

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 800 168**

51 Int. Cl.:

C12N 15/09 (2006.01)

C07K 19/00 (2006.01)

C12N 9/24 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.11.2015 PCT/JP2015/080958**

87 Fecha y número de publicación internacional: **12.05.2016 WO16072399**

96 Fecha de presentación y número de la solicitud europea: **02.11.2015 E 15857325 (3)**

97 Fecha y número de publicación de la concesión europea: **15.04.2020 EP 3216867**

54 Título: **Procedimiento de modificación de una secuencia genómica para introducir una mutación específica en una secuencia de ADN diana por reacción de eliminación de base, y complejo molecular usado en el mismo**

30 Prioridad:

04.11.2014 JP 2014224745

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

28.12.2020

73 Titular/es:

**NATIONAL UNIVERSITY CORPORATION KOBE
UNIVERSITY (100.0%)
1-1 Rokkodai-cho Nada-ku
Kobe-shi, Hyogo 657-8501, JP**

72 Inventor/es:

**NISHIDA, KEIJI y
KONDO, AKIHIKO**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 800 168 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

Procedimiento de modificación de una secuencia genómica para introducir una mutación específica en una secuencia de ADN diana por reacción de eliminación de base, y complejo molecular usado en el mismo

Campo técnico

5 La presente invención se refiere a un método de modificación de una secuencia genómica que posibilita la modificación de una base de ácido nucleico en una región particular de un genoma sin cortar el ADN bicatenario (sin corte o corte de una hebra) ni la inserción de un fragmento de ADN extraño, pero utilizando una reacción de escisión de base, y a un complejo de módulo de reconocimiento de secuencia de ácido nucleico y ADN glicosilasa para uso para ello.

Antecedentes de la técnica

10 En los últimos años, la edición genómica está atrayendo atención como una técnica para modificar un gen y región genómica objeto en diversas especies. Convencionalmente, como método de edición genómica se ha propuesto un método que utiliza una nucleasa artificial que comprende una molécula que tiene una capacidad de corte de ADN independiente de secuencia y una molécula que tiene una capacidad de reconocimiento de secuencia en combinación (documento no de patente 1).

15 Por ejemplo, se ha reseñado un método para efectuar la recombinación en un locus génico diana en ADN de una célula vegetal o célula de insecto como hospedador, usando una nucleasa de dedo de cinc (ZFN) en la que están ligados un dominio de unión a ADN de dedo de cinc y un dominio de corte de ADN no específico (documento de patente 1), un método de corte o modificación de un gen diana en una secuencia nucleotídica particular o un sitio adyacente a la misma usando TALEN, en el que están ligados un efector de tipo activador de la transcripción (TAL) que es un módulo de unión a ADN que tiene la bacteria patogénica vegetal *Xanthomonas* y una ADN endonucleasa (documento de patente 2), un método que utiliza el sistema CRISPR-Cas9 en el que se combinan CRISPR (repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas) de secuencia de ADN que funciona en un sistema inmunitario adquirido poseído por eubacterias y arqueobacterias y la familia de proteína nucleasa Cas (asociada a CRISPR) que tienen una función importante junto con CRISPR (documento de patente 3) y similares. Además, se ha reseñado también un método de corte de un gen diana en los alrededores de una secuencia particular usando una nucleasa artificial en la que están ligadas una proteína PPR, constituida para reconocer una secuencia nucleotídica particular por una continuación de motivos de PPR consistente cada uno en 35 aminoácidos y que reconoce una base de ácido nucleico, y una nucleasa (documento de patente 4).

20 Estas técnicas de edición genómica presuponen básicamente roturas de ADN bicatenario (DSB). Sin embargo, puesto que incluyen modificaciones genómicas inesperadas, pueden aparecer efectos secundarios tales como fuerte citotoxicidad, reorganización cromosómica y similares, y tienen los problemas comunes de fiabilidad alterada en terapia génica, número extremadamente bajo de células supervivientes por modificación nucleotídica y dificultad en la modificación genética misma en óvulos de primate y microorganismos unicelulares.

25 Por otro lado, se ha propuesto como método para efectuar modificación nucleotídica sin causar DSB la utilización de ADN glicosilasa. Por ejemplo, el documento de patente 5 describe que se obtuvo una ADN glicosilasa mutante que tiene actividad para eliminar una base de timina o citosina de un ADN monocatenario o bicatenario (actividad TDG o actividad CDG) introduciendo una mutación en uracilo-ADN glicosilasa (UDG) humana y que, usando la enzima, mejoró la eficacia de inducción de mutación en *Escherichia coli*.

30 Además, el documento de patente 6 y el documento no de patente 2 describen que es posible la modificación nucleotídica orientada en una región del genoma que tiene una secuencia de ADN del operador Tet (TetO) reconocida específicamente por TetR usando una proteína de fusión de 3-metiladenina-ADN glicosilasa de levadura (MAG1) y la proteína represora de Tet (TetR). Sin embargo, la MAG1 retira esencialmente una base de ADN especial que se dañó por alquilación y similar, que es principalmente 3-metiladenina, y la capacidad de retirar una base de bases normales, particularmente pares de bases sin desapareamiento, es extremadamente baja. Por lo tanto, es difícil observar el efecto de la introducción de mutación por MAG1 en células normales, y se ha obtenido una tasa de mutación detectable solo sobreexpresando MAG1 en células con genes alterados del sistema de reparación de escisión de base. Además, cuando se debilita el sistema de reparación de ADN, aumenta también la tasa de mutación de todo el genoma, lo que hace su implementación un objetivo lejano. Sin embargo, cuando se usó una UDG mutante que tiene actividad CDG para actuar sobre una base de citosina normal para este sistema en lugar de MAG1, no aumentó la eficacia de la inducción de mutación (documento de patente 6, documento no de patente 2).

Lista de documentos**Documentos de patente**

Documento de patente 1: JP-B-4968498

Documento de patente 2: publicación nacional de solicitud de patente internacional n.º 2013-513389

Documento de patente 3: publicación nacional de solicitud de patente internacional n.º 2010-519929

Documento de patente 4: JP-A-2013-128413

Documento de patente 5: WO 97/25416

Documento de patente 6: WO 2014/127287

5 Documentos no de patente

Documento no de patente 1: Kelvin M Esvelt, Harris H Wang (2013) Genome-scale engineering for systems and synthetic biology, *Molecular Systems Biology* 9: 641

Documento no de patente 2: Shwan P. Finney-Manchester, Narendra Maheshri, (2013) Harnessing mutagenic homologous recombination for targeted mutagenesis in vivo by TaGTEAM, *Nucleic Acids Research* 41(9): e99

10 **Compendio de la invención**

Problemas para resolver por la invención

Es un objeto de la presente divulgación proporcionar un método novedoso de edición genómica para modificar una base de ácido nucleico de una secuencia particular de un gen sin DSB ni inserción de un fragmento de ADN extraño, concretamente, sin corte de un ADN bicatenario o corte monocatenario, y un complejo de un módulo de reconocimiento de secuencia de ácido nucleico y una enzima que introduce mutación para ello.

Medio para resolver los problemas

Los presentes inventores han reseñado ya que han modificado exitosamente, sin DSB, una secuencia genómica por conversión de base de ácido nucleico en una región que contiene una secuencia de ADN particular usando una desaminasa que cataliza una reacción de desaminación y ligando la enzima y una molécula que tiene la capacidad de reconocimiento de secuencia de ADN (documento WO 2015/133554). Esperando procurar una tendencia de introducción de mutación diferente de la de desaminasa y similares, la presente invención está basada en la idea de causar una reacción de escisión de base por hidrólisis del enlace N-glicosídico del ADN, e inducir entonces la introducción de mutación en un proceso de reparación de células. Por tanto, se han hecho intentos de usar una enzima que tiene actividad CDG o actividad TDG, que es un mutante de uracilo-ADN glicosilasa mitocondrial de levadura (UNG 1), como enzima que efectúa tal reacción de escisión de base.

Los presentes inventores supusieron que una de las razones para el fallo en la mejora de la eficacia de inducción de mutación por métodos convencionales (p. ej., el documento de patente 6 y el documento no de patente 2) incluso usando UDG mutante que tiene actividad CDG es que la enzima causa escisión de base en todas partes del ADN bicatenario, lo que a su vez causa alta citotoxicidad y hace difícil expresar la proteína enzimática misma. Según tal hipótesis, los presentes inventores modificaron una secuencia genómica por conversión de base de ácido nucleico en una región que contiene una secuencia de ADN particular introduciendo una mutación que rebaja la acción sobre un ADN que tiene una estructura de doble hélice no relajada (ADN (de doble hélice) no relajado) en una UDG mutante que tiene actividad CDG o actividad TDG, y ligando la enzima mutante y una molécula que tiene una capacidad de reconocimiento de secuencia de ADN.

Para ser específico, se usó el sistema CRISPR-CAS (CRISPR-Cas mutante). Es decir, se produjo un ADN que codifica un molécula de ARN en la que se liga CRISPR-ARN:ARNcr (ARNg) específico de genoma que contiene una secuencia complementaria de una secuencia diana de un gen para modificar con un ARN (ARNcr transactivador; ARNtracr) para agrupar proteína Cas, se produjo un ADN en el que se ligan un ADN que codifica una proteína Cas mutante (dCas) en la que se inactiva la capacidad de corte de ambas hebras de un ADN bicatenario y un gen CDG o TDG mutante, y se introdujeron estos ADN en una célula de levadura hospedadora que contiene un gen para modificar. Como resultado, pudo introducirse mutación aleatoriamente dentro del intervalo de varios cientos de nucleótidos del gen objeto incluyendo la secuencia diana. Además, pudo introducirse mutación de forma extremadamente eficaz por orientación a múltiples regiones en el gen objeto. Es decir, se sembró una célula hospedadora con ADN introducido en un medio no selectivo y se examinó la secuencia del gen objeto en colonias seleccionadas aleatoriamente. Como resultado, se confirmó la introducción de mutación en casi todas las colonias. La eficacia de la inducción de mutación se mejoró además coexpresando endonucleasa AP mutante que carece de su función como enzima de un sistema de reparación de escisión de base, aumentando así la frecuencia de errores de reparación en un sitio abásico (sitio AP).

Se confirmó también que un sistema que usa una UDG mutante heterogénea funciona también en levadura. Inesperadamente, la UDG derivada de virus Vaccinia (vvUDG) no causa citotoxicidad incluso en ausencia de introducción de la mutación que rebaja la acción sobre una estructura de ADN de doble hélice no relajada, y la UDG mutante únicamente con mutaciones introducidas que cambian la especificidad de sustrato mostraba bastante mayor eficacia de inducción de mutación. Es decir, se sugería que la vvUDG era una enzima que tenía una acción nativa baja sobre ADN que tiene una estructura de doble hélice no relajada. Esto se apoyaba también por el hecho de que los sitios mutados por vvUDG están concentrados en una base particular dentro de la secuencia diana (concretamente,

la parte monocatenaria). Además, la actividad de vvUDG pudo aumentarse coexpresando A20, que es conocido como cofactor.

Los presentes inventores tuvieron la idea de utilizar el sistema de Cas9 dividida como medio para disminuir la mutación no específica por UDG, además de la introducción de mutación en la enzima. Las Cas9 dividida es una variante de Cas9 que se diseña para funcionar solo cuando el fragmento N-terminal y el fragmento C-terminal divididos de la proteína Cas9 se asocian con ARNg (Nat Biotechnol. 33(2): 139-142 (2015); PNAS 112(10): 2984-2989 (2015)). Los presentes inventores construyeron un sistema que expresa UNG1 mutante de dCas9 dividida como una proteína de fusión que consiste en: fragmento N-terminal de dCas9/fragmento N-terminal de UNG1 mutante/fragmento C-terminal de dCas9/fragmento C-terminal de UNG1 mutante y un sistema que expresa separadamente el fragmento N-terminal de dCas9/fragmento C-terminal de UNG1 mutante y el fragmento N-terminal de UNG1 mutante/fragmento C-terminal de dCas9, y examinaron entonces la frecuencia de mutación y mutación no específica en el sitio diana usando una secuencia del gen CanI como diana, así como proliferación celular. Como resultado, en cualquier sistema la tasa de supervivencia celular en un medio no selectivo es apenas diferente entre UNG1 mutante solo con mutación (N222D) introducida que confiere actividad CDG y UNG1 mutante con mutación (R308C) introducida adicional que disminuye la acción sobre una estructura de ADN de doble hélice no relajada, y la citotoxicidad se redujo notablemente por la utilización de la enzima dividida. Cuando se usó la enzima dividida, se redujo también la frecuencia de mutación no específica usando resistencia a tialisina como índice, sin depender de la presencia o ausencia de la mutación R308. La eficacia de la inducción de mutación en el sitio diana original disminuyó bastante, puesto que la adición de la mutación R308C disminuye también la actividad CDG. Se sugirió, por lo tanto, que preferiblemente no se añade una mutación que reduce la acción sobre una estructura de ADN de doble hélice no relajada a UDG cuando se usa una enzima dividida.

Los presentes inventores han realizado estudios adicionales basados en estos hallazgos y han completado la presente invención.

Por lo tanto, la presente invención es como se describe a continuación.

[1] Un método de modificación de un sitio diana de un ADN bicatenario en una célula, que comprende una etapa de puesta en contacto de un complejo en el que se unen un módulo de reconocimiento de secuencia de ácido nucleico que se une específicamente a una secuencia nucleotídica diana en un ADN bicatenario dado y una ADN glicosilasa con reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada (ADN no relajado), con dicho ADN bicatenario, para convertir uno o más nucleótidos en el sitio diana en uno u otros nucleótidos o eliminar uno o más nucleótidos, o insertar uno o más nucleótidos en dicho sitio diana, sin cortar al menos una hebra de dicho ADN bicatenario en el sitio diana.

[2] El método de [1] anteriormente mencionado, en el que se selecciona el módulo de reconocimiento de secuencia de ácido nucleico anteriormente mencionado del grupo consistente en un sistema CRISPR-Cas en el que se inactiva al menos una capacidad de corte de ADN de Cas, un motivo de dedo de cinc, un efector TAL y un motivo PPR.

[3] El método de [1] anteriormente mencionado, en el que el módulo de reconocimiento de secuencia de ácido nucleico anteriormente mencionado es un sistema CRISPR-Cas en el que se inactiva al menos una capacidad de corte de ADN de Cas.

[4] El método de [1] o [2] anteriormente mencionado, en el que se pone en contacto adicionalmente el ADN bicatenario con un factor que cambia la estructura bicatenaria de ADN.

[5] El método de cualquiera de los [1]-[4] anteriormente mencionados, que usa dos o más clases de módulos de reconocimiento de secuencia de ácido nucleico que se unen cada uno específicamente a una secuencia nucleotídica diana diferente.

[6] El método de [5] anteriormente mencionado, en el que las secuencias nucleotídicas diana diferentes anteriormente mencionadas están presentes en diferentes genes.

[7] El método de cualquiera de [1]-[6] anteriormente mencionados, en el que la ADN glicosilasa anteriormente mencionada tiene actividad citosina-ADN glicosilasa (CDG) o actividad timina-ADN glicosilasa (TDG).

[8] El método de [7] anteriormente mencionado, en el que la ADN glicosilasa anteriormente mencionada que tiene actividad CDG o actividad TDG es un mutante de uracilo-ADN glicosilasa (UDG).

[9] El método de cualquiera de [1]-[8] anteriormente mencionados, que comprende además poner en contacto el ADN bicatenario con una endonucleasa AP que tiene capacidad de unión a un sitio abásico pero que carece de actividad nucleasa.

[10] El método de cualquiera de [1]-[9] anteriormente mencionados, en el que la ADN glicosilasa anteriormente mencionada tiene reactividad nativa baja con una estructura de ADN de doble hélice no relajada.

- [11] El método de [10] anteriormente mencionado, en el que la ADN glicosilasa anteriormente mencionada es un mutante de uracilo-ADN glicosilasa (UDG) derivada de un virus perteneciente a Poxviridae y que tiene actividad CDG o actividad TDG.
- 5 [12] El método de [11] anteriormente mencionado, en el que se pone en contacto adicionalmente el ADN bicatenario con proteína A20.
- [13] El método de cualquiera de [1]-[9] anteriormente mencionados, en el que la ADN glicosilasa anteriormente mencionada es un mutante que tiene reactividad reducida con un ADN que tiene una estructura de doble hélice no relajada (ADN no relajado) en comparación con uno de tipo silvestre.
- 10 [14] El método de cualquiera de [1]-[9] anteriormente mencionados, en el que la ADN glicosilasa anteriormente mencionada y un elemento del módulo de reconocimiento de secuencia de ácido nucleico anteriormente mencionado que está unido directamente a la ADN glicosilasa se dividen respectivamente en dos fragmentos, estando los fragmentos de cualquiera de ADN glicosilasa y elemento respectivamente ligados a los fragmentos del otro proporcionando dos complejos parciales, y cuando los complejos parciales se repliegan entre sí, el módulo de reconocimiento de secuencia de ácido nucleico es capaz de unirse específicamente a la secuencia nucleotídica diana y el enlace específico posibilita a la ADN glicosilasa exhibir actividad enzimática.
- 15 [15] El método de [14] anteriormente mencionado, en el que el elemento del módulo de reconocimiento de secuencia de ácido nucleico que está directamente unido a la ADN glicosilasa anteriormente mencionada es una proteína Cas en la que está inactivada al menos una de las capacidades de corte de ADN.
- [16] El método de [14] o [15] anteriormente mencionado, en el que los dos complejos parciales anteriormente mencionados se proporcionan como complejos moleculares separados, y se repliegan por asociación de los mismos en la célula.
- 20 [17] El método de cualquiera de [1]-[16] anteriormente mencionados, en el que se pone en contacto el ADN bicatenario con el complejo introduciendo un ácido nucleico que codifica el complejo en una célula que tiene ADN bicatenario.
- 25 [18] El método de [17] anteriormente mencionado, en el que la célula anteriormente mencionada es una célula procariótica.
- [19] El método de [17] anteriormente mencionado, en el que la célula anteriormente mencionada es una célula eucariótica.
- 30 [20] El método de [17] anteriormente mencionado, en el que la célula anteriormente mencionada es una célula microbiana.
- [21] El método de [17] anteriormente mencionado, en el que la célula anteriormente mencionada es una célula vegetal.
- [22] El método de [17] anteriormente mencionado, en el que la célula anteriormente mencionada es una célula de insecto.
- 35 [23] El método de [17] anteriormente mencionado, en el que la célula anteriormente mencionada es una célula animal.
- [24] El método de [17] anteriormente mencionado, en el que la célula anteriormente mencionada es una célula de vertebrado.
- 40 [25] El método de [17] anteriormente mencionado, en el que la célula anteriormente mencionada es una célula de mamífero.
- [26] El método de cualquiera de [18]-[25] anteriormente mencionados, en el que la célula anteriormente mencionada es una célula poliploide, y todos los sitios diana en alelos de un cromosoma homólogo están modificados.
- 45 [27] Un complejo enzimático modificador de ácido nucleico en el que se unen un módulo de reconocimiento de secuencia de ácido nucleico que se une específicamente a una secuencia nucleotídica diana en un ADN bicatenario dado y una ADN glicosilasa con reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada (ADN no relajado), que convierte uno o más nucleótidos en el sitio diana en otro o más nucleótidos, o elimina uno o más nucleótidos, o inserta uno o más nucleótidos en dicho sitio diana sin cortar al menor una hebra de dicho ADN bicatenario en el sitio diana.
- 50 [28] Un ácido nucleico que codifica el complejo enzimático modificador de ácido nucleico de [27] anteriormente mencionado.

Efecto de la invención

Según la técnica de edición genómica de la presente divulgación, puesto que no está acompañada de inserción de ADN extraño o roturas de ADN bicatenario, la técnica es superior en seguridad y tiene una posibilidad no pequeña de procurar una solución a casos que causan disputas biológicas o legales en métodos convencionales relativas a la recombinación génica. También es posible establecer un intervalo apropiado de introducción de mutación hasta el intervalo de varios cientos de bases incluyendo la secuencia diana, y la técnica puede aplicarse también a la inducción de evolución tópica mediante la introducción de mutación aleatoria en una región restringida particular, lo que ha sido casi imposible hasta ahora.

Breve descripción de los dibujos

10 La Fig. 1 es un esquema que muestra el mecanismo de orientación específica de secuencia de ADN del mutante de uracilo-ADN glicosilasa.

La Fig. 2 es un esquema que muestra el mecanismo del método de modificación genética de la presente invención que usa ADN glicosilasa y el sistema CRISPR-Cas.

15 La Fig. 3 muestra los resultados de la verificación, usando una levadura de gemación, del efecto del método de modificación genética de la presente invención que comprende un sistema CRISPR-Cas y uracilo-ADN glicosilasa mutante derivada de levadura de gemación (que tiene actividad CDG y reactividad disminuida con ADN que tiene estructura de doble hélice no relajada) en combinación.

20 La Fig. 4 muestra los resultados de la verificación, usando una levadura de gemación, del efecto del método de modificación genética de la presente invención que comprende un sistema CRISPR-Cas y uracilo-ADN glicosilasa mutante derivada de levadura de gemación (que tiene actividad CDG o actividad TDG y reactividad disminuida con ADN que tiene estructura de doble hélice no relajada) en combinación.

La Fig. 5 muestra los resultados del análisis de sitios mutados en la colonia resistente a canavanina obtenida en la Fig. 4.

25 La Fig. 6 muestra los resultados cuando se construye un producto de construcción de expresión de tal forma que se unen uracilo-ADN glicosilasa mutante derivada de levadura de gemación y dCas9 a través del dominio SH3 y se introduce un ligando de unión del mismo en una levadura de gemación junto con dos clases de ARNg.

La Fig. 7 muestra que la eficacia de inducción de mutación por UNG1 mutante ("Ung" en la Figura significa UNG1 derivada de levadura, de aquí en adelante es lo mismo) mejora por la coexpresión de mutantes de endonucleasa AP (Apel) (E96Q, D210N).

30 La Fig. 8 muestra que la citotoxicidad en una levadura hospedadora con UNG1 mutante introducida que tiene actividad CDG (N222D) o actividad TDG (Y164A) puede evitarse introduciendo una mutación (L304A) que disminuye la reactividad con un ADN que tiene una estructura de doble hélice no relajada.

35 La Fig. 9 muestra la tasa de supervivencia y la eficacia de inducción de mutación en una levadura con una uracilo-ADN glicosilasa mutante heterogénea derivada de Escherichia coli o virus Vaccinia (EcUDG o vvUDG) introducida. La vvUDG no causaba citotoxicidad en una levadura hospedadora incluso sin introducción de una mutación (R187C) que disminuye la reactividad con un ADN que tiene una estructura de doble hélice no relajada. La eficacia de inducción de mutación por vvUDG mejoraba notablemente por coexpresión de la proteína A20.

40 La Fig. 10 muestra la comparación de sitios de introducción de mutación entre UNG1 mutante derivada de levadura y UDG mutante derivada de virus Vaccinia (vvUDG) en las secuencias nucleotídicas diana y en los alrededores de las mismas.

La Fig. 11 muestra que la mutación no específica por UNG1 mutante puede reducirse utilizando una enzima dividida.

Descripción de realizaciones

45 La presente invención proporciona un método de modificación de un nucleótido en un sitio diana de una ADN bicatenario utilizando un error en la reparación por escisión de base (REB) de células, al escindir una base de un nucleótido en una región monocatenaria o en una región en que la estructura de doble hélice está relajada en la secuencia nucleotídica diana del ADN bicatenario y en los alrededores de la misma, sin cortar al menos una hebra del ADN bicatenario para modificar. El método se caracteriza porque contiene una etapa de puesta en contacto de un complejo en el que se unen el módulo de reconocimiento de secuencia de ácido nucleico que se une específicamente a la secuencia nucleotídica diana en el ADN bicatenario y la ADN glicosilasa con reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada con el ADN bicatenario, convirtiendo el sitio diana, concretamente la secuencia nucleotídica diana y los nucleótidos en los alrededores de la misma, en otros nucleótidos, o eliminando los mismos, o insertando uno o más nucleótidos en el sitio diana.

