

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 979 326**

51 Int. Cl.:

C12N 9/02 (2006.01)
A61K 38/17 (2006.01)
C07K 14/435 (2006.01)
C07K 19/00 (2006.01)
C12N 15/62 (2006.01)
A61K 48/00 (2006.01)
A61P 35/00 (2006.01)
A61P 25/18 (2006.01)
A61P 9/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **29.05.2013** **E 20183740 (8)**

97 Fecha y número de publicación de la concesión europea: **31.01.2024** **EP 3744835**

54 Título: **Proteínas de fusión modificadoras del ADN y método de utilización de las mismas**

30 Prioridad:

29.05.2012 US 201261652711 P
15.03.2013 US 201313838520

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
25.09.2024

73 Titular/es:

THE GENERAL HOSPITAL CORPORATION
(100.0%)
55 Fruit Street
Boston, MA 02114, US

72 Inventor/es:

JOUNG, KEITH, J;
MAEDER, MORGAN y
ANGSTMAN, JAMES

74 Agente/Representante:

ISERN JARA, Jorge

Observaciones:

Véase nota informativa (Remarks, Remarques o
Bemerkungen) en el folleto original publicado por la
Oficina Europea de Patentes

ES 2 979 326 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

Proteínas de fusión modificadoras del ADN y método de utilización de las mismas

Campo técnico

La presente invención se refiere a proteínas de fusión que comprenden un dominio de unión a ADN, que se une específicamente a una secuencia diana preseleccionada que se encuentra a menos de 250 pb de un dirresiduo CpG en una célula de mamífero, en la que el dominio de unión a ADN es o comprende un dedo de zinc diseñado y un dominio catalítico que comprende una secuencia que cataliza la desmetilación de citosinas en el ADN, en el que la secuencia comprende los aminoácidos 1580 a 2052 de Tet1 (SEC ID n.º 2), en el que la proteína de fusión se une a la secuencia diana y reduce la metilación del dirresiduo CpG, a métodos *ex vivo* de utilización y dichas proteínas de fusión para la utilización en el tratamiento del cáncer.

Antecedentes

TET1 es un enzima que cataliza la conversión de la 5-metilcitosina (5mC) en 5-hidroximetilcitosina (5hmC) (Tahiliani, M. et al. Science 324, 930-935 (2009)). Aunque el papel de 5hmC no está totalmente claro, se ha propuesto que podría ser un intermediario en el proceso de desmetilación de la 5-metilcitosina en citosina. Lo anterior está respaldado por observaciones de que la sobreexpresión de TET1 en células en cultivo conduce a una reducción global de los niveles de 5mC (Tahiliani, M. et al. Science 324, 930-935 (2009)). Se han propuesto varios mecanismos de desmetilación. Existen pruebas de que 5hmC puede desaminarse y que la base no coincidente es reconocida por las ADN glucosilasas y seguidamente reparada a citosina mediante la ruta de reparación de excisión de bases (Guo et al., Cell 145, 423-434 (2011)). Alternativamente, también hay pruebas de que la oxidación iterativa de 5hmC por TET1 rinde 5-formilcitosina (fC) y 5-carboxilcitosina (caC), que seguidamente pueden ser reconocidas por la timina ADN glucosilasa y revertidas a citosina mediante reparación por excisión de bases (He et al., Science 333, 1303-1307 (2011)). En cualquier caso, todas las observaciones subrayan que TET1 es un catalizador primario para la desmetilación del ADN.

El documento n.º WO2009/134409 da a conocer una proteína de fusión que comprende un dominio de unión al ADN y un dominio de desmetilación de la citosina en el ADN.

El documento n.º WO2007/128982 da a conocer una proteína de fusión que comprende un dominio de unión al ADN y un dominio que muestra una actividad de citidina desaminasa.

Gregory et al., Epigenetics, vol. 7, 20125, páginas 344 a 349 dan a conocer una proteína de fusión que comprende un dominio de unión al ADN y una timina ADN glucosilasa (TDG, por sus siglas en inglés) fusionados mediante un conector corto rico en glicinas.

El documento n.º WO2010/037001 da a conocer que TET1, TET2, TET3 y CXXC4 presentan actividad de hidrolasa que convierte el nucleótido de citosina 5-metilcitosina en 5-hidroximetilcitosina.

Foley et al., Nature protocols, vol. 4, n.º 12, páginas 1855 a 1868 dan a conocer vectores de expresión de dedos de zinc.

Neville, et al., Nature protocols, vol. 7, 2012, páginas 171 a 192 dan a conocer el diseño y utilización de proteínas de fusión de matriz TALE, incluyendo factores de transcripción TALE y TALENS.

Frauer, et al., PloS ONE, volumen 6, número 2, páginas e16627 dan a conocer la actividad de unión a ADN de los dominios CXXC de las proteínas Dnmt1 y Tel1 de ratón.

Rodenhiser et al., CMAJ, 174(3), páginas 341 a 348 da a conocer mecanismos epigenéticos y su contribución a la salud humana.

Branco et al., Nature Reviews | Genetics | volumen 13, páginas 7 a 13 dan a conocer el papel de TET1 en la catálisis de la oxidación de la 5-metilcitosina.

Sumario

La invención se define en las reivindicaciones. Por lo menos en parte la presente invención se basa en la identificación de proteínas de fusión que comprenden un dominio diseñado de unión a ADN personalizado, fusionado a una proteína y capaz de catalizar la hidroxilación de las citosinas metiladas en ADN *in vitro*, en células o en organismos completos. Dicha proteína híbrida conduciría a la formación de citosinas hidroximetiladas a partir de citosinas metiladas en el ADN y la posterior desmetilación de las citosinas en células u organismos vivos.

De esta manera se proporcionan en la presente memoria proteínas de fusión que comprenden un dominio de unión a

ADN diseñado que se une específicamente a una secuencia diana preseleccionado que se encuentra a menos de 250 pb de un dirresiduo CpG en la célula de mamífero, en la que el dominio de unión al ADN es o comprende un dedo de zinc diseñado, y un dominio catalítico que comprende una secuencia que cataliza la desmetilación de citosinas en el ADN, es decir, que cataliza la hidroxilación o desmetilación de citosinas metiladas, en el que la secuencia comprende los aminoácidos 1580 a 2052 de Tet1 (SEC ID n.º 2), en el que la proteína de fusión se une a la secuencia diana y reduce la metilación del dirresiduo CpG. En algunas realizaciones, las proteínas de fusión comprenden, además, un conector entre el dominio de unión a ADN y el dominio catalítico.

En algunas realizaciones, el conector se selecciona del grupo que consiste en Gly3Ser (SEC ID n.º 5), Gly3SerGly4SerArgSer (SEC ID n.º 6), Gly3Ser(Gly4Ser)2ArgSer (SEC ID n.º 7), Gly3Ser(Gly4Ser)3ArgSer (SEC ID n.º 8), Gly3Ser(Gly4Ser)4ArgSer (SEC ID n.º 9) y Gly3Ser(Gly4Ser)5ArgSer (SEC ID n.º 10).

El dominio de unión a ADN puede ser o comprender una matriz diseñada de repetición de efectores de tipo activador de transcripción (TAL, por sus siglas en inglés) o un dedo de zinc, en donde la última alternativa forma parte de la invención reivindicada.

En algunas realizaciones, las proteínas de fusión comprenden una pluralidad de dominios catalíticos.

En algunas realizaciones, la proteína de fusión comprende, además, un dominio/secuencia de localización nuclear.

En un aspecto adicional, la invención proporciona métodos *ex vivo* para reducir la metilación de una secuencia de ADN seleccionada en una célula de mamífero, en donde el método comprende poner en contacto la célula con una proteína de fusión, tal como se indica en las reivindicaciones, que comprende un dominio de unión a ADN diseñado que se une específicamente a una secuencia diana, en el que la secuencia diana se encuentra a menos de 250 pb de la secuencia de ADN seleccionada, y un dominio catalítico que comprende una secuencia que cataliza la hidroxilación de citosinas metiladas en el ADN.

En otro aspecto, la invención proporciona métodos *ex vivo* para reducir la metilación de una secuencia de ADN seleccionada en una célula de mamífero, en donde el método comprende poner en contacto la célula con un ácido nucleico, tal como se indica en las reivindicaciones, codificante de una proteína de fusión que comprende un dominio de unión a ADN diseñado que se une específicamente a una secuencia diana, en el que la secuencia diana se encuentra a menos de 250 pb de la secuencia de ADN seleccionada, y un dominio catalítico que comprende una secuencia que cataliza la hidroxilación de citosinas metiladas en el ADN.

En otro aspecto, la proteína de fusión o ácido nucleico codificante de la proteína de fusión, tal como se define en las reivindicaciones, para la utilización en un método de reducción de la metilación de un dirresiduo CpG en una secuencia de ADN seleccionada en una célula de mamífero para tratar el cáncer, en donde la proteína de fusión comprende un dedo de zinc diseñado que se une específicamente a una secuencia diana que se encuentra a menos de 250 pb del dirresiduo CpG en la secuencia de ADN seleccionada en la célula de mamífero, y un dominio catalítico que comprende una secuencia que cataliza la hidroxilación de citosina metilada en el ADN, en donde el método comprende poner en contacto la célula con la proteína de fusión o ácido nucleico codificante de la proteína de fusión.

En algunas realizaciones, el dominio catalítico comprende los aminoácidos 1418 a 2136 de TET1.

En algunas realizaciones, la célula es una célula humana.

En algunas realizaciones, la célula se encuentra en un animal vivo, p. ej., un mamífero o pez, p. ej., el pez cebra.

En algunas realizaciones, la secuencia de DNA seleccionada es una secuencia de un gen indicado en la figura 7.

En algunas realizaciones, la secuencia de ADN se selecciona de una secuencia de los genes RASSF1A, MGMT, BRCA1, p16, CDH1, pf6, TIMP-3, p15, p73, DAPK, APC, BRCA2, PAX3, hMLHI y RELN.

En algunas realizaciones, la célula está presente en un individuo con diagnóstico de una enfermedad indicada en la fig. 7y la secuencia de ADN seleccionada es un gen asociado a esa enfermedad tal como se indica en la fig. 7.

A menos que se definan de otro modo, todos los términos técnicos y científicos utilizados en la presente memoria presentan los mismos significados entendidos comúnmente por el experto habitual en la materia a la que se refiere la presente invención. En la presente memoria se describen métodos y materiales para la utilización en la presente invención; también pueden utilizarse otros métodos y materiales adecuados que son conocidos de la técnica. Los materiales, métodos y ejemplos son meramente ilustrativos y no pretenden ser limitativos. En caso de conflicto, prevalecerá la presente memoria, incluyendo las definiciones.

Otras características y ventajas de la invención resultarán evidentes a partir de la descripción detallada y figuras siguientes, y a partir de las reivindicaciones adjuntas.

Descripción de los dibujos

Figuras 1A-C. Estructura de TET1 y las proteínas de fusión TAL-TET1.

A) El esquema muestra la arquitectura de dominio predicho de la proteína TET1 (TET1LC), el dominio catalítico de la proteína TET1 (TET1DC) y la versión truncada de TET1 en la que se ha eliminado el dominio CXXC (TET1NZ). Entre las características mostradas se incluyen el dominio de unión a zinc de tipo CXXC (CXXC), tres señales de localización nuclear (NLS, por sus siglas en inglés), la región rica en cisteínas (cys-rich) y el dominio de hélice β de doble cadena (DSBH, por sus siglas en inglés) (Tahiliani et al., Science 2009). B) Esquema que muestra una proteína de fusión TAL-TET1 de ejemplo en la que TET1 está fusionada con el extremo C-terminal de la proteína TALE mediante un conector Gly₃Ser. En algunas realizaciones, la TET1 está fusionada con el extremo N-terminal. La proteína TALE diseñada se une a la secuencia diana de ADN y de esta manera localiza la proteína TET1 en esta región. C) Actividades de desmetilasa de proteínas de fusión TAL-TET1 que portan conectores interdominio de diversas longitudes. Los gráficos muestran la fracción de CpG metilada (eje y) para diferentes posiciones a lo largo de la longitud del segundo intrón KLF4 (eje x, numerado respecto al inicio del intrón) en células K562 transfectadas con plásmidos que expresan proteínas KL1 TALE-TET1 portadores de conectores estándares o variantes entre los dominios TALE y TET1 o un plásmido de control que expresa GFP. La flecha a lo largo del eje x indica la localización y dirección (5' a 3') del sitio de unión a la diana TALE. Cada punto representa datos de un único experimento de transfección. Se evaluó el estado de metilación utilizando la secuenciación de bisulfito de alto rendimiento.

Figura 2. Matrices TAL con diana en Klf4 humana. Secuencia de Klf4 humana intrón 1-2. Se muestran las secuencias de intrón en letra minúscula, con las secuencias de exón vecinas en letra mayúscula. Los cuatro sitios para los que se generaron dominios diseñados de matriz de repetición de efector TAL personalizado se muestran subrayados y numerados. La secuencia directa es la SEC ID n.º 1.

Figura 3. Desmetilación dirigida en el gen KLF4 humano por proteínas de fusión TAL-TET1DC que alojan un conector Gly₃Ser. Los gráficos muestran la desmetilación observada tras la transfección de 4 proteínas de fusión TAL-TET1DC diferentes en células K562 humanas. La desmetilación se calcula para cada CpG mediante la resta del % de CpG no metilado en células simuladamente transfectadas o transfectadas con GFP respecto del % de CpG no metilado en células en que se expresaba la proteína TAL-TET1DC. La posición del sitio de unión de TAL se representa mediante una flecha y la dirección de la flecha indica la orientación N-terminal a C-terminal de la proteína. El sitio de unión de efector de TAL específico que es la diana se indica mediante un subíndice, en donde TAL_i se refiere al sitio de unión n.º 1, etc., y correspondiente a la numeración de sitios mostrada en la figura 2. En donde era posible se calcularon las barras de error, mediante determinación del error estándar a partir de las medias de las muestras experimentales y de control.

Figura 4A. Desmetilación dirigida en el gen KLF4 humano utilizando proteínas de fusión TAL-TET1DC que alojan variantes de conector más extendidas. Los gráficos muestran la desmetilación observada tras la transfección de 5 proteínas de fusión TAL-TET1DC diferentes en células K562 humanas. Cada fusión utiliza el mismo dominio de unión de ADN de TAL y TET1DC, aunque difiere en la longitud y secuencia de aminoácidos utilizados como conector entre los dos dominios. La desmetilación se calcula para cada CpG mediante la resta del % de CpG no metilado en células simuladamente transfectadas o transfectadas con GFP respecto del % de CpG no metilado en células en que se expresaba la proteína TAL-TET1DC. La posición del sitio de unión de TAL se representa mediante una flecha y la dirección de la flecha indica la orientación N-terminal a C-terminal de la proteína.

Figura 4B. Actividades de desmetilación de las proteínas de fusión TALE-TET1 CD (fila superior) y TALE-TET1 FL (fila inferior) con diana en primer sitio (columna izquierda) y segundo sitio (columna derecha) en células K562 humanas. Los gráficos muestran la fracción de CpG metilados (eje y) para diferentes posiciones a lo largo de la longitud del segundo intrón *KLF4* (eje x, numeración respecto al inicio del intrón). La flecha horizontal indica la localización y la dirección (5' a 3') del sitio de unión a la diana TALE. Cada punto representa la media de tres experimentos de transfección independientes, en donde las barras indican los errores estándares de la media. El estado de metilación para cada experimento se evaluó utilizando la secuenciación de bisulfito de alto rendimiento. Las fusiones de TALE fuera de diana se unen a una secuencia no relacionada en el gen informador *EGFP*. Observar que los puntos de datos del control de GFP mostrados son los mismos en ambos gráficos y se ilustran dos veces para facilitar la comparación.

Figuras 5A-C. Secuencias de aminoácidos de TET1 de longitud completa (TET1LC; 5A (SEC ID n.º 2)), dominio catalítico de TET1 (TET1DC; 5B (SEC ID n.º 3)) y proteína TET1 con eliminación del dominio CXXC (TET1NZ; 5C (SEC ID n.º 4)).

Figuras 6A-D. Grupo de ilustraciones esquemáticas que muestran estrategias de multimerización de TET1. A) Fusión simple de múltiples dominios de TET1 con un dominio de unión de ADN de TALE en orientación aminoterminal, orientación carboxiterminal o ambas. B) Esquemas de interacción no covalente de ADN, que muestran la unión dependiente de ligando (multimerización inducible) o la unión directa mediante dominios de interacción proteína-proteína. C) Esquema de interacción covalente del ADN que muestra inteínas divididas, en la presente memoria designadas como «Int-N» en la proteína aminoterminal e Int-C en la proteína carboxiterminal, antes y después del corte y empalme. D) Un ejemplo de la utilización de varios enfoques de multimerización a la vez que muestra la utilización de ambas inteínas divididas, así como la multimerización de dominios TET1 en orientación aminoterminal o carboxiterminal con respecto a la proteína de unión al ADN de TALE.

Figura 7. Lista de enfermedades y afecciones humanas asociadas a hipermetilación. Adaptada de la Tabla 1 de Rodenhiser y Mann, CMAJ 174(3):341-348 (2006). Las referencias señaladas en la Tabla son las siguientes:

1. Esteller M, Corn PG, Baylin SB, et al. Cancer Res. 2001;61:3225-9.
2. Uhlmann K, Rohde K, Zeller C, et al. Int. J. Cancer 2003;106:52-9.
3. Gao Y, Guan M, Su B, et al. Clin. Chim. Acta 2004;349:173-9.
4. Bello MJ, Alonso ME, Aminos C, et al. Mutat. Res. 2004;554:23-32.
5. Mancini DN, Rodenhiser DJ, Ainsworth PJ, et al. Oncogene 1998;16:1161-9.
6. Szyf M, Pakneshan P, Rabbani SA. Biochem. Pharmacol. 2004;68:1187-97.
7. Widschwendter M, Jones PA. Oncogene 2002;21:5462-82.
8. Lea JS, Coleman R, Kurien A, et al. Am. J. Obstet. Gynecol. 2004;190:674-9.
9. De Zhu J. Cell Res 2005;15:272-80.
10. Chim CS, Kwong YL, Fung TK, et al. Leuk. Res. 2004;28:379-85.
11. Esteller M, Silva JM, Dominguez G, et al. J. Natl. Cancer Inst. 2000;92:564-9.
12. Li LC, Okino ST, Dahiya R. Biochim. Biophys. Acta 2004;1704:87-102.
13. Kurmasheva et al. Pediatr. Blood Cancer 2005;44:328-37.
14. Sharma RP. Schizophr. Res. 2005;72:79-90.
15. Costa E, Chen Y, Davis J, et al. Mol. Interv. 2002;2:47-57.
16. Lund G, Anderson L, Lauria M, et al. J. Biol. Chem. 2004;279:29147-54.

