

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 715 666**

51 Int. Cl.:

C12N 15/90 (2006.01)

C12N 15/85 (2006.01)

C12N 15/86 (2006.01)

C12N 5/074 (2010.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **25.07.2014 PCT/US2014/048140**

87 Fecha y número de publicación internacional: **29.01.2015 WO15013583**

96 Fecha de presentación y número de la solicitud europea: **25.07.2014 E 14829377 (2)**

97 Fecha y número de publicación de la concesión europea: **19.12.2018 EP 3024964**

54 Título: **Ingeniería genómica**

30 Prioridad:

26.07.2013 US 201361858866 P

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

05.06.2019

73 Titular/es:

**PRESIDENT AND FELLOWS OF HARVARD
COLLEGE (100.0%)**

**17 Quincy Street
Cambridge, MA 02138, US**

72 Inventor/es:

CHURCH, GEORGE M.;
YANG, LUHAN;
CARGOL, MARC GUELL y
YANG, JOYCE LICHI

74 Agente/Representante:

DURAN-CORRETJER, S.L.P

ES 2 715 666 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

Ingeniería genómica

5 ANTECEDENTES

Se conoce la edición del genoma por medio de nucleasas específicas de secuencia. Véanse las referencias 1, 2 y 3. Una rotura en ADN bicatenario (ADNbc) mediada por nucleasa en el genoma puede repararse mediante dos mecanismos principales: unión de extremos no homólogos (NHEJ), que frecuentemente da como resultado la introducción de inserciones y deleciones (indels) no específicas, o reparación dirigida por homología (HDR), que incorpora una hebra homóloga como molde de reparación. Véase la referencia 4. Cuando se suministra una nucleasa específica de secuencia junto con una construcción de ADN donador homólogo que contiene las mutaciones deseadas, las eficacias de direccionamiento génico aumentan en 1.000 veces en comparación con tan solo la construcción donadora sola. Véase la referencia 5. Se ha notificado la utilización de oligodesoxirribonucleótidos monocatenarios ("ODNmc") como donadores de ADN. Véanse las referencias 21 y 22.

A pesar de los grandes avances en herramientas de edición genética, sigue habiendo muchos desafíos y cuestiones con respecto a la utilización de nucleasas diseñadas por ingeniería genética a medida en la modificación por ingeniería genética de células madre pluripotentes inducidas humanas ("hiPSC"). En primer lugar, a pesar de su simplicidad de diseño, las nucleasas efectoras de tipo activador de la transcripción (TALEN) seleccionan como diana secuencias de ADN particulares con copias en tándem de dominios de dirresiduo variable repetido (RVD). Véase la referencia 6. Aunque la naturaleza modular de los RVD simplifica el diseño de TALEN, sus secuencias repetitivas complican los procedimientos para sintetizar sus construcciones de ADN (véanse las referencias 2, 9 y 15-19) y también perjudican su utilización con vehículos de suministro de genes lentivirales. Véase la referencia 13.

En la práctica actual, se evalúan NHEJ y HDR frecuentemente utilizando ensayos diferenciados. A menudo se utilizan ensayos de endonucleasas sensibles a apareamientos erróneos (véase la referencia 14) para evaluar NHEJ, pero la exactitud cuantitativa de este procedimiento es variable y la sensibilidad se limita a frecuencias de NHEJ mayores de ~3%. Véase la referencia 15. HDR se evalúa frecuentemente mediante clonación y secuenciación, un procedimiento completamente diferente y a menudo engorroso. La sensibilidad es todavía un problema porque, aunque se notifican frecuentemente altas frecuencias de edición del orden del 50% para algunos tipos de células, tales como U2OS y K562 (véanse las referencias 12 y 14), las frecuencias son generalmente inferiores en hiPSC. Véase la referencia 10. Recientemente, se han notificado altas frecuencias de edición en hiPSC y hESC utilizando TALEN (véase la referencia 9), y frecuencias incluso superiores con el sistema de CRISPR Cas9-ARNg (véanse las referencias 16-19). Sin embargo, las tasas de edición en diferentes sitios parecen variar ampliamente (véase la referencia 17), y algunas veces no puede detectarse edición en todos o algunos sitios (véase la referencia 20).

Los sistemas de CRISPR-Cas bacterianos y de arqueas se basan en ARN de guía cortos en complejo con proteínas Cas para dirigir la degradación de secuencias complementarias presentes dentro de ácido nucleico foráneo invasor. Véase Deltcheva, E. y otros. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602-607 (2011); Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109, E2579-2586 (2012); Jinek, M. y otros. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821 (2012); Sapranaukas, R. y otros. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic acids research* 39, 9275-9282 (2011); y Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics* 45, 273-297 (2011). Una reconstitución *in vitro* reciente del sistema de CRISPR de tipo II de *S. pyogenes* demostró que ARNcr ("ARN de CRISPR") fusionado a un ARNtracr codificado en trans normalmente ("ARN de CRISPR de transactivación") es suficiente para dirigir la proteína Cas9 a secuencias de ADN diana de escisión específica de secuencia que se aparean con el ARNcr. La expresión de ARNg homólogo a un sitio diana da como resultado el reclutamiento de Cas9 y la degradación del ADN diana. Véase H. Deveau y otros., Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of Bacteriology* 190, 1390 (febrero, 2008).

55 CARACTERÍSTICAS

Aspectos de la presente divulgación se refieren a la utilización de nucleasas efectoras de tipo activador de la transcripción (TALEN) para modificar genéticamente una célula, tal como una célula somática o una célula madre. Se sabe que las TALEN incluyen secuencias repetidas. Aspectos de la presente divulgación se refieren a un procedimiento de alteración de ADN diana en una célula que incluye introducir en una célula una TALEN que carece de secuencias repetidas de 100 pb o más en el que la TALEN escinde el ADN diana y la célula experimenta unión de extremos no homólogos para producir ADN alterado en la célula. Según determinados aspectos, se han eliminado secuencias repetidas de longitud deseada de una TALEN. Según determinados aspectos, la TALEN carece de secuencias repetidas de una determinada longitud deseada. Según determinados aspectos, se proporciona una TALEN con secuencias repetidas de longitud deseada eliminadas. Según determinados aspectos, se modifica una TALEN para eliminar secuencias repetidas de longitud deseada. Según determinados aspectos, se

modifica por ingeniería genética una TALEN para eliminar secuencias repetidas de longitud deseada.

Los aspectos de la presente divulgación incluyen procedimientos de alteración de ADN diana en una célula que incluyen combinar dentro de una célula una TALEN que carece de secuencias repetidas de 100 pb o más y una secuencia de ácido nucleico donadora en la que la TALEN escinde el ADN diana y la secuencia de ácido nucleico donadora se inserta en el ADN en la célula. Aspectos de la presente divulgación se refieren a un virus que incluye una secuencia de ácido nucleico que codifica una TALEN que carece de secuencias repetidas de 100 pb o más. Aspectos de la presente divulgación se refieren a una célula que incluye una secuencia de ácido nucleico que codifica una TALEN que carece de secuencias repetidas de 100 pb o más. Según determinados aspectos descritos en el presente documento, la TALEN carece de secuencias repetidas de 100 pb o más, 90 pb o más, 80 pb o más, 70 pb o más, 60 pb o más, 50 pb o más, 40 pb o más, 30 pb o más, 20 pb o más, 19 pb o más, 18 pb o más, 17 pb o más, 16 pb o más, 15 pb o más, 14 pb o más, 13 pb o más, 12 pb o más, 11 pb o más o 10 pb o más.

Aspectos de la presente divulgación se refieren a preparar una TALE que incluyen combinar una endonucleasa, una ADN polimerasa, una ADN ligasa, una exonucleasa, una pluralidad de bloques de dímeros de ácido nucleico que codifican dominios de dirresiduo variable repetido y un vector de estructura principal de TALE-N/TF que incluye un sitio de corte de endonucleasa, activar la endonucleasa para que corte el vector de estructura principal de TALE-N/TF en el sitio de corte de endonucleasa para producir un primer extremo y un segundo extremo, activar la exonucleasa para crear una proyección en 3' y una en 5' en el vector de estructura principal de TALE-N/TF y la pluralidad de bloques de dímeros de ácido nucleico y para aparear el vector de estructura principal de TALE-N/TF y la pluralidad de bloques de dímeros de ácido nucleico en un orden deseado, activar la ADN polimerasa y la ADN ligasa para conectar el vector de estructura principal de TALE-N/TF y la pluralidad de bloques de dímeros de ácido nucleico. Un experto en la materia, basándose en la presente divulgación, será capaz fácilmente de identificar endonucleasas, ADN polimerasas, ADN ligasas, exonucleasas, bloques de dímeros de ácido nucleico que codifican dominios de dirresiduo variable repetido y vectores de estructura principal de TALE-N/TF adecuados.

Aspectos de la presente divulgación se refieren a un procedimiento de alteración de ADN diana en una célula madre que expresa una enzima que forma un complejo de colocalización con ARN complementario al ADN diana y que escinde el ADN diana de manera específica de sitio que incluye (a) introducir en la célula madre un primer ácido nucleico foráneo que codifica un ARN complementario al ADN diana y que guía a la enzima al ADN diana, en el que el ARN y la enzima son miembros de un complejo de colocalización para el ADN diana, introducir en la célula madre un segundo ácido nucleico foráneo que codifica una secuencia de ácido nucleico donadora, en el que el ARN y las secuencias de ácido nucleico donadoras se expresan, en el que el ARN y la enzima se localizan conjuntamente en el ADN diana, la enzima escinde el ADN diana y el ácido nucleico donador se inserta en el ADN diana para producir ADN alterado en la célula madre.

Aspectos de la presente divulgación se refieren a una célula madre que incluye un primer ácido nucleico foráneo que codifica una enzima que forma un complejo de colocalización con ARN complementario al ADN diana y que escinde el ADN diana de manera específica de sitio.

Aspectos de la presente divulgación se refieren a una célula que incluye un primer ácido nucleico foráneo que codifica una enzima que forma un complejo de colocalización con ARN complementario al ADN diana y que escinde el ADN diana de manera específica de sitio y que incluye un promotor inducible para promover la expresión de la enzima. De esta manera, puede regularse la expresión, por ejemplo, puede iniciarse y puede detenerse.

Aspectos de la presente divulgación se refieren a una célula que incluye un primer ácido nucleico foráneo que codifica una enzima que forma un complejo de colocalización con ARN complementario al ADN diana y que escinde el ADN diana de manera específica de sitio, en el que el primer ácido nucleico foráneo puede eliminarse del ADN genómico de la célula utilizando una enzima de eliminación, tal como una transposasa.

Aspectos de la presente divulgación se refieren a un procedimiento de alteración de ADN diana en una célula que expresa una enzima que forma un complejo de colocalización con ARN complementario al ADN diana y que escinde el ADN diana de manera específica de sitio que incluye (a) introducir en la célula un primer ácido nucleico foráneo que codifica una secuencia de ácido nucleico donadora, introducir en la célula, a partir del medio que rodea a la célula, un ARN complementario al ADN diana y que guía a la enzima al ADN diana, en el que el ARN y la enzima son miembros de un complejo de colocalización para el ADN diana, en el que la secuencia de ácido nucleico donadora se expresa, en el que el ARN y la enzima se localizan conjuntamente en el ADN diana, la enzima escinde el ADN diana y el ácido nucleico donador se inserta en el ADN diana para producir ADN alterado en la célula.

Aspectos de la presente divulgación se refieren a la utilización de una proteína de unión a ADN guiada por ARN para modificar genéticamente una célula madre. En un aspecto, la célula madre se ha modificado genéticamente para incluir un ácido nucleico que codifica la proteína de unión a ADN guiada por ARN y la célula madre expresa la proteína de unión a ADN guiada por ARN. Según un determinado aspecto, se optimizan ácidos nucleicos donadores para introducir mutaciones específicas para la edición del genoma utilizando o las TALEN modificadas o la proteína de unión a ADN guiada por ARN.

- Aspectos de la presente divulgación se refieren a la modificación de ADN, tal como modificación múltiplex de ADN, en una célula madre utilizando uno o más ARN de guía (ácidos ribonucleicos) para dirigir una enzima que tiene actividad nucleasa expresada por la célula madre, tal como una proteína de unión a ADN que tiene actividad nucleasa, a una ubicación diana en el ADN (ácido desoxirribonucleico) en los que la enzima corta el ADN y se inserta un ácido nucleico donador exógeno en el ADN, tal como mediante recombinación homóloga. Los aspectos de la presente divulgación incluyen ciclar o repetir las etapas de modificación de ADN en una célula madre para crear una célula madre que tiene múltiples modificaciones de ADN dentro de la célula. Las modificaciones pueden incluir la inserción de ácidos nucleicos donadores exógenos.
- Pueden lograrse múltiples inserciones de ácido nucleico exógeno mediante una única etapa de introducción en una célula madre, que expresa la enzima, de ácidos nucleicos que codifican una pluralidad de ARN y una pluralidad de ácidos nucleicos donadores exógenos, tal como mediante cotransformación, en la que los ARN se expresan y en la que cada ARN en la pluralidad guía a la enzima a un sitio particular del ADN, la enzima corta el ADN y se inserta uno de la pluralidad de ácidos nucleicos exógenos en el ADN en el sitio de corte. Según este aspecto, se crean muchas alteraciones o modificación del ADN en la célula en un único ciclo.
- Pueden lograrse múltiples inserciones de ácido nucleico exógeno en una célula mediante etapas o ciclos repetidos de introducción en una célula madre, que expresa la enzima, de uno o más ácidos nucleicos que codifican uno o más ARN o una pluralidad de ARN y uno o más ácidos nucleicos exógenos o una pluralidad de ácidos nucleicos exógenos en los que el ARN se expresa y guía a la enzima a un sitio particular del ADN, la enzima corta el ADN y se inserta el ácido nucleico exógeno en el ADN en el sitio de corte, para dar como resultado una célula que tiene múltiples alteraciones o inserciones de ADN exógeno en el ADN dentro de la célula madre. Según un aspecto, la célula madre que expresa la enzima se ha alterado genéticamente para expresar la enzima tal como introduciendo en la célula un ácido nucleico que codifica la enzima y que la célula madre puede expresar. De esta manera, los aspectos de la presente divulgación incluyen ciclar las etapas de introducir ARN en una célula madre que expresa la enzima, introducir ácido nucleico donador exógeno en la célula madre, expresar el ARN, formar un complejo de colocalización del ARN, la enzima y el ADN, cortar enzimáticamente el ADN mediante la enzima e insertar el ácido nucleico donador en el ADN. El ciclado o la repetición de las etapas anteriores da como resultado modificación genética multiplexada de una célula madre en múltiples loci, es decir, una célula madre que tiene múltiples modificaciones genéticas.
- Según determinados aspectos, las enzimas o proteínas de unión a ADN dentro del alcance de la presente divulgación incluyen una proteína que forma un complejo con el ARN de guía y con el ARN de guía que guía al complejo a una secuencia de ADN bicatenario en la que el complejo se une a la secuencia de ADN. Según un aspecto, la enzima puede ser una proteína de unión a ADN guiada por ARN, tal como una proteína de unión a ADN guiada por ARN de un sistema de CRISPR de tipo II, que se une al ADN y está guiado por ARN. Según un aspecto, la proteína de unión a ADN guiada por ARN es una proteína Cas9.
- Este aspecto de la presente divulgación puede denominarse localización conjunta de ARN y proteína de unión a ADN en o con el ADN bicatenario. De esta manera, puede utilizarse un complejo de proteína de unión a ADN-ARN de guía para cortar múltiples sitios del ADN bicatenario para crear una célula madre con múltiples modificaciones genéticas, tales como múltiples inserciones de ADN donador exógeno.
- Según determinados aspectos, se proporciona un procedimiento de producción de múltiples alteraciones en ADN diana en una célula madre que expresa una enzima que forma un complejo de colocalización con ARN complementario al ADN diana y que escinde el ADN diana de manera específica de sitio que incluye (a) introducir en la célula madre un primer ácido nucleico foráneo que codifica uno o más ARN complementarios al ADN diana y que guía a la enzima al ADN diana, en el que el uno o más ARN y la enzima son miembros de un complejo de colocalización para el ADN diana, introducir en la célula madre un segundo ácido nucleico foráneo que codifica una o más secuencias de ácido nucleico donadoras, en el que se expresan el uno o más ARN y la una o más secuencias de ácido nucleico donadoras, en el que el uno o más ARN y la enzima se localizan conjuntamente en el ADN diana, la enzima escinde el ADN diana y el ácido nucleico donador se inserta en el ADN diana para producir ADN alterado en la célula madre, y repetir la etapa (a) múltiples veces para producir múltiples alteraciones en el ADN en la célula madre.
- Según un aspecto, el ARN tiene entre aproximadamente 10 y aproximadamente 500 nucleótidos. Según un aspecto, el ARN tiene entre aproximadamente 20 y aproximadamente 100 nucleótidos.
- Según un aspecto, el uno o más ARN es un ARN de guía. Según un aspecto, el uno o más ARN es una fusión de ARNtracr-ARNcr.
- Según un aspecto, el ADN es ADN genómico, ADN mitocondrial, ADN viral o ADN exógeno.
- Según un aspecto, una célula puede modificarse genéticamente para que incluya de manera reversible un ácido nucleico que codifica una enzima de unión a ADN utilizando un vector que puede eliminarse fácilmente utilizando una enzima. Los expertos en la materia conocen procedimientos de vectores útiles e incluyen lentivirus, virus

adenoasociados, procedimientos de inserción dirigida mediados por nucleasa e integrasa y procedimientos de inserción mediados por transposón. Según un aspecto, el ácido nucleico que codifica una enzima de unión a ADN que se ha añadido, tal como utilizando un casete o vector, puede eliminarse en su totalidad junto con el casete y el vector y sin dejar una porción de tal ácido nucleico, casete o vector en el ADN genómico, por ejemplo. Tal eliminación se denomina en la materia eliminación "sin cicatrices", ya que el genoma queda igual que antes de la adición del ácido nucleico, casete o vector. Una realización a modo de ejemplo para la inserción y eliminación sin cicatrices es un vector PiggyBac disponible en el mercado de System Biosciences.

La presente invención da a conocer un procedimiento para alterar ADN diana en una célula madre tal como se expone en las reivindicaciones adjuntas.

Características y ventajas adicionales de determinadas realizaciones de la presente invención resultarán más evidentes de manera completa en la siguiente descripción de realizaciones y dibujos de la misma, y a partir de las reivindicaciones.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Las características y ventajas anteriores y otras de las presentes realizaciones se entenderán más completamente a partir de la siguiente descripción detallada de realizaciones ilustrativas tomadas conjuntamente con los dibujos adjuntos en los que:

la figura 1 se refiere a pruebas funcionales de re-TALEN en células madre y somáticas humanas.

(a) Representación esquemática de un diseño experimental para someter a prueba la eficacia de selección como diana del genoma. Una secuencia que codifica GFP integrada genómicamente se altera mediante la inserción de un codón de terminación y un fragmento genómico de 68 pb derivado del locus AAVS1 (parte inferior). La restauración de la secuencia de GFP mediante recombinación homóloga mediada por nucleasa con donador de tGFP (parte superior) da como resultado células GFP+ que pueden cuantificarse mediante FACS. Re-TALEN y TALEN seleccionan como diana secuencias idénticas dentro de fragmentos de AAVS1.

(b) Gráfico de barras que representa el porcentaje de células GFP+ introducidas por donador de tGFP solo, TALEN con donador de tGFP y re-TALEN con donador de tGFP en el locus diana, tal como se mide mediante FACS. (N=3, barra de error = DE). Debajo se muestran gráficos de FACS representativos.

(c) Visión general esquemática que representa la estrategia de selección como diana para el locus AAVS1 nativo. Se describieron el plásmido donador, que contiene el aceptor de corte y empalme (SA)-2A (péptidos de autoescisión), gen de resistencia a puromicina (PURO) y GFP (véase la referencia 10). Las ubicaciones de los cebadores de PCR utilizados para detectar acontecimientos de edición satisfactorios se representan como flechas azules.

(d) Se seleccionaron clones seleccionados como diana satisfactoriamente de hiPSC PGP1 con puromicina (0,5 ug/ml) durante 2 semanas. Se muestran imágenes de microscopía de tres clones GFP+ representativos. Las células se tiñeron también para los marcadores de pluripotencia TRA-1-60. Barra de escala: 200 μm.

(e) Ensayos de PCR realizados sobre estos clones de hiPSC GFP+ monoclonales demostraron inserciones satisfactorias de los casetes donadores en el sitio AAVS1 (carril 1,2,3), mientras que en hiPSC simples no muestran evidencias de inserción satisfactoria (carril C).

La figura 2 se refiere a una comparación de la eficacia de selección como diana del genoma de reTALEN y Cas9-ARNg sobre CCR5 en iPSC.

(a) Representación esquemática del diseño experimental de la modificación por ingeniería genética del genoma. En el sitio de selección como diana del par de re-TALEN o Cas9-ARNg, se suministró un ODNmc de 90 mers que portaba un apareamiento erróneo de 2 pb contra el ADN genómico junto con las construcciones de reTALEN o Cas9-ARNg en hiPSC PGP1. Los sitios de corte de las nucleasas se representan como flechas rojas en la figura.

(b) Análisis de secuenciación profunda de las eficacias de HDR y NHEJ para pares de re-TALEN (CCR5 n.º 3) y ODNmc, o el Cas9-ARNg y ODNmc. Se analizaron alteraciones en el genoma de hiPSC a partir de datos de secuencia de alto rendimiento mediante GEAS. Parte superior: se cuantificó HDR a partir de la fracción de lecturas que contenían una mutación puntual de 2 pb construida en el dentro del ODNmc (color azul), y se cuantificó la actividad de NHEJ a partir de la fracción de deleciones (color gris)/inserciones (color rojo) en cada posición específica en el genoma. Para los gráficos de reTALEN y ODNmc, se representan gráficamente líneas discontinuas verdes para marcar el límite externo de los sitios de unión de pares de re-TALEN, que están en las posiciones -26 pb y +26 pb en relación con el centro de los dos sitios de unión a re-TALEN. Para los gráficos de Cas9-ARNg y ODNmc, las líneas discontinuas verdes marcan el límite externo del sitio de selección como diana de ARNg, que están en las posiciones -20 y -1 pb en relación con la secuencia de PAM. Parte inferior: distribución de tamaño de deleción/inserción en hiPSC analizadas a partir de toda la población de NHEJ con tratamientos indicados anteriormente.

(c) La eficacia de edición del genoma de re-TALEN y Cas9-gRNA que seleccionan como diana CCR5 en HiPSC PGP1. Parte superior: representación esquemática de los sitios de edición del genoma seleccionados como

diana en CCR5. Los 15 sitios de selección como diana se ilustran mediante flechas azules por debajo. Para cada sitio, se cotranfectaron las células con un par de re-TALEN y su correspondiente donador de ODNmc que porta apareamientos erróneos de 2 pb contra el ADN genómico. Se sometieron a ensayo las eficacias de edición del genoma 6 días tras la transfección. De manera similar, se transfectaron 15 Cas9-ARNg con sus correspondientes ODNmc individualmente en PGP1-hiPSC para seleccionar como diana los mismos 15 sitios y se analizó la eficacia 6 días tras la transfección. Parte inferior: la eficacia de edición del genoma de re-TALEN y Cas9-ARNg que seleccionan como diana CCR5 en HiPSC PGP1. Los paneles 1 y 2 indican las eficacias de NHEJ y HDR mediadas por reTALEN. Los paneles 3 y 4 indican las eficacias de NHEJ y HDR mediadas por Cas9-ARNg. Se calcularon las tasas de NHEJ mediante la frecuencia de alelos genómicos que portan deleciones o inserciones en la región de selección como diana; se calcularon las tasas de HDR mediante la frecuencia de alelos genómicos que portan apareamientos erróneos de 2 pb. Panel 5, el perfil de ADNasa I HS de una línea celular hiPSC de la base de datos ENCODE (ADNasa HS Duke, iPS NIH7 DS). Cabe señalar que las escalas de los diferentes paneles son diferentes.

(f) La secuenciación de Sanger del amplicón de PCR a partir de las tres colonias de hiPSC seleccionadas como diana confirmó que están presentes las bases de ADN esperadas en el límite de inserción del genoma. M: marcador de tamaño molecular de ADN. C: control con ADN genómico de hiPSC simple.

La figura 3 se refiere a un estudio de parámetros funcionales que rigen HDR mediada por ODNmc con re-TALEN y Cas9-ARNg en HiPSC PGP1.

(a) Se cotransfectaron hiPSC PGP1 con un par de re-TALEN (n.º 3) y ODNmc de diferentes longitudes (50, 70, 90, 110, 130, 150, 170 nt). Todos los ODNmc presentaban un apareamiento erróneo idéntico de 2 pb contra el ADN genómico en el medio de su secuencia. Un ODNmc de 90 meros logró HDR óptima en el genoma seleccionado como diana. La evaluación de la eficacia de deleción e inserción producidas por NHEJ, HDR es tal como se describe en el presente documento.

(b) Se utilizaron ODNmc de 90 meros correspondientes al par de re-TALEN n.º 3 que contenían cada uno un apareamiento erróneo de 2 pb (A) en el centro y un apareamiento erróneo de 2 pb adicional (B) en diferentes posiciones desviadas de A (en donde las desviaciones variaban desde -30 pb→30 pb) para someter a prueba los efectos de desviaciones de la homología a lo largo del ODNmc. Se evaluó la eficacia de edición del genoma de cada ODNmc en HiPSC PGP1. El gráfico de barras inferior muestra la frecuencia de incorporación de A solo, B solo y A + B en el genoma seleccionado como diana. Las tasas de HDR disminuyen a medida que la distancia de desviaciones de la homología con respecto al centro aumenta.

(c) Se sometieron a prueba ODNmc dirigidos a sitios con distancias variables (-620 pb ~ 480 pb) lejos del sitio diana del par de re-TALEN n.º 3 para evaluar la distancia máxima dentro de la cual pueden colocarse ODNmc para introducir mutaciones. Todos los ODNmc portaban un apareamiento erróneo de 2 pb en el medio de sus secuencias. Se observó una eficacia de HDR mínima ($\leq 0,06\%$) cuando el apareamiento erróneo del ODNmc estaba situado 40 pb lejos del medio del sitio de unión del par de re-TALEN.

(d) Se cotransfectaron hiPSC PGP1 con Cas9-ARNg (AAVS1) y ODNmc de diferente orientación (Oc: complementario a ARNg; On: no complementario a ARNg) y diferentes longitudes (30, 50, 70, 90, 110 nt). Todos los ODNmc presentaban un apareamiento erróneo de 2 pb idéntico contra el ADN genómico en el medio de su secuencia. Un Oc de 70 meros logró HDR óptima en el genoma seleccionado como diana.

La figura 4 se refiere a la utilización de re-TALEN y ODNmc para obtener hiPSC con genoma editado monoclonal sin selección.

(a) Calendario del experimento.

(b) Eficacia de modificación por ingeniería genética del genoma del par de re-TALEN y ODNmc (n.º 3) evaluada mediante la plataforma de NGS descrita en la figura 2b.

(c) Resultados de la secuenciación de Sanger de colonias de hiPSC monoclonales tras la edición del genoma. El genotipo heterogéneo de 2 pb (CT/CT→TA/CT) se introdujo satisfactoriamente en el genoma de colonias de PGP1-iPS-3-11, PGP1-iPS-3-13.

(d) Tinción de inmunofluorescencia de PGP1-iPS-3-11 seleccionadas como diana. Se tiñeron las células para los marcadores de pluripotencia Tra-1-60 y SSEA4.

(e) Tinción de hematoxilina y eosina de secciones de teratoma generadas a partir de células PGP1-iPS-3-11 monoclonales.

Figura 5. Diseño de reTALE. (a) Alineación de secuencias del monómero de RVD de TALE original con monómeros en re-TALE-16.5 (re-TALE-M 1→re-TALE-M17). Las alteraciones de nucleótidos respecto a la secuencia original se resaltan en gris. (b) Prueba de repetitividad de re-TALE mediante PCR. El panel superior ilustra la estructura de re-TALE/TALE y las posiciones de los cebadores en la reacción del PCR. El panel inferior ilustra bandas de PCR con la condición indicada por debajo. Los marcadores de masa molecular de PCR presentes con el molde de TALE original (carril derecho).

Figura 6. Diseño y puesta en práctica del ensamblaje de ensamblaje de una única incubación de TALE (TASA).

(A) Representación esquemática de la biblioteca de bloques de dímeros de re-TALE para el ensamblaje de TASA. Hay una biblioteca de 10 bloques de dímeros de re-TALE que codifican dos RVD. Dentro de cada bloque,

los 16 dímeros comparten la misma secuencia de ADN excepto las secuencias que codifican RVD; los dímeros en diferentes bloques tienen secuencias distintas pero están diseñados de manera que compartan solapamientos de 32 pb con los bloques adyacentes. Se enumeran las secuencias de aminoácidos y ADN de un dímero (bloque 6_AC) a la derecha.

5 (B) Representación esquemática del ensamblaje de TASA. El panel izquierdo ilustra el procedimiento de ensamblaje de TASA: se realiza una reacción de incubación en un solo recipiente con una mezcla de enzimas/bloques de re-TALE/vectores de estructura principal de re-TALE-N/TF. El producto de reacción puede utilizarse directamente para transformación bacteriana. El panel derecho ilustra el mecanismo de TASA. El vector de destinación se linealiza mediante una endonucleasa a 37°C para eliminar mediante corte el casete de contraselección ccdB; la exonucleasa, que procesa el extremo de los bloques y vectores linealizados, expone proyecciones de ADNmc en el extremo de los fragmentos para permitir que los bloques y las estructuras principales de vector se apareen en un orden designado. Cuando la temperatura se eleva hasta 50°C, las polimerasas y ligasas funcionan conjuntamente sellando el hueco, produciendo las construcciones finales listas para la transformación.

10 (C) Eficacia de ensamblaje de TASA para re-TALE que presentan diferentes longitudes de monómero. Los bloques utilizados para el ensamblaje se ilustran a la izquierda y la eficacia de ensamblaje se presenta a la derecha.

Figura 7. La funcionalidad e integridad de secuencia de Lenti-reTALE.

20

Figura 8. La sensibilidad y reproducibilidad de GEAS.

(A) Análisis basado en información del límite de detección de HDR. Dado el conjunto de datos de re-TALEN (n.º 10)/ODNmc, se identificaron las lecturas que contenían la edición esperada (HDR) y estas lecturas de HDR se eliminaron sistemáticamente para generar diferentes conjuntos de datos artificiales con una señal de edición "diluida". Se generaron conjuntos de datos con el 100, el 99,8, el 99,9, el 98,9, el 97,8, el 89,2, el 78,4, el 64,9, el 21,6, el 10,8, el 2,2, el 1,1, el 0,2, el 0,1, el 0,02 y el 0% de eliminación de HDR para generar conjuntos de datos artificiales con una eficacia de HR que oscila entre el 0~0,67%. Para cada conjunto de datos individual, se estimó la información mutua (MI) de la señal de fondo (en color púrpura) y la señal obtenida en el sitio de selección como diana (en color verde). MI en el sitio de selección como diana es notablemente mayor que el fondo cuando la eficacia de HDR está por encima del 0,0014%. Se estimó un límite de detección de HDR de entre el 0,0014% y el 0,0071%. En el presente documento se describe el cálculo de MI.

(B) La prueba de reproducibilidad del sistema de evaluación de la edición del genoma. Los pares de gráficos (superior e inferior) muestran los resultados de evaluación de HDR y NHEJ de dos réplicas con el par de re-TALEN y el tipo de célula indicados anteriormente. Para cada experimento, se realizaron independientemente nucleofección, amplificación genómica dirigida, secuenciación profunda y análisis de datos. La variación de evaluación de la edición del genoma de réplicas se calculó como $\sqrt{2} (|HDR1-HDR2|)/((HDR1+HDR2)/2) = \Delta HDR/HDR$ y $\sqrt{2} (|NHEJ1-NHEJ2|)/((NHEJ1+NHEJ2)/2) = \Delta NHEJ/NHEJ$ y los resultados de la variación se enumeran por debajo de los gráficos. La variación promedio del sistema fue de $(19\%+11\%+4\%+9\%+10\%+35\%)/6=15\%$. Los factores que pueden contribuir a las variaciones incluyen el estado de las células bajo nucleofección, la eficacia de nucleofección y la calidad y cobertura de la secuenciación.

Figura 9. Análisis estadístico de las eficacias de NHEJ y HDR mediante reTALEN y Cas9-ARNg en CCR5.

45 (a) La correlación de las eficacias de HR y NHEJ mediadas por reTALEN en sitios idénticos en iPSC ($r=0,91$, $P<1 \times 10^{-5}$).
 (b) La correlación de las eficacias de HR y NHEJ mediadas por Cas9-ARNg en sitios idénticos en iPSC ($r=0,74$, $P=0,002$).
 50 (c) La correlación de las eficacias de NHEJ mediadas por Cas9-ARNg y la temperatura T_m del sitio de selección como diana de ARNg en iPSC ($r=0,52$, $P=0,04$).

Figura 10. El análisis de correlación de la eficacia de edición del genoma y estado epigenético. Se utilizó correlación de Pearson para estudiar posibles asociaciones entre la sensibilidad a ADNasa I y las eficacias de modificación por ingeniería genética del genoma (HR, NHEJ). Se comparó la correlación observada con un conjunto aleatorizado (N=100.000). Correlaciones observadas mayores del percentil 95, o menores del percentil 5 de la distribución simulada se consideraron como posibles asociaciones. No se observó correlación notable entre la sensibilidad a ADNasa I y las eficacias de NHEJ/HR.

Figura 11. El impacto del apareamiento de homología en la edición del genoma mediada por ODNmc.

60 (a) En el experimento descrito en la figura 3b, la HDR global tal como se mide mediante la tasa a la que se incorporó el apareamiento erróneo de 2 pb central (A) disminuyó a medida que los apareamientos erróneos secundarios B aumentaban su distancia desde el A (la posición relativa de B con respecto a A varía entre -30→30 pb). Las tasas de incorporación superiores cuando B está sólo 10 pb lejos de A (-10 pb y +10 pb) pueden reflejar una necesidad menor de apareamiento del ODNmc contra ADN genómico proximal a la rotura del ADNbc.
 65 (b) Distribución de longitudes de conversión génica a lo largo del ODNmc. A cada distancia de B con respecto a

A, una fracción de acontecimientos de HDR incorpora solo A mientras que otra fracción incorpora tanto A como B. Estos dos acontecimientos pueden ser interpretables en cuanto a tramos de conversión génica (Elliott y otros., 1998), mediante lo cual los acontecimientos A+B representan tramos de conversión largos que se extienden más allá de los acontecimientos de B y los acontecimientos de sólo A representan unos más cortos que no alcanzan hasta B. Según esta interpretación, puede estimarse una distribución de longitudes de conversión génica en ambas direcciones a lo largo del oligonucleótido (el medio del ODNmc se define como 0, la conversión sigue hacia el extremo 5' de ODNmc como sentido -, y el extremo 3' como sentido +). La incidencia de los tramos de conversión génica disminuye progresivamente a medida que sus longitudes aumentan, un resultado muy similar a las distribuciones de tramos de conversión génica observadas con donadores de ADNbc, pero a una escala de distancia altamente comprimida de decenas de pb para el oligo de ADNmc frente a cientos de bases para donadores de ADNbc.

(c) Ensayos para tramos de conversión génica utilizando un único ODNmc que contiene una serie de mutaciones y que miden series contiguas de incorporaciones. Se utilizó un donador de ODNmc con tres pares de apareamientos erróneos de 2 pb (color naranja) espaciados a intervalos de 10 nt a ambos lados del apareamiento erróneo de 2 pb central (parte superior). Se detectaron pocas lecturas de secuenciación genómica (véase la referencia 62) que portaban ≥ 1 apareamiento erróneo definido por ODNmc entre >300.000 lecturas de secuenciación en esta región. Todas estas lecturas se representaron gráficamente (parte inferior) y la secuencia de las lecturas se codificó mediante colores. Naranja: apareamientos erróneos definidos; verde: secuencia de tipo natural. La edición del genoma con este ODNmc dio lugar a un patrón en el que la mutación central sola se incorporó el 85% (53/62) de las veces, con múltiples apareamientos erróneos B incorporados las otras veces. Aunque los números de acontecimientos de incorporación de B eran demasiado bajos como para estimar una distribución de longitudes de tramo > 10 pb, queda claro que predomina la región de tramo corto de desde -10-10 pb.

Figura 12. Eficacias de edición del genoma de Cas9-ARNg nucleasas y nickasas.

Se cotransfectaron iPSC PGP1 con una combinación de nucleasa (C2) (Cas9-ARNg) o nickasa (Cc) (Cas9D10A-ARNg) y ODNmc de diferente orientación (Oc y On). Todos los ODNmc presentaban un apareamiento erróneo de 2 pb idéntico contra el ADN genómico en el medio de su secuencia. La evaluación de HDR se describe en el presente documento.

Figura 13. El diseño y la optimización de la secuencia de re-TALE.

Se hizo evolucionar la secuencia de re-TALE en varios ciclos de diseño para eliminar repeticiones. En cada ciclo, se evalúan secuencias sinónimas de cada repetición. Se seleccionan aquéllas con la distancia Hamming más grande. La secuencia final con $\Delta G = -9,8$ kcal/mol. Se proporcionó un paquete R para llevar a cabo este marco general para el diseño de proteínas sintéticas.

La figura 14 es una imagen de gel que muestra la validación por PCR de la inserción genómica de Cas 9 en células PGP1. Los carriles 3, 6, 9, 12 son productos de PCR de líneas celulares PGP1 simples.

La figura 15 es un gráfico del nivel de expresión de ARNm de Cas9 bajo la inducción.

La figura 16 es un gráfico que muestra la eficacia de selección como diana del genoma mediante diferentes diseños de ARN.

La figura 17 es un gráfico que muestra la eficacia de selección como diana del genoma del 44% de recombinación homóloga lograda mediante una fusión de ARN de guía - ADN donador.

La figura 18 es un diagrama que muestra el genotipo de líneas celulares PGP1 isógenas generadas mediante el sistema descrito en el presente documento. PGP1-iPS-BTHH tiene el fenotipo de delección de nucleótidos individuales que el paciente BTHH. PGP1-NHEJ tiene delecciones de 4 pb que generaron mutaciones del marco de lectura de un modo diferente.

La figura 19 es un gráfico que muestra que cardiomiocitos derivados de PGP1 iPS isógenas recapitulaban la producción de ATP defectuosa y la actividad específica de ATPasa F1F0 tal como se demostró en células específicas del paciente.

DESCRIPCIÓN DETALLADA

Aspectos de la presente divulgación se refieren a la utilización de una TALEN que carece de determinadas secuencias repetidas, para la modificación por ingeniería genética de ácido nucleico, por ejemplo cortando ácido nucleico bicatenario. La utilización de la TALEN para cortar ácido nucleico bicatenario puede dar como resultado unión de extremos no homólogos (NHEJ) o recombinación homóloga (HR). Aspectos de la presente divulgación también contemplan la utilización de una TALEN que carece de secuencias repetidas para la modificación por ingeniería genética de ácido nucleico, por ejemplo cortando ácido nucleico bicatenario, en presencia de un ácido nucleico donador e inserción del ácido nucleico donador en el ácido nucleico bicatenario, tal como mediante unión de extremos no homólogos (NHEJ) o recombinación homóloga (HR).

