

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 576 128**

51 Int. Cl.:

C12N 15/63 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **12.12.2013 E 13814368 (0)**

97 Fecha y número de publicación de la concesión europea: **09.03.2016 EP 2931898**

54 Título: **Modificación por tecnología genética y optimización de sistemas, métodos y composiciones para la manipulación de secuencias con dominios funcionales**

30 Prioridad:

12.12.2012 US 201261736527 P
02.01.2013 US 201361748427 P
30.01.2013 US 201361758468 P
25.02.2013 US 201361769046 P
15.03.2013 US 201361802174 P
15.03.2013 US 201361791409 P
28.03.2013 US 201361806375 P
20.04.2013 US 201361814263 P
06.05.2013 US 201361819803 P
28.05.2013 US 201361828130 P
17.06.2013 US 201361835931 P
17.06.2013 US 201361835936 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.07.2016

73 Titular/es:

THE BROAD INSTITUTE, INC. (33.3%)
415 Main Street
Cambridge, MA 02142, US;
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
(33.3%) y
PRESIDENT AND FELLOWS OF HARVARD
COLLEGE (33.3%)

72 Inventor/es:

ZHANG, FENG;
CONG, LE;
PLATT, RANDALL, JEFFREY;
RAN, FEI y
SANJANA, NEVILLE, ESPI

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 576 128 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

Modificación por tecnología genética y optimización de sistemas, métodos y composiciones para la manipulación de secuencias con dominios funcionales

CAMPO DE LA INVENCIÓN

5 La presente invención se refiere, en general, a la modificación por tecnología genética y la optimización de sistemas, métodos y composiciones utilizados para el control de la expresión génica que implican la focalización de la secuencia, tales como la perturbación del genoma o la edición de genes, que se refieren a Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Interespaciadas (CRISPR) y componentes de los mismos.

ANTECEDENTES DE LA INVENCIÓN

10 Recientes progresos en las técnicas de secuenciación del genoma y métodos de análisis han acelerado significativamente la capacidad para catalogar y mapear factores genéticos asociados con una diversa gama de funciones biológicas y enfermedades. Se necesitan tecnologías de focalización del genoma, precisas para permitir una modificación por tecnología genética inversa sistemática de las variaciones genéticas causales, permitiendo una perturbación selectiva de elementos genéticos individuales, así como para avanzar en la biología de síntesis,
 15 biotecnológica y aplicaciones médicas. Aunque las técnicas de edición del genoma, tales como los dedos de zinc diseñadores, efectores del tipo activador de la transcripción (TALEs), o meganucleasas migratorias están disponibles para la producción de perturbaciones del genoma focalizado, sigue habiendo una necesidad de nuevas tecnologías de la ingeniería del genoma que son asequibles, fáciles de instalar, expansibles y susceptibles de fijar como objetivo múltiples posiciones dentro del genoma eucariota.

SUMARIO DE LA INVENCIÓN

El CRISPR/Cas o el sistema CRISPR-cas (ambos términos se utilizan indistintamente a lo largo de toda esta solicitud) no requieren la generación de proteínas modificadas para fijar como objetivo secuencias diana específicas, sino más bien una sola enzima Cas puede ser programada por una molécula de ARN corta para reconocer una diana de ADN específica, en otras palabras, la enzima Cas puede ser reclutada a una diana de ADN específica
 25 utilizando dicha molécula de ARN corta. Añadiendo el sistema CRISPR-Cas al repertorio de técnicas de secuenciación del genoma y métodos de análisis puede simplificar significativamente la metodología y acelerar la capacidad de catalogar y mapear los factores genéticos asociados con una diversa gama de funciones biológicas y enfermedades. Para utilizar el sistema CRISPR-Cas de manera efectiva para la edición del genoma sin efectos perjudiciales, es fundamental entender aspectos de la modificación genética y la optimización de estas herramientas de modificación genética del genoma, que son aspectos de la presente descripción.
 30

De acuerdo con la presente invención se proporciona una composición que no se produce de forma natural o generada por tecnología genética de acuerdo con la reivindicación 1. Aspectos adicionales de la invención se describen en las reivindicaciones dependientes. La invención también proporciona métodos y usos de la composición, así como células huésped eucariotas de acuerdo con las restantes reivindicaciones independientes.
 35 Cualesquiera aspectos que caen fuera del alcance de las reivindicaciones se proporcionan como descripción.

Existe una necesidad urgente de sistemas y técnicas, alternativos y robustos, para la focalización de la secuencia de dirección con una amplia gama de aplicaciones. Aspectos de esta invención satisfacen esta necesidad y proporcionan ventajas relacionadas. Un complejo de CRISPR a modo de ejemplo puede comprender una enzima CRISPR que forma complejo con una secuencia guía hibridada con una secuencia diana dentro del polinucleótido diana. La secuencia guía está enlazada a una secuencia de emparejamiento tracr que, a su vez, se hibrida a una secuencia tracr.
 40

En un aspecto, la invención se refiere a uno o más elementos de un sistema CRISPR que tiene una funcionalidad mejorada o modificada. El complejo CRISPR de la invención proporciona un medio eficaz para la modificación de un polinucleótido diana. El complejo CRISPR de la invención tiene una amplia diversidad de utilidades, incluyendo modificar (por ejemplo, suprimir, insertar, translocar, inactivar, activar) un polinucleótido diana en una multiplicidad de tipos de células. Como tal, el complejo CRISPR de la invención tiene un amplio espectro de aplicaciones, p. ej., en la edición del gen o genoma, la terapia génica, el descubrimiento de fármacos, la detección de fármacos, el diagnóstico de enfermedades, y el pronóstico. Un complejo CRISPR a modo de ejemplo puede comprender una enzima de CRISPR que forma complejo con una secuencia guía hibridada con una secuencia diana dentro del
 45

polinucleótido diana. La secuencia guía está enlazada a una secuencia de emparejamiento tracr, que a su vez se hibrida a una secuencia tracr.

En algunas realizaciones, un sistema CRISPR/Cas puede comprender: (a) un primer elemento regulador enlazado operativamente a una secuencia de emparejamiento tracr y uno o más sitios de inserción para insertar una o más secuencias de guía aguas arriba de la secuencia de emparejamiento tracr, en donde, cuando se expresa, la secuencia guía dirige la unión específica para la secuencia de un complejo CRISPR a una secuencia diana en una célula eucariota, en donde el complejo CRISPR puede comprender una enzima de CRISPR que forma complejo con (1) la secuencia guía que se hibrida a la secuencia diana y (2) la secuencia de emparejamiento tracr que se hibrida a la secuencia de tracr; y (b) un segundo elemento regulador enlazado de forma operativa a una secuencia codificadora de la enzima que codifica dicha enzima CRISPR que puede comprender una secuencia de localización nuclear; en donde los componentes (a) y (b) están situados en el mismo o en diferentes vectores del sistema. En algunas realizaciones, el componente (a) puede comprender, además, la secuencia tracr aguas abajo de la secuencia de emparejamiento tracr bajo el control del primer elemento regulador. En algunas realizaciones, el componente (a) puede comprender, además, dos o más secuencias guía enlazadas operativamente al primer elemento regulador, en donde, cuando se expresan, cada una de las dos o más secuencias guía dirigen la unión específica para la secuencia de un complejo CRISPR a una secuencia diana diferente en una célula eucariota. En algunas realizaciones, el sistema puede comprender la secuencia de tracr bajo el control de un tercer elemento regulador tal como un promotor de polimerasa III. En algunas realizaciones, la secuencia de tracr exhibe al menos 50%, 60%, 70%, 80%, 90%, 95% o 99% de complementariedad de la secuencia a lo largo de la longitud de la secuencia de emparejamiento tracr cuando se alinean de manera óptima. En algunas realizaciones, la enzima CRISPR puede comprender una o más secuencias de localización nuclear de una fuerza suficiente para impulsar la acumulación de dicha enzima CRISPR en una cantidad detectable en el núcleo de una célula eucariota. Sin desear estar limitado por la teoría, se cree que una secuencia de localización nuclear no es necesaria para la actividad de CRISPR en eucariotas, pero que la inclusión de tales secuencias aumenta la actividad del sistema. En algunas realizaciones, la enzima CRISPR es una enzima del sistema CRISPR tipo II. En algunas realizaciones, la enzima CRISPR es una enzima Cas9. En algunas realizaciones, la enzima Cas9 es Cas9 de *S. pyogenes* o *S. thermophilus* y puede incluir Cas9 mutada, derivada de uno de estos organismos. La enzima puede ser un homólogo u ortólogo de Cas9. En algunas realizaciones, la enzima CRISPR está optimizada en el codón para la expresión en una célula eucariota. En algunas realizaciones, la enzima CRISPR dirige la escisión de una o dos cadenas en la localización de la secuencia diana. En algunas realizaciones, la enzima CRISPR carece de la actividad de escisión de la cadena de ADN. En algunas realizaciones, el primer elemento regulador es un promotor de la polimerasa III. En algunas realizaciones, el segundo elemento regulador es un promotor de la polimerasa II. En algunas realizaciones, la secuencia guía es de al menos 15, 16, 17, 18, 19, 20, 25 nucleótidos, o entre 10 y 30, o entre 15 y 25, o entre 15 y 20 nucleótidos de longitud. En general, y a lo largo de esta memoria, el término "vector" se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que está enlazada. Vectores incluyen, pero no se limitan a moléculas de ácido nucleico que son de cadena sencilla, de doble cadena, o parcialmente de doble cadena; moléculas de ácido nucleico que comprenden uno o más extremos libres, sin extremos libres (p. ej., circulares); moléculas de ácido nucleico que comprenden ADN, ARN, o ambos; y otras variedades de polinucleótidos conocidas en la técnica. Un tipo de vector es un "plásmido", que se refiere a un bucle de ADN de doble cadena circular, en el que pueden insertarse segmentos adicionales de ADN, tal como mediante técnicas de clonación molecular estándares. Otro tipo de vector es un vector viral, en donde secuencias de ADN o de ARN derivadas de virus están presentes en el vector para el empaquetamiento en un virus (p. ej., retrovirus, retrovirus defectuosos en la replicación, adenovirus, adenovirus defectuosos en la replicación y virus adeno-asociados). Los vectores virales también incluyen polinucleótidos portados por un virus para la transfección en una célula huésped. Determinados vectores son capaces de replicación autónoma en una célula huésped en la que se introducen (p. ej., vectores bacterianos que tienen un origen bacteriano de replicación y vectores episomales de mamíferos). Otros vectores (p. ej., vectores no episomales de mamíferos) se integran en el genoma de una célula huésped tras la introducción en la célula huésped, y con ello se replican junto con el genoma del huésped. Además de ello, determinados vectores son capaces de dirigir la expresión de genes a los que están operativamente enlazados. A vectores de este tipo se les alude en esta memoria como "vectores de expresión". Vectores de expresión comunes de utilidad en técnicas de ADN recombinante están a menudo en forma de plásmidos.

Los vectores de expresión recombinante pueden comprender un ácido nucleico de la invención en una forma adecuada para la expresión del ácido nucleico en una célula huésped, lo que significa que los vectores de expresión recombinante incluyen uno o más elementos reguladores, que pueden seleccionarse sobre la base de las células huéspedes a utilizar para la expresión, que están operativamente ligadas a la secuencia de ácido nucleico a expresar. Dentro de un vector de expresión recombinante, "operativamente ligada" pretende significar que la secuencia de nucleótidos de interés está enlazada al o a los elementos reguladores de una manera que permita la expresión de la secuencia de nucleótidos (p. ej., en un sistema de transcripción/traducción in vitro o en una célula huésped cuando el vector se introduce en la célula huésped).

La expresión "elemento regulador" pretende incluir promotores, potenciadores, sitios de entrada al ribosoma interno (IRES), y otros elementos de control de la expresión (p. ej., señales de terminación de la transcripción tales como señales de poliadenilación y secuencias de poli-U). Elementos reguladores de este tipo se describen, por ejemplo, en Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Los elementos reguladores incluyen aquellos que dirigen la expresión constitutiva de una secuencia de nucleótidos en muchos tipos de células huéspedes y aquellos que dirigen la expresión de la secuencia de nucleótidos sólo en ciertas células huéspedes (p. ej., secuencias reguladoras específicas para el tejido). Un promotor específico para el tejido puede dirigir la expresión principalmente en un tejido de interés deseado tal como el músculo, neurona, hueso, piel, sangre, órganos específicos (p. ej., hígado, páncreas) o tipos de células particulares (p. ej., linfocitos). Los elementos reguladores también pueden dirigir la expresión de una manera dependiente del tiempo, tal como de una manera dependiente del ciclo celular o dependiente de la fase de desarrollo, que puede o puede no ser también ser específico para el tejido o tipo de célula. En algunas realizaciones, un vector puede comprender uno o más promotores pol III (p. ej., 1, 2, 3, 4, 5 o más promotores pol III), uno o más promotores pol II (p. ej., 1, 2, 3, 4, 5 o más promotores pol II), uno o más promotores pol I (p. ej., 1, 2, 3, 4, 5 o más promotores pol I), o combinaciones de los mismos. Ejemplos de promotores pol III incluyen, pero no se limitan a promotores U6 y H1. Ejemplos de promotores pol II incluyen, pero no se limitan al promotor LTR del virus del sarcoma de Rous retroviral (RSV) (opcionalmente con el potenciador RSV), el promotor de citomegalovirus (CMV) (opcionalmente con el potenciador CMV) [véase, p. ej., Boshart et al, Cell, 41:521-530 (1985)], el promotor SV40, el promotor de dihidrofolato reductasa, el promotor β -actina, el promotor fosfoglicero1 quinasa (PGK) y el promotor EFl α . También se incluyen en la expresión "elemento regulador" elementos potenciadores tales como WPRE; potenciadores de CMV; el segmento R-U5' en LTR de HTLV-1 (Mol. Cell. Biol., Vol. 8(1), págs. 466-472, 1988); potenciador de SV40; y la secuencia del intrón entre los exones 2 y 3 de β -globina de conejo (Proc. Natl. Acad. Sci. USA, Vol. 78(3), págs. 1527-31, 1981). Se apreciará por los expertos en la técnica que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped a transformar, el nivel de expresión deseado, etc. Un vector puede introducirse en células huésped para de ese modo producir transcritos, proteínas o péptidos, incluyendo proteínas o péptidos de fusión, codificados por ácidos nucleicos tal como se describe en esta memoria (p. ej., transcritos de repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR), proteínas, enzimas, formas mutantes de los mismos, proteínas de fusión de los mismos, etc.).

Aspectos de la invención se refieren a métodos para mejorar o modificar la especificidad de la diana de una enzima de CRISPR (preferentemente una enzima Cas9) que puede comprender: a) seleccionar la enzima de CRISPR que tiene un tamaño más pequeño para un fácil empaquetamiento en vectores de suministro; o b) generar enzimas de CRISPR quiméricas; o c) utilizar enzimas de CRISPR mutadas (preferiblemente Cas9). Otros aspectos de la invención también se refieren a métodos y composiciones para mejorar la especificidad de la diana que implica dos enzimas de CRISPR (doble nickase), mejorando la estructura química de sgrRNA haciendo mejora en otros componentes aparte de la guía ARN.

La invención comprende una composición que no se produce de forma natural o generada por tecnología genética que comprende un sistema de vectores que puede comprender uno o más vectores que comprende I. un primer elemento regulador enlazado operativamente a una secuencia de polinucleótidos de ARN quimérico (chiRNA) del sistema CRISPR/Cas, en el que la secuencia de polinucleótidos puede comprender (a) una secuencia guía capaz de hibridarse a una secuencia diana en una célula, (b) una secuencia de emparejamiento tracr, y (c) una secuencia tracr, y II. un segundo elemento regulador enlazado operativamente a una secuencia codificadora de la enzima que codifica una enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear, en donde (a), (b) y (c) están dispuestos en una orientación 5' a 3', en donde los componentes I y II están situados en el mismo o en diferentes vectores del sistema, en donde, cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión, específica para la secuencia, de un complejo CRISPR a la secuencia diana, en donde el complejo CRISPR comprende la enzima de CRISPR que forma complejo con (1) la secuencia guía que se hibrida a la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia tracr, en donde la enzima de CRISPR puede comprender dos o más mutaciones, de manera que la enzima tiene una actividad nucleasa alterada en comparación con la enzima de tipo salvaje, y en donde la secuencia codificadora de la enzima codifica, además, uno o más dominios funcionales heterólogos.

La invención comprende, además, una composición de sistema de enzimas CRISPR de dos componentes multiplexado, que comprende un sistema de vectores que puede comprender uno o más vectores que comprende I. un primer elemento regulador enlazado operativamente a una secuencia de polinucleótidos de ARN quimérico (chiRNA) del sistema CRISPR/Cas, en el que la secuencia de polinucleótidos comprende (a) una secuencia guía capaz de hibridarse a una secuencia diana en una célula, (b) una secuencia de emparejamiento tracr, y (c) una secuencia tracr, y II. un segundo elemento regulador enlazado operativamente a una secuencia codificadora de la enzima que codifica una enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear, en donde (a), (b) y (c) están dispuestos en una orientación 5' a 3', en donde los componentes I y II están

situados en el mismo o en diferentes vectores del sistema, en donde, cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión, específica para la secuencia, de un complejo CRISPR a la secuencia diana, en donde el complejo CRISPR comprende la enzima de CRISPR que forma complejo con (1) la secuencia guía que se hibrida a la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia tracr, en donde la enzima de CRISPR puede comprender dos o más mutaciones, de manera que la enzima tiene una actividad nucleasa alterada en comparación con la enzima de tipo salvaje, en donde la secuencia codificadora de la enzima codifica, además, uno o más dominios funcionales heterólogos, y en donde en la composición del sistema multiplexado se utilizan múltiples secuencias de polinucleótidos chiRNA.

La invención comprende, además, una composición que no se produce de forma natural o generada por tecnología genética que comprende un sistema de vectores que puede comprender uno o más vectores que comprende I. un primer elemento regulador enlazado operativamente a (a) una secuencia guía capaz de hibridarse a una secuencia diana en una célula, y (b) al menos una o más secuencias de emparejamiento tracr, II. un segundo elemento regulador enlazado operativamente a una secuencia codificadora de la enzima que codifica una enzima de CRISPR y III. un tercer elemento regulador enlazado operativamente a una secuencia tracr, en donde los componentes I, II y III están localizados en el mismo o en diferentes vectores del sistema, en el que, cuando se transcriben, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, en el que el complejo de CRISPR comprende la enzima de CRISPR formando complejo con (1) la secuencia guía que se hibrida a la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida a la secuencia de tracr, en donde la enzima de CRISPR puede comprender dos o más mutaciones, de forma que la enzima tiene una actividad nucleasa alterada en comparación con la enzima de tipo salvaje, y en donde la secuencia que codifica la enzima codifica, además, uno o más dominios funcionales heterólogos.

La invención también comprende una composición de sistema de enzimas de CRISPR de tres componentes multiplexada que comprende un sistema de vectores que puede comprender uno o más vectores que comprende I. un primer elemento regulador enlazado operativamente a (a) una secuencia guía capaz de hibridarse a una secuencia diana en una célula y (b) al menos una o más secuencias de emparejamiento tracr, II. un segundo elemento regulador enlazado operativamente a una secuencia codificadora de la enzima que codifica una enzima de CRISPR y III. un tercer elemento regulador enlazado operativamente a una secuencia tracr, en donde los componentes I, II y III están localizados en el mismo o en diferentes vectores del sistema, en el que, cuando se transcriben, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, en el que el complejo de CRISPR comprende la enzima de CRISPR formando complejo con (1) la secuencia guía que se hibrida a la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida a la secuencia de tracr, en donde la enzima de CRISPR puede comprender dos o más mutaciones, de forma que la enzima tiene una actividad nucleasa alterada en comparación con la enzima de tipo salvaje, y en donde la secuencia que codifica la enzima codifica, además, uno o más dominios funcionales heterólogos, y en donde en la composición del sistema multiplexado se utilizan múltiples secuencias guía capaces de hibridarse a las múltiples secuencias diana.

En realizaciones de la invención, la enzima de CRISPR puede comprender una o más mutaciones en dos o más dominios catalíticamente activos. En otra realización la enzima de CRISPR tiene una actividad nucleasa reducida o abolida en comparación con la enzima de tipo salvaje. En otra realización, las dos mutaciones son D10A SpCas9 en un primer dominio catalíticamente activo y H840A SpCas9 en un segundo dominio catalíticamente activo o residuos correspondientes de otras enzimas de CRISPR. En otra realización, la enzima de CRISPR puede comprender dos o más mutaciones en un residuo seleccionado del grupo que consiste en D10, E762, H840, N854, N863 o D986. En otra realización, la enzima de CRISPR puede comprender dos o más mutaciones seleccionadas del grupo que comprende D10A, E762A, H840A, N854A, N863A o D986A. En una realización preferida, la enzima de CRISPR es una proteína de unión a ADN que no dirige la escisión de cualquiera de las cadenas en la ubicación de la secuencia diana. En una realización, cada una de las dos o más mutaciones es en un dominio catalíticamente activo de la enzima de CRISPR seleccionado del grupo que comprende el dominio RuvCI, RuvCII, RuvCIII o HNH. Se entenderá que a lo largo de esta solicitud muchas referencias a residuos aminoácidos específicos son los de la enzima SpCas9. La persona experta comprenderá que la invención es aplicable a otras enzimas Cas, incluyendo enzimas Cas9 de otras fuentes, y que en los casos en los que se hace referencia a residuos específicos de enzimas SpCas9 se pueden hacer alteraciones y mutaciones correspondientes en esas otras enzimas Cas. Por ejemplo, la persona experta será capaz de comparar secuencias de la SpCas9 y otras enzimas, e identificar residuos y dominios correspondientes y, por lo tanto, hacer las modificaciones adecuadas a esas enzimas.

En realizaciones adicionales, las composiciones de la invención pueden comprender al menos dos o más secuencias de localización nuclear. En un aspecto de la invención, el dominio funcional es un dominio de activación transcripcional, p. ej., VP64. En otro aspecto, el dominio funcional es un dominio represor transcripcional, p. ej., un

- dominio KRAB, un dominio SID o un dominio SID4X. En otra realización, la secuencia codificadora de la enzima codifica uno, dos, tres, cuatro, cinco o más dominios funcionales heterólogos fusionados a la enzima de CRISPR. La invención también comprende una o más secuencias de enlazador entre dos dominios. Realizaciones de la invención incluyen uno o más dominios funcionales que tienen una o más de las siguientes actividades: actividad metilasa, actividad desmetilasa, actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de la transcripción, actividad de modificación de histonas, actividad de escisión del ARN y actividad de unión a ácido nucleico. En otra realización, el dominio funcional une el ADN y/o afecta a la transcripción del ácido nucleico diana. En aspectos de la invención, la célula es una célula procariota o eucariota. En una realización preferida, la célula es una célula de mamífero o una célula de ser humano. En algunas realizaciones, la célula de mamífero puede ser una célula de roedor (por ejemplo, ratón o rata), una célula de ungulados o una célula de primates. En algunas realizaciones, la célula eucariota puede ser una célula de artrópodo (p. ej., insecto) o de nematodos. En otras realizaciones la célula puede ser una célula vegetal (incluyendo algas) o una célula fúngica. En realizaciones adicionales de la invención, la enzima de CRISPR está optimizada en codones para la expresión en una célula eucariota; la enzima de CRISPR es una enzima de CRISPR tipo II; la enzima de CRISPR es una enzima Cas9 y la enzima Cas9 es de un organismo seleccionado del grupo que comprende el género *Streptococcus*, *Campylobacter*, *Nitratifactor*, *Staphylococcus*, *Parvibaculum*, *Roseburia*, *Neisseria*, *Gluconacetobacter*, *Azospirillum*, *Sphaerochaeta*, *Lactobacillus*, *Eubacterium* o *Corynebacter*. En un aspecto adicional, los vectores del sistema son vectores virales seleccionados del grupo que comprende de un vector lentiviral, un vector adenoviral o un vector de AAV.
- La invención también comprende métodos de modular, es decir, alterar, activar, reprimir la expresión génica en un locus genómico de interés en una célula, poniendo en contacto la célula con composiciones de la invención.
- Un aspecto de la descripción se refiere a una composición que puede comprender un sistema de vectores que puede comprender uno o más vectores que pueden comprender
- I. un primer elemento regulador enlazado operativamente a una secuencia de polinucleótidos de ARN quimérico (chiRNA) del sistema CRISPR/Cas, en el que la secuencia de polinucleótidos puede comprender
 - (a) una secuencia guía capaz de hibridarse a una secuencia diana en una célula eucariota,
 - (b) una secuencia de emparejamiento tracr, y
 - (c) una secuencia tracr, y
 - II. un segundo elemento regulador enlazado operativamente a una secuencia codificadora de la enzima que codifica una enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear, en donde (a), (b) y (c) están dispuestos en una orientación 5' a 3', en donde los componentes I y II están situados en el mismo o en diferentes vectores del sistema, en donde, cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión, específica para la secuencia, de un complejo CRISPR a la secuencia diana, en donde el complejo CRISPR puede comprender la enzima de CRISPR que forma complejo con (1) la secuencia guía que se hibrida a la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia tracr, y en donde la enzima que codifica la secuencia codificadora de la enzima de CRISPR incluye, además, un dominio funcional heterólogo.
- La secuencia codificadora puede codificar uno o más dominios funcionales heterólogos (p. ej., aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más dominios, además de la enzima de CRISPR). Una proteína de fusión de la enzima de CRISPR puede comprender cualquier secuencia de proteínas adicional, y opcionalmente una secuencia de enlazador entre dos dominios. Ejemplos de dominios de proteínas que pueden fusionarse con una enzima de CRISPR incluyen, sin limitación, marcadores de epítomos, secuencias de gen informador, y dominios de proteínas que tienen una o más de las siguientes actividades: actividad metilasa, actividad desmetilasa, actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de la transcripción, actividad de modificación de histonas, actividad de escisión de ARN y actividad de unión a ácido nucleico. Ejemplos no limitantes de marcadores de epítomos incluyen marcadores de histidina (His), marcadores V5, marcadores FLAG, marcadores de hemaglutinina de la gripe (HA), marcadores Myc, marcadores VSV-G y marcadores de tiorredoxina (Trx). Ejemplos de genes informadores incluyen, pero no se limitan a, glutatión-S-transferasa (GST), peroxidasa de rábano picante (HRP), cloranfenicol acetiltransferasa (CAT), beta-galactosidasa, beta-glucuronidasa, luciferasa, proteína fluorescente verde (GFP), HcRed, DsRed, proteína fluorescente cian (CFP), proteína amarilla fluorescente (YFP) y proteínas autofluorescentes incluyendo la proteína fluorescente azul (BFP). Una enzima de CRISPR puede fusionarse a una secuencia de gen que codifica una proteína o un fragmento de una proteína que une moléculas de ADN o une otras moléculas celulares, incluyendo pero no limitada a la proteína de unión a maltosa (MBP), S-tag, fusiones del dominio de unión a ADN (DBD) de Lex A, fusiones del dominio de unión a ADN de GAL4 y fusiones de la proteína BP16 del virus herpes simplex (HSV). Dominios adicionales que pueden formar parte de una proteína de fusión que puede comprender una enzima de

CRISPR se describen en el documento US20110059502. En algunas realizaciones, se utiliza una enzima de CRISPR marcada para identificar la ubicación de una secuencia diana.

En un aspecto de la invención, el dominio funcional afecta a la transcripción del ácido nucleico diana.

5 Ejemplos no limitantes de enzimas de CRISPR incluyen Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (también conocida como Csn1 y Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homólogos de las mismas, o versiones modificadas de las mismas. Estas enzimas se conocen; por ejemplo, la secuencia de aminoácidos de la proteína Cas9 de *S. pyogenes* se puede encontrar en la base de datos SwissProt (disponible en el sitio de web uniprot.org) bajo el número de acceso Q99ZW2. En algunas realizaciones, la enzima es una enzima Cas9 de CRISPR. En algunas realizaciones, la enzima Cas9 es Cas9 de *S. pneumoniae*, *S. pyogenes* o *S. thermophilus*, y puede incluir Cas9 mutada, derivada de estos organismos. En algunas realizaciones, la enzima de CRISPR está optimizada en codones para la expresión en una célula eucariota. En algunas realizaciones, la enzima CRISPR dirige la escisión de una o dos cadenas en la ubicación de la secuencia diana. En algunas formas de realización, la enzima de CRISPR carece de actividad de escisión de la cadena de ADN.

En algunos aspectos de la invención, una enzima de CRISPR puede comprender una o más mutaciones y puede utilizarse como una proteína de unión a ADN genérico con o sin fusión a un dominio funcional. Las mutaciones pueden incluir, pero no se limitan a mutaciones catalíticas, por ejemplo, mutaciones en uno de los dominios catalíticos o mutación de residuos catalíticos. Ejemplos preferidos de mutaciones catalíticas adecuadas son el o los residuos catalíticos en el dominio RuvC I en el extremo N de Cas9 o el o los residuos catalíticos en el dominio HNH interno. En algunas realizaciones, la Cas9 es (o se deriva de) SpCas9. En tales realizaciones, las mutaciones preferidas son en cualquiera o en todas las posiciones 10, 762, 840, 854, 863 y/o 986 de SpCas9 o posiciones correspondientes en otras Cas9s (que pueden determinarse, por ejemplo, mediante herramientas de comparación de secuencias estándar). En particular, cualquiera o todas las siguientes mutaciones se prefieren en SpCas9: D10A, E762A, H840A, N854A, N863A y/o D986A; así como también se prevé la sustitución conservadora de cualquiera de los aminoácidos de sustitución. También se prefieren las mismas (o sustituciones conservativas de estas mutaciones) en las posiciones correspondientes en otras Cas9s. Particularmente preferidas son D10 y H840 en SpCas9. Sin embargo, en otras Cas9s, también se prefieren los residuos correspondientes a SpCas9 D10 y H840.

En un aspecto más ventajoso de la invención la enzima de CRISPR mutada puede fusionarse a un dominio de activación transcripcional. En un aspecto de la invención, el dominio de activación transcripcional puede ser VP64. Otros aspectos de la invención se refieren a la enzima de CRISPR mutada que se fusiona a los dominios que incluyen pero no están limitados a un represor de la transcripción, una recombinasa, una transposasa, un remodelador de histona, una ADN metiltransferasa, un criptocromo, un dominio inducible/controlable por la luz o un dominio químicamente inducible/controlable.

35 En algunos aspectos de la invención, la enzima de CRISPR se compone de menos de cuatro mil, y en algunos aspectos menos de mil aminoácidos. Enzimas de este tipo se pueden proporcionar fusionadas a un dominio funcional heterólogo, o no. Cuando la enzima se fusiona a un dominio funcional heterólogo, el tamaño de la enzima (menos de cuatro mil, menos de mil aminoácidos) se refiere a la porción de CRISPR de la proteína de fusión. En determinadas realizaciones descritas en esta memoria, la invención o el método se pone en práctica en un organismo o sujeto. En determinadas realizaciones, el organismo o sujeto es un eucariota o un eucariota no humano. En determinadas realizaciones, el organismo o sujeto es una planta. En determinadas realizaciones, el organismo o sujeto es un mamífero o un mamífero no humano. En determinadas realizaciones, el organismo o sujeto son algas. En algunas realizaciones, el organismo o sujeto puede ser un roedor (por ejemplo, ratón o rata), un ungulado o un primate. En algunas realizaciones, el organismo o sujeto puede ser un artrópodo (p. ej., insecto) o nematodo. En otras realizaciones el organismo o sujeto puede ser una planta o un hongo.

Estas y otras realizaciones se describen o son evidentes a partir de y quedan abarcadas por la siguiente Descripción Detallada.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

50 Las características nuevas de la invención se exponen con particularidad en las reivindicaciones adjuntas. Una mejor comprensión de las características y ventajas de la presente invención se obtendrá haciendo referencia a la siguiente descripción detallada que recoge realizaciones ilustrativas, en las que se utilizan los principios de la invención, y los dibujos adjuntos de los cuales:

- 5 La **Figura 1** muestra un esquema de nucleasa Cas9 guiada por ARN. La nucleasa Cas9 de *Streptococcus pyogenes* (amarillo) fija como objetivo el ADN genómico mediante un ARN guía sintético (SgRNA) que consiste en una secuencia guía de 20 nt (azul) y un armazón (rojo). Los pares de bases de la secuencia guía la diana de ADN (azul), directamente aguas arriba de un motivo adyacente al protoespaciador 5'-NGG requerido (PAM; magenta) y Cas9 media en una rotura de doble cadena (DSB) ~ 3 pb aguas arriba del PAM (triángulo rojo).
- Las **Figuras 2A-2F** muestran un sistema CRISPR a modo de ejemplo, un posible mecanismo de acción, una adaptación de ejemplo para la expresión en células eucariotas, y los resultados de las pruebas que evalúan la localización nuclear y la actividad de CRISPR.
- La **Figura 3A-D** es un árbol filogenético de genes Cas.
- 10 Las **Figuras 4A-4F** muestran el análisis filogenético que revela cinco familias de Cas9s, incluyendo tres grupos de grandes Cas9s (~ 1400 aminoácidos) y dos de pequeñas Cas9s (~ 1100 aminoácidos).
- La **Figura 5** muestra una construcción esquemática, en la que el dominio de activación de la transcripción (VP64) está fusionado a Cas9 con dos mutaciones en los dominios catalíticos (D10 y H840).
- 15 La **Figura 6** muestra una representación gráfica de la activación transcripcional tras co-transfección de la Cas9-VP64 con ARN crispr quimérico generado por PCR (chiRNA) en células 293. La evaluación se llevó a cabo 72 horas después de la transfección utilizando RT-qPCR.
- La **Figura 7** muestra un cierto número de vectores que incorporan genes Cas9 mutantes con marcadores VP64, NLS y GFP.
- 20 La **Figura 8** muestra la ubicación de construcciones Cas9-VP64-GFP en células 293 tal como se evaluó por un microscopio de fluorescencia 12 horas post-transfección.
- La **Figura 9** muestra la ubicación de 16 fusiones dCas9-GFP con la misma secuencia NLS alfa importina en cualquiera de los extremos N o C mirando a cero hasta tres repeticiones en tándem. Cada una de las construcciones se transfectó en células HEK 293FT utilizando Lipofectame 2000 y se tomaron imágenes 24 horas post-transfección.
- 25 La **Figura 10** muestra seis versiones de marcador 6xHis añadido a dCas9, transfectado en células 293 FT y teñido con un anticuerpo anti-6xHis. Tres de ellas se construyeron para la activación transcripcional (fusiones VP64), y las otras tres eran para la represión de la transcripción (sin dominio funcional).
- La **Figura 11** muestra una relación titulada de chiRNA (Sox2.1 y Sox2.5) para Cas9 (NLS-VP64-NLS-hSpCas9-NLS-VP64-NLS), transfectada en células 293 y cuantificada utilizando RT-qPCR.
- 30 La **Figura 12** muestra el posicionamiento de sitios diana en el locus Sox2 humano, siendo cada una de las dianas de 20 pb de largo con un motivo adyacente protoespaciador NGG (PAM) vecino.
- La **Figura 13** muestra la co-transfección de cada una de las construcciones que contienen dCas9 con plásmidos pA6 en células HEK 293FT utilizando Lipofectame 2000. 72 horas post-transfección ARN total se extrajo de las células. 1 ug de ARN se transcribió de forma inversa en ADNc (qScript Supermix) en una reacción de 40 ul. Se añadieron 2 ul de producto de reacción en una sola reacción qPCR de ensayo TaqMan de 20 ul. Cada uno de los experimentos se realizó por triplicatas biológicas y técnicas. Ningún control RT y ninguna reacción de control del molde mostraron amplificación.
- 35 La **Figura 14** muestra el test de construcciones (pXRP011, pXRP013, pXRP015) utilizando las mismas dianas Sox2 que la Figura 13.
- 40 Las **Figuras 15A-15B** muestran una lista de 31 construcciones para explorar diferentes enlazadores, dominios funcionales y fusiones en los extremos N y C y elementos críticos.
- La **Figura 16** muestra la co-transfección de cada uno de los plásmidos represores dCas9 con dos ARNs guía dirigidos a la cadena codificadora del gen beta-catenina. Se aisló el ARN 72 horas después de la transfección y la expresión de los genes se cuantificó por RT-qPCR. El gen de control endógeno era GAPDH. Se utilizaron dos

shRNAs validados como controles positivos. Los controles negativos eran determinados plásmidos transfectados sin ARNg, éstos se como "control pXRP ##".

5 La **Figura 17** muestra una representación gráfica de la activación transcripcional adicional a la relación de chiRNA (Sox2.1 y Sox2.5) a Cas9 (NLS-VP64-NLS-hSpCas9-NLS-VP64-NLS) que se titula y transfecta en células 293. Los resultados se cuantificaron mediante Rt-qPCR.

La **Figura 18** muestra datos del informador de luciferasa para el activador (panel superior) y represor (panel inferior) de Cas9. En comparación con controles "Sin Cas9", la activación de más de 3 veces que se logró cuando se fijó como objetivo el promotor de Sox2. Cuando se fijó como objetivo el cuerpo del gen de beta-catenina (CTNNB1), se logró una represión de aproximadamente 3 veces.

10 La **Figura 19** muestra la expresión génica de beta-catenina en células HEK 293FT 72 horas después de la transfección. Construcciones del represor Cas9 fueron fijadas como objetivo al locus beta-catenina y se comparan con los ShRNA estándar dorados. Una represión similar pudo observarse con represores Cas9 y el shRNA.

15 La **Figura 20** muestra una representación gráfica del nivel de expresión de Neurog2 de las secuencias SgRNA de 20 pb fijadas como objetivo al locus Neurog2 en células Neuro 2A de ratón. Los niveles de ARNm de Neurog2 se midieron utilizando RT-qPCR.

La **Figura 21** muestra la modulación de la expresión génica basal de dCas9-VP64 que se midió para las dianas de genes indicadas. La cursiva representa dianas de genes de ratones testadas en células Neuro 2 A, todas las mayúsculas representan dianas de genes humanos testadas en células 293FT. En cada caso, la expresión del gen en las células transfectadas con GFP se comparó con la expresión en células transfectadas con dCas9-VP64.

20 La **Figura 22** muestra los cambios en el nivel de expresión de los genes indicados cuando las muestras fueron transfectadas con la construcción dCas9-VP64, pero sin un sgRNA.

25 La **Figura 23A-G** muestra a, La nucleasa Cas9 guiada por ARN del sistema CRISPR/Cas de Streptococcus pyogenes de tipo II puede ser convertida en una proteína de unión a ADN guiada por ARN nucleolíticamente inactiva (Cas9**) mediante la introducción de dos sustituciones de alanina (D10A y H840A). Se muestra esquemáticamente que un ARN guía sintético (sgRNA) puede dirigir la fusión Cas9**- efector a un locus específico en el genoma humano. El sgRNA contiene una secuencia guía de 20 pb en el extremo 5' que especifica la secuencia diana. En el ADN genómico diana, el sitio diana de 20 pb necesita ser seguido por un motivo 5'-NGG PAM, b, c, Esquemas que muestran los sitios diana de sgRNA en los loci humanos KLF4 y SOX2, respectivamente. Cada uno de los sitios diana se indica por la barra azul y la secuencia de PAM correspondiente se indica por la barra de magenta. d, e, Esquemas del activador de la transcripción Cas9**-VP64 y las construcciones del represor de la transcripción SID4X-Cas9**, f, g, Activación de KLF4 mediada por Cas9**-VP64 y SID4X-Cas9** y represión de SOX2, respectivamente. Todos los niveles de ARNm fueron medidos con relación a células 293FT transfectadas de forma simulada (media \pm s.e.m.; n = 3 réplicas biológicas).

La **Figura 24** muestra la secuencia de activador Cas9.

35 La **Figura 25** muestra la secuencia de represor Cas9.

La **Figura 26** muestra la representación gráfica del nivel de activación de diferentes genes fijados como objetivo por el activador de Cas9 (pXRP57) y el ARN guía.

La **Figura 27** muestra la representación gráfica del nivel de represión del gen hSox2 fijado como objetivo por el activador de Cas9 (pXRP57) y el ARN guía.

40 La **Figura 28** muestra un mapa del plásmido pAAV-EF1a-dCas9-GS-CIB1 (mNLS d318-334)_ WPRE_hGHpoliA.

La **Figura 29** muestra un mapa del plásmido pAAV-EF1a-dCas9-GS-NLS-cib1-WPRE-hGHpA.

La **Figura 30** muestra un mapa del plásmido pAAV-EF1a-dCas9-GS-NLS-NLS-cib1-WPRE-hGHpA.

La **Figura 31** muestra una representación gráfica de construcciones CasLITE que exhiben diferentes niveles de activación de la transcripción inducible por la luz.

La **Figura 32** muestra la validación de células madre embrionarias de ratón Cas9 inactivado Rosa26 Cre-dependiente.

5 La **Figura 33** muestra una imagen de gel que indica los resultados de genotipado para ratones Cas9.

La **Figura 34** muestra el silenciamiento de ARN mediado por CRISPR/Cmr. Las proteínas Cmr forman un complejo con crRNA maduro para fijar específicamente como objetivo y escindir ARN. crRNAs maduros de *P. furiosus* existen en dos formas de diferentes longitudes, larga (45 nt) y corta (39 nt), que consiste en un 5' mango y una secuencia guía. Se requiere el 5' mango para que el crRNA se incluya en el complejo Cmr. La secuencia de guía programa el sitio diana y pueden ser larga (37 nt) o corta (31 nt). La escisión de ARN se produce 14 nucleótidos del extremo 3' del crRNA. Esta plataforma puede ser reutilizada para fijar como objetivo genes de mamífero expresados a partir de un locus genómico con solo cambiar la secuencia guía.

15 La **Figura 35A-B** muestra que proteínas Cmr se expresan en células de mamíferos. Seis genes Cmr (Cmr 1-6) y un gen Cas (Cas6) de *P. furiosus* fueron clonados en vectores de expresión de mamíferos y fueron transfectados en células HEK 293FT. Se tomaron imágenes de fluorescencia 72 horas después de la transfección y se prepararon lisados de proteínas. A) se observa fuertemente una expresión de EGFP, lo que sugiere que las proteínas exógenas se expresan robustamente. B) Imágenes de transferencia Western muestran bandas de proteínas ubicadas en el tamaño esperado. La segunda banda dentro de cada una de las pistas es ~ 30 kDa mayor en cada caso y es probablemente el resultado de una secuencia de P2A no escindida.

20 La **Figura 36** muestra vectores de expresión CRISPR/Cmr.

Las figuras en esta memoria son sólo para fines ilustrativos y no están necesariamente dibujadas a escala.

DESCRIPCIÓN DETALLADA DE LA INVENCION

25 La invención se refiere a la modificación por tecnología genética y la optimización de sistemas, métodos y composiciones utilizados para el control de la expresión génica que implican la focalización de la secuencia, tales como la perturbación del genoma o la edición de genes, que se refieren al sistema CRISPR/Cas y componentes de los mismos. En realizaciones ventajosas, la enzima Cas es Cas9; por ejemplo, Cas9 de *S. pyogenes* o *S. thermophilus*.

30 Los términos y expresiones "polinucleótido", "nucleótido", "secuencia de nucleótidos", "ácido nucleico" y "oligonucleótido" se utilizan indistintamente. Se refieren a una forma polimérica de nucleótidos de cualquier longitud, ya sean desoxirribonucleótidos o ribonucleótidos, o análogos de los mismos. Los polinucleótidos pueden tener cualquier estructura tridimensional y pueden realizar cualquier función, conocida o desconocida. Los siguientes son ejemplos no limitativos de polinucleótidos: regiones codificadoras o no codificadoras de un gen o fragmento de gen, loci (locus) definidos a partir de análisis de ligamiento, exones, intrones, ARN mensajero (ARNm), ARN de transferencia, ARN ribosomal, ARN de interferencia corto (siRNA), ARN de horquilla corto (shRNA), micro-ARN (miRNA), ribozimas, ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plásmidos, vectores, ADN aislado de cualquier secuencia, ARN aislado de cualquier secuencia, sondas de ácido nucleico y cebadores. El término también abarca estructuras de tipo ácido nucleico con cadenas principales sintéticas, véase, p. ej., Eckstein, 1991; Baserga et al., 1992; Milligan, 1993; documentos WO97/03211; WO 96/39154; Mata, 1997; Strauss-Soukup, 1997; y Samstag, 1996. Un polinucleótido puede comprender uno o más nucleótidos modificados tales como nucleótidos metilados y análogos de nucleótidos. Si están presentes, modificaciones en la estructura de nucleótidos se pueden impartir antes o después del ensamblaje del polímero. La secuencia de nucleótidos puede estar interrumpida por componentes no nucleótidos. Un polinucleótido puede modificarse adicionalmente después de la polimerización tal como por conjugación con un componente de marcaje.

45 En aspectos de la invención, las expresiones "ARN quimérico", "ARN guía quimérico", "ARN guía", "ARN guía sencillo" y "ARN guía sintético" se utilizan indistintamente y se refieren a la secuencia de polinucleótidos que comprende la secuencia guía, la secuencia tracr y la secuencia de emparejamiento tracr. La expresión "secuencia guía" se refiere a la secuencia de aproximadamente 20 pb dentro del ARN guía que especifica el sitio diana y puede utilizarse indistintamente con los términos "guía" o "espaciador". La expresión "secuencia de emparejamiento tracr" también se puede utilizar de forma indistinta con el término "repetición o repeticiones directas".

Tal como se utiliza en esta memoria, la expresión "tipo salvaje" es un término de la técnica entendido por las personas expertas y significa la forma típica de un organismo, cepa, gen o característica tal como se produce en la naturaleza que se distingue de las formas mutantes o variantes. Por ejemplo, "StCas9 de tipo salvaje" se refiere a Cas9 de tipo salvaje de *S. thermophilus*, cuya secuencia de proteína se da en la base de datos SwissProt con el número de acceso G3ECR1. De manera similar, *S. pyogenes* Cas9 está incluida en SwissProt con el número de acceso Q99ZW2.

Tal como se utiliza en esta memoria, el término "variante" se deben tomar en el sentido de la exhibición de cualidades que tienen un patrón que se desvía de lo que ocurre en la naturaleza.

Las expresiones "que se produce de forma no natural" o "modificado por ingeniería genética" se utilizan indistintamente e indican la implicación de la mano del hombre. Las expresiones, cuando se refieren a moléculas de ácidos nucleicos o polipéptidos significan que la molécula de ácido nucleico o el polipéptido está al menos sustancialmente libre de al menos otro componente con el que están asociados de forma natural en la naturaleza y como se encuentra en la naturaleza.

"Complementariedad" se refiere a la capacidad de un ácido nucleico de formar enlace(s) hidrógeno con otra secuencia de ácido nucleico por apareamiento de bases de Watson-Crick tradicional u otros tipos no tradicionales. A porcentaje de complementariedad indica el porcentaje de residuos en una molécula de ácido nucleico que puede formar enlaces hidrógeno (p. ej., apareamiento de bases de Watson-Crick) con una segunda secuencia de ácido nucleico (p. ej., siendo 5, 6, 7, 8, 9, 10 de 10 50%, 60%, 70%, 80%, 90%, 90% y 100% complementarias). "Perfectamente complementarios" significa que todos los residuos contiguos de una secuencia de ácidos nucleicos se enlazará por hidrógeno con el mismo número de residuos contiguos en una segunda secuencia de ácido nucleico. "Sustancialmente complementarios", tal como se utiliza en esta memoria, se refiere a un grado de complementariedad que es al menos 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% o 100% sobre una región de 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, o más nucleótidos, o se refiere a dos ácidos nucleicos que se hibridan en condiciones rigurosas.

