonación molecular común y técnicas químicas de síntesis de ácidos nucleicos). También se pueden usar técnicas similares para preparar análogos de polinucleótidos tales como morfolinos o derivados de fosforotioato. Además, los polinucleótidos y los análogos de polinucleótidos se pueden obtener del mercado. Para los oligonucleótidos, los ejemplos de composiciones farmacéuticamente aceptables son sales que incluyen, por ejemplo, (a) sales formadas con cationes tales como sodio, potasio, amonio, etc.; (b) sales de adición de ácido formadas con ácidos inorgánicos, por ejemplo, ácido clorhídrico, ácido bromhídrico; (c) sales formadas con ácidos orgánicos, por ejemplo, ácido acético, ácido oxálico, ácido tartárico; y (d) sales formadas con aniones elementales, por ejemplo, cloro, bromo y yodo.

10 Kits

Otro aspecto de la presente divulgación se refiere a kits para llevar a cabo la administración de una molécula de fusión, la administración de un ácido nucleico que codifique un polipéptido de fusión, la administración de una nucleoproteína que comprende una molécula de fusión o la administración de una nucleoproteína y un ácido nucleico. En una realización, el kit comprende una nucleoproteína que comprende un ácido nucleico que es complementario u homólogo a una secuencia de nucleótidos de interés, opcionalmente, formulada en un vehículo farmacéutico. En otra realización, el kit comprende una nucleoproteína que comprende un ácido nucleico que es complementario u homólogo a una secuencia de nucleótidos de interés, y al menos un ácido nucleico para la inserción en un genoma en el entorno de la secuencia de interés, formulada cuando sea apropiado, en una o más preparaciones farmacéuticas separadas. La nucleoproteína y el ácido nucleico se pueden formular juntos en una sola preparación o se pueden suministrar como preparaciones separadas. Los kits pueden incluir componentes y/o instrucciones para las realizaciones como se expone en el presente documento, incluyendo el apartado titulado "Descripción adicional".

25 En el presente documento, se citan diversas publicaciones.

Ejemplos

Ejemplo 1: Construcción de la molécula de fusión NLS/RecA/Gal4

30 Se construyó una proteína RecA de longitud completa, fusionada en su extremo N con una señal de localización nuclear (NLS) y, en su extremo C, con el dominio de unión al ADN de Gal4 (NLS/RecA/Gal4) mediante la fusión de la secuencia de longitud completa que codifica la RecA bacteriana y el dominio de unión al ADN de Gal4 de levadura. Las secuencias se amplificaron con la polimerasa Pfx50 de prueba de lectura (Invitrogen, Carlsbad, CA) del plásmido de RecA pBEU14 (Uhlín y Clark, 1981) y del plásmido de Gal4 pGBT9 (Clontech, Mountain View, CA.). El extremo N de RecA se fusionó con una señal de localización nuclear (NLS) del antígeno T grande de SV40 para potenciar la dirección nuclear (Keller *et al.*, 2003). La NLS del extremo 5' de cada proteína de fusión se añadió usando el siguiente cebador: 5'-CATATGCCACCTAAAAAGAAGAGAAAGGTAGAAGACCCCAAGATGGCTATCGACGAAAACAA-3' (SEC ID N° 7). La NLS tenía la secuencia de aminoácidos PPKKKRKVEDPK (SEC ID N° 1). El dominio de unión al ADN de Gal4 contenía los aminoácidos 1 a 147 de la proteína Gal4 y tenía la secuencia de aminoácidos MKLLSSIEQACDICRLKLLKCSKEKPKCAKCLKNNWECRYSPKTKRSPLTRAHLTEVESRLERLEQLFLLIFPREDLDMI LKMDSLQDIKALLTGLFVQDNVNKDAVTDRLASVETDMPPLTRQHRISATSSSEESSNKGQRQLTVS (SEC ID N° 2). La proteína de fusión NLS/RecA/Gal4 también contenía un marcador His₆ en el extremo carboxi (aminoácidos 521-528) para ayudar en la purificación. La secuencia de aminoácidos completa de la proteína de fusión se muestra en la Figura 1 (SEC ID N° 3). La secuencia de nucleótidos que codifica la proteína de fusión NLS/RecA/Gal4 se da en la Figura 2 (SEC ID N° 4).

50 La proteína de fusión se expresó en BL21(DE3)pLysS_Cam⁺ de *E. coli*. El análisis de transferencia de proteínas de los lisados usando un anticuerpo de cabra anti-RecA mostró que la proteína de fusión de 58 kD se expresó abundantemente después de la inducción con IPTG 1 mM durante 2 horas o 16 horas. La proteína se purificó de lisados de células inducidas usando una columna de Ni²⁺.

Ejemplo 2: Construcción de NLS/RecA

55 También se construyó una fusión NLS/RecA como se ha explicado anteriormente para NLS/RecA/Gal4, a excepción del uso del siguiente cebador 3': 5'-GATCGCGGCCGCAAATCTTCGTTAGTTTCTG-3' (SEC ID N° 8). La secuencia de aminoácidos de la proteína de fusión NLS/RecA se muestra en la Figura 3 (SEC ID N° 5). La secuencia de nucleótidos que codifica la proteína de fusión NLS/RecA se da en la Figura 4 (SEC ID N° 6).

Ejemplo 3: Formación de filamentos de nucleoproteína con moléculas de fusión

65 Para ensayar si las moléculas de fusión conservaban la capacidad mediada por RecA para formar filamentos de nucleoproteína en el ADN monocatenario, se usó una reacción *in vitro* que contenía el análogo de ATP no hidrolizable ATP- γ -S. Stasiak *et al.* (1994) *Experientia* 50: 192-203, Baliga *et al.* (1995) *Proc. Natl. Acad. Sci. EE.UU.* 92:10393-10397. Se diluyeron en 4,0 μ l de agua de 10 a ~20 ng de ADN bicatenario (~250 pb) del gen de la cabeza

flotante del pez cebra. Este ADN no se expuso a bromuro de etidio (EtBr). Se desnaturizó este ADNbc mediante calentamiento hasta 95 °C en un ciclador de temperatura (MASTERCYCLER EPGRADIENT S, Eppendorf, Hamburgo, Alemania) durante 12 minutos y se enfrió en hielo durante dos minutos. A continuación, se añadieron 0,8 µl de tampón de recubrimiento (TrisOAc 100 mM, pH 7,5; NaOAc 500 mM; DTT 10 mM, Mg(OAc)₂ 10 mM), 0,6 µl de ATP_γS 16,2 mM (de Sigma) y 100-200 ng de RecA, NLS-RecA o NLS-RecA-Gal4 en 4,0 µl de sondas de ADNmcc. A continuación, se añadió agua a un volumen de reacción 7,0 µl, que se incubó de inmediato a 37 °C durante 30 minutos para crear filamentos de ADN-RecA.

Se ensayó la capacidad de la proteína NLS-RecA-Gal4 para recubrir ADN monocatenario (mc). Se usó una reacción *in vitro* con la proteína purificada, ADN monocatenario y una forma no hidrolizable de ATP, ATP- γ -S, para ensayar la actividad de recubrimiento. Para ello, se usó ADNmc desnaturizado complementario (c) (ADNmcc) correspondiente al locus *flh* (250 nucleótidos para cada cadena). Después del recubrimiento, se analizó el cambio de movilidad del ADN en un gel de agarosa convencional. La RecA recubrió de manera eficaz el ADNmcc, produciendo un cambio de movilidad predicho del ADN monocatenario tras la electroforesis. Por el contrario, la incubación con una cantidad equivalente de BSA en lugar de RecA no produjo un cambio de movilidad del ADN (datos no mostrados). Al igual que la RecA nativa, la proteína NLS-RecA-Gal4 también recubrió el ADNmcc y provocó un cambio de movilidad, lo que indica que la proteína de fusión NLS-RecA-Gal4 conserva la capacidad de unirse ADNmcc. Sin embargo, la mayor parte del ADNmcc recubierto no suele migrar en el gel de agarosa, lo que sugiere la formación de estructuras de orden superior entre los filamentos de NLS-RecA-Gal4. Estos complejos podrían ser el resultado del dominio de dimerización en la región de unión al ADN de Gal4.

Ejemplo 4: Proteínas de fusión para la interrupción de genes dirigidos

Los filamentos de RecA de ADNmcc producidos con diferentes proteínas de fusión RecA estimularon la interrupción de genes dirigidos de regiones cromosómicas homólogas al ADNmc.

Diseño de experimentos con filamentos de RecA. Se ensayó la capacidad de diferentes proteínas de fusión RecA para inducir la pérdida de heterocigosidad (LOH) en locus específicos del pez cebra. Una fusión de los extremos N y C entre la señal de localización nuclear (NLS) de SV40 del antígeno T grande de SV40, RecA y el dominio de unión al ADN de Gal4, la proteína NLS-RecA-Gal4 (Figura 5, panel A), resultó tener una actividad notable para inducir la LOH en un locus específico del pez cebra. La secuencia NLS estaba presente para mejorar la dirección nuclear de los filamentos de ADNmc-RecA. Se comparó la actividad de la proteína recombinante NLS-RecA-Gal4 con NLS-RecA, que carecía del dominio Gal4 (Figura 5, panel B). La región de unión al ADN de Gal4 contenía tanto un dominio de dimerización como un dominio de reconocimiento del ADN/unión a metal que puede contribuir a la actividad observada para inducir la LOH en locus específicos. Por esta razón, se comparó la actividad de NLS-RecA-Gal4 con una proteína de fusión NLS-RecA-Gal4 que carecía del dominio de dimerización de Gal4, pero que conservaba el dominio de unión al ADN, denominada NLS-RecA-Gal4 Δ DD (Figura 5, Panel C). Se inyectaron filamentos de ADNmcc-RecA, complementarios al locus *golden* (*gol*) en embriones en etapa de una sola célula de pez cebra que eran heterocigotos para el alelo *gol*^{b1} (Figura 6, panel A). Después de la inyección, se examinó la LOH de los embriones en el locus *gol* a los 3 días de la fertilización (dpf).