Las nucleasas efectoras de tipo activador de la transcripción (TALEN) se conocen en la materia e incluyen enzimas de restricción artificiales generadas fusionando un dominio de unión a ADN efector TAL con un dominio de escisión de ADN. Las enzimas de restricción son enzimas que cortan hebras de ADN en una secuencia específica. Pueden diseñarse por ingeniería genética efectoras de tipo activador de la transcripción (TALE) para que se unan a una secuencia de ADN deseada. Véase Boch, Jens (febrero de 2011). "TALEs of genome targeting". Nature Biotechnology 29 (2): 135-6. Al combinar un TALE diseñado por ingeniería genética de este tipo con un dominio de

escisión de ADN (que corta hebras de ADN), se produce una TALEN que es una enzima de restricción que es específica para cualquier secuencia de ADN deseada. Según determinados aspectos, la TALEN se introduce en una célula para la edición de ácido nucleico diana *in situ*, tal como edición del genoma *in situ*.

5 Según un aspecto, puede utilizarse el dominio de escisión de ADN no específico del extremo de la endonucleasa FokI para construir nucleasas híbridas que son activas en células de levadura, células vegetales y células animales. El dominio FokI funciona como un dímero, que requiere dos construcciones con dominios de unión a ADN únicos para sitios en el genoma diana con orientación y separación apropiadas. Tanto el número de residuos de aminoácido entre el dominio de unión a ADN de TALE como el dominio de escisión de FokI como el número de bases entre los dos sitios de unión a TALEN individuales afectan a la actividad.

La relación entre la secuencia de aminoácidos y el reconocimiento de ADN del dominio de unión a TALE permite lograr proteínas que pueden diseñarse. Pueden utilizarse programas de software tales como DNABworks para diseñar construcciones de TALE. Los expertos en la materia conocen otros procedimientos de diseño de construcciones de TALE. Véanse Cermak, T.; Doile, E. L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J. A.; Somia, N. V. y otros. (2011). "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting". *Nucleic Acids Research*. doi:10.1093/nar/gkr218; Zhang, Feng; y otros (febrero de 2011). "Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription", *Nature Biotechnology* 29 (2): 149-53; Morbitzer, R.; Elsaesser, J.; Hausner, J.; Lahaye, T. (2011). "Assembly of custom TALE-type DNA binding domains by modular cloning". *Nucleic Acids Research*. doi:10.1093/nar/gkr151; Li, T.; Huang, S.; Zhao, X.; Wright, D. A.; Carpenter, S.; Spalding, M. H.; Weeks, D. P.; Yang, B. (2011). "Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes". *Nucleic Acids Research*. doi:10.1093/nar/gkr188; Geißler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011). "Transcriptional Activators of Human Genes with Programmable DNA-Specificity". En Shiu, Shin-Han. *PLoS ONE* 6 (5): e19509; Weber, E.; Gruetzner, R.; Werner, S.; Engler, C.; Marillonnet, S. (2011). "Assembly of Designer TAL Effectors by Golden Gate Cloning". En Bendahmane, Mohammed. *PLoS ONE* 6 (5): e19722.

Según un aspecto a modo de ejemplo, una vez que los genes de TALEN se han ensamblado, pueden insertarse en plásmidos, según determinadas realizaciones; los plásmidos se utilizan entonces para transfectar la célula diana en donde los productos génicos se expresan y entran en el núcleo para acceder al genoma. Según aspectos a modo de ejemplo, pueden utilizarse TALEN, tal como se describe en el presente documento, para editar ácidos nucleicos diana, tales como genomas, induciendo roturas bicatenarias (DSB), a las que las células responden con mecanismos de reparación. Los mecanismos de reparación a modo de ejemplo incluyen unión de extremos no homólogos (NHEJ) que reconecta el ADN a partir de cualquier lado de una rotura bicatenaria en donde hay muy poco o ningún solapamiento de secuencia para el apareamiento. Este mecanismo de reparación induce errores en el genoma por medio de inserción o delección (indels), o transposición cromosómica; cualquiera de tales errores puede hacer que los productos génicos codificados en esa ubicación no sean funcionales. Véase Miller, Jeffrey; y otros (febrero de 2011). "A TALE nuclease architecture for efficient genome editing". *Nature Biotechnology* 29 (2): 143-8. Debido a que esta actividad puede variar dependiendo de la especie, el tipo de célula, el gen diana y la nucleasa utilizada, la actividad puede monitorizarse utilizando un ensayo de escisión de heterodúplex que detecta cualquier diferencia entre dos alelos amplificados por PCR. Los productos de escisión pueden visualizarse en geles de agarosa sencillos o sistemas de gel plano.

Alternativamente, puede introducirse ADN en un genoma a través de NHEJ en presencia de fragmentos de ADN bicatenario exógeno. La reparación dirigida por homología también puede introducir ADN foráneo en la DSB ya que se utilizan secuencias bicatenarias transfectadas como moldes para las enzimas de reparación. Según determinados aspectos, las TALEN descritas en el presente documento pueden utilizarse para generar clones de células madre embrionarias humanas modificadas de manera estable y células madre pluripotentes inducidas (iPSC). Según determinados aspectos, las TALEN descritas en el presente documento pueden utilizarse para generar especies deficientes tales como *C. elegans*, ratas deficientes, ratones deficientes o pez cebra deficiente.

Según un aspecto de la presente divulgación, las realizaciones se refieren a la utilización de ADN exógeno, enzimas nucleasa tales como proteínas de unión a ADN y ARN de guía para localizarse conjuntamente con el ADN dentro de una célula madre y digerir o cortar el ADN con inserción del ADN exógeno. Los expertos en la materia saben fácilmente que tales proteínas de unión a ADN se unen a ADN para diversos propósitos. Tales proteínas de unión a ADN pueden producirse de manera natural. Las proteínas de unión a ADN incluidas dentro del alcance de la presente divulgación incluyen las que pueden guiarse mediante ARN, denominado en el presente documento ARN de guía. Según este aspecto, el ARN de guía y la proteína de unión a ADN guiada por ARN forman un complejo de colocalización en el ADN. Tales proteínas de unión a ADN que tienen actividad nucleasa las conocen los expertos en la materia, e incluyen proteínas de unión a ADN que se producen de manera natural que tienen actividad nucleasa, tales como proteínas Cas9 presentes, por ejemplo, en sistemas de CRISPR de tipo II. Tales proteínas Cas9 y sistemas de CRISPR de tipo II están bien documentados en la materia. Véase Makarova y otros., *Nature Reviews, Microbiology*, vol. 9, junio de 2011, págs. 467-477 incluyendo toda la información complementaria.

Proteínas de unión a ADN a modo de ejemplo que tienen actividad nucleasa funcionan mellando o cortando ADN

bicatenario. Tal actividad nucleasa puede resultar de la proteína de unión a ADN que tiene una o más secuencias de polipéptido que presentan actividad nucleasa. Tales proteínas de unión a ADN a modo de ejemplo pueden tener dos dominios nucleasa independientes, siendo responsable cada dominio de cortar o mellar una hebra particular del ADN bicatenario. Secuencias de polipéptido a modo de ejemplo que tienen actividad nucleasa conocidas por los expertos en la materia incluyen el dominio relacionado con nucleasa McrA-HNH y el dominio de nucleasa de tipo RuvC. Por consiguiente, proteínas de unión a ADN a modo de ejemplo son aquellas que en la naturaleza contienen uno o más del dominio relacionado con nucleasa McrA-HNH y el dominio de nucleasa de tipo RuvC.

Una proteína de unión a ADN a modo de ejemplo es una proteína de unión a ADN guiada por ARN de un sistema de CRISPR de tipo II. Una proteína de unión a ADN a modo de ejemplo es una proteína Cas9.

En *S. pyogenes*, Cas9 genera una rotura bicatenaria de extremos romos 3 pb cadena arriba del motivo adyacente de protoespaciador (PAM) por medio de un proceso mediado por dos dominios catalíticos en la proteína: un dominio HNH que escinde la hebra complementaria del ADN y un dominio de tipo RuvC que escinde la hebra no complementaria. Véase Jinke y otros., Science 337, 816-821 (2012). Se sabe que las proteínas Cas9 existen en muchos sistemas de CRISPR de tipo II incluyendo los siguientes tal como se identifican en la información complementaria de Makarova y otros., Nature Reviews, Microbiology, vol. 9, junio de 2011, págs. 467-477: *Methanococcus maripaludis* C7; *Corynebacterium diphtheriae*; *Corynebacterium efficiens* YS-314; *Corynebacterium glutamicum* ATCC 13032 Kitasato; *Corynebacterium glutamicum* ATCC 13032 Bielefeld; *Corynebacterium glutamicum* R; *Corynebacterium kroppenstedtii* DSM 44385; *Mycobacterium abscessus* ATCC 19977; *Nocardia farcinica* IFM10152; *Rhodococcus erythropolis* PR4; *Rhodococcus jostii* RHA1; *Rhodococcus opacus* B4 uid36573; *Acidothermus cellulolyticus* 11B; *Arthrobacter chlorophenolicus* A6; *Kribbella flavida* DSM 17836 uid43465; *Thermomonospora curvata* DSM 43183; *Bifidobacterium dentium* Bd1; *Bifidobacterium longum* DJO10A; *Slackia heliotrinireducens* DSM 20476; *Persephonella marina* EX H1; *Bacteroides fragilis* NCTC 9434; *Capnocytophaga ochracea* DSM 7271; *Flavobacterium psychrophilum* JIP02 86; *Akkermansia muciniphila* ATCC BAA 835; *Roseiflexus castenholzii* DSM 13941; *Roseiflexus* RS1; *Synechocystis* PCC6803; *Elusimicrobium minutum* Pei191; filotipo de bacterias de grupo 1 termita sin cultivar Rs D17; *Fibrobacter succinogenes* S85; *Bacillus cereus* ATCC 10987; *Listeria innocua*; *Lactobacillus casei*; *Lactobacillus rhamnosus* GG; *Lactobacillus salivarius* UCC118; *Streptococcus agalactiae* A909; *Streptococcus agalactiae* NEM316; *Streptococcus agalactiae* 2603; *Streptococcus dysgalactiae equisimilis* GGS 124; *Streptococcus equi zooepidemicus* MGCS10565; *Streptococcus gallolyticus* UCN34 uid46061; *Streptococcus gordonii* Challis sust CHI; *Streptococcus mutans* NN2025 uid46353; *Streptococcus mutans*; *Streptococcus pyogenes* M1 GAS; *Streptococcus pyogenes* MGAS5005; *Streptococcus pyogenes* MGAS2096; *Streptococcus pyogenes* MGAS9429; *Streptococcus pyogenes* MGAS 10270; *Streptococcus pyogenes* MGAS6180; *Streptococcus pyogenes* MGAS315; *Streptococcus pyogenes* SSI-1; *Streptococcus pyogenes* MGAS10750; *Streptococcus pyogenes* NZ131; *Streptococcus thermophilus* CNRZ1066; *Streptococcus thermophilus* LMD-9; *Streptococcus thermophilus* LMG 18311; *Clostridium botulinum* A3 Loch Maree; *Clostridium botulinum* B Eklund 17B; *Clostridium botulinum* Ba4 657; *Clostridium botulinum* F Langeland; *Clostridium cellulolyticum* H10; *Fingoldia magna* ATCC 29328; *Eubacterium rectale* ATCC 33656; *Mycoplasma gallisepticum*; *Mycoplasma mobile* 163K; *Mycoplasma penetrans*; *Mycoplasma synoviae* 53; *Streptobacillus moniliformis* DSM 12112; *Bradyrhizobium* BTail; *Nitrobacter hamburgensis* X14; *Rhodopseudomonas palustris* BisB18; *Rhodopseudomonas palustris* BisB5; *Parvibaculum lavamentivorans* DS-1; *Dinoroseobacter shibae* DFL 12; *Gluconacetobacter diazotrophicus* Pal 5 FAPERJ; *Gluconacetobacter diazotrophicus* Pal 5 JGI; *Azospirillum* B510 uid46085; *Rhodospirillum rubrum* ATCC 11170; *Diaphorobacter* TPSY uid29975; *Verminephrobacter eiseniae* EF01-2; *Neisseria meningitidis* 053442; *Neisseria meningitidis* alpha14; *Neisseria meningitidis* Z2491; *Desulfobivrio salexigens* DSM 2638; *Campylobacter jejuni* doilei 269 97; *Campylobacter jejuni* 81116; *Campylobacter jejuni*; *Campylobacter lari* RM2100; *Helicobacter hepaticus*; *Wolinella succinogenes*; *Tolomonas auensis* DSM 9187; *Pseudoalteromonas atlantica* T6c; *Shewanella pealeana* ATCC 700345; *Legionella pneumophila* Paris; *Actinobacillus succinogenes* 130Z; *Pasteurella multocida*; *Francisella tularensis* novicida U112; *Francisella tularensis* holarctica; *Francisella tularensis* FSC 198; *Francisella tularensis* tularensis; *Francisella tularensis* WY96-3418; y *Treponema denticola* ATCC 35405. Por consiguiente, aspectos de la presente divulgación se refieren a una proteína Cas9 presente en un sistema de CRISPR de tipo II.

Un experto en la materia puede hacer referencia a una proteína Cas9 en la bibliografía como CsnI. La proteína Cas9 de *S. pyogenes* se muestra a continuación. Véase Deltcheva y otros., Nature 471, 602-607 (2011).

MDKKYSIGLDIGTNSVGVAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAE
 ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFG
 NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPNDS

VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGN
 LIALSLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
 AILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
 VVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
 SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
 RLSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDSSLTFKEDIQKAQVSGQGDSL
 HEHIANLAGSPAIKKILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRER
 MKRIEIGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDH
 IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
 KLVSDFRKDFQFYKREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVR
 K
 MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
 ATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
 YSVLVVAKVEKGKSKLKSVELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIILKPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLLKGSPEDEQKQLFVE
 QHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENIHLFTLTNLGA
 PAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD-

5 Según un aspecto, la proteína de unión a ADN guiada por ARN incluye homólogos y ortólogos de Cas9 que conservan la capacidad de la proteína para unirse al ADN, ser guiada por el ARN y cortar el ADN. Según un aspecto, la proteína Cas9 incluye la secuencia tal como se expone para Cas9 que se produce de manera natural a partir de *S. pyogenes* y secuencias proteicas que tienen, como mínimo, el 30%, el 40%, el 50%, el 60%, el 70%, el 80%, el 90%, el 95%, el 98% o el 99% de homología con la misma y que son una proteína de unión a ADN, tal como una proteína de unión a ADN guiada por ARN.

10 Según un aspecto, se proporciona un sistema de Cas9-ARNg modificado por ingeniería genética que permite el corte del genoma guiado por ARN de manera específica de sitio en una célula madre, si se desea, y la modificación del genoma de la célula madre mediante inserción de ácidos nucleicos donadores exógenos. Los ARN de guía son complementarios a sitios diana o loci diana en el ADN. Los ARN de guía pueden ser quimeras de ARNcr-ARNtracr.
 15 Los ARN de guía pueden introducirse a partir del medio circundante de la célula. De esta manera se proporciona un procedimiento de modificación continua de una célula hasta el grado de que se proporcionan diversos ARN de guía al medio circundante y con la captación por la célula de los ARN de guía y con complementación del medio con ARN de guía adicionales. La complementación puede ser de una manera continua. La Cas9 se une en o cerca del ADN genómico diana. El uno o más ARN de guía se unen en o cerca del ADN genómico diana. La Cas9 corta el ADN genómico diana y se inserta ADN donador exógeno en el ADN en el sitio de corte.

20 Por consiguiente, los procedimientos se refieren a la utilización de un ARN de guía con una proteína Cas9 y un ácido nucleico donador exógeno para inserciones múltiplex de ácidos nucleicos donadores exógenos en ADN dentro de una célula madre que expresa Cas9 ciclando la inserción de ácido nucleico que codifica el ARN (o proporcionando ARN del medio circundante) y ácido nucleico donador exógeno, expresando el ARN (o captando el ARN),
 25 localizando conjuntamente el ARN, Cas9 y ADN de una manera para cortar el ADN, e insertando el ácido nucleico donador exógeno. Las etapas del procedimiento pueden ciclarse en cualquier número deseado para dar como resultado cualquier número deseado de modificaciones del ADN. Los procedimientos de la presente divulgación se refieren por consiguiente a editar genes diana utilizando las proteínas Cas9 y ARN de guía descritos en el presente documento para proporcionar modificación por ingeniería genética y epigenética múltiplex de células madre.

Aspectos adicionales de la presente divulgación se refieren a la utilización de sistemas o proteínas de unión a ADN (tales como las TALEN o Cas9 modificadas descritas en el presente documento) en general para la inserción múltiplex de ácidos nucleicos donadores exógenos en el ADN, tal como ADN genómico, de una célula madre, tal como una célula madre humana. Un experto en la materia identificará fácilmente sistemas de unión a ADN a modo de ejemplo basándose en la presente divulgación.

Las células, según la presente divulgación, a menos que se especifique otra cosa, incluyen cualquier célula en la que puedan introducirse y expresarse ácidos nucleicos foráneos, tal como se describe en el presente documento. Ha de entenderse que los conceptos básicos de la presente divulgación descritos en el presente documento no están limitados por el tipo de célula. Las células, según la presente divulgación, incluyen células somáticas, células madre, células eucariotas, células procariotas, células animales, células vegetales, células fúngicas, células de arqueas, células de eubacterias y similares. Las células incluyen células eucariotas tales como células de levaduras, células vegetales y células animales. Las células particulares incluyen células de mamífero, tales como células humanas. Además, las células incluyen cualquiera en la que sería beneficioso o deseable modificar el ADN.

Los ácidos nucleicos diana incluyen cualquier secuencia de ácido nucleico para la que una TALEN o proteína de unión a ADN guiada por ARN que tiene actividad nucleasa, tal como se describe en el presente documento, puede ser útil para mellar o cortar. Los ácidos nucleicos diana incluyen cualquier secuencia de ácido nucleico para la que un complejo de colocalización, tal como se describe en el presente documento, puede ser útil para mellar o cortar. Los ácidos nucleicos diana incluyen genes. Para los propósitos de la presente divulgación, el ADN, tal como ADN bicatenario, puede incluir el ácido nucleico diana y un complejo de colocalización que puede unirse a o localizarse conjuntamente de otro modo con el ADN o una TALEN que puede unirse de otro modo con el ADN en o adyacente o cerca del ácido nucleico diana y de una manera en la que el complejo de colocalización o la TALEN pueden tener un efecto deseado sobre el ácido nucleico diana. Tales ácidos nucleicos diana pueden incluir ácidos nucleicos endógenos (o que se producen de manera natural) y ácidos nucleicos exógenos (o foráneos). Un experto, basándose en la presente divulgación, podrá identificar o diseñar fácilmente ARN de guía y proteínas Cas9 que se localizan conjuntamente en un ADN o una TALEN que se une a un ADN, incluyendo un ácido nucleico diana. Un experto podrá, además, identificar dominios o proteínas reguladoras de la transcripción, tales como activadores de la transcripción o represores de la transcripción, que se localizan conjuntamente asimismo en un ADN que incluye un ácido nucleico diana. El ADN incluye ADN genómico, ADN mitocondrial, ADN viral o ADN exógeno. Según un aspecto, los materiales y procedimientos útiles en la puesta en práctica de la presente divulgación incluyen los descritos en Di Carlo, y otros., *Nucleic Acids Research*, 2013, vol. 41, n.º 7 4336-4343 incluyendo cepas y medios a modo de ejemplo, construcción de plásmidos, transformación de plásmidos, electroporación de casete de ARNq transitorio y ácidos nucleicos donadores, transformación de plásmido de ARNq con ADN donador en células que expresan Cas9, inducción con galactosa de Cas9, identificación de dianas de CRISPR-Cas en el genoma de levadura, etc. Se proporcionan referencias adicionales incluyendo información, materiales y procedimientos útiles para que un experto lleve a cabo la invención en Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. y Church, G.M. (2013) RNA-Guided human genome engineering via Cas9. *Science*, 10,1126fscience.1232033; Storici, F., Durham, C.L., Gordenin, D.A. y Resnick, M.A. (2003) Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *PNAS*, 100, 14994-14999 y Jinek, M., Chilinski, K., Fonfara, I., Hauer, M., Doudna, J.A. y Charpentier, E. (2012) A programmable dual-RNA-Guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-821.

Pueden introducirse ácidos nucleicos foráneos (es decir, aquellos que no son parte de la composición de ácidos nucleicos naturales de una célula) en una célula utilizando cualquier procedimiento conocido por los expertos en la materia para tal introducción. Tales procedimientos incluyen transfección, transducción, transducción viral, microinyección, lipofección, nucleofección, bombardeo con nanopartículas, transformación, conjugación y similares. Un experto en la materia entenderá y adaptará fácilmente tales procedimientos utilizando fuentes bibliográficas fácilmente identificables.

Los ácidos nucleicos donadores incluyen cualquier ácido nucleico que va a insertarse en una secuencia de ácido nucleico tal como se describe en el presente documento.

Los siguientes ejemplos se exponen como representativos de la presente divulgación. Estos ejemplos no han de interpretarse como limitativos del alcance de la presente divulgación ya que estas y otras realizaciones equivalentes resultarán evidentes en vista de la presente divulgación, figuras y reivindicaciones adjuntas.

EJEMPLO I

Ensamblaje de ARN de guía

Se incorporaron 19 pb de la secuencia diana seleccionada (es decir, 5'-N19 de 5'-N19-NGG-3') en dos oligonucleótidos de 100 meros complementarios (TTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGN19GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAG GCTAGTCC). Se suspendió cada oligonucleótido de 100 meros a 100 mM en agua, se mezcló con un volumen igual

y se apareó en una máquina de termociclado (95°C, 5 min; elevación hasta 4°C, 0,1°C/s). Para preparar el vector de destinación, se linealizó el vector de clonación de ARNg (plásmido de Addgene ID 41824) utilizando AfIII y se purificó el vector. Se llevó a cabo la reacción de ensamblaje de ARNg (10 ul) con 10 ng de fragmento de 100 pb apareado, 100 ng de estructura principal de destinación, mezcla de reacción de ensamblaje de Gibson 1X (New England Biolabs) a 50°C durante 30 min. La reacción puede procesarse directamente para transformación bacteriana para colonizar ensamblajes individuales.

EJEMPLO II

10 Diseño y ensamblaje de TALE recodificados

Se optimizaron re-TALE a diferentes niveles para facilitar el ensamblaje, y mejorar la expresión. En primer lugar se optimizaron conjuntamente secuencias de ADN de re-TALE para una utilización de codones humanos, y baja energía de plegamiento de ARNm en el extremo 5' (GeneGA, Bioconductor). La secuencia obtenida se hizo evolucionar a través de varios ciclos para eliminar repeticiones (directas o invertidas) mayores de 11 pb (véase la figura 12). En cada ciclo, se evalúan secuencias sinónimas para cada repetición. Se seleccionan aquellas con la mayor distancia de Hamming con respecto al ADN en evolución. La secuencia de un re-TALE que presenta 16,5 monómeros es tal como sigue

CTAACCCCTGAACAGGTAGTCGCTATAGCTTCAAATATCGGGGGCAAGCAAGCACTTG
 AGACCGTTCAACGACTCCTGCCAGTGCTCTGCCAAGCCATGGATTGACTCCGGAGCA
 AGTCGTCGCGATCGCGAGCAACGGCGGGGGGAAGCAGGCGCTGGAAACTGTTTCAGAG
 ACTGCTGCCTGTACTTTGTCAGGCGCATGGTCTCACCCCCGAACAGGTTGTCGCAATA
 GCAAGTAATATAGGCGGTAAGCAAGCCCTAGAGACTGTGCAACGCCTGCTCCCCGTGC
 TGTGTCAGGCTCACGGTCTGACACCTGAACAAGTTGTGCGGATAGCCAGTCACGACGG
 GGGAAAACAAGCTCTAGAAACGGTTCAAAGGTTGTTGCCCGTTTCTGTGCCAAGCACAT
 GGGTTAACACCCGAACAAGTAGTAGCGATAGCGTCAAATAACGGGGGTAAACAGGCT
 TTGGAGACGGTACAGCGGTTATTGCCGGTCCTCTGCCAGGCCACGGACTTACGCCAG
 AACAGGTGGTTGCAATTGCCTCCAACATCGGCGGGAAACAAGCGTTGGAAACTGTGC
 AGAGACTCCTTCCTGTTTTGTGTCAAGCCCACGGCTTGACGCCTGAGCAGGTTGTGGC
 20 CATCGCTAGCCACGACGGAGGGAAGCAGGCTCTTGAAACCGTACAGCGACTTCTCCCA

GTTTTGTGCCAAGCTCACGGGCTAACCCCCGAGCAAGTAGTTGCCATAGCAAGCAACG
 GAGGAGGAAAACAGGCATTAGAAACAGTTCAGCGCTTGCTCCCGGTA CTCTGTCAGG
 CACACGGTCTAACTCCGGAACAGGTCGTAGCCATTGCTTCCCATGATGGCGGCAAACA
 GGCGCTAGAGACAGTCCAGAGGCTCTTGCCCTGTGTTATGCCAGGCACATGGCCTCACC
 CCGGAGCAGGTCGTTGCCATCGCCAGTAATATCGGCGGAAAGCAAGCTCTCGAAACA
 GTACAACGGCTGTTGCCAGTCCTATGTCAAGCTCATGGACTGACGCCCAGCAGGTAG
 TGGCAATCGCATCTCACGATGGAGGTAAACAAGCACTCGAGACTGTCCAAAGATTGTT
 ACCCGTACTATGCCAAGCGCATGGTTTAAACCCAGAGCAAGTTGTGGCTATTGCATCT
 AACGGCGGTGGCAAACAAGCCTTGGAGACAGTGCAACGATTACTGCCTGTCTTATGTC
 AGGCCCATGGCCTTACTCCTGAGCAAGTCGTAGCTATCGCCAGCAACATAGGTGGGAA
 ACAGGCCCTGGAAACCGTACAACGTCTCCTCCCAGTACTTTGTCAAGCACACGGGTTG
 ACACCGGAACAAGTGGTGGCGATTGCGTCCAACGGCGGAGGCAAGCAGGCACTGGAG
 ACCGTCCAACGGCTTCTTCCGGTTCTTTGCCAGGCTCATGGGCTCACGCCAGAGCAGG
 TGGTAGCAATAGCGTCGAACATCGGTGGTAAGCAAGCGCTTGAAACGGTCCAGCGTCT
 TCTGCCGGTGTTGTGCCAGGCGCACGGACTCACACCAGAACAAGTGGTTGCTATTGCT
 AGTAACAACGGTGGAAAGCAGGCCCTCGAGACGGTGCAGAGGTTACTTCCCGTCTCT
 GTCAAGCGCACGGCCTCACTCCAGAGCAAGTGGTTGCGATCGCTTCAAACAATGGTGG
 AAGACCTGCCCTGGAA

5 Según determinados aspectos, pueden utilizarse TALE que tienen, como mínimo, el 80% de identidad de secuencia, como mínimo, el 85% de identidad de secuencia, como mínimo, el 90% de identidad de secuencia, como mínimo, el 95% de identidad de secuencia, como mínimo, el 98% de identidad de secuencia o, como mínimo, el 99% de identidad de secuencia con respecto a la secuencia anterior. Un experto entenderá fácilmente dónde puede variar la secuencia anterior mientras que mantiene todavía la actividad de unión a ADN del TALE.

10 Se generaron bloques de dímeros de re-TALE que codificaban dos RVD (véase la figura 6A) mediante dos rondas de PCR en condiciones de PCR de Kapa HIFI (KPAP) convencionales, en las que la primera ronda de PCR introdujo la secuencia codificante de RVD y la segunda ronda de PCR generó los bloques de dímero completos con solapamientos de 36 pb con los bloques adyacentes. Se purificaron productos de PCR utilizando el kit de purificación de PCR QIAquick 96 (QIAGEN) y se midieron las concentraciones mediante Nano-drop. Las secuencias de cebadores y moldes se enumeran en la tabla 1 y tabla 2 a continuación.

15

Tabla 1. Secuencias de bloques de re-TALE

bloque 0	CGCAATGCGCTCACGGGAGCACCCCTCAACCTAACCCCTGAACAGGTA GTCGCTATAGCTTCANNNNNNGGGGGCAAGCAAGCACTTGAGACCGT TCAACGACTCCTGCCAGTGCTCTGCCAAGCCCATGGATTGACTCCGGA GCAAGTCGTCGCGATCGCGAGCNNNNNNGGGGGGAAGCAGGCGCTGG AACTGTTTCAGAGACTGCTGCCTGTACTTTGTTCAGGCGCATGGTCTC
----------	--

ES 2 715 666 T3

bloque 1	<p>AGACTGCTGCCTGTACTTTGTCAGGCGCATGGTCTCACCCCCGAACAG GTTGTGCGCAATAGCAAGTNNNNNNGGCGGTAAGCAAGCCCTAGAGAC TGTGCAACGCCTGCTCCCCGTGCTGTGTCAGGCTCACGGTCTGACACC TGAACAAGTTGTCGCGATAGCCAGTNNNNNNGGGGAAAACAAGCTC TAGAAACGGTTCAAAGGTTGTTGCCCGTTCTGTGCCAAGCACATGGGT TA</p>
bloque 1'	<p>TGCGCTCACGGGAGCACCCCTCAACCTCACCCCCGAACAGGTTGTGCGC AATAGCAAGTNNNNNNGGCGGTAAGCAAGCCCTAGAGACTGTGCAAC GCCTGCTCCCCGTGCTGTGTCAGGCTCACGGTCTGACACCTGAACAAG TTGTCGCGATAGCCAGTNNNNNNGGGGAAAACAAGCTCTAGAAACG GTTCAAAGGTTGTTGCCCGTTCTGTGCCAAGCACATGGGTTA</p>
bloque 2	<p>AGGTTGTTGCCCGTTCTGTGCCAAGCACATGGGTAAACACCCGAACAA GTAGTAGCGATAGCGTCANNNNNNGGGGTAAACAGGCTTTGGAGAC GGTACAGCGGTTATTGCCGCTCCTCTGCCAGGCCACGGACTTACGCC AGAACAGGTGGTTGCAATTGCCTCCNNNNNNGGCGGGAAAACAAGCGT TGAAACTGTGCAGAGACTCCTTCCTGTTTTGTGTCAAGCCCACGGCTT GACGCCT</p>
bloque 3	<p>AGACTCCTTCCTGTTTTGTGTCAAGCCCACGGCTTGACGCCTGAGCAG GTTGTGGCCATCGCTAGCNNNNNNGGAGGGAAGCAGGCTCTTGAAAC CGTACAGCGACTTCTCCAGTTTTGTGCCAAGCTCACGGGCTAACCCC CGAGCAAGTAGTTGCCATAGCAAGCNNNNNNGGAGGAAAACAGGCAT TAGAAACAGTTCAGCGCTTGCTCCCGTACTCTGTGTCAGGCACACGGTC TA</p>
bloque 4	<p>CGCTTGCTCCCGTACTCTGTGTCAGGCACACGGTCTAACTCCGGAACAG GTCGTAGCCATTGCTTCCNNNNNNGGCGGCAAACAGGCGCTAGAGAC CGTCCAGAGGCTCTTGCCGTGTTATGCCAGGCACATGGCCTCACCCC GGAGCAGGTCGTTGCCATCGCCAGTNNNNNNGGCGGAAAGCAAGCTC TCGAAACAGTACAACGGCTGTTGCCAGTCCTATGTCAAGCTCATGGAC TG</p>
bloque 5	<p>CGGCTGTTGCCAGTCCTATGTCAAGCTCATGGACTGACGCCCAGCAG GTAGTGGCAATCGCATCTNNNNNNGGAGGTAACAAGCACTCGAGAC TGTCCAAGATTGTTACCCGTAATAAGCAAGCGCATGGTTTAAACCC AGAGCAAGTTGTGGCTATTGCATCTNNNNNNGGTGGCAAACAAGCCTT GGAGACCGTGCAACGATTACTGCCTGTCTTATGTGTCAGGCCCATGGCCT T</p>

bloque 6	CGATTACTGCCTGTCTTATGTCAGGCCCATGGCCTTACTCCTGAGCAGG TGGTCGCTATCGCCAGCNNNNNNGGGGGCAAGCAAGCAACTGGAAACA GTCCAGCGTTTGTCTCCAGTACTTTGTCAGGCGCATGGATTGACACCG GAACAAGTGGTGGCTATAGCCTCANNNNNNGGAGGAAAGCAGGCGCT GGAAACCGTCCAACGTCTTTTACCGGTGCTTTGCCAGGCGCACGGGCT C
bloque 6'	CGATTACTGCCTGTCTTATGTCAGGCCCATGGCCTTACTCCTGAGCAAG TCGTAGCTATCGCCAGCNNNNNNGGTGGGAAACAGGCCCTGGAAACC GTACAACGTCTCCTCCCAGTACTTTGTCAAGCACACGGGTTGACACCG GAACAAGTGGTGGCGATTGCGTCCNNNNNNGGAGGCAAGCAGGCACT GGAGACCGTCCAACGGCTTCTTCCGGTCTTTGCCAGGCTCATGGGCT C
bloque 7	CGGCTTCTTCCGGTCTTTGCCAGGCTCATGGGCTCACGCCAGAGCAG GTGGTAGCAATAGCGTCGNNNNNNGGTGGTAAGCAAGCGCTTGAAAC GGTCCAGCGTCTTCTGCCGGTGTGTGCCAGGCGCACGGACTCACACC AGAACAAGTGGTTGCTATTGCTAGTNNNNNNGGTGGAAAGCAGGCC TCGAGACGGTGCAGAGGTTACTTCCCGTCCTCTGTCAAGCGCACGGCC TC

Tabla 2. Secuencias de cebadores de bloques de re-TALE

bloque 0-F	CGCAATGCGCTCACGGGAGCACCCCTCAACctAACCCCTGAACAGGT* A*G
bloque 0-R	GAGACCATGCGCCTGACAAAGTACAGGCAGCAGTCTCTGAACAG*T* T
bloque 1'-F	TGGCGCAATGCGCTCACGGGAGCACCCCTCA*A*C
bloque 1-F	AGACTGCTGCCTGTACTTTGTCAGGCGCATGGTCTCACCCCGAACA* G*G
bloque 1-R/bloque 1'-R	TAACCCATGTGCTTGGCACAGAACGGGCAACAACCTTTGAACCG*T*T
bloque 2-F	AGGTTGTTGCCCGTTCTGTGCCAAGCACATGGGTTAACACCCgaac*a*a
bloque 2-R	AGGCGTCAAGCCGTGGGCTTGACACAAAACAGGAAGGAGTCTCTGCA CAG*T*t
bloque 3-F	AGACTCCTTCTGTTTTGTGTCAAGCCCACGGCTTGACGCCTG*A*G
bloque 3-R	TAGACCGTGTGCCTGACAGAGTACCGGGAGCAAGCGCT*G*A
bloque 4-F	CGCTTGCTCCCGTACTCTGTGTCAGGCACACGGTCTAA*C*T
bloque 4-R	CAGTCCATGAGCTTGACATAGGACTGGCAACAGCCGTT*G*T
bloque 5-F	CGGCTGTTGCCAGTCTATGTCAAGCTCATGGACTGA*C*G
bloque 5-R	AAGGCCATGGGCCTGACATAAGACAGGCAGTAATCGTT*G*C

bloque 6-F	CGATTACTGCCTGTCTTATGTCAGGCCCATGGCCTTA*C*T
bloque 6-R	GAGCCCGTGCGCCTGGCAAAGCACCGGTAAGACGTTGGA*C*G
bloque 6'-F	CGATTACTGCCTGTCTTATGTCAGGCCCATGGCCTTACTCCTGAGCAA *G*T
bloque 6'-R	GAGCCCATGAGCCTGGCAAAGAACCGGAAGAAGCCGTT*G*G
bloque 7-F	CGGCTTCTTCCGGTTCTTTGCCAGGCTCATGGGCTCACGCCAGAGCAG G*T*G
bloque 7-R	GAGGCCGTGCGCTTGACAGAGGACGGGAAGTAACCTCT*G*C

Se construyeron vectores de destinación de re-TALE-TF y re-TALEN modificando las estructuras principales de clonación de TALE-TF y TALEN (véase la referencia 24). Las regiones de 0,5 RVD en los vectores se recodificaron y se incorporó el sitio de corte de Sapl en el sitio de clonación de re-TALE designado. Las secuencias de re-TALEN y estructuras principales de re-TALE-TF se proporcionan en la figura 20. Los plásmidos pueden tratarse previamente con Sapl (New England Biolabs) con condiciones recomendadas por el fabricante y purificarse con el kit de purificación de PCR QIAquick (QIAGEN).

Se llevó a cabo una reacción de ensamblaje de TASA en un solo recipiente (10 ul) con 200 ng de cada bloque, 500 ng de estructura principal de destinación, mezcla enzimática de TASA 1X (2 U de Sapl, 100 U de Ampligase (Epicentre), 10 mU de exonucleasa T5 (Epicentre), 2,5 U de ADN polimerasa Phusion (New England Biolabs)) y tampón de reacción de ensamblaje isotérmico 1X tal como se describió anteriormente (véase la referencia 25) (PEG-8000 al 5%, Tris-HCl 100 mM pH 7,5, MgCl₂ 10 mM, DTT 10 mM, 0,2 mM de cada uno de los cuatro dNTP y NAD 1 mM). Se realizaron las incubaciones a 37°C durante 5 min y 50°C durante 30 min. La reacción de ensamblaje de TASA puede procesarse directamente para transformación bacteriana para colonizar ensamblajes individuales. La eficacia de obtención de construcción de longitud completa es de ~20% con este enfoque. Alternativamente, puede lograrse >90% de eficacia mediante un ensamblaje de tres etapas. En primer lugar, se realizan reacciones de ensamblaje de re-TALE de 10 ul con 200 ng de cada bloque, mezcla enzimática de re-TALE 1X (100 U de Ampligase, 12,5 mU de exonucleasa T5, 2,5 U de ADN polimerasa Phusion) y tampón de ensamblaje isotérmico 1X a 50°C durante 30 min, seguido por reacción de PCR Kapa HIFI estandarizada, electroforesis en gel de agarosa y extracción de gel QIAquick (Qiagen) para enriquecer los re-TALE de longitud completa. Entonces pueden mezclarse 200 ng amplicones de re-TALE con 500 ng de estructura principal de destinación pretratada con Sapl, mezcla de ensamblaje de re-TALE 1X y tampón de reacción de ensamblaje isotérmico 1X e incubarse a 50°C durante 30 min. La reacción de ensamblaje final de re-TALE puede procesarse directamente para transformación bacteriana para colonizar ensamblajes individuales. Un experto en la materia podrá seleccionar fácilmente endonucleasas, exonucleasas, polimerasas y ligasas de entre las conocidas para poner en práctica los procedimientos descritos en el presente documento. Por ejemplo, pueden utilizarse endonucleasas de tipo II, tales como: Fok I, Bts I, Ear I, Sap I. Pueden utilizarse exonucleasas titulables, tales como exonucleasa lamda, exonucleasa T5 y exonucleasa III. Pueden utilizarse polimerasas sin inicio en caliente, tales como ADN polimerasa Phusion, ADN polimerasa Taq y ADN polimerasa VentR. Pueden utilizarse ligasas termoestables en esta reacción, tales como Ampligase, ADN ligasa pfu, ADN ligasa Taq. Además, pueden utilizarse diferentes condiciones de reacción para activar tales endonucleasas, exonucleasas, polimerasas y ligasas dependiendo de la especie particular utilizada.