Tal como se utiliza en esta memoria, "condiciones rigurosas" para la hibridación se refieren a condiciones en las que un ácido nucleico que tiene complementariedad con una secuencia diana se hibrida predominantemente con la secuencia diana, y sustancialmente no se hibrida con secuencias no diana. Las condiciones rigurosas son generalmente dependientes de la secuencia, y varían dependiendo de un cierto número de factores. En general, cuanto mayor sea la secuencia, más alta será la temperatura a la que la secuencia se hibrida específicamente con su secuencia diana. Ejemplos no limitantes de las condiciones restrictivas se describen en detalle en Tijssen (1993), *Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Parte I, Segundo Capítulo "Overview of principles of hybridization and the strategy of nucleic acid probe assay"*, Elsevier, N.Y.

"Hibridación" se refiere a una reacción en la que uno o más polinucleótidos reaccionan para formar un complejo que se estabiliza a través de enlaces hidrógeno entre las bases de los residuos de nucleótidos. El enlace hidrógeno se puede producir por apareamiento de bases de Watson Crick, la unión Hoogstein, o en cualquier otra forma específica para la secuencia. El complejo puede comprender dos cadenas que forman una estructura dúplex, tres o más cadenas que forman un complejo multi-cadena, una sola cadena auto-hibridante, o cualquier combinación de éstas. Una reacción de hibridación puede constituir una etapa de un proceso más extenso tal como el inicio de la PCR, o la escisión de un polinucleótido por una enzima. A una secuencia capaz de hibridarse con una secuencia dada se la alude como el "complemento" de la secuencia dada.

Tal como se utiliza en esta memoria, la expresión "locus genómico" o "locus" (plural loci) es la ubicación específica de un gen o secuencia de ADN en un cromosoma. Un "gen" se refiere a tramos de ADN o ARN que codifican un polipéptido o una cadena de ARN que tiene que jugar un papel funcional en un organismo y, por lo tanto, es la unidad molecular hereditaria en los organismos vivos. Para los fines de esta invención, se puede considerar que los genes incluyen regiones que regulan la producción del producto génico, estén o no estas secuencias reguladoras adyacentes a las secuencias codificadoras y/o transcritas. Por consiguiente, un gen incluye, pero no se limita necesariamente a secuencias de promotor, terminadores, secuencias reguladoras de la traducción tales como sitios de unión a ribosomas y sitios de entrada al ribosoma interno, potenciadores, silenciadores, aisladores, elementos de borde, orígenes de replicación, sitios de unión a la matriz y regiones de control del locus.

Tal como se utiliza en esta memoria, "expresión de un locus genómico" o "expresión génica" es el proceso por el cual la información de un gen se utiliza en la síntesis de un producto génico funcional. Los productos de la expresión génica son a menudo proteínas, pero en genes no codificadores de proteínas tales como genes de ARNr o genes de ARNt, el producto es ARN funcional. El proceso de la expresión génica se utiliza por todos los eucariotas vivos

5 conocidos (incluyendo organismos multicelulares), procariotas (bacterias y arqueas) y virus para generar productos funcionales para sobrevivir. Tal como se utiliza en esta memoria, "expresión" de un gen o ácido nucleico abarca no sólo la expresión del gen celular, sino también la transcripción y la traducción de ácido o ácidos nucleicos en los sistemas de clonación y en cualquier otro contexto. Tal como se utiliza en esta memoria, "expresión" también se refiere al proceso mediante el cual un polinucleótido se transcribe a partir de un molde de ADN (como en y ARNm u otro transcrito de ARN) y/o el proceso mediante el cual un ARNm transcrito se traduce posteriormente en péptidos, polipéptidos o proteínas. A transcritos y polipéptidos codificados se les puede aludir selectivamente como "producto génico". Si el polinucleótido se deriva de ADN genómico, la expresión puede incluir el corte y empalme de ARNm en una célula eucariota.

10 Los términos "polipéptido", "péptido" y "proteína" se utilizan indistintamente en esta memoria para referirse a polímeros de aminoácidos de cualquier longitud. El polímero puede ser lineal o ramificado, puede comprender aminoácidos modificados y puede estar interrumpido por no aminoácidos. Los términos también abarcan un polímero de aminoácidos que ha sido modificado; por ejemplo, formación de enlaces disulfuro, glicosilación, lipídación, acetilación, fosforilación o cualquier otra manipulación tal como conjugación con un componente de marcate. Tal como se utiliza en esta memoria, el término "aminoácido" incluye aminoácidos naturales y/o no naturales o sintéticos, incluyendo glicina y tanto los isómeros ópticos D o L y análogos de aminoácidos y peptidomiméticos.

Tal como se utiliza en esta memoria, el término "dominio" o la expresión "dominio de proteína" se refiere a una parte de una secuencia de proteína que pueden existir y funcionar independientemente del resto de la cadena de proteína.

20 Tal como se describe en aspectos de la invención, la identidad de secuencia está relacionada con la homología de secuencia. Las comparaciones de homología pueden realizarse a simple vista, o más habitualmente, con la ayuda de programas de comparación de secuencias fácilmente disponibles. Estos programas informáticos disponibles comercialmente pueden calcular el porcentaje (%) de homología entre dos o más secuencias y también pueden calcular la identidad de secuencia compartida por dos o más secuencias de aminoácidos o de ácido nucleico. Las homologías de secuencia pueden generarse mediante cualquiera de un cierto número de programas de ordenador conocidos en la técnica, por ejemplo BLAST o FASTA, etc. Un programa informático adecuado para llevar a cabo dicha alineación es el paquete GCG Wisconsin Bestfit (Universidad de Wisconsin, EE.UU.; Devereux et al., 1984, Nucleic Acids Research 12:387). Ejemplos de otros software que pueden realizar comparaciones de secuencias incluyen, pero no se limitan a, el paquete BLAST (véase Ausubel et al., 1999 *ibid.* - Capítulo 18), FASTA (Atschul et al., 1990, J. Mol. Biol, 403-410) y el conjunto GENWORKS de herramientas de comparación. Tanto BLAST como FASTA están disponibles para la búsqueda fuera de línea y en línea (véase Ausubel et al., 1999 *ibid.*, páginas 7-58 a 7-60). Sin embargo, se prefiere utilizar el programa GCG Bestfit.

35 El % de homología puede calcularse sobre secuencias contiguas, es decir, una secuencia se alinea con la otra secuencia y cada uno de los aminoácidos o nucleótidos en una secuencia se compara directamente con el aminoácido o nucleótido correspondiente en la otra secuencia, un residuo cada vez. Esto se denomina alineamiento "sin huecos". Típicamente, estos alineamientos sin huecos se realizan sólo sobre un número relativamente corto de residuos.

40 Aunque este es un método muy simple y consistente, no tiene en consideración que, por ejemplo, en un par de secuencias de otro modo idénticas, una inserción o deleción puede provocar que los siguientes restos de aminoácidos se coloquen fuera de alineamiento, por lo tanto, resultando potencialmente en una gran reducción en el % de homología cuando se realiza un alineamiento global. En consecuencia, la mayoría de los métodos de comparación de secuencias se diseñan para producir alineamientos óptimos que tienen en cuenta posibles inserciones y deleciones sin penalizar indebidamente la puntuación de homología o identidad global. Esto se consigue insertando "huecos" en el alineamiento de secuencias para intentar maximizar la homología o identidad local.

45 Sin embargo, estos métodos más complejos asignan "penalizaciones por hueco" a cada uno de los huecos que se produce en el alineamiento, de modo que, para el mismo número de aminoácidos idénticos, un alineamiento de la secuencia con tan pocos huecos como sea posible – lo que refleja una mayor relatividad entre las dos secuencias comparadas – puede alcanzar una puntuación más alta que con muchos huecos. Se utilizan típicamente "costos de huecos de afinidad" que cargan un coste relativamente alto para la existencia de un hueco y una penalización más pequeña para cada uno de los residuos posteriores en el hueco. Este es el sistema de puntuación de huecos más comúnmente utilizado. Penalizaciones de huecos elevadas pueden producir, por supuesto, alineamientos optimizados con menos huecos. La mayoría de los programas de alineamiento permiten modificar las penalizaciones por hueco. Sin embargo, se prefiere utilizar los valores por defecto cuando se utiliza un software de este tipo para las

comparaciones de secuencias. Por ejemplo, al utilizar el paquete GCG Wisconsin Bestfit la penalización por hueco por defecto para secuencias de aminoácidos es -12 para un hueco y -4 para cada extensión,

El cálculo del % máximo de homología, por tanto, requiere primero la producción de una alineación óptima, teniendo en consideración las penalizaciones por hueco. Un programa informático adecuado para llevar a cabo dicho alineamiento es el paquete GCG Wisconsin Bestfit (Devereux et al., 1984 *Nuc. Acids Research* 12 p 387). Ejemplos de otros software que pueden realizar comparaciones de secuencias incluyen, pero no se limitan a, el paquete BLAST (véase Ausubel et al., 1999 *Short Protocols in Molecular Biology*, 4ª Ed. - Capítulo 18), FASTA (Altschul et al., 1990, *J. Mol. Biol.* 403-410) y el conjunto GENEWORKS de herramientas de comparación. Tanto BLAST como FASTA están disponibles para la búsqueda en línea y fuera de línea (véase Ausubel et al., 1999, *Short Protocols in Molecular Biology*, páginas 7-58 a 7-60). Sin embargo, para algunas aplicaciones, se prefiere utilizar el programa GCG Bestfit. Una nueva herramienta, denominada BLAST 2 Sequences, está también disponible para la comparación de secuencias de proteínas y nucleótidos (véase *FEMS Microbiol Lett.* 1999 174(2):247-50; *FEMS Microbiol Lett.* 1999 177(1):187-8 y el sitio web del Centro Nacional de información sobre Biotecnología en la página web de National Institutes for Health).

Aunque el % de homología final puede medirse en términos de identidad, el proceso de alineamiento en sí no se basa típicamente en una comparación de pares de todo o nada. En su lugar, se utiliza generalmente una matriz de puntuación de similitud a escala que asigna puntuaciones a cada comparación por pares basada en la similitud química o la distancia evolutiva. Un ejemplo de una matriz de este tipo comúnmente utilizada es la matriz BLOSUM62 - la matriz por defecto para el conjunto de programas BLAST. Los programas GCG Wisconsin generalmente utilizan cualquiera de los valores por defecto públicos o una tabla de comparación de símbolos personalizada, si se suministra (véase el manual del usuario para más detalles). Para algunas aplicaciones, se prefiere utilizar los valores por defecto públicos para el paquete GCG, o en el caso de otro software, la matriz por defecto, tal como BLOSUM62.

Alternativamente, las homologías porcentuales pueden calcularse utilizando la característica de alineamiento múltiple en DNASIS™ (Hitachi Software), basado en un algoritmo, análogo a CLUSTAL (Higgins DG y Sharp PM (1988), *Gene* 73 (1), 237-244). Una vez que el software ha producido un alineamiento óptimo, es posible calcular el % de homología, preferentemente el % de identidad de secuencia. El software hace normalmente esto como parte de la comparación de secuencias y genera un resultado numérico.

Las secuencias también pueden tener deleciones, inserciones o sustituciones de residuos de aminoácidos que producen un cambio silencioso y resultan en una sustancia funcionalmente equivalente. Sustituciones de aminoácidos deliberadas pueden realizarse sobre la base de la similitud en las propiedades de aminoácidos (tales como polaridad, carga, solubilidad, hidrofobicidad, hidrofiliicidad y/o la naturaleza anfipática de los residuos) y, por lo tanto, es útil para agrupar juntos aminoácidos en grupos funcionales. Los aminoácidos pueden agruparse basándose en las propiedades de sus cadenas laterales solas. Sin embargo, es más útil incluir los datos de mutación también. Los conjuntos de aminoácidos así obtenidos es probable que se conserven por razones estructurales. Estos conjuntos pueden describirse en forma de un diagrama de Venn (Livingstone CD y Barton GJ (1993) "Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation" *Comput. Appl. Biosci.* 9: 745-756) (Taylor W.R. (1986) "The classification of amino acid conservation" *J. Theor. Biol.* 119; 205-218). Sustituciones conservadoras se pueden hacer, por ejemplo, de acuerdo con la tabla que figura a continuación que describe un diagrama de Venn generalmente aceptado de agrupación de aminoácidos.

Conjunto		Sub-conjunto	
Hidrofóbico	F W Y H K M I L V A G C	Aromático	F W Y H
		Alifático	I L V
Polar	W Y H K R E D C S T N Q	Cargado	H K R E D
		Positivamente cargado	H K R
		Negativamente cargado	E D
Pequeño	V C A G S P T N D	Minúsculo	A G S

Realizaciones de la invención incluyen secuencias (tanto de polinucleótidos o polipéptidos) que puede comprender la sustitución homóloga (sustitución y reemplazo se utilizan ambos en esta memoria para dar a entender el intercambio de un residuo de aminoácido o nucleótido existente, con un residuo o nucleótido alternativo) que puede producirse, es decir, una sustitución comparable en el caso de aminoácidos tal como básico por básico, ácido por ácido, polar por polar, etc. También puede producirse una sustitución no homóloga, es decir, de una clase de

residuo a otro o, alternativamente, que implica la inclusión de aminoácidos no naturales tales como ornitina (a la que se alude en lo sucesivo como Z), ornitina del ácido diaminobutírico (a la que se alude en lo sucesivo como B), ornitina norleucina (a la que se alude en lo sucesivo como O), pirilalanina, tienilalanina, naftilalanina y fenilglicina.

5 Secuencias de aminoácidos variantes pueden incluir grupos espaciadores adecuados que pueden insertarse entre cualesquiera dos residuos de aminoácidos de la secuencia, incluyendo grupos alquilo tales como grupos metilo, etilo o propilo, además de espaciadores de aminoácidos tales como residuos glicina o β -alanina. Una forma adicional de variación, que implica la presencia de uno o más residuos aminoácidos en forma peptoide, pueda ser bien comprendida por los expertos en la técnica. Para evitar dudas, "la forma peptoide" se utiliza para referirse a residuos aminoácidos variantes, en donde el grupo de sustituyentes carbono α está en átomo de nitrógeno del residuo en lugar del carbono α . Procedimientos para preparar péptidos en forma peptoide son conocidos en la técnica, por ejemplo Simon RJ et al., *PNAS* (1992) 89(20), 9367-9371 y Horwell DC, *Trends Biotechnol.* (1995) 13(4), 132-134.

15 La práctica de la presente invención emplea, a menos que se indique lo contrario, técnicas convencionales de inmunología, bioquímica, química, biología molecular, microbiología, biología celular, genómica y ADN recombinante, que están dentro de la experiencia de la técnica. Véase Sambrook, Fritsch y Maniatis, *MOLECULAR CLONING: A LABORATORY MANUAL*, 2ª edición (1989); *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY* (F. M. Ausubel, et al. comps., (1987)); la serie *METHODS IN ENZYMOLOGY* (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames y G.R. Taylor comps. (1995)), Harlow y Lane, comps. (1988) *ANTIBODIES, A LABORATORY MANUAL*, y *ANIMAL CELL CULTURE* (R.I. Freshney, comp. (1987)).

20 En un aspecto, la invención proporciona vectores que se utilizan en la ingeniería y optimización de los sistemas CRISPR/Cas.

Tal como se utiliza en esta memoria, un "vector" es una herramienta que permite o facilita la transferencia de una entidad de un entorno a otro. Es un replicón, tal como un plásmido, fago, o cósmido, en el que se puede insertar otro segmento de ADN con el fin de llevar a cabo la replicación del segmento insertado. Generalmente, un vector es capaz de replicarse cuando se asocia con los elementos de control apropiados. En general, el término "vector" se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que se ha unido. Vectores incluyen, pero no se limitan a, moléculas de ácido nucleico que son de cadena sencilla, de cadena doble, o parcialmente de doble cadena; moléculas de ácido nucleico que comprenden uno o más extremos libres, sin extremos libres (p. ej., circulares); moléculas de ácido nucleico que comprenden ADN, ARN, o ambos; y otras variedades de polinucleótidos conocidos en la técnica. Un tipo de vector es un "plásmido", que se refiere a un bucle de ADN de doble cadena circular en el que se pueden insertar segmentos de ADN adicionales tal como por técnicas de clonación molecular estándares. Otro tipo de vector es un vector viral, en el que las secuencias de ADN o ARN derivadas de virus están presentes en el vector para el empaquetamiento en un virus (p. ej., retrovirus, retrovirus defectuosos en la replicación, adenovirus, adenovirus defectuosos en la replicación, y virus adeno-asociados). Los vectores virales también incluyen polinucleótidos portados por un virus para la transfección en una célula huésped. 25 Determinados vectores son capaces de replicación autónoma en una célula huésped en la que se introducen (p. ej., vectores bacterianos que tienen un origen bacteriano de replicación y vectores episomales de mamífero). Otros vectores (p. ej., vectores no episomales de mamífero) se integran en el genoma de una célula huésped tras la introducción en la célula huésped y, con ello, se replican junto con el genoma del huésped. Además, determinados vectores son capaces de dirigir la expresión de genes a los que están enlazados operativamente. Tales vectores se denominan en esta memoria "vectores de expresión". Vectores de expresión comunes de utilidad en técnicas de ADN recombinante están a menudo en forma de plásmidos.

Los vectores de expresión recombinantes pueden comprender un ácido nucleico de la invención en una forma adecuada para la expresión del ácido nucleico en una célula huésped, lo que significa que los vectores de expresión recombinantes incluyen uno o más elementos reguladores, que pueden seleccionarse sobre la base de las células huésped a utilizar para la expresión, que están operativamente enlazadas a la secuencia de ácido nucleico a expresar. Dentro de un vector de expresión recombinante, "enlazada operativamente" pretende significar que la secuencia de nucleótidos de interés está enlazada al o a los elementos reguladores de una manera que permite la expresión de la secuencia de nucleótidos (p. ej., en un sistema de transcripción/traducción in vitro o en una célula huésped cuando el vector se introduce en la célula huésped). Con respecto a métodos de recombinación y de clonación, se hace mención a la publicación de la solicitud de patente de EE.UU. 2004-0171156 A1.

Aspectos de la invención se refieren a vectores bicistrónicos para ARN quimérico y Cas9. Se prefieren vectores de expresión bicistrónicos para ARN quimérico y Cas9. En general, y en particular en esta forma de realización, Cas9 es impulsado preferiblemente por el promotor CBh. El ARN quimérico puede ser impulsado preferentemente por un promotor U6. De manera ideal los dos se combinan. El ARN guía quimérico consiste típicamente en una secuencia guía de 20 pb (Ns) y esto puede estar unido a la secuencia tracr (que va desde la primera "U" de la cadena inferior al

extremo del transcrito). La secuencia tracr se puede estar truncada en varias posiciones tal como se indica. Las secuencias guía y tracr están separadas por la secuencia de emparejamiento tracr, que puede ser GUUUUAGAGCUA. Esto puede ser seguido por la secuencia de bucle GAAA tal como se muestra. Ambas son ejemplos preferidos. La solicitante ha demostrado indeles mediados por Cas9 en los loci *EMX1* y *PVALB* humanos mediante ensayos SURVEYOR. Los ChiRNAs se indican por su designación "+n", y crRNA se refiere a un ARN híbrido en el que las secuencias guía y tracr se expresan como transcritos separados. A lo largo de esta solicitud, el ARN quimérico también se puede denominar guía única, o ARN guía sintético (sgRNA). El bucle es preferiblemente GAAA, pero no se limita a esta secuencia o, de hecho, a ser sólo de 4 pb de longitud. De hecho, secuencias formadoras de bucle preferidas para su uso en estructuras de horquilla son de cuatro nucleótidos de longitud, y lo más preferiblemente tienen la secuencia GAAA. Sin embargo, se pueden utilizar secuencias de bucle más largas o más cortas, al igual que secuencias alternativas. Las secuencias incluyen preferiblemente un triplete de nucleótidos (por ejemplo, AAA), y un nucleótido adicional (por ejemplo C o G). Ejemplos de secuencias formadoras de bucle incluyen CAAA y AAAG.

La expresión "elemento regulador" pretende incluir promotores, potenciadores, sitios de entrada al ribosoma interno (IRES), y otros elementos de control de la expresión (p. ej., señales de terminación de la transcripción tales como señales de poliadenilación y secuencias poli-U). Se describen estos elementos reguladores, por ejemplo, en Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Elementos reguladores incluyen los que dirigen la expresión constitutiva de una secuencia de nucleótidos en muchos tipos de células huéspedes y los que dirigen la expresión directa de la secuencia de nucleótidos sólo en determinadas células huéspedes (p. ej., secuencias reguladoras específicas para el tejido). Un promotor específico para el tejido puede dirigir la expresión principalmente en un tejido deseado de interés tal como el músculo, neurona, hueso, piel, sangre, órganos específicos (p. ej., hígado, páncreas), o tipos de células particulares (p. ej., linfocitos). Los elementos reguladores también pueden dirigir la expresión de una manera dependiente del tiempo tal como de una manera dependiente del ciclo celular o dependiente de la fase de desarrollo, que puede o puede no ser también específico para el tejido o tipo de célula. En algunas realizaciones, un vector puede comprender uno o más promotores pol III (p. ej., 1, 2, 3, 4, 5 o más promotores pol III), uno o más promotores pol II (p. ej., 1, 2, 3, 4, 5 o más promotores pol II), uno o más promotores pol I (p. ej., 1, 2, 3, 4, 5 o más promotores pol I), o combinaciones de los mismos. Ejemplos de promotores pol III incluyen, pero no se limitan a promotores U6 y H1. Ejemplos de promotores pol II incluyen, pero no se limitan al promotor del virus del sarcoma de Rous retroviral (RSV) LTR (opcionalmente con el potenciador RSV), el promotor de citomegalovirus (CMV) (opcionalmente con el potenciador CMV) [véase, p. ej., Boshart et al., Cell, 41:521-530 (1985)], el promotor SV40, el promotor de dihidrofolato reductasa, el promotor β -actina, el promotor fosfoglicerol quinasa (PGK) y el promotor EF1 α . También quedan abarcados por la expresión "elemento regulador" elementos potenciadores tales como WPRE; potenciadores de CMV; el segmento R-U5' en LTR de HTLV-I (Mol. Cell. Biol., Vol 8(1), págs. 466-472, 1988); el potenciador SV40; y la secuencia de intrón entre los exones 2 y 3 de β -globina de conejo (Proc. Natl. Acad. Sci. USA, Vol, 78(3), págs. 1527-31, 1981). Se apreciará por los expertos en la técnica que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped a transformar, el nivel de expresión deseado, etc. Un vector puede ser introducido en células huéspedes para con ello producir transcritos, proteínas o péptidos, incluyendo proteínas o péptidos de fusión, codificados por ácidos nucleicos tal como se describe en esta memoria (p. ej., repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR), transcritos, proteínas, enzimas, formas mutantes de los mismos, proteínas de fusión de los mismos, etc.). Con respecto a secuencias reguladoras, se hace mención de la publicación de la solicitud de patente de EE.UU. 2004-0242517 A1. Con respecto a los promotores, se hace mención a la publicación PCT WO 2011/028929 y la publicación de la solicitud de patente de EE.UU. 2011-0027239 A1.

Los vectores pueden ser diseñados para la expresión de transcritos de CRISPR (p. ej., transcritos de ácidos nucleicos, proteínas o enzimas) en células procariontas o eucariotas. Por ejemplo, los transcritos de CRISPR pueden expresarse en células bacterianas tales como *Escherichia coli*, células de insectos (utilizando vectores de expresión de baculovirus), células de levaduras o células de mamífero. Células huéspedes adecuadas se discuten en Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, California (1990). Alternativamente, el vector de expresión recombinante puede transcribirse y traducirse in vitro, por ejemplo utilizando secuencias reguladoras del promotor T7 y T7 polimerasa.

Los vectores se pueden introducir y propagar en una célula procariota o procariótica. En algunas realizaciones, se utiliza un procariota para amplificar copias de un vector que se ha de introducir en una célula eucariota o como un vector intermedio en la producción de un vector a ser introducido en una célula eucariota (por ejemplo, una ampliación de un plásmido como parte de un sistema de empaquetado de un vector viral). En algunas realizaciones, se utiliza un procariota para amplificar copias de un vector y expresar uno o más ácidos nucleicos tal como para proporcionar una fuente de una o más proteínas para el suministro a una célula huésped u organismo huésped. La expresión de proteínas en procariontas se lleva a cabo más a menudo en *Escherichia coli* con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de proteínas de fusión o de no fusión. Los vectores de

fusión añaden un cierto número de aminoácidos a una proteína codificada en el mismo tal como al extremo amino de la proteína recombinante. Tales vectores de fusión pueden cumplir uno o más fines tales como: (i) aumentar la expresión de la proteína recombinante; (ii) aumentar la solubilidad de la proteína recombinante; y (iii) ayudar en la purificación de la proteína recombinante, actuando como un ligando en purificación por afinidad. A menudo, en vectores de expresión de fusión, un sitio de escisión proteolítica se introduce en la unión del resto de fusión y la proteína recombinante para permitir la separación de la proteína recombinante del resto de fusión después de la purificación de la proteína de fusión. Tales enzimas, y sus secuencias de reconocimiento relacionadas, incluyen Factor Xa, trombina y enteroquinasa. Ejemplos de vectores de expresión de fusión incluyen pGEX (Pharmacia Biotech Inc; Smith y Johnson, 1988. *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) y pRIT 5 (Pharmacia, Piscataway, N.J.) que fusionan la glutatión S transferasa (GST), proteína de unión a maltosa, o proteína A, respectivamente, a la proteína recombinante diana.

Ejemplos de vectores de expresión de *E. coli* sin fusión inducibles adecuados incluyen pTrc (Amrann et al., (1988) *Gene* 69: 301-315) y pET 11d (Studier et al., *GENE EXPRESSION TECHNOLOGY: METHODS IN ENZIMOLGY* 185, Academic Press, San Diego, Calif. (1990) 60-89).

En algunas realizaciones, un vector es un vector de expresión de levadura. Ejemplos de vectores para la expresión en la levadura *Saccharomyces cerevisiae* incluyen pYepSecl (Baldari, et al., 1987. *EMBO J.* 6:229-234), pMFa (Kuijan y Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz et al., 1987. *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.) y picZ (Invitrogen Corp, San Diego, Calif).

En algunas realizaciones, un vector impulsa la expresión de proteínas en células de insecto utilizando vectores de expresión de baculovirus. Vectores de baculovirus disponibles para la expresión de proteínas en células de insecto cultivadas (p. ej., células SF9) incluyen la serie pAc (Smith, et al., 1983. *Mol. Cell. Biol.* 3:2156-2165) y la serie pVL (Lucklow y Summers, 1989. *Virology* 170:31-39).

En algunas realizaciones, un vector es capaz de impulsar la expresión de una o más secuencias en células de mamífero utilizando un vector de expresión de mamífero. Ejemplos de vectores de expresión de mamífero incluyen pCDM8 (Seed, 1987. *Nature* 329:840) y pMT2PC (Kaufman, et al., 1987. *EMBO J.* 6:187-195). Cuando se utiliza en células de mamífero, las funciones de control del vector de expresión son proporcionadas típicamente por uno o más elementos reguladores. Por ejemplo, promotores comúnmente utilizados se derivan de poliovirus, adenovirus 2, citomegalovirus, virus de simio 40, y otros descritos en esta memoria y conocidos en la técnica. Para otros sistemas de expresión adecuados tanto para células procariontas como eucariotas véanse, p. ej., los Capítulos 16 y 17 de Sambrook, et al., *MOLECULAR CLONING: A LABORATORY MANUAL*. 2ª ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., 1989.

En algunas realizaciones, el vector de expresión de mamífero recombinante es capaz de dirigir la expresión del ácido nucleico preferentemente en un tipo de célula particular (p. ej., elementos reguladores específicos para tejidos se utilizan para expresar el ácido nucleico). Elementos reguladores específicos para tejidos son conocidos en la técnica. Ejemplos no limitantes de promotores específicos para tejidos, adecuados, incluyen el promotor de albúmina (específico para el hígado; Pinkert, et al., 1987, *Genes Dev.* 1: 268-277), promotores linfocitos-específicos (Calame y Eaton, 1988. *Adv. Immunol.* 43:235-275), en particular promotores de receptores de células T (Winoto y Baltimore, 1989. *EMBO J.* 8:729-733) e inmunoglobulinas (Baneiji, et al., 1983. *Cell* 33: 729-740; Queen y Baltimore, 1983 *Cell* 33: 741-748), promotores específicos para neuronas (p. ej., el promotor para neurofilamentos; Byrne y Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), promotores específicos para páncreas (Edlund, et al., 1985. *Science* 230:912-916) y promotores específicos para glándulas mamarias (p. ej., promotor de suero lácteo; patente de EE.UU. N° 4.873.316 y la Publicación de Solicitud Europea N° 264.166.). Promotores regulados en el desarrollo también están comprendidos, p. ej., los promotores *hox* murinos (Kessel y Gruss, 1990. *Science* 249:374-379) y el promotor de α -fetoproteína (Campes y Tilghman, 1989. *Genes Dev.* 3: 537-546). Con respecto a estos vectores procariontas y eucariotas, se hace mención a la patente de EE.UU. 6.750.059. Otras realizaciones de la invención pueden estar relacionadas con el uso de vectores virales, en lo que respecta a los que se hace mención en la publicación de la solicitud de patente de EE.UU. 2012-0003201 A1. Elementos reguladores específicos para tejidos son conocidos en la técnica y, en este sentido, se hace mención a la patente de EE.UU. 7.776.321.

En algunas realizaciones, un elemento regulador está enlazado operativamente a uno o más elementos de un sistema CRISPR con el fin de impulsar la expresión de los uno o más elementos del sistema de CRISPR. En general, los CRISPRs (Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Interespaciadas), también conocidos como SPIDRs (Repeticiones Directas Intercaladas Separadas), constituyen una familia de loci de ADN que habitualmente son específicos para una especie bacteriana particular. El locus CRISPR puede comprender una clase distinta de Repeticiones de Secuencia Corta Intercaladas (SSRs) que fueron reconocidas en *E. coli* (Ishino et al, *J. Bacteriol.*, 169:5429-5433 [1987]; y Nakata et al, *J. Bacteriol.*, 171: 3553-3556 [1989]), y genes asociados.

SSRs intercaladas similares han sido identificadas en *Haloferax mediterranei*, *Streptococcus pyogenes*, *Anabaena* y *Mycobacterium tuberculosis* (Véase, Groenen et al, *Mol. Microbiol.*, 10:1057-1065 [1993]; Hoe et al., *Emerg. Infect. Dis.*, 5:254-263 [1999]; Masepohl et al., *Biochim. Biophys. Acta* 1307:26-30 [1996]; y Mojica et al., *Mol. Microbiol.*, 17:85-93 [1995]). Los loci CRISPR difieren típicamente de otros SSRs por la estructura de las repeticiones, que se han denominado repeticiones cortas regularmente espaciadas (SRSRs) (Janssen et al., *OMICS J. Integ. Biol.*, 6:23-33 [2002]; y Mojica et. al., *Mol Microbiol*, 36:244-246 [2000]). En general, las repeticiones son elementos cortos que se producen en grupos que están espaciados regularmente por secuencias intermedias únicas con una longitud sustancialmente constante (Mojica et al., [2000], supra). Aunque las secuencias de repetición están altamente conservadas entre las cepas, el número de repeticiones intercaladas y las secuencias de las regiones del espaciador difieren típicamente de una cepa a otra (van Embden et al., *J. Bacteriol.*, 182:2393-2401 [2000]). Loci CRISPR han sido identificados en más de 40 procariontes (Véase, p. ej., Jansen et al, *Mol. Microbiol.*, 43:1565-1575 [2002]; y Mojica et al., [2005]), incluyendo, pero no limitado a *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halocarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacterium*, *Azarcus*, *Chromobacterium*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Myxococcus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photobacterium*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Treponema* y *Thermotoga*.

En general, "sistema de CRISPR" se refiere colectivamente a transcritos y otros elementos implicados en la expresión de o dirigir la actividad de CRISPR-asociado ("Cas") genes, incluyendo secuencias que codifican un gen Cas, una secuencia tracr (CRISPR trans-activante) (p. ej., tracrRNA o un tracrRNA parcial activo), una secuencia tracr-emparejamiento (que abarca una "repetición directa" y una repetición directa parcial procesada por tracrRNA en el contexto de un sistema de CRISPR endógeno), una secuencia guía(a la que también se alude como un "espaciador" en el contexto de un sistema de CRISPR endógeno), u otras secuencias y transcritos de un locus CRISPR. En realizaciones de la invención, las expresiones secuencia guía y ARN guía se utilizan indistintamente. En algunas realizaciones, uno o más elementos de un sistema de CRISPR se derivan de un sistema de CRISPR de tipo I, tipo II o de tipo III. En algunas realizaciones, uno o más elementos de un sistema de CRISPR se derivan de un organismo particular que puede comprender un sistema de CRISPR endógeno tal como *Streptococcus pyogenes*. En general, un sistema de CRISPR se caracteriza por elementos que fomentan la formación de un complejo de CRISPR en el sitio de una secuencia diana (al que también se alude como un protoespaciador en el contexto de un sistema de CRISPR endógeno). En el contexto de la formación de un complejo de CRISPR, "secuencia diana" se refiere a una secuencia a la que una secuencia guía está diseñada para que tenga complementariedad, en donde la hibridación entre una secuencia diana y una secuencia guía fomenta la formación de un complejo de CRISPR. Una secuencia diana puede comprender cualquier polinucleótido tal como polinucleótidos de ADN o ARN. En algunas realizaciones, una secuencia diana se encuentra en el núcleo o citoplasma de una célula.

En realizaciones preferidas de la invención, el sistema de CRISPR es un sistema de CRISPR de tipo II y la enzima Cas es Cas9, que cataliza la escisión de ADN. La acción enzimática por Cas9 derivada de *Streptococcus pyogenes* o cualquier Cas9 estrechamente relacionada genera roturas de doble cadena en las secuencias del sitio diana que se hibridan a 20 nucleótidos de la secuencia guía y que tienen una secuencia NGG del motivo adyacente al protoespaciador (PAM) después de los 20 nucleótidos de la secuencia diana. La actividad de CRISPR a través de Cas9 para el reconocimiento de ADN específico para el sitio y la escisión se definen por la secuencia guía, la secuencia tracr que se hibrida en parte a la secuencia guía y la secuencia de PAM. Se describen más aspectos del sistema de CRISPR en Karginov y Hannon. *The CRISPR system: small RNA-guided defense in bacteria and archaea*, *Mole Cell* 2010, 15 de enero; 37(1): 7.

El locus de CRISPR de tipo II de *Streptococcus pyogenes* SF370 contiene un racimo de cuatro genes Cas9, Cas1, Cas2 y Csn1, así como dos elementos de ARN no codificantes, tracrRNA y una matriz característica de secuencias repetitivas (repeticiones directas) interespaciadas por tramos cortos de secuencias no repetitivas (espaciadores, aproximadamente de 30 pb cada uno). En este sistema, la rotura de doble cadena de ADN (DSB) fijada como objetivo se genera en cuatro etapas secuenciales (Figura 2A). En primer lugar, dos ARN no codificantes, la matriz pre-crRNA y tracrRNA se transcriben a partir del locus de CRISPR. En segundo lugar, tracrRNA se hibrida a las repeticiones directas de pre-crRNA, que se procesa luego en crRNAs maduros que contienen secuencias de espaciador individuales. En tercer lugar, el complejo crRNA:tracrRNA maduro dirige Cas9 a la diana de ADN que consiste en el protoespaciador y el PAM correspondiente a través de la formación de heteroduplex entre la región del espaciador del crRNA y el ADN del protoespaciador. Por último, Cas9 media en la escisión del ADN diana aguas arriba del PAM para crear una DSB dentro del protoespaciador (Figura 2A). La Figura 2B muestra la ubicación nuclear de Cas9 optimizado en codones. Para fomentar la iniciación de la transcripción precisa, se seleccionó el promotor U6 basado en ARN polimerasa III para impulsar la expresión de tracrRNA (Figura 2C). De manera similar, se desarrolló una construcción basada en el promotor U6 para expresar una matriz de pre-crRNA que consiste en un único espaciador flanqueado por dos repeticiones directas (DRs, también abarcados por la expresión "secuencias de

emparejamiento de tracr"; Figura 2C). El espaciador inicial se diseñó para fijar como objetivo un sitio diana de 33 pares de bases (pb) (protoespaciador de 30 pb más una secuencia del motivo de CRISPR (PAM) de 3 pb que satisface el motivo de reconocimiento de NGG de Cas9) en el locus EMX1 humano (Figura 2C), un gen clave en el desarrollo de la corteza cerebral.

5 Típicamente, en el contexto de un sistema de CRISPR endógeno, la formación de un complejo de CRISPR (que puede comprender una secuencia guía hibridada a una secuencia diana y formando complejo con una o más proteínas Cas) resulta en la escisión de una o las dos cadenas en o cerca de (p. ej., dentro de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, o más pares de bases) de la secuencia diana. Sin desear estar limitados por la teoría, la secuencia tracr, que puede comprender o consistir en la totalidad o una porción de una secuencia tracr de tipo salvaje (p. ej., aproximadamente o más de aproximadamente 20, 26, 32, 45, 48, 54, 63, 67, 85, o más nucleótidos de una secuencia tracr de tipo salvaje), también pueden formar parte de un complejo de CRISPR, tal como mediante la hibridación a lo largo de al menos una porción de la secuencia de tracr a la totalidad o una parte de una secuencia de emparejamiento tracr que está enlazada operativamente a la secuencia guía. En algunas realizaciones, uno o más vectores que impulsan la expresión de uno o más elementos de un sistema de CRISPR se introducen en una célula huésped de manera que la expresión de los elementos del sistema de CRISPR dirigen la formación de un complejo de CRISPR en uno o más sitios diana. Por ejemplo, una enzima Cas, una secuencia guía enlazada a una secuencia tracr-emparejamiento, y una secuencia de tracr podrían ser unidas operativamente en cada caso para separar elementos reguladores en vectores separados. Alternativamente, dos o más de los elementos expresados a partir de los mismos o diferentes elementos reguladores, se pueden combinar en un solo vector, con uno o más vectores adicionales que proporcionan cualesquiera componentes del sistema de CRISPR no incluidos en el primer vector. Elementos del sistema de CRISPR que se combinan en un único vector pueden estar dispuestos en cualquier orientación adecuada, tal como un elemento situado en 5' con respecto a ("aguas arriba" de) o 3' con respecto a ("aguas abajo" de) un segundo elemento. La secuencia codificadora de un elemento puede estar situada en la misma o en una cadena diferente de la secuencia codificadora de un segundo elemento, y estar orientada en la misma o en dirección opuesta. En algunas realizaciones, un único promotor dirige la expresión de un transcrito que codifica una enzima de CRISPR y una o más de la secuencia guía, secuencia de emparejamiento tracr (opcionalmente enlazada operativamente a la secuencia de guía) y una secuencia de tracr incrustada dentro de una o más secuencias de intrones (p. ej., cada una en un intrón diferente, dos o más en al menos un intrón, o todas en un solo intrón). En algunas realizaciones, la enzima de CRISPR, la secuencia guía, la secuencia de emparejamiento tracr y la secuencia tracr están enlazadas operativamente a y son expresada a partir del mismo promotor.

En algunas realizaciones, un vector puede comprender uno o más sitios de inserción, tales como una secuencia de reconocimiento de endonucleasa de restricción (al que también se alude como un "sitio de clonación"). En algunas realizaciones, uno o más sitios de inserción (p. ej., aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más sitios de inserción) están situados aguas arriba y/o aguas abajo de uno o más elementos de secuencia de uno o más vectores. En algunas realizaciones, un vector puede comprender un sitio de inserción aguas arriba de una secuencia de emparejamiento tracr y, opcionalmente, aguas abajo de un elemento regulador enlazado operativamente a la secuencia de emparejamiento tracr, de tal manera que después de la inserción de una secuencia guía en el sitio de inserción y tras la expresión de la secuencia guía dirige la unión específica para la secuencia de un complejo CRISPR a una secuencia diana en una célula eucariota. En algunas realizaciones, un vector puede comprender dos o más sitios de inserción, estando situado cada uno de los sitios de inserción entre dos secuencias de emparejamiento tracr con el fin de permitir la inserción de una secuencia guía en cada uno de los sitios. En una disposición de este tipo, las dos o más secuencias guía pueden comprender dos o más copias de una única secuencia guía, dos o más secuencias guía diferentes, o combinaciones de éstas. Cuando se utilizan múltiples secuencias guía diferentes, se puede utilizar una única construcción de expresión para fijar como objetivo la actividad de CRISPR a múltiples secuencias diana diferentes, correspondiente, dentro de una célula. Por ejemplo, un único vector puede comprender aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, o más secuencias guía. En algunas realizaciones, pueden proporcionarse aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más de dichos vectores que contienen la secuencia guía, y opcionalmente suministrarse a una célula.

50 En algunas realizaciones, un vector puede comprender un elemento regulador enlazado operativamente a una secuencia codificadora de la enzima que codifica una enzima de CRISPR tal como una proteína Cas. Ejemplos no limitantes de proteínas Cas incluyen Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (también conocida como Csn1 y Csx12), Cas 10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologías de las mismas, o versiones modificadas de las mismas. En algunas realizaciones, la enzima de CRISPR no modificada tiene actividad de escisión de ADN, tal como Cas9. En algunas realizaciones, la enzima de CRISPR dirige la escisión de una o ambas cadenas en la ubicación de una secuencia diana, tal como dentro de la secuencia diana y/o dentro del complemento de la secuencia diana. En algunas realizaciones, la enzima de CRISPR dirige la escisión de una o ambas cadenas dentro de aproximadamente

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, o más pares de bases desde el primer o el último nucleótidos de una secuencia diana. En algunas realizaciones, un vector codifica una enzima de CRISPR que está mutada con respecto a una enzima de tipo salvaje correspondiente, de tal manera que la enzima de CRISPR mutada carece de la capacidad de escindir una o ambas cadenas de un polinucleótido diana que contiene una secuencia diana. Por ejemplo, una sustitución-aspartato-a alanina (D10A) en el dominio catalítico RuvC I de Cas9 de *S. pyogenes* convierte Cas9 de una nucleasa que escinde ambas hebras en una nicasa (escinde una sola cadena). Otros ejemplos de mutaciones que hacen de Cas9 una nicasa incluyen, sin limitación, H840A, N854A y N863A. Como ejemplo adicional, dos o más dominios catalíticos de Cas9 (RuvC I, RuvC II, y RuvC III o el dominio HNH) pueden mutarse para producir una Cas9 mutada que carece sustancialmente de toda la actividad de escisión de ADN. En algunas realizaciones, una mutación D10A se combina con uno o más mutaciones de H840A, N854A o N863A para producir una enzima Cas9 que carece sustancialmente de toda la actividad de escisión de ADN. En algunas realizaciones, una enzima de CRISPR se considera que carece sustancialmente de toda la actividad de escisión de ADN cuando la actividad de escisión de ADN de la enzima mutada es menor que aproximadamente 25%, 10%, 5%, 1%, 0,1%, 0,01%, o más baja con respecto a su forma no mutada. En los casos en los que la enzima no es SpCas9, las mutaciones se pueden realizar en cualquiera o en todos los residuos correspondientes a las posiciones 10, 762, 840, 854, 863 y/o 986 de SpCas9 (que puede determinarse, por ejemplo, mediante herramientas de comparación de secuencias estándar. En particular, cualquiera o todas de las siguientes mutaciones se prefieren en SpCas9: D10A, E762A, H840A, N854A, N863A y/o D986A; así como también está prevista la sustitución conservadora para cualquiera de los aminoácidos de reemplazo. También se prefieren las mismas (o sustituciones conservativas de estas mutaciones) en posiciones correspondientes en otras Cas9s. Son particularmente preferidas D10 y H840 en SpCas9. Sin embargo, en otros Cas9s, también se prefieren los residuos correspondientes a SpCas9 D10 y H840.

Una sustitución aspartato-a-alanina (D10A) en el dominio catalítico RuvC I de SpCas9 fue diseñado para convertir la nucleasa en una nicasa (SpCas9n) (véase, p. ej., Sapranaukas et al., 2011, *Nucleic Acid Research*, 39: 9275; Gasiunas et al., 2012, *Proc. Natl. Acad. Sci. USA*, 109:E2579), de manera que el ADN genómico mellado sufre la reparación dirigida a homología de alta fidelidad (HDR). El ensayo Surveyor confirmó que SpCas9n no genera indeles en la diana protoespaciador EMX1. La co-expresión de crRNA quimérico que fija como objetivo EMX1 (que tiene asimismo el componente tracrRNA) con SpCas9 produjo indeles en el sitio diana, mientras que la co-expresión con SpCas9n no lo hizo ($n = 3$). Además de ello, la secuenciación de 327 amplicones no detectó indeles inducidos por SpCas9n. El mismo locus fue seleccionado para testar HR mediada por CRISPR mediante co-transfección de células HEK 293FT con el ARN quimérico que fija como objetivo EMX1, hSpCas9 o hSpCas9n, así como un molde de HR para introducir un par de sitios de restricción (HindIII y NheI) cerca del protoespaciador.

En esta memoria se describen ortólogos preferidos. Una enzima Cas puede ser identificada Cas9, ya que ésta puede referirse a la clase general de enzimas que comparten homología con la mayor nucleasa con múltiples dominios nucleasa del sistema de CRISPR tipo II. Lo más preferiblemente, la enzima Cas9 es, o se deriva de spCas9 o saCas9. Por derivarse, la solicitante quieren dar a entender que la enzima derivada se basa ampliamente, en el sentido de tener un alto grado de homología de secuencia con una enzima de tipo salvaje, pero que ha sido mutada (modificada) de alguna manera tal como se describe en esta memoria.

En algunas realizaciones, una secuencia codificadora de la enzima que codifica una enzima CRISPR está optimizada en codones para la expresión en células particulares, tales como células eucariotas. Las células eucariotas pueden ser las de o derivadas de un organismo particular, tal como un mamífero, incluyendo, pero no limitado a, ser humano, ratón, rata, conejo, perro o primate no humano. En general, la optimización de codones se refiere a un proceso de modificar una secuencia de ácido nucleico para la expresión potenciada en las células huéspedes de interés mediante el reemplazo de al menos un codón (p. ej., aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, o más codones) de la secuencia nativa con codones que son con más frecuencia o lo más frecuentemente utilizados en los genes de esa célula huésped, al tiempo que se mantiene la secuencia de aminoácidos nativa. Diversas especies exhiben una desviación particular para ciertos codones de un aminoácido particular. Desviación del codón (diferencias en el uso de codones entre organismos) se correlaciona a menudo con la eficacia de la traducción de ARN mensajero (ARNm) que, es a su vez, se cree que depende, entre otros, de las propiedades de los codones que están siendo traducidos y la disponibilidad de moléculas de ARN de transferencia (ARNt) particulares. El predominio de los ARNt seleccionados en una célula es, por lo general, un reflejo de los codones más frecuentemente utilizados en la síntesis de péptidos. Por consiguiente, los genes pueden adaptarse para la expresión génica óptima en un organismo dado sobre la base de la optimización de codones. Tablas de uso de codones están fácilmente disponibles, por ejemplo, en la base de datos "Codon Usage Database", disponible en www.kazusa.or.jp/codon/ (visitada el 9 de julio de 2002), y estas tablas pueden ser adaptados en un cierto número de maneras. Véase Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence database status for the year 2000" *Nucl. Acids Res.* 28:292 (2000). También están disponibles algoritmos informáticos para la optimización de codones de una secuencia particular para la expresión en una célula huésped particular, tal como Gene Forge (Aptagen; Jacobus, PA). En algunas realizaciones, uno o más codones (p. ej., 1, 2,

3, 4, 5, 10, 15, 20, 25, 50, o más, o todos los codones) en una secuencia que codifica una enzima de CRISPR corresponden al codón utilizado con mayor frecuencia para un aminoácido particular.