Figuras 8A-D. Desmetilación dirigida y activación del gen de HBB humano por proteínas de fusión TALE-TET1DC. A) Esquema que ilustra el locus de HBB humano con CpG indicados con flechas negras. La numeración indica la posición en el ADN respecto al sitio de inicio de la transcripción (flecha en ángulo recto). Las flechas indican la localización y dirección (5' a 3') de 10 sitios de unión de proteína de fusión TALE-TET1DC. B) Actividades de desmetilación de proteínas de fusión TALE-TET1DC HBB-1, -2, -3, -4, -5 y -6 en células K562. Los gráficos muestran la fracción de las CpG metiladas (eje y) para diferentes posiciones a lo largo de la longitud del promotor de HBB (eje x, numeración respecto al sitio de inicio de transcripción) en células K562 transfectadas con plásmidos que expresan una de las seis proteínas de fusión HBB TALE-TET1DC (círculos), una fusión fuera de diana TALE-TET1DC a un sitio en el gen de KLF4 (triángulos negros) o GFP (triángulos grises). Cada punto de datos representa la media de tres experimentos de transfección independientes, en donde las barras indican los errores estándares de la media. El estado de metilación para cada experimento se evaluó utilizando la secuenciación de bisulfito de alto rendimiento. Observar que los puntos de datos de GFP y TALE-TET1DC fuera de diana mostrados son iguales en ambos paneles y se representan dos veces para facilitar la comparación con las muestras experimentales. Las flechas numeradas indican la localización y la dirección (5' a 3') de los diversos sitios de unión HB TALE-TET1 DC. C) Actividades de desmetilación de proteínas de fusión TALE-TET1DC HB-7, -8, -9 y -10 en células K562. Los datos se representan tal como en (B). D) Niveles de expresión del ARNm de HBB (normalizado respecto a los niveles de ARNm de β -actina) en células K562 transfectadas con los plásmidos de expresión de proteína de fusión TALE-TET1DC indicados o un plásmido de expresión de GFP de control. Se muestran las medias de tres muestras independientes sometidas a ensayo mediante RT-PCR cuantitativa en donde las barras representan los errores estándares de la media. Los asteriscos indican muestras con valores significativamente mayores que los obtenidos con una proteína de fusión TALE-TET1DC de control fuera de diana (dirigido a un sitio en el gen KLF4 humano) según se determina mediante una prueba T de una cola ($P < 0,05$).

Figuras 9A-B. Proteínas de fusión dedo de zinc-TET1 diseñadas. A) Desmetilación inducida por la proteína de fusión DZ-TET1DC del gen KLF4 humano. Los gráficos muestran la fracción de CpG metilados (eje y) para diferentes posiciones a lo largo de la longitud del segundo intrón KLF4 (eje x, numeración respecto al inicio del intrón) en células K562 transfectadas con plásmidos que expresan una de las dos fusiones DZ-TET1DC con diana en KLF4, un DZ-TET1DC fuera de diana (dirigido a un sitio no relacionado en el gen VEGFA humano), o GFP. Las flechas horizontales indican la localización y la dirección (3' a 5') del sitio de unión a la diana ZF. Cada punto representa la media de tres experimentos de transfección independientes, en donde las barras indican los errores estándares de la media. El estado de metilación para cada experimento se evaluó utilizando la secuenciación de bisulfito de alto rendimiento. Observar que los puntos de datos de la fusión DZ-TET1DC fuera de diana y de control de GFP mostrados son los mismos en ambos gráficos y se ilustran dos veces para facilitar la comparación. B) (parte superior) Esquema que ilustra el locus de HBB humano con CpG indicados con flechas negras. La numeración indica la posición en el ADN respecto al sitio de inicio de la transcripción (flecha en ángulo recto). Las flechas coloreadas indican la localización y dirección (3' a 5') de seis sitios de unión de proteína de fusión DZ-TET1DC. (parte inferior) Gráficos que muestran la fracción de CpG metiladas (eje y) para diferentes posiciones a lo largo de la longitud del gen de HBB (eje x, numeración respecto al sitio de inicio de transcripción del promotor de HBB) en células K562 transfectadas con plásmidos que expresan una de seis fusiones DZ-TET1DC con diana en HBB (los círculos corresponden a las flechas en (parte superior)), un DZ-TET1DC fuera de diana (con diana en un sitio no relacionado en el gen de VEGFA humano) (triángulos negros) o GFP (triángulos grises). Las flechas indican la localización y la dirección (3' a 5') del sitio de unión a la diana ZF. La posición de los sitios de unión de DZ-TET1DC se ilustra con flechas que corresponden a los círculos en el gráfico. Se muestran datos para las proteínas DZ-TET1DC con diana en CpG -266 y +83/+110 en los paneles izquierdo y derecho, respectivamente. Cada punto representa la media de tres experimentos de transfección independientes, en donde las barras indican los errores estándares de la media. El estado de metilación para cada experimento se evaluó utilizando la secuenciación de bisulfito de alto rendimiento.

Figuras 10A-B. Experimento de curso temporal extendido tras la metilación del promotor en CpG y la expresión génica de HBB humano en células transfectadas con plásmidos que expresan proteínas de fusión TALE-TET1. A) Se midieron niveles de expresión de ARNm de HBB (normalizado respecto a los niveles de ARNm de β -actina) en células K562 transfectadas con plásmidos que expresan HB-4, HB-5, HB-6 o una proteína de fusión TALE-TET1 fuera de diana o con un plásmido de expresión de GFP de control, 4, 7, 14 y 30 días después de la transfección. (La fusión TALE-TET1 fuera de diana presenta como diana un sitio en el gen de KLF4 humano). Se muestran las medias de tres muestras independientes sometidas a ensayo mediante RT-PCR cuantitativa en donde las barras representan los errores estándares de la media. B) Los gráficos muestran la fracción de CpG metilados (eje y) para diferentes posiciones a lo largo de la longitud del promotor de HBB (eje x, numeración respecto al sitio de inicio de transcripción) para la mismas células transfectadas descritas en (A). Cada punto de datos representa la media de tres experimentos de transfección independientes, en donde las barras indican los errores estándares de la media. El estado de metilación para cada experimento se evaluó utilizando la secuenciación de bisulfito de alto rendimiento. Rombos, día 4; cuadrados, día 7; triángulos, día 14; círculos, día 30.

Descripción detallada

La invención se define en las reivindicaciones. En la presente memoria se describen proteínas de fusión que comprenden un dominio de unión a ADN (es decir, un dominio diseñado de unión a ADN personalizado) y una o más copias de un dominio catalítico (de una proteína diferente) que comprende una secuencia que cataliza la hidroxilación de citosinas metiladas en ADN con un conector opcional entre los dominios, p. ej., entre los dominios de unión a ADN y dominios catalíticos, y/o entre dominios catalíticos en proteínas de fusión que incluyen múltiples dominios catalíticos. Tal como se indica en la presente memoria, la expresión de una proteína de fusión de matriz de repetición efectora de TAL-TET1 en células K562 humanas resulta en la desmetilación de CpG en estrecha proximidad al sitio de unión al que se une la parte de matriz de repetición de efector TAL de la proteína.

En un aspecto de la exposición es una proteína híbrida que consiste en una matriz de repetición de efector diseñada de tipo activador de transcripción (TAL) fusionada con la proteína TET1 de longitud completa (TET1LC), el dominio catalítico de TET1 (TET1DC) o una versión truncada de la proteína TET1LC en la que el dominio putativo de unión a ADN nativo se ha eliminado (TET1NZ). La especificidad de unión a ADN se define por la matriz diseñada de repetición de efector TAL. Dichas proteínas de unión a ADN pueden diseñarse para la unión a esencialmente cualquier secuencia de ADN y el trabajo publicado de diversos laboratorios, así como trabajo publicado y no publicado del laboratorio de los presentes inventores, ha demostrado que dichos dominios personalizables pueden presentar como diana eficiente una diversidad de dominios fusionados con localizaciones genómicas específicas (Reyon et al., FLASH assembly of TALENs for high-throughput genome editing. Nat. Biotechnol. (2012).doi:10.1038/nbt.2170; Moscou y Bogdanove, Science 326, 1501-1501 (2009); Boch et al., Science 326, 1509-1512 (2009); Miller et al., Nat. Biotechnol. 29, 143-148 (2010)). Por ejemplo, se han fusionado matrices de repetición diseñadas de efector TAL con el dominio de escisión de la endonucleasa FokI, así como con activadores y represores, y actúan dirigiendo dichos dominios a una secuencia definida por el usuario dentro del contexto del genoma.

Dominios de unión al ADN

Las proteínas de fusión descritas en la presente memoria pueden incluir cualquier dominio de unión a ADN (DBD, por sus siglas en inglés) conocido de la técnica o diseñado para un sitio de unión específico. Entre los DBD de ejemplo se incluyen matrices de repetición diseñadas o nativas de efector TAL, dedos de zinc de diseño o nativos, meganucleasas dirigidas, nucleasas química y DBD nativos.

Matrices de repetición de efector TAL

En aspectos de la exposición, los efectores TAL de bacterias patógenas vegetales en el género *Xanthomonas* desempeñan importantes funciones en enfermedades, o inducen la inmunidad, mediante la unión a ADN del huésped y activando genes del huésped específicos de efector. La especificidad depende de un número variable de repeticiones de aminoácidos en el efector, imperfectas, típicamente -33-35. Los polimorfismos están presentes principalmente en las posiciones de repetición 12 y 13, a las que se hace referencia en la presente memoria como dirresiduos variables de repetición (RVD, por sus siglas en inglés). Las RVD de efectores TAL corresponden a los nucleótidos en sus sitios diana de una manera lineal directa, una RVD para un nucleótido, con cierta degeneración y sin dependencia aparente del contexto. En algunas realizaciones, la región polimórfica que proporciona especificidad de nucleótidos puede expresarse como un trirresiduo o triplete (que representa las posiciones 11, 12 y 13 en la repetición).

Cada repetición de unión al ADN puede incluir una RVD que determina el reconocimiento de un par de bases en la secuencia de ADN diana, en la que cada repetición de unión a ADN es responsable de reconocer un par de bases en la secuencia de ADN diana. En algunas realizaciones, la RD puede comprender uno o más de: HA para reconocer C; ND para reconocer C; HI para reconocer C; HN para reconocer G; NA para reconocer G; NH para reconocer G; SN para reconocer G o A; YG para reconocer T y NK para reconocer G, y uno o más de: HD para reconocer C; NG para reconocer T; NI para reconocer A; NN para reconocer G o A; NS para reconocer A o C o G o T; N* para reconocer C o T, en donde * representa un hueco en la segunda posición de la RVD; HG para reconocer T; H* para reconocer T, en donde * representa un hueco en la segunda posición de la RVD e IG para reconocer T.

Las proteínas TALE pueden resultar útiles en investigación y biotecnología como nucleasas quiméricas dirigidas que pueden facilitar la recombinación homóloga, la reparación dirigida a homología o la reparación de unión de extremos no homóloga con tendencia a errores para la ingeniería del genoma (p. ej., para añadir o potenciar características útiles para biocombustibles o biorrenovables en plantas). Dichas proteínas también pueden resultar útiles como, por ejemplo, factores de transcripción, y especialmente para aplicaciones terapéuticas que requieren un nivel muy elevado de especificidad, tal como terapéuticos contra patógenos (p. ej., virus) a modo de ejemplos no limitativos.

Se conocen de la técnica métodos para generar matrices TALE diseñadas; ver, p. ej., el sistema de alto rendimiento en fase sólida automatizable basado en la ligación rápida (FLASH, por sus siglas en inglés) descrito en el documento n.º USSN 61/610,212 y en Reyon et al., *Nature Biotechnology* 30,460-465 (2012); así como los métodos descritos en Bogdanove y Voytas, *Science* 333, 1843-1846 (2011); Bogdanove et al., *Curr. Opin. Plant Biol.* 13, 394-401 (2010); Scholze y Boch, *J. Curr. Opin. Microbiol.* (2011); Boch et al., *Science* 326, 1509-1512 (2009); Moscou y Bogdanove, *Science* 326, 1501 (2009); Miller et al., *Nat. Biotechnol.* 29, 143-148 (2011); Morbitzer et al., *T. Proc. Natl. Acad. Sci. USA* 107, 21617-21622 (2010); Morbitzer et al., *Nucleic Acids Res.* 39, 5790-5799 (2011); Zhang et al., *Nat. Biotechnol.* 29, 149-153 (2011); Geissler et al., *PLoS ONE* 6, e19509 (2011); Weber et al., *PLoS ONE* 6, e19722 (2011); Christian et al., *Genetics* 186, 757-761 (2010); Li et al., *Nucleic Acids Res.* 39, 359-372 (2011); Mahfouz et al., *Proc. Natl. Acad. Sci. USA* 108, 2623-2628 (2011); Mussolino et al., *Nucleic Acids Res.* (2011); Li et al., *Nucleic Acids Res.* 39, 6315-6325 (2011); Cermak et al., *Nucleic Acids Res.* 39, e82 (2011); Wood et al., *Science* 333, 307 (2011); Hockemeyer et al., *Nat. Biotechnol.* 29, 731-734 (2011); Tesson et al., *Nat. Biotechnol.* 29, 695-696 (2011); Sander et al., *Nat. Biotechnol.* 29, 697-698 (2011); Huang et al., *Nat. Biotechnol.* 29, 699-700 (2011) y Zhang et al., *Nat. Biotechnol.* 29, 149-153 (2011).

Dedos de zinc

Las proteínas de dedos de zinc son proteínas de unión al ADN que contienen uno o más dedos de zinc, minidominios que contienen zinc independientemente doblados, la estructura de los cuales es bien conocido de la técnica y se define en, por ejemplo, Miller et al., 1985, *EMBO J.*, 4:1609; Berg, 1988, *Proc. Natl. Acad. Sci. USA*, 85:99; Lee et al., 1989, *Science*, 245:635 y Klug, 1993, *Gene*, 135:83. Las estructuras cristalinas de la proteína de dedo de zinc Zif268 y sus variantes unidas al ADN muestran un patrón semiconservador de interacciones, en el que normalmente tres aminoácidos de la hélice alfa del dedo de zinc entran en contacto con tres pares de bases contiguos o un «subsitio» en el ADN (Pavletich et al., 1991, *Science*, 252:809; Elrod-Erickson et al., 1998, *Structure*, 6:451). De esta manera, la estructura cristalina de Zif268 sugiere que los dominios de unión al ADN del dedo de zinc podrían funcionar de una manera modular con una interacción uno a uno entre un dedo de zinc y un «subsitio» de tres pares de bases en la secuencia de ADN. En factores de transcripción de dedo de zinc naturales, normalmente se unen múltiples dedos de zinc entre sí en una serie en tándem para conseguir el reconocimiento específico de secuencia de una secuencia de ADN contiguo (Klug, 1993, *Gene* 135:83).

Múltiples estudios han mostrado que resulta posible diseñar artificialmente las características de unión del ADN de dedos de zinc individuales mediante la aleatorización de los aminoácidos en las posiciones alfa-helicoidales implicadas en la unión del ADN y utilizando metodologías de selecciones, tales como la exposición fágica, para identificar las variantes deseadas capaces de unirse a los sitios diana de interés del ADN (Rebar et al., 1994, *Science*, 263:671; Choo et al., 1994 *Proc. Natl. Acad. Sci. USA*, 91:11163; Jamieson et al., 1994, *Biochemistry* 33:5689; Wu et al., 1995 *Proc. Natl. Acad. Sci. USA*, 92: 344). Dichas proteínas de dedo de zinc recombinantes pueden fusionarse con dominios funcionales, tales como activadores transcripcionales, represores de transcripción, dominios de metilación y nucleasas para regular la expresión génica, alterar la metilación del ADN e introducir alteraciones dirigidas en genomas de organismos modelo, plantas y células humanas (Carroll, 2008, *Gene Ther.*, 15:1463-68; Cathomen, 2008, *Mol. Ther.*, 16:1200-07; Wu et al., 2007, *Cell. Mol. Life Sci.*, 64:2933-44).

La adopción generalizada y uso a gran escala de la tecnología de proteínas de dedos de zinc se han visto dificultados por la falta continua de un método robusto, fácil de utilizar y disponible públicamente para el diseño de matrices de dedos de zinc. Un enfoque existente, conocido como «ensamblaje modular» defiende la unión simple de módulos de dedo de zinc preseleccionados formando matrices (Segal et al., 2003, *Biochemistry*, 42:2137-48; Beerli et al., 2002, *Nat. Biotechnol.*, 20:135-141; Mandell et al., 2006, *Nucleic Acids Res.*, 34:W516-523; Carroll et al., 2006, *Nat. Protoc.* 1:1329-41; Liu et al., 2002, *J. Biol. Chem.*, 277:3850-56; Bae et al., 2003, *Nat. Biotechnol.*, 21:275-280; Wright et al., 2006, *Nat. Protoc.*, 1:1637-52). Aunque es suficientemente sencillo para que sea puesto en práctico por cualquier investigador, los últimos informes han demostrado una elevada tasa de fracaso para este método, particularmente en el contexto de las nucleasas de dedo de zinc (Ramirez et al., 2008, *Nat. Methods*, 5:374-375; Kim et al., 2009, *Genome Res.* 19:1279-88), una limitación que normalmente requiere la construcción y ensayo celular de un número muy elevado de proteínas de dedo de zinc para cualquier gen diana (Kim et al., 2009, *Genome Res.* 19:1279-88).

Se ha mostrado que los métodos basados en la selección combinatorial que identifican las matrices de dedos de zinc de bibliotecas aleatorizadas presentan tasas de éxito más altas que el ensamblaje modular (Maeder et al., 2008, *Mol. Cell*, 31:294-301; Joung et al., 2010, *Nat. Methods*, 7:91-92; Isalan et al., 2001, *Nat. Biotechnol.*, 19:656-660). En realizaciones preferentes, las matrices de dedo de zinc se describen o se generan tal como se describe en los

documentos n.º WO 2011/017293 y n.º WO 2004/099366. Se describen DBD de dedo de zinc adecuados en las patentes US n.º 6.511.808, n.º 6.013.453, n.º 6.007.988 y n.º 6.503.717 y la solicitud de patente US n.º 2002/0160940.

DBD nativos

En algunos aspectos de la exposición, puede utilizarse un DBD nativo (p. ej., una parte de una proteína de unión a ADN no diseñada de tipo salvaje que se une a una secuencia diana específica). Por ejemplo, puede utilizarse el DBD de un factor de transcripción, nucleasa, histona, telomerasa u otra proteína de unión a ADN. Habitualmente entre los DBD se incluye una estructura que facilita la interacción específica con una secuencia de ácido nucleico diana; entre las estructuras de DBD comunes se incluyen hélice-giro-hélice, dedo de zinc, cremallera de leucinas, hélice alada, hélice alada giro hélice, hélice-bucle-hélice y caja hmg. El DBD nativo puede proceder de cualquier organismo. Ver, p. ej., Kummerfeld y Teichmann, *Nucleic Acids Res.* 34 (número dedicado a bases de datos): D74-81 (2006). Los residuos en una proteína de unión a ADN que entran en contacto con el ADN y que, de esta manera, forman parte del DBD, pueden determinarse empíricamente o predecirse computacionalmente, p. ej., tal como se describe en Tjong y Zhou, *Nucl. Acids Res.* 35:1465-1477 (2007). Puede utilizarse una base de datos de proteínas de unión a ADN y DBD para la utilización en las presentes composiciones y métodos; ver, p. ej., Harrison, *Nature*, 353, 715-719 (1991); Karmirantzou y Hamodrakas, *Protein Eng.* 14(7): 465-472 (2001); Kumar et al., *BMC Bioinformatics*. 8:463 (2007); Kumar et al., *J Biomol Struct Dyn.* 26(6):679-86 (2009); Lin et al., *PLoS One*. 6(9):e24756 (2011).

En el caso de que se utilice un DBD nativo en una proteína de fusión descrita en la presente memoria, el dominio catalítico procedente de una proteína diferente.

Meganucleasas de reconocimiento

Las meganucleasas son endonucleasas específicas de secuencia que se originan en una diversidad de organismos, tales como bacterias, levaduras, algas y orgánulos vegetales. Las meganucleasas de reconocimiento presentan sitios de reconocimiento de 12 a 30 pares de bases; se han descrito sitios de unión a ADN personalizados con sitios de reconocimiento de meganucleasa de 18 pb y 24 pb de longitud, y cualquiera de ellos puede utilizarse en los presentes métodos y constructos. Ver, p. ej., Silva, G., et al., *Current Gene Therapy*, 11:11-27, (2011); Arnould et al., *Journal of Molecular Biology*, 355:443-58 (2006); Arnould et al., *Protein Engineering Design & Selection*, 24:27-31 (2011) y Stoddard, Q. *Rev. Biophys.* 38, 49 (2005); Grizot et al., *Nucleic Acids Research*, 38:2006-18 (2010).

