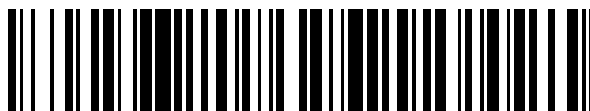


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 565 216**

51 Int. Cl.:

**C07K 14/47** (2006.01)

**G01N 33/58** (2006.01)

**C12N 15/90** (2006.01)

**G01N 33/50** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.04.2011 E 11769489 (3)**

97 Fecha y número de publicación de la concesión europea: **16.12.2015 EP 2558575**

54 Título: **Métodos de generación de proteínas marcadas endógenamente**

30 Prioridad:

**12.01.2011 US 431957 P**

**01.11.2010 US 408856 P**

**07.10.2010 US 390668 P**

**23.07.2010 US 367017 P**

**13.04.2010 US 323698 P**

**13.04.2010 US 323719 P**

**13.04.2010 US 323702 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**01.04.2016**

73 Titular/es:

**SIGMA-ALDRICH CO. LLC (100.0%)**

**3050 Spruce Street**

**St Louis MO 63103, US**

72 Inventor/es:

**MALKOV, DMITRY;**

**ZENSER, NATHAN;**

**VASSAR, DEBORAH;**

**ZHANG, FAN y**

**ZHANG, HONGYI**

74 Agente/Representante:

**ZEA CHECA, Bernabé**

**ES 2 565 216 T3**

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

## DESCRIPCIÓN

Métodos de generación de proteínas marcadas endógenamente

## 5 Referencia cruzada a solicitudes relacionadas

[0001] La presente solicitud reivindica la prioridad de la solicitud provisional de EE.UU. número 61/323.702, presentada el 13 de abril de 2010, la solicitud provisional de EE.UU. número 61/323.719, presentada el 13 de abril de 2010, la solicitud provisional de EE.UU. número 61/323.698, presentada el 13 de abril de 2010, la solicitud provisional de EE.UU. número 61/367.017, presentada el 23 de julio de 2010, la solicitud provisional de EE.UU. número 61/390.668, presentada el 7 de octubre de 2010, la solicitud provisional de EE.UU. número 61/408.856, presentada el 1 de noviembre de 2010, y la solicitud provisional de EE.UU. número 61/431.957, presentada el 12 de enero de 2011, cada una de las cuales se incorpora en el presente documento por referencia en su totalidad.

## 15 Campo de la invención

[0002] La presente divulgación se refiere a métodos de marcaje de proteínas endógenas.

## Antecedentes de la invención

20

[0003] El marcaje de proteínas se usa ampliamente para proporcionar una lectura visual de la proteína de interés en la célula. Entre otros usos, las proteínas marcadas se usan para estudiar la abundancia y la ubicación de las proteínas, la regulación de la transcripción y de la traducción, las modificaciones posteriores a la traducción, las interacciones entre proteínas, el corte y empalme alternativo, la desactivación de ARN y proteínas mediante ARNi y los sitios de unión a los factores de transcripción. Sin embargo, los métodos actuales de expresión de proteínas marcadas en la célula dan lugar a una expresión distorsionada, que no refleja el patrón de expresión de la proteína endógena. Esto se debe a que la expresión de las proteínas marcadas suele basarse en promotores heterólogos para la expresión. Además, algunas proteínas marcadas se expresan ectópicamente a partir de vectores epigenéticos o vectores integrados aleatoriamente en el genoma de la célula y, por lo tanto, no están controladas por las vías de regulación endógenas. Así pues, existe una gran necesidad de un método que pueda dirigir la integración específica en el cromosoma de una célula para producir una proteína marcada controlada por vías de regulación endógenas.

## Resumen de la invención

35

[0004] En un aspecto, la presente divulgación proporciona un método de marcaje de al menos una proteína endógena. El método comprende (a) introducir en una célula (i) al menos una endonucleasa dirigida o un ácido nucleico que codifique una endonucleasa dirigida, uniéndose la endonucleasa dirigida a un sitio diana y siendo capaz de escindir un sitio de escisión de una secuencia cromosómica que codifique la proteína endógena; e (ii) al menos un polinucleótido donante que comprenda una secuencia de marcaje, estando la secuencia de marcaje flanqueada por una secuencia cadena arriba y una secuencia cadena abajo, compartiendo la secuencia cadena arriba y la secuencia cadena abajo una identidad de secuencia sustancial con cualquier lado del sitio de escisión de la secuencia cromosómica; y (b) mantener la célula en condiciones de modo que una rotura de doble cadena introducida en el sitio de escisión por la endonucleasa dirigida se repare mediante un proceso dirigido por la homología de manera que la secuencia de marcaje del polinucleótido donante se integre en fase en la secuencia cromosómica que codifica la proteína endógena, seleccionada entre actina, tubulina, lamina, receptor del factor de crecimiento epidérmico humano (HER2) y proteína del grupo de alta movilidad A (HMGA), en la que se produce una proteína endógena marcada.

[0005] En otro aspecto, la presente divulgación proporciona una célula que comprende al menos una secuencia de marcaje integrada en fase en una secuencia cromosómica que codifica una proteína endógena, de manera que la célula expresa al menos una proteína endógena marcada.

[0006] En otro aspecto más, la presente divulgación proporciona un kit para controlar la ubicación de una proteína endógena. El kit comprende una célula que tiene al menos una secuencia de marcaje integrada en fase en una secuencia cromosómica que codifica una proteína endógena, de modo que la célula expresa al menos una proteína endógena marcada.

[0007] A continuación, se describen con más detalle otros aspectos e iteraciones de la divulgación.

60

## Referencia a las figuras en color

[0008] El archivo de la solicitud contiene al menos una fotografía realizada en color. Las copias de la presente publicación de solicitud de patente con fotografías en color serán proporcionadas por la Oficina bajo petición y pago de la tasa necesaria.

65

**Breve descripción de las figuras**

**[0009]** La **Fig. 1** muestra el diseño de la integración de la secuencia de marcaje en el locus de TUBA1B. (A) Es un esquema que muestra la secuencia cromosómica (SEQ ID NO: 29) en la región diana para la integración de la secuencia de marcaje, los sitios de unión a ZFN (nucleótidos en caja) de la región cromosómica diana, el sitio de corte de ZFN (flecha amarilla) y el sitio de integración de la secuencia de marcaje (flecha verde). (B) Es un esquema que representa la región diana genómica de TUBA1B, que muestra la región codificante (rojo), la región no traducida (azul) y el sitio de corte de ZFN (flecha amarilla). (C) Es un esquema del fragmento de ADN de la región genómica de TUBA1B antes de la integración. (D) Es un esquema del fragmento de ADN de la región genómica de TUBA1B con la secuencia de GFP integrada en fase con la secuencia de codificación de TUBA1B. (E) Es un esquema de la proteína  $\alpha$ -tubulina endógena condensada a la GFP marcadora en el extremo N-terminal creado tras la integración con éxito de la secuencia de marcaje.

**[0010]** La **Fig. 2** muestra el mapa de un plásmido donante que comprende la GFP marcadora flanqueada por las secuencias genómicas de la tubulina.

**[0011]** La **Fig. 3** representa la secuencia de ADN (SEQ ID NO: 4) de la región genómica de TUBA1B en células U2OS, que demuestra que la secuencia de codificación de GFP2 se integró en la región de codificación de la tubulina. El texto subrayado indica la región secuenciada, el texto en negrita indica la secuencia de codificación de GFP2, el texto en cursiva indica el sitio de restricción o el enlazador, y el texto en negrita y mayúscula indica el codón Met para la unión de corte y empalme.

**[0012]** La **Fig. 4** representa la secuencia de ADN (SEQ ID NO: 5) de la región genómica de TUBA1B en células U2OS, que demuestra que la secuencia de codificación de RFP se integró en la región de codificación de la tubulina. El texto subrayado indica la región secuenciada, el texto en negrita indica la secuencia de codificación de RFP, el texto en cursiva indica el sitio de restricción o el enlazador, y el texto en negrita y mayúscula indica el codón Met para la unión de corte y empalme.

**[0013]** La **Fig. 5** presenta el análisis de electroforesis en gel de agarosa de la PCR de unión de 14 clones de células usando cebadores específicos de la integración dirigida de GFP en el locus de TUBA1B. También se muestran los marcadores del tamaño molecular y un control de GFP.

**[0014]** La **Fig. 6** muestra varios ejemplos de imágenes de contraste diferencial de interferencia (DIC) y de microscopía de fluorescencia de clones de células individuales aislados que expresan la proteína de la isoforma de la  $\alpha$ -tubulina 1B endógena marcada con GFP. (A) proteína de la isoforma de la  $\alpha$ -tubulina 1B marcada con GFP en células U2OS; (B) proteína de la isoforma de la  $\alpha$ -tubulina 1B marcada con GFP en células U2OS; (C) proteína de la isoforma de la  $\alpha$ -tubulina 1B marcada con GFP en células U2OS; (D) proteína de la isoforma de la  $\alpha$ -tubulina 1B marcada con GFP en células A549, (E) proteína de la isoforma de la  $\alpha$ -tubulina 1B marcada con GFP en células A549, (F) proteína de la isoforma de la  $\alpha$ -tubulina 1B marcada con GFP en células K562; (G) proteína de la isoforma de la  $\alpha$ -tubulina 1B marcada con GFP en células HEK293; y (H) proteína de la isoforma de la  $\alpha$ -tubulina 1B marcada con GFP en células HEK293T.

**[0015]** La **Fig. 7** muestra el mapa de un plásmido donante que comprende la RFP marcadora flanqueada por las secuencias genómicas de tubulina.

**[0016]** La **Fig. 8** muestra la verificación de la integración de RFP en la región de TUBA1B en la línea celular MCF10a. La integración se comprobó mediante PCR genómica y PCR de la unión usando cebadores de la tubulina. (A) Transferencia de Southern que muestra la presencia de la banda de fusión de RFP/tubulina de 1.945 pb y (B) PCR genómica que muestra la integración positiva de la secuencia de la RFP marcadora en TUBA1B en varios clones (T.I. = integración dirigida). Se usaron la célula MCF10a de tipo silvestre y la línea celular U2SO con la integración de RFP como controles.

**[0017]** La **Fig. 9** representa la secuencia confirmada de la región de TUBA1B en células MCF10a, que demuestra la integración de la secuencia de RFP (SEQ ID NO: 8). El texto subrayado indica la región secuenciada, el texto en negrita indica la secuencia de codificación de GFP2, el texto en cursiva indica el sitio de restricción o el enlazador, y el texto en negrita y mayúscula indica el codón Met para la unión de corte y empalme.

**[0018]** La **Fig. 10** representa la verificación por PCR de la integración de RFP en el locus de TUBA1B de células MCF10a, así como la integración de RFP y GFP en el mismo locus de células U2OS. La banda de tipo silvestre fue de 452 pb y la banda integrada dirigida (T.I.) fue de 1.190 pb.

**[0019]** La **Fig. 11** muestra que las uniones en el sitio de inserción de RFP en clon 5 de MCF10a fueron de los tamaños esperados. El tamaño esperado de la unión de la izquierda es de 453 pb, y el tamaño esperado de la unión de la derecha es de 4.089 pb.

**[0020]** La **Fig. 12** representa la transferencia Western para detectar RFP y la expresión de la tubulina en el clon 5 de MCF10a con la tubulina marcada con RFP.

**[0021]** La **Fig. 13** demuestra que > 99 % de las células MCF10a de tipo silvestre carecen de fluorescencia de color rojo, mientras que > 99 % de las células del clon 5 de MCF10a que comprenden tubulina marcada con RFP resultaron tener fluorescencia de color rojo.

**[0022]** La **Fig. 14** representa la estabilidad fenotípica de las células MCF10a transfectadas que comprenden tubulina marcada con RFP. (A) Expresión en P2 y P18 (B). imágenes de DIC a la izquierda e imágenes de fluorescencia a la derecha.

**[0023]** La **Fig. 15** representa el mapa de un plásmido donante que comprende la GFP marcadora flanqueada por las secuencias genómicas de STAT3.

**[0024]** La **Fig. 16** representa un esquema que muestra la secuencia cromosómica (SEQ ID NO: 27) en la región de STAT3 para la integración de la secuencia de marcaje, los sitios de unión a ZFN (secuencia amarilla) de la región cromosómica diana, el sitio de corte de ZFN (flecha amarilla), y el sitio de integración de la secuencia de marcaje (flecha verde). "M" simboliza el aminoácido metionina del codón de inicio.

**[0025]** La **Fig. 17** representa un ensayo de Cel-1 que confirma la eficacia de las ZFN en el corte de la secuencia cromosómica de STAT3 en el sitio previsto (tercer carril). También se muestran los resultados de Cel-1 para el control del polinucleótido donante solo y ZFN con el control del polinucleótido donante.

**[0026]** La **Fig. 18** presenta el análisis de electroforesis en gel de agarosa del ARN sintetizado que codifica las ZFN específicas del locus de STAT3.

**[0027]** La **Fig. 19** representa los datos de clasificación de células para las células transfectadas con ZFN y polinucleótido donante para la integración de GFP en locus de STAT3 (A). También se muestran los datos de clasificación de células para las células de control negativo (B).

**[0028]** La **Fig. 20** representa un análisis de electroforesis en gel de agarosa de la PCR de unión de 2 regiones diana diferentes del genoma: la región ACTB que codifica la  $\beta$ -actina fue diana con una secuencia de marcaje codificante bien de GFP o de RFP, mientras que STAT3 fue diana con una secuencia de marcaje codificante de GFP. La STAT3 se analizó usando 2 conjuntos de cebadores de unión diferentes ("cebador 1" y "cebador 2"). La PCR confirmó la integración en el locus de actina, pero no en el locus de STAT3. También se muestran los marcadores del tamaño molecular y un control de GFP.

**[0029]** La **Fig. 21** representa el mapa de un plásmido donante que comprende las secuencias genómicas de MAPRE3 que flanquean la secuencia de la GFP marcadora.

**[0030]** La **Fig. 22** representa un ensayo de Cel-1 que muestra la eficacia de una serie de pares de ZFN en el corte de la secuencia cromosómica de MAPRE3 en el sitio de integración del extremo N-terminal. El carril 1 es un marcador del tamaño del ADN, los carriles 2 y 11 son control de GFP, y los carriles 3 a 10 representan el ensayo de Cel-1 que usa diversos pares de ZFN mostrados sobre cada carril.

**[0031]** La **Fig. 23** representa un ensayo de Cel-1 que muestra la eficacia de una serie de pares de ZFN en el corte de la secuencia cromosómica dirigida de MAPRE3 en el sitio de integración del extremo N-terminal (carriles 4-7) y los resultados del ensayo de Cel-1 de los pares de ZFN en el corte de la secuencia cromosómica dirigida de LMNB1 (carriles 10-13). Los carriles 1 y 2 son un marcador del tamaño de ADN, los carriles 3 y 8 son controles de GFP-MAPRE3, y los carriles 9 y 14 son controles de GFP-lamina.

**[0032]** La **Fig. 24** presenta un análisis de electroforesis en gel de agarosa de la PCR de unión en el sitio diana de MAPRE3. Los círculos señalan la posible integración de las secuencias de marcaje.

**[0033]** La **Fig. 25** representa el análisis de clasificación de células para las células transfectadas con ZFN y polinucleótido donante para la integración de secuencias de GFP marcadora en el locus de MAPRE3. (A) células de control transfectadas con polinucleótido donante solo; y (B) células transfectadas con ZFN + polinucleótido donante.

**[0034]** La **Fig. 26** representa el diseño de la integración de la secuencia de marcaje en el locus de ACTB. (A) Es un esquema que muestra la secuencia cromosómica (SEQ ID NO: 24) en la región diana para la integración de la secuencia de marcaje, los sitios de unión a ZFN (secuencia amarilla) de la región cromosómica diana, el sitio de corte de ZFN (flecha amarilla) y los sitios de integración de la secuencia de marcaje (flechas verde, y verde y amarilla). (B) Es un esquema que representa la región diana genómica de ACTB, que muestra la región codificante (rojo), la región no traducida (azul) y el sitio de corte de ZFN (flecha amarilla). (C) Es un esquema de la región genómica de ACTB con la secuencia de GFP integrada en fase con la secuencia de codificación de ACTB. (D) Es un

esquema de la proteína  $\beta$ -actina endógena condensada a la GFP marcadora en el extremo N-terminal creado tras la integración con éxito de la secuencia de marcaje.

5 [0035] La Fig. 27 muestra la selección de un ensayo de Cel-1 para las ZFN dirigidas al locus de ACTB en células K562. El carril 1 es un marcador, y los números que figuran sobre los carriles se refieren a pares de ZFN.

[0036] La Fig. 28 representa el mapa de un plásmido donante que comprende la GFP marcadora flanqueada por las secuencias genómicas de ACTB, cuyo sitio de integración está representado por "v.2" en la Fig. 26A.

10 [0037] La Fig. 29 muestra imágenes de microscopio de fluorescencia de clones de células aislados individuales que expresan la proteína  $\beta$ -actina endógena marcada con GFP. La posición del pocillo está marcada sobre cada imagen.

15 [0038] La Fig. 30 representa la secuencia de ADN (SEQ ID NO: 16) de la región genómica de ACTB1 en células U2OS, que demuestra que la secuencia de codificación de GFP2 se integró en la región de codificación de la actina. El texto subrayado indica la región secuenciada, el texto en negrita indica la secuencia de codificación de GFP2, el texto en cursiva indica el sitio de restricción o el enlazador, y el texto en negrita y mayúscula indica el codón Met para la unión de corte y empalme.

20 [0039] La Fig. 31 representa la secuencia de ADN (SEQ ID NO: 17) de la región genómica de ACTB1 en células U2OS, que demuestra que la secuencia de codificación de RFP se integró en la región de codificación de la actina. El texto subrayado indica la región secuenciada, el texto en negrita indica la secuencia de codificación de RFP, el texto en cursiva indica el sitio de restricción o el enlazador, y el texto en negrita y mayúscula indica el codón Met para la unión de corte y empalme.

25 [0040] La Fig. 32 representa el mapa de un plásmido donante para la integración de las secuencias de GFP marcadora y el intercambio de la secuencia genómica que codifica los 15 primeros aminoácidos de la proteína  $\beta$ -actina con una secuencia de ácido nucleico que codifica el uso de un codón alterno cuyo sitio de integración se representa como "v.1" en la Fig. 26A.

30 [0041] La Fig. 33 es un esquema del fragmento de ADN mostrado en la Fig. 32 de la región genómica de ACTB en el polinucleótido donante usado para reemplazar la secuencia genómica que codifica los 15 primeros aminoácidos de la proteína  $\beta$ -actina con una secuencia de ácido nucleico que codifica el uso de un codón alterno.

35 [0042] La Fig. 34 representa el diseño de la integración secuencia de marcaje en el locus de LMNB1. (A) es un esquema que muestra la secuencia cromosómica (SEQ ID NO: 20) en la región diana para la integración de la secuencia de marcaje, los sitios de unión a ZFN (secuencia amarilla) en la región cromosómica diana, el sitio de corte de ZFN (flecha amarilla), y el sitio de integración de la secuencia de marcaje (flecha verde). (B) es un esquema que representa la región genómica diana de LMNB1 que muestra la región codificante (rojo), la región no traducida (azul) y el sitio de corte de ZFN (flecha amarilla). (C) es un esquema del sitio diana de la integración en la región genómica de LMNB1. (D) es un esquema de la región genómica de LMNB1 con la secuencia de GFP integrada en la secuencia de codificación de LMNB1. (E) es un esquema de la proteína endógena Lamina B1 condensada a la GFP marcadora en el extremo N-terminal creado tras la integración con éxito de la secuencia de marcaje.

45 [0043] La Fig. 35 muestra las imágenes de contraste diferencial de interferencia (DIC) y de microscopía de fluorescencia de células que expresan la proteína endógena Lamina B1 marcada con GFP.

[0044] La Fig. 36 representa la secuencia de ADN de la región genómica de LAMNB1 en células U2OS, que demuestra que la secuencia de codificación de RFP se integró en la región de codificación de la lamina (SEQ ID NO: 50 21). El texto subrayado indica la región secuenciada, el texto en negrita indica la secuencia de codificación de GFP2, el texto en cursiva indica el sitio de restricción o el enlazador, y el texto en negrita y mayúscula indica el codón de inicio para la unión de corte y empalme.

[0045] La Fig. 37 muestra imágenes de las células iPS que comprenden Lamina marcada con RFP. (A) Imagen de 55 DIC del campo de las células. (B) Imagen de fluorescencia de color rojo que muestra la expresión de la lamina marcada con RFP. (C) Núcleos de las células teñidas con DAPI.

[0046] La Fig. 38 representa el diseño de la integración de la secuencia de marcaje (SEQ ID: 15) en el locus de 60 ERBB2. La figura esquemática muestra la secuencia cromosómica en la región diana para la integración de la secuencia de marcaje, sitios de unión a ZFN en la región cromosómica diana, el sitio de corte de ZFN y el sitio de integración de la secuencia de marcaje.

[0047] La Fig. 39 representa el mapa de un plásmido donante para la integración de las secuencias de la GFP 65 marcadora. La secuencia codificante de GFP está flanqueada por secuencias genómicas de ERBB2.

[0048] La Fig. 40 representa la PCR de unión de la unión izquierda para confirmar la integración de GFP2 en el locus de ERBB2 en las células SKOV3.

[0049] La Fig. 41 muestra la expresión de HER2 marcada con GFP en células SKOV3. Imágenes superiores: DIC; 5 Imágenes inferiores: microscopía fluorescente.

[0050] La Fig. 42 muestra el diseño de la integración de la secuencia de marcaje en el locus de HMGA. La figura esquemática muestra la secuencia cromosómica (SEQ ID NO: 3) en la región diana para la integración de la secuencia de marcaje, los sitios de unión a ZFN en la región cromosómica diana, el sitio de corte de ZFN y el sitio de integración de la secuencia de marcaje, y la ubicación pertinente de la región de codificación, la región no traducida y el sitio de inserción de GFP en el locus de HMGA. 10

[0051] La Fig. 43 representa el mapa de un plásmido donante para la integración de las secuencias de la GFP marcadora. La secuencia codificante de GFP está flanqueada por secuencias cromosómicas de HMG1. 15

[0052] La Fig. 44 representa la verificación por (A) PCR genómica y (B) transferencia de Southern (con una sonda de GFP) de la integración de la GFP marcadora en locus de HMGA1 en clones seleccionados.

[0053] La Fig. 45 representa la secuencia de ADN de la región genómica de HMGA1 en células U2OS, que demuestra que la secuencia de codificación de GFP2 se integró en la región de codificación de HMGA (SEQ ID NO: 17). El texto subrayado indica la región secuenciada, el texto en negrita indica la secuencia de codificación de GFP2, el texto en cursiva indica el sitio de restricción o el enlazador, y el texto en negrita y mayúscula indica el codón de inicio para la unión de corte y empalme. 20

[0054] La Fig. 46 muestra imágenes de las células U2OS que expresan la proteína HMGA1 marcada con GFP. Izquierda: imagen DIC; derecha: imagen fluorescente. 25

#### Descripción detallada de la invención

[0055] La presente divulgación engloba un método de marcaje de una proteína endógena en una célula. El método comprende poner en contacto una célula con una endonucleasa de dirección y un polinucleótido donante que comprende una secuencia de marcaje. La endonucleasa de dirección introduce una rotura de doble cadena en un sitio específico de la secuencia cromosómica que codifica la proteína endógena. La rotura de doble cadena induce el proceso de reparación del ADN celular que da lugar a la recombinación homóloga y a la reparación de la rotura de doble cadena usando un polinucleótido donante como molde. Por consiguiente, la secuencia de marcaje del polinucleótido donante está integrada en fase en la secuencia cromosómica que codifica la proteína endógena. Debido a que la secuencia de marcaje está integrada en fase con la secuencia de codificación endógena, la proteína endógena comprende una secuencia de marcaje cuando se produce. 30 35

[0056] Ventajosamente, como se ilustra en los ejemplos, el método se puede utilizar para expresar proteínas marcadas bajo el control de vías de regulación endógenas que reflejan el patrón de expresión de la proteína endógena. 40

[0057] La presente divulgación también proporciona células que comprenden al menos una secuencia de marcaje integrada en fase en una secuencia cromosómica que codifica una proteína endógena, de manera que la célula expresa al menos una proteína endógena marcada. También se proporciona en el presente documento un kit de control de la ubicación de al menos una proteína endógena, en el que el kit comprende una célula que tiene al menos una secuencia de marcaje integrada en fase en una secuencia cromosómica que codifica una proteína endógena. 45 50

#### I. Célula que comprende proteína/s endógena/s marcada/s

[0058] Un aspecto de la presente divulgación engloba una célula que comprende al menos una secuencia de marcaje integrada en fase en una secuencia cromosómica que codifica una proteína endógena, de manera que la célula expresa al menos una proteína endógena marcada. Los ejemplos de proteínas endógenas adecuadas se detallan a continuación como ejemplos de marcadores adecuados. 55

##### (a) proteína endógena

[0059] La expresión "proteína endógena", en el presente documento, se refiere a una proteína codificada por el material genético de la célula. En general, cualquier proteína endógena de interés se puede marcar con una variedad de secuencias de marcaje. 60

[0060] En una realización, la proteína endógena puede ser una proteína tubulina. En diversas realizaciones, la proteína tubulina puede ser una proteína tubulina humana tal como una proteína  $\alpha$ -tubulina codificada por los genes 65

TUBA1A, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBA3E, TUBA4A y TUBA8; una proteína  $\beta$ -tubulina codificada por los genes TUBB, TUBB1, TUBB2A, TUBB2B, TUBB2C, TUBB3, TUBB4, TUBB4Q y TUBB6; una proteína  $\gamma$ -tubulina codificada por los genes TUBG1, TUBG2, TUBGCP2, TUBGCP3, TUBGCP4, TUBGCP5 y TUBGCP6; una proteína  $\delta$ -tubulina codificada por el gen TUBD1, o una proteína de  $\epsilon$ -tubulina codificada por el gen TUBE1. En una realización ilustrativa, la tubulina endógena puede ser la proteína de la isoforma de la  $\alpha$ -tubulina 1B humana codificada por el gen TUBA1B en el cromosoma humano número 12 (número de acceso NM\_006082).

**[0061]** En otra realización, la proteína endógena puede ser una proteína actina. En alguna realización, la proteína actina puede ser una proteína actina humana tal como  $\alpha$ -actina codificada por el gen ACTA1, la proteína  $\beta$ -actina codificada por el gen ACTB, o la proteína  $\gamma$ -actina codificada por el gen ACTG1. En una realización ilustrativa, la proteína endógena puede ser la proteína  $\beta$ -actina humana codificada por el gen ACTB en el cromosoma 7 humano (número de acceso NM\_001101).

**[0062]** En otra realización más, la proteína endógena puede ser una proteína lamina. En ciertas realizaciones, la proteína lamina puede ser una proteína lamina humana, tal como laminas B1 y B2, expresadas por los genes LMNB1 y LMNB2, o proteínas Lamina A y C, las variantes de corte y empalme del gen LMNA. En una realización ilustrativa, la proteína endógena puede ser la proteína Lamina B1 humana codificada por el gen LMNB1 en el cromosoma 5 humano (número de acceso NM\_005573).

**[0063]** En otra realización más, la proteína endógena puede ser el receptor del factor de crecimiento epidérmico humano 2 (proteína HER2) que está codificado por el gen ERBB2. HER2 es un receptor tirosina quinasa unido a la superficie de la membrana celular y participa en las vías de transducción de señales que conducen al crecimiento y a la diferenciación celular. La amplificación del gen ERBB2 o la sobreexpresión de su producto proteico se asocia con el cáncer de mama, el cáncer de ovario y el cáncer de estómago. La proteína HER2 endógena puede ser la proteína HER2 humana (UniProtKB/Swiss-Prot, número de acceso: P04626).

**[0064]** En una realización alternativa, la proteína endógena puede ser HMGA. HMGA se refiere al grupo de alta movilidad de las proteínas cromosómicas que regulan la expresión génica mediante el cambio de la configuración del ADN a través de la unión a regiones ricas en AT. Se encuentran entre el grupo más grande y mejor caracterizado de proteínas nucleares no histonas. El gen HMGA1 regula una gran variedad de procesos biológicos normales, incluyendo el crecimiento, la proliferación, la diferenciación y la muerte celular. Para este gen, se han encontrado al menos siete variantes de transcripción que codifican dos isoformas diferentes. En algunas realizaciones, la proteína endógena puede ser una proteína HMGA humana. Los ejemplos no limitantes de proteínas HMGA humanas que se pueden usar en la invención incluyen la isoforma a y la isoforma b de HMGA, expresadas por el gen HMGA1 (número de acceso NM\_145899).

**[0065]** En realizaciones adicionales, la proteína endógena puede ser una proteína que aparece en la TABLA A.

<b>Tabla A: Otras proteínas marcadas endógenamente</b>				
	<b>Símbolo del gen</b>	<b>Nombre de la proteína</b>	<b>Símbolo de la proteína</b>	<b>N.º de acceso de la proteína</b>
1	HiF1a	factor inducible por la hipoxia-1	HIF1	Q16665
2	VEGF (A, B, C)	factor de crecimiento endotelial vascular (A, B, C)	VEGFA, VEGFB, VEGFC	P15692, P49765, P49767
3	GLUT1 (SLC2A1)	familia de vehículos de solutos 2 (transportador facilitado de la glucosa)	GTR1	P11166
4	LDHA	lactato deshidrogenasa A	LDHA	P00338
5	IL-1 (A, B)	Interleucina 1 (alfa, beta)	IL1A, IL1B	P01538, P01584
6	IL-8	Interleucina 8	IL8	P10145
7	Cox-2 (PTGS2)	prostaglandina-endoperóxido sintasa 2	PTGS2	P35354
8	CCND1	Ciclina D1	CCND1	P24385
9	CDKN1B (p27)	inhibidor de la quinasa dependiente de la ciclina 1B	CDKN1B	P46527
10	CREB1	proteína de unión al elemento sensible a cAMP 1	CREB1	P16220
11	Bcl2	CLL/linfoma 2 de linfocitos B	BCL2	P10415
12	MDM2	proteína de unión a p53	MDM2	Q00987

<b>Tabla A: Otras proteínas marcadas endógenamente</b>				
	<b>Símbolo del gen</b>	<b>Nombre de la proteína</b>	<b>Símbolo de la proteína</b>	<b>N.º de acceso de la proteína</b>
13	p70S6K (RPS6KB1)	proteína quinasa S6 ribosomal, 70 kDa, polipéptido 1	RPS6KB1	P23443
14	FKHR (FOXO1)	Forkhead box O1	FOXO1	Q12778
15	$\beta$ -catenina (Ctnnb1)	catenina (proteína asociada a la cadherina), beta 1	CTNNB1	P35222
16	MMP7	metalopectidasa de la matriz 7 (matrilisina, uterina)	MMP7	P09237
17	Vim	Vimentina	VIM	P08670
18	BIRC5	5 que contiene la repetición IAP de baculovirus (variante 3 alfa de la survivina)	BIRC5	O15392
19	CCND2	Ciclina D2	CCND2	P30279
20	BCLXL (BCL2L1)	BCL2 de tipo 1	BCL2L1	Q07817
21	p21 (CIP1,CDKN1A)	inhibidor de la quinasa dependiente de la ciclina 1A (p21, Cip1)	CDKN1A	P38936
22	STAT1	transductor de señales y activador de la transcripción 1	STAT1	P42224
23	STAT2	transductor de señales y activador de la transcripción 2	STAT2	P52630
24	STAT3	transductor de señales y activador de la transcripción 3	STAT3	P40763
25	STAT4 (SLEB11)	transductor de señales y activador de la transcripción 4	STAT4	Q14765
26	EGFR (ERBB1)	receptor del factor de crecimiento epidérmico	EGFR	P00533
27	SOCS1	supresor de la señalización de las citocinas 1	SOCS1	O15524
28	SOCS2	supresor de la señalización de las citocinas 2	SOCS2	O14508
29	SOCS3	supresor de la señalización de las citocinas 3	SOCS3	O14543
30	Viperina (RSAD2, ciq5)	2 que contiene el dominio de la S-adenosil metionina radical (Viperina)	RSAD2	Q8WXG1
31	GLUT4 (SLC2A4)	familia de vehículos de solutos 2 (transportador facilitado de la glucosa), miembro 4	GTR4	P14672
32	COL1A1	colágeno, tipo I, alfa 1	COL1A1	P02452
33	PPARG	receptor gamma activado por el factor proliferador de peroxisomas	PPARG	P37231
34	SMAD3	miembro 3 de la familia SMAD	SMAD3	P84022
35	SMAD4	miembro 4 de la familia SMAD	SMAD4	Q13485
36	JNK (MAPK8)	proteína quinasa activada por mitógeno 8	MAPK8	P45983
37	TP53	proteína tumoral p53	TP53	P04637
38	NF-kB (NFKB1, p50)	factor nuclear 1 potenciador de las cadenas ligeras kappa de los linfocitos B	NFKB1	P19838
39	Notch1	Notch1	NOTC1	P46531
40	ATF-2	factor de transcripción activador 2	ATF2	P15336
41	c-JUN (Jun)	proto-oncogén Jun	JUN	P05412
42	AKT1	homólogo del oncogén viral de timoma murino v-akt	AKT1	P31749
43	p38 $\alpha$ (MAPK14)	proteína quinasa activada por mitógeno 14	MK14	Q16539
44	p38 $\beta$ (MAPK11)	proteína quinasa activada por mitógeno 11	MK11	Q15759
45	p38 $\gamma$ (MAPK12)	proteína quinasa activada por mitógeno 12	MK12	P53778
46	ERK (MAPK1)	proteína quinasa activada por mitógeno 1	MK01	P28482
47	AhR	receptor de hidrocarburo de arilo	AHR	P35869
48	PXR	subfamilia de receptores nucleares 1, grupo I, miembro 2	NR1L2	O75469
49	CAR	subfamilia de receptores nucleares 1, grupo I miembro 3	NR1L3	Q14994



<b>Tabla A: Otras proteínas marcadas endógenamente</b>				
	<b>Símbolo del gen</b>	<b>Nombre de la proteína</b>	<b>Símbolo de la proteína</b>	<b>N.º de acceso de la proteína</b>
50	CYP1A2	citocromo P450, familia 1, subfamilia A, polipéptido 2	CP1A2	P05177
51	CYP3A4	citocromo P450, familia 3, subfamilia A, polipéptido 4	CP3A4	P08684
52	CYP2B6	citocromo P450, familia 2, subfamilia B, polipéptido 6	CP2B6	P20813
53	Nrf2	factor 2 relacionado con el factor nuclear eritroide 2	NF2L2	Q16236
54	Hmox1	hemo oxigenasa (desciclado) 1	HMOX1	P09601
55	GSTA2	glutaciona S-transferasa alfa 2	GSTA2	P09210
56	Prdx1	peroxiredoxina 1	PRDX1	Q06830
57	Keap1	proteína 1 asociada con ECH de tipo kelch	KEAP1	Q14145
58	Grp78	receptor 78 acoplado a la proteína G	GPR78	Q96P69
59	ATF4	factor de transcripción activador 4 (elemento potenciador sensible a tax B67)	ATF4	P18848
60	ATF6	factor de transcripción activador 6	ATF6	P18850
61	XBP1	proteína 1 de unión a la caja X	XBP1	P17861
62	Gadd45a	detención del crecimiento e inducible de daño del ADN, alfa	GADD45A	P24522
63	p21	subunidad de 21 kDa de la ribonucleasa P/MRP	RPP21	Q9H633
64	Bax	proteína X asociada con BCL2	BAX	Q07812
65	RAD51c	homólogo 3 de la proteína RAD51 de reparación del ADN	RA51C	O43502
66	BTG2	miembro 2 de la familia de BTG	BTG2	P78543
67	OATP1B1	familia de transportadores de aniones orgánicos en vehículos de solutos, miembro 1B1	OATP2	Q9Y6L6
68	OATP1B3	familia de transportadores de aniones orgánicos en vehículos de solutos, miembro 1B3	OATP8	Q9NPD5
69	OAT1	familia 22 de vehículos de solutos (transportador de aniones orgánicos), miembro 6	OAT1	Q4U2R8
70	OAT3	familia 22 de vehículos de solutos (transportador de aniones orgánicos), miembro 8	OAT3	Q8TCC7
71	OCT2	familia 22 de vehículos de solutos (transportador de cationes orgánicos), miembro 2	OCT2	O15244
72	BSEP	casete de unión a ATP, subfamilia B (MDR/TAP), miembro 11	BSEP	O95342
73	MATE1	familia 47 de vehículos de solutos, miembro 1	MATE1	Q96FL8
74	BCRP	proteína de resistencia al cáncer de mama	BCRP	Q9UNQ0
75	ABCB1	proteína 1 de resistencia a múltiples fármacos	MDR1	P08183
76	ABCC2	miembro 2 de la subfamilia C de casetes de unión al ATP	MRP2	Q92887
77	Pdk1	proteína quinasa 1 dependiente del fosfoinosítido 3	PDK1	O15530
78	HSF-1	proteína de factor de choque térmico 1	HSF1	Q00613
79	HSP90 (AA1, AB1)	proteína de choque térmico HSP 90 (-alfa, -beta)	HSP90	P07900 P08238
80	HSPA1A/1B	proteína de choque térmico 1A/1B de 70 kDa	HSP70	P08107
81	HSPB1	proteína de choque térmico beta-1	Hsp27	P04792
82	p65	Factor de transcripción p65	TP65	Q04206
83	IL2	Interleucina-2	IL-2	P60568
84	NOS2	Óxido nítrico sintasa, inducible	iNOS	P35228

<b>Tabla A: Otras proteínas marcadas endógenamente</b>				
	<b>Símbolo del gen</b>	<b>Nombre de la proteína</b>	<b>Símbolo de la proteína</b>	<b>N.º de acceso de la proteína</b>
85	iCAM (1,2,3,4,5)	Molécula de adhesión intercelular (1,2,3,4,5)	iCAM (1,2,3,4,5)	P05362, P13598, P32942, Q14773, Q9UMF0
86	JUN	Factor de transcripción AP-1	AP1	P05412
87	Fbx15	proteína 15 de caja F sola	FBX15	Q8NCQ5
88	TUBB3	Cadena beta-3 de la tubulina	TUBB3	Q13509
89	UCHL1	Isozima L1 de la hidrolasa carboxilo-terminal de ubiquitina	UCHL1	P09936
90	SERPIN1			
91	SV2A	glicoproteína de vesícula sináptica 2A	SV2A	Q7L0J3
92	GRIA2	receptor de glutamato, ionotrópico, AMPA2	GRIA2	P42262
93	MAP2	proteína 2 asociada a los microtúbulos	MAP2	P11137
94	GFAP	proteína ácida fibrilar glial	GFAP	P14136
95	PEA15	fosfoproteína enriquecida en astrocitos 15	PEA15	Q15121
96	PLP	proteína proteolipídica 1	PLP	P60
97	GALC	galactosilceramidasa	GALC	P54803
98	MBP	proteína básica de la mielina	MBP	P02686
99	CNP	2',3'-nucleótido cíclico-3'-fosfodiesterasa	CNP	P09543
100	Olig2	factor de transcripción de oligodendrocitos 2	Olig2	Q13516
101	NES	Nestina	Nestin	Q48681
102	Sox2	Factor de transcripción SOX-2	SOX2	P48431
103	FoxG1B	Proteína de la caja Foxhead G1	FOXG1B	P55316
104	Pax6	Proteína Pax-6	PAX6	P26367
105	TH	Tirosina 3-monooxigenasa	TH	P07101
106	CLDN6	Claudina-6	CLDN6	P56747
107	GATA4	Factor de transcripción GATA-4	GATA4	P43694
108	PDX1	Proteína de homeocaja de páncreas/duodeno 1	PDX-1	P52945
109	Krt20	Queratina, citoesqueleto 20 de tipo I	KRT20	P35900
110	KLF4	Factor 4 de tipo Krueppel	KLF4	O43474
111	Sox17	Factor de transcripción SOX-17	Sox17	Q9H6I2
112	FoxA2	Factor nuclear de hepatocitos 3-beta	FOXA2	Q9Y261
113	CXCR4	receptor de quimioquina C-X-C de tipo 4	CXCR4	P61073
114	HNF4A	Factor nuclear de hepatocitos 4-alfa	HNF4	P41235
115	DPP4	Dipeptidil peptidasa 4	DPP4	P27487
116	AFM	Afamina	ALB2	P43652
117	KRT19	Queratina, citoesqueleto 19 de tipo I	KRT19	P08727
118	KRT18	Queratina, citoesqueleto 18 de tipo I	KRT18	P05783
119	CYP7A1	Colesterol 7-alfa-monooxigenasa	CYP7A1	P22680
120	CYP3A4	Citocromo P450 3A4	CYP3A4	P08684
121	Cyp2B6	Citocromo P450 2B6	CYP2B6	P20813
122	PCK1	Fosfoenolpiruvato carboxiquinasa, citosólica [GTP]	PCK1	P35558
123	PCK2	Fosfoenolpiruvato carboxiquinasa [GTP], mitocondrial	PCK2	Q16822
124	TAT	Tirosina aminotransferasa	TAT	P17735
125	TD02	Triptófano 2,3-dioxigenasa	TDO	P48775
126	GalC	Galactocerebrosidasa	GALC	P54803
127	Mafa	Factor de transcripción MafA	MAFA	Q8NHW3
128	NEUROG3	Neurogenina-3	NGN-3	Q9Y4Z2

<b>Tabla A: Otras proteínas marcadas endógenamente</b>				
	<b>Símbolo del gen</b>	<b>Nombre de la proteína</b>	<b>Símbolo de la proteína</b>	<b>N.º de acceso de la proteína</b>
129	RUNX1	Factor de transcripción relacionado con Runt 1	RUNX1	Q01196
130	myb (c-myb)	Activador de la transcripción Myb	c-Myb	P10242
131	VAV1	Proto-oncogén vav	VAV1	P15498
132	GATA1	Factor de transcripción eritroide	GATA1	P15976
133	LCLAT1	Lisocardiopina aciltransferasa 1	LCLAT1	Q6UWP7
134	CD34	Antígeno de células progenitoras hematopoyéticas CD34	CD34	P28906
135	PTPRC	Proteína asociada al receptor de la proteína tirosina fosfatasa de tipo C	CD45	Q14761
136	MNX1	Proteína 1 de la homeocaja de las neuronas motoras y del páncreas	HOXHB9	P50219
137	CD34			
138	ICA1	autoantígeno 1 de las células de los islotes	ICAp69	Q05084
139	MYEF-2	Factor de expresión de la mielina 2	MYEF-2	Q9P2K5
140	ChAT	colina O-acetiltransferasa	ChAT	P28329
141	ISLET (ISL1)	Proteína potenciadora del gen de la insulina ISL-1	ISLET	P61371
142	NKX2-5	Factor de transcripción de NK2 relacionado, locus 5	NKX2-5	P52952
143	EHMT1 (Brachyury)	histona-lisina N-metiltransferasa eucromática 1	EHMT1	Q9H9B1
144	MyH6	miosina, cadena pesada 6, músculo cardiaco, alfa	MYH6	P13533
145	TNNT2	troponina T de tipo 2 (cardiaca)	TNNT2	P45379
146	Mixl1	Proteína de la homeocaja MIXL1	MIXL	Q9H2W2
147	MLC2a	Cadena ligera reguladora de la miosina 2, isoforma atrial	MLC-2a	Q01449
148	MLC2v	Cadena ligera reguladora de la miosina 2, isoforma del músculo ventricular/cardiaco	MLC-2v	P10916
149	HCN4	Canal 4 abierto por nucleótidos cíclicos activados por la hiperpolarización del potasio/sodio	HCN4	Q9Y3Q4
150	Hey1	Vellosa/potenciadora de la división relacionada con la proteína 1 del motivo YRPW	CHF-2	Q9Y5J3
151	Hey2	Vellosa/potenciadora de la división relacionada con la proteína 2 del motivo YRPW	CHF-1	Q9UBP5
152	Mesp1	Proteína del mesodermo posterior 1	Mesp1	Q9BRJ9
153	GRE (elemento de respuesta a glucocorticoides)			
1-46 genes relacionados con las vías de señalización celular				
47-86 genes relacionados con la ADEM/toxicidad				
89-152 genes relacionados con la medicina regenerativa/células madre				

**(b) secuencia de marcaje**

**[0066]** El término “marcador” se refiere, en el presente documento, a una proteína que se fusiona con la proteína endógena para crear las proteínas endógenas marcadas. La secuencia de marcaje se fusiona en fase con la secuencia de codificación de la proteína endógena, de manera que se genera una proteína de fusión. La expresión “en fase” significa que el marco de lectura abierto (ORF) de la secuencia cromosómica que codifica la proteína se mantiene tras la inserción de la secuencia de marcaje. Las inserciones en fase se producen cuando el número de nucleótidos insertados es divisible entre tres, lo que se puede lograr mediante la adición de un enlazador de cualquier número de nucleótidos a la secuencia que codifica la proteína de marcaje según el caso. La proteína endógena se puede marcar en cualquier lugar dentro de la secuencia polipeptídica de la proteína, siempre que la función de la proteína endógena no se vea afectada. En general, el marcaje se realiza en el extremo N- o C-terminal de la proteína. La proteína endógena se puede marcar, por ejemplo, en el extremo N-terminal de la proteína. Como

alternativa, la proteína endógena se puede marcar en el extremo C-terminal de la proteína.