En la presente invención, la “modificación” de un ADN bicatenario significa que se convierte un nucleótido (p. ej., dC) en una hebra de ADN en otro nucleótido (p. ej., dT, dA o dG), o se elimina o se inserta un nucleótido o una secuencia nucleotídica entre ciertos nucleótidos en una hebra de ADN. Aunque el ADN bicatenario para modificar no está particularmente limitado, es preferiblemente un ADN genómico. El “sitio diana” de un ADN bicatenario significa la “secuencia nucleotídica diana” total o parcial que un módulo de reconocimiento de secuencia de ácido nucleico reconoce específicamente y a la que se une, o los alrededores de la secuencia nucleotídica diana (uno o ambos de en dirección 5’ y en dirección 3’) y el intervalo del mismo puede ajustarse apropiadamente entre 1 base y varios cientos de bases según el objeto.

En la presente invención, el “módulo de reconocimiento de secuencia de ácido nucleico” significa una molécula o complejo molecular que tiene la capacidad de reconocer específicamente y unirse a una secuencia nucleotídica particular (concretamente, secuencia nucleotídica diana) en una hebra de ADN. La unión del módulo de reconocimiento de secuencia de ácido nucleico a una secuencia nucleotídica diana posibilita a la ADN glicosilasa ligada al módulo actuar específicamente sobre un sitio diana en un ADN bicatenario.

En la presente invención, “ADN glicosilasa” significa una enzima que hidroliza el enlace N-glicosídico del ADN. La ADN glicosilasa desempeña originalmente el papel de eliminar bases dañadas del ADN en REB. En la presente invención, es preferible una ADN glicosilasa capaz de actuar sobre bases normales (concretamente, dC, dT, dA y dG, o aquellas después de modificación epigenética) en el ADN. Una ADN glicosilasa mutante que no reacciona de forma nativa con una base normal o que tiene baja reactividad, pero que tiene reactividad adquirida o reactividad mejorada con una base normal por mutación, está englobada en la ADN glicosilasa de la presente invención y puede usarse preferiblemente. Se trata un sitio abásico (sitio apurínico/apirimidínico (AP)) resultante de la reacción de escisión de base por la enzima con una enzima en dirección 3’ de la ruta de REB, tal como endonucleasa AP, ADN polimerasa, ADN ligasa y similares.

“Reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice relajada” significa que se efectúa una reacción de escisión de base en una región donde se forma ADN que tiene una estructura de doble hélice no relajada solo a una frecuencia suficiente para suprimir la citotoxicidad a un nivel no influyente sobre la supervivencia de las células. Como se usa en la presente memoria, un “ADN que tiene una estructura de doble hélice no relajada” hace referencia a formar una estructura de doble hélice consistente (a saber, ADN de doble hélice no relajado (o simplemente ADN no relajado)) y no incluye el estado de ADN monocatenario de bases de apareamiento complementemente disociadas, ni el estado de dobles hebras en pares de bases que forman una estructura de doble hélice desenrollada y relajada (ADN bicatenario relajado). Los ejemplos de ADN glicosilasa con reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada incluyen ADN glicosilasa con reactividad nativa suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada, ADN glicosilasa mutante con una mutación introducida para disminuir la reactividad con un ADN que tiene una estructura de doble hélice no relajada en comparación con el tipo silvestres y similares. Además, la ADN glicosilasa que es una enzima dividida diseñada para dividirse en dos fragmentos, en la que se unen los fragmentos respectivos a cada uno del módulo de reconocimiento de secuencia de ácido nucleico dividido en dos para formar dos complejos, cuando se repliegan ambos complejos, el módulo de reconocimiento de secuencia de ácido nucleico puede unirse específicamente a la secuencia nucleotídica diana y, debido al enlace específico, la ADN glicosilasa puede catalizar la reacción de escisión de base, está también englobada en “ADN glicosilasa con reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada” de la presente invención.

En la presente invención, el “complejo enzimático modificador de ácido nucleico” significa un complejo molecular que comprende un complejo en el que se ligan el módulo de reconocimiento de secuencia de ácido nucleico anteriormente mencionado y una ADN glicosilasa con reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada, confiriendo a dicho complejo molecular una capacidad de reconocimiento de secuencia nucleotídica particular y una actividad para catalizar una reacción de escisión de base de ácido nucleico. El “complejo” engloba aquí no solo el constituido por múltiples moléculas, sino también el que tiene un módulo de reconocimiento de secuencia de ácido nucleico y ADN glicosilasa en una sola molécula, como una proteína de fusión. Además, el “complejo enzimático modificador de ácido nucleico” de la presente invención engloba también un complejo molecular en que se ligan entre sí el módulo de reconocimiento de secuencia de ácido nucleico dividido en dos fragmentos y cualquier fragmento de ADN glicosilasa dividida en dos fragmentos formando dos “complejos parciales”, y el complejo molecular adquiere la capacidad de reconocimiento de secuencia nucleotídica y una actividad catalizadora de reacción de escisión de base cuando los complejos parciales se repliegan entre sí.

Aunque la ADN glicosilasa para usar en la presente invención no está particularmente limitada, a condición de que pueda hidrolizar el enlace N-glicosídico del ADN para catalizar una reacción para liberar una base, es preferible una ADN glicosilasa capaz de actuar sobre bases normales (concretamente, dC, dT, dA y dG, o aquellas después de modificación epigenética, por ejemplo 5-metilcitosina, etc.) en el ADN para aumentar la versatilidad como técnica de edición genómica. Los ejemplos de tal enzima incluyen una enzima que tiene actividad CDG para catalizar una reacción para liberar citosina, una enzima que tiene actividad TDG para catalizar una reacción para liberar timina, una enzima que tiene actividad para catalizar una reacción para liberar 5-metilcitosina (actividad 5-mCDG) y similares. Aunque no se ha reseñado hasta la fecha una ADN glicosilasa que tenga de forma nativa actividad CDG, en casi todas las especies puede obtenerse una UDG mutante que tiene actividad CDG o TDG que reconoce a citidina o timidina como sustrato introduciendo una mutación en un residuo aminoácido particular de UNG, conocida como la UDG

principal responsable de la retirada de uracilo incorporado a ADN (en el caso de eucariotas, tiene dos variantes de corte y empalme de localización mitocondrial UNG1 y localización nuclear UNG2, que tienen una secuencia aminoacídica común excepto la secuencia N-terminal que contiene cada señal de localización de órgano) (véase la Fig. 1, panel superior). Más específicamente, por ejemplo en el caso de UNG1 derivada de levadura (SEQ ID NO: 2; N.º de acceso a GenBank NP_013691), la actividad CDG puede conferirse sustituyendo la asparagina 222 del extremo N-terminal por ácido aspártico (también se hace referencia al mutante como N222D), y la actividad TDG puede conferirse sustituyendo la tirosina 164 por alanina (también se hace referencia al mutante como Y164A) (Kavli B. et al., EMBO J. (1996) 15(13): 3442-7). Los presentes inventores han encontrado recientemente que puede conferirse una actividad TDG mayor que la del mutante Y164A sustituyendo la tirosina 164 por glicina (también se hace referencia al mutante como Y164G). Además, puesto que la citosina se obtiene sustituyendo el grupo carbonilo en posición 4 del uracilo por un grupo amino y la timina es equivalente a uracilo en que la posición 5 está metilada, puede obtenerse ADN glicosilasa que tiene actividad para liberar 5-metilcitosina derivada de la metilación de la posición 5 de citosina del ADN (actividad 5-mCDG) introduciendo una doble mutación en el residuo de asparagina 222 y el residuo de tirosina 164 (p. ej., N222D/Y164A o N222D/Y164G). Puesto que la escisión de base de 5-metilcitosina puede cambiar el estado de metilación del ADN genómico, la técnica de edición genómica posibilita también edición epigenómica para cambiar la modificación epigenética.

Cuando se usa UNG2, puede obtenerse UNG mutante que tiene actividad CDG, actividad TDG o actividad 5-mCDG introduciendo una mutación similar en el residuo aminoacídico correspondiente al sitio mutado anteriormente mencionado.

En la UNG mutante, el sitio anteriormente mencionado puede sustituirse por un aminoácido distinto del aminoácido anteriormente mencionado, o puede introducirse la mutación en un sitio distinto del sitio anteriormente mencionado, a condición de que pueda actuar sobre una base normal. Que la UNG mutante puede actuar sobre una base normal puede confirmarse por el método descrito, por ejemplo, en Kavli B. et al., EMBO J. (1996) 15(13): 3442-7.

Aunque la derivación de UNG no está particularmente limitada, puede usarse por ejemplo un derivada de *Escherichia coli* (Varshney, U. et al. (1988) J. Biol. Chem., 263, 7776-7784), UNG1 o UNG2 derivadas de levadura, mamífero (p. ej., humana, de ratón, porcina, bovina, de caballo, mono, etc.) y similares, o UDG derivada de virus (p. ej., Poxviridae (virus Vaccinia, etc.), Herpesviridae y similares). Por ejemplo, puede hacerse referencia a UniprotKB N.º P13051-2 y P13051-1 para las secuencias aminoacídicas de UNG1 y UNG2 humanas respectivamente. Además, puede hacerse referencia a UniprotKB N.º P12295 para la secuencia aminoacídica de un derivada de *Escherichia coli* (cepa K-12) y puede hacerse referencia a UniprotKB N.º P20536 para la secuencia aminoacídica de UDG de virus Vaccinia (cepa Copenhagen). La secuencia aminoacídica de UNG está ampliamente conservada entre especies, y el sitio correspondiente para mutación puede identificarse alineando la secuencia aminoacídica de UNG1 o UNG2 derivadas de un organismo deseado con la de UNG1 de levadura mencionada anteriormente. Por ejemplo, en el caso de UNG1 humana, el aminoácido correspondiente a N222 de UNG1 de levadura es la asparagina 204 (N204) y el aminoácido correspondiente a Y164 es la tirosina 147 (Y147). En el caso de un derivada de *Escherichia coli*, el aminoácido correspondiente a N222 de UNG1 de levadura es la asparagina 123 (N123) y el aminoácido correspondiente a Y164 es la tirosina 66 (Y66). En el caso de UDG derivada de virus Poxviridae tales como virus Vaccinia, virus de la viruela, virus de la viruela de simio, virus de la viruela aviar, virus de la viruela porcina y virus de fibroma de conejo, el aminoácido correspondiente a N222 de UNG1 de levadura es la asparagina 120 (N120), y el aminoácido correspondiente a Y164 es la tirosina 70 (Y70).

Aunque UNG puede retirar uracilo de ADN monocatenario y ADN bicatenario, tiene mayor afinidad por ADN monocatenario. Esta tendencia se encuentra también en la UNG mutante anteriormente mencionada dotada de actividad CDG y actividad TDG. Sin embargo, puesto que la citosina y timina existen en todas partes en el ADN genómico, la UNG mutante que tiene actividad CDG o actividad TDG actúa, al contrario que la UNG de tipo silvestre que usa exclusivamente como sustrato uracilo, que raramente se introduce en ADN genómico por un error de replicación o desaminación de citosina, sobre cualquier sitio en la secuencia nucleotídica diana de un módulo de reconocimiento de secuencia de ácido nucleico y ADN bicatenario en los alrededores del mismo para retirar citosina o timina, causando por tanto una fuerte citotoxicidad. Por lo tanto, se requiere que la ADN glicosilasa para usar en la presente invención muestre una reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada. Cuando se usa UNG mutante como ADN glicosilasa, puede evitarse la citotoxicidad por la enzima introduciendo adicionalmente una mutación que disminuye la reactividad con un ADN que tiene una estructura de doble hélice no relajada para dar una reacción de escisión de base por UNG mutante que tiene actividad CDG o actividad TDG más selectiva por la región de un ADN bicatenario relajado o ADN monocatenario (véase la Fig. 1, panel medio). Específicamente, por ejemplo en el caso de UNG, puede mencionarse una mutación que disminuye la reactividad con ADN bicatenario de 25 unidades U-A a no más de 1/20, más preferiblemente no más de 1/50, aún más preferiblemente no más de 1/100 in vitro que la de tipo silvestre. Sin embargo, a condición de disminuir la reactividad con un ADN que tiene una estructura de doble hélice no relajada a un nivel libre de citotoxicidad letal in vivo, no está limitada por la reactividad in vitro. Que la ADN glicosilasa para usar tenga una "reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada" puede confirmarse, por ejemplo, insertando la ADN glicosilasa en el producto de construcción descrito en la Fig. 2, introduciendo el producto de construcción junto con el ARN de guía en la célula objeto mediante el método descrito en el Ejemplo 1, cultivando el transformante obtenido y verificando la capacidad de supervivencia del mismo. Como alternativa, puede cribarse también un candidato a ADN glicosilasa con una "reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada"

evaluando, como uno de los criterios, la reactividad con un oligómero de ADN bicatenario in vitro mediante el método descrito en Chen, C.Y. et al. DNA Repair (Amst) (2005) 4(7): 793-805.

Como ADN glicosilasa que satisface las condiciones anteriormente mencionadas en el caso, por ejemplo, de UNG1 derivada de levadura (SEQ ID NO: 2), puede mencionarse un mutante en que la leucina 304 del extremo N-terminal se sustituye con alanina (se hace referencia al mutante también como L304A) (Slupphaug, G. et al. Nature (1996) 384(6604): 87-92). Como alternativa, un mutante en que la arginina 308 se sustituye por ácido glutámico o cisteína (se hace referencia a los mutantes también como R308E, R308C, respectivamente) muestra también una reactividad notablemente disminuida con un ADN que tiene una estructura de doble hélice no relajada (Chen, C.Y. et al. DNA Repair (Amst) (2005) 4(7): 793-805). Cuando se usa una UNG derivada de una especie diferente, el sitio correspondiente para mutación puede identificarse alineando la secuencia aminoacídica de UNG1 o UNG2 derivadas de un organismo deseado con la de la UNG1 de levadura mencionada anteriormente. Por ejemplo, en el caso de UNG1 humana, el aminoácido correspondiente a L304 de UNG1 de levadura es la leucina 272 (L272), y el aminoácido correspondiente a R308 es la arginina 276 (R276). En el caso de un *Escherichia coli*, el aminoácido correspondiente a L304 de UNG1 de levadura es la leucina 191 (L191) y el aminoácido correspondiente a R308 es la arginina 195 (R195). En el caso de UDG de virus Vaccinia, el aminoácido correspondiente a R308 de UNG1 de levadura es la arginina 187 (R187).

Se muestran en la Tabla 1 la especificidad de sustrato de UNG1 (ung) de levadura (*Saccharomyces cerevisiae*), *Escherichia coli*, humana y de virus Poxviridae, y el sitio de mutación que cambia la reactividad con un ADN que tiene una estructura de doble hélice no relajada. El aminoácido mutante se indica con una letra, y el número muestra la posición del aminoácido mutante cuando el aminoácido N-terminal es 1.

Tabla 1

Levadura	<i>Escherichia coli</i>	Humano	Poxvirus	Cambio de la forma mutante y el fenotipo
Y164	Y66	Y147	Y70	reconoce tiamina por mutación a A, G
N222	N123	N204	N120	reconoce citosina por mutación a D
L304	L191	L272	-	la reactividad con ADN que tiene estructura de doble hélice no relajada disminuye por mutación a A
R308	R195	R276	R187 ^{a)}	la reactividad con ADN que tiene estructura de doble hélice no relajada disminuye por mutación a E, C

a) En UDG de virus Vaccinia etc., el tipo silvestre mismo se considera que tiene una reactividad nativa baja con ADN que tiene estructura de doble hélice no relajada.

A partir de lo anterior, los ejemplos preferidos de “ADN glicosilasa con reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada” para usar en la presente invención incluyen el mutante doble N222D/L304A, el mutante doble N222D/R308E, el mutante doble N222D/R308C, el mutante doble Y164A/L304A, el mutante doble Y164A/R308E, el mutante doble Y164A/R308C, el mutante doble Y164G/L304A, el mutante doble Y164G/R308E, el mutante doble Y164G/R308C, el mutante triple N222D/Y164A/L304A, el mutante triple N222D/Y164A/R308E, el mutante triple N222D/Y164A/R308C, el mutante triple N222D/Y164G/L304A, el mutante triple N222D/Y164G/R308E, el mutante triple N222D/Y164G/R308C y similares de UNG1 de levadura. Cuando se usa una UNG diferente en lugar de UNG1 de levadura, puede usarse un mutante que tiene un aminoácido correspondiente a cada uno de los mutantes anteriormente mencionados, en que se introduce una mutación similar.

Como alternativa, como una “ADN glicosilasa con reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada”, puede usarse también una ADN glicosilasa con reactividad nativa baja con un ADN que tiene una estructura de doble hélice no relajada y alta selectividad por ADN bicatenario relajado o monocatenario. Los ejemplos de tal ADN glicosilasa incluyen SMUG1 (uracilo-ADN glicosilasa monofuncional selectiva de una hebra) que tiene actividad UDG. Aunque no es conocida una mutación que confiera actividad CDG o actividad TDG a SMUG1, se ha reseñado que SMUG1 es importante para la retirada del uracilo resultante de la desaminación de citosina (Nilsen, H. et al. EMBO J. (2001) 20: 4278-4286). Los presentes inventores han desarrollado ya un método para introducir específicamente una mutación en la secuencia nucleotídica diana y los alrededores de la misma combinando citidina desaminasa capaz de convertir citosina en uracilo y un módulo de reconocimiento de secuencia de ácido nucleico similar al de la presente invención (documento WO2015/133554). Al combinar con la técnica, la citosina puede convertirse artificialmente en uracilo en el sitio diana en ADN genómico, después de lo cual SMUG1 puede reaccionar adicionalmente liberando el uracilo del ADN. Aunque la derivación de SMUG1 no está particularmente limitada, por ejemplo, pueden usarse la SMUG1 derivada de *Escherichia coli*, levadura, mamífero (p. ej., humana, de ratón, de cerdo, bovina, de caballo, de mono, etc.) y similares. Por ejemplo, puede hacerse referencia a UniprotKB N.º Q53HC7-1 y Q53HV7-2 para las dos secuencias aminoacídicas de las dos isoformas de SMUG1 humana. Además, aunque la derivación de citidina desaminasa no está particularmente limitada, por ejemplo pueden usarse PmCDAI derivada de *Petromyzon marinus* (citosina desaminasa 1 de *Petromyzon marinus*), o AID (citidina

desaminasa inducida por activación; AICDA) derivada de mamífero (p. ej., humana, de cerdo, bovina, de caballo, de mono, etc.). Por ejemplo, puede hacerse referencia a los n.º de acceso a GenBank EFQ94822 y ABO15149 para la secuencia de bases y la secuencia aminoacídica de ADNc de PmCDAI, y puede hacerse referencia a los n.º de acceso a GenBank NM_020661 y NP_065712 para la secuencia de bases y la secuencia de aminoácidos de ADNc de AID humana.

En otra realización preferible, puede mencionarse como ADN glicosilasa con baja reactividad nativa con un ADN que tiene una estructura de doble hélice no relajada la UDG derivada de virus pertenecientes a los Poxviridae tales como virus Vaccinia y similares. Como se muestra en los ejemplos mencionados a continuación, la UDG derivada de virus Vaccinia (vvUDG) se considera que es de reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada a un nivel libre de toxicidad en una célula hospedadora, porque muestra un crecimiento equivalente al de UNG1 de levadura y un *Escherichia coli* con una mutación introducida que disminuye la reactividad con un ADN que tiene una estructura de doble hélice no relajada (p. ej. R187C), en un medio no selectivo, incluso cuando no se introduce tal mutación. Cuando se usa UDG derivada de virus Poxviridae tales como vvUDG y similares como ADN glicosilasa, puede introducirse adicionalmente una mutación (p. ej., R187C) correspondiente a la mutación que disminuye la reactividad con un ADN que tiene una estructura de doble hélice no relajada en UNG1 de levadura. Sin embargo, cuando tal mutación disminuye la actividad CDG (N120D) y la actividad TDG (Y70A, Y70G) de UDG, deseablemente se evita. A partir de lo anterior, los ejemplos preferibles de UDG derivada de virus Poxviridae tales como vvUDG para usar en la presente invención incluyen el mutante N120D, el mutante Y70G o Y70A, el mutante doble N120D/Y70G o el mutante doble N120D/Y70A y similares.

Cuando se usa como ADN glicosilasa UDG derivada de virus Poxviridae tal como vvUDG mutante y similares, es preferible poner en contacto la proteína A20, que interacciona con UDG formando un heterodímero que funciona como factor de procesividad de ADN polimerasa vírica, con ADN bicatenario junto con UDG. Puesto que el uso combinado con proteína A20 aumenta la actividad CDG o la actividad TDG de UDG mutante, por ejemplo, la eficacia de inducción de mutación a tímica de vvUDG mutante que tiene baja actividad TDG (Y70G) en comparación con la actividad CDG (N120D) puede mejorarse mediante el uso combinado con proteína A20. Aunque la derivación de A20 no está particularmente limitada, por ejemplo, puede usarse A20 derivada de virus pertenecientes a los Poxviridae tales como virus Vaccinia, virus de la viruela, virus de la viruela del simio, virus de la viruela aviar, virus de la viruela porcina y virus de fibroma de conejo. Por ejemplo, puede hacerse referencia a UniprotKB N.º P20995 para la secuencia aminoacídica de A20 de virus Vaccinia (cepa Copenhague).

En aún otra realización preferible, puede usarse una enzima dividida diseñada de tal forma que el módulo de reconocimiento de secuencia de ácido nucleico y la ADN glicosilasa están cada uno dividido en dos fragmentos, cualquier fragmento está ligado entre sí formando dos complejos parciales, estos complejos se asocian para reconstituir un módulo de reconocimiento de secuencia de ácido nucleico funcional y el módulo se une a la secuencia nucleotídica diana para reconstituir una ADN glicosilasa funcional que puede usarse como una ADN glicosilasa que tiene baja reactividad con un ADN que tiene una estructura de doble hélice no relajada. En la enzima dividida, puesto que la actividad enzimática se exhibe solo cuando está unida a la secuencia nucleotídica diana, incluso cuando la ADN glicosilasa misma para reconstituir no tiene una reactividad reducida con un ADN que tiene una estructura de doble hélice no relajada, actúa consiguientemente de forma selectiva sobre la parte de ADN monocatenario o ADN bicatenario relajado en la secuencia nucleotídica diana y los alrededores de la misma. Por ejemplo, el módulo de reconocimiento de secuencia de ácido nucleico y la ADN glicosilasa se dividen cada uno en fragmentos de lado N-terminal y fragmentos de lado C-terminal, por ejemplo los fragmentos de lado N-terminal se ligan entre sí dando un complejo parcial, los fragmentos de lado C-terminal se ligan entre sí dando un complejo parcial (o los fragmentos de lado N-terminal del módulo de reconocimiento de secuencia de ácido nucleico y los fragmentos de lado C-terminal de ADN glicosilasa se ligan dando un complejo parcial y los fragmentos de lado N-terminal de ADN glicosilasa y los fragmentos de lado C-terminal de módulo de reconocimiento de secuencia de ácido nucleico se ligan dando un complejo parcial) y se asocian, con lo que pueden reconstituirse el módulo de reconocimiento de secuencia de ácido nucleico funcional y la ADN glicosilasa funcional. La combinación de los fragmentos para ligar no está particularmente limitada a condición de que se reconstituya un complejo de módulo de reconocimiento de secuencia de ácido nucleico funcional y ADN glicosilasa funcional cuando se asocian los dos complejos parciales. Los dos complejos parciales pueden proporcionarse como moléculas separadas, o pueden proporcionarse como una única proteína de fusión ligándolos directamente o a través de un ligador adecuado. El sitio de división en la ADN glicosilasa no está particularmente limitado a condición de que puedan reconstituirse dos fragmentos divididos como ADN glicosilasa funcional, y la ADN glicosilasa puede dividirse en un sitio proporcionando un fragmento de lado N-terminal y un fragmento de lado C-terminal, o pueden ligarse apropiadamente no menos de 3 fragmentos obtenidos dividiendo en dos o más sitios dando dos fragmentos. Son conocidas las estructuras tridimensionales de diversas proteínas UDG, y los especialistas en la materia pueden seleccionar apropiadamente los sitios de división basándose en tal información. Por ejemplo, la UNG1 de levadura (SEQ ID NO: 2) puede dividirse entre el aminoácido 258 y el aminoácido 259 del extremo N dando un fragmento de lado N-terminal (1-258) y un fragmento de lado C-terminal (259-359).