Nucleasas químicas

Los dominios de unión al ADN de las denominadas «nucleasas químicas» (Pingoud y Silva, *Nat. Biotechnol.* 25:743-4 (2007)), p. ej., oligonucleótidos formadores de triplex o ácidos péptido-nucleicos también pueden utilizarse en las presentes composiciones y métodos; ver, p. ej., Schleifman et al., *Methods Mol Biol.* 2008;435:175-90; Arimondo et al., *Mol Cell Biol.* 2006 Jan; 26(1):324-33; Majumdar et al., *J Biol Chem.* 2008 Apr 25; 283(17):11244-52; Simon et al., *Nucleic Acids Res.* 2008 Jun; 36(11):3531-8 o Eisenschmidt et al., *Nucleic Acids Res.* 2005; 33(22):7039-47.

Dominios catalíticos

Entre las proteínas de fusión se incluyen un dominio catalítico que comprende una secuencia que cataliza la hidroxilación de las citoquinas metiladas en el ADN. Entre las proteínas de ejemplo se incluyen la familia de proteínas de traslocación diez-once (TET, por sus siglas en inglés) 1-3, enzimas que convierten la 5-metilcitosina (5-mC) en 5-hidroximetilcitosina (5-hmC) en el ADN.

Las secuencias para la TET1-3 humana son conocidas de la técnica y se muestran en la tabla a continuación:

GenBank n.º de acceso		
Gen	Aminoácido	Ácido nucleico
TET1	NP_085128.2	NM_030625.2
TET2*	NP_001120680.1 (var 1)	NM_001127208.2
	NP_060098.3 (var 2)	NM_017628.4
TET3	NP_659430.1	NM_144993.1
* Variante (1) representa el transcrito más largo y codifican la isoforma más larga (a). La variante (2) difiere en la región no traducida (RNT) 5', la RNT 3' y la secuencia codificante respecto a la variante 1. La isoforma resultante (b) es más corta y presenta un extremo C-terminal diferente a la isoforma a.		

En algunas realizaciones, puede incluirse la totalidad o parte de la secuencia de longitud completa del dominio catalítico, p. ej., un módulo catalítico que comprende la extensión rica en cisteínas y el dominio 2OGFeDO codificado por 7 exones altamente conservados, p. ej., el dominio catalítico TET1 que comprende los aminoácidos 1580 a 2052; TET2 que comprende los aminoácidos 1290 a 1905 y TET3 que comprende los aminoácidos 966 a 1678. Ver, p. ej., la fig. 1 de Iyer et al., *Cell Cycle*. 2009, 1 de junio;8(11):1698-710. Epub. 2009, 27 de junio. para una alineación que ilustra los residuos catalíticos clave en la totalidad de las tres proteínas Tet, y los materiales suplementarios de las

mismas (disponibles en el sitio de ftp://ftp.ncbi.nih.gov/pub/aravind/DONS/supplementary_material_DONS.html) para las secuencias de longitud completa (ver, p. ej., seq 2c); en algunas realizaciones, la secuencia incluye los aminoácidos 1418 a 2136 de TET1 o la región correspondiente de TET2/3.

Otros módulos catalíticos pueden ser de proteínas identificadas en Iyer et al., 2009.

Construcción de proteínas de fusión

Para generar una proteína recombinante funcional, el dominio de unión al ADN se fusiona con por lo menos un dominio catalítico. La fusión de dominios catalíticos con DBD para formar proteínas de fusión funcionales implica solo técnicas rutinarias de biología molecular que ponen en práctica comúnmente los expertos en la materia; ver, por ejemplo, la patente US n.º 6.511.808, n.º 6.013.453, n.º 6.007.988, n.º 6.503.717 y la solicitud de patente US n.º 2002/0160940). Los dominios catalíticos pueden estar asociados al dominio DBD en cualquier posición adecuada, incluyendo el extremo C-terminal o N-terminal del DBD. En algunas realizaciones, entre las proteínas de fusión se incluyen múltiples dominios catalíticos, p. ej., en uno o ambos extremos del DBD, p. ej., concatenados junto con un conector intermedio opcional; de esta manera, puede haber uno o más dominios catalíticos en cada extremo del DBD.

Alternativamente, los dominios catalíticos, p. ej., las unidades de TET1, podrían multimerizarse mediante dominios de interacción proteína-proteína constitutivos o inducibles específicos (tales como dominios de cremallera de leucinas o el sistema iDimerize de ClonTech, sistemas y ligandos de homodimerización y heterodimerización (p. ej., AP20187 y AP21967), que anteriormente eran proporcionados por ARID bajo el nombre comercial ARGENT. El homodimerizador B/B (AP20187) induce la dimerización de dos proteínas, cada una de las cuales contiene el dominio de homodimerización DmrB (ver la tabla, posteriormente, que explica la nomenclatura de los dominios). El heterodimerizador A/C (AP21967) induce la dimerización de una proteína que posee el dominio DmrA y una segunda proteína que contiene el dominio DmrC. El solubilizador D/D (alternativa a AP21998) induce la disociación/desagregación de proteínas que posee dominios DmrD. DmrD causa la autoasociación automática de proteínas fusionadas con el mismo; ver, p. ej., Burnett et al., J. Leukoc. Biol. 75(4):612-623 (2004); Freeman et al., Cancer Res. 63(23):8256-8563 (2003); Castellano et al., Curr. Biol. 9(7): 351-360 (1999); Crabtree y Schreiber, Trends Biochem. Sci. 21(11): 418-422 (1996); Graef et al., Embo. J. 16(18): 5618-5628 (1997); Muthuswamy et al., Mol. Cell. Biol. 19(10): 6845-6857 (1999)). Alternativamente, la multimerización podría conseguirse mediante la utilización de inteínas divididas, una clase de péptidos inteína autocatalíticos que permiten el corte y empalme covalente perfecto de dos proteínas separadas de una manera predecible y eficiente (d'Avignon, et al., Biopolymers. 2006, 15 de oct.;83(3):255-67; Zitzewitz, et al., Biochemistry. 1995, 3 de oct.;34(39):12812-9; Li et al., Hum Gene Ther. 2008 sep.;19(9):958-64). Tanto la interacción proteína-proteína como los enfoques de inteína podrían optimizarse para producir cadenas multimerizadas muy largas de dominios catalíticos. Las figuras 6A-D muestran esquemas de ejemplo para la multimerización.

Entre los conectores que pueden utilizarse en dichas proteínas de fusión (o entre proteínas de fusión en una estructura concatenada) se puede incluir cualquier secuencia que no interfiera con la función de las proteínas de fusión. En realizaciones preferentes, los conectores son cortos, p. ej., de 2 a 20 aminoácidos, y normalmente son flexibles (es decir, comprenden aminoácidos con un elevado grado de libertad, tal como glicina, alanina y serina). En algunas realizaciones, el conector comprende una o más unidades que consisten en la unidad GGGS (SEC ID n.º 17) o GGGGS (SEC ID n.º 18), p. ej., dos, tres, cuatro o más repeticiones de GGGS (SEC ID n.º 17) o GGGGS (SEC ID n.º 18). También pueden utilizarse otras secuencias de conector.

Utilización de proteínas de fusión

Las proteínas de fusión diseñadas y producidas utilizando los métodos descritos en la presente memoria pueden utilizarse para realizar cualquier función en la que se desee desmetilar una secuencia de ADN específica. Las proteínas de fusión pueden resultar útiles para el tratamiento de enfermedades; por ejemplo, las proteínas de fusión pueden direccionarse a una región hipermetilada de un gen indicado en la Tabla 1 de Rodenhiser y Mann, CMAJ 174(3):341-348 (2006) como hipermetilado en una enfermedad, que se reproduce en la fig. 7 en la presente memoria, o una referencia expuesta en la misma, o de otro modo, alterando la estructura o función de un gen dado *in vivo*. Las proteínas de fusión de la presente invención también resultan útiles como herramientas de investigación, por ejemplo en la realización de estudios genómicos funcionales *in vivo* o *in vitro* (ver, por ejemplo, la patente US n.º 6.503.717 y en la solicitud de patente US n.º 2002/0164575).

Sistemas de expresión de polipéptidos

Con el fin de utilizar las proteínas de fusión descritas, puede resultar deseable expresar las proteínas diseñadas a partir de un ácido nucleico codificante de las mismas. Lo anterior puede llevarse a cabo de una diversidad de formas. Por ejemplo, el ácido nucleico codificante de la proteína de fusión puede clonarse en un vector intermedio para la transformación en células procarióticas o eucarióticas para la replicación y/o la expresión. Los vectores intermedios normalmente son vectores procarióticos, p. ej., plásmidos, o vectores lanzadera, o vectores de insecto, para el almacenamiento o la manipulación del ácido nucleico codificante de la proteína de fusión o para la producción de la proteína de fusión. El ácido nucleico codificante de la proteína de fusión también puede clonarse en un vector de

expresión para la administración en una célula vegetal, célula animal, preferentemente una célula de mamífero o una célula humana, célula fúngica, célula bacteriana o célula protozoaria.

Para conseguir la expresión, la proteína de fusión normalmente se subclona en un vector de expresión que contiene un promotor para dirigir la transcripción. Los promotores bacterianos y eucarióticos adecuados son bien conocidos de la técnica y se describen en, p. ej., Sambrook et al., *Molecular Cloning, A Laboratory Manual* (3a ed., 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990) y *Current Protocols in Molecular Biology* (Ausubel et al., eds., 2010). Los sistemas de expresión bacteriana para expresar la proteína de repetición TALE diseñada están disponibles en, p. ej., *E. coli*, *Bacillus* sp. y *Salmonella* (Palva et al., 1983, *Gene* 22:229-235). Los kits para dichos sistemas de expresión se encuentran disponibles comercialmente. Los sistemas de expresión eucarióticos para células de mamífero, levaduras y células de insecto son bien conocidos de la técnica y también se encuentran disponibles comercialmente.

El promotor utilizado para dirigir la expresión del ácido nucleico de proteína de fusión depende de la aplicación particular. Por ejemplo, normalmente se utiliza un promotor constitutivo fuerte para la expresión y la purificación de las proteínas de fusión. En contraste, en el caso de que la proteína de fusión vaya a administrarse *in vivo* para la regulación génica, puede utilizarse un promotor constitutivo o inducible, dependiendo del uso particular de la proteína de fusión. Además, el promotor preferente para la administración de la proteína de fusión puede ser un promotor débil, tal como TK del VHS, o un promotor que presente una actividad similar. El promotor puede incluir, además, elementos que son sensibles a la transactivación, p. ej., elementos de respuesta a la hipoxia, elementos de respuesta a Gal4, elementos de respuesta del represor lac y sistemas de control de molécula pequeña, tales como sistemas regulados por tetraciclina y el sistema RU-486 (ver, p. ej., Gossen y Bujard, 1992, *Proc. Natl. Acad. Sci. USA*, 89:5547; Oligino et al., 1998, *Gene Ther.*, 5:491-496; Wang et al., 1997, *Gene Ther.*, 4:432-441; Neering et al., 1996, *Blood*, 88:1147-55 y Rendahl et al., 1998, *Nat. Biotechnol.*, 16:757-761).

Además del promotor, el vector de expresión normalmente contiene una unidad de transcripción o casete de expresión que contiene todos los elementos adicionales requeridos para la expresión del ácido nucleico en células huésped, ya sea procarióticos o eucarióticos. Un casete de expresión típico, de esta manera, contiene un promotor operablemente ligado, p. ej., a la secuencia de ácido nucleico codificante de la proteína de fusión, y cualesquiera señales requeridas, p. ej., para la poliadenilación eficiente del transcrito, terminación transcripcional, sitios de unión ribosómica o terminación de la traducción. Entre los elementos adicionales del casete pueden incluirse, p. ej., intensificadores y señales intrónicas procesadas heterólogas.

El vector de expresión particular utilizado para transportar la información genética hasta el interior de la célula se selecciona con respecto a la utilización pretendida de la proteína de fusión, p. ej., la expresión en plantas, animales, bacterias, hongos, protozoos, etc. Entre los vectores de expresión bacteriana estándares se incluyen plásmidos, tales como los plásmidos basados en pBR322, pSKF, pET23D y los sistemas de expresión de fusión con etiqueta disponibles comercialmente, tales como GST y LacZ. Una proteína de fusión con etiqueta preferente es la proteína de unión a maltosa, «MBP» (por sus siglas en inglés). Dichas proteínas de fusión con etiqueta pueden utilizarse para la purificación de la proteína de repetición TALE diseñada. También pueden añadirse etiquetas de epítipo a las proteínas recombinantes para proporcionar métodos convenientes de aislamiento, para el seguimiento de la expresión y para el seguimiento de la localización celular y subcelular, p. ej., c-myc o FLAG.

Los vectores de expresión que contienen elementos reguladores de virus eucarióticos se utilizan con frecuencia en vectores de expresión eucarióticos, p. ej., vectores de SV40, vectores del virus del papiloma y vectores derivados del virus de Epstein-Barr. Entre otros vectores eucarióticos de ejemplo se incluyen pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE y cualquier otro vector que permita la expresión de proteínas bajo la dirección del promotor temprano de SV40, el promotor de metalotioneína, el promotor del virus tumoral mamario murino, el promotor del virus del sarcoma de Rous, el promotor de polihedrina u otros promotores que se ha mostrado que resultan eficaces para la expresión en células eucarióticas.

Algunos sistemas de expresión presentan marcadores para la selección de líneas celulares transfectadas establemente, tales como timidina quinasa, higromicina B fosfotransferasa y dihidrofolato reductasa. También resultan adecuados los sistemas de expresión de alto rendimiento, tales como la utilización de un vector baculovirus en células de insecto, con la secuencia codificante de la proteína de fusión bajo la dirección del promotor de polihedrina u otros promotores de baculovirus fuertes.

Entre los elementos que se incluyen típicamente en los vectores de expresión se incluye también un replicó que funciona en *E. coli*, un gen codificante de resistencia a antibiótico para permitir la selección de bacterias que alojan plásmidos recombinantes, y sitios de restricción únicos en regiones no esenciales del plásmido para permitir la inserción de secuencias recombinantes.

Se utilizan métodos de transfección estándares para producir líneas celulares bacterianas, de mamífero, de levadura o de insecto que expresan grandes cantidades de proteínas, que después se purifican utilizando técnicas estándares (ver, p. ej., Colley et al., 1989, *J. Biol. Chem.*, 264:17619-22; *Guide to Protein Purification*, in *Methods in Enzymology*, vol. 182 (Deutscher, ed., 1990)). La transformación de las células eucarióticas y procarióticas se lleva a cabo de

acuerdo con técnicas estándares (ver, p. ej., Morrison, 1977, J. Bacteriol. 132:349-351; Clark-Curtiss y Curtiss, Methods in Enzymology 101:347-362 (Wu et al., eds, 1983).

Puede utilizarse cualquiera de los procedimientos conocidos para introducir secuencias de nucleótidos foráneos en las células huésped. Entre ellos se incluyen la utilización de la transfección con fosfato cálcico, polibreno, fusión de protoplastos, electroporación, nucleofección, liposomas, microinyección, ADN desnudo, vectores plasmídicos, vectores víricos, tanto episómicos como integrativos, y cualquiera de los demás métodos bien conocidos para introducir ADN genómico clonado, ADNc, ADN sintético u otro material genético foráneo en una célula huésped (ver, p. ej., Sambrook et al., *supra*). Solo resulta necesario que el procedimiento particular de ingeniería genética utilizado sea capaz de introducir con éxito por lo menos un gen en la célula huésped capaz de expresar la proteína seleccionada.

En algunas realizaciones, la proteína de fusión incluye un dominio de localización nuclear que permite que la proteína sea traslocada al núcleo. Se conocen varias secuencias de localización nuclear (SLN), y puede utilizarse cualquier SLN adecuada. Por ejemplo, muchas SLN presentan una pluralidad de aminoácidos básicos, denominados repeticiones básicas bipartitas (revisión en Garcia-Bustos et al, 1991, Biochim. Biophys. Acta, 1071:83-101). Puede introducirse un NLC que contenga repeticiones básicas bipartitas en cualquier parte de la proteína quimérica y resulta en la localización de la proteína quimérica dentro del núcleo. En realizaciones preferentes, se incorpora un dominio de localización nuclear en la proteína de fusión final, ya que las funciones últimas de las proteínas de fusión descritas en la presente memoria típicamente requerirán que las proteínas estén localizadas en el núcleo. Sin embargo, podría no resultar necesario añadir un dominio de localización nuclear separado en los casos en que el dominio DBD mismo, u otro dominio funcional dentro de la proteína quimérica final, presente una función intrínseca de traslocación nuclear.

Utilización de las proteínas de fusión en la terapia génica

Las proteínas de fusión descritas en la presente memoria pueden utilizarse para regular la expresión génica o alterar la secuencia génica en aplicaciones de terapia génica en la misma. Ver, por ejemplo, la patente US. 6.511.808, la patente US 6.013.453, la patente US 6.007.988, la patente US n.º 6.503.717, solicitud de patente US n.º 2002/0164575 y solicitud de patente US n.º 2002/0160940. Entre los métodos puede incluirse la administración de una o más de las proteínas de fusión descritas en la presente memoria con diana en uno o más genes. Debido a que la metilación de muchas CpGs en cientos de pares de bases de ADN en promotores o regiones impresas puede influir en la expresión génica (Han et al., Mol Cell Biol. 2008 Feb;28(3):1124-35, Mabaera, et al., Blood. 2007, 15 de agosto;110(4):1343-52), puede resultar deseable reducir la metilación en secuencias más largas. En el caso de que se desee la desmetilación de una región más grande de un gen, puede administrarse una pluralidad de proteínas de fusión la totalidad de las cuales tiene como diana posiciones en el mismo gen, p. ej., posiciones a aproximadamente 1000, 500, 300, 250, 150, 100, 50, 40, 30 o 20 pb del sitio que debe desmetilarse. Alternativa o adicionalmente, puede administrarse una o una pluralidad de proteínas de fusión que se multimerizan tal como se indica en la presente memoria (ver, p. ej., las figs. 6A-D).

Pueden utilizarse métodos de transferencia génica víricos y no víricos convencionales para introducir ácidos nucleicos codificantes de la proteína de fusión en células de mamífero o tejidos diana. Dichos métodos pueden utilizarse para administrar ácidos nucleicos codificantes de proteínas de fusión en células *in vitro*. Preferentemente, los ácidos nucleicos codificantes de las proteínas de fusión se administran para los usos de terapia génica *in vivo* o *ex vivo*. Entre los sistemas de administración de vector no vírico se incluyen plásmidos de ADN, ácidos nucleicos desnudos y ácidos nucleicos acomplexados con un vehículo de administración, tal como un liposoma. Entre los sistemas de administración de vector vírico se incluyen virus ADN y ARN, que presentan genomas episómicos o integrados tras la administración en la célula. Para una revisión de los procedimientos de terapia génica, ver Anderson, 1992, Science, 256:808-813; Nabel y Felgner, 1993, TIBTECH, 11:211-217; Mitani y Caskey, 1993, TIBTECH, 11:162-166; Dillon, 1993, TIBTECH, 11:167-175; Miller, 1992, Nature, 357:455-460; Van Brunt, 1988, Biotechnology, 6:1149-54; Vigne, 1995, Restorat. Neurol. Neurosci., 8:35-36; Kremer y Perricaudet, 1995, Br. Med. Bull., 51:31-44; Haddada et al., en: Current Topics in Microbiology and Immunology Doerfler and Bohm (eds.) (1995) y Yu et al., 1994, Gene Ther., 1:13-26.