**[0067]** Una secuencia de marcaje puede ser cualquier secuencia peptídica codificada por una secuencia de ácido nucleico. La secuencia de marcaje puede codificar una variedad de marcadores, incluyendo, pero sin limitación, epítopos marcadores, marcadores de afinidad, indicadores o combinaciones de los mismos.

**[0068]** El marcador puede ser, por ejemplo, un epítipo marcador. El epítipo marcador puede comprender una secuencia de aminoácidos aleatoria o una secuencia de aminoácidos conocida. Una secuencia de aminoácidos conocida puede tener, por ejemplo, anticuerpos generados contra ella, o puede no haber anticuerpos conocidos generados contra la secuencia. El epítipo marcador puede ser una epítipo marcador de anticuerpo para el que haya anticuerpos comerciales disponibles. Los ejemplos no limitantes de epítopos marcadores de anticuerpos adecuados son myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, HA, proteína de unión a la maltosa, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6xHis, BCCP y calmodulina.

**[0069]** Un marcador ilustrativo puede ser un indicador. Los indicadores adecuados son conocidos en la técnica. Los ejemplos no limitantes de indicadores incluyen marcadores de afinidad, indicadores visuales o indicadores marcadores seleccionables. Los ejemplos no limitantes de marcadores de afinidad incluyen la proteína de unión a la quitina (CBP), la tiorredoxina (TRX), poli(NANP), el marcador de purificación por afinidad en tándem (TAP) y la glutatión-S-transferasa (GST). Los indicadores visuales suelen dar lugar a una señal visual, tal como un cambio de color en la célula, o la fluorescencia o luminiscencia de la célula. Por ejemplo, el indicador LacZ, que codifica  $\beta$ -galactosidasa, volverá azul una célula en presencia de un sustrato adecuado, tal como X-gal. Otros ejemplos no limitantes de indicadores visuales incluyen una proteína fluorescente, luciferasa, fosfatasa alcalina, beta-galactosidasa, beta-lactamasa, peroxidasa de rábano picante y variantes de las mismas. Además, se puede usar la luciferasa. Los indicadores marcadores seleccionables normalmente confieren un rasgo seleccionable a la célula, tal como resistencia a fármacos (por ejemplo, resistencia a los antibióticos).

**[0070]** Un marcador ilustrativo es un indicador visual de proteína fluorescente. Los ejemplos no limitantes de indicadores visuales de proteína fluorescente incluyen proteínas verdes fluorescentes (por ejemplo, GFP, GFP-2, tagGFP, turboGFP, EGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), proteínas amarillas fluorescentes (por ejemplo, YFP, EYFP, citrina, Venus, YPet, PhiYFP, ZsYellow1), proteínas azules fluorescentes (por ejemplo, EBFP, EBFP2, Azurite, mKalama1, GFPuv, Sapphire, T-sapphire), proteínas azul verdosas fluorescentes (por ejemplo, ECFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), proteínas rojas fluorescentes (mKate, mKate2, mPlum, DsRed monomer, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred) y proteínas naranjas fluorescentes (mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato) o cualquier otra proteína fluorescente adecuada. Son marcadores ilustrativos una proteína verde fluorescente o una proteína roja fluorescente.

**[0071]** Los ejemplos no limitantes también incluyen permutaciones circulares de proteínas verdes fluorescentes, en las que las partes amino y carboxilo se intercambian y se vuelven a unir con un espaciador corto que conecta los extremos originales, sin que dejen de ser fluorescentes. Estas permutaciones circulares de proteína fluorescente tienen alterados los valores de pKa y las orientaciones del cromóforo. Además, ciertos lugares dentro de algunas proteínas fluorescentes toleran la inserción de proteínas enteras, y los cambios de configuración en la inserción pueden tener efectos profundos en la fluorescencia, tal como el aumento o el cambio de color. Por ejemplo, las inserciones de calmodulina o un dominio de dedos de cinc en el lugar de Tyr-145 de una mutante amarilla (EYFP, proteína amarilla fluorescente mejorada) de GFP dan lugar a proteínas indicadoras cuya fluorescencia se puede mejorar varias veces tras la unión del metal. El injerto de calmodulina en una proteína amarilla fluorescente mejorada puede controlar el  $\text{Ca}^{2+}$  citosólico en células de mamífero individuales.

**[0072]** La proteína endógena se puede fusionar, por ejemplo, con el marcador a través de un enlace peptídico. La secuencia del enlazador peptídico se selecciona basándose en contribuciones estructurales y conformacionales conocidas de segmentos peptídicos para permitir el plegamiento apropiado y evitar un posible impedimento estérico de la proteína que se vaya a marcar y del polipéptido marcador. Los enlazadores peptídicos se usan comúnmente y se conocen en la técnica, y pueden ser de aproximadamente 3 a aproximadamente 40 aminoácidos de longitud.

**[0073]** La proteína endógena también se puede marcar con más de un marcador. Por ejemplo, una proteína endógena se puede marcar con al menos uno, dos, tres, cuatro, cinco, seis, siete, ocho o nueve marcadores. Se puede expresar más de un marcador como un solo polipéptido fusionado a una proteína endógena de interés. Más de un marcador fusionado a una proteína endógena se puede expresar como un solo polipéptido que se escinde en los polipéptidos marcadores individuales después de la traducción. A modo de ejemplo no limitante, los péptidos 2A de los picornavirus insertados entre polipéptidos marcadores o entre el polipéptido marcador y la proteína endógena pueden dar lugar a la "escisión" junto con la traducción de un marcador, y conducir a la expresión de múltiples proteínas a niveles equimolares.

**[0074]** En una realización ilustrativa, la célula expresa una proteína endógena que está marcada con una proteína

fluorescente. En otra realización ilustrativa, la célula expresa dos proteínas endógenas marcadas por fluorescencia. En otra realización ilustrativa más, la célula expresa tres proteínas endógenas marcadas por fluorescencia. En una realización adicional, la célula expresa cuatro o más proteínas endógenas marcadas.

### 5 (c) Tipo de célula

**[0075]** En general, la célula será una célula eucariota. Las células adecuadas incluyen hongos o levaduras, tales como *Pichia pastoris* o *Saccharomyces cerevisiae*; células de insectos, tales como células SF9 de *Spodoptera frugiperda* o células S2 de *Drosophila melanogaster*; células vegetales; y células animales, tales como células de ratón, rata, hámster, primate no humano o ser humano. Las células ilustrativas son de mamífero. Las células de mamífero pueden ser células primarias. En general, se puede usar cualquier célula primaria que sea sensible a las roturas de doble cadena. Las células pueden ser de una variedad de tipos de células, por ejemplo, fibroblastos, mioblastos, linfocitos T o B, macrófagos, células epiteliales, etcétera.

**[0076]** La célula de mamífero puede ser una célula de línea celular de mamífero. La línea celular puede ser cualquier línea celular establecida o una línea celular primaria. La línea celular puede ser adherente o no adherente, o la línea celular se puede cultivar en condiciones que potencien el crecimiento adherente, no adherente u organotípico usando técnicas convencionales conocidas por los expertos en la materia. Los ejemplos no limitantes de líneas celulares de mamífero adecuadas incluyen células de ovario de hámster chino (CHO), línea CVI de riñón de mono transformada por SV40 (COS7); línea de riñón embrionario humano 293; células de riñón de cría de hámster (BHK); células de Sertoli de ratón (TM4); células de riñón de mono (CVI-76); células de riñón de mono verde africano (VERO-76); células humanas de carcinoma de cuello uterino (HeLa); células de riñón canino (MDCK); células de hígado de rata búfalo (BRL 3A); células de pulmón humanas (W138); células de hígado humanas (Hep G2); células de tumor de mama de ratón (MMT); células de hepatoma de rata (HTC); células HIH/3T3, la línea celular de osteosarcoma humana U2-OS, la línea celular A549 humana, la línea celular K562 humana, la línea celular HEK293 humana, la línea celular HEK293T humana y células TRI. Para una extensa lista de líneas celulares de mamíferos, los expertos en la materia pueden consultar el catálogo de la Colección Americana de Cultivos Tipo (ATCC<sup>®</sup>, Mamassas, VA). En general, las células pueden ser de una variedad de tipos de células, por ejemplo, fibroblastos, mioblastos, linfocitos T o B, macrófagos, células epiteliales, etcétera. Una línea celular ilustrativa de acuerdo con la presente divulgación es la línea celular de osteosarcoma U2OS humana. Otras líneas celulares humanas ilustrativas alternativas son la línea celular A549, la línea celular K562, la línea celular HEK293 y la línea celular HEK293T. Otra línea celular humana ilustrativa es la MCF10a, una línea celular de cáncer epitelial de mama. Otra línea celular humana ilustrativa es la SKOV3, una línea de células epiteliales. Otras líneas celulares ilustrativas alternativas incluyen células iPS, que son células madre pluripotentes inducidas generadas a partir de fibroblastos u otros tipos de células.

**[0077]** En otras realizaciones más, la célula puede ser una célula madre. Las células madre adecuadas incluyen, sin limitación, células madre embrionarias, células madre de tipo ES, células madre fetales, células madre adultas, células madre pluripotentes, células madre pluripotentes inducidas, células madre multipotentes, células madre oligopotentes y células madre unipotentes.

**[0078]** En realizaciones adicionales, la célula puede ser un embrión monocelular. El embrión puede ser un vertebrado o un invertebrado. Los vertebrados adecuados incluyen mamíferos, aves, reptiles, anfibios y peces. Los ejemplos de mamíferos adecuados incluyen, sin limitación, roedores, animales de compañía, animales de granja y no primates. Los ejemplos no limitantes de roedores incluyen ratones, ratas, hámsteres, jerbos y cobayas. Los animales de compañía adecuados incluyen, pero sin limitación, gatos, perros, conejos, erizos y hurones. Los ejemplos no limitantes de animales de granja incluyen caballos, cabras, ovejas, cerdos, vacas, llamas y alpacas. Los no primates adecuados incluyen, pero sin limitación, monos capuchinos, chimpancés, lemures, macacos, monos tíf, tamarinos, monos araña, monos ardilla y monos verdes. Los ejemplos de aves no limitantes incluyen pollos, pavos, patos y gansos. Como alternativa, el animal puede ser un invertebrado tal como un insecto, un nematodo, y similares. Los ejemplos no limitantes de insectos incluyen *Drosophila* y mosquitos.

## II. Método de marcaje de proteína/s endógenas/s

**[0079]** Otro aspecto de la presente divulgación engloba un método de marcaje de al menos una proteína endógena en una célula. El método comprende el uso de una endonucleasa de dirección para mediar en la integración de un marcaje en fase con una secuencia codificante endógena. Más concretamente, el método comprende introducir en una célula al menos una nucleasa de dedos de cinc o ácido nucleico que codifique una nucleasa de dedos de cinc y al menos un polinucleótido donante. El polinucleótido donante comprende una secuencia de marcaje que se va a integrar en fase en la secuencia cromosómica endógena, una secuencia cadena arriba y una secuencia cadena abajo flanqueando la secuencia de marcaje, en el que las secuencias cadena arriba y cadena abajo comparten una identidad de secuencia sustancial con cualquiera de los lados del sitio de escisión de la secuencia cromosómica endógena que codifica la proteína. A continuación, se mantienen las células en condiciones que permitan la reparación de una rotura de doble cadena introducida en el sitio de escisión por la nucleasa de dedos de cinc mediante un proceso de homología dirigida, de manera que la secuencia de marcaje del

polinucleótido donante se integre en fase en la secuencia cromosómica que codifica la proteína endógena. Las células generadas por el método que expresan al menos una proteína endógena marcada se detallan anteriormente en el apartado (I). Los componentes del método se describen más detalladamente a continuación.

#### 5 (a) Endonucleasa de dirección

[0080] El método comprende, en parte, introducir en una célula al menos una endonucleasa de dirección o ácido nucleico que codifique una endonucleasa de dirección. La endonucleasa de dirección puede ser una proteína de origen natural o una proteína diseñada por ingeniería genética. En algunas realizaciones, la endonucleasa de dirección puede ser una meganucleasa o una endonucleasa buscadora. En otras realizaciones, la endonucleasa de dirección puede ser una nucleasa efectora de tipo activador de la transcripción (TALE). En realizaciones preferidas, la endonucleasa de dirección puede ser una nucleasa de dedos de cinc. Por lo general, una nucleasa de dedos de cinc comprende un dominio de unión a ADN (es decir, dedo de cinc) y un dominio de escisión (es decir, nucleasa), que se describen a continuación.

15

##### (i) Dominio de unión de dedos de cinc

[0081] Los dominios de unión de dedos de cinc pueden estar diseñados para reconocer y unirse a cualquier secuencia de ácido nucleico de elección. Véase, por ejemplo, Beerli *et al.* (2002) *Nat. Biotechnol.* 20:135-141; Pabo *et al.* (2001) *Ann. Rev. Biochem.* 70:313-340; Isalan *et al.* (2001) *Nat. Biotechnol.* 19:656-660; Segal *et al.* (2001) *Curr. Opin. Biotechnol.* 12:632-637; Choo *et al.* (2000) *Curr. Opin. Struct. Biol.* 10:411-416; Zhang *et al.* (2000) *J. Biol. Chem.* 275(43):33850-33860; Doyon *et al.* (2008) *Nat. Biotechnol.* 26:702-708; y Santiago *et al.* (2008) *Proc. Natl. Acad. Sci. EE.UU.* 105:5809-5814. Un dominio de unión de dedos de cinc diseñado puede tener una nueva especificidad de unión en comparación con una proteína de dedos de cinc de origen natural. Los métodos de diseño incluyen, pero sin limitación, el diseño racional y diversos tipos de selección. El diseño racional incluye, por ejemplo, el uso de bases de datos que comprenden secuencias de nucleótidos doblete, triplete y/o cuadruplete y secuencias de aminoácidos de dedos de cinc individuales, en las que cada secuencia de nucleótidos de doblete, triplete o cuadruplete se asocia con una o más secuencias de aminoácidos de dedos de cinc que se unen a la secuencia triplete o cuadruplete en particular. Véanse, por ejemplo, las patentes de EE.UU. n.º 6.453.242 y 6.534.261, cuyas divulgaciones se incorporan en el presente documento como referencia en su totalidad. Como ejemplo, se puede usar el algoritmo descrito en la patente de EE.UU. n.º 6.453.242 para diseñar un dominio de unión de dedos de cinc para dirigirse a una secuencia preseleccionada. También se pueden usar métodos alternativos, tales como el diseño racional usando una tabla de códigos de reconocimiento no degenerado para diseñar un dominio de unión de dedos de cinc para dirigirse a una secuencia específica (Sera *et al.* (2002) *Biochemistry* 41: 7074-7081). Se pueden encontrar herramientas basadas en la Web disponibles para el público para identificar los posibles sitios diana en las secuencias de ADN y diseñar los dominios de unión de dedos de cinc en <http://www.zincfingertools.org> y <http://bindr.qdcb.iastate.edu/ZiFiT/>, respectivamente (Mandell *et al.* (2006) *Nuc. Acid Res.* 34:W516-W523; Sander *et al.* (2007) *Nuc. Acid Res.* 35:W599-W605).

[0082] Se puede diseñar un dominio de unión de dedos de cinc que reconozca y se una a una secuencia de ADN que varíe de aproximadamente 3 nucleótidos a aproximadamente 21 nucleótidos de longitud, o de aproximadamente 8 a aproximadamente 19 nucleótidos de longitud. En general, los dominios de unión de dedos de cinc de las nucleasas de dedos de cinc desveladas en la presente documento comprenden al menos tres regiones de reconocimiento de dedos de cinc (es decir, dedos de cinc). En una realización, el dominio de unión de dedos de cinc puede comprender cuatro regiones de reconocimiento de dedos de cinc. En otra realización, el dominio de unión de dedos de cinc puede comprender cinco regiones de reconocimiento de dedos de cinc. En otra realización más, el dominio de unión de dedos de cinc puede comprender seis regiones de reconocimiento de dedos de cinc. Se puede diseñar un dominio de unión de dedos de cinc que se una a cualquier secuencia de ADN diana adecuada. Véase, por ejemplo, las patentes de EE.UU. n.º 6.607.882; 6.534.261 y 6.453.242, cuyas divulgaciones se incorporan por referencia en el presente documento en su totalidad.

[0083] Los métodos ilustrativos de selección de una región de reconocimiento de dedos de cinc pueden incluir la presentación de fagos y los sistemas de dos híbridos, y se desvelan en las patentes de EE.UU. n.º 5.789.538; 5.925.523; 6.007.988; 6.013.453; 6.410.248; 6.140.466; 6.200.759; y 6.242.568; así como los documentos WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 y GB 2.338.237, cada uno de los cuales se incorpora por referencia en el presente documento en su totalidad. Además, la mejora de la especificidad de unión de los dominios de unión de dedos de cinc se ha descrito, por ejemplo, en el documento WO 02/077227.

[0084] Los expertos en la materia conocen los dominios de unión de dedos de cinc y los métodos de diseño y construcción de proteínas de fusión (y los polinucleótidos que las codifican), y se describen en detalle en las publicaciones de solicitud de patente de EE.UU. n.º 20050064474 y 20060188987, cada una incorporada por referencia en el presente documento en su totalidad. Las regiones de reconocimiento de dedos de cinc y/o las proteínas de varios dedos de cinc se pueden unir entre sí usando secuencias enlazadoras adecuadas, incluyendo, por ejemplo, enlazadores de cinco o más aminoácidos de longitud. Véanse, las patentes de EE.UU. n.º 6.479.626; 6.903.185 y 7.153.949, cuyas divulgaciones se incorporan en el presente documento por referencia en su totalidad,

para consultar ejemplos no limitantes de secuencias enlazadoras de seis o más aminoácidos de longitud. El dominio de unión de dedos de cinc descrito en el presente documento puede incluir una combinación de enlazadores adecuados entre los dedos de cinc individuales de la proteína.

5 **[0085]** En algunas realizaciones, la nucleasa de dedos de cinc puede comprender además una señal o secuencia de localización nuclear (NLS). Una NLS es una secuencia de aminoácidos que facilita la dirección de la proteína nucleasa de dedos de cinc hacia el núcleo para introducir una rotura de doble cadena en la secuencia diana del cromosoma. Las señales de localización nuclear son conocidas en la técnica. Véase, por ejemplo, Makkerh *et al.* (1996) *Current Biology* 6:1025-1027.

10

**[0086]** Un dominio de unión de ADN de dedos de cinc ilustrativo reconoce y se une a una secuencia que tiene al menos aproximadamente el 80 % de identidad de secuencia con una secuencia seleccionada entre SEQ ID NO: 1, 2, 13, 14, 18, 19, 22, 23, 25 y 26. En otras realizaciones, la identidad de secuencia puede ser del aproximadamente 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 15 99 % o 100 %.

(ii) Dominio de escisión

**[0087]** Una nucleasa de dedos de cinc también incluye un dominio de escisión. La parte del dominio de escisión de las nucleasas de dedos de cinc desveladas en el presente documento puede obtenerse de cualquier endonucleasa o exonucleasa. Los ejemplos no limitantes de endonucleasas de las que se puede obtener un dominio de escisión incluyen, pero sin limitación, endonucleasas de restricción y endonucleasas buscadoras. Véase, por ejemplo, 2002-2003 Catalog, New England Biolabs, Beverly, Mass.; y Belfort *et al.* (1997) *Nucleic Acids Res.* 25:3379-3388 o [www.neb.com](http://www.neb.com). Se conocen enzimas adicionales que escinden el ADN (por ejemplo, nucleasa S1; nucleasa de judía mungo; DNasa pancreática I; nucleasa microcócica; endonucleasa de levadura HO). Véase también Linn *et al.* (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993. Una o más de estas enzimas (o fragmentos funcionales de las mismas) se pueden usar como fuente de dominios de escisión.

**[0088]** También se puede obtener un dominio de escisión de una enzima o parte de la misma, como se describe anteriormente, que requiera la dimerización para la actividad de escisión. Se pueden requerir dos nucleasas de dedos de cinc para la escisión, ya que cada nucleasa comprende un monómero del dímero de enzima activa. Como alternativa, una sola nucleasa de dedo de cinc puede comprender ambos monómeros para crear un dímero de enzima activa. Como se usa en el presente documento, un "dímero de enzima activa" es un dímero enzimático capaz de escindir una molécula de ácido nucleico. Los dos monómeros de escisión pueden derivar de la misma endonucleasa (o fragmentos funcionales de la misma) o cada monómero puede derivar de una endonucleasa diferente (o fragmentos funcionales de la misma).

**[0089]** Cuando se usan dos monómeros de escisión para formar un dímero de enzima activa, los sitios de reconocimiento para las dos nucleasas de dedos de cinc se disponen preferentemente de manera que la unión de las dos nucleasas de dedos de cinc con sus respectivos sitios de reconocimiento coloque a los monómeros de escisión en una orientación espacial entre sí que permita que los monómeros de escisión formen un dímero de enzima activa, por ejemplo, por dimerización. Como resultado de ello, los bordes cercanos de los sitios de reconocimiento pueden estar separados por aproximadamente 5 a aproximadamente 18 nucleótidos. Por ejemplo, los bordes cercanos pueden estar separados por aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 o 18 nucleótidos. No obstante, se entenderá que cualquier número entero de nucleótidos o pares de nucleótidos puede intercalarse entre dos sitios de reconocimiento (por ejemplo, de aproximadamente 2 a aproximadamente 50 pares de nucleótidos o más). Los bordes cercanos de los sitios de reconocimiento de las nucleasas de dedos de cinc, tales como, por ejemplo, los descritos en detalle en el presente documento, pueden estar separados por 6 nucleótidos. En general, el sitio de escisión se encuentra entre los sitios de reconocimiento.

50

**[0090]** Las endonucleasas de restricción (enzimas de restricción) están presentes en muchas especies y son capaces de unirse con especificidad de secuencia al ADN (en un sitio de reconocimiento), y escindir el ADN en o cerca del sitio de unión. Ciertas enzimas de restricción (por ejemplo, de tipo IIS) escinden el ADN en sitios retirados del sitio de reconocimiento, y tienen dominios de unión y de escisión separables. Por ejemplo, la enzima de Tipo IIS FokI cataliza la escisión de doble cadena del ADN, a 9 nucleótidos de su sitio de reconocimiento en una cadena y 13 nucleótidos de su sitio de reconocimiento en la otra. Véanse, por ejemplo, las patentes de EE.UU. n.º 5.356.802; 5.436.150 y 5.487.994; así como Li *et al.* (1992) *Proc. Natl. Acad. Sci. EE.UU.* 89: 4275-4279; Li *et al.* (1993) *Proc. Natl. Acad. Sci. EE.UU.* 90: 2764-2768; Kim *et al.* (1994a) *Proc. Natl. Acad. Sci. EE.UU.* 91:883-887; Kim *et al.* (1994b) *J. Biol. Chem.* 269:31, 978-31, 982. Por lo tanto, una nucleasa de dedos de cinc puede comprender el dominio de escisión de al menos una enzima de restricción de tipo IIS y uno o más dominios de unión de dedos de cinc, que pueden o no estar diseñados. Se describen enzimas de restricción de tipo ISS ilustrativas, por ejemplo, en la publicación internacional WO 07/014.275, cuya divulgación se incorpora por referencia en el presente documento en su totalidad. Otras enzimas de restricción adicionales también contienen dominios de unión y de escisión separables, y también están contempladas por la presente divulgación. Véase, por ejemplo, Roberts *et al.* (2003) *Nucleic Acids Res.* 31: 418-420.

65

- [0091]** Un ejemplo de enzima de restricción de tipo IIS, cuyo dominio de escisión es separable del dominio de unión, es FokI. Esta enzima en particular es activa como dímero (Bitinaite *et al.* (1998) *Proc. Natl. Acad. Sci. EE.UU.* 95:10, 570-10, 575). Por consiguiente, para los fines de la presente divulgación, la parte de la enzima FokI usada en una nucleasa de dedos de cinc se considera un monómero de escisión. Por lo tanto, para la escisión de doble cadena dirigida usando un dominio de escisión de FokI, se pueden usar dos nucleasas de dedos de cinc, que comprenden cada una un monómero de escisión de FokI, para reconstituir un dímero de enzima activa. Como alternativa, también se puede usar una molécula de un solo polipéptido que contenga un dominio de unión de dedos de cinc y dos monómeros de escisión de FokI.
- 10 **[0092]** En ciertas realizaciones, el dominio de escisión puede comprender uno o más monómeros de escisión diseñados que minimicen o eviten la homodimerización, como se describe, por ejemplo, en las patentes de EE.UU. n.º 20050064474, 20060188987 y 20080131962, cada una de las cuales se incorpora por referencia en el presente documento en su totalidad. A modo de ejemplo no limitante, todos los restos de aminoácidos de las posiciones 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537 y 538 de FokI son dianas para influir en la dimerización de las mitades de los dominios de escisión de FokI. Los ejemplos de monómeros de escisión de FokI diseñados que forman heterodímeros obligados incluyen un par en el que un primer monómero de escisión incluye mutaciones en las posiciones de los restos de los aminoácidos 490 y 538 de FokI y un segundo monómero de escisión incluye mutaciones en las posiciones de resto de aminoácido 486 y 499.
- 15 **[0093]** Por lo tanto, en una realización, una mutación en la posición del aminoácido 490 reemplaza Glu (E) por Lys (K); una mutación en el resto de aminoácido 538 reemplaza Iso (I) por Lys (K); una mutación en el resto de aminoácido 486 reemplaza Gln (Q) por Glu (E); y una mutación en la posición 499 reemplaza Iso (I) por Lys (K). En concreto, los monómeros de escisión diseñados se pueden preparar mediante la mutación de las posiciones 490 de E a K y 538 de I a K en un monómero de escisión para producir un monómero de escisión diseñado designado "E490K:I538K" y mediante la mutación de las posiciones 486 de Q a E y 499 de I a L en otro monómero de escisión para producir un monómero de escisión diseñado designado "Q486E:I499L". Los monómeros de escisión diseñados descritos anteriormente son heterodímeros mutantes obligados en los que se reduce al mínimo o se elimina la escisión aberrante. Los monómeros de escisión diseñados se pueden preparar usando un método adecuado, por ejemplo, por mutagénesis dirigida de monómeros de escisión de tipo silvestre (FokI) como se describe en la publicación de patente de EE.UU. n.º 20050064474 (véase el Ejemplo 5).
- 20 **[0094]** La nucleasa de dedos de cinc descrita anteriormente se puede diseñar para introducir una rotura de doble cadena en el sitio de integración diana. La rotura de doble cadena puede estar en el sitio de integración diana, o puede estar hasta a 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100 o 1.000 nucleótidos de distancia del sitio de integración. En algunas realizaciones, la rotura de doble cadena puede estar hasta a 1, 2, 3, 4, 5, 10, 15 o 20 nucleótidos de distancia del sitio de integración. En otras realizaciones, la rotura de doble cadena puede estar hasta a 10, 15, 20, 25, 30, 35, 40, 45 o 50 nucleótidos de distancia del sitio de integración. En otras realizaciones más, la rotura de doble cadena puede estar hasta a 50, 100 o 1.000 nucleótidos de distancia del sitio de integración.
- 25 **(iv) Métodos adicionales para la escisión dirigida**
- [0095]** En los métodos desvelados en el presente documento, se puede usar cualquier nucleasa que tenga un sitio diana en una secuencia cromosómica. Por ejemplo, las endonucleasas buscadoras y las meganucleasas tienen secuencias de reconocimiento muy largas, algunas de las cuales tienden a estar presentes, basándose en la estadística, una vez en un genoma de tamaño humano. Cualquiera de dichas nucleasas que tenga un sitio diana único en un genoma celular se puede usar en lugar de, o además de, una nucleasa de dedos de cinc, para la escisión dirigida de un cromosoma celular.
- 30 **[0096]** Los ejemplos no limitantes de endonucleasas buscadoras incluyen I-SceI, I-CeuI, PI-PspI, PI-Sce, I-SceIV, I-CsmI, I-PanI, I-Scell, I-PpoI, I-ScellI, I-Crel, I-TevI, I-TevII e I-TevIII. Las secuencias de reconocimiento de estas enzimas son conocidas en la técnica. Véase también la patente de EE.UU. n.º 5.420.032; la patente de EE.UU. n.º 6.833.252; Belfort *et al.* (1997) *Nucleic Acids Res.* 25:3379-3388; Dujon *et al.* (1989) *Gene* 82:115-118; Perler *et al.* (1994) *Nucleic Acids Res.* 22, 1125-1127; Jasin (1996) *Trends Genet.* 12:224-228; Gimble *et al.* (1996) *J. Mol. Biol.* 263:163-180; Argast *et al.* (1998) *J. Mol. Biol.* 280:345-353 y el catálogo de New England Biolabs.
- 35 **[0097]** Aunque la especificidad de la escisión de la mayoría de las endonucleasas buscadoras no es absoluta con respecto a sus sitios de reconocimiento, los sitios son de longitud suficiente para que se pueda obtener una sola escisión por genoma de tamaño de mamífero mediante la expresión de una endonucleasa buscadora en una célula que contenga una sola copia de su sitio de reconocimiento. También se ha informado que la especificidad de las endonucleasas buscadoras y de las meganucleasas se puede diseñar para que se unan a sitios diana no naturales. Véase, por ejemplo, Chevalier *et al.* (2002) *Molec. Cell* 10:895-905; Epinat *et al.* (2003) *Nucleic Acids Res.* 31:2952-2962; Ashworth *et al.* (2006) *Nature* 441:656-659; Paques *et al.* (2007) *Current Gene Therapy* 7:49-66.
- 40
- 45
- 50
- 55
- 60



(v) Ácido nucleico que codifica una nucleasa de dedos de cinc

[0098] La nucleasa de dedos de cinc se puede introducir en la célula como un ácido nucleico que codifique la nucleasa de dedos de cinc. El ácido nucleico que codifica una nucleasa de dedos de cinc puede ser ADN o ARN. En una realización, el ácido nucleico que codifica una nucleasa de dedos de cinc puede ser ADN. Por ejemplo, puede introducirse ADN de plásmido que comprenda una secuencia de codificación de nucleasa de dedos de cinc en la célula. En otra realización, el ácido nucleico que codifica una nucleasa de dedos de cinc puede ser ARN o ARNm. Cuando el ácido nucleico que codifica una nucleasa de dedos de cinc es ARNm, la molécula de ARNm puede estar protegida terminalmente en 5'. Del mismo modo, cuando el ácido nucleico que codifica una nucleasa de dedos de cinc es ARNm, la molécula de ARNm puede estar poliadenilada. Por lo tanto, un ácido nucleico de acuerdo con el método puede ser una molécula de ARNm protegida terminalmente y poliadenilada, que codifique una nucleasa de dedos de cinc. Los métodos de protección terminal y de poliadenilación de ARNm son conocidos en la técnica.

**(b) Polinucleótido donante**

[0099] El método de integración de la secuencia de marcaje en fase en una secuencia cromosómica diana comprende además introducir en la célula al menos un polinucleótido donante que comprenda la secuencia de marcaje. Un polinucleótido donante no solo comprende la secuencia de marcaje, como se detalla anteriormente en el apartado (l)(b), sino que también comprende una secuencia cadena arriba y una secuencia cadena abajo. Las secuencias cadena arriba y cadena abajo flanquean la secuencia de marcaje en el polinucleótido donante. Además, las secuencias cadena arriba y cadena abajo comparten identidad de secuencia sustancial con cualquiera de los lados del sitio de integración de la secuencia cromosómica endógena.

[0100] Las secuencias cadena arriba y cadena abajo del polinucleótido donante se seleccionan para potenciar la recombinación entre la secuencia cromosómica diana y el polinucleótido donante. La secuencia cadena arriba, como se usa en el presente documento, se refiere a una secuencia de ácido nucleico que comparte similitud de secuencia con la secuencia cromosómica cadena arriba del sitio de integración diana. Del mismo modo, la secuencia cadena abajo se refiere a una secuencia de ácido nucleico que comparte similitud de secuencia con la secuencia cromosómica cadena abajo del sitio de integración diana. Las secuencias cadena arriba y cadena abajo del polinucleótido donante pueden tener aproximadamente el 75 %, 80 %, 85 %, 90 %, 95 % o 100 % de identidad de secuencia con la secuencia cromosómica diana. En otras realizaciones, las secuencias cadena arriba y cadena abajo del polinucleótido donante pueden tener aproximadamente 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con la secuencia cromosómica diana. En una realización ilustrativa, las secuencias cadena arriba y cadena abajo del polinucleótido donante pueden tener aproximadamente el 99 % o 100 % de identidad de secuencia con la secuencia cromosómica diana.

[0101] Una secuencia cadena arriba o cadena abajo puede comprender de aproximadamente 20 pb a aproximadamente 2.500 pb. En una realización, una secuencia cadena arriba o cadena abajo puede comprender aproximadamente 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000, 1.100, 1.200, 1.300, 1.400, 1.500, 1.600, 1.700, 1.800, 1.900, 2.000, 2.100, 2.200, 2.300, 2.400 o 2.500 pb. Una secuencia cadena arriba o cadena abajo ilustrativa puede comprender de aproximadamente 200 pb a aproximadamente 2.000 pb, de aproximadamente 600 pb a aproximadamente 1.000 pb, o más particularmente, de aproximadamente 700 pb a aproximadamente 1.000 pb.

[0102] Por lo general, el polinucleótido donante será ADN. El polinucleótido donante puede ser un plásmido de ADN, un cromosoma artificial bacteriano (BAC), un cromosoma artificial de levadura (YAC), un vector viral, un trozo lineal de ADN, un fragmento de PCR, un ácido nucleico desnudo o un ácido nucleico complejo con un vehículo de administración tal como un liposoma o poloxámero. En una realización, el polinucleótido donante que comprende la secuencia de marcaje puede ser un plásmido de ADN. En otra realización, el polinucleótido donante que comprende la secuencia de marcaje puede ser un BAC.

[0103] Un experto en la materia sería capaz de construir un polinucleótido donante como se describe en el presente documento usando técnicas recombinantes convencionales bien conocidas (véase, por ejemplo, Sambrook *et al.*, 2001 y Ausubel *et al.*, 1996).

**(c) Administración a la célula**

[0104] El método comprende la introducción de la endonucleasa de dirección o el ácido nucleico que codifica la endonucleasa de dirección y el polinucleótido donante en una célula. Las células adecuadas se detallan anteriormente en el apartado (l)(c).