Como se menciona anteriormente, en el mecanismo de reparación por escisión de base (REB), cuando se escinde una base por ADN glicosilasa, la endonucleasa AP pone una muesca en el sitio abásico (sitio AP) y la exonucleasa escinde completamente el sitio AP. Cuando se escinde el sitio AP, la ADN polimerasa produce una nueva base usando la base de la hebra opuesta como molde, y la ADN ligasa sella finalmente la muesca para completar la reparación. La endonucleasa AP mutante que ha perdido la actividad enzimática, pero mantiene la capacidad de unión al sitio AP, es

conocida por inhibir competitivamente la REB. Por lo tanto, cuando la endonucleasa AP mutante se pone en contacto con ADN bicatenario junto con ADN glicosilasa, se inhibe la reparación del sitio AP por el mecanismo de REB endógeno en la célula hospedadora, y mejora la frecuencia de errores de reparación, a saber, la eficacia de la inducción de mutación. Por ejemplo, la eficacia de la inducción de mutación a timina en UNG1 de levadura mutante que tiene menor actividad TDG (Y164G) en comparación con la actividad CDG (N222D) puede mejorarse usando endonucleasa AP mutante en combinación. Aunque la derivación de endonucleasa AP no está particularmente limitada, por ejemplo, pueden usarse endonucleasa AP derivada de *Escherichia coli*, levadura, mamífero (p. ej., humana, de ratón, de cerdo, bovina, de caballo, de mono, etc.) y similares. Por ejemplo, puede hacerse referencia a UniprotKB N.º P27695 para la secuencia aminoacídica de Ape1 humana. Los ejemplos de endonucleasa AP mutante que ha perdido la actividad enzimática, pero mantiene la capacidad de unión al sitio AP, incluyen proteínas que tienen un sitio de actividad mutado y un sitio de unión a Mg (cofactor) mutado. Por ejemplo, pueden mencionarse E96Q, Y171A, Y171F, Y171H, D210N, D210A, N212A y similares para Ape1 humana.

La secuencia nucleotídica diana en un ADN bicatenario para reconocer por el módulo de reconocimiento de secuencia de ácido nucleico en el complejo enzimático modificador de ácido nucleico de la presente invención no está particularmente limitada, a condición de que el módulo se una específicamente a, y pueda ser cualquier secuencia del ADN bicatenario. La longitud de la secuencia nucleotídica diana solo tiene que ser suficiente para la unión específica del módulo de reconocimiento de secuencia de ácido nucleico. Por ejemplo, cuando se introduce una mutación en un sitio particular en el ADN genómico de un mamífero, no es de menos de 12 nucleótidos, preferiblemente no menos de 15 nucleótidos, más preferiblemente no menos de 17 nucleótidos, según el tamaño de genoma del mismo. Aunque el límite superior de longitud no está particularmente limitado, preferiblemente es de no más de 25 nucleótidos, más preferiblemente no más de 22 nucleótidos.

Como módulo de reconocimiento de secuencia de ácido nucleico en el complejo enzimático modificador de ácido nucleico de la presente invención, pueden usarse: sistema CRISPR-Cas en el que está inactivada al menos una capacidad de corte de ADN de Cas (Cas mutante de CRISPR), motivo de dedo de cinc, efector TAL y motivo PPR y similares, así como un fragmento que contiene un dominio de unión a ADN de una proteína que se une específicamente a ADN, tal como enzima de restricción, factor de transcripción, ARN polimerasa y similares, y libre de capacidad de corte de doble hebra de ADN y similares, pero el módulo no está limitado a los mismos. Preferiblemente, pueden mencionarse Cas mutante de CRISPR, motivo de dedo de cinc, efector Tal, motivo PPAR y similares.

Un motivo de dedo de cinc está constituido por el ligamiento de 3-6 unidades de dedo de cinc de tipo Cys2His2 diferentes (1 dedo reconoce aproximadamente 3 bases), y puede reconocer una secuencia nucleotídica diana de 9-18 bases. Un motivo de dedo de cinc puede producirse mediante un método conocido tal como el método de ensamblado modular (Nat Biotechnol (2002) 20: 135-141), el método OPEN (Mol Cell (2008) 31: 294-301), el método CoDA (Nat Methods (2011) 8: 67-69), el método de un híbrido de *Escherichia coli* (Nat Biotechnol (2008) 26: 695-701) y similares. Puede hacerse referencia al documento de patente 1 anteriormente mencionado para detalles de la producción de motivo de dedo de cinc.

Un efector Tal tiene una estructura repetida modular con aproximadamente 34 aminoácidos como unidad, y los residuos aminoacídicos 12 y 13 (llamados RVD) de un módulo determinan la estabilidad de unión y especificidad de base. Puesto que cada módulo es altamente independiente, puede producirse el efector TAL específico de una secuencia nucleotídica diana simplemente conectando el módulo. Para el efector TAL, se ha establecido un método de producción que utiliza un recurso libre (método REAL (Curr Protoc Mol Biol (2012) Capítulo 12: Unidad 12.15), método FLASH (Nat Biotechnol (2012) 30: 460-465) y método Golden Gate (Nucleic Acids Res (2011) 39: e82) etc.), y puede diseñarse comparativamente de forma conveniente un efector TAL para una secuencia nucleotídica diana. Puede hacerse referencia al documento de patente 2 anteriormente mencionado para los detalles de la producción del efector TAL.

El motivo PPR está constituido de tal forma que se reconoce una secuencia nucleotídica particular por una continuación de motivos PPR consistente cada uno en 35 aminoácidos y que reconocen una base de ácido nucleico, y reconoce una base diana solo a 1, 4 y ii (-2) aminoácidos de cada motivo. La constitución de motivos no tiene dependencia y está libre de interferencia de motivos en ambos lados. Por lo tanto, como el efector TAL, puede producirse una proteína PPR específica de la secuencia nucleotídica diana simplemente conectando los motivos PPR. Puede hacerse referencia al documento de patente 4 anteriormente mencionado para detalles de la producción del motivo PPR.

Cuando se usa un fragmento de enzima de restricción, factor de transcripción, ARN polimerasa y similares, puesto que los dominios de unión a ADN de estas proteínas son bien conocidos, puede diseñarse y construirse fácilmente un fragmento que contiene el dominio y está libre de la capacidad de corte de doble hebra de ADN.

Como se menciona anteriormente, la ADN glicosilasa usada para el complejo enzimático modificador de ácido nucleico de la presente invención es preferiblemente UDG mutante a la que se ha conferido actividad CDG o actividad TDG, más preferiblemente UNG, y tiene que ser de reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada de forma que la actividad CDG o actividad TDG no actúe sobre cualquier lugar en el sitio diana (véase la Fig. 1, panel medio). Por lo tanto, el sitio diana está deseablemente en estado de ADN monocatenario, o la estructura de ADN relajada resultante del desenrollamiento de al menos una estructura de doble hélice consistente,

lo que posibilita a la ADN glicosilasa actuar eficazmente cuando se pone en contacto con ADN glicosilasa. Cuando se usa el sistema CRISPR-Cas como módulo de reconocimiento de secuencia de ácido nucleico, el ARN de guía complementario de la secuencia nucleotídica diana reconoce la secuencia del ADN bicatenario objeto y forma específicamente un híbrido con la secuencia nucleotídica diana, con lo que el sitio diana adquiere el estado de estructura monocatenaria o el estado de doble hélice desenrollada (bicatenario relajado). Por lo tanto, una ADN glicosilasa con reactividad suficientemente baja con un ADN que tiene estructura de doble hélice no relajada actúa selectivamente sobre citosina o timina en el sitio diana y puede escindir la base (véase la Fig. 1, panel inferior). Por otro lado, cuando se usan motivo de dedo de cinc, efector TAL, motivo PPR y similares como módulo de reconocimiento de secuencia de ácido nucleico, puesto que el módulo mismo no tiene la función de cambiar la estructura de ADN bicatenario (causar la distorsión de la estructura de doble hélice), es deseable poner en contacto el complejo enzimático modificador de ácido nucleico de la presente invención en combinación con un factor (p. ej., girasa, topoisomerasa, helicasa, etc.) que cambie la estructura del ADN bicatenario objeto, con el ADN bicatenario.

Cualquiera de los módulos de reconocimiento de secuencia de ácido nucleico anteriormente mencionados puede proporcionarse como una proteína de fusión con la ADN glicosilasa anteriormente mencionada, o un dominio de unión a proteína tal como un dominio SH3, dominio PDZ, dominio GK, dominio GB y similares y puede fusionarse un copartícipe de unión del mismo con un módulo de reconocimiento de secuencia de ácido nucleico y una ADN glicosilasa, respectivamente, y proporcionarse como un complejo proteico a través de la interacción del dominio y un copartícipe de unión del mismo. Como alternativa, pueden fusionarse cada uno del módulo de reconocimiento de secuencia de ácido nucleico y ADN glicosilasa con inteína, y pueden ligarse por ligadura después de síntesis de proteína.

Se pone en contacto el complejo enzimático modificador de ácido nucleico de la presente invención que contiene un complejo (incluyendo proteína de fusión) en el que se unen un módulo de reconocimiento de secuencia de ácido nucleico y ADN glicosilasa con un ADN bicatenario (p. ej., ADN genómico) introduciendo el complejo o un ácido nucleico que codifica el complejo en una célula que tiene el ADN bicatenario objeto. En consideración de la eficacia de introducción y expresión, es deseable introducir el complejo enzimático modificador de ácido nucleico en la célula en forma de un ácido nucleico que codifica el complejo, en lugar del complejo mismo, y permitir la expresión del complejo en la célula. También, cuando se usan en combinación endonucleasa AP mutante, proteína A20 y similares, es deseable introducir las en la célula en forma de un ácido nucleico que las codifica, y permitir la expresión del complejo en la célula.

Por lo tanto, el módulo de reconocimiento de secuencia de ácido nucleico y la ADN glicosilasa se preparan preferiblemente como un ácido nucleico que codifica una proteína de fusión de los mismos, o en una forma capaz de formar un complejo en una célula hospedadora después de traducción en una proteína utilizando un dominio de unión, inteína y similares, o como un ácido nucleico que codifica cada una de ellas. El ácido nucleico aquí puede ser un ADN o un ARN. Cuando es un ADN, es preferiblemente un ADN bicatenario, y se proporciona en forma de un vector de expresión dispuesto bajo la regulación de un promotor funcional en una célula hospedadora. Cuando es un ARN, es preferiblemente un ARN monocatenario.

Cuando se dividen cada uno del módulo de reconocimiento de secuencia de ácido nucleico y la ADN glicosilasa en dos fragmentos, se ligan respectivamente los fragmentos de cualquiera de ellos con los fragmentos del otro proporcionando dos complejos parciales, por ejemplo, se preparan respectivamente un ADN que codifica el fragmento de lado N-terminal y un ADN que codifica el fragmento de lado C-terminal del módulo de reconocimiento de secuencia de ácido nucleico mediante el método de PCR usando cebadores adecuados; y se preparan un ADN que codifica el fragmento de lado N-terminal y un ADN que codifica el fragmento de lado C-terminal de la ADN glicosilasa de la misma manera y, por ejemplo, se ligan los ADN que codifican los fragmentos de lado N-terminal y los ADN que codifican los fragmentos de lado C-terminal entre sí mediante un método convencional, con lo que puede producirse un ADN que codifica los dos complejos parciales. Como alternativa, se ligan un ADN que codifica el fragmento de lado N-terminal del módulo de reconocimiento de secuencia de ácido nucleico y un ADN que codifica el fragmento de lado C-terminal de la ADN glicosilasa; y se ligan un ADN que codifica el fragmento de lado N-terminal de la ADN glicosilasa y un ADN que codifica el fragmento de lado C-terminal del módulo de reconocimiento de secuencia de ácido nucleico, con lo que puede producirse también un ADN que codifica los dos complejos parciales. La combinación de los fragmentos para ligar no está particularmente limitada, a condición de que se reconstituya el complejo de módulo de reconocimiento de secuencia de ácido nucleico funcional y la ADN glicosilasa funcional cuando se asocian los dos complejos parciales. Los dos complejos parciales no solo se expresan como moléculas separadas, sino que pueden expresarse también como una única proteína de fusión ligando ácidos nucleicos que los codifican directamente o través de un ligador adecuado, cuya proteína forma un complejo de módulo de reconocimiento de secuencia de ácido nucleico funcional y ADN glicosilasa funcional por asociación intramolecular.

Puesto que el complejo de la presente invención en el que se unen un módulo de reconocimiento de secuencia de ácido nucleico y una ADN glicosilasa no está acompañado de roturas de ADN bicatenario (DSB), es posible una edición genómica con baja toxicidad, y el método de modificación genética de la presente invención puede aplicarse a un amplio intervalo de materiales biológicos. Por lo tanto, las células para introducir ácido nucleico que codifica módulo de reconocimiento de secuencia de ácido nucleico y/o ADN glicosilasa pueden englobar células de cualquier especie, de bacteria *Escherichia coli* y similares que sean procariontes, células de microorganismos tales como levaduras y similares que sean eucariotas inferiores, hasta células de vertebrados incluyendo mamíferos tales como seres

humanos y similares, y células de eucariotas superiores tales como insectos, plantas y similares.

5 Puede obtenerse un ADN que codifica un módulo de reconocimiento de secuencia de ácido nucleico, tal como motivo de dedo de cinc, efector TAL, motivo PPR y similares, mediante cualquier método mencionado anteriormente para cada módulo. Puede clonarse un ADN que codifica un módulo de reconocimiento de secuencia de enzima de restricción, factor de transcripción, ARN polimerasa y similar, por ejemplo, sintetizando un cebador de oligoADN que cubre una región que codifica la parte deseada de la proteína (parte que contiene el dominio de unión a ADN) basándose en la información de secuencia de ADNc del mismo, y amplificando mediante el método de RT-PCR usando, como molde, el ARN total o la fracción de ARNm preparada a partir de las células productoras de proteína.

10 Puede clonarse también de forma similar un ADN que codifica ADN glicosilasa sintetizando un cebador de oligoADN basado en la información de secuencia de ADNc del mismo y amplificando mediante el método de RT-PCR usando, como molde, el ARN total o la fracción de ARNm preparada a partir de las células productoras de enzima. Por ejemplo, puede clonarse un ADN que codifica UNG1 de levadura diseñando cebadores adecuados en la dirección 5' y dirección 3' de CDS basándose en la secuencia de ADNc (nº de acceso NM_001182379) registrada en la base de datos NCBI, y clonando a partir de ARN derivado de levadura mediante el método RT-PCR.

15 Puede obtenerse un ácido nucleico que codifica ADN glicosilasa con reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada mediante un método de mutagénesis específica de sitio conocido por se usando el ADNc obtenido como molde, e introduciendo una mutación que confiere actividad CDG, actividad TDG o actividad 5-mCDG, y una mutación que disminuye la reactividad con un ADN que tiene una estructura de doble hélice no relajada. Cuando la ADN glicosilasa tiene una reactividad nativa suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada, tal como vVUDG y similares, puede introducirse solo una mutación que confiere actividad CDG, actividad TDG o actividad 5-mCDG.

20 El ADN clonado puede ligarse directamente, o después de digestión con una enzima de restricción cuando se desee, o después de adición de un ligador adecuado (p. ej., ligador GS, ligador GGGAR, etc.), espaciador (p. ej., secuencia FLAG, etc.) y/o una señal de localización nuclear (NLS) (cada señal de localización de orgánulo cuando el ADN bicatenario objeto es ADN mitocondrial o de cloroplasto), con un ADN que codifica un módulo de reconocimiento de secuencia de ácido nucleico para preparar un ADN que codifica una proteína de fusión. Puesto que UNG1 y UNG2 tienen cada una una señal de localización mitocondrial y una señal de localización nuclear en el extremo N, pueden utilizarse también tal cual. Como alternativa, por ejemplo, cuando se usa UNG1 para modificación nucleotídica orientada a ADN genómico nuclear, es posible retirar la señal de localización mitocondrial y ligar separadamente una señal de localización nuclear.

25 Como alternativa, pueden fusionarse cada uno de un ADN que codifica un módulo de reconocimiento de secuencia de ácido nucleico y un ADN que codifica una ADN glicosilasa con un ADN que codifica un dominio de unión a ADN o un copartícipe de unión del mismo, o ambos ADN pueden fusionarse con un ADN que codifica una inteína de separación, con lo que el módulo de conversión de reconocimiento de secuencia de ácido nucleico y la ADN glicosilasa se traducen en una célula hospedadora formando un complejo. En estos casos, puede ligarse un ligador y/o una señal de localización nuclear en una posición adecuada de uno o ambos ADN cuando se desee.

30 Pueden obtenerse un ADN que codifica un módulo de reconocimiento de secuencia de ácido nucleico y un ADN que codifica una ADN glicosilasa sintetizando químicamente la hebra de ADN, o conectando hebras cortas de oligoADN parcialmente superpuestas sintetizadas utilizando el método de PCR y el método de ensamblaje de Gibson para construir un ADN que codifica toda su longitud. La ventaja de construir un ADN de longitud completa mediante síntesis química o una combinación del método de PCR o el método de ensamblaje de Gibson es que el codón para usar puede diseñarse en la longitud completa de CDS según el hospedador en que se introduce el ADN. En la expresión de un ADN heterólogo, se espera que aumente el nivel de expresión de proteína al convertir la secuencia de ADN de la misma en un codón usado con alta frecuencia en el organismo hospedador. Como los datos de frecuencia de uso de codón en el hospedador para usar, por ejemplo, puede usarse la base de datos de frecuencia de uso del código genético (<http://www.kazusa.or.jp/codon/index.html>) divulgada en la página del Kazusa DNA Research Institute, o puede hacerse referencia a documentos que muestran la frecuencia de uso de codón en cada hospedador. Como referencia a los datos obtenidos y la secuencia de ADN para introducir, los codones que muestran baja frecuencia de uso en el hospedador entre los usados para la secuencia de ADN pueden convertirse en un codón que codifica el mismo aminoácido y muestra alta frecuencia de uso.

35 Puede producirse un vector de expresión que contiene un ADN que codifica un módulo de reconocimiento de secuencia de ácido nucleico y/o una ADN glicosilasa, por ejemplo, ligando el ADN en dirección 3' de un promotor en un vector de expresión adecuado.

40 Como vector de expresión, se usan plásmidos derivados de Escherichia coli (p. ej., pBR322, pBR325, pUC12, pUC13); plásmidos derivados de Bacillus subtilis (p. ej., pUB110, pTP5, pC194); plásmidos derivados de levadura (p. ej., pSH19, pSH15); plásmidos de expresión de células de insecto (p. ej., pFast-Bac); plásmidos de expresión de células animales (p. ej., pA1-11, pXT1, pRc/CMV, pRc/RSV, pcDNA1/Neo); bacteriófagos tales como fago λ y similares; vectores de virus de insectos tales como baculovirus y similares (p. ej., BmNPV, AcNPV); vectores de virus animales tales como retrovirus, virus Vaccinia, adenovirus y similares, y similares.

- 5 Como promotor, puede usarse cualquier promotor adecuado para usar por un hospedador para expresión génica. En un método convencional que usa DSB, puesto que la tasa de supervivencia de la célula hospedadora a veces disminuye notablemente debido a la toxicidad, es deseable aumentar el número de células al inicio de la inducción usando un promotor inductivo. Sin embargo, puesto que puede procurarse también suficiente proliferación celular expresando el complejo enzimático modificador de ácido nucleico de la presente invención, puede usarse también un promotor constitutivo sin limitación.
- 10 Por ejemplo, cuando el hospedador es una célula animal, se usan el promotor SR α , el promotor SV40, el promotor LTR, el promotor CMV (citomegalovirus), el promotor RSV (virus de sarcoma de Rous), el promotor LTR de MoMuLV (virus de leucemia de ratón de Moloney), el promotor HSV-TK (timidina cinasa de herpesvirus simple) y similares. De estos, son preferibles el promotor CMV, el promotor SR α y similares.
- 15 Cuando el hospedador es *Escherichia coli*, son preferibles el promotor trp, el promotor lac, el promotor recA, el promotor AP, el promotor lpp, el promotor T7 y similares.
- 20 Cuando el hospedador es del género *Bacillus*, son preferibles el promotor SPO1, el promotor SPO2, el promotor penP y similares.
- 25 Cuando el hospedador es una levadura, son preferibles el promotor Gall/10, el promotor PHO5, el promotor PGK, el promotor GAP, el promotor ADH y similares.
- 30 Cuando el hospedador es una célula de insecto, son preferibles el promotor de polihedrina, el promotor P10 y similares.
- 35 Cuando el hospedador es una célula vegetal, son preferibles el promotor CaMV35S, el promotor CaMV19S, el promotor NOS y similares.
- 40 Como vector de expresión, aparte de aquellos mencionados anteriormente, pueden usarse uno que contenga potenciador, señal de corte y empalme, terminador, señal de adición de poliA, marcador de selección tal como gen de resistencia a fármaco, gen complementario auxotrófico y similares, origen de replicación y similares a petición.
- 45 Pueden prepararse un ARN que codifica un módulo de reconocimiento de secuencia de ácido nucleico y/o una ADN glicosilasa, por ejemplo, mediante transcripción a ARNm en un sistema de transcripción in vitro conocido per se por usar un vector que codifica ADN que codifica el módulo de reconocimiento de secuencia de ácido nucleico mencionado anteriormente y/o una ADN glicosilasa como molde.
- 50 Puede expresarse intracelularmente un complejo de módulo de reconocimiento de secuencia de ácido nucleico y una ADN glicosilasa introduciendo un vector de expresión que contiene un ADN que codifica un módulo de reconocimiento de secuencia de ácido nucleico y/o una ADN glicosilasa en una célula hospedadora, y cultivando la célula hospedadora.
- 55 Como hospedador, se usan células del género *Escherichia*, del género *Bacillus*, de levadura, insectos, células animales y similares.
- 60 Como del género *Escherichia*, se usan *Escherichia coli* K12:DH1 [Proc.Natl. Acad. Sci. USA, 60, 160 (1968)], *Escherichia coli* JM103 [Nucleic Acids Research, 9, 309 (1981)], *Escherichia coli* JA221 [Journal of Molecular Biology, 120, 517 (1978)], *Escherichia coli* HB101 [Journal of Molecular Biology, 41, 459 (1969)], *Escherichia coli* C600 [Genetics, 39, 440 (1954)] y similares.
- 65 Como del género *Bacillus*, se usan *Bacillus subtilis* M1114 [Gene, 24, 255 (1983)], *Bacillus subtilis* 207-21 [Journal of Biochemistry, 95, 87 (1984)] y similares.
- 70 Como levadura, se usan *Saccharomyces cerevisiae* AH22, AH22R, NA87-11A, DKD-5D, 20B-12, *Schizosaccharomyces pombe* NCYC1913, NCYC2036, *Pichia pastoris* KM71 y similares.
- 75 Como célula de insecto cuando el virus es AcNPV, se usan células de la estirpe establecida derivada de larva de cogollero del maíz (célula de *Spodoptera frugiperda*; célula Sf), células MG1 derivadas del intestino medio de *Trichoplusia ni*, células High Five™ derivadas de un huevo de *Trichoplusia ni*, células derivadas de *Mamestra brassicae*, células derivadas de *Estigmene acrea* y similares. Cuando el virus es BmNPV, se usan células de una estirpe establecida derivada de *Bombyx mori* (célula N de *Bombyx mori*; célula BmN) y similares como células de insecto. Como célula Sf, se usan por ejemplo célula Sf9 (ATCC CRL1711), célula Sf21 [todos los anteriores, In Vivo, 13, 213-217 (1977)] y similares.
- 80 Como insecto, se usan por ejemplo larvas de *Bombyx mori*, *Drosophila*, grillo y similares [Nature, 315, 592 (1985)].
- 85 Como célula animal, se usan estirpes celulares tales como célula COS-7 de mono, célula Vero de mono, célula de ovario de hámster chino (CHO), célula CHO deficiente en el gen dhfr, célula L de ratón, célula AtT-20 de ratón, célula de mieloma de ratón, célula GH3 de rata, célula FL humana y similares, citoblastos pluripotentes tales como célula iPS, célula ES y similares de seres humanos y otros mamíferos, y células cultivadas primarias preparadas a partir de diversos tejidos. Además, pueden usarse también embrión de pez cebra, ovocitos de *Xenopus* y similares.