Entre los métodos de administración no vírica de ácidos nucleicos codificantes de las proteínas de fusión se incluyen lipofección, microinyección, biolísticos, virosomas, liposomas, inmunoliposomas, policonjugados o conjugados de lípidos:ácidos nucleicos, ADN o ARN desnudo, viriones artificiales e incorporación potenciada por un agente del ADN o ARN. La lipofección se describe en, p. ej., la patente US n.º 5.049.386, n.º 4.946.787 y n.º 4.897.355) y los reactivos de lipofección están comercialmente disponibles (p. ej., Transfectam™ y Lipofectin™). Entre los lípidos catiónicos y neutros que resultan adecuados para la lipofección de reconocimiento de receptores eficiente de polinucleótidos se incluyen los de Felgner, documentos n.º WO 91/17424 y n.º WO 91/16024. La administración puede ser en células (administración *ex vivo*) o en tejidos diana (administración *in vivo*).

La preparación de complejos de lípidos:ácidos nucleicos, incluyendo liposomas dirigidos, tales como complejos de lípidos inmunitarios, es bien conocida por el experto en la materia (ver, p. ej., Crystal, 1995, Science, 270:404-410; Blaese et al., 1995, Cancer Gene Ther., 2:291-297; Behr et al., 1994, Bioconjugate Chem. 5:382-389; Remy et al., 1994, Bioconjugate Chem., 5:647-654; Gao et al., Gene Ther., 2:710-722; Ahmad et al., 1992, Cancer Res.,

52:4817-20; patentes US n.º 4.186.183, n.º 4.217.344, n.º 4.235.871, n.º 4.261.975, n.º 4.485.054, n.º 4.501.728, n.º 4.774.085, n.º 4.837.028 y n.º 4.946.787).

La utilización de sistemas basados en virus ARN o ADN para la administración de ácidos nucleicos codificantes de las proteínas de fusión aprovecha los procedimientos altamente desarrollados para dirigir un virus a células específicas en el cuerpo y el tráfico de la carga vírica hasta el núcleo. Los vectores víricos pueden administrarse directamente en los pacientes (*in vivo*) o pueden utilizarse para tratar las células *in vitro* y las células modificadas se administran en los pacientes (*ex vivo*). Entre los sistemas basados en virus convencionales para la administración de proteínas de fusión podrían incluirse vectores de retrovirus, lentivirus, adenovirus, virus adenoasociados, virus Sendai y virus herpes simplex para la transferencia génica. Los vectores víricos son actualmente el método más eficiente y versátil de transferencia génica en las células y tejidos diana. La integración en el genoma huésped es posible con los métodos de transferencia génica con retrovirus, lentivirus y virus adenoasociados, que con frecuencia resulta en la expresión a largo plazo del transgén insertado. Además, se han observado elevadas eficiencias de transducción en muchos tipos celulares y tejidos diana diferentes.

El tropismo de un retrovirus puede alterarse mediante la incorporación de proteínas de cubierta foráneas, expandiendo la población diana potencial de células diana. Los vectores lentivíricos son vectores retrovíricos que son capaces de transducir o infectar células que no están en división y producir habitualmente elevados títulos víricos. La selección de un sistema de transferencia génica retrovírico, por lo tanto, dependerá del tejido diana. Los vectores retrovíricos comprenden repeticiones terminales largas de acción en cis con capacidad de empaquetamiento de hasta 6-10 kb de secuencia foránea. Las repeticiones terminales largas (RTL) de acción en cis resultan suficientes para replicar y empaquetar los vectores, que seguidamente se utilizan para integrar el gen terapéutico en la célula diana para proporcionar una expresión permanente del transgén. Entre los vectores retrovíricos utilizados ampliamente se incluyen los basados en el virus de la leucemia murina (VLMu), el virus de la leucemia del simio gibón (VLSG), el virus de la inmunodeficiencia del simio (VIS), el virus de la inmunodeficiencia humana (VIH) y combinaciones de los mismos (ver, p. ej., Buchscher et al., 1992, J. Virol., 66:2731-39; Johann et al., 1992, J. Virol., 66:1635-40; Sommerfelt et al., 1990, Virology, 176:58-59; Wilson et al., 1989, J. Virol., 63:2374-78; Miller et al., 1991, J. Virol., 65:2220-24; documento n.º WO 94/26877).

En aplicaciones en las que resulta preferente la expresión transitoria de la proteína de fusión, pueden utilizarse sistemas basados en adenovirus. Los vectores basados en adenovirus son capaces de eficiencias de transducción muy altas en muchos tipos celulares y no requieren la división celular. Con dichos vectores, se han obtenido títulos y niveles elevados de expresión. Dicho vector puede producirse en grandes cantidades en un sistema relativamente simple. Los vectores virus adenoasociados («VAA») también se utilizan para transducir células con ácidos nucleicos diana, p. ej., en la producción *in vitro* de ácidos nucleicos y péptidos, y para los procedimientos de terapia génica *in vivo* y *ex vivo* (ver, p. ej., West et al., 1987, Virology 160:38-47; patente US n.º 4.797.368, documento n.º WO 93/24641; Kotin, 1994, Hum. Gene Ther., 5:793-801; Muzyczka, 1994, J. Clin. Invest., 94:1351). La construcción de vectores VAA recombinante se describe en varias publicaciones, incluyendo la patente US n.º 5.173.414; Tratschin et al., 1985, Mol. Cell. Biol. 5:3251-60; Tratschin et al., 1984, Mol. Cell. Biol., 4:2072-81; Hermonat y Muzyczka, 1984, Proc. Natl. Acad. Sci. USA, 81:6466-70 y Samulski et al., 1989, J. Virol., 63:3822-28.

En particular, actualmente se encuentran disponibles por lo menos seis enfoques de vector vírico para la transferencia génica en ensayos clínicos, donde los vectores retrovíricos son claramente el sistema utilizado más frecuentemente. La totalidad de dichos vectores víricos utiliza enfoques que implican la complementación de vectores defectuosos por genes insertados en líneas celulares ayudante para generar el agente transductor.

pLASN y MFG-S son ejemplos de vectores retrovíricos que se han utilizado en ensayos clínicos (Dunbar et al., 1995, Blood, 85:3048; Kohn et al., 1995, Nat. Med., 1:1017; Malech et al., 1997, Proc. Natl. Acad. Sci. USA, 94:12133-38). PA317/pLASN fue el primer vector terapéutico utilizando en un ensayo de terapia génica. (Blaese et al., 1995, Science, 270:475-480). Se han observado eficiencias de transducción de 50 % o superiores para los vectores empaquetados MFG-S (Ellem et al., 1997, Immunol Immunother., 44:10-20; Dranoff et al., 1997, Hum. Gene Ther., 1:111-112).

Los vectores de virus adenoasociado recombinante (VAAr) son un sistema alternativo de administración génica prometedor que se basa en el virus parvovirus adenoasociado de tipo 2 defectuoso y no patogénico. Normalmente, los vectores se derivan de un plásmido que retiene solo las repeticiones terminales invertidas de 145 pb del VAA que flanquean el casete de expresión del transgén. La transferencia génica eficiente y la administración estable del transgén debidas a la integración en los genomas de la célula transducida son características clave para este sistema de vector (Wagner et al., 1998, Lancet, 351:1702-1703; Kearns et al., 1996, Gene Ther., 9:748-55).

Los vectores adenovíricos (Ad) recombinantes deficientes en replicación se utilizan predominantemente para la terapia génica del cáncer de colon debido a que pueden producirse a un título más alto e infectan fácilmente varios tipos celulares diferentes. La mayoría de vectores adenovirus se diseña de manera que un transgén sustituya los genes Ad E1a, E1b y E3; posteriormente, el vector defectuoso para la replicación se propaga en las células 293 humanas que suministran la función génica delecionada en trans. Los vectores Ad pueden transducir múltiples tipos de tejido *in vivo*, incluyendo células diferenciadas que no se dividen, tales como las que se encuentran en los tejidos de hígado, riñones y sistema muscular. Los vectores Ad convencionales presentan una gran capacidad de carga. Un ejemplo de la

utilización de un vector Ad en un ensayo clínico implicó el tratamiento con polinucleótidos para la inmunización antitumoral mediante inyección intramuscular (Sternan et al., 1998, Hum. Gene Ther. 7:1083-89). Entre los ejemplos adicionales de la utilización de vectores adenovíricos para la transferencia génica en ensayos clínicos se incluyen Rosenecker et al., 1996, Infection, 24:15-10; Sternan et al., 1998, Hum. Gene Ther., 9:7 1083-89; Welsh et al., 1995, Hum. Gene Ther., 2:205-218; Alvarez et al., 1997, Hum. Gene Ther. 5:597-613; Topf et al., 1998, Gene Ther., 5:507-513; Sternan et al., 1998, Hum. Gene Ther., 7:1083-89.

Se utilizan células empaquetadoras par formar partículas víricas que son capaces de infectar una célula huésped. Entre dichas células se incluyen las células 293, que empaquetan adenovirus, y las células Ψ 2 y las células PA317, que empaquetan retrovirus. Los vectores víricos utilizados en la terapia génica habitualmente son generadas por la línea celular productora que empaqueta un vector de ácido nucleico en una partícula vírica. Los vectores habitualmente contienen las secuencias víricas mínimas requeridas para empaquetar y posteriormente integrarse en un huésped, en donde otras secuencias víricas son sustituidas por un casete de expresión para la proteína que debe expresarse. Las funciones víricas faltantes son suministradas en trans por la línea celular empaquetadora. Por ejemplo, los vectores VAA utilizados en terapia génica normalmente solo poseen secuencias de repeticiones terminales invertidas (RTI) procedentes del genoma del VAA que resultan necesarias para el empaquetamiento e integración en el genoma del huésped. El ADN vírico se empaqueta en una línea celular que contiene un plásmido ayudante codificante de los demás genes del VAA, es decir, *rep* y *cap*, pero que carece de las secuencias de RTI. La línea celular también está infectada por adenovirus como ayudante. El virus ayudante promueve la replicación del vector VAA y la expresión de los genes del VAA a partir del plásmido ayudante. El plásmido ayudante no se empaqueta en cantidades significativas debido a la falta de secuencias de RTI. La contaminación por adenovirus puede reducirse mediante, p. ej., tratamiento térmico, al que los adenovirus son más sensibles que los VAA.

En muchas aplicaciones de terapia génica, resulta deseable que el vector de terapia génica se administre con un alto grado de especificidad en un tipo de tejido particular. Un vector vírico normalmente se modifica para que presente especificidad para un tipo celular dado mediante la expresión de un ligando en forma de una proteína de fusión con una proteína de cubierta vírica sobre la superficie externa de los virus. El ligando se selecciona para que presente afinidad para un receptor que es conocido que está presente sobre el tipo celular de interés. Por ejemplo, Han et al., 1995, Proc. Natl. Acad. Sci. USA, 92:9747-51, informaron de que el virus de la leucemia murina de Moloney puede modificarse para expresar heregulina humana fusionada con gp70, y el virus recombinante infecta determinadas células de cáncer de mama humano que expresan el receptor del factor de crecimiento epidérmico humano. Este principio puede extenderse a otros pares de virus que expresan una proteína de fusión de ligando y células diana que expresan un receptor. Por ejemplo, el fago filamentososo puede diseñarse para exponer fragmentos de anticuerpos (p. ej., Fab o Fv) que presentan una afinidad de unión específica para prácticamente cualquier receptor celular seleccionado. Aunque la descripción anterior se aplica principalmente a los vectores víricos, pueden aplicarse los mismos principios a los vectores no víricos. Dichos vectores pueden diseñarse para que contengan secuencias de incorporación específicas que se cree favorecen la incorporación por células diana específicas.

Los vectores de terapia génica pueden administrarse *in vivo* mediante la administración en el paciente individual, normalmente mediante administración sistémica (p. ej., infusión intravenosa, intraperitoneal, intramuscular, subdérmica o intracraneal) o aplicación tópica, tal como se indica posteriormente. Alternativamente, pueden administrarse vectores en las células *ex vivo*, tal como células explantadas de un paciente individual (p. ej., linfocitos, aspirados de médula ósea, biopsia de tejido) o células madre (p. ej., células madre hematopoyéticas de donante universal, células madre embrionarias (ME), células madre parcialmente diferenciadas, células madre no pluripotentes, células madre pluripotentes, células madre pluripotentes inducidas (células MPi) (ver, p. ej., Sipione et al., Diabetologia, 47:499-508, 2004)), seguido de la reimplantación de las células en el paciente, normalmente después de la selección para las células en las que se ha incorporado el vector.

La transfección celular *ex vivo* con fines diagnósticos, de investigación o para la terapia génica (p. ej., mediante reinfusión de las células transfectadas en el organismo huésped) es bien conocida por el experto en la materia. En una realización preferente, se aíslan las células a partir del organismo sujeto, se transfectan con ácido nucleico (un gen o ADNc) codificante de la proteína de fusión y se infunden de vuelta nuevamente en el organismo sujeto (p. ej., un paciente). El experto en la materia conoce diversos tipos celulares adecuados para la transfección *ex vivo* (ver, p. ej., Freshney et al., Culture of Animal Cells, A Manual of Basic Technique (5a ed. 2005)) y las referencias citadas en la misma para un análisis sobre cómo aislar y cultivar células de pacientes).

En una realización, las células madre (p. ej., células madre hematopoyéticas de donante universal, células madre embrionarias (ME), células madre parcialmente diferenciadas, células madre no pluripotentes, células madre pluripotentes, células madre pluripotentes inducidas (células MPi) (ver, p. ej., Sipione et al., Diabetologia, 47:499-508, 2004)) se utilizan en procedimientos *ex vivo* para la transfección celular y la terapia génica. La ventaja de utilizar células madre es que pueden diferenciarse en los demás tipos celulares *in vitro*, o pueden introducirse en un mamífero (tal como el donante de las células), donde se injertarán en la médula ósea. Se conocen métodos para diferenciar las células CD34+ *in vitro* en tipos celulares inmunitarios clínicamente importantes utilizando citoquinas, tales como GM-CSF, IFN-gamma y TNF-alfa (ver Inaba et al., 1992, J. Exp. Med., 176:1693-1702).

Las células madre pueden aislarse para la transducción y diferenciación utilizando métodos conocidos. Por ejemplo,

pueden aislarse células madre a partir de células de médula ósea mediante enriquecimiento cíclico («panning», en inglés) de las células de médula ósea con anticuerpos que se unen a células no deseadas, tales como células T CD4⁺ y CD8⁺, células CD45⁺ (células panB), GR-1 (granulocitos) y 1ad (células presentadoras de antígenos diferenciadas) (ver Inaba et al., 1992, J. Exp. Med., 176:1693-1702).

Los vectores (p. ej., retrovirus, adenovirus, liposomas, etc.) que contienen ácidos nucleicos codificantes de la proteína de fusión también pueden administrarse directamente en el organismo para la transducción de las células *in vivo*. Alternativamente, puede administrarse ADN o ARN desnudo. El ADN o ARN puede modificarse para mejorar sus propiedades, p. ej., para incrementar la semivida o para reducir la respuesta inmunitaria (ver, p. ej., los documentos n.º US20130123481, n.º 20130115272 y n.º 20120251618). La administración se lleva a cabo mediante cualquiera de las vías normalmente utilizadas para introducir una molécula en contacto íntimo con células sanguíneas o de tejidos. Se encuentran disponibles métodos adecuados de administración de dichos ácidos nucleicos y son bien conocidos por el experto en la materia y, aunque puede utilizarse más de una vía para administrar una composición particular, una vía en particular con frecuencia puede proporcionar una reacción más inmediata y más eficaz que otras vías. Alternativamente, también pueden administrarse formulaciones estables de la proteína de fusión.

Los portadores farmacéuticamente aceptables están determinados en parte por la composición particular que se administra, así como por el método particular utilizado para administrar la composición. De acuerdo con lo anterior, existe una amplia variedad de formulaciones adecuadas de composiciones farmacéuticas disponibles, tal como se indica posteriormente (ver, p. ej., Remington: The Science and Practice of Pharmacy, 21a ed., 2005).

Vehículos de administración

Un factor importante en la administración de compuestos polipeptídicos, tales como las proteínas de fusión de la presente invención, es garantizar que el polipéptido presenta la capacidad de atravesar la membrana plasmática de la célula, o la membrana de un compartimiento intracelular, tal como el núcleo. Las membranas celulares están compuestas de bicapas de lípidos-proteínas que son libremente permeables para compuestos lipofílicos no iónicos pequeños y son inherentemente impermeables a los compuestos polares, macromoléculas y agentes terapéuticos o diagnósticos. Sin embargo, se han descrito proteínas y otros compuestos, tales como los liposomas, que presentan la capacidad de traslocar polipéptidos, tales como proteínas de fusión, a través de una membrana celular.

Por ejemplo, los «polipéptidos de traslocación membranar» presentan subsecuencias de aminoácidos anfífilas o hidrofóbicas que poseen la capacidad de actuar como portadores traslocadores de membrana. En una realización, las proteínas de homeodominio poseen la capacidad de traslocar a través de las membranas celulares. El péptido internalizable más corto de una proteína de homeodominio llamada Antennapedia se ha encontrado que es la tercera hélice de la proteína, entre las posiciones aminoácidas 43 y 58 (ver, p. ej., Prochiantz, 1996, Curr. Opin. Neurobiol., 6:629-634). Otra subsecuencia, el dominio h (hidrofóbico) de péptidos de señal, se ha encontrado que presentan características similares de traslocación de la membrana celular (ver, p. ej., Lin et al., 1995, J. Biol. Chem., 270:14255-58).

Entre los ejemplos de secuencias peptídicas que pueden unirse a una proteína para facilitar la incorporación de la proteína en las células se incluyen, aunque sin limitarse a ellos, fragmentos peptídicos de la proteína tat del VIH (Endoh et al., 2010, Methods Mol. Biol., 623:271-281; Schmidt et al., 2010, FEBS Lett., 584:1806-13; Futaki, 2006, Biopolymers, 84:241-249); una secuencia peptídica de 20 residuos que corresponde a los aminoácidos 84 a 103 de la proteína p16 (ver Fahraeus et al., 1996, Curr. Biol., 6:84); la tercera hélice del homeodominio de 60 aminoácidos de longitud de Antennapedia (Derossi et al., 1994, J. Biol. Chem., 269:10444); la región h de un péptido de señal, tal como la región h del factor de crecimiento fibroblástico de Kaposi (K-FGF, por sus siglas en inglés) (Lin et al., supra); el dominio de traslocación VP22 del VHS (Elliot & O'Hare, 1997, Cell, 88:223-233); o proteínas supercargadas o intrafilinas, p. ej., tal como se describen en los documentos n.º US20120100569 y n.º US20110112040; y en Cronican et al (2011) Chem Biol. 18, 833; Cronican et al (2010) ACS Chem. Biol. 5, 747; McNaughton et al (2009) Proc. Natl. Acad. Sci. USA 106, 6111 y Lawrence et al. (2007) J. Am. Chem. Soc. 129, 10110. Ver también, p. ej., Caron et al., 2001, Mol Ther., 3:310-318; Langel, Cell-Penetrating Peptides: Processes and Applications (CRC Press, Boca Raton FL 2002); El-Andaloussi et al., 2005, Curr. Pharm. Des., 11:3597-3611 y Deshayes et al., 2005, Cell. Mol. Life Sci., 62:1839-49. Otras fracciones químicas adecuadas que proporcionan una incorporación celular potenciada también pueden unirse químicamente a las proteínas de fusión descritas en la presente memoria.

Las moléculas de toxinas también poseen la capacidad de transportar polipéptidos a través de las membranas celulares. Con frecuencia, dichas moléculas están compuestas de por lo menos dos partes (denominadas «toxinas binarias»): un dominio o polipéptido de translocación o unión y un dominio o polipéptido toxina separado. Normalmente, el dominio o polipéptido toxina se une a un receptor celular y después la toxina es transportada hasta el interior de la célula. Se utilizaron varias toxinas bacterianas, incluyendo la toxina iota de *Clostridium perfringens*, toxina diftérica (TD), exotoxina A de *Pseudomonas* (PE), toxina pertussis (TP), toxina de *Bacillus anthracis* y adenilato ciclase de pertussis (CYA), intentando administrar péptidos en el citosol celular como fusiones internas o aminoterminales (Arora et al., 1993, J. Biol. Chem., 268:3334-41; Perelle et al., 1993, Infect. Immun., 61:5147-56; Stenmark et al., 1991, J. Cell Biol., 113:1025-32; Donnelly et al., 1993, Proc. Natl. Acad. Sci. USA, 90:3530-34; Carbonetti et al., 1995, Abstr.