EJEMPLO III

Línea celular y cultivo celular

Se mantuvieron células PGP1 iPS en placas recubiertas con Matrigel (BD Biosciences) en mTeSR1 (Stemcell Technologies). Se realizaron pases de los cultivos cada 5-7 días con TrypLE Express (Invitrogen). Se hicieron crecer células 293T y 293FT y se mantuvieron en medio de Eagle modificado por Dulbecco (DMEM, Invitrogen) con alto contenido en glucosa complementado con suero bovino fetal al 10% (FBS, Invitrogen), penicilina/estreptomicina (pen./estrep., Invitrogen) y aminoácidos no esenciales (NEAA, Invitrogen). Se hicieron crecer células K562 y se mantuvieron en RPMI (Invitrogen) complementado con suero bovino fetal al 10% (FBS, Invitrogen 15%) y penicilina/estreptomicina (pen./strep., Invitrogen). Todas las células se mantuvieron a 37°C y el 5% de CO₂ en un incubador humidificado.

Se estableció una línea de células 293T estable para detectar la eficacia de HDR tal como se describe en la referencia 26. Específicamente, las líneas celulares indicadoras llevan secuencias codificantes de GFP integradas genómicamente interrumpidas por la inserción de un codón de terminación y un fragmento genómico de 68 pb derivado del locus AAVS1.

EJEMPLO IV

Prueba de actividad de re-TALEN

5 Se sembraron células indicadoras 293T a densidades de 2×10^5 células por pocillo en una placa de 24 pocillos y se transfectaron con 1 μg de cada plásmido de re-TALEN y 2 μg de plásmido donador de ADN utilizando Lipofectamine 2000 siguiendo los protocolos del fabricante. Se recogieron las células utilizando TrypLE Express (Invitrogen) ~18 h después de la transfección y se resuspendieron en 200 μl de medio para el análisis por citometría de flujo utilizando un analizador de células LSRFortessa (BD Biosciences). Se analizaron los datos de citometría de flujo utilizando FlowJo (FlowJo). Se analizaron, como mínimo, 25.000 acontecimientos para cada muestra de transfección. Para el experimento de selección como diana del locus AAVS1 endógeno en 293T, los procedimientos de transfección fueron idénticos a los descritos anteriormente y se realizó selección con puromicina con una concentración de fármaco a 3 $\mu\text{g}/\text{ml}$ 1 semana después de la transfección.

15 EJEMPLO V

Evaluación de la generación de lentivirus funcionales

20 Se crearon vectores lentivirales mediante técnicas de PCR y clonación convencionales. Se transfectaron plásmidos lentivirales mediante Lipofectamine 2000 con mezcla de empaquetamiento lentiviral (Invitrogen) en células 293FT (Invitrogen) cultivadas para producir lentivirus. Se recogió el sobrenadante 48 y 72 h tras la transfección, se esterilizó por filtración y se añadieron 100 μl de sobrenadante filtrado a 5×10^5 células 293T recientes con polibreno. Se calculó el título de lentivirus basándose en la siguiente fórmula: título de virus = (porcentaje de células 293T GFP+ * número de células inicial bajo transducción) / (el volumen de sobrenadante de recogida de virus original utilizado en el experimento de transducción). Para someter a prueba la funcionalidad de lentivirus, 3 días tras la transducción, se transfectaron células 293T transducidas con lentivirus con 30 ng de plásmidos que portaban el indicador mCherry y 500 ng de plásmidos pUC19 utilizando Lipofectamine 2000 (Invitrogen). Se analizaron imágenes de células utilizando Axio Observer Z.1 (Zeiss) 18 horas después de la transfección y se recogieron utilizando TrypLE Express (Invitrogen) y se resuspendieron en 200 μl de medio para el análisis por citometría de flujo utilizando un analizador de células LSRFortessa (BD Biosciences). Se analizaron los datos de citometría de flujo utilizando BD FACSDiva (BD Biosciences).

EJEMPLO VI

35 Prueba de eficacia de edición del genoma de Cas9-ARNg y re-TALEN

40 Se cultivaron iPSC PGP1 en inhibidor de Rho cinasa (ROCK) Y-27632 (Calbiochem) 2 h antes de la nucleofección. Se realizaron las transfecciones utilizando el kit P3 Primary Cell 4D-Nucleofector X (Lonza). Específicamente, se recogieron células utilizando TrypLE Express (Invitrogen) y se resuspendieron 2×10^6 células en 20 μl de mezcla de nucleofección que utilizando 16,4 μl de solución de P3 Nucleofector, 3,6 μl de complemento, 1 μg de cada plásmido de re-TALEN o 1 μg de Cas9 y 1 μg de construcción de ARNg, 2 μl de ODNmc 100 μM . Posteriormente, se transfirieron las mezclas a tiras Nucleocuvette de 20 μl y se realizó la nucleofección utilizando el programa CB150. Se sembraron en placa las células en placas recubiertas con Matrigel en medio mTeSR1 complementado con inhibidor de ROCK durante las primeras 24 h. Para el experimento de selección como diana del locus AAVS1 endógeno con donador de ADNbc, se siguió el mismo procedimiento excepto porque se utilizaron 2 μg de ADNbc y se complementó el medio mTeSR1 con puromicina a la concentración de 0,5 $\mu\text{g}/\text{ml}$ 1 semana después de la transfección.

50 La información de reTALEN, ARNg y ODNmc utilizados en este ejemplo se enumera en la tabla 3 y tabla 4 a continuación.

Tabla 3. Información de pares de re-TALEN/Cas9-ARNg que seleccionan como diana CCR5

N.º de sitio de selección como diana	re-TALEN	re-TALEN	Secuencia de selección como diana de re-TALEN-L		Secuencia de selección como diana de re-TALEN-R	Secuencia de selección como diana de ARNg	Posición de inicio de la secuencia de selección como diana de ARNg
	Sitio de selección como diana del par (inicio)	par	sitio de selección como diana (fin)				
	/chr3 :	/chr3:					
1	46409942	46409993	TCCCCACTTTCT TGTGAA	TAACCACTCAG GACAGGG	CACTTTCCTTGTG AATCCTT	46409946	
2	46410227	46410278	TCACACAGCAA GTCAGCA	TAGGGAGCAG GCTCGGA	TGGGCTAGCGG AGCAGGCT	46410264	
3	46411260	46411311	TACCCAGACGA GAAAGCT	TCAGACTGCCA AGCTTGA	ACCCAGACGAG AAAAGCTGA	46411261	
4	46411464	46411515	TCTTGTGGCTC GGGAGTA	TATTGTCAGCA GAGCTGA	AGAGGGCATCTT GTGGCTC	46411456	
5	46411517	46411568	TTGAGATTTTC AGATGTC	TATACAGTCAT ATCAAGC	ATCAAAGTCTCT TGGCGGT	46411538	
6	46411634	46411685	TTCAGATAGAT TATATCT	TGCCAGATACA TAGTGG	GCTTCAGATAGA TTATATC	46411632	
7	46412396	46412447	TTATACTGTCT ATATGAT	TCAGCTCTTCT GGCCAGA	ACGGATGCTCA GCTCTTC	46412437	
8	46412432	46412483	TGGCCAGAAGA GCTGAGA	TTACCGGGGAG AGTTTCT	CCGGGGAGAGT TTCCTTGA	46412461	

(continuación)

N.º de sitio de selección como diana	re-TALEN	re-TALEN	Secuencia de selección como diana de re-TALEN-L		Secuencia de selección como diana de re-TALEN-R	Secuencia de selección como diana de ARNg	Posición de inicio de la secuencia de selección como diana de ARNg
	Sitio de selección como diana del par (inicio)	par	Secuencia de selección como diana de re-TALEN-L				
	/chr3 :	sitio de selección como diana (fin)	ATGAGTC	TAAAGATT			
9	46412750	46412801	TTTGCAGAGAG	TTAGCAGAAGA	GAAAATCTTATCT	46412782	
10	46413152	46413203	TATAAAGACTAA ACTACCC	TCGTCTGCCAC CACAGAT	AATGCATGACAT TCATCTG	46413172	
11	46414305	46414356	TAAAACAGTTT GCATTCA	TATAAAGTCCT AGAATGT	AACAGTTTGCAT TCATGGA	46414308	
12	46414608	46414659	TGGCCATCTCT GACCTGT	TAGTGAGCCCA GAAGGGG	CCAGAAGGGGA CAGTAAGA	46414632	
13	46414768	46414820	TAGGTACCTGG CTGTTCGT	TGACCCGTCCTG GCTTTTA	CTGACAATCGAT AGGTACC	46414757	
14	46415017	46415068	TGTCATGGTCA TCTGCTA	TCGACACCGAA GCAGAGT	ACACCGAAGCA GAGTTTTT	46415046	
15	46420034	46420084	TGCCCCCGCGA GGCCACA	TCTGGAAAGTTG AACACCC	GGAAAGTTGAAC ACCCCTTGC	46420062	

Tabla 4. Diseño de ODNmc para estudiar la edición del genoma mediada por ODNmc

Figura 3b	Distancia entre la mutación secundaria y DSB	90-*1	CTACTGTCATTTCAGGGCAATACCCAGACGAGAAAAGCTGAGGGTAT AACAGGTTTCAAGCTTGGCAGTCTGACTACAGAGGCCACTGGCTT
		90-*2	CTACTGTCATTTCAGCCCAATACCCTAACGAGAAAAGCTGAGGGTATA ACAGGTTTCAAGCTTGGCAGTCTGACTACAGAGGCCACTGGCTT
		90-*3	CTACTGTCATTTCAGCCCAATACCCAGACGAGAAAAGCTGAGGGTATA ACAGGTTTCAAGCTTGGCAGTCTGACTACAGAGGCCACTGGCTT
		90M-0	CTACTGTCATTTCAGCCCAATACCCAGACGAGAAAAGCTGAGGGTATA ACAGGTTTCAAGCTTGGCAGTCTGACTACAGAGGCCACTGGCTT
		90-*4	CTACTGTCATTTCAGCCCAATACCCAGACGAGAAAAGCTGAGGGTATA ACAGGTTTGTAGCTTGGCAGTCTGACTACAGAGGCCACTGGCTT
		90-*5	CTACTGTCATTTCAGCCCAATACCCAGACGAGAAAAGCTGAGGGTATA ACAGGTTTCAAGCTTGGCTCTCTGACTACAGAGGCCACTGGCTT
		90-*6	CTACTGTCATTTCAGCCCAATACCCAGACGAGAAAAGCTGAGGGTATA ACAGGTTTCAAGCTTGGCAGTCTGACTAGTGAGGCCACTGGCTT
Figura 3c	Distancia entre ODNmc y la DSB	L670bp_90M	CACTTTATATTTCCCTGCTTAAACAGTCCCCGAGGGTGGGTGCGG AAAAGGCTCTACACTTGTATCATTCCCTCTCCACCACAGGCAT
		L570bp_90M	TTTGTATTTGGGTTTTTTTAAAACCTCCACTCTACAGTTAAGAATTC TAAGGCACAGAGCTTCAATAATTTGGTCAGAGCCAAGTAGCAG
		L480bp_90M	GGAGGTAAACCCAGCAGCATGACTGCAGTTCTTAATCAATGCCCC TTGAATTGCACATATGGGATGAACTAGAACATTTTCTCGATGAT
		L394bp_90M	CTCGATGATTGCTGCTTGTATGATTATGTTACTGAGCTCTACT GTAGCACAGACATATGTCCCTATATGGGGCGGGGTGGGGGTG
		L290bp_90M	GGTGTCTTGATCGCTGGGCTATTTCTATACTGTTCTGGCTTTTCGGA AGCAGTCATTTCTTCTATTCTCCAAGCACCAGCAATTAGCTT
		L200bp_90M	GCTTCTAGTTTGCTGAAACTAATCTGCTATAGACAGAGACTCCGAC GAACCAATTTTATTAGGATTTGATCAAATAAACTCTCTCTGACA
		L114bp_90M	GAAAGAGTAACTAAGAGTTTGATGTTTACTGAGTGCATAGTATGCA CTAGATGCTGGCCGTGGATGCCTCATAGAATCCTCCCAACAACCT
		L45bp_90M	GCTAGATGCTGGCCGTGGATGCCTCATAGAATCCTCCCAACAACCG ATGAAATGACTACTGTCATTCAGCCCAATACCCAGACGAGAAAAG
		R40bp_90M	ACAGGTTTCAAGCTTGGCAGTCTGACTACAGAGGCCACTGGCTT CCCCTGGGTTAGTCTGCCTCTGTAGGATTGGGGGCACGTAATTT
		R100bp_90M	TTAGTCTGCCTCTGTAGGATTGGGGGCACGTAATTTGCTGTTTAAG GTCTCATTTGCCTTCTTAGAGATCACAAGCCAAAGCTTTTAT
		R200bp_90M	GGAAGCCCAGAGGGCATCTTGTGGCTCGGGAGTAGCTCTCTGCTAC CTTCTCAGCTCTGCTGACAATACTTGAGATTTTCAGATGTCACC

		R261bp_90M	TCAGCTCTGCTGACAATACTTGAGATTTTCAGATGTCACCAACCAG CAAGAGAGCTTGATATGACTGTATATAGTATAGTCATAAAGAAC
		R322bp_90M	CATAAAGAACCTGAACTTGACCATATACTTATGTCATGTGAAAATC TTCTCATAGCTTCAGATAGATTATATCTGGAGTGAAGAATCCTG
		R375M_90M	GTGGAAAATTTTCATAGCTTCAGATAGATTATATCTGGAGTGAGC AATCCTGCCACCTATGTATCTGGCATAAGTGTGAGTCCTCATAAA
		R448bp_90M	GGTTTGAAGGGCAACAAAATAGTGAACAGAGTGAAAATCCCCACC TAGATCCTGGGTCCAGAAAAAGATGGGAAACCTGTTAGCTCACC
		Complemento-30 meros	GGCCACT AGGGACAAAA TTGGTGAcagaaa
Figura 3d	Longitud de ODNmc y orientación para la selección como diana de Cas 9-ARNg	Complemento-50 meros	CCCACAGTGGGGCCACTAGGGACAAAATTGGTGAcagaaaagc cccatcc
		Complemento-70 meros	TCCCCTCCACCCACAGTGGGGCCACTAGGGACAAAATTGGTGAcag aaaagcccatccttaggctcc
		Complemento-90 meros	cttTTATCTGTCCCCTCCACCCACAGTGGGGCCACTAGGGACAAAAT TGGTGAcagaaaagcccatccttaggctcctcctcctag
		Complemento-110 meros	gtctgggtactTTATCTGTCCCCTCCACCCACAGTGGGGCCACTAGGGA CAAAATTGGTGAcagaaaagcccatccttaggctcctcctcctagctcctgata
		Sin complemento-30 meros	TTTCTGTCACCAATGGTGTCCCTAGTGGCC
		Sin complemento-50 meros	GGATGGGGCTTTTCTGTCACCAATGGTGTCCCTAGTGGCCCCACTG TGGG
		Sin complemento-70 meros	GGAGGCCTAAGGATGGGGCTTTTCTGTCACCAATGGTGTCCCTAGT GGCCCCACTGTGGGGTGGAGGGGA
		Sin complemento-90mer	CTAGGAAGGAGGAGGCCTAAGGATGGGGCTTTTCTGTCACCAATG GTGTCCCTAGTGGCCCCACTGTGGGGTGGAGGGGACAGATAAAAAG
		Sin complemento-110 meros	TATCAGGAGACTAGGAAGGAGGAGGCCTAAGGATGGGGCTTTTCT GTCACCAATGGTGTCCCTAGTGGCCCCACTGTGGGGTGGAGGGGAC AGATAAAAAGTACCCAGAAC
Figura 2c	ODNmc on	Cas9-ARNg	TTCTAGTAACCACTCAGGACAGGGGGTTAGCCCAAAAATTACACA AGAAAGTGGGGACCCATGGGAAAT
	o para selección como diana de	A-CCR5-1	
	Cas 9-ARNg-	Cas9-ARNg-CCR5-2	CAGCAAGTCAGCAGCACAGCGTGTGTGACTCCGAGGGTGCTCCGCT AGCCCACATTGCCCTCTGGGGGTG
	CCR5	Cas9-ARNg-CCR5-3	GTCAGACTGCCAAGCTTGAAACCTGTCTTACCCTCTACTTTCTCGTC TGGGTATTGGGCTGAATGACAGT
		Cas9-ARNg-CCR5-4	CAGAGCTGAGAAGACAGCAGAGAGCTACTCCCGAAGCACAAGATC CCCTCTGGGCTTCCGTGACCTTGGC

ES 2 715 666 T3

		Cas9-ARNg-CCR5-5	CTGACAATACTTGAGATTTTCAGATGTCACCAACGACCAAGAGAGC TTGATATGACTGTATATAGTATAG
		Cas9-ARNg-CCR5-6	CAGATACATAGGTGGCAGGATTCTTCACTCCAGACTTAATCTATCT GAAGCTATGAGAAAATTTCCACAT
		Cas9-ARNg-CCR 5-7	TATATGATTGATTTGCACAGCTCATCTGGCCAGATAAGCTGAGACA TCCGTTCCCCTACAAGAACTCTC
		Cas9-ARNg-CCR5-8	ATCTGGCCAGAAGAGCTGAGACATCCGTTCCCCTTGAAGAACTCT CCCCGGTAAGTAACCTCTCAGCTG
		Cas9-ARNg-CCR 5-9	AGGCATCTCACTGGAGAGGGTTTAGTTCTCCTTAAGAGAAGATAAG ATTTCAAGAGGGAAGCTAAGACTC
		Cas9-ARNg-CCR5-10	ATAATATAATAAAAAATGTTTCGTCTGCCACCACTAATGAATGTCA TGCATTCTGGGTAGTTTAGTCTTA
		Cas9-ARNg-CCR5-11	TTTATAAAGTCTTAGAATGTATTTAGTTGCCCTCGTTGAATGCAAAC TGTTTTATACATCAATAGTTTT
		Cas9-ARNg-CCR5-12	GCTCAACCTGGCCATCTCTGACCTGTTTTTCCTTCCCCTGTCCT TCTGGGCTCACTATGCTGCCGCC
		Cas9-ARNg-CCR5-13	TTTTAAAGCAAAACAGCATGGACGACAGCCAGGCTCCTATCGATT
			GTCAGGAGGATGATGAAGAAGATT
		Cas9-ARNg-CCR5-14	GCTTGTTCATGGTCATCTGCTACTCGGGAATCCTAATTACTCTGCTTC GGTGTGCGAAATGAGAAGAAGAGG
		Cas9-ARNg-CCR5-15	ATACTGCCCCCGCAGGCCACATTGGCAAACCAGCTTGGGTGTTCA ACTTCCAGACTTGGCCATGGAGAA
	Donador de ODNmc para reTALEN que seleccionan como diana CCR5	reTALEN-CCR5-1	CTGAAGAATTTCCCATGGGTCCCCTTTCTTGTGAATCCTTGGAGT GAACCCCCCTGTCTGAGTGGTTACTAGAACACACCTCTGGAC
		reTALEN-CCR5-2	TGGAAGTATCTTGGCAGGTCACACAGCAAGTCAGCAGCACAGCC AGTGTGACTCCGAGCCTGCTCCGCTAGCCACATTGCCCTCTGGG
		reTALEN-CCR5-3	CTACTGTCATTTCAGCCCAATACCCAGACGAGAAAGCTGAGGGTATA ACAGGTTTCAAGCTTGGCAGTCTGACTACAGAGGCCACTGGCTT
		reTALEN-CCR5-4	GGAAGCCCAGAGGGCATCTTGTGGCTCGGGAGTAGCTCTCTGCTAC CTTCTCAGCTCTGCTGACAATACTTGAGATTTTCAGATGTCACC
		reTALEN-CCR5-5	TCAGCTCTGCTGACAATACTTGAGATTTTCAGATGTCACCAACGCC CAAGAGAGCTTGATATGACTGTATATAGTATAGTCATAAAGAAC
		reTALEN-CCR5-6	GTGGAAAATTTTCATAGCTTCAGATAGATTATATCTGGAGTGAGC AATCCTGCCACCTATGTATCTGGCATAAGTGTGAGTCTCATAAA
		reTALEN-CCR5-7	GAAACAGCATTTCCTACTTTTATACTGTCTATATGATTGATTTGGTC AGCTCATCTGGCCAGAAGAGCTGAGACATCCGTTCCCCTACAA
		reTALEN-CCR5-8	TTGATTTGCACAGCTCATCTGGCCAGAAGAGCTGAGACATCCGTAT CCCTACAAGAACTCTCCCCGGTAAGTAACCTCTCAGCTGCTTG
		reTALEN-CCR5-9	GGAGAGGGTTTAGTTCTCCTTAGCAGAAGATAAGATTTCAAGATGA GAGCTAAGACTCATCTCTGCAAATCTTTCTTTTGGAGAGGTAA

		reTALEN-CCR5-10	TAATATAATAAAAAATGTTTCGTCTGCCACCACAGATGAATGTCGA GCATTCTGGGTAGTTTAGTCTTATAACCAGCTGTCTTGCCCTAGT
		reTALEN-CCR5-11	TTAAAAACCTATTGATGTATAAAACAGTTTGCATTCATGGAGGGTG ACTAAATACATTCTAGGACTTTATAAAAGATCACTTTTTATTTA
		reTALEN-CCR5-12	GACATCTACCTGCTCAACCTGGCCATCTCTGACCTGTTTTTCCTATT TACTGTCCCCTTCTGGGCTCACTATGCTGCCGCCAGTGGGAC
		reTALEN-CCR5-13	TCATCCTCCTGACAATCGATAGGTACCTGGCTGTGCTCCATGCTAC GTTTGCTTTAAAAGCCAGGACGGTCACCTTTGGGGTGGTGACAA
		reTALEN-CCR5-14	GGCTGGTCCTGCCGCTGCTTGTTCATGGTCATCTGCTACTCGGGAGA CCTAAAAACTCTGCTTCGGTGTGCAAAATGAGAAGAAGAGGCACA
		reTALEN-CCR5-15	GGCAAGCCTTGGGTCATACTGCCCCGCGAGGCCACATTGGCAAGT CAGCAAGGGTGTCAACTTCCAGACTTGGCCATGGAGAAGACAT

EJEMPLO VII

Preparación de bibliotecas de amplicones de las regiones de selección como diana

5 Se recogieron células 6 días después de la nucleofección y se añadieron 0,1 µl de enzima proteasa tisular prepGEM (ZyGEM) y 1 µl de tampón prepGEM gold (ZyGEM) a 8,9 µl de las 2-5 x 10⁵ células en el medio. Entonces se añadieron 1 ul de las reacciones a 9 1 de mezcla de PCR que contenía 5 ul de 2X KAPA Hifi Hotstart Readymix (KAPA Biosystems) y 100 nM de los correspondientes pares de cebadores de amplificación. Se incubaron las reacciones a 95°C durante 5 min seguido por 15 ciclos de 98°C, 20 s; 65°C, 20 s y 72°C, 20 s. Para añadir el adaptador de secuencia de Illumina, se añadieron entonces 5 µl de productos de reacción a 20 µl de mezcla de PCR que contenía 12,5 µl de 2X KAPA HIFI Hotstart Readymix (KAPA Biosystems) y cebadores 200 nM que portaban adaptadores de secuencia de Illumina. Se incubaron las reacciones a 95°C durante 5 min seguido por 25 ciclos de 98°C, 20 s; 65°C, 20 s y 72°C, 20 s. Se purificaron los productos de PCR mediante el kit de purificación de PCR QIAquick, se mezclaron a aproximadamente la misma concentración y se secuenciaron con el secuenciador personal MiSeq. Los cebadores de PCR se enumeran en la tabla 5 a continuación.

Tabla 5. Secuencias de cebadores de PCR del sitio de selección como diana de CCR5

N.º de selección como diana en CCR5	Nombre	Secuencia de cebador
1	sitio 1-F1	ACACTCTTTCCTACACGACGCTCTTCCGATCTCGTGATTTT GCAGTGTGCGTTACTCC
	sitio 1-F2	ACACTCTTTCCTACACGACGCTCTTCCGATCTACATCGTTT GCAGTGTGCGTTACTCC
	sitio 1-F3	ACACTCTTTCCTACACGACGCTCTTCCGATCTGCCTAATTT GCAGTGTGCGTTACTCC
	sitio 1-F4	ACACTCTTTCCTACACGACGCTCTTCCGATCTTGGTCATTT GCAGTGTGCGTTACTCC
	sitio 1-R	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTCCAAGCAA CTAAGTCACAGCA
	2	Sitio 2-F1

ES 2 715 666 T3

N.º de selección como diana en CCR5	Nombre	Secuencia de cebador
	Sitio 2-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACATCGAT GAGGAAATGGAAGCTTG
	Sitio 2-F3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCTAAAT GAGGAAATGGAAGCTTG
	Sitio 2-F4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGTCAATG AGGAAATGGAAGCTTG
	Sitio 2-R	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTCATTAGGG TATTGGAGGA
	3	sitio 3-F1
sitio 3-F2		ACACTCTTTCCCTACACGACGCTCTTCCGATCTACATCGAA TCCTCCCAACAACATCAT
sitio 3-F3		ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCTAAAA TCCTCCCAACAACATCAT
sitio 3-F4		ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGTCAAAT CCTCCCAACAACATCAT
sitio 3_R		CTCGGCATTCTGCTGAACCGCTCTTCCGATCTCCAATCCT ACAGAGGCAG
4		sitio 4-F1
	sitio 4-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACATCGAA GCCAAAGCTTTTTATTCT
	sitio 4-F3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCTAAAA GCCAAAGCTTTTTATTCT
	sitio 4-F4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGTCAAA GCCAAAGCTTTTTATTCT
	sitio 4_R	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGCCAAA GCTTTTTATTCT
5	sitio 5-F1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGATATC TTGTGGCTCGGGAGTAG
	sitio 5-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACATCGATC TTGTGGCTCGGGAGTAG
	sitio 5-R	CTCGGCATTCTGCTGAACCGCTCTTCCGATCTTGGCAGGA TTCTTCACTCCA

ES 2 715 666 T3

N.º de selección como diana en CCR5	Nombre	Secuencia de cebador
6	sitio 6-F 1	ACACTCTTTCCTACACGACGCTCTTCCGATCTCGTGATCTA TTTTGTTGCCCTTCAA
	sitio 6-F2	ACACTCTTTCCTACACGACGCTCTTCCGATCTACATCGCTA TTTTGTTGCCCTTCAA
	sitio 6-R	CTCGGCATTCCCTGCTGAACCGCTCTTCCGATCTAACCTGAA CTTGACCATATACT
7	sitio 7-F1	ACACTCTTTCCTACACGACGCTCTTCCGATCTCGTGATCA GCTGAGAGGTTACTTACC
	sitio 7-F2	ACACTCTTTCCTACACGACGCTCTTCCGATCTACATCGCA GCTGAGAGGTTACTTACC
	sitio 7-R	CTCGGCATTCCCTGCTGAACCGCTCTTCCGATCTAATGATTTA ACTCCACCCTC
8	sitio 8-F1	ACACTCTTTCCTACACGACGCTCTTCCGATCTCGTGATACT CCACCCTCCTTCAAAGA
	sitio 8-F2	ACACTCTTTCCTACACGACGCTCTTCCGATCTACATCGACT CCACCCTCCTTCAAAGA
	sitio 8-R	CTCGGCATTCCCTGCTGAACCGCTCTTCCGATCTGGTGTTTG CCAAATGTCT
9	sitio 9_F1	ACACTCTTTCCTACACGACGCTCTTCCGATCTCGTGATGG GCACATATTCAGAAGGCA
	sitio 9_F2	ACACTCTTTCCTACACGACGCTCTTCCGATCTACATCGGG GCACATATTCAGAAGGCA
	sitio 9_R	CTCGGCATTCCCTGCTGAACCGCTCTTCCGATCTAGTGAAAG ACTTTAAAGGGAGCA
10	sitio 10-F1	ACACTCTTTCCTACACGACGCTCTTCCGATCTCGTGATCAC AATTAAGAGTTGTCATA
	sitio 10-F2	ACACTCTTTCCTACACGACGCTCTTCCGATCTACATCGCA CAATTAAGAGTTGTCATA
	sitio 10-R	CTCGGCATTCCCTGCTGAACCGCTCTTCCGATCTCTCAGCTA GAGCAGCTGAAC
11	sitio 11-F1	CTCGGCATTCCCTGCTGAACCGCTCTTCCGATCTGACACTTG ATAATCCATC
	sitio 11-F2	ACACTCTTTCCTACACGACGCTCTTCCGATCTACATCGTCA
		ATGTAGACATCTATGTAG

ES 2 715 666 T3

N.º de selección como diana en CCR5	Nombre	Secuencia de cebador
	sitio 11-R	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGATTCA ATGTAGACATCTATGTAG
12	sitio 12-F1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGATACT GCAAAGGCTGAAGAGC
	sitio 12-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACATCGACT GCAAAGGCTGAAGAGC
	sitio 12-F3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCTAAACT GCAAAGGCTGAAGAGC
	sitio 12-F4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGTCAACT GCAAAGGCTGAAGAGC
	sitio 12-R	CTCGGCATTCCCTGCTGAACCGCTCTTCCGATCTGCCTATAA AATAGAGCCCTGTCAA
13	sitio 13-F1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGATCTC TATTTTATAGGCTTCTTC
	sitio 13-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACATCGCTC TATTTTATAGGCTTCTTC
	sitio 13-R	CTCGGCATTCCCTGCTGAACCGCTCTTCCGATCTAGCCACCA CCCAAGTGATC
14	sitio 14-F1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACATCGTTC CAGACATTAAGATAGTC
	sitio 14-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGATTTC CAGACATTAAGATAGTC
	sitio 14-R	CTCGGCATTCCCTGCTGAACCGCTCTTCCGATCTAATCATGA TGGTGAAGATAAG
15	sitio 15-F1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGATCCG GCAGAGACAAACATTAAA
	sitio 15-F2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGGCAGA GACAAACATTAAA
	sitio 15-R	CTCGGCATTCCCTGCTGAACCGCTCTTCCGATCTAGCTAGGA AGCCATGGCAAG
Adaptador de Illumina	PE-PCR-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACA cgac*g*c
	PE-PCR-R	CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCCCT

N.º de selección como diana en CCR5	Nombre	Secuencia de cebador
		GCTGAACc*g*c
Cebador de PCR de secuenciación múltiplex		
3	sitio 3-M-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTGCATA GTATGTGCTAGATGCTG
	sitio 3-M-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGATCTC TAAGAAGGCAAATGAGAC
Adaptador de Illumina	PCR index	CAAGCAGAAGACGGCATAACGAGATN1N2N3N4N5N6GTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCT
	PCR universal	AATGATACGGCGACCACCGAGATCTACTACTTTCCCTACA CGACGCTCTTCCGATCT
*Los cebadores de PCR index se adquirieron de Epicentre (cebadores de Index PCR ScriptSeq™)		

EJEMPLO VIII

Sistema de evaluación de edición del genoma (GEAS)

5 Se ha utilizado secuenciación de última generación para detectar alteraciones genómicas poco comunes. Véanse las referencias 27-30. Para permitir una utilización amplia de este enfoque para evaluar rápidamente la eficacia de HDR y NHEJ en hiPSCs, se creó un software, denominado “canalización”, para analizar los datos de modificación por ingeniería genética del genoma. Esta canalización se integra en un único módulo de Unix, que utiliza diferentes herramientas, tales como R, BLAT y FASTX Toolkit.

10 Separación con código de barras: se agruparon entre sí grupos de muestras y se secuenciaron utilizando extremos apareados de 150 pb MiSeq (PE150) (Illumina Next Gen Sequencing), y posteriormente se separaron basándose en códigos de barras de ADN utilizando FASTX Toolkit.

15 Filtración de calidad: se cortaron los nucleótidos con menor calidad de secuencia (puntuación de phred <20). Tras cortar, se descartaron lecturas menores de 80 nucleótidos.

20 Mapeo: se utilizó BLAT para mapear las lecturas apareadas independientemente del genoma de referencia y se generaron archivos .psl como salida.

25 Designación de indels: se definieron indels como las lecturas de longitud completa que contenían 2 bloques de coincidencias en la alineación. Sólo se consideraron las lecturas que seguían este patrón en ambos extremos apareados. Como control de calidad, se requería que las lecturas de indels presentaran una coincidencia de 70 nt mínima con el genoma de referencia y que ambos bloques tuvieran, como mínimo, 20 nt de largo. Se calcularon el tamaño y la posición de los indels mediante las posiciones de cada bloque con respecto al genoma de referencia. Se ha estimado la unión de extremos no homólogos (NHEJ) como el porcentaje de lecturas que contenían indels (véase la ecuación 1 a continuación). La mayoría de los acontecimientos de NHEJ se han detectado en las proximidades del sitio de selección como diana.

30 Eficacia de la recombinación dirigida por homología (HDR): se utilizó la coincidencia de patrones (grep) dentro de una ventana de 12 pb centrada sobre la DSB para contar firmas específicas correspondientes a lecturas que contenían la secuencia de referencia, modificaciones de la secuencia de referencia (apareamientos erróneos previstos de 2 pb) y lecturas que contenían sólo una mutación de 1 pb dentro de los apareamientos erróneos previstos de 2 pb (véase la ecuación 1 a continuación).

Ecuación 1. Estimación de NHEJ y HDR

40 A = lecturas idénticas a la referencia: XXXXXABXXXXX
 B = lecturas que contienen un apareamiento erróneo de 2 pb programado mediante ODNmc: XXXXXabXXXXX
 C = lecturas que contienen sólo una mutación de 1 pb en el sitio diana: tales como XXXXXaBXXXXX o

XXXXXAbXXXXX

D = lecturas que contienen indels tal como se describió anteriormente

$$\text{Eficacia de NHEJ} = \left(100 \times \frac{D}{A + B + C + D}\right)\%$$

5

$$\text{Eficacia de HDR} = \left(100 \times \frac{B}{A + B + C + D}\right)\%$$

EJEMPLO IX

10 Examen de genotipo de hiPSC colonizadas

Se pretrataron células iPS humanas en cultivos libres de alimentador con medio mTesr-1 complementado con SMC4 (tizovivina 5 uM, CHIR99021 1 uM, PD0325901 0,4 uM, SB431542 2 uM) (véase la referencia 23) durante, como mínimo, 2 h antes de la clasificación por FACS. Se disociaron los cultivos utilizando Accutase (Millipore) y se resuspendieron en medio mTesr-1 complementado con SMC4 y el tinte de viabilidad ToPro-3 (Invitrogen) a una concentración de $1\text{--}2 \times 10^7$ /ml. Las células hiPS vivas se clasificaron individualmente utilizando un instrumento BD FACSAria II SORP UV (BD Biosciences) con una boquilla de 100 um en condiciones estériles en placas de 96 pocillos recubiertas con fibroblastos embrionarios de ratón CF-1 (Global Stem). Cada pocillo contenía medio celular hES (véase la referencia 31) con 100 ng/ml de factor de crecimiento de fibroblastos básico humano recombinante (bFGF) (Millipore) complementado con SMC4 y 5 ug/ml de fibronectina (Sigma). Tras la clasificación, se centrifugaron las placas a 70 x g durante 3 min. Se observó la formación de colonias 4 días tras la clasificación, y se reemplazó el medio de cultivo por medio celular hES con SMC4. SMC4 puede eliminarse del medio celular hES 8 días tras la clasificación.

Se recogieron unos pocos miles de células 8 días tras la clasificación celular activada por fluorescencia (FACS) y se añadieron 0,1 ul de enzima proteasa celular prepGEM (ZyGEM) y 1 ul de tampón prepGEM gold (ZyGEM) a 8,9 ul de células en el medio. Entonces se añadieron las reacciones a 40 ul de mezcla de PCR que contenía 35,5 ml de Platinum 1,1X Supermix (Invitrogen), 250 nM de cada dNTP y cebadores 400 nM. Se incubaron las reacciones a 95°C durante 3 min seguido por 30 ciclos de 95°C, 20 s; 65°C, 30 s y 72°C, 20 s. Se secuenciaron los productos según Sanger utilizando uno cualquiera de los cebadores de PCR de la tabla 5 y se analizaron las secuencias utilizando DNASTAR (DNASTAR).

EJEMPLO X

35 Inmunotinción y ensayos de teratomas de hiPSC

Se incubaron las células en el medio DMEM/F-12 de desactivación a 37°C durante 60 minutos utilizando los siguientes anticuerpos: Anti-SSEA-4 PE (Millipore) (diluido 1:500); Tra-1-60 (BD Pharmingen) (diluido 1:100). Después de la incubación, se lavaron las células tres veces con DMEM/F-12 de desactivación y se obtuvieron imágenes en el instrumento Axio Observer Z.1 (ZEISS).

Para realizar el análisis de formación de teratomas, se recogieron iPSC humanas utilizando colagenasa tipo IV (Invitrogen) y se resuspendieron las células en 200 ul de Matrigel y se inyectaron por vía intramuscular en las extremidades traseras de ratones deficientes Rag2gamma. Se aislaron los teratomas y se fijaron en formalina entre 4 - 8 semanas después de la inyección. Posteriormente se analizaron los teratomas mediante tinción con hematoxilina y eosina.

EJEMPLO XI

50 Loci genómicos de selección como diana en células somáticas humanas y células madre humanas utilizando reTALEN

Según determinados aspectos, se modifican TALE conocidos por los expertos en la materia o se recodifican para eliminar secuencias repetidas. Tales TALE adecuados para su modificación y utilización en los procedimientos de edición del genoma en vehículos de suministro virales y en diversas líneas celulares y organismos descritos en el presente documento se dan a conocer en las referencias 2, 7-12. Se han desarrollado varias estrategias para ensamblar las secuencias de alineamientos de RVD de TALE repetitivas (véanse las referencias 14 y 32-34). Sin embargo, una vez ensambladas, las repeticiones de secuencias de TALE siguen siendo inestables, lo que limita la amplia utilidad de esta herramienta, especialmente para vehículos de suministro génico virales (véanse las referencias 13 y 35). Por consiguiente, un aspecto de la presente divulgación se refiere a TALE que carecen de

repeticiones, tales como que carecen completamente de repeticiones. Un TALE recodificado de este tipo es ventajoso porque permite una síntesis más rápida y más sencilla de alineamientos de RVD de TALE extendidos.

5 Para eliminar repeticiones, las secuencias de nucleótidos de alineamientos de RVD de TALE se hicieron evolucionar informáticamente para minimizar el número de repeticiones de secuencia al tiempo que se mantenía la composición de aminoácido. TALE recodificados (Re-TALE) que codifican 16 monómeros de reconocimiento de ADN de RVD en tándem, más la mitad de la repetición de RVD final, carecen de cualquier repetición de 12 pb (véase la figura 5a). Notablemente, este nivel de recodificación es suficiente como para permitir la amplificación por PCR de cualquier subsección o monómero específico a partir de una construcción de re-TALE de longitud completa (véase la figura 10 5b). El diseño mejorado de re-TALE puede sintetizarse utilizando tecnología de síntesis de ADN convencional (véase la referencia 36) sin incurrir en los costes adicionales o procedimientos asociados con secuencias con muchas repeticiones. Además, el diseño de secuencias recodificadas permite un ensamblaje eficaz de construcciones de re-TALE utilizando una reacción de ensamblaje isotérmica modificada tal como se describe en los procedimientos en el presente documento y con referencia a la figura 6.