Como célula vegetal, se usan células cultivadas suspendidas, callo, protoplasto, segmento de hoja, segmento de raíz y similares preparados a partir de diversas plantas (p. ej., grano tal como arroz, trigo, maíz y similares, cosechas de producto tales como tomate, pepino, berenjena y similares, plantas de jardín tales como clavel, *Eustoma russellianum* y similares, plantas experimentales tales como tabaco, *Arabidopsis thaliana* y similares).

5 Todas las células hospedadoras anteriormente mencionadas pueden ser haploides (monoploides) o poliploides (p. ej., diploides, triploides, tetraploides y similares). En los métodos de introducción de mutación convencionales, la mutación se introduce en principio en solo un cromosoma homólogo para producir un tipo heterogénico. Por lo tanto, el fenotipo deseado no se expresa a menos que aparezca una mutación dominante, y la homocigosidad requiere inconvenientemente trabajo y tiempo. En contraposición, según la presente invención, puesto que la mutación puede
10 introducirse en cualquier alelo del cromosoma homólogo del genoma, el fenotipo deseado puede expresarse en una sola generación incluso en caso de mutación recesiva, lo que es extremadamente útil puesto que puede resolver el problema del método convencional.

Puede introducirse un vector de expresión mediante un método conocido (p. ej., método de lisozima, método competente, método de PEG, método de coprecipitación con CaCl_2 , método de electroporación, método de microinyección, método de pistola de partículas, método de lipofección, método de *Agrobacterium* y similares) según
15 la clase del hospedador.

Escherichia coli puede transformarse según los métodos descritos, por ejemplo, en Proc. Natl. Acad. Sci. USA, 69, 2110 (1972), Gene, 17, 107 (1982) y similares.

20 El género *Bacillus* puede introducirse en un vector según los métodos descritos, por ejemplo, en Molecular & General Genetics, 168, 111 (1979) y similares.

Una levadura puede introducirse en un vector según los métodos descritos, por ejemplo, en Methods in Enzymology, 194, 182-187 (1991), Proc. Natl. Acad. Sci. USA, 75, 1929 (1978) y similares.

Una célula de insecto y un insecto pueden introducirse en un vector según los métodos descritos, por ejemplo, en Bio/Technology, 6, 47-55 (1988) y similares.

25 Una célula animal puede introducirse en un vector según los métodos descritos, por ejemplo, en Cell Engineering, volumen adicional 8, New Cell Engineering Experiment Protocol, 263-267 (1995) (publicado por Shujunsha) y Virology, 52, 456 (1973).

Una célula con vector introducido puede cultivarse según un método conocido según la clase de hospedador.

30 Por ejemplo, cuando se cultiva *Escherichia coli* o el género *Bacillus*, es preferible un medio líquido como medio para usar para cultivo. El medio contiene preferiblemente una fuente de carbono, fuente de nitrógeno, sustancias inorgánicas y similares necesarios para el crecimiento del transformante. Los ejemplos de fuente de carbono incluyen glucosa, dextrina, almidón soluble, sacarosa y similares; los ejemplos de fuente de nitrógeno incluyen sustancias inorgánicas u orgánicas tales como sales de amonio, sales nitrato, licor de maíz fermentado, peptona, caseína, extracto de carne, torta de soja, extracto de patata y similares; y los ejemplos de sustancias inorgánicas incluyen
35 cloruro de calcio, dihidrogenofosfato de sodio, cloruro de magnesio y similares. El medio puede contener extracto de levadura, vitaminas, factor promotor del crecimiento y similares. El pH del medio es preferiblemente de aproximadamente 5 a aproximadamente 8.

40 Como medio para cultivar *Escherichia coli*, por ejemplo, es preferible medio M9 que contiene glucosa y casaminoácidos [Journal of Experiments in Molecular Genetics, 431-433, Cold Spring Harbor. Laboratory, Nueva York 1972]. Cuando sea necesario, pueden añadirse por ejemplo agentes tales como ácido 3 β -indolilacrílico al medio para asegurar una función eficaz de un promotor. *Escherichia coli* se cultiva generalmente a aproximadamente 15 a aproximadamente 43 °C. Cuando sea necesario, pueden efectuarse aireación y agitación.

El género *Bacillus* se cultiva generalmente a aproximadamente 30 a aproximadamente 40 °C. Cuando sea necesario, pueden efectuarse aireación y agitación.

45 Los ejemplos de medio para cultivar levadura incluyen medio mínimo de Burkholder [Proc. Natl. Acad. Sci. USA, 77, 4505 (1980)], medio SD que contiene 0,5 % de casaminoácidos [Proc. Natl. Acad. Sci. USA, 81, 5330 (1984)] y similares. El pH del medio es preferiblemente de aproximadamente 5 a aproximadamente 8. El cultivo se efectúa generalmente a aproximadamente 20 a aproximadamente 35 °C. Cuando sea necesario, pueden efectuarse aireación y agitación.

50 Como medio para cultivar una célula de insecto o insecto, se usa por ejemplo medio de insecto de Grace [Nature, 195, 788 (1962)] que contiene un aditivo tal como suero bovino al 10 % inactivado y similares según sea apropiado, y similares. El pH del medio es preferiblemente de aproximadamente 6,2 a aproximadamente 6,4. El cultivo se efectúa generalmente a aproximadamente 27 °C. Cuando sea necesario, puede efectuarse aireación y agitación.

Como medio para cultivar una célula animal, se usan por ejemplo medio esencial mínimo (MEM) que contiene

aproximadamente 5 a aproximadamente 20 % de suero fetal bovino [Science, 122, 501 (1952)], medio Eagle modificado por Dulbecco (DMEM) [Virology, 8, 396 (1959)], medio RPMI 1640 [The Journal of the American Medical Association, 199, 519 (1967)], medio 199 [Proceeding of the Society for the Biological Medicine, 73, 1 (1950)] y similares. El pH del medio es preferiblemente de aproximadamente 6 a aproximadamente 8. El cultivo se efectúa generalmente a aproximadamente 30° C a aproximadamente 40 °C. Cuando sea necesario, pueden efectuarse aireación y agitación.

Como medio para cultivar una célula vegetal, se usan por ejemplo medio M5, medio LS, medio B5 y similares. El pH del medio es preferiblemente de aproximadamente 5 a aproximadamente 8. El cultivo se efectúa generalmente a aproximadamente 20 a aproximadamente 30 °C. Cuando sea necesario, puede efectuarse aireación y agitación.

Como se menciona anteriormente, puede expresarse intracelularmente un complejo de un módulo de reconocimiento de secuencia de ácido nucleico y una ADN glicosilasa, concretamente complejo enzimático modificador de ácido nucleico.

Puede introducirse un ARN que codifica un módulo de reconocimiento de secuencia de ácido nucleico y/o una ADN glicosilasa en una célula hospedadora mediante el método de microinyección, el método de lipofección y similares. La introducción de ARN puede efectuarse una vez o múltiples veces repetidas (p. ej., 2-5 veces) a intervalos adecuados.

Cuando se expresa un complejo de un módulo de reconocimiento de secuencia de ácido nucleico y una ADN glicosilasa por un vector de expresión o molécula de ARN introducida en la célula, el módulo de reconocimiento de secuencia de ácido nucleico reconoce específicamente y se une a una secuencia nucleotídica diana en el ADN bicatenario (p. ej., ADN genómico) de interés y, debido a la acción de la ADN glicosilasa ligada al módulo de reconocimiento de secuencia de ácido nucleico, aparece una reacción de escisión de base en la hebra codificante o la hebra anticodificante del sitio diana (la secuencia nucleotídica diana total o parcial o ajustada apropiadamente dentro de varios cientos de bases que incluyen los alrededores de la misma) y se produce un sitio abásico (sitio AP) en una de las hebras del ADN bicatenario. Entonces, actúa el sistema de reparación de escisión de base (REB) de la célula, la endonucleasa AP reconoce primero el sitio AP y corta el enlace de ácido fosfórico en una de las hebras de ADN y la exonucleasa retira el nucleótido sometido a escisión de base. Entonces, la ADN polimerasa inserta un nuevo nucleótido usando la hebra opuesta de ADN como molde y finalmente la ADN ligasa repara la unión. Se introducen diversas mutaciones por un error de reparación que aparezca en cualquier paso de esta REB. Como se menciona anteriormente, el mecanismo de REB en la célula está inhibido, y la frecuencia de error de reparación, y por tanto la eficacia de inducción de mutación, puede mejorarse usando una endonucleasa AP mutante que perdió la actividad enzimática, pero retiene la capacidad de unión al sitio AP en combinación.

En cuanto al motivo de dedo de cinc, la producción de muchos motivos de dedo de cinc realmente funcionales no es fácil, puesto que la eficacia de producción de un dedo de cinc que se una específicamente a una secuencia nucleotídica diana no es alta y la selección de un dedo de cinc que tenga alta especificidad de unión es complicada. Aunque el efector TAL y el motivo PPR tienen un alto grado de libertad de reconocimiento de secuencia de ácido nucleico diana en comparación con el motivo de dedo de cinc, permanece el problema de la eficacia, puesto que tiene que diseñarse y construirse cada vez una proteína grande según la secuencia nucleotídica diana. Además, puesto que estos módulos de reconocimiento de secuencia de ácido nucleico no tienen la función de cambiar la estructura de ADN bicatenario (causando torsión en la estructura de doble hélice) para que una ADN glicosilasa con reactividad suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada actúe eficazmente sobre el sitio diana, es necesario poner en contacto separadamente un factor que cambie la estructura del ADN bicatenario con el ADN bicatenario objeto, haciendo así complicada la operación.

En contraposición, puesto que el sistema CRISPR-Cas reconoce la secuencia de ADN bicatenario objeto por un ARN de guía complementario de la secuencia nucleotídica diana, puede orientarse a cualquier secuencia simplemente sintetizando un oligoADN capaz de formar específicamente un híbrido con la secuencia nucleotídica diana. Además, en el sitio diana, puesto que el ADN bicatenario se desenrolla para generar una región con estructura monocatenaria, y una región adyacente a la misma que tiene una estructura de ADN bicatenario relajado, puede hacerse actuar eficazmente la ADN glicosilasa de manera específica de sitio diana, sin combinar factores que cambien la estructura del ADN bicatenario.

Por lo tanto, en una realización más preferible de la presente invención, se usa un sistema CRISPR-Cas en el que se inactiva al menos una capacidad de unión de ADN de Cas (Cas mutante de CRISPR) como módulo de reconocimiento de secuencia de ácido nucleico.

La Fig. 2 es un esquema que muestra un método de modificación de ADN bicatenario de la presente invención que usa Cas mutante de CRISPR como módulo de reconocimiento de secuencia de ácido nucleico.

El módulo de reconocimiento de secuencia de ácido nucleico que usa Cas mutante de CRISPR se proporciona como un complejo de una molécula de ARN consistente en un ARN de guía complementario de la secuencia nucleotídica diana y el ARNtracr necesario para agrupar la proteína Cas mutante, y una proteína Cas mutante.

La proteína Cas para usar en la presente invención no está particularmente limitada, a condición de que pertenezca al sistema CRISPR, y se prefiere Cas9. Los ejemplos de Cas9 incluyen, pero sin limitación, Cas9 derivada de

Streptococcus pyogenes (SpCas9), Cas9 derivada de Streptococcus thermophilus (StCas9) y similares. Se prefiere SpCas9. Como Cas mutante para usar en la presente invención, puede usarse cualquier Cas en la que la capacidad de corte de ambas hebras del ADN bicatenario esté inactivada y una que tenga actividad nicasa en la que al menos una capacidad de corte de una hebra sola esté inactivada. Por ejemplo, en el caso de SpCas9, puede usarse un mutante D10A en que el residuo Asp 10 se convierte en un residuo Ala y carece de la capacidad de corte de la hebra opuesta a la hebra que forma una hebra complementaria con un ARN de guía, o un mutante H840A en que el residuo His 840 se convierte en un residuo Ala y carece de la capacidad de corte de hebra complementaria al ARN de guía o un doble mutante de los mismos, y pueden usarse de forma similar otras Cas mutantes.

La ADN glicosilasa se proporciona como un complejo con Cas mutante mediante un método similar al esquema de acoplamiento con dedo de cinc anteriormente mencionado y similares. Como alternativa, pueden unirse también una ADN glicosilasa y Cas mutante utilizando los aptámeros de ARN MS2F6, PP7 y similares y el armazón de ARN uniendo proteínas al mismo. El ARN de guía forma una hebra complementaria con la secuencia nucleotídica diana, la Cas mutante se agrupa por el ARNtracr enlazado y la Cas mutante reconoce la secuencia de reconocimiento PAM del sitio de corte de ADN (motivo adyacente al protoespaciador) (cuando se usa SpCas9, PAM es las 3 bases NGG (N es cualquier base) y, teóricamente, puede orientarse a cualquier posición en el genoma). Uno o ambos ADN no pueden cortarse y, debido a la acción de la ADN glicosilasa ligada a Cas mutante, aparece escisión de base en el sitio diana (ajustada apropiadamente dentro de varios cientos de bases que incluyen la secuencia nucleotídica diana total o parcial) y aparece un sitio AP en el ADN bicatenario. Se introducen diversas mutaciones debido a los errores cometidos por el sistema REB de la célula para reparar (véase, por ejemplo, la Fig. 5).

Incluso cuando se usa Cas mutante de CRISPR como módulo de reconocimiento de secuencia de ácido nucleico, se introducen deseablemente un módulo de reconocimiento de secuencia de ácido nucleico y una ADN glicosilasa en forma de un ácido nucleico que codifica los mismos en una célula que tiene un ADN bicatenario de interés, de forma similar a cuando se usan dedo de cinc y similares como módulo de reconocimiento de secuencia de ácido nucleico.

Puede clonarse un ADN que codifica Cas mediante un método similar al método anteriormente mencionado para un ADN que codifica una ADN glicosilasa, a partir de una célula productora de la enzima. Puede obtenerse una Cas mutante introduciendo una mutación para convertir un residuo aminoacídico de la parte importante para actividad de corte de ADN (p. ej., el residuo Asp 10 y el residuo His 840 para Cas9, aunque sin limitación) en otro aminoácido en un ADN que codifica Cas clonada, mediante un método de inducción de mutación específico de sitio conocido per se.

Como alternativa, puede construirse también un ADN que codifica una Cas mutante como un ADN que muestra un uso de codón adecuado para expresión en la célula hospedadora para usar, mediante un método similar a los mencionados anteriormente para un ADN que codifica un módulo de reconocimiento de secuencia de ácido nucleico y un ADN que codifica una ADN glicosilasa, y mediante una combinación de síntesis química o método de PCR o método de ensamblaje de Gibson. Por ejemplo, se muestran la secuencia de CDS y la secuencia aminoacídica optimizada para expresión de SpCas9 en células eucarióticas en las SEQ ID NO: 3 y 4. En la secuencia mostrada en la SEQ ID NO: 3, cuando "A" se convierte en "C" en la base n.^o 29, puede obtenerse un ADN que codifica un mutante D10A, y cuando "CA" se convierte en "GC" en la base n.^o 2518-2519, puede obtenerse un ADN que codifica un mutante H840A.

Pueden ligarse un ADN que codifica una Cas mutante y un ADN que codifica una ADN glicosilasa para permitir la expresión como proteína de fusión, o diseñarse para expresarse separadamente usando un dominio de unión, inteína y similar, y formar un complejo en una célula hospedadora mediante interacción proteína-proteína y ligadura de proteína. Como alternativa, puede emplearse un diseño en que un ADN que codifica Cas mutante y un ADN que codifica ADN glicosilasa se dividen cada uno en dos fragmentos a una tasa de división adecuada, cualquiera de los fragmentos se ligan entre sí directamente o a través de un ligador adecuado para expresar un complejo enzimático modificador de ácido nucleico como dos complejos parciales, que se asocian y repliegan en la célula para reconstituir Cas mutante funcional que tiene una capacidad de reconocimiento de secuencia de ácido nucleico particular, y se reconstituye una ADN glicosilasa funcional que tiene una actividad catalizadora de reacción de escisión de base cuando se une la Cas mutante a la secuencia nucleotídica diana. Por ejemplo, se preparan respectivamente un ADN que codifica el fragmento de lado N-terminal y un ADN que codifica el fragmento de lado C-terminal de Cas mutante mediante el método de PCR usando cebadores adecuados; se preparan de la misma manera un ADN que codifica el fragmento de lado N-terminal y un ADN que codifica el fragmento de lado C-terminal de ADN glicosilasa; por ejemplo, se ligan los ADN que codifican los fragmentos de lado N-terminal entre sí y se ligan los ADN que codifican los fragmentos de lado C-terminal entre sí mediante un método convencional, con lo que puede producirse un ADN que codifica dos complejos parciales. Como alternativa, se ligan un ADN que codifica el fragmento de lado N-terminal de Cas mutante y un ADN que codifica el fragmento de lado C-terminal de ADN glicosilasa; y se ligan un ADN que codifica el fragmento de lado N-terminal de ADN glicosilasa y un ADN que codifica el fragmento de lado C-terminal de Cas mutante, con lo que puede producirse también un ADN que codifica dos complejos parciales. Los complejos parciales respectivos pueden ligarse para permitir la expresión como proteína de fusión, o diseñarse para expresarse separadamente usando un dominio de unión, inteína y similar, y formar un complejo en una célula hospedadora mediante interacción proteína-proteína y ligadura de proteína. Pueden ligarse dos complejos parciales para expresarse como una proteína de fusión. El sitio de división de Cas mutante no está particularmente limitado, a condición de que los dos fragmentos de división puedan reconstituirse de forma que reconozcan y se unan a la secuencia nucleotídica diana, y puedan dividirse en un sitio para proporcionar un fragmento de lado N-terminal y un fragmento de lado C-

terminal, o pueden ligarse apropiadamente no menos de 3 fragmentos obtenidos dividiendo en dos o más sitios dando dos fragmentos. Son conocidas las estructuras tridimensionales de diversas proteínas Cas, y los especialistas en la materia pueden seleccionar apropiadamente el sitio de división basándose en tal información. Por ejemplo, puesto que la región consistente en los aminoácidos 94 a 718 del extremo N de SpCas9 es un dominio (REC) implicado en el reconocimiento de la secuencia nucleotídica diana y el ARN de guía, y la región consistente en el aminoácido 1099 al aminoácido C-terminal es el dominio (PI) implicado en la interacción con PAM, el fragmento de lado N-terminal y el fragmento de lado C-terminal pueden dividirse en cualquier sitio del dominio REC o el dominio PI, preferiblemente en una región libre de estructura (p. ej., entre los aminoácidos 204 y 205 desde el extremo N (204..205), entre los aminoácidos 535 y 536 desde el extremo N (535..536) y similares) (véase, por ejemplo, Nat Biotechnol. 33(2): 139-142 (2015)).

El ADN obtenido que codifica una Cas mutante y/o una ADN glicosilasa puede insertarse en dirección 3' de un promotor de un vector de expresión similar al mencionado anteriormente, según el hospedador.

Por otro lado, puede obtenerse un ADN que codifica un ARN de guía y ARNtracr diseñando una secuencia de oligoADN que liga la secuencia de ARN de guía complementaria de la secuencia nucleotídica diana y una secuencia de ARNtracr conocida (p. ej. gttttagagctagaaatagcaaggttaaataaggctagtcggttatcaacttgaaaagtgggcaccgagtcgggtgg tgctttt; SEQ ID NO: 5) y sintetizando químicamente usando un sintetizador de ADN/ARN. Aunque puede insertarse también un ADN que codifica un ARN de guía y ARNtracr en un vector de expresión similar al mencionado anteriormente, según el hospedador. Como promotor, se usan preferiblemente el promotor sistémico pol III (p. ej., promotor SNR6, SNR52, SCR1, RPR1, U6, H1, etc.) y terminador (p. ej., secuencia T₆).

Puede prepararse un ARN que codifica Cas mutante y/o una ADN glicosilasa, por ejemplo, mediante transcripción en ARNm en un sistema de transcripción in vitro conocido por se usando un vector que codifica la Cas mutante anteriormente mencionada y/o ADN que codifica una ADN glicosilasa como molde.

El ARN de guía-ARNtracr puede obtenerse diseñando una secuencia de oligoADN que liga una secuencia complementaria de la secuencia nucleotídica diana y una secuencia de ARNtracr conocida y sintetizando químicamente usando un sintetizador de ADN/ARN.

Puede introducirse un ADN o ARN que codifica Cas mutante y/o una ADN glicosilasa, ARN de guía-ARNtracr o un ADN que codifica los mismos en una célula hospedadora mediante un método similar al anterior, según el hospedador.

Puesto que la nucleasa artificial convencional está acompañada de roturas de ADN bicatenario (DSB), aparecen inhibición del crecimiento y muerte celular supuestamente causadas por un corte desordenado del cromosoma (corte fuera de diana) al orientar a una secuencia del genoma. El efecto de esto es particularmente mortal para muchos microorganismos y procariontes, y evita su aplicabilidad. En la presente invención, la mutación se introduce no por rotura de ADN, sino por una reacción de escisión de base en la base de ADN y, por lo tanto, puede conseguirse una drástica reducción de la toxicidad.

La modificación del ADN bicatenario en la presente invención no evita la aparición de corte del ADN bicatenario en un sitio distinto del sitio diana (ajustado apropiadamente dentro de varios cientos de bases incluyendo la secuencia nucleotídica diana total o parcial). Sin embargo, una de las mayores ventajas de la presente invención es evitar la toxicidad por corte fuera de diana, lo que es generalmente aplicable a cualquier especie. En una realización preferible, por lo tanto, la modificación del ADN bicatenario de la presente invención no está acompañada de corte de la hebra de ADN no solo en un sitio diana de un ADN bicatenario dado, sino en un sitio distinto del mismo.

Como se muestra en los Ejemplos mencionados a continuación, cuando se producen módulos de reconocimiento de secuencia correspondientes a secuencias nucleotídicas diana múltiples adyacentes, y se usan simultáneamente, la introducción de mutación puede aumentar eficazmente más que usando una única secuencia nucleotídica como diana. Como efecto de esto, se consigue una inducción de mutación similar incluso cuando ambas secuencias nucleotídicas diana se solapan parcialmente o cuando ambas están separadas por aproximadamente 600 pb. Puede aparecer tanto cuando las secuencias nucleotídicas diana están en la misma dirección (las secuencias nucleotídicas diana están presentes en la misma hebra) como cuando están opuestas (la secuencia nucleotídica diana está presente en cada hebra de ADN bicatenario).