Annu. Meet. Am. Soc. Microbiol. 95:295; Sebo et al., 1995, Infect. Immun., 63:3851-57; Klimpel et al., 1992, Proc. Natl. Acad. Sci. USA, 89:10277-81 y Novak et al., 1992, J. Biol. Chem., 267:17186-93).

Dichas subsecuencias pueden utilizarse para traslocar proteínas de fusión a través de la membrana celular. Las proteínas de fusión pueden fusionarse convenientemente o derivatizarse con dichas secuencias. Habitualmente, la secuencia de traslocación se proporciona como parte de una proteína de fusión. Opcionalmente, puede utilizarse un conector para unir la proteína de fusión y la secuencia de traslocación. Puede utilizarse cualquier conector adecuado, p. ej., un conector peptídico.

La proteína de fusión también puede introducirse en una célula animal, preferentemente una célula de mamífero, mediante liposomas y derivados de liposomas, tales como inmunoliposomas. El término «liposoma» se refiere a vesículas comprendidas de una o más bicapas lipídicas ordenadas concéntricamente, que encapsulan una fase acuosa. La fase acuosa normalmente contiene el compuesto que debe administrarse en la célula, es decir, la proteína de fusión.

El liposoma se fusiona con la membrana plasmática, liberando de esta manera el compuesto en el citosol. Alternativamente, el liposoma resulta fagocitado o incorporado por la célula en una vesícula de transporte. Una vez en el endosoma o fagosoma, el liposoma se degrada o se fusiona con la membrana del vesícula de transporte y libera su contenido.

En los métodos actuales de administración de fármaco mediante liposomas, el liposoma finalmente se vuelve permeable y libera el compuesto encapsulado (p. ej., la proteína de fusión o el ácido nucleico codificante de la misma) en el tejido o célula diana. Para la administración sistémica o específica de tejido, lo anterior puede conseguirse, por ejemplo, de una manera pasiva, en donde la bicapa del liposoma se degrada durante el tiempo mediante la acción de diversos agentes en el organismo. Alternativamente, la liberación del compuesto activo implica utilizar un agente para inducir un cambio de permeabilidad en la vesícula liposómica. Las membranas de los liposomas pueden construirse de manera que se desestabilicen cuando el medio se vuelve ácido en proximidad a la membrana liposómica (ver, p. ej., Proc. Natl. Acad. Sci. USA, 84:7851 (1987); Biochemistry, 28:908 (1989)). Cuando los liposomas son endocitados por una célula diana, por ejemplo, se desestabilizan y liberan su contenido. Dicha desestabilización se denomina fusogénesis. La dioleoilfosfatidiletanolamina (DOPE, por sus siglas en inglés) es la base de muchos sistemas «fusogénicos».

Dichos liposomas habitualmente comprenden la proteína de fusión y un componente lipídico, p. ej., un lípido neutro y/o catiónico, que opcionalmente incluye una molécula de reconocimiento de receptor, tal como un anticuerpo que se une a un receptor o ligando de superficie celular predeterminado (p. ej., un antígeno). Se encuentra disponible una variedad de métodos para la preparación de liposomas tal como se describen en, p. ej., Szoka et al., 1980, Annu. Rev. Biophys. Bioeng., 9:467, patentes US n.º 4.186.183, n.º 4.217.344, n.º 4.235.871, n.º 4.261.975, n.º 4.485.054, n.º 4.501.728, n.º 4.774.085, n.º 4.837.028, n.º 4.235.871, n.º 4.261.975, n.º 4.485.054, n.º 4.501.728, n.º 4.774.085, n.º 4.837.028, n.º 4.946.787, publicación de patente PCT n.º WO 91/17424, Deamer y Bangham, 1976, Biochim. Biophys. Acta, 443:629-634; Fraley, et al., 1979, Proc. Natl. Acad. Sci. USA, 76:3348-52; Hope et al., 1985, Biochim. Biophys. Acta, 812:55-65; Mayer et al., 1986, Biochim. Biophys. Acta, 858:161-168; Williams et al., 1988, Proc. Natl. Acad. Sci. USA, 85:242-246; Liposomes (Ostro (ed.), 1983, capítulo 1); Hope et al., 1986, Chem. Phys. Lip., 40:89; Gregoriadis, Liposome Technology (1984) y Lasic, Liposomes: from Physics to Applications (1993)). Entre dichos métodos se incluyen, por ejemplo, sonicación, extrusión, homogenización a alta presión, microfluidificación, diálisis con detergente, fusión inducida con calcio de vesículas liposómicas pequeñas y métodos de fusión con éter, la totalidad de los cuales son bien conocidos de la técnica.

En determinadas realizaciones, resulta deseable el direccionamiento de liposomas utilizando fracciones de direccionamiento que son específicas de un tipo celular, de tejido y similar particular. El direccionamiento de liposomas utilizando una diversidad de fracciones de direccionamiento (p. ej., ligandos, receptores y anticuerpos monoclonales) ha sido descrito anteriormente (ver, p. ej., la patente US n.º 4.957.773 y n.º 4.603.044).

Entre los ejemplos de fracciones de direccionamiento se incluyen anticuerpos monoclonales específicos de antígenos asociados a neoplasmas, tales como antígeno específico de cáncer prostático y MAGE. Los tumores también pueden diagnosticarse mediante la detección de productos génicos que resultan de la activación o sobreexpresión de oncogenes, tales como ras o c-erbB2. Además, muchos tumores expresan antígenos normalmente expresados por el tejido fetal, tales como la alfafetoproteína (AFP) y el antígeno carcinoembrionario (CEA, por sus siglas en inglés). Los sitios de infección vírica pueden diagnosticarse utilizando diversos antígenos víricos, tales como antígenos nucleares y de superficie de la hepatitis B (VHBc, VHBs), antígenos de la hepatitis C, antígenos del virus de Epstein-Barr, antígenos del virus de la inmunodeficiencia humana de tipo 1 (VIH1) y del virus del papiloma. Puede detectarse la inflamación utilizando moléculas reconocidas específicamente por moléculas de superficie que se expresan en sitios de inflamación, tales como integrinas (p. ej., VCAM-1), receptores de selectina (p. ej., ELAM-1) y similares.

Pueden utilizarse métodos estándares para el acoplamiento de agentes de direccionamiento con liposomas. Estos métodos generalmente implican la incorporación en componentes lipídicos de los liposomas, p. ej., fosfatidiletanolamina, que pueden activarse para la unión de agentes de direccionamiento, o compuestos lipofílicos

derivatizados, tales como la bleomicina derivatizada con lípidos. Pueden construirse liposomas con diana en anticuerpos, utilizando, por ejemplo, liposomas que incorporan proteína A (ver Renneisen et al., 1990, J. Biol. Chem., 265:16337-42 y Leonetti et al., 1990, Proc. Natl. Acad. Sci. USA, 87:2448-51).

Dosis

Para aplicaciones terapéuticas, la dosis de la proteína de fusión que debe administrarse en el paciente puede calcularse de una manera similar a la descrita para las proteínas de dedo de zinc; ver, por ejemplo, la patente US 6.511.808, la patente US 6.492.117, la patente US n.º 6.453.242, solicitud de patente US n.º 2002/0164575 y solicitud de patente US n.º 2002/0160940. En el contexto de la presente exposición, la dosis debería ser suficiente para proporcionar un efecto a una respuesta terapéutica beneficiosa en el paciente durante el tiempo. Además, unos regímenes de dosis particulares pueden resultar útiles para determinar los cambios fenotípicos en un contexto experimental, p. ej., en estudios de genómica funcional, y en modelos celulares o animales. La dosis será determinada por la eficacia, especificidad y K_D de la proteína de fusión particular utilizada, el volumen nuclear de la célula diana, y la afección del paciente, así como el peso corporal o superficie del paciente que debe tratarse. El tamaño de la dosis también estará determinada a partir de la existencia, naturaleza y extensión de cualesquiera efectos secundarios adversos que acompañan a la administración de un compuesto o vector particular en un paciente particular.

Composiciones farmacéuticas y administración

Las composiciones farmacéuticas apropiadas para la administración de las proteínas de fusión de la presente invención pueden determinarse tal como se ha descrito para las proteínas de dedo de zinc; ver, por ejemplo, la patente US 6.511.808, la patente US 6.492.117, la patente US n.º 6.453.242, solicitud de patente US n.º 2002/0164575 y solicitud de patente US n.º 2002/0160940. Las proteínas de fusión y los vectores de expresión codificantes de proteínas de fusión pueden administrarse directamente en el paciente para la modulación de los patrones de metilación, p. ej., y la expresión génica, y para aplicaciones terapéuticas o profilácticas, por ejemplo, para el tratamiento de las enfermedades enumeradas en la fig. 7 como asociadas a hipermetilación, incluyendo el cáncer (p. ej., vejiga, cerebro (p. ej., glioma o glioblastoma), cáncer de mama, cuello uterino, colon, colorrectal, esófago, cabeza/cuello, riñón, leucemia, hígado, pulmón, linfoma, mieloma, ovario, páncreas, próstata, rhabdomyosarcoma y cáncer de útero); esquizofrenia, formación de memoria y aterosclerosis. De esta manera, los métodos pueden incluir la identificación de un sujeto que presenta una enfermedad asociada a la hipermetilación (p. ej., que incluye opcionalmente la obtención de una muestra y la detección de la metilación de ADN, p. ej., del gen asociado que se indica en la fig. 7 y la selección del sujeto en caso de que el ADN esté hipermetilado) y administrar una cantidad terapéuticamente eficaz de una proteína de fusión, o un ácido nucleico codificante de una proteína de fusión, tal como se indica en la presente memoria, en el sujeto.

La administración de cantidades terapéuticamente eficaces se lleva a cabo mediante cualquiera de las vías normalmente utilizadas para introducir proteínas de fusión en contacto íntimo con el tejido que va a tratarse. Las proteínas de fusión se administran de cualquier manera adecuada, preferentemente con portadores farmacéuticamente aceptables. Se encuentran disponibles métodos adecuados de administración de dichos moduladores y son bien conocidos por el experto en la materia y, aunque puede utilizarse más de una vía para administrar una composición particular, una vía en particular con frecuencia puede proporcionar una reacción más inmediata y más eficaz que otras vías.

Los portadores farmacéuticamente aceptables están determinados en parte por la composición particular que se administra, así como por el método particular utilizado para administrar la composición. De acuerdo con lo anterior, existe una amplia diversidad de formulaciones adecuadas de composiciones farmacéuticas: que se encuentran disponibles (ver, p. ej., Remington: The Science and Practice of Pharmacy, 21a ed., 2005).

Las proteínas de fusión, juntas o en combinación con otros componentes adecuados, pueden prepararse en formulaciones de aerosol (es decir, pueden «nebulizarse») para la administración mediante inhalación. Las formulaciones de aerosol pueden añadirse a propelentes aceptables presurizados, tales como diclorodifluorometano, propano, nitrógeno y similares.

Entre las formulaciones adecuadas para la administración parenteral, tal como, por ejemplo, por vía intravenosa, intramuscular, intradérmica y subcutánea, se incluyen las soluciones para inyección estériles isotónicas, acuosas y no acuosas, que pueden contener antioxidantes, tampones, bacteriostatos y solutos que hacen que la formulación sea isotónica con la sangre del receptor pretendido, y suspensiones estériles acuosas y no acuosas que pueden incluir agentes de suspensión, solubilizadores, agentes espesantes, estabilizantes y conservantes. Las composiciones dadas a conocer pueden administrarse, por ejemplo, mediante infusión intravenosa, o por vía oral, tópica, intraperitoneal, intravesical o intratecal. Las formulaciones de compuestos pueden presentarse en recipientes sellados unidos o multidosis, tales como ampollas y viales. Pueden prepararse soluciones y suspensiones para inyección a partir de polvos, gránulos y tabletas estériles del tipo anteriormente indicado.

Ejemplos

La invención se describe en mayor detalle en los ejemplos siguientes, que no son limitativos del alcance de la invención descrita en las reivindicaciones. Cualesquiera ejemplos no comprendidos dentro del alcance según las reivindicaciones se proporciona con fines meramente comparativos.

Ejemplo 1. Generación de proteínas de fusión TAL-TET1 para loci genómicos específicos en el gen KLF4 humano endógeno.

Se construyeron plásmidos que expresaban tres versiones diferentes de la proteína TET1 humana, basándose en la arquitectura de domino predicha que se ha sido descrita por Tahiliani et al. (Science 324, 930-935 (2009)) (figura 1A).

La secuencia codificante completa de TET1 se sintetizó como gBlocks (Integrated DNA Technologies) y se ensambló mediante digestión y ligación estándar con enzimas de restricción para construir los vectores de expresión de TET1LC y TET1DC. Todos los TALE se ensamblaron utilizando el método FLASH y se clonaron en los vectores de expresión de TALE-TET1 (pJA344C7, pJA345D4, pJA344E9 y pJA247) que contienen una a cuatro terminaciones TALE 0,5 C-terminales, una señal de localización nuclear N-terminal, el dominio N-terminal TALE $\Delta 152$ y el dominio C-terminal de TALE +95 tal como se ha descrito anteriormente (Miller et al., Nat Biotechnol 29, 143-148 (2010); Reyon et al., Nature Biotechnology 30:460-465 (2012)). En todos los plásmidos de expresión utilizados, se fusionó TET1LC o TET1DC con el extremo C-terminal del dominio de unión a ADN derivado de TALE mediante un conector GlyGlyGlySer (SEC ID n.º 18) y la expresión de dicha fusión estaba controlada por un promotor EF1alfa. Para los experimentos de transferencia western, los vectores de expresión presentaban una etiqueta triple-flag adicional clonada cadena arriba de la señal de localización nuclear.

TET1LC es la proteína TET1 de longitud completa, mientras que TET1DC comprende solo el dominio catalítico de la proteína, incluyendo la región rica en Cys y el dominio de hélice β de doble cadena (figura 5). TET1NZ es una variante de truncado intermedia que carece de un dominio de unión a zinc de tipo CXXC, que podría ser responsable de dirigir la unión al ADN de la proteína endógena (Zhang et al., Cell Res 20, 1390-1393 (2010); Xu et al., Mol. Cell 42, 451-464 (2011); figura 5). Observar que TET1NZ mantiene más de la proteína nativa que TET1DC por sí solo. En resumen, se sintetizaron fragmentos de ADN codificantes de TET1 de longitud completa (TET1LC, aminoácidos 1 a 2136) y TET1 dominio catalítico (TET1DC, aminoácidos 1418 a 2136) en 8 fragmentos (Integrated DNA Technologies) y se clonaron iterativamente juntos utilizando clonación estándar basada en digestión/ligación de restricción.

Estos fragmentos codificantes de TET1LC y TET1DC seguidamente se clonaron en un vector plasmídico que permitía su expresión como fusiones con el extremo carboxiterminal de cuatro dominios de proteínas de matriz de repetición de TAL diseñados para unirse a diferentes sitios diana de 18 pb dentro del segundo intrón del gen Klf4 humano (ver la figura 2). Las secuencias de conector utilizadas para unir la proteína TET1 o fragmentos de proteína TET1 a las matrices de repetición de efector TAL estaban compuestas de uno de los conectores siguientes: Gly₃Ser (SEC ID n.º 5), Gly₃SerGly₄SerArgSer (SEC ID n.º 6), Gly₃Ser(Gly₄Ser)₂ArgSer (SEC ID n.º 7), Gly₃Ser(Gly₄Ser)₃ArgSer (SEC ID n.º 8), Gly₃Ser(Gly₄Ser)₄ArgSer (SEC ID n.º 9), o Gly₃Ser(Gly₄Ser)₅ArgSer (SEC ID n.º 10). El alargamiento del conector entre la matriz de repetición de TALE y TET1DC no alteró apreciablemente las eficiencias de desmetilación observadas (figura 1C). Basándose en estos resultados, se construyeron todas las proteínas de fusión adicionales descritas en el presente informe, utilizando la arquitectura TALE-TET1C con un conector corto GGGS.

El plásmido de expresión final se introdujo en las proteínas de fusión bajo el control transcripcional de un promotor EF1 α constitutivamente activo. Se construyó TET1NZ mediante eliminación del dominio CXXC situado en los aminoácidos 584 a 624 de la proteína de longitud completa. Lo anterior se consiguió mediante la utilización de PCR con cola de un pequeño segmento de TET1LC que introdujo un sitio BamHI inmediatamente después del dominio CXXC e incluía un sitio PmlI cadena arriba del dominio CXXC que era único en el plásmido de expresión de TET1LC. Dicho producto de PCR seguidamente se clonó en el plásmido de expresión de TET1LC utilizando una doble digestión BamHI/PmlI tanto del producto de PCR como del plásmido de TET1LC, seguido de ligación y transformación.

Los cuatro loci genómicos seleccionados como dianas para la desmetilación (y para los que se diseñaron las matrices de repetición de efector TAL) se seleccionaron basándose en varios criterios. La totalidad de los cuatro loci dentro del intrón 1-2 del gen KLF4 humano, que muestran marcas de cromatina abierta basándose en los datos del proyecto ENCODE (con acceso mediante el buscador genómico de UCSC), sugiere su probable accesibilidad para la unión por dominios de unión al ADN del efector TAL. Dicho intrón también resultaba atractivo porque alojaba múltiples dirresiduos CpG que están metilados en las células K562 pero que no están metilados en las células madre embrionarias humanas, lo que sugiere que la metilación de estas citoquinas no es necesaria para la supervivencia o viabilidad celular.

Se diseñaron matrices de repetición de efector TAL utilizando el método de ensamblaje FLASH anteriormente descrito (Reyon et al., Nature Biotechnology 30:460-465(2012)). Se clonaron fragmentos de ADN codificantes de matrices ensambladas, en los constructos de expresión que contenían los aminoácidos 153 a 288 de la proteína TAL13 y el dominio C-terminal +95, que comprendía los aminoácidos 715 a 809 del extremo C-terminal de TALE13 (Miller et al., Nat. Biotechnol. 29, 143-148 (2010)).

Ejemplo 2. Desmetilación específica de sitio de CpG en el locus de KLF4 humano mediada por proteínas de fusión TAL-TET1.

A continuación, se sometieron a ensayo en células K562 humanas las capacidades de dichas proteínas de desmetilar CpG contiguos. Se cultivaron células K562 en medio RPMI 1640 (Gibco) con HI-FBS al 10 % (Gibco), Glutamatx al 1 % (Gibco) y Pen/Strep al 1 % (Gibco) y se transfectaron con plásmidos codificantes de diversas proteínas de fusión TAL-TET1DC que alojaban el conector Gly3Ser mediante nucleofección de 1×10^6 células en división con 10 μ g de ADN en 100 μ l de solución Kit V (Lonza) utilizando el programa T-016 en el instrumento Nucleofector 2b Device (Lonza). Las células nucleofectadas se incubaron en placas de 6 pocillos a 37 °C durante 4 días tras la nucleofección. Se recolectó el ADN genómico 4 días después de la transfección. La secuenciación del ADN con bisulfito de KLF4 intrón 1-2 de dichas poblaciones celulares transfectadas se llevó a cabo del modo siguiente. Se aisló el ADN genómico a partir de las células transfectadas utilizando el minikit para sangre de Qiagen. Se trataron con bisulfito 200 a 1000 ng de ADN genómico utilizando el kit de metilación de ADN EZ (Zymo), el kit de metilación del ADN «EZ DNA Methylation-Lightning» (Zymo) o el kit de conversión de células a CpG de bisulfito (Applied Biosystems) siguiendo los protocolos recomendados. La amplificación por PCR de Bis-ADN se llevó a cabo utilizando el sistema de ADN polimerasa Taq HiFi Accuprime (Invitrogen). Los productos de PCR se sometieron a clonación TA utilizando el protocolo original de clonación TA (Invitrogen) o el sistema de vector pGEM-T Easy (Promega) y se transformaron en células competentes Top10 o XL1-Blue. Los clones individuales se prepararon en minipreps y se secuenciaron, y las secuencias se analizaron utilizando el software BISM disponible en línea en (Rohde et al., BMC Bioinformatics 2010, 11:230), utilizando parámetros por defecto.