[0105] Los métodos de administración adecuados incluyen microinyección, electroporación, sonoporación, biolística, transfección mediada por fosfato de calcio, transfección catiónica, transfección de liposomas, transfección de dendrímeros, transfección de choque térmico, transfección de nucleofección, magnetofección, lipofección, impalefacción, transfección óptica, captación mejorada con agente patentado de ácidos nucleicos y administración a través de liposomas, inmunoliposomas, virosomas o viriones artificiales. En una realización, las moléculas se

pueden introducir en una célula por nucleofección. En otra realización, las moléculas se pueden introducir por microinyección. Las moléculas se pueden microinyectar en el núcleo o en el citoplasma de la célula.

**[0106]** La proporción del polinucleótido donante que comprende la secuencia de marcaje con respecto a la endonucleasa de dirección o el ácido nucleico que codifica la endonucleasa de dirección puede variar y variará. En una realización preferida, la endonucleasa de dirección puede ser una nucleasa de dedos de cinc. En general, la proporción del polinucleótido donante con respecto a la molécula nucleasa de dedos de cinc puede variar de aproximadamente 1:10 a aproximadamente 10:1. En diversas realizaciones, la proporción de polinucleótido donante con respecto a las moléculas nucleasas de dedos de cinc puede ser de aproximadamente 1:10, 1:9, 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, 1:2, 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1 o 10:1. En una realización, la proporción puede ser de aproximadamente 1:1.

**[0107]** En realizaciones en las que se introducen más de una molécula de endonucleasa de dirección y más de un polinucleótido donante en una célula, las moléculas se pueden introducir simultánea o secuencialmente. Por ejemplo, las moléculas de endonucleasas de dirección, cada una específica de una secuencia de reconocimiento distinta, así como los correspondientes polinucleótidos donantes, se pueden introducir al mismo tiempo. Como alternativa, cada molécula de endonucleasa de dirección, así como el polinucleótido donante correspondiente, se puede introducir secuencialmente.

#### 20 (d) Cultivo de la célula

**[0108]** El método comprende además el mantenimiento de la célula en condiciones apropiadas que permitan que tenga lugar la integración mediada por la endonucleasa de dirección. La célula se puede cultivar usando procedimientos convencionales para permitir la expresión de la endonucleasa de dirección, si es necesario. Las técnicas de cultivo celular convencionales se describen, por ejemplo, en Santiago *et al.* (2008) *PNAS* 105:5809-5814; Moehle *et al.* (2007) *PNAS* 104:3055-3060; Umov *et al.* (2005) *Nature* 435:646-651; y Lombardo *et al.* (2007) *Nat. Biotechnology* 25:1298-1306. Los expertos en la materia apreciarán que se conocen métodos para el cultivo de células en la técnica, y que pueden variar y variarán dependiendo del tipo de célula. La optimización habitual se puede usar, en todos los casos, para determinar las mejores técnicas para un tipo de célula en particular.

**[0109]** En realizaciones en las que la célula es un embrión monocelular, el embrión se puede cultivar *in vitro* (por ejemplo, en cultivo celular). Por lo general, el embrión se cultiva a una temperatura apropiada y en los medios apropiados con la proporción de O<sub>2</sub>/CO<sub>2</sub> necesaria para permitir la expresión de la nucleasa de dedos de cinc. Los ejemplos no limitantes adecuados de los medios incluyen los medios M2, M16, KSOM, BMOC y HTF. Un experto en la materia apreciará que las condiciones de cultivo pueden variar y variarán dependiendo de la especie de embrión. Se puede usar la optimización habitual, en todos los casos, para determinar las mejores condiciones de cultivo para una especie de embrión en particular. En algunos casos, el embrión también se puede cultivar *in vivo* mediante la transferencia del embrión al útero de una hembra hospedadora. En términos generales, la hembra hospedadora es de la misma especie o de especie similar a la del embrión. Preferentemente, la hembra hospedadora está pseudo-preñada. Los métodos de preparación de las hembras hospedadoras pseudo-preñadas son conocidos en la técnica. Además, se conocen métodos de transferencia de un embrión a una hembra hospedadora. El cultivo de un embrión *in vivo* permite que el embrión se desarrolle y puede dar lugar al nacimiento de un animal vivo derivado del embrión.

**[0110]** Durante esta etapa del proceso, la endonucleasa de dirección (que, en algunos casos se expresa a partir del ácido nucleico introducido) reconoce, se une y escinde la secuencia diana en el cromosoma. La rotura de doble cadena introducida por la endonucleasa de dirección se repara, por medio de recombinación homóloga con el polinucleótido donante, de forma que la secuencia de marcaje del polinucleótido donante se integra en fase en la ubicación cromosómica. El polinucleótido donante se puede integrar físicamente o, de forma alternativa, el polinucleótido donante se puede usar como un molde para la reparación de la rotura, lo que resulta en la integración de la secuencia de marcaje, así como todas o parte de las secuencias cadena arriba y cadena abajo del polinucleótido donante en el cromosoma. Un experto en la materia apreciará que se conocen métodos de cultivo de células en la técnica, y que pueden variar y variarán dependiendo del tipo de célula. Se puede usar la optimización habitual, en todos los casos, para determinar las mejores técnicas para un tipo de célula en particular.

#### 55 (e) Integraciones múltiples

**[0111]** Una realización adicional de la invención anterior comprende la realización de un método de la invención en serie, de manera que se desarrolla una célula con más de una integración dirigida, de modo que se tiene como diana a más de una proteína endógena. Por ejemplo, una célula con una primera integración dirigida se puede usar después en un método de la invención para crear una segunda integración dirigida. Se puede repetir el mismo proceso para crear una célula con tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más de diez integraciones dirigidas.

**[0112]** Como alternativa, una célula se puede desarrollar con múltiples integraciones mediante la introducción de más de una endonucleasa de dirección, cada una específica de un sitio distinto de integración, y la introducción de

un número correspondiente de polinucleótidos donantes. Cada polinucleótido donante comprendería una secuencia de ácido nucleico que se fuera a integrar y una secuencia de cadena arriba y cadena abajo homóloga al sitio cromosómico de integración como se ha detallado anteriormente. El número de endonucleasas de dirección y polinucleótidos donantes correspondientes inyectados en una célula pueden ser dos, tres, cuatro, cinco o más de cinco.

### III. Kit de marcaje de la proteína endógena

**[0113]** La presente divulgación también engloba un kit para el seguimiento de la ubicación de al menos una proteína endógena en una célula. El kit comprende una célula que tiene al menos una secuencia de marcaje integrada en fase en una secuencia cromosómica que codifica una proteína endógena, de modo que la célula expresa al menos una proteína endógena marcada. La célula puede ser una célula de mamífero. Preferentemente, la célula es una célula humana. La célula humana puede ser una célula de una línea celular seleccionada entre una célula U2OS humana, una MCF10A humana, una SKOV3 humana o una iPS humana. La proteína endógena marcada se puede seleccionar entre tubulina, actina, lamina, HER2 y HMGA. Como alternativa, el kit puede expresar al menos una proteína endógena marcada seleccionada entre las enumeradas en la Tabla A. En realizaciones preferidas, el marcador de la proteína endógena puede ser una proteína fluorescente seleccionada entre una proteína verde fluorescente, una proteína azul fluorescente, una proteína azul-verdosa fluorescente, una proteína amarilla fluorescente, una proteína naranja fluorescente y una proteína roja fluorescente. Los marcadores ilustrativos son la proteína verde fluorescente y la proteína roja fluorescente.

### Definiciones

**[0114]** A menos que se defina lo contrario, todos los términos técnicos y científicos usados en el presente documento tienen el significado comúnmente entendido por un experto en la materia a la que pertenece la presente invención. Las siguientes referencias proporcionan a un experto una definición general de muchos de los términos usados en la presente invención: Singleton *et al.*, "Dictionary of Microbiology and Molecular Biology" (2ª ed. 1994); "The Cambridge Dictionary of Science and Technology" (Walker ed., 1988); "The Glossary of Genetics", 5ª Ed., R. Rieger *et al.* (eds.), Springer Verlag (1991); y Hale y Marham, "The Harper Collins Dictionary of Biology" (1991). Como se usan en el presente documento, los siguientes términos tienen los significados que se les atribuyen a no ser que se especifique lo contrario.

**[0115]** Cuando se introducen elementos de la presente divulgación o de las realizaciones preferidas de la misma, los artículos "un/a", "el/la" y "dicho/a" pretenden significar que hay uno o más de los elementos. Los términos "que comprende", "que incluye" y "que tiene" pretenden ser inclusivos, y significan que puede haber elementos adicionales distintos de los elementos enumerados.

**[0116]** Un "gen", como se usa en el presente documento, se refiere a una región de ADN (incluyendo exones e intrones) que codifica un producto génico, así como a todas las regiones de ADN que regulan la producción del producto génico, sean o no dichas secuencias reguladoras adyacentes a secuencias de codificación y/o transcritas. Por consiguiente, un gen incluye, pero sin limitarse necesariamente a, secuencias promotoras, terminadores, secuencias reguladoras de la traducción tales como sitios de unión a ribosomas y sitios de entrada a ribosomas internos, potenciadores, silenciadores, aislantes, elementos del límite, orígenes de replicación, sitios de unión a la matriz y regiones de control de locus.

**[0117]** Una "proteína heteróloga" es una proteína que no es natural (es decir, foránea) para la célula u organismo de interés.

**[0118]** Los términos "ácido nucleico" y "polinucleótido" se refieren a un polímero de desoxirribonucleótidos o ribonucleótidos, en una configuración lineal o circular, y en forma bien monocatenaria o bicatenaria. Para los fines de la presente divulgación, estos términos no deben interpretarse como limitantes con respecto a la longitud de un polímero. Los términos pueden abarcar análogos conocidos de nucleótidos naturales, así como nucleótidos que estén modificados en la base, azúcar y/o restos fosfato (por ejemplo, cadenas principales de fosforotioato). En general, un análogo de un determinado nucleótido tiene la misma especificidad de emparejamiento de bases; es decir, un análogo de A se emparejará con T.

**[0119]** Los términos "polipéptido" y "proteína" se usan indistintamente para referirse a un polímero de restos de aminoácidos.

**[0120]** El término "recombinación" se refiere a un proceso de intercambio de información genética entre dos polinucleótidos. A los efectos de la presente divulgación, "recombinación homóloga" se refiere a la forma especializada de dicho intercambio, que tiene lugar, por ejemplo, durante la reparación de las roturas de la doble cadena en las células. Este proceso requiere la similitud de secuencia entre los dos polinucleótidos, usa una molécula "donante" o de "intercambio" para la reparación con mole de una molécula "diana" (es decir, la que experimentó la rotura de doble cadena), y se conoce indistintamente como "conversión génica sin cruzamiento" o

"conversión génica de extensión corta", ya que conduce a la transferencia de información genética del donante a la diana. Sin quedar ligados a ninguna teoría en particular, dicha transferencia puede implicar la corrección del desapareamiento de ADN heterodúplex que se forma entre la diana rota y el donante, y/o "la hibridación de cadenas dependiente de la síntesis", en la que el donante se usa para volver a sintetizar la información genética que formará parte de la diana y/o los procesos relacionados. Dicha recombinación homóloga especializada a menudo produce la alteración de la secuencia de la molécula diana de modo que parte o toda la secuencia del polinucleótido donante se incorpore al polinucleótido diana.

**[0121]** La expresión "identidad de secuencia" se refiere al grado en el que dos secuencias de nucleótidos son invariables, es decir, las dos secuencias tienen el mismo nucleótido en la misma posición. En general, la identidad de secuencia se expresa como un porcentaje. Dos secuencias de nucleótidos que son idénticas en secuencia y longitud tienen un 100 % de identidad de secuencia.

**[0122]** Como se usan en el presente documento, las expresiones "sitio diana" o "secuencia diana" se refieren a una secuencia de ácido nucleico que define una parte de una secuencia cromosómica que se va a editar y para la que una nucleasa de dedos de cinc está diseñada para reconocer y unirse, siempre que existan las condiciones suficientes para la unión.

**[0123]** Las técnicas para determinar la identidad de las secuencias de ácido nucleico y aminoácidos son conocidas en la materia. Por lo general, dichas técnicas incluyen determinar la secuencia de nucleótidos del ARNm para un gen y/o determinar la secuencia de aminoácidos codificada por el mismo, y comparar estas secuencias con una segunda secuencia de nucleótidos o de aminoácidos. Las secuencias genómicas también se pueden determinar y comparar de esta forma. En general, la identidad se refiere a una correspondencia exacta de nucleótido a nucleótido o de aminoácido a aminoácido de dos polinucleótidos o secuencias polipeptídicas, respectivamente. Dos o más secuencias (polinucleótido o aminoácido) se pueden comparar determinando su porcentaje de identidad. El porcentaje de identidad de dos secuencias, ya sean secuencias de ácido nucleico o de aminoácidos, es el número de coincidencias exactas entre dos secuencias alineadas dividido entre la longitud de las secuencias más cortas y multiplicado por 100. El algoritmo de homología local de Smith y Waterman proporciona un alineamiento aproximado para secuencias de ácido nucleico, "Advances in Applied Mathematics" 2: 482-489 (1981). Este algoritmo se puede aplicar a secuencias de aminoácidos usando la matriz de puntuación desarrollada por Dayhoff, "Atlas of Protein Sequences and Structure", M. O. Dayhoff ed., 5 supl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., EE.UU., y normalizada por Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). Genetics Computer Group (Madison, Wis.) proporciona una implementación ilustrativa de este algoritmo para determinar el porcentaje de identidad de una secuencia en la aplicación de la utilidad "BestFit". En la técnica, se conocen en general otros programas adecuados para calcular el porcentaje de identidad o de similitud entre secuencias, por ejemplo, otro programa de alineamiento es BLAST, usado con los parámetros por defecto. Por ejemplo, BLASTN y BLASTP se pueden usar usando los siguientes parámetros por defecto: código genético = convencional; filtro = ninguno; cadena = ambas; punto de corte = 60; esperado = 10; Matriz = BLOSUM62; Descripciones = 50 secuencias; ordenar por = PUNTUACIÓN ALTA; Bases de datos = no redundantes, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Los detalles de estos programas se pueden encontrar en la página Web del GenBank. Con respecto a las secuencias descritas en el presente documento, el intervalo de grados deseados de identidad de secuencia es del aproximadamente 80 % al 100 % y cualquier valor entero entre ellos. Por lo general, los porcentajes de identidad entre las secuencias son del al menos 70 a 75 %, preferentemente 80-82 %, más preferentemente 85-90 %, incluso más preferentemente del 92 %, aún más preferentemente del 95 %, y lo más preferentemente del 98 % de identidad de secuencia.

**[0124]** Como alternativa, el grado de similitud de secuencia entre polinucleótidos se puede determinar mediante la hibridación de polinucleótidos en condiciones que permitan la formación de dúplex estables entre regiones que compartan un grado de identidad de secuencia, seguida de la digestión con nucleasa/s específica/s de una sola cadena, y la determinación del tamaño de los fragmentos digeridos. Dos ácidos nucleicos o dos secuencias de polipéptidos son sustancialmente similares entre sí cuando las secuencias presentan al menos aproximadamente 70 %-75 %, preferentemente 80 %-82 %, más preferentemente 85 %-90 %, incluso más preferentemente 92 %, aún más preferentemente 95 %, y lo más preferentemente 98 % de identidad de secuencia en una longitud definida de las moléculas, según lo determinado usando los métodos anteriores. Como se usa en el presente documento, sustancialmente similares también se refiere a secuencias que muestran identidad completa con una secuencia de ADN o polipéptido especificada. Las secuencias de ADN que son sustancialmente similares se pueden identificar en un experimento de hibridación Southern, por ejemplo, en condiciones rigurosas, según lo definido para ese sistema en particular. La definición de las condiciones de hibridación apropiadas pertenece al alcance de la técnica. Véase, por ejemplo, Sambrook *et al.*, *supra*; "Nucleic Acid Hybridization: A Practical Approach", editores B. D. Hames y S. J. Higgins, (1985) Oxford; Washington, D. C.; IRL Press).

**[0125]** La hibridación selectiva de dos fragmentos de ácido nucleico se puede determinar de la siguiente manera. El grado de identidad de secuencia entre dos moléculas de ácido nucleico afecta a la eficacia y a la fuerza de las hibridaciones entre dichas moléculas. Una secuencia de ácido nucleico parcialmente idéntica inhibirá al menos parcialmente la hibridación de una secuencia completamente idéntica a una molécula diana. La inhibición de la

hibridación de la secuencia completamente idéntica se puede evaluar usando ensayos de hibridación que sean bien conocidos en la técnica (por ejemplo, transferencia de Southern (ADN), transferencia Northern (ARN), hibridación en solución, o similares, véase Sambrook, *et al.*, "Molecular Cloning: A Laboratory Manual", Segunda edición, (1989) Cold Spring Harbor, N.Y.). Dichos ensayos se pueden realizar usando grados variables de selectividad, por ejemplo, usando condiciones que varíen de baja a alta rigurosidad. Si se emplean condiciones de baja rigurosidad, se puede evaluar la ausencia de unión inespecífica usando una sonda secundaria que carezca, incluso en un grado parcial, de identidad de secuencia (por ejemplo, una sonda que tenga menos del aproximadamente 30 % de identidad de secuencia con la molécula diana), de modo que, en ausencia de uniones inespecíficas, la sonda secundaria no se hibride a la diana.

10

**[0126]** Cuando se utiliza un sistema de detección basado en la hibridación, se selecciona una sonda de ácido nucleico que sea complementaria a una secuencia de ácido nucleico de referencia, y luego, mediante la selección de las condiciones apropiadas, la sonda y la secuencia de referencia se hibridan selectivamente, o se unen, entre sí para formar una molécula dúplex. Una molécula de ácido nucleico que es capaz de hibridarse selectivamente a una secuencia de referencia en condiciones de hibridación moderadamente rigurosas normalmente se hibrida en condiciones que permiten la detección de una secuencia de ácido nucleico diana de al menos aproximadamente 10-14 nucleótidos de longitud que tenga al menos aproximadamente el 70 % de identidad de secuencia con la secuencia de la sonda de ácido nucleico seleccionada. Las condiciones de hibridación rigurosas normalmente permiten la detección de secuencias de ácido nucleico diana de al menos aproximadamente 10-14 nucleótidos de longitud que tengan una identidad de secuencia de más del aproximadamente 90-95 % con la secuencia de la sonda de ácido nucleico seleccionada. Las condiciones de hibridación útiles para la hibridación de la sonda/secuencia de referencia, en las que la sonda y la secuencia de referencia tienen un grado específico de identidad de secuencia, se pueden determinar como se conoce en la técnica (véase, por ejemplo, "Nucleic Acid Hybridization: A Practical Approach", editores B. D. Hames y S. J. Higgins, (1985) Oxford; Washington, D. C.; IRL Press). Las condiciones para la hibridación son bien conocidas por los expertos en la materia.

**[0127]** Rigurosidad de la hibridación se refiere al grado en que las condiciones de hibridación desfavorecen la formación de híbridos que contienen nucleótidos desapareados, estando la mayor rigurosidad correlacionada con una menor tolerancia para los híbridos desapareados. Los factores que afectan a la rigurosidad de la hibridación son bien conocidos por los expertos en la materia e incluyen, pero sin limitación, temperatura, pH, fuerza iónica y concentración de disolventes orgánicos tales como, por ejemplo, formamida y dimetilsulfóxido. Como es sabido por los expertos en la materia, la rigurosidad de la hibridación se aumenta al aumentarse las temperaturas, reducirse la fuerza iónica y reducirse las concentraciones de los disolventes. Con respecto a las condiciones de rigurosidad para la hibridación, es bien sabido en la técnica que se pueden emplear numerosas condiciones equivalentes para establecer una determinada rigurosidad mediante la variación, por ejemplo, de los siguientes factores: la longitud y la naturaleza de las secuencias, la composición de las bases de las diversas secuencias, las concentraciones de sales y otros componentes de la solución de hibridación, la presencia o ausencia de agentes bloqueantes en las soluciones de hibridación (por ejemplo, sulfato de dextrano y polietilenglicol), la temperatura de la reacción de hibridación y los parámetros de tiempo, así como mediante la variación de las condiciones de lavado. Se puede seleccionar un determinado conjunto de condiciones de hibridación siguiendo los métodos convencionales de la técnica (véase, por ejemplo, Sambrook, *et al.*, "Molecular Cloning: A Laboratory Manual", Segunda edición, (1989) Cold Spring Harbor, N.Y.).

#### Ejemplos

45

**[0128]** Los siguientes ejemplos se incluyen para demostrar las realizaciones preferidas de la invención.

#### Ejemplo 1. Marcaje de la proteína de la isoforma de $\alpha$ -tubulina 1B endógena

**[0129]** Se marcó la proteína de la isoforma de  $\alpha$ -tubulina 1B endógena con GFP usando recombinación homóloga inducida por ZFN. En resumen, se usaron ZFN para introducir una rotura de doble cadena en la región del cromosoma que codifica la isoforma de  $\alpha$ -tubulina 1B codificada por el locus TUBA1B. La rotura de doble cadena induce la recombinación homóloga con un polinucleótido donante que comprende la secuencia que codifica GFP flanqueada por secuencias de ácido nucleico homólogas a la región cromosómica del locus TUBA1B, y que produce la integración de la región de codificación de GFP en el cromosoma. El polinucleótido donante se construyó para fusionar la GFP marcadora en fase con la secuencia de codificación de la isoforma de  $\alpha$ -tubulina 1B para producir una proteína marcada con GFP en el extremo N-terminal. La proteína de la isoforma de  $\alpha$ -tubulina 1B marcada con GFP se expresó bajo el control del promotor de tubulina endógeno.

**[0130]** Se diseñó un par de ZFN para la integración dirigida de un marcador en el lugar de la diana TUBA1B. Para más información, véase *Science* (2009) 325: 433, incorporada en el presente documento por referencia en su totalidad. Se determinó la frecuencia de la generación de roturas de doble cadena del par de ZFN dirigido en las combinaciones de células tratadas con ZFN usando el ensayo de nucleasa Cel-1. Dicho ensayo detecta los alelos del locus diana que se desvían del tipo silvestre como resultado de la reparación imperfecta mediada por la unión de extremos no homólogos (NHEJ) de las roturas de doble cadena de ADN inducidas por ZFN. La amplificación por

65

PCR de la región diana de una combinación de células tratadas con ZFN genera una mezcla de amplicones de tipo silvestre y mutantes. La fusión y la rehibridación de esta mezcla da lugar a la formación de desapareamientos entre los heterodúplex de los alelos de tipo silvestre y mutantes. Una "burbuja" de ADN formada en el sitio del desapareamiento es escindida por la nucleasa Cel-1 Surveyor, y los productos de la escisión se pueden resolver mediante electroforesis en gel. La intensidad relativa de los productos de escisión en comparación con la banda parental es una medida del nivel de escisión de Cel-1 del heterodúplex. Esto, a su vez, refleja la frecuencia de la escisión mediada por ZFN del locus diana endógeno que posteriormente sufre la reparación imperfecta por parte de NHEJ. Para el par de ZFN usado para marcar la proteína de la isoforma de  $\alpha$ -tubulina 1B, una ZFN se diseñó para unirse a la secuencia 5' CTTGCGCTCCTAATC 3' (SEQ ID NO:1), y la otra ZFN se diseñó para unirse a la secuencia 5' CACTATGGTGAGTAA 3' (SEQ ID NO: 2) (**Fig. 1A**). A continuación, se diseñaron ARNm poliadenilados, protegidos terminalmente, codificantes del par de ZFN, usando técnicas de biología molecular conocidas. Tras la unión, el par de ZFN introduce una rotura de doble cadena en la secuencia cromosómica CCTAGC entre los sitios de reconocimiento (**Fig. 1A y 1B**) para inducir la recombinación homóloga.

15 **[0131]** Se construyó un plásmido (**Fig. 2**) como un donante de polinucleótido para la integración dirigida de una GFP marcadora en el locus TUBA1B de la línea celular humana U2OS. El plásmido comprendía la secuencia codificante de GFP flanqueada por pares de bases de 1 Kb y 700 de la secuencia del locus TUBA1B cadena arriba y cadena abajo del sitio de corte introducido por el par de ZFN (**Fig. 1C y D**). Se fusionó la secuencia de marcaje del plásmido con el locus TUBA1B de cadena arriba y cadena debajo de modo que, cuando se expresara el locus TUBA1B, se produjera una proteína de la isoforma de  $\alpha$ -tubulina 1B fusionada a la GFP marcadora en el extremo N-terminal como se detalla en la **Fig 1E**. La fusión de GFP-tubulina también se diseñó de modo que la señal de corte y empalme del primer exón del locus TUBA1B, donde se introdujo la secuencia codificante de GFP, se mantuviera intacta.

#### 25 Marcaje de la tubulina en células U2OS

**[0132]** Se transfectaron el plásmido donante y el par de ARN que codifican las ZFN en células U2OS, A549, K562, HEK293, MCF10a o HEK293T. La mezcla de ácido nucleico comprendía una parte de ADN donante por una parte de ARN de ZFN. A continuación, se cultivaron las células transfectadas y se analizaron los clones de las células individuales. Se usó la PCR de la unión realizada a 37 °C y 30 °C para confirmar que el ADN donante se había integrado en el locus TUBA1B de la tubulina. El análisis de la secuencia confirmó que la secuencia de GFP2 se integró en el locus TUBA1B de las células USOS, como se muestra en la **Fig. 3** (SEQ ID NO: 4). En la **Fig. 4** (SEQ ID NO: 5), se muestra la integración confirmada de la secuencia de RFP en la región de TUBA1B de las células U2OS.

35 **[0133]** El análisis de PCR usando cebadores que flanqueaban la unión derecha confirmó la integración. Para ello, se amplificaron 100 ng de ADN molde en una mezcla de reacción de 25  $\mu$ l (26 ciclos de 95 °C, 5 min; 95 °C, 30 s; 51 °C, 30 s; 70 °C, 1,1 min; 70 °C, 7 min, 4 °C, mantenimiento). La **Fig 5** muestra que catorce clones de células comprendían el tamaño del fragmento de PCR indicativo de la integración de la GFP. A continuación, se usó microscopía de fluorescencia para visualizar las células U2OS (**Fig. 6A-C**), las células A549 (**Fig. 6D-E**), células K562 (**Fig. 6F**) y células HEK293 (**Fig. 6G-H**) con la proteína de la isoforma de  $\alpha$ -tubulina 1B marcada con GFP.

#### Marcaje de la tubulina en la línea celular MCF10a

45 **[0134]** Se construyó un plásmido (**Fig. 7**) como un polinucleótido donante para la integración dirigida de una RFP marcadora en el locus TUBA1B de la línea celular humana MCF10a. La integración de la RFP marcadora en las células MCF10a de locus TUBA1B se verificó por PCR genómica y PCR de unión usando los cebadores de tubulina: 5' CCCCTCCGCGAGCCGCTACT 3' (SEQ ID NO:6; tub80U) y 5' GGACCGCACCCAGGACACAGT 3' (SEQ ID NO:7; tub511L). La PCR genómica y la transferencia de Southern indicaron la integración de la RFP marcadora en TUBA1B en varios clones (**Fig. 8**). El análisis de secuencia que confirmaba la integración de la secuencia de marcaje en el locus TUBA1B de la línea celular MCF10a se muestra en la **Fig. 9** (SEQ ID NO: 8). Se seleccionó el Clon 5 de la MCF10a transfectada para su posterior verificación (**Fig. 10**). En la verificación por PCR Jumpstart de la integración de RFP: 95 ng de ADN genómico (tipo silvestre y clon 5 de las células MCF10a, y tipo silvestre y clones 9, 5 de las células U2OS), se amplificó (x 35, hibridación a 69 °C y extensión a 72 °C usando los cebadores tub80U y tub522L). Se confirmó que el Clon5 de MCF10a transfectadas tenía la secuencia integrada (véase la **Fig. 10**). Se confirmaron los tamaños de las uniones izquierda y derecha del clon 5 de MCF10A usando cebadores específicos de la RFP y específicos de la tubulina, y se encontró que eran los tamaños esperados de 452 pares de bases y 408 pares de bases, respectivamente (**Fig. 11**). La expresión de la proteína RFP-tubulina se verificó a través de la transferencia Western (**Fig. 12**). Las transferencias se sondaron con anticuerpos bien anti-RFP o anti-tubulina. La expresión de RFP también se observó con microscopía fluorescente, y se observó que se ubicaba junto con la expresión de TUBA1B endógena (**Fig 13**). Se compararon las características de crecimiento de las células MCF10a transfectadas con la línea celular parental. El tiempo de duplicación de las células MCF10a transfectadas fue  $\pm$  20 % del de la línea celular parental. Se evaluó la estabilidad del fenotipo de las células MCF10a transfectadas. Se observó que, después de 8 semanas y 16 divisiones, el 99 % de las células mantuvo la señal de RFP (**Tabla 1**). La microscopía de fluorescencia confirmó la expresión de la tubulina marcada con RFP en las células MCF10a del Clon

5 (Fig. 14).

Tabla 1		
	clon 5 con RFP	MCF10a de tipo silvestre
Viabilidad a P0	90 %	100 %
Viabilidad a P10	99 %	98 %
Viabilidad a P17	98 %	99 %
Análisis de flujo de "RED" en P17	99 %	2 %

**Ejemplo 2. Intento de marcaje de la proteína transductor de señales y activador de la transcripción 3 codificada por STAT3**

[0135] El intento de producir una proteína transductor de señales y activador de la transcripción 3 codificada por STAT3 marcada por GFP o RFP no tuvo éxito. Se produjo un plásmido donante que comprendía las secuencias del locus STAT3 de cadena arriba y cadena abajo flanqueantes de un polinucleótido codificante de GFP o RFP fusionada al extremo N-terminal de la proteína transductor de señales (Fig. 15). Las ZFN se diseñaron como se describe en el ejemplo anterior. Una ZFN se diseñó para unirse a la secuencia 5' AGCTACAGCAGCTTG 3' (SEQ ID NO: 9), y otra ZFN se diseñó para unirse a la secuencia 5' CGGTACCTGGAGCAG 3' (SEQ ID NO:10), que comprende el locus STAT3 (Fig. 16). Se usó el ensayo Cel-1 descrito anteriormente para confirmar el par de ZFN cortado eficazmente del locus STAT3 en el lugar adecuado (Fig. 17).

[0136] Se transfectaron el plásmido donante y el par de ZFN codificante de ARN (Fig. 18) a las células. El análisis de clasificación de células activada por fluorescencia (FACS) demostró que no se detectó ninguna señal fluorescente y, por lo tanto, la integración dirigida no tuvo éxito (Fig. 19). Estos resultados se confirmaron mediante el análisis de PCR de unión, que no pudo detectar ninguna integración dirigida de GFP en el locus STAT3, mientras que sí detectó la integración dirigida de secuencias de marcaje codificantes de GFP y RFP en el locus ACTB codificante de la proteína β-actina. (Fig. 20).

[0137] Por lo tanto, a pesar de que el par ZFN diseñado fue capaz de introducir una rotura de doble cadena en la ubicación cromosómica correcta, no se logró la integración de la GFP marcadora.

**Ejemplo 3. Intento de marcaje de la proteína miembro 3 de la familia de proteínas RP/EB asociadas a microtúbulos codificada por MAPRE3**

[0138] Un intento de producir una proteína miembro 3 de la familia de proteínas RP/EB asociadas a microtúbulos codificada por MAPRE3 marcada con GFP no tuvo éxito.

[0139] En primer lugar, se intentó el marcaje de la proteína asociada a microtúbulos en el extremo N-terminal. Se diseñaron múltiples ZFN como se describe en el Ejemplo 1 anterior para integrar secuencias de marcaje en el extremo N-terminal de la proteína asociada a microtúbulos. Se encontraron ZFN que cortaron con éxito el ADN cromosómico cerca del extremo N-terminal de MAPRE3 (par 6/8 y 16/17; Fig. 22 y Tabla 2). Sin embargo, ninguno de los pares de ZFN cortó el cromosoma en una ubicación que fuera adecuada para la producción de la proteína de fusión marcada deseada.

Tabla 2					
Par de ZFN	Total	Parental	Banda uno	Banda dos	% eficiencia
1/2	3.282	2.495	787	0	24 % Bandas extra fuera de lo esperado
3/5	10.187				0 Ninguna banda detectada por densitometría
6/8	3.210	2.803	210	197	13 %
9/10					0 Ninguna banda detectada por densitometría
11/12					0 Ninguna banda detectada por densitometría
16/17	2.647	2.185	131	331	17 %
21/22	2.496	2.056	160	280	18 % Ligero bandeado múltiple
23/24					0 Ninguna banda detectada por densitometría

[0140] Dado que el marcaje de la proteína asociada a microtúbulos en el extremo N-terminal no tuvo éxito, a continuación, se intentó el marcaje de la proteína en el extremo C-terminal. Se diseñaron múltiples pares de ZFN para integrar secuencias de marcaje en el extremo C-terminal de la proteína asociada a microtúbulos. Como control, también se diseñaron pares de ZFN para integrar secuencias de marcaje en el extremo N-terminal de una proteína Lamina (Fig. 23 y Tabla 4). Se encontró un par de ZFN que cortó con éxito el ADN cromosómico en o cerca del

extremo C-terminal de MAPRE3 (par 31/32; **Fig. 23 y Tabla 3**). En este par, una ZFN se diseñó para que se uniera a la secuencia 5' TTCCTCTCTCTCCCAC 3' (SEQ ID NO: 11), y la otra ZFN se diseñó para que se uniera a la secuencia 5' AGGAAGGATTTCGCAC 3' (SEQ ID NO: 12) que comprendía el locus MAPRE3.

Par de ZFN	Total	Parental	Banda uno	Banda dos	% eficiencia	
26/27					0	Ninguna banda detectada por densitometría
29/30					0	Ninguna banda detectada por densitometría
31/32	3.448	2.728	579	141	21 %	Banda más pequeña detectada fácilmente
						Banda de 317 pb apenas detectada
33/35					0	Ninguna banda detectada por densitometría

5

Par de ZFN	Total	Parental	Banda uno	Banda dos	% eficiencia	
12/13-L	5.476	5.172	304		0	Ninguna banda detectada por densitometría
14/16-L	4.093	3.463	377	253	015 %	Ninguna banda detectada por densitometría
50/51-L	4.722	3.512	726	484	26 %	Banda más pequeña detectada fácilmente, banda de 317 pb apenas detectada
59/60-L	5.726	4.022	983	721	30 %	Ninguna banda detectada por densitometría

**[0141]** Se produjo un plásmido donante que comprendía secuencias del locus MAPRE3 de cadena arriba y cadena abajo que flanqueaban un polinucleótido codificante de GFP (**Fig. 21**). Se transfectaron el plásmido donante y el par 31/32 de ARN codificante de ZFN a las células, y la PCR de unión demostró una posible inserción de la GFP marcadora en el locus MAPRE3 (**Fig. 24**). Sin embargo, el análisis FACS demostró que no se detectó señal fluorescente y, por lo tanto, que la integración dirigida no tuvo éxito (**Fig. 25**).

#### Ejemplo 4. Marcaje de la proteína $\beta$ -actina endógena

**[0142]** Se marcó la proteína  $\beta$ -actina endógena con GFP usando recombinación homóloga inducida por ZFN. En resumen, se usaron ZFN para introducir una rotura de doble cadena en la región cromosómica codificante de la  $\beta$ -actina codificada por el locus ACTB. La rotura de doble cadena induce la recombinación homóloga con un polinucleótido donante que comprende la secuencia codificante de GFP flanqueada por secuencias de ácido nucleico homólogas a la región cromosómica del locus ACTB, y que da lugar a la integración de la región codificante de GFP en el cromosoma. El polinucleótido donante (**Fig. 28**) se construyó para que tuviera integrada la GFP marcadora en fase con la secuencia codificante de la  $\beta$ -actina (**Fig. 26**, "v.2") para producir una proteína marcada con GFP en el extremo N-terminal (**Fig. 26D**). La proteína  $\beta$ -actina marcada con GFP se expresó bajo el control del promotor de la actina endógena.

**[0143]** Se diseñó un par de ZFN para la integración dirigida de un marcador en el sitio diana ACTB, como se ha detallado anteriormente. Para el par de ZFN usado para marcar la proteína  $\beta$ -actina, una ZFN se diseñó para que se uniera a la secuencia 5' GTCGTCGACAACGGCTCC 3' (SEQ ID NO: 13), y la otra ZFN se diseñó para que se uniera a la secuencia 5' TGCAAGGCCGGCTTCGCGG 3' (SEQ ID NO: 14) (**Fig. 26A**). Tras la unión, el par de ZFN introduce una rotura de doble cadena en la secuencia cromosómica GGATG entre los sitios de reconocimiento (**Fig. 26A y 26B**) para inducir la recombinación homóloga. A continuación, se produjo ARNm poliadenilado, protegido terminalmente, codificante del par de ZFN, usando técnicas de biología molecular conocidas.

**[0144]** Se determinó la frecuencia de la generación de roturas de doble cadena del par de ZFN dirigido en combinaciones de células tratadas ZFN, usando el ensayo de nucleasa Cel-1 (**Fig. 27**). Dicho ensayo detecta los alelos del locus diana que se desvían del tipo silvestre como resultado de la reparación imperfecta mediada por la unión de extremos no homólogos (NHEJ) de las roturas de doble cadena de ADN inducidas por ZFN. La amplificación por PCR de la región diana de una combinación de células tratadas con ZFN genera una mezcla de amplicones de tipo silvestre y mutantes. La fusión y la rehibridación de esta mezcla da lugar a la formación de desapareamientos entre los heterodúplex de los alelos de tipo silvestre y mutantes. Una "burbuja" de ADN formada en el sitio del desapareamiento es escindida por la nucleasa Cel-1 Surveyor, y los productos de la escisión se pueden resolver mediante electroforesis en gel. La intensidad relativa de los productos de escisión en comparación con la banda parental es una medida del nivel de escisión de Cel-1 del heterodúplex. Esto, a su vez, refleja la frecuencia de la escisión mediada por ZFN del locus diana endógeno que posteriormente sufre la reparación imperfecta por parte de NHEJ.

**[0145]** Se construyó un plásmido (**Fig. 28**) como un polinucleótido donante para la integración dirigida de una GFP marcadora en el locus ACTB de la línea celular humana. El plásmido comprendía la secuencia codificante de GFP



flanqueada por 861 y 593 nucleótidos de las secuencias del locus ACTB cadena arriba y cadena abajo del sitio de corte introducido por el par ZFN (**Fig. 26C**). Se fusionó la secuencia de marcaje del plásmido a las secuencias cadena arriba y cadena abajo del locus ACTB de modo que, cuando se expresara el locus ACTB, se produjera una proteína  $\beta$ -actina fusionada a la GFP marcadora en el extremo N-terminal como se detalla en la **Fig. 26D**. La fusión de GFP-actina también se diseñó de modo que la señal de corte y empalme del primer exón del locus ACTB, donde se introdujo la secuencia codificante de GFP, se mantuviera intacta.

**[0146]** Se transfectaron el plásmido donante y el par de ARN codificantes de ZFN a las células. La mezcla de ácido nucleico comprendía una parte de ADN donante por una parte de ARN de ZFN. A continuación, se cultivaron las células transfectadas y se analizaron los clones de células individuales. Se usó la microscopía de fluorescencia para visualizar la proteína  $\beta$ -actina marcada con GFP (**Fig. 29**). En la **Fig. 30** (SEQ ID NO: 16), se muestra la secuencia confirmada del locus ACTB con la integración de GFP2 en células U2OS. En la **Fig. 31** (SEQ ID NO: 17), se muestra la secuencia confirmada del locus ACTB con la integración de RFP en las células U2OS.

#### 15 Ejemplo 5. $\beta$ -Actina marcada con GFP utilizando el péptido 2A

**[0147]** Se marcó también la  $\beta$ -actina en el extremo N-terminal con GFP mientras se reemplazaba simultáneamente la secuencia de ácido nucleico codificante de los 15 primeros aminoácidos de la  $\beta$ -actina con una secuencia de ácido nucleico con el uso de codones alternativos.