El método de modificación de secuencia genómica puede introducir mutación en casi todas las células en que se ha expresado el complejo enzimático modificador de ácido nucleico de la presente invención, al seleccionar una secuencia nucleotídica diana adecuada. Por tanto, la inserción y selección de un gen marcador de selección, que son esenciales en la edición genómica convencional, no son necesarias. Esto facilita y simplifica drásticamente la manipulación génica y amplía la aplicabilidad al mejoramiento vegetal y similares, puesto que no se produce un organismo recombinante con ADN extraño.

Puesto que el método de modificación de secuencia genómica muestra una eficacia extraordinariamente alta de inducción de mutación, y no requiere selección por marcadores, puede efectuarse la modificación de múltiples regiones de ADN en posiciones completamente diferentes como diana. Por lo tanto, en una realización preferible de la presente invención, pueden usarse dos o más clases de módulos de reconocimiento de secuencia de ácido nucleico que se unen específicamente a diferentes secuencias nucleotídicas diana (que pueden estar presentes en un gen objeto, o

dos o más genes objeto diferentes, pudiendo dichos genes objeto estar presentes en el mismo cromosoma o diferentes cromosomas). En este caso, cada uno de estos módulos de reconocimiento de secuencia de ácido nucleico y ADN glicosilasa forma un complejo enzimático modificador de ácido nucleico. Aquí, puede usarse una ADN glicosilasa común. Por ejemplo, cuando se usa el sistema CRISPR-Cas como módulo de reconocimiento de secuencia de ácido nucleico, se usa un complejo común de proteína Cas y ADN glicosilasa (incluyendo una proteína de fusión), y se producen dos o más clases de ARN quiméricos o ARNtracr y cada uno de dos o más ARN de guía que forman respectivamente una hebra complementaria con una secuencia nucleotídica diana diferente y se usan como ARN de guía-ARNtracr. Por otro lado, cuando se usan el motivo de dedo de cinc, efector TAL y similares como módulos de reconocimiento de secuencia de ácido nucleico, puede fusionarse una ADN glicosilasa con un módulo de reconocimiento de secuencia de ácido nucleico que se une específicamente a un nucleótido diana diferente.

Para expresar el complejo enzimático modificador de ácido nucleico de la presente invención en una célula hospedadora, como se menciona anteriormente, se introduce en una célula hospedadora un vector de expresión que contiene un ADN que codifica el complejo enzimático modificador de ácido nucleico, o un ARN que codifica el complejo enzimático modificador de ácido nucleico. Para una introducción de mutación eficaz, es deseable mantener la expresión de complejo enzimático modificador de ácido nucleico a un nivel dado o superior durante no menos de un periodo dado. A partir de tal aspecto, se asegura introducir un vector de expresión (plásmido, etc.) autónomamente replicable en una célula hospedadora. Sin embargo, puesto que el plásmido, etc. son ADN extraños, se retiran preferiblemente con rapidez después de la introducción exitosa de mutación. Cuando se usa endonucleasa AP mutante en combinación, puesto que la enzima mutante inhibe el mecanismo de REB en la célula hospedadora, puede inducir mutaciones espontáneas indeseables fuera de la región diana. Por tanto, es preferible retirar también un plásmido que contiene un ADN que codifica la enzima mutante inmediatamente después de la introducción de la mutación deseada. Por lo tanto, aunque sujeto a cambio dependiendo de la clase de célula hospedadora y similares, por ejemplo, el plásmido introducido se retira deseablemente de la célula hospedadora después de un periodo de 6 h-2 días desde la introducción de un vector de expresión usando diversos métodos de retirada de plásmido bien conocidos en la materia.

Como alternativa, a condición de obtener una expresión de un complejo enzimático modificador de ácido nucleico que sea suficiente para la introducción de mutación, es preferible introducir mutación en el ADN bicatenario objeto por expresión transitoria usando un vector de expresión sin replicabilidad autónoma en una célula hospedadora (p. ej., vector etc. que carece de origen de replicación que funciona en la célula hospedadora y/o gen que codifica una proteína necesaria para replicación) o ARN.

La presente invención se explica a continuación por referencia a ejemplos, que no han de considerarse limitantes.

Ejemplos

En los ejemplos mencionados a continuación, los experimentos se efectuaron como sigue:

Estirpe celular, cultivo, transformación e inducción de expresión

Se cultivó levadura de gemación *Saccharomyces cerevisiae* cepa BY4741 (que requiere leucina y uracilo) en un medio YPDA o medio SD estándar con una composición Dropout que satisface la auxotroficidad. El cultivo efectuado era cultivo en reposo en placa de agar o cultivo agitado en un medio líquido a entre 25 y 30 °C. Se efectuó la transformación mediante un método de ácido acético y litio y se realizó la selección en medio SD que satisface la auxotroficidad apropiada. Para inducción de expresión por galactosa, después de precultivo durante una noche en un medio SD apropiado, se realizaron para la inducción de expresión cultivo en medio SR con fuente de carbono cambiada de 2 % de glucosa a 2 % de rafinosa durante una noche y cultivo adicional en medio Sgal con fuente de carbono cambiada a 0,2 % de galactosa durante 3 h a aproximadamente 2 noches.

Para la medida del número de células supervivientes y la tasa de mutación de Can1, se diluyó apropiadamente una suspensión celular, se aplicó a medio de placa SD o medio de placa SD-Arg+60 mg/l de canavanina o medio de placa SD+300 mg/l de canavanina, y se contó el número de colonias que surgen 3 días después como el número de células supervivientes. Usando el número de colonias supervivientes en placa SD como el número total de células, y el número de colonias supervivientes en placa de canavanina como el número de cepas mutantes resistentes, se calculó y evaluó la tasa de mutación. Se identificó el sitio de introducción de mutación amplificando fragmentos de ADN que contienen la región génica diana de cada cepa mediante un método de PCR de colonia, efectuando la secuenciación de ADN y efectuando un análisis de alineamiento basado en la secuencia de la base de datos genómica de *Saccharomyces* (<http://www.yeastgenome.org/>).

Operación de ácido nucleico

Se procesó ADN o se construyó mediante cualquiera del método de PCR, tratamiento de enzima de restricción, ligadura, método de ensamblaje de Gibson y síntesis química artificial. Para plásmido, se usaron como vector lanzadera de *Escherichia coli* en levadura pRS315 para selección de leucina y pRS426 para selección de uracilo como esqueleto. Se amplificó el plásmido por la estirpe XL-10 Gold o DH5 α de *Escherichia coli* y se introdujo en levadura mediante el método de ácido acético y litio.

Producto de construcción

Para expresión inducible, se usó pGal1/10 (SEQ ID NO: 6) de levadura de gemación, que es un promotor bidireccional inducido por galactosa. En dirección 3' del promotor, se añadió una señal de localización nuclear (ccc aag aag aag agg aag gtg; SEQ ID NO: 7 (PKKKRV; codifica la SEQ ID NO: 8)) a ORF del gen Cas9 derivado de *Streptococcus pyogenes* que tiene un codón optimizado para expresión eucariótica (SEQ ID NO: 3) y la secuencia del ORF (el ORF del gen de tipo silvestre se muestra en la SEQ ID NO: 1, la mutación Y164A es la sustitución de las bases números 490-491 ta por gc (ta490gc); la mutación Y164G es la sustitución de las bases números 490-491 ta por gg (ta490gg); la mutación N222D es la sustitución de la base número 664 a por g (a664g); la mutación L304A es la sustitución de las bases números 910-911 tt por gc (tt910gc); la mutación R308E es la sustitución de las bases números 922-923 ag por ga (ag922ga); la mutación R308C es la sustitución de las bases números 922-924 aga por tgt (aga922tgt)) de genes de uracilo-ADN glicosilasa de tipo silvestre o diversos mutantes (UNG1 derivada de levadura *Saccharomyces cerevisiae*), excluyendo una región (bases números 1-60) que codifica la señal de localización mitocondrial, se ligó a través de una secuencia ligadora y se expresó como proteína de fusión. Para comparación, se ligó el gen UNG1 en lugar del gen de desaminasa (PmCDA1 derivado de *Petromyzon marinus*) y se expresó como proteína de fusión. Como secuencia ligadora, se usó el ligador 2xGS (dos repeticiones de ggt gga gga ggt tct; SEQ ID NO: 9 (GGGGS; que codifica la SEQ ID NO: 10)). Como terminador, se ligaron el terminador ADH1 derivado de levadura de gemación (SEQ ID NO: 11) y el terminador Top2 (SEQ ID NO: 12). En el método de integración de dominio, se ligó el ORF del gen Cas9 al dominio SH3 (SEQ ID NO: 13 y 14) a través del ligador 2xGS dando una proteína, UNG1 de levadura mutante con adición de la secuencia de ligando SH3 (SEQ ID NO: 15 y 16) como otra proteína y se ligaron con el promotor Gal1/10 en ambas direcciones y se expresaron simultáneamente. Se incorporaron estos en el plásmido pRS315.

En Cas9, se introdujeron la mutación para convertir el ácido aspártico 10 en alanina (D10A, correspondiente a la mutación de secuencia de ADN a29c) y la mutación para convertir la histidina 840 en alanina (H840A, correspondiente a la mutación de secuencia de ADN ca2518gc) para retirar la capacidad de corte de cada lado de la hebra de ADN.

Se dispuso ARNg como una estructura quimérica con ARNtracr (derivado de *Streptococcus pyogenes*; SEQ ID NO: 5) entre el promotor SNR52 (SEQ ID NO: 17) y el terminador Sup4 (SEQ ID NO: 18), y se incorporó al plásmido pRS426. Como secuencia de bases diana de ARNg, se usaron las secuencias de hebra complementaria 793-812 (aaccagggtgcctgggtcc; SEQ ID NO: 19) y 767-786 (ataacggaatccaactgggc; SEQ ID NO: 20) de ORF del gen CAN1. Para expresión simultánea de múltiples dianas, se incorporaron una secuencia desde un promotor a un terminador como un conjunto y una pluralidad de las mismas al mismo plásmido. Se introdujeron en células junto con el plásmido de expresión Cas9-UNG1, se expresaron intracelularmente y se formó un complejo de ARNg-ARNtracr y Cas9-UNG1.

Ejemplo 1: Modificación de la secuencia del genoma ligando la capacidad de reconocimiento de secuencia de ADN de CRISPR-Cas a uracilo-ADN glicosilasa mutante (1)

Para probar el efecto de la técnica de modificación de secuencia genómica utilizando uracilo-ADN glicosilasa mutante y la capacidad de reconocimiento de ácido nucleico de CRISPR-Cas, se probó la introducción de mutación en el gen CAN1 que codifica un transportador de canavanina que adquiere resistencia a canavanina debido a deficiencia génica. Como ARNg, se usaron una secuencia complementaria de 793-812 de ORF del gen CAN1 y una secuencia complementaria de 767-786 de la secuencia de hebra complementaria, se construyeron un vector de expresión de ARN quimérico obtenido ligando al mismo ARNtracr derivado de *Streptococcus pyogenes* y un vector que expresa una proteína obtenida fusionando dCas9 con actividad nucleasa alterada mediante la introducción de mutaciones (D10A y H840A) en Cas9 derivada de *Streptococcus pyogenes* (SpCas9) y UNG1 derivada de levadura de tipo silvestre o UNG1 derivada de levadura con diversas mutaciones introducidas (mutación simple N222D y mutación doble N222D y L304A, mutación R308E o R308C), se introdujeron en la levadura de gemación mediante el método de ácido acético-litio y se coexpresaron. Se muestran los resultados en la Fig. 3 Cuando se cultivan en una placa SD que contiene canavanina, solo las células sometidas a introducción y expresión de ARNg-ARNtracr y UNG1 mutante de dCas9 (mutante doble de la mutación N222D que confiere actividad CDG y mutación L304A, R308E o R308C que disminuye la reactividad con ADN que tiene una estructura de doble hélice no relajada) formaban colonias resistentes a canavanina. Con la mutación única N222D, la citotoxicidad era fuerte y el cultivo celular y evaluación eran difíciles, y por lo tanto los resultados no se muestran. A partir de lo anterior, se mostró que la introducción de mutación específica de diana se vuelve posible al disminuir la reactividad de ADN glicosilasa con un ADN que tiene una estructura de doble hélice no relajada.

Ejemplo 2: Modificación de secuencia genómica ligando la capacidad de reconocimiento de secuencia de ADN de CRISPR-Cas a uracilo-ADN glicosilasa mutante (2)

Usando UNG1 de levadura con mutación doble introducida de mutación R308C que disminuye la reactividad con un ADN que tiene estructura de doble hélice no relajada y mutación N222D que confiere actividad CDG, o mutación Y164A o Y164G que confiere actividad TDG, y mediante un método similar al del Ejemplo 1, se probó la introducción de mutación en el gen CAN1. Los resultados se muestran en la Fig. 4. Se mostró que R308C N222D puede conseguir una eficacia de inducción de mutación comparable a la de la desaminasa PmCDA1, y la cepa mutante puede obtenerse incluso sin selección. Se mostró que la base timina podía editarse, puesto que se obtenía colonia resistente a canavanina también en R308C Y164A. La mutación Y164G mejoraba la eficacia de inducción de mutación.

Entonces, se sometió cada clon resistente a canavanina a análisis de secuencia de la región del gen Can1. Se muestran los resultados en la Fig. 5. Las mutaciones estaban centradas algo aleatoriamente alrededor de dos sitios diana adyacentes (767-786, 793-812). Esto es diferente de la introducción localizada de mutación por desaminasa (documento WO 2015-133554), y sugiere que la técnica de edición genómica es adecuada para la introducción aleatoria de mutación en la secuencia nucleotídica diana y en los alrededores de la misma. Como se suponía, se encontraba principalmente la mutación puntual de C o G en N222D, y se encontraba principalmente la mutación puntual a T o A en Y164A y Y164G.

Ejemplo 3: Uso de diferentes esquemas de acoplamiento

Se examinó si la mutación puede introducirse en un gen diana incluso cuando no se usan Cas9 y ADN glicosilasa como proteína de fusión, sino cuando se forma un complejo enzimático modificador de ácido nucleico a través de un dominio de unión y un ligando del mismo. Como Cas9, se usó dCas9 del Ejemplo 1, se usó UNG1 de levadura mutante (mutante doble de mutación N222D o Y164A, y mutante R308E o R308C) como ADN glicosilasa, se fusionó el dominio SH3 con el primero y se fusionó un ligando de unión del mismo con el último produciendo diversos productos de construcciones mostrados en la Fig. 6. De la misma manera que en el Ejemplo 1, se usaron secuencias del gen CAN1 como dianas de ARNg, y se introdujeron estos productos de construcciones en una levadura de gemación. Como resultado, incluso cuando dCas9 y ADN glicosilasa estaban unidas a través del dominio de unión, se introducía eficazmente la mutación en el sitio diana del gen CAN1 (Fig. 6).

Ejemplo 4: Mejora de la eficacia de inducción de mutación por coexpresión de endonucleasa AP mutante

Usando UNG1 de levadura con doble mutación introducida de mutación R308C que disminuye la reactividad con un ADN que tiene una estructura de doble hélice no relajada y mutación N222D que confiere actividad CDG o mutación Y164G que confiere actividad TDG, y APE1 humana mutante (E960, D210N) que perdía la actividad enzimática pero retenía capacidad de unión al sitio AP, y mediante un método similar al del Ejemplo 1, se probó la introducción de mutación en el gen CAN1. Se muestran los resultados en la Fig. 7. Cuando se coexpresaba APE1 mutante, el número de colonias resistentes a canavanina aumentaba incluso en Y164G, R308C, que mostraba baja eficacia cuando se usaba sola, y la eficacia de introducción de mutación orientada a timina mejoraba notablemente.

Ejemplo 5: Reducción de la citotoxicidad por introducción de una mutación que disminuye la reactividad con ADN que tiene una estructura de doble hélice no relajada

Se examinó la influencia de la presencia o ausencia de mutación (L304A) que disminuye la reactividad con un ADN que tiene una estructura de hélice doble no relajada en UNG1 sobre la tasa de supervivencia de la levadura hospedadora. Se muestran los resultados en la Fig. 8. La levadura hospedadora con UNG1 introducida mutante que tiene solo la mutación que confiere actividad CDG (N222D) o actividad TDG (Y164A) mostraba una notable disminución de la tasa de supervivencia en comparación con la levadura con UNG2 introducida de tipo silvestre. Esto se supone que es debido a que la UNG1 de tipo silvestre retira el uracilo, que es una base aberrante que aparece raramente en el ADN, mientras que la UNG1 mutante que tiene actividad CDG o actividad TDG retira citosina o timina de cualquier lugar del ADN genómico y produce mutaciones indeseables para la supervivencia de la célula. Por otro lado, cuando la reactividad con un ADN que tiene una estructura de doble hélice no relajada disminuye al introducir la mutación L304, se recuperaba notablemente la tasa de supervivencia de la levadura hospedadora y podía evitarse la citotoxicidad.

Ejemplo 6. Utilización de uracilo-ADN glicosilasa heterogénea

Se examinó si la introducción de una mutación diana en la levadura hospedadora es posible incluso cuando se usa UNG1 mutante heterogénea en lugar de UNG1 mutante derivada de levadura. Se usaron dos clases de ung mutantes derivadas de Escherichia coli (EcUDG) (mutante doble N123D/L191A, mutante doble Y66G/L191A) y cuatro clases de UDG mutantes derivadas de virus Vaccinia (vvUDG) (doble mutante N120D/R187C, doble mutante Y70G/R187C, mutante N120D, mutante Y70G). Se muestran los resultados en la Fig. 9. Aunque tanto EcUDG como vvUDG eran funcionales en levadura, la eficacia de inducción de mutación era baja en comparación con UNG1 de levadura, y se mostró que el uso de ADN glicosilasa alogénica era ventajoso. Sorprendentemente, se aclaró que la citotoxicidad estaba ausente en vvUDG, independientemente de la presencia o ausencia de mutación R187C correspondiente a la mutación R308C de UNG1 de levadura. Como resultado del análisis de secuencia, puesto que la mutación por vvUDG se concentraba en una base específica en la secuencia nucleotídica diana independientemente de la presencia o ausencia de mutación R187C (véase la Fig. 10), se sugirió que vvUDG era una ADN glicosilasa con reactividad nativa suficientemente baja con un ADN que tiene una estructura de doble hélice no relajada. La eficacia de inducción de mutación por vvUDG aumentaba notablemente en ADN polimerasa de virus al coexpresar A20, que interacciona con vvUDG y actúa como factor de procesividad (Fig. 9).

Ejemplo 7. Reducción de la mutación no específica mediante la utilización de enzima dividida

Además de la utilización de una mutación que disminuye la reactividad con un ADN que tiene una estructura de doble hélice no relajada y una ADN glicosilasa con actividad nativa baja con un ADN que tiene una estructura de doble hélice tal como vvUDG, se probó la utilización de la técnica de enzima dividida como medio diferente para reducir la mutación no específica por ADN glicosilasa. Se introdujeron los plásmidos mostrados en la Fig. 11 que contienen ADN que

codifican diversas enzimas divididas en la levadura hospedadora junto con un plásmido que contiene un ADN que codifica ARN de guía-ARNtracr mediante un método similar al del Ejemplo 1, y se examinaron el número de células, el número de colonias resistentes a canavanina (mutación en el sitio diana) y el número de colonias resistentes a tialisina (mutación no específica) en un medio no selectivo. La tasa de supervivencia de la levadura hospedadora en un medio no selectivo era equivalente a cuando se introducía UNG1 mutante con mutación introducida (R308C) que disminuye la reactividad con un ADN que tiene una estructura de doble hélice no relajada, incluso cuando no se usaba enzima dividida. Por tanto, se mostró que la citotoxicidad puede disminuir suficientemente mediante la utilización de una enzima dividida, incluso sin introducción de una mutación que disminuya la reactividad con un ADN que tiene una estructura de doble hélice no relajada, con lo que se sugirió que puede suprimirse la mutación no específica (Fig. 11). De hecho, se disminuyó la frecuencia de mutación no específica por el uso de resistencia a tialisina como índice mediante la utilización de una enzima dividida (Fig. 11).

Aplicabilidad industrial

La presente invención hace posible introducir con seguridad una mutación específica de sitio en cualquier especie sin la inserción acompañante de un ADN extraño o roturas de ADN bicatenario. También es posible establecer un amplio intervalo de introducción de mutación en una secuencia nucleotídica diana y varios cientos de bases en sus alrededores, y la técnica puede aplicarse también a la inducción de evolución tópica mediante la introducción de mutación aleatoria en una región restringida particular, lo que ha sido casi imposible hasta ahora, y es extremadamente útil. Además, cuando se confiere a UNG la mutación que confiere actividad CDG y la mutación que confiere actividad TDG, se vuelve posible la escisión de base usando 5-metilcitosina como sustrato. Según la presente invención, por lo tanto, puede reescribirse la información epigenómica, por ejemplo, la liberación específica de región del estado de metilación para cambiar el patrón de expresión génica y similares. Por lo tanto, se vuelven posibles la diferenciación celular artificial, inhibición de células cancerosas y modificación de la función génica sin reescribir la secuencia genómica y similares.