5 En algunas realizaciones, un vector codifica una enzima de CRISPR que puede comprender una o más secuencias de localización nuclear (NLSs), tal como aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más NLSs. En algunas realizaciones, la enzima de CRISPR puede comprender aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más NLSs en o cerca del extremo amino, aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más NLSs en o cerca del extremo carboxi, o una combinación de éstas (p. ej., una o más NLS en el extremo amino y una o más NLS en el extremo carboxi). Cuando está presente más de una NLS, cada una se puede seleccionar independientemente de las otras, de manera que una sola NLS
10 puede estar presente en más de una copia y/o en combinación con una o más de otras NLSs presentes en una o más copias. En una realización preferida de la invención, la enzima de CRISPR puede comprender como a lo sumo 6 NLSs. En algunas realizaciones, una NLS se considera cerca del extremo N o C cuando el aminoácido más cercano de la NLS está dentro de aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50 o más aminoácidos a lo largo de la cadena de polipéptido desde el extremo N o C. Ejemplos no limitantes de NLSs incluyen una secuencia NLS derivada de: la NLS del antígeno T grande del virus SV40, que tiene la secuencia de aminoácidos PKKKRKV; la NLS del nucleoplasma (p. ej., la NLS bipartita nucleoplasmina con la secuencia KRPAATKKAGQAKKKK); las NLS c-myc que tienen la secuencia de aminoácidos PAAKRVKLD o QRRLNELKRSP; la NLS hRNPAL M9 que tiene la secuencia
15 NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY; la secuencia RMRIZFKNKKGKDTAELRRRVEVSVELRKAKKDEQILKRRNV del dominio IBB de importina-alfa; las secuencias VSRKRPRP y PPKKARED de la proteína T mioma; la secuencia POPKKKPL de p53 humana; la secuencia SALIKKKKKMAP de ratón c-abl IV; las secuencias DRLRR y PQKKRK del virus NS1 de la gripe; la secuencia RKLKKKIKKL del antígeno delta del virus de la hepatitis; la secuencia REKKKFLKRR de la proteína Mx1 de ratón; la secuencia KRKGDEV DGVDEVAKKKS K de la poli(ADP-ribosa)polimerasa humana; y la secuencia RKCLQAGMNL EARKTKK del glucocorticoide de los receptores de hormonas esteroides (humanas).

25 En general, la una o más NLSs son de una fuerza suficiente para impulsar la acumulación de la enzima de CRISPR en una cantidad detectable en el núcleo de una célula eucariota. En general, la fuerza de la actividad de localización nuclear puede derivar del número de NLSs en la enzima de CRISPR, la o las NLSs particulares utilizadas, o una combinación de estos factores. La detección de la acumulación en el núcleo puede ser realizada por cualquier técnica adecuada. Por ejemplo, un marcador detectable puede estar fusionado a la enzima CRISPR, de manera que
30 puede ser visualizada la ubicación dentro de una célula, tal como en combinación con medios para detectar la ubicación del núcleo (p. ej., un colorante específico para el núcleo, tal como DAPI). Los núcleos de las células también pueden aislarse de las células, cuyos contenidos pueden luego ser analizados mediante cualquier procedimiento adecuado para la detección de proteínas, tal como inmunohistoquímica, transferencia Western o un ensayo de actividad enzimática. La acumulación en el núcleo también puede determinarse indirectamente tal como
35 mediante un ensayo para el efecto de la formación de complejos CRISPR (p. ej., ensayo para la escisión o mutación de ADN en la secuencia diana, o ensayo para la actividad de la expresión génica alterada, afectada por la formación de complejos de CRISPR y/o actividad de la enzima de CRISPR), en comparación con un control no expuesto a la enzima o un complejo CRISPR, o expuesto a una enzima de CRISPR que carece de la una o más NLSs.

40 En general, una secuencia guía es cualquier secuencia de polinucleótidos que tiene suficiente complementariedad con una secuencia de polinucleótidos diana para hibridarse con la secuencia diana y dirigir la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana. En algunas realizaciones, el grado de complementariedad entre una secuencia guía y su secuencia diana correspondiente, cuando se alinean óptimamente utilizando un algoritmo de alineación adecuado, se trata de o más de aproximadamente 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97,5%, 99%, o más. Un alineamiento óptimo puede ser determinado con el uso de cualquier algoritmo adecuado para el alineamiento de secuencias, ejemplos no limitativos de los cuales incluyen el algoritmo de Smith-Waterman, el algoritmo de Needleman-Wunsch, algoritmos basados en la Transformada de Burrows-Wheeler (p. ej., el Alineador de Burrows Wheeler), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; disponible en www.novocraft.com), ELAND (Ilumina, San Diego, CA), SOAP (disponible en soap.genomics.org.cn) y Maq (disponible en maq.sourceforge.net). En algunas realizaciones, una secuencia guía
50 está relacionada en aproximadamente o más de aproximadamente 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75 o más nucleótidos de longitud. En algunas realizaciones, una secuencia guía es menor que aproximadamente 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, o menos nucleótidos de longitud. La capacidad de una secuencia de guía para dirigir unión específica para la secuencia de un complejo CRISPR a una secuencia diana se puede determinar mediante cualquier ensayo adecuado. Por ejemplo, los
55 componentes de un sistema de CRISPR, suficientes para formar un complejo de CRISPR, incluyendo la secuencia guía a ensayar, se pueden proporcionar a una célula huésped que tiene la secuencia diana correspondiente, tal como por transfección con vectores que codifican los componentes de la secuencia CRISPR, seguido de un estudio de escisión preferencial dentro de la secuencia diana tal como el ensayo Surveyor como se describe en esta memoria. De manera similar, la escisión de una secuencia de polinucleótido diana puede ser evaluada en un tubo de

ensayo proporcionando la secuencia diana, componentes de un complejo de CRISPR, que incluye la secuencia guía a ensayar y una secuencia guía de control diferente de la secuencia guía de ensayo, y comparando la unión o la tasa de escisión en la secuencia diana entre las reacciones de ensayo y de la secuencia guía de control. Son posibles otros ensayos, y se les ocurrirán a los expertos en la técnica.

- 5 Una secuencia guía se puede seleccionar para fijar como objetivo cualquier secuencia diana. En algunas realizaciones, la secuencia diana es una secuencia dentro de un genoma de una célula. Secuencias diana a modo de ejemplo incluyen las que son únicas en el genoma diana. Por ejemplo, para la Cas9 de *S. pyogenes*, una secuencia diana única en un genoma puede incluir un sitio diana Cas9 de la forma MMMMMMMMNNNNNNNNNNNNXGG, en que NNNNNNNNNNNXGG (N es A, G, T, o C; y X puede ser cualquier cosa) tiene una sola aparición en el genoma.
- 10 Una secuencia diana única en un genoma puede incluir un sitio diana Cas9 de *S. pyogenes* de la forma MMMMMMMMNNNNNNNNNNNNXGG, en que NNNNNNNNNNNXGG (N es A, G, T, o C; y X puede ser cualquier cosa) tiene una sola aparición en el genoma. Para la Cas9 de CRISPR1 de *S. thermophilus*, una secuencia diana única en un genoma puede incluir un sitio diana Cas9 de la forma MMMMMMMMNNNNNNNNNNNNXXAGAAW, en que NNNNNNNNNNNXXAGAAW (N es A, G, T, o C; X puede ser cualquier cosa, y W es A o T) tiene una sola aparición en el genoma.
- 15 Una secuencia diana única en un genoma puede incluir un sitio diana Cas9 de CRISPR1 de *S. thermophilus* de la forma MMMMMMMMNNNNNNNNNNNNXXAGAAW, en que NNNNNNNNNNNXXAGAAW (N es A, G, T, o C; X puede ser cualquier cosa, y W es A o T) tiene una sola aparición en el genoma. Para la Cas9 de *S. pyogenes*, una secuencia diana única en un genoma puede incluir un sitio diana Cas9 de la forma MMMMMMMMNNNNNNNNNNNNXGGXG, en que NNNNNNNNNNNXGGXG (N es A, G, T, o C; y X puede ser cualquier cosa) tiene una sola aparición en el genoma.
- 20 Una secuencia diana única en un genoma puede incluir un sitio diana Cas9 de *S. pyogenes* de la forma MMMMMMMMNNNNNNNNNNNNXGGXG, en que NNNNNNNNNNNXGGXG (N es A, G, T, o C; y X puede ser cualquier cosa) tiene una sola aparición en el genoma. En cada una de estas secuencias "M" puede ser A, G, T o C, y no necesita ser considerada en la identificación de una secuencia como la única.
- 25 En algunas realizaciones, se selecciona una secuencia guía para reducir el grado de estructura secundaria dentro de la secuencia guía. En algunas realizaciones, aproximadamente de o menos de aproximadamente 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, o menos de los nucleótidos de la secuencia guía participa en el apareamiento de bases auto-complementario cuando se pliega de manera óptima. El plegamiento óptimo puede ser determinado por cualquier algoritmo de plegamiento de polinucleótidos adecuado. Algunos programas se basan en el cálculo de la energía libre mínima de Gibbs. Un ejemplo de un algoritmo de este tipo es mFold, tal como se describe por Zuker y Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Otro ejemplo de algoritmo de plegamiento está en el servidor web en línea RNAfold, desarrollado en el Instituto de Química Teórica de la Universidad de Viena, utilizando el algoritmo de predicción de estructura centroide (véase, p. ej., A.R. Gruber *et al.*, 2008, *Cell* 106(1): 23-24, y PA Carr y GM Church 2009, *Nature Biotechnology* 27(12):1151-62).
- 30 En general, una secuencia de emparejamiento tracr incluye cualquier secuencia que tiene una suficiente complementariedad con una secuencia tracr para fomentar uno o más de: (1) la escisión de una secuencia guía flanqueada por secuencias de emparejamiento tracr en una célula que contiene la secuencia tracr correspondiente; y (2) la formación de un complejo de CRISPR en una secuencia diana, en donde el complejo CRISPR puede comprender la secuencia de emparejamiento tracr hibridada a la secuencia de tracr. En general, el grado de complementariedad es con referencia al alineamiento óptimo de la secuencia de emparejamiento tracr y la secuencia tracr, a lo largo de la longitud de la más corta de las dos secuencias. El alineamiento óptimo puede ser determinado por cualquier algoritmo de alineamiento adecuado, y puede representar, además, estructuras secundarias tales como la auto-complementariedad dentro de la secuencia de tracr o la secuencia de emparejamiento tracr. En algunas realizaciones, el grado de complementariedad entre la secuencia tracr y la secuencia de emparejamiento tracr a lo largo de la longitud de la más corta de las dos cuando se alinean óptimamente es de aproximadamente o más de aproximadamente 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97,5%, 99%, o superior. En algunas realizaciones, la secuencia de tracr es aproximadamente o más de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, o más nucleótidos de longitud. En algunas realizaciones, la secuencia tracr y la secuencia de emparejamiento tracr están contenidas dentro de un único transcrito, de modo que la hibridación entre los dos produce un transcrito que tiene una estructura secundaria, tal como una horquilla. En una realización de la invención, el transcrito o secuencia de polinucleótido transcrita tiene al menos dos o más horquillas. En realizaciones preferidas, el transcrito tiene dos, tres, cuatro o cinco horquillas. En una realización adicional de la invención, el transcrito tiene a lo sumo cinco horquillas. En una estructura de horquilla, la parte de la secuencia 5' del final "N" y aguas arriba del bucle corresponde a la secuencia de emparejamiento de tracr, y la parte de la secuencia 3' del bucle corresponde a la secuencia de tracr. Ejemplos no limitantes adicionales de polinucleótidos sencillos que pueden comprender una secuencia guía, una secuencia de emparejamiento de tracr y una secuencia de tracr son como sigue (listados de 5' a 3'), en donde "N" representa una base de una secuencia guía, el primer bloque de letras en minúscula representa la secuencia de
- 35
- 40
- 45
- 50
- 55

emparejamiento de tracr y el segundo bloque de letras en minúscula representa la secuencia tracr, y la secuencia poli-T final representa el terminador de la transcripción:

(1)

NNNNNNNNNNNNNNNNNNNNNNNgttttgtactctcaagatttaGAAAtaaactctgcagaagctacaaagataaggctt

5 catgccgaaatcaacaccctgcatcttatggcagggtgttttcggtatttaaTTTTTT; (2)

NNNNNNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatca

acaccctgcatcttatggcagggtgttttcggtatttaaTTTTTT; (3)

NNNNNNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatca

acaccctgcatcttatggcagggtgtTTTTTT; (4)

10 NNNNNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaaggttaaataaggctagtcggtatcaactgaaaa
agtgccaccgagtcggtgcTTTTTT; (5)

NNNNNNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaaggttaaataaggctagtcggtatcaactgaa aaagtgTTTTTTT;

y (6)

15 NNNNNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaaggttaaataaggctagtcggtatcaTTTTT TTT. En algunas realizaciones, las secuencias (1) a (3) se utilizan en combinación con Cas9 de *S. thermophilus* CRISPR1. En algunas realizaciones, las secuencias (4) a (6) se utilizan en combinación con Cas9 de *S. pyogenes*. En algunas realizaciones, la secuencia de tracr es un transcrito separado de un transcrito que puede comprender la secuencia de emparejamiento de tracr.

20 En algunas realizaciones, también se proporciona un molde de recombinación. Un molde de recombinación puede ser un componente de otro vector tal como se describe en esta memoria, contenido en un vector separado, o proporcionado como un polinucleótido separado. En algunas realizaciones, un molde de recombinación está diseñado para servir como un molde en la recombinación homóloga tal como dentro o cerca de una secuencia diana mellada o escindida por una enzima de CRISPR como parte de un complejo de CRISPR. Un polinucleótido molde puede ser de cualquier longitud adecuada, tal como aproximadamente o más de aproximadamente 10, 15, 20, 25, 25, 50, 75, 100, 150, 200, 500, 1000, o más nucleótidos de longitud. En algunas realizaciones, el polinucleótido molde es complementario a una parte de un polinucleótido que puede comprender la secuencia diana. Cuando se alinean óptimamente, un polinucleótido molde podría solaparse con uno o más nucleótidos de una secuencia diana (p. ej., aproximadamente o más de aproximadamente 1, 5, 10, 15, 20, o más nucleótidos). En algunas realizaciones, cuando una secuencia de un molde y un polinucleótido, que puede comprender una secuencia diana, se alinean de manera óptima, el nucleótido más próximo del polinucleótido del molde está dentro de aproximadamente 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, o más nucleótidos de la secuencia diana.

35 En algunas realizaciones, la enzima de CRISPR es parte de una proteína de fusión que puede comprender uno o más dominios de proteínas heterólogos (p. ej., aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más dominios, además de la enzima de CRISPR). Una proteína de fusión enzima de CRISPR puede comprender cualquier secuencia de la proteína adicional, y opcionalmente una secuencia de unión entre cualquiera de dos dominios. Ejemplos de dominios de proteínas que pueden fusionarse con una enzima de CRISPR incluyen, sin limitación, macadores de epítomos, secuencias de gen indicador y dominios de proteínas que tienen una o más de las siguientes actividades: actividad metilasa, actividad desmetilasa, actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de la transcripción, actividad de la modificación de histonas, actividad de escisión del ARN y actividad de unión a ácido nucleico. Ejemplos no limitantes de marcadores de epítomos incluyen marcadores de histidina (His), marcadores V5, marcadores FLAG, marcadores de hemaglutinina de la gripe (HA), marcadores Myc, marcadores VSV-G y marcadores de tiorredoxina (Trx). Ejemplos de genes informadores incluyen, pero no se limitan a, glutatión-S-transferasa (GST), peroxidasa de rábano picante (HRP), cloranfenicol acetiltransferasa (CAT), beta-galactosidasa, beta-glucuronidasa, luciferasa, proteína fluorescente verde (GFP), HcRed, DsRed, proteína fluorescente cian (CFP), proteína amarilla fluorescente (YFP) y proteínas autofluorescentes incluyendo la proteína fluorescente azul (BFP). Una enzima de CRISPR puede fusionarse a una secuencia de gen que codifica una proteína o un fragmento de una proteína que une moléculas de ADN o une otras moléculas celulares, incluyendo pero no limitada a la proteína de unión a maltosa (MBP), S-tag, fusiones del dominio de unión a ADN (DBD) de Lex A, fusiones del dominio de unión a ADN de GAL4 y fusiones de la proteína BP16 del virus herpes simplex (HSV). Dominios adicionales que pueden formar parte de una proteína de

fusión que puede comprender una enzima de CRISPR se describen en el documento US20110059502. En algunas realizaciones, se utiliza una enzima de CRISPR marcada para identificar la ubicación de una secuencia diana.

En algunos aspectos, la invención proporciona métodos que pueden comprender suministrar uno o más polinucleótidos, tales como o uno o más vectores tal como se describen en esta memoria, uno o más transcritos de los mismos, y/o una o más proteínas transcritas de los mismos, a una célula huésped. En algunos aspectos, la invención proporciona, además, células producidas por tales métodos, y animales que pueden comprender o producidos a partir de tales células. En algunas realizaciones, una enzima de CRISPR en combinación con (y opcionalmente formando complejo con) una secuencia guía se suministra a una célula. Métodos de transferencia génica basados en virus y no virales convencionales pueden utilizarse para introducir ácidos nucleicos en células de mamífero o tejidos diana. Tales métodos se pueden utilizar para administrar ácidos nucleicos que codifican componentes de un sistema de CRISPR para células en cultivo, o en un organismo huésped. Sistemas de suministro de vectores no virales incluyen plásmidos de ADN, ARN (p. ej., un transcrito de un vector descrito en esta memoria), ácido nucleico desnudo y ácido nucleico formando complejo con un vehículo de suministro, tal como un liposoma. Sistemas de suministro de vectores virales incluyen virus de ADN y ARN, que tienen genomas episomales o integrados después de suministro a la célula. Para una revisión de procedimientos de terapia génica, véase Anderson, *Science* 256: 808- 813 (1992); Nabel y Felgner, *TIBTECH* 11: 211-217 (1993); Mitani y Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357: 455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer y Perricaudet. *British Medical Bulletin* 51(1):31-44 (1995); Haddada et al., en *Current Topics in Microbiology and Immunology Doerfler y Böhm (comps.)* (1995); y Yu et al, *Gene Therapy* 1:13-26 (1994).

Métodos de suministro no viral de ácidos nucleicos incluyen lipofección, microinyección, biolística, virosomas, liposomas, inmunoliposomas, poliatión o conjugados de lípido:ácido nucleico, ADN desnudo, viriones artificiales y absorción potenciada por agente de ADN. La lipofección se describe, p. ej., en las patentes de EE.UU. N.ºs. 5.049.386, 4.946.787; y 4.897.355) y reactivos de lipofección se venden comercialmente (p. ej., Transfectam™ y Lipofectin™). Lípidos catiónicos y neutros que son adecuados para una lipofección de reconocimiento de receptores eficiente de polinucleótidos incluyen los de Felgner, documento WO 91/17424; documento WO 91/16024. El suministro puede ser a células (p. ej., administración in vitro o ex vivo) o tejidos diana (p. ej., la administración in vivo).

La preparación de complejos de lípido:ácido nucleico, incluyendo liposomas fijados como objetivo tales como complejos de inmunolípidos, es bien conocida para un experto en la técnica (véase, p. ej., Crystal, *Science* 270: 404-410 (1995); Blaese et al, *Cancer Gene Ther.* 2: 291-297 (1995); Behr et al, *Bioconjugate Chem.* 5: 382-389 (1994); Remy et al, *Bioconjugate Chem* 5: 647-654 (1994); Gao et al., *Gene Therapy* 2:710-722 (1995); Ahmad et al., *Cancer Res.* 52: 4817-4820 (1992); patentes de EE.UU. N.ºs 4.186.183, 4.217.344, 4.235.871, 4.261.975, 4.485.054, 4.501.728, 4.774.085, 4.837.028 y 4.946.787).

El uso de sistemas basados en virus de ARN o ADN para el suministro de ácidos nucleicos aprovecha los procesos altamente evolucionados para fijar como objetivo un virus a células específicas en el cuerpo y el tráfico de la carga útil viral al núcleo. Vectores virales pueden administrarse directamente a los pacientes (in vivo) o pueden utilizarse para tratar células in vitro y las células modificadas se pueden opcionalmente administrar a pacientes (ex vivo). Sistemas basados en virus convencionales podrían incluir vectores retroviral, de lentivirus, de virus adeno-asociados y de virus del herpes simplex para la transferencia de genes. La integración en el genoma huésped es posible con métodos de transferencia de genes de retrovirus, lentivirus y virus adeno-asociados, a menudo resulta en la expresión a largo plazo del transgén insertado. Adicionalmente, las elevadas eficiencias de la transducción se han observado en muchos tipos de células y tejidos diana diferentes.

El tropismo de un retrovirus puede alterarse por incorporación de proteínas de la envoltura extraña y expansión de la población diana potencial de células diana. Los vectores lentivirales son vectores retrovirales que son capaces de transducir o infectar células que no se dividen y que producen típicamente altos títulos virales. Por lo tanto, la selección de un sistema de transferencia de genes retroviral dependería del tejido diana. Los vectores retrovirales se componen de repeticiones cis-terminales de larga acción con capacidad de empaquetamiento para hasta 6-10 kb de secuencia extraña. LTRs que actúan en cis mínimos son suficientes para la replicación y el empaquetamiento de los vectores, que se utilizan luego para integrar el gen terapéutico en la célula objetivo para proporcionar una expresión del transgén permanente. Vectores retrovirales ampliamente utilizados incluyen los basados en virus de la leucemia murina (MuLV), virus de la leucemia del mono gibón (GaLV), virus de la inmunodeficiencia de simios (SIV), virus de la inmunodeficiencia humana (VIH) y combinaciones de los mismos (véase, p. ej., Buchscher et al., *J. Virol.* 66:2731-2739 (1992); Johann et al, *J. Virol.* 66: 1635-1640 (1992); Sommerfelt et al, *Virol.* 176:58-59 (1990); Wilson et al., *J. Virol.* 63: 2374-2378 (1989); Miller et al., *J. Virol.* 65:2220-2224 (1991); documento PCT/US94/05700).

En aplicaciones en las que se prefiere la expresión transitoria, se pueden utilizar sistemas basados en adenovirus. Los vectores basados en adenovirus son capaces de una muy alta eficiencia de transducción en muchos tipos de células y no requieren la división celular. Con este tipo de vectores se han obtenido un título y niveles de expresión altos. Este vector puede producirse en grandes cantidades en un sistema relativamente simple. Vectores de virus adeno-asociados ("AAV") también pueden utilizarse para transducir células con ácidos nucleicos diana, p. ej., en la producción in vitro de ácidos nucleicos y péptidos, y para procedimientos de terapia génica in vivo y ex vivo (véase, p. ej., West et al., *Virology* 160:38-47 (1987); patente de EE.UU. Nº. 4.797.368; documento WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). La construcción de vectores AAV recombinantes se describe en un cierto número de publicaciones, incluyendo la patente de EE.UU. Nº. 5,173,414; Tratschin et al., *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, et al., *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat y Muzyczka, *PNAS* 81:6466-6470 (1984); y Samulski et al., *J. Virol.* 63: 03822-3828 (1989).

Células de empaquetamiento se utilizan típicamente para formar partículas de virus que son capaces de infectar una célula huésped. Tales células incluyen células 293, que empaquetan adenovirus, y células ψ 2 o células PA317, que empaquetan retrovirus. Los vectores virales utilizados en terapia génica se generan habitualmente por el productor de una línea celular que empaqueta un vector de ácido nucleico en una partícula viral. Los vectores contienen típicamente las secuencias virales mínimas requeridas para el empaquetamiento y la posterior integración en un huésped, siendo reemplazadas otras secuencias virales por una casete de expresión para el o los polinucleótidos a expresar. Las funciones virales que faltan se suministran normalmente en trans por la línea celular de empaquetamiento. Por ejemplo, los vectores AAV utilizados en terapia génica poseen típicamente sólo secuencias ITR del genoma de AAV que se requieren para el empaquetamiento y la integración en el genoma huésped. El ADN viral se empaqueta en una línea celular, que contiene un plásmido auxiliar que codifica los otros genes AAV, a saber rep y cap, pero que carece de secuencias ITR. La línea celular también puede estar infectada con adenovirus como un helper (cooperador). El virus helper fomenta la replicación del vector AAV y la expresión de genes AAV a partir del plásmido helper. El plásmido helper no está empaquetado en cantidades significativas debido a una carencia de secuencias ITR. La contaminación con adenovirus puede reducirse mediante, p. ej., tratamiento térmico al que el adenovirus es más sensible que AAV. Métodos adicionales para el suministro de ácidos nucleicos a las células son conocidos por los expertos en la técnica. Véase, por ejemplo, el documento US20030087817.

En algunas realizaciones, una célula huésped es transfectada de forma transitoria o no transitoria con uno o más vectores descritos en esta memoria. En algunas realizaciones, una célula se transfecta tal como se produce de forma natural en un sujeto. En algunas realizaciones, una célula que se transfecta se toma de un sujeto. En algunas realizaciones, la célula se deriva de células tomadas de un sujeto tal como una línea celular. En la técnica se conoce una amplia diversidad de líneas celulares para el cultivo de tejidos. Ejemplos de líneas celulares incluyen, pero no se limitan a, C8161 CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEK_n, HEK_a, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRGB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 epitelial del riñón de mono, fibroblastos de embrión de ratón BALB/ 3T3, 3T3 Swiss, 3T3-L1, fibroblastos fetales humanos 132-d5; fibroblastos de ratón 10.1, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, células BCP-1, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr -/-, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, células JY, células K562, Ku812, KCL22, KG1, KEYO1, LNCap, Ma-Me1 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN / OPCT cell lines, Peer, PNT-1A / PNT 2, RenCa, RIN-5F, RMA/RMAS, células Saos-2, Sf-9, SkBr3, T2, T-47D, T84, línea celular THP1, U373, U87, U937, VCaP, células Vero, WM39, WT-49, X63, YAC-1, YAR, y variedades transgénicas de las mismas. Las líneas celulares están disponibles de una diversidad de fuentes conocidas por los expertos en la técnica (véase, p. ej., American Type Culture Collection (ATCC) (Manassus, Va.)). En algunas realizaciones, una célula transfectada con uno o más vectores descritos en esta memoria se utiliza para establecer una nueva línea celular que puede comprender una o más secuencias derivadas del vector. En algunas realizaciones, una célula transfectada transitoriamente con los componentes de un sistema de CRISPR, tal como se describe en esta memoria (por ejemplo, por transfección transitoria de uno o más vectores, o la transfección con ARN), y modificadas a través de la actividad de un complejo de CRISPR, se utiliza para establecer una nueva línea celular que puede comprender células que contienen la modificación, pero que carecen de cualquier otra secuencia exógena. En algunas realizaciones, las células transfectadas de forma transitoria o no transitoria con uno o más vectores descritos en esta memoria, o líneas celulares derivadas de estas células se utilizan en la evaluación de uno o más compuestos de ensayo.

En algunas realizaciones, uno o más vectores descritos en esta memoria se utilizan para producir un animal transgénico no humano o planta transgénica. En algunas realizaciones, el animal transgénico es un mamífero, tal como un ratón, rata o conejo. Métodos para producir plantas y animales transgénicos son conocidos en la técnica, y generalmente comienzan con un método de transfección celular tal como se describe en esta memoria.

5 Con los recientes progresos en genómica de cultivos, la capacidad de utilizar los sistemas de CRISPR-Cas para llevar a cabo la edición y la manipulación de genes eficiente y rentable permitirá la rápida selección y la comparación de manipulaciones genéticas individuales y multiplexadas para transformar dichos genomas para una producción mejorada y rasgos potenciados. A este respecto se hace referencia a las patentes y publicaciones de EE.UU. Patente de EE.UU. N° 6.603.061 - Agrobacterium-Mediated Plant Transformation Method; patente de EE.UU. N° 7.868.149 - Plant Genome Sequences and Uses Thereof y de EE.UU. 2009/0100536 – Transgenic Plants with Enhanced Agronomic Traits. En la práctica de la invención también se mencionan el contenido y la descripción de Morrell et al "Crop genomics: advances and applications" Nat Rev Genet. 29 dic. 2011;13(2):85-96 En una forma de realización ventajosa de la invención, el sistema CRISPR/Cas9 se utiliza para modificar genéticamente microalgas. Por consiguiente, referencia en esta memoria a células animales también se puede aplicar, mutatis mutandis, a células vegetales, a menos que de otro modo resulte evidente.

En plantas, los patógenos son a menudo específicos para el huésped. Por ejemplo, *Fusarium oxysporum* f. sp. *lycopersici* provoca el marchitamiento del tomate, pero ataca sólo al tomate, y *F. oxysporum* f. *dianthii* *Puccinia graminis* f. sp. *tritici* ataca solamente al trigo. Las plantas tienen defensas existentes e inducidas para resistir frente a la mayoría de los patógenos. Mutaciones y eventos de recombinación a través de generaciones de plantas conducen a la variabilidad genética que da lugar a la susceptibilidad, especialmente ya que los patógenos se reproducen con más frecuencia que las plantas. En las plantas puede no haber resistencia al huésped, p. ej., el huésped y el patógeno son incompatibles. También puede haber Resistencia Horizontal, p. ej., la resistencia parcial contra todas las razas de un patógeno, típicamente controladas por muchos genes, y la Resistencia Vertical, p. ej., resistencia completa a algunas razas de un patógeno pero no a otras razas, típicamente controlada por unos pocos genes. En un nivel de Gen-a-Gen, las plantas y los patógenos evolucionan juntos, y los cambios genéticos en un equilibrio cambian a otro. En consecuencia, utilizando la Variabilidad Natural, los criadores combinan los genes más útiles para el Rendimiento, la Calidad, Uniformidad, Dureza, Resistencia. Las fuentes de genes de resistencia incluyen Variedades nativas o extrañas, variedades de Reliquia, Parientes de Plantas Silvestres y Mutaciones Inducidas, p. ej. tratando el material vegetal con agentes mutagénicos. Utilizando la presente invención, los criadores de plantas están provistos de una nueva herramienta para inducir mutaciones. Por consiguiente, un experto en la técnica puede analizar el genoma de fuentes de genes de resistencia, y en las Variedades que tengan características o rasgos deseados, emplear la presente invención para inducir el aumento de genes de resistencia, con más precisión que los agentes mutagénicos anteriores y, por lo tanto, para acelerar y mejorar los programas de cultivo.

35 En un aspecto, la invención proporciona métodos de modificar un polinucleótido diana en una célula eucariota, que puede ser in vivo, ex vivo o in vitro. En algunas realizaciones, el método comprende el muestreo de una célula o población de células de un animal humano o no humano o planta (incluyendo micro-algas), y la modificación de la célula o células. El cultivo puede producirse en cualquier fase ex vivo. La célula o las células pueden incluso ser reintroducidos en el animal no humano o la planta (incluyendo micro-algas). Para las células reintroducidas se prefiere particularmente que las células sean células madre.

En un aspecto, la invención proporciona métodos de modificar un polinucleótido diana en una célula eucariota. En algunas realizaciones, el método puede comprender permitir que un complejo de CRISPR se una al polinucleótido diana para efectuar la escisión de dicho polinucleótido diana, modificando así el polinucleótido diana, en el que el complejo de CRISPR puede comprender una enzima de CRISPR formando complejo con una secuencia guía hibridada a una secuencia diana dentro de dicho polinucleótido diana, en el que dicha secuencia guía está unida a una secuencia de emparejamiento tracr que a su vez se hibrida a una secuencia de tracr.

En un aspecto, la invención proporciona un método de modificar la expresión de un polinucleótido en una célula eucariota. En algunas realizaciones, el método puede comprender permitir que un complejo de CRISPR se una al polinucleótido de manera que dicha unión resulte en una expresión incrementada o reducida de dicho polinucleótido; en el que el complejo de CRISPR puede comprender una enzima de CRISPR formando complejo con una secuencia guía hibridada a una secuencia diana dentro de dicho polinucleótido, en el que dicha secuencia guía está enlazada a una secuencia de emparejamiento tracr que a su vez se hibrida a una secuencia de tracr.

En un aspecto, la invención proporciona kits que contienen uno o más de los elementos descritos en los métodos y las composiciones anteriores. Los elementos pueden proporcionarse individualmente o en combinaciones, y se

pueden proporcionar en cualquier recipiente adecuado, tal como un vial, una botella, o un tubo. En algunas realizaciones, el kit incluye instrucciones en uno o más idiomas, por ejemplo, en más de un idioma.

5 En algunas realizaciones, un kit puede comprender uno o más reactivos para su uso en un proceso que utiliza uno o más de los elementos descritos en esta memoria. Los reactivos pueden proporcionarse en cualquier recipiente adecuado. Por ejemplo, un kit puede proporcionar uno o más tampones de reacción o de almacenamiento. Los reactivos se pueden proporcionar en una forma que sea utilizable en un ensayo particular, o en una forma que requiere la adición de uno o más de otros componentes antes de su uso (p. ej., en forma de concentrado o liofilizada). Un tampón puede ser cualquier tampón, incluyendo, pero no limitado a un tampón carbonato de sodio, un tampón bicarbonato de sodio, un tampón borato, un tampón Tris, un tampón MOPS, un tampón HEPES, y
10 combinaciones de los mismos. En algunas realizaciones, el tampón es alcalino. En algunas realizaciones, el tampón tiene un pH de aproximadamente 7 a aproximadamente 10. En algunas realizaciones, el kit puede comprender uno o más oligonucleótidos correspondientes a una secuencia guía para la inserción en un vector con el fin de enlazar operativamente la secuencia guía y un elemento regulador. En algunas realizaciones, el kit puede comprender un polinucleótido molde de recombinación homóloga.

15 En un aspecto, la invención proporciona métodos para usar uno o más elementos de un sistema de CRISPR. El complejo de CRISPR de la invención proporciona un medio eficaz para la modificación de un polinucleótido diana. El complejo de CRISPR de la invención tiene una amplia variedad de utilidad, incluyendo modificar (por ejemplo, suprimir, insertar, translocar, inactivar, activar) un polinucleótido diana en una multiplicidad de tipos de células. Como tal, el complejo de CRISPR de la invención tiene un amplio espectro de aplicaciones en, p. ej., la terapia génica, el
20 rastreo de fármacos, el diagnóstico de enfermedades y el pronóstico. Un complejo de CRISPR a modo de ejemplo puede comprender una enzima de CRISPR formando complejo con una secuencia guía hibridada a una secuencia diana dentro del polinucleótido diana. La secuencia guía está enlazada a una secuencia de emparejamiento tracr que, a su vez, se hibrida a una secuencia de tracr.

25 En una realización, esta invención proporciona un método para escindir un polinucleótido diana. El método puede comprender modificar un polinucleótido diana utilizando un complejo de CRISPR que se une al polinucleótido diana y efectúa la escisión de dicho polinucleótido diana. Típicamente, el complejo de CRISPR de la invención, cuando se introduce en una célula, crea una ruptura (p. ej., una rotura de la cadena sencilla o doble) en la secuencia del genoma. Por ejemplo, el método puede utilizarse para escindir un gen de la enfermedad en una célula.

30 La ruptura creada por el complejo de CRISPR puede ser reparada por unos procesos de reparación tales como la vía de unión de extremos no homólogos (NHEJ) propensa a errores o la reparación dirigida a la homología de alta fidelidad (HDR). Durante estos procesos de reparación, un molde de polinucleótido exógeno puede ser introducido en la secuencia del genoma. En algunos métodos, el proceso de HDR se utiliza para modificar la secuencia del genoma. Por ejemplo, un molde de polinucleótido exógeno que puede comprender una secuencia que se ha de integrar flanqueada por una secuencia aguas arriba y una secuencia aguas abajo se introduce en una célula. Las
35 secuencias aguas arriba y aguas abajo comparten similitud de secuencia con cualquiera de los lados del sitio de integración en el cromosoma.

40 Cuando se desee, un polinucleótido donante puede ser ADN, p. ej., un plásmido de ADN, un cromosoma bacteriano artificial (BAC), un cromosoma artificial de levadura (YAC), un vector viral, un trozo lineal de ADN, un fragmento de la PCR, un ácido nucleico desnudo, o un ácido nucleico formando complejo con un vehículo de suministro tal como un liposoma o poloxámero.

45 El molde de polinucleótido exógeno puede comprender una secuencia para ser integrada (p. ej., un gen mutado). La secuencia de integración puede ser una secuencia endógena o exógena a la célula. Ejemplos de una secuencia a ser integrada incluyen polinucleótidos que codifican una proteína o un ARN no codificante (p. ej., un microARN). Por lo tanto, la secuencia de la integración puede estar operativamente enlazada a una secuencia o secuencias de control apropiadas. Alternativamente, la secuencia a ser integrada puede proporcionar una función reguladora.

50 Las secuencias de aguas arriba y aguas abajo del molde de polinucleótido exógeno se seleccionan para fomentar la recombinación entre la secuencia cromosómica de interés y el polinucleótido donante. La secuencia de aguas arriba es una secuencia de ácido nucleico que comparte similitud de secuencia con la secuencia del genoma de aguas arriba del sitio diana para la integración. De manera similar, la secuencia de aguas abajo es una secuencia de ácido nucleico que comparte similitud de secuencia con la secuencia cromosómica aguas abajo del sitio diana de la integración. Las secuencias aguas arriba y aguas abajo en el molde de polinucleótido exógeno puede tener 75%, 80%, 85%, 90%, 95% o 100% de identidad de secuencia con la secuencia del genoma de destino. Preferiblemente, las secuencias de aguas arriba y aguas abajo en la plantilla de polinucleótido exógeno tienen aproximadamente 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia del genoma fijada como objetivo. En

algunos métodos, las secuencias de aguas arriba y aguas abajo en el molde de polinucleótido exógeno tienen aproximadamente el 99% o el 100% de identidad de la secuencia con la secuencia del genoma fijada como objetivo.

5 Una secuencia de aguas arriba o aguas abajo puede comprender de aproximadamente 20 pb a aproximadamente 2500 pb, por ejemplo, aproximadamente 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400 ó 2500 pb. En algunos métodos, la secuencia aguas arriba o aguas abajo a modo de ejemplo tiene aproximadamente 200 pb a aproximadamente 2000 pb, aproximadamente 600 pb a aproximadamente 1000 pb, o más particularmente aproximadamente 700 pb a aproximadamente 1000 pb.

10 En algunos métodos, el molde de polinucleótido exógeno puede comprender, además, un marcador. Este marcador puede hacer más fácil rastrear integraciones fijadas como objetivo. Ejemplos de marcadores adecuados incluyen sitios de restricción, proteínas fluorescentes o marcadores seleccionables. El molde de polinucleótido exógeno de la invención puede construirse utilizando técnicas recombinantes (véase, por ejemplo, Sambrook et al., 2001 y Ausubel et al., 1996).

15 En un método a modo de ejemplo, para la modificación de un polinucleótido diana mediante la integración de un molde de polinucleótido exógeno, se introduce en la secuencia del genoma una rotura de la cadena doble por el complejo de CRISPR, la ruptura se repara mediante recombinación homóloga de un molde de polinucleótido exógeno de tal forma que el molde se integra en el genoma. La presencia de una rotura de cadena doble facilita la integración del molde.

20 En otras realizaciones, esta invención proporciona un método para modificar la expresión de un polinucleótido en una célula eucariota. El método puede comprender aumentar o disminuir la expresión de un polinucleótido diana utilizando un complejo de CRISPR que se une al polinucleótido,

25 En algunos métodos, un polinucleótido diana puede ser inactivado para efectuar la modificación de la expresión en una célula. Por ejemplo, tras la unión de un complejo de CRISPR a una secuencia diana en una célula, el polinucleótido diana se inactiva de manera que la secuencia no se transcribe, la proteína codificada no se produce o la secuencia no funciona como la secuencia de tipo salvaje. Por ejemplo, una proteína o secuencia codificadora de microARN puede inactivarse de manera que no se produzca la proteína.

30 En algunos métodos, una secuencia de control puede ser inactivada de manera que ya no funcione como una secuencia de control. Tal como se utiliza en esta memoria, "secuencia de control" se refiere a cualquier secuencia de ácido nucleico que efectúa la transcripción, traducción o la accesibilidad de una secuencia de ácido nucleico. Ejemplos de una secuencia de control incluyen un promotor, un terminador de la transcripción, y un promotor son secuencias de control.

35 La secuencia diana inactivada puede incluir una mutación por delección (es decir, la delección de uno o más nucleótidos), una mutación por inserción (es decir, la inserción de uno o más nucleótidos), o una mutación sin sentido (es decir, la sustitución de un único nucleótido con otro nucleótido de modo que se introduce un codón de parada). En algunos métodos, la inactivación de una secuencia diana da como resultado la "inactivación" de la secuencia diana.

40 Un método de la invención puede utilizarse para crear una planta, un animal o una célula que puede ser usado como un modelo de enfermedad. Tal como se utiliza en esta memoria, "enfermedad" se refiere a una enfermedad, trastorno o indicación en un sujeto. Por ejemplo, un método de la invención puede utilizarse para crear un animal o célula que puede comprender una modificación en una o más secuencias de ácido nucleico asociadas con una enfermedad, o un animal o célula en el que se altera la expresión de una o más secuencias de ácido nucleico asociada con una enfermedad. Una secuencia de ácido nucleico de este tipo puede codificar una secuencia de proteína asociada a la enfermedad o puede ser una secuencia de control asociada a la enfermedad.

45 En algunos métodos, el modelo de la enfermedad se puede utilizar para estudiar los efectos de las mutaciones en el animal o la célula y el desarrollo y/o progreso de la enfermedad utilizando medidas de uso común en el estudio de la enfermedad. Alternativamente, un modelo de enfermedad de este tipo es útil para estudiar el efecto de un compuesto farmacéuticamente activo en la enfermedad.

50 En algunos métodos, el modelo de la enfermedad se puede utilizar para evaluar la eficacia de una potencial estrategia de terapia génica. Es decir, un gen o polinucleótido asociado a la enfermedad puede ser modificado de manera que se inhiba o se reduzca el desarrollo y/o el progreso de la enfermedad. En particular, el método puede

comprender modificar un gen asociado a la enfermedad o polinucleótido de modo que se produzca una proteína alterada y, como resultado, el animal o célula tiene una respuesta alterada. Por consiguiente, en algunos métodos, un animal modificado genéticamente puede ser comparado con un animal predispuesto al desarrollo de la enfermedad de manera que puede evaluarse el efecto del evento terapia génica.

5 En otra realización, esta invención proporciona un método de desarrollar un agente biológicamente activo que modula un evento de señalización celular asociado con un gen de la enfermedad. El método puede comprender poner en contacto un compuesto de ensayo con una célula que puede comprender uno o más vectores que impulsan la expresión de una o más de una enzima de CRISPR, una secuencia guía enlazada a una secuencia de emparejamiento tracr y una secuencia de tracr; y detectar un cambio en una lectura que es indicativo de una
10 reducción o un aumento de un evento de señalización celular asociada con, p. ej., una mutación en un gen de la enfermedad contenido en la célula.

Un modelo de célula o modelo animal se puede construir en combinación con el método de la invención para el rastreo de un cambio de la función celular. Un modelo de este tipo se puede utilizar para estudiar los efectos de una secuencia del genoma modificada por el complejo de CRISPR de la invención en una función celular de interés. Por
15 ejemplo, un modelo de la función celular se puede utilizar para estudiar el efecto de una secuencia de genoma modificado en la señalización intracelular o la señalización extracelular. Alternativamente, un modelo de la función celular se puede utilizar para estudiar los efectos de una secuencia del genoma modificado en la percepción sensorial. En algunos de este tipo de modelos, se modifican las una o más secuencias del genoma asociadas con una vía bioquímica de señalización en el modelo.

20 Una expresión alterada de una o más secuencias del genoma asociadas con una vía de señalización bioquímica puede ser determinada por ensayo de una diferencia en los niveles de ARNm de los genes correspondientes entre la célula modelo de ensayo y una célula control, cuando están en contacto con un agente candidato. Alternativamente, la expresión diferencial de las secuencias asociadas con una vía de señalización bioquímica se determina mediante la detección de una diferencia en el nivel del producto génico o polipéptido codificado.

25 Para ensayar para una alteración inducida por el agente en el nivel de transcritos de ARNm o polinucleótidos correspondientes, ácido nucleico contenido en una muestra se extrae primero de acuerdo con métodos estándares en la técnica. Por ejemplo, el ARNm puede aislarse utilizando diversas enzimas líticas o disoluciones químicas de acuerdo con los procedimientos establecidos en Sambrook et al. (1989), o puede extraerse mediante resinas de unión a ácido nucleico siguiendo las instrucciones adjuntas proporcionadas por los fabricantes. El ARNm contenido
30 en la muestra de ácido nucleico extraído se detecta a continuación mediante procedimientos de amplificación o ensayos de hibridación convencionales (p. ej., análisis de transferencia Northern) de acuerdo con métodos ampliamente conocidos en la técnica o en base a los métodos ejemplificados en el presente documento.

Para los fines de esta invención, amplificación significa cualquier método que emplea un cebador y una polimerasa capaz de replicar una secuencia diana con fidelidad razonable. La amplificación puede ser portada por el ADN-
35 polimerasas naturales o recombinantes tales como TaqGold™, T7 ADN polimerasa, el fragmento Klenow de la ADN polimerasa de E. coli y transcriptasa inversa. Un método de amplificación preferido es la PCR. En particular, el ARN aislado se puede someter a un ensayo de transcripción inversa que está acoplado con una reacción en cadena de la polimerasa cuantitativa (RT-PCR) con el fin de cuantificar el nivel de expresión de una secuencia asociada con una vía de señalización bioquímica.

40 La detección del nivel de expresión génica puede llevarse a cabo en tiempo real en un ensayo de amplificación. En un aspecto, los productos amplificados pueden ser directamente visualizados con agentes de unión a ADN fluorescentes, incluyendo, pero no limitados a, intercaladores de ADN y aglomerantes de de la muesca de ADN. Debido a que la cantidad de los intercaladores incorporados en las moléculas de ADN de doble cadena es típicamente proporcional a la cantidad de los productos de ADN amplificados, se puede determinar
45 convenientemente la cantidad de los productos amplificados mediante la cuantificación de la fluorescencia del colorante intercalado utilizando sistemas ópticos convencionales en la técnica. Colorante de unión a ADN adecuado para esta aplicación incluyen verde SYBR, azul SYBR, DAPI, yoduro de propidio, Hoeste, oro SYBR, bromuro de etidio, acridinas, proflavina, naranja de acridina, acriflavina, fluorcoumanina, elipticina, daunomicina, cloroquina, distamicina D, cromomicina, homidio, mitramicina, polipiridilos rutenio, antramycin, y similares.

50 En otro aspecto, otros marcadores fluorescentes tales como sondas específicas para la secuencia se pueden emplear en la reacción de amplificación para facilitar la detección y cuantificación de los productos amplificados. La amplificación cuantitativa basada en la sonda se basa en la detección específica para la secuencia de un producto amplificado deseado. Utiliza, sondas específicas para la diana, fluorescentes (p. ej., sondas TaqMan®) resultando

en una mayor especificidad y sensibilidad. Métodos para realizar la amplificación cuantitativa basada en la sonda están bien establecidos en la técnica y se enseña en la Patente de EE.UU. Nº 5.210.015.

5 Todavía en otro aspecto, se pueden realizar ensayos de hibridación convencionales utilizando sondas de hibridación que comparten homología de secuencia con las secuencias asociadas con una vía de señalización bioquímica. Típicamente, se deja que las sondas formen complejos estables con las secuencias asociadas con una vía de
 10 señalización bioquímica contenida dentro de la muestra biológica derivada del sujeto de ensayo en una reacción de hibridación. Se apreciará por un experto en la técnica que, cuando se utilice antisentido como el ácido nucleico sonda, los polinucleótidos diana proporcionados en la muestra se eligen para que sean complementarios a las secuencias de los ácidos nucleicos antisentido. En cambio, cuando la sonda de nucleótidos es un ácido nucleico
 15 sentido, el polinucleótido diana se selecciona para que sea complementario a las secuencias del ácido nucleico sentido.