**[0148]** Para integrar una secuencia de marcaje cerca del sitio de corte de ZFN (**Fig. 26**, "v.1") que daría lugar a  $\beta$ -actina de longitud completa fusionada con traducción con GFP, se creó un nuevo plásmido donante en el que se cambiaron los 15 primeros aminoácidos ácidos de la  $\beta$ -actina (**Fig. 32**). El plásmido donante comprendía las secuencias del locus ACTB de cadena arriba y cadena abajo flanqueantes de un polinucleótido codificante del péptido 2a fusionado a GFP que, a su vez, se fusionó a través de un enlazador de 3 restos de aminoácido alanina con los 15 primeros aminoácidos de la  $\beta$ -actina codificada por codones alternativos (**Fig. 33**). La escisión con traducción del péptido 2a elimina los 15 primeros aminoácidos de la  $\beta$ -actina codificada por los nuevos codones, produciendo una proteína  $\beta$ -actina marcada con GFP en el extremo N-terminal (**Fig. 26D**).

**[0149]** Las ZFN fueron como se describe en el Ejemplo 4. Se transfectaron el plásmido donante y el par de ARN codificante de las ZFN a las células. La mezcla de ácido nucleico comprendía una parte de ADN donante por una parte de ARN de ZFN. Entonces se cultivaron las células transfectadas y se analizaron los clones de células individuales. Se usó la microscopía de fluorescencia para confirmar la expresión de la proteína  $\beta$ -actina marcada con GFP (**Fig. 29**).

#### 35 Ejemplo 6. Marcaje de la proteína Lamina B1 endógena

**[0150]** La proteína Lamina B1 endógena se marcó con GFP usando recombinación homóloga inducida por ZFN. En resumen, se usaron ZFN para introducir una rotura de doble cadena en la región cromosómica codificante de la Lamina B1 codificada por el locus LMNB1. La rotura de doble cadena induce la recombinación homóloga con un polinucleótido donante que comprende la secuencia que codifica GFP flanqueada por secuencias de ácido nucleico homólogas a la región cromosómica del locus LMNB1, y que resulta en la integración de la región codificante de GFP en el cromosoma. El polinucleótido donante se construyó para fusionar la GFP marcadora en fase con la secuencia codificante de la Lamina B1 para producir una proteína marcada con GFP en el extremo N-terminal. La proteína Lamina B1 marcada con GFP se expresó bajo el control del promotor de la Lamina endógena.

**[0151]** Se diseñó un par de ZFN como se describe anteriormente. Se determinó la frecuencia de la generación de roturas de doble cadena del par de ZFN dirigido en las combinaciones de células tratadas con ZFN usando el ensayo de nucleasa Cel-1. Del par de ZFN usado para marcar la proteína Lamina B1, una ZFN se diseñó para que se uniera a la secuencia 5' CCTCGCCGCCCGCT 3' (SEQ ID NO:18), y la otra ZFN se diseñó para que se uniera a la secuencia 5' GCCGCCGCCATGGCG 3' (SEQ ID NO:19) (**Fig. 34A**). Tras la unión, el par de ZFN introduce una rotura de doble cadena en la secuencia del cromosoma GTCTCC entre los sitios de reconocimiento (**Fig. 34A y 34B**) para inducir la recombinación homóloga. Entonces, se produjeron los ARNm poliadenilados, protegidos terminalmente, codificantes del par de ZFN, usando técnicas de biología molecular conocidas.

**[0152]** Se construyó un plásmido como polinucleótido donante para la integración dirigida de una GFP marcadora en el locus LMNB1 de la línea celular humana U2OS. El plásmido comprendía la secuencia codificante de GFP flanqueada por pares de bases de 633 y 629 Kb de la secuencia del locus LMNB1 cadena arriba y cadena abajo del sitio de corte introducido por el par de ZFN (**Fig. 34C y 34D**). Se fusionó la secuencia de marcaje del plásmido con el locus LMNB1 de cadena arriba y cadena abajo de modo que, cuando se expresó el locus LMNB1, se produjo una proteína Lamina B1 fusionada con la GFP marcadora en el extremo N-terminal como se detalla en la **Fig. 34E**.

**[0153]** Se transfectaron el plásmido donante y el par de ARN codificantes de ZFN a las células. La mezcla de ácido nucleico comprendía una parte de ADN donante por una parte de ARN de ZFN. A continuación, se cultivaron las células transfectadas y se analizaron los clones de células individuales. Se usó la PCR de unión realizada a

37 °C y 30 °C para confirmar que el ADN donante se había integrado en el locus LMNB1 de la Lamina. A continuación, se usó microscopía de fluorescencia para visualizar la proteína Lamina B1 marcada con GFP (**Fig. 35**). La secuencia confirmada en el sitio de integración de GFP2 en la región de codificación de la lamina en células U2OS se muestra en la **Fig. 36** (SEQ ID NO: 21).

**[0154]** También se transfectaron un plásmido donante comprendía la secuencia de codificación de RFP y secuencias flanqueantes de lamina, y el par de ARN codificante de ZFN a células iPS, que son células madre pluripotentes inducidas generadas a partir de fibroblastos u otros tipos de células. Las imágenes de células iPS que comprendían Lamina marcada con RFP se muestran en la **Fig. 37**.

#### 10 Ejemplo 7. Marcaje de la proteína HER2 endógena

**[0155]** La proteína HER2 endógena se marcó con GFP usando recombinación homóloga inducida por ZFN. En resumen, se usaron ZFN para introducir una rotura de doble cadena en la región cromosómica codificante de la HER2 codificada por el locus del gen ERBB2. La rotura de doble cadena induce la recombinación homóloga con un polinucleótido donante que comprende la secuencia que codifica GFP flanqueada por secuencias de ácido nucleico homólogas a la región cromosómica del locus ERBB2, y que resulta en la integración de la región codificante de GFP en el cromosoma. El polinucleótido donante se construyó para fusionar la GFP marcadora en fase con la secuencia codificante de HER2 para producir una proteína marcada con GFP en el extremo N-terminal. La proteína HER2 marcada con GFP se expresó bajo el control del promotor de la HER2 endógeno.

**[0156]** Se diseñó un par de ZFN como se describe anteriormente. Se determinó la frecuencia de la generación de roturas de doble cadena del par de ZFN dirigido en las combinaciones de células tratadas con ZFN usando el ensayo de nucleasa Cel-1. Del par de ZFN usado para marcar la proteína HER2, una ZFN se diseñó para que se uniera a la secuencia 5' TACCTGGGTCTGGAC 3' (SEQ ID NO:22), y la otra ZFN se diseñó para que se uniera a la secuencia 5' AGTGTGAACCAGAAGGCC 3' (SEQ ID NO:23). Tras la unión, el par de ZFN introduce una rotura de doble cadena en la secuencia del cromosoma GTGCC entre los sitios de reconocimiento (**Fig. 38**) para inducir la recombinación homóloga. Entonces, se produjeron los ARNm poliadenilados, protegidos terminalmente, codificantes del par de ZFN, usando técnicas de biología molecular conocidas.

**[0157]** Se construyó un plásmido como polinucleótido donante para la integración dirigida de una GFP marcadora en el locus ERBB2 (**Fig. 39**). Se fusionó la secuencia de marcaje del plásmido con el locus ERBB2 de cadena arriba y cadena abajo de modo que, cuando se expresó el locus ERBB2, se produjo una proteína HER2 fusionada con la GFP marcadora en el extremo N-terminal.

**[0158]** Se transfectaron el plásmido donante y el par de ARN codificantes de ZFN a células SKOV3. La mezcla de ácido nucleico comprendía una parte de ADN donante por una parte de ARN de ZFN. A continuación, se cultivaron las células transfectadas y se analizaron los clones de células individuales. Se usó la PCR de unión realizada a 37 °C y 30 °C para confirmar que el ADN donante se había integrado en el locus ERBB2 de las células SKOV3 transfectadas (**Fig. 40**). A continuación, se usó microscopía de fluorescencia para visualizar la proteína HER2 marcada con GFP (**Fig. 41**).

#### Ejemplo 8. Marcaje de la proteína HMGA endógena

**[0159]** La proteína HMGA endógena se marcó con GFP usando recombinación homóloga inducida por ZFN. En resumen, se usaron ZFN para introducir una rotura de doble cadena en la región cromosómica codificante de la HMGA codificada por el locus HMGA1. La rotura de doble cadena induce la recombinación homóloga con un polinucleótido donante que comprende la secuencia que codifica GFP flanqueada por secuencias de ácido nucleico homólogas a la región cromosómica del locus HMGA1, y que resulta en la integración de la región codificante de GFP en el cromosoma. El polinucleótido donante se construyó para fusionar la GFP marcadora en fase con la secuencia codificante de HMGA1 para producir una proteína marcada con GFP en el extremo N-terminal. La proteína HMGA1 marcada con GFP se expresó bajo el control del promotor de la HMGA1 endógeno.

**[0160]** Se diseñó un par de ZFN como se describe anteriormente para marcar la proteína HMG1 endógena. Una ZFN se diseñó para que se uniera a la secuencia 5' CACACCAACAACACTGCCCA 3' (SEQ ID NO:25), y la otra ZFN se diseñó para que se uniera a la secuencia 5' GGAGAAGGAGGAAGA 3' (SEQ ID NO:26) (**Fig. 42**). Tras la unión, el par de ZFN introduce una rotura de doble cadena en la secuencia del cromosoma CCTCACA entre los sitios de reconocimiento (**Fig. 44**) para inducir la recombinación homóloga. Entonces, se produjeron los ARNm poliadenilados, protegidos terminalmente, codificantes del par de ZFN, usando técnicas de biología molecular conocidas.

**[0161]** Se construyó un plásmido como polinucleótido donante para la integración dirigida de una GFP marcadora en el locus HMGA1 (**Fig. 43**). El plásmido comprendía la secuencia codificante de GFP flanqueada por 806 pares de bases y 747 pares de bases de la secuencia del locus HMGA1 cadena arriba y cadena abajo del sitio de corte introducido por el par de ZFN (**Fig. 43**). Se fusionó la secuencia de marcaje del plásmido con el locus HMGA1 de

cadena arriba y cadena abajo de modo que, cuando se expresó el locus HMGA1, se produjo una proteína HMGA fusionada con la GFP marcadora en el extremo N-terminal.

**[0162]** Se transfectaron el plásmido donante y el par de ARN codificantes de ZFN a células U2OS. La mezcla de ácido nucleico comprendía una parte de ADN donante por una parte de ARN de ZFN. A continuación, se cultivaron las células transfectadas y se analizaron los clones de células individuales. La PCR genómica y la transferencia de Southern indicaron la integración de la secuencia de marcaje en el locus HMGA1 de los clones seleccionados (**Fig. 44A y Fig. 44B**). El análisis de secuencia confirmó la integración en la región cromosómica diana (**Fig. 45**) (SEQ ID NO: 28). A continuación, se usó la microscopía de fluorescencia para visualizar la proteína HMGA1 marcada con GFP (**Fig. 46**).

LISTADO DE SECUENCIAS

**[0163]**

15 <110> SIGMA-ALDRICH CO.  
MALKOV, Dmitry  
ZENSER, Nathan  
VASSAR, Deborah L  
20 ZHANG, Fan  
ZHANG, Hongyi

<120> MÉTODOS DE GENERACIÓN DE PROTEÍNAS MARCADAS ENDÓGENAMENTE

25 <130> 047497-424729

<150> US 61/323.702  
<151> 13-04-2010

30 <150> US 61/323.719  
<151> 13-04-2010

<150> US 61/323.698  
<151> 13-04-2010

35 <150> US 61/367.017  
<151> 23-07-2010

40 <150> US 61/390.668  
<151> 07-10-2010

<150> US 61/408.856  
<151> 01-11-2010

45 <150> US 61/431.957  
<151> 12-01-2011

<160> 29

50 <170> PatentIn versión 3.5

<210> 1  
<211> 15  
<212> ADN

55 <213> *Homo sapiens*

<400> 1  
cttcgcctcc taatc

15

60 <210> 2  
<211> 15  
<212> ADN  
<213> *Homo sapiens*

65 <400> 2

ES 2 565 216 T3

	cactatggg agtaa	15
	<210> 3	
	<211> 80	
5	<212> ADN	
	<213> <i>Homo sapiens</i>	
	<400> 3	
	<b>cacaccaaca actgccacc tcacaggaga aggaggaaga ggagggcatc tcgcaggagt</b>	<b>60</b>
10	<b>cctcggagga ggagcagtga</b>	<b>80</b>
	<210> 4	
	<211> 1197	
	<212> ADN	
15	<213> Secuencia artificial	
	<220>	
	<223> SINTETIZADO	
20	<400> 4	
	<b>ccgcccctcc gcagccgcta cttaagaggc tccagcgcgc gccccgccct agtgcgttac</b>	<b>60</b>
	<b>ttacctcgac tcttagcttg tcggggacgg taaccgggac ccggtgtctg ctctgtcgc</b>	<b>120</b>
	<b>cttcgcctcc taatccctag ccactatggg taccgcaagc gggggcgagg agctgttcgc</b>	<b>180</b>
	<b>cgccatcgtg ccggtgctga tcgagctgga cggcgacgtg cacggccaca agttcagcgt</b>	<b>240</b>
	<b>gcgcggcgag ggcgagggcg acgcccacta cggcaagctg gagatcaagt tcatctgcac</b>	<b>300</b>
	<b>caccggcaag ctgcccgctc cctggcccac cctggtgacc accctctgct acggcatcca</b>	<b>360</b>
	<b>gtgcttcgcc cgctaccccg agcacatgaa gatgaacgac ttcttcaaga gcgccatgcc</b>	<b>420</b>
	<b>cgagggctac atccaggagc gcaccatcca gttccaggac gacggcaagt acaagacccg</b>	<b>480</b>
	<b>cggcgaggtg aagttcgagg gcgacaccct ggtgaaccgc atcgagctga agggcaagga</b>	<b>540</b>
	<b>cttcaaggag gacggcaaca tcctgggcca caagctggag tacagcttca acagccacaa</b>	<b>600</b>
	<b>cgtgtacatc cgccccgaca aggccaacaa cggcctggag gctaacttca agacccgcc</b>	<b>660</b>
	<b>caacatcgag ggcggcggcg tgcagctggc cgaccactac cagaccaacg tgcccctggg</b>	<b>720</b>
	<b>cgacggcccc gtgctgatcc ccatcaacca ctacctgagc actcagacca agatcagcaa</b>	<b>780</b>
	<b>ggaccgcaac gaggcccgcg accacatggt gtcctggag tccttcagcg cctgctgcca</b>	<b>840</b>
	<b>caccacggc atggacgagc tgtacagggc cgctggggat cccactatgg tgagtaagcc</b>	<b>900</b>
	<b>gtgcccctcc cggtgcttt caggaagca gggaaaagcg agccggcggg gcgctggggc</b>	<b>960</b>
	<b>cctgtataca gccgggaagg gctggcctca gagccgtccg tttggagggc ggaaaacgag</b>	<b>1020</b>
	<b>gcgagaggcc agggcgggag tggtagacc tcggtgtgtg taaatagcgg gggcccggaa</b>	<b>1080</b>
	<b>aggtcgaggg gcgccaggat ttcttctcgg actctggaag ggatgggggg ctcgggctgc</b>	<b>1140</b>
	<b>cctccgccgt atccggagct ctctttgtc gcgtaactgt gtccctgggtg cggctcc</b>	<b>1197</b>

ES 2 565 216 T3

<210> 5  
 <211> 1198  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> SINTETIZADO

<400> 5

10

```

gcccctccgc agccgctact taagaggctc cagcgcggc cccgccctag tgcgttactt      60
acctcgactc ttagcttgtc ggggacggta accgggaccc ggtgtctgct cctgtcgcct      120
tcgcctccta atccctagcc actatgggta cgcagtgctc taagggcgaa gagctgatta      180
aggagaacat gcacatgaag ctgtacatgg agggcacctg gaacaaccac cacttcaagt      240
gcacatccga gggcgaaggc aagccctacg agggcaccca gaccatgaga atcaagggtg      300
tcgagggcgg ccctctcccc ttgccttcg acatcctggc taccagcttc atgtacggca      360
gcagaacctt catcaaccac acccagggca tccccgactt ctttaagcag tccttcctg      420
agggcttcac atgggagaga gtcaccacat acgaagacgg gggcgtgctg accgctaccc      480
aggacaccag cctccaggac ggctgcctca tctacaacgt caagatcaga ggggtgaact      540
tcccatccaa cggccctgtg atgcagaaga aaacactcgg ctgggaggcc aacaccgaga      600
tgtgtaccc cgctgacggc ggctggaag gcagaagcga catggccctg aagctcgtgg      660
gcgggggcca cctgatctgc aacttcaaga ccacatacag atccaagaaa cccgctaaga      720
acctcaagat gcccggcgtc tactatgtgg accacagact ggaaagaatc aaggaggccg      780
acaaagagac ctacgtcgag cagcacgagg tggctgtggc cagatactgc gacctcccta      840
gcaaactggg gcacaaactt aatgccgctg gggatcccac tatggtgagt aagccgtgcg      900
gctcccggct gctttcaggg aagcagggaa aagcgagccg gcggggcgct ggggccctgt      960
atacagccgg gaagggctgg cctcagagcc gtccgtttgg agggcgaaa acgagggcag      1020
aggccagggc gggagtgggtg agacctcggg gtgtgtaaata agcgggggccc cggaaaggtc      1080
gaggggcgcc aggatttctt ctggactct ggaagggatg gggggctcgg gctgcccctcc      1140
gccgtatccg gagctctctt ttgtcgcgta actgtgtcct gggtcgggtc cctcgagt      1198
    
```

<210> 6  
 <211> 19  
 <212> ADN  
 <213> *Homo sapiens*

15

<400> 6  
 cccctccgca gccgctact

19

<210> 7  
 <211> 19  
 <212> ADN  
 <213> *Homo sapiens*

25

ES 2 565 216 T3

<400> 7  
cccctccgca gccgctact 19

5 <210> 8  
<211> 1198  
<212> ADN  
<213> Secuencia artificial

10 <220>  
<223> SINTETIZADO

<400> 8

**gccctccgc agccgctact taagaggctc cagcgccggc cccgcctag tgcgttactt 60**

**acctcgactc ttagcttgtc ggggacggta accgggaccc ggtgtctgct cctgtcgect 120**

**tcgcctccta atccctagcc actatgggta ccgcagtgtc taagggcgaa gagctgatta 180**

**aggagaacat gcacatgaag ctgtacatgg agggcacctg gaacaaccac cacttcaagt 240**

**gcacatccga gggcgaaggc aagccctacg agggcaccca gaccatgaga atcaaggtgg 300**

**tcgagggcgg ccctctcccc ttcgecttcg acatcctggc taccagcttc atgtacggca 360**

**gcagaacctt catcaaccac acccagggca tccccgactt ctttaagcag tcttccctg 420**

**agggcttcac atgggagaga gtcaccacat acgaagacgg gggcgtgctg accgctaccc 480**

**aggacaccag cctccaggac ggctgcctca tctacaacgt caagatcaga ggggtgaact 540**

**toccatcaa cggccctgtg atgcagaaga aaactctgg ctgggaggcc aacaccgaga 600**

**tgctgtaccc cgctgacggc ggcctggaag gcagaagcga catggccctg aagctcgtgg 660**

**gcggggggcca cctgatctgc aacttcaaga ccacatacag atccaagaaa cccgctaaga 720**

**acctcaagat gcccgcgctc tactatgtgg accacagact ggaaagaatc aaggaggccg 780**

**acaagagac ctacgtcgag cagcacgagg tggctgtggc cagatactgc gacctcccta 840**

**gcaaactggg gcacaaactt aatgccgctg gggatcccac tatggtgagt aagccgtgcg 900**

**gctcccggct gctttcaggg aagcagggaa aagcgagccg gcggggcgct ggggccctgt 960**

**atacagccgg gaagggtggt cctcagagcc gtccgtttgg agggcgaaa acgaggcgag 1020**

**aggccagggc gggagtgtgt agacctcggg gtgtgtaaat agcggggggc cggaaaggtc 1080**

**gagggcgcc aggatttctt ctcgactct ggaagggatg gggggctcgg gctgcctcc 1140**

15 **gcogtatccg gagctctctt ttgtcgcgta actgtgtcct ggggtcggtc cctcgagt 1198**

<210> 9  
<211> 15  
<212> ADN  
20 <213> *Homo sapiens*

<400> 9  
agctacagca gcttg 15

25 <210> 10  
<211> 15  
<212> ADN

# ES 2 565 216 T3

	<213> <i>Homo sapiens</i>	
5	<400> 10 cggtacctgg agcag	15
	<210> 11 <211> 16 <212> ADN <213> <i>Homo sapiens</i>	
10	<400> 11 ttcctctctc tcccac	16
15	<210> 12 <211> 15 <212> ADN <213> <i>Homo sapiens</i>	
20	<400> 12 aggaaggatt cgcac	15
25	<210> 13 <211> 18 <212> ADN <213> <i>Homo sapiens</i>	
30	<400> 13 gtcgtcgaca acggctcc	18
35	<210> 14 <211> 19 <212> ADN <213> <i>Homo sapiens</i>	
40	<400> 14 tgcaaggccg gcttcgagg	19
	<210> 15 <211> 47 <212> ADN <213> <i>Homo sapiens</i>	
45	<400> 15 ccagagtacc tgggtctgga cgtgccagtg tgaaccagaa ggccaag	47
50	<210> 16 <211> 1191 <212> ADN <213> Secuencia artificial	
55	<220> <223> SINTETIZADO  <400> 16	

ES 2 565 216 T3

ggacgcctcc gaccagtgtt tgccttttat ggtaataacg cggccggccc ggcttccttt 60  
 gtccccaatc tgggcgcgcg ccggcgcccc ctggcggcct aaggactcgg cgcgcggaa 120  
 gtggccaggg cgggggcgac ctccgctcac agcgcgcccg gctattctcg cagctacca 180  
 tgggatccgg gggcgaggag ctgttogccg gcatcgtgcc cgtgctgac gagctggaag 240  
 gcgacgtgca cggccacaag ttcagcgtgc gcggcgaggg cgagggcgac gccgactacg 300  
 gcaagctgga gatcaagttc atctgcacca ccggcaagct gcccggtccc tggcccaccc 360  
 tggtgaccac cctctgctac ggcattccagt gcttcgcccg ctaccccag cacatgaaga 420  
 tgaacgactt cttcaagagc gccatgcccg agggctacat ccaggagcgc accatccagt 480  
 tccaggacga cggcaagtac aagaccgcg gcgaggtgaa gttcgagggc gacaccctgg 540  
 tgaaccgcat cgagctgaag ggcaaggact tcaaggagga cggcaacatc ctggggcaca 600  
 agctggagta cagcttcaac agccacaacg tgtacatccg ccccgacaag gccacaacg 660  
 gcctggaggc taacttcaag acccgccaca acatcgaggg cggcggcgtg cagctggccg 720  
 accactacca gaccaacgtg cccctgggcg acggcccctg gctgatcccc atcaaccact 780  
 acctgagcac tcagaccaag atcagcaagg accgcaacga ggcccgcgac cacatggtgc 840  
 tcctggagtc cttcagcgcg tgctgccaca cccacggcat ggacgagctg tacagggccg 900  
 gctccggtac cgatgatgat atcgccgcg tcgtcgtcga caacggcagc ggcattgtgca 960  
 aggcgcgctt cgcgggcgac gatgcccccc gggccttctt cccctccatc gtggggcgcc 1020  
 ccaggcacca ggtaggggag ctggctgggt ggggcagccc cgggagcggg cgggaggcaa 1080  
 gggcgctttc tctgcacagg agcctcccgg tttccggggt ggggctgcg cccgtgctca 1140  
 gggcttcttg tcctttcctt cccagggcgt gatggtgggc atgggtcaga a 1191

- <210> 17
- 5 <211> 1192
- <212> ADN
- <213> Secuencia artificial
- <220>
- 10 <223> SINTETIZADO
- <400> 17



ES 2 565 216 T3

ggacgcctcc gaccagtgtt tgccttttat ggtaataacg cggccggccc ggettccttt 60  
 gtccccaatc tggggcgcg cgggcgcccc ctggcggcct aaggactcgg cgcgccggaa 120  
 gtggccaggg cggggcgac ctgggctcac agcgcgcccg gctattctcg cagctcacca 180  
 tgggatccgt gtctaagggc gaagagctga ttaaggagaa catgcacatg aagctgtaca 240  
 tggagggcac cgtgaacaac caccacttca agtgcacatc cgagggcgaa ggcaagccct 300  
 acgagggcac ccagaccatg agaatcaagg tggtcgaggg cggccctctc cccttcgct 360  
 tgcacatcct ggctaccagc ttcattgtacg gcagcagaac cttcatcaac cacaccagg 420  
 gcatccccga cttctttaag cagtccctcc ctgagggctt cacatgggag agagtacca 480  
 catacgaaga cggggcgctg ctgaccgcta cccaggacac cagcctccag gacggctgcc 540  
 tcatctacaa cgtcaagatc agaggggtga acttcccac ccaacggccct gtgatgcaga 600  
 agaaaacact cggctgggag gccaacaccg agatgctgta ccccgctgac ggcgccctgg 660  
 aaggcagaag cgacatggcc ctgaagctcg tgggcggggg ccacctgatc tgcaacttca 720  
 agaccacata cagatccaag aaaccgcta agaacctcaa gatgcccggc gtctactatg 780  
 tggaccacag actggaaga atcaaggagg ccgacaaaga gacctacgtc gagcagcacg 840  
 aggtggctgt ggccagatac tgcgacctcc ctagcaaact ggggcacaaa cttaatgccg 900  
 gctccggtac cgatgatgat atcgccgcgc tcgtcgtcga caacggcagc ggcatgtgca 960  
 aggccggctt cgcggcgac gatgcccccc gggcctctt cccctccac gtggggcgcc 1020  
 ccaggcacca ggtaggggag ctggctgggt ggggcagccc cgggagcggg cgggagggaa 1080  
 gggcgcttct tctgcacagg agcctcccgg tttccggggt gggggctgcg cccgtgctca 1140  
 gggcttcttg tccttccct cccagggcgt gatggtgggc atgggtcaga ag 1192

5 <210> 18  
 <211> 15  
 <212> ADN  
 <213> *Homo sapiens*

10 <400> 18  
 cctcgcgcc cgcct 15

15 <210> 19  
 <211> 16  
 <212> ADN  
 <213> *Homo sapiens*

<400> 19  
 gccgcccgcc atggcg 16

20 <210> 20  
 <211> 37  
 <212> ADN  
 <213> *Homo sapiens*

25 <400> 20  
 cctcgcgcc cgcctgtctc cgcccccgc catggcg 37

ES 2 565 216 T3

<210> 21  
 <211> 1996  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> SINTETIZADO

10 <400> 21

<b>ccctgagcct ggtccgggaa ccgcccagcc gggagggccg agctgacggt tgcccaaggg</b>	<b>60</b>
<b>ccagatttta aatttacagg cccggccccc gaaccgccga agcgcgctgc ctgctcccca</b>	<b>120</b>
<b>ttggcccatg gtagtcacgt ggaggcgcgc gggcgtgccg gccatgttgg ggagtgcggc</b>	<b>180</b>
<b>gccgcggccc gcgccacctc cgcgcccccgc ggcttgccctc cagcccgcc ctcceggccc</b>	<b>240</b>
<b>tcctcccccc gcccgccgct ccgtgcagcc tgagaggaaa caaagtgctg cgagcaggag</b>	<b>300</b>
<b>acggcggcgg cgcgaaccct gctgggcctc cagtcaacct cgtcttgcct tttcccgct</b>	<b>360</b>
<b>gcgtgtgtga gtgggtgtgt gtgttttctt acaaagggtta tttcgcgac gatcgattga</b>	<b>420</b>
<b>ttcgtagttc cccccgcgc gcctttgcc tttgtgctgt aatcgagctc ccgccatccc</b>	<b>480</b>
<b>aggtgcttct ccgttctct aaacgccagc gtctggacgt gagcgcaggt cgcgggttg</b>	<b>540</b>
<b>tgcttcggt ccccgcttcg ccccctgcc tcccctcctt atcacggtcc cgcctcgggc</b>	<b>600</b>
<b>ctcgcgcccc cgctgtctcc gccgccgcc atgggatccg ggggcgagga gctgttcgcc</b>	<b>660</b>

ES 2 565 216 T3

ggcatcgtgc ccgtgctgat cgagctggac ggcgacgtgc acggccacaa gttcagcgtg 720  
 cgcggcgagg gcgagggcga cgccgactac ggcaagctgg agatcaagtt catctgcacc 780  
 accggcaagc tgcccgtgcc ctggcccacc ctggtgacca ccctctgcta cggcatccag 840  
 tgcttcgccc gctaccccga gcacatgaag atgaacgact tcttcaagag cgccatgccc 900  
 gagggctaca tccaggagcg caccatccag ttccaggacg acggcaagta caagacccgc 960  
 ggcgaggtga agttcgaggg cgacaccctg gtgaaccgca tcgagctgaa gggcaaggac 1020  
 ttcaaggagg acggcaacat cctgggccac aagctggagt acagcttcaa cagccacaac 1080  
 gtgtacatcc gccccgacaa ggccaacaac ggcctggagg ctaacttcaa gacccgccac 1140  
 aacatcgagg gcggcggcgt gcagctggcc gaccactacc agaccaacgt gccctgggc 1200  
 gacggccccg tgctgatccc catcaaccac tacctgagca ctacagaccaa gatcagcaag 1260  
 gaccgcaacg agggcccgga ccacatggtg ctctggagt ccttcagcgc ctgctgccac 1320  
 acccacggca tggacgagct gtacagggga tctggatcag gtaccgagc tgcgaccccc 1380  
 gtgcccgcgc ggatgggcag ccgcgctggc ggccccacca cgccgctgag ccccacgcgc 1440  
 ctgtcgcggc tccaggagaa ggaggagctg cgcgagctca atgaccggt ggcggtgtac 1500  
 atcgacaagg tgcgcagcct ggagacggag aacagcgcgc tgcagctgca ggtgacggag 1560  
 cgcgaggagg tgccgggccg tgagctcacc ggcctcaagg cgctctacga gaccgagctg 1620  
 gccgacgcgc gacgcgcgct cgacgacacg gcccgcgagc gcgccaagct gcagatcgag 1680  
 ctgggcaagt gcaaggcgga acacgaccag ctgctcctca agtgagtgtct agctggcggc 1740  
 cgcgttagcg ccaaggaggg gcgggggcgc aaccgcggcg accagctcac cgggttctgc 1800  
 cgtggggagg gagcagaggc caggatgcac gcgtccttct gaaggaacag ggtctcggtc 1860  
 tccggaaaag agaaagaatc tagagttcat agcggagcag ggtcgcgga gggggctcga 1920  
 gctgtagcgc tggggggccg tgatgcccat ttctagattt tggatacccg ctgggacgtg 1980  
 gtaagtgcgc gcctgg 1996

5 <210> 22  
 <211> 15  
 <212> ADN  
 <213> *Homo sapiens*

10 <400> 22  
 tacctgggtc tggac 15

15 <210> 23  
 <211> 18  
 <212> ADN  
 <213> *Homo sapiens*

20 <400> 23  
 agtgtgaacc agaaggcc 18

<210> 24  
 <211> 69

ES 2 565 216 T3

	<212> ADN	
	<213> <i>Homo sapiens</i>	
	<400> 24	
5	<b>atggatgatg atatacggcgc gctcgtcgtc gacaacggct ccggcatgtg caaggccggc</b>	<b>60</b>
	<b>ttcgcgggc</b>	<b>69</b>
	<210> 25	
	<211> 18	
10	<212> ADN	
	<213> <i>Homo sapiens</i>	
	<400> 25	
15	cacaccaaca actgccca	18
	<210> 26	
	<211> 15	
	<212> ADN	
	<213> <i>Homo sapiens</i>	
20	<400> 26	
	ggagaaggag gaaga	15
	<210> 27	
25	<211> 51	
	<212> ADN	
	<213> <i>Homo sapiens</i>	
	<400> 27	
30	atggccaat ggaatcagct acagcagctt gacacacggg acctggagca g	51
	<210> 28	
	<211> 3343	
	<212> ADN	
35	<213> Secuencia artificial	
	<220>	
	<223> SINTETIZADO	
40	<400> 28	
	<b>ggtctgcccc ccatcactat tgggcatcgg gtgagcactg atgagcattt tggacttagg</b>	<b>60</b>
	<b>agatattttc tctaaccctc tagaaaacca ccacaactcc aggaaggaaa ccaaggggca</b>	<b>120</b>
	<b>gacccaaaaa actggtaggt gaagaagcag actgctgctt gcctcctggg ctcttttgag</b>	<b>180</b>
	<b>ttgaggggtg tgagtacaca gtacctgatg ctatgcaccc cctatggaaa ggctctcctt</b>	<b>240</b>
	<b>gacctgctgg gacatcagat ttacagaag tccagagagg ggaaggtacc tggcctgggc</b>	<b>300</b>
	<b>tgtgcccattg agaagtgagg ggtcccaggt ataatcaga ccacatcccc ctgccctgcc</b>	<b>360</b>
	<b>ctgccctagt tgtgtgtggg ggtcccctcc tctcctgctc ctagaatact cagaacttct</b>	<b>420</b>
	<b>aggggagatc ttggaagtca tctagcctgt gtccccctca attagagatg aggaaaggaa</b>	<b>480</b>
	<b>gccatagggg gaaggtttgt ccttctctatg agcctctgca gaagagaaac agcgaaggag</b>	<b>540</b>

ES 2 565 216 T3

ctgggcccctg ggaggggtcg gtgctggagt tctgatgtga cccaccacac tgcaactggag 600  
 ggcacatcc aattctgggt ccccaaacag ctggtgaaa ggctcgggtg gctgagtcaa 660  
 gaagctgcct ctagggggcc actgcagtta gggtcacccc agccttcag ctctggccc 720  
 tctcctaccc ccagcctgcc ccctcaaata cctgaagctg tcattccttg agctgagcca 780  
 ctgctggggt gggggggtta gggggtgctg ctggccaggc cccaagagtg agtaacagga 840  
 aacaagttgt tttggagttt gtgcctggca cgggggctgt agccccgtgt ggtgtcccga 900  
 cattcccgcc cagtgagtga gccccggcgg cacacacttc cccttcctcc ccaccccggc 960  
 ctagggtcag ccctcggcca ccccgagggg ccagggcacc acagcacagc atctgcccct 1020  
 gtgggccaag gacctggttc ccctgcaccc accagcgggc tcttgacact tccagccacc 1080  
 ccttcccatt tcctcccca gccacctctt cccccacctc ctcttctccc ctagggagtc 1140  
 agtcacatcc tgaagctcat tgctgccctg agctctgcc tcctgccctc cctgggctg 1200  
 ggggccaagg gggcttggct cctggctctg ggtgagagca gcatgtgtgt ggggttttt 1260  
 cctcctttta aattcttttt atgaatgaag cggggccgtg gaggttgctg agtcacccac 1320  
 aactcagcc ctgactcatc cctcttcagg agagccaggg agtgaggga gggggtggg 1380  
 ccagcctctg ggggtggaag agggggaccg ggccagagct cacaccaaca actgcccacc 1440  
 tcacaggaga aggaggaaga ggagggcatc tcgcaggagt cctcggagga ggagcagga 1500  
 tctggatcag gtgctagcgg gggcgaggag ctgttcgccg gcatcgtgcc cgtgctgatc 1560  
 gagctggacg ggcagctgca cggccacaag ttcagcgtgc gggcgaggg cgagggcgac 1620  
 gccgactacg gcaagctgga gatcaagttc atctgcacca cggcaagct gcccggtgcc 1680  
 tggcccacc tggtgaccac cctctgctac ggcatccagt gcttcgcccg ctaccccag 1740  
 cacatgaaga tgaacgactt cttcaagagc gccatgcccg agggctacat ccaggagcgc 1800  
 accatccagt tccaggacga cggcaagtac aagaccgcg gcgaggtgaa gttcgagggc 1860  
 gacaccctgg tgaaccgat cgagctgaag ggcaaggact tcaaggagga cggcaacatc 1920  
 ctgggcccaca agctggagta cagcttcaac agccacaacg tgtacatccg ccccgacaag 1980  
 gccacaacg gcctggaggc taacttcaag acccgccaca acatcgaggg cggcggtgctg 2040  
 cagctggcgg accactacca gaccaacgtg cccctggcgg acggccccgt gctgatcccc 2100  
 atcaaccact acctgagcac tcagaccaag atcagcaagg accgcaacga ggcccgcgac 2160  
 cacatggtgc tcctggagtc cttcagcgc tgctgccaca cccacggcat ggacgagctg 2220  
 tacaggtgaa gtacctgacc catgctgcc gcctgctcct cactggagga gcagcttctc 2280  
 tctgggactg gacagctttg ctccgtccc accgccccca ccccttccc aggccacca 2340  
 tcaccacggc ctctggcggc ccccccatc ttccacctgt gccctacca ccacactaca 2400  
 cagcacacca gccgctgcag ggctcccatg ggctgagtgg ggagcagttt tcccctggc 2460

ES 2 565 216 T3

tcagttccca gctcccccg cccacccacg catacacaca tgccctcctg gacaaggcta 2520  
 acatcccact tagccgcacc ctgcacctgc tgcgtcccca ctcccttggg ggtggggaca 2580  
 ttgctctctg ggcttttggg ttggggggcg cctctctgct ccttcaactgt tccctctggc 2640  
 ttcccatagt ggggcctggg agggttcccc tggccttaaa aggggcccga gccccatctc 2700  
 atcctggcac gccctactcc actgccttgg cagcagcagg tgtggccaat ggaggggggt 2760  
 gctggccccc aggattcccc cagccaaact gtctttgtca ccacgtgggg ctcaactttc 2820  
 atccttcccc aacttcocct gtccccgtac tagggtggac agcccccttc ggttacagga 2880  
 aggcaggagg ggtgagtccc ctactccctc ttcactgtgg ccacagcccc cttgccctcc 2940  
 gcctgggacg tgagtacata ttgtgggatg ggagatgcag tcaactattg tccagggtgag 3000  
 gcccaagagc cctgtggccg ccacctgagg tgggctgggg ctgctcccct aaccctactt 3060  
 tgcttccgcc actcagccat ttccccctcc tcagatgggg caccaataac aaggagetca 3120  
 ccctgcccgc tcccaacccc cctcctgctc ctccctgcc cccaaggttc tggttccatt 3180  
 tttcctctgt tcacaaacta cctctggaca gttgtgtgtg tttttgttca atgttccatt 3240  
 cttcgacatc cgtcattgct gctgctacca gcgccaaatg ttcatactca ttgcctcctg 3300  
 ttctgcccac gatccccctc cccaagatac tctttgtggg gaa 3343

<210> 29  
 <211> 42  
 <212> ADN  
 <213> *Homo sapiens*

5

<400> 29  
 tgtcgccctc gcctcctaat ccctagccac tatggtagt aa

42

10

**REIVINDICACIONES**

1. Un método de marcaje de al menos una proteína endógena, comprendiendo el método:

- 5 a) introducir en una célula de mamífero (i) al menos una endonucleasa de dirección o ácido nucleico que codifique una endonucleasa de dirección, uniéndose la endonucleasa de dirección a un sitio diana y siendo capaz de escindir un sitio de escisión en una secuencia cromosómica que codifica la proteína endógena, y (ii) al menos un polinucleótido donante que comprende una secuencia de marcaje, la secuencia de marcaje estando  
 10 flanqueada por una secuencia dirección cadena arriba y una secuencia dirección cadena abajo, la secuencia cadena arriba y la secuencia cadena abajo compartiendo una identidad de secuencia sustancial con cualquier lado del sitio de escisión de la secuencia cromosómica; y  
 b) mantener la célula en condiciones que reparan la rotura de doble cadena introducida en el sitio de escisión por la endonucleasa de dirección mediante un proceso de homología dirigida, de modo que la secuencia de marcaje del polinucleótido donante se integra en fase en la secuencia cromosómica codificante de la proteína endógena  
 15 de modo que se produce una proteína endógena marcada;

en el que la proteína endógena se selecciona entre actina, tubulina, lamina, receptor 2 del factor de crecimiento epidérmico humano (HER2) y proteína del grupo de alta movilidad A (HMGA).