El locus *gol* fue la diana por varias razones. En primer lugar, el locus *gol* es necesario de una manera autónoma de la célula para la pigmentación (Streisinger *et al.*, 1989), proporcionando una relación directa entre el fenotipo y el genotipo, y permitiendo el examen de la producción de pigmento de las células individuales. La pigmentación de embriones de pez cebra es visible desde el segundo día del desarrollo (Feitsma *et al.*, 2008). En segundo lugar, los mutantes *gol* homocigotos recesivos son viables, pero carecen de pigmentación oscura en las células epiteliales de la retina y los melanocitos. Los heterocigotos para *gol* muestran niveles normales o de tipo silvestre de pigmentación (Figura 6, panel B). La LOH en los embriones heterocigotos para *gol* da lugar a parches claros en el ojo que carecen de pigmentación (Figura 6, panel B) (Moore *et al.*, 2006). Esto permite la detección rápida de un fenotipo visible en un gen mutado que no es esencial para la embriogénesis. En tercer lugar, el gen correspondiente a la mutación de *gol*, denominado *slc24a5*, está clonado, y la región genómica que contiene el locus *gol* está bien caracterizada (Lamason *et al.*, 2005).

Resultados

Se usó un fragmento de 1.300 pb de ADN complementario a una región que abarcaba desde el intrón 3 hasta el intrón 5 del gen *gol* que contenía un exón 4 mutado (denominado *gol*-1300 en la Figura 6, Panel C). La mutación en el exón 4 da como resultado un codón de detección prematura, y se diseñó para reemplazar un sitio de reconocimiento de la enzima de restricción *Hgal* endógeno en el exón 4 con el fin de crear un polimorfismo de longitud de fragmentos de restricción (RFLP) si se recombinara en el locus *gol* endógeno. Se desnaturizó el ADN de 1.300 pb y se recubrió con NLS-RecA-Gal4 para producir filamentos de *cssgol*-1300-NLS-RecA-Gal4. Después de la inyección de estos filamentos en embriones heterocigotos para *gol*^{b1}, el 2,9 % de los embriones inyectados presentaron LOH en el locus *gol*, según lo mostrado por la pérdida de pigmentación en los parches del epitelio de la retina (Tabla I). Como control, se mezcló la proteína de NLS-RecA-Gal4 con *gol*-1300 que no se desnaturizó, y se inyectó en embriones heterocigotos para *gol*^{b1}. Esta condición no mostró niveles detectables de LOH en el locus *gol*. Estos resultados indican que los filamentos de *cssgol*-1300-NLS-RecA-Gal4 fueron capaces de mutar la copia

de tipo silvestre de *gol*. Los presentes inventores esperan que la frecuencia a la que esto ocurre, que es el doble de lo que se detectó como alteración del alelo *golb1*, no dé lugar a un fenotipo detectable. El análisis del ADN aislado de embriones que contienen patrones de pigmentación en mosaico no mostró RFLP (datos no mostrados). Este resultado indicó que el ADN monocatenario del filamento no estaba sustituido en el ADN cromosómico, es decir, la LOH observada en el locus *gol* no fue inducida por la sustitución de ADN homólogo.

También se ensayaron filamentos menores de *cssgol*-NLS-RecA-Gal4, y también se encontró que inducían la LOH en el locus *gol*. Se diseñaron oligonucleótidos de 60 pb que eran complementarios al exón 6, denominados *ssgol**ex6m*-60 sentido (s) y *ssgol**ex6m*-60 antisentido (a) (Figura 6). Al igual que el fragmento *gol*-1300, los oligonucleótidos fueron diseñados para contener codones de terminación situados en el centro de la secuencia de 60 pb y crear un RFLP en el caso de que los oligonucleótidos fueran a reemplazarse por recombinación con el gen *gol* endógeno. El RFLP se detectaría como un cambio desde un sitio de restricción de *AfeI* endógeno a un sitio de restricción de *HindIII* artificial. Cuando los oligonucleótidos de 60 pb se recubrieron con la proteína NLS-RecA-Gal4, se mezclaron entre sí y se inyectaron en embriones heterocigotos para *gol*, se observó LOH en el 2,5 % de los embriones inyectados como falta de pigmentación en los ojos, al igual que la frecuencia observada con los filamentos de *cssgol*-1300-NLS-RecA-Gal4. Una vez más, el análisis del ADN aislado de embriones que presentaban LOH en el locus *gol* después de la inyección de filamentos de *cssgol**ex6m*-60-NLS-RecA-Gal4 no reveló RFLP (datos no mostrados). Estos resultados indicaron que los filamentos de *cssgol*-NLS-RecA-Gal4 tan cortos como de 60 pb son capaces de mutar el gen diana, pero no son inducidos por recombinación.

Para descartar que la LOH inducida observada fuera el resultado de la recombinación de los filamentos de ADN_{mcc}-NLS-RecA-Gal4, se ensayó la capacidad de los oligonucleótidos monocatenarios complementarios sin codón de terminación para potenciar la LOH en el locus *gol*. Para ello, se diseñaron dos pares adyacentes de oligonucleótidos *gol* *mcc* cortos, denominados *gol**ex6*-60-1 y -2, complementarios al exón 6 en el gen *gol* sin mutaciones puntuales (Figura 6, Panel B). La inyección de cualquier par complementario de filamentos de *gol**ex6*-60-1 y -2-NLS-RecA-Gal4 en embriones heterocigotos para *gol* dio lugar a LOH a una frecuencia similar a la de los filamentos de *cssgol**ex6m*-60-NLS-RecA-Gal4 (Figura 6, panel B y Tabla I). Esta frecuencia no se mejoró mediante la inyección de filamentos de *gol**ex6*-60-1- ni -2-NLS-RecA-Gal4, y era dependiente de NLS-RecA-Gal4 (Tabla I). Estos datos apoyan aún más que los filamentos de ADN_{mcc}-NLS-RecA-Gal4 potencian la LOH en el locus *gol* mediante un mecanismo independiente de la recombinación que da lugar a la sustitución de genes.

Se realizó un ensayo para determinar si se requerían ambos filamentos de ADN_{mc}-NLS-RecA-Gal4 complementarios para inducir LOH. La inyección de filamentos de *ssgol**ex6m*-60-NLS-RecA-Gal4 bien sentido o antisentido solos en embriones heterocigotos para *golb1* no induce LOH detectable en el locus *gol* (Tabla I). Además, cuando se inyectaron filamentos no apareados, pero adyacentes, producidos con NLS-RecA-Gal4 (ya sea *ssgol**ex6*-60-1 sentido con *ssgol**ex6*-60-2 antisentido o *ssgol**ex6*-60-1 antisentido con *ssgol**ex6*-60-2 sentido) en embriones heterocigotos para *gol*, no se observó LOH (Tabla I). Estos resultados muestran que se requieren filamentos de ADN_{mc}-NLS-RecA-Gal4 complementarios para la inducción de alta frecuencia de LOH en el locus *gol*.

Para ensayar si se requieren el grupo fosfato 5' y el grupo OH 3' para esta dirección génica, se sintetizaron dos pares modificados de oligonucleótidos *gsg2* y se inyectaron con NLSRecAGal4. Uno de ellos es el grupo C6 modificado con amino 5' (5'AmC6), que bloquea el grupo fosfato en el extremo 5' de los oligonucleótidos. Se demostró que 5'AmC6 no tiene ningún efecto en la capacidad de unión de la proteína RecA, aunque afectó a la formación de tren del filamento de RecA (una forma concatenada de filamento de RecA que puede conectarse hasta al menos 50 oligonucleótidos) (Simonson *et al.*, 1994). El otro es el modificador de extremos 3' desoxitimidina invertido (3' InvdT), que bloquea el grupo hidroxilo 3' de los oligonucleótidos e inhibe la extensión del cebador impulsada por la ADN polimerasa desde el extremo 3' (Dames *et al.*, 2007). Como se muestra en la Tabla I, la inyección mediante cualquiera de oligonucleótidos 5'AmC6 o 3'InvdT con NLSRecAGal4, conserva la actividad para inducir el suceso de LOH en el locus *gol* en una proporción constante sin compromiso, independientemente del extremo que esté bloqueado.

Finalmente, para confirmar la especificidad de la dirección génica causada por la actividad de la búsqueda de homólogos de RecA, se diseñaron oligonucleótidos no *gol* complementarios con la misma longitud de 60 nt para el experimento de control. Este par de sondas no específicas fue incapaz de causar el mismo fenotipo de LOH en el locus *gol* que los filamentos oligo-NLS-RecA-Gal4 de *gol* (Tabla I). Este resultado muestra el requisito de especificidad dirigida por sonda en este ensayo de NLS-RecA-Gal4.

Ejemplo 5: Inserción específica de sitio de un gen exógeno

Los datos anteriores indicaron que ADN_{mcc}-NLS-RecA-Gal4 potenció la LOH en el locus *gol*. Sin quedar ligados a una teoría específica, se cree que este efecto fue mediado por las roturas de las cadenas dobles (DSB) de la región diana. Este ejemplo indica que las DSB creadas por los filamentos de NLS-RecA-Gal4 se pueden usar para potenciar la inserción de ADN exógeno suministrado en los locus específicos. Sin quedar ligados a una determinada teoría, se cree que los filamentos de NLSRecAGal4 causan DSB en las regiones diana, pudiéndose luego incorporar ADN lineal exógeno en el sitio de la rotura durante la vía de unión de extremos no homólogos (NHEJ). En este ejemplo, los sucesos de inserción dirigidos se demostraron mediante la expresión específica de tejido del gen EGFP

tras la inyección conjunta con filamentos de ADNmc-NLS-RecA-Gal4 complementarios a los locus *gol*, *prominin1* y *floating head (flh)* (Figura 7). El gen *prominin1* se seleccionó como un gen previamente no caracterizado y por su expresión específica en el diencéfalo dorsal y la retina durante el desarrollo embrionario. Los genes se insertaron sin promotores, por lo que la expresión requiere la inserción del sitio específica del sitio que se acopla a promotores endógenos que proporcionan la expresión génica.