Los resultados revelaron una robusta desmetilación en los CpG situados en contigüidad al sitio al que se unían cada una de las proteínas de fusión TAL-TET1DC. La figura 3 muestra las tasas de desmetilación para los CpG en proximidad a los cuatro sitios diferentes de unión de TAL. La proteína TAL-TET1DC con diana en el sitio 1 mostró los niveles más altos de desmetilación, con 32,5 % y 45,4 % de desmetilación de los CpG situados 9 y 15 pb cadena abajo del sitio de unión a la diana, respectivamente. La proteína TAL-TET1DC con diana en el sitio 2 no mostró desmetilación de los CpG situados 1 y 3 pb cadena abajo del sitio de unión, pero sí mostró una desmetilación de 36,5 % y 28,2 % en los CpG situados 17 y 27 pb cadena abajo, respectivamente. La proteína TAL-TET1 con diana en el sitio 3 mostró varias desmetilaciones superiores al 25 % en los CpG situados 22, 36, 116 y 143 pb cadena abajo, y una proteína TAL-TET1DC con diana en el sitio 4 mostró una desmetilación de 37 % en un CpG situado 13 pb cadena abajo, aunque no consiguió mostrar ninguna desmetilación significativa en ningún otro CpG. En experimentos adicionales, la proteína de fusión KL1 TALE-TET1DC redujo la metilación de los CpG situados 10 y 16 pb desde el límite 3' del sitio de unión de TALE en 54 % y 66 %, respectivamente. Resulta interesante que esta fusión de K-1 también indujo la desmetilación de los CpG situados a ~150 pb del borde 3' de su sitio de unión, aunque en menor grado que los CpG más próximos.

Las proteínas de fusión de control portadoras de una matriz de repetición de TALE con diana en una secuencia de gen informador EGFP no relacionado no desmetilaron los CpG en el intrón KLF4, demostrando que la desmetilación requiere una unión específica al locus diana por las repeticiones de TALE y no se debe simplemente a la sobreexpresión de proteínas que alojan actividad de TET1 hidroximetilasa.

Globalmente lo anterior demuestra la capacidad de las proteínas de fusión TAL-TET1DC diseñadas que alojan un conector Gly3Ser de inducir una desmetilación de 30 % a 50 % de los CpG situados aproximadamente 10 a 30 pb cadena abajo del sitio de unión de TAL. (Observar que para la totalidad de estos casos, «cadena abajo» se refiere a bases situadas 3' respecto a la última base del sitio de unión contactado por la repetición de efector de TAL carboxiterminal en la proteína de fusión). Las proteínas de fusión TAL-TET1DC que alojan conectores más extendidos (Gly3SerGly4SerArgSer (SEC ID n.º 6), Gly3Ser(Gly4Ser)2ArgSer (SEQ ID n.º 7), Gly3Ser(Gly4Ser)3ArgSer (SEC ID n.º 8), Gly3Ser(Gly4Ser)4ArgSer (SEC ID n.º 9), o Gly3Ser(Gly4Ser)5ArgSer (SEC ID n.º 10) también mostraban niveles similares de desmetilación (figura 4A).

En experimentos de comparación, las fusiones de TALE portadoras del dominio TET1DC indujeron reducciones significativamente mayores de la metilación de los CpG próximos al sitio de unión de TALE que las portadoras de la proteína TET1 de longitud completa, a juzgar por la secuenciación profunda con bisulfito (figura 4B). Debido a que las fusiones de TALE-TET1 de longitud completa eran menos activas que las fusiones TALE-TET1DC, se realizaron experimentos adicionales centrados en las fusiones TALE-TET1DC.

Ejemplo 3. Proteínas TALE-TET1DC diseñadas con diana en el locus HBB humano.

A fin de extender y además generalizar el resultado de que las proteínas de fusión TALE-TET1DC pueden utilizarse para identificar las marcas de metilación de CpG que influyen sobre la expresión génica endógena, se diseñó una proteína de fusión TALE-TET1DC para desmetilar los CpG en un locus adicional: el promotor del gen de beta-globina humano (HBB, por sus siglas en inglés).

El trabajo publicado ha identificado un pequeño número de CpG (aproximadamente cuatro) en los promotores de β -globina y γ -globina humanos que se metilan diferentemente en células hepáticas fetales y células de médula ósea adulta, lo que sugiere que la metilación de estos CpG podrían gobernar la expresión de estos genes durante el

desarrollo (Mabaera et al., Blood. 2007, 15 de agosto;110(4):1343-52). Se diseñaron diez matrices de TALE con diana en secuencias próximas a dichos CpG diferencialmente metilados en el locus de β -globina humana: cuatro TALE con diana en un grupo de dos CpG ~100 pb cadena abajo del sitio de inicio de transcripción y seis TALE con diana en un grupo de dos CpG ~300 pb cadena arriba del sitio de inicio (figura 8A).

Se cultivaron células K562 en medio RPMI 1640 (Gibco) con HI-FBS al 10 % (Gibco), Glutamax al 1 % (Gibco) y Pen/Strep al 1 % (Gibco) y se transfectaron con plásmidos codificantes de diversas proteínas de fusión TALE-TET1DC que alojaban el conector Gly₃Ser mediante nucleofección de 1×10^6 células en división con 10 μ g de ADN en 100 μ l de solución Kit V (Lonza) utilizando el programa T-016 en el instrumento Nucleofector 2b Device (Lonza). Las células nucleofectadas se incubaron en placas de 6 pocillos a 37 °C durante 4 días tras la nucleofección. Se recolectó el ADN genómico 4 días después de la transfección. La secuenciación del ADN con bisulfito del locus HBB a partir de estas poblaciones celulares transfectadas se llevó a cabo del modo siguiente. Se aisló el ADN genómico a partir de las células transfectadas utilizando el minikit para sangre de Qiagen. Se trataron con bisulfito 200 a 1000 ng de ADN genómico utilizando el kit de metilación de ADN EZ (Zymo), el kit de metilación de ADN «EZ DNA Methylation-Lightning» (Zymo) o el kit de conversión de células a CpG de bisulfito (Applied Biosystems) siguiendo los protocolos recomendados. La amplificación por PCR de Bis-ADN se llevó a cabo utilizando el kit de PCR Pyromark (Qiagen). Se añadieron adaptadores y códigos de barra Illumina mediante PCR con el enzima de PCR Phusion High-Fidelity (NEB) y se secuenciaron los amplicones en un sistema Illumina MiSeq. Se aisló el ARN total a partir de las mismas células con el minikit de ARN PureLink (Ambion) siguiendo las instrucciones del fabricante. Se llevó a cabo la transcripción inversa con el kit de TI SuperscriptIII (Invitrogen) y se realizaron ensayos Taqman en un aparato de PCR en tiempo real Applied Biosystems 7500Fast.

Los niveles de expresión del gen de HBB se sometieron a ensayo del modo siguiente. El ARN se sometió a transcripción inversa utilizando la supermezcla de síntesis de primera cadena SuperScript III y oligo-dT (Life Technologies). La PCR cuantitativa se llevó a cabo con una mezcla maestra de PCR Taqman Universal (Applied Biosystems) en un sistema de PCR en tiempo real ABI 7500 Fast con los juegos de cebadores/sondas siguientes: Cebador de HBB directo: 5'-CAAGGGCACCTTTGCCACAC-3' (SEC ID n.º 11); cebador de HBB inverso: 5'-TTTGCCAAAGTGATGGGCCA-3' (SEC ID n.º 12); sonda Taqman de HBB: 5'-/56-FAM/CCTGGGCAA/ZEN/CGTGCTGGTCTGTG/3IABkFQ/-3' (SEC ID n.º 13); cebador de β -actina directo (ACTB): 5'-GGCACCCAGCACAATGAAG-3' (SEC ID n.º 14); cebador ACTB inverso 5'-GCCGATCCACACGGAGTACT-3' (SEC ID n.º 15); ACTB Taqman Probe 5'-/SMA550-Y/TCAAGATCA/ZEN/TTGCTCCTCCTGAGCGC/3IABik_FQ/-3' (SEC ID n.º 16).

Aunque la totalidad de las diez fusiones TALE-TET1DC (denominadas HB-1 a HB-10) indujeron una desmetilación significativa de los CpG en proximidad a sus sitios de unión respectivos en las células K562 humanas (figs. 8B y 8C), se observaron incrementos significativos de la expresión del gen HBB medidos mediante RT-PCR cuantitativa, con solo cuatro de dichas proteínas (HB-3, HB-4, HB-5 y HB-6) (fig. 8D). Cabe destacar que las tres proteínas (HB-4, HB-5 y HB-6) que indujeron el mayor factor de activación del promotor eran las únicas fusiones que indujeron una desmetilación significativa de CpG en la posición -266 (numeración respecto al sitio de inicio de transcripción, fig. 8D); además, la magnitud de la activación del gen HBB observada con dichas tres fusiones TALE-TET1DC se correlacionaba con el grado de desmetilación (fig. 8D). Los experimentos de control mostraron que las proteínas HB-4, HB-5 y HB-6 portadoras de las mutaciones H1671Y/D1673A (que inactivan la actividad de dominio catalítico de TET1) no consiguieron desmetilar el CpG -266 y tampoco consiguieron activar eficientemente la expresión del gen HBB en las células K562.

Se llevó a cabo el análisis de transferencia western de la manera siguiente: Se lisaron las células transfectadas en tampón de extracción de proteínas en ensayo de radioinmunoprecipitación (Cell Signaling) y se centrifugaron durante 5 min a 4 °C para recolectar el sobrenadante. La concentración de las proteínas extraídas se determinó utilizando un kit de ensayo de proteínas de Thermo Scientific. Se cargaron diez microgramos de las proteínas extraídas por carril en geles de tris-glicina al 4 % a 12 % (Life Technologies) para la electroforesis y se electrotransfirieron a una membrana de nitrocelulosa (Life Technologies). Se bloqueó esta membrana con leche seca desnatada al 5 % en solución salina tamponada con Tris-tampón Tween [Tris 25 mM (pH 7,4), NaCl 137 mM, KCl 2,7 mM, Tween-20 al 0,1 %] durante 1 hora a temperatura ambiente, seguido de la incubación con anticuerpo monoclonal de FLAG (Sigma) y anticuerpo monoclonal-HRP de β -actina (C4) (Santa Cruz Biotechnology) en tampón de bloqueo durante 2 horas a temperatura ambiente. Tras el lavado con tampón TBS-T, la membrana se incubó con anticuerpo IgG (H+L) de cabra antirratón-HRP (Bio-Rad Laboratories) a una dilución de 1:7500 en tampón de bloqueo durante 1 hora a temperatura ambiente. La membrana se reveló utilizando el sistema de quimioluminiscencia ImmunoCruz (Santa Cruz Biotechnology) y se expuso a película BioMax XAR (Kodak). Estos experimentos de transferencia western confirmaron que la pérdida de actividades de desmetilación y activación génica observada con los mutantes catalíticamente inactivos de HB-5 y HB-6 no se debía a una expresión reducida de las proteínas en células K562.

Para someter a ensayo si la desmetilación CpG -266 y la expresión incrementada del gen HBB inducida por HB-4, HB-5 y HB-6 era estable durante el tiempo, estas variables se sometieron a ensayo en un experimento de curso temporal más extendida. Las células K562 transfectadas con plásmidos de expresión codificantes de HB-4, HB-5, HB-6 o una fusión TALE-TET1DC de control con diana en el locus KLF4 se evaluaron para la desmetilación de CpG -266 y la expresión del gen HBB 4, 7, 14 y 30 días después de la transfección. Las fusiones de HB-4, HB-5 y HB-6 mostraron

todas sus niveles más altos de activación en el día 4 después de la transfección, reduciéndose la expresión de HBB constantemente hasta el mismo nivel el día 30 observado con la proteína TALE-TET1DC de control con diana en KLF4. Los resultados se muestran en las figuras 10A-B. Inesperadamente, para la totalidad de las proteínas TALE-TET1DC con diana en HBB, el grado de metilación de CpG-266 durante el tiempo en la población de células transfectadas reflejó directamente el nivel de activación de la expresión del gen HBB, mientras que el estado de metilación de otro CpG en la posición -306 no cambió significativamente durante el curso del experimento.

Conjuntamente, estos resultados sugieren fuertemente que la desmetilación de CpG -266 podría requerir la expresión continua de una proteína TAL-TET1DC diana y que la desmetilación de este CpG particular resulta necesaria para la activación observada de la expresión del gen HBB. Estos resultados demuestran que las proteínas de fusión TALE-TET1DC pueden utilizarse para desmetilar residuos CpG específicos en un promotor génico y, de esta manera, regular positivamente la expresión de ese gen en las células vivas.

Ejemplo 4. Proteínas de fusión diseñadas DZ-TET1

Además de generar fusiones TALE-TET1, también se diseñaron proteínas de fusión dedo de zinc (DZ)-TET1. Las matrices de dedo de zinc se construyeron con diana en dos sitios en el gen KLF4 y en seis sitios en el gen HBB. Cada una de dichas matrices de dedo de zinc consistía en seis dedos y se ensamblaron utilizando unidades de dos dedos según el método originalmente descrito por Moore et al., Proc. Natl. Acad. Sci. USA 98, 1432-1436 (2001), y similar al enfoque utilizado por Sangamo Biosciences (Doyon et al., Nat. Biotechnol. 26, 702-708 (2008)). Las matrices de seis dedos de zinc se ensamblaron utilizando unidades de dos dedos derivadas de proteínas de tres dedos seleccionadas mediante el método OPEN (Maeder et al., Mol. Cell 31, 294-301 (2008)). Las proteínas de seis dedos ensambladas se fusionaron con el dominio catalítico de TET1 utilizando la secuencia de conector Leu-Arg-Gly-Ser (SEC ID n.º 19) y con una señal de localización nuclear N-terminal con expresión de la proteína de fusión DZ-TET1DC controlada por el promotor EF1alfa.

De las dos fusiones DZ-TETCD con diana en KLF4, una no consiguió desmetilar los CpG próximos, mientras que la otra fue capaz de inducir la desmetilación de los CpG situados 6 y 24 pb cadena abajo del sitio de unión del DZ, en 61 % y 18 %, respectivamente (figura 9A). Una fusión DZ-TET1DC con diana en el CpG -266 en el promotor de HBB no consiguió inducir la desmetilación, pero la totalidad de las cinco fusiones con diana en la región aproximadamente 100 pb cadena abajo del sitio de inicio de transcripción fueron capaces de desmetilar los CpG con eficiencias comparables a las observadas con las proteínas TALE-TET1DC (figura 9B).

Estos datos demuestran que el dominio catalítico de TET1 puede fusionarse con dominios de unión a ADN alternativos con el fin de generar enzimas de desmetilación de ADN específica de sitio, resultando en la regulación positiva de la expresión de los genes diana en las células vivas.

Otras realizaciones

Debe entenderse que, aunque la invención se ha descrito junto con la descripción detallada de la misma, la descripción anteriormente proporcionada tiene un fin ilustrativo y no limitativo del alcance de la invención, que está definida por el alcance según las reivindicaciones adjuntas. Otros aspectos, ventajas y modificaciones se encuentran comprendidas dentro del alcance según las reivindicaciones siguientes.

REIVINDICACIONES

1. Proteína de fusión, que comprende:

un dominio de unión a ADN que se une específicamente a una secuencia diana preseleccionada que está a menos de 250 pb de un dirresiduo CpG en una célula de mamífero, en la que el dominio de unión a ADN es o comprende un dedo de zinc diseñado, y

un dominio catalítico que comprende una secuencia que cataliza la desmetilación de citosinas en ADN, en la que la secuencia comprende los aminoácidos 1580 a 2052 de Tet1 (SEC ID n.º 2), en el que la proteína de fusión se une a la secuencia diana y reduce la metilación del dirresiduo CpG.

2. Proteína de fusión según la reivindicación 1, que comprende, además, un conector entre el dominio de unión a ADN y el dominio catalítico.

3. Proteína de fusión según la reivindicación 2, en la que el conector se selecciona del grupo que consiste en Gly₃Ser (SEC ID n.º 5), Gly₃SerGly₄SerArgSer (SEC ID n.º 6), Gly₃Ser(Gly₄Ser)₂ArgSer (SEC ID n.º 7), Gly₃Ser(Gly₄Ser)₃ArgSer (SEC ID n.º 8), Gly₃Ser(Gly₄Ser)₄ArgSer (SEC ID n.º 9) y Gly₃Ser(Gly₄Ser)₅ArgSer (SEC ID n.º 10).

4. Proteína de fusión según cualquiera de las reivindicaciones 1 a 3, que comprende una pluralidad de dominios catalíticos.

5. Proteína de fusión según cualquier reivindicación anterior, que comprende, además, un dominio/secuencia de localización nuclear.

6. Método *ex vivo* de reducción de la metilación de una secuencia de ADN seleccionada en una célula de mamífero, donde el método comprende poner en contacto la célula con una proteína de fusión según cualquiera de las reivindicaciones 1 a 5, o un ácido nucleico codificante de la proteína de fusión según cualquiera de las reivindicaciones 1 a 5.

7. Proteína de fusión según cualquiera de las reivindicaciones 1 a 5, o ácido nucleico codificante de la proteína de fusión según cualquiera de las reivindicaciones 1 a 5, para la utilización en un método de reducción de la metilación de un dirresiduo CpG en una secuencia de ADN seleccionada, en una célula de mamífero para el tratamiento del cáncer, donde el método comprende poner en contacto la célula con la proteína de fusión o un ácido nucleico codificante de la proteína de fusión.

8. Método según la reivindicación 6 o proteína de fusión o ácido nucleico codificante de la proteína de fusión para la utilización según la reivindicación 7, en el que el dominio catalítico comprende los aminoácidos 1418 a 2136 de Tet1.

9. Método según la reivindicación 6 o proteína de fusión o ácido nucleico codificante de la proteína de fusión para la utilización según la reivindicación 7, en el que la célula es una célula humana.

10. Método según la reivindicación 6, en el que la secuencia de ADN seleccionada se selecciona de una secuencia de los genes RASSF1A, MGMT, BRCA1, p16, CDH1, TIMP-3, p15, p73, DAPK, APC y RELN.