20 2. El método de la reivindicación 1, en el que la endonucleasa de dirección es una nucleasa de dedo de cinc.

3. El método de la reivindicación 1, en el que la célula es una célula U2OS humana, una célula MCF10A humana, una célula SKOV3 humana o una célula iPS humana.

25 4. El método de la reivindicación 1, en el que la proteína endógena se marca en el extremo C o en el extremo N.

5. El método de la reivindicación 2, en el que la nucleasa de dedo de cinc se une a una secuencia que tiene al menos aproximadamente 80 % de identidad de secuencia con una secuencia seleccionada entre SEQ ID NO: 13, 14, 1, 2, 18, 19, 22, 23, 25 y 26.

30 6. Una célula de mamífero que comprende al menos una secuencia de marcaje integrada en fase en una secuencia cromosómica que codifica una proteína endógena, de modo que la célula expresa al menos una proteína endógena marcada, en la que la proteína endógena se selecciona entre actina, tubulina, lamina, receptor 2 del factor de crecimiento epidérmico humano (HER2) y proteína del grupo de alta movilidad A (HMGA).

35 7. La célula de la reivindicación 6, en la que la célula es una célula U2OS humana, una célula MCF10A humana, una célula SKOV3 humana o una célula iPS humana.

8. La célula de la reivindicación 6, en la que la proteína endógena se marca en el extremo C o en el extremo N.

40 9. La célula de la reivindicación 6, en el que la célula expresa una o más proteínas endógenas marcadas con fluorescencia.

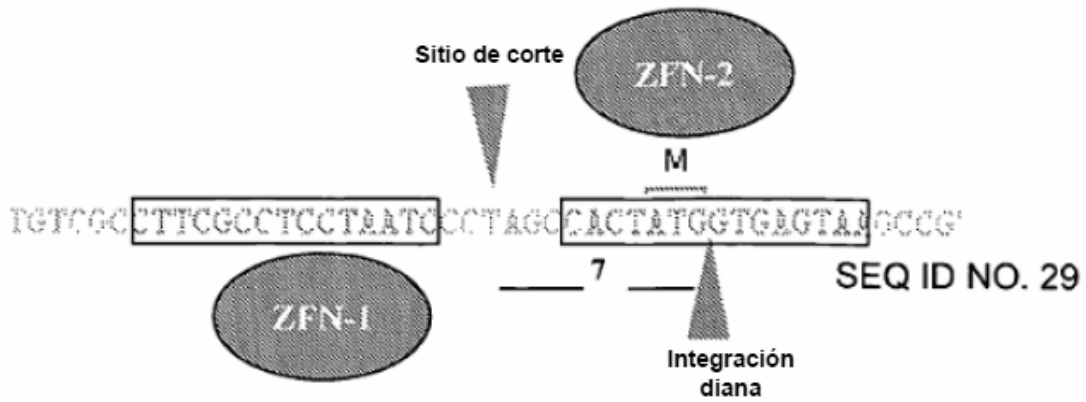
10. La célula de la reivindicación 6, en la que la célula se produce mediante:

- 45 a) introducción en una célula parental de (i) al menos una endonucleasa de dirección o ácido nucleico que codifica una endonucleasa de dirección, uniéndose la endonucleasa de dirección a un sitio diana y siendo capaz de escindir un sitio de escisión en una secuencia cromosómica que codifique la proteína endógena, y (ii) al menos un polinucleótido donante que comprende una secuencia de marcaje, la secuencia de marcaje estando  
 50 flanqueada por una secuencia dirección cadena arriba y una secuencia dirección cadena abajo, la secuencia cadena arriba y la secuencia cadena abajo compartiendo una identidad de secuencia sustancial con cualquier lado del sitio de escisión de la secuencia cromosómica; y  
 b) mantenimiento de la célula en condiciones que reparen la rotura de doble cadena introducida en el sitio de escisión por la endonucleasa de dirección mediante un proceso de homología dirigida, de modo que la secuencia de marcaje del polinucleótido donante se integra en fase en la secuencia cromosómica codificante de la proteína endógena.  
 55

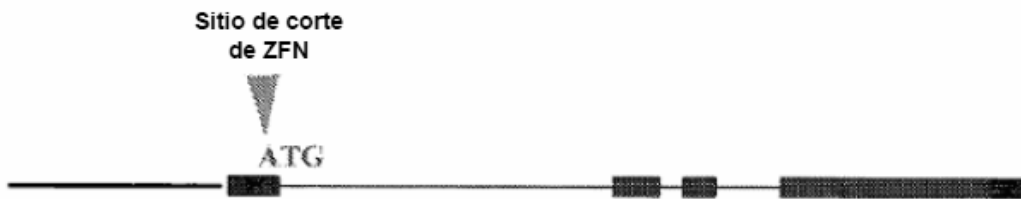
11. La célula de la reivindicación 10, en la que la endonucleasa de dirección es una nucleasa de dedo de cinc.

60 12. La célula de la reivindicación 11, en la que la nucleasa de dedo de cinc se une a una secuencia que tiene al menos aproximadamente 80 % de identidad de secuencia con una secuencia seleccionada entre SEQ ID NO: 13, 14, 1, 2, 18, 19, 22, 23, 25 y 26.

13. El método de la reivindicación 5 o la célula de la reivindicación 12, en donde la identidad de secuencia es de  
 65 aproximadamente el 85 %, 90 %, 95 %, 99 % ó 100 %.



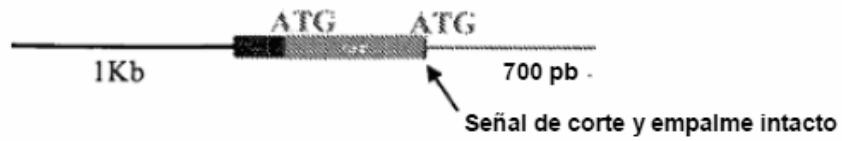
**FIG. 1A**



**FIG. 1B**



**FIG. 1C**



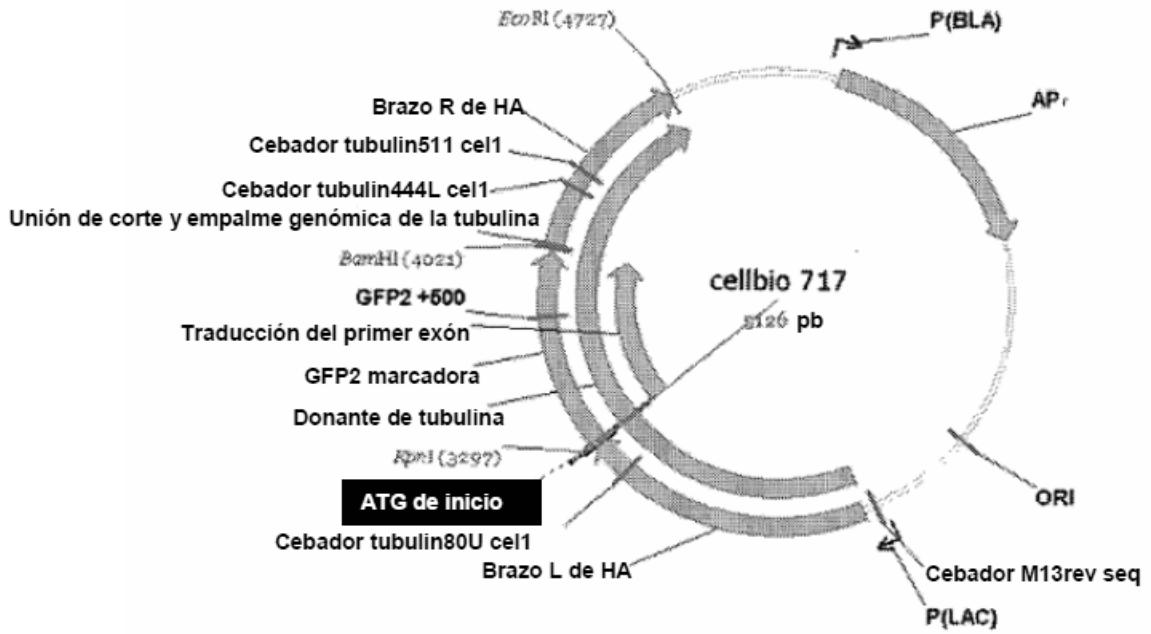
**FIG. 1D**



**FIG. 1E**



### Mapa del donante



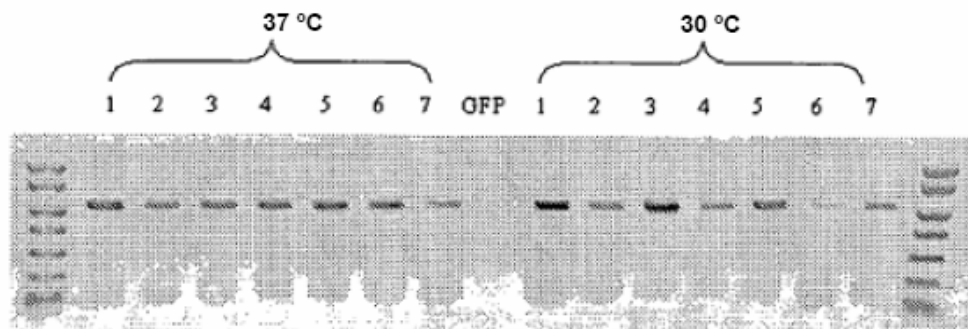
**FIG. 2**

ccgccccctccgcagccgctacttaagaggctccagcgccggccccgccttaqtgctta  
cttacctcgactcttaqcttgcggggacggtaaccgggacccgggtgctgctcctgctgcc  
ttcgcctcctaaticcctagccaactatgggtaccgcaagcgggggcgaggagctgttc  
gccggcatcgtgccgctgctgatcgagctggacggcgacgtgcacggccaca  
agttcaqcgctgcggcgaggcgaggcgacgccgactacggcaagctgg  
agatcaagttcatctgcaccaccggcaagctgccgctgccctggcccaccctgg  
tgaccaccctctgctacggcatccagtgcttcgcccgctaccccgagcacatga  
agatgaacgacttctcaagagcgccatgcccgagggtacatccaggagcgc  
accatccagttccaggacgacggcaagtacaagaccgcggcgagggtgaagtt  
cgagggcgacaccctgggtgaaccgcatcgagctgaagggcaaggacttcaag  
gaggacggcaacatcctgggccacaagctggagtagagcttcaacagccacaa  
cgtgtacatccgccccgacaaggccaacaacggcctggaggctaacttcaaga  
ccgcccacaacatcgagggcgggcggcgtgcagctggccgaccactaccagac  
caacgtgccctggggcgacggccccgtgctgatccccatcaaccactacctga  
gcactcaqaccaagatcagcaaggaccqcaacgagggcccgcaccacatggt  
gctcctggagtccttcagcgctgctgccacaccacggcatggacgagctgta  
cagggcccgtgggatcccactATGgtgagtaagccgtgcggctcccggctgcttca  
gggaagcagggaaaagcgagccggcgggggcctggggccctgtatacagccggga  
agggctggcctcagagccgtccgttgaggggcggaacgaggcgagaggccagg  
gcgggagtggtgagacctcgggtgtgtgtaaatagcgggggcccggaaaggtcgaggg  
gcgccaggatttctctcggactctggaagggatggggggctcgggctgccctccgccgt  
atccggagctctctttgtcgcgtaactgtgtcctgggtgcgggtccc (SEQ ID NO: 4)

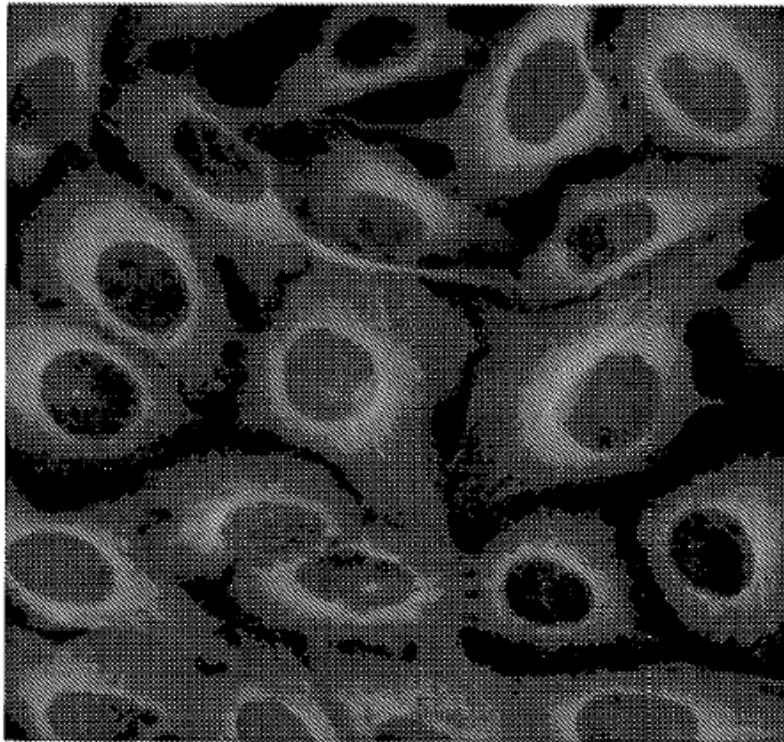
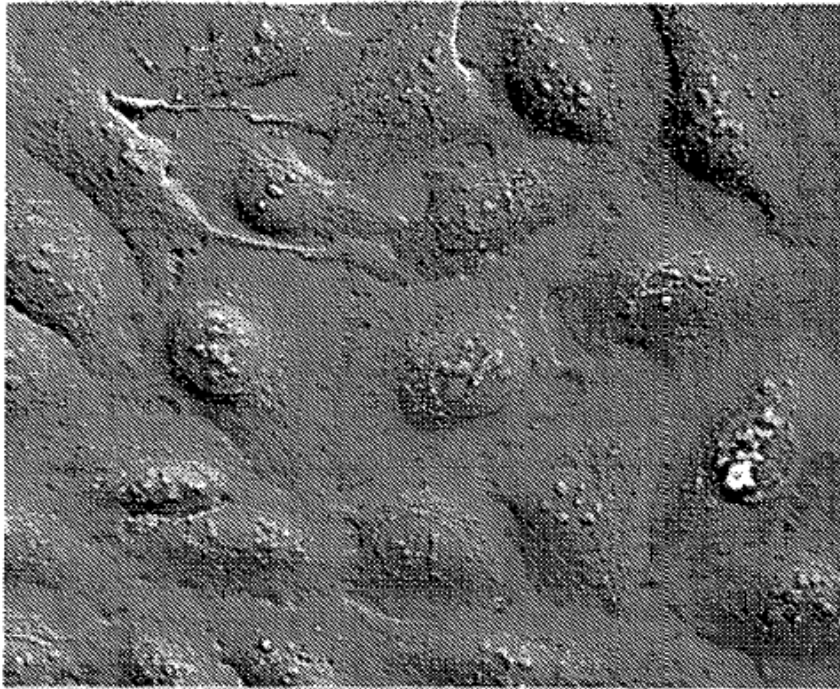
### **FIG. 3**

gcccctccgcagccgctacttaagaagctccagcgccgccccgcccctagtcggttacttacctc  
gactcttagcttgcggggacggtaaccgggacccgggtgctgctcctgctgccttcgctccta  
ccctagccactatgggtaaccgacgtgtctaagggcgaagagctgattaaggagaacatgc  
acatgaagctgtacatggagggcaccgtgaacaaccaccacttcaagtgccacatccga  
gggcgaaggcaagccctacgagggcaccagaccatgagaatcaaggtggtcgagg  
gcggccctctccccttcgcttcgacatcctggctaccagcttcatgtacggcagcagaa  
ccttcatcaaccacaccaggggcatccccgacttctttaagcagtccttccctgagggctt  
cacatgggagagagtcaccacatacgaagacggggcgctgctgaccgctaccagg  
acaccagcctccaggacggctgcctcatctacaacgtcaagatcagaggggtgaactt  
ccatccaacggccctgtgatgcagaagaaaactcggctgggaggccaacaccga  
gatgctgtaccccgctgacggcggcctggaaggcagaagcgacatggccctgaagct  
cgtagggcggggccacctgatctgcaacttcaagaccacatacagatccaagaacc  
gctaagaacctcaagatgcccggtctactatgtggaccacagactggaaagaatca  
aggaggccgacaaagagacctacgtcgagcagcacgaggtggctgtggccagatact  
gcgacctccctagcaaaactggggcacaacttaataccgctggggatcccactATGgtg  
agtaagccgtgcggctcccggctgcttcaagggaaacagggaaaagcgaagccggcggggc  
ctggggccctgtatacagccgggaaaggctggcctcagaagccgtccggttgagggcggaaa  
acgagggcgaagggccagggcgggagtggtgagacctcgggtgtatataaatagcggggccc  
ggaaaggctcagggggcggcaggatctctcggactctggaagggatggggggctcgggctg  
ccctccgccgtatccggagctctctttgtcgcgtaactgtctcttgggtacggtccctcgagt  
 (SEQ ID NO: 5)

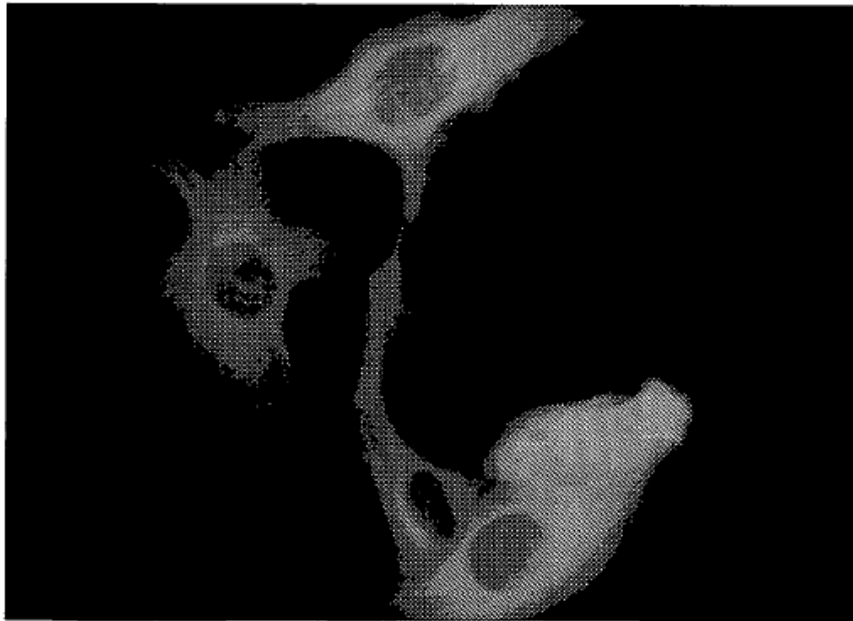
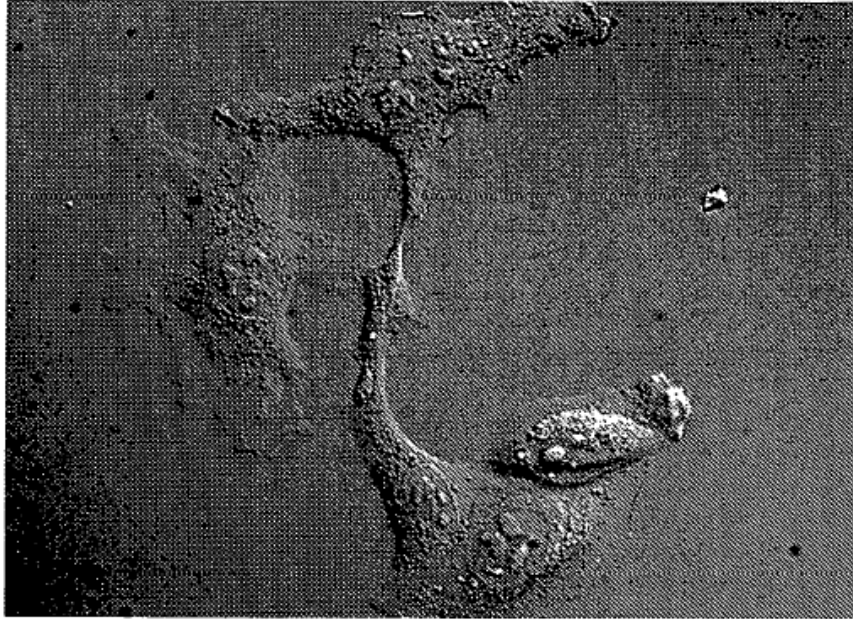
**FIG. 4**



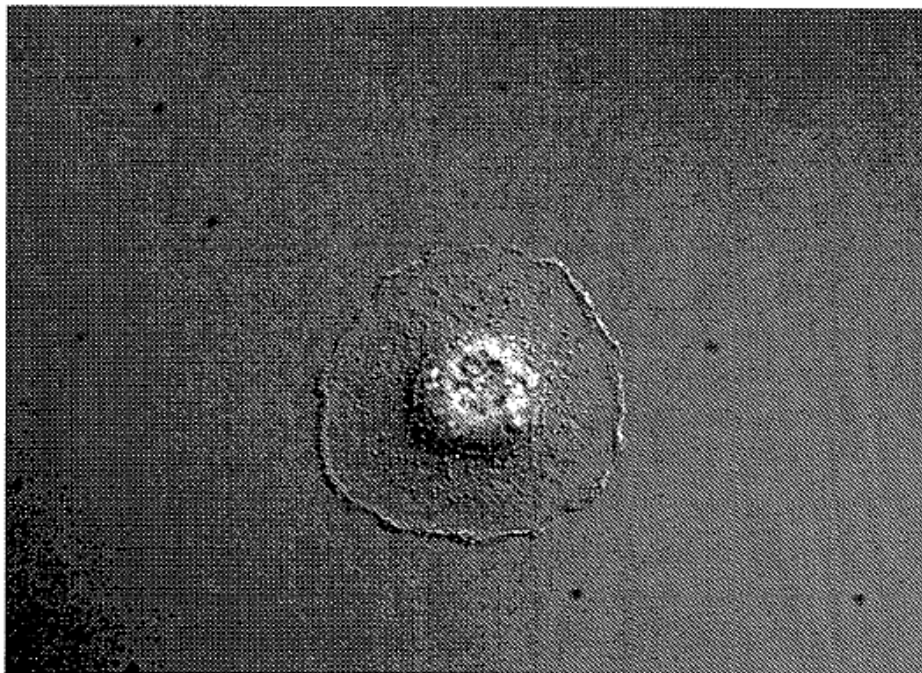
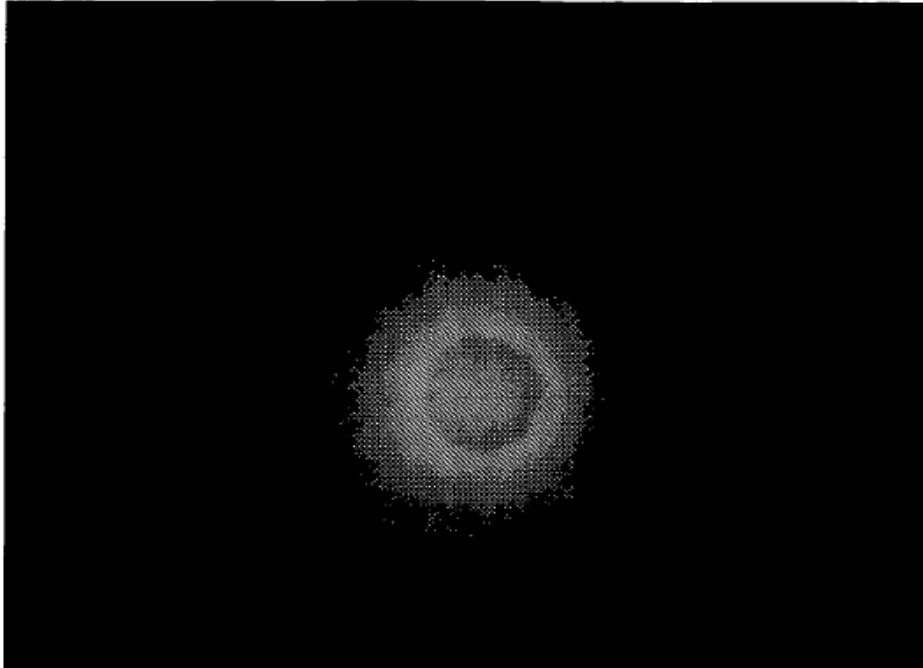
**FIG. 5**



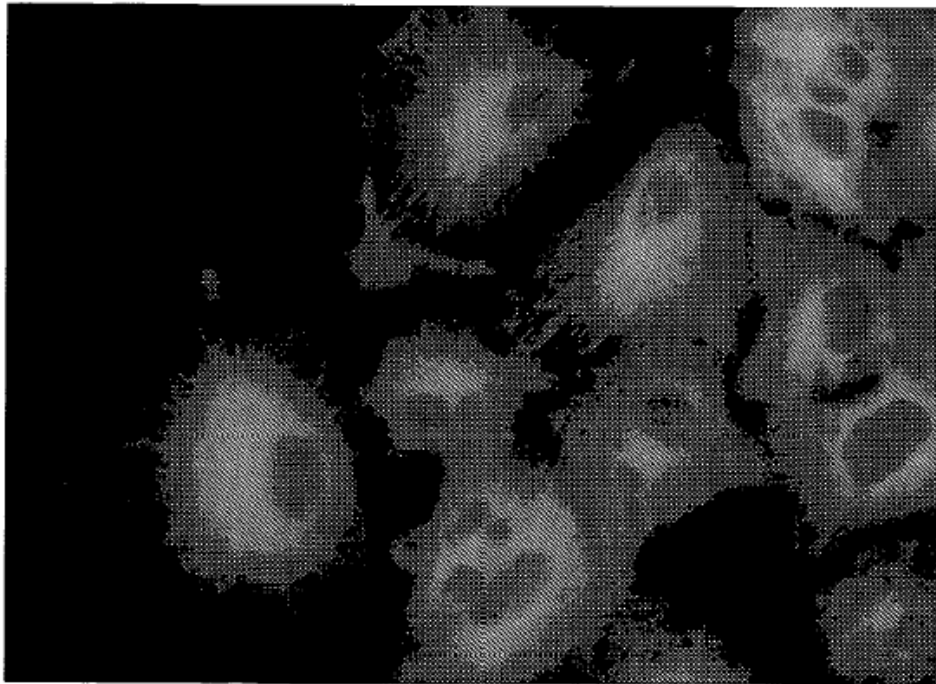
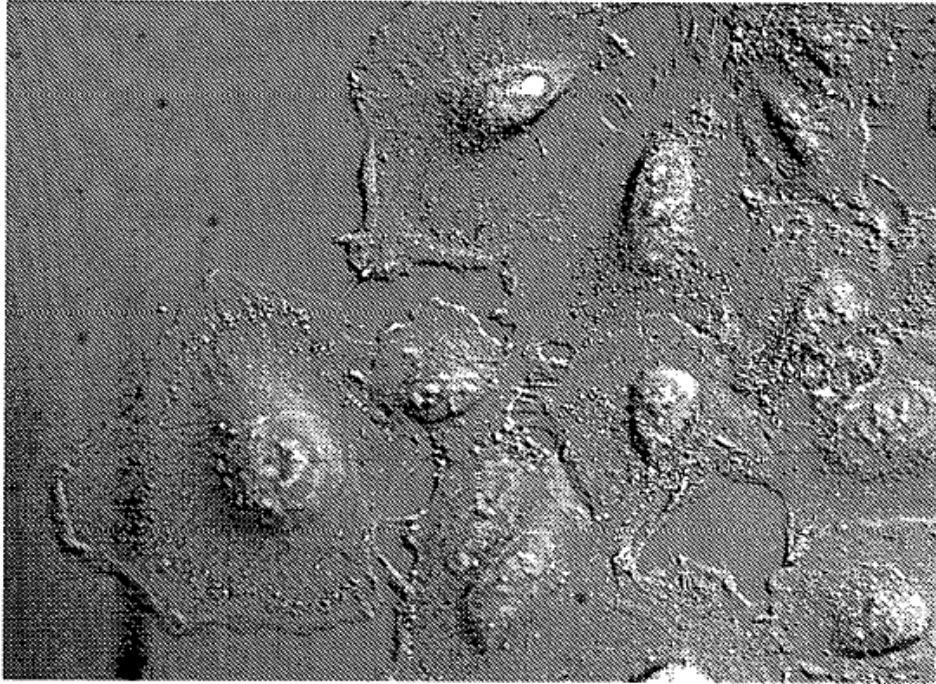
**FIG. 6A**



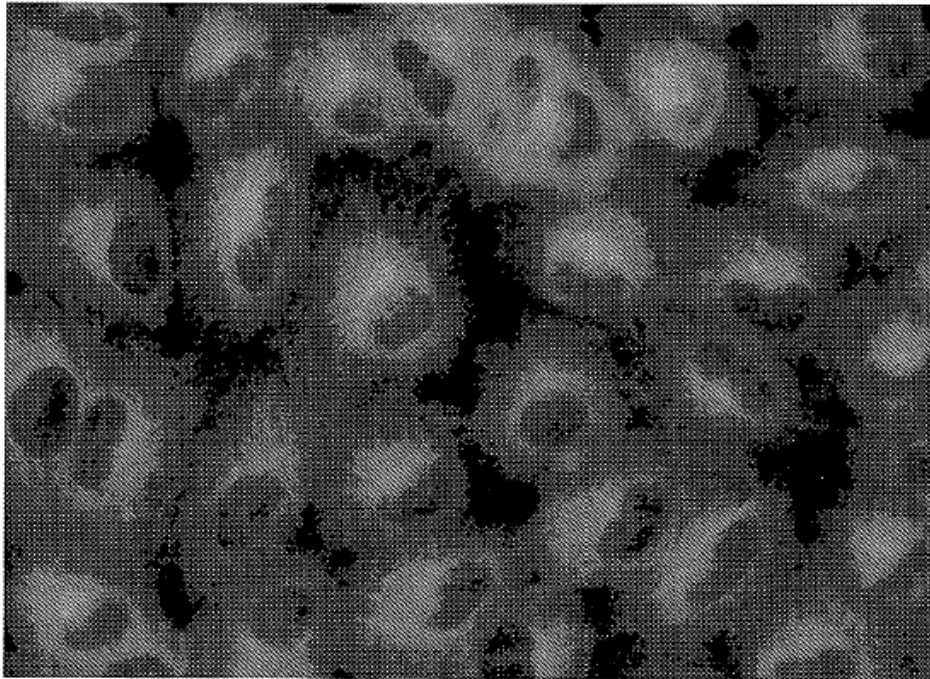
**FIG. 6B**



**FIG. 6C**

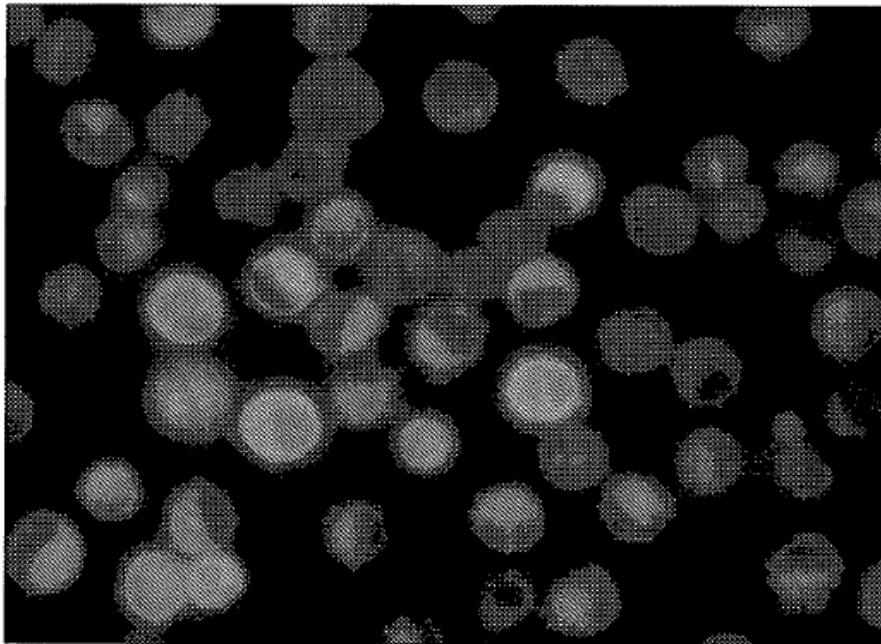
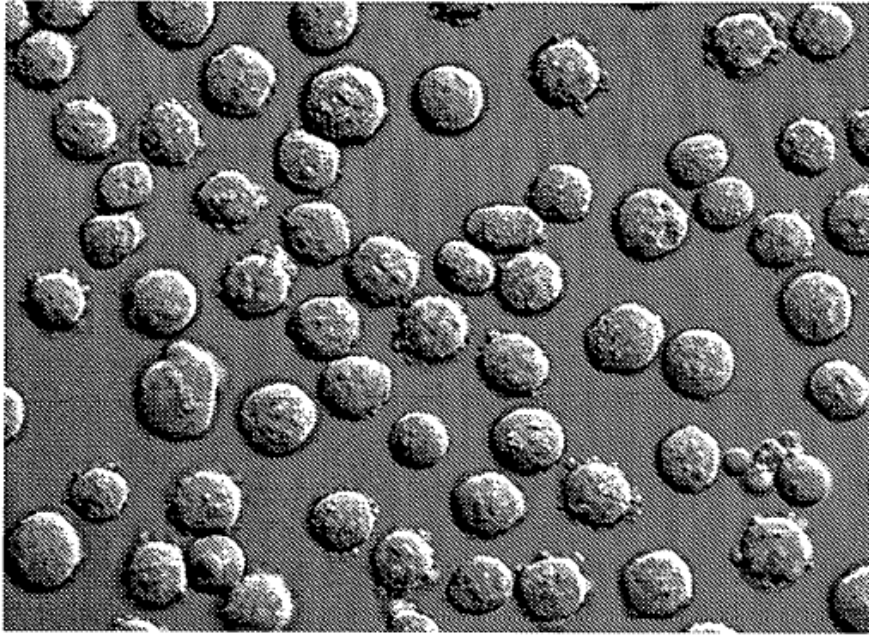


**FIG. 6D**

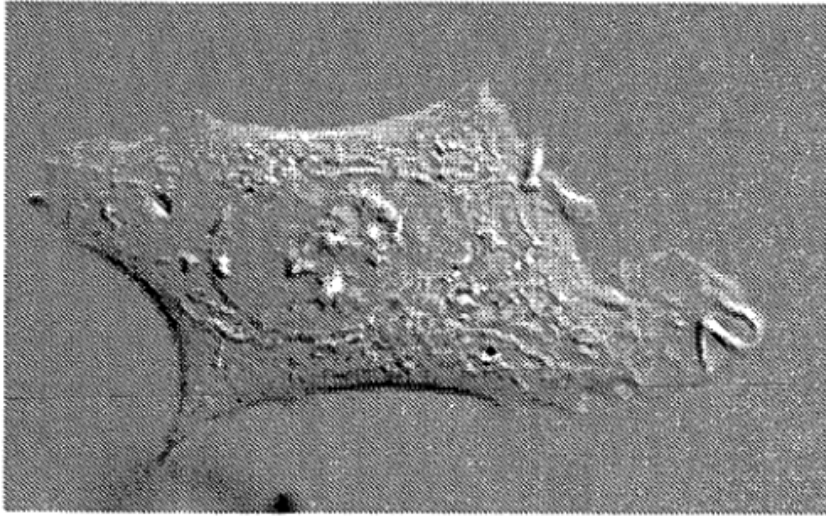


**FIG. 6E**

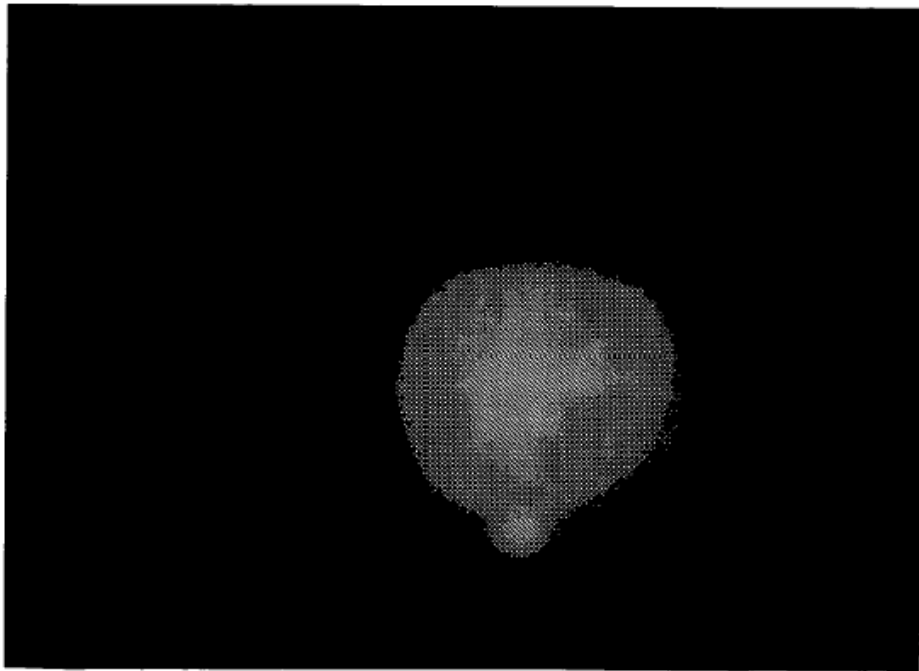
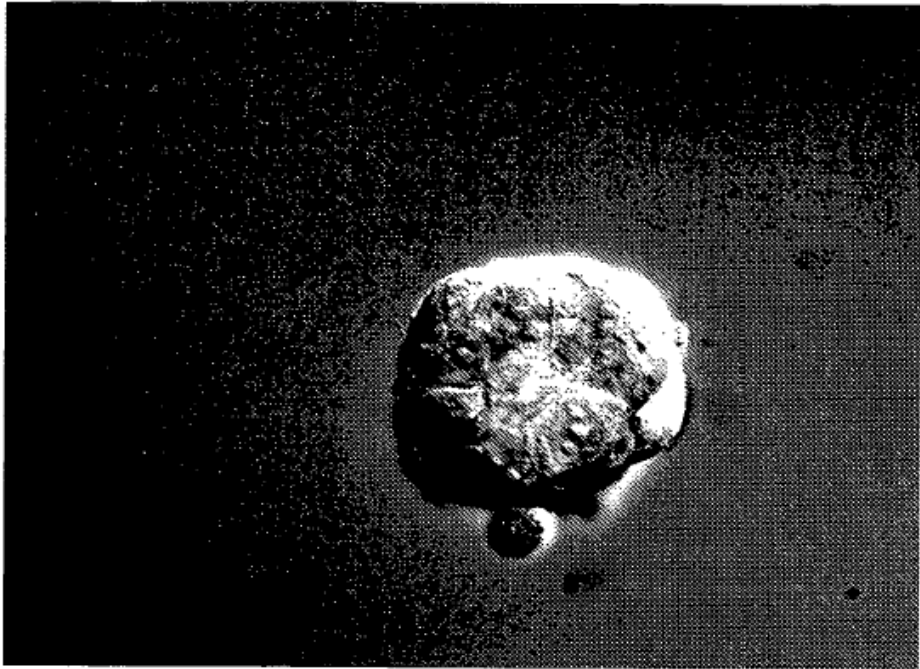