Para seguir la dirección específica de sitio de la expresión génica impulsada por el promotor endógeno, la estrategia de dirección es similar para una dirección génica como la usada en la interrupción del gen *gol* como ya se ha descrito. Para ello, se amplifican simplemente de 200 a 300 pb de la región diana del gen por PCR, se desnaturaliza el producto de amplificación por calentamiento, y se recubren las cadenas simples por NLSRecAGal4 con la inyección conjunta adicional con un fragmento lineal de ADN que contiene un indicador de EGFP (Figura 8). Este tipo de expresión de EGFP específica de tejido se observó para del 5 al 19 % de los embriones inyectados de diferente gen de dirección (Tabla II). En todos los casos, se observó muy poca expresión fuera de la diana (datos no mostrados).

Para los experimentos que se muestran en el presente documento, el gen indicador se usa como un casete con un aceptor de corte y empalme seguido de EGFP en los tres marcos seguidos por la secuencia terminal de la transcripción poliA sin ninguna secuencia promotora (Figura 8, Panel A). Dado que estos genes indicadores se podrían insertar en cualquier dirección, cabría esperar la observación de la expresión en solo 1 de cada 6 inserciones en el gen diana. Por lo tanto, se cree que la carga mutagénica en el locus diana puede ser al menos 6 veces superior a la que se puede observar mediante la expresión de EGFP tras la inyección. La expresión de EGFP en este ejemplo es un marcador para indicar el suceso de interrupción génica en el pez de tipo silvestre. En comparación con los peces heterocigotos, los peces de tipo silvestre homocigotos carecen de fenotipo LOH para la selección. Por consiguiente, este método de inserción de EGFP permitirá la detección previa mediante un suceso de dirección hasta el crecimiento hasta la edad adulta. Se supone que este método no solo crea mutaciones en locus específicos, sino que también las permite después de la expresión del locus endógeno.

Se analizaron las pruebas moleculares de los sucesos de dirección e inserción. Dos regiones del gen *gol* (Figura 8, panel A) y una región de cada gen *flh* y *prominin1* fueron las dianas (datos no mostrados, pero se encontraron resultados similares). Como se muestra a continuación, se realizaron inserciones dirigidas altamente específicas en dos regiones distintas del gen *gol* mediante la selección de dos filamentos de ADNmc-NLS-RecA-Gal4 diferentes. En este ejemplo, la ubicación entre las dos sondas diferentes es de aproximadamente 1 kb (Figura 8, Panel A). Se amplificaron los fragmentos de unión entre el ADN de indicador EGFP suministrado exógenamente y el locus endógeno mediante análisis de PCR. Este análisis mostró la inserción en el locus *gol* en dos sitios distintos (Figura 8, panel B). Estos productos de amplificación se verificaron mediante secuenciación de ADN y mostraron las secuencias correctas desde la unión del locus endógeno y el ADN suministrado exógenamente. Las uniones se encontraron cerca de los extremos de los filamentos de ADNmc en muchos casos (Figura 8, Panel C). Como se muestra en la Figura 8, Panel C, el sitio de inserción estaba en aproximadamente 500 bases de la sonda. Por consiguiente, la inserción se puede realizar en aproximadamente 500 pares de bases de un sitio deseado mediante la elección de una sonda.

Ejemplo 6: La inyección conjunta de filamentos de ADNmc-NLS-RecA-Gal4 de *gol* con un gen exógeno da lugar a mutaciones que se transmiten a través de la línea germinal a la siguiente generación

Se seleccionaron los embriones que mostraban la expresión de EGFP en el ojo (Figura 7) tras la inyección conjunta de filamentos de ADNmc-NLS-RecA-Gal4 de *gol* con el gen de indicador EGFP (Figura 8) hasta el crecimiento hasta la edad adulta. Se usaron peces homocigotos Gol^{b1} para realizar un ensayo cruzado de complementación con este pez F0 fundador. De 21 adultos F0 examinados, solo dos peces F0 produjeron descendencia que no complementó el alelo *b1* (Figura 9). Estos resultados demuestran que este método se puede usar para dirigir genes, seleccionar peces y transmitir a la línea germinal. La línea germinal de estos dos peces fundadores es altamente de mosaico, mostrando el 0,7 y el 3,7 % de la descendencia de los dos fundadores la falta de complementación del alelo *b1* (Tabla II). Se inyectó uno de los fundadores F0 con filamentos de ADNmc-NLS-RecA-Gal4 correspondientes a la sonda A, y el otro se inyectó con filamentos de sonda B (Figura 8). La recuperación de dos inserciones independientes en el locus *gol* usando diferentes sondas muestra que este método se puede usar para dirigir cualquier gen del genoma. Aunque la fluorescencia del gen de indicador EGFP no se presentó en la descendencia no complementaria, la razón podría ser consecuente con al menos una posibilidad 5 veces inferior de expresión de EGFP en el marco que la interrupción de genes provocada por la eliminación o inserción de ADN en el locus *gol*.

Teorías de acción

Según los ejemplos, se ha demostrado que los filamentos de ADNmc-NLS-RecA-Gal4 complementarios (c) dirigidos al locus *gol* son capaces de inducir la pérdida de heterocigosidad en este locus después de la inyección en embriones de pez cebra. Esta actividad requiere tanto los dominios de unión/dimerización de ADN de Gal4 en NLS-RecA-Gal4 como la complementariedad entre los filamentos. Sin quedar ligados a una determinada teoría, en el presente documento, se propone un modelo para esta actividad, donde los filamentos de ADNmc-NLS-RecA-Gal4 se dirigen al locus *gol* mediante la creación de horquillas de replicación detenidas que dan lugar a roturas de

cadenas dobles (DSB) (Fig. 10).

La proteína NLS-RecA-Gal4 es capaz de recubrir ADNmc (Fig. 8). Cuando este complejo de ADN-proteína se inyecta en los embriones de pez cebra, la señal de NLS aparentemente guía al complejo hacia el núcleo de las células embrionarias. La actividad de RecA en los filamentos potencia una búsqueda de homología para encontrar ADN cromosómico homólogo. Una vez que la diana cromosómica es localizada por el apareamiento homólogo, los filamentos de ADNmcc-NLS-RecA-Gal4 inician la invasión de las cadenas de ADN. Durante la invasión de las cadenas y las etapas de intercambio de cadenas, el filamento de ADNmc aparentemente invadirá y desenrollará su ADN genómico de doble cadena homólogo, produciendo la formación de estructuras de bucle D. Debido al dominio de dimerización de Gal4, se propone NLS-RecA-Gal4 para formar un dímero y estabilizar el complejo de filamentos de cadena sencilla directa (fss)-NLS-RecA-Gal4 y cadena sencilla inversa (rss)-NLS-RecA-Gal4 en la región genómica diana (Fig. 10). Debido a que el filamento de ADNmc-RecA se desmonta tras la hidrólisis de ATP (Sigurdsson *et al.*, 2002), se usó una forma no hidrolítica de ATP, ATP γ S, para la fabricación de filamentos de ADNmcc-NLS-RecA-Gal4 estables. Se ha observado que algunas proteínas que se unen fuertemente al ADN, tal como la maquinaria de transcripción mutada, pueden impedir la progresión de la horquilla de replicación en los cromosomas, generando horquillas de replicación atascadas o colapsadas (Michel *et al.*, 2001; Aguilera y Gómez-González, 2008). Las horquillas de replicación detenidas pueden aumentar la presión sobre el cromosoma, causando la formación de DSB de ADN y el colapso de las horquillas de replicación (Michel *et al.*, 1997).

Las horquillas de replicación detenidas se pueden procesar de manera independiente de la reparación y dependiente de la reparación. Si dicha horquilla de replicación no se repara, el fragmento de ADN libre se puede perder, causando una gran eliminación. Como alternativa, el fragmento de ADN libre se puede mover a una región genómica diferente, dando lugar a una translocación cromosómica (Michel *et al.*, 2001). Una gran eliminación o translocación pueden causar un alelo haploide deficiente durante la meiosis, en la que el fenotipo resultante se puede detectar mediante un ensayo de complementación o un rastreo de haploides de una sola generación (Imai *et al.*, 2000).

Las horquillas de replicación detenidas se pueden reparar mediante una variedad de mecanismos que también podría coincidir con los estudios preliminares de los presentes inventores. A menudo, las horquillas de replicación atascadas pueden desencadenar la detención del ciclo celular y estimular el mecanismo de reparación de DSB ya sea por la vía HR o la vía NHEJ. Si la horquilla de replicación detenida está cerca del telómero, uno de los extremos de la rotura de la cadena doble (no de la rotura de la cadena doble de dos extremos) creado por el colapso de la horquilla de replicación se puede reparar por la vía de replicación inducida por la rotura (BIR) (Smith *et al.*, 2007; Llorente *et al.*, 2008). Durante este proceso, el extremo del cromosoma es reparado por una forma de recombinación homóloga usando el cromosoma hermano como molde. Esto da lugar a un tramo largo de LOH.

También se pueden realizar la DSB de dos extremos tras un colapso de la horquilla de replicación (Shrivastav *et al.*, 2008). Las DSB de dos extremos generadas por la resolución de una molécula de unión de Holliday (HJ) tras una estructura intermedia de "pata de pollo" se forma mediante la regresión de la horquilla de replicación durante el inicio de la reparación (Lundin *et al.*, 2002). Se ha demostrado que las proteínas RecA y RecG son capaces de potenciar la regresión de la horquilla de replicación del ADN para la reparación del ADN (Robu *et al.*, 2001 y 2004). Se puede usar cualquiera de las vías HR o NHEJ para reparar este tipo de DSB de ADN de dos extremos (Shrivastav *et al.*, 2008).