15 Los datos de NGS de edición del genoma se analizaron estadísticamente tal como sigue. Para el análisis de la especificidad de HDR, se utilizó una prueba binomial exacta para calcular las probabilidades de observar diversos números de lecturas de secuencia que contenían el apareamiento erróneo de 2 pb. Basándose en los resultados de secuenciación de ventanas de 10 pb antes y después del sitio de selección como diana, se estimaron las tasas de cambio de bases máximas de las dos ventanas (P1 y P2). Utilizando la hipótesis nula de que los cambios de cada uno de los dos pb diana eran independientes, se calculó la probabilidad esperada de observar apareamiento erróneo de 2 pb en el sitio de selección como diana por casualidad como el producto de estas dos probabilidades (P1*P2). Dado un conjunto de datos que contenía N números de lecturas totales y n números de lecturas de HDR, se calculó el valor de p de la eficacia de HDR observada. Para el análisis de sensibilidad de HDR, los donadores de ADN de ODNmc contenían un apareamiento erróneo de 2 pb contra el genoma de selección como diana, lo que hacía probable la presencia conjunta de los cambios de bases en los dos pb diana si el ODNmc se incorporaba en el genoma de selección como diana. Otros cambios de secuencia observados no previstos no cambiarían probablemente al mismo tiempo. Por consiguiente, los cambios no previstos eran mucho menos interdependientes. Basándose en estas suposiciones, se utilizó la información mutua (MI) para medir la dependencia mutua de cambios de dos pares de bases en los demás pares de posiciones, y se estimó el límite de detección de HDR como la HDR más pequeña en donde la MI del sitio de 2 pb de selección como diana es mayor que la MI de los demás pares de posiciones. Para un experimento dado, se identificaron lecturas de HDR con apareamiento erróneo previsto de 2 pb a partir del archivo fastq original y se simuló un conjunto de archivos fastq con eficacias de HDR diluidas eliminando sistemáticamente diferentes números de lecturas de HDR el conjunto de datos original. Se calculó la información mutua (MI) entre todos los pares de posiciones dentro de una ventana de 20 pb centrada sobre el sitio de selección como diana. En estos cálculos, se calcula la información mutua de la composición de bases entre dos cualesquiera posiciones. A diferencia de la medida de especificidad de HDR descrita anteriormente, esta medida no evalúa la tendencia de pares de posiciones a cambiar a cualquier par particular de bases diana, sólo su tendencia a cambiar al mismo tiempo (véase la figura 8A). La tabla 6 muestra la eficacia de HDR y NHEJ de re-TALEN/ODNmc que seleccionan como diana CCR5 y la eficacia de NHEJ de Cas9-ARNg. Se codificó este análisis en R y se calculó la MI utilizando el paquete infotheo.

Tabla 6

N.º de sitio de selección como diana	Tipo de célula	HDR (reTALEN) (%)	NHEJ (reTALE) (%)	Límite de detección de HDR basándose en análisis de información	NHEJ (Cas9-ARNg)	HDR (Cas9-ARNg)
1	PCP1-iPS	0,06%	0,80%	0,04%	0,58%	0,38%
2	PGP1-iPS	0,48%	0,26%	0,01%	16,02%	3,71%
3	PGP1-iPS	1,71%	0,07%	0,03%	3,44%	3,20%
4	PGP1-iPS	0,02%	1,20%	0,02%*	1,50%	0,14%
5	PGP1-iPS	0,80%	0,04%	0,00%	3,70%	0,39%
6	PGP1-iPS	0,20%	0,73%	0,00%	1,12%	0,49%

N.º de sitio de selección como diana	Tipo de célula	HDR (reTALEN) (%)	NHEJ (reTALE) (%)	Límite de detección de HDR basándose en análisis de información	NHEJ (Cas9-ARNg)	HDR (Cas9-ARNg)
7	PGP1-iPS	0,01%	0,15%	0,01%*	1,98%	1,78%
8	PGP1-iPS	0,03%	0,00%	0,00%	1,85%	0,03%
9	PGP1-iPS	1,60%	0,06%	0,00%	0,50%	0,13%
10	PGP1-iPS	0,68%	1,25%	0,01%	8,77%	1,32%
11	PGP1-iPS	0,06%	0,27%	0,00%	0,62%	0,44%
12	PGP1-iPS	1,60%	0,03%	0,04%	0,18%	0,99%
13	PGP1-iPS	0,00%	1,47%	0,00%	0,65%	0,02%
14	PGP1-iPS	0,47%	0,13%	0,02%	2,50%	0,31%
15	PGP1-iPS	0,8	0,14	0,08%	1,50	1,10%

* El grupo en donde el límite de detección de HDR excede la HDR real detectada

Se abordaron las correlaciones entre eficacia de edición del genoma y estado epigenético tal como sigue. Se calcularon los coeficientes de correlación de Pearson para estudiar posibles asociaciones entre parámetros epigenéticos (ocupación de nucleosomas o ADNasa I HS) y eficacias de modificación por ingeniería genética del genoma (HDR, NHEJ). El conjunto de datos de hipersensibilidad a ADNasa I se descargó del servidor de genoma de UCSC. ADNasa I HS de hiPSC: /gdb/hg19/bbi/wgEncodeOpenChromDnaselpsihi7Sig.bigWig.

Para calcular los valores de P, se comparó la correlación observada con una distribución simulada que se construyó aleatorizando la posición del parámetro epigenético (N=100.000). Correlaciones observadas mayores del percentil 95 o menores del percentil 5 de la distribución simulada se consideraron como posibles asociaciones.

Se determinó la función de reTALEN en comparación con la TALEN no recodificada correspondiente en células humanas. Se utilizó una línea celular HEK 293 que contenía un casete indicador de GFP que portaba una inserción de desplazamiento del marco tal como se describe en la referencia 37. Véase también la figura 1a. El suministro de TALEN o reTALEN que seleccionan como diana la secuencia de inserción, junto con una construcción donadora de GFP sin promotor, conduce a reparación HDR inducida por DSB del casete de GFP, de modo que puede utilizarse la eficacia de reparación de GFP para evaluar la eficacia de corte de nucleasas. Véase la referencia 38. Las reTALEN inducen reparación de GFP en el 1,4% de las células transfectadas, similar a la lograda mediante TALEN (1,2%) (véase la figura 1b). Se sometió a prueba la actividad de reTALEN en el locus AAVS1 en HiPSC PGP1 (véase la figura 1c) y se recuperaron satisfactoriamente clones celulares que contenían inserciones específicas (véase la figura 1d,e), confirmando que las reTALEN son activas tanto en células humanas somáticas como pluripotentes.

La eliminación de repeticiones permitió la generación de lentivirus funcionales con una carga de re-TALE. Específicamente, se empaquetaron partículas lentivirales que codificaban re-TALE-2A-GFP y se sometieron a prueba para detectar la actividad del re-TALE-TF codificado por partículas virales transfectando un indicador de mCherry en un conjunto de células 293T infectadas con lenti-reTALE-2A-GFP. Células 293T transducidas mediante lenti-re-TALE-TF mostraban una activación de la expresión del indicador de 36X en comparación con el indicador sólo negativo (véase la figura 7a,b,c). Se comprobó la integridad de secuencia del re-TALE-TF en las células infectadas con lentivirus y se detectaron reTALE de longitud completa en los 10 clones sometidos a prueba. (Véase la figura 7d).

EJEMPLO XII

Comparación de la eficacia de ReTALE y Cas9-ARNg en hiPSC con el sistema de evaluación de la edición del genoma (GEAS)

Para comparar las eficacias de edición de re-TALEN frente a Cas9-ARNg en hiPSC, se desarrolló una plataforma de secuenciación de última generación (sistema de evaluación de la edición del genoma) para identificar y cuantificar los acontecimientos de edición del genoma de tanto NHEJ como HDR. Se diseñaron y construyeron un par de re-TALEN y un Cas9-ARNg, que seleccionaban como diana ambos la región cadena arriba de CCR5 (re-TALEN, Cas9-ARNg par n.º 3 en la tabla 3), junto con un donador de ODNmc de 90 nt idéntico al sitio diana excepto por un apareamiento erróneo de 2 pb (véase la figura 2a). Las construcciones de nucleasa y ODNmc donador se transflectaron en hiPSC. Para cuantificar la eficacia de edición del genoma, se realizó secuenciación profunda de extremos apareados en la región genómica diana 3 días después de la transfección. Se midió la eficacia de HDR mediante el porcentaje de lecturas que contenían el apareamiento erróneo de 2 pb preciso. Se midió la eficacia de NHEJ mediante el porcentaje de lecturas que portaban indels.

El suministro del ODNmc solo en hiPSC dio como resultado tasas de HDR y NHEJ mínimas, mientras que el suministro de las re-TALEN y el ODNmc condujo a eficacias del 1,7% de HDR y el 1,2% de NHEJ (véase la figura 2b). La introducción del Cas9-ARNg con el ODNmc condujo a eficacias del 1,2% de HDR y el 3,4% de NHEJ. Notablemente, la tasa de deleciones e inserciones genómicas alcanzó un pico en el medio de la región espaciadora entre el sitio de unión de las dos reTALEN, pero alcanzó un pico de 3-4 pb cadena arriba de la secuencia de motivo asociado de protoespaciador (PAM) del sitio de selección como diana de Cas9-ARNg (véase la figura 2b), tal como se esperaría puesto que las roturas bicatenarias tienen lugar en estas regiones. Se observó una mediana de tamaño de deleción genómica de 6 pb y de tamaño de inserción de 3 pb generados por las re-TALEN y se observó una mediana de tamaño de deleción de 7 pb y de inserción de 1 pb mediante el Cas9-ARNg (véase la figura 2b), consecuente con los patrones de lesión del ADN generados habitualmente mediante NHEJ (véase la referencia 4). Varios análisis de la plataforma de secuenciación de última generación revelaron que el GEAS puede detectar tasas de detección de HDR de tan sólo el 0,007%, lo que es tanto altamente reproducible (coeficiente de variación entre réplicas = $\pm 15\%$ * eficacia medida) como 400 veces más sensible que la mayoría de los ensayos de endonucleasas sensibles a apareamientos erróneos comúnmente utilizados (véase la figura 8).

Se construyeron pares de re-TALEN y Cas9-ARNg dirigidos a quince sitios en el locus genómico CCR5 para determinar la eficacia de edición (véase la figura 2c, véase la tabla 3). Se seleccionaron estos sitios para representar un amplio intervalo de sensibilidades a ADNasa I (véase la referencia 39). Se transflectaron las construcciones de nucleasa con los correspondientes donadores de ODNmc (véase la tabla 3) en hiPSC PGP1. Seis días después de la transfección, se obtuvo el perfil de las eficacias de edición del genoma en estos sitios (tabla 6). Para 13 de 15 pares de re-TALEN con donadores de ODNmc, se detectó NHEJ y HDR a niveles por encima de los umbrales de detección estadísticos, con una eficacia de NHEJ promedio del 0,4% y una eficacia de HDR promedio del 0,6% (véase la figura 2c). Además, se encontró una correlación positiva estadísticamente significativa ($r^2 = 0,81$) entre la eficacia de HR y NHEJ en los mismos loci de selección como diana ($P < 1 \times 10^{-4}$) (véase la figura 9a), lo que sugiere que la generación de DSB, la etapa anterior común de tanto HDR como NHEJ, es una etapa limitante de la velocidad para la edición del genoma mediada por reTALEN.

En cambio, los 15 pares de Cas9-ARNg mostraron niveles significativos de NHEJ y HR, con una eficacia de NHEJ promedio del 3% y una eficacia de HDR promedio del 1,0% (véase la figura 2c). Además, se detectó también una correlación positiva entre la eficacia de NHEJ y HDR introducida por Cas9-ARNg (véase la figura 9b) ($r^2 = 0,52$, $p = 0,003$), consecuente con las observaciones para reTALEN. La eficacia de NHEJ lograda mediante Cas9-ARNg fue significativamente mayor que la lograda mediante reTALEN (prueba de la t, extremos apareados, $P = 0,02$). Se observó una correlación moderada pero estadísticamente significativa entre la eficacia de NHEJ y la temperatura de fusión de la secuencia de selección como diana de ARNg (véase la figura 9c) ($r^2 = 0,28$, $p = 0,04$), lo que sugiere que la fuerza del apareamiento de bases entre el ARNg y su diana genómica podría explicar hasta el 28% de la variación en la eficacia de la generación de DSB mediada por Cas9-ARNg. Aun cuando Cas9-ARNg produjo niveles de NHEJ a un promedio de 7 veces mayor que la correspondiente reTALEN, Cas9-ARNg sólo logró niveles de HDR (promedio = 1,0%) similares a los de las correspondientes reTALEN (promedio = 0,6%). Sin querer restringirse a la teoría científica, estos resultados pueden sugerir o bien que la concentración de ODNmc en la DSB es el factor limitante de HDR o que la estructura de rotura genómica creada por el Cas9-ARNg no es favorable para lograr una HDR eficaz. No se observó correlación entre ADNasa I HS y las eficacias de selección como diana del genoma para cualquier procedimiento (véase la figura 10).

EJEMPLO XIII

Optimización del diseño de donadores de ODNmc para HDR

Se diseñaron ODNmc de alto rendimiento en hiPSC tal como sigue. Se diseñó un conjunto de donadores de ODNmc de diferentes longitudes (50-170 nt), que portaban todos el mismo apareamiento erróneo de 2 pb en el medio de la región espaciadora de los sitios diana n.º 3 del par de re-TALEN de CCR5. Se observó que la eficacia de HDR variaba con la longitud de ODNmc, y se observó una eficacia de HDR óptima de ~1,8% con un ODNmc de 90 nt, mientras que ODNmc más largos disminuían la eficacia de HDR (véase la figura 3a). Puesto que regiones de homología más largas mejoran las tasas de HDR cuando se utilizan donadores de ADNbc con nucleasas (véase la referencia 40), los posibles motivos de este resultado pueden ser que se utilizan ODNmc en un proceso de

reparación del genoma alternativo; que ODNmc más largos están menos disponibles para el aparato de reparación del genoma; o que ODNmc más largos pueden incurrir en efectos negativos que compensan cualquier mejora obtenida por una homología más larga, en comparación con donadores de ADNbc (véase la referencia 41). Además, si fuera el caso de cualquiera de los dos primeros motivos, entonces las tasas de NHEJ o bien no deben verse afectadas o bien aumentarían con ODNmc más largos puesto que la reparación de NHEJ no implica el donador de ODNmc. Sin embargo, se observó que las tasas de NHEJ disminuían junto con HDR (véase la figura 3a), lo que sugiere que los ODNmc más largos presentan efectos de compensación. Posibles hipótesis serían que ODNmc más largos son tóxicos para la célula (véase la referencia 42), o que la transfección de ODNmc más largos satura la maquinaria de procesamiento de ADN, provocando de ese modo una captación de ADN molar disminuida, y la reducción de la capacidad de las células para captar o expresar plásmidos de re-TALEN.

Se examinó cómo varía la tasa de incorporación de un apareamiento erróneo portado por el donador de ODNmc con su distancia a la rotura bicatenaria ("DSB"). Se diseñó una serie de ODNmc de 90 nt que presentaban todos el mismo apareamiento erróneo de 2 pb (A) en el centro de la región espaciadora del par de re-TALEN n.º 3. Cada ODNmc contenía también un segundo apareamiento erróneo de 2 pb (B) a distancias variables desde el centro (véase la figura 3b). Se utilizó como control un ODNmc que presentaba solo el apareamiento erróneo de 2 pb central. Cada uno de estos ODNmc se introdujo individualmente con el par de re-TALEN n.º 3 y se analizaron los desenlaces con GEAS. Se encontró que la HDR global, tal como se mide mediante la tasa a la que se incorporaba el apareamiento erróneo A (A sólo o A+B), disminuía a medida que los apareamientos erróneos B se alejaban del centro (véase la figura 3b, véase la figura 11a). La tasa de HDR global superior observada cuando B está sólo 10 pb alejada de A puede reflejar una menor necesidad de apareamiento del ODNmc contra ADN genómico inmediatamente proximal a la rotura del ADNbc.

Para cada distancia de B con respecto a A, una fracción de acontecimientos de HDR incorporaba solo el apareamiento erróneo A, mientras que otra fracción incorporaba ambos apareamientos erróneos A y B (véase la figura 3b (A solo y A+B)). Estos dos desenlaces pueden deberse a tramos de conversión génica (véase la referencia 43) a lo largo de la longitud del oligonucleótido de ADNmc, mediante lo cual la incorporación de los apareamientos erróneos A+B resultaba de tramos de conversión largos que se extendían más allá del apareamiento erróneo B, y la incorporación de solo el apareamiento erróneo A resultaba de tramos más cortos que no alcanzaban B. Con esta interpretación, se estimó una distribución de longitudes de conversión génica en ambas direcciones a lo largo del ODNmc (véase la figura 11b). La distribución estimada implica que los tramos de conversión génica se vuelven progresivamente menos frecuentes a medida que sus longitudes aumentan, un resultado muy similar a las distribuciones de tramos de conversión génica observadas con donadores de ADNbc, pero en una escala de distancia altamente comprimida de decenas de bases para el donador de ADNmc frente a cientos de bases para donadores de ADNbc. Consecuente con este resultado, un experimento con un ODNmc que contenía tres pares de apareamientos erróneos de 2 pb espaciados a intervalos de 10 nt en cualquier lado del apareamiento erróneo de 2 pb "A" central dio lugar a un patrón en el que A solo se incorporaba el 86% de las veces, incorporándose múltiples apareamientos erróneos B otras veces (véase la figura 11c). Aunque los números de acontecimientos de incorporación de B solo eran demasiado bajos como para estimar una distribución de longitudes de tramo menores de 10 pb, queda claro que predomina la región de tramo corto dentro de 10 pb del sitio de nucleasa (véase la figura 11b). Finalmente, en todos los experimentos con apareamientos erróneos B individuales, se observa una pequeña fracción de acontecimientos de incorporación de B sólo (0,04%~0,12%) que es aproximadamente constante a través de todas las distancias de B con respecto a A.

Además, se llevó a cabo un análisis de la lejanía a la que puede colocarse el donador de ODNmc de la rotura del ADNbc inducida por re-TALEN al tiempo que todavía se observa incorporación. Se sometieron a prueba un conjunto de ODNmc de 90 nt con apareamientos erróneos de 2 pb centrales que seleccionaban como diana un intervalo de mayores distancias (de -600 pb a +400 pb) lejos de la rotura del ADNbc inducida por re-TALEN. Cuando los ODNmc se apareaban a una distancia de ≥ 40 pb, se observaron eficacias de HDR $>30x$ menores en comparación con el ODNmc de control situado centralmente sobre la región de corte (véase la figura 3c). El bajo nivel de incorporación que se observó puede deberse a procesos no relacionados con el corte del ADNbc, tal como se observa en experimentos en los que se alteran genomas mediante un donador de ADNmc solo, véase la referencia 42. Mientras tanto, el bajo nivel de HDR presente cuando el ODNmc está a una distancia de ~ 40 pb puede deberse a una combinación de homología debilitada en el lado que contiene apareamiento erróneo del corte del ADNbc junto con una longitud de oligonucleótido de ODNmc insuficiente en el otro lado de la rotura del ADNbc.

Se sometió a prueba el diseño de donadores de ADN de ODNmc para la selección como diana mediada por Cas9-ARNg. Se construyó Cas9-ARNg (C2) que selecciona como diana el locus AAVS1 y se diseñaron donadores de ODNmc de orientaciones (Oc: complementario al ARNg y On: no complementario al ARNg) y longitudes (30, 50, 60, 70, 90, 110 nt) variables. Oc logró una mejor eficacia que On, logrando un Oc de 70 mers una tasa de HDR óptima del 1,5% (véase la figura 3d). Se detectó el mismo sesgo de la hebra de ODNmc utilizando una nickasa derivada de Cas9 (Cc: Cas9_D10A), a pesar del hecho de que las eficacias de HDR mediada por Cc con ODNmc eran significativamente menores que C2 (prueba de la t, extremos apareados, $P=0,02$) (véase la figura 12).

EJEMPLO XIV

Aislamiento clonal de hiPSC de células corregidas

5 El GEAS reveló que el par de re-TALEN n.º 3 logró una edición del genoma precisa con una eficacia de ~1% en hiPSC, un nivel al que pueden aislarse habitualmente células editadas correctamente mediante el examen de clones. Las hiPSC tienen escasa viabilidad como células individuales. Se utilizaron los protocolos optimizados descritos en la referencia 23 junto con un procedimiento de clasificación por FACS de células individuales para establecer una plataforma robusta para el mantenimiento y la clasificación de hiPSC individuales, en donde pueden recuperarse clones de hiPSC con tasas de supervivencia de >25%. Se combinó este procedimiento con un sistema de genotipado rápido y eficaz para realizar extracción de ADN cromosómico y amplificación del genoma seleccionado como diana en reacciones en tubos individuales de 1 hora, permitiendo el genotipado a gran escala de hiPSC editadas. Conjuntamente, estos procedimientos comprenden una canalización para obtener de manera robusta hiPSC con genoma editado sin selección.

15 Para demostrar este sistema (véase la figura 4a), se transfectaron hiPSC PGP1 con un par de re-TALEN y un ODNmc que selecciona como diana CCR5 en el sitio n.º 3 (véase la tabla 3). Se realizó GEAS con una porción de las células transfectadas, hallando una frecuencia de HDR del 1,7% (véase la figura 4b). Esta información, junto con la recuperación del 25% de clones de células individuales clasificadas, permite la estimación de la obtención de, como mínimo, un clon editado correctamente a partir de cinco placas de 96 pocillos con una probabilidad de Poisson del 98% (suponiendo $\mu=0,017 * 0,25 * 96 * 5 * 2$). Seis días tras la transfección, se clasificaron por FACS las hiPSC y ocho días después de la clasificación, se examinaron 100 clones de hiPSC. La secuenciación de Sanger reveló que 2 de 100 de estas colonias de hiPSC no seleccionadas contenían un genotipo heterocigoto que presentaba la mutación de 2 pb introducida mediante el donador de ODNmc observado (figura 4c). La eficacia de selección como diana del 1% ($1\%=2/2*100$, 2 clones corregidos monoalélicos de 100 células examinadas) era consecuente con el análisis de secuenciación de última generación (1,7%) (véase la figura 4b). La pluripotencia de las hiPSC resultantes se confirmó con inmunotinción para SSEA4 y TRA-1-60 (véase la figura 4d). Los clones de hiPSC seleccionados como diana satisfactoriamente podían generar teratomas maduros con características de las tres capas germinales (véase la figura 4e).

EJEMPLO XV

Procedimiento para la edición continua del genoma celular

35 Según determinados aspectos, se proporciona un procedimiento para la edición del genoma en células, incluyendo una célula humana, por ejemplo una célula madre humana, en el que la célula se modifica genéticamente para incluir un ácido nucleico que codifica una enzima que forma un complejo de colocación con ARN complementario al ADN diana y que escinde el ADN diana de manera específica de sitio. Una enzima de este tipo incluye una proteína de unión a ADN guiada por ARN, tal como una proteína de unión a ADN guiada por ARN de un sistema de CRISPR de tipo II. Una enzima a modo de ejemplo es Cas9. Según este aspecto, la célula expresa la enzima y se proporciona ARN de guía a la célula a partir del medio circundante de la célula. El ARN de guía y la enzima forman un complejo de colocación en el ADN diana donde la enzima corta el ADN. Opcionalmente, puede estar presente un ácido nucleico donador para su inserción en el ADN en el sitio de corte, por ejemplo mediante unión de extremos no homólogos o recombinación homóloga. Según un aspecto, el ácido nucleico que codifica una enzima que forma un complejo de colocación con ARN complementario en el ADN diana y que escinde el ADN diana de manera específica de sitio, tal como Cas9, está bajo la influencia de un promotor, de manera que el ácido nucleico puede activarse y silenciarse. Tales promotores los conocen bien los expertos en la materia. Un promotor a modo de ejemplo es el promotor inducible dox. Según un aspecto, la célula se modifica genéticamente al haberse insertado reversiblemente en su genoma el ácido nucleico que codifica una enzima que forma un complejo de colocación con ARN complementario al ADN diana y que escinde el ADN diana de manera específica de sitio. Una vez insertado, el ácido nucleico puede eliminarse mediante la utilización de un reactivo, tal como una transposasa. De esta manera, el ácido nucleico puede eliminarse fácilmente tras su utilización.

55 Según un aspecto, se proporciona un sistema de edición continua del genoma en células madre pluripotentes inducidas humanas (hiPSC) utilizando un sistema de CRISPR. Según un aspecto a modo de ejemplo, el procedimiento incluye la utilización de una línea de hiPSC con Cas9 insertada de manera reversible en el genoma (Cas9-hiPSC); y ARNg que se ha modificado con respecto a su forma nativa para permitir su paso desde el medio circundante de las células al interior de las células para su utilización con la Cas9. Tal ARNg se ha tratado con una fosfatasa de una manera para eliminar grupos fosfato. La edición del genoma se lleva a cabo con Cas9 suplementando ARNg tratado con fosfatasa en el medio de cultivo tisular. Este enfoque permite la edición del genoma sin cicatrices en HiPSC con eficacias de hasta el 50% con días de tratamiento individuales, 2-10 veces más eficaz que las mejores eficacias notificadas hasta la fecha. Además, el procedimiento es fácil de utilizar y con toxicidad celular significativamente inferior. Las realizaciones de la presente divulgación incluyen la edición individual de hiPSC para investigación biológica y aplicaciones terapéuticas, edición múltiples de hiPSC para investigación biológica y aplicaciones terapéuticas, evolución de hiPSC direccional y examen del fenotipo de hiPSC y sus células derivadas.

Según determinados aspectos, pueden utilizarse otras líneas celulares y organismos descritos en el presente documento además de células madre. Por ejemplo, el procedimiento descrito en el presente documento puede utilizarse para células animales tales como células de ratón o rata de modo que pueden generarse células de ratón y células de rata con Cas9 integrada estable y puede realizarse edición del genoma específica de tejido introduciendo localmente ARNg tratado con fosfatasa a partir del medio circundante de las células. Además, pueden insertarse otros derivados de Cas9 en muchas líneas celulares y organismos, y pueden realizarse manipulaciones genómicas seleccionadas como diana, tales como mellado específico de secuencia, activación génica, supresión y modificación epigenética.

Aspectos de la presente divulgación se refieren a preparar hiPSC estables con Cas9 insertada en el genoma. Aspectos de la presente divulgación se refieren a modificar ARN para permitir su entrada en una célula a través de la pared celular y su colocalización con Cas9 al tiempo que se evita la respuesta inmunitaria de la célula. Tal ARN de guía modificado puede lograr eficacias de transfección óptimas con toxicidad mínima. Aspectos de la presente divulgación se refieren a la edición del genoma optimizada en Cas9-hiPSC utilizando ARNg tratado con fosfatasa. Los aspectos de la presente divulgación incluyen eliminar Cas9 de hiPSC para lograr la edición del genoma sin cicatrices, en donde el ácido nucleico que codifica Cas9 se ha colocado reversiblemente en el genoma celular. Los aspectos de la presente divulgación incluyen ingeniería biomédica utilizando hiPSC con Cas9 insertada en el genoma para crear mutaciones genéticas deseadas. Tales hiPSC modificadas por ingeniería genética mantienen la pluripotencia y pueden diferenciarse satisfactoriamente en diversos tipos de células, incluyendo cardiomiocitos, que pueden recapitular completamente el fenotipo de líneas celulares del paciente.

Los aspectos de la presente divulgación incluyen bibliotecas de ARNg tratado con fosfatasa para edición del genoma múltiple. Los aspectos de la presente divulgación incluyen generar una biblioteca de líneas celulares PGP portando, cada una, de 1 a unas cuantas mutaciones designadas en el genoma, que pueden servir como recurso para el examen de fármacos. Los aspectos de la presente divulgación incluyen generar líneas celulares PGP1 con todos los retrotransposones codificados con código de barras con diferentes secuencias para rastrear la ubicación y actividad de este elemento.

EJEMPLO XVI

Generación de hiPSC estables con Cas9 insertada en el genoma

Se codificó Cas9 bajo el promotor inducible dox y se colocó la construcción en un vector Piggybac que puede insertarse en y eliminarse del genoma con la ayuda de la transposasa Piggybac. Una reacción PCR validó la inserción estable del vector (véase la figura 14). Se determinó la expresión de Cas9 inducible por medio de RT-QPCR. El nivel de ARNm de Cas9 aumentó 1.000X tras 8 horas de suplementación con DOX 1 ug/ml en los medios de cultivo y el nivel de ARNm de Cas9 descendió hasta el nivel normal ~ 20 horas tras la retirada del DOX (véase la figura 15).

Según un aspecto, la edición del genoma basada en el sistema de Cas9-hiPSC sorteaba el procedimiento de transfección de ARN/plásmido de Cas9, una construcción grande habitualmente con una eficacia de transfección de < 1% en hiPSC. El presente sistema de Cas9-hiPSC puede servir como plataforma para realizar modificación por ingeniería genómica altamente eficaz en células madre humanas. Además, el casete de Cas9 introducido en las hiPSC utilizando el sistema Piggybac puede eliminarse del genoma fácilmente tras la introducción de transposasas.

EJEMPLO XVII

ARN de guía tratado con fosfatasa

Para permitir la edición continua del genoma en Cas9-hiPSC, se generaron una serie de ARN modificados que codifican ARNg y se suplementaron en medio de cultivo de Cas9-iPS en complejo con liposomas. El ARN nativo tratado con fosfatasa sin ninguna ocupación de extremos logró la eficacia de HDR óptima del 13%, 30X más que el ARN 5'Cap-Mod notificado previamente (véase la figura 16).

Según un aspecto, el ARN de guía se une físicamente al ADN donador. De esta manera, se proporciona un procedimiento de acoplamiento del corte genómico mediado por Cas9 y la HDR mediada por ODNmc, estimulando por tanto la edición genómica específica de secuencia. ARNg unido con donador de ODNmc de ADN con concentración optimizada logró un 44% de HDR y NHEJ inespecífica del 2% (véase la figura 17). Cabe destacar que este procedimiento no incurrió en toxicidad visible tal como se observó con nucleofección o electroporación.

Según un aspecto, la presente divulgación proporciona una estructura de ARN modificada por ingeniería genética *in vitro* que codifica ARNg, que logró una alta eficacia de transfección, eficacia de edición del genoma en colaboración con Cas9 insertada genómicamente. Además, la presente divulgación proporciona una construcción quimérica de ARNg-ADN para acoplar un acontecimiento de corte genómico con la reacción de recombinación dirigida por homología.

EJEMPLO XVIII

5 Eliminación de Cas9 modificada por ingeniería genética de manera reversible de hiPSC para lograr edición del genoma sin cicatrices

Según determinados aspectos, se inserta un casete de Cas9 en el genoma de células hiPSC utilizando un vector reversible. Por consiguiente, se insertó de manera reversible un casete de Cas9 en el genoma de células hiPSC utilizando un vector PiggyBac. El casete de Cas9 se eliminó de las hiPSC con el genoma editado transfectando la
 10 célula con plásmido que codifica transposasa. Por consiguiente, los aspectos de la presente divulgación incluyen la utilización de un vector reversible, que conocen los expertos en la materia. Un vector reversible es uno que puede insertarse en un genoma, por ejemplo, y luego eliminarse con una enzima de eliminación de vector correspondiente. Tales vectores y enzimas de eliminación de vector correspondientes los conocen los expertos en la materia. Se realizó un examen en células iPS colonizadas y se recuperaron colonias que carecían de casete de Cas9 tal como
 15 se confirmó mediante reacción de PCR. Por consiguiente, la presente divulgación da a conocer un procedimiento de edición del genoma sin afectar al resto del genoma al tener un casete de Cas9 permanente presente en la célula.

EJEMPLO XIX

20 Edición del genoma en células iPGP1

La investigación sobre la patogenia de la miocardiopatía se ha visto dificultada históricamente por la falta de sistemas modelo adecuados. La diferenciación de cardiomiocitos de células madre pluripotentes inducidas (iPSC) derivadas de pacientes ofrece una vía prometedora para vencer esta barrera, y han comenzado a surgir informes de
 25 modelos de iPSC de miocardiopatía. Sin embargo, la materialización de esta promesa requerirá enfoques para superar la heterogeneidad de líneas de iPSC derivadas de pacientes.

Se utilizaron líneas celulares Cas9-iPGP1 y ARN de guía tratado con fosfatasa unido a ADN para generar tres líneas de iPSC que son isógenas excepto para la secuencia en TAZ exón 6, que se identificó que portaba una deleción de
 30 un solo nucleótido en pacientes con síndrome de Barth. Una única ronda de transfección de ARN logró una eficacia de HDR de ~30%. Se colonizaron células Cas9-iPGP1 modificadas con mutaciones deseadas (véase la figura 18) y las líneas celulares se diferenciaron en cardiomiocitos. Los cardiomiocitos derivados de la Cas9-iPGP1 modificada por ingeniería genética recapitulaban completamente las deficiencias de ATP, mitocondriales y de cardiolipina observadas en iPSC derivadas de pacientes y en el modelo de inactivación de TAZ de ratas neonatales (véase la
 35 figura 19). Por consiguiente, se proporcionan procedimientos para corregir mutaciones que provocan enfermedades en células pluripotentes seguido por diferenciación de las células en tipos de células deseadas.

EJEMPLO XX

40 Materiales y procedimientos

1. Establecimiento de líneas de iPS/ES humanas estables inducibles con dox con Cas9 PiggyBac

- 45 1. Tras alcanzar las células el 70% de confluencia, pretratar el cultivo con inhibidor de ROCK Y27632 a una concentración final de 10 μ M durante la noche.
2. El día siguiente preparar la solución de nucleofección combinando los 82 μ l de solución Nucleofector de células madre humanas y 18 μ l suplemento 1 en un tubo eppendorf de 1,5 ml estéril. Mezclar bien. Incubar la solución a 37°C durante 5 min.
- 50 3. Aspirar mTeSR1; enjuagar suavemente las células con DPBS a 2 ml/pocillo de una placa de seis pocillos.
4. Aspirar el DPBS, añadir 2 ml/pocillo de Versene y poner el cultivo de nuevo en el incubador a 37°C hasta que se redondean y quedan adheridas de manera floja, pero no desprendidas. Esto requiere 3-7 min.
5. Aspirar suavemente el Versene y añadir mTeSR1. Añadir 1 ml de mTeSR1 y extraer las células haciendo fluir suavemente mTeSR1 sobre las mismas con una micropipeta de 1.000 μ l.
- 55 6. Recoger las células extraídas, triturarlas suavemente para dar una suspensión de células individuales y cuantificarlas mediante un hemocitómetro y ajustar la densidad celular a 1 millón de células por ml.
7. Añadir 1 ml de suspensión celular a un tubo eppendorf de 1,5 ml y centrifugar a 1.100 RPM durante 5 min en una centrífuga de sobremesa.
8. Resuspender las células en 100 μ l de solución Nucleofector de células madre humanas de la etapa 2.
- 60 9. Transferir las células a una cubeta Nucleofector utilizando una punta de pipeta de 1 ml. Añadir 1 μ g de plásmido de transposonasa y 5 μ g de plásmidos de PB Cas9 en la suspensión celular en la cubeta. Mezclar las células y el ADN removiendo suavemente.
10. Poner la cubeta en el Nucleofector. Se seleccionó el programa B-016 y se sometieron a nucleofección las células pulsando el botón X.
- 65 11. Añadir 500 μ l de medio mTeSR1 con inhibidor de ROCK en la cubeta tras la nucleofección.
12. Aspirar las células sometidas a nucleofección de la cubeta utilizando la pipeta Pasteur de plástico proporcionada y transferir las células gota a gota a un pocillo recubierto con Matrigel de una placa de 6 pocillos

ES 2 715 666 T3

de medio mTeSR1 con inhibidor de ROCK. Incubar las células a 37°C durante la noche.

13. Cambiar el medio a mTesr1 el día siguiente y tras 72 horas de transfección; añadir puromicina a una concentración final a 1 ug/ml. La línea se establecerá en el plazo de 7 días.

5 2. Preparación de ARN

1. Preparar un molde de ADN con promotor de T7 cadena arriba de la secuencia codificante de ARNg.
2. Purificar el ADN utilizando Mega Clear Purification y normalizar la concentración.
3. Preparar mezclas de NTPS a medida para la producción de diferentes ARNg.

N.º 1 Mezcla de ARN nativo	[Final] (mM)
GTP	7,5
ATP	7,5
CTP	7,5
UTP	7,5
Volumen total	

10

N.º 2 Mezcla de ARN nativo con caperuza	[Final] (mM)
Análogo de estructura de caperuza de 3'-O-Me-m7G (NEB)	6
GTP	1,5
ATP	7,5
CTP	7,5
UTP	7,5
Volumen total	

N.º 3 Mezcla de ARN modificado	[Final] (mM)
GTP	7,5
ATP	7,5
5-Me-CTP (Tri-Link)	7,5
Pseudo-UTP (Tri-Link)	7,5
Volumen total	

N.º 4 Mezcla de ARN con caperuza/modificado	[Final] (mM)
Análogo de estructura de caperuza de 3'-O-Me-m7G (NEB)	6
GTP	1,5
ATP	7,5
5-Me-CTP (Tri-Link)	7,5
Pseudo-UTP (Tri-Link)	7,5
Volumen total	

4. Preparar la mezcla de transcripción *in vitro* a temperatura ambiente.

15

	Amt (ul)
NTPS a medida (*añadir vol/IVT rxn tal como se indicó anteriormente) sobre hielo	NA
Producto de PCR (100 ng/ul) = 1.600 ng en total ([final]=40 ng/ul)	16
Tampón X10 (kit MEGAscript de Ambion) a TA	4
Enzima de T7 (kit MEGAscript de Ambion)	4

5. Incubar durante 4 horas (3-6 h apropiado) a 37°C (termociclador).

6. Añadir 2 µl de ADNasa Turbo (kit MEGAscript de Ambion) a cada muestra. Mezclar suavemente e incubar a

37°C durante 15'.

7. Purificar la reacción tratada con ADNasa utilizando MegaClear de Ambion según las instrucciones del fabricante.

8. Purificar ARN utilizando MEGAclean. (El ARN purificado puede almacenarse a -80 durante varios meses).

5 9. Eliminar los grupos fosfato para evitar la reacción inmunitaria de Toll 2 a partir de la célula huésped.

Tratamiento con fosfatasa del ARN	1X	12
Para cada muestra de ARN	~ 100 ul	NA
Tampón fosfatasa Antarctic 10X	11 ul	132
Fosfatasa Antarctic	2 ul	24
Mezclar suavemente la muestra e incubar a 37°C durante 30' (30'-1 h apropiado)		

3. Transfección de ARN

10 1. Sembrar en placa 10K-20K células por 48 pocillos sin antibióticos. Las células deben ser confluentes al 30-50% para la transfección.

2. Cambiar el medio celular a medio con B18R (200 ng/ml), DOX (1 ug/ml), puromicina (2 ug/ml), como mínimo, dos horas antes de la transfección.

15 3. Preparar el reactivo de transfección que contiene ARNg (0,5 ug~2 ug), ADN donador (0,5 ug~2 ug) y RNAiMax, incubar la mezcla en temperatura rm durante 15 minutos y transferir a la célula.