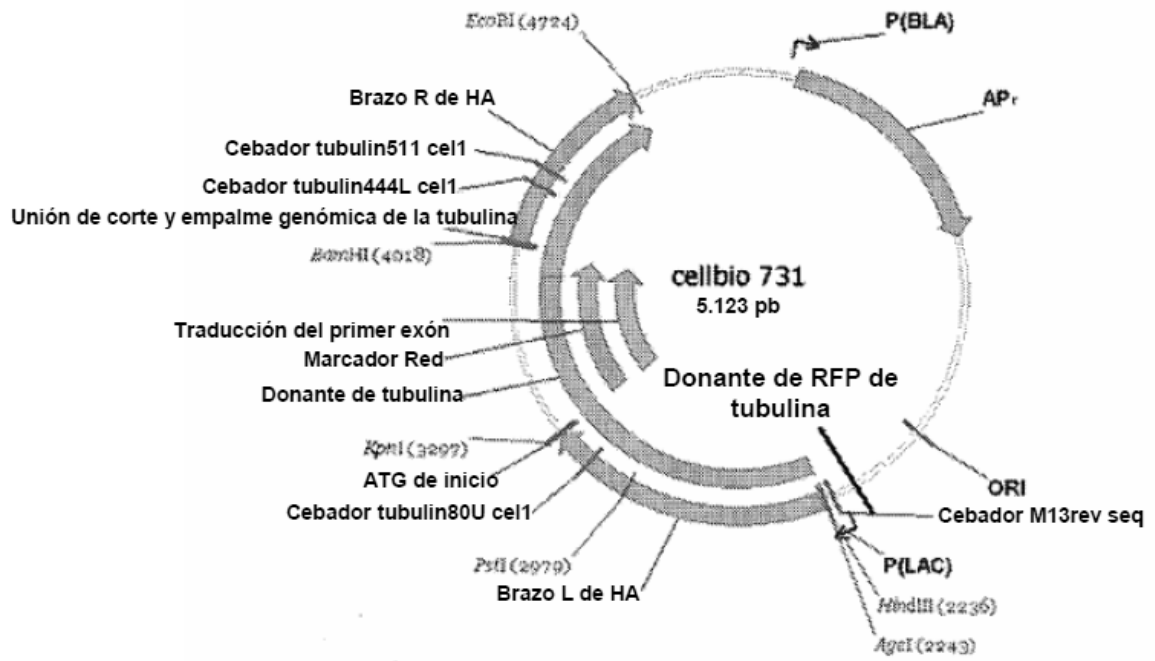
**FIG. 6F**



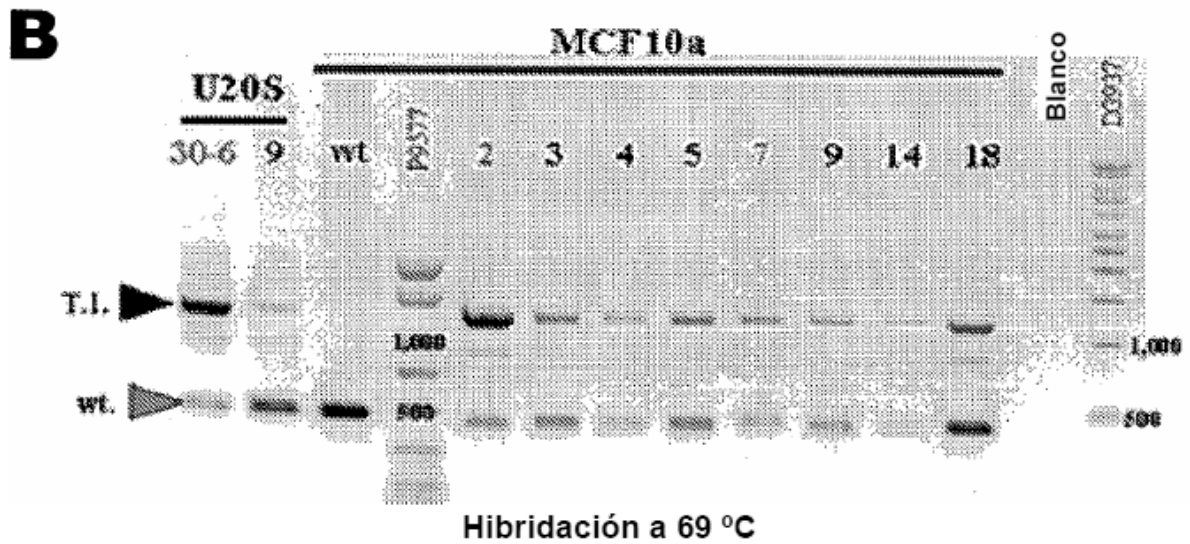
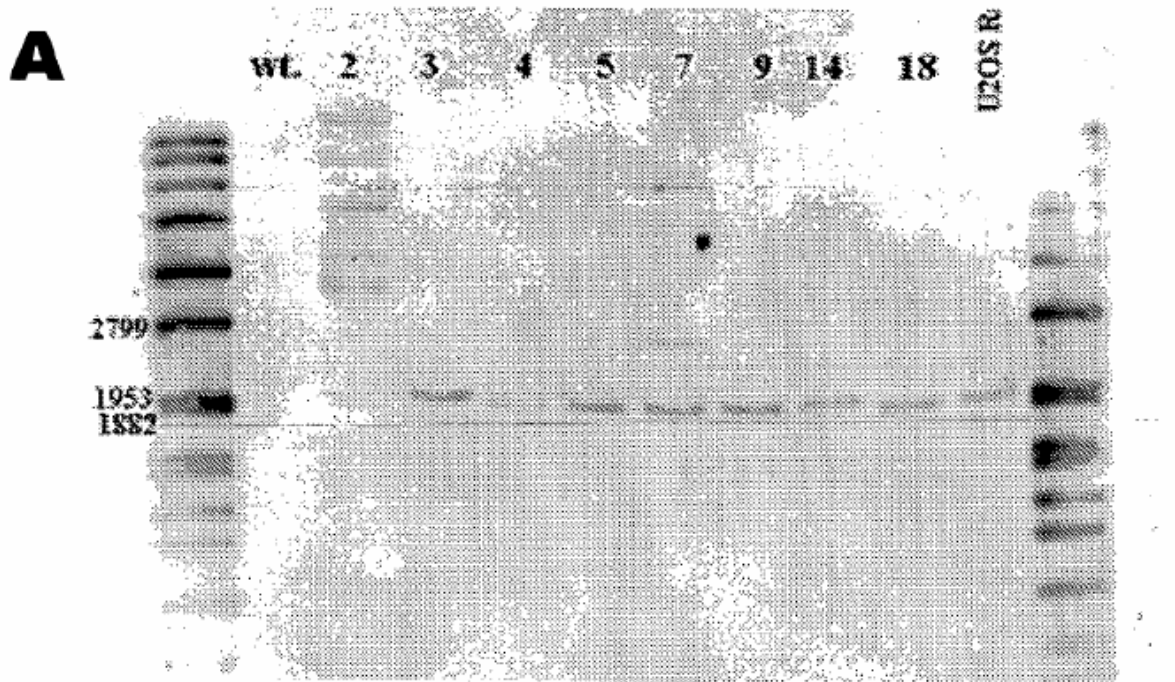
**FIG. 6G**



**FIG. 6H**



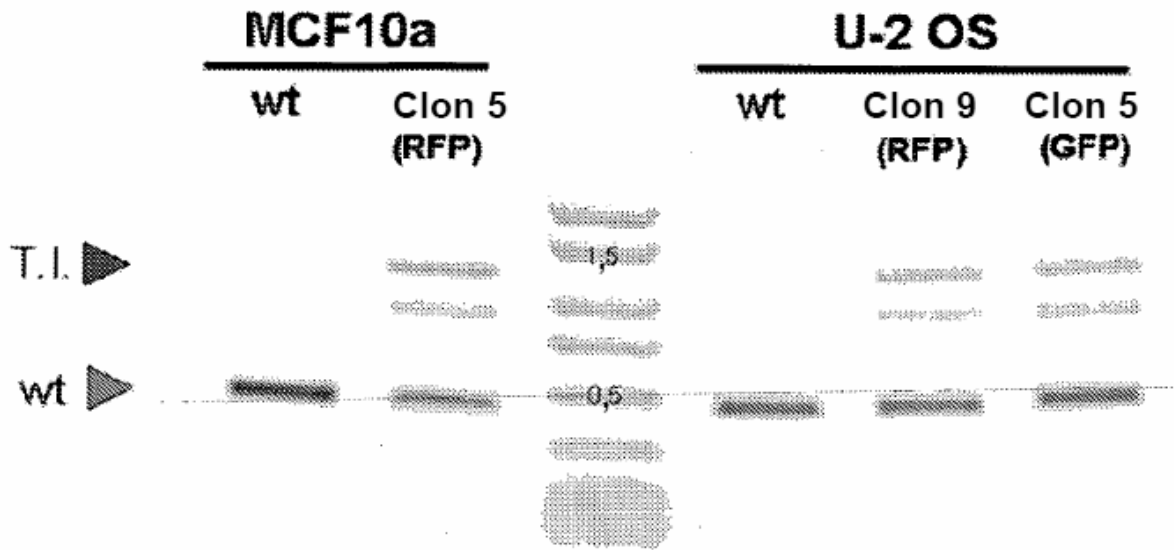
**FIG. 7**



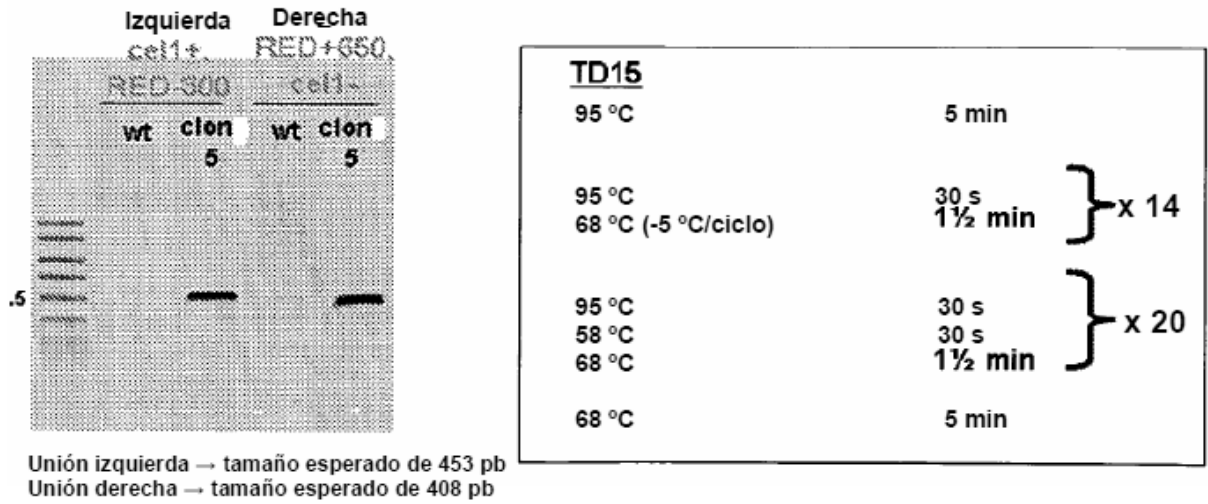
**FIG. 8**

gccccctccgcagccgctacttaagaggctccagcgcggcccccgcctagtcggttacttacctc  
gactcttagcttgcggggacggtaaccgggacccggtgtctgctcctgtcgccttcgcctcctaac  
cctaagccactatgggtaccgcagtgtctaagggcggaaggagctgattaaggagaacatgcac  
atgaagctgtacatggagggcaccgtgaacaaccaccacttcaagtgacacatccgagggg  
cgaaggcaagccctacgagggcaccagaccatgagaatcaaggtggtcgaggggcg  
gccctctccccttcgccttcgacatcctggctaccagcttcatgtacggcagcagaacctt  
catcaaccacaccagggcatccccgacttctttaagcagtccttccctgagggcttcaca  
tgggagagagtcaccacatacgaagacgggggcgtgctgaccgctaccagggacacc  
agcctccaggacggctgcctcatctacaacgtcaagatcagaggggtgaacttcccatcc  
aacggccctgtgatgcagaagaaaacactcggctgggagggccaacaccgagatgctgt  
accccgctgacggcggcctggaaggcagaagcagacatggccctgaagctcgtgggcg  
ggggccacctgatctgcaacttcaagaccacatacagatccaagaaaccgctaagaac  
ctcaagatgccggcgtctactatgtggaccacagactggaaagaatcaaggaggccg  
acaaagagacctacgtcgagcagcagcaggggtggctgtggccagatactgcgacctccc  
tagcaaaactggggcacaacttaatgccgctggggatcccactATGgtgagtaaqccctgc  
ggctcccggctgcttcagggaaagcagggaaaagcagcggccggcggggcgtggggccctgat  
acagccgggaagggtggcctcagagccgtccgtttggagggcggaaaacgagggcgagagg  
ccagggcgggagtggtgagacctcgggtgtgtgtaaatagcggggcccgaaaggtcagggg  
gcgccaggatttctctcggactctggaagggatggggggctcgggctgccctccgccgatccg  
gagctctctttgtcgcgtaactgtgtcctgggtgcggctccctcgagt (SEQ ID NO: 8)

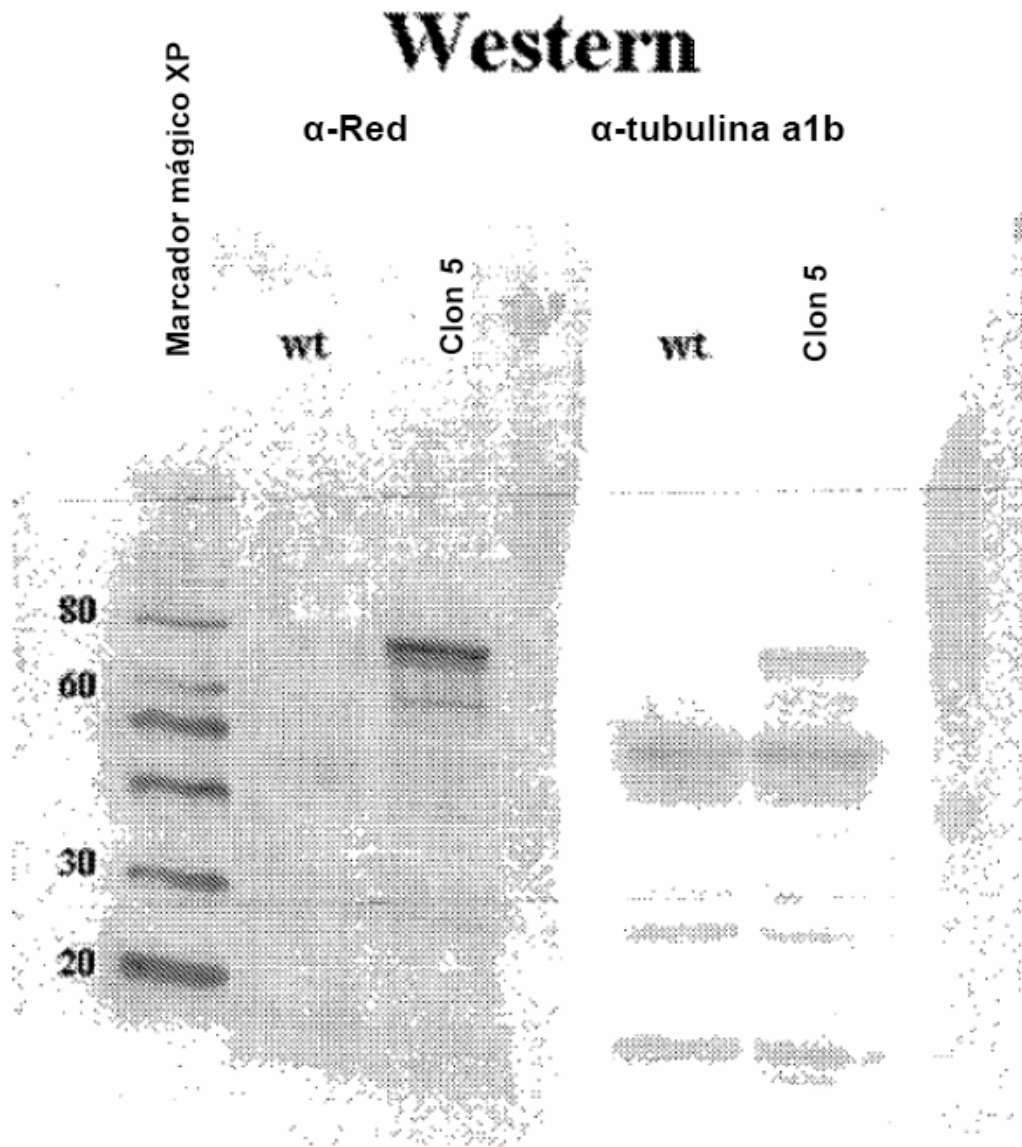
## FIG. 9



**FIG. 10**

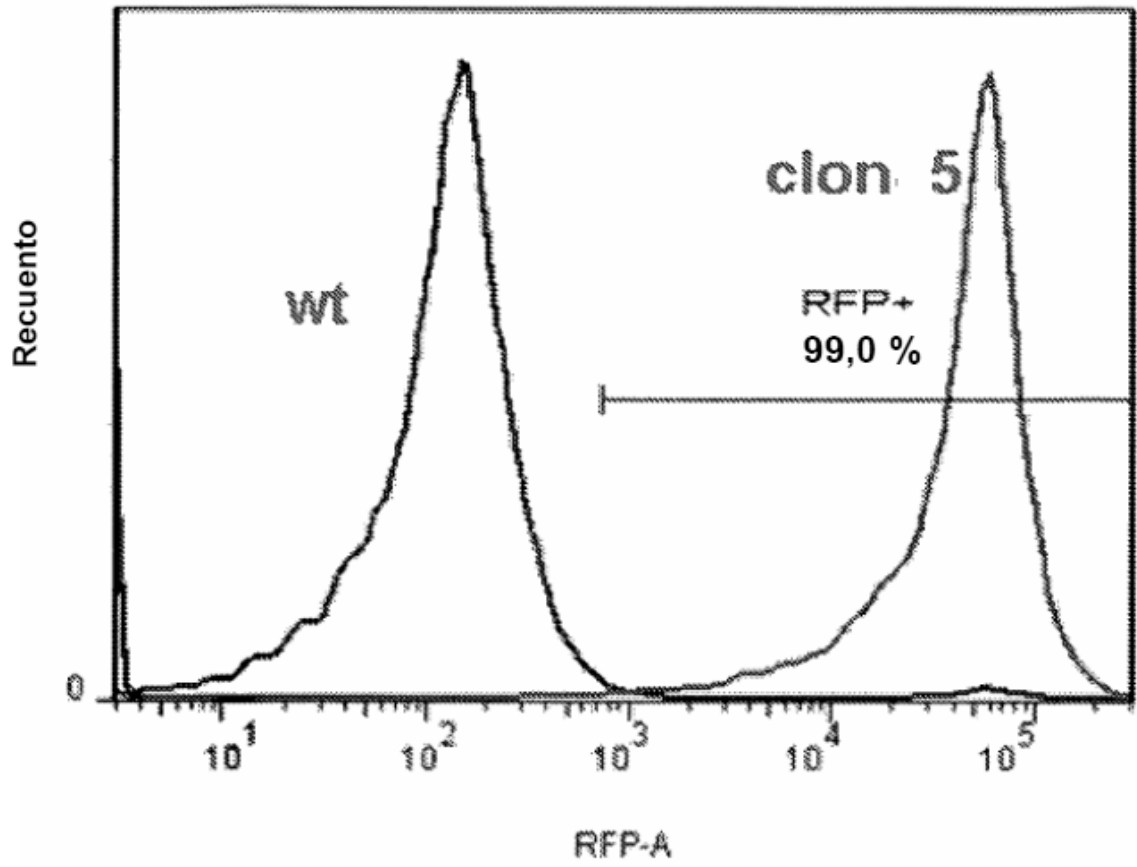


**FIG. 11**



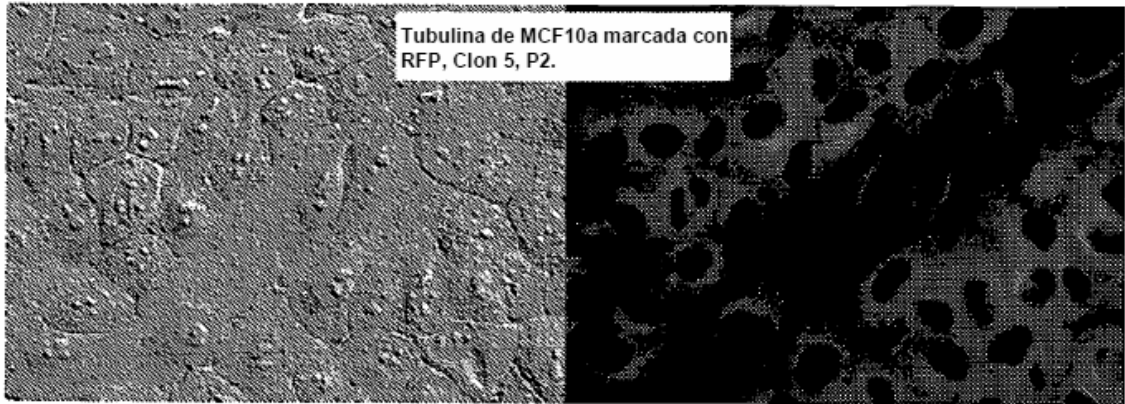
**FIG. 12**



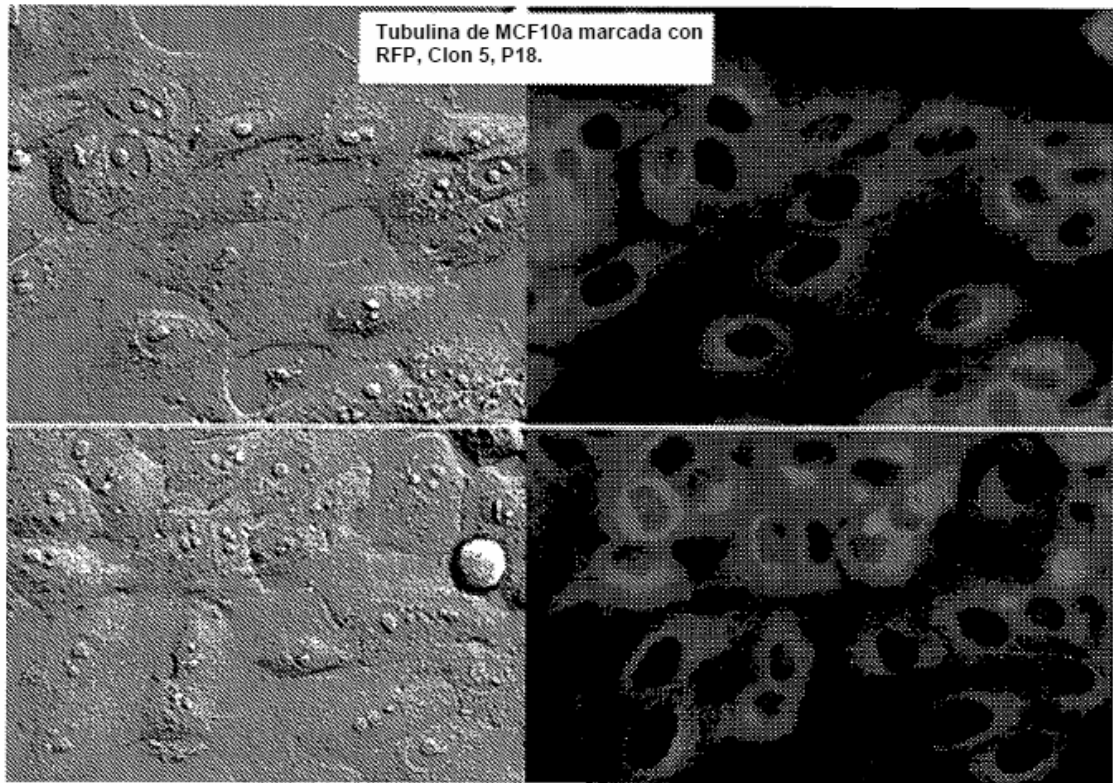


**FIG. 13**

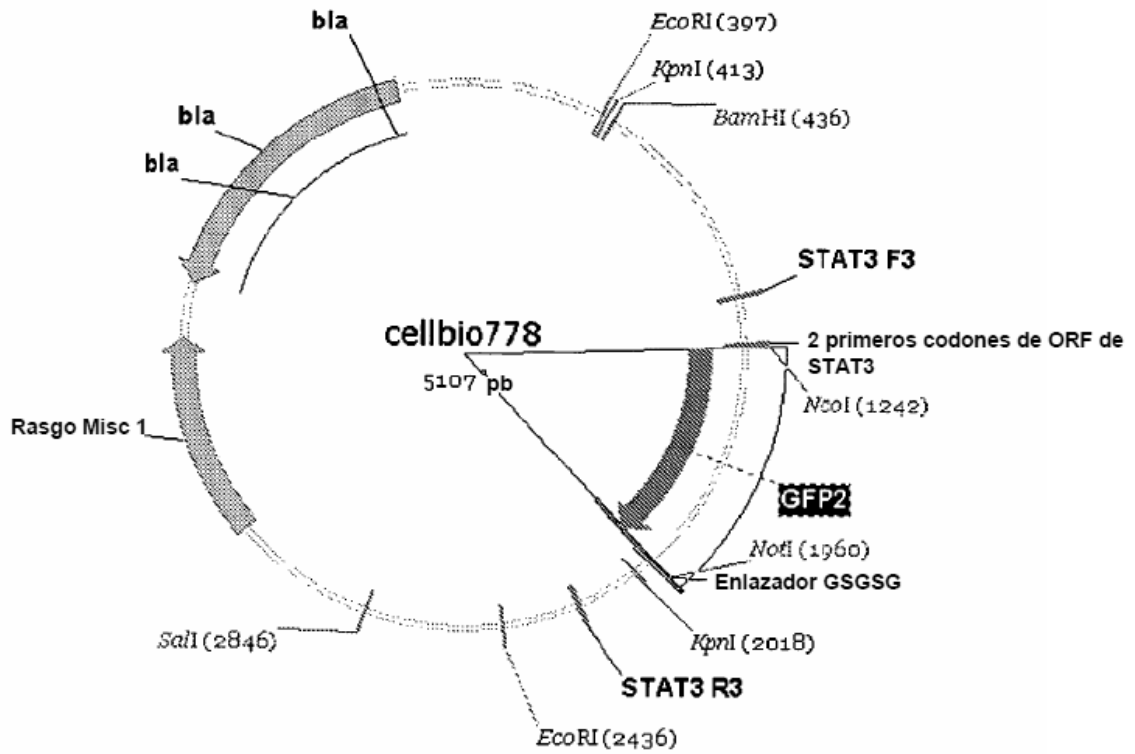
**A**



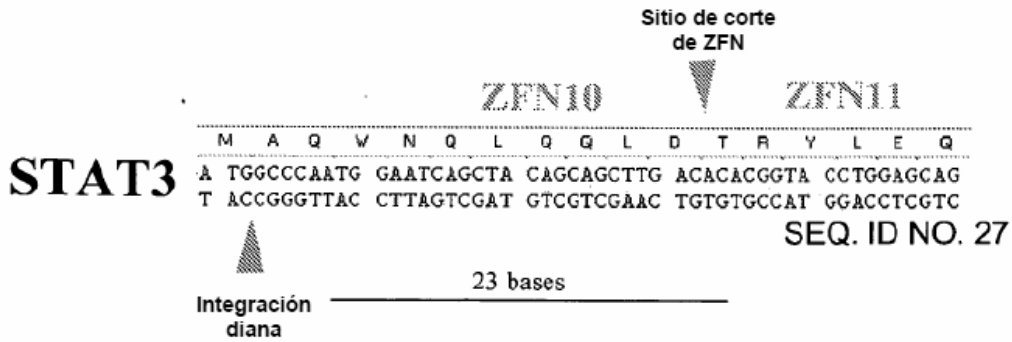
**B**



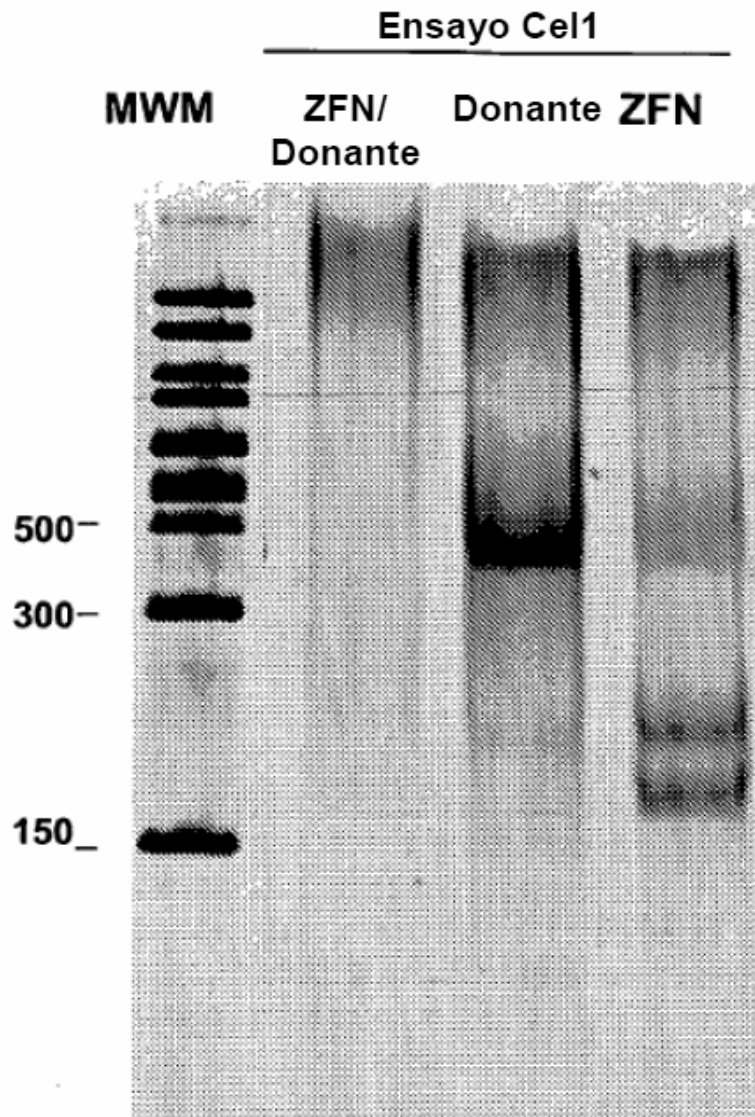
**FIG. 14**



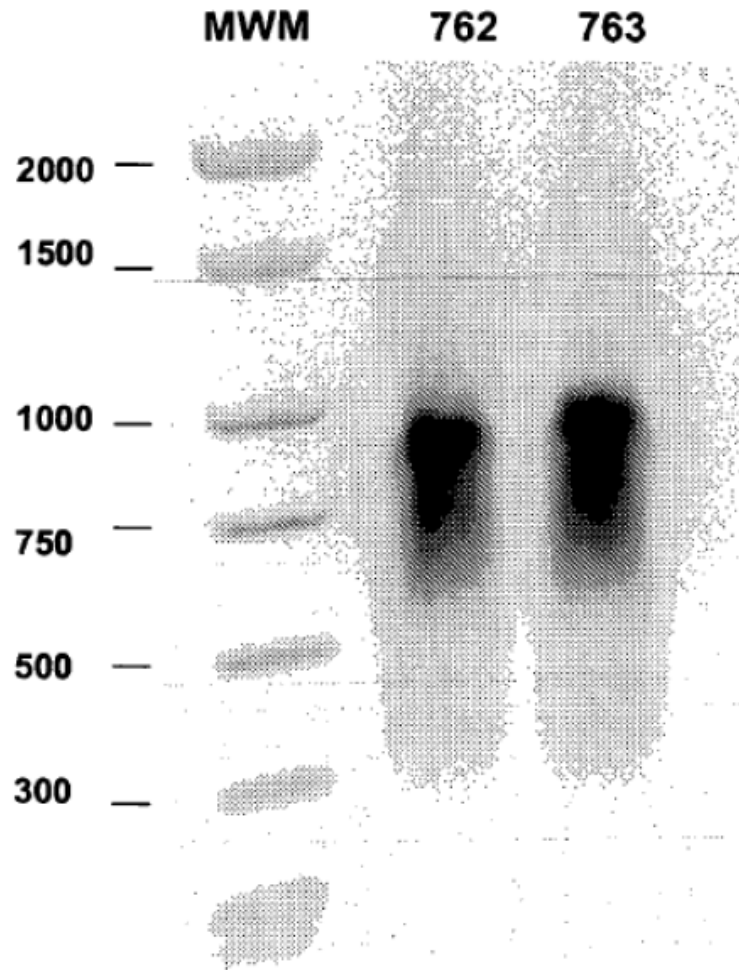
**FIG. 15**



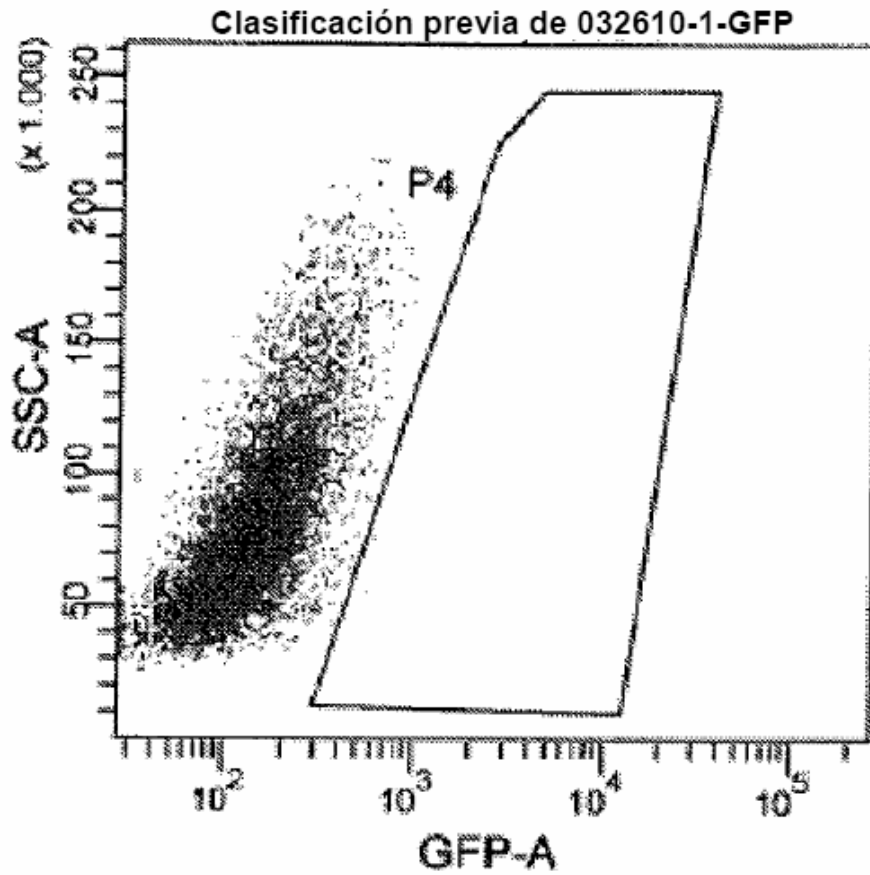
**FIG. 16**



**FIG. 17**



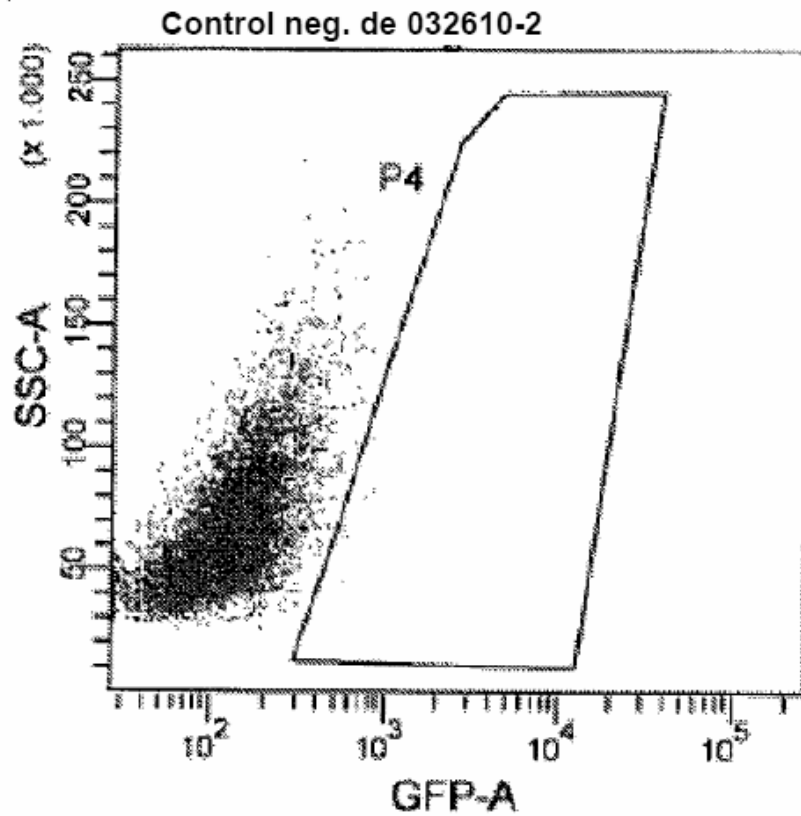
**FIG. 18**



Tubo: clasificación previa de 1-GFP

Población	n.º de acontecimientos	% Parental	% Total
☐ Todos los acontecimientos	11.847		100,0
■ P1	11.084	93,6	93,6
■ P2	10.425	94,1	88,0
■ P3	10.000	95,9	84,4
■ P4	3	0,0	0,0

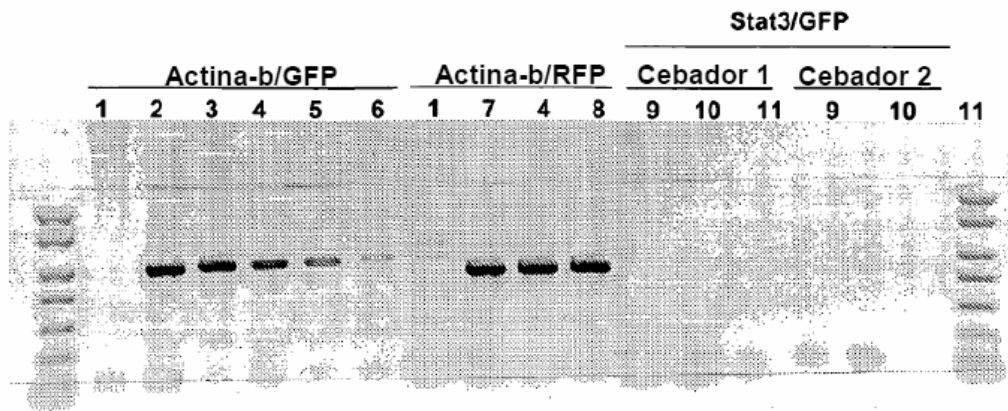
**FIG. 19A**



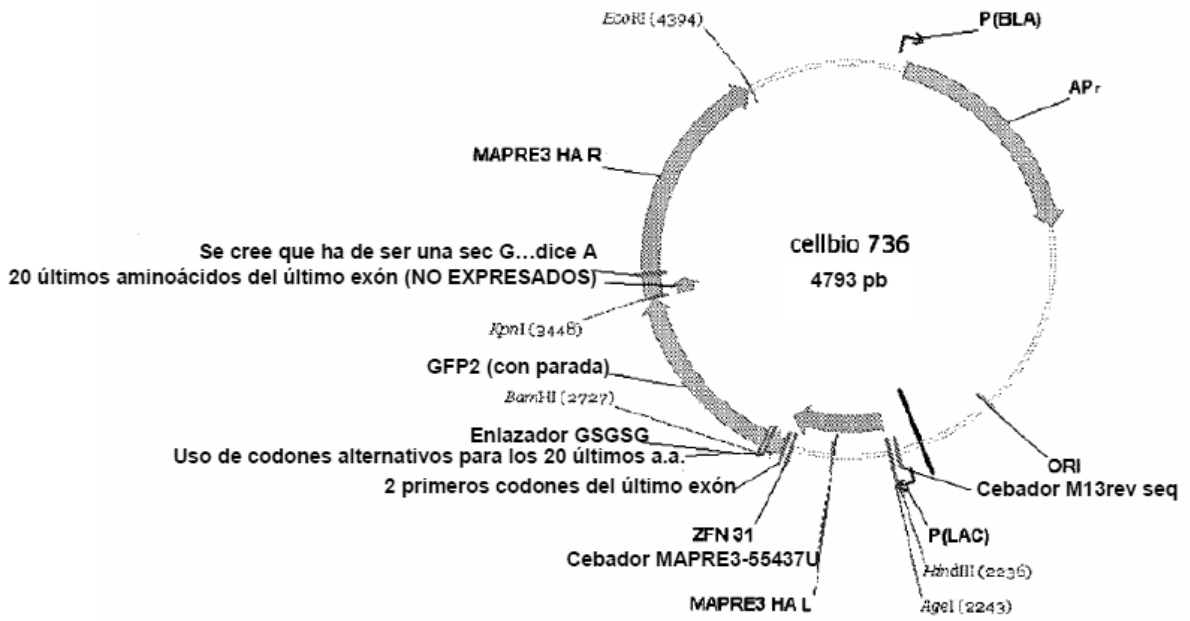
Tubo: Control negativo 2

Población	n.º de acontecimientos	% Parental	% Total
□ Todos los acontecimientos	10.702		100,0
■ P1	10.428	97,4	97,4
■ P2	10.155	97,4	94,9
■ P3	10.000	98,5	93,4
■ P4	9	0,1	0,1