Las horquillas de replicación detenidas también son una diana para la digestión con nucleasas que da lugar a la formación de DSB (Michel, *et al.*, 2001). Por ejemplo, en un modelo de bacteriófago T4, la endonucleasa VII de T4 escinde las horquillas de replicación detenidas inducidas a partir de un complejo de fármaco antitumoral-topoisomerasa (Hong y Kreuzer, 2003). En la levadura, Mus81 y Mms4/Eme1 forman una endonucleasa específica de estructuras heterodiméricas que puede escindir las estructuras de ADN ramificadas formadas por uniones de Holliday en las horquillas de replicación atascadas (Boddy *et al.*, 2001; Lundin *et al.*, 2002). Este tipo de DSB crea una hendidura más corta que la causada por el colapso de la horquilla de replicación. Por consiguiente, una DSB corta se aparea de manera más eficaz ya sea por la vía HR clásica o la vía NHEJ.

Por lo tanto, en este modelo, se propone la posibilidad de usar un complejo de ADNmcc-NLS-RecA-Gal4 u otra proteína de fusión de unión a NLS-recombinasa-ADN para inducir la interrupción génica dirigida por el bloqueo de la progresión de la horquilla de replicación del ADN. Esto se traduce en DSB específica de sitio que se repara ya sea por la vía HR o la vía NHEJ endógenas. La actividad de búsqueda de homólogos de RecA puede proporcionar especificidad de diana, y el dominio de dimerización de Gal4 puede estabilizar el complejo molecular de unión complementaria en la región diana del cromosoma. Este complejo de ADN-proteína también se puede apilar con otros filamentos de ADNmcc-NLS-RecA-Gal4 y formar una estructura de alto orden para aumentar la resistencia del bloqueo de la horquilla de replicación del ADN. La horquilla de replicación detenida puede ya sea causar el colapso de la horquilla de replicación o potenciar la accesibilidad de las endonucleasas de ADN para inducir DSB de ADN. Este tipo de DSB dirigida se diferencia de las inducidas aleatoriamente por las horquillas de replicación atascadas de la inhibición química de la replicación (Feitsma *et al.*, 2008) o proteínas de unión a ADN mutantes (Michel *et al.*, 2001). Basándose en los resultados descritos en el presente documento y en el modelo propuesto, se espera potenciar la mutación génica específica de sitio mediante la inducción de DSB durante la resolución de las horquillas

de replicación atascadas. Si no se repara la DSB, cabe esperar grandes eliminaciones y translocaciones. Si se repara por la vía NHEJ, se observarán eliminaciones o inserciones predominantemente pequeñas en el sitio de la reparación. Si además se suministra ADN exógeno, se insertará en la DSB durante su reparación por la vía de NHEJ.

5

Descripción adicional

Las realizaciones de la divulgación incluyen un método, un kit, un uso o un sistema que comprende una proteína de fusión o una molécula de fusión proteica que comprende un polipéptido que posee actividad recombinasa y un dominio de unión al ADN polipeptídico. Otra realización es el sistema montado para estar exento de cualquier fragmento de ADN que se una específicamente al dominio de unión al ADN polipeptídico y/o que comprende además un ADN exógeno que no está unido específicamente a la proteína de fusión. Las realizaciones pueden comprender además un ácido nucleico monocatenario que forma un filamento de nucleoproteína mediante la unión específica al polipéptido que posee actividad recombinasa. La recombinasa puede ser una recombinasa como la establecida en el presente documento, por ejemplo, RecA, recA803, uvsX, RecA mutantes, recombinasas de tipo RecA, RuvC, DST2, KEM1 y XRN1, Stpa/DST1 y HPP-1. El dominio de unión al ADN polipeptídico puede ser como el establecido en la presente divulgación, por ejemplo, se puede seleccionar del grupo que consiste en Gal4, una nucleasa, una nucleasa de dedo de cinc, un dedo de cinc y una proteína de hélice-giro-hélice. La molécula de fusión puede comprender además una secuencia de localización nuclear (NLS). El ADN exógeno puede comprender una secuencia de gen marcador de ADN. El ADN exógeno puede codificar un polipéptido que vaya a ser expresado por una célula que reciba el sistema. El ADN exógeno puede ser un marcador para la identificación después de la inserción en un cromosoma de una célula hospedadora que reciba el sistema. La molécula de fusión puede comprender además un enlazador sintético, opcionalmente dispuesto entre el polipéptido que posee actividad recombinasa y el dominio de unión a ADN polipeptídico, por ejemplo, un óxido de polietileno. La proteína de fusión puede comprender RecA y/o Gal4.

Las realizaciones incluyen un método, un kit, un uso o un sistema para la transfección de un locus diana de una célula con un fragmento de ADN exógeno que comprende: una molécula de fusión que comprende un polipéptido que posee actividad recombinasa y un dominio de unión al ADN polipeptídico, un fragmento de ADN monocatenario con una homología considerable con el locus (una sonda), uniéndose el fragmento de manera específica con el polipéptido que posee actividad recombinasa para formar así un filamento, y un fragmento de ADN exógeno que no se une de manera específica con la proteína de fusión, opcionalmente en la que dicho ADN exógeno codifica un polipéptido para ser expresado por la célula. El sistema puede estar exento de cualquier fragmento de ADN que se una específicamente al dominio de unión al ADN polipeptídico. Se puede proporcionar un sistema en el que el fragmento de ADN exógeno esté presente a una concentración molar que supere la concentración molar de la proteína de fusión, siendo el exceso opcionalmente de al menos el doble o de entre aproximadamente el doble y 500 veces. Los expertos en la materia apreciarán de inmediato que se contemplan todos los intervalos y valores de entre los valores indicados explícitamente, por ejemplo, un exceso de 10 veces más o de aproximadamente 5 veces a aproximadamente 500 veces. El fragmento de ADN exógeno puede codificar un polipéptido para la expresión celular y/o estar exento de una secuencia promotora. El fragmento de ADN exógeno puede codificar un polipéptido para la expresión celular y, opcionalmente, incluir un casete de expresión. La recombinasa puede ser una recombinasa como la expuesta en el presente documento. El dominio de unión al ADN polipeptídico se puede seleccionar del grupo que consiste en Gal4, una nucleasa, una nucleasa de dedo de cinc, un dedo de cinc y una proteína de hélice-giro-hélice. La proteína de fusión puede comprender además una secuencia de localización nuclear (NLS). Se puede proporcionar ADN exógeno que comprenda una secuencia de gen marcador de ADN, codifique un polipéptido que vaya a ser expresado por una célula que reciba el sistema o sea un marcador para la identificación después de la inserción en un cromosoma de una célula hospedadora que reciba el sistema. La proteína de fusión puede comprender además un enlazador sintético, opcionalmente dispuesto entre el polipéptido que posee actividad recombinasa y el dominio de unión al ADN polipeptídico, por ejemplo, un óxido de polietileno.

Las realizaciones incluyen un método, un kit, un uso o un sistema para transfectar una célula que comprenden exponer la célula al sistema de cualquiera de 12-21, en el que un usuario elige un sitio diana de un cromosoma de la célula, forma el filamento, y administra el filamento y el ADN exógeno a la célula, en la que el ADN exógeno se dispone de manera eficaz en menos de aproximadamente 5.000 pares de bases del sitio diana. Los expertos en la materia apreciarán de inmediato que se contemplan todos los intervalos y valores de los intervalos indicados explícitamente, por ejemplo, 0-5.000, aproximadamente 100 a aproximadamente 1.000, aproximadamente 0 a aproximadamente 500, menos de 2.000.

Las realizaciones incluyen un método, un kit, un uso o un sistema para la transfección de una célula con un fragmento de ADN exógeno y considerablemente homólogo que comprenden: una molécula de fusión que comprende un polipéptido que posee actividad recombinasa y un dominio de unión al ADN polipeptídico, un fragmento de ADN bicatenario con al menos una porción que tiene una secuencia de al menos aproximadamente 20 restos que tiene una homología considerable con el locus, estando el fragmento de ADN bicatenario exento de unión específica con la proteína de fusión. El fragmento de ADN bicatenario puede tener al menos dos secuencias de al menos 20 restos que tienen una homología considerable con el locus. La homología considerable puede ser una identidad. El sistema puede estar exento de cualquier fragmento de ADN que se una específicamente al dominio de

unión al ADN polipeptídico. Se puede proporcionar un sistema en el que el fragmento de ADN bicatenario esté presente a una concentración molar que supere la concentración molar de la proteína de fusión, siendo el exceso opcionalmente de al menos el doble o de entre aproximadamente el doble y 500 veces. Los expertos en la materia apreciarán de inmediato que se contemplan todos los intervalos y valores de entre los valores indicados explícitamente, por ejemplo, un exceso de 10 veces o de aproximadamente 5 veces a aproximadamente 50 veces. El fragmento de ADN exógeno puede codificar un polipéptido para la expresión celular. El polipéptido que posee recombinasa puede ser una recombinasa como la establecida en el presente documento. El dominio de unión al ADN polipeptídico puede ser como el expuesto en el presente documento, por ejemplo, seleccionado del grupo que consiste en Gal4, una nucleasa, una nucleasa de dedo de cinc, un dedo de cinc y una proteína de hélice-giro-hélice. La proteína de fusión puede comprender además una secuencia de localización nuclear (NLS). El ADN bicatenario puede comprender una secuencia de gen marcador de ADN, codificar un polipéptido que vaya a ser expresado por una célula que reciba el sistema o un marcador para la identificación después de la inserción en un cromosoma de una célula hospedadora que reciba el sistema. La proteína de fusión puede comprender además un enlazador sintético, opcionalmente dispuesto entre el polipéptido que posee actividad recombinasa y el dominio de unión al ADN polipeptídico, por ejemplo, un óxido de polietileno. La homología puede comprender secuencias homólogas situadas en los extremos del ácido nucleico y/o internamente en el primer ácido nucleico. Las realizaciones pueden incluir el caso en el que la molécula de ADN bicatenario tenga extremos 3' monocatenarios que sobresalen. El dominio de unión al ADN polipeptídico puede comprender un dominio de unión al ADN de Gal4.