La hibridación puede realizarse en condiciones de diferentes rigurosidad. Condiciones de hibridación adecuadas para la práctica de la presente invención son tales que la interacción de reconocimiento entre la sonda y secuencias
 15 asociadas con una vía de señalización bioquímica es a la vez lo suficientemente específica y lo suficientemente estable. Condiciones que aumentan la rigurosidad de una reacción de hibridación son ampliamente conocidas y están publicadas en la técnica. Véase, por ejemplo, Sambrook, et al., (1989).; Nonradioactive In Situ Hybridization Application Manual, Boehringer Mannheim, segunda edición. El ensayo de hibridación se puede formar utilizando
 20 sondas inmovilizadas sobre cualquier soporte sólido, incluyendo, pero no limitadas a nitrocelulosa, vidrio, silicio, y una diversidad de matrices de genes. Un ensayo de hibridación preferido se lleva a cabo en chips de genes de alta densidad tal como se describe en la Patente de EE.UU. Nº 5.445.934.

Para una detección conveniente de los complejos de sonda-diana que se forman durante el ensayo de hibridación, las sondas de nucleótidos se conjugan a un marcador detectable. Marcadores detectables adecuados para su uso
 25 en la presente invención incluyen cualquier composición detectable por medios fotoquímicos, bioquímicos, espectroscópicos, inmunoquímicos, eléctricos, ópticos o químicos. Una amplia variedad de marcadores detectables adecuados son conocidos en la técnica, que incluyen marcadores fluorescentes o quimioluminiscentes, marcadores de isótopos radiactivos, enzimáticos u otros ligandos. En realizaciones preferidas, probablemente se deseará emplear un marcador fluorescente o una etiqueta de enzima tal como digoxigenina, β -galactosidasa, ureasa, fosfatasa alcalina o peroxidasa, complejo de avidina/biotina.

Los métodos de detección utilizados para detectar o cuantificar la intensidad de hibridación dependerán típicamente
 30 del marcador arriba seleccionado. Por ejemplo, los marcadores radiactivos se pueden detectar utilizando una película fotográfica o un phosphoimager. Los marcadores fluorescentes pueden ser detectados y cuantificados utilizando un fotodetector para detectar la luz emitida. Los marcadores enzimáticos se detectan típicamente proveyendo a la enzima de un sustrato y midiendo el producto de reacción producido por la acción de la enzima sobre el sustrato; y finalmente marcadores colorimétricos son detectadas simplemente visualizando el marcador
 35 coloreado.

Un cambio inducido por agente en la expresión de secuencias asociadas con una vía de señalización bioquímica también se puede determinar mediante el examen de los productos génicos correspondientes. La determinación del nivel de proteína implica típicamente a) poner en contacto la proteína contenida en una muestra biológica con un agente que se une específicamente a una proteína asociada con una vía bioquímica de señalización; y (b) identificar
 40 cualquier complejo de agente:proteína así formado. En un aspecto de esta realización, el agente que une específicamente una proteína asociada con una vía de señalización bioquímica es un anticuerpo, preferiblemente un anticuerpo monoclonal.

La reacción se lleva a cabo poniendo en contacto el agente con una muestra de las proteínas asociadas con una vía
 45 de señalización bioquímica derivada de las muestras de ensayo en condiciones que permitan que se forme un complejo entre el agente y las proteínas asociadas con una vía de señalización bioquímica. La formación del complejo se puede detectar directa o indirectamente de acuerdo con procedimientos estándares en la técnica. En el método de detección directo, los agentes se suministran con un marcador detectable y agentes que no han reaccionado pueden separarse del complejo; la cantidad de marcador restante indicando con ello la cantidad de complejo formado. Para un método de este tipo, es preferible seleccionar los marcadores que permanecer unidos a los agentes, incluso durante condiciones de lavado rigurosas. Es preferible que el marcador no interfiera con la
 50 reacción de unión. En la alternativa, un procedimiento de detección indirecta puede utilizar un agente que contiene un marcador introducido química o enzimáticamente. Un marcador deseable no interfiere generalmente con la unión o la estabilidad del complejo agente:polipéptido resultante. Sin embargo, el marcador se diseña típicamente para que sea accesible a un anticuerpo para una unión eficaz y, por lo tanto, para generar una señal detectable.

Una amplia diversidad de marcadores adecuados para la detección de los niveles de proteína son conocidos en la técnica. Ejemplos no limitantes incluyen radioisótopos, enzimas, metales coloidales, compuestos fluorescentes, compuestos bioluminiscentes y compuestos quimioluminiscentes.

5 La cantidad de complejos agente:polipéptido formados durante la reacción de unión se puede cuantificar mediante ensayos cuantitativos estándares. Tal como se ilustra anteriormente, la formación de complejo agente:polipéptido se puede medir directamente por la cantidad de marcador permaneció en el sitio de unión. En una alternativa, la proteína asociada con una vía de señalización bioquímica se testa en cuanto a su capacidad para competir con un análogo marcado para sitios de unión en el agente específico. En este ensayo competitivo, la cantidad de marcador capturado es inversamente proporcional a la cantidad de secuencias de proteínas asociadas con una vía de
10 señalización bioquímica presente en una muestra de ensayo.

En la técnica está disponible un cierto número de técnicas para el análisis de proteínas sobre la base de los principios generales esbozados anteriormente. Estos incluyen, pero no se limitan a radioinmunoensayos, ELISA (ensayos inmunoradiométricos ligados a enzima), inmunoensayos "sándwich", ensayos inmunoradiométricos, inmunoensayos in situ (utilizando, p. ej., oro coloidal, enzimas o radioisótopos etiquetas), análisis de transferencia
15 Western, ensayos de inmunoprecipitación, ensayos de inmunofluorescencia y SDS-PAGE.

Anticuerpos que reconocen específicamente o se unen a proteínas relacionadas con una vía de señalización bioquímica son preferibles para la realización de los análisis de proteínas mencionados anteriormente. En los casos en los que se desee, se pueden utilizar los anticuerpos que reconocen un tipo específico de modificaciones post-traduccionales (p. ej., modificaciones inducibles en la vía bioquímica de señalización). Modificaciones post-traduccionales incluyen, pero no se limitan a glicosilación, lipidación, acetilación y fosforilación. Estos anticuerpos pueden adquirirse de proveedores comerciales. Por ejemplo, los anticuerpos anti-fosfotirosina que reconocen específicamente las proteínas tirosina-fosforiladas están disponibles de un número de proveedores, incluyendo Invitrogen y Perkin Elmer. Anticuerpos anti-fosfotirosina son particularmente útiles en la detección de proteínas que son fosforiladas diferencialmente en sus residuos de tirosina en respuesta a un estrés ER. Tales proteínas incluyen, pero no se limitan al factor 2 alfa de iniciación de la traducción eucariótica (eIF-2 α). Alternativamente, estos anticuerpos pueden generarse utilizando tecnologías de anticuerpos policlonales o monoclonales convencionales mediante la inmunización de un animal huésped o una célula productora de anticuerpos con una proteína diana que exhibe la modificación post-traduccionales deseada.
20

En la práctica del método en cuestión, puede ser deseable discernir el patrón de expresión de una proteína asociada con una vía de señalización bioquímica en diferentes tejidos corporales, en diferentes tipos de células y/o en diferentes estructuras subcelulares. Estos estudios se pueden realizar con el uso de anticuerpos específicos para el tejido, específicos para la célula o específicos para la estructura subcelular, capaces de unirse a marcadores de proteínas que se expresan preferentemente en determinados tejidos, tipos de célula o estructuras subcelulares.
30

Una expresión alterada de un gen asociado con una vía de señalización bioquímica también se puede determinar examinando un cambio en la actividad del producto génico con relación a una célula de control. El ensayo para un cambio inducido por agente en la actividad de una proteína asociada con una vía de señalización bioquímica dependerá de la actividad biológica y/o la vía de transducción de señales que está bajo investigación. Por ejemplo, en los casos en los que la proteína es una quinasa, un cambio en su capacidad para fosforilar el o los sustratos aguas abajo puede ser determinado por una diversidad de ensayos conocidos en la técnica. Ensayos representativos incluyen, pero no se limitan a inmunotransferencia e inmunoprecipitación con anticuerpos tales como anticuerpos anti-fosfotirosina que reconocen las proteínas fosforiladas. Además, la actividad quinasa se puede detectar por ensayos quimioluminiscentes de alto rendimiento tales como los ensayos AlphaScreen™ (disponible de Perkin Elmer) y eTag™ (Chan-Hui, et al. (2003) Clinical Immunology 111:162-174).
35

En los casos en los que la proteína asociada con una vía de señalización bioquímica es parte de una cascada de señalización que conduce a una fluctuación de la condición de pH intracelular, moléculas sensibles al pH tales como colorantes de pH fluorescentes se pueden utilizar como moléculas informadoras. En otro ejemplo en el que la proteína asociada a una vía de señalización bioquímica es un canal de iones, se pueden vigilar las fluctuaciones en el potencial de membrana y/o en la concentración intracelular de iones. Un cierto número de kits comerciales y dispositivos de alto rendimiento son particularmente adecuados para un rastreo rápido y robusto para moduladores de canales de iones. Instrumentos representativos incluyen FLIPRTM (Molecular Devices, Inc.) y VIPR (Aurora Biosciences). Estos instrumentos son capaces de detectar reacciones en pocillos de más de 1000 muestras de una microplaca de forma simultánea, y proporcionan una medición en tiempo real y datos funcionales en el espacio de un segundo o incluso un milisegundo.
45
50

- En la práctica de cualquiera de los métodos descritos en esta memoria, un vector adecuado se puede introducir en una célula o un embrión a través de uno o más métodos conocidos en la técnica, incluyendo, sin limitación, microinyección, electroporación, sonoporación, biolística, transfección mediada por fosfato de calcio, transfección catiónica, transfección de liposomas, transfección de dendrímeros, transfección de choque térmico, transfección por nucleofección, magnetofección, lipofección, impalefacción, transfección óptica, captación de propiedad reforzada por el agente de ácidos nucleicos y el suministro a través de liposomas, inmunoliposomas, virosomas o viriones artificiales. En algunos métodos, el vector se introduce en un embrión mediante microinyección. El vector o los vectores se pueden microinyectar en el núcleo o en el citoplasma del embrión. En algunos métodos, el vector o los vectores se pueden introducir en una célula por nucleofección.
- 5 El polinucleótido diana de un complejo de CRISPR puede ser cualquier polinucleótido endógeno o exógeno a la célula eucariota. Por ejemplo, el polinucleótido diana puede ser un polinucleótido que reside en el núcleo de la célula eucariota. El polinucleótido diana puede ser una secuencia codificadora un producto génico (p. ej., una proteína) o una secuencia no codificadora (p. ej., un polinucleótido regulador o un ADN basura).
- 10 Ejemplos de polinucleótidos diana incluyen una secuencia asociada con una vía de señalización bioquímica, p. ej., un gen o polinucleótido asociado a la vía de señalización bioquímica. Ejemplos de polinucleótidos diana incluyen un gen o polinucleótido asociado a la enfermedad. Un gen o polinucleótido "asociado a la enfermedad" se refiere a cualquier gen o polinucleótido que está proporcionando productos de transcripción o traducción en un nivel anormal o en una forma anormal en células derivadas de tejidos afectados por la enfermedad en comparación con los tejidos o células de un control no enfermo . Puede ser un gen que se expresa a un nivel anormalmente alto; puede ser un gen que se expresa a un nivel anormalmente bajo, en donde la expresión alterada se correlaciona con la aparición y/o el progreso de la enfermedad. Un gen asociado a la enfermedad también se refiere a un gen que posee mutación o mutaciones o una variación genética que es directamente responsable o está en desequilibrio de enlace con un gen o genes que es responsable de la etiología de una enfermedad. Los productos transcritos o traducidos pueden ser conocidos o desconocidos, y pueden estar en un nivel normal o anormal.
- 15 20 El polinucleótido diana de un complejo de CRISPR puede ser cualquier polinucleótido endógeno o exógeno a la célula eucariota. Por ejemplo, el polinucleótido diana puede ser un polinucleótido que reside en el núcleo de la célula eucariota. El polinucleótido diana puede ser una secuencia codificadora de un producto génico (p. ej., una proteína) o una secuencia no codificadora (p. ej., un polinucleótido regulador o un ADN basura). Sin desear estar limitados por la teoría, se cree que la secuencia diana debería estar asociada con un PAM (motivo adyacente al protoespaciador); es decir, una secuencia corta reconocida por el complejo de CRISPR. Los requisitos de secuencia y longitud precisas para el PAM difieren dependiendo de la enzima utilizada en el CRISPR, pero PAMs son típicamente de 2-5 secuencias de bases adyacentes al protoespaciador (es decir, la secuencia diana). Ejemplos de secuencias de PAM se dan en la sección de ejemplos que figura más adelante, y la persona experta será capaz de identificar secuencias adicionales de PAM para su uso con una enzima de CRISPR dada.
- 25 30 Ejemplos de polinucleótidos diana incluyen una secuencia asociada con una vía de señalización bioquímica, p. ej., un gen o polinucleótido asociado a la vía de señalización bioquímica. Ejemplos de polinucleótidos diana incluyen un gen o polinucleótido asociado a la enfermedad. Un gen o polinucleótido "asociado a la enfermedad" se refiere a cualquier gen o polinucleótido que está proporcionando productos de la transcripción o traducción en un nivel anormal o en una forma anormal a células derivadas de tejidos afectados por la enfermedad en comparación con los tejidos o células de un control no enfermo. Puede ser un gen que se expresa a un nivel anormalmente alto; puede ser un gen que se expresa a un nivel anormalmente bajo, en donde la expresión alterada se correlaciona con la aparición y/o progreso de la enfermedad. Un gen asociado a la enfermedad también se refiere a un gen que posee la o las mutaciones o variación genética que es directamente responsable o está en desequilibrio de enlace con un gen o genes que son responsables de la etiología de una enfermedad. Los productos transcritos o traducidos pueden ser conocidos o desconocidos, y pueden estar en un nivel normal o anormal.
- 35 40 45 El polinucleótido diana de un complejo de CRISPR puede incluir un cierto número de genes asociados a la enfermedad y polinucleótidos, así como genes y polinucleótidos asociados a la vía de señalización bioquímica tal como se enumeran en las solicitudes de patente de EE.UU. provisionales 61/736.527 y 61/748.427, ambas tituladas SISTEMAS, MÉTODOS Y COMPOSICIONES PARA LA MANIPULACIÓN DE SECUENCIAS presentadas el 12 de diciembre de 2012 y el 2 de enero de 2013 , respectivamente.
- 50 Ejemplos de polinucleótidos diana incluyen una secuencia asociada con una vía de señalización bioquímica, p. ej., un gen o polinucleótido asociado a la vía de señalización bioquímica. Ejemplos de polinucleótidos diana incluyen un gen o polinucleótido asociado a la enfermedad. Un "gen o polinucleótido asociado a la enfermedad" se refiere a cualquier gen o polinucleótido que está proporcionando productos de transcripción o traducción en un nivel anormal o en una forma anormal en células derivadas de tejidos afectados por la enfermedad en comparación con tejidos o
- 55

- 5 células de un control no enfermo. Puede ser un gen que se expresa a un nivel anormalmente alto; puede ser un gen que se expresa a un nivel anormalmente bajo, en donde la expresión alterada se correlaciona con la aparición y/o el progreso de la enfermedad. Un gen asociado a la enfermedad también se refiere a un gen que posee mutación o mutaciones o una variación genética que es directamente responsable o está en desequilibrio de enlace con un gen o genes que son responsables de la etiología de una enfermedad. Los productos transcritos o traducidos pueden ser conocidos o desconocidos, y pueden estar en un nivel normal o anormal.

Ejemplos de genes y polinucleótidos asociados a una enfermedad están disponibles en McKusick-Nathans Institute for Genetic Medicine, Universidad Johns Hopkins (Baltimore, Md.) y el Centro Nacional de Información sobre Biotecnología, Biblioteca Nacional de Medicina (Bethesda, Md.), disponible en la World Wide Web.

- 10 Ejemplos de genes y polinucleótidos asociados con la enfermedad se enumeran en las Tablas A y B. Información específica de la enfermedad está disponible en McKusick-Nathans Institute for Genetic Medicine, Universidad Johns Hopkins (Baltimore, Md.) y el Centro Nacional de Información sobre Biotecnología, Biblioteca Nacional de Medicina (Bethesda, Md.), disponible en la World Wide Web. Ejemplos de señalización de genes y polinucleótidos asociados a la vía de señalización bioquímica se enumeran en la Tabla C.
- 15 Mutaciones en estos genes y estas vías pueden resultar en la producción de proteínas inadecuadas o proteínas en cantidades inadecuadas que afectan a la función. Genes, proteínas y vías de este tipo pueden ser el polinucleótido diana de un complejo de CRISPR.

Tabla A

ENFERMEDAD/TRASTORNOS	GENE(S)
Neoplasia	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4;
	Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF;
	HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alfa; PPAR
	gamma; WT1 (Tumor de Wilms); miembros de la Familia de receptors FGF
	(5 miembros: 1, 2, 3, 4, 5); CDKN2a; APC; RB
	(retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR
	(Receptor de Andr6geno); TSG101; IGF; Receptor de IGF; Igf1 (4
	variantes); Igf2 (3 variantes); Receptor de Igf1; Receptor de Igf2;
	Bax; Bcl2; familia de caspasas (9 miembros:
	1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
Macular relacionada con la edad	Abcr; Ccl2; Cc2; cp (ceruloplasmina); Timp3; catepsinaD;
Degeneraci6n	Vldlr; Ccr2
Esquizofrenia	Neuregulina1 (Nrg1); Erb4 (receptor para Neuregulina);

ES 2 576 128 T3

ENFERMEDAD/TRASTORNOS	GENE(S)
	Complexina1 (Cplx1); Tph1 Triptofanohidroxilasa; Tph2
	Triptófano hidroxilasa 2; Neurexina 1; GSK3; GSK3a;
	GSK3b
Trastornos	5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA;
	DTNBP1; Dao (Dao1)
Repetición de Trinucleótidos	HTT (Enfermedad de Huntington); SBMA/SMAX1/AR (Enfermedad de Kennedy);
Trastornos	FXN/X25 (Ataxia de Friedrich); ATX3 (Enfermedad de Machado-
	Joseph); ATXN1 and ATXN2 (ataxias espinocerebelares);
	DMPK (distrofia miotónica); Atrofin-1 y Atn1
	(Enfermedad DRPLA); CBP (Creb-BP – inestabilidad global); VLDLR
	(Alzheimer); Atxn7; Atxn10
Síndrome X Frágil	FMR2; FXR1; FXR2; mGLUR5
Relacionado con Secretasa	APH-1 (alfa y beta); Presenilina (Psen1); nicastrina
Trastornos	(Ncstn); PEN-2
Otros	Nos1; Parp1; Nat1; Nat2
Trastornos relacionados con Priones	Prp
ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a;
	VEGF-b; VEGF-c)
Adicción a fármacos	Prkce (alcohol); Drd2; Drd4; ABAT (alcohol); GRIA2;
	Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1 (alcohol)
Autismo	Mecp2; BZRAP1; MDGA2; Sema5A; Neurexina 1; Frágil X
	(FMR2 (AFF2); FXR1; FXR2; Mglur5)

ES 2 576 128 T3

ENFERMEDAD/TRASTORNOS	GENE(S)
Enfermedad de Alzheimer	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterina; PS1;
	SORL1; CR1; Vldlr; Uba1; Uba3; CHIP28 (Aqp1,
	Aquaporina 1); Uchl1; Uchl3; APP
Inflamación	IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a(CTLA8); IL-
	17b; IL-17c; IL,-17d; IL-17f); IL-23; Cx3cr1; ptpn22; TNFa;
	NOD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b);
	CTLA4; Cx3c11
Enfermedad de Parkinson	x-Sinucleína; DJ-1; LRRK2; Parkin; PINK1

Tabla B:

Enfermedades y trastornos de la sangre y coagulación	Anemia (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); síndrome de linfocito desnudo (TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5); Trastornos de hemorragia (TBXA2R, P2RX1, P2X1); Factor H y tipo factor H 1 (HF 1, CFH, HUS); Factor V y factor VIII (MCFD2); Deficiencia de factor VII (F7); Deficiencia de factor X (F10); Deficiencia de factor XI (F11); Deficiencia de factor XII (F 12, HAF); Deficiencia de factor XIIIa (F13A1, F13A); Deficiencia de factor XIIIb (F13B); anemia de Fanconi (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCI, PHF9, FANCL, FANCM, KIAA1596); Trastornos de linfocitosis hemofagocítica (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemofilia A (F8, F8C, HEMA); Hemofilia B (F9, HEMB); Trastornos hemorrágicos (PI, ATT, F5); Deficiencias y trastornos de leucocitos (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); anemia de células de Sickle (HBB); Talasemia (HBA2, HBB, HBD, LCRB, HBA1).
Enfermedades y trastornos de disregulación celular y oncología	Linfoma no-Hodgkin B-cel1 (BCL7A, BCL7); Leucemia (TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL 11, ARLTS 1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN).
Enfermedades y trastornos relacionados	AIDS (KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1); Síndrome linfoproliferativo autoinmune (TNFRSF6, APT1, FAS, CD95, ALPS1A); Inmunodeficiencia

<p>con la inflamación y el sistema inmune</p>	<p>combinada, (IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228), Susceptibilidad o infección por VIH (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCKR5 (CCR5)); Inmunodeficiencias (CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TAC1); Inflamación (IL-10, IL-1 (IL-1a, IL-1b), IL-13, IL-17 (IL-17a (CTLA8), IL-17b, IL-17c, IL-17d, IL-17f), IL-23, Cx3cl1, ptpn22, TNFa, NOD2/CARD15 for IBD, IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3cl1); Inmunodeficiencias combinadas graves (SCIDs)(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).</p>
<p>Enfermedades y trastornos metabólicos, del hígado, riñon y proteínas</p>	<p>Neuropatía amiloide (TTR, PALB); Amiloidosis (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); Cirrosis (KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); Fibrosis quística (CFTR, ABCC7, CF, MRP7); Enfermedades de almacenamiento de glucógeno (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Adenoma hepático, 142330 (TCF1, HNF1A, MODY3), Insuficiencia hepática, brote temprano y trastorno neurológico (SCOD1, SCO1), Deficiencia de lipasa hepática (LIPC), Hepatoblastoma, cáncer y carcinomas (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; Enfermedad de riñón quístico medular (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); fenilcetonuria (PAH, PKU1, QDPR, DHPR, PTS); Riñón poliquístico y enfermedad hepática (FCYT, PKHD 1, ARPKD, PKD 1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63).</p>
<p>Enfermedades y trastornos Musculares/ del Esqueleto</p>	<p>Distrofia muscular de Becker (DMD, BMD, MYF6), Distrofia muscular de Duchenne (DMD, BMD); Distrofia muscular de Emery-Dreifuss (LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); Distrofia muscular facioescapulohumeral (FSHMD1A, FSHD1A); Distrofia muscular (FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT 1, CAV3, LGMD1C, SEPN 1, SELN, RSMD1, PLEC1, PLTN, EBS1); Osteopetrosis (LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); Atrofia muscular (VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1).</p>
<p>Enfermedades y trastornos neurológicos y neuronales</p>	<p>ALS (SOD1, ALS2, STEX, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c); enfermedad de Alzheimer (APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLAU, URK, ACE, DCP 1, ACE 1, MPO, PACIP 1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); Autismo (MECP2, BZRAP1, MDGA2, Sema5A, Neurexina 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); Síndrome X frágil (FMR2, FXR1, FXR2, mGLUR55); Enfermedad y trastornos tipo enfermedad de Huntington (HD, IT 15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); Enfermedad de Parkinson (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK5, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); Síndrome de Rett (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Sinucleína, DJ-1); Esquizofrenia (Neuregulina1 (Nrg1), Erb4 (receptor para Neuregulina), Complexina1 (Cplx1), Tph1 Triptófano hidroxilasa, Tph2, Triptófano hidroxilasa 2, Neurexina 1, GSK3, GSK3a, GSK3b, 5-HTT (Slc6a4), COMT, DRD (Drd1a), SLC6A3, DAOA, DTNBP1, Dao (Dao1)); Trastornos relacionados con secretasa (APH-1 (alfa y beta), Presenilina (Psen1), nicastrina, (Ncstn), PEN-2, Nos1,</p>

	Parp1, Nat1, Nat2); Trastornos de la repetición de trinucleótidos (HTT (Enfermedad de Huntington), SBMA/SMAX1/AR (Enfermedad de Kennedy), FXN/X25 (Ataxia de Friedrich), ATX3 (Enfermedad de Machado-Joseph), ATXN1 y ATXN2 (ataxias espinocerebelares), DMPK (distrofia miotónica), Atrofia-1 y Atn1 (DRPLA Dx), CBP (Creb-BP – inestabilidad global), VLDLR (Alzheimer), Atn7, Atn10).
Enfermedades y trastornos oculares	Degeneración macular relacionada con la edad (Abcr, Cc12, Cc2, cp (ceruloplasmina), Timp3, catepsinaD, Vldlr, Ccr2); Cataratas (CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); Opacidad y distrofia de la córnea (APOA1, TGFB1, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD); Córnea plana congénita (KERA, CNA2); Glaucoma (MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A); Amaurosis congénita del hígado (CRB1, RP 12, CRX, CORD2, CRD, RPGRIP 1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3); Distrofia macular (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2).

Tabla C:

FUNCIÓN CELULAR	GENES
Señalización PI3K/AKT	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2;
	PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1;
	AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2;
	PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2;
	ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3;
	PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7;
	YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A;
	CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1;
	CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1;
	PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2;
	TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK;
	HSP90AA1; RPS6KB1
Señalización ERK/MAPK	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2;

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
	EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6;
	MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1;
	PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A;
	PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN;
	EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC;
	CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ;
	PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1;
	PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1;
	CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
Señalización del Receptor de Glucocorticoides	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1;
	MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I;
	PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2;
	MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1;
	MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13;
	RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1;
	PIK3C2A; CDKN1C; TRAF2; SERPINA1; NCOA3;
	MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP;
	CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2;
	PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1;
	ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1;
	STAT1; IL6; HSP90AA1

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
Señalización de Guía Axonal	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12;
	IGF 1; RAC 1; RAP1A; EIF4E; PRKCZ; NRP 1; NTRK2;
	ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2;
	PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2;
	CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11;
	PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1;
	FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1;
	GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3;
	CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B;
	AKT3; PRKCA
Señalización del Receptor Efrina	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1;
	PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2;
	MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2;
	DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14;
	CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1;
	KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2;
	PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10;
	MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2;
	EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4;
	AKT3; SGK

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
Señalización del Citoesqueleto de Actina	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1;
	PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6;
	ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8;
	PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8;
	F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD;
	PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7;
	PPP1CC; PXN; VIL2; RAF 1; GSN; DYRK1A; ITGB 1;
	MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3;
	ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL;
	BRAF; VAV3; SGK
Señalización de la Enfermedad de Huntington	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2;
	MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2;
	PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST;
	GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1;
	GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2;
	HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A;
	HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1;
	PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX;
	ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3
Señalización de la Apoptosis	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1;
	BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB;
	CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8;

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
	BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF;
	RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2;
	CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2;
	BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK;
	CASP3; BIRC3; PARP1
Señalización del Receptor de Células B	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11;
	AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1;
	MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9;
	EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB;
	MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1;
	NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN;
	GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Señalización de la Extravasación de Leucocitos	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA;
	RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11;
	MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12;
	PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB;
	MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK;
	MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2;
	CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK;
	CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
Señalización de Integrina	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A;
	TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2;
	CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8;
	CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP;
	RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1;
	TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2;
	CRKL; BRAF; GSK3B; AKT3
Señalización de la Respuesta de Fase Aguda	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11;
	AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14;
	PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS;
	MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1;
	TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1;
	IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1;
	CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN;
	AKT3; IL 1R1; IL6
Señalización de PTEN	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11;
	MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA;
	CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1;
	MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR;
	RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2;
	AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1;

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
	NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2;
	GSK3B; AKT3; FOXO1; CASP3; RPS6KB1
Señalización de p53	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A;
	BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2;
	PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1;
	PMAIP 1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9;
	CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A;
	HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1;
	SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN;
	SNAI2; GSK3B; BAX; AKT3
Señalización del Receptor de Aril Hidrocarbano	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1;
	NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1;
	SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1;
	MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1;
	SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF;
	CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1;
	CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1;
	HSP90AA1
Señalización del Metabolismo Xenobiótico	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1;
	NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A;
	PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1;
	ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD;

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
	GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL;
	NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1;
	CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1;
	NFKB 1; KEAP 1; PRKCA; EIF2AK3; IL6; CYP1B1;
	HSP90AA1
Señalización de SAPK/JNK	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1;
	GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA;
	FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1;
	GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A;
	TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2;
	PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1;
	CRKL; BRAF; SGK
Señalización de PPAR/RXR	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN;
	RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2;
	ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8;
	IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A;
	NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7;
	CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1;
	TGFBR1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1;
	ADIPOQ
Señalización de NF-KB	IRAK1; EIF2AK2; EP300; INS; MYD88; PRK CZ; TRAF6;

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
	TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2;
	KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF;
	INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1;
	PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10;
	GSK3B; AKT3; TNFAIP3; IL 1R1
Señalización de Neuregulina	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1;
	MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI;
	CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS;
	PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2;
	ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3;
	EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL;
	AKT3; PRKCA; HSP90AA1; RPS6KB1
Señalización de Wnt y Beta catenina	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO;
	AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A;
	WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK;
	LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1;
	PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1; CCND1;
	GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B;
	AKT3; SOX2
Señalización del Receptor de Insulina	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1;
	PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3;

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
	MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1;
	SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN;
	MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1;
	GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK;
	RPS6KB1
Señalización de IL-6	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3;
	MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1;
	MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG;
	RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3;
	MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6
Colestasis Hepática	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA;
	RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8;
	PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1;
	TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8;
	CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4;
	JUN; IL1R1; PRKCA; IL6
Señalización de IGF-1	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2;
	PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8;
	IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A;
	YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1;
	PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3;

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
	FOXO1; SRF; CTGF; RPS6KB1
Respuesta de Estrés Oxidativo mediada por NRF2	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1;
	NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL;
	NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP;
	MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1;
	GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
Fibrosis Hepática/Activación de Células Estelares	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF;
	SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9;
	IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8;
	PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX;
	IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
Señalización de PPAR	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB;
	NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3;
	NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2;
	PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG;
	RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA;
	MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Señalización de Fc Epsilon RI	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11;
	AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD;
	MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN;

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
	MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3;
	VAV3; PRKCA
Señalización del Receptor Acoplado a Proteína G	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB;
	PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB;
	PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1;
	IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK;
	PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3;
	PRKCA
Metabolismo de Inositol Fosfato	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6;
	MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2;
	PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5KIA; PIK3R1;
	MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
Señalización de PDGF	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB;
	PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC;
	PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2;
	PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC;
	JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
Señalización de VEGF	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF;
	AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3;
	BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN;
	RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN;

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
	VEGFA; AKT3; FOXO1; PRKCA
Señalización de Células Asesinas Naturales	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11;
	KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB;
	PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6;
	PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1;
	PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
Ciclo Celular: G1/S	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC;
Regulación del punto de verificación	ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11;
	HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1;
	E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1;
	GSK3B; RBL1; HDAC6
Señalización del Receptor de Células T	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS;
	NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN;
	MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10;
	JUN; VAV3
Señalización del Receptor de Muerte	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD;
	FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8;
	DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB;
	CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3;
	BIRC3
Señalización de FGF	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11;

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
	AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8;
	MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1;
	AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4;
	AKT3; PRKCA; HGF
Señalización de GM-CSF	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A;
	STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3;
	ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3;
	STAT1
Señalización de la Esclerosis Lateral Amiotrófica	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2;
	PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1;
	PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1;
	APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
Señalización de JAK/Stat	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B;
	PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A;
	PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3;
	STAT1
Metabolismo de Nicotinato y Nicotinamida	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1;
	PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1;
	PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2;
	MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
Señalización de Quimioquinas	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ;
	CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13;
	RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1;
	MAP2K2; MAP2K1; JUN; CCL2; PRKCA
Señalización de IL-2	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS;
	STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2;
	JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
Depresión Sináptica a Largo Plazo	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS;
	PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3;
	KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA;
	YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Señalización del Receptor de Estrógenos	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2;
	SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1;
	HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP;
	MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Vía de la Ubiquitinación de Proteínas	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4;
	CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7;
	USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8;
	USP1; VHL; HSP90AA1; BIRC3
Señalización de IL-10	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2;
	MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF;

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
	IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1;
	JUN; IL1R1; IL6
Activación de VDR/RXR	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1;
	NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD;
	RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1;
	LRP5; CEBPB; FOXO1; PRKCA
Señalización de TGF-beta	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1;
	FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2;
	SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2;
	MAP2K1; TGFBR1; SMAD4; JUN; SMAD5
Señalización del Receptor de tipo Toll	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13;
	RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK;
	NFKB1; TLR2; JUN
Señalización de p38 MAPK	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS;
	CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2;
	MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL 1R1;
	SRF; STAT1
Señalización de Neurotrofina/TRK	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS;
	PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A;
	RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1;
	CDC42; JUN; ATF4

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
Activación de FXR/RXR	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8;
	APOB; MAPK10; PPARG; MTTP; MAPK9; PPARGC1A;
	TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
Potenciación Sináptica a Largo Plazo	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1;
	PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS;
	PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1;
	ATF4; PRKCA
Señalización de Calcio	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1;
	CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11;
	HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4;
	HDAC6
Señalización de EGF	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1;
	STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Señalización de Hipoxia en el	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT;
Sistema Cardiovascular	HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM;
	VEGFA; JUN; ATF4; VHL; HSP90AA1
Inhibición Mediada por LPS/IL-1	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1;
de la Función RXR	MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2;
	TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
Activación de LXR/RXR	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA;
	NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1;

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
	SREBF1; IL1R1; CCL2; IL6; MMP9
Procesamiento Amiloide	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2;
	CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1;
	PSEN1; CSNK1A1; GSK3B; AKT3; APP
Señalización de IL-4	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1;
	PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1;
	FRAP1; AKT3; RPS6KB1
Ciclo Celular: G2/M Regulación del Punto	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC;
de Verificación de Daño a ADN	CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A;
	PRKDC; ATM; SFN; CDKN2A
Señalización de Óxido Nítrico en el	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3;
Sistema Cardiovascular	CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1;
	VEGFA; AKT3; HSP90AA1
Metabolismo de Purina	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4;
	PKM2; ENTPD 1; RAD51; RRM2B; TJP2; RAD51C;
	NT5E; POLD1; NME1
Señalización mediada por cAMP	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3;
	SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Disfunción Mitocondrial	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9;
	PARK7; PSEN1; PARK2; APP; CASP3
Señalización de Notch	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2;
	PSEN1; NOTCH3; NOTCH1; DLL4

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
Vía del Estrés del Retículo Endoplásmico	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4;
	EIF2AK3; CASP3
Metabolismo de Pirimidina	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B;
	NT5E; POLD1; NME1
Señalización de Parkinson	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7;
	PARK2; CASP3
Señalización Cardíaca y Beta Adrenérgica	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC;
	PPP2R5C
Glicolisis/Gluconeogénesis	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Señalización de Interferón	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Señalización Sonic Hedgehog	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B
Metabolismo de Glicerofosfolípidos	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Degradación de Fosfolípidos	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Metabolismo de triptófano	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Degradación de Lisina	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Vía de la Reparación de la Escisión de Nucleótidos	ERCC5; ERCC4; XPA; XPC; ERCC1
Metabolismo de Almidón y Sacarosa	UCHL1; HK2; GCK; GPI; HK1
Metabolismo de Aminoazúcares	NQO1; HK2; GCK; HK1
Metabolismo del Ácido Araquidónico	PRDX6; GRN; YWHAZ; CYP1B1

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
Señalización del Ritmo Circadiano	CSNK1E; CREB1; ATF4; NR1D1
Sistema de Coagulación	BDKRB1; F2R; SERPINE1; F3
Señalización del Receptor de Dopamina	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Metabolismo de Glutacion	IDH2; GSTP1; ANPEP; IDH1
Metabolismo de Glicerolípidos	ALDH1A1; GPAM; SPHK1; SPHK2
Metabolismo del Ácido Linoleico	PRDX6; GRN; YWHAZ; CYP1B1
Metabolismo de Metionina	DNMT1; DNMT3B; AHCY; DNMT3A
Metabolismo de Piruvato	GLO1; ALDH1A1; PKM2; LDHA
Metabolismo de Arginina y Prolina	ALDH1A1; NOS3; NOS2A
Señalización de Eicosanoides	PRDX6; GRN; YWHAZ
Metabolismo de Fructosa y Manosa	HK2; GCK; HK1
Metabolismo de Galactosa	HK2; GCK; HK1
Biosíntesis de Estilbeno, Coumarina y	PRDX6; PRDX1; TYR
Lignina	
Vía de la Presentación de Antígenos	CALR; B2M
Biosíntesis de Esteroides	NQO1; DHCR7
Metabolismo de Butanoato	ALDH1A1; NLGN1

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
Ciclo del Citrato	IDH2; IDH1
Metabolismo de Ácidos Grasos	ALDH1A1; CYP1B1
Metabolismo de Glicerofosfolípidos	PRDX6; CHKA
Metabolismo de Histidina	PRMT5; ALDH1A1
Metabolismo de Inositol	ERO1L; APEX1
Metabolismo de Xenobióticos	GSTP1; CYP1B1
por el Citocromo p450	
Metabolismo del Metano	PRDX6; PRDX1
Metabolismo de Fenilalanina	PRDX6; PRDX1
Metabolismo de Propanoato	ALDH1A1; LDHA
Metabolismo de Selenoaminoácido	PRMT5; AHCY
Metabolismo de Esfingolípidos	SPHK1; SPHK2
Metabolismo de Aminofosfonato	PRMT5
Metabolismo de Andrógenos y Estrógenos	PRMT5
Metabolismo de Ascorbato y Aldarato	ALDH1A1
Biosíntesis de Ácidos Biliares	ALDH1A1
Metabolismo de Cisteína	LDHA

ES 2 576 128 T3

FUNCIÓN CELULAR	GENES
Biosíntesis de Ácidos Grasos	FASN
Señalización del Receptor de Glutamato	GNB2L1
Respuesta al Estrés Oxidativo medida por NRF2	PRDX1
Vía del Fosfato de Pentosa	GPI
Interconversiones de Pentosa y Glucuronato	UCHL1
Metabolism de Retinol	ALDH1A1
Metabolismo de Riboflavina	TYR
Metabolismo de Tirosina	PRMT5, TYR
Biosíntesis de Ubiquinona	PRMT5
Degradación de Valina, Leucina e	ALDH1A1
Isoleucina	
Metabolismo de Glicina, Serina y	CHKA
Treonina	
Degradación de Lisina	ALDH1A1
Dolor/Sabor	TRPM5; TRPA1
Dolor	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2;
	Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca;
	Prkacb; Prkar1a; Prkar2a

FUNCIÓN CELULAR	GENES
Función Mitocondrial	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Neurología del desarrollo	BMP-4; Cordina (Chrd); Nogina (Nog); WNT (Wnt2;
	Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b;
	Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenina;
	Dkk-1; Proteínas relacionadas con Frizzled; Otx-2; Gbx2; FGF-8;
	Reelina; Dab1; unc-86 (Pou4f1 o Brn3a); Numb; Reln

Realizaciones de la invención también se refieren a métodos y composiciones relacionados con la inactivación de genes, la amplificación de genes y la reparación de mutaciones particulares asociadas con la inestabilidad en la repetición de ADN y trastornos neurológicos (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Segunda Edición, Academic Press 13 de oct. de 2011 - Medical). Se ha encontrado que aspectos específicos de secuencias repetidas en tándem son los responsables de más de una veintena de enfermedades humanas (New insights into repeat instability: role of RNA•DNA hybrids. Mclvor EI, Polak U, Napierala M. RNA Biol. 2010 Sep.-Oct.;7(5):551-8). El sistema de CRISPR-Cas puede ser aprovechado para corregir estos defectos de la inestabilidad genómica.

Un aspecto adicional de la invención se refiere a la utilización del sistema de CRISPR-Cas para la corrección de defectos en los genes EMP2A y EMP2B que han sido identificados que están asociados con la enfermedad de Lafora. La enfermedad de Lafora es una afección autosómica recesiva que se caracteriza por la epilepsia mioclónica progresiva que puede comenzar como convulsiones epilépticas en la adolescencia. Unos pocos casos de la enfermedad pueden ser provocados por mutaciones en genes aún por identificar. La enfermedad provoca convulsiones, espasmos musculares, dificultad para caminar, demencia y, finalmente, la muerte. Actualmente no existe una terapia que haya demostrado ser eficaz contra el progreso de la enfermedad. Otras anomalías genéticas asociadas con la epilepsia también pueden ser fijadas como objetivo por el sistema de CRISPR-Cas y la genética subyacente se describe con más detalle en Genetics of Epilepsy and Genetic Epilepsies, editado por Giuliano Avanzini, Jeffrey L. Noebels, Mariani Foundation Paediatric Neurology:20; 2009).

Todavía en otro aspecto de la invención, el sistema de CRISPR-Cas puede utilizarse para corregir defectos oculares que surgen de varias mutaciones genéticas descritas adicionalmente en Genetic Diseases of the Eye, Segunda Edición, editado por Elias I. Traboulsi, Oxford University Press, 2012.

Varios aspectos adicionales de la invención se refieren a la corrección de los defectos asociados con una amplia gama de enfermedades genéticas que se describe con más detalle en la página web de National Institutes of Health, bajo la subsección Genetic Disorders (sitio web en health.nih.gov/topic/GeneticDisorders). Las enfermedades genéticas del cerebro pueden incluir, pero no se limitan a adrenoleucodistrofia, agenesia del cuerpo caloso, síndrome de Aicardi, enfermedad de Alper, enfermedad de Alzheimer, síndrome de Barth, enfermedad de Batten, CADASIL, Degeneración cerebelosa, enfermedad de Fabry, enfermedad de Gerstmann-Straussier-Scheinker, enfermedad de Huntington y otros trastornos de la repetición de tripletes, enfermedad de Leigh, síndrome de Lesch-Nyhan, enfermedad de Menkes, miopatías mitocondriales y colpocefalia NINDS. Estas enfermedades se describen con más detalle en la página web de National Institutes of Health, bajo la subsección Genetic Brain Disorders.

En algunas realizaciones, la afección puede ser la neoplasia. En algunas realizaciones, en los casos en los que la afección es la neoplasia, los genes a ser fijados como objetivo son cualquiera de los enumerados en la Tabla A (en este caso PTEN, etcétera). En algunas realizaciones, la afección puede ser la degeneración macular relacionada con la edad. En algunas realizaciones, la afección puede ser un trastorno esquizofrénico. En algunas realizaciones, la afección puede ser un trastorno de repeticiones de trinucleótidos. En algunas realizaciones, la afección puede ser el síndrome X frágil. En algunas realizaciones, la afección puede ser un trastorno relacionado con secretasa. En algunas realizaciones, la afección puede ser un trastorno relacionado con Priones. En algunas realizaciones, la

afección puede ser ALS. En algunas realizaciones, la afección puede ser una adicción a los fármacos. En algunas realizaciones, la afección puede ser el autismo. En algunas realizaciones, la afección puede ser la enfermedad de Alzheimer. En algunas realizaciones, la afección puede ser la inflamación. En algunas realizaciones, la afección puede ser la enfermedad de Parkinson.

- 5 Ejemplos de proteínas asociadas con la enfermedad de Parkinson incluyen, pero no se limitan a α -sinucleína, DJ-1, LRRK2, Pink1, Parkin, UCHL1, Sinfilina-1 y NURR1.

Ejemplos de proteínas relacionadas con la adicción pueden incluir ABAT, por ejemplo.

- 10 Ejemplos de proteínas relacionadas con la inflamación pueden incluir la proteína quimioatrayente de monocitos 1 (MCP1) codificada por el gen Ccr2, el receptor de quimioquina C-C tipo 5 (CCR5) codificado por el gen Ccr5, el receptor IIB de IgG (FCGR2b, también denominado CD32) codificado por el gen Fcgr2b o la proteína Fc epsilon R1g (FCER1g) codificada por el gen Fcer1g, por ejemplo.

- 15 Ejemplos de proteínas asociadas a enfermedades cardiovasculares pueden incluir IL1B (interleuquina 1, beta), XDH (xantina deshidrogenasa), TP53 (proteína tumoral p53), PTGIS (prostaglandina I2 (prostaciclina) sintasa), MB (mioglobina), IL4 (interleuquina 4), ANGPT1 (angiopoyetina 1), ABCG8 (casete de unión a ATP, subfamilia G (WHITE) miembro, 8), o CTSK (catepsina K), por ejemplo.

Ejemplos de proteínas asociadas a la enfermedad de Alzheimer pueden incluir la proteína receptor de lipoproteínas de muy baja densidad (VLDLR) codificada por el gen VLDLR, la enzima activante del modificador de tipo ubiquitina 1 (UBA1) codificada por el gen UBA1 o la proteína de la subunidad catalítica E1 de la enzima activante de NEDD8 (UBE1C) codificada por el gen UBA3, por ejemplo.

- 20 Ejemplos de proteínas asociadas al trastorno del espectro del autismo pueden incluir la proteína 1 asociada al receptor de benzodiazepina (periférico) (BZRAP1) codificada por el gen BZRAP1, la proteína del miembro 2 de la familia AF4/FMR2 (AFF2) codificada por el gen AFF2 (también denominada MFR2), la proteína homóloga 1 autosómica del retardo mental X frágil (FXR1) codificada por el gen FXR1 o la proteína homóloga 2 autosómica del retardo mental X frágil (FXR2) codificada por el gen FXR2, por ejemplo.

- 25 Ejemplos de proteínas asociadas con la degeneración macular puede incluir la proteína de la casete de unión a ATP, sub-familia A (ABC1) miembro 4 (ABCA4) codificada por el gen ABCR, la proteína apolipoproteína E (APOE) codificada por el gen APOE o la proteína quimioquina (motivo C-C) Ligando 2 (CCL2) codificada por el gen CCL2, por ejemplo.

- 30 Ejemplos de proteínas asociadas con la esquizofrenia pueden incluir NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, DISC1, GSK3B, y combinaciones de las mismas.

- 35 Ejemplos de proteínas implicadas en la supresión de tumores pueden incluir ATM (ataxia telangiectasia mutada), ATR (ataxia telangiectasia y relacionada con Rad3), EGFR (receptor del factor de crecimiento epidérmico), ERBB2 (homólogo 2 del oncogén de la leucemia eritroblástica v-erb-b2), ERBB3 (homólogo 3 del oncogén de la leucemia eritroblástica v-erb-b2), ERBB4 (homólogo 4 del oncogén de la leucemia eritroblástica v-erb-b2), Notch 1, Notch2, Notch 3 o Notch 4, por ejemplo.

Ejemplos de proteínas asociadas con un trastorno secretasa pueden incluir PSENEN (homólogo de potenciador de presenilina 2 (C. elegans)), CTSB (catepsina B), PSEN1 (presenilina 1), APP (precursor de la proteína amiloide beta (A4)), APH1B (homólogo B defectuoso de la faringe anterior 1 (C. elegans)), PSEN2 (presenilina 2 (enfermedad de Alzheimer 4)) o BACE1 (enzima de escisión de APP del sitio beta 1), por ejemplo.

- 40 Ejemplos de proteínas asociadas con la esclerosis lateral amiotrófica pueden incluir SOD1 (superóxido dismutasa 1), ALS2 (esclerosis lateral amiotrófica 2), FUS (fusionado en sarcoma), TARDBP (proteína de unión a ADN TAR), VAGFA (factor de crecimiento vascular endotelial A), VAGFB (factor de crecimiento vascular endotelial B) y VAGFC (factor de crecimiento vascular endotelial C), y cualquier combinación de las mismas.

- 45 Ejemplos de proteínas asociadas con las enfermedades priónicas pueden incluir SOD1 (superóxido dismutasa 1), ALS2 (esclerosis lateral amiotrófica 2), FUS (fusionado en sarcoma), TARDBP (proteína de unión a ADN TAR), VAGFA (factor de crecimiento vascular endotelial A), VAGFB (factor de crecimiento vascular endotelial B) y VAGFC (factor de crecimiento vascular endotelial C), y cualquier combinación de las mismas.