Listado de secuencias

<110> National University Corporation Kobe University

<120> Método de modificación de una secuencia genómica que comprende introducir específicamente la mutación en la secuencia de ADN diana por reacción de escisión de base, y complejo molecular utilizado para ello

<130> 092377

<150> JP 2014-224745

<151> 2014-11-04

<160> 20

<170> PatentIn versión 3.5

<210> 1

<211> 1077

<212> ADN

<213> Saccharomyces cerevisiae

<220>

<221> CDS

<222> (1)..(1077)

<220>

<221> péptido_tránsito

<222> (1)..(60)

<400> 1

ES 2 800 168 T3

atg	tgg	tgc	atg	aga	aga	ttg	cca	aca	aat	tca	gtt	atg	acg	gtg	gct	48
Met	Trp	Cys	Met	Arg	Arg	Leu	Pro	Thr	Asn	Ser	Val	Met	Thr	Val	Ala	
1			5					10					15			
cga	aag	aga	aag	caa	act	acg	atc	gaa	gac	ttc	ttt	ggt	aca	aag	aaa	96
Arg	Lys	Arg	Lys	Gln	Thr	Thr	Ile	Glu	Asp	Phe	Phe	Gly	Thr	Lys	Lys	
			20					25					30			
agc	act	aat	gag	gcg	ccc	aat	aag	aaa	ggt	aaa	tcg	ggc	gct	act	ttc	144
Ser	Thr	Asn	Glu	Ala	Pro	Asn	Lys	Lys	Gly	Lys	Ser	Gly	Ala	Thr	Phe	
		35					40					45				
atg	acc	ata	aca	aat	ggc	gca	gct	atc	aag	aca	gag	aca	aag	gcg	gtc	192
Met	Thr	Ile	Thr	Asn	Gly	Ala	Ala	Ile	Lys	Thr	Glu	Thr	Lys	Ala	Val	
		50				55					60					
gca	aag	gag	gca	aac	acg	gac	aaa	tat	cct	gct	aat	tca	aat	gca	aaa	240
Ala	Lys	Glu	Ala	Asn	Thr	Asp	Lys	Tyr	Pro	Ala	Asn	Ser	Asn	Ala	Lys	
65				70					75					80		
gat	gta	tat	tcg	aag	aac	ttg	agc	agc	aat	tta	cgt	aca	cta	ctg	tcg	288
Asp	Val	Tyr	Ser	Lys	Asn	Leu	Ser	Ser	Asn	Leu	Arg	Thr	Leu	Leu	Ser	
				85					90					95		
ctg	gag	cta	gaa	acg	att	gat	gat	tcg	tgg	ttt	cca	cat	tta	atg	gat	336
Leu	Glu	Leu	Glu	Thr	Ile	Asp	Asp	Ser	Trp	Phe	Pro	His	Leu	Met	Asp	
			100					105					110			
gaa	ttt	aaa	aag	cca	tat	ttt	gta	aag	ttg	aaa	cag	ttt	gtc	act	aaa	384
Glu	Phe	Lys	Lys	Pro	Tyr	Phe	Val	Lys	Leu	Lys	Gln	Phe	Val	Thr	Lys	
		115					120					125				

ES 2 800 168 T3

gag caa gct gat cat aca gta ttt ccg cct gca aag gat att tac tca 432
 Glu Gln Ala Asp His Thr Val Phe Pro Pro Ala Lys Asp Ile Tyr Ser
 130 135 140

tgg acc agg cta acg cct ttt aat aaa gtt aaa gtc gtg att atc ggt 480
 Trp Thr Arg Leu Thr Pro Phe Asn Lys Val Lys Val Val Ile Ile Gly
 145 150 155 160

caa gat ccc tac cac aac ttt aat cag gcg cat ggc ttg gct ttt agc 528
 Gln Asp Pro Tyr His Asn Phe Asn Gln Ala His Gly Leu Ala Phe Ser
 165 170 175

gtc aaa ccc cct aca cca gca cca ccg tca tta aag aac ata tat aag 576
 Val Lys Pro Pro Thr Pro Ala Pro Pro Ser Leu Lys Asn Ile Tyr Lys
 180 185 190

gaa ctg aag caa gag tat cct gat ttt gtt gaa gat aat aaa gtg gga 624
 Glu Leu Lys Gln Glu Tyr Pro Asp Phe Val Glu Asp Asn Lys Val Gly
 195 200 205

gat tta act cac tgg gct tca caa ggg gtt tta ttg ctt aat acc tcg 672
 Asp Leu Thr His Trp Ala Ser Gln Gly Val Leu Leu Leu Asn Thr Ser
 210 215 220

tta act gta aga gca cat aat gca aac tcg cat tct aag cat ggt tgg 720
 Leu Thr Val Arg Ala His Asn Ala Asn Ser His Ser Lys His Gly Trp
 225 230 235 240

gaa act ttt aca aaa agg gta gtt cag ttg ttg atc cag gac aga gag 768
 Glu Thr Phe Thr Lys Arg Val Val Gln Leu Leu Ile Gln Asp Arg Glu
 245 250 255

gcc gac ggt aag agt tta gta ttc ctc tta tgg ggg aac aat gct atc 816
 Ala Asp Gly Lys Ser Leu Val Phe Leu Leu Trp Gly Asn Asn Ala Ile
 260 265 270

aaa tta gta gag tct ctg ttg gga tct act tcc gtt gga agc ggt agt 864
 Lys Leu Val Glu Ser Leu Leu Gly Ser Thr Ser Val Gly Ser Gly Ser
 275 280 285

aag tac cct aat atc atg gtg atg aag tca gtg cat ccg tct cca tta 912
 Lys Tyr Pro Asn Ile Met Val Met Lys Ser Val His Pro Ser Pro Leu
 290 295 300

agt gca agt aga gga ttt ttt ggt acc aac cat ttc aaa atg ata aac 960
 Ser Ala Ser Arg Gly Phe Phe Gly Thr Asn His Phe Lys Met Ile Asn
 305 310 315 320

gat tgg cta tac aat acc cgc gga gag aaa atg ata gac tgg agt gtt 1008
 Asp Trp Leu Tyr Asn Thr Arg Gly Glu Lys Met Ile Asp Trp Ser Val
 325 330 335

gtt cct gga acg tca tta aga gaa gtt cag gag gca aat gct cgc tta 1056
 Val Pro Gly Thr Ser Leu Arg Glu Val Gln Glu Ala Asn Ala Arg Leu
 340 345 350

gag tca gaa tca aag gac cct 1077
 Glu Ser Glu Ser Lys Asp Pro
 355

<210> 2

<211> 359

<212> PRT

5 <213> *Saccharomyces cerevisiae*

<400> 2

ES 2 800 168 T3

Met Trp Cys Met Arg Arg Leu Pro Thr Asn Ser Val Met Thr Val Ala
 1 5 10 15

Arg Lys Arg Lys Gln Thr Thr Ile Glu Asp Phe Phe Gly Thr Lys Lys
 20 25 30

Ser Thr Asn Glu Ala Pro Asn Lys Lys Gly Lys Ser Gly Ala Thr Phe
 35 40 45

Met Thr Ile Thr Asn Gly Ala Ala Ile Lys Thr Glu Thr Lys Ala Val
 50 55 60

Ala Lys Glu Ala Asn Thr Asp Lys Tyr Pro Ala Asn Ser Asn Ala Lys
 65 70 75 80

Asp Val Tyr Ser Lys Asn Leu Ser Ser Asn Leu Arg Thr Leu Leu Ser
 85 90 95

Leu Glu Leu Glu Thr Ile Asp Asp Ser Trp Phe Pro His Leu Met Asp
 100 105 110

Glu Phe Lys Lys Pro Tyr Phe Val Lys Leu Lys Gln Phe Val Thr Lys
 115 120 125

Glu Gln Ala Asp His Thr Val Phe Pro Pro Ala Lys Asp Ile Tyr Ser
 130 135 140

Trp Thr Arg Leu Thr Pro Phe Asn Lys Val Lys Val Val Ile Ile Gly
 145 150 155 160

Gln Asp Pro Tyr His Asn Phe Asn Gln Ala His Gly Leu Ala Phe Ser
 165 170 175

Val Lys Pro Pro Thr Pro Ala Pro Pro Ser Leu Lys Asn Ile Tyr Lys
 180 185 190

Glu Leu Lys Gln Glu Tyr Pro Asp Phe Val Glu Asp Asn Lys Val Gly
 195 200 205

Asp Leu Thr His Trp Ala Ser Gln Gly Val Leu Leu Leu Asn Thr Ser
 210 215 220

Leu Thr Val Arg Ala His Asn Ala Asn Ser His Ser Lys His Gly Trp

ES 2 800 168 T3

225 230 235 240
 Glu Thr Phe Thr Lys Arg Val Val Gln Leu Leu Ile Gln Asp Arg Glu
 245 250 255
 Ala Asp Gly Lys Ser Leu Val Phe Leu Leu Trp Gly Asn Asn Ala Ile
 260 265 270
 Lys Leu Val Glu Ser Leu Leu Gly Ser Thr Ser Val Gly Ser Gly Ser
 275 280 285
 Lys Tyr Pro Asn Ile Met Val Met Lys Ser Val His Pro Ser Pro Leu
 290 295 300
 Ser Ala Ser Arg Gly Phe Phe Gly Thr Asn His Phe Lys Met Ile Asn
 305 310 315 320
 Asp Trp Leu Tyr Asn Thr Arg Gly Glu Lys Met Ile Asp Trp Ser Val
 325 330 335
 Val Pro Gly Thr Ser Leu Arg Glu Val Gln Glu Ala Asn Ala Arg Leu
 340 345 350
 Glu Ser Glu Ser Lys Asp Pro
 355

<210> 3
 <211> 4116
 <212> ADN
 <213> Secuencia artificial

5

<220>
 <223> CDS de Cas9 derivada de Streptococcus pyogenes optimizada para expresión eucariótica.

<220>
 <221> CDS
 <222> (1)..(4116)

10

<400> 3
 atg gac aag aag tac tcc att ggg ctc gat atc ggc aca aac agc gtc 48
 Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
 1 5 10 15
 ggt tgg gcc gtc att acg gac gag tac aag gtg ccg agc aaa aaa ttc 96
 Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
 20 25 30
 aaa gtt ctg ggc aat acc gat cgc cac agc ata aag aag aac ctc att 144
 Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
 35 40 45
 ggc gcc ctc ctg ttc gac tcc ggg gag acg gcc gaa gcc acg cgg ctc 192

ES 2 800 168 T3

Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu	Thr	Ala	Glu	Ala	Thr	Arg	Leu		
50						55					60						
aaa	aga	aca	gca	cgg	cgc	aga	tat	acc	cgc	aga	aag	aat	cgg	atc	tgc		240
Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Arg	Lys	Asn	Arg	Ile	Cys		
65				70					75					80			
tac	ctg	cag	gag	atc	ttt	agt	aat	gag	atg	gct	aag	gtg	gat	gac	tct		288
Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys	Val	Asp	Asp	Ser		
				85					90					95			
ttc	ttc	cat	agg	ctg	gag	gag	tcc	ttt	ttg	gtg	gag	gag	gat	aaa	aag		336
Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	Glu	Asp	Lys	Lys		
			100					105					110				
cac	gag	cgc	cac	cca	atc	ttt	ggc	aat	atc	gtg	gac	gag	gtg	gcg	tac		384
His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp	Glu	Val	Ala	Tyr		
			115				120					125					
cat	gaa	aag	tac	cca	acc	ata	tat	cat	ctg	agg	aag	aag	ctt	gta	gac		432
His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys	Lys	Leu	Val	Asp		
			130				135				140						
agt	act	gat	aag	gct	gac	ttg	cgg	ttg	atc	tat	ctc	gcg	ctg	gcg	cat		480
Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu	Ala	Leu	Ala	His		
145				150						155				160			
atg	atc	aaa	ttt	cgg	gga	cac	ttc	ctc	atc	gag	ggg	gac	ctg	aac	cca		528
Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly	Asp	Leu	Asn	Pro		
				165				170						175			
gac	aac	agc	gat	gtc	gac	aaa	ctc	ttt	atc	caa	ctg	gtt	cag	act	tac		576
Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe	Ile	Gln	Leu	Val	Gln	Thr	Tyr		
				180				185					190				
aat	cag	ctt	ttc	gaa	gag	aac	cgg	atc	aac	gca	tcc	gga	ggt	gac	gcc		624
Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile	Asn	Ala	Ser	Gly	Val	Asp	Ala		
			195				200					205					
aaa	gca	atc	ctg	agc	gct	agg	ctg	tcc	aaa	tcc	cgg	cgg	ctc	gaa	aac		672
Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg	Arg	Leu	Glu	Asn		
			210			215					220						
ctc	atc	gca	cag	ctc	cct	ggg	gag	aag	aag	aac	ggc	ctg	ttt	ggt	aat		720
Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu	Lys	Lys	Asn	Gly	Leu	Phe	Gly	Asn		
225					230					235				240			
ctt	atc	gcc	ctg	tca	ctc	ggg	ctg	acc	ccc	aac	ttt	aaa	tct	aac	ttc		768
Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr	Pro	Asn	Phe	Lys	Ser	Asn	Phe		
				245					250					255			
gac	ctg	gcc	gaa	gat	gcc	aag	ctt	caa	ctg	agc	aaa	gac	acc	tac	gat		816
Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp	Thr	Tyr	Asp		
			260				265						270				
gat	gat	ctc	gac	aat	ctg	ctg	gcc	cag	atc	ggc	gac	cag	tac	gca	gac		864
Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln	Ile	Gly	Asp	Gln	Tyr	Ala	Asp		
			275				280					285					
ctt	ttt	ttg	gcg	gca	aag	aac	ctg	tca	gac	gcc	att	ctg	ctg	agt	gat		912
Leu	Phe	Leu	Ala	Ala	Lys	Asn	Leu	Ser	Asp	Ala	Ile	Leu	Leu	Ser	Asp		
			290			295				300							

ES 2 800 168 T3

att ctg cga gtg aac acg gag atc acc aaa gct ccg ctg agc gct agt Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser 305 310 315 320	960
atg atc aag cgc tat gat gag cac cac caa gac ttg act ttg ctg aag Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys 325 330 335	1008
gcc ctt gtc aga cag caa ctg cct gag aag tac aag gaa att ttc ttc Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe 340 345 350	1056
gat cag tct aaa aat ggc tac gcc gga tac att gac ggc gga gca agc Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser 355 360 365	1104
cag gag gaa ttt tac aaa ttt att aag ccc atc ttg gaa aaa atg gac Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp 370 375 380	1152
ggc acc gag gag ctg ctg gta aag ctt aac aga gaa gat ctg ttg cgc Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg 385 390 395 400	1200
aaa cag cgc act ttc gac aat gga agc atc ccc cac cag att cac ctg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu 405 410 415	1248
ggc gaa ctg cac gct atc ctc agg cgg caa gag gat ttc tac ccc ttt Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe 420 425 430	1296
ttg aaa gat aac agg gaa aag att gag aaa atc ctc aca ttt cgg ata Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile 435 440 445	1344
ccc tac tat gta ggc ccc ctc gcc cgg gga aat tcc aga ttc gcg tgg Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp 450 455 460	1392
atg act cgc aaa tca gaa gag acc atc act ccc tgg aac ttc gag gaa Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu 465 470 475 480	1440
gtc gtg gat aag ggg gcc tct gcc cag tcc ttc atc gaa agg atg act Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr 485 490 495	1488
aac ttt gat aaa aat ctg cct aac gaa aag gtg ctt cct aaa cac tct Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser 500 505 510	1536
ctg ctg tac gag tac ttc aca gtt tat aac gag ctc acc aag gtc aaa Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys 515 520 525	1584
tac gtc aca gaa ggg atg aga aag cca gca ttc ctg tct gga gag cag Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln 530 535 540	1632
aag aaa gct atc gtg gac ctc ctc ttc aag acg aac cgg aaa gtt acc Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr 545 550 555 560	1680

ES 2 800 168 T3

gtg aaa cag ctc aaa gaa gac tat ttc aaa aag att gaa tgt ttc gac Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp 565 570 575	1728
tct gtt gaa atc agc gga gtg gag gat cgc ttc aac gca tcc ctg gga Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly 580 585 590	1776
acg tat cac gat ctc ctg aaa atc att aaa gac aag gac ttc ctg gac Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp 595 600 605	1824
aat gag gag aac gag gac att ctt gag gac att gtc ctc acc ctt acg Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr 610 615 620	1872
ttg ttt gaa gat agg gag atg att gaa gaa cgc ttg aaa act tac gct Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala 625 630 635 640	1920
cat ctc ttc gac gac aaa gtc atg aaa cag ctc aag agg cgc cga tat His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr 645 650 655	1968
aca gga tgg ggg cgg ctg tca aga aaa ctg atc aat ggg atc cga gac Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp 660 665 670	2016
aag cag agt gga aag aca atc ctg gat ttt ctt aag tcc gat gga ttt Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe 675 680 685	2064
gcc aac cgg aac ttc atg cag ttg atc cat gat gac tct ctc acc ttt Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe 690 695 700	2112
aag gag gac atc cag aaa gca caa gtt tct ggc cag ggg gac agt ctt Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu 705 710 715 720	2160
cac gag cac atc gct aat ctt gca ggt agc cca gct atc aaa aag gga His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly 725 730 735	2208
ata ctg cag acc gtt aag gtc gtg gat gaa ctc gtc aaa gta atg gga Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly 740 745 750	2256
agg cat aag ccc gag aat atc gtt atc gag atg gcc cga gag aac caa Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln 755 760 765	2304
act acc cag aag gga cag aag aac agt agg gaa agg atg aag agg att Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile 770 775 780	2352
gaa gag ggt ata aaa gaa ctg ggg tcc caa atc ctt aag gaa cac cca Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro 785 790 795 800	2400
gtt gaa aac acc cag ctt cag aat gag aag ctc tac ctg tac tac ctg Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu	2448

ES 2 800 168 T3

805	810	815	
cag aac ggc agg gac atg tac gtg gat	cag gaa ctg gac atc aat cgg		2496
Gln Asn Gly Arg Asp Met Tyr Val Asp	Gln Glu Leu Asp Ile Asn Arg		
820	825	830	
ctc tcc gac tac gac gtg gat cat atc gtg ccc cag tct ttt ctc aaa			2544
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys			
835	840	845	
gat gat tct att gat aat aaa gtg ttg aca aga tcc gat aaa aat aga			2592
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg			
850	855	860	
ggg aag agt gat aac gtc ccc tca gaa gaa gtt gtc aag aaa atg aaa			2640
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys			
865	870	875	880
aat tat tgg cgg cag ctg ctg aac gcc aaa ctg atc aca caa cgg aag			2688
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys			
885	890	895	
ttc gat aat ctg act aag gct gaa cga ggt ggc ctg tct gag ttg gat			2736
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp			
900	905	910	
aaa gcc ggc ttc atc aaa agg cag ctt gtt gag aca cgc cag atc acc			2784
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr			
915	920	925	
aag cac gtg gcc caa att ctc gat tca cgc atg aac acc aag tac gat			2832
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp			
930	935	940	
gaa aat gac aaa ctg att cga gag gtg aaa gtt att act ctg aag tct			2880
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser			
945	950	955	960
aag ctg gtc tca gat ttc aga aag gac ttt cag ttt tat aag gtg aga			2928
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg			
965	970	975	
gag atc aac aat tac cac cat gcg cat gat gcc tac ctg aat gca gtg			2976
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val			
980	985	990	
gta ggc act gca ctt atc aaa aaa tat ccc aag ctt gaa tct gaa ttt			3024
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe			
995	1000	1005	
gtt tac gga gac tat aaa gtg tac gat gtt agg aaa atg atc gca			3069
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala			
1010	1015	1020	
aag tct gag cag gaa ata ggc aag gcc acc gct aag tac ttc ttt			3114
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe			
1025	1030	1035	
tac agc aat att atg aat ttt ttc aag acc gag att aca ctg gcc			3159
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala			
1040	1045	1050	
aat gga gag att cgg aag cga cca ctt atc gaa aca aac gga gaa			3204

ES 2 800 168 T3

Asn	Gly	Glu	Ile	Arg	Lys	Arg	Pro	Leu	Ile	Glu	Thr	Asn	Gly	Glu			
1055						1060					1065						
aca	gga	gaa	atc	gtg	tgg	gac	aag	ggt	agg	gat	ttc	gcg	aca	gtc			3249
Thr	Gly	Glu	Ile	Val	Trp	Asp	Lys	Gly	Arg	Asp	Phe	Ala	Thr	Val			
1070						1075					1080						
cgg	aag	gtc	ctg	tcc	atg	ccg	cag	gtg	aac	atc	gtt	aaa	aag	acc			3294
Arg	Lys	Val	Leu	Ser	Met	Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Thr			
1085						1090					1095						
gaa	gta	cag	acc	gga	ggc	ttc	tcc	aag	gaa	agt	atc	ctc	ccg	aaa			3339
Glu	Val	Gln	Thr	Gly	Gly	Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	Lys			
1100						1105					1110						
agg	aac	agc	gac	aag	ctg	atc	gca	cgc	aaa	aaa	gat	tgg	gac	ccc			3384
Arg	Asn	Ser	Asp	Lys	Leu	Ile	Ala	Arg	Lys	Lys	Asp	Trp	Asp	Pro			
1115						1120					1125						
aag	aaa	tac	ggc	gga	ttc	gat	tct	cct	aca	gtc	gct	tac	agt	gta			3429
Lys	Lys	Tyr	Gly	Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala	Tyr	Ser	Val			
1130						1135					1140						
ctg	gtt	gtg	gcc	aaa	gtg	gag	aaa	ggg	aag	tct	aaa	aaa	ctc	aaa			3474
Leu	Val	Val	Ala	Lys	Val	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	Lys			
1145						1150					1155						
agc	gtc	aag	gaa	ctg	ctg	ggc	atc	aca	atc	atg	gag	cga	tca	agc			3519
Ser	Val	Lys	Glu	Leu	Leu	Gly	Ile	Thr	Ile	Met	Glu	Arg	Ser	Ser			
1160						1165					1170						
ttc	gaa	aaa	aac	ccc	atc	gac	ttt	ctc	gag	gcg	aaa	gga	tat	aaa			3564
Phe	Glu	Lys	Asn	Pro	Ile	Asp	Phe	Leu	Glu	Ala	Lys	Gly	Tyr	Lys			
1175						1180					1185						
gag	gtc	aaa	aaa	gac	ctc	atc	att	aag	ctt	ccc	aag	tac	tct	ctc			3609
Glu	Val	Lys	Lys	Asp	Leu	Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser	Leu			
1190						1195					1200						
ttt	gag	ctt	gaa	aac	ggc	cgg	aaa	cga	atg	ctc	gct	agt	gcg	ggc			3654
Phe	Glu	Leu	Glu	Asn	Gly	Arg	Lys	Arg	Met	Leu	Ala	Ser	Ala	Gly			
1205						1210					1215						
gag	ctg	cag	aaa	ggt	aac	gag	ctg	gca	ctg	ccc	tct	aaa	tac	gtt			3699
Glu	Leu	Gln	Lys	Gly	Asn	Glu	Leu	Ala	Leu	Pro	Ser	Lys	Tyr	Val			
1220						1225					1230						
aat	ttc	ttg	tat	ctg	gcc	agc	cac	tat	gaa	aag	ctc	aaa	ggg	tct			3744
Asn	Phe	Leu	Tyr	Leu	Ala	Ser	His	Tyr	Glu	Lys	Leu	Lys	Gly	Ser			
1235						1240					1245						
ccc	gaa	gat	aat	gag	cag	aag	cag	ctg	ttc	gtg	gaa	caa	cac	aaa			3789
Pro	Glu	Asp	Asn	Glu	Gln	Lys	Gln	Leu	Phe	Val	Glu	Gln	His	Lys			
1250						1255					1260						
cac	tac	ctt	gat	gag	atc	atc	gag	caa	ata	agc	gaa	ttc	tcc	aaa			3834
His	Tyr	Leu	Asp	Glu	Ile	Ile	Glu	Gln	Ile	Ser	Glu	Phe	Ser	Lys			
1265						1270					1275						
aga	gtg	atc	ctc	gcc	gac	gct	aac	ctc	gat	aag	gtg	ctt	tct	gct			3879
Arg	Val	Ile	Leu	Ala	Asp	Ala	Asn	Leu	Asp	Lys	Val	Leu	Ser	Ala			
1280						1285					1290						

ES 2 800 168 T3

tac aat aag cac agg gat aag ccc atc agg gag cag gca gaa aac 3924
 Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
 1295 1300 1305

att atc cac ttg ttt act ctg acc aac ttg ggc gcg cct gca gcc 3969
 Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
 1310 1315 1320

ttc aag tac ttc gac acc acc ata gac aga aag cgg tac acc tct 4014
 Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
 1325 1330 1335

aca aag gag gtc ctg gac gcc aca ctg att cat cag tca att acg 4059
 Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
 1340 1345 1350

ggg ctc tat gaa aca aga atc gac ctc tct cag ctc ggt gga gac 4104
 Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
 1355 1360 1365

agc agg gct gac 4116
 Ser Arg Ala Asp
 1370

<210> 4
 <211> 1372
 <212> PRT
 <213> Secuencia artificial

5

<220>
 <223> Construcción sintética

<400> 4
 Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
 1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
 20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
 35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
 50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
 65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
 85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
 100 105 110

10

ES 2 800 168 T3

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
 115 120 125
 His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
 130 135 140
 Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
 145 150 155 160
 Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
 165 170 175
 Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
 180 185 190
 Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
 195 200 205
 Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
 210 215 220
 Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
 225 230 235 240
 Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 245 250 255
 Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
 260 265 270
 Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
 275 280 285
 Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
 290 295 300
 Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
 305 310 315 320
 Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 325 330 335
 Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
 340 345 350
 Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
 355 360 365

ES 2 800 168 T3

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620

ES 2 800 168 T3

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
 625 630 635 640
 His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
 645 650 655
 Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
 660 665 670
 Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
 675 680 685
 Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
 690 695 700
 Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
 705 710 715 720
 His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
 725 730 735
 Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
 740 745 750
 Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
 755 760 765
 Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
 770 775 780
 Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
 785 790 795 800
 Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
 805 810 815
 Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
 820 825 830
 Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
 835 840 845
 Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
 850 855 860
 Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

ES 2 800 168 T3

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
 885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
 900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
 915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
 965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
 980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
 995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
 1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
 1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
 1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
 1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
 1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
 1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
 1100 1105 1110

ES 2 800 168 T3

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
 1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
 1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
 1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
 1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
 1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
 1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
 1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
 1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
 1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
 1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
 1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
 1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
 1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
 1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
 1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
 1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
 1355 1360 1365

Ser Arg Ala Asp
 1370

<210> 5
 <211> 83
 <212> ADN
 <213> Streptococcus pyogenes

ES 2 800 168 T3

<400> 5
 gtttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt 60
 ggcaccgagt cgggtggtgct ttt 83

<210> 6
 <211> 665
 <212> ADN
 <213> Saccharomyces cerevisiae

5
 <400> 6
 tttcaaaaat tcttactttt tttttggatg gacgcaaaga agtttaataa tcatattaca 60
 tggcattacc accatataca tatccatata catatccata tctaacttta cttatatggt 120
 gtggaaatgt aaagagcccc attatcttag cctaaaaaaa ccttctcttt ggaactttca 180
 gtaatacgct taactgctca ttgctatatt gaagtacgga ttagaagccg ccgagcgggt 240
 gacagccctc cgaaggaaga ctctcctcgg tgcgtcctcg tcttcaccgg tcgcttctct 300
 gaaacgcaga tgtgcctcgc gccgcactgc tccgaacaat aaagattcta caatactagc 360
 ttttatgggt atgaagagga aaaattggca gtaacctggc cccacaaacc ttcaaatgaa 420
 cgaatcaaat taacaacccat aggatgataa tgcgattagt tttttagcct tatttctggg 480
 gtaattaatc agcgaagcga tgatthttga tctattaaca gatatataaa tgcaaaaact 540
 gcataaccac ttaactaat actttcaaca ttttcggttt gtattacttc ttattcaaat 600
 gtaataaaag tatcaacaaa aaattgttaa tatacctcta tactttaacg tcaaggagaa 660
 aaaac 665

10
 <210> 7
 <211> 21
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Señal de transición nuclear.