Las realizaciones incluyen un método, un kit, un uso o un sistema para la mutagénesis dirigida en o cerca de una región de interés en una secuencia de ácido nucleico celular, método que comprende: (a) proporcionar una molécula de ácido nucleico que tenga homología con la región de interés; (b) unir una molécula de fusión con el ácido nucleico, molécula de fusión que comprende: (i) secuencias polipeptídicas que tienen actividad RecA/Rad51, y (ii) secuencias polipeptídicas que comprenden un dominio de unión al ADN específico de secuencia; y (c) introducir el ácido nucleico unido a proteína en la célula. La secuencia de ácido nucleico celular puede estar en un cromosoma. La molécula de ácido nucleico puede ser ADN. El ADN puede ser monocatenario. El ADN puede ser bicatenario. La molécula de ácido nucleico puede ser ARN. La molécula de fusión comprende secuencias polipeptídicas que tienen actividad RecA. El dominio de unión al ADN específico de secuencia puede comprender un dominio de unión al ADN de Gal4. La proteína de fusión puede comprender además una secuencia de localización nuclear (NLS). La mutagénesis dirigida puede realizarse de modo que genere la conversión de una secuencia mutante a una secuencia de tipo silvestre. La mutagénesis dirigida se puede realizar de modo que genere la conversión de un primer alelo/haplotipo a un segundo alelo/haplotipo. La mutagénesis dirigida se puede realizar para genere una conversión de una secuencia de tipo silvestre a una secuencia mutante. La mutación se puede seleccionar del grupo que consiste en una mutación puntual, una inserción, una eliminación, una translocación y una inversión.

Las realizaciones incluyen un método, un kit, un uso o un sistema para la recombinación homóloga dirigida entre una secuencia de interés de ADN celular y un ácido nucleico bicatenario exógeno, método que comprende: (a) proporcionar un ácido nucleico bicatenario lineal que contiene una o más regiones homólogas a la secuencia de interés; (b) unir una proteína de fusión con el ácido nucleico, proteína de fusión que comprende: (i) secuencias polipeptídicas que tienen actividad RecA/Rad51, y (ii) secuencias polipeptídicas que comprenden un dominio de unión al ADN específico de secuencia y (c) introducir el ácido nucleico unido a proteína en la célula. El ácido nucleico de la etapa (a) puede contener las regiones homólogas a la secuencia de interés en ambos de sus extremos. Cada una de las regiones de homología puede ser de al menos 10, 20 o 50 nucleótidos de longitud. Una o más regiones homólogas a la secuencia de interés pueden estar situadas internamente en el ácido nucleico de la etapa (a). Además, todo el ácido nucleico exógeno puede estar integrado en el ADN celular. El ADN celular puede estar, por ejemplo, en un cromosoma, en un episoma, compuesto por secuencias que codifican una proteína. El ácido nucleico exógeno puede comprender además secuencias reguladoras. Se puede proporcionar una proteína de fusión que comprenda secuencias polipeptídicas que tengan actividad RecA. El dominio de unión al ADN específico de secuencia puede comprender el dominio de unión al ADN de Gal4. La proteína de fusión puede comprender además una secuencia de localización nuclear (NLS). El ácido nucleico de la etapa (a) se puede proporcionar de manera que no contenga un sitio de reconocimiento para una recombinasa, transposasa o integrasa. El ácido nucleico de la etapa (a) puede estar exento de cualquier transposón o un genoma viral. La recombinación se puede diseñar por ingeniería genética de modo que produzca la conversión de una secuencia mutante en una secuencia de tipo silvestre. La recombinación puede dar lugar a la conversión de un primer alelo/haplotipo en un segundo alelo/haplotipo. La recombinación puede dar lugar a la conversión de una secuencia de tipo silvestre en una secuencia mutante. La mutación se puede seleccionar, por ejemplo, del grupo que consiste en una mutación puntual, una inserción, una eliminación, una translocación y una inversión.

Una realización es una proteína de fusión que comprende: (a) secuencias polipeptídicas que tienen actividad RecA/Rad51, y (b) secuencias polipeptídicas que comprenden un dominio de unión al ADN específico de la secuencia. La proteína de fusión puede comprender además una secuencia de localización nuclear (NLS). Como alternativa, la molécula de fusión se puede colocar directamente en la célula como se describe en el presente documento.

Las realizaciones incluyen un método, un kit, un uso o un sistema para la estimulación de la conversión de genes en una célula, método que comprende introducir, en la célula, una molécula de fusión como la expuesta en el presente documento. Un ejemplo es una molécula de fusión proteica que comprende: (a) RecA; (b) una NLS; y (c) Gal4; en el que dicha conversión génica no requiere la hidrólisis de ATP.

Las realizaciones incluyen un método, un kit, un uso o un sistema para transfectar una célula que comprende exponer la célula a una realización de un sistema como el expuesto en el presente documento. Las realizaciones incluyen un método, un kit, un uso o un sistema para transfectar una célula que comprende exponer la célula al sistema como ya se ha descrito. Son ejemplos de células una célula de vertebrado, una célula de mamífero, una célula porcina, una célula humana, una célula vegetal y una célula madre. Las realizaciones incluyen un animal transgénico formado por los métodos o sistemas descritos, por ejemplo, un cerdo u ortodáctilo, o cerdo enano, cabra, conejo o ratón. El método, kit, uso o sistema puede estar exento de un transposón y/o un genoma viral. Los materiales se pueden proporcionar en una forma farmacéuticamente aceptable o con un excipiente farmacéuticamente aceptable. Las moléculas de fusión se pueden usar para producir proteínas terapéuticas *in vitro* o *in vivo*.

Una realización es un método de transfección de una célula que comprende introducir en la célula: un ácido nucleico exógeno y un filamento de nucleoproteína de una molécula de fusión proteica y una sonda de ácido nucleico complementaria a un sitio diana de DNA de la célula, en el que la proteína de fusión comprende un dominio de recombinasa que contribuye al filamento, y un dominio de unión al ADN, en el que el ácido nucleico exógeno se incorpora en el ADN de la célula y es expresado por la célula. Una realización es una composición purificada para la transfección de ADN exógeno en ADN cromosómico de una célula, composición que comprende un filamento de nucleoproteína de una sonda y una molécula de fusión proteica, en la que la sonda comprende ADN desnaturalizado bicatenario complementario a un sitio de ADN cromosómico, y la molécula de fusión comprende un dominio de recombinasa y un dominio de unión al ADN, estando composición exenta de secuencias de ADN que se unen específicamente al dominio de unión al ADN. Una realización es un método de tratamiento de una enfermedad genética en un animal que comprende introducir en una célula del animal: un ácido nucleico exógeno y un filamento de nucleoproteína de una molécula de fusión proteica y una sonda de ácido nucleico complementaria a un sitio diana de ADN de la célula, en el que la molécula de fusión comprende un dominio de recombinasa que contribuye al filamento, y un dominio de unión al ADN en el momento de la introducción, en el que el ácido nucleico exógeno es expresado por la célula para proporcionar una proteína terapéutica al animal para tratar la enfermedad, el método se realiza sin un vector viral y sin un vector transposón, y la célula se transfecta mediante un método seleccionado del grupo que consiste en *in vitro* y *ex vivo*. El ácido nucleico exógeno puede ser ADN bicatenario y también puede estar exento de secuencias promotoras. La sonda de ácido nucleico puede proporcionarse de forma que comprenda una sola cadena de ADN complementaria (ADNmcc) y el dominio de unión al ADN no se una específicamente al ADN en el momento de la introducción. El sitio de ácido nucleico exógeno de la inserción puede proporcionarse para que esté en aproximadamente 1.000 bases del sitio diana. Los sistemas se proporcionan con una eficacia de transfección de la célula de al menos aproximadamente el 1 %, medida por un ensayo *in vitro* con inyección directa. La recombinasa puede ser o comprender RecA, o un fragmento funcional de la misma. El dominio de unión al ADN puede comprender Gal4. La recombinasa puede ser una recombinasa expuesta en el presente documento, por ejemplo, seleccionada del grupo que consiste en recombinasa Cre, recombinasa Hin, recombinasa Tre, enzima de recombinación de flipasa, uvsX, RuvC, DST2, KEM1 y XRN1, Stpa/DST1 y HPP-1. El dominio de unión al ADN puede comprender un polipéptido que se una específicamente al ADN, y se selecciona del grupo que consiste en enlazadores al surco menor, enlazadores al surco mayor, antibióticos, agentes intercalantes, poliamidas y una secuencia polipeptídica de un factor de transcripción, nucleasa, nucleasas de dedo de cinc, dedos de cinc y proteínas de hélice-giro-hélice. La molécula de fusión se puede seleccionar del grupo que consiste en SEC ID N° 3, SEC ID N° 5 y sustituciones conservadoras de las mismas. La señal de localización nuclear puede ser, por ejemplo, un miembro de la familia de SV40. La molécula de fusión puede comprender un enlazador no peptídico sintético. La sonda se puede dirigir a una mutación del ADN de la célula, y el ácido nucleico exógeno comprende una secuencia de tipo silvestre correspondiente a la mutación. El ácido nucleico exógeno puede ser no homólogo con respecto al ADN celular. La molécula de fusión puede comprender además un dominio de señal de localización nuclear. Las células se pueden transfectar con los sistemas o métodos, por ejemplo, una célula de vertebrado, una célula de mamífero, una célula porcina, una célula humana, una célula vegetal y una célula madre. Se puede formar un animal transgénico mediante dichos métodos, por ejemplo, a partir de la progenie de una célula de la línea germinal transfectada mediante el método o los sistemas; por ejemplo, un cerdo o artiodáctilo o cerdo enano, cabra, conejo o ratón. El método de introducción se puede seleccionar, por ejemplo, del grupo que consiste en electroporación, liposomas, trasplante nuclear, microinyección pronuclear y transferencia nuclear de células somáticas. La sonda se puede dirigir, por ejemplo, a un ADN mutado del animal que contribuye a la enfermedad.