4. Siembra de células iPS humanas individuales y recogida de clones individuales

20 1. Tras 4 días de inducción con dox y 1 día de retirada de dox, aspirar el medio, enjuagar suavemente con el DPBS. Añadir 2 ml/pocillo de Versene y poner el cultivo de nuevo en el incubador a 37°C hasta que se redondean y quedan adheridas de manera floja, pero no desprendidas. Esto requiere 3-7 min.

25 2. Aspirar suavemente el Versene y añadir mTeSR1. Añadir 1 ml de mTeSR1 y extraer las células haciendo fluir suavemente mTeSR1 sobre ellas con una micropipeta de 1.000 ul.

3. Recoger las células extraídas, triturarlas suavemente para dar una suspensión de células individuales y cuantificarlas mediante un hemocitómetro y ajustar la densidad celular a 100K células por ml.

4. Sembrar las células en placas de 10 cm recubiertas con Matrigel con mTeSR1 más inhibidor de ROCK a una densidad celular de 50K, 100K y 400K por placa de 10 cm.

30 5. Examen de los clones de células individuales formados

35 1. Tras 12 días cultivo en la placa de 10 cm, los clones son lo suficientemente grandes como para identificarse a simple vista y se marcaron mediante marcador de colon. No permitir que los clones se vuelvan demasiado grandes y se adhieran entre sí.

2. Poner la placa de 10 cm en la campana de cultivo y utilizar una pipeta P20 (ajustada a 10 ul) con puntas de filtro. Aspirar 10 ul de medio para un pocillo de placas de 24 pocillos. Recoger un clon raspando el clon para dar pequeños trozos y transferirlos a un pocillo de la placa de 24 pocillos. Cada punta de filtro para cada clon.

40 3. Tras 4-5 días, los clones dentro de un pocillo de la placa de 24 pocillos se volvieron lo suficientemente grandes como para dividirse.

4. Aspirar el medio y enjuagar con 2 ml/pocillo de DPBS.

5. Aspirar el DPBS, reemplazar por 250 ul/pocillo de dispasa (0,1 U/ml) e incubar las células en dispasa a 37°C durante 7 min.

6. Reemplazar la dispasa por 2 ml de DPBS.

45 7. Añadir 250 ul de mTeSR1. Utilizar un raspador de células para extraer las células y recoger las células.

8. Transferir 125 ul de suspensión celular a un pocillo de placas de 24 pocillos recubiertas con Matrigel.

9. Transferir 125 ul de suspensión celular a un tubo eppendorf de 1,5 ml para la extracción de ADN genómico.

6. Examen de clones

50 1. Centrifugar el tubo de la etapa 7.7

2. Aspirar el medio y añadir 250 ul de tampón de lisis por pocillo (10 mM+Tris pH 7,5+(o+8,0), EDTA 10 mM, 10 mM.

3. NaCl, +SDS al 10%, +proteinasas K 40 ug/ml (añadida de manera reciente antes de utilizar el tampón).

55 4. Incubar a 55 durante la noche.

5. Precipitar el ADN añadiendo 250 ul de isopropanol.

6. Centrifugar durante 30 minutos a la máxima velocidad. Lavar con etanol al 70%.

7. Eliminar suavemente el etanol. Secar al aire durante 5 min.

8. Resuspender ADNg con 100-200 ul de dH₂O.
9. Amplificación por PCR de la región genómica seleccionada como diana con cebadores específicos.
10. Secuenciación de Sanger del producto de PCR con el cebador respectivo.
11. Análisis de los datos de secuencia de Sanger y expansión de los clones seleccionados como diana.

- 5
7. Eliminación del vector Piggybac
1. Repetir la etapa 2.1-2.9
 2. Transferir las células a una cubeta Nucleofector utilizando una punta de pipeta de 1 ml. Añadir 2 µg de plásmido de transposonasa a la célula suspensión en la cubeta. Mezclar las células y el ADN removiendo suavemente.
 3. Repetir la etapa 2.10-2.11
 4. Aspirar las células sometidas a nucleofección de la cubeta utilizando la pipeta Pasteur de plástico proporcionada y transferir las células gota a gota a un pocillo recubierto con Matrigel de una placa de 10 cm con medio mTeSR1 más inhibidor de ROCK. Incubar las células a 37°C durante la noche.
 5. El día siguiente cambiar el medio a mTesr1 y cambiar el medio cada día durante los 4 días siguientes.
 6. Después de que los clones se vuelvan lo suficientemente grandes, recoger 20-50 clones y sembrarlos en 24 pocillos.
 7. Genotipar los clones con cebadores de vector PB Cas9 PiggyBac y expandir los clones negativos.

20 Referencias

Las referencias se designan a lo largo de toda la memoria descriptiva mediante su número a continuación.

- 25
1. Carroll,D. (2011) Genome engineering with zinc-finger nucleases. *Genetics*, 188, 773-82.
 2. Wood,A.J., Lo,T.-W., Zeitler,B., Pickle,C.S., Ralston,E.J., Lee,A.H., Amora,R., Miller,J.C., Leung,E., Meng,X., y otros. (2011) Targeted genome editing across species using ZFNs and TALENs. *Science (Nueva York, N.Y.)*, 333, 307.
 3. Perez-Pinera,P., Ousterout,D.G. y Gersbach,C.A. (2012) Advances in targeted genome editing. *Current opinion in chemical biology*, 16, 268-77.
 4. Symington,L.S. y Gautier,J. (2011) Double-strand break end resection and repair pathway choice. *Annual review of genetics*, 45, 247-71.
 5. Urnov,F.D., Miller,J.C., Lee,Y.-L., Beausejour,C.M., Rock,J.M., Augustus,S., Jamieson,A.C., Porteus,M.H., Gregory, P.D. y Holmes,M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, 435, 646-51.
 6. Boch,J., Scholze,H., Schornack,S., Landgraf,A., Hahn,S., Kay,S., Lahaye,T., Nickstadt,A. y Bonas,U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (Nueva York, N.Y.)*, 326, 1509-12.
 7. Cell,P., Replacement,K.S., Talens,A., Type,A., Collection,C., Ccl,-A. and Quickextract, E. Genetic engineering of human pluripotent cells using TALE nucleases.
 8. Mussolino,C., Morbitzer,R., Lütge,F., Dannemann,N., Lahaye,T. y Cathomen,T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic acids research*, 39, 9283-93.
 9. Ding,Q., Lee,Y., Schaefer,E.A.K., Peters,D.T., Veres,A., Kim,K., Kuperwasser,N., Motola,D.L., Meissner,T.B., Hendriks,W.T., y otros. (2013) Resource A TALEN Genome-Editing System for Generating Human Stem Cell-Based Disease Models.
 10. Hockemeyer,D., Wang,H., Kiani,S., Lai,C.S., Gao,Q., Cassady,J.P., Cost,G.J., Zhang,L., Santiago,Y., Miller, J.C., y otros. (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nature biotechnology*, 29, 731-4.
 11. Bedell,V.M., Wang,Y., Campbell,J.M., Poshusta,T.L., Starker,C.G., Krug li,R.G., Tan,W., Penheiter,S.G., Ma,A.C., Leung,A.Y.H., y otros. (2012) In vivo genome editing using a high-efficiency TALEN system. *Nature*, 490, 114-118.
 12. Miller,J.C., Tan,S., Qiao,G., Barlow,K. a, Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J., y otros. (2011) A TALE nuclease architecture for efficient genome editing. *Nature biotechnology*, 29, 143-8.
 13. Holkers,M., Maggio,I., Liu,J., Janssen,J.M., Miselli,F., Mussolino,C., Recchia,A., Cathomen,T. y Gonçalves,M. a F. V (2012) Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into

- human cells. *Nucleic acids research*, 10.1093/nar/gks 1446.
- 5 14. Reyon,D., Tsai,S.Q., Khayter,C., Foden,J. a, Sander,J.D. y Joung,J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nature Biotechnology*, 30, 460-465.
- 10 15. Qiu,P., Shandilya,H., D'Alessio,J.M., O'Connor,K., Durocher,J. y Gerard,G.F. (2004) Mutation detection using Surveyor nuclease. *BioTechniques*, 36, 702-7.
16. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. y Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science (Nueva York, N.Y.)*, 339, 823-6.
17. Ding,Q., Regan,S.N., Xia,Y., Ostrom,L.A., Cowan,C.A. y Musunuru,K. (2013) Enhanced Efficiency of Human Pluripotent Stem Cell Genome Editing through Replacing TALENs with CRISPRs. *Cell Stem Cell*, 12, 393-394.
- 15 18. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L. a, y otros. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science (Nueva York, N.Y.)*, 339, 819-23.
19. Cho,S.W., Kim,S., Kim,J.M. y Kim,J.-S. (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology*, 31, 230-232.
- 20 20. Hwang,W.Y., Fu,Y., Reyon,D., Maeder,M.L., Tsai,S.Q., Sander,J.D., Peterson,R.T., Yeh,J.-R.J. y Joung,J.K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology*, 31, 227-229.
- 25 21. Chen,F., Pruett-Miller,S.M., Huang,Y., Gjoka,M., Duda,K., Taunton,J., Collingwood,T.N., Frodin,M. y Davis, G.D. (2011) High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nature methods*, 8, 753-5.
- 30 22. Soldner,F., Laganieri,J., Cheng,A.W., Hockemeyer,D., Gao,Q., Alagappan,R., Khurana,V., Golbe,L.I., Myers, R.H., Lindquist,S., y otros. (2011) Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell*, 146, 318-31.
- 35 23. Valamehr,B., Abujarour,R., Robinson,M., Le,T., Robbins,D., Shoemaker,D. y Flynn,P. (2012) A novel platform to enable the high-throughput derivation and characterization of feeder-free human iPSCs. *Scientific reports*, 2, 213.
24. Sanjana,N.E., Cong,L., Zhou,Y., Cunniff,M.M., Feng,G. y Zhang,F. (2012) A transcription activator-like effector toolbox for genome engineering. *Nature protocols*, 7, 171-92.
- 40 25. Gibson,D.G., Young,L., Chuang,R., Venter,J.C., Iii,C.A.H., Smith,H.O. y America,N. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *6*, 12-16.
26. Zou,J., Maeder,M.L., Mali,P., Pruett-Miller,S.M., Thibodeau-Beganny,S., Chou,B.-K., Chen,G., Ye,Z., Park,I.-H., Daley,G.Q., y otros. (2009) Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell stem cell*, 5, 97-110.
- 45 27. Perez,E.E., Wang,J., Miller,J.C., Jouvenot,Y., Kim,K. a, Liu,O., Wang,N., Lee,G., Bartsevich,V. V, Lee,Y.-L., y otros. (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nature biotechnology*, 26, 808-16.
- 50 28. Bhakta,M.S., Henry,I.M., Ousterout,D.G., Das,K.T., Lockwood,S.H., Meckler,J.F., Wallen,M.C., Zykovich,A., Yu,Y., Leo,H., y otros. (2013) Highly active zinc-finger nucleases by extended modular assembly. *Genome research*, 10.1101/gr.143693.112.
- 55 29. Kim,E., Kim,S., Kim,D.H., Choi,B.-S., Choi,I.-Y. y Kim,J.-S. (2012) Precision genome engineering with programmable DNA-nicking enzymes. *Genome research*, 22, 1327-33.
30. Gupta,A., Meng,X., Zhu,L.J., Lawson,N.D. y Wolfe,S. a (2011) Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic acids research*, 39, 381-92.
- 60 31. Park,I.-H., Lerou,P.H., Zhao,R., Huo,H. y Daley,G.Q. (2008) Generation of human-induced pluripotent stem cells. *Nature protocols*, 3, 1180-6.
32. Cermak,T., Doyle,E.L., Christian,M., Wang,L., Zhang,Y., Schmidt,C., Baller,J.A., Somia,N. V, Bogdanove,A.J. y Voytas,D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research*, 39, e82.
- 65 33. Briggs,A.W., Rios,X., Chari,R., Yang,L., Zhang,F., Mali,P. y Church,G.M. (2012) Iterative capped assembly: rapid

- and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic acids research*, 10.1093/nar/gks624.
- 5 34. Zhang,F., Cong,L., Lodato,S., Kosuri,S., Church,G.M. y Arlotta,P. (2011) LETTErs Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *29*, 149-154.
- 10 35. Pathak,V.K. y Temin,H.M. (1990) Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proceedings of the National Academy of Sciences of the United States of America*, 87, 6019-23.
- 10 36. Tian,J., Ma,K. y Saaem,I. (2009) Advancing high-throughput gene synthesis technology. *Molecular bioSystems*, 5, 714-22.
- 15 37. Zou,J., Mali,P., Huang,X., Dowe,y,S.N. y Cheng,L. (2011) Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood*, 118, 4599-608.
- 15 38. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., Dicarlo,J.E., Norville,J.E. y Church,G.M. (2013) RNA-Guided Human Genome.
- 20 39. Boyle,A.P., Davis,S., Shulha,H.P., Meltzer,P., Margulies,E.H., Weng,Z., Furey,T.S. y Crawford,G.E. (2008) High-resolution mapping and characterization of open chromatin across the genome. *Cell*, 132, 311-22.
- 25 40. Orlando,S.J., Santiago,Y., DeKolver,R.C., Freyvert,Y., Boydston,E. a, Moehle,E. a, Choi,V.M., Gopalan,S.M., Lou,J.F., Li,J., y otros. (2010) Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic acids research*, 38, e152.
41. Wang,Z., Zhou,Z.-J., Liu,D.-P. y Huang,J.-D. (2008) Double-stranded break can be repaired by single-stranded oligonucleotides via the ATM/ATR pathway in mammalian cells. *Oligonucleotides*, 18, 21-32.
- 30 42. Rios,X., Briggs,A.W., Christodoulou,D., Gorham,J.M., Seidman,J.G. y Church,G.M. (2012) Stable gene targeting in human cells using single-strand oligonucleotides with modified bases. *PLoS one*, 7, e36697.
- 35 43. Elliott,B., Richardson,C., Winderbaum,J., Jac,A., Jasin,M. y Nickoloff,J.A.C.A. (1998) Gene Conversion Tracts from Double-Strand Break Repair in Mammalian Cells Gene Conversion Tracts from Double-Strand Break Repair in Mammalian Cells. *18*.
- 40 44. Lombardo,A., Genovese,P., Beausejour,C.M., Colleoni,S., Lee,Y.-L., Kim,K. a, Ando,D., Urnov,F.D., Galli,C., Gregory,P.D., y otros. (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nature biotechnology*, 25, 1298-306.
- 45 45. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J. a y Charpentier,E. (2012) A programmable dual-RNA guided DNA endonuclease in adaptive bacterial immunity. *Science (Nueva York, N.Y.)*, 337, 816-21.
- 45 46. Shrivastav,M., De Haro,L.P. y Nickoloff,J. a (2008) Regulation of DNA double-strand break repair pathway choice. *Cell research*, 18, 134-47.
47. Kim,Y.G., Cha,J. y Chandrasegaran,S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 1156-60.
- 50 48. Mimitou,E.P. y Symington,L.S. (2008) Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature*, 455, 770-4.
- 55 49. Doyon,Y., Choi,V.M., Xia,D.F., Vo,T.D., Gregory,P.D. y Holmes,M.C. (2010) Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. *Nature methods*, 7, 459-60.

REIVINDICACIONES

1. Procedimiento de alteración de ADN diana en una célula madre, en el que la célula madre se modifica genéticamente para incluir un ácido nucleico que codifica una enzima Cas que forma un complejo de colocación con un ARN de guía complementario al ADN diana y que escinde el ADN diana de manera específica de sitio que comprende
- 5
- introducir en la célula madre un ARN de guía complementario al ADN diana y que guía a la enzima al ADN diana, en el que el ARN de guía y la enzima son miembros de un complejo de colocación para el ADN diana,
- 10
- introducir en la célula madre una secuencia de ácido nucleico donadora,
- en el que el ARN de guía y la enzima Cas se localizan conjuntamente en el ADN diana, la enzima Cas escinde el ADN diana y la secuencia de ácido nucleico donadora se inserta en el ADN diana para producir ADN alterado en la célula madre, en el que el ARN de guía comprende ARN de guía tratado con fosfatasa proporcionado a la célula a partir del medio circundante que se suplementa de manera continua con el ARN de guía tratado con fosfatasa para permitir la edición de ADN diana continua, y en el que el procedimiento tiene lugar *in vitro* o *ex vivo*.
- 15
2. Procedimiento, según la reivindicación 1, en el que la enzima es Cas9.
- 20
3. Procedimiento, según la reivindicación 1, en el que el ARN tiene entre aproximadamente 10 y aproximadamente 500 nucleótidos.
4. Procedimiento, según la reivindicación 1, en el que el ARN tiene entre aproximadamente 20 y aproximadamente 100 nucleótidos.
- 25
5. Procedimiento, según la reivindicación 1, en el que el ARN de guía es una fusión de ARNtracr-ARNcr.
6. Procedimiento, según la reivindicación 1, en el que el ADN es ADN genómico, ADN mitocondrial, ADN viral o ADN exógeno.
- 30
7. Procedimiento, según la reivindicación 1, en el que la secuencia de ácido nucleico donadora se inserta mediante recombinación homóloga o mediante unión de extremos no homólogos.
- 35
8. Procedimiento, según la reivindicación 1, en el que la enzima Cas se introduce en la célula madre como ácido nucleico foráneo que la célula madre expresa.
9. Procedimiento, según la reivindicación 1, que comprende además introducir múltiples ácidos nucleicos donadores y múltiples ARN de guía para producir múltiples alteraciones en el ADN en la célula.
- 40
10. Procedimiento, según la reivindicación 1, en el que tras producir ADN alterado en una célula madre, se elimina un ácido nucleico que codifica la enzima Cas del genoma de la célula madre.
11. Procedimiento, según la reivindicación 1, en el que la célula madre se modifica genéticamente insertando un vector o casete eliminable enzimáticamente que incluye un ácido nucleico que codifica la enzima Cas en el ADN genómico de la célula madre.
- 45
12. Procedimiento, según la reivindicación 11, en el que:
- 50
- i) el vector es un vector piggybac;
 - ii) el vector o casete eliminable enzimáticamente que incluye el ácido nucleico que codifica la enzima Cas se elimina utilizando una enzima; o
 - iii) el vector o casete eliminable enzimáticamente que incluye el ácido nucleico que codifica la enzima Cas se elimina utilizando una transposasa.
- 55
13. Procedimiento, según la reivindicación 11, en el que:
- 60
- i) el ácido nucleico que codifica la enzima Cas es inducible;
 - ii) el ácido nucleico que codifica la enzima Cas está bajo la influencia de un promotor inducible por doxiciclina; o
 - iii) el ácido nucleico que codifica la enzima Cas puede inducirse utilizando doxiciclina.
14. Procedimiento, según cualquiera de las reivindicaciones anteriores, en el que el ARN de guía en el medio circundante se compleja con un liposoma.
- 65
15. Procedimiento, según cualquiera de las reivindicaciones anteriores, en el que el ARN de guía se ha tratado con una fosfatasa para evitar una reacción inmunitaria de Toll 2 a partir de la célula madre.

FIG. 1A

Donador de tGFP

Indicador cromosómico

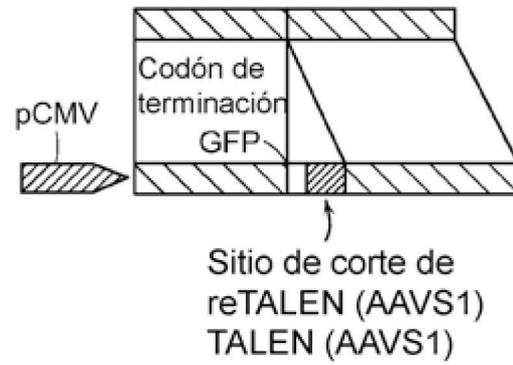


FIG. 1B

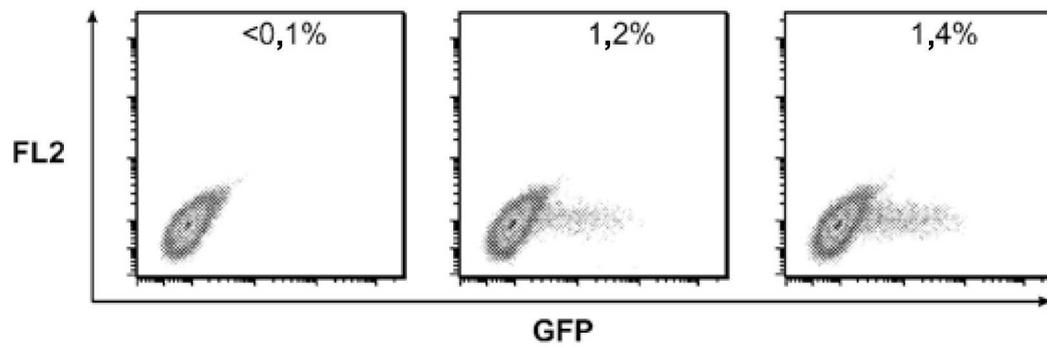
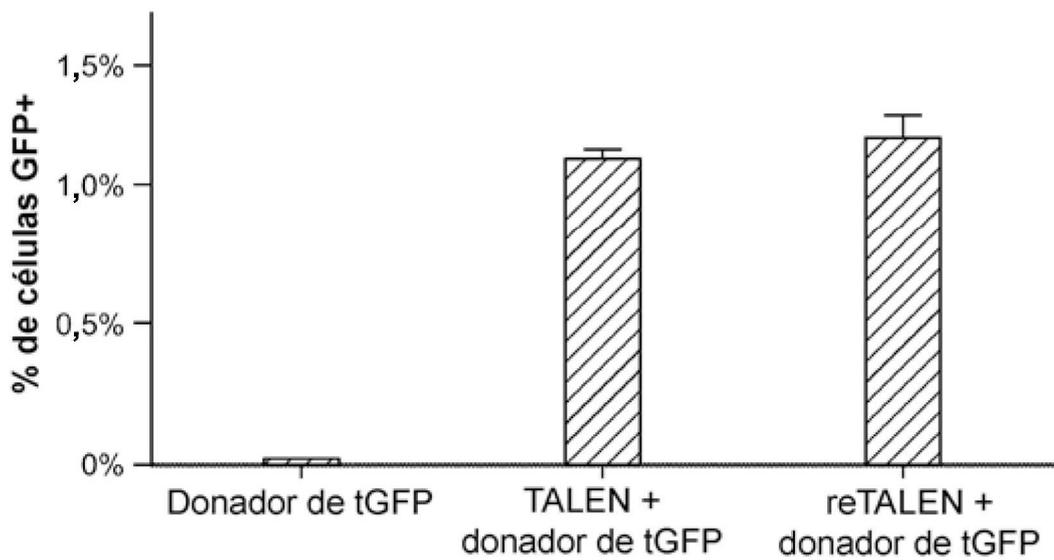
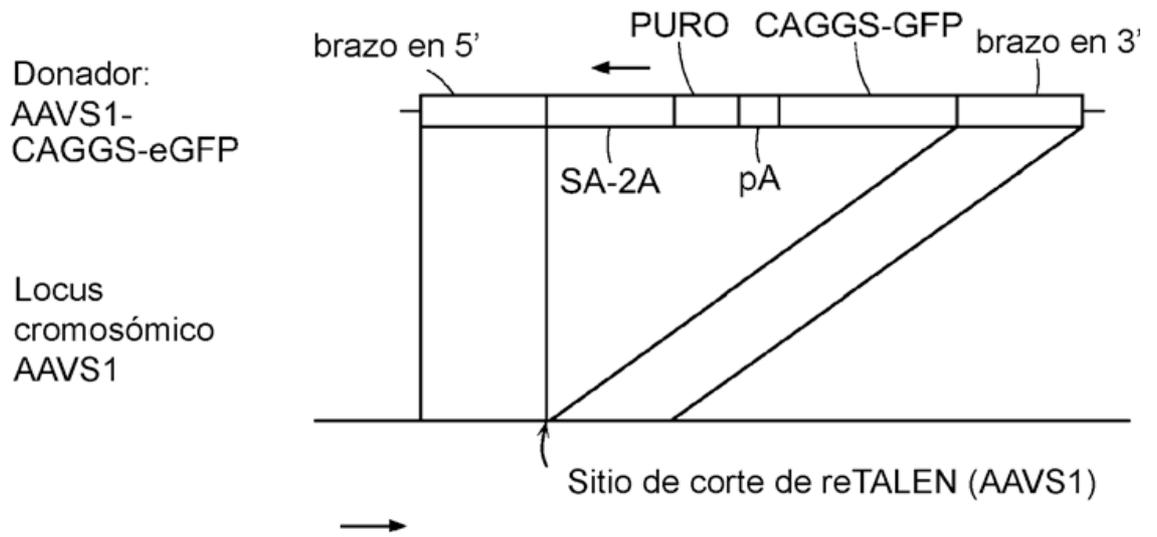


FIG. 1C



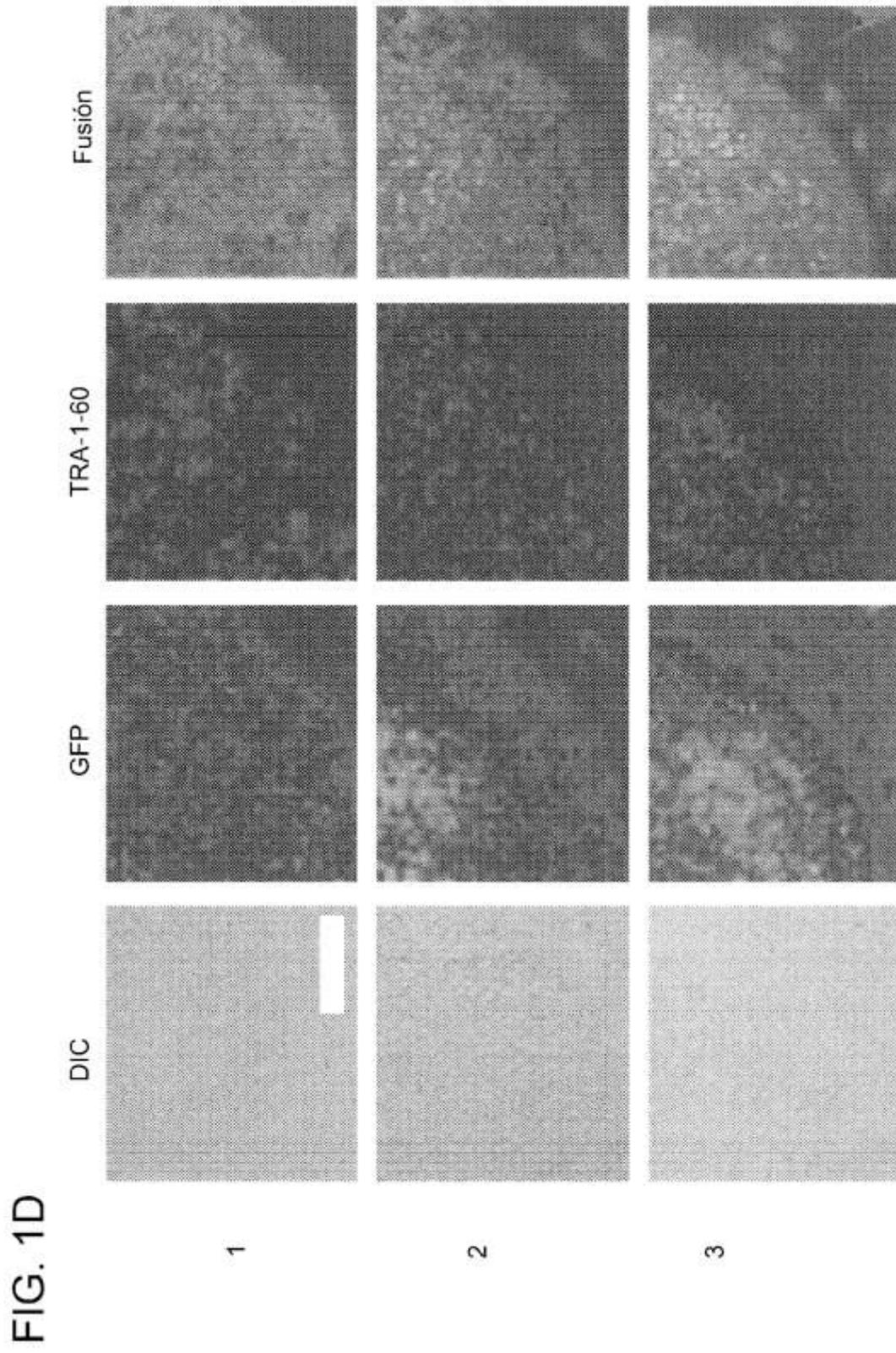


FIG. 1E

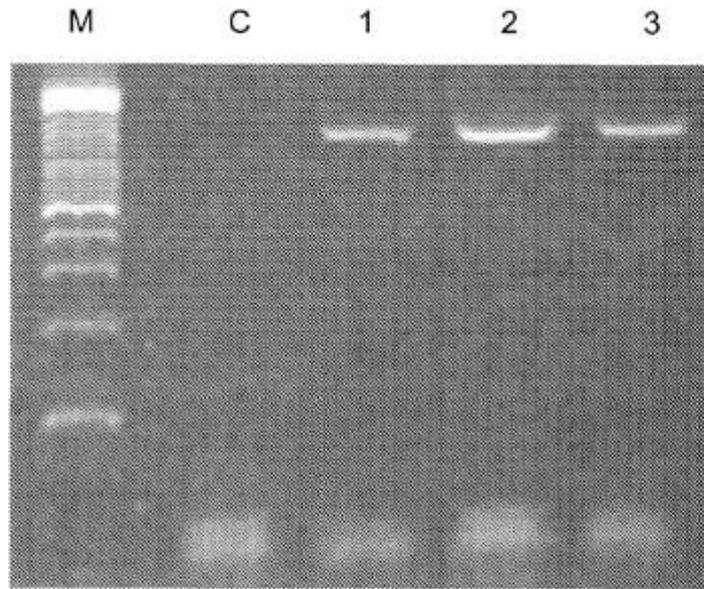


FIG. 1F

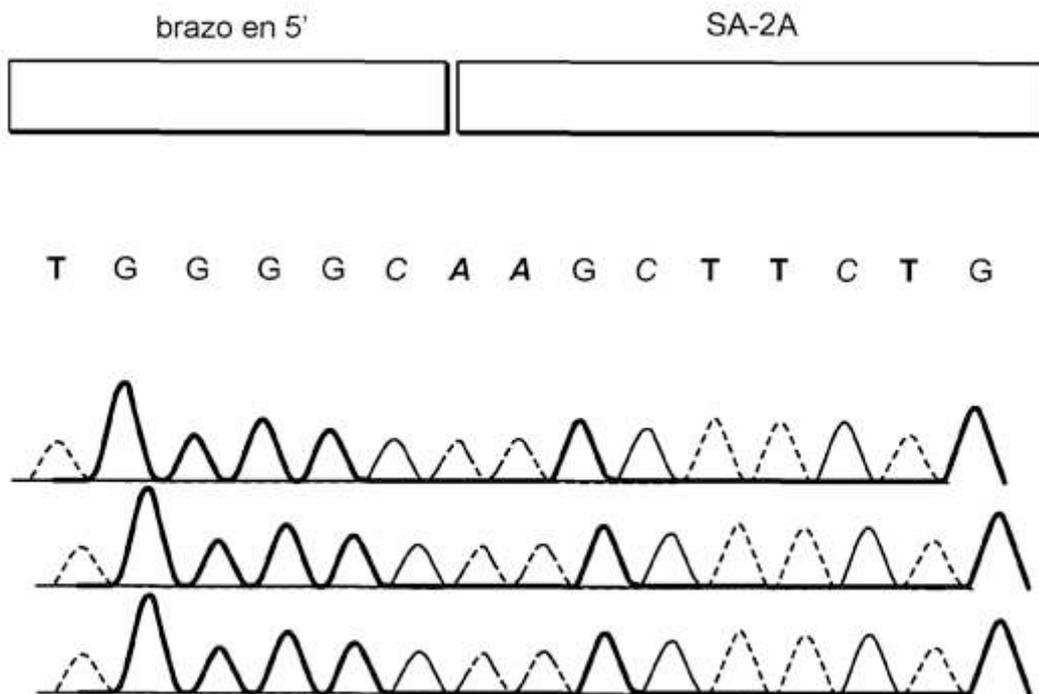
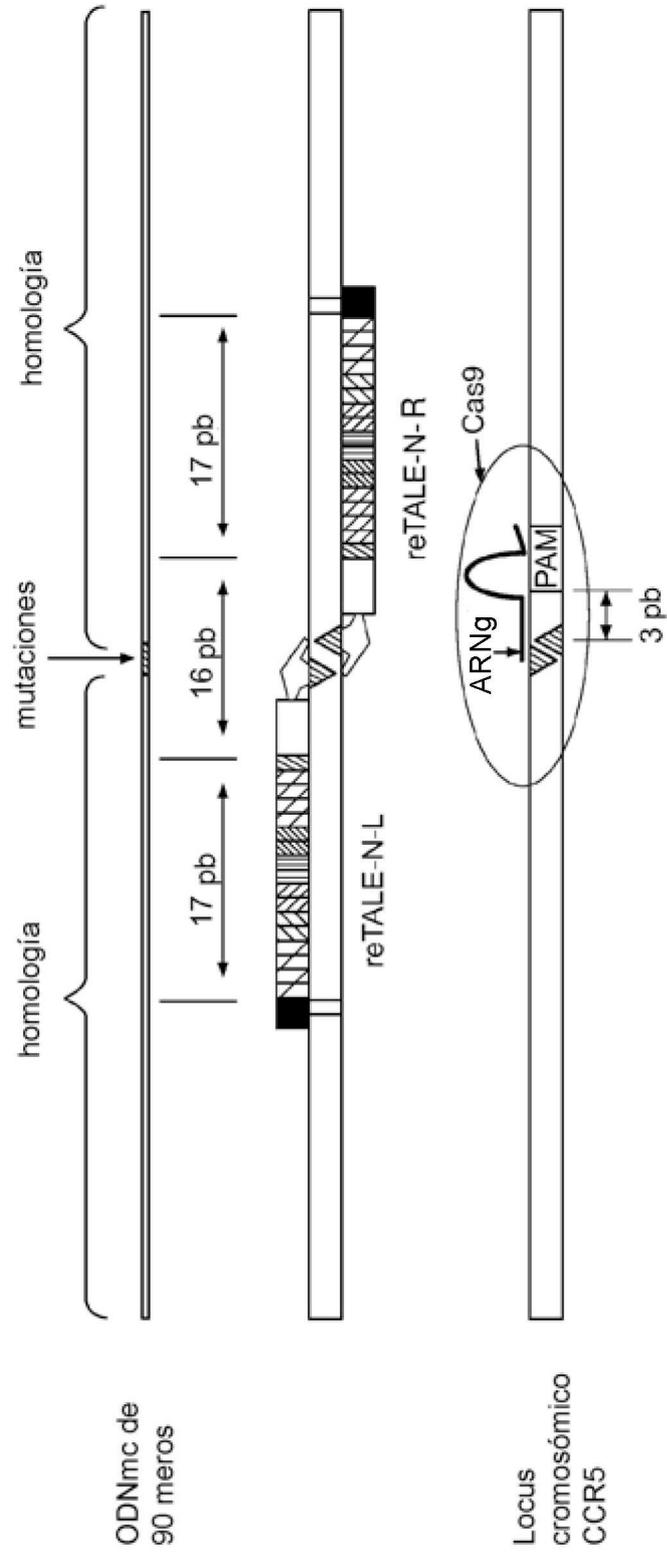
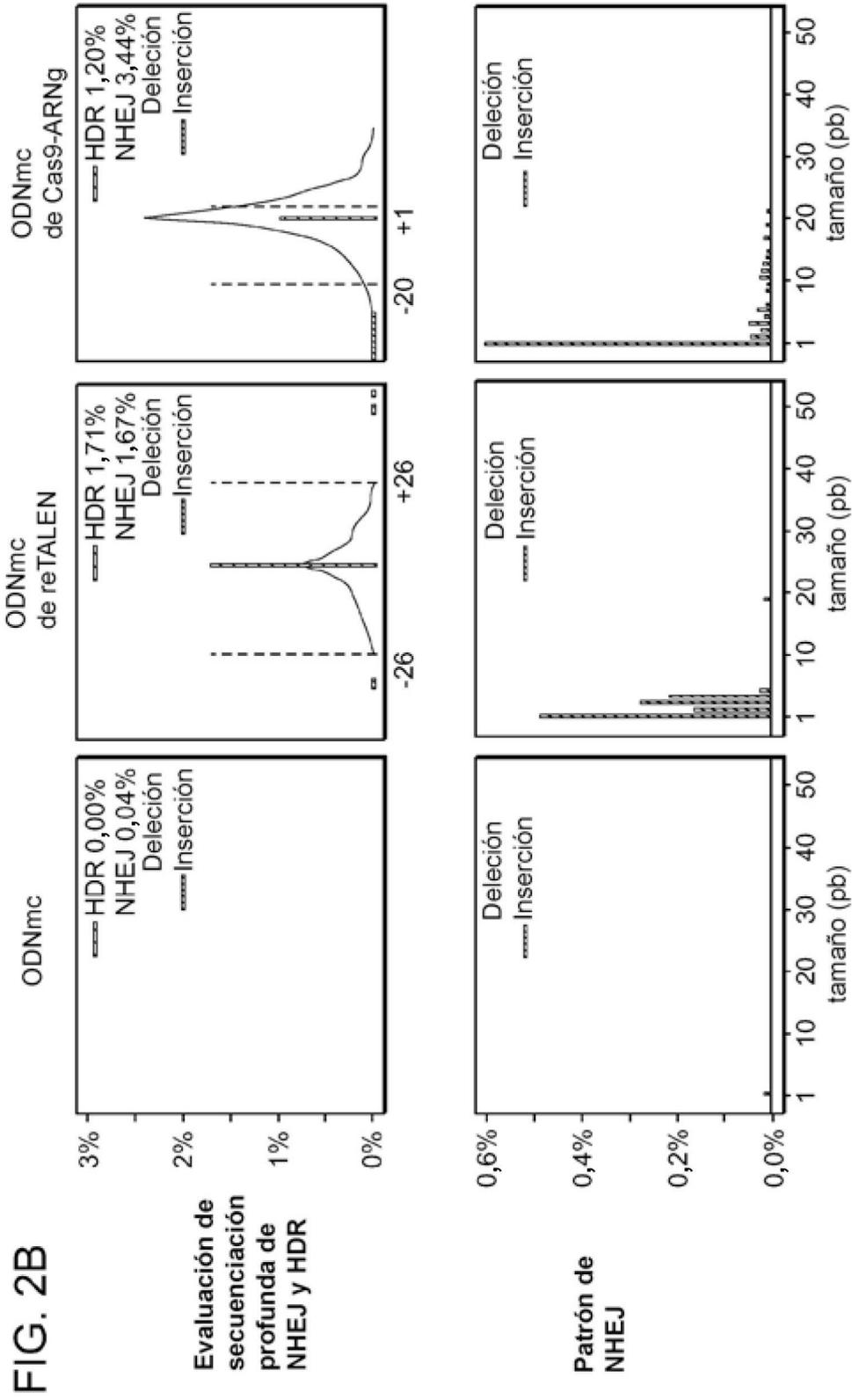