Ejemplos de proteínas relacionadas con enfermedades neurodegenerativas en los trastornos por priones pueden incluir A2M (alfa-2-macroglobulina), AATF (factor de transcripción antagonizante de la apoptosis), ACPP (fosfatasa ácida prostática), ACTA2 (actina alfa 2 aorta músculo liso), ADAM22 (dominio de metalopeptidasa ADAM), ADORA3 (receptor de adenosina A3) o ADRA1D (receptor adrenérgico alfa-1D para el adrenorreceptor alfa-1D), por ejemplo.

5 Ejemplos de proteínas asociadas con inmunodeficiencia puede incluir A2M [alfa-2- macroglobulina]; AANAT [arilalquilamina N-acetiltransferasa]; ABCA1 [casete de unión a ATP, subfamilia A (ABC1), miembro 1]; ABCA2 [casete de unión a ATP, subfamilia A (ABC1), miembro 2]; o ABCA3 [casete de unión a ATP, subfamilia A (ABC1), miembro 3]; por ejemplo.

10 Ejemplos de proteínas asociadas con Trastornos por Repetición de Trinucleótidos incluyen AR (receptor de andrógenos), FMR1 (retardo mental X frágil 1), HTT (huntingtina) o DMPK (distrofia miotónica-proteína quinasa), FXN (frataxina), ATXN2 (ataxina 2), por ejemplo.

Ejemplos de proteínas asociadas con los Trastornos de Neurotransmisión incluyen SST (somatostatina), NOS1 (óxido nítrico sintasa 1 (neuronal)), ADRA2A (receptor adrenérgico alfa -2A), ADRA2C (receptor adrenérgico alfa-2C), TACR1 (receptor de taquiquinina 1) o HTR2C (receptor de 5-hidroxitriptamina (serotonina) 2C), por ejemplo.

15 Ejemplos de secuencias asociadas al neurodesarrollo incluyen A2BP1 [proteína de unión 1 ataxina 2], AADAT [aminoadipato aminotransferasa], AANAT [arilalquilamina N-acetiltransferasa], ABAT [4-aminobutirato aminotransferasa], ABCA1 [casete de unión a ATP, subfamilia A (ABC1), miembro 1] o ABCA13 [casete de unión a ATP, subfamilia A (ABC1), miembro 13], por ejemplo.

20 Ejemplos adicionales de afecciones preferidas tratables con el presente sistema se pueden seleccionar de: Síndrome de Aicardi-Goutieres; Enfermedad de Alexander; Síndrome de Allan-Herndon-Dudley; Trastornos Relacionados con POLG; Alfa-Manosidosis (Tipos I y III); Síndrome de Alström; Síndrome de Angelman; Ataxia-Telangiectasia; Lipofuscinosis Ceroide Neuronal; Beta-Talasemia; Atrofia Óptica Bilateral y Atrofia Óptica (Infantil) Tipo 1; Retinoblastoma (bilateral); Enfermedad de Canavan; Síndrome Cerebro-oculofacioesquelético 1 [COFS1]; Xantomatosis Cerebrotendinosa; Síndrome de Cornelia de Lange; Trastornos Relacionados con MAPT; Enfermedades Priónicas Genéticas; Síndrome de Dravet; Enfermedad de Alzheimer Familiar de Aparición Temprana; Ataxia de Friedreich [FRDA]; Síndrome de Frysns; Fucosidosis; Distrofia Muscular Congénita de Fukuyama; Galactosialidosis; Enfermedad de Gaucher; Acidemias Orgánicas; Linfocitosis Hemofagocítica; Síndrome de Progeria de Hutchinson-Gilford; Mucopolidosis II; Enfermedad de Almacenamiento de Ácido Siálico Libre Infantil; Neurodegeneración Asociada a PLA2G6; Síndrome de Jervell y Lange-Nielsen; Epidermolisis Bullosa de la Unión; Enfermedad de Huntington; Enfermedad de Krabbe (Infantil); Síndrome de Leigh y NARP asociado a ADN mitocondrial; Síndrome de Lesch-Nyhan; Lisencefalia Asociada a LIS1; Síndrome de Lowe; Enfermedad de la Orina de Jarabe de Arce; Síndrome de Duplicación de MECP2; Trastornos de Transporte de Cobre Relacionados con ATP7A; Distrofia Muscular Relacionada con LAMA2; Deficiencia de Arilsulfatasa A; Mucopolisacaridosis Tipos I, II o III; Trastornos de la Biogénesis de Peroxisomas, Espectro del Síndrome de Zellweger; Neurodegeneración con Trastornos de Acumulación de Hierro en el Cerebro; Deficiencia de Esfingomielinasa Ácida; Enfermedad de Niemann-Pick Tipo C; Encefalopatía de Glicina; Trastornos Relacionados con ARX; Trastornos del Ciclo de la Urea; Osteogénesis Imperfecta Relacionada con COL1A1/2; Síndromes Mitocondriales de Deleción de ADN; Trastornos relacionados con PLP1; Síndrome de Perry; Síndrome de Phelan-McDermid; Enfermedad del Almacenamiento de Glucógeno Tipo II (Enfermedad de Pompe) (Infantil); Trastornos Relacionados con MAPT; Trastornos Relacionados con MECP2; Condrosplasia Punctata Rizomérica Tipo 1; Síndrome de Roberts; Enfermedad de Sandhoff; Enfermedad de Schindler - Tipo 1; Deficiencia de Adenosina Desaminasa; Síndrome de Smith-Lemli-Opitz; Atrofia Muscular Espinal; Ataxia Espinocerebelosa de Brote Infantil; Deficiencia de Hexosaminidasa A; Displasia Tanatofórica Tipo 1; Trastornos Relacionados con Colágeno Tipo VI; Síndrome de Usher Tipo I; Distrofia Muscular Congénita; Síndrome de Wolf-Hirschhorn; Deficiencia Lisosomal de Lipasa Ácida; y Xenoderma Pigmentosum.

45 Como resultará evidente, se prevé que el presente sistema se pueda utilizar para fijar como objetivo a cualquier secuencia de polinucleótidos de interés. Algunos ejemplos de afecciones o enfermedades que podrían ser tratadas útilmente utilizando el presente sistema se incluyen en las Tablas anteriores y también se proporcionan allí ejemplos de genes asociados actualmente con esas afecciones. Sin embargo, los genes ejemplificados no son exhaustivos.

EJEMPLOS

50 Los siguientes ejemplos se dan con el propósito de ilustrar diversas realizaciones de la invención.

Ejemplo 1: Mejora del sistema Cas9 para la aplicación in vivo

La solicitante llevó a cabo una búsqueda metagenómica para una Cas9 con pequeño peso molecular. La mayoría de los homólogos de Cas9 son bastante grandes. Por ejemplo, la SpCas9 es de alrededor de 1368aa de larga, que es demasiado grande para ser empaquetada fácilmente en vectores virales para el suministro.

5 A través de análisis computacional, la solicitante encontró que en la cepa bacteriana *Campylobacter* existen dos proteínas Cas9 con menos de 1000 aminoácidos. La secuencia para una Cas9 de *Campylobacter jejuni* se presenta a continuación. En esta longitud, CjCas9 puede empaquetarse fácilmente en AAV, lentivirus, adenovirus y otros vectores virales para el suministro sólido en células primarias y en modelos animales *in vivo*. En una realización preferida de la invención, se utiliza la proteína Cas9 de *S. aureus*.

>Cas9 de *Campylobacter jejuni* (CjCas9)

MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARR
 KARLNHLKHLIANEFKLNEDYQSFDESLAKAYKGLISPYELRFRALNELLSKQDFAR
 VILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSK
 EFTNVRNKKESYERCIAQSFLKDELKLFKKQREFGFSFSKFFEEVLSVAFYKRALKDFS
 HLVGNCSEFTDEKRAPKNSPLAFMFVALTRIINLLNKNTEGILYTKDDLNALLNEVLK
 NGTLTYKQTKKLLGLSDDYEFKGEKGTIFYEFKKEFIKALGEHNLSQDDLNEIAKDIT
 LIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNE
 LNLKVAINEDKKDFLPAFNETYYKDEVTPVVLRAIKEYRKVLNALLKKYGVHVKINIE
 LAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGKINSKNILKLRLFKEQKEFCA
 YSGEKIKISDLQDEKMLEIDHIYPYSRFDSDSYMKNVLFVFTKQNQEKLNQTPFEAFGNS
 AKWQKIEVLAKNLPTKKQKRILDKNYKDKQKFNKDRNLNDTRYIARLVNLYTKDYL
 10 DFLPLSDDENTKLNQKQKGVHVEAKSGMLTSALRHTWGFSKDRNNHLHHAIDAVI
 IAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIF
 VSKPERKKPSGALHEETFRKEEFYQSYGGKEGVKALELGKIRKVNKIVKNGDMFR
 VDIFKHKKTNFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILMDENYEFCSLYK
 DSLILIQTKDMQEPEFVYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAS
 IGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK.

El elemento tracrRNA putativo para esta CjCas9 es:

TATAATCTCATAAGAAATTTAAAAAGGGACTAAAATAAAGAGTTTGCGGG
 ACTCTGCGGGGTTACAATCCCCTAAAACCGCTTTTAAAATT

La secuencia de Repetición Directa es:

15 ATTTTACCATAAAGAAATTTAAAAAGGGACTAAAAC

Un ejemplo de un ARN químa para CjCas9 es:

NNNNNNNNNNNNNNNNNNNNNGUUUUAGUCCCGAAAGGGACUAAAAUAAAGAGUU
 UGCGGGACUCUGCGGGGUUACAAUCCCUAAAACCGCUUUU

Ejemplo 2: Optimización de Cas9

20 Para potenciar la función o para desarrollar nuevas funciones, la solicitante genera proteínas quiméricas Cas9 mediante la combinación de fragmentos de diferentes homólogos de Cas9. Por ejemplo, dos proteínas quiméricas Cas9 de ejemplo:

Por ejemplo, la solicitante fusionó el extremo N de St1Cas9 (el fragmento de esta proteína está en **negrita**) con el extremo C de SpCas9 (el fragmento de esta proteína está subrayado).

>St1(N)Sp(C)Cas9

MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRRTNRQGRRLAR
 RKKHRRVRLNRLFEEGLITDFTKISINLNPNYQLRVKGLTDELSNEELFIALKNMVK
 HRGISYLLDDASDDGNSSVDYAQIVKENSQKLETKTPGQIQLERYQTYGQLRGDFT
 VEKDGKKHRLINVFPTSA YRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYH
 GPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDL
 NNLTVPTETKLSKEQKNQIINYVNEKAMGPAKLFKYIAKLLSCDVADIKGYRID
 KSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTNTEREGIQEAEHEF
 ADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTR
 LGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNVIEM
 ARENQTTQKGQKNSRERMKRIE EGKELGSQILKEHPVENTQLQNEKLYLYLQNGRD
 MYVDQELDINRLSDYDV DHIVPOSFLKDDSIDNKVLRSDKNRGKSDNVPSEEVVKKM
 KNYWRQLLNAKLITQRKFDNLTKAERGGGLSELDKAGFIKQRLVETROITKHVAQILDSR
 MNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGTA
 LIKKYPKLESEFVYGDYKVVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
 IRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPK KYGGFDSPTVAYSVLVVAKEVEGKSKLKS VKELLGITIMERSSFE
 KNPIDFLEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVN
 FLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKR VILADANLDKVL SAY
 NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGL
 YETRIDLSQLGGD

>Sp1(N)St1(C)Cas9

MDKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETA
 EATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIE
 GNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
 DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRENLI AQLPGEKKNGLFG
 NLI ALSGLTPNFKSNFDLAEDAQLSKD TYDDDLNLLAQIGDQYADFLAAKNLSD
 AILLSDILRVNTEITKAPLSAMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY
 AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
 HAILRRQEDFY PFLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
 VVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPA
 FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
 KIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG
 WGRLSRKLLINGIRDKQSGKTILDFLKSDFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
 DSLHEHIANLAGSPA IKKGILOTVKVVDELVKVMGRHKPENIVIEMARETNEDEKKAI
 QKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERC
 LYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQAL
 DSMDDAWSFRELKAFVRESKTL SNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYA
 SRVVLNALQEHFRAHKIDTKVS VVRGQFTS QLRRHWGIEKTRD TYHHHAVDALIA
 ASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFDLTKSK
 EFEDSILFSYQVDSKFN RKISDATIYATRQAKV GKDKADETYVLGKIKDIYTQDGYD
 AFMKIYKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKE
 EHG YIRKYSKKGNGPEIKSLKYYSKLG NHIDITPKDSNNKVV LQSVSPWRADVYF
 NKT TGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDS DSEFKFTLYKND
 LLLVKDETETKEQQLFRFLSRTMPKQKHVELKPYDKQKFEGGEALIKVLGNVANS
 GQCKKGLGKSNISYKVRTDVLGNQHIIKNEGD KPKLDF

El beneficio de producir Cas9 química incluye:

- 5 a. reducir la toxicidad.
- b. mejorar la expresión en células eucariotas
- c. mejorar la especificidad

d. reducir el peso molecular de la proteína, producir una proteína más pequeña mediante la combinación de los dominios más pequeños de diferentes homólogos de Cas9.

e. alterar el requisito de la secuencia PAM

Ejemplo 3: Utilización de Cas9 como una proteína de unión a ADN genérica

5 La solicitante utilizó Cas9 como una proteína de unión a ADN genérica mediante la mutación de los dos dominios catalíticos (D10 y H840) responsables de la escisión de ambas cadenas de la diana de ADN. Con el fin de regular positivamente la transcripción de genes a un locus diana, la solicitante fusionó el dominio de activación transcripcional (VP64) a Cas9 (Figura 5). La solicitante planteó la hipótesis de si sería importante ver una fuerte localización nuclear de la proteína de fusión VP64-Cas9, ya que la fuerza de activación del factor de transcripción es una función del tiempo de estancia en la diana. Por lo tanto, la solicitante clonó un conjunto de construcciones Cas9-VP64-GFP, las transfeció en células 293 y evaluó su ubicación bajo un microscopio de fluorescencia 12 horas después de la transfección (Figura 8).

15 Las mismas construcciones se clonaron como un 2A-GFP en lugar de una fusión directa con el fin de probar funcionalmente las construcciones sin un GFP voluminoso presente que interfiriera. La solicitante eligió fijar como objetivo el locus Sox2 con el transactivador Cas9, ya que podría ser útil para la reprogramación celular y el lugar ya ha sido validado como una diana para la activación transcripcional mediada por TALE-TF. Para el locus Sox2, la solicitante eligió ocho dianas cerca del sitio de inicio de la transcripción (TSS). Cada una de las dianas era de 20 pb de longitud con un motivo adyacente al protoespaciador vecino NGG (PAM) (Figura 12). Cada una de las construcciones Cas9-VP64 se co-transfectó con cada uno de los ARN crispr quiméricos generados por PCR (chiRNA) en células 293. 72 horas después de la transfección, la activación transcripcional se evaluó mediante RT-qPCR (Figura 6).

25 Para optimizar adicionalmente el activador de la transcripción, la solicitante tituló la relación de chiRNA (Sox2.1 y Sox2.5) para Cas9 (NLS-VP64-NLS-hSpCas9-NLS-VP64-NLS), transfectado en células 293, y cuantificó el uso de RT-qPCR (Figura 11). Estos resultados indican que Cas9 se puede utilizar como un dominio de unión a ADN genérico para regular positivamente la transcripción de genes en un locus diana.

La solicitante diseñó una segunda generación de construcciones. (Tabla 1 a continuación).

pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-hSpCsn1(D10A, H840A)-NLS
pLenti-EF 1a-GFP-2A-6xHis-NLS-VP64-NLS-hSpCsn1(D10A, H840A)
pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-NLS-hSpCsn1(D10A, H840A)
pLenti-EF1a-GFP-2A-6xHis-NLS-hSpCsn1(D10A, H840A)-NLS
pLenti-EF1a-GFP-2A-6xHis-NLS-hSpCsn1(D10A, H840A)
pLenti-EF1a-GFP-2A-6xHis-NLS-NLS-hSpCsn1(D10A, H840A)

30 La solicitante utiliza estas construcciones para evaluar la activación transcripcional (construcciones VP64 fusionadas) y la represión (Cas9 solamente) por RT-qPCR. La solicitante evaluar la ubicación celular de cada una de las construcciones utilizando anticuerpo anti-His, la actividad nucleasa utilizando un ensayo de nucleasa Surveyor y la afinidad de unión a ADN utilizando un ensayo de desplazamiento en gel. En una realización preferida de la invención, el ensayo de desplazamiento en gel es un ensayo de desplazamiento en gel EMSA.

Ejemplo 4: Represor de Cas9

5 Se ha demostrado previamente que dCas9 se puede utilizar como un dominio de unión a ADN genérico para reprimir la expresión de genes. La solicitante informan de un diseño mejorado de dCas9, así como fusiones dCas9 a los dominios represores KRAB y SID4x. A partir del banco de plásmidos creado para modular la transcripción utilizando Cas9 en la Tabla 2 que figura más adelante, y la Figura 15, se caracterizaron funcionalmente por qPCR los siguientes plásmidos represores: pXRP27, pXRP28, pXRP29, pXRP48, pXRP49, pXRP50, pXRP51, pXRP52, pXRP53, pXRP56, pXRP58, pXRP59, pXRP61 y pXRP62.

10 Cada uno de los plásmidos represores dCas9 fue co-transfectado con dos ARNs guía dirigidos a la cadena codificadora del gen beta-catenina (los ARNg fueron diseñados por mi colaborador en el Broad, Joseph Rosenbluh). Se aisló el ARN 72 horas después de la transfección y expresión de genes se cuantificó por RT-qPCR (Figura 16). El gen endógeno de control era GAPDH. Dos shRNAs validados se utilizaron como controles positivos. Los controles negativos eran determinados plásmidos transfectados sin ARNg, éstos se designan como " pXRP ## control " en la Figura 16. Los plásmidos pXRP28, pXRP29, pXRP48 y pXRP49 pudieron reprimir el gen de la beta-catenina cuando se utiliza la estrategia de focalización especificada. Estos plásmidos corresponden a dCas9 sin un dominio funcional (pXRP28 y pXRP28) y dCas9 fusionado a SID4x (pXRP48 y pXRP49).

15 Tabla 2

pXRP024-pLenti2-EF1a-VP64-NLS-FLAG-Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP025-pLenti2-EF1a-VP64-NLS-GGGGS ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP026-pLenti2-EF1a-VP64-NLS-EAAAK ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP027-pLenti2-EF1a-NLS-FLAG-Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP028-pLenti2-EF1a-NLS-GGGGS ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP029-pLenti2-EF1a-NLS-EAAAK ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP030-pLenti2-pSV40-VP64-NLS-FLAG-Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP031-pLenti2-pPGK-VP64-NLS-FLAG-Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP032-pLenti2-LTR-VP64-NLS-FLAG-Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP033-pLenti2-pSV40-VP64-NLS-GGGGS ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP034-pLenti2-pPGK-VP64-NLS-GGGGS ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP035-pLenti2-LTR-VP64-NLS-GGGGS ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP036-pLenti2-pSV40-VP64-NLS-EAAAK ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP037-pLenti2-pPGK-VP64-NLS-EAAAK ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP038-pLenti2-LTR-VP64-NLS-EAAAK ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP048-pLenti2-EFla-SID4x-NLS-FLAG-Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE

ES 2 576 128 T3

pXRP049-pLenti2-EF1a-SID4X-NLS-GGGGS ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP050-pLenti2-EF1a-SID4X-NLS-EAAAK ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP051-pLenti2-EF1a-KRAB-NLS-FLAG-Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP052-pLenti2-EF1a-KRAB-NLS-GGGGS ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP053-pLenti2-EF1a-KRAB-NLS-EAAAK ₃ Enlazador-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP054-pLenti2-EF1a-dCas9-Enlazador-FLAG-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP055-pLenti2-EF1a-dCas9-Enlazador-FLAG-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP056-pLenti2-EF1a-dCas9-Enlazador-FLAG-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP057-pLenti2-EF1a-dCas9-GGGGS ₃ -NLS-VP64-gLuc-2A-GFP-WPRE
pXRP058-pLenti2-EF1a-dCas9-GGGGS ₃ -NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP059-pLenti2-EF1a-dCas9-GGGGS ₃ -NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP060-pLenti2-EF1a-dCas9-EAAAK ₃ -NLS-VP64-gLuc-2A-GFP-WPRE
pXRP061-pLenti2-EF1a-dCas9-EAAAK ₃ -NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP062-pLenti2-EF1a-dCas9-EAAAK ₃ -NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP024-pLenti2-EF1a-VP64-NLS-FLAG-Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP025-pLenti2-EF1a-VP64-NLS-GGGGS ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP026-pLenti2-EF1a-VP64-NLS-EAAAK ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP027-pLenti2-EF1a-NLS-FLAG-Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP028-pLenti2-EF1a-NLS-GGGGS ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP029-pLenti2-EF1a-NLS-EAAAK ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP030-pLenti2-pSV40-VP64-NLS-FLAG-Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP031-pLenti2-pPGK-VP64-NLS-FLAG-Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP032-pLenti2-LTR-VP64-NLS-FLAG-Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE

ES 2 576 128 T3

pXRP033-pLenti2-pSV40-VP64-NLS-GGGGS ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP034-pLenti2-pPGK-VP64-NLS-GGGGS ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP035-pLenti2-LTR-VP64-NLS-GGGGS ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP036-pLenti2-pSV40-VP64-NLS-EAAAK ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP037-pLenti2-pPGK-VP64-NLS-EAAAK ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP038-pLenti2-LTR-VP64-NLS-EAAAK ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP048-pLenti2-EF1a-SID4x-NLS-FLAG-Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP049-pLenti2-EF1a-SID4X-NLS-GGGGS ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP050-pLenti2-EF1a-SID4X-NLS-EAAAK ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP051-pLenti2-EF1a-KRAB-NLS-FLAG-Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP052-pLenti2-EF1a-KRAB-NLS-GGGGS ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP053-pLenti2-EF1a-KRAB-NLS-EAAAK ₃ Enlazador-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP054-pLenti2-EF1a-Cas9-Enlazador-FLAG-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP055-pLenti2-EF1a-Cas9-Enlazador-FLAG-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP056-pLenti2-EF1a-Cas9-Enlazador-FLAG-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP057-pLenti2-EF1a-Cas9-GGGGS ₃ -NLS-VP64-gLuc-2A-GFP-WPRE
pXRP058-pLenti2-EF1a-Cas9-GGGGS ₃ -NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP059-pLenti2-EF1a-Cas9-GGGGS ₃ -NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP060-pLenti2-EF1a-Cas9-EAAAK ₃ -NLS-VP64-gLuc-2A-GFP-WPRE
pXRP061-pLenti2-EF1a-Cas9-EAAAK ₃ -NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP062-pLenti2-EF1a-Cas9-EAAAK ₃ -NLS-KRAB-gLuc-2A-GFP-WPRE

Ejemplo 5: Modulador transcripcional de Cas9 y ubicación nuclear

Modulador transcripcional de Cas9. La solicitante se ha propuesto cambiar el sistema de Cas9/gRNA CRISPR en un sistema de unión a ADN generalizado en la que la solicitante pueda ejecutar funciones más allá de la escisión del ADN. Con lo cual, al fusionar dominios funcionales en una Cas9 catalíticamente inactiva, dCas9, la solicitante puede impartir nuevas funciones tales como la activación/represión transcripcional, la metilación/desmetilación o modificaciones de cromatina. Para lograr este objetivo, la solicitante hizo un mutante Cas9 catalíticamente inactivo cambiando dos residuos esenciales para la actividad nucleasa, D10 y H840, a alanina. Se ha demostrado previamente que mediante la mutación de estos dos residuos se suprime la actividad nucleasa de Cas9, al tiempo que se mantiene la capacidad de unir ADN diana. Los dominios funcionales en los que la solicitante decidió centrarse para poner a prueba la hipótesis de la solicitante son el activador transcripcional VP64 y los represores transcripcionales SID y KRAB.

Ubicación nuclear de dCas9. La solicitante planteó la hipótesis de que el modulador de la transcripción más eficaz dCas9 estaría fuertemente localizado en el núcleo en donde tendría su mayor influencia en la transcripción. Por otra parte, cualquier dCas9 residual en el citoplasma podría tener efectos no deseados. En trabajos anteriores, la solicitante determinó que Cas9 de tipo salvaje no se localiza en el núcleo sin incluir múltiples señales de localización nuclear (NLSs). Debido a que se requerían múltiples secuencias NLS, la solicitante razonó que es difícil conseguir Cas9 en el núcleo y que cualquier dominio adicional que esté fusionado a Cas9 podría interrumpir la ubicación nuclear. Por lo tanto, la solicitante construyó cinco construcciones de fusión dCas9-VP64-GFP para optimizar la posición y el número de secuencias de NLS. Una segunda versión de cada una de las construcciones se hizo para el ensayo funcional, y estas construcciones no se fusionaron directamente a GFP. Un esquema de estos diseños se muestra en la Figura 7. Estas construcciones se clonaron en una cadena principal de penti2 bajo la expresión del promotor EF1a humano. El elemento WPRE también se añadió para la expresión más sólida de proteínas. Cada una de las construcciones se transfeció a células HEK 293FT utilizando Lipofectame 2000 y se tomaron imágenes 24 horas después de la transfección (Figura 8). La Figura 8 muestra que la mejor ubicación nuclear se obtiene cuando las proteínas de fusión tienen secuencias NLS tanto en los extremos N como C término de la proteína de fusión. La ubicación nuclear más alta observada se produjo en la construcción con cuatro elementos NLS.

Para entender más sólidamente la influencia de elementos NLS en Cas9, la solicitante realizó 16 fusiones dCas9-GFP mediante la adición de la misma secuencia NLS de alfa importina en cualquiera de los extremos N o C mirando cero a tres repeticiones en tándem. Se utilizó la misma estrategia de clonación que la que se explicó anteriormente. Cada una de las construcciones se transfeció en células HEK 293FT utilizando Lipofectame 2000 y se tomaron imágenes 24 horas después de la transfección (Figura 9). Cabe destacar que el número de elementos de NLS no se correlaciona directamente con el grado de ubicación nuclear. La adición de una NLS al extremo C tiene una mayor influencia en la ubicación nuclear que la adición en el extremo N. Para aplicaciones futuras es más favorable tener proteínas más pequeñas. Por lo tanto, la arquitectura óptima de NLS para dCas9 era tener una NLS en el extremo N y el extremo C (pXRPNLS006).

Un factor de confusión potencial de usar una fusión GFP para determinar la ubicación nuclear es que el propio GFP podría tener una gran influencia en donde se localiza la proteína en la célula. Por lo tanto, la solicitante determinó que un buen enfoque sería incluir un marcador 6xHis a dCas9, transfectar en células 293FT, y luego teñir con un anticuerpo anti-6xHis. La solicitante hizo seis versiones utilizando esta idea (Figura 10). Tres de ellas fueron construidas para la activación transcripcional (fusiones VP64), y las otras tres fueron para la represión transcripcional (sin dominio funcional). También se hicieron estas construcciones de modo que la solicitante podía purificar la proteína recombinante utilizando una columna de níquel. Hay muchas aplicaciones de las construcciones 6xHis-dCas9 tales como, por ejemplo, para llevar a cabo un ensayo de desplazamiento en gel de electromovilidad para demostrar que Cas9 mutada puede de hecho unirse a su diana de ADN. Las construcciones de represión (pXRP013, pXRP014, pXRP015) se transfecieron en células HEK 293FT y se inmunotñieron y se tomaron imágenes después de 24 horas (Figura 11). Bajo este contexto específico, dCas9 requiere dos señales NLS en el extremo N a ser ubicadas en el núcleo. Curiosamente, una NLS en cada extremo no provocó que dCas9 estuviera fuertemente ubicada en el núcleo. Tomados en conjunto, lo que sugieren estos datos es que no existe una norma generalizada que se pueda hacer para decir de manera concluyente que una proteína de fusión dCas9 quedará ubicada en el núcleo. La construcción de interés debe ser validada en el sistema se va a utilizar para determinar si de hecho está ubicada en el núcleo.

Ensayo Funcional del Modulador Transcripcional de Cas9. La solicitante sometió a ensayo funcional la proteína dCas9-VP64 fijando como objetivo el locus Sox2 y cuantificando la activación transcripcional mediante RT-qPCR. Se eligieron ocho sitios diana de ADN para abarcar el promotor de Sox2 (Figura 12). Las casetas de expresión de ARNg se clonaron en un plásmido de la cadena principal pAAV-U6-gRNA/tracrRNA, pA6. El plásmido utiliza la enzima BbsI de tipo IIs para revelar extremos adherentes para el oligo ligamiento reasociado. Así, para cada una de las dianas todo lo que se necesita son dos oligonucleótidos complementarios con extremos adherentes apropiados.

Cada una de las construcciones que contenía dCas9 fue co-transfectada con plásmidos pA6 en células HEK 293FT utilizando Lipofectame 2000. 72 horas post-transfección, el ARN total se extrajo de las células. 1 ug de ARN se transcribió de forma inversa en ADNc (qScript Supermix) en una reacción de 40 ul. 2 ul del producto de reacción se añadió en una sola reacción qPCR de 20 ul de ensayo TaqMan. Cada uno de los experimentos se realizó por triplicado biológico y técnico. Ninguna reacción de control RT y de control del molde mostró amplificación. Los resultados de este experimento se muestran en la Figura 13. Las construcciones que no muestran una fuerte ubicación nuclear, pXRP02 y pXRP04, no dan lugar a activación alguna. Para la construcción que mostró una fuerte ubicación nuclear, pXRP08, se observó una activación moderada. Estadísticamente se observó una activación significativa en el caso de los ARNg Sox2.4 y Sox2.5. De las construcciones 6xHis-dCas9 (pXRP011, pXRP013, pXRP015) también se ensayaron utilizando las mismas dianas Sox2 (Figura 14). En comparación con los controles simulados no ARNg, la solicitante pudo reprimir o activar Sox2.

Para optimizar adicionalmente el activador y represor transcripcional dCas9 la solicitante construyó 31 construcciones para explorar diferentes enlazadores, dominios funcionales y fusiones en los extremos N y C. La lista de construcciones y los elementos críticos se muestran en la Figura 15. Se testaron tres enlazadores y el primero era un enlazador inusual que fue publicado por Shen y colegas (Shen et al. 2013, Documento de investigación celular del laboratorio de Xingxu Huang: Ceil Research 23, 720-723 (mayo de 2013)) donde demostraron que mejoró drásticamente la ubicación nuclear y la actividad nucleasa de Cas9. El segundo y tercer enlazadores eran enlazadores muy comunes conocidos por ser flexibles (GGGGS3) y rígidos (EAAAK3). Los dominios funcionales testados fueron VP64 para la activación y SID4x y KRAB para la represión, También se hizo dCas9 sin un dominio funcional para testar la represión. Los promotores testados eran EF1a humano, pPGK humano, SV40, y la ausencia de un promotor, en donde la expresión es impulsada por el LTR viral.

Todas las construcciones de activador transcripcional fueron fijadas como objetivo al locus Sox2 utilizando una combinación de dos ARNg, hSox2.1 y hSox2.4. Las construcciones de activador dCas9 fueron co-transfectadas con los ARNg en células HEK 293FT y analizadas por qPCR después de 72 horas (Figura 16).

Se adoptó un ensayo informador de luciferasa Sox2 para simplificar los métodos asociados para la determinación de los niveles de activación. El ensayo funciona mediante la co-transfección en tres plásmidos, 1) dCas9, 2) ARNg, 3) elemento sensible a Sox2 que impulsa luciferasa. Así el nivel de luciferasa es proporcional a la cantidad de Sox2. El plásmido sensible a Sox2 se adquirió de BioCat. El locus Sox2 fue fijado como objetivo utilizando diferentes combinaciones de ARNg y las construcciones pXRP024 y pXRP025. Se utilizó un control "Sin dCas9" para determinar el nivel basal de activación. Los resultados mostrados en el panel superior de la Figura 11 sugieren que la activación triple podría alcanzarse en comparación con controles "Sin dCas9" y "Sin ARNg".

Un ensayo de luciferasa similar se adoptó para testar la represión transcripcional de dCas9. El candidato de gen que eligió la solicitante para la represión era beta-catenina y las dianas estaban ubicadas cerca del extremo 5' del gen cuerpo en la cadena no codificadora. El primer experimento que utiliza este ensayo comparó el represor pXRPNLS006 con el represor "dCas9", publicado por Qi y colegas (Qi et al 2003, Documento de la célula del laboratorio de Wendell Lim: <http://www.sciencedirect.com/science/article/pii/S0092867413002110>). Los resultados del ensayo de luciferasa (panel inferior de la Figura 11) muestran que se puede alcanzar la represión triple y que el plásmido pXRPNLS006 aquí reseñado funciona significativamente mejor que el estándar actual, "dCas9".

Ejemplo 6: Activador transcripcional de Cas9 en líneas celulares murinas y humanas

La solicitante demostró previamente la capacidad de CRISPR/Cas9 de actuar como un activador transcripcional. En este ejemplo, la solicitante testaron adicionalmente esta capacidad en una línea celular murina. La solicitante también examinó la actividad basal sin guía de dCas9-VP64 (H840A D10A Cas9 doble mutante fusionada al dominio de activación transcripcional de VP64). La solicitante suministró el sistema activador Cas9 a células de ratón y, por separado, caracterizó los efectos no guiados de dCas9-VP64 para varios genes humanos.

Experimentos con la Línea Celular de Ratones: La solicitante seleccionó el locus para el gen de ratón *Neurog2* para testar la activación transcripcional de dCas9-VP64 en células Neuro 2A. La solicitante también seleccionó 7 secuencias que fijan como objetivo SgrNA de una región de 300 pb aguas arriba a 100 pb aguas abajo del sitio de inicio transcripcional de *Neurog2* anotado. Estas secuencias guía se clonaron por el método Golden Gate en el plásmido px362 que contenía el promotor U6 aguas arriba de la cadena principal de clonación un sgRNA quimérico.

Células Neuro 2A se cultivaron en medio que consistía en una relación 1:1 de D5 (100:5:1 DMEM de alto contenido en glucosa con Glutamax y piruvato de sodio: suero bovino fetal: penicilina/estretomicina). Para la transfección, células 120K en 0,5 mL de medio de cultivo se sembraron en cada uno de los pocillos de una placa de 24 pocillos.

Después de 22 h en cultivo, cada uno de los pocillos de células Neuro 2A se transfectaron con 1 ug de ADN utilizando 5 uL de Lipofectamine siguiendo los pasos recomendados por el fabricante (Life Technologies). Para todas las muestras de dCas9-VP64 + sgRNA, se utilizaron 0,5 ug de plásmido dCas9-VP64 y 0,5 ug de plásmido de expresión sgRNA. Para las muestras control, las masas de plásmido reemplazaron al plásmido que expresa GFP según fuese necesario. 4 h después de la transfección, el medio se reemplazó con 1 mL de medio de cultivo. 48 h después de la transfección, el ARN fue purificado utilizando el kit de ARN Macherey Nagel Nucleospin 96. La transcripción inversa se realizó con qScript cDNA supermix de acuerdo con el protocolo del fabricante. La qPCR se realizó utilizando sondas Taqman y Taqman Fast Advanced Master Mix de Life Technologies en una máquina de PCR en tiempo real LightCycler 480 II de Roche.

10 *Experimentos con la Línea Celular Humana:* La solicitante seleccionó 8 secuencias de sgRNA previamente disponibles para el ensayo de células 293FT utilizando la construcción de activador pXRP057 dCas9-VP64 de la solicitante. Secuencias guía se clonaron por el método Golden Gate en el plásmido px362 que contiene el promotor U6 aguas arriba de una cadena principal de clonación de sgRNA quimérico.

15 Células 293FT se cultivaron en medio D10 (relación 100:10:1 de DMEM de alto contenido en glucosa con Glutamax y piruvato de sodio: suero bovino fetal: disolución 100 x HEPES). Para la transfección, células 100K en 0,5 mL de medio de cultivo se sembraron en cada uno de los pocillos de una placa de 24 pocillos. Después de 22 h de cultivo, cada uno de los pocillos de células 293FT se transfectó con 1 ug de ADN utilizando 5 uL de Lipofectamine siguiendo los pasos recomendados por el fabricante (Life Technologies). Para todas las muestras de dCas9-VP64 + sgRNA, se utilizaron 0,5 ug de plásmido dCas9-VP64 y 0,5 ug de plásmido de expresión sgRNA. Para las muestras control, incluyendo las analizadas para efectos de transcripción no guiados, las masas de plásmido fueron reemplazadas por el plásmido que expresa GFP según fuese necesario. 4 h después de la transfección, el medio se reemplazó con 1 mL de medio de cultivo. 48 h después de la transfección, el ARN fue purificado utilizando el kit de ARN Macherey Nagel Nucleospin 96. La transcripción inversa se realizó con qScript cDNA supermix de acuerdo con el protocolo del fabricante. La qPCR se realizó utilizando sondas Taqman y Taqman Fast Advanced Master Mix de Life Technologies en una máquina de PCR en tiempo real LightCycler 480 II de Roche.

Resultados experimentales para Experimentos con la Línea Celular de Ratón: Tal como se muestra en la Fig. 20, 6 de los 7 sgRNAs fijados como objetivo al locus Neurog2 de ratón fueron capaces de inducir la regulación positiva de ARNm de Neurog2, medida por qPCR. Con estos resultados, la solicitante demostró la utilidad del sistema de activador Cas9 para la modulación génica en un modelo celular de ratón. Las secuencias sgRNA 7 se enumeran a continuación:

<u>sgRNA</u>	<u>SECUENCIA</u>
SGRNA0001	TGGTTCAGTGGCTGCGTGTC
sgRNA0002	TGTTTTCTTGGTGGTATATA
sgRNA0003	ATACGATGAAAAGAATAAGC
sgRNA0004	GGGGGAGAGGGACTAAAGAA
sgRNA0005	GGGCGGGGGAAGGGTAGGTG
sgRNA0006	ATTAGATAAAGGGGGGACGG
sgRNA0007	CGGCTTTAACTGGAGTGCCT

Resultados Experimentales para Experimentos de Línea Celular Humana: Tal como se muestra en la Fig. 21, sgRNAs fijados como objetivo a varios genes humanos (ASCL1, MYOD1, VEGFa y NTF3) indujeron con éxito la regulación positiva del ARNm deseado utilizando la construcción pXRP057 dCas9-VP64 de la solicitante (Koneremann et al., Optical control of mammalian endogenous transcription and epigenetic states. Nature. 22 de

agosto de 2013; 500(7463):472-6). La solicitante demostró que muestras transfectadas con la construcción dCas9-VP64, pero sin un sgRNA tenían cambios significativos en el nivel de expresión de varios genes y se ha de realizar una caracterización adicional (Fig. 22). Los recuentos de PAM (-100 a +100 pb de TSS) se indican a continuación:

Gen	Nivel de Cambio Basal	Recuentos de NGG + NAG
Neurog2	0,60	213
VEGFA	1,25	211
NTF3	0,94	203
ASCL1	4,19	235
MYOD1	0,90	204

Ejemplo 7: Estudios adicionales sobre datos sobre el modulador-activador y represor de Cas9

- 5 La solicitante llevó a cabo estudios para desarrollar adicionalmente el activador (pXRP57) y el represor (pXRP48) de dCas9 (Fig. 23). La solicitante clonó los plásmidos pXRP57 y pXRP48 (Koneremann et al, Optical control of mammalian endogenous transcription and epigenetic states. Nature. 22 de agosto de 2013; 500(7463):472-6). La secuencia de activador Cas9 se indica en la Fig. 24 y la secuencia de represor de Cas9 se indica en la Fig. 25. Se recogieron datos adicionales para el activador (Fig. 26), y el represor (Fig. 27). Una lista de los ARN de guía utilizados en el experimento se indica a continuación:
- 10

Gen	ARN guía
hSox2-1	GTGGCTGGCAGGCTGGCTCT
hSox2-2	GGCCTCCCCGCGCGGCCGG
hSox2-3	GCCCCCTTTCATGCAAACC
hSox2-4	GACAGCCCCCGTCACATGGA
hSox2-5	GGCAGGCGAGGAGGGGAGG
hSox2-6	GGCGGGGCCTCCCGCGCCGC
hSox2-7	GCTGCCGGGTTTTGCATGAA
hSox2-8	GGGGCTGTCAGGGAATAAAT
hKlf4-1	GAGAGAACGAACGTGTCTGC
half4-2	GAGGGTCACTCGGCGGCTCC
half4-3	GCGCGCTCCACACAACCTCAC
half4-4	GGGGCTGTGGCCGGGGCGGT

Gen	ARN guía
hKlf4-5	GGCGACCGCGACAGTGGTGG
half4-6	GCAAAAATAGACAATCAGCA
half4-7	GAAGGATCTCGGCCAATTTG
half4-8	GTGGGGGCCAGAAAGGTCCT
hNanog-1	GCCACGGCCTCCCAATTTAC
hNanog-2	GGAATATGGTTCAACAGGAA
hNanog-3	GCTGCAGAGTAACCCAGACT
hNanog-4	GCCTTGGTGAGACTGGTAGA
hNanog-5	GATTAAGTGAATAATTCACAA
hNanog-6	GTGTGCCCGCCAGGAGGGGT
hNanog-7	GTTGCCTGCATAATAACATG
hNanog-8	GGAGGAAAAAATTTAAGAGG
hOct4-1	GGACCGGGATTGTCCAGCCA
hOct4-2	GAGTGATAAGACACCCGCTT
hOct4-3	GCAGCTGGCCATTGTGCTTA
hOct4-4	GGAGAGGGGGTCAAGCACCT
hOct4-5	GCGGGTTGGGAGTTGAAAGT
hOct4-6	GCTCCAGCCTCCTAAGTGGC
hOct4-7	GGAGGTGGGGGAGAACTG
hOct4-8	GGTGAAATGAGGGCTTGCGA

Ejemplo 8: Sistema activador de CRISPR/Cas9 por la inducibilidad con luz del sistema LITE

La solicitante combinó la orientación de ARN del sistema activador de CRISPR/Cas9 con la inducibilidad de la luz del sistema LITE. Un sistema de expresión inducible basado en CRISPR/ Cas9 puede hacer que la pantalla inducible de muchos genes fije objetivos mucho más rápido y más fácil que lo que antes era posible, al tiempo que también

5

proporciona la oportunidad de multiplexar fácilmente objetivos diferentes mediante la combinación de sgRNAs fijados como objetivo a dispares loci genéticos,

5 Materiales y métodos: El sistema CasLITE consiste en 3 componentes: una proteína de fusión dCas9-CIB1, la construcción CRY2PHR-VP64 y una secuencia guía sgRNA. La solicitante sintetizó 3 versiones diferentes de dCas9-CIB1: (1) dCas9-GS-NLS-CIB1 que comprendía dCas9 doble mutante (D10A H840A) de su plásmido pXRP057, un enlazador de glicina serina, señal de ubicación nuclear de SV40, y CIB1 de LITE1.0. (2) dCas9-GS-NLS-NLS-CIB1 que comprende dCas9 doble mutante, el enlazador de glicina serina, 2 secuencias de SV40 NLS, y CIB1 de LITE1.0. (3) dCas9-GS-CIB1 (mNLS d318-334) que comprende la dCas9 doble mutante, enlazador de glicina serina del LITE2.0 y CIB1 de LITE2.0. Los vectores utilizados en la práctica de la invención se ilustran en las Figs. 28, 29, 10 30.

La solicitante seleccionó secuencias de sgRNA fijadas como objetivo a ASCL1 y MYOD1, previamente validadas con un activador constitutivo dCas9-VP64, para someter a ensayo los sistemas de CasLITE en células humanas. Las secuencias guía se clonaron por el método de Golden Gate en el plásmido px362 que contiene el promotor U6 aguas arriba de la cadena principal de clonación de un sgRNA quimérico.

15 Células 293FT se cultivaron en medio D10 sin HEPES (relación 10:1 de DMEM de alto contenido en glucosa con Glutamax y piruvato de sodio: suero bovino fetal). Para la transfección, células 100K en 0,5 mL de medio de cultivo se sembraron en cada uno de los pocillos de una placa de 24 pocillos. Después de 22 h de cultivo, cada uno de los pocillos de células 293FT se transfectó con 1,5 ug de ADN utilizando 7,5 uL de Lipofectamine siguiendo los pasos recomendados por el fabricante (Life Technologies). Para todas las muestras de CasLITE + sgRNA, se utilizaron 0,5 20 ug de plásmido dCas9-CIB1, 0,5 ug de plásmido CRY2PHR-VP64 (LITE1.0 o LITE2.0 para emparejarse con CIB1) y 0,5 ug de plásmido de expresión sgRNA. Los experimentos con 4 sgRNAs utilizaron masas iguales de cada uno de los 4 sgRNAs, totalizando 0,5 ug. Para las muestras control, las masas de plásmido fueron reemplazadas por el plásmido que expresa GFP según fuese necesario. 4 h después de la transfección, el medio se reemplazó con 1 mL de medio de cultivo. 48 h después de la transfección, la excitación se inició en muestras estimuladas con luz, 25 utilizando luz azul de 5 mW/cm² 475 nm. 12 h después de la estimulación, se recogieron todas las muestras, incluyendo todos los controles para el ARN, utilizando el kit de ARN Macherey Nagel Nucleospin 96. La transcripción inversa se realizó con qScript cDNA supermix de acuerdo con el protocolo del fabricante. La qPCR se realizó utilizando sondas Taqman y Taqman Fast Advanced Master Mix de Life Technologies en una máquina de PCR en tiempo real LightCycler 480 II de Roche.

30 Resultados: Las construcciones CasLITE exhibían diferentes niveles de activación de la transcripción inducible por la luz (Fig. 31). Aunque la funcionalidad inicial del sistema parece ser modesto, llamando a la experimentación repetida, estos resultados sugieren la posibilidad de la activación transcripcional guiada por ARN e inducible por la luz.

Ejemplo 9: Modelo de ratón Cas9 y validación de la línea de células madre embrionarias de ratón

35 La solicitante validó una línea de células madre embrionarias de ratón que se modificó para incluir la casete Cas9 Cre-dependiente (pCM2) en el locus Rosa26 (Fig. 32) y obtuvo resultados de genotipado para los fundadores transmitidos por la línea germinal (Fig. 33).

40 Para validar que la casete Cas9 era funcional, la línea mESC modificado se electropolarizó con las construcciones designadas en la Fig. 32. La línea mESC fue utilizada para la inyección de blastocistos en receptores C57BL/6J (realizado por la DCM en MIT). Se produjeron seis progenies quiméricas y luego se retrocruzaron al fondo C57BL/6J. La transmisión de la línea germinal se identificó con éxito sobre la base del color de la capa agouti. Dos de los seis quiméricos produjeron crías de transmisión de la línea germinal. ADN genómico fue extraído de tejido de los ratones Cas9 y se realizó una PCR de genotipado de Rosa 26 de brazo corto (Fig. 33). Una banda de 1,5 kb es el tamaño correcto para los transgenes insertados correctamente en el locus Rosa26. Estos resultados demuestran 45 la transmisión con éxito de la línea germinal del transgén Cas9 Cre-dependiente en el locus Rosa26.

El beneficio de la línea de activación de Cas9 Rosa 26 Cre-dependiente es que puede ser cruzada a cualquier línea impulsora de Cre para permitir la expresión de Cas9. Además de ello, Cas9 puede ser activada por el suministro de Cre por el vector viral u otros medios de suministro. Esto permite la expresión del tipo de célula, tejido y la especificidad de desarrollo de Cas9. Un beneficio adicional es que el ratón se puede utilizar para el aislamiento de 50 células primarias. Esto sería particularmente útil en el caso de un tipo de célula primaria que no es susceptible al suministro de ácidos nucleicos o que no se puede cultivar el tiempo suficiente para la utilización de vectores virales.