**FIG. 19B**

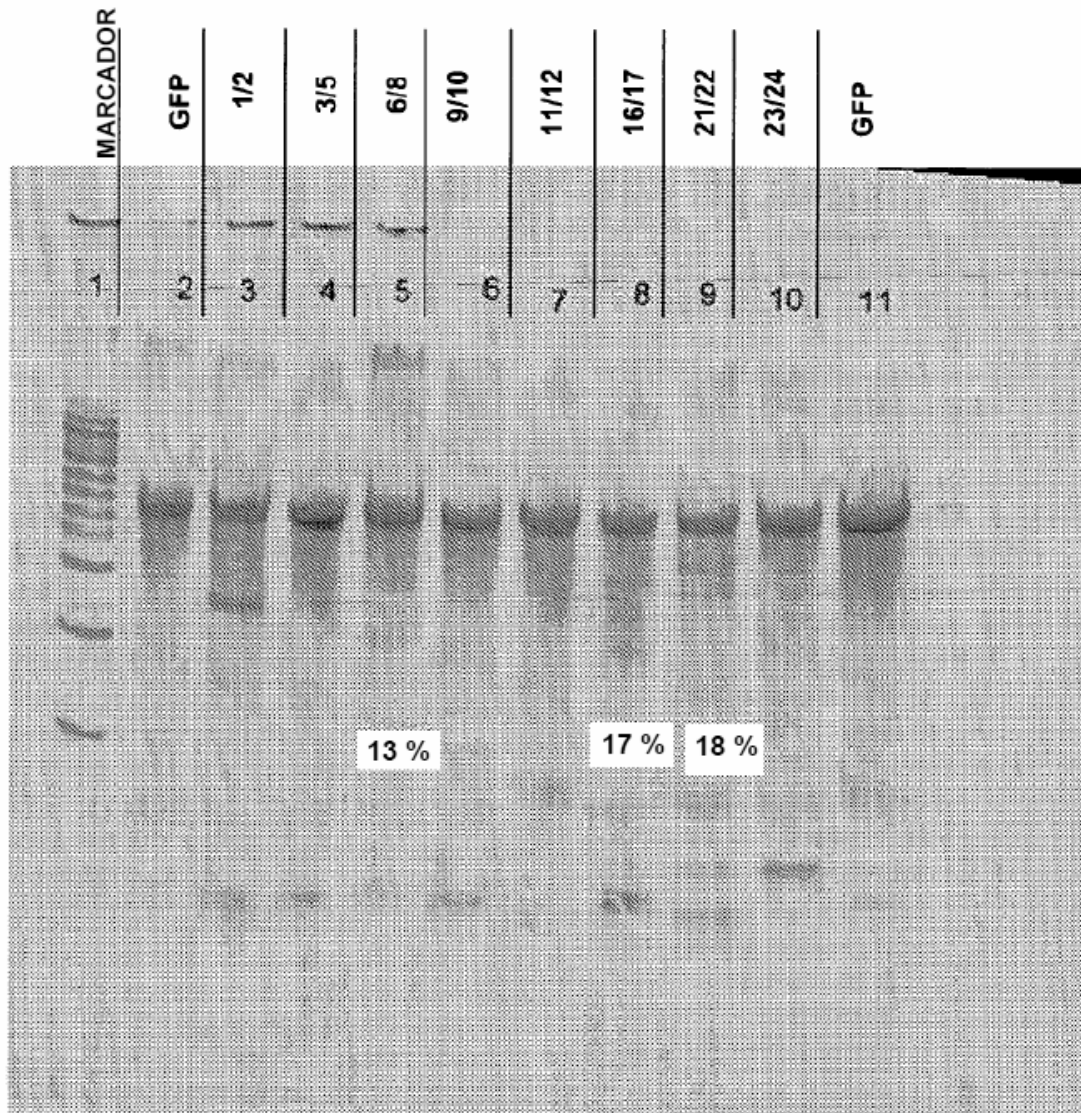


**FIG. 20**

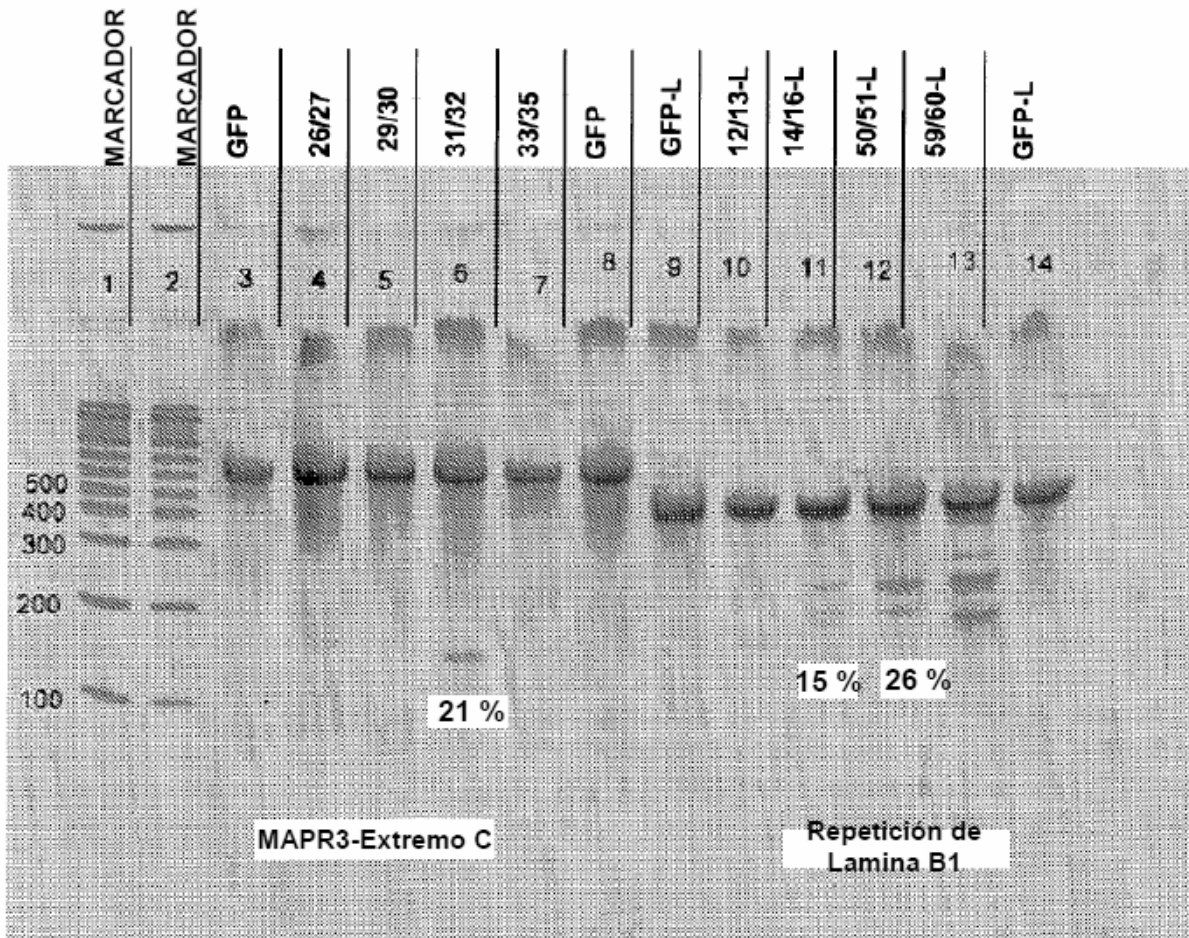


**FIG. 21**

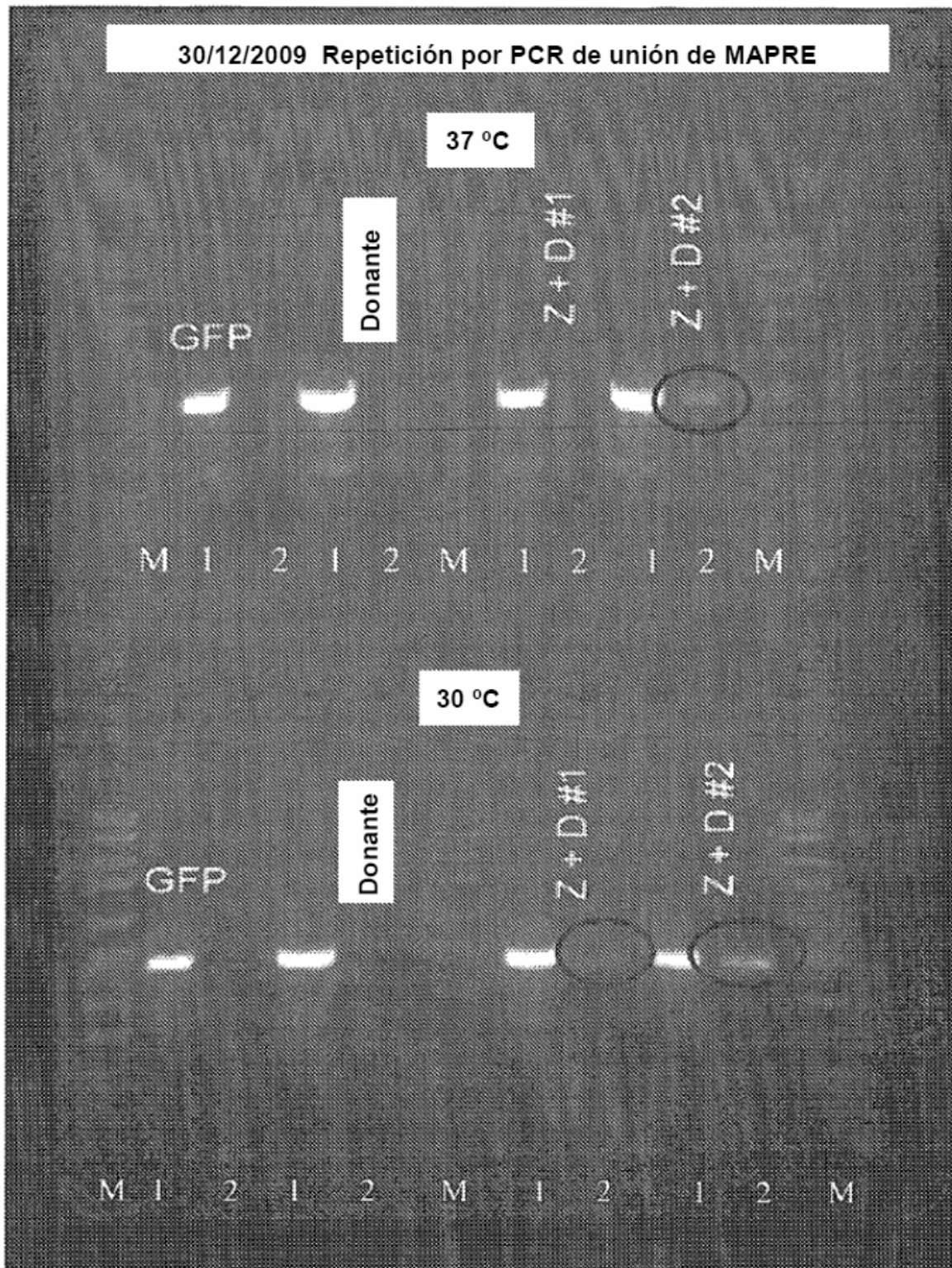




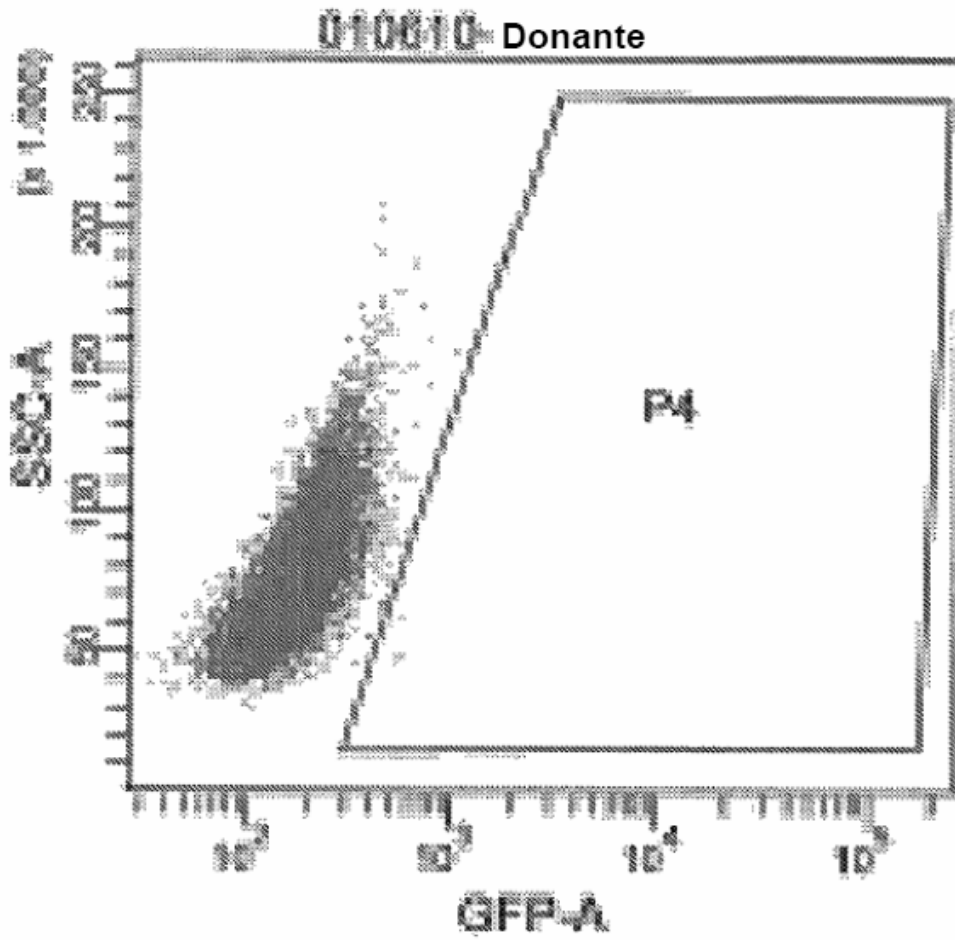
**FIG. 22**



**FIG. 23**



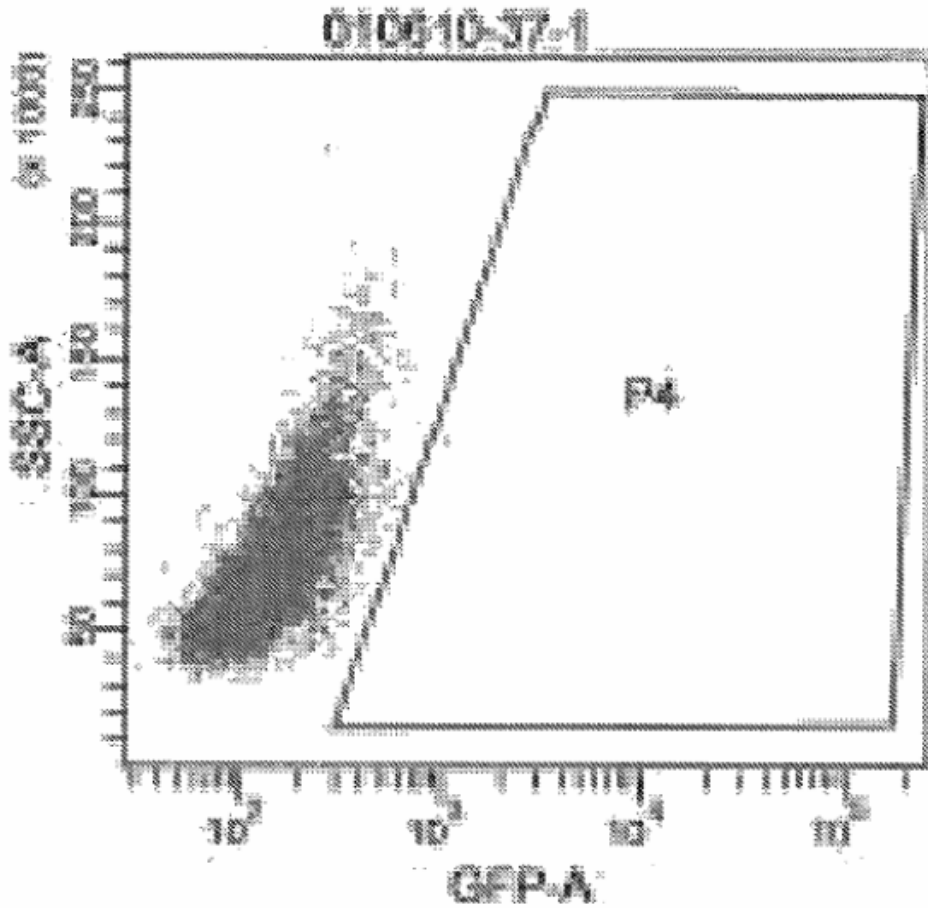
**FIG. 24**



Tubo: Donante

Población	n.º de acontecimientos	% Parental	% Total
<input type="checkbox"/> Todos los acontecimientos	11.696		100,0
<input checked="" type="checkbox"/> P1	11.524	98,5	98,5
<input checked="" type="checkbox"/> P2	10.291	89,3	86,0
<input checked="" type="checkbox"/> P3	10.002	97,2	85,5
<input checked="" type="checkbox"/> P4	3	0,0	0,0
<input checked="" type="checkbox"/> P5	7	0,1	0,1

**FIG. 25A**



Tubo: 37-1

Población	n.º de acontecimientos	% Parental	% Total
<input type="checkbox"/> Todos los acontecimientos	12.050		100,0
<input checked="" type="checkbox"/> P1	11.876	98,6	98,6
<input checked="" type="checkbox"/> P2	10.250	85,4	85,1
<input checked="" type="checkbox"/> P3	8.940	75,9	72,5
<input checked="" type="checkbox"/> P4	3	0,0	0,0
<input checked="" type="checkbox"/> P5	1	0,1	0,1

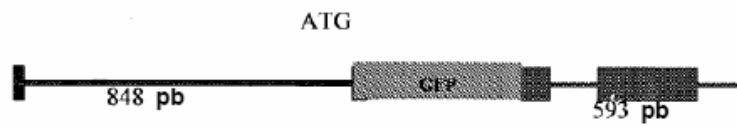
**FIG. 25B**



**FIG. 26A**



**FIG. 26B**

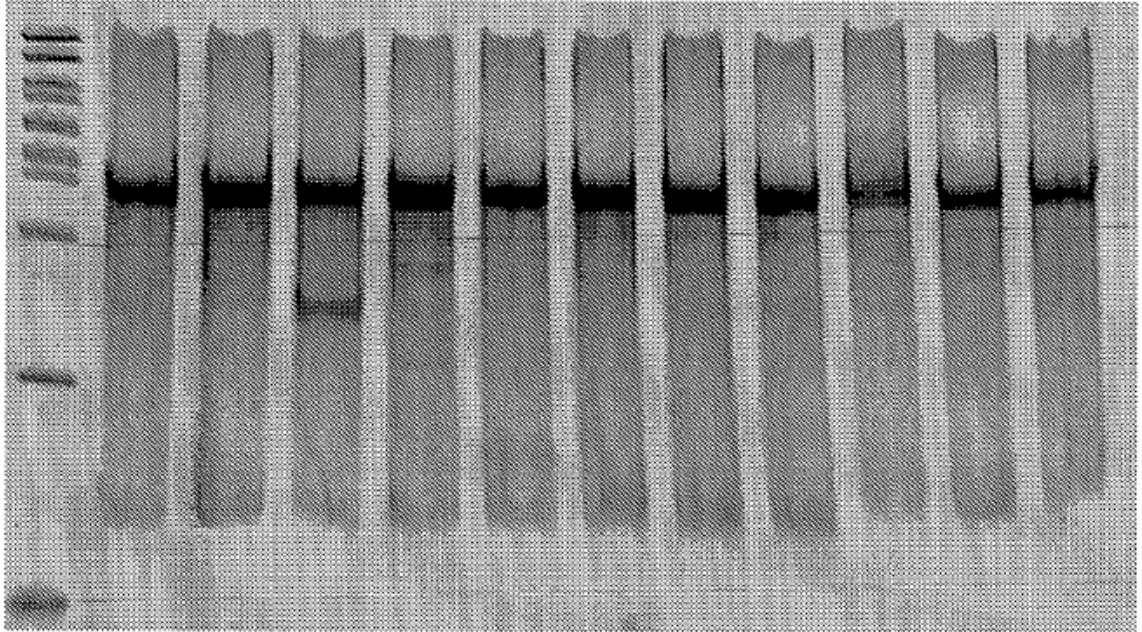


**FIG. 26C**

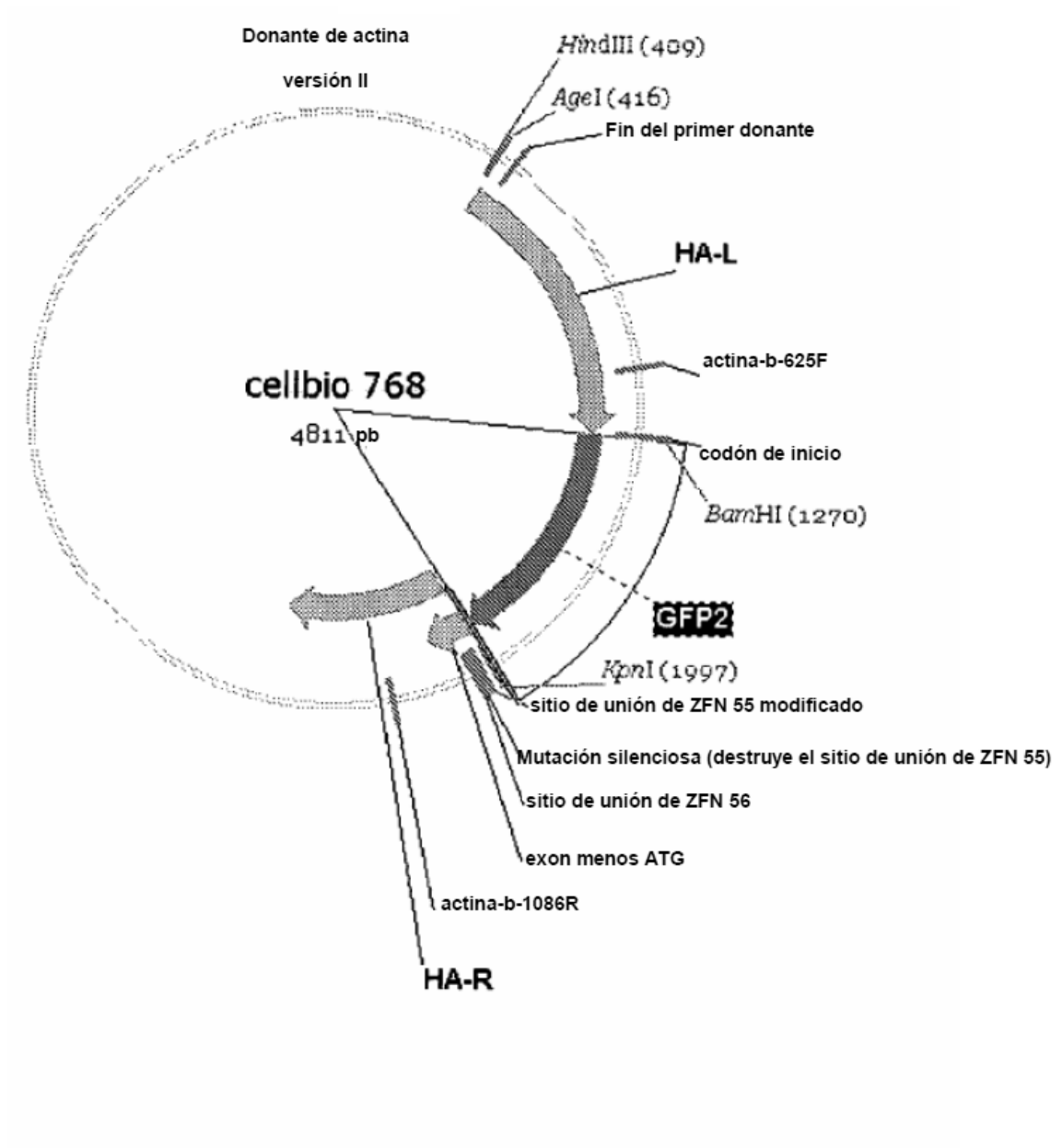


**FIG. 26D**

Diseño n.º GFP 47/48 55/56 78/79 81/82 83/84 83/85 86/88 91/92 91/93 97/99

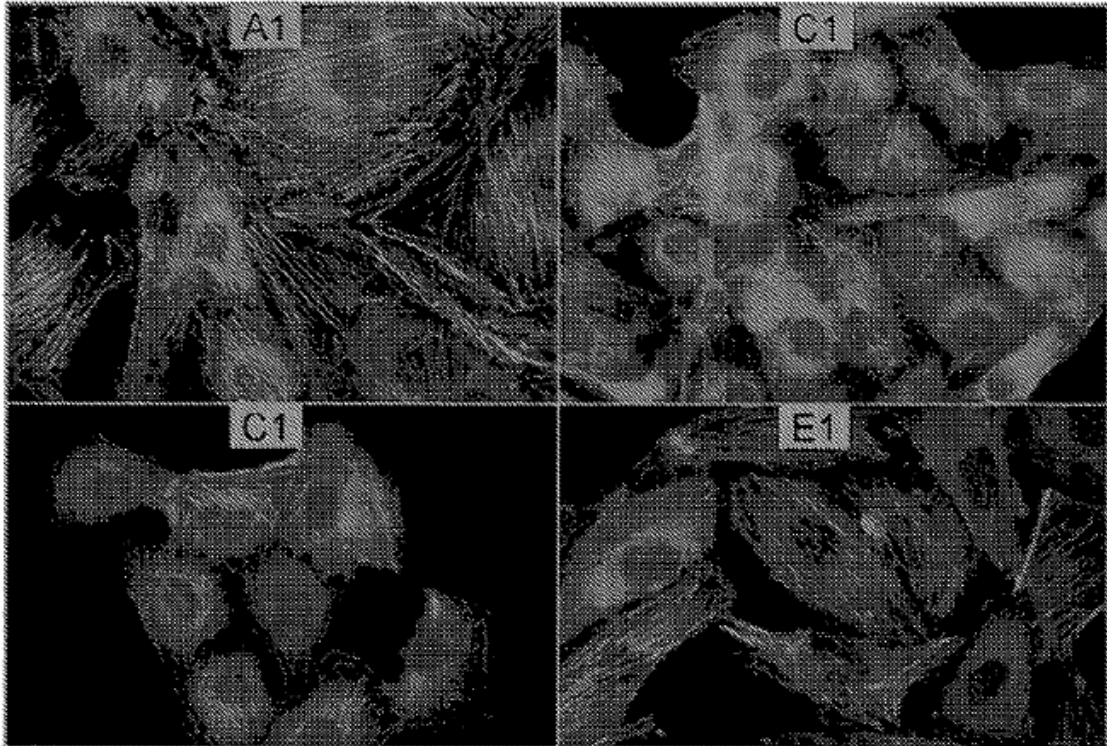


**FIG. 27**



**FIG. 28**



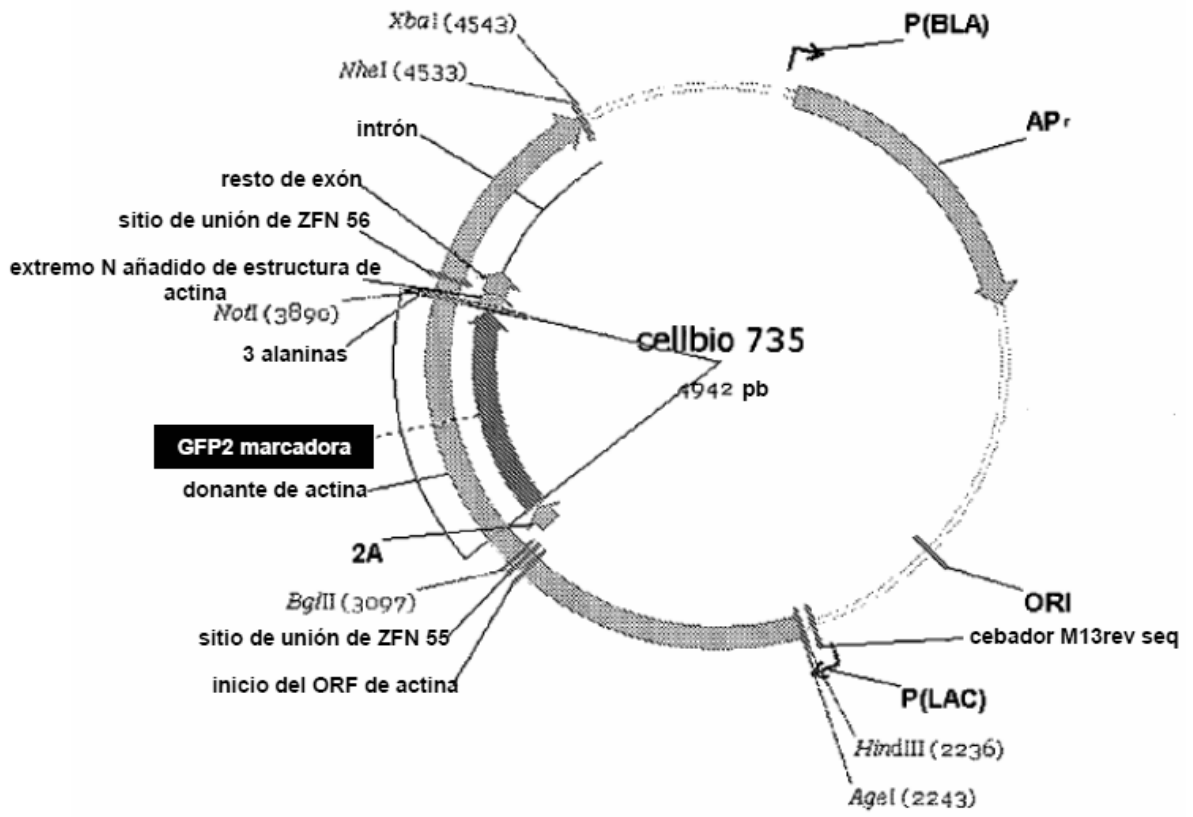


**FIG. 29**

ggacgcctccgaccagtgcttgcctttatggtaataacgcggccggcccgcttctttgtccccaatctg  
ggcgcgccggcccccctggcgccctaaggactcggcgcgcccgaagtggccagggccgggg  
cgacctcggctcacagcgcgcccggctattctcgcagctcaccatggatccggggcgaggagct  
gttcgcccgcacgtgccgtgctgatcgagctggacggcgacgtgcacggccacaagttca  
gcgtgcgcggcgagggcgagggcgacgcccactacggcaagctggagatcaagttcatct  
gcaccaccggcaagctgccctgcccaccctggtgaccaccctctctacggcatc  
cagtgcttcgcccgtaccccgagcacatgaagatgaacgacttctcaagagcgccatgcc  
gagggctacatccaggagcgcaccatccagttccaggacgacggcaagtacaagaccgc  
ggcgaggtgaagttcgagggcgacaccctgggtgaaccgcatcgagctgaagggcaaggac  
ttcaaggaggacggcaacatcctgggccacaagctggagtacagcttcaacagccacaacgt  
gtacatccgccccgacaaggccaacaacggcctggaggctaacttcaagaccgcccacaac  
atcgagggcgccgctgcagctggccgaccactaccagaccaacgtgccctgggcgac  
ggccccgtgctgatccccatcaaccactacctgagcactcagaccaagatcagcaaggaccg  
caacgagggcccgcgaccacatgggtgctcctggagtccttcagcgcctgctgccacaccacg  
gcatggacgagctgtacaggccggctccggtagccatgATGatatcggcgcctcgtcgtcgc  
aacggcagcggcatgtgcaaggccgcttcgcgcccgaagatccccccggcccgtcttcccctcca  
tcgtggggcggcccaggcaccaggtaggggagctggctgggtggggcagccccgggagcggcg  
ggaggaaggccgcttctctgcacaggagcctcccggttccgggggtggggctgcgcccgtctca  
ggcttctgtccttccctcccaggcgctgatggtaggcatgggtcagaa (SEQ ID NO:16)

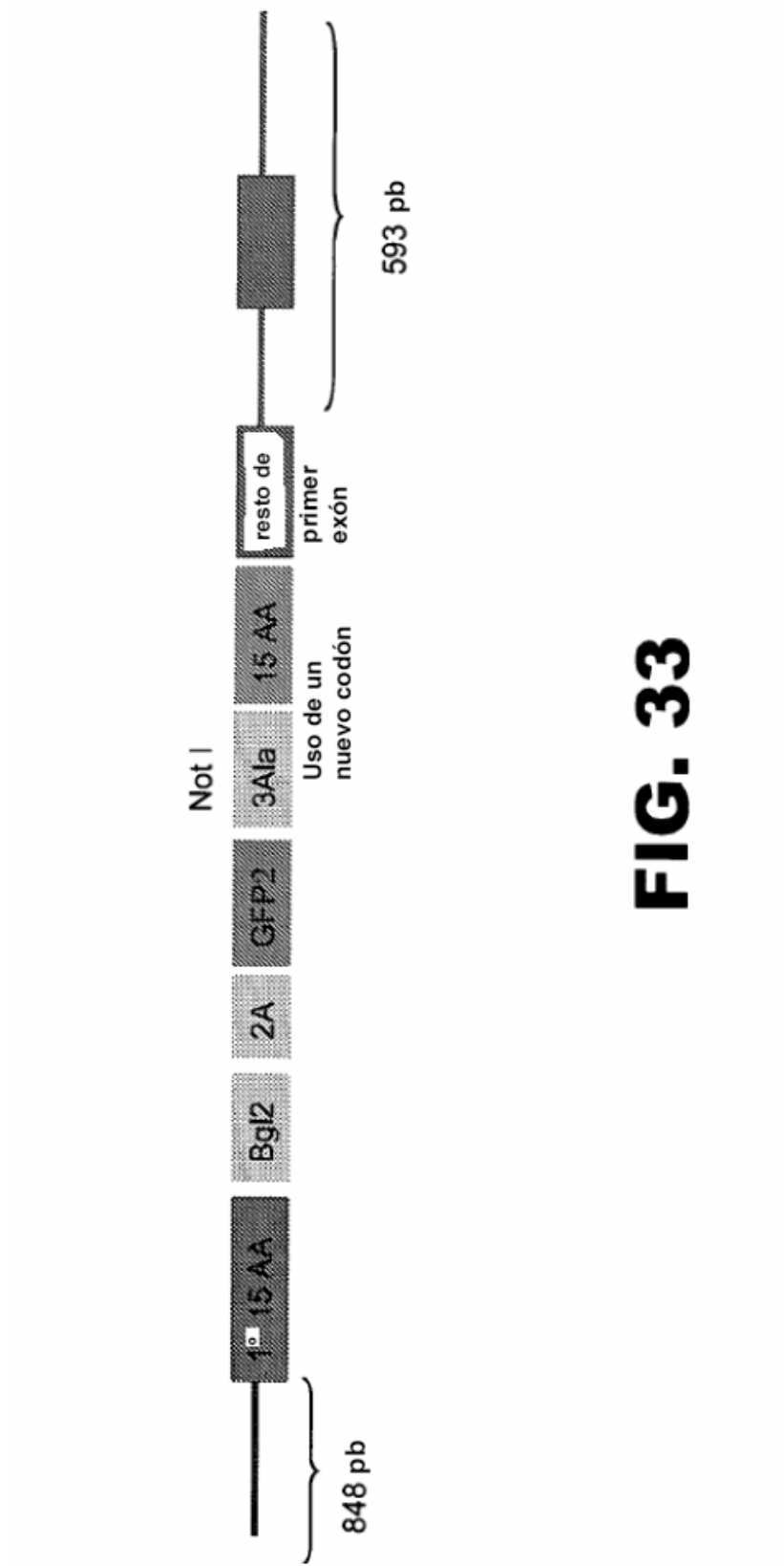
## FIG. 30



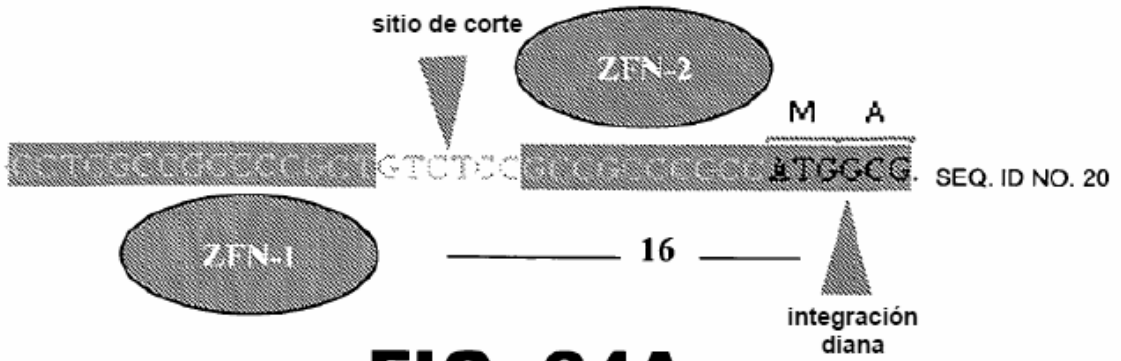


Donante de actina  
versión I

**FIG. 32**



**FIG. 33**



**FIG. 34A**



**FIG. 34B**



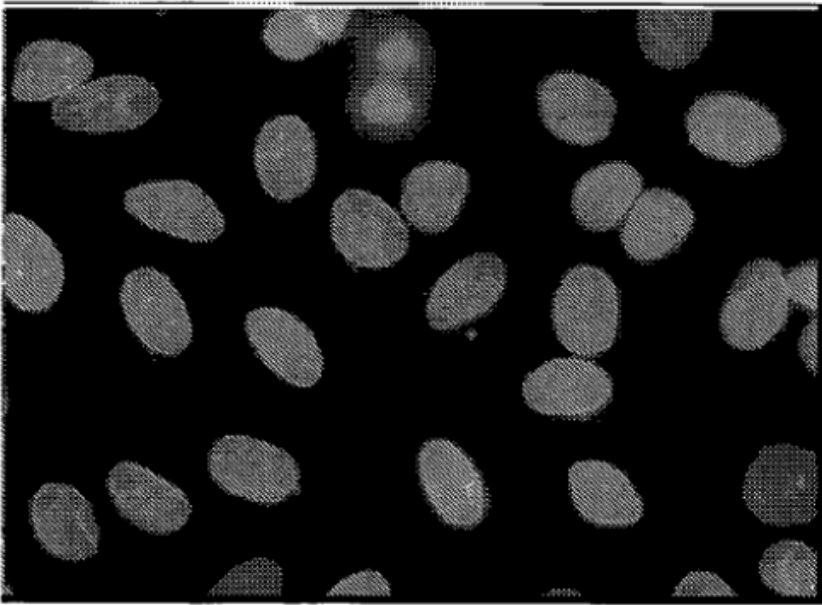
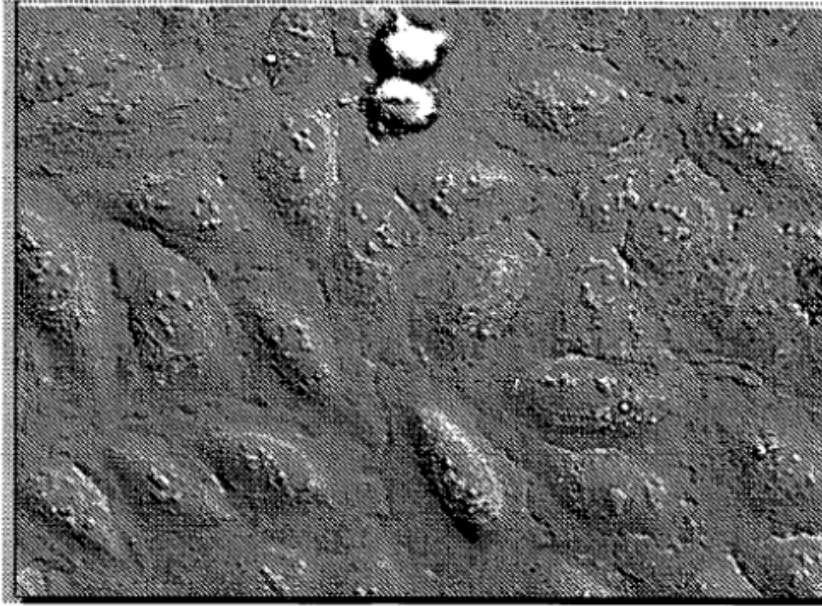
**FIG. 34C**