FIG. 2A





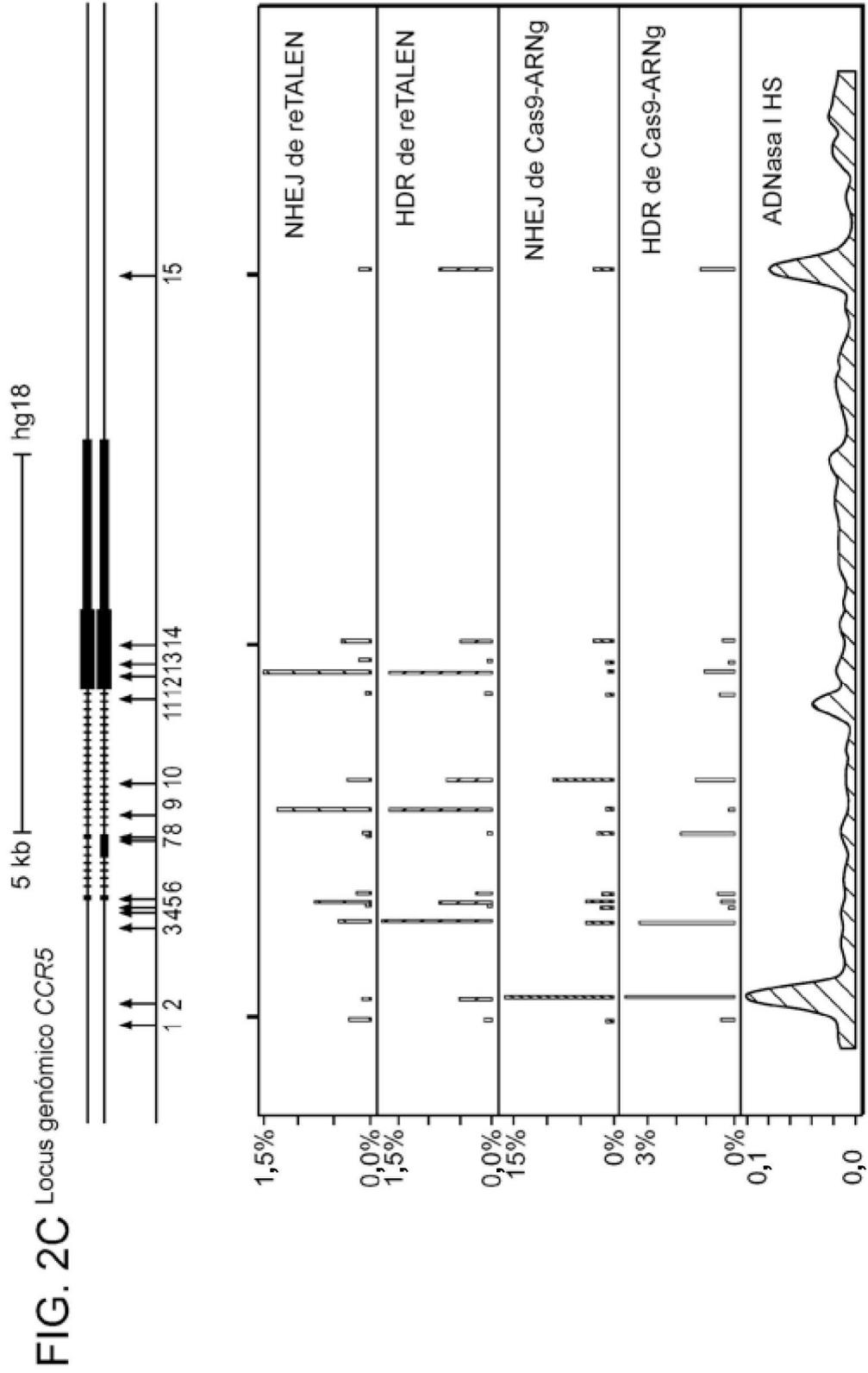


FIG. 3A

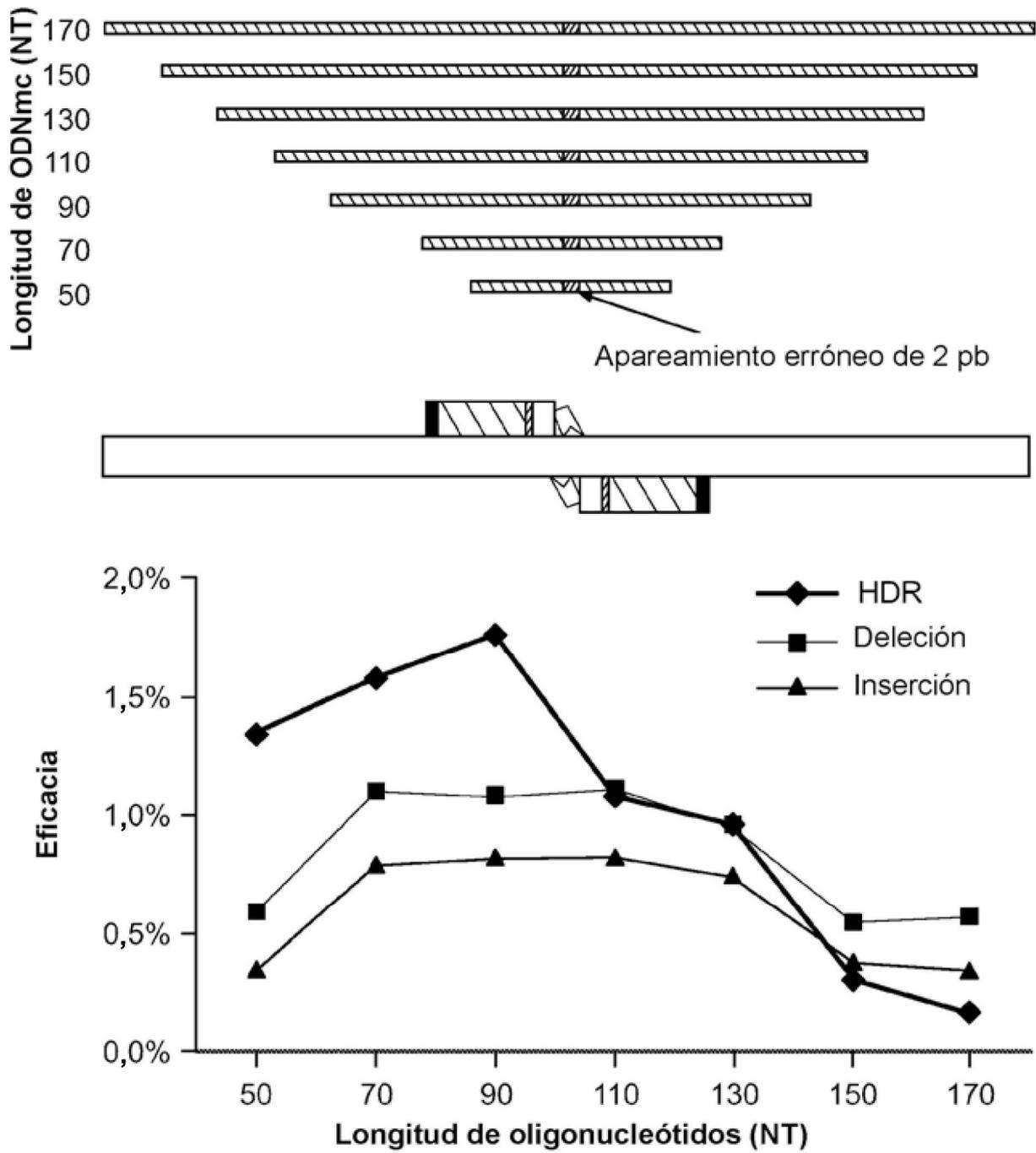


FIG. 3B

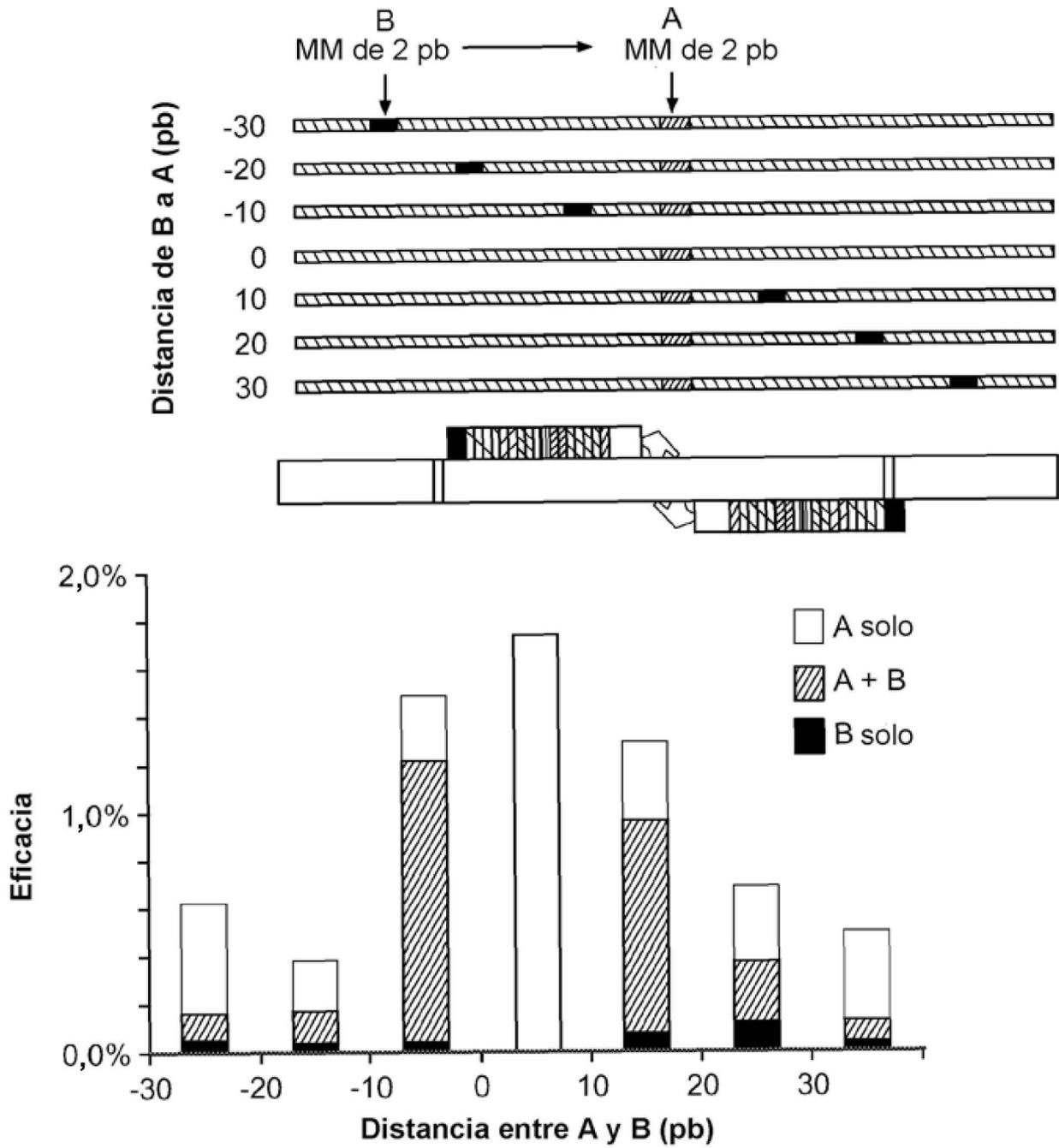


FIG. 3C

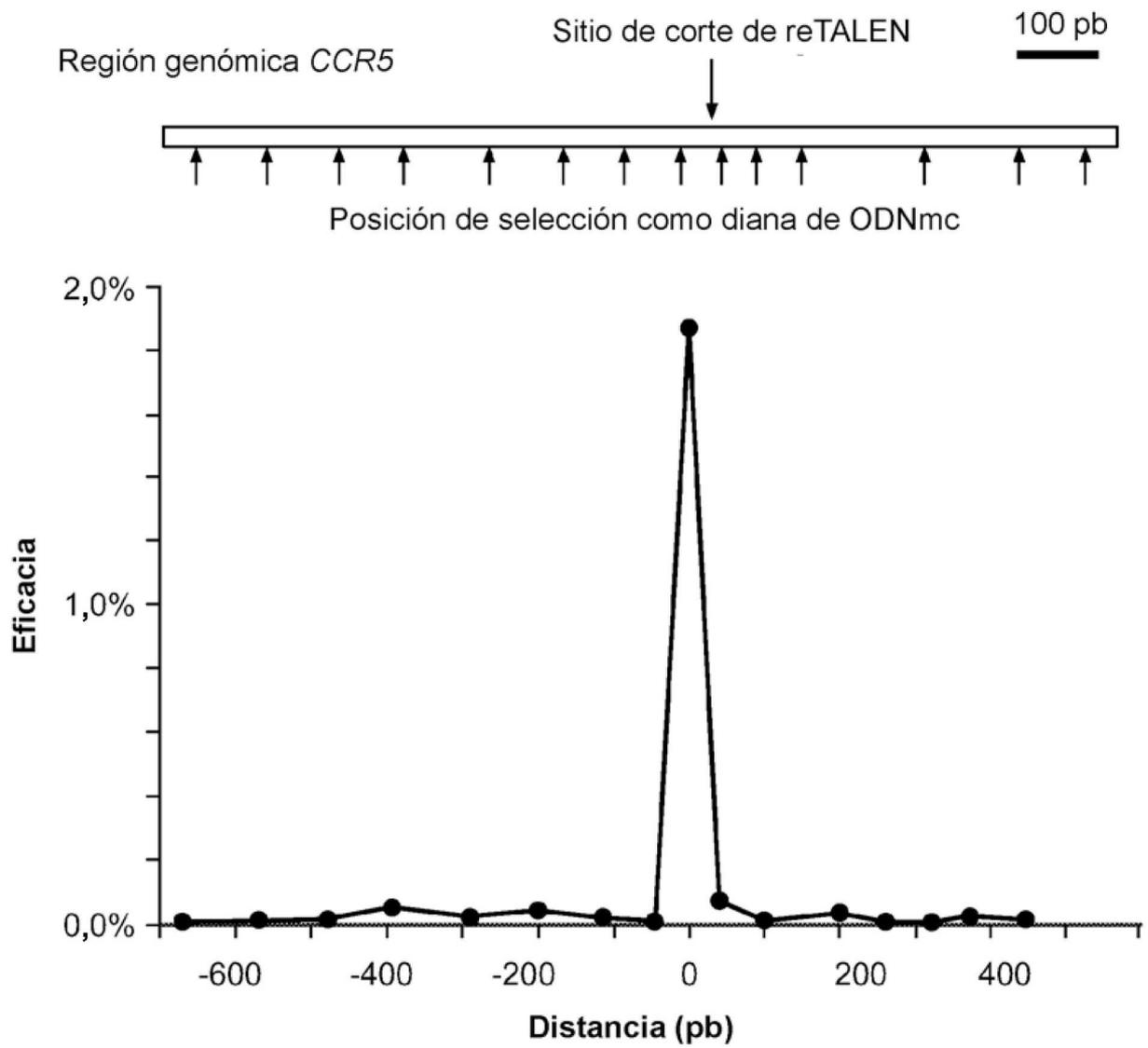
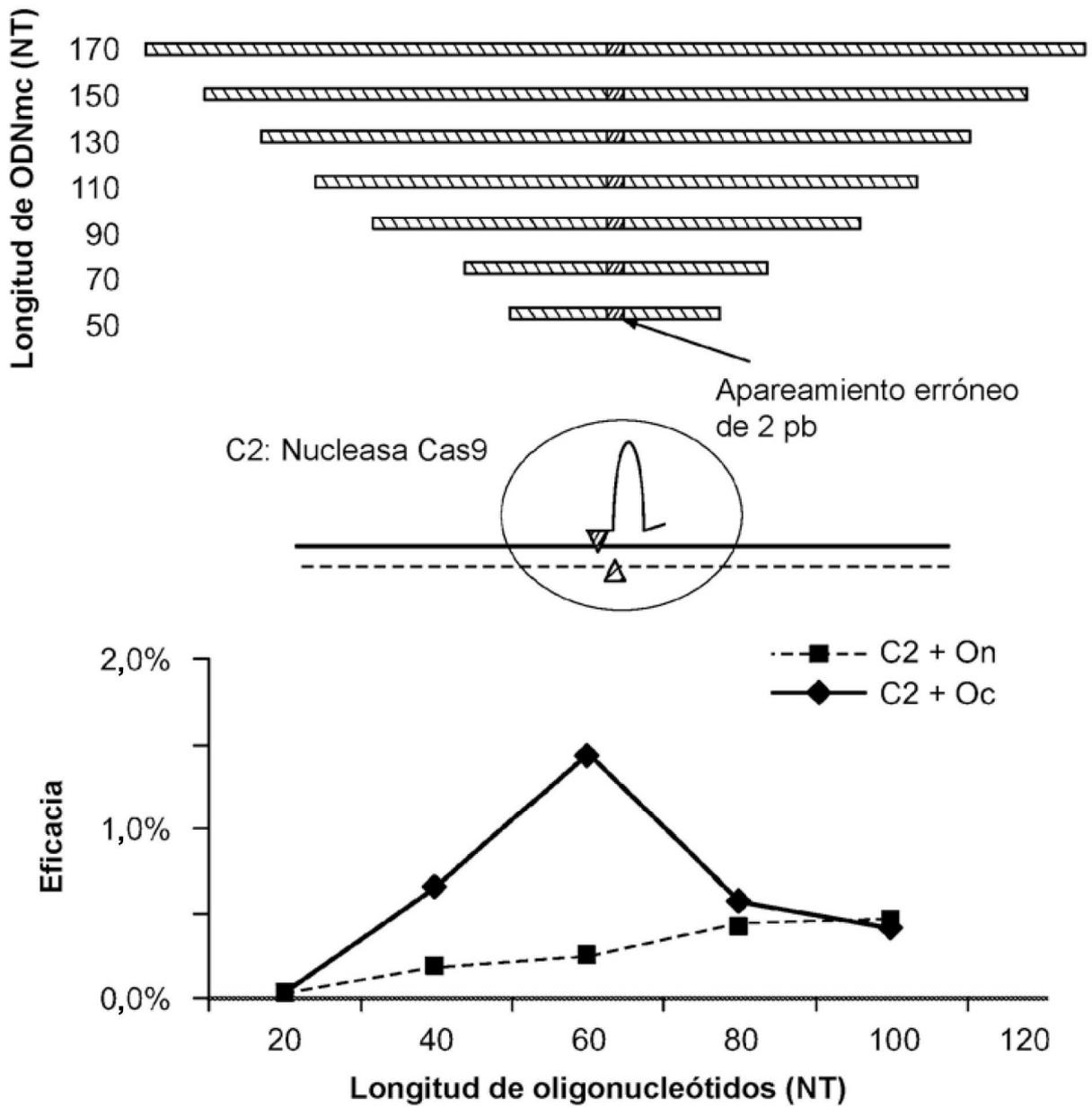


FIG. 3D



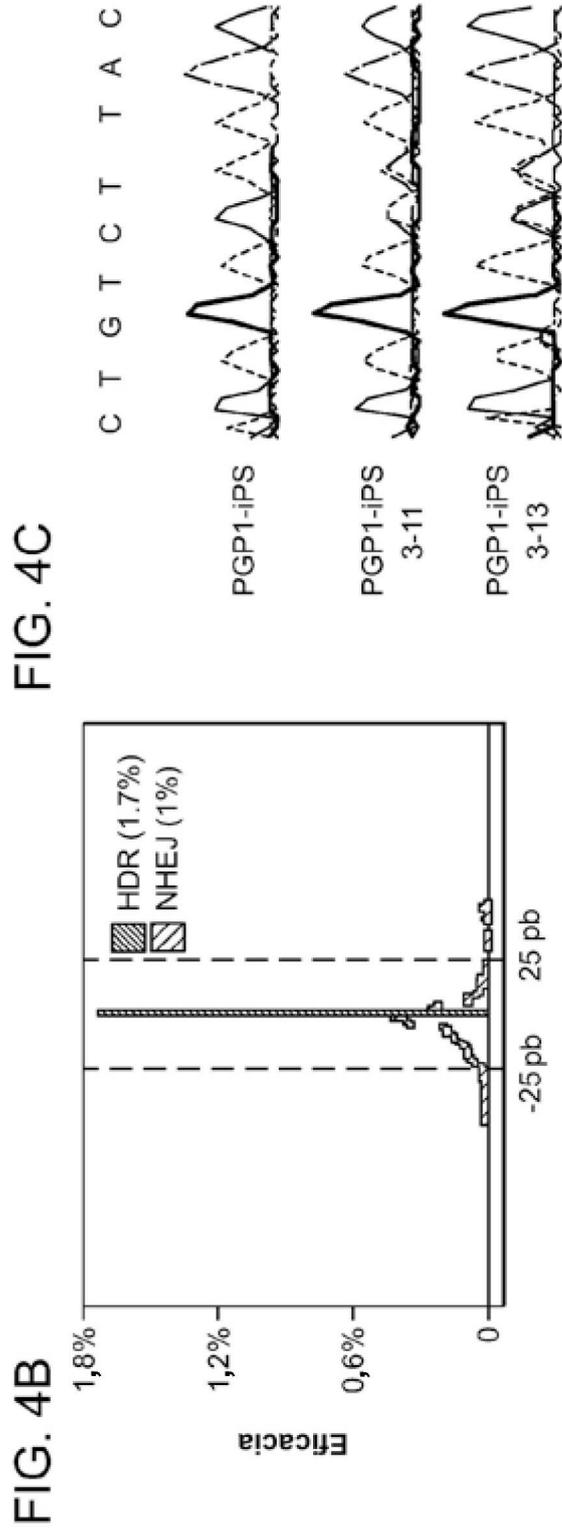
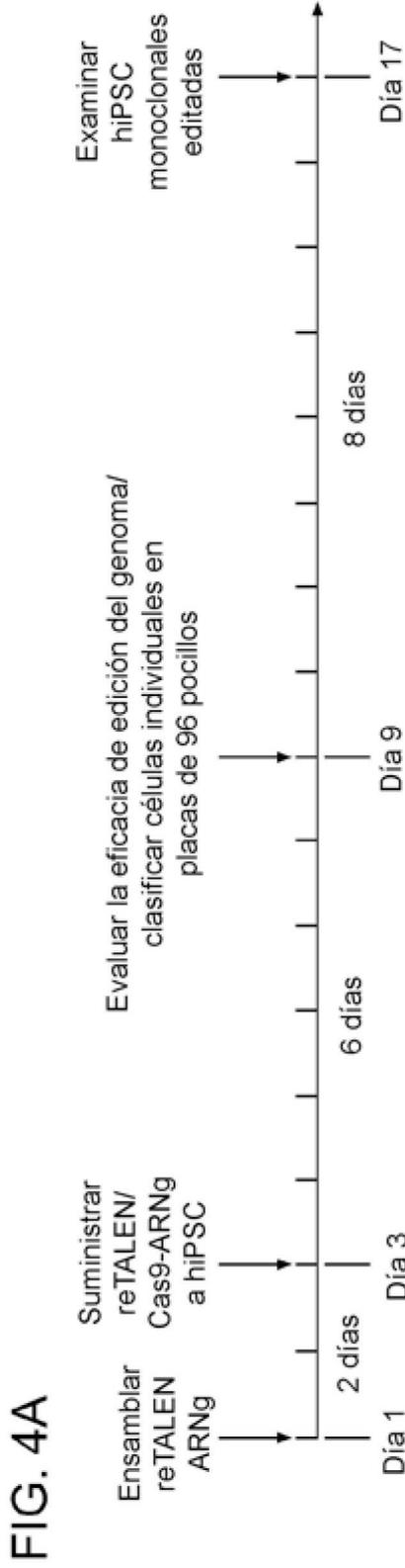
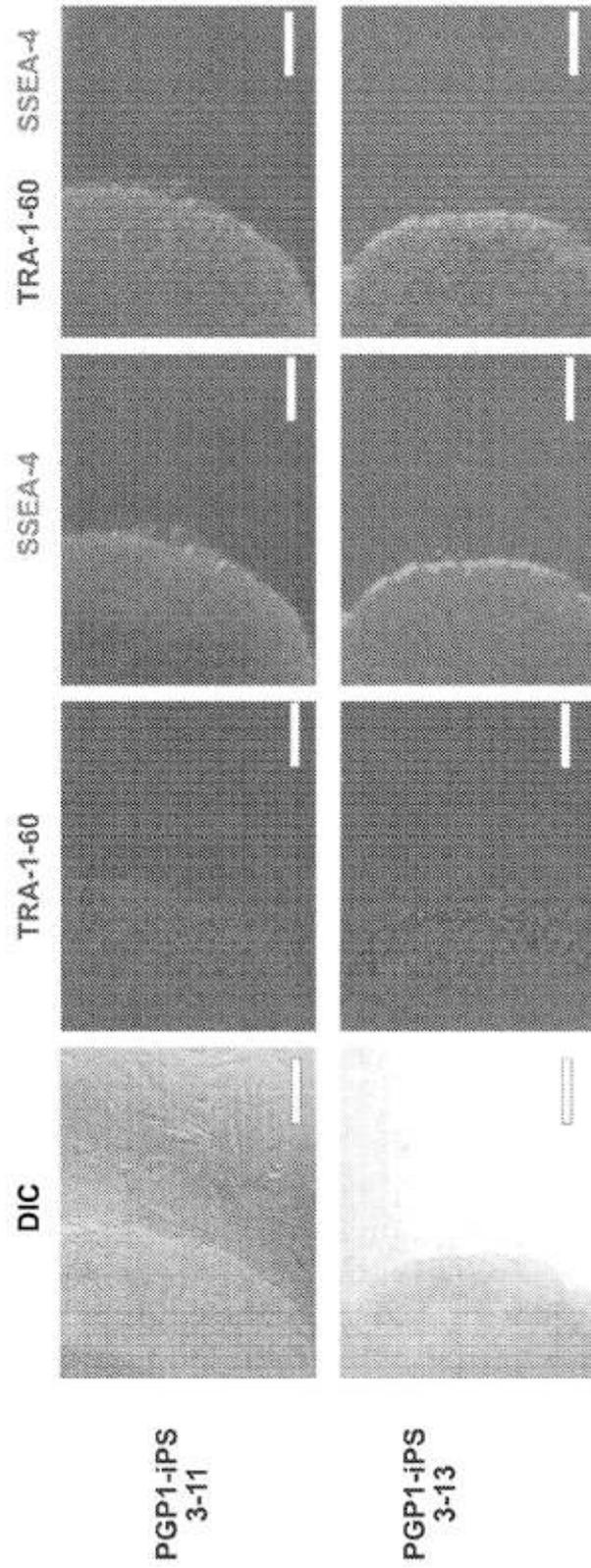


FIG. 4D



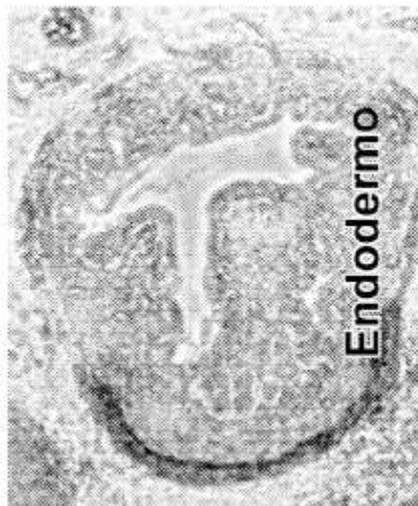
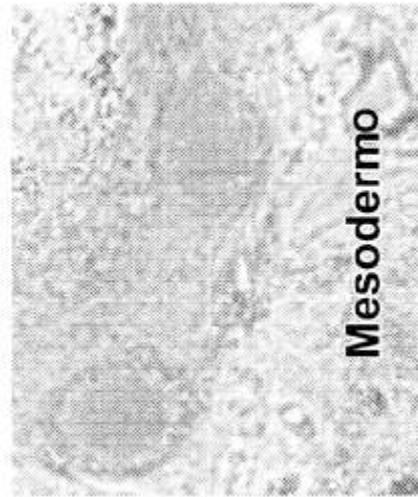
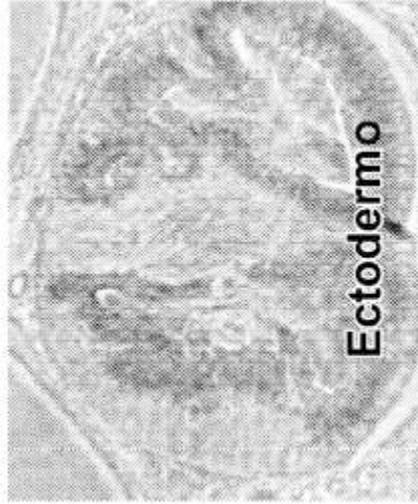


FIG. 4E

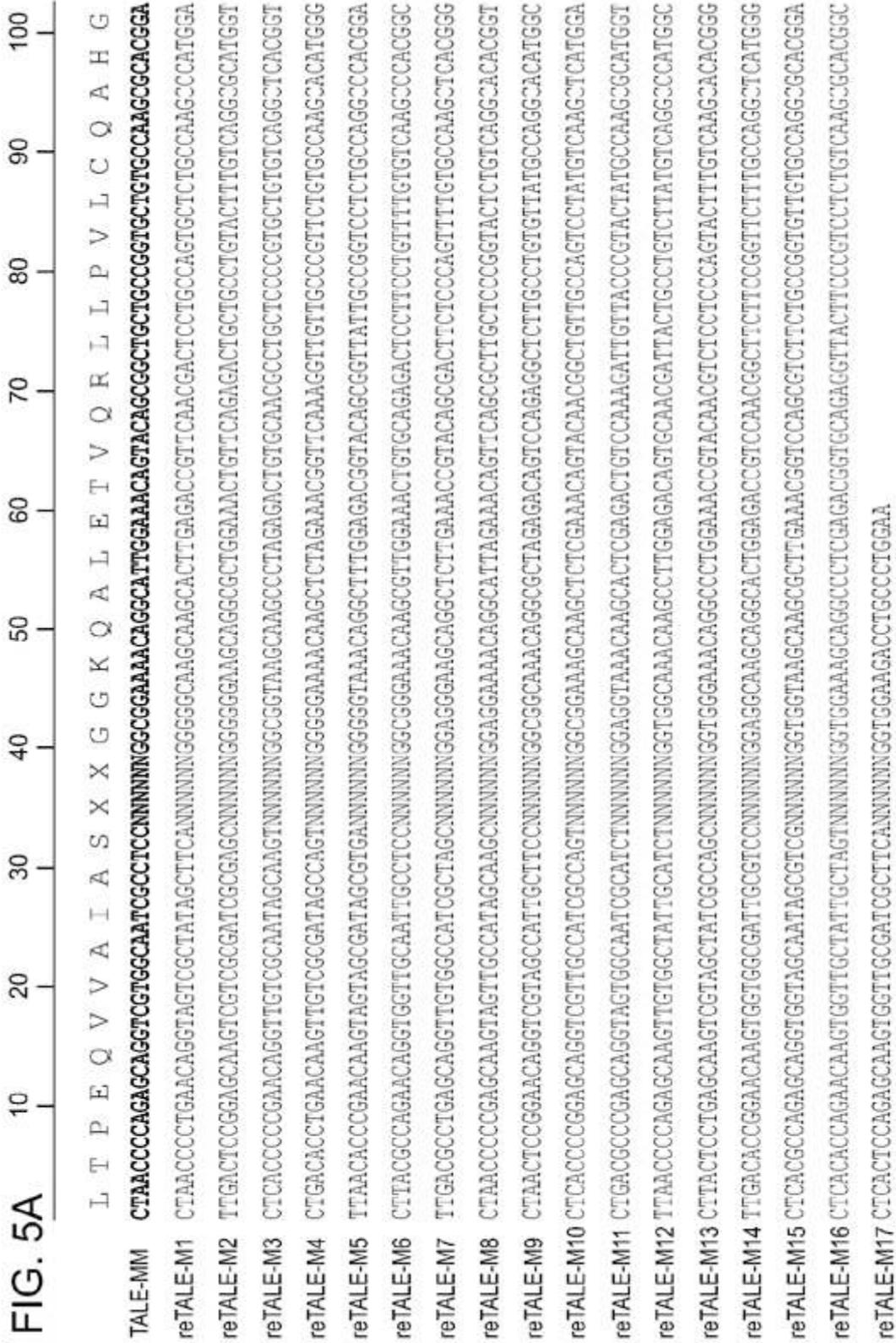
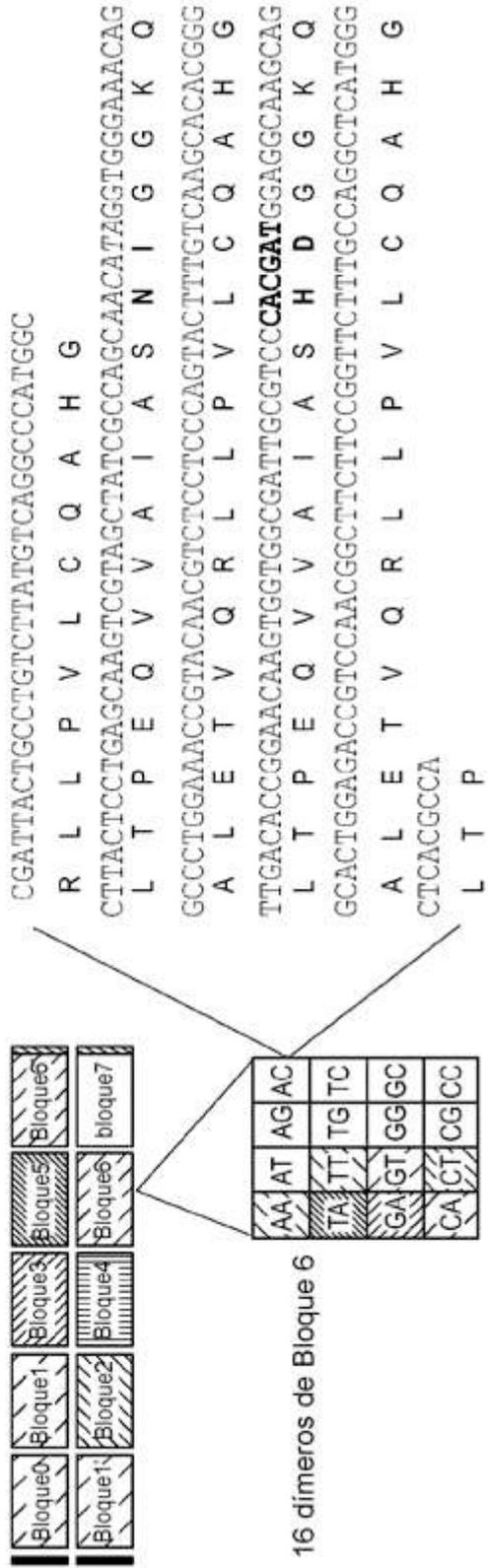


FIG. 6A

Una biblioteca de bloques de dímeros de reTALE Secuencia de dímero (Bloque 6_AC)