Ejemplo 10: Plataforma basada en CRISPR fijada como objetivo de ARN para la modificación por ingeniería genética del epigenoma de mamíferos

5 La solicitante desarrolla una plataforma basada en CRISPR fijada como objetivo de ARN para la modificación por ingeniería genética del epigenoma de mamíferos. Dos sistemas inmunes mediados por CRISPR fijados como
 10 objetivo de ARN, los loci CRISPR/Cmr, se caracterizaron recientemente de *Pyrococcus furiosus* y *Sulfolobus solfataricus*. Sobre la base de homología de proteínas, estos loci CRISPR/Cmr son parte de una familia mucho más grande que existe dentro de una diversidad de genomas bacterianos y arqueales. Al igual que en el ARNi y Cas9 el mecanismo de interferencia de un locus CRISPR/Cmr se basa en la homología de un corto tramo de nucleótidos y, por lo tanto, es susceptible a rastreo de todo el genoma. Por otra parte, la región de homología con el sitio diana se incrementa significativamente, haciendo mucho mayor el potencial para la especificidad.

15 La solicitante realizó un análisis metagenómico de la familia CRISPR/Cmr, el desarrollo de un sistema de ensayo *in vitro* para identificar loci funcionales, y la caracterización de la especificidad y la capacidad de multiplexar en células de mamífero. Una plataforma de ingeniería genética en el genoma basada en CRISPR fijada como objetivo ARN (Fig. 34) puede reemplazar a muchas utilidades de ARNi y proporcionar un sistema para perturbar la biología del ARN que nunca antes había sido posible.

20 Trabajo preliminar con el locus CRISPR/Cmr de *P. furiosus*: Basado en la extensa base de conocimientos sobre el locus CRISPR/Cmr de *P. furiosus*, la solicitante investigó la viabilidad de expresar los genes de Cmr dentro de células de mamíferos. ADN genómico de *P. furiosus* se obtuvo de la ATCC y se utilizó como un molde de la PCR para amplificar los seis genes Cmr (Cmr 1-6). La solicitante también amplificó otra proteína dentro del locus, Cas6. Cas6 está involucrado en el procesamiento de crRNA en su forma madura. Aunque sólo son esenciales cinco genes Cmr (Cmr1,2,3,4,6), la solicitante consideró que Cas6 puede ayudar en la solución de problemas si los cinco genes Cmr pasan a ser insuficientes en el contexto de los mamíferos.

25 Cmr1-6 y Cas6 se clonaron en plásmidos de expresión de mamíferos y se transfectaron en células HEK 293FT. Para facilitar la visualización, el plásmido de clonación contenía un marcador HA y una secuencia P2A-EGFP. La secuencia P2A provoca un salto ribosomal, dando como resultado dos proteínas maduras del mismo transcrito. 72 horas post-transfección se observó fluorescencia EGFP para cada uno de los plásmidos de expresión de Cmr (Fig. 35 A), lo que sugiere que Cmr1-6 y Cas6 se expresan con éxito. Para validar que las proteínas son estables en el entorno de células de mamíferos y que también son de tamaño correcto, la solicitante llevó a cabo una transferencia Western (Fig. 35B). Dentro de cada una de las pistas existe una banda del tamaño esperado, lo que sugiere que cada una de las proteínas se expresa y es estable dentro de las células de mamíferos. La segunda banda más grande que aparece dentro de cada una de las pistas puede ser el resultado de la insuficiencia de la secuencia de P2A para provocar un salto ribosomal y, por lo tanto, se crea una proteína de fusión y no dos productos de proteína individuales. La banda situada por encima del tamaño predicho es ~30 kDa más grande que el tamaño esperado, que corresponde a P2A-EGFP. La solicitante confirmará esto por inmunotinción el mismo borrón de transferencia utilizando un anticuerpo anti-EGFP y debería ver dos bandas por pista, una que es el tamaño de P2A-EGFP y otra banda más grande que es el tamaño de Cmr-P2A-EGFP. Los resultados de la transferencia Western también muestran una expresión variable para las siete proteínas, mientras que, Cmr1, Cmr3 y Cmr6 tienen una baja expresión, mientras Cmr2, Cmr4 y Cmr5 tienen una alta expresión. La solicitante ha demostrado que de hecho es posible expresar proteínas Cmr de *P. furiosus* dentro de células de mamíferos.

40 En este ejemplo, la solicitante identifica loci CRISPR/Cmr fijados como objetivo a ARN putativos y valida la expresión de proteínas individuales y de correspondientes crRNAs en células de mamífero. Utilizando un enfoque basado en la homología, la solicitante identifica loci CRISPR/Cmr fijados como objetivo a ARN putativos a través de una diversidad de genomas de bacterias y arqueas. Proteínas Cmr y crRNAs de cada uno de los locus se identifican y validan a través de la clonación, la transfección en células de mamíferos y la detección por transferencia western y transferencia Northern, respectivamente. La metodología es la siguiente:

45 1. Identificar loci CRISPR fijados como objetivo a ARN: La identificación de loci CRISPR fijados como objetivo a ARN del trabajo publicado es el primer paso en la comprensión de los componentes esenciales de un sistema CRISPR/Cmr de Tipo III-B funcional. Dos loci fijados como objetivo a ARN a partir de dos organismos diferentes se han publicado y se han caracterizado en diferentes grados.

50 2. Identificar loci CRISPR fijados como objetivo a ARN putativos a través de una diversidad de especies de bacterias y arqueas: Un extenso trabajo que se ha hecho hacia la clasificación de todos los conocidos loci CRISPR en grupos filogenéticos y, por lo tanto, se ha traducido en las bases de datos de alta confianza de proteínas orólogas. Utilizando las bases de datos más extensas y manualmente conservadas la solicitante identifica loci CRISPR/Cmr putativos basados en la presencia de múltiples ortólogos de proteínas. Se recogerá un subconjunto que abarca la

diversidad de loci CRISPR/Cmr para la caracterización molecular adicional en un amplio intento de probar la diversidad natural.

3. Identificar proteínas Cmr y crRNAs dentro de los loci CRISPR/Cmr: loci CRISPR/Cmr están bien estudiados y muchos genomas de bacterias y arqueas han sido ampliamente comentados, por lo que la identificación de las proteínas Cmr debería ser sencilla en estos casos. Sin embargo, en ausencia de una buena anotación, el problema será más complicado. Por esta razón, la solicitante preferentemente selecciona loci CRISPR/CMR de organismos en los que están disponibles buenas anotaciones. Si estas estrategias no proporcionan una cobertura completa de la diversidad de loci CRISPR/Cmr, la solicitante utiliza un enfoque basado en la homología combinado con algoritmos de predicción ORF para identificar las proteínas Cmr. Matrices de repetición CRISPR tienen estructuras predecibles y se encuentran cerca de las proteínas Cmr asociadas a CRISPR. La matriz de CRISPR consiste en una secuencia líder que impulsa la expresión de toda la matriz de repeticiones y espaciadores, creando un transcrito primario que es ampliamente procesado para hacer ARNs de CRISPR maduros (crRNA)s. Herramientas computacionales existentes (CRISPRFinder) permitirán la identificación de las matrices de repetición. La predicción de crRNA maduro de cada locus puede no ser tan sencilla y puede requerir la optimización. Hay características generales conocidas acerca de los crRNAs maduros de loci CRISPR/Cmr de Tipo III-B que la solicitante utiliza para predecir nuevos crRNAs. Todos los crRNAs maduros de loci CRISPR de Tipo III-B tienen un 5' mango que es parte de la secuencia de repetición, seguido de un espaciador. Las funciones 5' mango en el reconocimiento del crRNA por el complejo Cmr y el espaciador guía el complejo Cmr al ARN diana a través de apareamiento de bases de Watson-Crick. La solicitante diseña un panel de posibles arquitecturas de crRNA que son de diferente longitud total y que tienen 5' mangos de diferente longitud, ya que es difícil predecir las características únicas específicas de los nuevos crRNAs a priori.

4. Clonar proteínas Cmr y crRNAs en vectores de expresión de mamíferos: proteínas Cmr se clonan por separado en dos vectores de expresión de mamíferos diferentes (Fig. 36). El primer vector (V1.0) expresa proteínas Cmr individuales utilizando un promotor CMV, la secuencia de cola de poliA de la hormona de crecimiento bovina, y un elemento WPRE para maximizar la expresión de proteínas. El segundo vector (V1.2) utiliza los mismos elementos que en V1.0, pero incluirá un marcador HA en el extremo N y una secuencia P2A-EGFP en el extremo C. El marcador de HA permite la detección y cuantificación de proteínas y P2A-EGFP permite un método visual sencillo para detectar la expresión en células de mamífero. La solicitante hace ambas versiones, porque el impacto de la adición de un marcador HA y un péptido P2A en proteínas Cmr y la forma en que influye sobre la formación correcta del complejo Cmr es desconocido. crRNAs se expresan utilizando el promotor U6 humano, que ha sido utilizado con éxito para expresar los ARN no codificadores utilizados para RNAi y CRISPR/Cas9 que fija como objetivo ADN.

5. Validar la proteína Cmr y la expresión crRNA: la expresión de la proteína Cmr se valida transfecando plásmidos V1.2 en células HEK 293FT y observando la expresión de EGFP. A fin de validar adicionalmente que la proteína se expresa y el tamaño correcto la solicitante prepara lisados de proteínas a partir de células transfectadas y realiza una transferencia Western utilizando un anticuerpo anti-marcador HA. Para validar la expresión de crRNA, la solicitante purifica pequeños ARN de células transfectadas y sonda, mediante transferencia Northern.

La solicitante demuestra la escisión de ARN en un lisado de células de mamífero *in vitro* utilizando crRNAs codificados de forma natural. La solicitante desarrolla un simple lisado de células de mamíferos en un ensayo *in vitro* para validar rápidamente loci CRISPR/Cmr putativos. Los lisados celulares de células de mamífero transfectadas con proteínas Cmr se preparan y se incuban con crRNAs codificados de forma natural y una diana de ARN correspondiente. Loci CRISPR/Cmr en funcionamiento se identifican en base a los productos de escisión del tamaño correcto utilizando electroforesis en gel. La metodología es la siguiente:

1. Expresión de proteínas Cmr en células de mamíferos y preparación de lisados de células: la solicitante transfecta plásmidos de expresión individuales para todas las proteínas Cmr de una especie en células HEK 293FT y espera 72 horas para que se produzca la expresión sólida de la proteína. Las células se lisan, homogeneizaron y se dividen en partes alícuotas para uso futuro. Aunque no todas las células se transfectan con cada uno de los plásmidos Cmr, una vez que se preparan los lisados, todas las proteínas Cmr estarán en la misma mezcla. Este es un beneficio sobre el modelo de células vivas, en donde la co-transfección de siete componentes puede ser un reto. Por otra parte, esto dará a la solicitante un control absoluto sobre la concentración final del crRNA y ARN diana en la reacción.

2. Preparación de crRNAs: para cada locus CRISPR/Cmr la solicitante hace crRNAs maduros que contienen el espaciador endógeno. Específicamente, la solicitante utiliza la secuencia de espaciador del crRNA más próximo al líder. Como se discutió previamente, la solicitante experimenta con diferentes arquitecturas de crRNA con diferentes extremos 5'. Un método directo para la creación de crRNA maduro de una secuencia definida es la transcripción *in vitro* utilizando kits de transcripción T7 comercialmente disponibles.

3. Transcribir *in vitro* ARN que alberga secuencias diana de crRNA: plásmidos de expresión de T7 se clonan para albergar únicamente secuencias fijadas como objetivo por el espaciador endógeno de cada uno de los locus CRISPR/Cmr. Esta diana se clonará en el medio de un trozo de ARN de ~ 1 kb, de modo que los productos de escisión resultantes serán fácilmente visualizados en un gel de agarosa.

5 4. Realizar y optimizar el experimento de escisión *in vitro* del lisado de células de mamífero: una vez que todos los componentes están preparados, la solicitante combina e incuba los lisados de células, crRNAs individuales y ARN diana, seguido de la visualización en un gel de agarosa. Sistemas de CRISPR funcionales se identifican sobre la base de productos de escisión del tamaño correcto. Se tienen en cuenta determinados parámetros. Los parámetros para la optimización son la identidad y la concentración de proteína Cmr, la arquitectura y la concentración de crRNA. Parámetros menos significativos para la optimización son el contenido y la concentración de tampón, la estabilidad de crRNA y ARN, tiempo de reacción y la visualización de ARN.

15 La solicitante demuestra y caracteriza el desmontaje de genes de mamíferos endógenos utilizando nuevos crRNAs. Los crRNAs están modificados genéticamente para fijar como objetivo los genes de mamíferos endógenos y cotransfectados con plásmidos de expresión de Cmr en células de mamífero. El desmontaje de ARNm y proteínas se cuantificó mediante PCR cuantitativa y transferencia Western. La capacidad de multiplexación se investiga mediante la fijación como objetivo de múltiples genes al mismo tiempo. La especificidad se caracteriza y se compara con ARNi utilizando ARN-seq. La metodología es la siguiente:

20 1. crRNAs diana de genes de mamíferos endógenos: Con loci CRISPR/Cmr fijados como objetivo a ARN funcionalmente validados, la solicitante diseña nuevos crRNAs que fijan como objetivo tres genes que son dianas de ARNi terapéuticamente validadas, a saber, P53, VEGF, y CTNNB1 (Beta-catenina). La solicitante recoge ocho sitios diana dentro de cada uno de los genes para ayudar a resolver cualquier limitación específica para la secuencia de la tecnología. Para diferentes loci CRISPR/Cmr la vista exacta de la diana puede variar en unas pocas bases debido a las características únicas requeridas de cada uno de los crRNA, pero la solicitante fija como objetivo una región constante para ayudar en las comparaciones.

25 2. Cuantificar y caracterizar el desmontaje de ARNm: la solicitante somete a ensayo los sistemas de CRISPR/Cmr en células HEK 293FT transfectando plásmidos de expresión de mamífero para componentes proteicos Cmr y crRNAs individuales. ARNm escindidos se degradan rápidamente en la célula, mientras que los fragmentos de ARN resultantes no tienen tanto un casquete 5' y una cola poliA. La solicitante cuantifica el grado de desmontaje del ARNm mediante la realización de una PCR cuantitativa. Cuando se observa un desmontaje de ARNm la solicitante realiza una transferencia Western para la proteína fijada como objetivo.

35 Además de simples experimentos de transfección de crRNA, la solicitante añade múltiples crRNAs que fijan como objetivo los mismos genes. Se observó previamente que otros moduladores de la transcripción actúan de forma sinérgica cuando se fijan como objetivo a múltiples sitios dentro de la misma región. Esto también es una estrategia útil para identificar sistemas de CRISPR/CMR subóptimos que no funcionan cuando se utilizan crRNAs individuales, pero que pueden funcionar en general y sólo necesitan ser optimizados por una razón u otra. La solicitante también suma múltiples crRNAs que fijan como objetivo diferentes genes en la misma condición para seguir considerando la multiplexación en este sistema. Esto tiene amplias implicaciones para la modificación por ingeniería genética de la vía y rastreos a escala del genoma.

40 3. Caracterizar especificidad por RNA-seq: una motivación principal de este trabajo es el desarrollo de una plataforma de fijar como objetivo ARN con mayor especificidad que el ARNi. Inherentemente hay una mayor oportunidad para que el sistema CRISPR/Cmr tenga una mayor especificidad debido a la complementariedad extendida al sitio diana. Sin embargo, cada par de bases dentro de la guía puede no ser necesario o puede haber cierta tolerancia a apareamientos erróneos. Para entender estas características del sistema CRISPR/Cmr la solicitante utiliza una línea celular derivada de HEK 293T que expresa una sola copia de EGFP. La solicitante fija como objetivo EGFP para el desmontaje, utilizando 24 crRNAs únicos. El uso de una línea celular HEK/GFP hace que el análisis fuera del objetivo sea sencillo, puesto que EGFP se puede considerar una proteína inocua que no tiene influencia alguna sobre los procesos celulares, por lo que es probable que cualquier transcrito alterado son efectos fuera del objetivo. Los resultados se comparan con un siRNA que fija como objetivo EGFP con efectos fuera de objetivo conocidos. Debido a la capacidad de apareamiento de bases extendida del complejo de CRISPR/Cmr, la solicitante espera que se reduzcan las dianas fuera de objetivo. La tolerancia a apareamientos erróneos también se explora y esto será importante para el desarrollo futuro de la plataforma de CRISPR/Cmr.

Teniendo en cuenta que un sistema CRISPR/Cmr que fija como objetivo ARN funciona en células de mamífero, la solicitante persiguen inicialmente dos aplicaciones: rastreos a escala del genoma y una plataforma de unión a ARN catalíticamente inactiva. Un rastreo de RNAi a escala del genoma es un poderoso método para sondear la biología.

5 Sin embargo, los efectos fuera de objetivo limitan seriamente el enfoque. Una plataforma con mayor especificidad podría reemplazar el rastreo genómico de RNAi o proporcionar un enfoque ortogonal que ayudaría en la identificación de los verdaderos positivos. Una plataforma de CRISPR/Cmr que fija como objetivo ARN es susceptible de rastreos genómicos, porque cada uno de los objetivos se programa por sólo unos pocos nucleótidos, con lo cual se podría sintetizar y suministrar fácilmente un banco complejo.

10 Una plataforma de unión al ARN catalíticamente inactiva es una tecnología que permite el desarrollo de complejos de efectores que ayudan en el sondeo de ARN. La solicitante identifica los componentes con actividad catalítica del complejo Cmr y muta los residuos específicos responsables. Esto se ha hecho con éxito en un cierto número de casos, p. ej., esto se hizo para CRISPR/Cas9 para crear un nicasa y un modulador transcripcional. Utilizando esta plataforma la solicitante desarrolla una tecnología para detectar y representar en imágenes transcritos de ARN dentro de la célula. Esto tiene aplicaciones para la comprensión de la influencia de la ubicación del ARN en los procesos celulares y la detección *in vivo* y la cuantificación de ARN. Alternativamente, la solicitante añade moléculas efectoras en una plataforma de unión al ARN hacia el sondeo y la comprensión de las áreas de la biología del ARN que no son bien entendidas, a saber, la edición de ARN y la epigenética.

15 Aunque se han mostrado y descrito en esta memoria realizaciones preferidas de la presente invención, resultará obvio para los expertos en la técnica de modo que tales realizaciones se proporcionan a modo de ejemplo solamente. Numerosas variaciones, cambios y sustituciones se les ocurrirán a los expertos en la técnica sin apartarse de la invención. Debe entenderse que diversas alternativas a las realizaciones de la invención descrita en esta memoria se pueden emplear en la práctica de la invención.

20 Referencias:

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, 315(5819), 1709-12. doi:10.1126/science.1138140

25 Birmingham, A., Anderson, E. M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Khvorova, A. (2006). 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nature methods*, 3(3), 199-204. doi:10.1038/nmeth854

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science (New York, N.Y.)*, 321(5891), 960-4. doi:10.1126/science.1159689

30 Burnett, J. C., Rossi, J. J., & Tiemann, K. (2011). Current progress of siRNA/shRNA therapeutics in clinical trials. *Biotechnology journal*, 6(9), 1130-46. doi:10.1002/biot.201100054

Cheong, C.-G., y Hall, T. M. T. (2006). Engineering RNA sequence specificity of Pumilio repeats. *Proceedings of the National Academy of Sciences of the United States of America*, 103(37), 13635-9. doi:10.1073/pnas.0606294103

35 Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (New York, N.Y.)*. doi: 10.1126/science.1231143

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., y Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, 411(6836), 494-8. doi:10.1038/35078107

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., y Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669), 806-11. doi:10.1038/35888

40 Hale, C., Kleppe, K., Terns, R. M., y Terns, M. P. (2008). Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA (New York, N.Y.)*, 14(12), 2572-9. doi:10.1261/rna.1246808

Hale, C. R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S. Terns, M. P. (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Molecular cell*, 45(3), 292-302. doi: 10.1016/j.molcel.2011.10.023

- Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, M. P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, 139(5), 945-956.
- 5 Hou, Z., Zhang, Y., Propson, N. E., Howden, S. E., Chu, L.-F., Sontheimer, E. J., y Thomson, J. A. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proceedings of the National Academy of Sciences of the United States of America*, 110(39), 15644-9. doi:10.1073/pnas.1313587110
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., y Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology*, 169(12), 5429-33.
- 10 Jackson, A. L., Bartz, S. R., Schelter, J., Kobayashi, S. V., Burchard, J., Mao, M., Linsley, P. S. (2003). Expression profiling reveals off-target gene regulation by RNAi. *Nature biotechnology*, 21(6), 635-7. doi:10.1038/nbt831
- Jackson, A. L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J., Linsley, P. S. (2006). Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA (New York, N.Y.)*, 12(7), 1197-205. doi:10.1261/rna.30706
- 15 Jansen, R., Embden, J. D. A. van, Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology*, 43(6), 1565-75.
- Konermann, S., Brigham, M. D., Trevino, A. E., Hsu, P. D., Heidenreich, M., Cong, L., Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. *Nature*, 500(7463), 472-6. doi:10.1038/nature12466
- 20 Mackay, J. P., Font, J., y Segal, D. J. (2011). The prospects for designer single-stranded RNA-binding proteins. *Nature structural & molecular biology*, 18(3), 256-61. doi:10.1038/nsmb.2005
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Koonin, E. V. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology*, 9(6), 467-77. doi:10.1038/nrmicro2577
- 25 Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*, 339(6121), 823-6. doi:10.1126/science.1232033
- Marraffini, L. A., y Sontheimer, E. J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature reviews. Genetics*, 11(3), 181-90. doi:10.1038/nrg2749
- Mojica, F. J., Diez-Villasenor, C., Soria, E., y Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology*, 36(1), 244-6.
- 30 Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J., y Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, 60(2), 174-82. doi:10.1007/s00239-004-0046-3
- Peabody, D. S. (1993). The RNA binding site of bacteriophage MS2 coat protein. *The EMBO journal*, 12(2), 595-600. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., y Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5), 1173-83. doi:10.1016/j.cell.2013.02.022
- 35 Sapranaukas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., y Siksny, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic acids research*, 39(21), 9275-82. doi:10.1093/nar/gkr606
- 40 SenGupta, D. J., Zhang, B., Kraemer, B., Pochart, P., Fields, S., y Wickens, M. (1996). A three-hybrid system to detect RNA-protein interactions in vivo. *Proceedings of the National Academy of Sciences*, 93(16), 8496-8501. doi:10.1073/pnas.93.16.8496

Valencia-Burton, M., McCullough, R. M., Cantor, C. R., y Broude, N. E. (2007). RNA visualization in live bacterial cells using fluorescent protein complementation. *Nature methods*, 4(5), 421-7. doi:10.1038/nmeth1023

Wang, X., McLachlan, J., Zamore, P. D., y Hall, T. M. T. (2002). Modular Recognition of RNA by a Human Pumilio-Homology Domain. *Cell*, 110(4), 501-512.

- 5 Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., White, M. F. (2012). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Molecular cell*, 45(3), 303-13. doi:10.1016/j.molcel.2011.12.013.

REIVINDICACIONES

1. Una composición que se produce de forma no natural o que se produce mediante ingeniería genética, que comprende:
 un sistema de suministro operativamente configurado para suministrar componentes de un complejo CRISPR-Cas o
 5 secuencias de polinucleótidos que comprenden o codifican dichos componentes en una célula eucariota,
 en donde dicho complejo CRISPR-Cas es operativo en la célula eucariota;
 I. una o más secuencias de polinucleótidos de complejo CRISPR-Cas que comprenden o codifican la
 expresión en la célula eucariota:
 (a) una secuencia guía capaz de hibridarse a una secuencia diana en una célula eucariota,
 10 (b) una secuencia de emparejamiento tracr, y
 (c) una secuencia tracr, y
 II. una enzima CRISPR que comprende uno o más dominios funcionales heterólogos o un polinucleótido
 que codifica una enzima de CRISPR que comprende uno o más dominios funcionales heterólogos para la expresión
 en la célula eucariota;
 15 en donde
 la secuencia de emparejamiento tracr se hibrida a la secuencia tracr,
 la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia
 diana,
 el complejo de CRISPR comprende la enzima de CRISPR formando complejo con (1) la secuencia guía
 20 que se hibrida a la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia
 tracr,
 la enzima CRISPR comprende una o más mutaciones, de manera que la enzima tiene actividad de
 nucleasa alterada en comparación con la enzima de tipo salvaje,
 en donde los uno o más dominios funcionales heterólogos comprende una o más Señales de Localización
 25 Nuclear NLS(s).
2. La composición de acuerdo con la reivindicación 1, en donde el sistema de suministro comprende un sistema de
 vector que comprende uno o más vectores, y en donde el componente (I) comprende un primer elemento regulador
 enlazado operativamente a una secuencia de polinucleótidos que comprende la secuencia guía, la secuencia de
 emparejamiento tracr y la secuencia tracr, y en donde el componente (II) comprende un segundo elemento regulador
 30 enlazado operativamente a una secuencia de polinucleótidos que codifica la enzima de CRISPR.
3. La composición de acuerdo con la reivindicación 1, en donde el sistema de suministro comprende un sistema de
 vector que comprende uno o más vectores, y en donde el componente (I) comprende un primer elemento regulador
 enlazado operativamente a la secuencia guía y la secuencia de emparejamiento tracr, y un tercer elemento
 regulador enlazado operativamente a la secuencia tracr, y en donde el componente (II) comprende un segundo
 35 elemento regulador enlazado operativamente a una secuencia de polinucleótidos que codifica la enzima de CRISPR.
4. La composición de acuerdo con una cualquiera de las reivindicaciones 2 ó 3, en donde los componentes I y II
 están ubicados en el mismo o en diferentes vectores del sistema.
5. La composición de acuerdo con la reivindicación 4, en donde los componentes I y II están ubicados en el mismo
 vector.
- 40 6. Una composición de acuerdo con una cualquiera de las reivindicaciones precedentes, que es una composición
 multiplexada que comprende múltiples secuencias guía capaces de hibridarse a múltiples secuencias diana.
7. La composición de acuerdo con una cualquiera de las reivindicaciones 2 a 6, en donde el uno o más vectores
 comprenden uno o más vectores virales.
- 45 8. Una composición de acuerdo con la reivindicación 7, en donde el uno o más vectores virales comprenden uno o
 más vectores de retrovirus, lentivirus, adenovirus, virus adeno-asociados o virus herpes simplex.
9. La composición de acuerdo con una cualquiera de las reivindicaciones 1 a 6, en donde el sistema de suministro
 comprende un sistema de levadura, un sistema de lipofección, un sistema de microinyección, un sistema biolístico,
 virosomas, liposomas, inmunoliposomas, policones, conjugados de lípido:ácido nucleico o viriones artificiales.
- 50 10. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la enzima de
 CRISPR comprende una o más mutaciones en dos o más dominios catalíticamente activos.

11. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la enzima de CRISPR tiene una actividad nucleasa reducida o suprimida en comparación con la enzima de tipo salvaje.
12. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la una o más mutaciones comprenden una mutación en D10 SpCas9, H840 SpCas9, N854 SpCas9 o N863 SpCas9, o residuos correspondientes de otras enzimas de CRISPR.
13. La composición de acuerdo con la reivindicación 12, en donde la una o más mutaciones comprenden D10A, H840A, N854A o N863A.
14. La composición de acuerdo con una cualquiera de las reivindicaciones 12 ó 13, en donde la una o más mutaciones comprenden dos mutaciones.
15. La composición de acuerdo con la reivindicación 14, en donde la mutación comprende D10A SpCas9 y H840A SpCas9, o residuos correspondientes de otras enzimas de CRISPR.
16. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la enzima CRISPR es una proteína de unión a ADN que no dirige la escisión de cualquier cadena en la ubicación de la secuencia diana.
17. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la una o más mutaciones es en un dominio catalíticamente activo de la enzima de CRISPR que comprende RuvCI, RuvCII o RuvCIII.
18. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la enzima CRISPR comprende dos o más dominios funcionales heterólogos.
19. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, que comprende al menos dos o más NLSs.
20. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde dichas al menos una o más o al menos dos o más secuencias de localización nuclear se encuentran en o próximas al extremo amino de la enzima y/o en o próximas al extremo carboxi de la enzima.
21. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde un dominio funcional comprende un dominio de activación transcripcional.
22. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde un dominio funcional comprende VP64.
23. La composición de acuerdo con una cualquiera de las reivindicaciones 1 a 20, en donde un dominio funcional es un dominio represor transcripcional.
24. La composición de acuerdo con la reivindicación 23, en donde el dominio funcional comprende un dominio KRAB o un dominio SID.
25. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la secuencia que codifica la enzima codifica dos o más dominios funcionales heterólogos.
26. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la secuencia que codifica la enzima codifica uno o más dominios funcionales heterólogos fusionados a la enzima CRISPR.
27. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la secuencia que codifica la enzima comprende una secuencia de enlazador entre cualquiera de dos dominios.
28. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde uno o más dominios funcionales tienen una o más de las siguientes actividades: actividad metilasa, actividad desmetilasa, actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de la transcripción, actividad de modificación de histonas, actividad de escisión del ARN y actividad de unión a ácido nucleico.

29. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde un dominio funcional se une a ADN.
30. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde un dominio funcional afecta a la transcripción del ácido nucleico diana.
- 5 31. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la célula eucariota comprende una célula de mamífero.
32. La composición de acuerdo con la reivindicación 31, en donde la célula de mamífero comprende una célula humana.
- 10 33. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la enzima CRISPR está optimizada en codones para la expresión en una célula eucariótica.
34. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la enzima CRISPR es una enzima CRISPR Cas9 de tipo II.
- 15 35. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la enzima Cas9 es de un organismo seleccionado del grupo que comprende el género *Streptococcus*, *Campylobacter*, *Nitratifactor*, *Staphylococcus*, *Parvibaculum*, *Roseburia*, *Neisseria*, *Gluconacetobacter*, *Azospirillum*, *Sphaerochaeta*, *Lactobacillus*, *Eubacterium* o *Corynebacter*.
36. La composición de acuerdo con una cualquiera de las reivindicaciones 1-34, en donde la Cas9 es de *Staphylococcus aureus*.
- 20 37. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la enzima CRISPR es una enzima Cas9 que es una proteína Cas9 quimérica que comprende fragmentos de diferentes homólogos de Cas9.
38. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la secuencia diana está adyacente al Motivo Adyacente al Protoespaciador (PAM) reconocido por la enzima CRISPR.
- 25 39. La composición de acuerdo con una cualquiera de las reivindicaciones 2 a 38, en donde dichos primer y segundo elemento reguladores comprende cada uno un promotor específico para tejidos.
40. La composición de acuerdo con la reivindicación 39, en donde el promotor específico para tejidos dirige la expresión de transcritos CRISPR en sangre.
41. La composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la secuencia tracr es de 30 o más nucleótidos de longitud.
- 30 42. Un método de modular la expresión génica en un locus genómico de interés en una célula de un organismo eucariota, al poner en contacto la célula con la composición de acuerdo con una cualquiera de las reivindicaciones precedentes, con la condición de que el organismo no sea un ser humano u otro animal.
- 35 43. Un método ex vivo o in vitro de modular la expresión génica en un locus genómico de interés en una célula de un organismo eucariota, al poner en contacto la célula con la composición de acuerdo con una cualquiera de las reivindicaciones 1 a 41; con la condición de que dicho método no comprenda un proceso para modificar la identidad genética de la línea germinal de un ser humano.
44. Uso de una composición de acuerdo con una cualquiera de las reivindicaciones 1 a 41 en la edición de un gen o genoma *ex vivo* o *in vitro*; con la condición de que dicha edición no comprenda un proceso para modificar la identidad genética de la línea germinal de un ser humano.
- 40 45. La composición de acuerdo con una cualquiera de las reivindicaciones 1 a 41, para uso como un producto terapéutico.
46. La composición para uso de acuerdo con la reivindicación 45, en donde dicho producto terapéutico es para la edición de un gen o genoma, o para la terapia génica.

47. Una célula huésped eucariota in vitro o ex vivo, que comprende la composición de acuerdo con una cualquiera de las reivindicaciones 1 a 41; con la condición de que la célula huésped no sea una célula de la línea germinal humana.
- 5 48. Una célula huésped eucariota in vitro o ex vivo, que comprende el complejo de CRISPR-Cas según se define en una cualquiera de las reivindicaciones 1 a 41; con la condición de que la célula huésped no sea una célula de la línea germinal humana.
49. Progenie de la célula huésped de la reivindicación 47 ó 48; con la condición de que la progenie no comprenda un cuerpo humano.

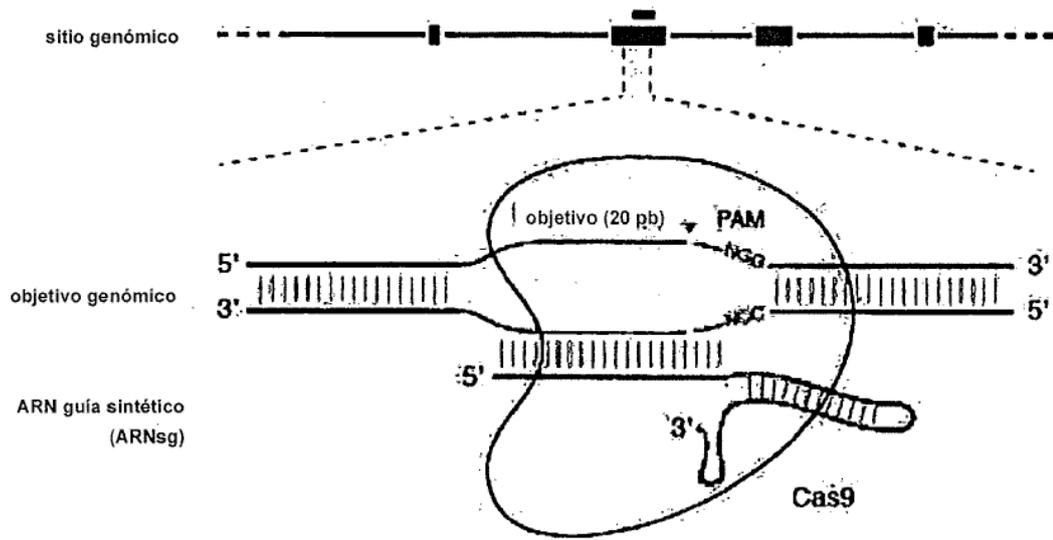


FIG. 1

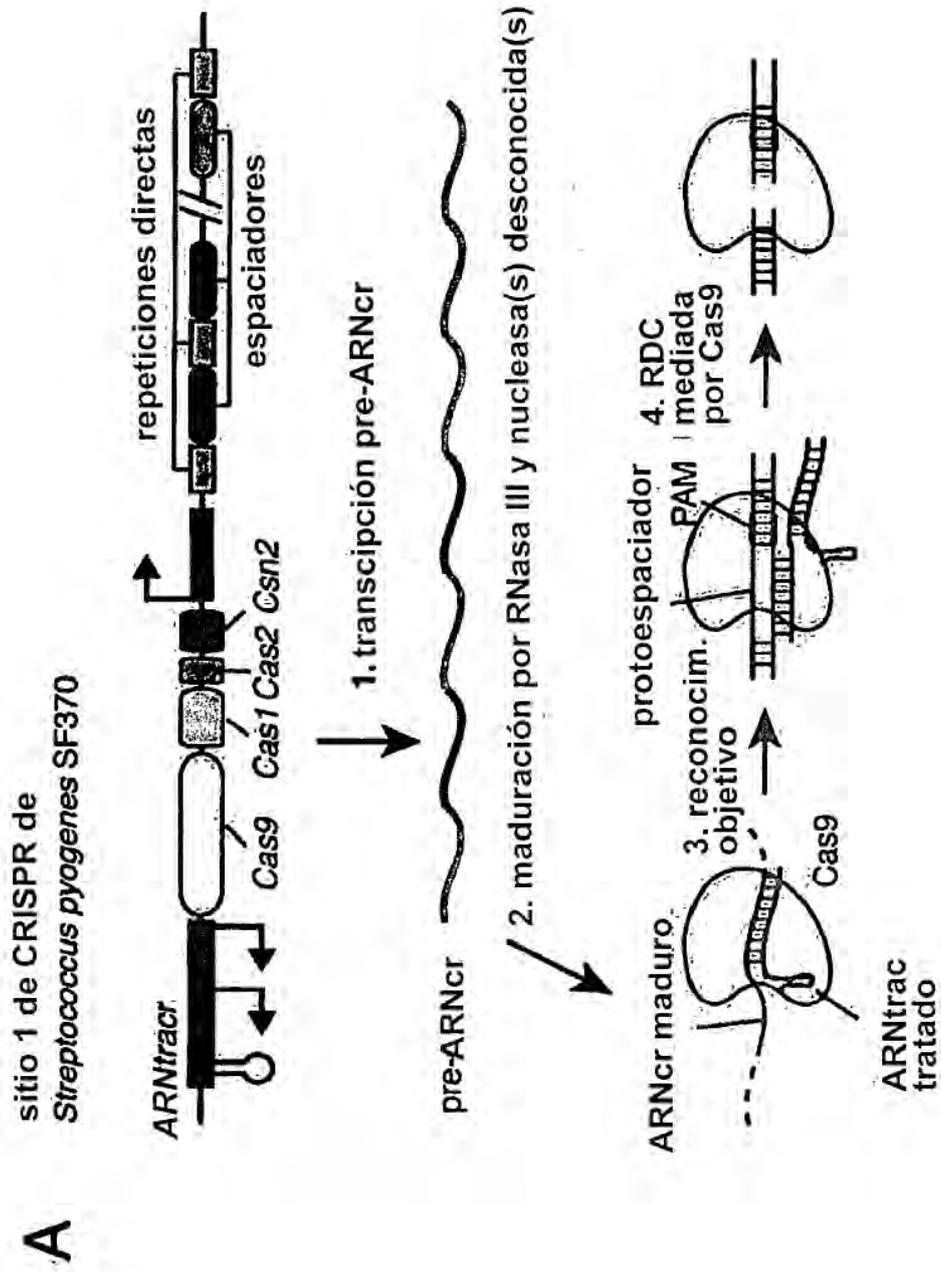


FIG. 2A

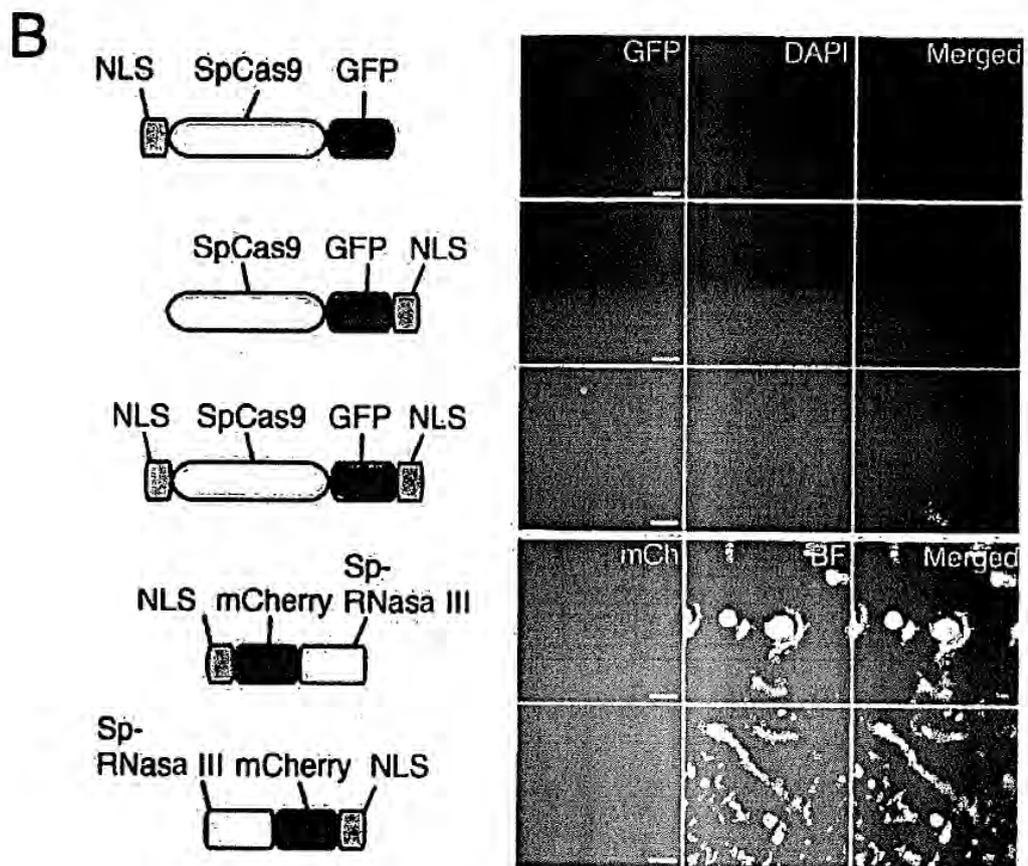


FIG. 2B

C

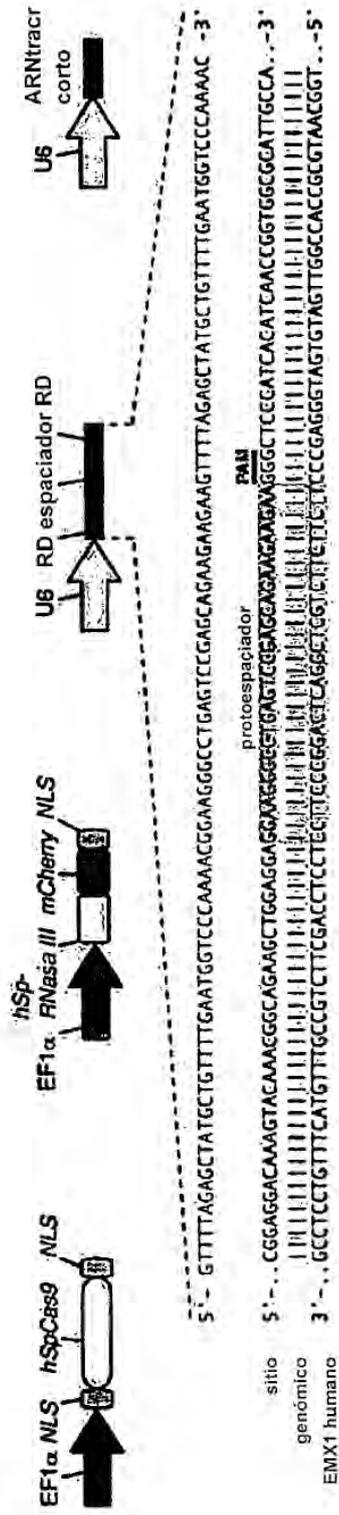
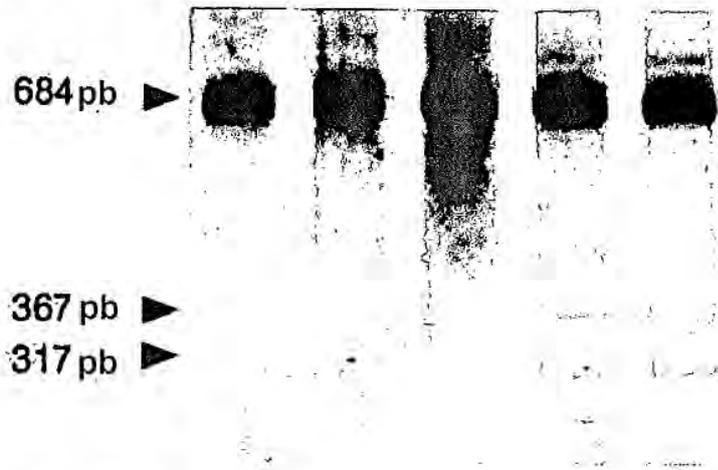


FIG. 2C

D

2xNLS-SpCas9	+	+	+	+	+
SpRNAsa III	-	+	+	-	+
ARNtracr corto	-	+	-	+	+
RD-EMX1-RD	+	-	+	+	+



inserción o supresión (%):

4,7 5,0

FIG. 2D

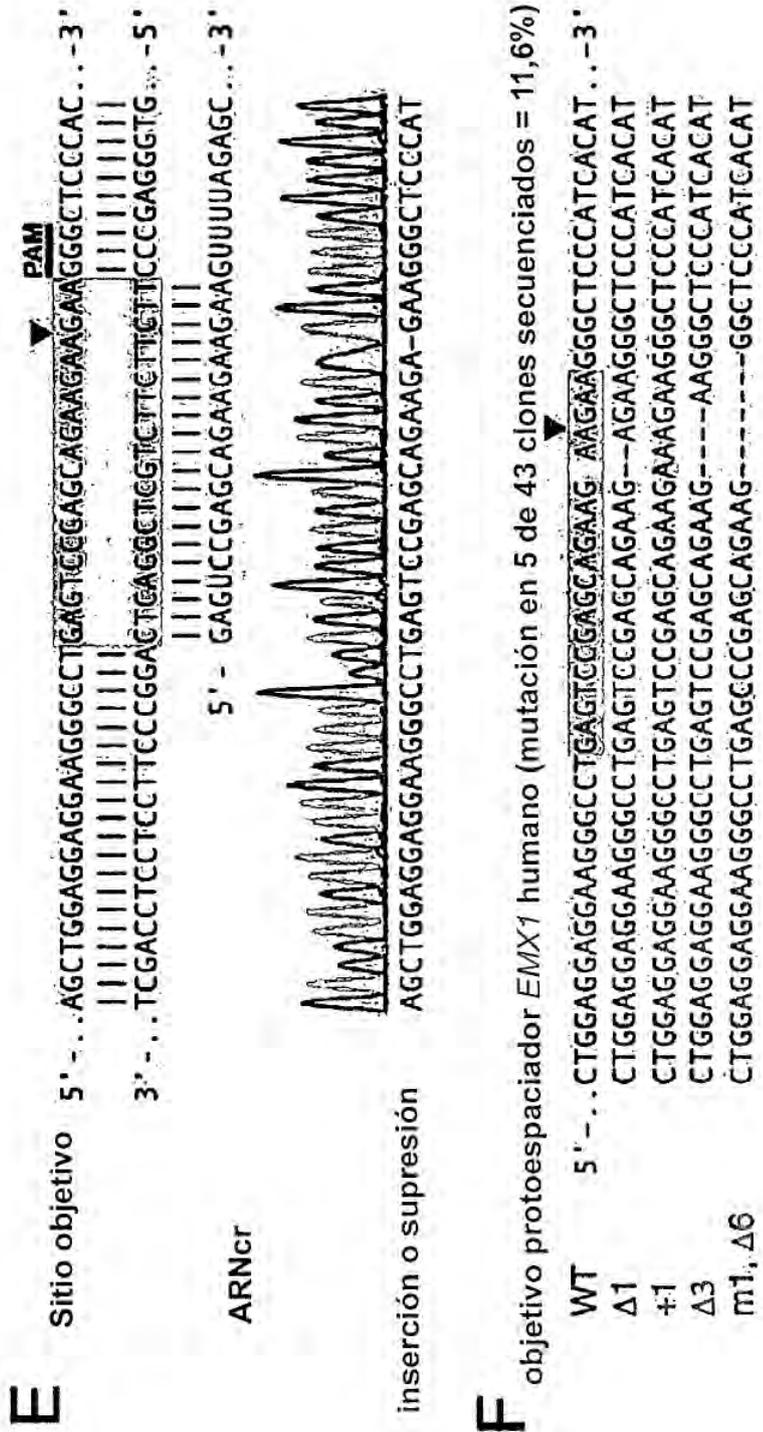


FIG. 2E-F

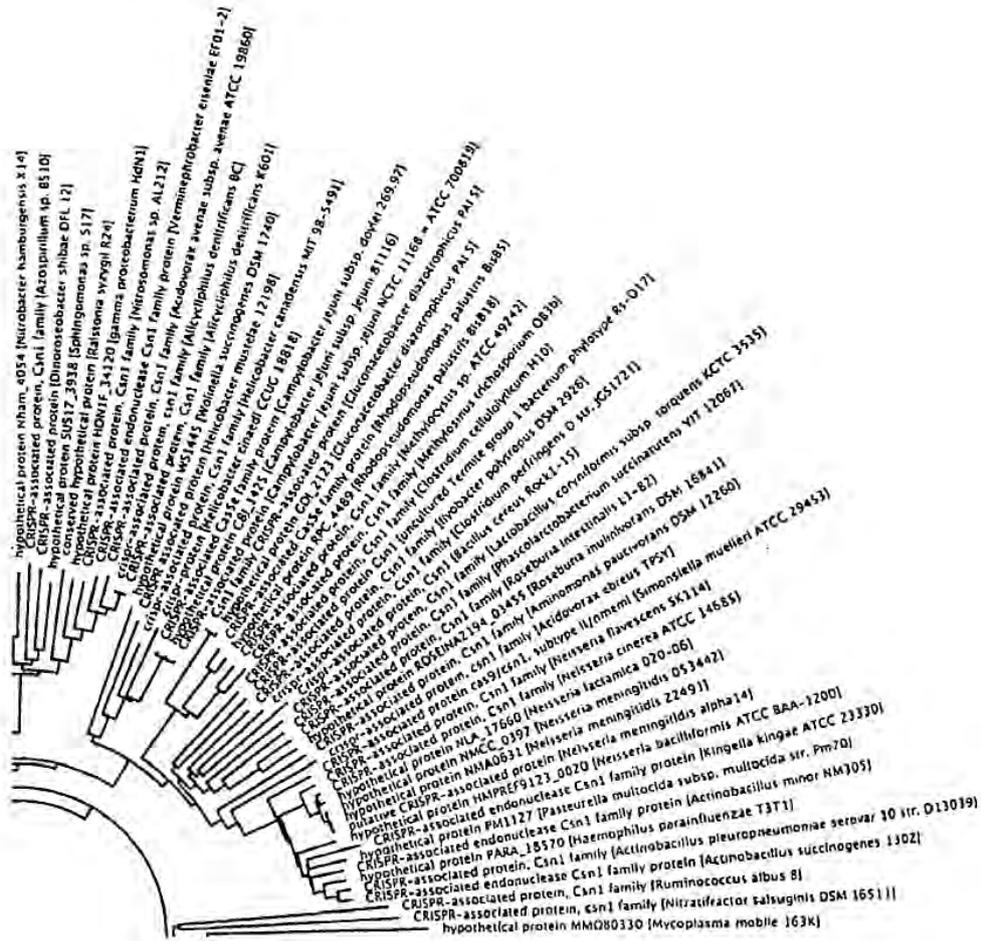


FIG. 3B

hypothetical protein = proteína hipotética; CRISPR-associated protein, Csn1 family = familia Csn1 de proteína asociada a CRISPR;
 CRISPR-associated endonuclease Csn1 family protein, = proteína de familia Csn1 de endonucleasa asociada a CRISPR;
 CRISPR-associated protein Cas9/Csn1, subtype II = proteína Cas9/Csn1 asociada a CRISPR, subtipo II; CRISPR-associated CasSe
 family protein = proteína de la familia CasSe asociada a CRISPR;

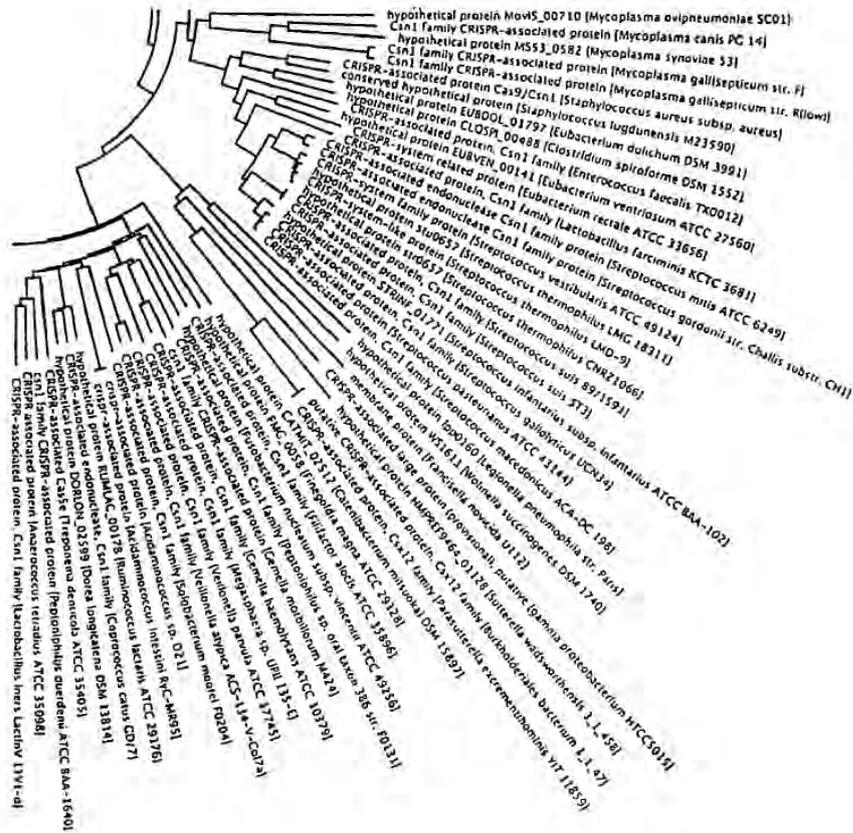


FIG. 3D

Putative CRISPR-associated Protein Csx12 family = familia Csx12 de proteína asociada a CRISPR putativa; hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated Protein Cas9/Csn1, subtype II = Proteína Cas9/Csn1 asociada a CRISPR, subtipo II; conserved hypothetical protein = proteína hipotética conservada; CRISPR-system related protein = proteína relacionada con el sistema CRISPR; CRISPR-system like protein = proteína de tipo sistema CRISPR; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; membrane protein = proteína de membrana; CRISPR-associated large protein (provisional), putative = proteína grande asociada a CRISPR (provisional) putativa; CRISPR-associated CasSe = CasSe asociada a CRISPR.