15
 <220>
 <221> CDS
 <222> (1)..(21)

<400> 7
 ccc aag aag aag agg aag gtg 21
 Pro Lys Lys Lys Arg Lys Val
 1 5

20
 <210> 8
 <211> 7
 <212> PRT
 <213> Secuencia artificial

25
 <220>
 <223> Construcción sintética

<400> 8
 Pro Lys Lys Lys Arg Lys Val
 1 5

30
 <210> 9
 <211> 15
 <212> ADN
 <213> Secuencia artificial

<220>

ES 2 800 168 T3

<223> conector GS

<220>
 <221> CDS
 <222> (1)..(15)

5 <400> 9
 ggt gga gga ggt tct 15
 Gly Gly Gly Gly Ser
 1 5

<210> 10
 <211> 5
 <212> PRT
 10 <213> Secuencia artificial

<220>
 <223> Construcción sintética

<400> 10
 Gly Gly Gly Gly Ser
 1 5

15 <210> 11
 <211> 188
 <212> ADN
 <213> Saccharomyces cerevisiae

<400> 11
 gcgaatttct tatgatttat gatTTTTatt attaaataag ttataaaaa aataagtgtg 60

20 taaaaatttt aaagtgactc ttaggtttta aaacgaaaat tcttattctt gagtaactct 120
 ttctgttagg tcaggttgct ttctcaggta tagcatgagg tcgctcttat tgaccacacc 180
 tctaccgg 188

<210> 12
 <211> 417
 <212> ADN
 25 <213> Saccharomyces cerevisiae

<400> 12
 ataccaggca tggagcttat ctgggtccgtt cgagtTTTTcg acgagtttgg agacattctt 60
 tatagatgtc cTTTTTTTT aatgatattc gttaaagaac aaaaagtcaa agcagtttaa 120
 cctaacacct gttgttgatg ctacttgaaa caaggcttct aggcgaatac ttaaaaaggT 180
 aatttcaata gcggtttata tatctgtttg cttttcaaga tattatgtaa acgcacgatg 240
 tttttcgccc aggctttatt tttttgttg ttgttgtctt ctogaagaat tttctcgggc 300
 agatctttgt cggaatgtaa aaaagcgcgt aattaaactt tctattatgc tgactaaaat 360
 ggaagtgatc accaaaggct atttctgatt atataatcta gtcattactc gctcgag 417

<210> 13
 <211> 171
 <212> ADN
 30 <213> Secuencia artificial

<220>
 <223> dominio SH3

35 <220>
 <221> CDS
 <222> (1)..(171)

<400> 13

ES 2 800 168 T3

gca gag tat gtg cgg gcc ctc ttt gac ttt aat ggg aat gat gaa gaa 48
 Ala Glu Tyr Val Arg Ala Leu Phe Asp Phe Asn Gly Asn Asp Glu Glu
 1 5 10 15

gat ctt ccc ttt aag aaa gga gac atc ctg aga atc cgg gat aag cct 96
 Asp Leu Pro Phe Lys Lys Gly Asp Ile Leu Arg Ile Arg Asp Lys Pro
 20 25 30

gaa gag cag tgg tgg aat gca gag gac agc gaa gga aag agg ggg atg 144
 Glu Glu Gln Trp Trp Asn Ala Glu Asp Ser Glu Gly Lys Arg Gly Met
 35 40 45

att cct gtc cct tac gtg gag aag tat 171
 Ile Pro Val Pro Tyr Val Glu Lys Tyr
 50 55

<210> 14
 <211> 57
 <212> PRT
 5 <213> Secuencia artificial

<220>
 <223> Construcción sintética

<400> 14
 Ala Glu Tyr Val Arg Ala Leu Phe Asp Phe Asn Gly Asn Asp Glu Glu
 1 5 10 15

Asp Leu Pro Phe Lys Lys Gly Asp Ile Leu Arg Ile Arg Asp Lys Pro
 20 25 30

Glu Glu Gln Trp Trp Asn Ala Glu Asp Ser Glu Gly Lys Arg Gly Met
 35 40 45

Ile Pro Val Pro Tyr Val Glu Lys Tyr
 50 55

10 <210> 15
 <211> 33
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> Ligando de unión a SH3

<220>
 <221> CDS
 <222> (1)..(33)

<400> 15 33
 cct cca cct gct ctg cca cct aag aga agg aga
 Pro Pro Pro Ala Leu Pro Pro Lys Arg Arg Arg
 1 5 10

<210> 16
 <211> 11
 <212> PRT
 <213> Secuencia artificial

25 <220>
 <223> Construcción sintética

<400> 16
 Pro Pro Pro Ala Leu Pro Pro Lys Arg Arg Arg
 1 5 10

30 <210> 17
 <211> 269

ES 2 800 168 T3

<212> ADN
 <213> *Saccharomyces cerevisiae*

<400> 17
 tctttgaaaa gataatgtat gattatgctt tcactcatat ttatacagaa acttgatggt 60
 5 ttctttcgag tatatacaag gtgattacat gtacgtttga agtacaactc tagattttgt 120
 agtgccctct tgggctagcg gtaaagggtgc gcattttttc acaccctaca atgttctggt 180
 caaaagattt tggtaaagc ctgtagaagt gaaagttggt gcgcatgttt cggcgttcga 240
 aacttctccg cagtgaaaga taaatgatc 269

<210> 18
 <211> 14
 <212> ADN
 10 <213> *Saccharomyces cerevisiae*

<400> 18
 tgtttttat gtct 14

<210> 19
 <211> 20
 15 <212> ADN
 <213> *Saccharomyces cerevisiae*

<400> 19
 aaccaggtg cctgggtcc 20

<210> 20
 <211> 20
 20 <212> ADN
 <213> *Saccharomyces cerevisiae*

<400> 20
 ataacggaat ccaactgggc 20
 25

REIVINDICACIONES

- 5 1. Un método de modificación de un sitio diana de un ADN bicatenario en una célula, que comprende la etapa de poner en contacto un complejo en el que se unen un módulo de reconocimiento de secuencia de ácido nucleico que se une específicamente a una secuencia nucleotídica diana en un ADN bicatenario dado y ADN glicosilasa, con dicho ADN bicatenario, para convertir uno o más nucleótidos en el sitio diana en otro o más nucleótidos o eliminar uno o más nucleótidos, o insertar uno o más nucleótidos en dicho sitio diana, sin cortar al menos una hebra de dicho ADN bicatenario en el sitio diana, en el que la ADN glicosilasa es
- 10 (i) un mutante de UNG1 de levadura que tiene las mutaciones N222D/L304A, las mutaciones N222D/R308E, las mutaciones N222D/R308C, las mutaciones Y164A/L304A, las mutaciones Y164A/R308E, las mutaciones Y164A/R308C, las mutaciones Y164G/L304A, las mutaciones Y164G/R308E, las mutaciones Y164G/R308C, las mutaciones N222D/Y164A/L304A, las mutaciones N222D/Y164A/R308E, las mutaciones N222D/Y164A/R308C, las mutaciones N222D/Y164G/L304A, las mutaciones N222D/Y164G/R308E o las mutaciones N222D/Y164G/R308C, o
- 15 (ii) un mutante de UNG distinta de UNG1 de levadura, que tiene mutaciones correspondientes a las mutaciones de cada uno de los mutantes de (i) y
- en el que el método no es un método de tratamiento del cuerpo humano o el cuerpo animal por terapia y no es un proceso para modificar la identidad genética de línea germinal de seres humanos.
- 20 2. El método según la reivindicación 1, en el que el módulo de reconocimiento de secuencia de ácido nucleico se selecciona del grupo consistente en un sistema CRISPR-Cas en el que se inactiva al menos una capacidad de corte de ADN de Cas, un motivo de dedo de cinc, un efector TAL y un motivo PPR, y es preferiblemente el sistema CRISPR-Cas.
3. El método según la reivindicación 1 o 2, en el que se pone en contacto adicionalmente el ADN bicatenario con un factor que cambia la estructura bicatenaria del ADN.
- 25 4. El método según una cualquiera de las reivindicaciones 1 a 3, que usa dos o más clases de módulos de reconocimiento de secuencia de ácido nucleico que se unen respectivamente de forma específica a diferentes secuencias nucleotídicas diana, preferiblemente en el que las diferentes secuencias nucleotídicas diana están presentes en diferentes genes.
5. El método según una cualquiera de las reivindicaciones 1 a 4, que comprende adicionalmente poner en contacto el ADN bicatenario con una endonucleasa AP que tiene capacidad de unión a un sitio abásico pero que carece de actividad nucleasa.
- 30 6. Un método de modificación de un sitio diana de un ADN bicatenario en una célula, que comprende la etapa de poner en contacto un complejo en el que están unidos un módulo de reconocimiento de secuencia de ácido nucleico que se une específicamente a una secuencia nucleotídica diana en un ADN bicatenario dado y ADN glicosilasa, con dicho ADN bicatenario, para convertir uno o más nucleótidos en el sitio diana en otro o más nucleótidos o eliminar uno o más nucleótidos, o insertar uno o más nucleótidos en dicho sitio diana, sin cortar al menos una hebra de dicho ADN bicatenario en el sitio diana, en el que la ADN glicosilasa es un mutante de uracilo-ADN glicosilasa (UDG) derivada de un virus perteneciente a los Poxviridae y que tiene la mutación N120D, la mutación Y70G, la mutación Y70A, las mutaciones N120D/Y70G o las mutaciones N120D/Y70A; y en el que el método no es un método de tratamiento del cuerpo humano o el cuerpo animal por terapia y no es un proceso para modificar la identidad genética de línea germinal de seres humanos.
- 35 40 7. El método según la reivindicación 6, en el que el ADN bicatenario se pone en contacto adicionalmente con proteína A20.
8. Un método de modificación de un sitio diana de un ADN bicatenario en una célula, que comprende la etapa de poner en contacto un complejo en el que se une un módulo de reconocimiento de secuencia de ácido nucleico que se une específicamente a una secuencia nucleotídica diana en un ADN bicatenario dado y ADN glicosilasa, con dicho ADN bicatenario, para convertir uno o más nucleótidos en el sitio diana en otro o más nucleótidos o eliminar uno o más nucleótidos, o insertar uno o más nucleótidos en dicho sitio diana, sin cortar al menos una hebra de dicho ADN bicatenario en el sitio diana, en el que la ADN glicosilasa es
- 45 (i) un mutante de UNG1 de levadura que tiene la mutación N222D, la mutación Y164G, las mutaciones N222D/L304A, las mutaciones N222D/R308E, las mutaciones N222D/R308C, las mutaciones Y164A/L304A, las mutaciones Y164A/R308E, las mutaciones Y164A/R308C, las mutaciones Y164G/L304A, las mutaciones Y164G/R308E, las mutaciones Y164G/R308C, las mutaciones N222D/Y164A/L304A, las mutaciones N222D/Y164A/R308E, las mutaciones N222D/Y164A/R308C, las mutaciones N222D/Y164G/L304A, las mutaciones N222D/Y164G/R308E o las mutaciones N222D/Y164G/R308C, o
- 50 (ii) un mutante de UNG1 distinta de UNG1 de levadura, que tiene una mutación o mutaciones correspondientes a las mutaciones de cada uno de los mutantes de (1),
- 55

- 5 en el que la ADN glicosilasa, y un elemento del módulo de reconocimiento de secuencia de ácido nucleico que se une directamente a la ADN glicosilasa, se dividen respectivamente en dos fragmentos, estando los fragmentos de cualquiera de la ADN glicosilasa y el elemento ligados respectivamente a los fragmentos del otro proporcionando dos complejos parciales, y cuando los complejos parciales se repliegan entre sí, el módulo de reconocimiento de secuencia de ácido nucleico es capaz de unirse específicamente a la secuencia nucleotídica diana y el enlace específico posibilita a la ADN glicosilasa exhibir actividad enzimática, preferiblemente en el que el elemento del módulo de reconocimiento de secuencia de ácido nucleico que está directamente unido a la ADN glicosilasa es una proteína Cas en la que se inactivan al menos una de las capacidades de corte de ADN; y
- 10 en el que el método no es un método de tratamiento del cuerpo humano o el cuerpo animal por terapia y no es un proceso para modificar la identidad genética de línea germinal de seres humanos.
9. El método según la reivindicación 8, en el que los dos complejos parciales se proporcionan como complejos moleculares separados, y se repliegan por asociación de los mismos en la célula.
- 15 10. El método según una cualquiera de las reivindicaciones 1 a 9, en el que se pone en contacto el ADN bicatenario con el complejo mediante la introducción de un ácido nucleico que codifica el complejo en una célula que tiene el ADN bicatenario.
11. El método según la reivindicación 10, en el que la célula se selecciona del grupo consistente en una célula procariótica, una célula eucariótica, una célula microbiana, una célula vegetal, una célula de insecto, una célula animal, una célula de vertebrado y una célula de mamífero.
- 20 12. Un complejo enzimático modificador de ácido nucleico en el que se unen un módulo de reconocimiento de secuencia de ácido nucleico que se une específicamente a una secuencia nucleotídica diana en un ADN bicatenario dado y ADN glicosilasa, que convierte uno o más nucleótidos en el sitio diana en otro o más nucleótidos, o elimina uno o más nucleótidos, o inserta uno o más nucleótidos en dicho sitio diana, sin cortar al menos una hebra de dicho ADN bicatenario en el sitio diana, en el que la ADN glicosilasa es
- 25 (i) un mutante de UNG1 de levadura que tiene las mutaciones N222D/L304A, las mutaciones N222D/R308E, las mutaciones N222D/R308C, las mutaciones Y164A/L304A, las mutaciones Y164A/R308E, las mutaciones Y164A/R308C, las mutaciones Y164G/L304A, las mutaciones Y164G/R308E, las mutaciones Y164G/R308C, las mutaciones N222D/Y164A/L304A, las mutaciones N222D/Y164A/R308E, las mutaciones N222D/Y164A/R308C, las mutaciones N222D/Y164G/L304A, las mutaciones N222D/Y164G/R308E o las mutaciones N222D/Y164G/R308C, o
- 30 (ii) un mutante de UNG distinta de UNG1 de levadura, que tiene mutaciones correspondientes a las mutaciones de cada uno de los mutantes de (i) o
- (iii) un mutante de UDG derivada de un virus perteneciente a los Poxviridae que tiene la mutación N120D, la mutación Y70G, la mutación Y70OA, las mutaciones N120D/Y70G o las mutaciones N120D/Y70A.
- 35 13. Un ácido nucleico que codifica el complejo enzimático modificador de ácido nucleico según la reivindicación 12.