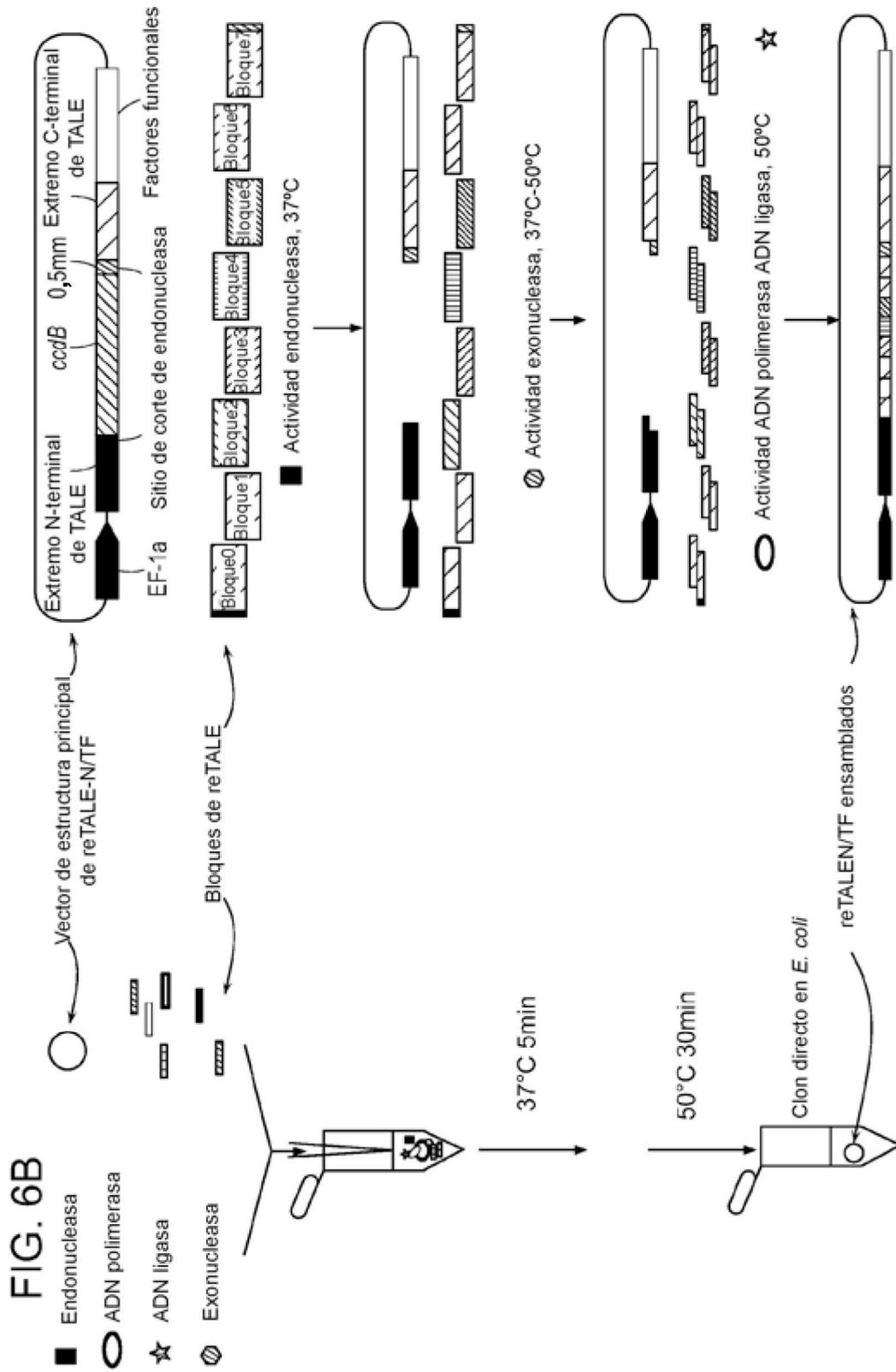


FIG. 6C

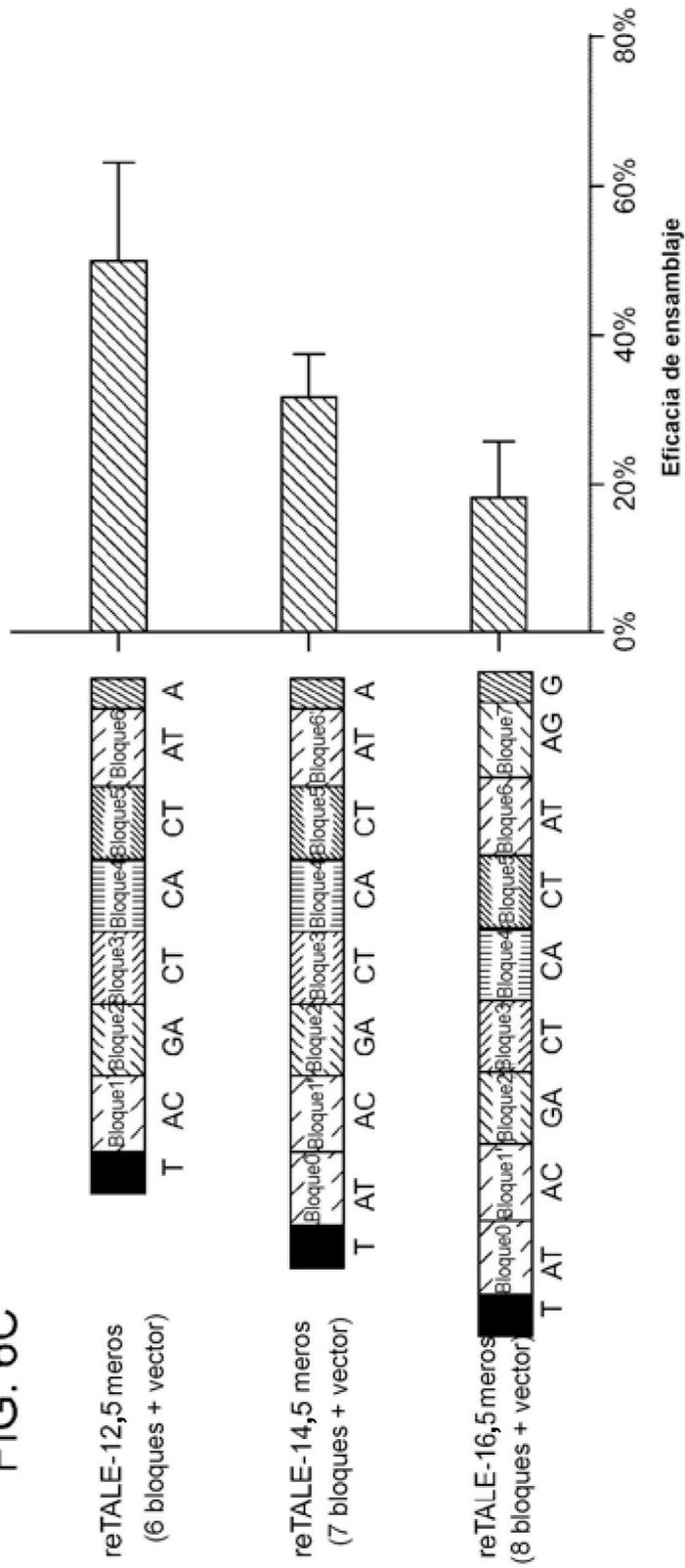


FIG. 7A

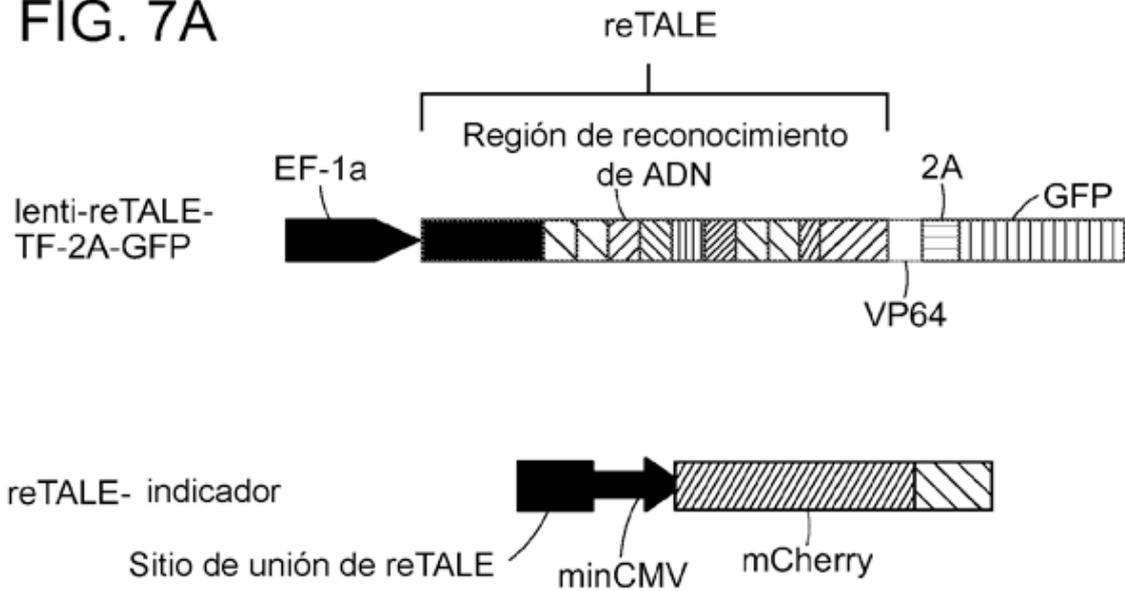


FIG. 7B

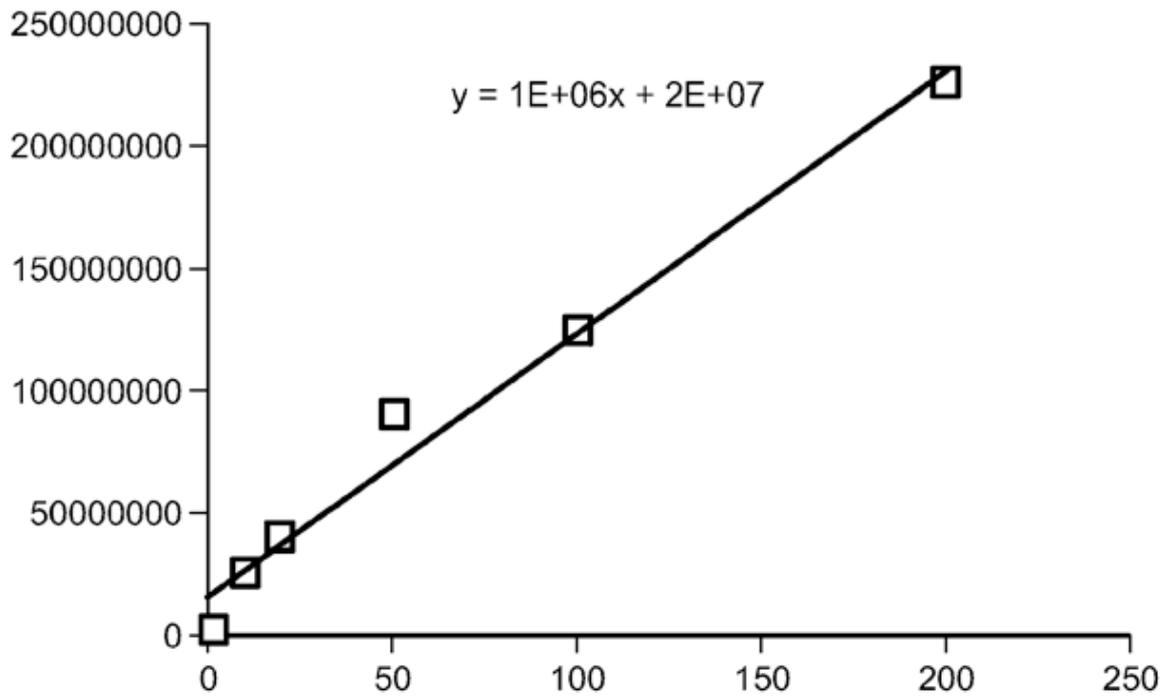


FIG. 7C

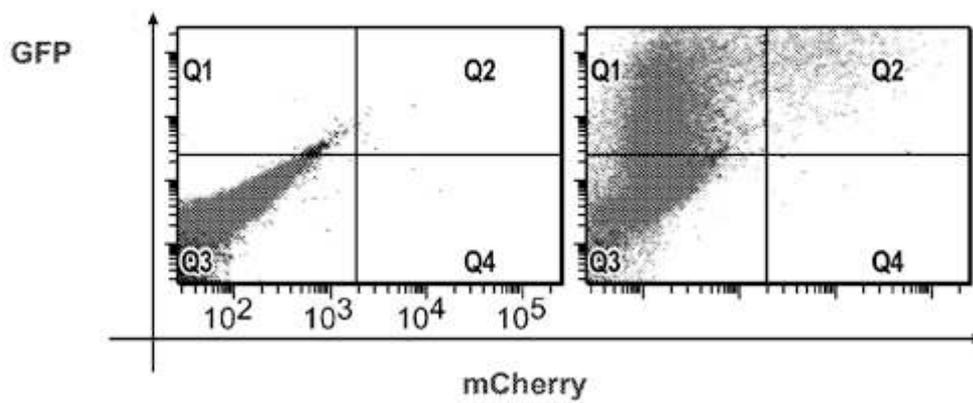
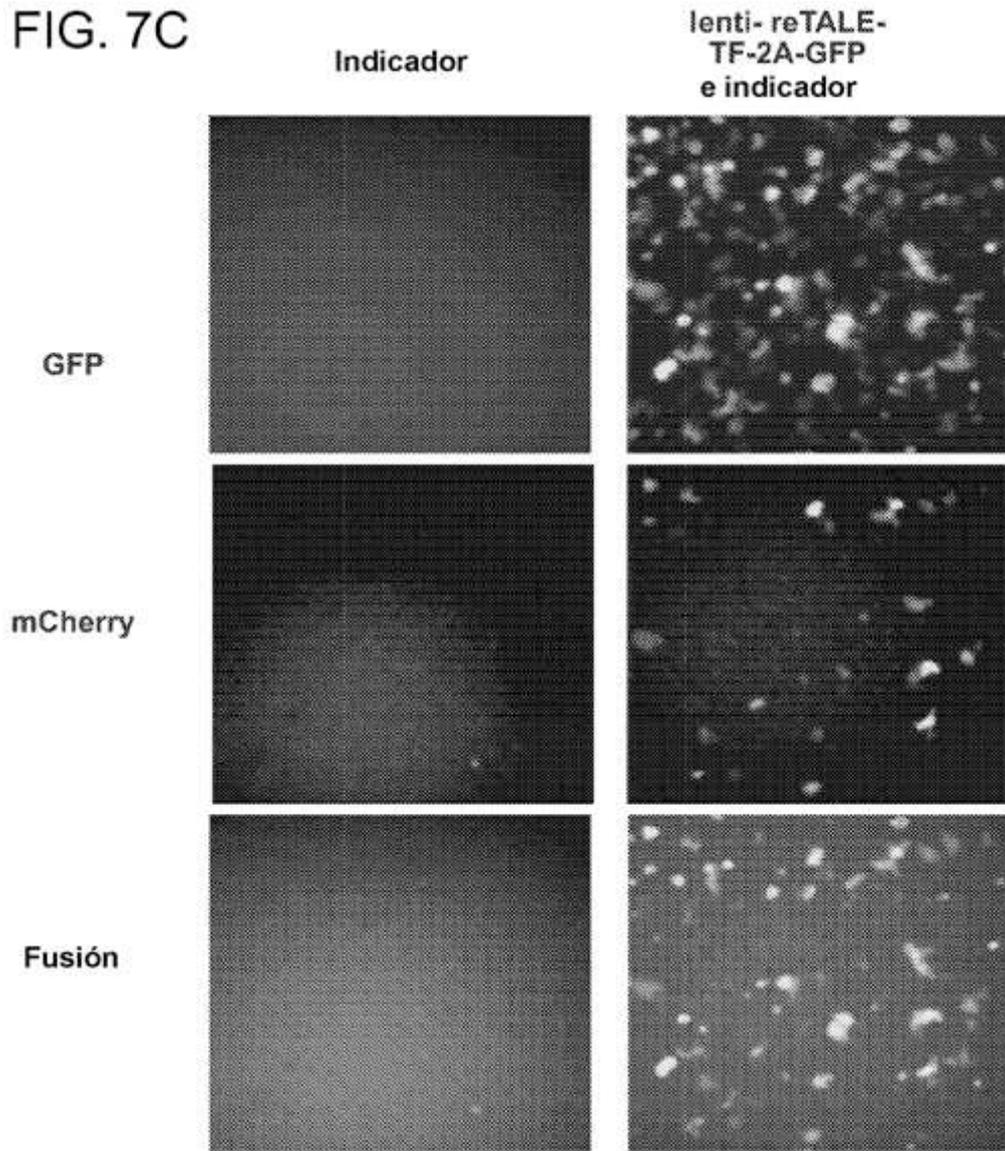
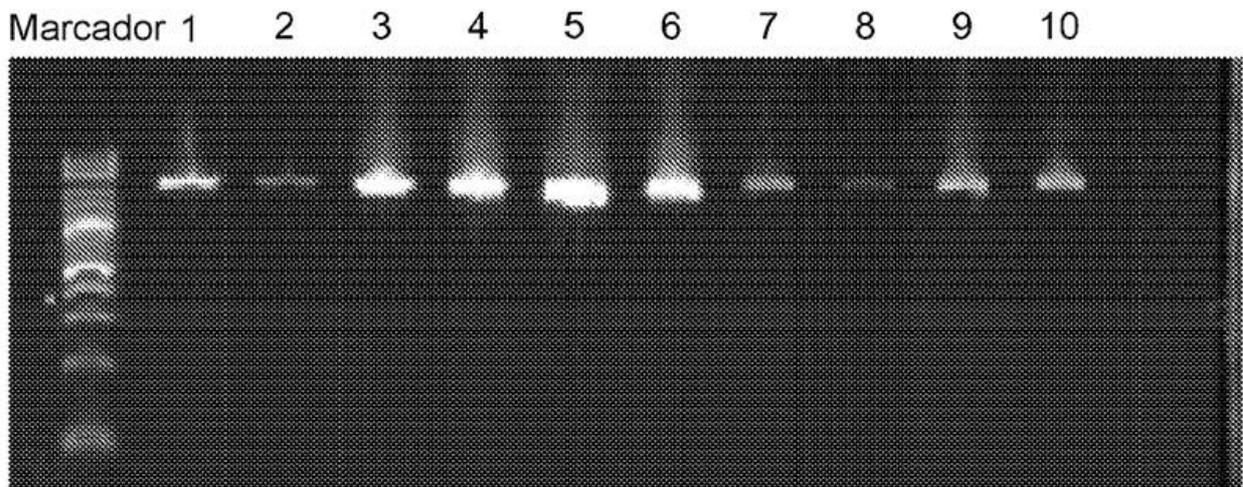