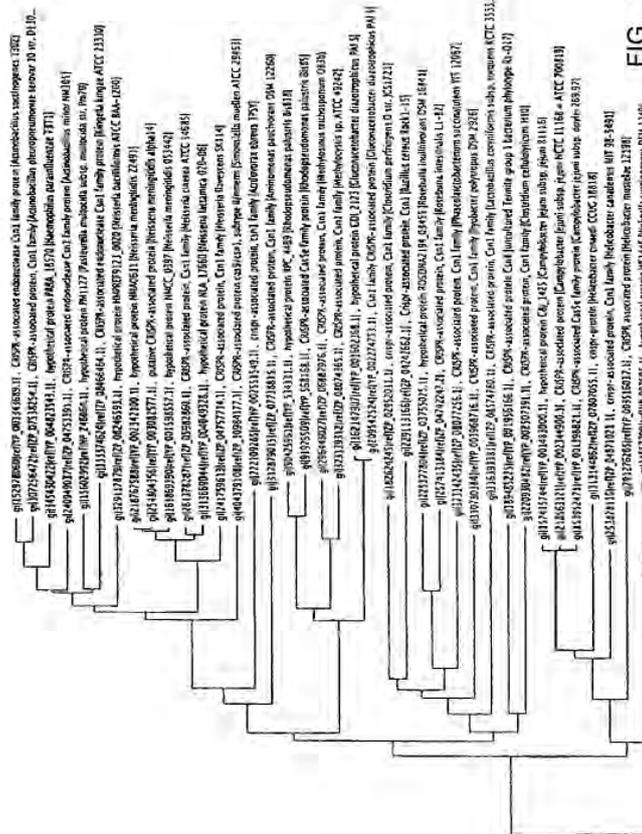


FIG. 4A

Putative CRISPR-associated Protein = proteína asociada a CRISPR putativa; hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated Protein Cas9/Csn1, subtype II = Proteína Cas9/Csn1 asociada a CRISPR, subtipo II; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; CRISPR-associated CasSe family protein = proteína de la familia CasSe asociada a CRISPR

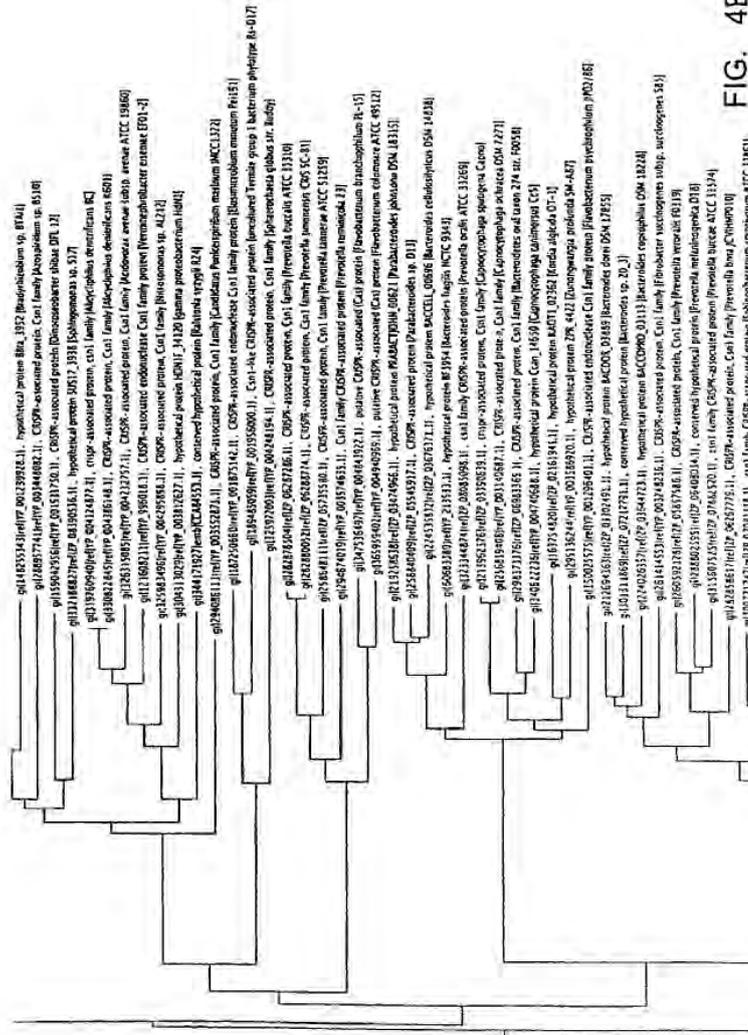


FIG. 4B

Putative CRISPR-associated (Cas) Protein = proteína (Cas) asociada a CRISPR putativa; hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; conserved hypothetical protein = proteína hipotética conservada; Csn1-like CRISPR-associated protein = proteína asociada a CRISPR de tipo Csn1;

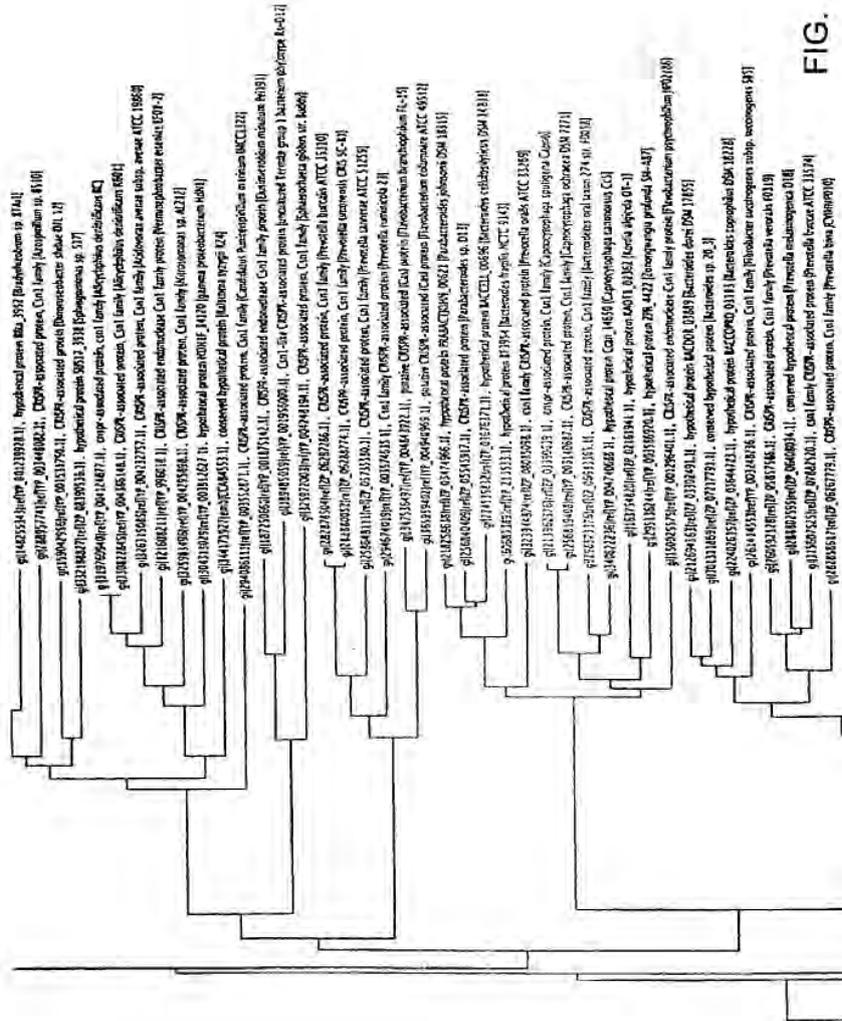


FIG. 4C

Putative CRISPR-associated (Cas) Protein = proteína (Cas) asociada a CRISPR putativa; hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; conserved hypothetical protein = proteína hipotética conservada; Csn1-like CRISPR-associated protein = proteína asociada a CRISPR de tipo Csn1;

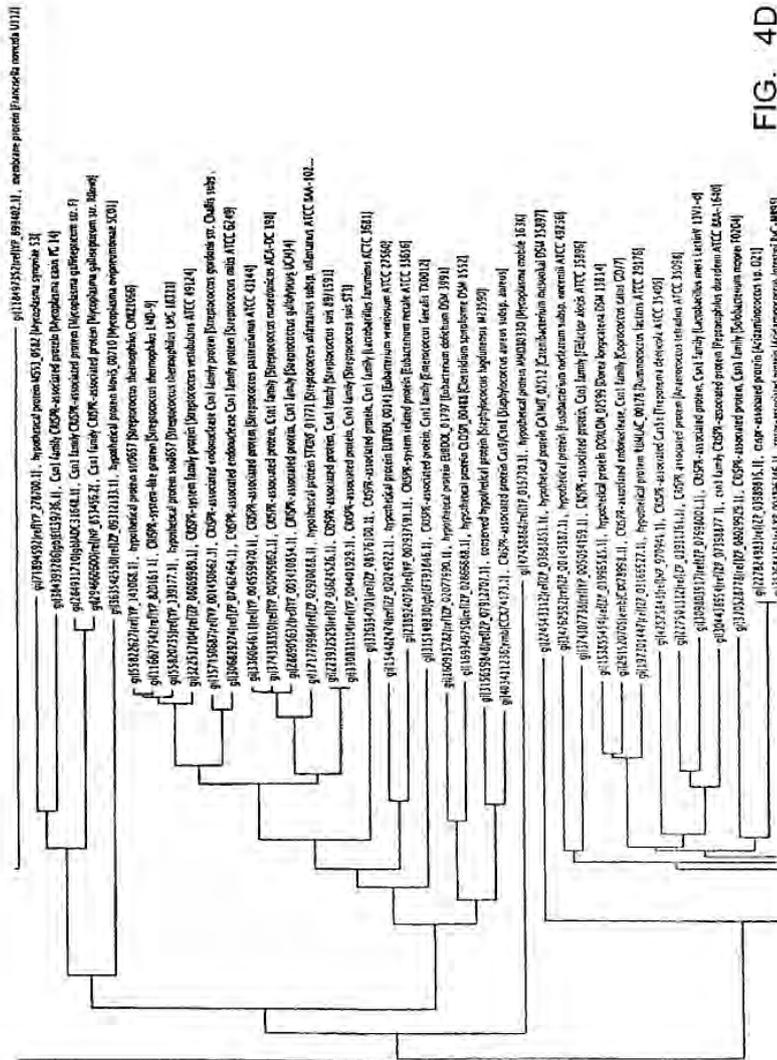


FIG. 4D

Hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; conserved hypothetical protein = proteína hipotética conservada; Csn1-like CRISPR-associated protein = proteína asociada a CRISPR de tipo Csn1; CRISPR-system-like protein = Proteína de tipo sistema CRISPR; CRISPR-associated protein Cas9/Csn1= proteína Cas9/Csn1 asociada a CRISPR

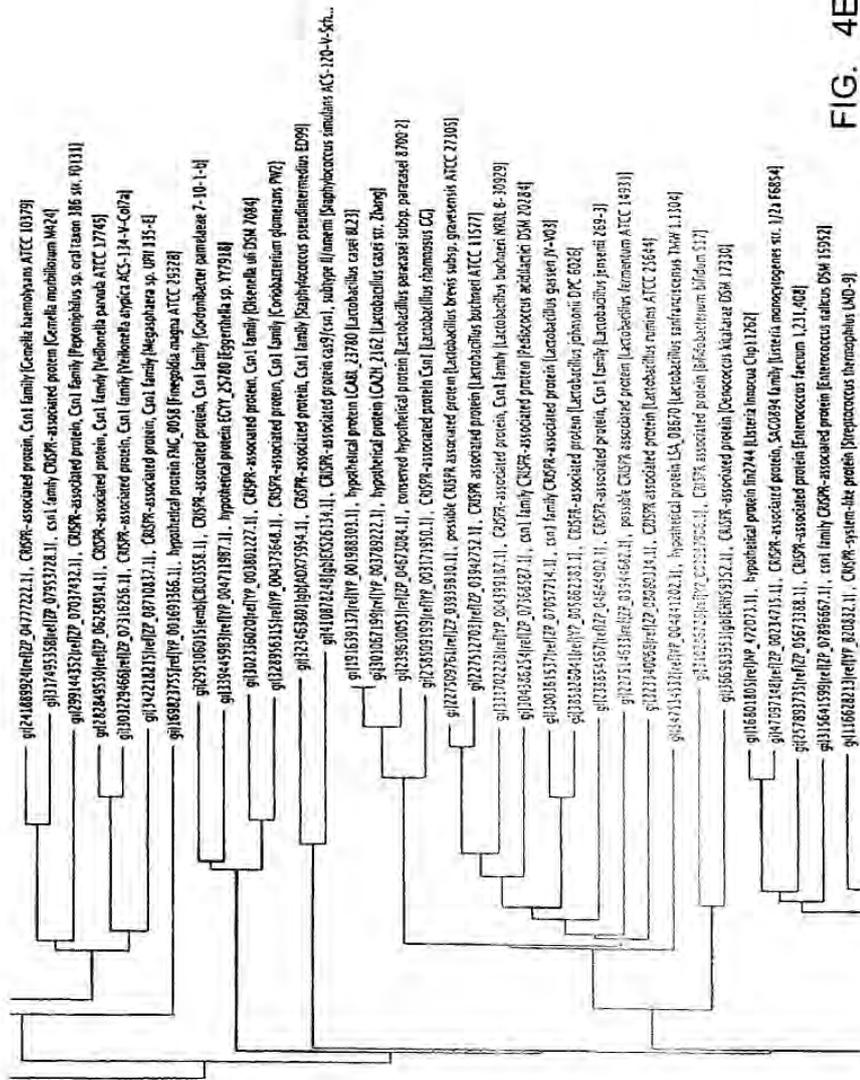


FIG. 4E

Hypothetical protein = proteína hipotética; CRISPR-associated Protein Csn1 family = familia Csn1 Proteína asociada a CRISPR; possible CRISPR-associated protein = Proteína asociada a CRISPR posible; conserved hypothetical protein = proteína hipotética conservada.

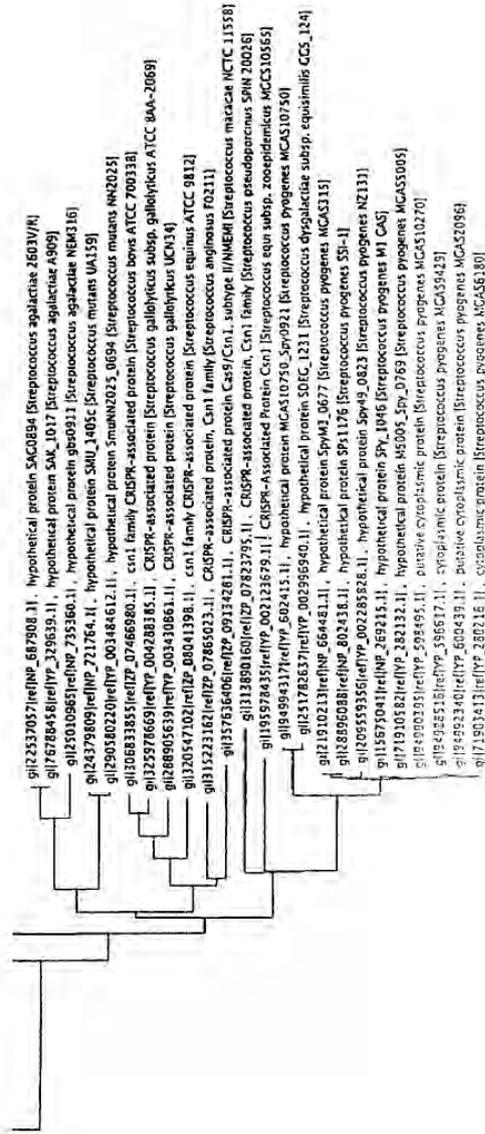


FIG. 4F

Hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated protein Cas9/Csn1, subtype II = proteína Cas9/Csn1 asociada a CRISPR, subtipo II; Putative cytoplasmic protein = proteína citoplasmática putativa; cytoplasmic protein = proteína citoplasmática

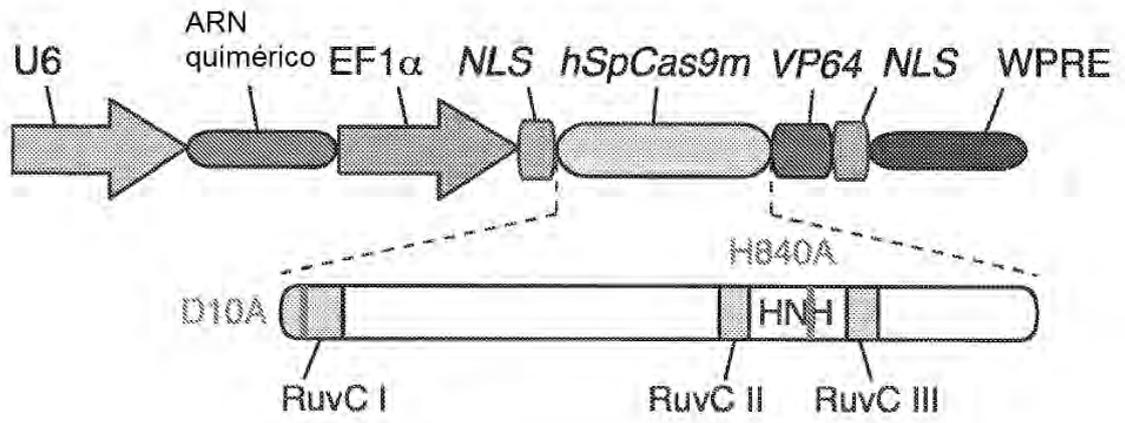


FIG. 5

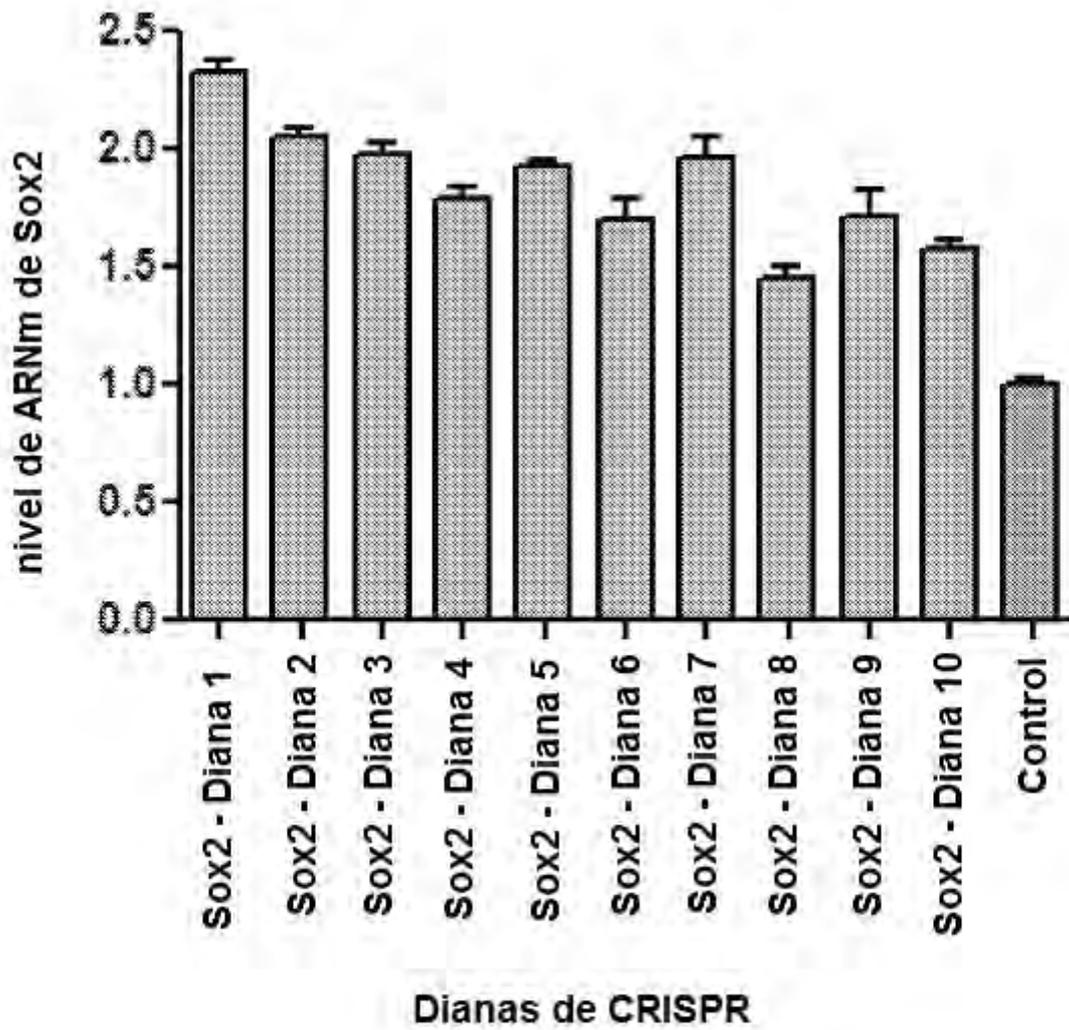


FIG. 6

Figura 1

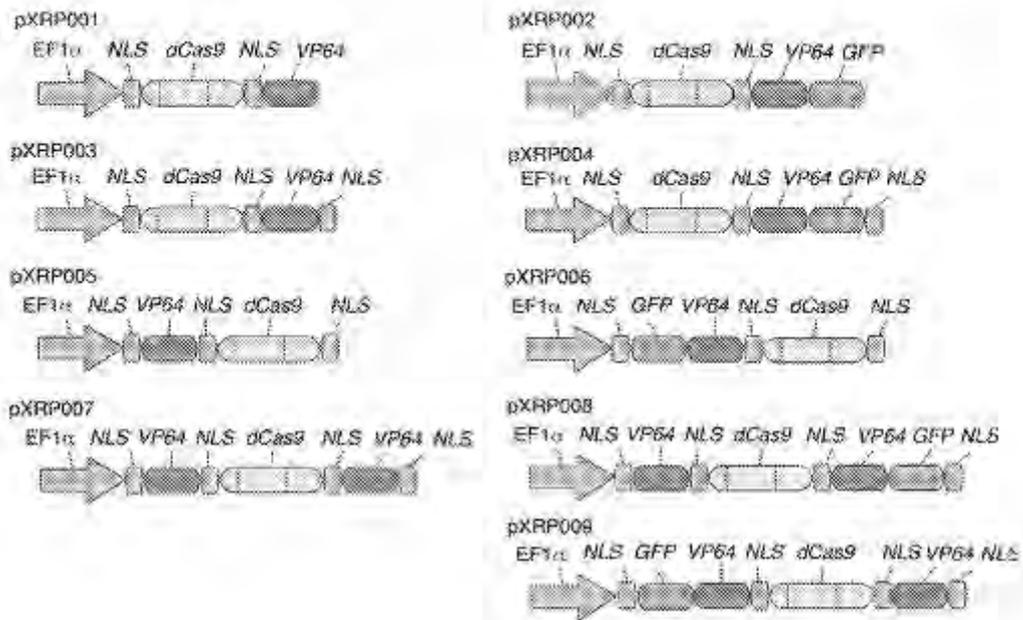
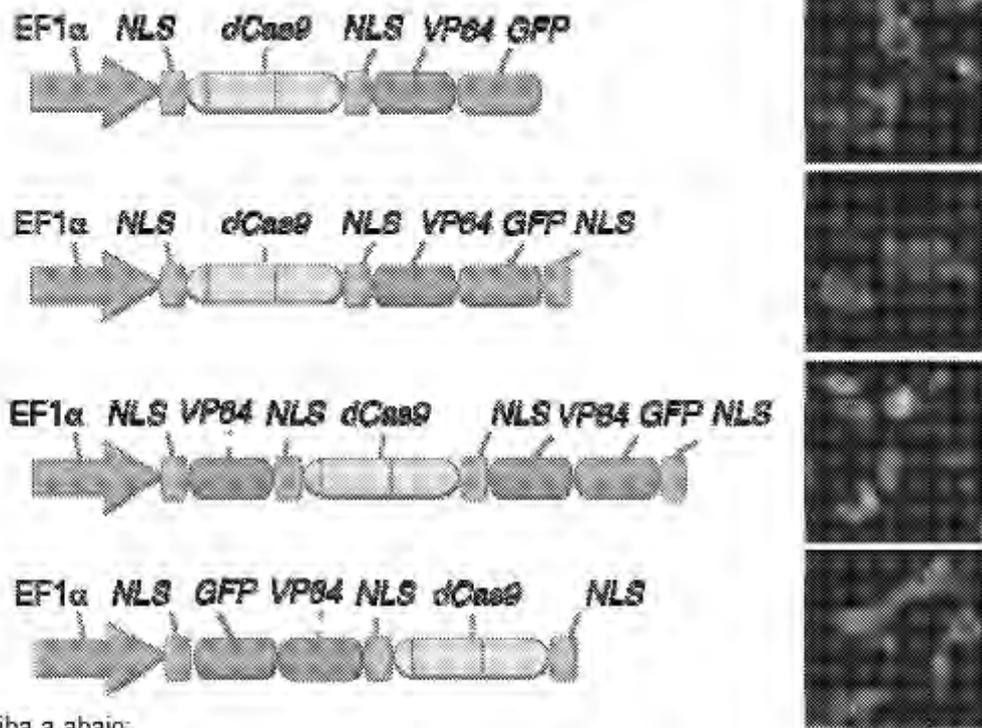


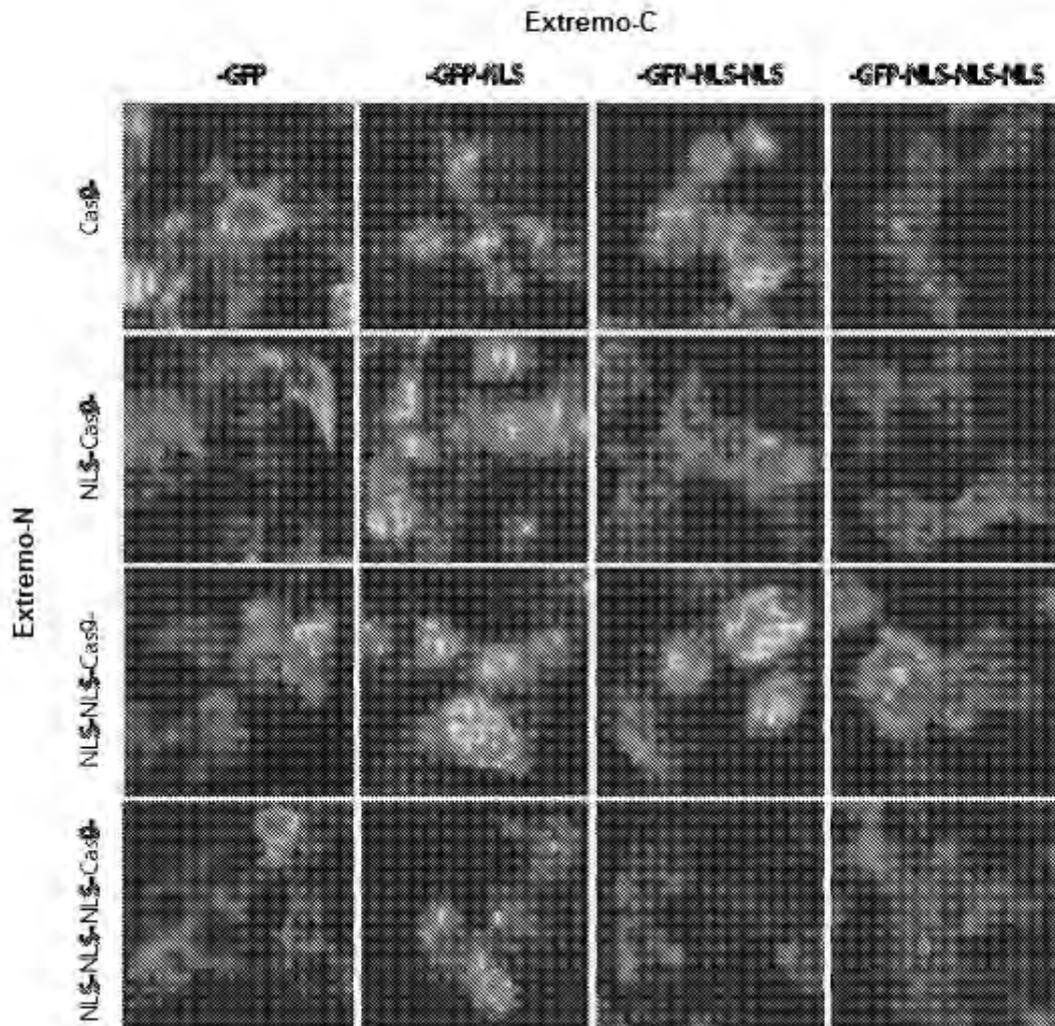
FIG. 7



De arriba a abajo:

- pXRP02- pLenti2-EF1 α -NLS-hSpCsn1(10A,840A)-NLS-VP64-EGFP
- pXRP04- pLenti2-EF1 α -NLS-hSpCsn1(10A,840A)-NLS-VP64-2A-EGFP-NLS
- pXRP06- pLenti2-EF1 α -NLS-EGFP-VP64-NLS-hSpCsn1(10A,840A)-NLS
- pXRP08- pLenti2-EF1 α -NLS-VP64-NLS-hSpCsn1(10A,840A)-NLS-VP64-EGFP-NLS

FIG. 8



Las construcciones tienen los siguientes nombres:

pXRPNLS001	pXRPNLS005	pXRPNLS009	pXRPNLS013
pXRPNLS002	pXRPNLS006	pXRPNLS010	pXRPNLS014
pXRPNLS003	pXRPNLS007	pXRPNLS011	pXRPNLS015
pXRPNLS004	pXRPNLS008	pXRPNLS012	pXRPNLS016

Secuencia de importina alfa NLS:

AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG

FIG. 9

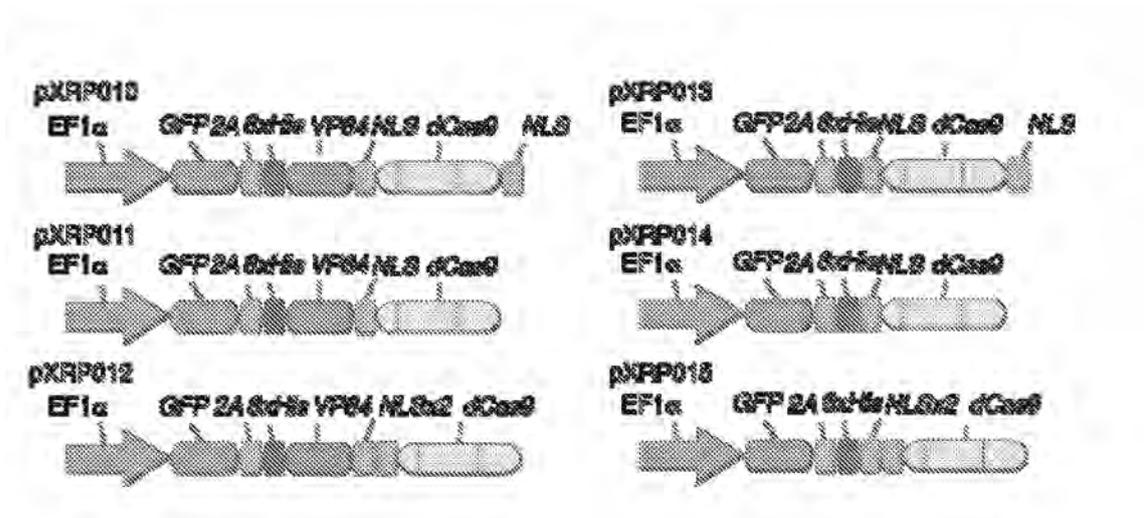
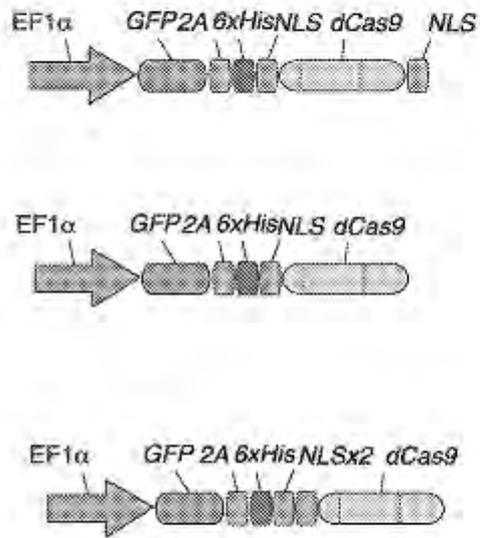


FIG. 10



Rojo = alexa-fluor 594
Azul = DAPI

FIG. 11

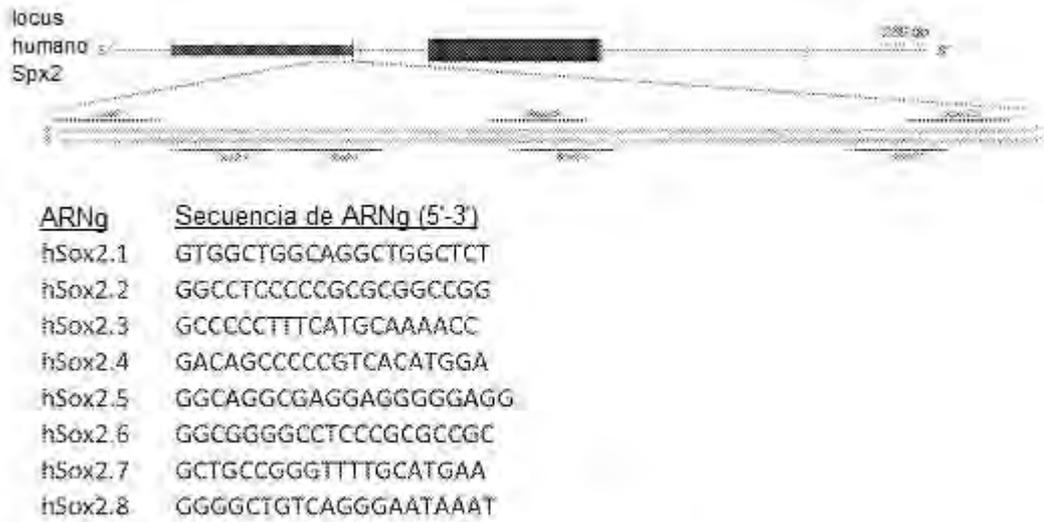
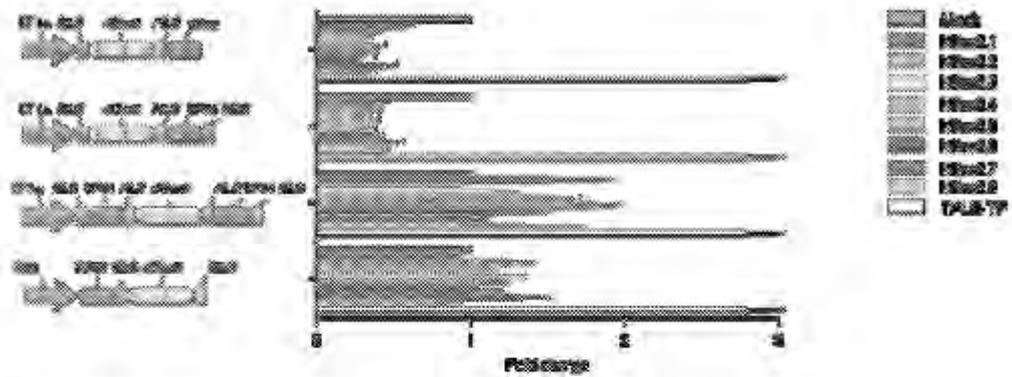


FIG. 12



De arriba a abajo:
 pXRP001
 pXRP005
 pXRP007
 Cbh-VP64-NLS-dCas9-NLS

FIG. 13

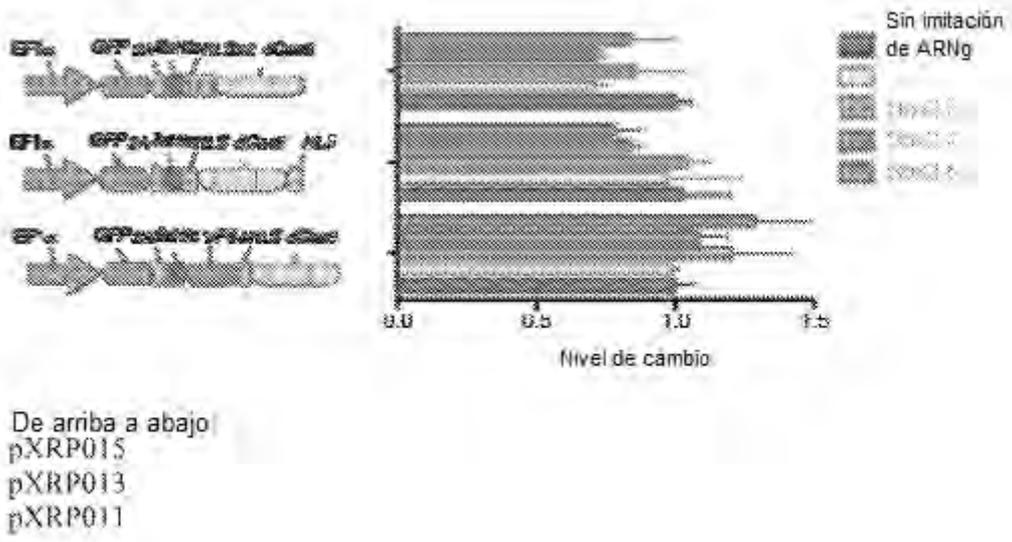


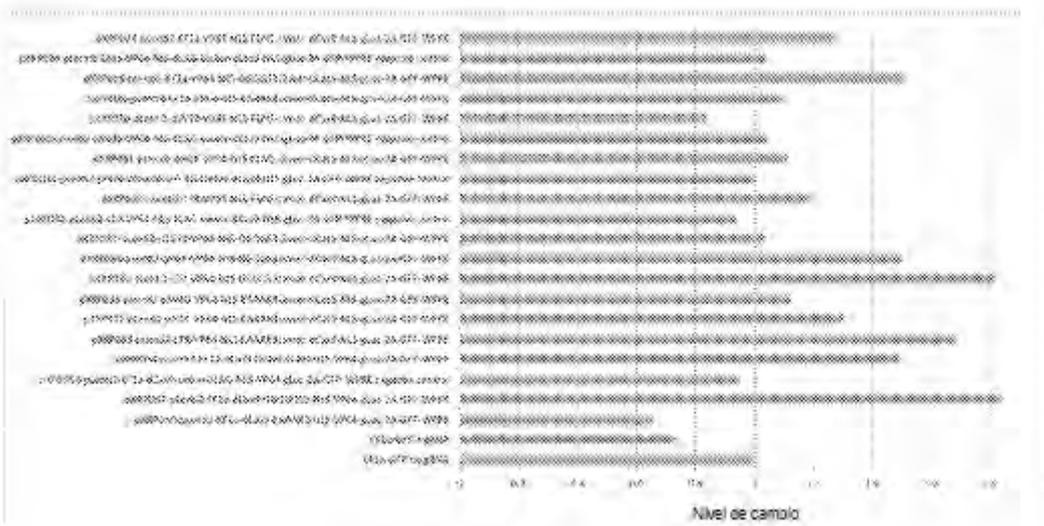
FIG. 14

pXRP024-pLenti2-EF1a-VP64-NLS-FLAG-Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP025-pLenti2-EF1a-VP64-NLS-GGGGS ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP026-pLenti2-EF1a-VP64-NLS-EAAAK ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP027-pLenti2-EF1a-NLS-FLAG-Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP028-pLenti2-EF1a-NLS-GGGGS ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP029-pLenti2-EF1a-NLS-EAAAK ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP030-pLenti2-pSV40-VP64-NLS-FLAG-Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP031-pLenti2-pPGK-VP64-NLS-FLAG-Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP032-pLenti2-LTR-VP64-NLS-FLAG-Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP033-pLenti2-pSV40-VP64-NLS-GGGGS ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP034-pLenti2-pPGK-VP64-NLS-GGGGS ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP035-pLenti2-LTR-VP64-NLS-GGGGS ₂ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP036-pLenti2-pSV40-VP64-NLS-EAAAK ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP037-pLenti2-pPGK-VP64-NLS-EAAAK ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP038-pLenti2-LTR-VP64-NLS-EAAAK ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP048-pLenti2-EF1a-SID4x-NLS-FLAG-Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP049-pLenti2-EF1a-SID4X-NLS-GGGGS ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP050-pLenti2-EF1a-SID4X-NLS-EAAAK ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP051-pLenti2-EF1a-KRAB-NLS-FLAG-Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP052-pLenti2-EF1a-KRAB-NLS-GGGGS ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP053-pLenti2-EF1a-KRAB-NLS-EAAAK ₃ Enlaz-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP054-pLenti2-EF1a-dCas9-Enlaz-FLAG-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP055-pLenti2-EF1a-dCas9-Enlaz-FLAG-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP056-pLenti2-EF1a-dCas9-Enlaz-FLAG-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP057-pLenti2-EF1a-dCas9-GGGGS ₃ -NLS-VP64-gLuc-2A-GFP-WPRE
pXRP058-pLenti2-EF1a-dCas9-GGGGS ₂ -NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP059-pLenti2-EF1a-dCas9-GGGGS ₂ -NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP060-pLenti2-EF1a-dCas9-EAAAK ₃ -NLS-VP64-gLuc-2A-GFP-WPRE
pXRP061-pLenti2-EF1a-dCas9-EAAAK ₃ -NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP062-pLenti2-EF1a-dCas9-EAAAK ₃ -NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP024-pLenti2-EF1a-VP64-NLS-FLAG-Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP025-pLenti2-EF1a-VP64-NLS-GGGGS ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP026-pLenti2-EF1a-VP64-NLS-EAAAK ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP027-pLenti2-EF1a-NLS-FLAG-Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP028-pLenti2-EF1a-NLS-GGGGS ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP029-pLenti2-EF1a-NLS-EAAAK ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE

FIG. 15A

pXRP030-pLenti2-pSV40-VP64-NLS-FLAG- Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP031-pLenti2-pPGK-VP64-NLS-FLAG- Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP032-pLenti2-LTR-VP64-NLS-FLAG- Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP033-pLenti2-pSV40-VP64-NLS-GGGGS ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP034-pLenti2-pPGK-VP64-NLS-GGGGS ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP035-pLenti2-LTR-VP64-NLS-GGGGS ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP036-pLenti2-pSV40-VP64-NLS-EAAAK ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP037-pLenti2-pPGK-VP64-NLS-EAAAK ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP038-pLenti2-LTR-VP64-NLS-EAAAK ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP048-pLenti2-EF1a-SID4X-NLS-FLAG- Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP049-pLenti2-EF1a-SID4X-NLS-GGGGS ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP050-pLenti2-EF1a-SID4X-NLS-EAAAK ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP051-pLenti2-EF1a-KRAB-NLS-FLAG- Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP052-pLenti2-EF1a-KRAB-NLS-GGGGS ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP053-pLenti2-EF1a-KRAB-NLS-EAAAK ₃ Enlaz-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP054-pLenti2-EF1a-Cas9- Enlaz-FLAG-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP055-pLenti2-EF1a-Cas9- Enlaz-FLAG-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP056-pLenti2-EF1a-Cas9- Enlaz-FLAG-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP057-pLenti2-EF1a-Cas9-GGGGS ₃ -NLS-VP64-gLuc-2A-GFP-WPRE
pXRP058-pLenti2-EF1a-Cas9-GGGGS ₃ -NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP059-pLenti2-EF1a-Cas9-GGGGS ₃ -NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP060-pLenti2-EF1a-Cas9-EAAAK ₃ -NLS-VP64-gLuc-2A-GFP-WPRE
pXRP061-pLenti2-EF1a-Cas9-EAAAK ₃ -NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP062-pLenti2-EF1a-Cas9-EAAAK ₃ -NLS-KRAB-gLuc-2A-GFP-WPRE

FIG. 15B



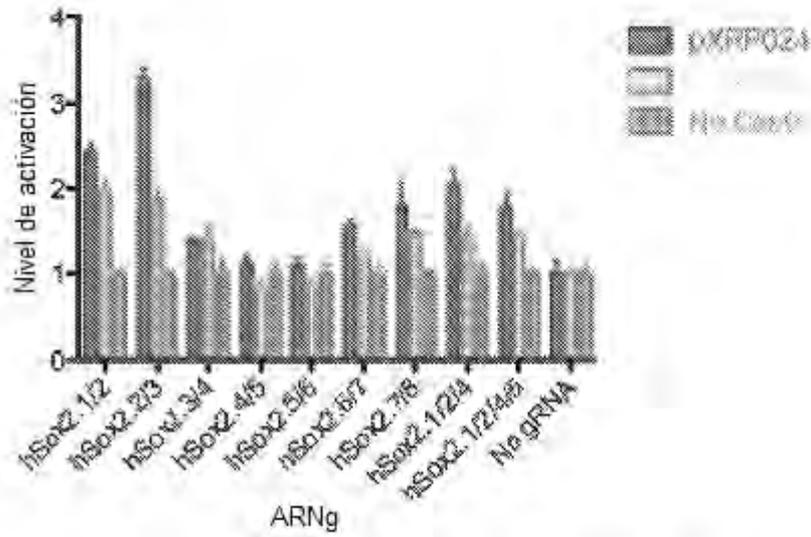


FIG. 17

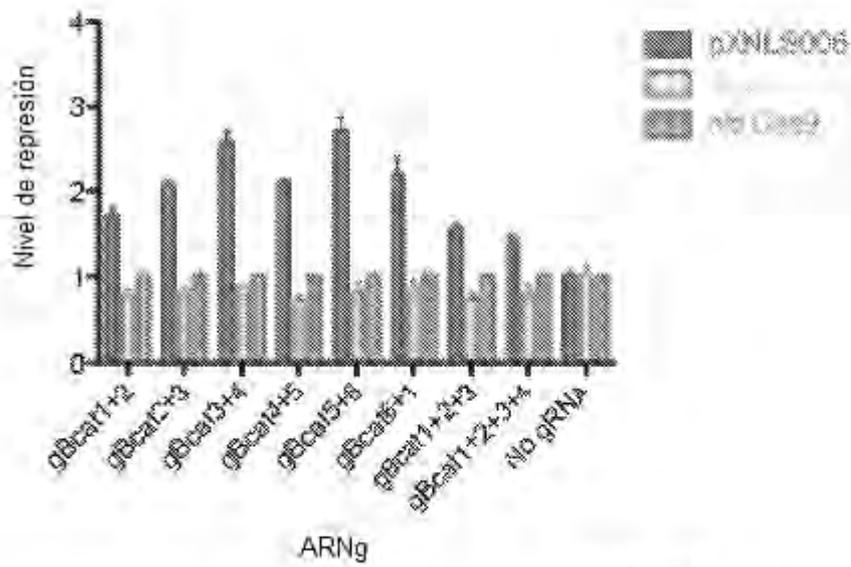


FIG. 18

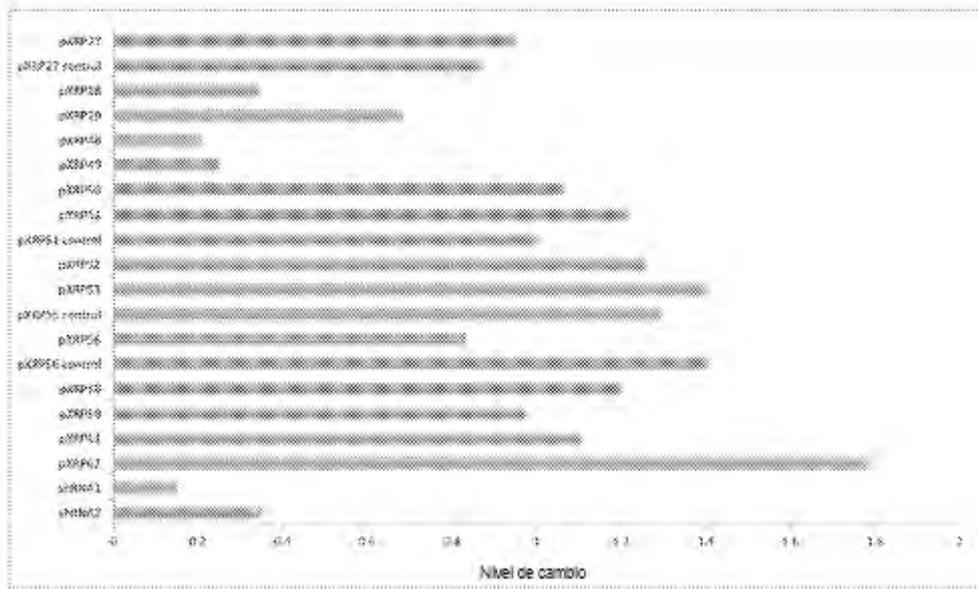


FIG. 19

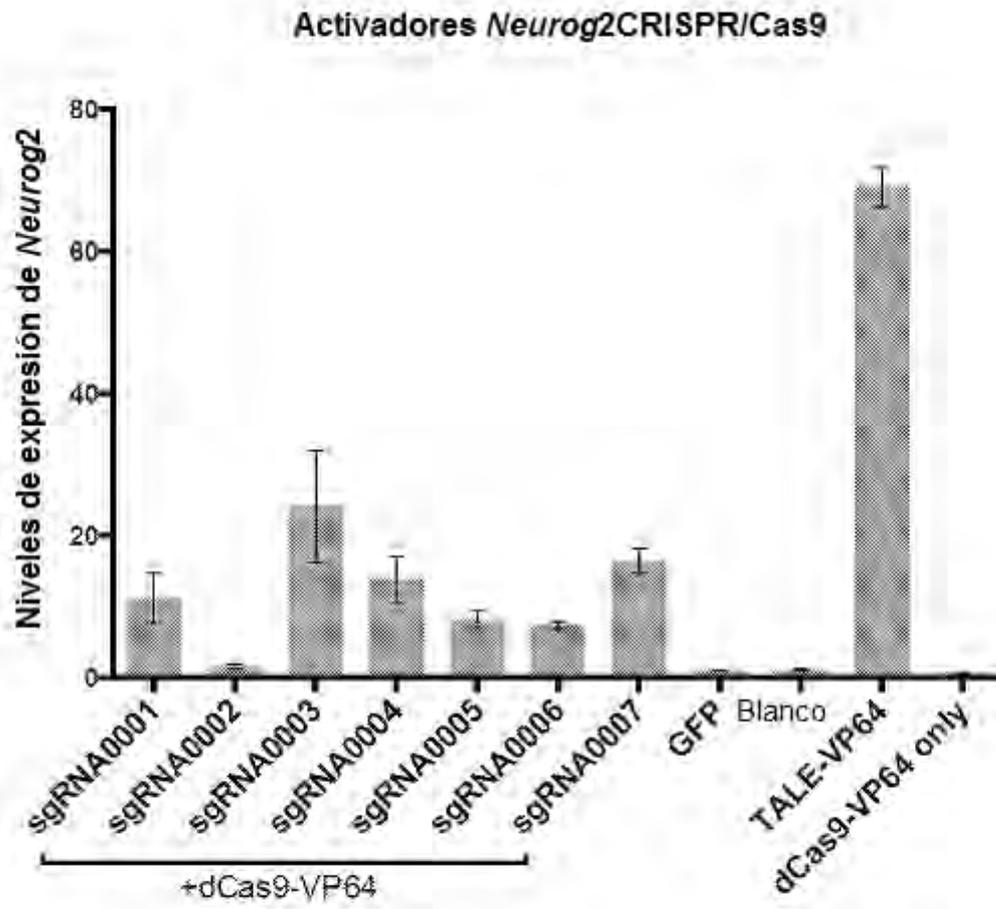


FIG. 20

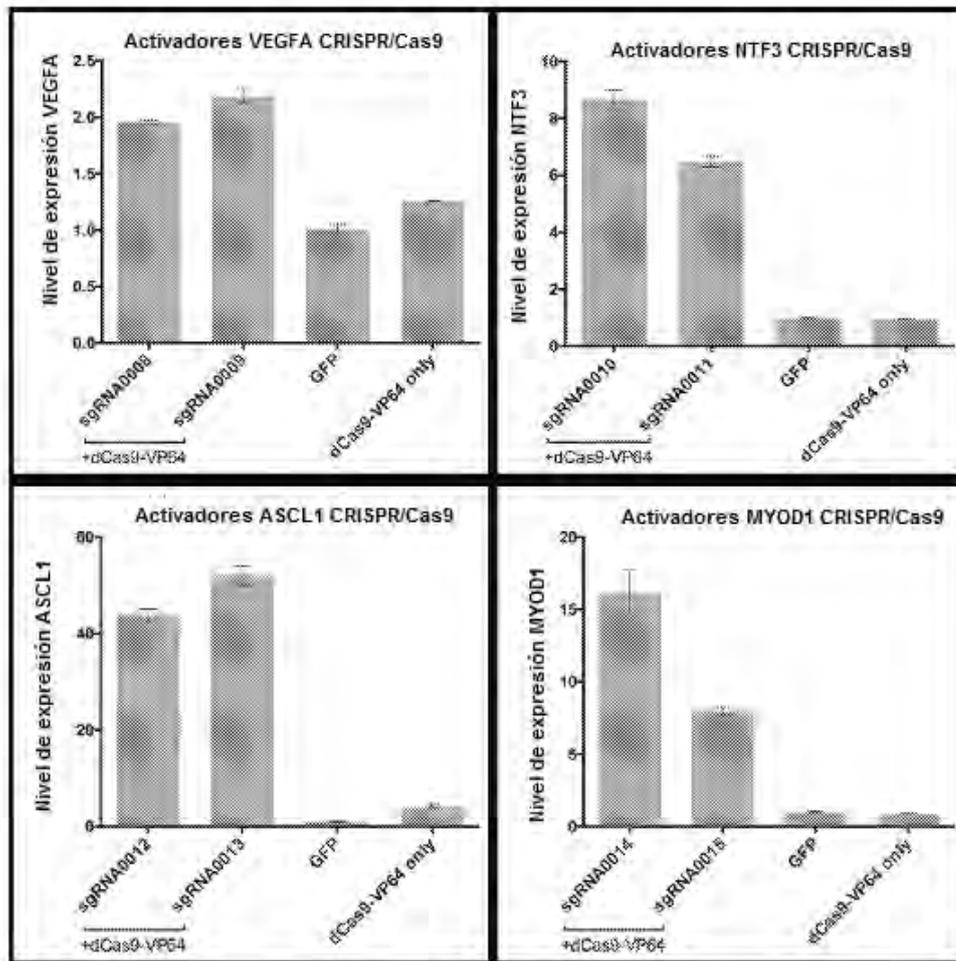


FIG. 21

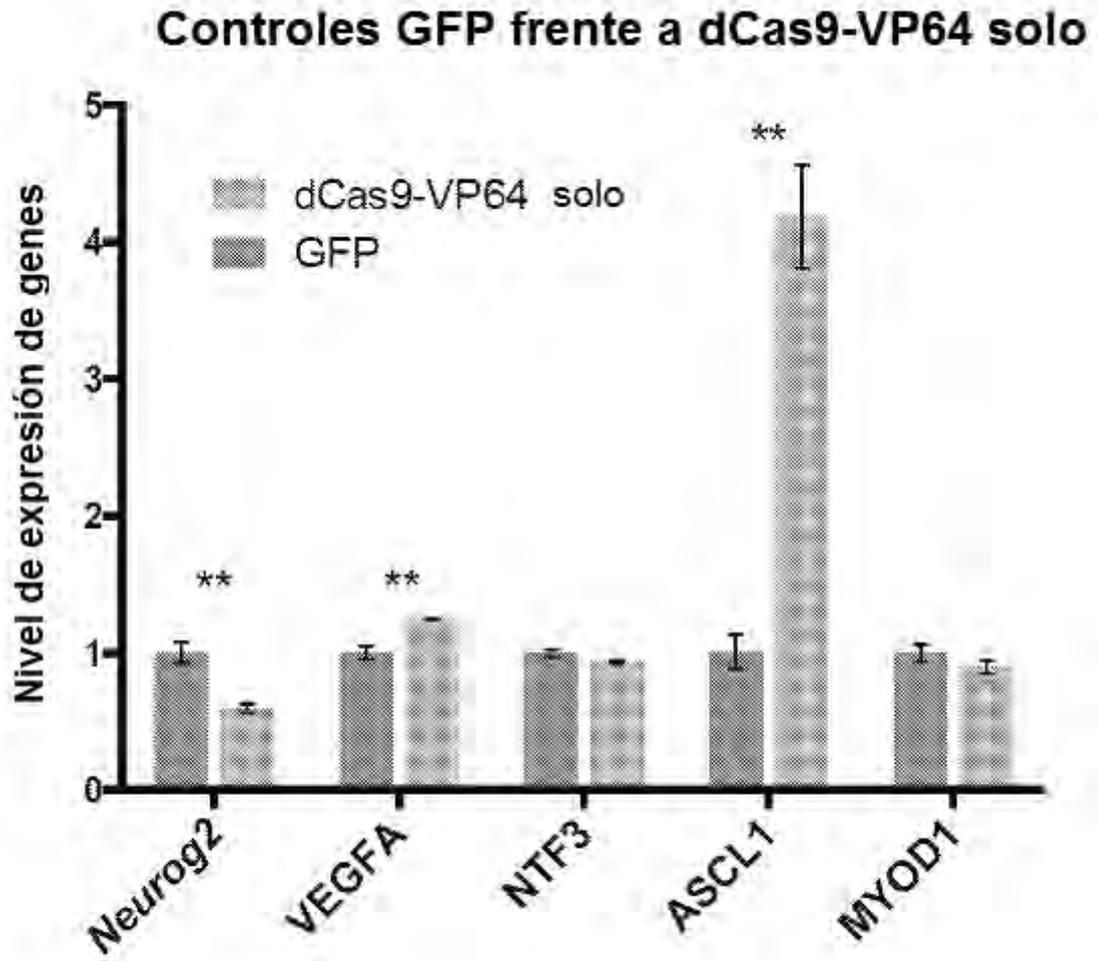


FIG. 22

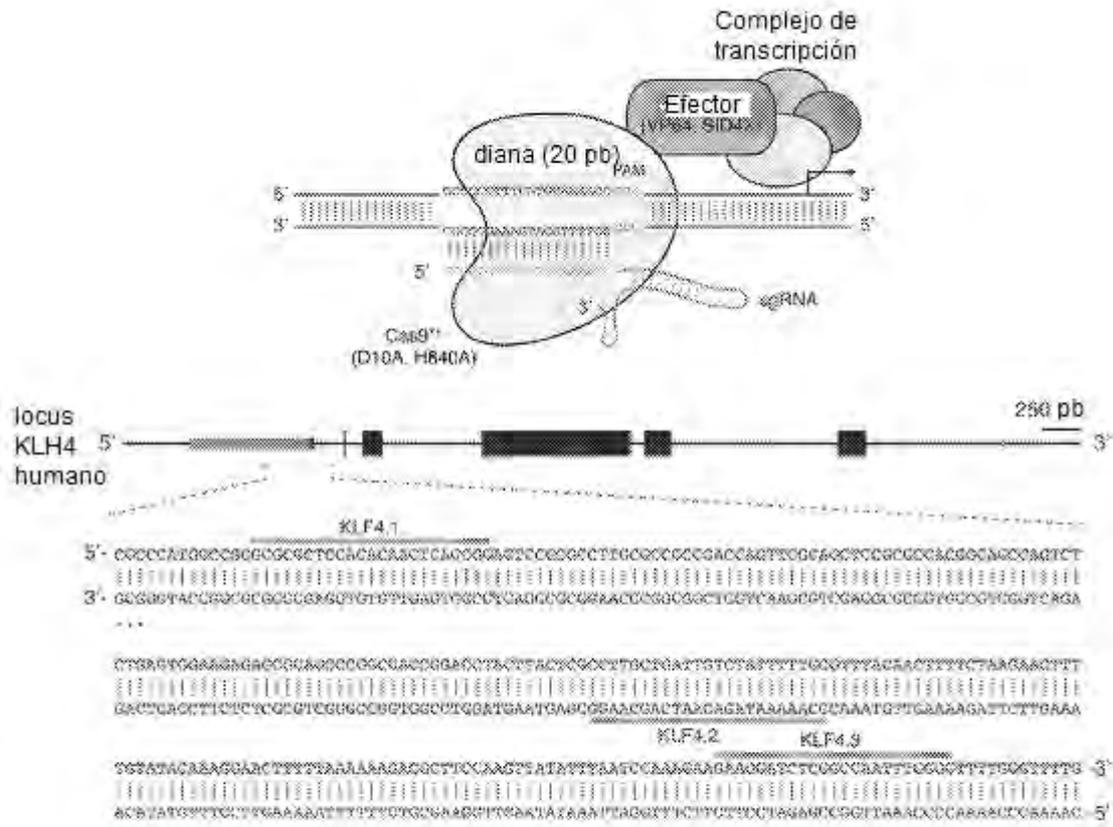


FIG. 23A-B

hSpCas9(D10A,B840A)- Enlaz **NLS-SP6**

MDKKYSIGLAIGTNSVGWAVITDEYKVPSCRKFKVLGNTDRHSIKRNLIGALLFDSGETA
 EATRLKRTFARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLEBSFLVEEDKKHERHPIF
 GNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
 DVDKLFIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFG
 NLIALSLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSLILRVNTEITKAPLSASMIKRYDEHHQDLTLIKALVRQQLPEKYKEIFFDQSKNGY
 AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
 HAILRRQEDFYPELKDNREKIEKILFRIPYYVGPLARGNSRFAMTRKSEETITPWNFEE
 VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLEYEFTVYNELTKVKYVTEGMRKPA
 PLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLI
 KIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG
 WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
 DSLHEHIANLAGSPAIKKILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQGGQKN
 SRERMKRIEIGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD
 YDVAIVPQSFLKDDSIDNKVLTNRSDKNRGSNDNVPSEEVVKKMKNYWRQLLNAKLIT
 QRKFDNLTKAERGGLELDKAGFIKRQLYETRQITKHVAQILDSRMNTKYDENDKLIRE
 VKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGTAIKKYPKLESEFVYG
 DYKVDVVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
 VWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKK
 YGGFDSPTVAYSVLVAKVEKGSKLLKSVKELLGITIMERSSEKPNIDFLEAKGYKE
 VKKDLIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFYLYASHYEKLGKS
 PEDNEQKQLFVEQHKHYLDEIHEQISEFSKRVILADANLDKVL SAYNKHRDKPIREQAENI
 IHLFTLNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSIITGLYETRIDLSQLGGDSA
**GGGGGGGSGGGGSGGKLRKVAAGSGLRADALDDEFDMLGSDALDDEFDLDMIA
 EDALDDEDLDMIGSDALDDEFDLDMIA**

FIG. 24

SIDIX-NLS-FLAG-Enlaz-hSpCas9(D10A,H840A)-NLS

MGSGMNIQMILLEAADYLFRRFRFAFHGYASMLPGSGMNIQMILLEAADYLFRRFRFAF
 HGYASMLPGSGMNIQMILLEAADYLFRRFRFAEHGYASMLPGSGMNIQMILLEAADYLE
 RRFRFAEHGYASMLPSRSAQPKKILFVAADYKDDDDKSRLDIPGFKPYKCPICGKSF
 QSGALTBHQRTHTTRDKKYSIGLAIGTNSVGWAVITDEYKVPKPKFKVLGNTDRHSIKKN
 LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSEFHRLEESFL
 VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR
 GHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDKAILSARLSKSRRLLENLI
 AQLPGEKKNGLPGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQ
 YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
 KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
 DNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYVVGPLARGNSRPAWMT
 RKSEETIPWNFEVVDKGAQAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTK
 VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
 RFNASLGTYHDLIKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDK
 VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMOIHHDDSLTFK
 EDIQAQVSGQGDSLHEHIANLAGSPAIKKILQTVKVVDELVKVMGRHKPENIYIEMA
 RENQTTQKQKNSRERMKRIEIGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDM
 YVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK
 NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
 NTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHDAYLNAVVGTAI
 KKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIR
 KRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTTGGFSKESILPKRNSDK
 LIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGGKSKLKSVKELGITIMERSSEFKN
 PIDFLEAKGYKEVKKDLIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFY
 LASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIBQISEFSKRVILADANLKVLSAYNKII
 RDKPIREQAENIHLFTLNLGAPAAFKYFDTTIDRKRYTSTKEVLDAITLHQSIITGIYETR
 IDLSQLGGDEKPTATCKAGQAKKKA

FIG. 25

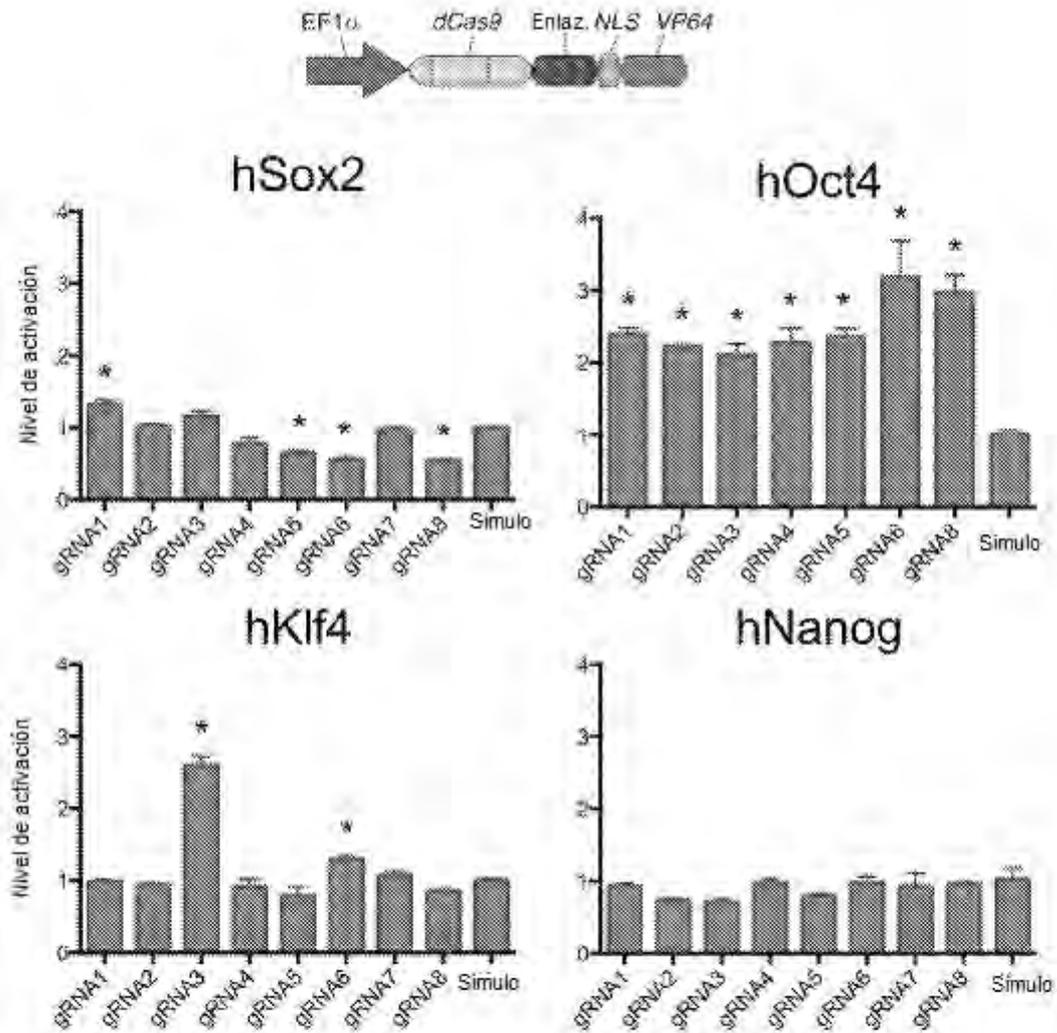


FIG. 26

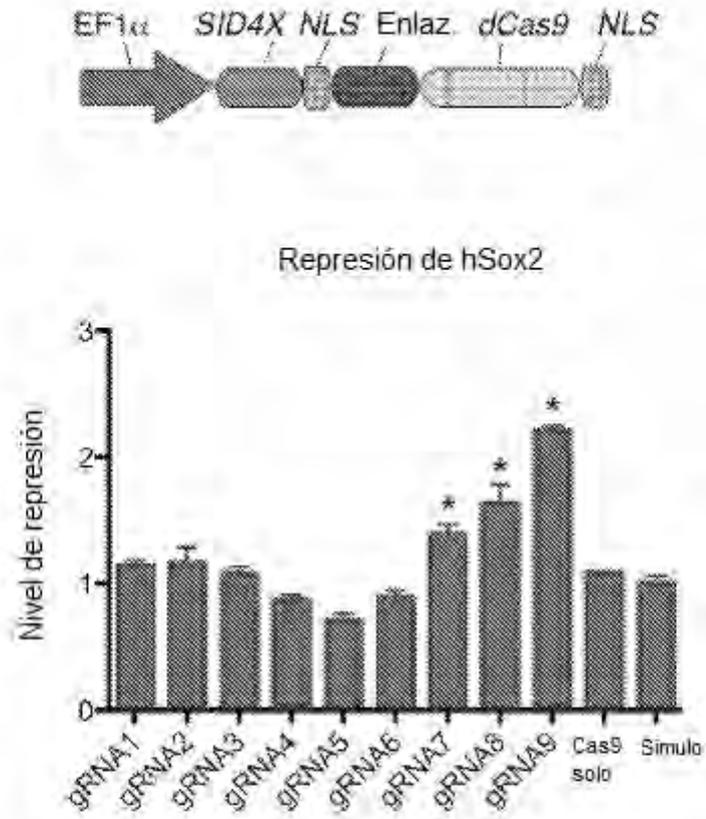


FIG. 27

pAAV-EF1a-dCas9-GS-CIB1(mNLS d318-334)_WPRE_hGH poli A

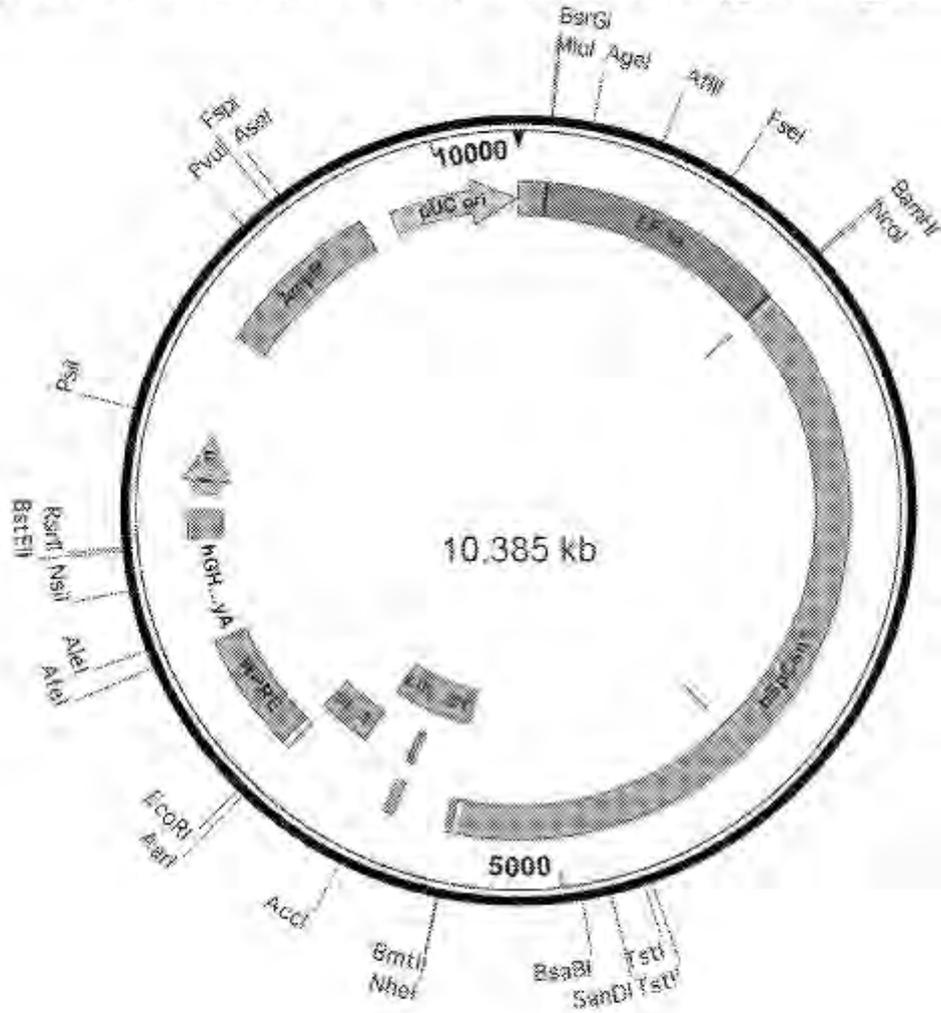


FIG. 28

pAAV-EF1a-dCas9-GS-NLS-cib1-WPRE-hGHpA

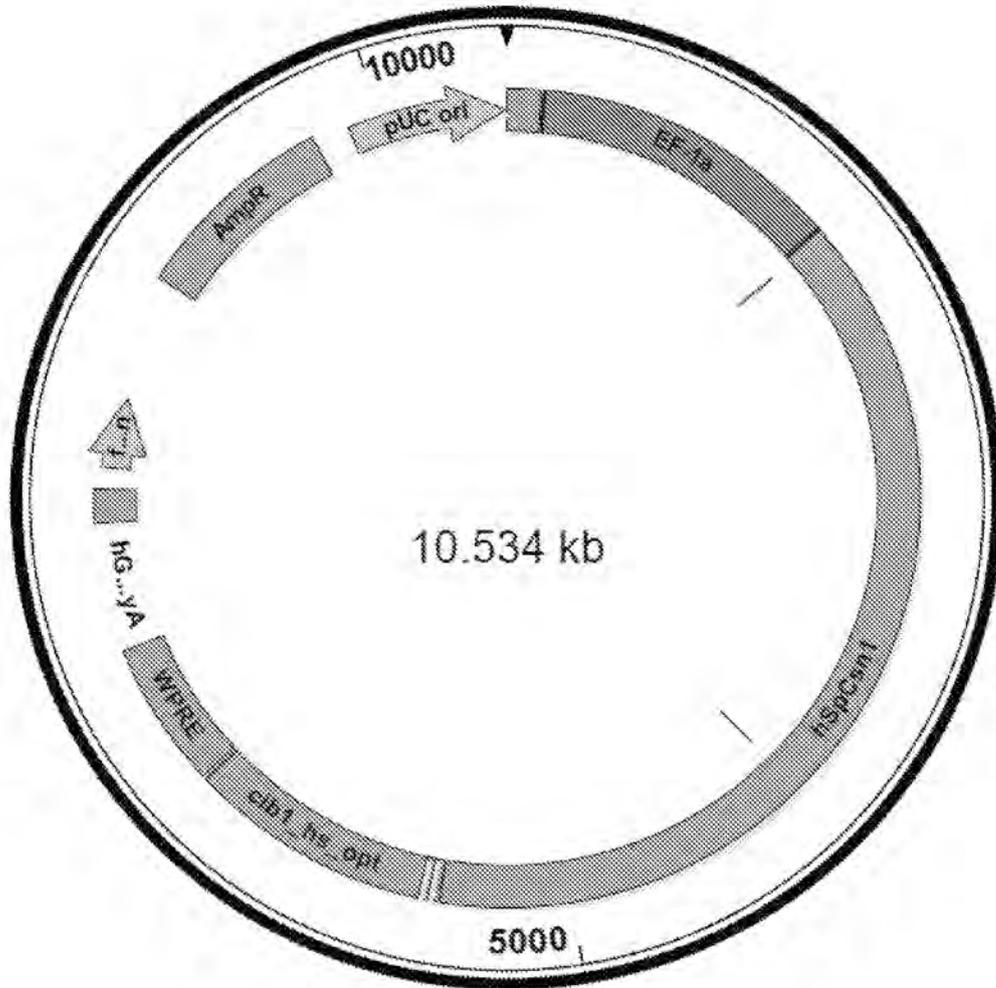


FIG. 29

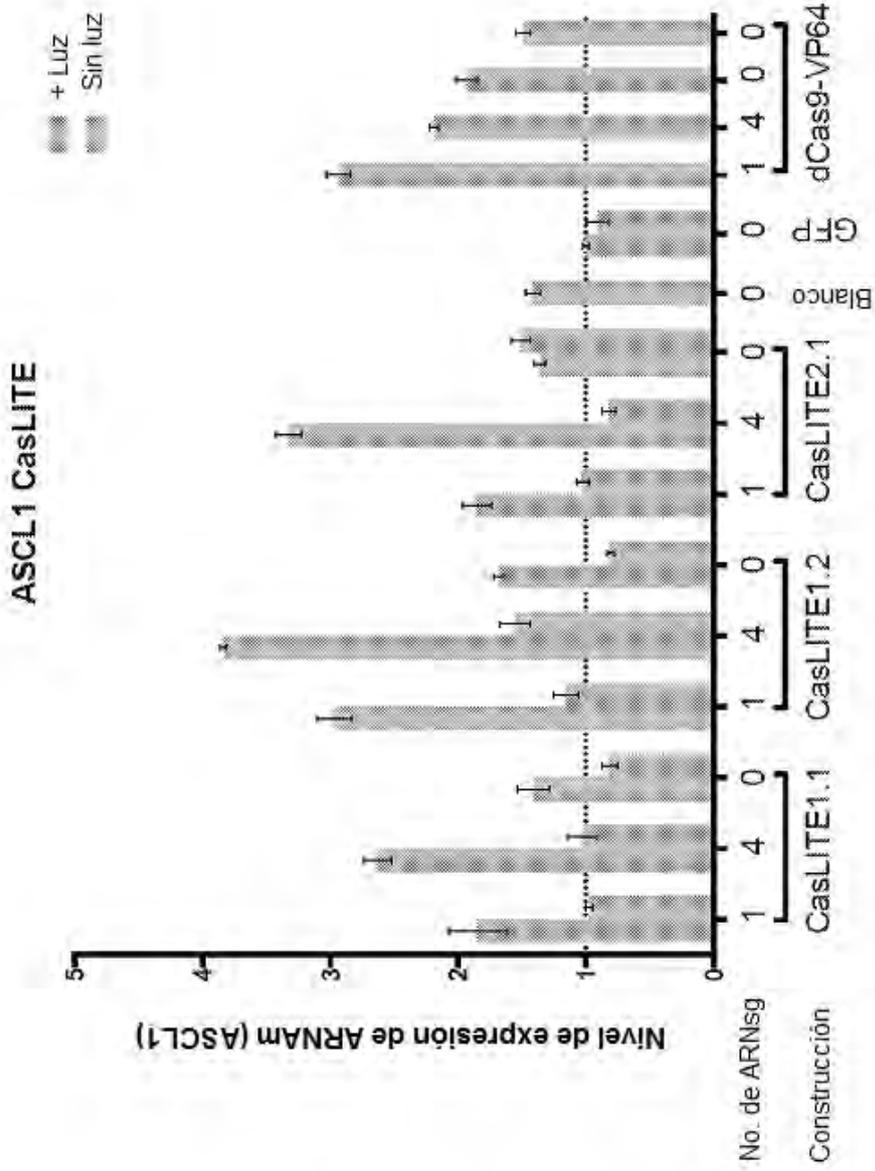
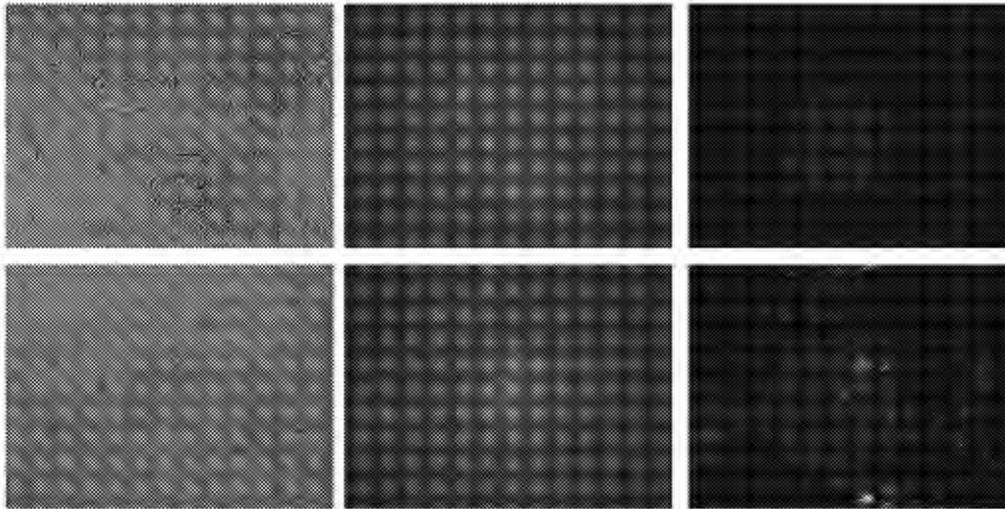


FIG. 31

mESCs (ROSA26-CAG-loxP-poli A-loxP-Cas9)
 72 horas post-magnetofección
 Formación de imágenes por fluorescencia y ensayo de supervivientes



Arriba: sgRNA_mShank3.1
 Abajo: sgRNA_mShank3.1 + Cre



Pista

- 1. Neg
- 2. GFP
- 3. Cre
- 4. sgRNA_mShank3.1 + Cre
- 5. sgRNA_mShank3.2 + Cre
- 6. sgRNA_mShank3.3 + Cre
- 7. sgRNA_mShank3.1
- 8. sgRNA_mShank3.2
- 9. sgRNA_mShank3.3

% de corte

- 18
- 18
- 22
- 32
- 28
- 27

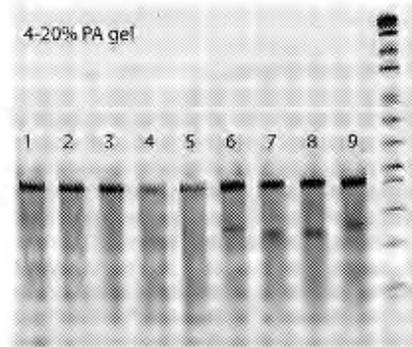


FIG. 32

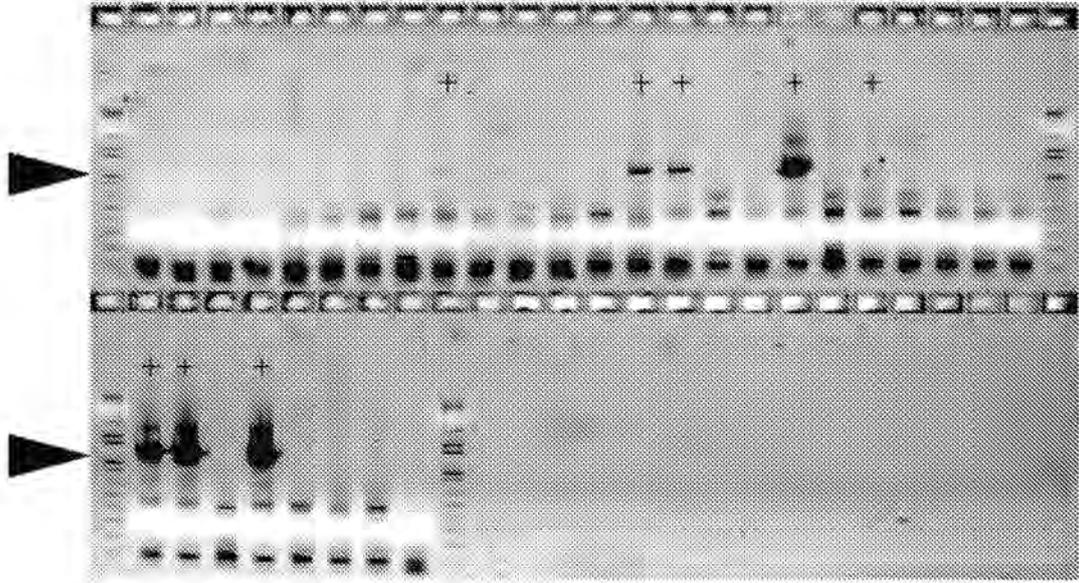


FIG. 33

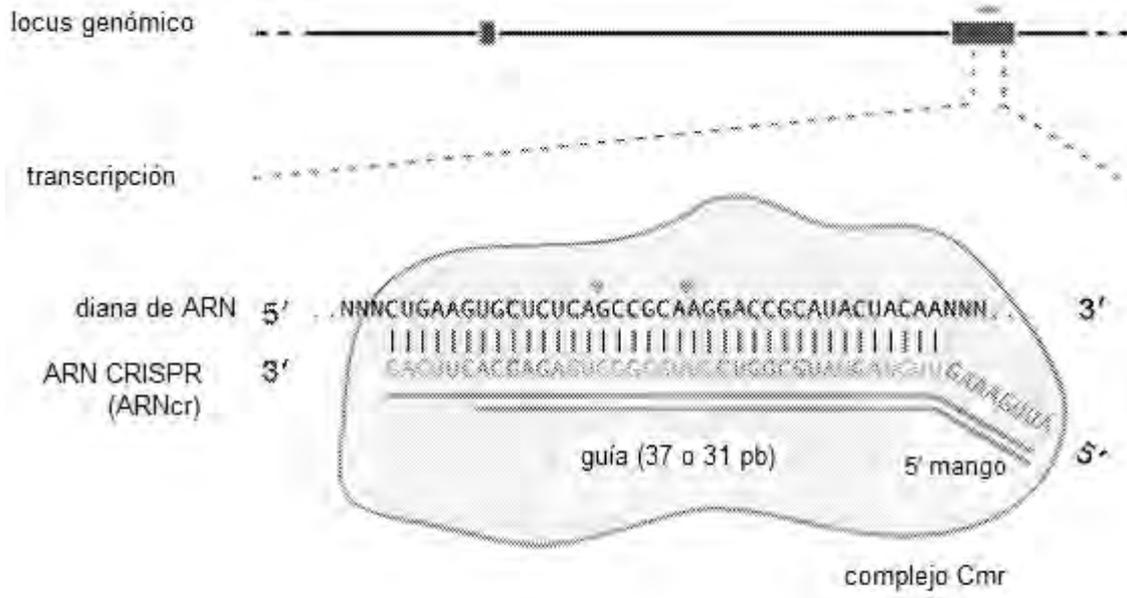


FIG. 34

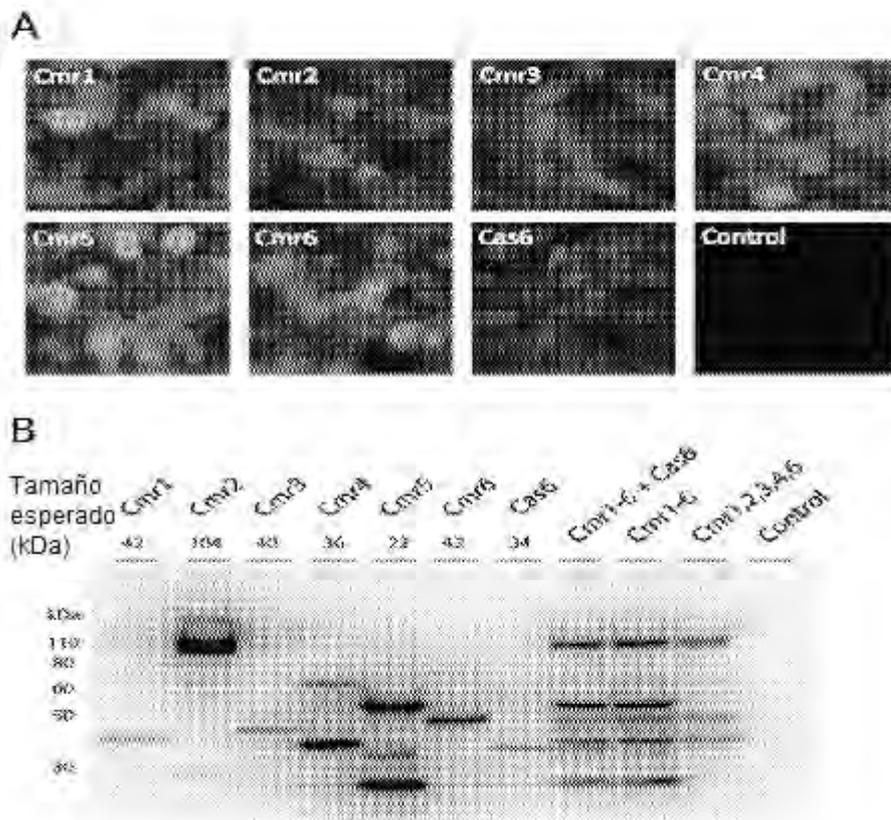


FIG. 35A-B

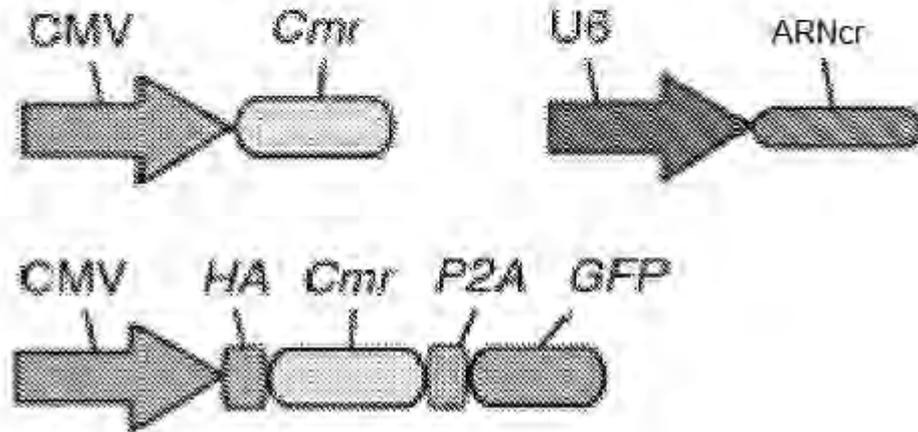


FIG. 36