FIG. 1

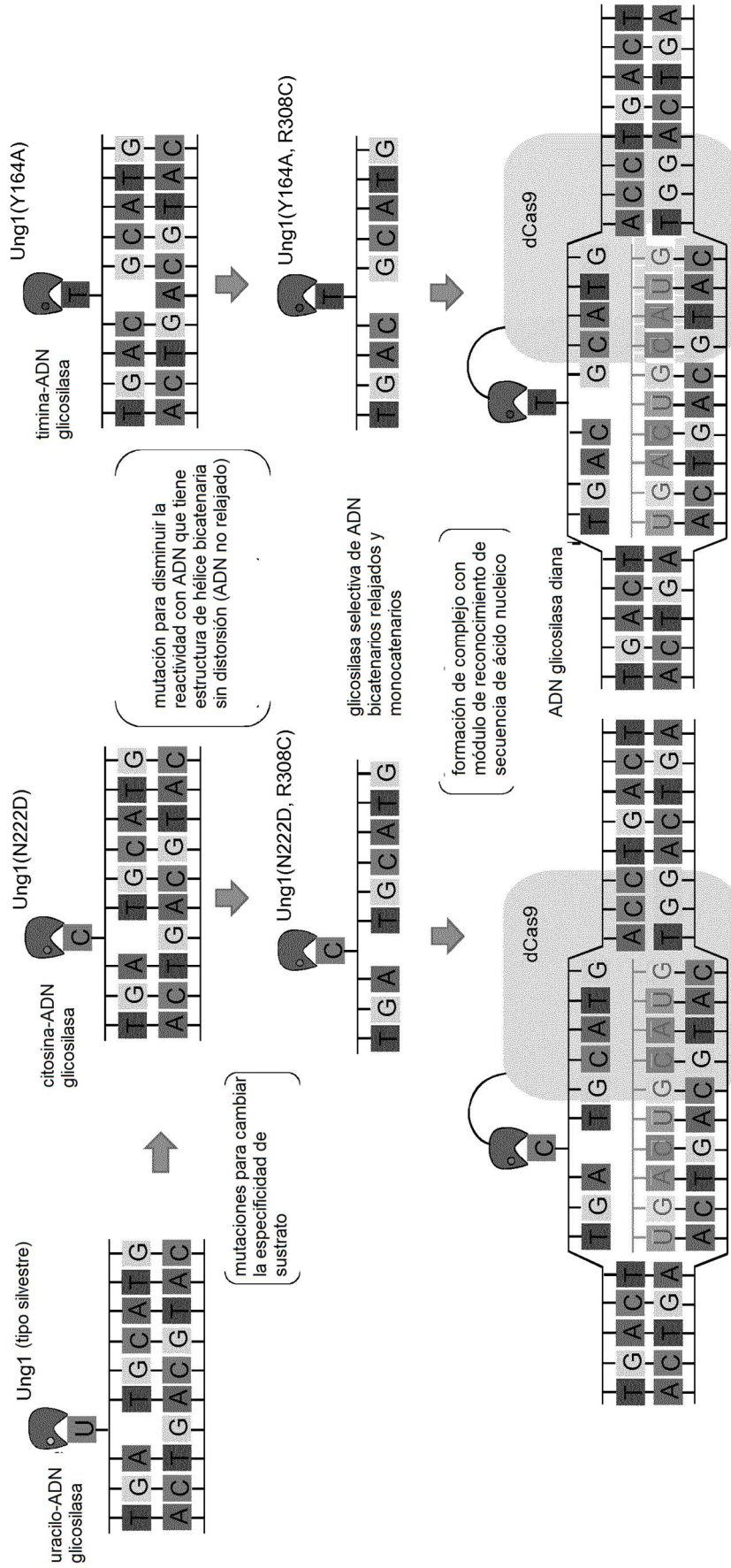


FIG. 2

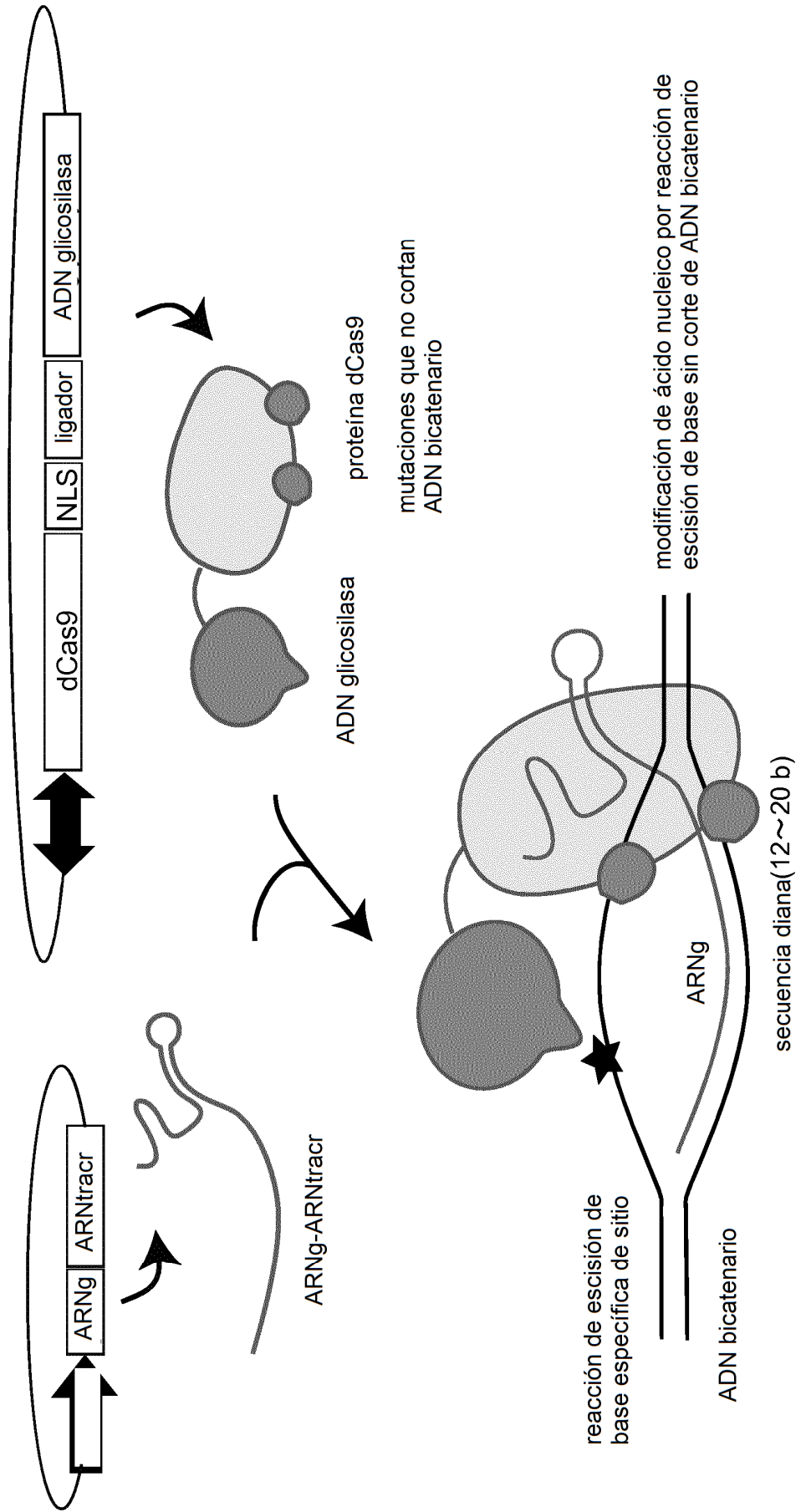


FIG. 3

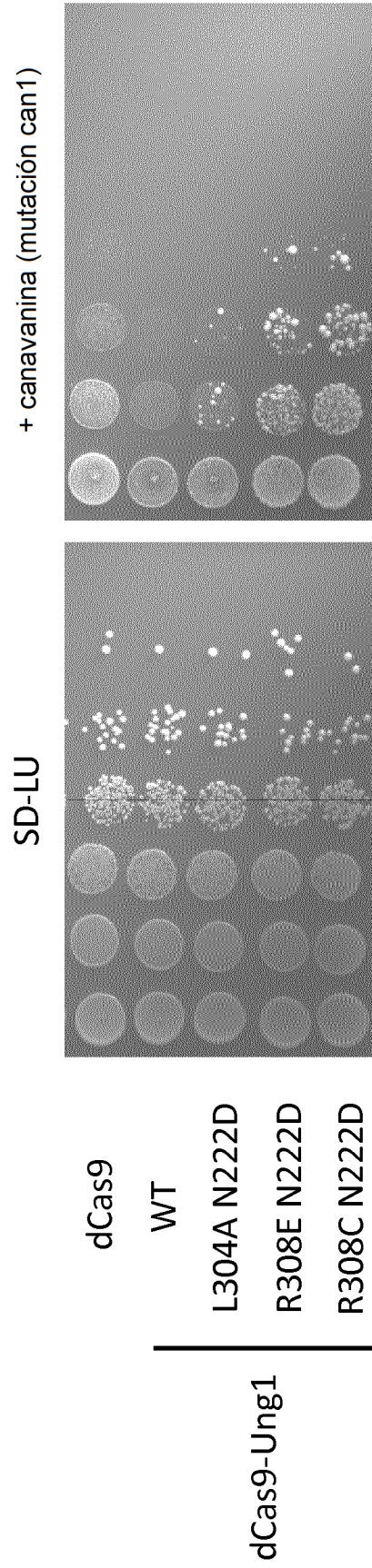


FIG. 4

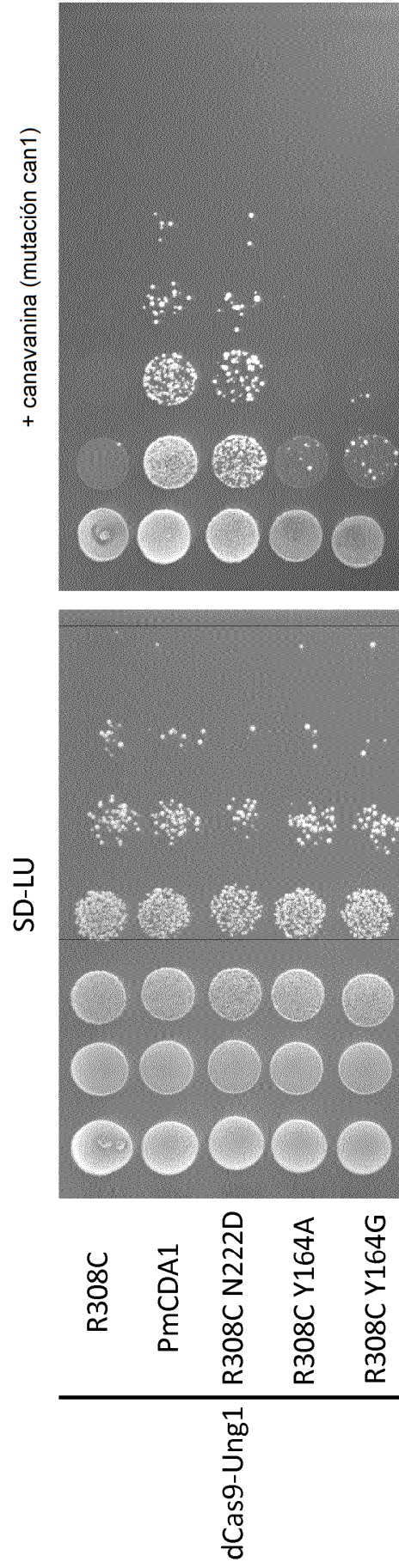


FIG. 5

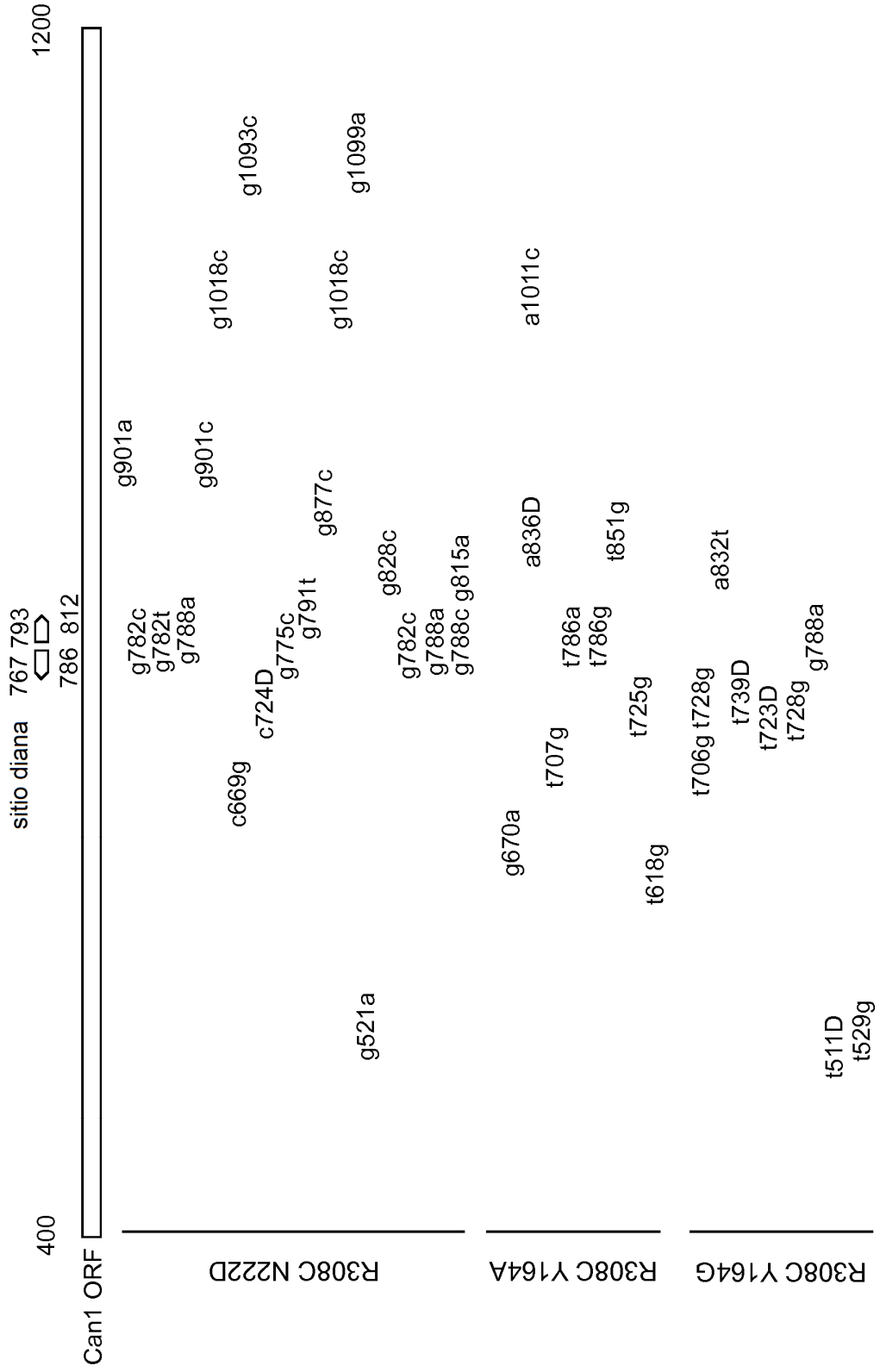


FIG. 6

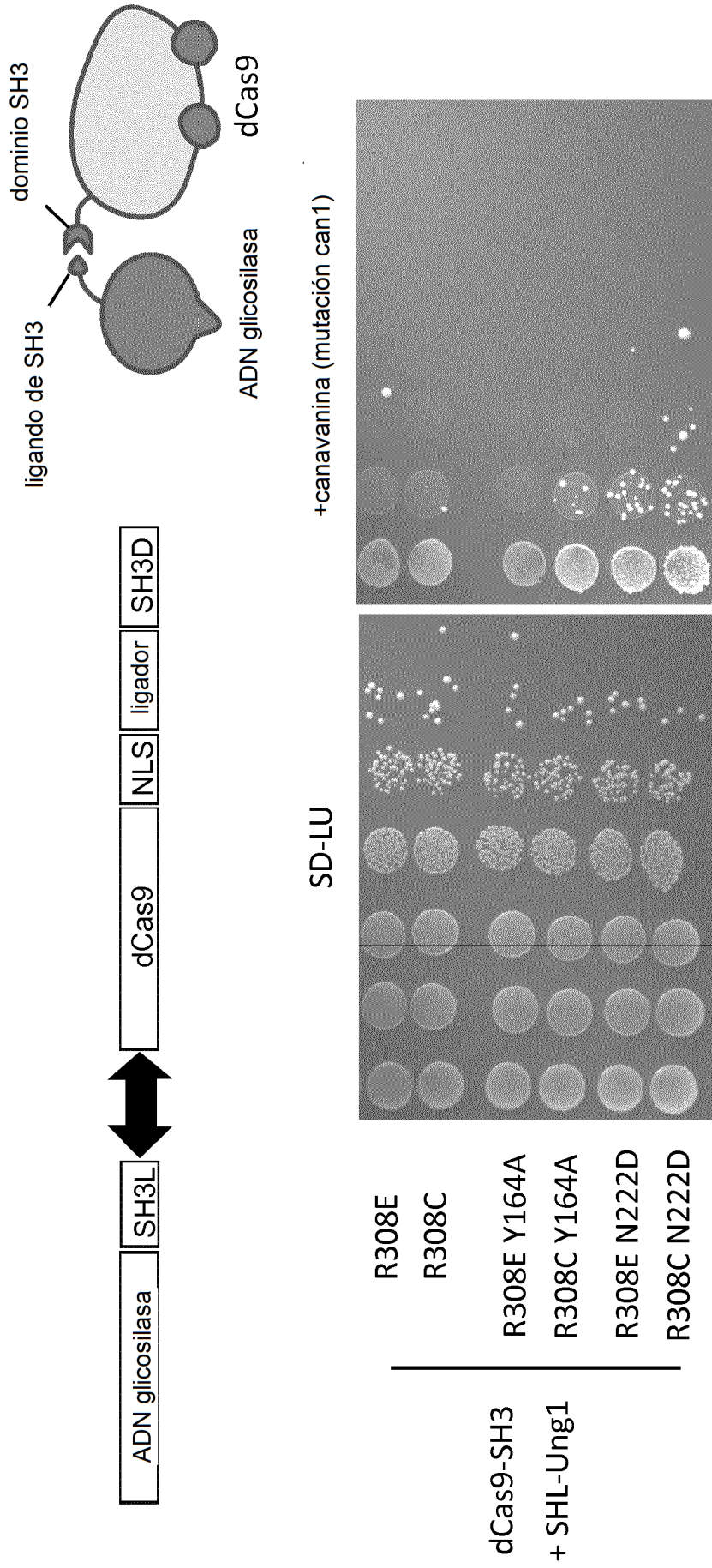


FIG. 7

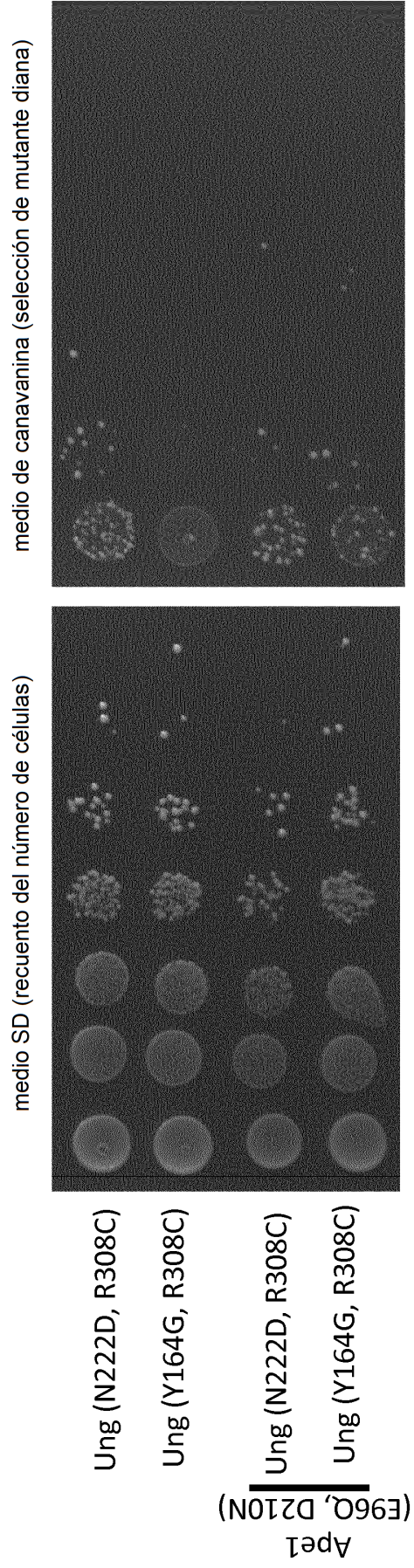


FIG. 8

Placa SD -Leu -Ura

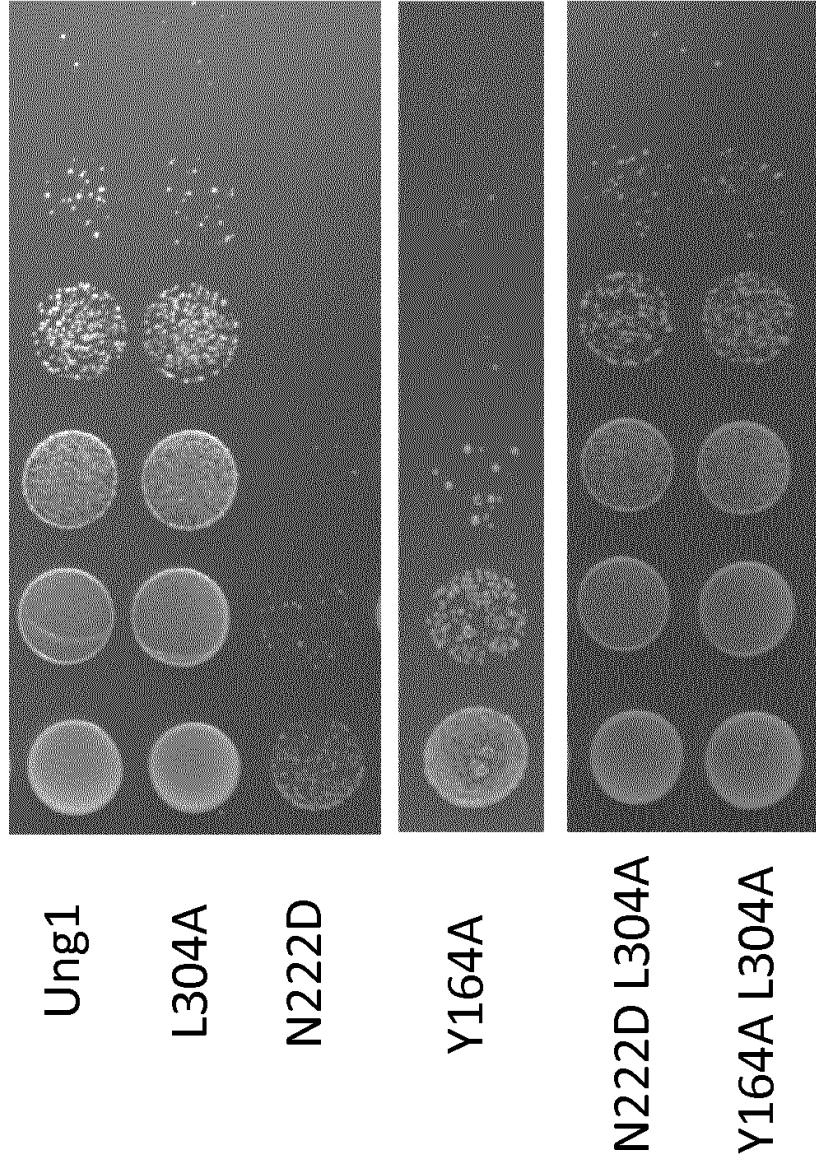
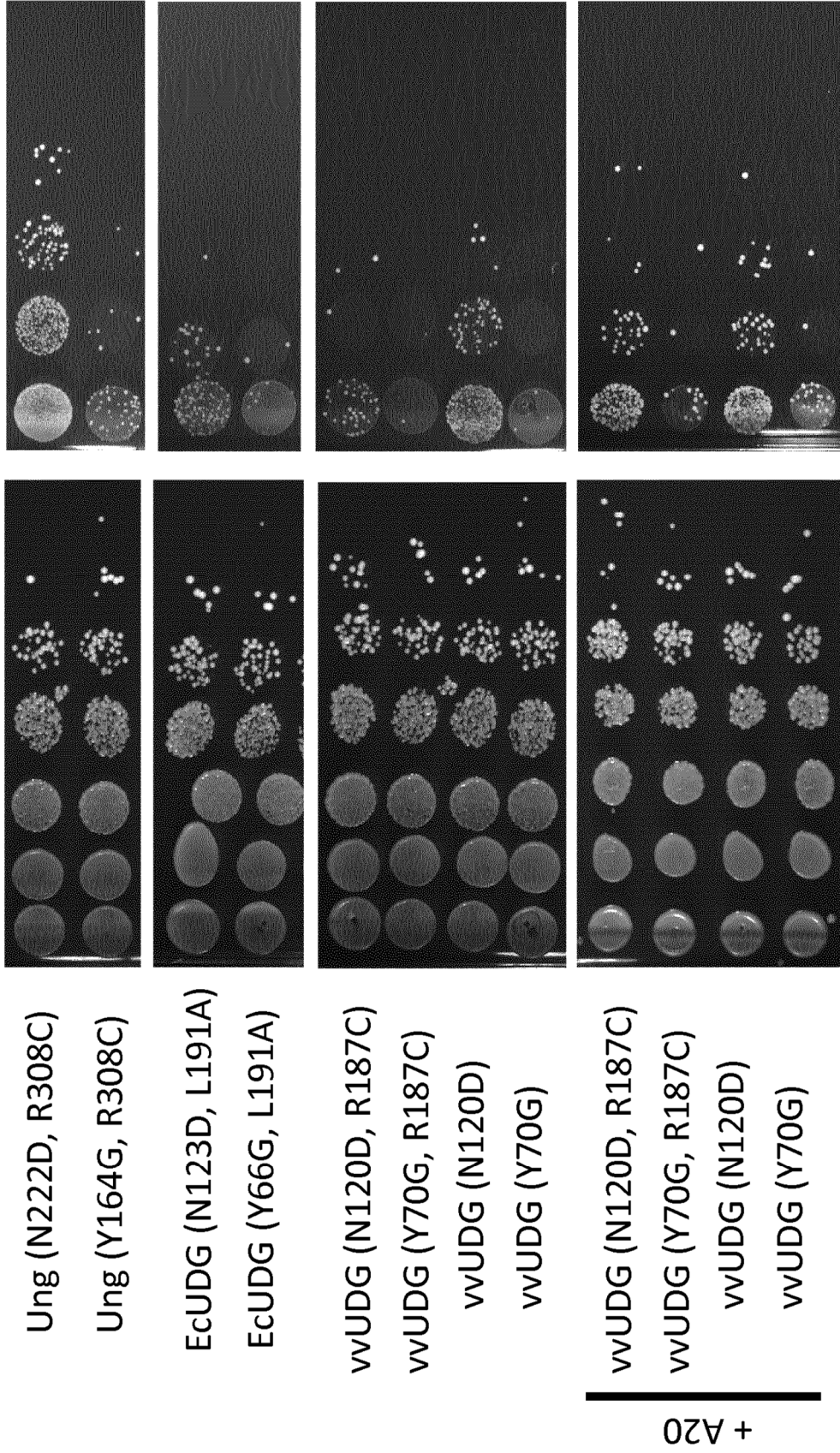


FIG. 9



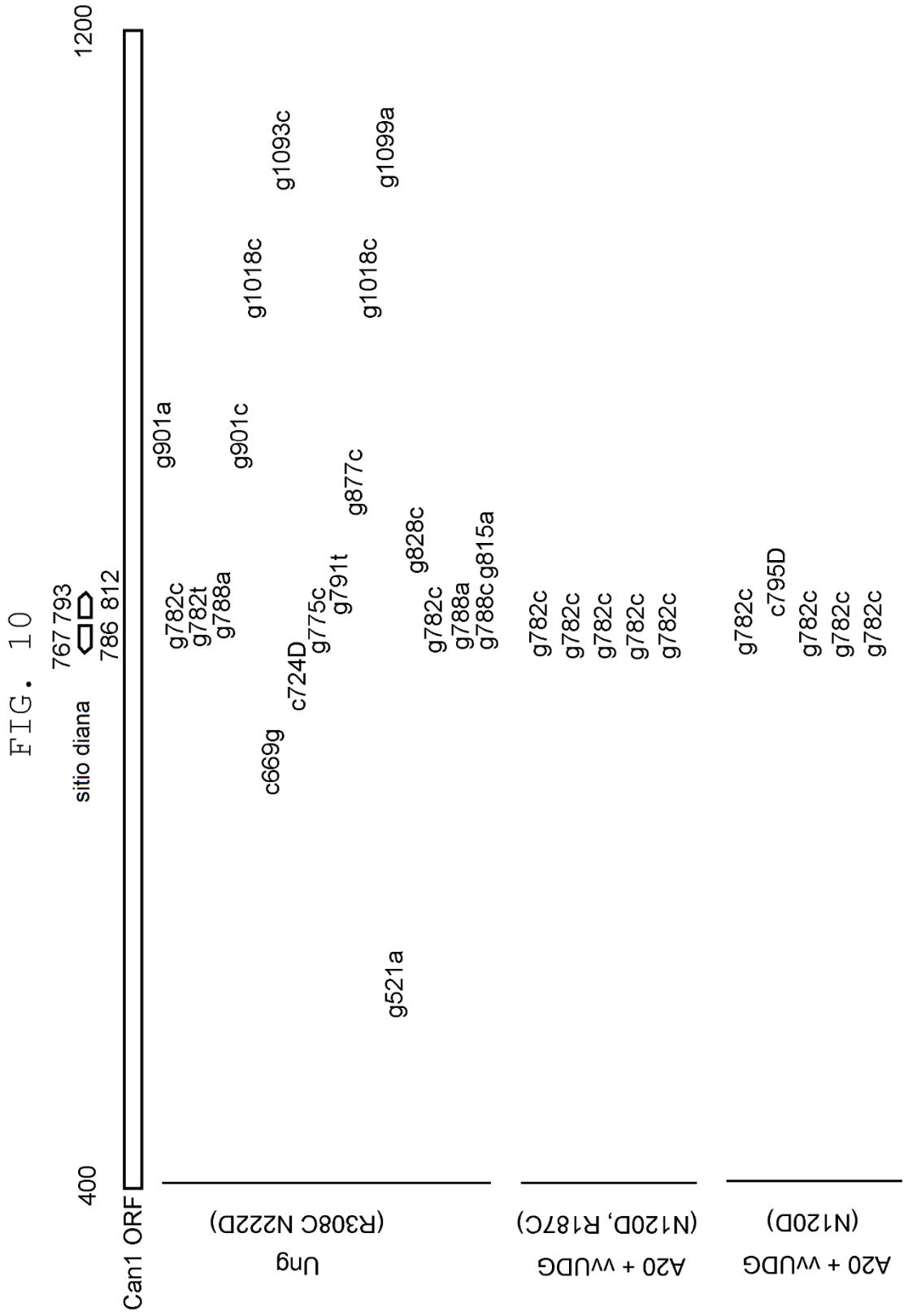


FIG. 11