11. Método según la reivindicación 7, en el que la célula está presente en un individuo:

- (i) con diagnóstico de cáncer cerebral (glioma), en el que la secuencia de ADN seleccionada es RASSF1A,
- (ii) con diagnóstico de cáncer cerebral (glioblastoma), en el que la secuencia de ADN seleccionada es MGMT,
- (iii) con diagnóstico de cáncer de mama, en el que la secuencia de ADN seleccionada es BRCA1,
- (iv) con diagnóstico de cáncer de cuello uterino, en el que la secuencia de ADN seleccionada es p16,
- (v) con diagnóstico de cáncer de esófago, en el que la secuencia de ADN seleccionada es CDH1,
- (vi) con diagnóstico de cáncer de cabeza y cuello, en el que las secuencias de ADN seleccionadas son p16 y MGMT,
- (vii) con diagnóstico de cáncer renal, en el que la secuencia de ADN seleccionada es TIMP-3,
- (viii) con diagnóstico de cáncer pulmonar, en el que la secuencia de ADN seleccionada es p16 y p73,
- (ix) con diagnóstico de cáncer ovárico, en el que la secuencia de ADN seleccionada es BRCA1,
- (x) con diagnóstico de cáncer pancreático, en el que la secuencia de ADN seleccionada es APC,
- (xi) con diagnóstico de cáncer prostático, en el que la secuencia de ADN seleccionada es BRCA2,
- (xii) con diagnóstico de cáncer uterino, en el que la secuencia de ADN seleccionada es hMLH1,
- (xiii) con diagnóstico de leucemia, en el que la secuencia de ADN seleccionada es p15,
- (xiv) con diagnóstico de linfoma, en el que la secuencia de ADN seleccionada es DAPK,
- (xv) con diagnóstico de mieloma, en el que la secuencia de ADN seleccionada es DAPK,
- (xvi) con diagnóstico de rhabdomyosarcoma, en el que la secuencia de ADN seleccionada es PAX3, o
- (xvii) con diagnóstico de esquizofrenia, en el que la secuencia de ADN seleccionada es RELN.

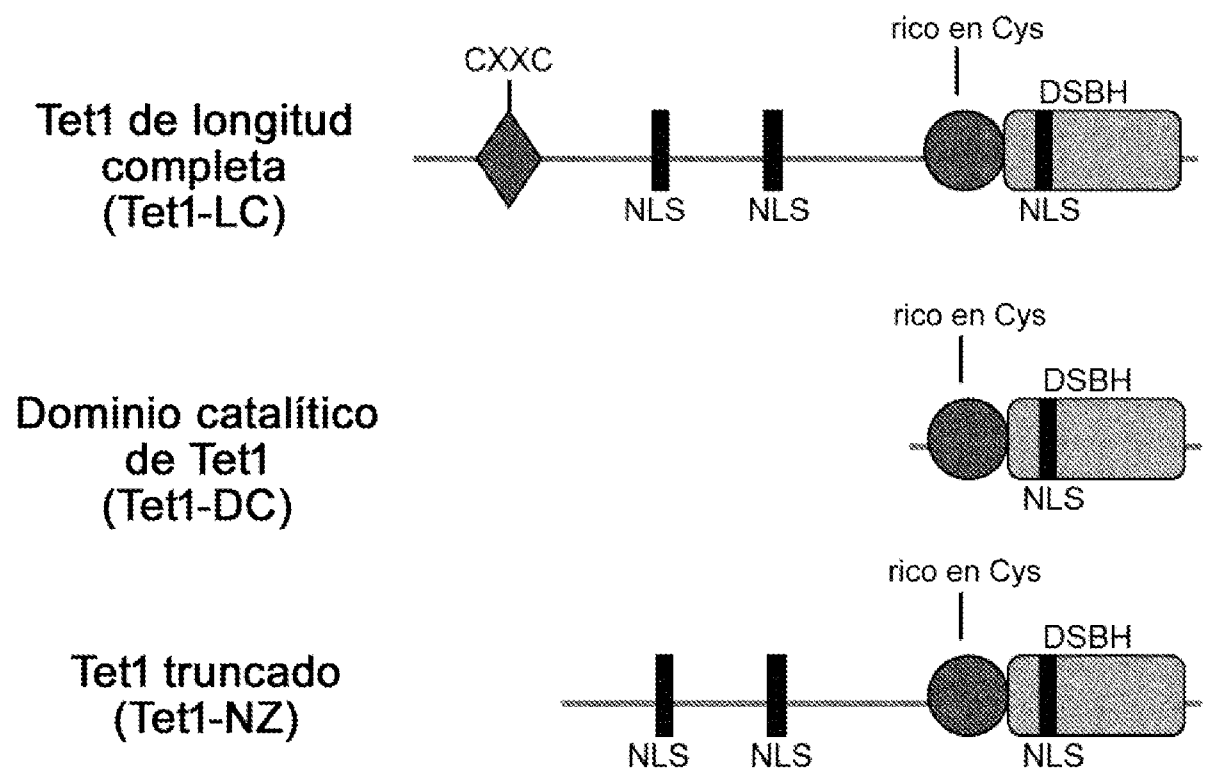


FIG. 1A

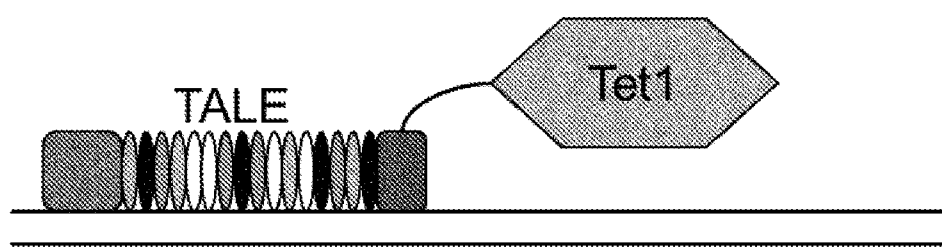


FIG. 1B

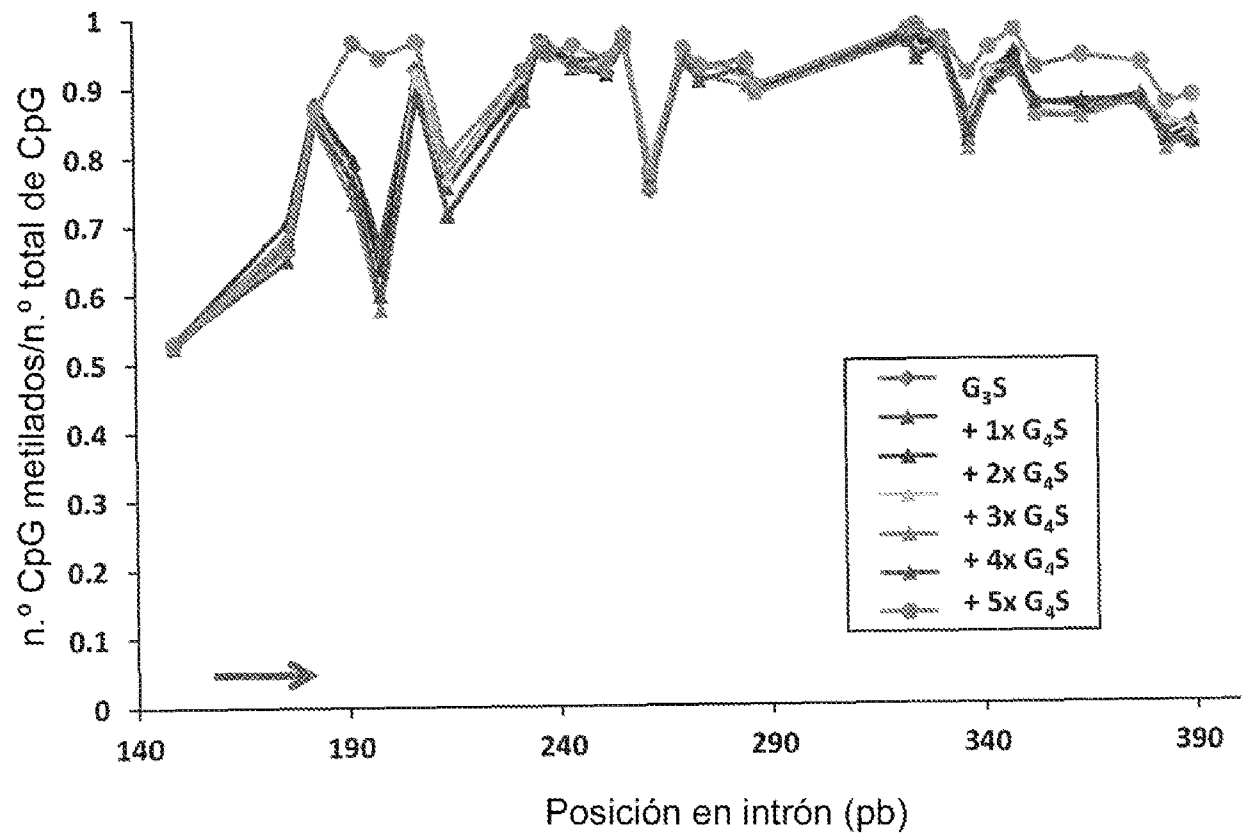


FIG. 1C



FIG. 2

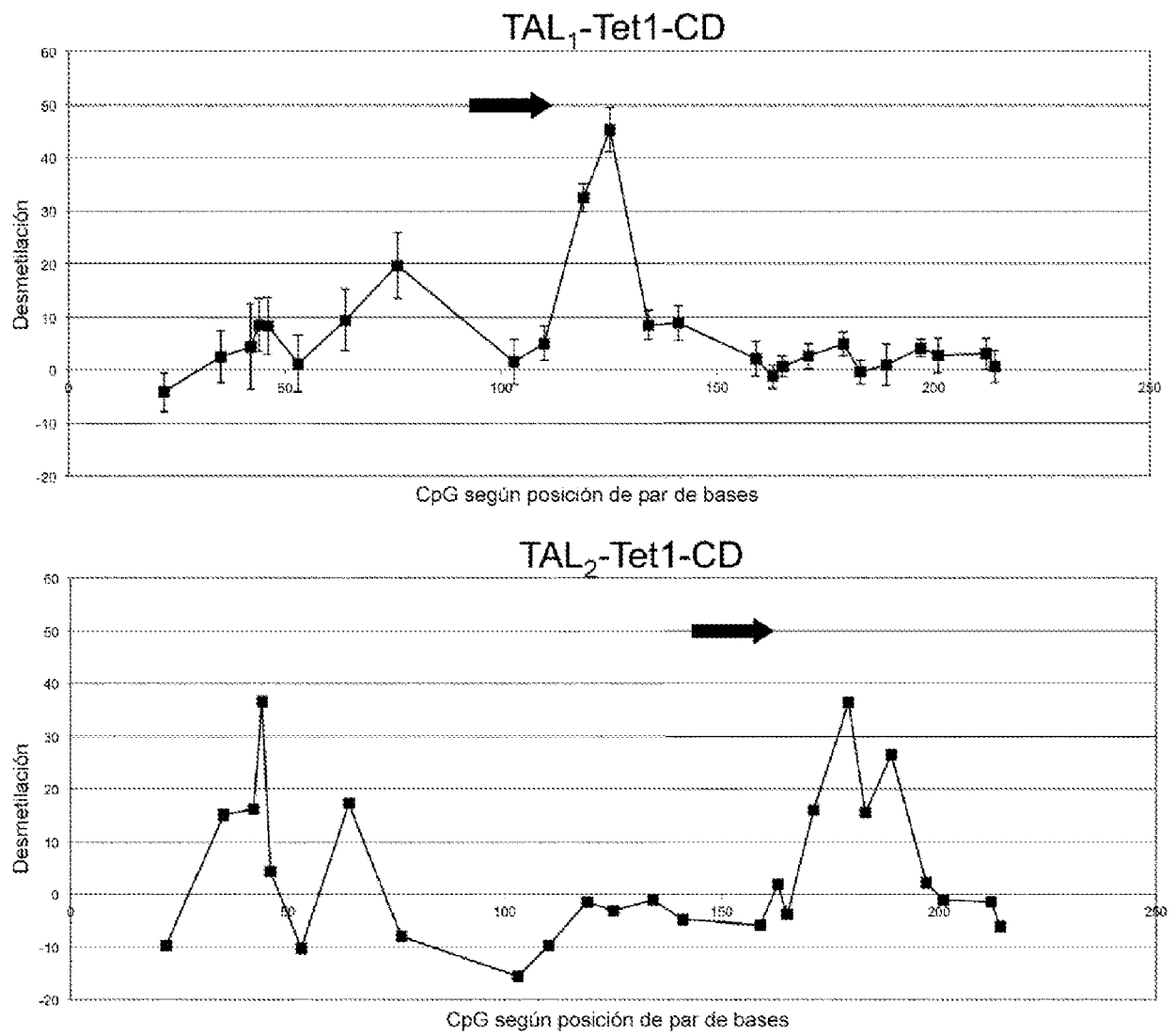


FIG. 3, página 1 de 2

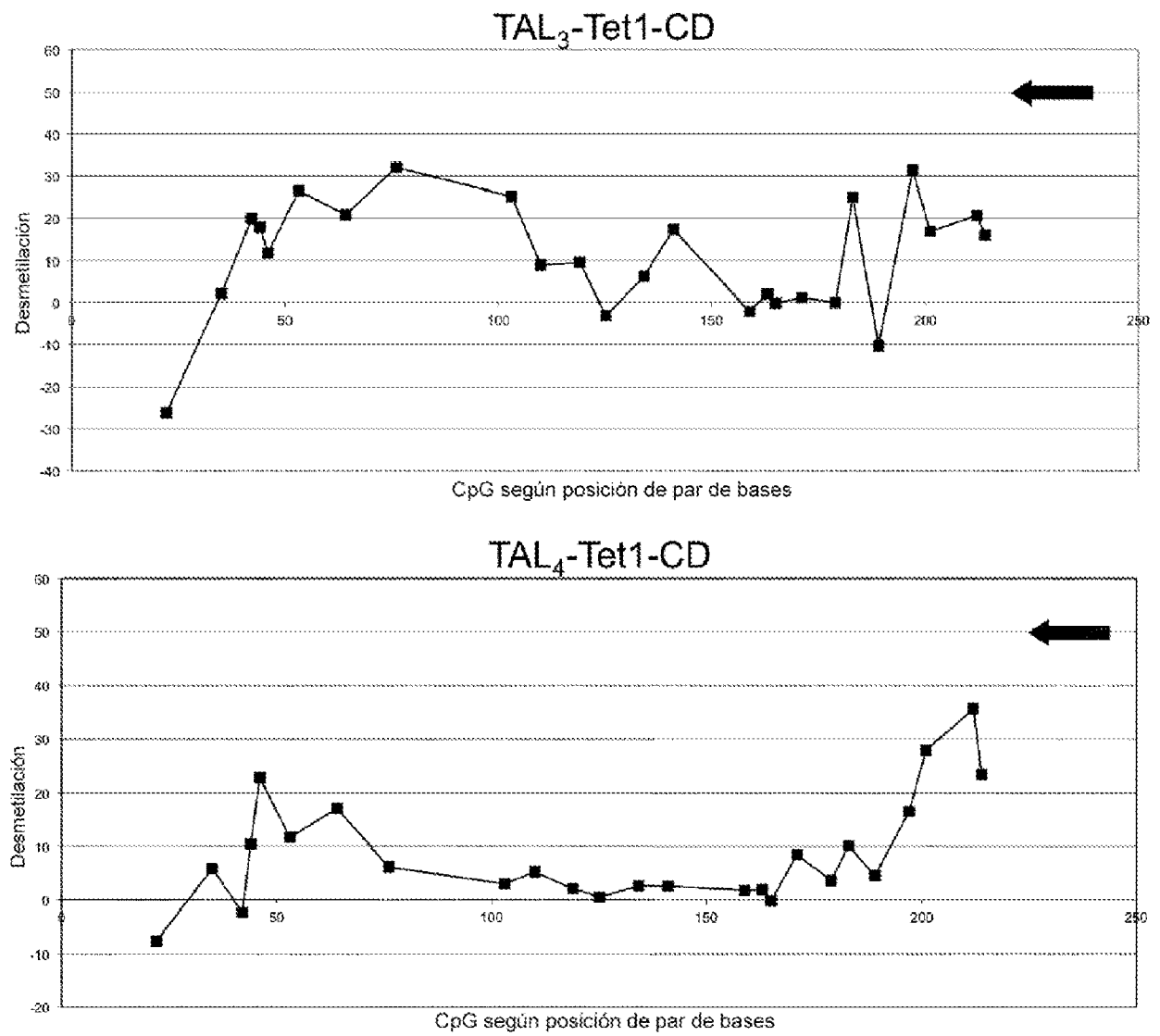


FIG. 3, página 2 de 2

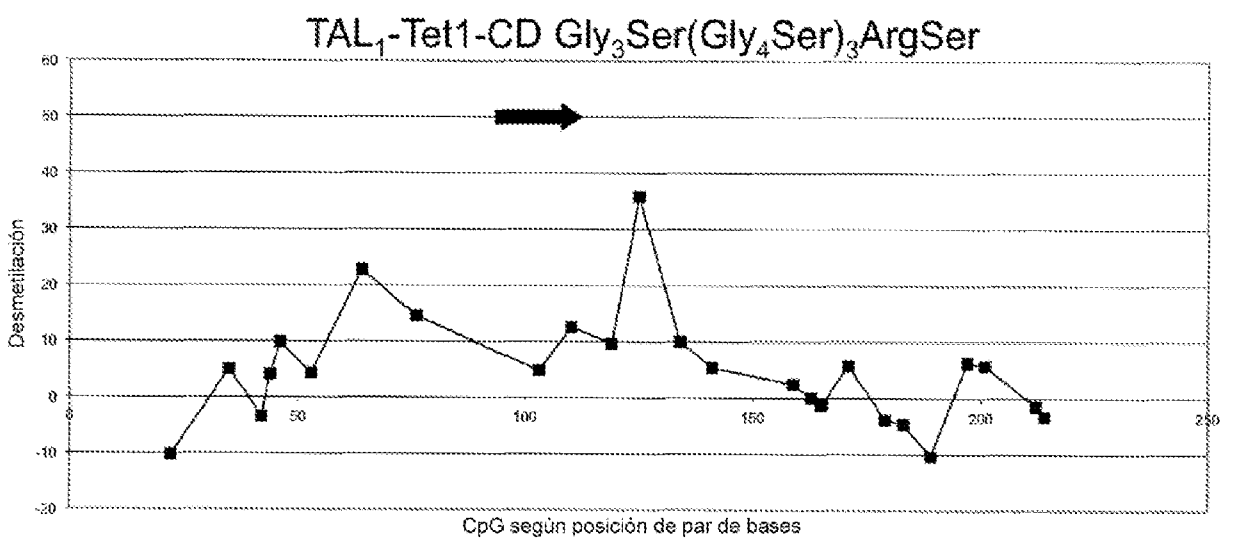
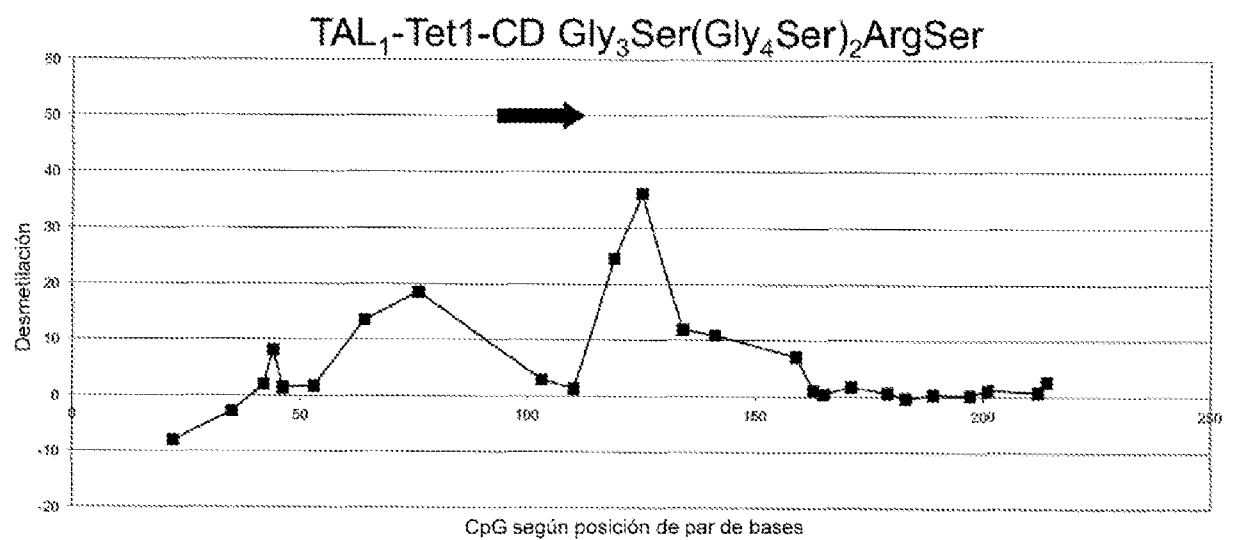
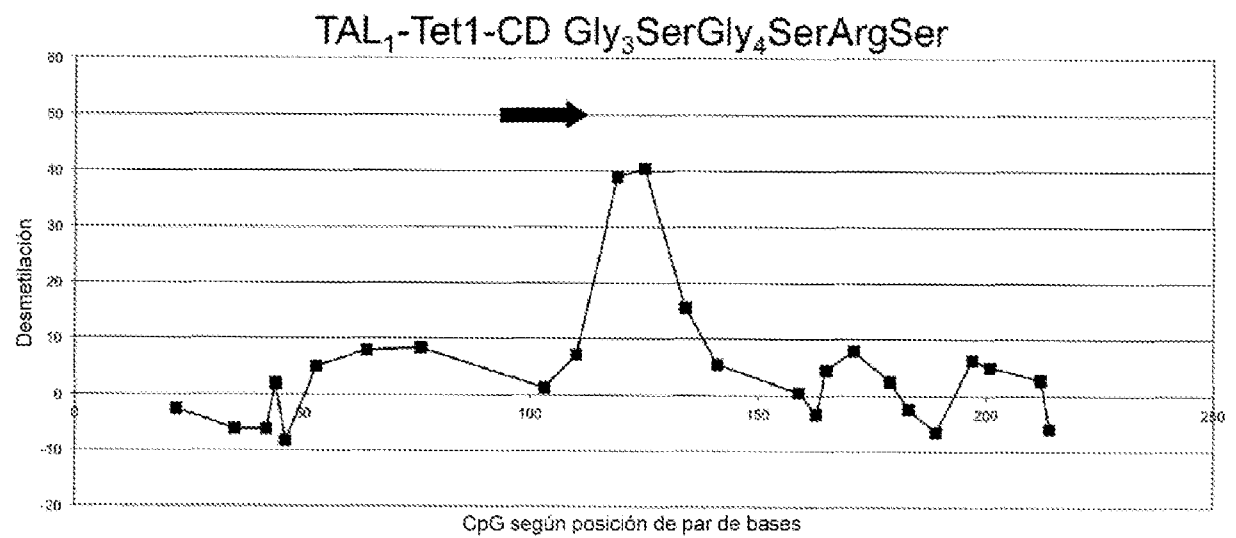