**FIG. 34D**



**FIG. 34E**

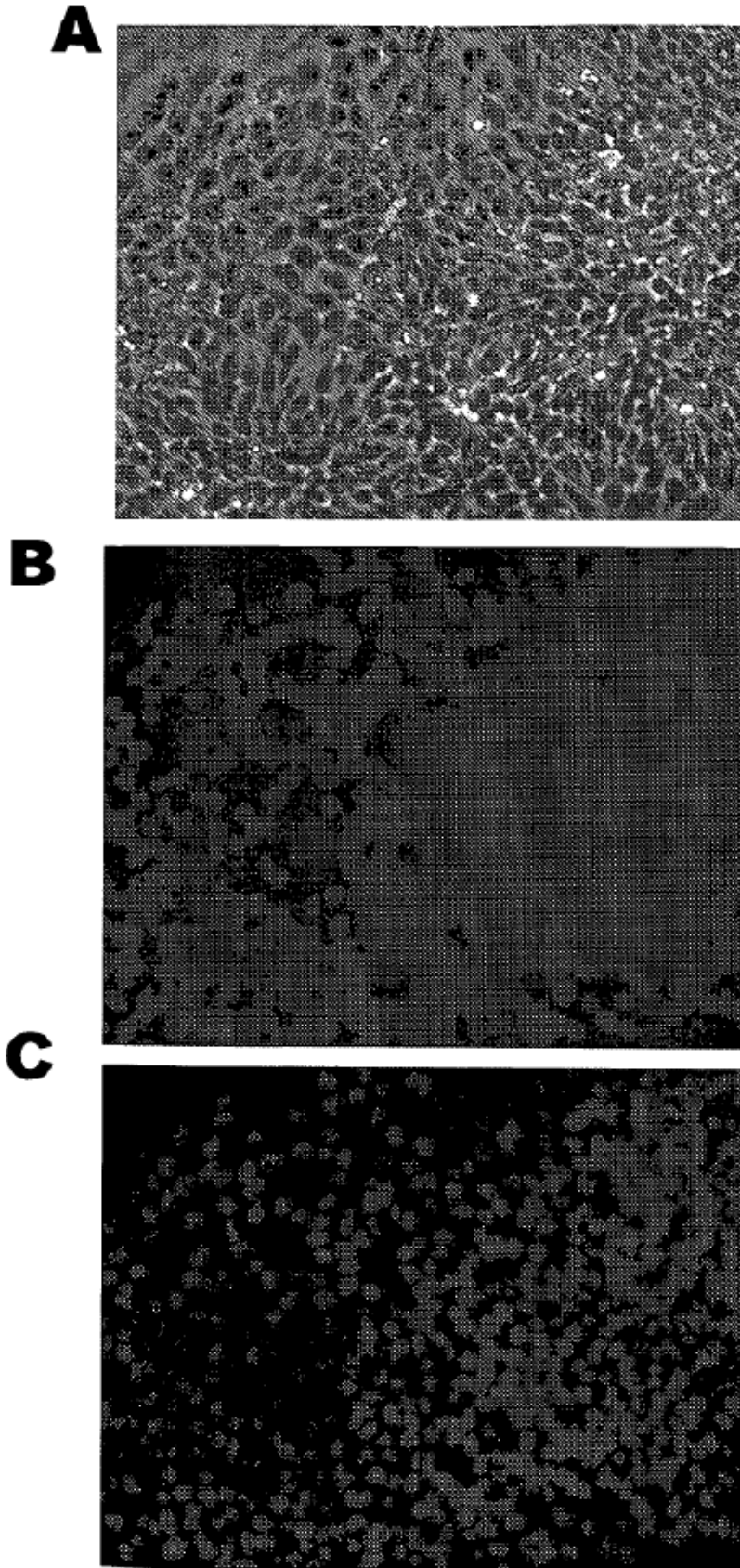


**FIG. 35**

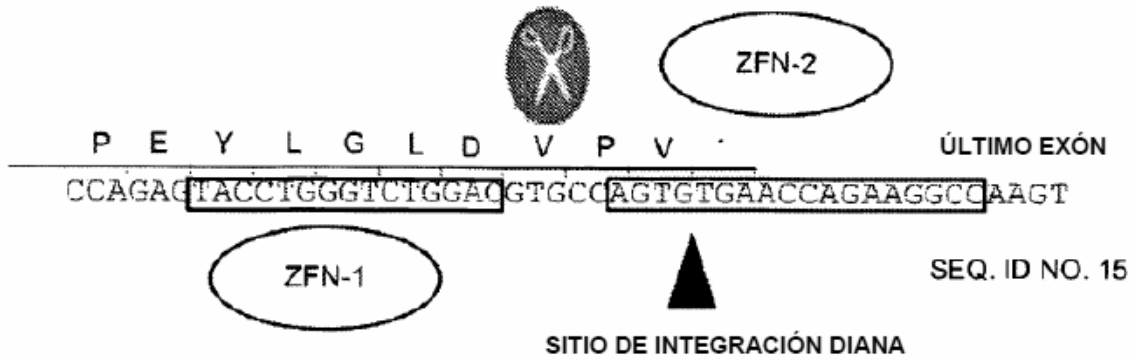
ccctgagcctggccgggaaccgcccagccgggagggccgagctgacgggtgcccaggggccagattt  
 taatttacaggcccggccccgaaccgcccgaagcgcgctgcctgctccccattggcccatggtagtcac  
 gtggaggcgcggggcgctgccggccatgtggggagtgccggcgccgcccgcacctccgccc  
 cccgcccgttgccctccagcccggccccccccggccctccccccgcccggcctccgtgcagcctgaga  
 ggaaacaaagtgtgctgcgagcaggagacggcggcgccggaaccctgctggcctccagtcaccctcg  
 tcttgcaatttcccgcgctgctgtgagtggtgtgtgtgtttcttacaaggggtatttcgcatcgatcgattg  
 attcgtagttccccccgcccgcctttgccccttgctgtaatcgagctcccgccatcccaggctcttccgtt  
cctctaaaccccagcgtctggaagcgtgagcggctgcgggtttgtcccttcggtccccgctcggcccctg  
ccgtccccctctatcacggtcccgtcgcggcctcgcggccccgctgtctccgcccggccatggatc  
cgggggcgaggagctgttcgcccgcacatcgtcccgtgctgacgagctggacggcgcagctgca  
cgccacaaagtccagcctgcgcggcgaaggcgaaggcgacgccgactaccggcaagctggag  
atcaagttcatctgcaccaccggcaagctgcccgtgcccctggcccaccctggtgaccaccctctg  
ctacggcatccagtgcttcgcccgtaccccgagcacatgaagatgaacgactcttcaagagcg  
 ccatgcccgagggtacatccaggagcgcacatccagttccaggacgacggcaagtacaag  
 acccgccgaggggtgaagttcagggcgacaccctggtgaaccgcatcgagctgaagggcaa  
 ggacttcaaggaggacggcaacatcctggggccacaagctggagtagcctcaacagccacaa  
 cgtgtacatccgccccgacaaggccaacaacggcctggaggctaacttcaagaccgcccacaa  
 catcgaggcgccggcgtgcagctggccgaccactaccagaccaacgtgcccctgggcgac  
ggcccgtgctgatccccatcaaccactacctgagcactcaqaccaagatcaqcaaggaccgc  
aacgaggcccgcgaccacatgggtgctcctggagtccttcagcgcctgctgccacaccacggc  
atggaccgagctgtacagggaatctggatcaggtaccgagctgcgacccccgtgcccggcgga  
Gggcagccgctgcccggccccaccagccgctgagccccacgcccctgtcgcggctccaggagaa  
ggaggagctgcggaactcaatgaccggctggcgggtacatcgaacaaggtcgcagcctggagacg  
gagaacaacgcccgtcagctgcaggtgacggagccgaggggtgcgcccggctgagctcaccggc  
 ctcaaggcgtctacgagaccgagctggccgacgcgcgacgcgcgtcgcagcagcggcccgcgag  
 cgcgccaagctgcagatcgagctgggcaagtgaaggcggaaacacgaccagctgctcctcaagtgagt  
 gctagctggcggccgcttagcggcaaggagggggcgggggcgcaaccgcccgcagcagctcaccgg  
 gttctgcccgtggggaggagcagaggccaggatgcacgcgtccttctgaaggaacagggtctcggctc  
 cggaaaggagaaagaatctagagttcalagcggagcaggggtcgcggagggggctcgagctgtagcg  
 ctggggggccgctgatgccatttctagattttggatacccgcctgggacgtggtaagtgcgcccctgg  
 (SEQ ID NO:21)

**FIG. 36**

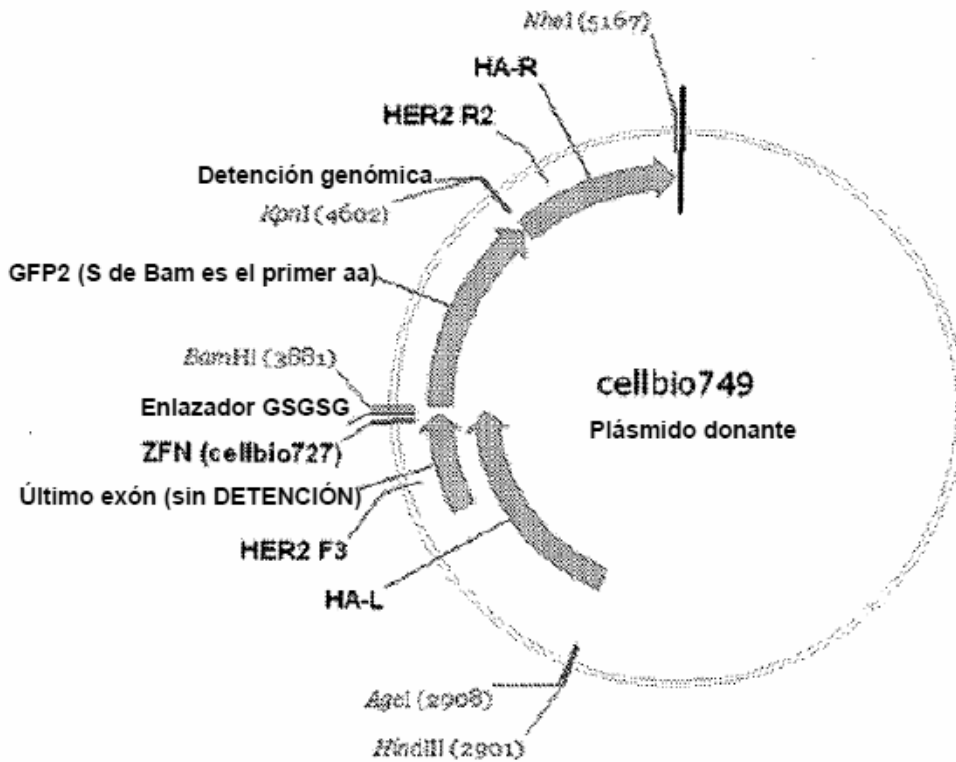




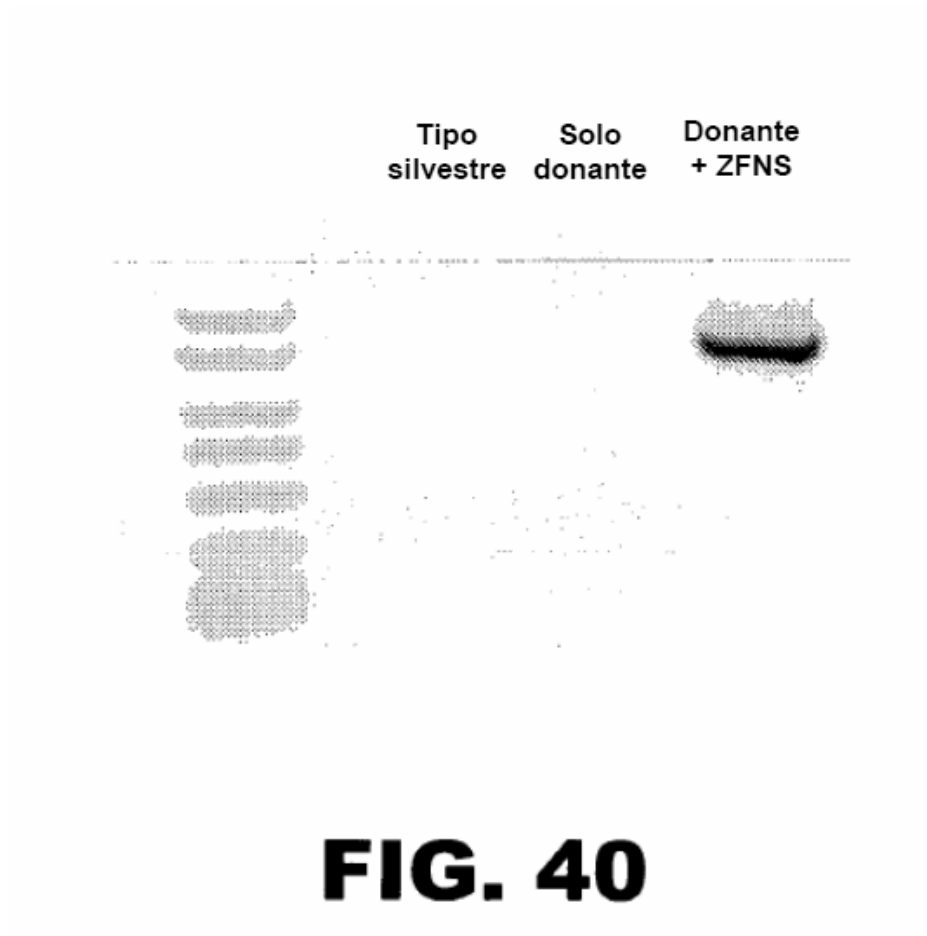
**FIG. 37**

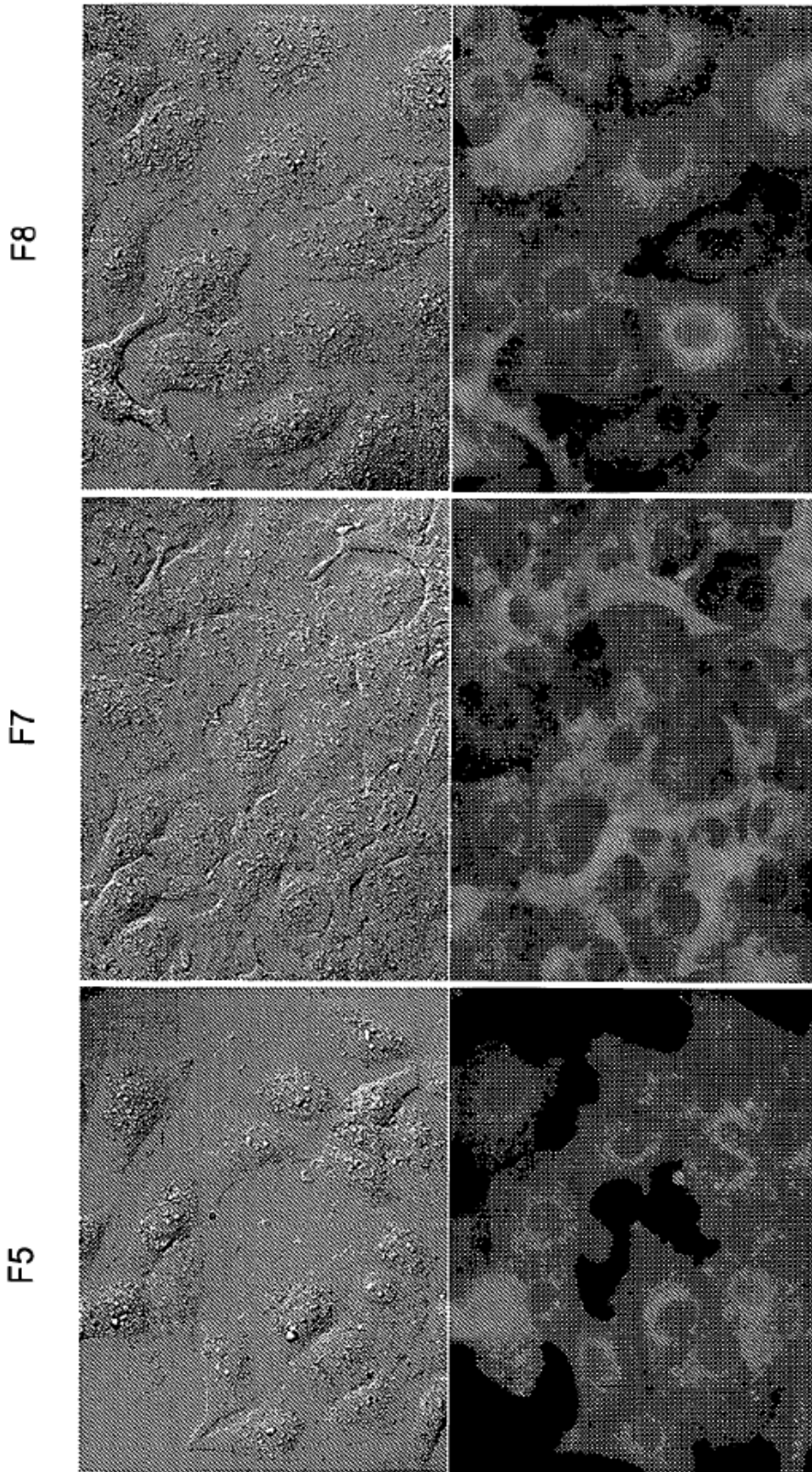


**FIG. 38**



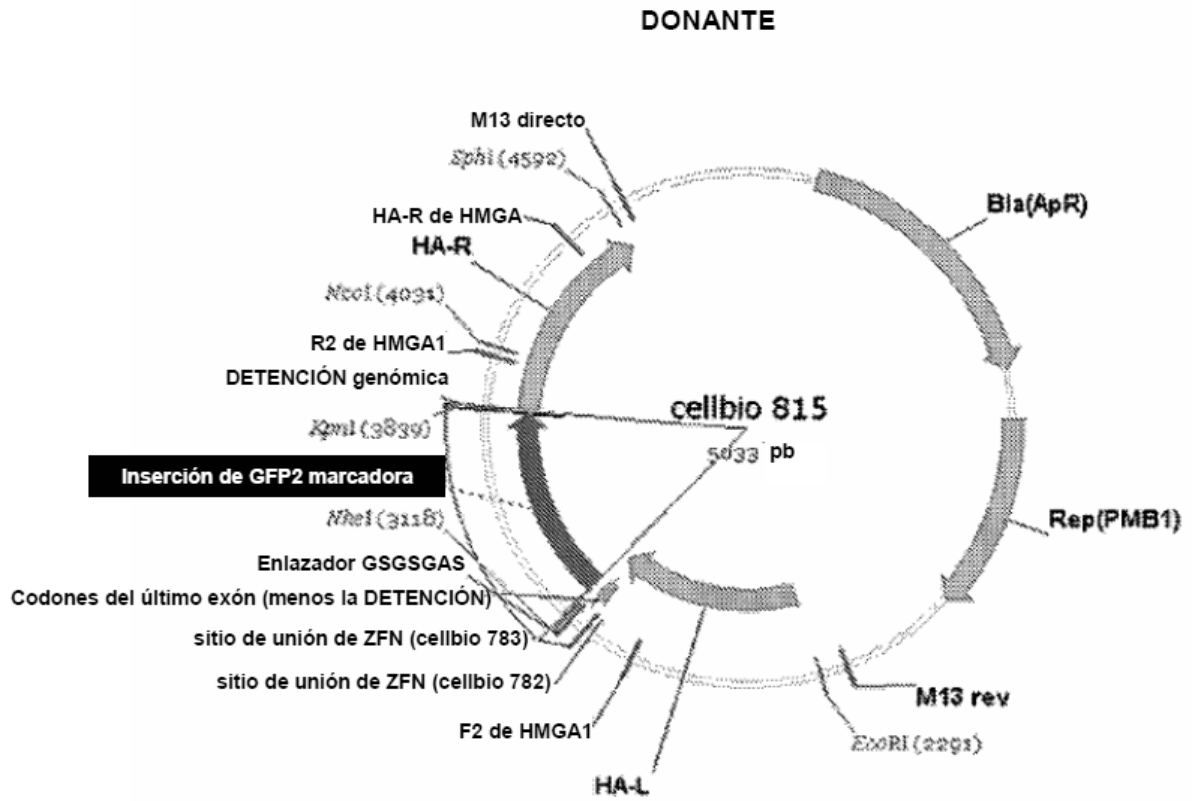
**FIG. 39**



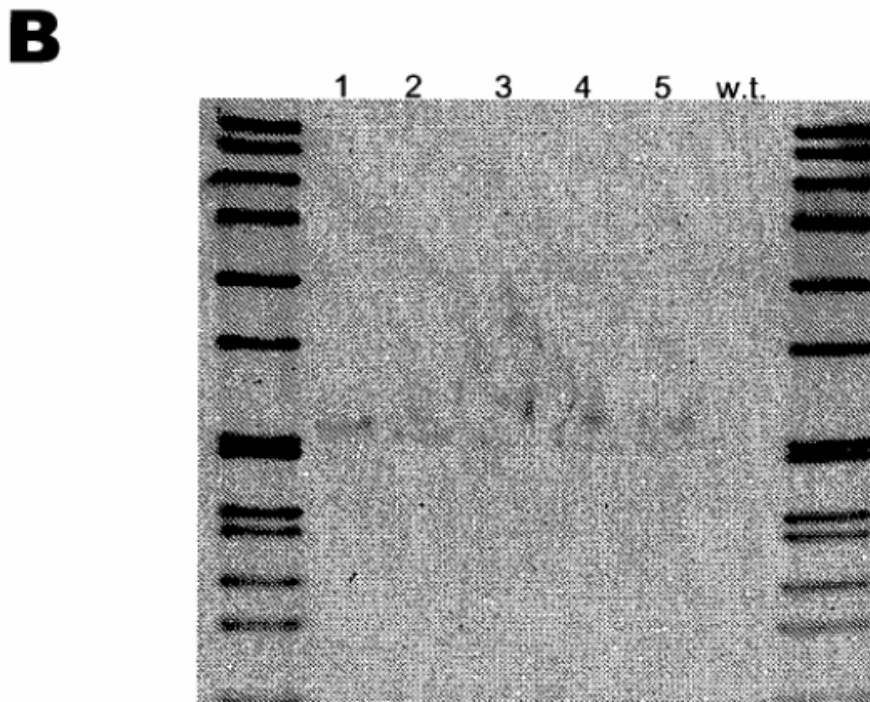
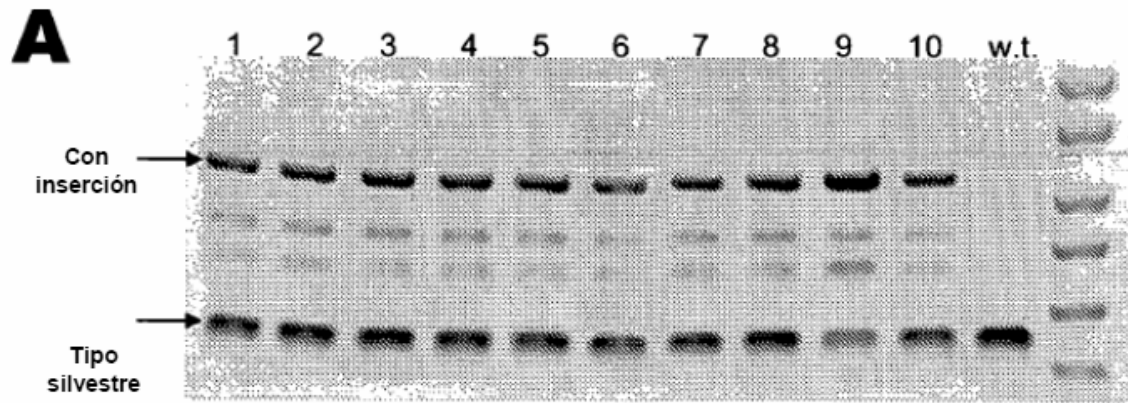


**FIG. 41**





**FIG. 43**

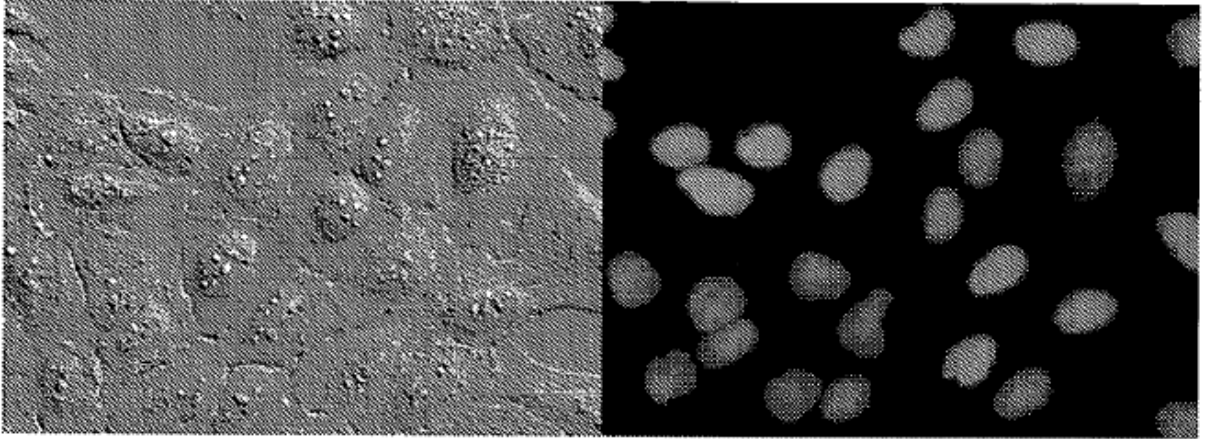


**FIG. 44**

ggctgcccccatcactattgggcaicgggtgagcactgatgagcatttggacttaggagatatttctctaaacctctaga  
 aaaccaccacaaciccaggaaggaaaccaaggggagacccaaaaactggtaggtgaagaagcagactgctgct  
 tgctctgggctctttgagttgaggggtgtgagtacacagtaacctgatgctatgcaccccctatggaaaggctctcctgac  
 ctgctgggacatcagatttacagaagtcagagaggggaaggtaacctggcctgggctgtgcccatagagaagtgagggg  
 tcccaggtataaatcagaccacalcccccgccctgcccctgcccagttgtgtgtgggggcccctcccctctctgctctaga  
 aactcagaactctaggggagatcttgaagtcactagcctgtgtccccctcaattagagatgaggaaaggaagccata  
 gggggaaggttgcctcctatgagcctctgcagaagagaaaacagcgaaggagctgggcccctgggaggggtcgggtgt  
 ggagttctgatgtgaccaccacactgcaactggagggcaccatccaaltctgggtccccaaacagctggtggaaaggct  
 cgggtggctgagicaagaagctgctctagggggccactgcaagttagggtcaccacagcctccagctctggcccctc  
 ctacccccagcctgccccctcaaatccctgaagctgtcattccttgagctgagccactgctggggggggggttagggg  
 tgcctgctggccaggccccaaagagtgagtaacaggaaacaagtgtttggagttgtgcccggcagggggcgtgtagccc  
 cgtgtgggtgcccacattcccggccagtgagtgagccccggcggcacacactccccttctccccaccccggcctagg  
 gtcagccccctggccaccccggagggccagggcaccacagcagcaictgccccgtgggccaaggacctggttccc  
 ctgaccccaccagcgggctcttgacccctccagccaccccctccattctccccagccacctctccccacctcctctc  
 tcccclagggagtcagtcacatccctgaagctcattgctgcccctgagctctgcccctctgcccctcccctgggcccctggggccaa  
 gggggctgtgctcctggcctggtgagagcagcatgtgtgtgggggttttctccttttaaattctttatgaatgaagccgg  
 gccgtggaggtgtgctgagtcaccacacactcagccctgactcatcccctctcaggagagccagggactgcaggagcg  
ggtgggccaagcctctgggggtggaagagggggaacggggccagagctcacaccaacaactcccacctcacagga  
gaaggagcaagagggagggcatctcagagagctcctcggagggagagcagggatctggatcaggtgctagcggggg  
cgagggagctgttcggcgcacatcgtgccctgctgacagctggacggcgcacatgcacggccacaagttcag  
cgtgcggggcagggcgagggcgacggcgcactacggcaagctggagatcaagttcatctgcaccaccggc  
aagctgcccgtgccctggcccaccctggtgaccaccctctgctacggcatccagtgcttgcggcctaccccc  
agcacatgaagatgaacgacttctcaagagcgccatgcccggggctacatccaggagcgcaccatccagt  
tccaggacgacggcaagtaacaagaccgcggcgaggtgaagttcagggcgacaccctggtgaaccgcat  
cgagctgaagggcaaggacttcaagggagcggcaaacatcctggccacaagctggagtacagcttcaaca  
gccacaacgtgtacatccgccccgacaaggccaacaacggcctggagggctaactcaagaccggccacaac  
atcgagggcgggcggtgagctggccgaccactaccagaccaacgtgcccctggggcagggcccctgtgc  
tgatccccatcaaccactacctgagcactcagaccaagatcagcaaggaccgcaacgaggcccgcgaccac  
atggtgtcctggagtcctcagcgcctgctgcccacaccacggcatggacgagctgtacaggtaaaglacctg  
accATGcgtgcccctgctcctcactggaggagcagctcctctgaggactggacagcttctcctcccaccgccc  
 ccaccccctcccaggcccacacalcaaccaccgcccctgcccggccacccccatctccaactgtgcccctaccaccacact  
 acacagcacaccagcccgtgaggggtcccattggggtgagtggggagcagtttcccctggcctcagttcccagctcccc  
 ccgcccacccacgcatacacacatgcccicctggacaaggctaacalcccacttagccgcaccctgcaactgtgctgctc  
 cccactcccctgggtggggacattgctctctgggctttgggtggggcgccctctctgctcctcactgttcccctggtc  
 ccatagtggggctgggaggggtcccctggccttaaaggggcccaagccccatctcatctggcagccctactcact  
 gcctggcagcagcaggtgtggccaatggaggggggtgctggccccaggaltccccagccaaactgtctttgtcacc  
 acgtggggctcactttatccttcccacactccctagctcccglactaggtggacagccccctcgggtacaggaaggca  
 ggaggggtgagctcccactcctctcactgtggccacagccccctgcccctcggcctgggatctgagtaacatattgtgtg  
 atggagatgcagtcactattgtccaggtgaggccaagagcccctgtggccgcccactgaggtgggctggggctgctccc  
 ctaaccctactttgctccgccaactcagccattccccctcctcagatggggcaccaataacaaggagctcaccctgccc  
 ctcccacccccctctgctcctcccggcccccaaggtctgggtccattttctctgttcaaaactacctctgacagttgt  
 gttgtttttgtcaatgttccattctcgacatccgtcattgtgctgctaccagcggcaaatgttcatcctcattgcccctgttctg  
 cccacgatcccctcccccaagatactctttgtgggaa (SEQ ID NO: 28)

**FIG. 45**





**FIG. 46**

## REFERENCIAS CITADAS EN LA DESCRIPCIÓN

Esta lista de referencias citadas por el solicitante es únicamente para la comodidad del lector. No forma parte del documento de la patente europea. A pesar del cuidado tenido en la recopilación de las referencias, no se pueden excluir errores u omisiones y la EPO niega toda responsabilidad en este sentido.

## Documentos de patentes citados en la descripción

- 10
- US 61323702 B [0001] [0118]
  - US 61323719 B [0001] [0118]
  - US 61323698 B [0001] [0118]
  - US 61367017 B [0001] [0118]
  - 15
  - US 61390668 B [0001] [0118]
  - US 61408856 B [0001] [0118]
  - US 61431957 B [0001] [0118]
  - US 6453242 B [0036] [0037]
  - 20
  - US 6534261 B [0036] [0037]
  - US 6607882 B [0037]
  - US 5789538 A [0038]
  - US 5925523 A [0038]
  - 25
  - US 6007988 A [0038]
  - US 6013453 A [0038]
  - US 6410248 B [0038]
  - US 6140466 A [0038]
  - 30
  - US 6200759 B [0038]
  - US 6242568 B [0038]
  - WO 9837186 A [0038]
  - WO 9853057 A [0038]
  - WO 0027878 A [0038]
  - WO 0188197 A [0038]
  - GB 2338237 A [0038]
  - WO 02077227 A [0038]
  - US 20050064474 A [0039] [0047] [0048]
  - US 20060188987 A [0039] [0047]
  - US 6479626 B [0039]
  - US 6903185 B [0039]
  - US 7153949 B [0039]
  - US 5356802 A [0045]
  - US 5436150 A [0045]
  - US 5487994 A [0045]
  - WO 07014275 A [0045]
  - US 20080131962 A [0047]
  - US 5420032 A [0051]
  - US 6833252 B [0051]

## Literatura diferente de patentes citada en la descripción

- 35
- BEERLI et al. *Nat. Biotechnol.*, 2002, vol. 20, 135-141 [0036]
  - PABO et al. *Ann. Rev. Biochem.*, 2001, vol. 70, 313-340 [0036]
  - 40
  - ISALAN et al. *Nat. Biotechnol.*, 2001, vol. 19, 656-660 [0036]
  - SEGAL et al. *Curr. Opin. Biotechnol.*, 2001, vol. 12, 632-637 [0036]
  - 45
  - CHOO et al. *Curr. Opin. Struct. Biol.*, 2000, vol. 10, 411-416 [0036]
  - ZHANG et al. *J. Biol. Chem.*, 2000, vol. 275 (43), 33850-33860 [0036]
  - DOYON et al. *Nat. Biotechnol.*, 2008, vol. 26, 702-708 [0036]
  - 50
  - SANTIAGO et al. *Proc. Natl. Acad. Sci. USA*, 2008, vol. 105, 5809-5814 [0036]
  - SERA et al. *Biochemistry*, 2002, vol. 41, 7074-7081 [0036]
  - 55
  - MANDELL et al. *Nuc. Acid Res.*, 2006, vol. 34, W516-W523 [0036]
  - SANDER et al. *Nuc. Acid Res.*, 2007, vol. 35, W599-W605 [0036]
  - 60
  - MAKKERH et al. *Current Biology*, 1996, vol. 6, 1025-1027 [0040]
  - BELFORT et al. *Nucleic Acids Res.*, 1997, vol. 25, 3379-3388 [0042] [0051]
  - 65
  - Nucleases. Cold Spring Harbor Laboratory Press, 1993 [0042]
  - LI et al. *Proc. Natl. Acad. Sci. USA*, 1992, vol. 89, 4275-4279 [0045]
  - LI et al. *Proc. Natl. Acad. Sci. USA*, 1993, vol. 90, 2764-2768 [0045]
  - KIM et al. *Proc. Natl. Acad. Sci. USA*, 1994, vol. 91, 883-887 [0045]
  - KIM et al. *J. Biol. Chem.*, 1982, vol. 269 (31), 978-31 [0045]
  - ROBERTS et al. *Nucleic Acids Res.*, 2003, vol. 31, 418-420 [0045]
  - BITINAITE et al. *Proc. Natl. Acad. Sci. USA*, 1998, vol. 95 (10), 570-10, 575 [0046]
  - DUJON et al. *Gene*, 1989, vol. 82, 115-118 [0051]
  - PERLER et al. *Nucleic Acids Res.*, 1994, vol. 22, 1125-1127 [0051]
  - JASIN. *Trends Genet.*, 1996, vol. 12, 224-228 [0051]
  - GIMBLE et al. *J. Mol. Biol.*, 1996, vol. 263, 163-180 [0051]
  - ARGAST et al. *J. Mol. Biol.*, 1998, vol. 280, 345-353 [0051]
  - CHEVALIER et al. *Molec. Cell*, 2002, vol. 10, 895-905 [0052]
  - EPINAT et al. *Nucleic Acids Res.*, 2003, vol. 31, 2952-2962 [0052]

- **ASHWORTH et al.** *Nature*, 2006, vol. 441, 656-659 [0052]
- 5 • **PAQUES et al.** *Current Gene Therapy*, 2007, vol. 7, 49-66 [0052]
- **SANTIAGO et al.** *PNAS*, 2008, vol. 105, 5809-5814 [0063]
- **MOEHLE et al.** *PNAS*, 2007, vol. 104, 3055-3060 [0063]
- 10 • **URNOV et al.** *Nature*, 2005, vol. 435, 646-651 [0063]
- **LOMBARDO et al.** *Nat. Biotechnology*, 2007, vol. 25, 1298-1306 [0063]
- 15 • **SINGLETON et al.** *Dictionary of Microbiology and Molecular Biology*. 1994 [0069]
- *The Cambridge Dictionary of Science and Technology*. 1988 [0069]
- 20 • *The Glossary of Genetics*. Springer Verlag, 1991 [0069]
- **HALE ; MARHAM.** *The Harper Collins Dictionary of Biology*, 1991 [0069]
- **SMITH ; WATERMAN.** *Advances in Applied Mathematics*, 1991, vol. 2, 482-489 [0078]
- **DAYHOFF.** *Atlas of Protein Sequences and Structure*. National Biomedical Research Foundation, vol. 3, 353-358 [0078]
- **GRIBSKOV.** *Nucl. Acids Res.*, 1986, vol. 14 (6), 6745-6763 [0078]
- *Nucleic Acid Hybridization: A Practical Approach*. IRL Press, 1985 [0079] [0081]
- **SAMBROOK et al.** *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, 1989 [0080] [0082]
- *Science*, 2009, vol. 325, 433 [0085]