FIG. 7D



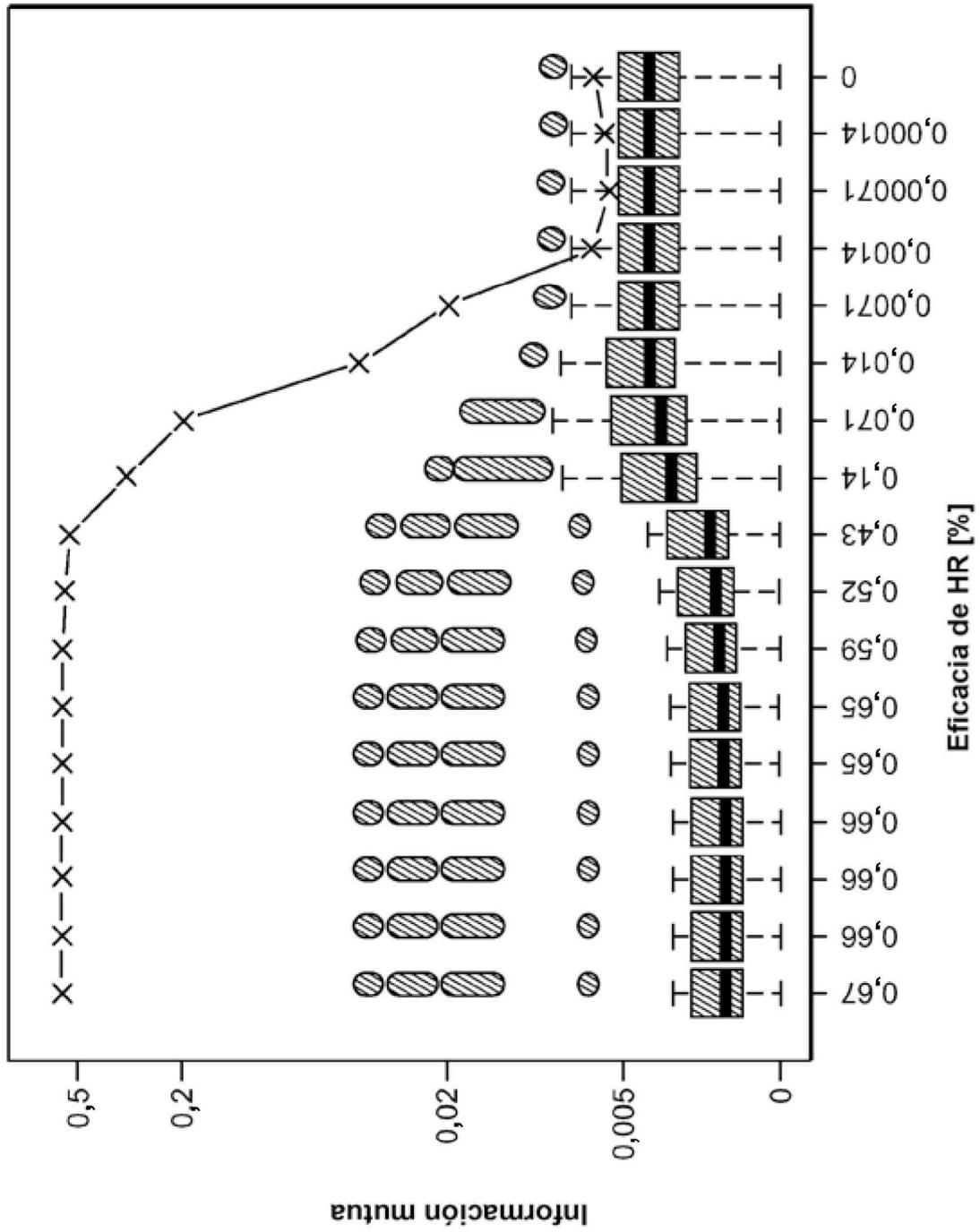


FIG. 8A

FIG. 8B

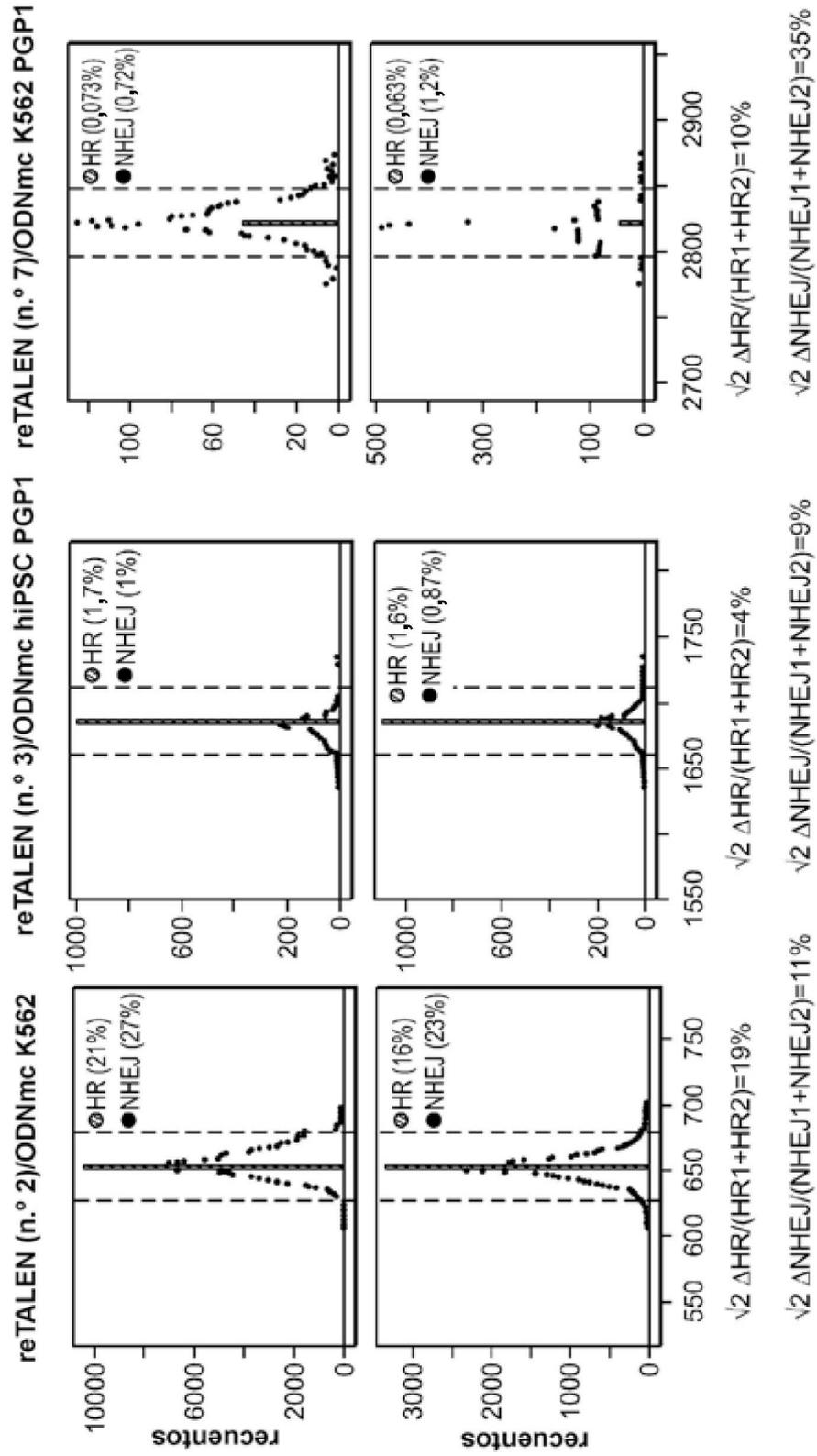


FIG. 9A

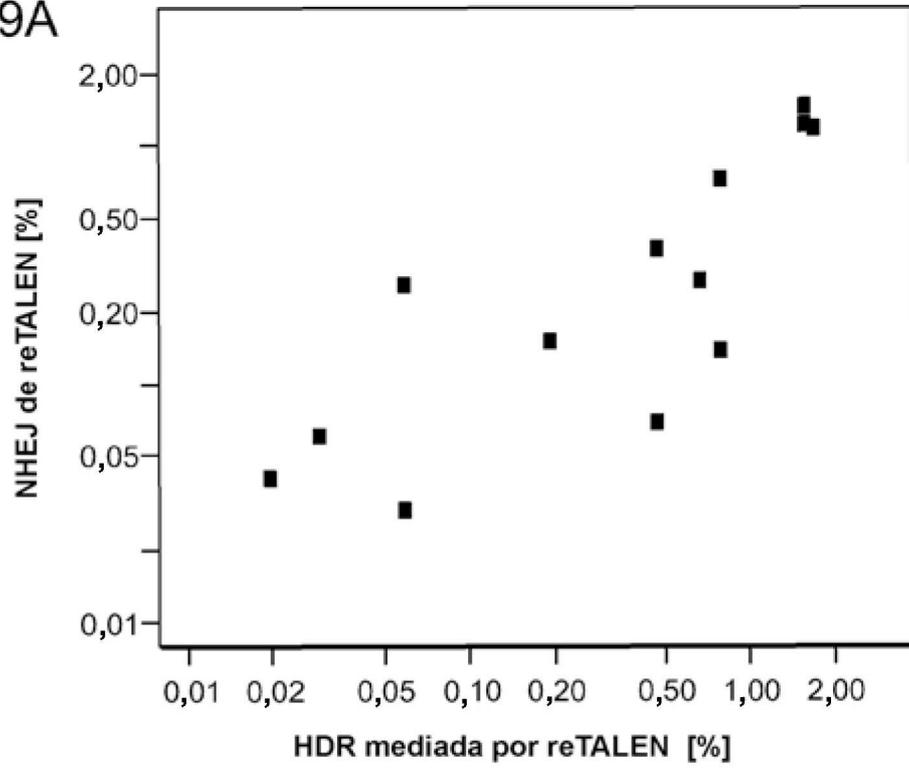


FIG. 9B

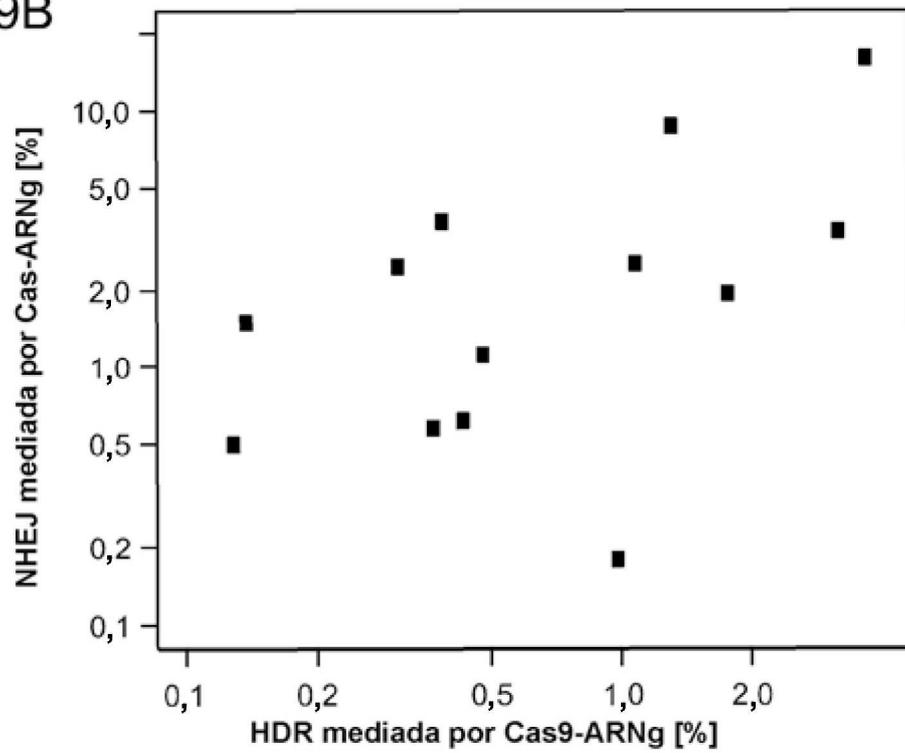


FIG. 9C

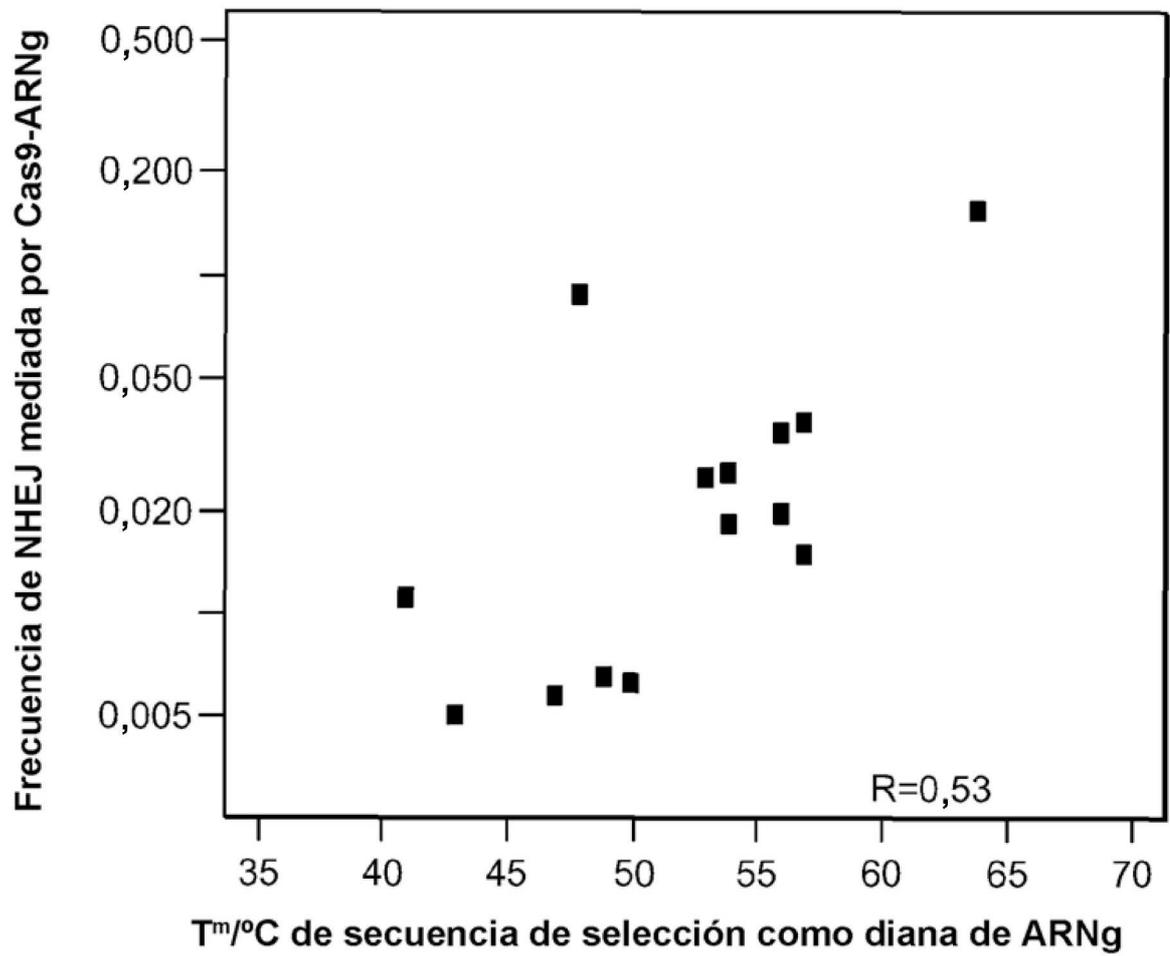


FIG. 10

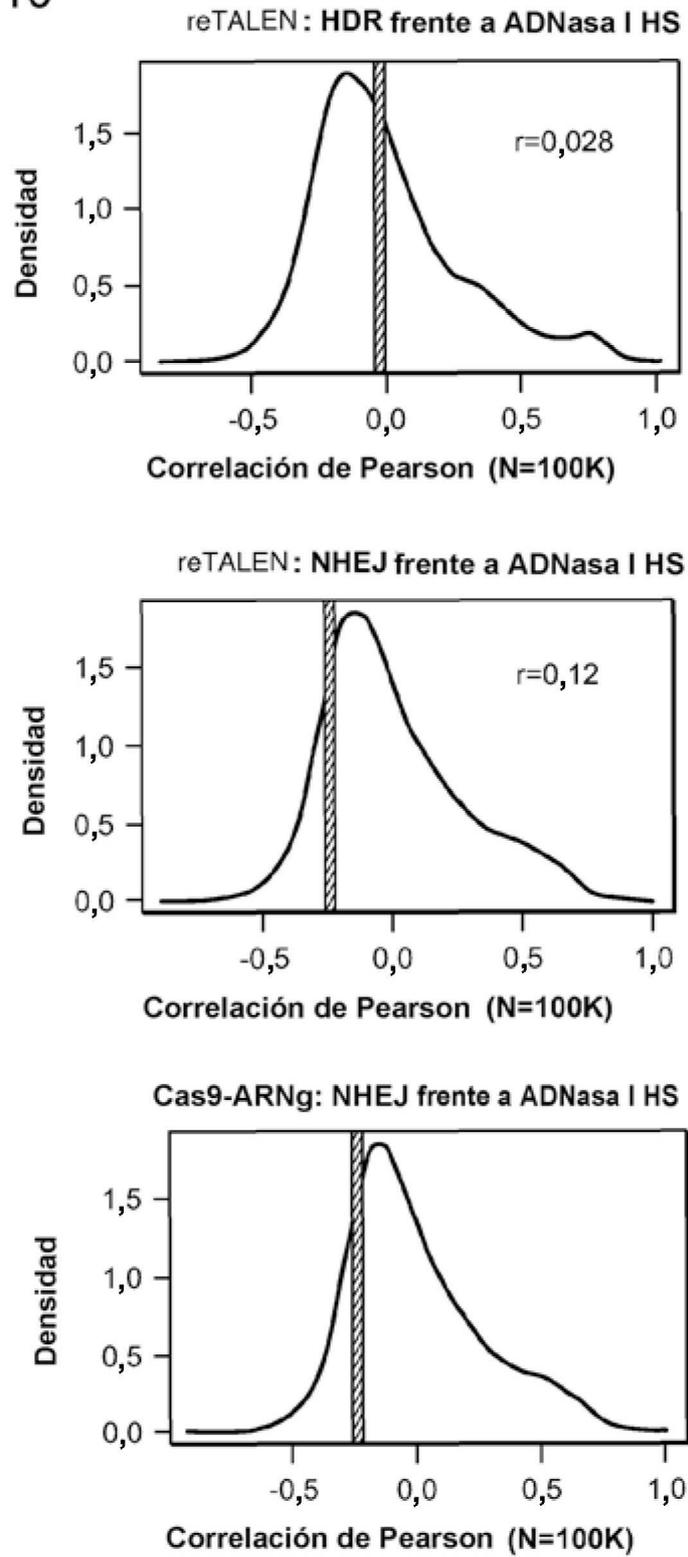


FIG. 11A

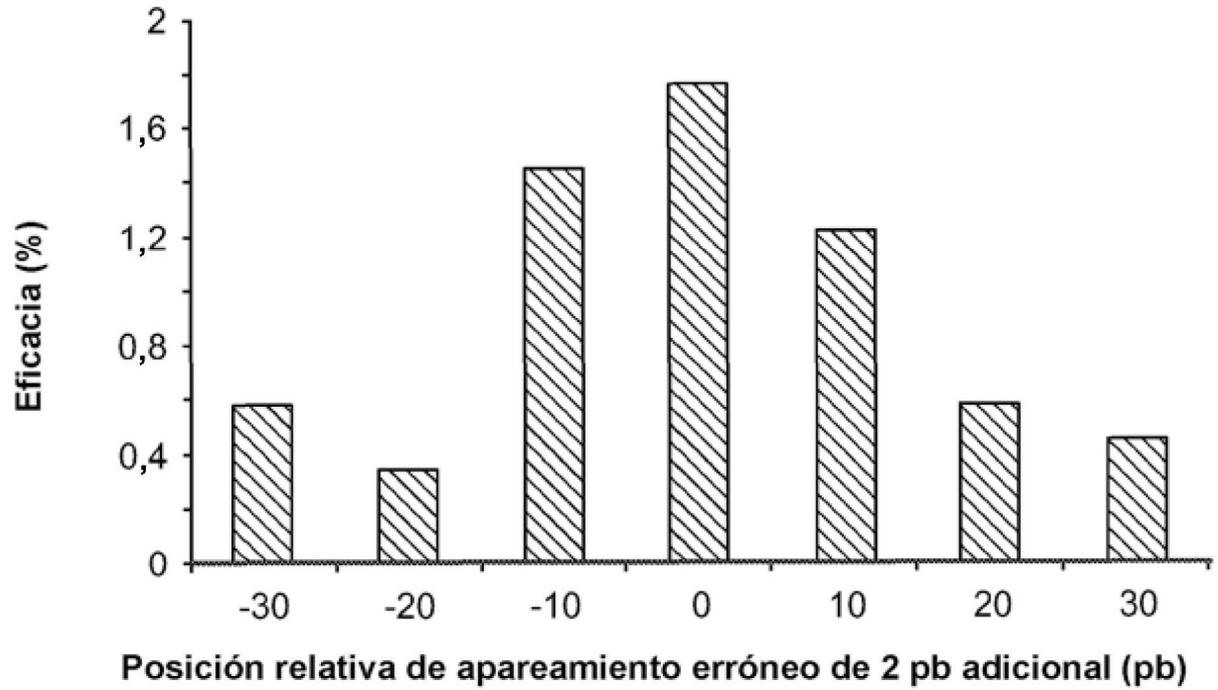


FIG. 11B

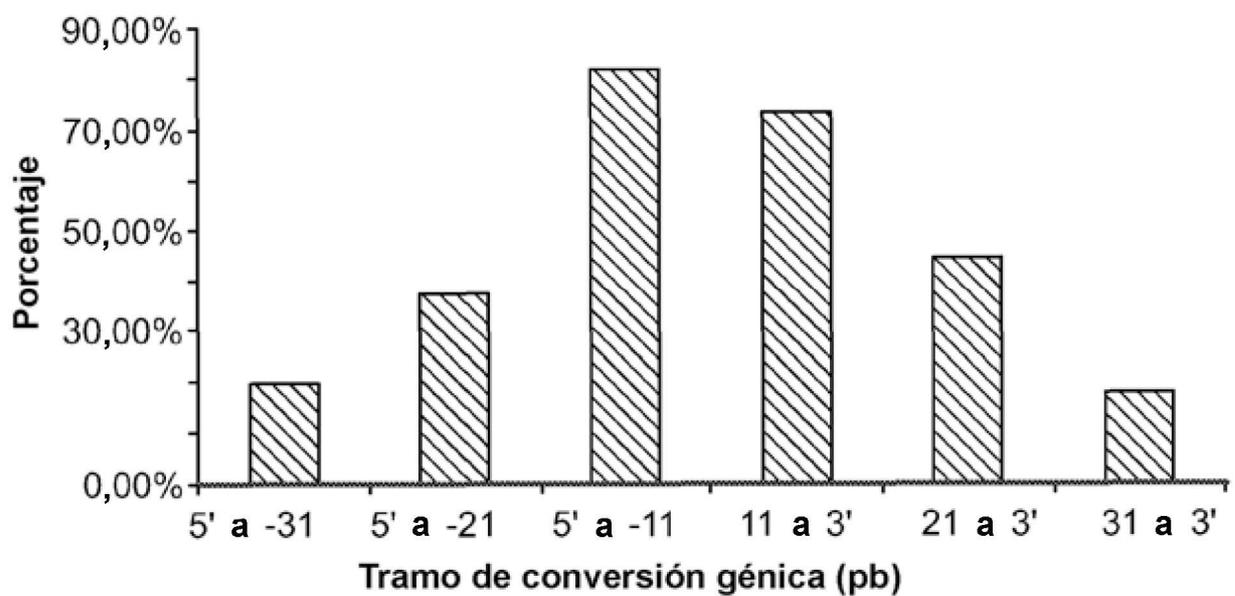


FIG. 11C

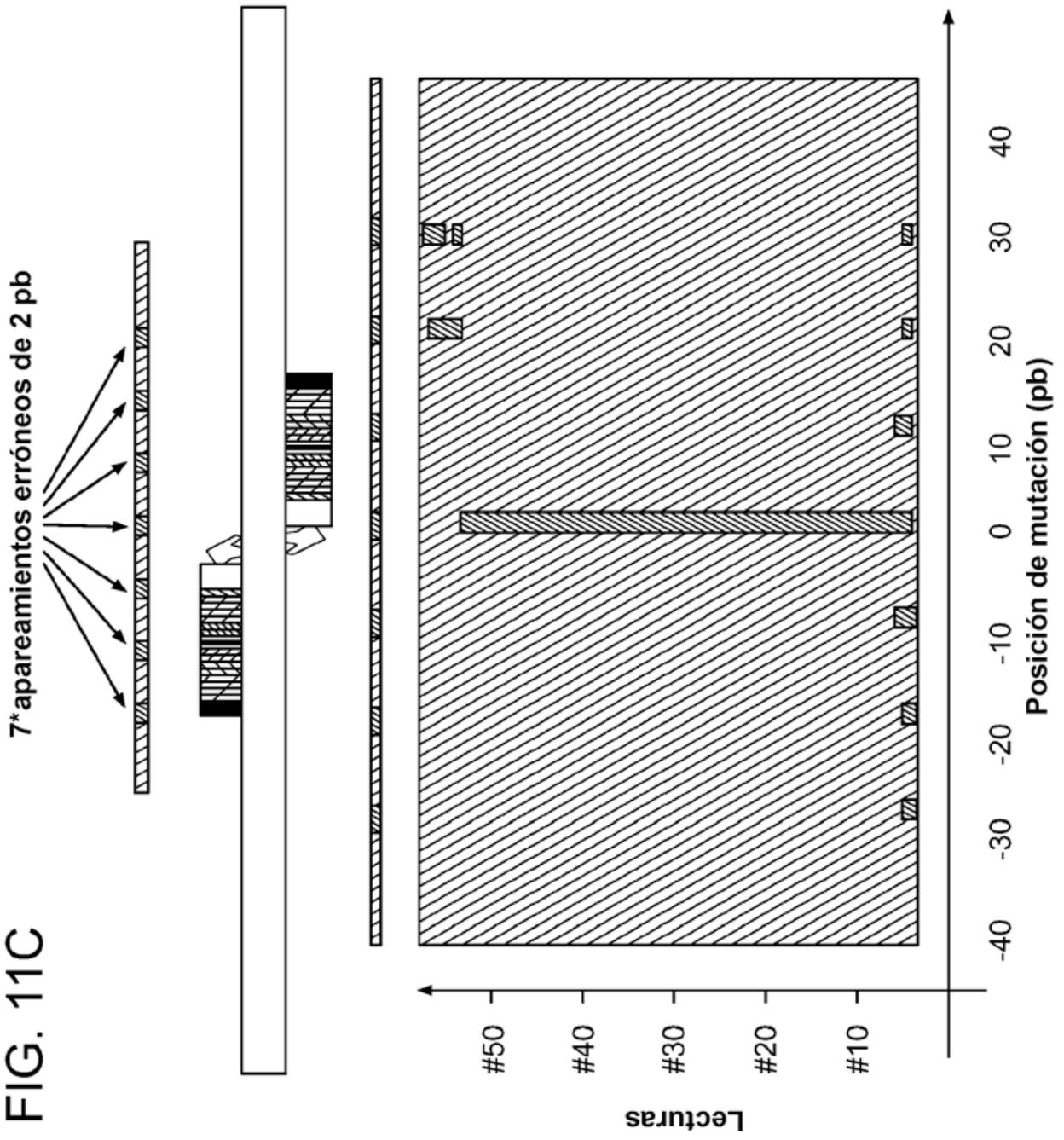


FIG. 12

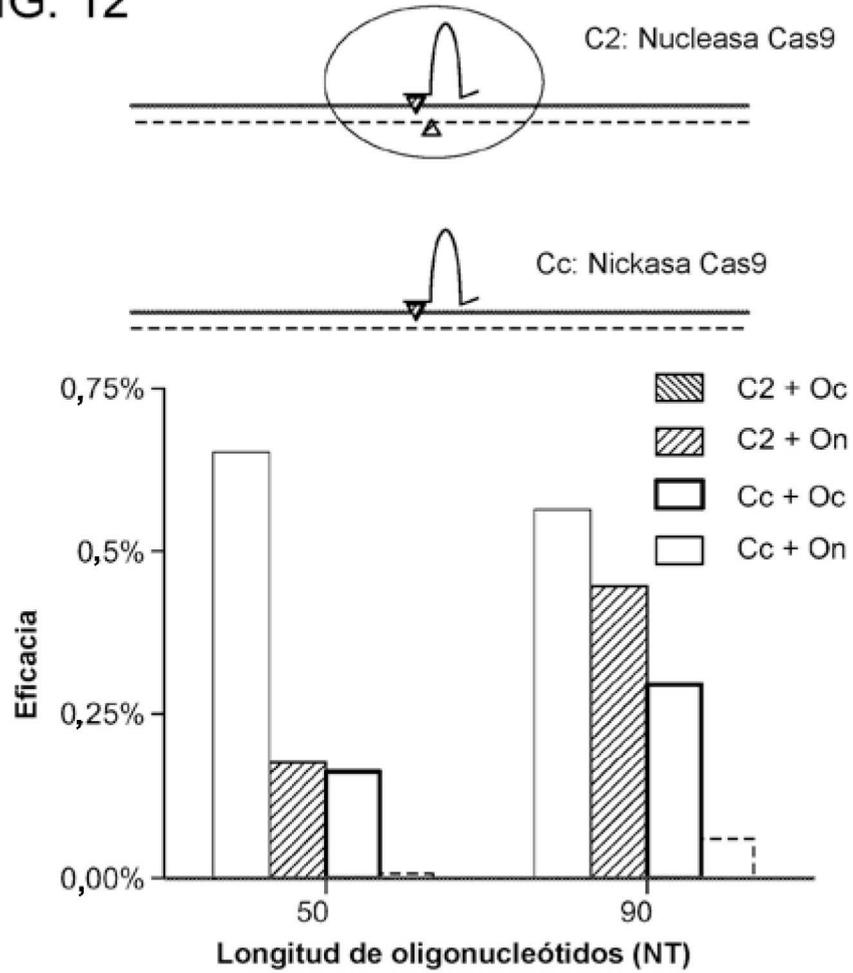
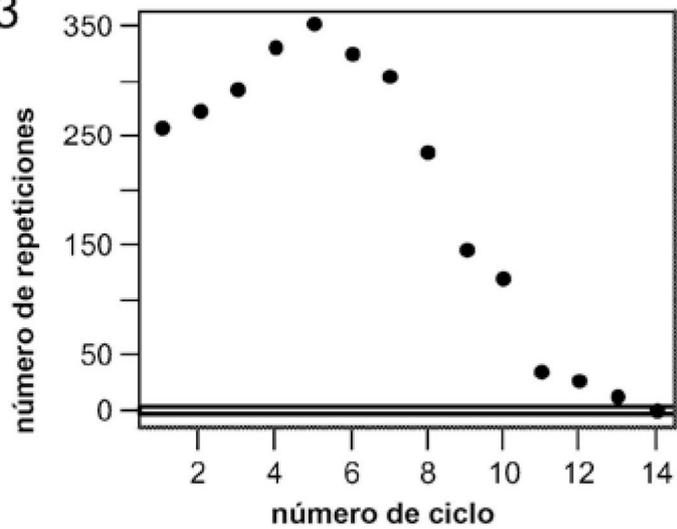


FIG. 13



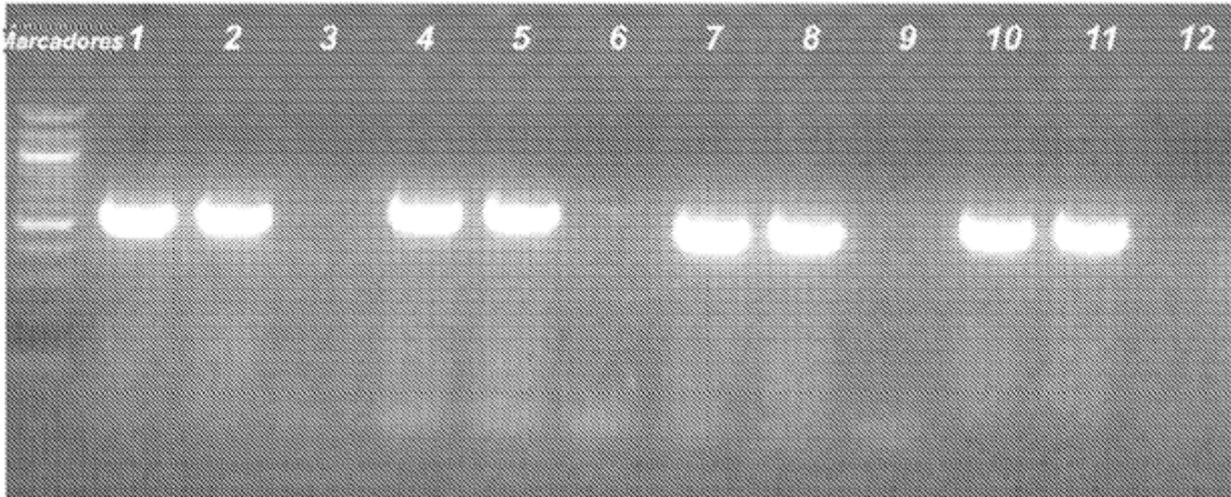


FIG. 14

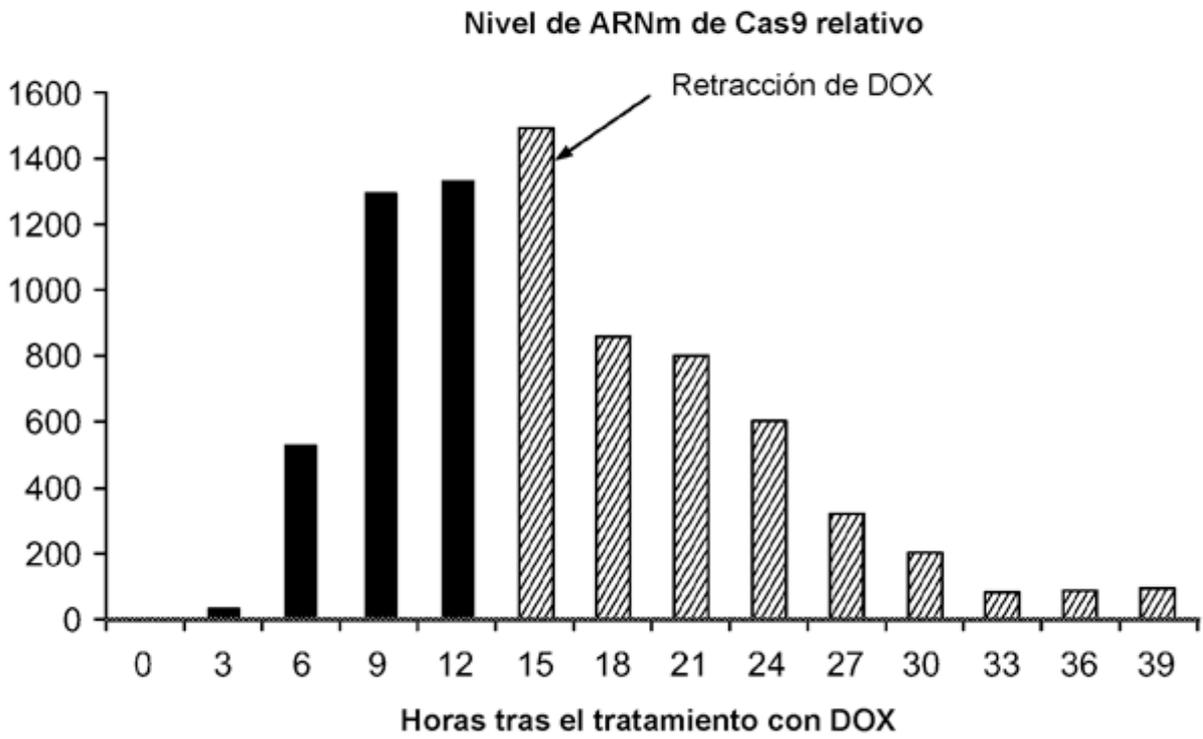


FIG. 15

FIG. 16

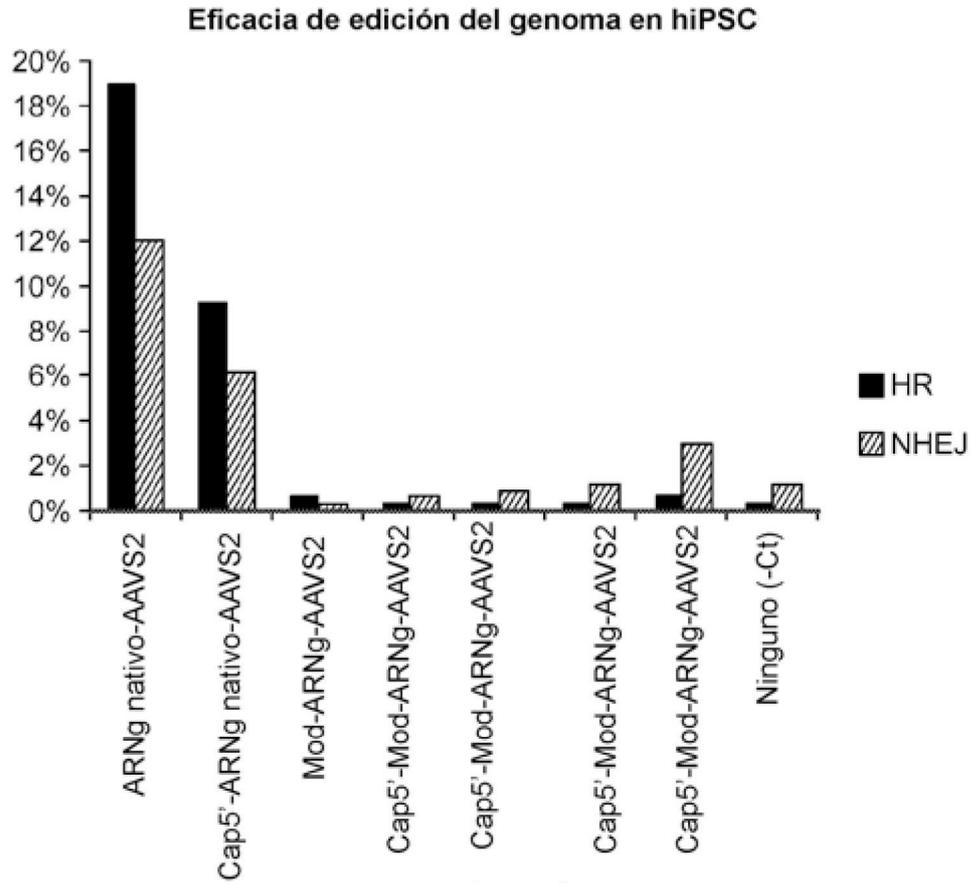


FIG. 17

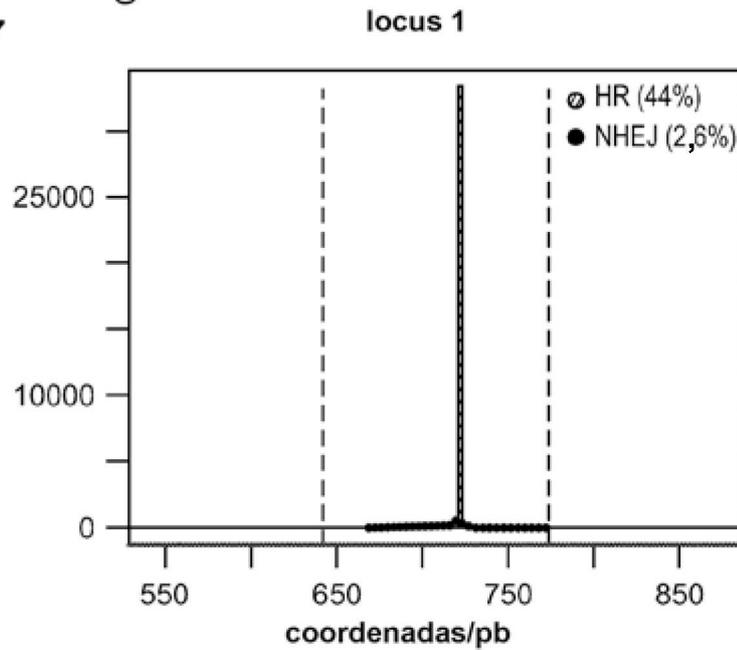
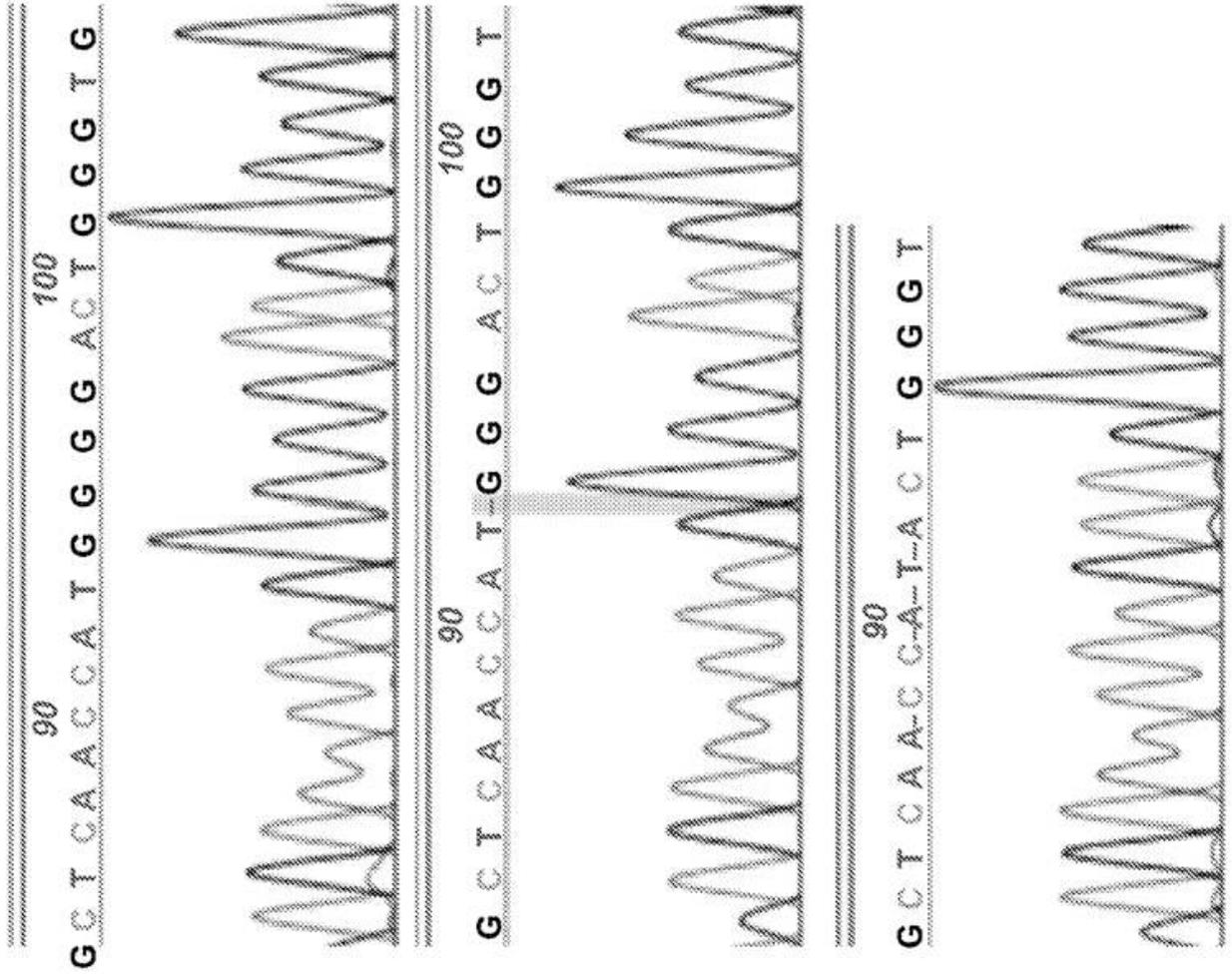


FIG. 18



PGP1-iPS-NT

PGP1-iPS-BTHH

PGP1-iPS-NHEJ

FIG. 19

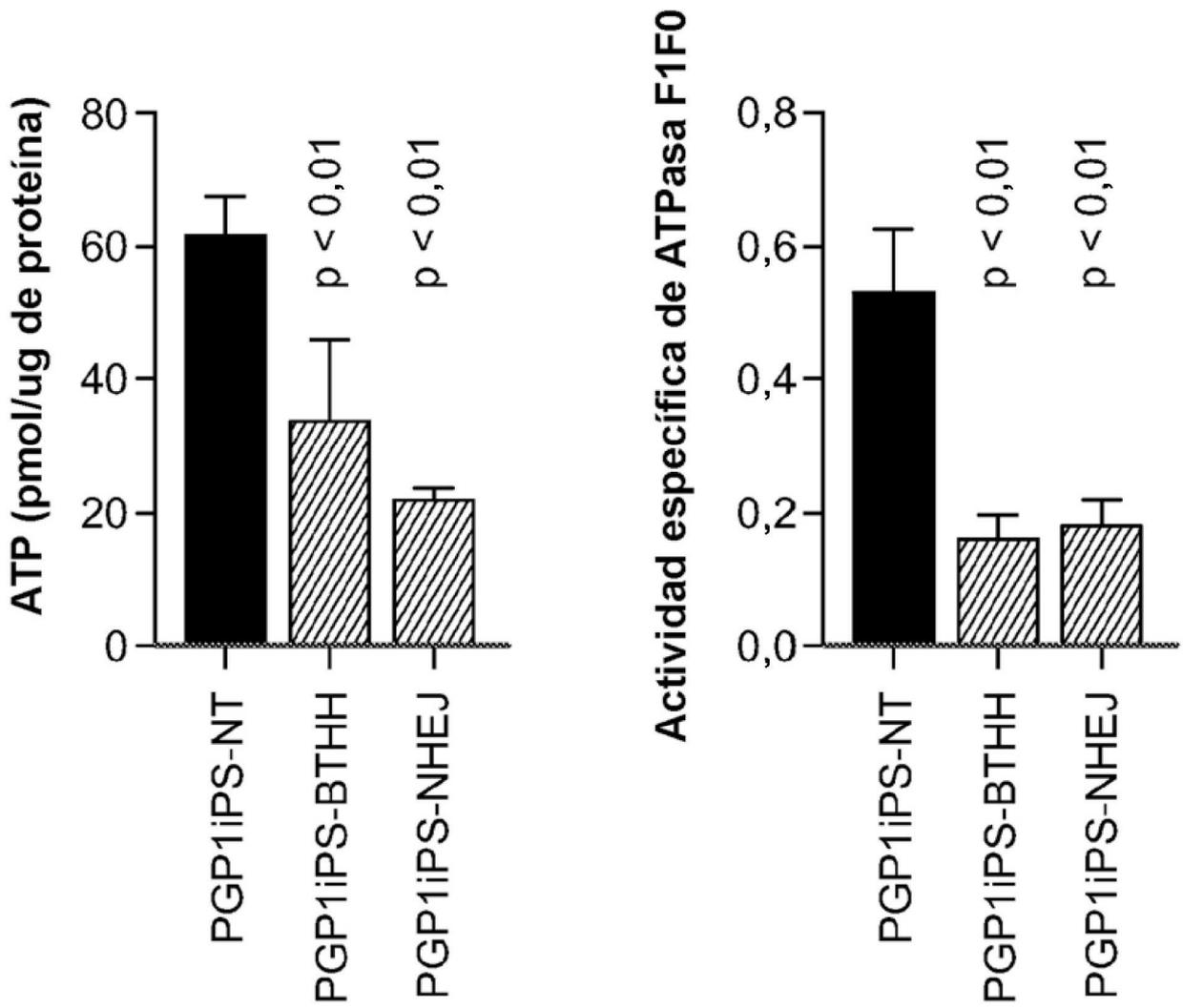


FIG. 20A

Secuencia de estructura principal de re-TALEN

(púrpura: re-TALE-N; rojo: *sitio SapI*; verde: monómero 0,5; azul: re-TALEN-C; naranja: Fok I)

ATGTCGGACCCGGCTCCCTTCCCCACCCCGCACCCAGCCAGCGTTTTCGGCCGAC
TCGTTCAGACCTGCTTAGGCAGTTCGACCCCTCACTGTTTAACACATCGTTGTTCG
ACTCCCTCCCTCCGTTTGGGGCCACCATACGGAGGCCACCCGGGGAGTGGGAT
GAGTGCACTCGGGATTGAGAGCTGCGGATGCACCCACCCCAACCATGCCGGTGGC
CGTACCGCTGCCCGACCCGAGGGCGAAGCCCGCACCAAGCGGAGGGCAGCCG
AACCGTCCGACGAAGCCCGCAGCGCAAGTAGATTGAGAACTTTGGGATATTCA
CAGCAGCAGGAAAGATCAAGCCCAAAGTGAGTGCAGTCGACAGTCGCCGAGCATC
ACGAAGCGTGGTGGTCAATGGGTTTACACATGCCCACATCGTAGCCTTGTCCGAGC
ACCCTGCAGCCCTTGGCACGGTCGCCGTCAGTACCAGGACATGATTCGGCGGTTGC
CGGAAGCCACATGAGCGGATCGTGGTGTGGGAAACAGTGGAGCGGAGCCCG
AGCCTTGAGCCCTGTTGACGGTCGCGGAGAGCTGAGAGGCCCTCCCCTCAGC
TGGACCGGCCAGTGTGTAAGATCGCGAAGCGGGAGGAGTCACTCGGGTCCGAG
GCGGTGCACGCGTGGCGCAATGCGCTCACGGGAGCACCCCTCAACAGTTCACGTG
 ACAGAGACCGGGCCGCAATTAGGCACCCAGGCTTTACACTTTATGCTTCCGGCTCG
 TATAATGTGTGATTTGAGTTAGGATCCGTCGAGATTTTCAGGAGCTAAGGAAGCT
 AAAATGGAGAAAAAATCACTGGATAATACCACCGTTGATATAATCCCAATGGCATCG
 TAAAGAACAATTTGAGGCATTTCAAGTCAAGTCAATGTAACCTATAACAGACCCT
 TCAGCTGGATAATTACGGCCTTTTAAAGACCGTAAAGAAAAAATAAGCACAAAGTTTA
 TCCGGCCTTATTACACATTCCTGGCCGCTGATGAATGCTCATCCGGAAATTCGGTATG
 GCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCCTTGTACACCCGTT

FIG. 20B

TTCCATGAGCAAAC TGAAACGTTTTCATCGCTCTGGAGTGAATACCCACGACGATTC
 CGGCAGTTCTACACATATATTCGCAAGATGTGGCGTTACGGTGAAAACCTGGCC
 TATTTCCCTAAAGGTTTATGAGAAATATGTTTTCGTCCTCAGCCAATCCCTGGGTGA
 GTTTCACCAGTTTGATTTAAACGTGGCCAATATGGACAACCTTTCGCCCCCGTTTT
 CACCATGGGCAAATATATACGCAAGCGACAAGTGCTGATGCCGCTGGCGATTC
 AGGTTCAATCATGCCGTTTGATGGCTTCCATGTCCGAGAAATGCTTAATGAATTAC
 AACAGTACTGCCGATGAGTGGCAGGGCGGTAAGATCTGGATCCGGCTTACTA
 AAAGCCAGATAACAGTATGCCGTTTTCGCGCTGATTTTTCGGGTAAGAATAAT
 ACTGATATGTATACCCGAAGTATGTCAAAGAGGATGCTATGAAGCAGCGTATT
 ACAGTGACAGTTGACAGCGACAGCTATCAGTTGCTCAAGGCATATATGATGTCAATA
 TCTCCGGTCTGGTAAGCACAAACCATGCAGAAATGAAGCCCGTCTCGGTGCCGAAC
 GCTGAAAGCGGAAAATCAGGAAGGATGGCTGAGGTCGCCCGTTTATTGAAAATG
 AACGGCTCTTTTGCTGACGAGAACAGGGCTGGTGAAATGCAGTTTAAAGTTTACAC
 CTATAAAGAGAGAGCCGTTATCGTCTGTGTGGATGTACAGAGTGATATTATGA
 CAGCCCGGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCGTGTCAGATA
 AAGTCTCCCGTGAACTTTACCCGGTGGTGCAATACGGGGATGAAAGCTGGCGCATGA
 TGACCACCGATATGGCCAGTGTGCCGCTCCGTTATCGGGGAAGAAGTGGCTGATC
 TCAGCCACCGGAAAATGACATCAAACAACGCCATTAACTGATGTTCTGGGGAATA
 TAAATGTCAGGCTCCCTTATACACAGCCAGTCTGCAGGTCGACGGTCTC**GCTCTTCG**
 AAGGTTACTTCCCGTCCCTGTCAAGCCACGGCCCTCACTCCAGAGCAAGTGGTTGC
 GATCGCTTCAAACAACGGTGGGAAGACCTGCCCTGGAAATCAATCGTGGCCAGCTTC

FIG. 20C

GAGGCCGACCCCGGCTGGCCGCACTCACTAATGATCATCTTGTAGCGCTGGCCCTG
CCTCGCGGACGACCCCGCCTTGATGCCGTGAAGAAGGGCTCCCGCACGCCGCCCTG
CATTGATTAAGCGGACCAACAGAAGGATTCCCGAGAGGACATCACATCGAGTGGCA
GGTCCCAACTCGTGAAGAGTGAACCTTGAGGAGAAAAAGTCGGAGCTGCGGCACAAA
ATTGAAATACGTACCGCATGAATACATCGAACTTATCGAAATTGCTAGGAACTCGAC
TCAAGACAGAAATCCTTGAGATGAAGGTAATGGAGTCTTTATGAAGGTTTATGGATA
CCGAGGGAAGCATCTCGGTGGATCACGAAAACCCGACGGAGCAATCTATACGGTGG
GGAGCCCGATTGATTACGGAGTGATCGTCGACACGAAAGCCTACAGCGGTGGGTAC
AATCTTCCCATCGGCAGGCAGATGAGATGCAACGTTATGTCGAAGAAAATCAGAC
CAGGAACAAACACATCAAATCCAAAATGAGTGGTGGAAAAGTGATCCCTTCATCAGTGA
CCGAGTTTAAGTTTTTGTTTTGTTCTCTGGGCATTTCAAAGGCAACTATAAGGCCCAGCT
CACACGTTGAATCACATTAACGAACTGCAAATGGTGCAGTTTGTCCGTAGAGGAACT
GCTCATTTGGTGAGAAAATGATCAAAGCGGGAACCTGTGACACTGGAAGAAGTCAGAC
GCAAGTTTAACAAATGGCGAGATCAAATTTCCCGC

Secuencia de estructura principal de re-TALE-TF

(púrpura: re-TALE-N; rojo: sitio SapI; verde: monómero 0,5; azul: re-TALEN-C; naranja: NLS-VP64.)

2A-GFP está resaltado en verde.

ATGTCGGGACCCGGCTCCCTTCCCAACCCGACCCAGCCAGCGTTTTCGGCCGAC
TCGTTCTCAGACCCTGCTTAGGCAGTTCGACCCCTCACTGTTTAAACACATCGTTGTTCCG

FIG. 20D

ACTCCCTTCCGTTTGGGGCGCACCATACGGAGGGGCCACCGGGGAGTGGGAT
 GAGGTGAGTCGGGATTGAGAGCTGCCGATGCACCAACCCCAACCATGCCGGGTGGC
 CGTCAACCGTCCCGACCCGAGGGCGAAGCCCGCACCAAGGCGGAGGGCAGCGC
 AACCGTCCGACGCAAGCCCCGAGCGCAAGTAGATTGAGAACTTTGGGATATTCA
 CAGCAGCAGCAGGAAAGATCAAGCCCCAAGTGAGGTCGACAGTCGCCGCAGCATC
 ACGAAGCGCTGGTGGTCAATGGGTTTACACATGCCCACATCGTAGCCTTGTCCGACG
 ACCCTGCAGCCCTTGGCACCGGTCGCCGTCGAAGTACCAGGACATGATGCCGGCGTTGC
 CGGAAGCCACACATGAGGCCGATCGTCGGTGTGGGGAACAGTGGAGCGGAGCCCCG
 AGCGCTTGAGGCCCTGTTGACGGTCGCCGGGAGAGCTGAGAGGGCTCCCCCTTCAGC
 TGGACACGGGCCAGTTGCTGAAGATCGGGAAGCGGGGAGGAGTCACGGCGGTCGAG
 GCGGTGCACGGTGGCGCAATGCGCTCACGGGAGCACCCCTCAACAGTTCACGCTG
 ACAGAGACCGGGCCGCAATTAGGCACCCAGGCTTTACACTTTATGCTTCCGGCTCG
 TATAATGTGTGATTGAGTTAGGATTAGGATCCGTCGAGATTTTCAGGAGCTAAGGAAGCT
 AAAATGGAGAAAAAATCACTGGATATACCACCCTTGATATATCCCAATGGCATCG
 TAAAGAACATTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGT
 TCAGCTGGATATTACGGCCTTTAAAGACCCTAAAGAAAAAATAGCACACAAGTTTAA
 TCCGGCCTTTATTCACATCTTGCCCGCTGATGAATGCTCATCCGGAATCCCGTATG
 GCAATGAAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCCTTGTACACCCGTT
 TTCCATGAGCAAACTGAAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCC
 CGGCAGTTTCTACACATAATTCGCAAGATGTGGCGTGTACGGTGAAAACCTGGCC
 TATTTCCCTAAAAGGGTTTATTGAGAATAATGTTTTTCGTCTCAGCCAATCCCTGGGTGA
 GTTTCACCAGTTTGTATTTAAACGTGGCCAAATATGGACAACCTTCTTCGCCCCCGTTTT

FIG. 20E

CACCATGGGCAAAATATTATACGCAAGGCGACAAAGGTGCTGATGCCGCTGGCGATTCC
 AGGTTCAATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAAATGCTTAAATGAATTAC
 AACAGTACTGCGATGAGTGGCAGGGCGGTAAGATCTGGATCCGGCTTACTA
 AAAGCCAGATAACAGTATGCGTATTTGCGCGCTGATTTTTCGGGTAAGAAATATAT
 ACTGATATGTATACCCGAAGTATGTCAAAAAGAGGTATGCTATGAAGCAGCGTATT
 ACAGTGACAGTTGACAGCGCACAGCTATCAGTTGCTCAAGGCATATATGATGTCAATA
 TCTCCGGTCTGGTAAGCACAAACCATGCAGAAATGAAGCCCGTCTGCTGCCGAAC
 GCTGGAAAGCGGAAAATCAGGAAGGATGGCTGAGGTCGCCCGTTTATTGAAAATG
 AACGGCTCTTTGCTGACGAGAACAGGGGCTGGTGAATGCAGTTTAAAGTTTACAC
 CTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTAATTGA
 CACGCCGGGACGGATGGTGATCCCCCTGGCCAGTGCCACGCTGCTGTCAGATA
 AAGTCTCCCGTGAACTTTACCCGGTGGTGATATCGGGGATGAAAAGCTGGCGCATGA
 TGACCAACCGATATGGCCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATC
 TCAGCCACCGGAAAATGACATCAAAAACGCCATTAAACCTGATGTTCTGGGGAATA
 TAAATGTCAGGCTCCCTTATACACAGCCAGTCTGCAGGTCGACGGTCTC**GCTCTTCG**
AAGTTACTTCCCGTCTCTGTCAAGCGCACGGCCTCACTCCAGAGCAAGTGGTTGC
GATCGCTTCAAAACAACGGTGAAGACCTGCCCTGGAATCAATCGTGGCCCCAGCTTTC
GAGCCGGACCCCGCTGGCCGCACTCACTAATGATCATCTTGTAGCGCTGGCCTG
CCTCGGCGGACCCCGCTTGGATGCGGTGAAGAAGGGGCTCCCGCACCGCCCTG
CATTGATTAAGCGGACCAACAGAAGGATTCGCCGAGAGGACATAGCCCCCAAGAAGAA
GAGAAAGGTGGAGGCCAGCGGTTCCGGACGGGCTGACGCATTTGGACGATTTTGATC

FIG. 20F

TGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCCGG
ATGCCCTTGATGACTTTGACCTCGACATGCTCGCAGTGACGCCCTTGATGATTTCCG
 \A\C\T\G\G\A\C\A\T\G\C\T\G\A\T\T\A\A\C\T\C\T\A\G\A\G\G\C\A\G\T\G\G\A\G\A\G\G\C\A\G\A\G\A\G\T\C\T\G\C\T\A
 \A\C\A\T\G\C\G\G\T\G\A\C\G\T\C\G\A\G\G\A\G\A\T\C\C\T\G\C\C\A\G\T\G\A\G\C\A\G\G\C\A\G\G\C\G\A\G\G\C\G\A\G\G\C\T\G\T
 \C\A\G\G\G\G\T\G\G\T\G\C\C\A\T\C\T\G\G\T\G\A\G\C\T\G\A\G\C\T\G\A\G\C\G\G\T\A\A\C\G\G\C\A\C\A\G\T
 \T\C\A\G\C\G\T\G\T\C\G\G\G\C\G\A\G\G\G\C\G\A\T\G\C\C\A\C\T\A\C\G\G\C\A\A\G\C\T\G\A\C\T\G\A\C\T\G\A\G
 \M\T\C\A\T\C\T\G\C\A\C\A\C\G\G\A\G\C\T\G\C\C\G\T\G\C\C\T\G\C\C\A\C\C\T\G\T\G\A\C\G\A\C\C\T\G
 \A\C\C\T\A\C\G\G\C\G\T\G\C\A\G\T\G\C\T\C\A\G\C\G\C\T\A\C\C\G\A\C\A\T\G\A\G\C\A\G\C\A\C\G\A\C\T\T\C
 \M\T\C\A\A\G\T\C\G\C\A\T\G\C\C\G\A\A\G\G\C\T\A\G\T\C\A\G\G\G\A\C\A\G\G\G\A\C\A\T\T\C\T\T\A\A\G\G\A\C
 \G\A\G\G\C\A\A\C\T\A\C\A\A\G\A\C\C\G\C\G\G\A\G\G\T\G\A\A\G\T\C\G\A\G\G\C\A\C\A\C\C\T\G\G\T\G\A\A
 \C\G\C\A\T\C\G\A\G\T\G\A\G\G\C\A\T\C\G\A\C\T\T\C\A\G\G\A\G\G\A\C\G\G\C\A\A\C\A\T\C\T\G\G\G\C\A\C\A
 \A\G\G\T\G\G\A\G\T\A\C\A\A\C\T\A\A\C\A\G\G\C\A\A\C\A\A\C\G\T\C\T\A\T\A\T\A\T\G\G\C\G\A\G\A\G\A\G\A\G
 \A\A\G\G\C\A\T\C\A\A\G\T\G\A\A\C\T\T\C\A\A\G\A\T\C\G\C\A\C\A\A\C\A\A\C\A\A\C\A\A\C\A\A\C\A\A\C\A
 \G\C\T\C\G\C\G\A\C\A\C\T\A\C\A\G\C\A\G\A\C\A\A\C\C\A\T\C\G\G\C\G\A\G\G\G\C\C\G\T\G\C\T\G\C\T\G\C
 \C\G\G\C\A\A\C\A\C\A\T\A\C\T\G\A\G\C\A\C\C\A\G\T\C\G\G\C\T\G\A\G\C\A\A\A\G\A\A\G\A\A\C\C\A\A\C\G\A\G\A\G
 \C\G\G\A\T\C\A\C\A\T\G\G\T\C\C\T\G\G\T\G\G\A\G\T\T\G\G\T\G\A\C\C\G\G\C\G\G\G\A\T\C\A\C\T\C\T\G\G\C\A\T\G
 \G\A\G\A\G\T\G\T\A\C\A\A\G,