Tabla I. La inyección de ADNmc-NLSRecAGal4 produce pérdida de heterocigosidad en el locus de *gol* (b1)

	Tipo de RecA	Sonda*	Dosis	Total	Vida total	Desv. normal	Clones con ojo <i>gol</i>	
Inyección	NLSRecAGal4	gsg1 mcc	45 pg	292	241	N.D	7/241	2,9 %
	NLSRecAGal4	gsg1 bc	45 pg	284	237	N.D	0/237	0 %
	NLSRecAGal4	gsg2-F mc	80 pg	155	136	N.D	0/136	0 %
	NLSRecAGal4	gsg2-R mc	80 pg	258	224	N.D	0/224	0 %
	NLSRecAGal4	gsg2 mcc	80 pg	278	201	N.D	5/201	2,5 %
	No	gbg1 mcc	160 pg	74	65	58	0/58	0 %
	No	gbg2 mcc	160 pg	227	208	150	0/150	0 %
	No	gbg1F+gbg2R mc	160 pg	75	68	68	0/68	0 %
	No	gbg2F+gbg1R mc	160 pg	82	71	70	0/70	0 %
	No	gbg1+gbg2 mcc	160 pg	597	379	342	0/342	0 %
	NLSRecAGal4	gbg1 mcc	160 pg	260	188	144	6/144	4,2 %
	NLSRecAGal4	gbg2 mcc	160 pg	139	135	107	3/107	2,8 %
	NLSRecAGal4	gbg1F-gbg2R mc	160 pg	222	197	190	0/190	0 %
	NLSRecAGal4	gbg2F+gbg1R mc	160 pg	147	129	129	0/129	0 %
	NLSRecAGal4	gbg1+gbg2 mcc	160 pg	979	745	591	20/591	3,4 %
	RecA	gbg1+gbg2 mcc	160 pg	317	233	206	0/206	0 %
	NLSRecA	gbg1+gbg2 mcc	160 pg	301	253	220	0/220	0 %
	NLSRecAGal4ΔDD	gbg1+gbg2 mcc	160 pg	396	341	296	3/296	1,0 %
	NLSRecAGal4	Oligos flh mcc	160 pg	298	268	255	0/255	0 %
	NLSRecAGal4	Oligos prim mcc	160 pg	411	314	300	0/300	0 %
	NLSRecAGal4	Oligos vegfa mcc	160 pg	542	337	307	0/307	0 %
	NLSRecAGal4	5'AmC6-gbg2 mcc	160 pg	152	109	67	2/67	3,0 %
	NLSRecAGal4	3'InvdT-gbg2 mcc	160 pg	393	318	236	9/236	3,8 %
Sin inyección				471	438	N.D	0/438	0 %

*mc: sonda de ADN monocatenario; bc: sonda de ADN bicatenario; mcc: sonda de ADN monocatenario complementario; gsg: sonda de *gol* con codón de terminación en la mitad; gbg: sonda de *gol* sin codón de terminación; 5'AmC6: bloque C6 modificador de amino 5'; 3'InvdT: bloque dT invertido en 3'; N.D: no determinado.

5

Tabla II. Expresión de gen indicador somático dirigida por NLSRecAGal4 en *golden*, *floating head* y *prominin-1*

Gen de dirección	Tipo de sonda	Expresión de gen indicador somático			Ensayo de complementariedad con línea germinal	
		Embriones totales	Supervivientes	Fluorescencia	Fundador (%)	Progenie de <i>gol</i> (%)
<i>golden</i>	Sonda I de <i>gol</i> (gol270pb)	791*	649	33 (5,1 %)	1/13 (7,7 %)	34/919 (3,7 %)
	Sonda II de <i>gol</i> (gol300pb)	207	165	14 (8,5 %)	1/8 (12,5 %)	1/139 (0,7 %)
<i>Floating head</i>	flh250pb	216	212	18 (8,5 %)		
<i>Prominin-1</i>	prim200pb	162	154	29 (18,8 %)		

* Los datos se recogieron de tres experimentos individuales.

10

REIVINDICACIONES

- 5 1. Un método de integración de un ácido nucleico exógeno en un sitio diana que comprende introducir en una célula: un ADN bicatenario lineal exógeno y un filamento de nucleoproteína de una molécula de fusión proteica y una sonda de ácido nucleico complementaria a un sitio diana de ADN de la célula; en el que la proteína de fusión comprende un dominio de RecA o Rad51 que tiene una actividad recombinasa que contribuye al filamento, un dominio de dimerización de Gal4 y un dominio de señal de localización nuclear; en el que el ácido nucleico exógeno se incorpora en el ADN de la célula y es expresado por la célula, comprendiendo la sonda de ácido nucleico un par de ADN monocatenarios complementarios (ADNmcc) correspondiente al sitio diana.
- 10 2. El método de la reivindicación 1, en el que el ácido nucleico exógeno está exento de secuencias promotoras.
- 15 3. El método de la reivindicación 1, en el que la molécula de fusión se selecciona del grupo que consiste en SEC ID N° 3, SEC ID N° 5 y sustituciones conservadoras de las mismas; o en el que la molécula de fusión comprende un enlazador no peptídico sintético.
- 20 4. El método de la reivindicación 1, en el que la sonda se dirige a una mutación en el ADN de la célula, y el ácido nucleico exógeno comprende una secuencia de tipo silvestre correspondiente a la mutación y/o en el que el ácido nucleico exógeno no es homólogo con respecto al ADN de la célula.
- 25 5. El método de la reivindicación 1, en el que la célula se selecciona del grupo que consiste en un célula de vertebrado, una célula de mamífero, una célula porcina, una célula humana, una célula vegetal y una célula madre.
- 30 6. Un método de producción de un animal transgénico no humano que comprende el método de la reivindicación 1.
- 35 7. Una composición purificada que comprende un filamento de nucleoproteína de una sonda y una molécula de fusión proteica, en la que la sonda comprende un par de ADN monocatenarios complementarios, y la molécula de fusión comprende un dominio de RecA o Rad51 con actividad recombinasa, un dominio de dimerización de Gal4 y un dominio de señal de localización nuclear, estando la composición exenta de secuencias de ADN que se unen específicamente al dominio de dimerización de Gal4.
- 40 8. La composición de la reivindicación 7 que comprende además ADN bicatenario exógeno.
9. Un uso de una molécula de fusión para crear *in vitro* o *in vivo* una rotura de cadena doble de un ADN bicatenario en un sitio diana de un cromosoma, comprendiendo la molécula de fusión un dominio de recombinasa de RecA o Rad51, un dominio de dimerización de Gal4 y un dominio de señal de localización nuclear.
10. El uso de la reivindicación 9 que comprende además la introducción de un ADN bicatenario exógeno en el cromosoma, en el sitio de la rotura de la cadena doble.

FIG. 1

Secuencia de aminoácidos de NLS/RecA/Gal 4 (SEC ID N° 3)

MPPKKKRKVEDPKMAIDENKQKALAAALGQIEKQFGKGS
IMRLGEDRSMDVETISTGSLSLDIALGAGGLPMGRIVEIYG
PESSGKTTTLTLQVIAAAQREGKTCAFIDAEHALDPIYARKL
GVDIDNLLCSQPDTGEQALEICDALARSGAVDVIVVDSVA
ALTPKAEIEGEIGDSHMGLAARMMSQAMRKLGNLQSN
TLLIFINQIRMKIGVMFGNPETTTGGNALKFYASVRLDIRR
IGAVKEGENVVGSETRVKVVKNKIAAPFKQAEFQILYGEG
INFYGELVDLGVKEKLIKAGAWYSYKGEKIGQGKANAT
AWLKDNPETAKEIEKKVRELLSNPNSTPDFSVDDSEGVA
ETNEDFASMKLLSSIEQACDICRLKCLKCSKEKPKCAKCL
KNNWECRYSPKTKRSPLTRAHLTEVESRLERLEQLFLLIFP
REDLDMILKMDSLQDIKALLTGLFVQDNVNKDAVTDRLA
SVETDMPLTLRQHRISATSSSEESSNKGQRQLTVSAAALE
HHHHHHHH

1-13 aa: NLS
14-366 aa: RecA
369-515 aa: Gal4
521-528 aa: marcador de His
otros aa: enlazador

FIG. 2

Secuencia de nucleótidos que codifica NLS/RecA/Gal 4 (SEC ID N° 4)

atgccacctaaaaagaagagaaaggtagaagaccccaagatggctatcgacgaaaacaaacagaaagcgttggcggcagca
ctgggccagattgagaaacaatttggtaaaggctccatcatgcgcctgggtgaagaccgttccatggatgtgaaaccatctcta
ccggttcgcttctactggatcgcgcttggggcaggtggtctgccgatgggcccgtatcgtcgaaatctacggaccggaatctc
cgttaaaaccacgctgacgctgcaggtgatccgcagcgcagcgtgaaggtaaaacctgtgcgtttatcgatctgaaacacg
cgctggaccaatctacgcacgfaaactggcgtcgatafcgataacctgctgtgctcccagccggacaccggcggagcaggc
actggaatctgtgacgcctggcgcgttctggcgcagtagacgttatcgtcgttgactccgtggcggcactgaccccgaaagc
ggaaatcgaaggcgaaatcggcgaactcacaatgggccttgcggcacgtatgatgagccaggcgatgcgtaagctggcgggt
aacctgaagcagtcacaacacgctgctgatctcatcaaccagatccgatgaaaattggtgfgatgttcggtaaccggaaacca
ctaccggtggtaacgcgctgaaattctacgcctctgttcgtctcgacatccgtcgtatcggcgcgggtgaaagagggcgaaaacg
tgggtggtagcgaaaccgcgfgaaagtggtagaagaacaaaatcgctgcgccgtttaaaccaggctgaattccagatcctctacg
gcgaaggatcaactctacggcgaactggtgacctggcgtgaaaagagaagctgatcgagaagcaggcgcgtggtfacag
ctacaaaggtagaagatcggtcagggtaaagcgaatgcgactgcctggctgaaagataaccggaaaccgcgaaagagat
cgagaagaaagtagctgagttgctgctgagcaaccgaaactcaacccggatttctctgtagatgatagcgaaggcgtagcag
aaactaacgaagatttctagcatgaagctactgtctctatcgaacaagcatgcatatttgcgacttaaaaagctcaagtgtc
caaagaaaaaccgaagtgcgccaagtgtctgagaacaactgggagtgctcactctccaaaacaaaaggctccgctg
actagggcacatctgacagaagtggaatcaaggctagaagactggaacagctatttctactgattttcctcgagaagacctga
catgatttgaaaatggattcttacaggatataaaagcattgttaacaggattatttgtacaagataatgtaataaagatgccgca
cagatagattggctttagtgagactgatatgccttaacattgagacagcatagaataagtgcgacatcatcatcgaagagag
tagtaacaaaggtaaaagacagttgactgtatcggcggccgcactcgagcaccaccaccaccaccaccac

FIG. 3

Secuencia de aminoácidos de NLS/RecA (SEC ID N° 5)

M P P K K K R K V E D P K M A I D E N K Q K A L A A A L G Q
I E K Q F G K G S I M R L G E D R S M D V E T I S T G S L S
L D I A L G A G G L P M G R I V E I Y G P E S S G K T T L T
L Q V I A A A Q R E G K T C A F I D A E H A L D P I Y A R K
L G V D I D N L L C S Q P D T G E Q A L E I C D A L A R S G
A V D V I V V D S V A A L T P K A E I E G E I G D S H M G L
A A R M M S Q A M R K L A G N L K Q S N T L L I F I N Q I R
M K I G V M F G N P E T T T G G N A L K F Y A S V R L D I R
R I G A V K E G E N V V G S E T R V K V V K N K I A A P F K
Q A E F Q I L Y G E G I N F Y G E L V D L G V K E K L I E K
A G A W Y S Y K G E K I G Q G K A N A T A W L K D N P E T A
K E I E K K V R E L L L S N P N S T P D F S V D D S E G V A
E T N E D F A S A A A L E H H H H H H H H

(SEC ID N° 5)

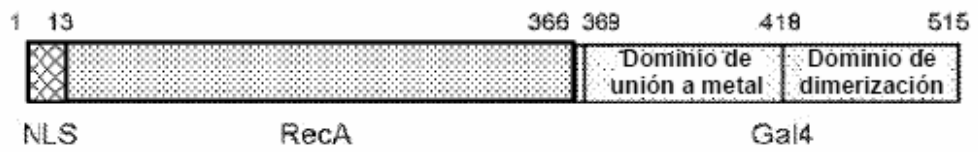
FIG. 4

Secuencia de aminoácidos de NLS/RecA (SEC ID N° 6)

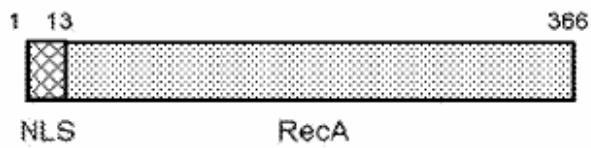
atgccacctaataaagaagagaaaggtagaagacccaagatggctatcgacgaaaacaa
acagaaagcgttggcggcagcactgggcccagattgagaaacaatttggtaaaggctcca
tcatgcgcctgggtgaagaccgttccatggatgtggaaccatctctaccggttcgctt
tcaactggatatacgcgcttggggcaggtggctctgccgatgggcccgatcgtcgaatcta
cggaccggaatcttccggtaaaaccacgctgacgctgcaggtgatcgccgcagcgcagc
gtgaaggtaaaacctgtgcgtttatcgatgctgaacacgcgctggaccaatctacgca
cgtaaactgggcgctgataatcgataacctgctgtgctcccagccggacaccggcgagca
ggcactggaaatctgtgacgccctggcgcgcttctggcgcagtagacgttatcgtcgttg
actccgtggcggcactgacgccgaaagcggaaatcgaaggcgaaatcggcgactctcac
atgggccttgcggcacgtatgatgagccaggcgtatgcgtaagctggcgggtaacctgaa
gcagtccaacacgctgctgatcttcatcaaccagatccgatatgaaaattgggtgtgatgt
tcggtaaccocggaaaccactaccggtggtaacgcgctgaaattctacgcctctgttcgt
ctcgacatccgtcgtatcggcgcggtgaaagagggcgaaaacgtgggtgggtagcgaac
ccgctgaaagtgggtgaagaacaaaatcgtcgcgcgctttaaaccaggctgaattccaga
tcctctacggcgaaggtatcaacttctacggcgaactggttgacctgggcgtaaaagag
aagctgatcgagaaagcaggcgcggtggtacagctacaaaggtgagaagatcggtcaggg
taaagcgaatgcgactgcctggctgaaagataaccggaaaccgcgaaagagatcgaga
agaaagtacgtgagttgctgctgagcaaccgaactcaacgccggatttctctgtagat
gatagcgaaggcgtagcagaaactaacgaagatttgcggccgcactcgagcaccacca
ccaccaccaccaccac

FIG. 5

A. NLS-RecA-Gal4



B. NLS-RecA



C. NLS-RecA-Gal4 Δ DD

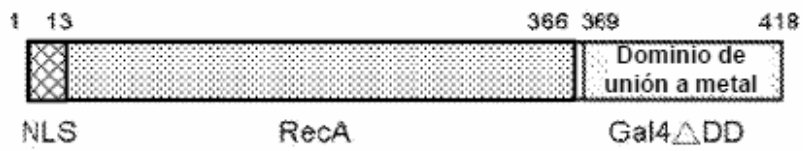
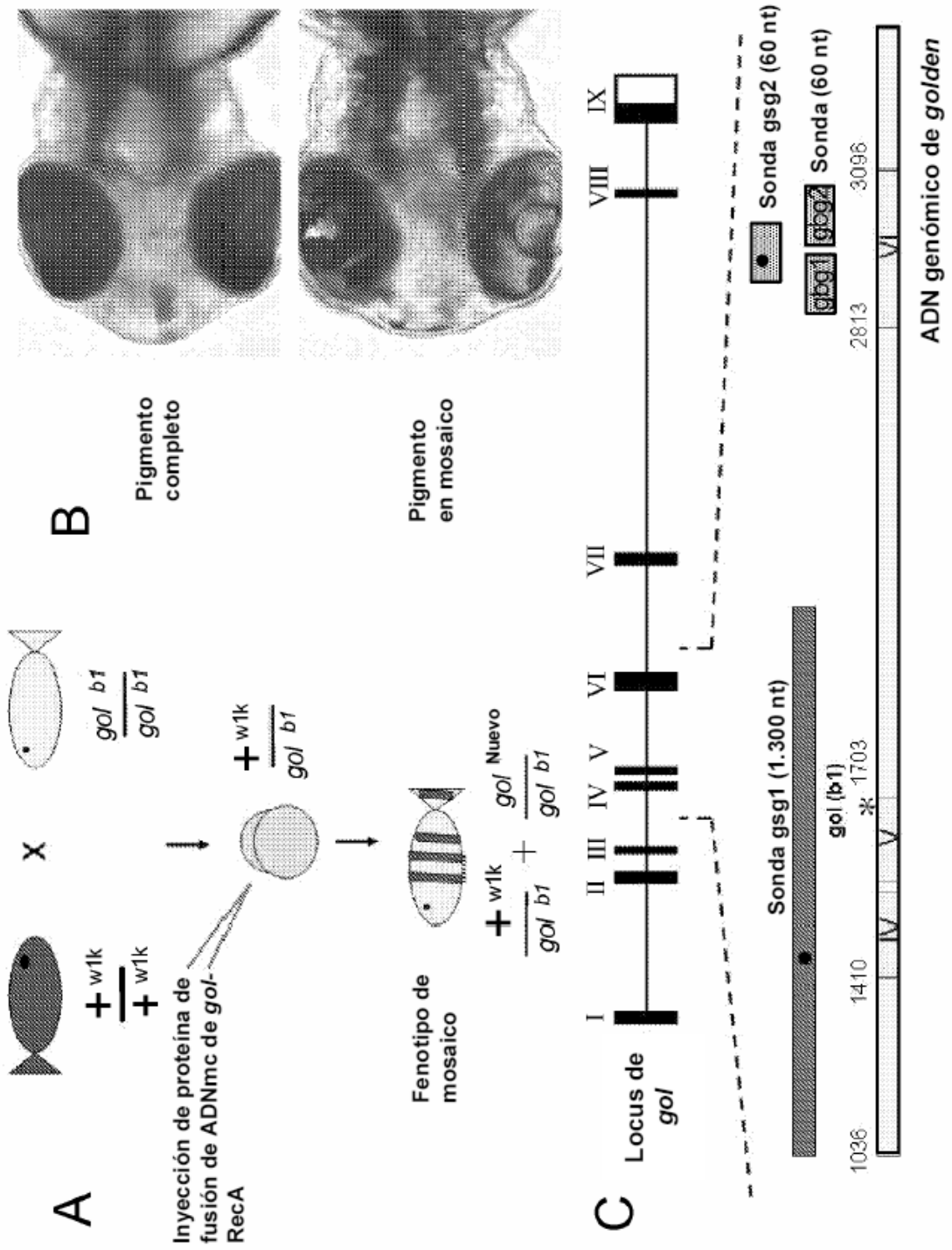


FIG. 6



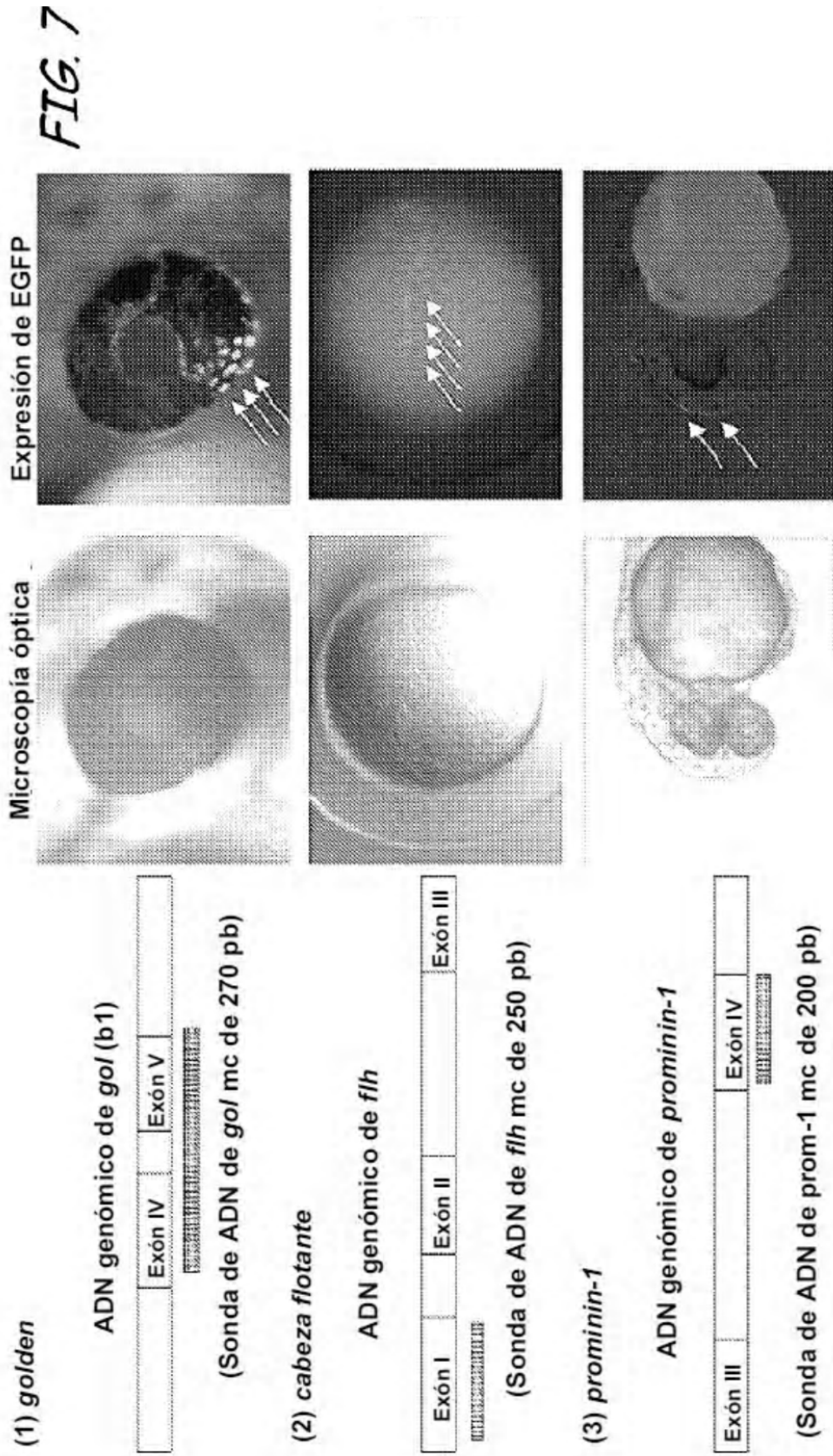


FIG. 8

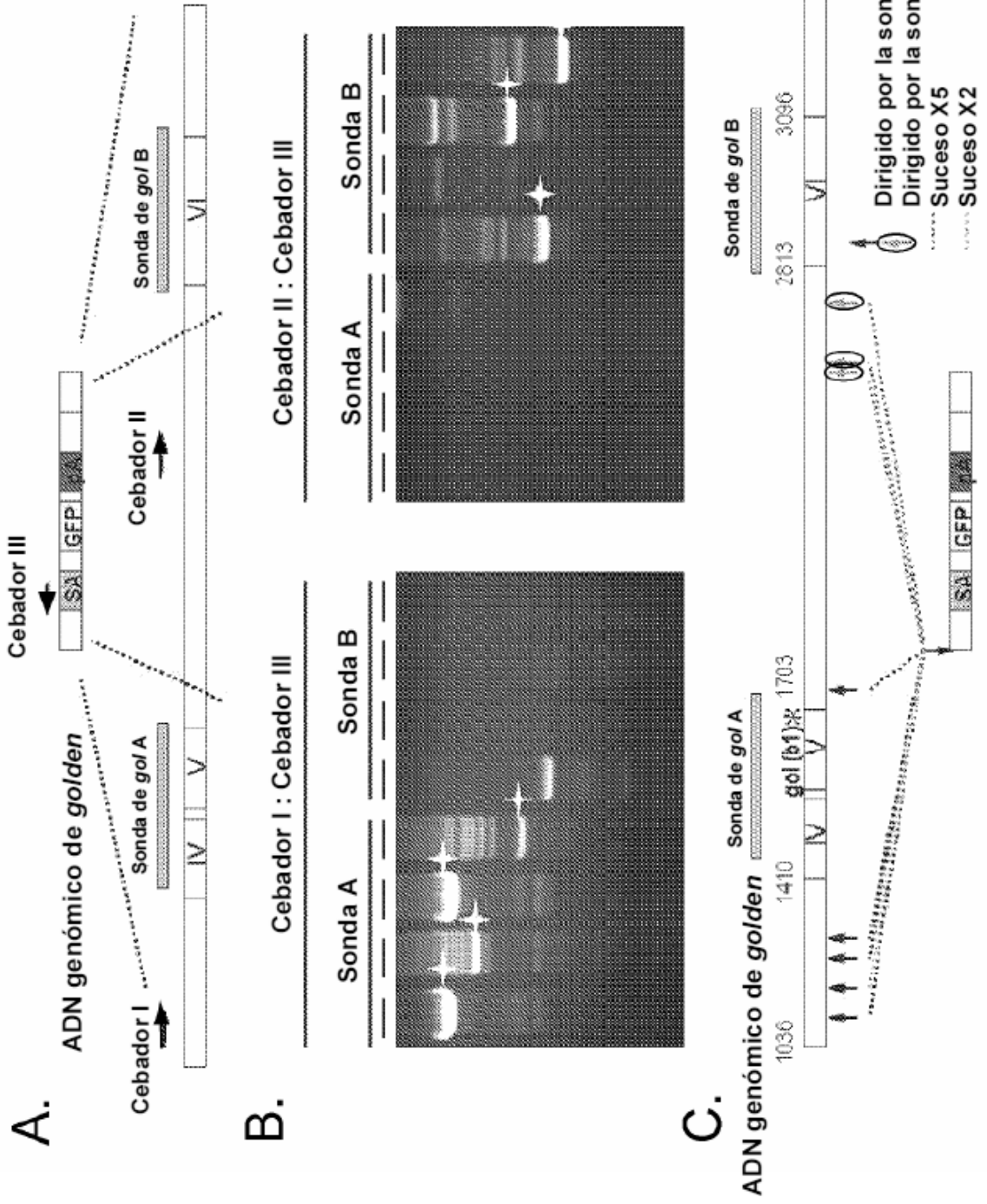
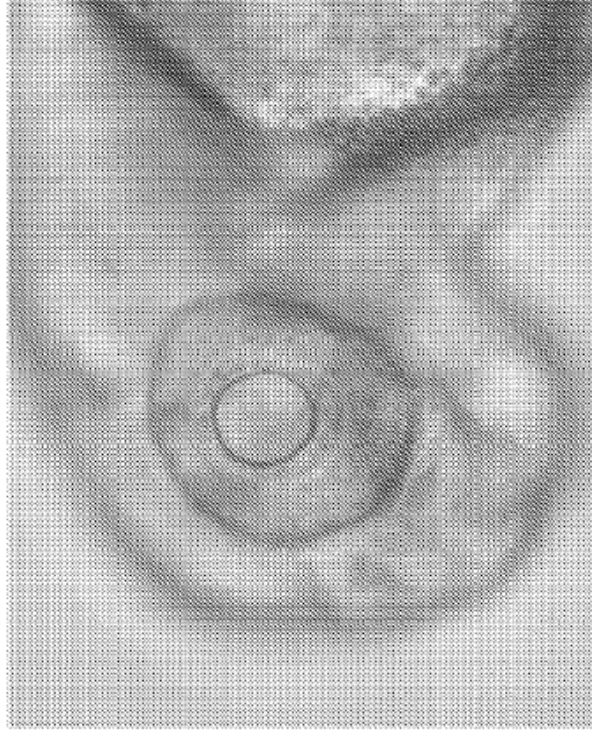
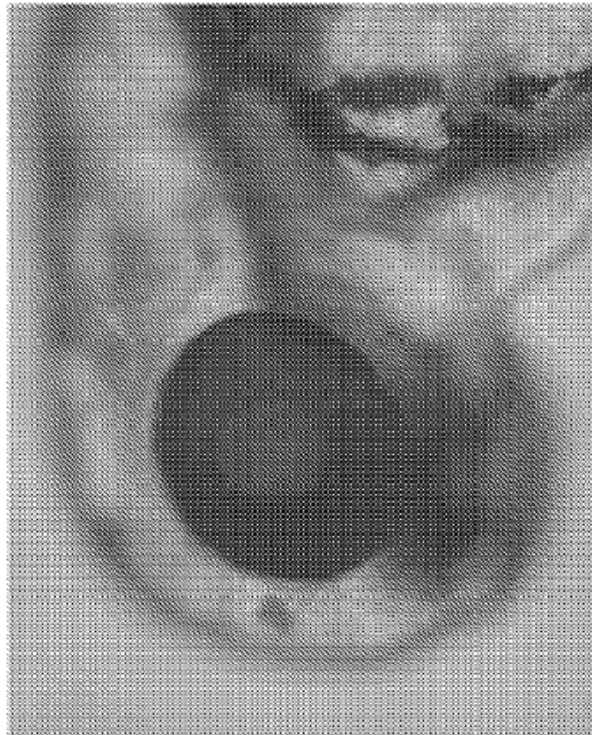


FIG. 9



**F1 de un cruce de ensayo
Pigmento luminoso**



**Control de tipo silvestre
Pigmento normal**