Fig. 4A, página 1 de 2

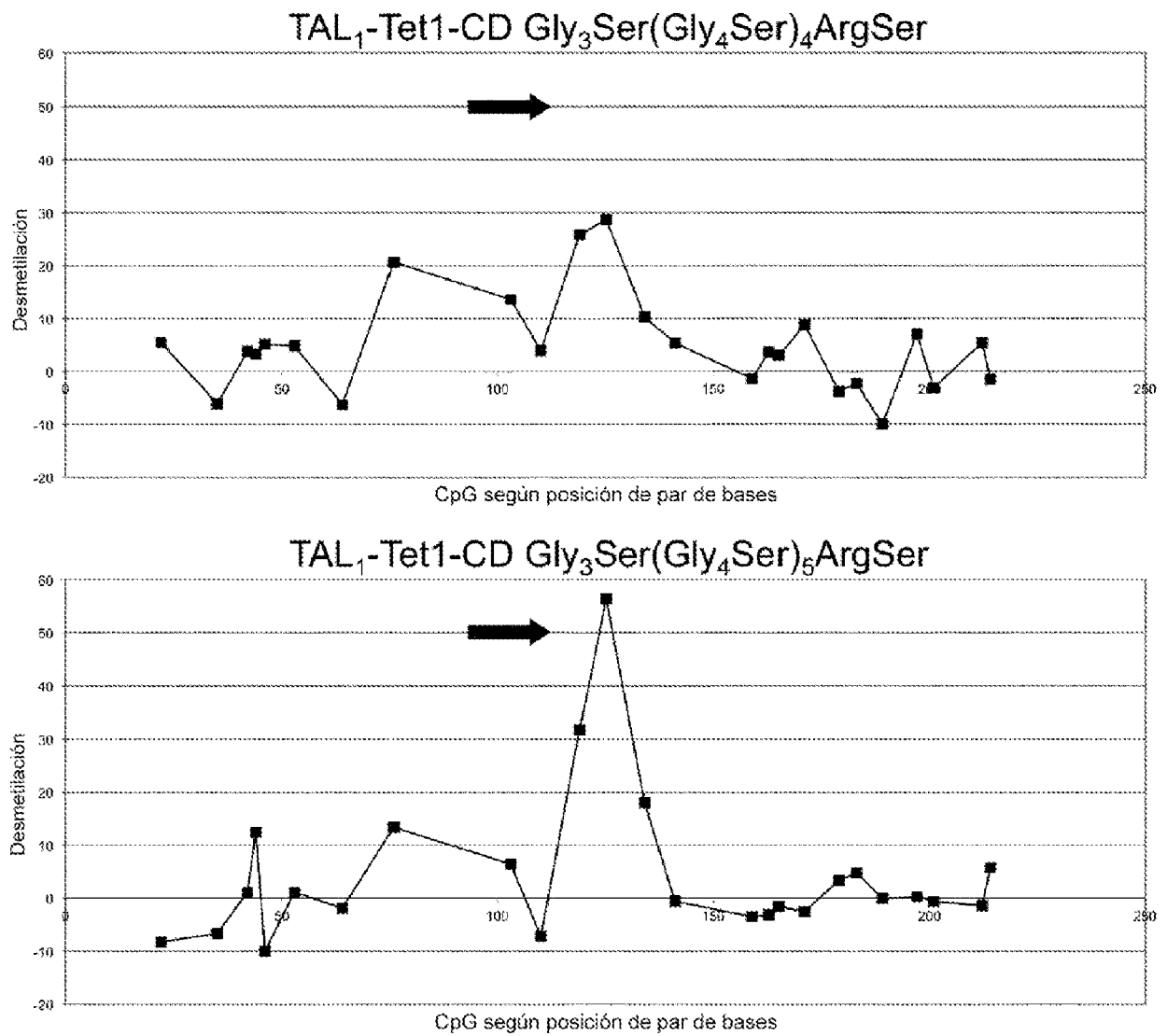
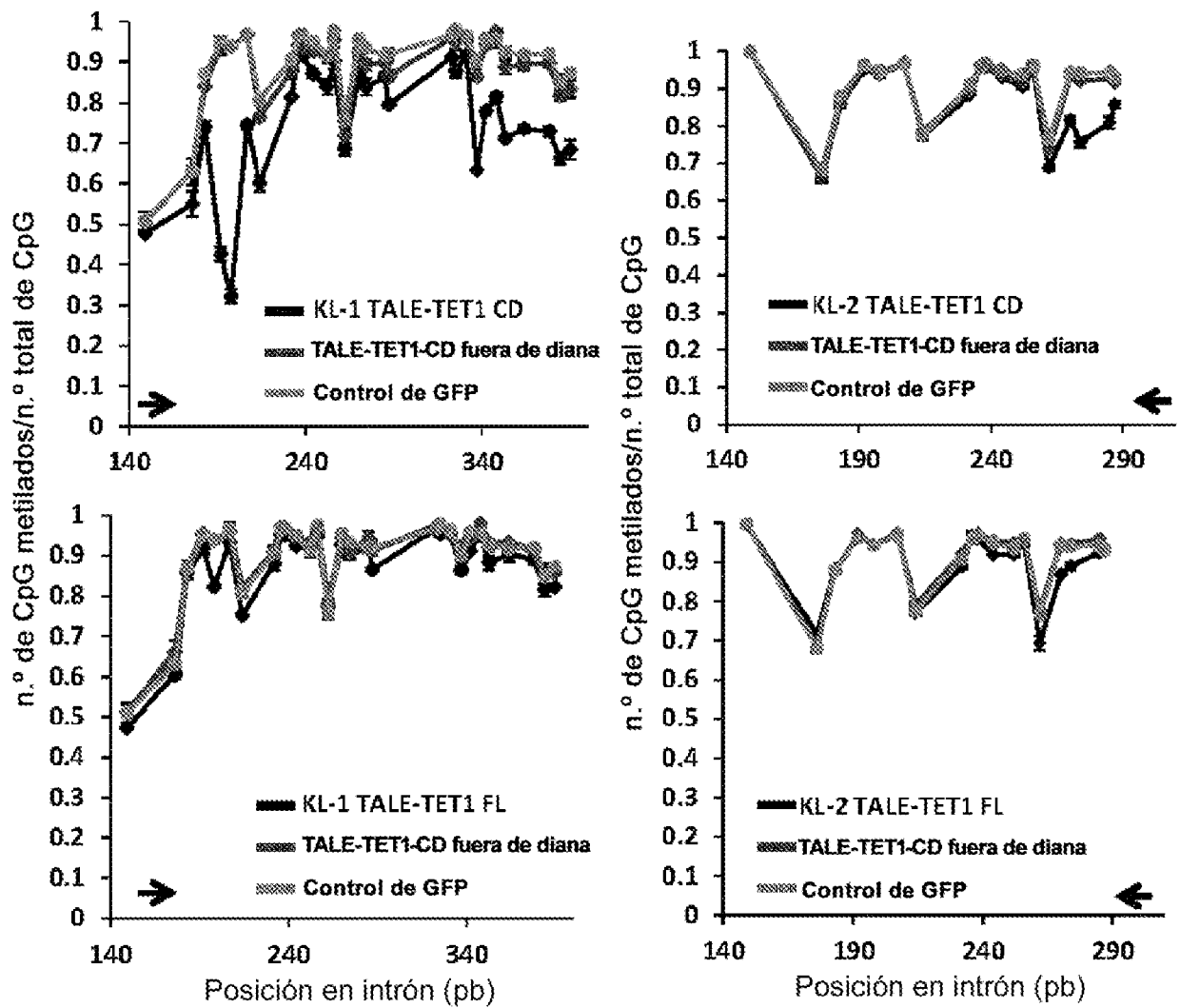


FIG. 4A, página 2 de 2



Tet1-LC

MSRSRHARFSLVRKEDVNKKKKNSQLRKTTKGANKNVASVKTLSPGKCLKQLIQERDVKKKT
 EPKPPVPVRSLLTRAGAARMNLDRTEVLVFQNPESLTCNGFTMALRSTSLSRRLSQPPLVVAK
 SKKVPLSKGLEKQHDQDYKILPALGVKHSENDSVPMQDTQVLPDIETLIGVQNPSSLKGKSQ
 ETTQFWSQRVEDSKINIPTHSGFAAEILPGPLEGTRCGEGLFSEETLNDTSGSPKMFAQDTV
 CAPFPQRATPKVTSQGNPSIQLEELGSRVESLKLSDSYLDPIKSEHDCYPTSSLNKVIPDLN
 LRNCLALGGSTSPTSVIKFLLAGSKQATLGAKPDHQEAFEATANQQEVSDTTSFLGQAFGAI
 PHQWELPGADPVHGEALGETPDLPEIPGAIPVQGEVFGTILDQQETLGMMSGSVVPDLPVFLP
 VPPNPIATFNAPSKWPEPQSTVSYGLAVQGAIQILPLGSGHTPQSSSNSEKNSLPPVMAISN
 VENEQVHISFLPANTQGFPLAPERGLFHASLGIAQLSQAGPSKSDRGSSQVSVTSTVHVVN
 TTVVTMPVPMVSTSSSSSYTTLLPTLEKKKKRRCRGVCEPCQOKTNCGECTYCKNRKNSHQICK
 KRKCEELKKKPSVVVPLEVIKENKRPQREKKPKVLKADFDNKPVNGPKSESMDYSRCGHGEE
 QKLELNPHTVENVTKNEDSMTGIEVEKWTQNKKSQLTDHVKGDFSANVPEAEKSKNSEVDKK
 RTKSPKLFVQTVRNGIKHVHCLPAETNVSEFKKFNIEEFGKTLENNSYKFLKDTANHKNAMSS
 VATDMSCDHLKGRSNVLVFQQPGFNCSSSIHSSHSIINHHSIHNEDGQPKTPENIPSKEPK
 DGSPVQPSLLSLMKDRRLTLEQVVAIEALTQLSEAPSENSSPSKSEKDEESEQRTASLLNSC
 KAILYTVRKDLQDPNLQGEPPKLNHCPSLEKQSSCNTTVVFNGQTTTLSNSHINSATNQASTK
 SHEYSKVTNSLSLFIKSNSSKIDTNKSIAGGIITLDNCNDLHQLPPRNNEVEYCNQLLDS
 SKKLDSDDLSCQDATHQIEEDVATQLTQLASIIKINYIKPEDKKVESTPTSLVTCNVQQKY
 NOEKGTMOQKPPSSVHNNHGSSLTQKKNPTQKKTSTPSRDRRKKKPTVVSQENDROKWEK
 LSYMYGTICDIWIASKFQNFQFCPHDFPTVFGKISSSTKIWKPLAQTRSIMQPKTVFPPLT
 QIKLQRYPESAEEKVKVEPLDLSLSLFHLKTESNGKAFTDKAYNSQVQLTVNANQKAHPLTQP
 SSPFNQCANVMAGDDQIREQQVVKEQLMHQRLPTLPGISHETPLPESALTLRNVNVVCSGGI
 TVVSTKSEEEVCSSSFGTSEFSTVDSAQKNFNDYAMNFFTNPNTKNLVSITKDSELPTCSCLD
 RVIQKDKGPFYYTHLGAGPSVAAVREIMENRYGQKGNARIEIVVYTGKEGKSSHGCPIAKWV
 LRRSSDEEKVLCLVRQRTGHHHCPTAVMVVLIMVWDGIPLPMADRLYTELTENLKSYNHGPTD
 RRCTLNENRTCTCQGIDPETCGASFSGCSWSMYFNGCKFGRSPSPRRFRIDPSSPLHEKNL
 EDNLQSLATRLAPIYKQYAPVAYQNOVEYENVARECRLGSKEGRPFSGVTACLDFCAHPHRD
 IHNMNNGSTVVCTLTREDNRSLGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEAKIKSGAIEV
 LAPRRKKRRTCFTQPVPRSGKKRAAMMTEVLAKIRAVEKKPIFRIKRKNNSTTTNNSKFPSSL
 PTLGSNTETVQPEVKSETEPHFILKSSDNKTYSLMPSPAPHPVKEASPGFSWSPKTASATPA
 PLKNDATASCGFSERSSTPHCTMPSGRLSGANAAAADGPGISQLGEVAPLPTLSAPVMEPLI
 NSEPSTGVTEPLTPHQPNHQPSFLTSPQDLASSPMEEDEQHSEADEPFSDEPLSDDPLSPA
 EKLPHIDEYWSDSEHIFLDANIGGVAIAPAHGSLVIECARRELHATTPVEHPNRNHPTRLSL
 VFYQHKNLKNKPQHGFELNKKIKFEAKEAKNKKMKASEQKDQAANEGPEQSSEVNELNQIPSHK
 ALTLTHDNVTVSPYALTHVAGPYNHWV (SEC ID n.º 2)

FIG. 5A

Tet1-NZ

EELKKKPSVVVPLEVIKENKRPQREKKPKVLKADFDNKPVNGPKSESMDYSRCGHGEEQKLE
 LNPHTVENVTKNEDSMTGIEVEKWTQNKKSQLTDHVKGDFSANVPEAEKSKNSEVDKKRTKS
 PKLFFVQTVRNGIKHVHCLPAETNVSFKKFNIIEFGKTLENNSYKFLKDTANHKNAMSSVATD
 MSCDHLKGRSNVLFVQQPGFNCSSIPHSSHSIINHHSIHNEGDQPKTPENIPSKEPKDGSF
 VQPSLLSLMKDRRLTLEQVVAIEALTQLSEAPSENSSPSKSEKDEESEQRTASLLNSCKAIL
 YTVRKDLQDPNLQGEPPKLNHCPSLEKQSSCNTVVVFNGQTTTLSNSHINSATNQASTKSHEY
 SKVTNSLSLFIIPKSNSSKIDTNKSIAQGIITLDNCSNDLHQLPPRNNEVEYCNQLLDSSKKL
 DSDDLSCQDATHTQIEEDVATQLTQLASIIKINYIKPEDKKVESTPTSLVTCNVQQKYNQEK
 GTMQQKPPSSVHNNHGSSSLTKQKNPTQKKTKSTPSRDRRKKKPTVVSYQENDRQKWEKLSYM
 YGTICDIWIASKFQNFQFCPHDFPTVFGKISSSTKIWKPLAQTRSIMQPKTVFPPLTQIKL
 QRYPESAEEKVKVEPLDSLFLHLKTESNGKAFTDKAYNSQVQLTVNANQKAHPLTQPSSPP
 NQCANVMAGDDQIRFQQVVKEQLMHQRLPTLPGISHETPLPESALTLRNVNVVCSGGITVVS
 TKSEEEVCSSSFGTSEFSTVDSAQKNFENDYAMNFFTNPTKNLVSITKDSELPTCSCLDRIQ
 KDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSHGCPIAKWVLRRS
 SDEEKVLCLVRQRTGHHCP TAVMVVLIMVWDGIPLPMADRLYTEL TENLKS YNGHPTDRRCT
 LNENRTCTCQGIDPETCGASFSGCSWSMYFNGCKFGRSPSPRRFRIDPSSPLHEKNLEDNL
 QSLATRLAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACLD FCAH PHRDIHNM
 NNGSTVVCTLTREDNRS LGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEAKIKSGAIEVLAPR
 RKKRTCFTQPVPRSGKKRAAMMTEVLAHKIRAVEKKPIPRIKRKNNSTTTNNSKPSSLPTLG
 SNTETVQPEVKSETEPHFILKSSDNTKTYSLMPSAPHPVKEASPGFSWSPKTASATPAPLKN
 DATASCGFSERSSTPHCTMPSGRLSGANAAAADGPGISQLGEVAPLPTLSAPVMEPLINSEP
 STGVTEPLTPHQPNHQPSFLTSPQDLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLP
 HIDEYWSDSEHIFLDANIGGVAIAPAHGSVLI ECARRELHATTPVEHPNRNHPTRL SLV FYQ
 HKNLNKPQHGFELNKIKFEAKEAKNKKMKASEQKDQAANEGPEQSSEVNELNQIPSHKALT
 LTHDNVTVSPYALTHVAGPYNHVW (SEC ID n.° 3)

FIG. 5B**Tet1-CD**

LPTCSCLDRIQKDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSH
 GCPIAKWVLRRSSDEEKVLCLVRQRTGHHCP TAVMVVLIMVWDGIPLPMADRLYTEL TENLK
 SYNGHPTDRRCTLNENRTCTCQGIDPETCGASFSGCSWSMYFNGCKFGRSPSPRRFRIDPS
 SPLHEKNLEDNLQSLATRLAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACLD
 FCAH PHRDIHNMNNGSTVVCTLTREDNRS LGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEAK
 IKSGAIEVLAPRRKKRTCFTQPVPRSGKKRAAMMTEVLAHKIRAVEKKPIPRIKRKNNSTTT
 NNSKPSSLPTLGSNTETVQPEVKSETEPHFILKSSDNTKTYSLMPSAPHPVKEASPGFSWSP
 KTASATPAPLKNDATASCGFSERSSTPHCTMPSGRLSGANAAAADGPGISQLGEVAPLPTLS
 APVMEPLINSEPSTGVTEPLTPHQPNHQPSFLTSPQDLASSPMEEDEQHSEADEPPSDEPLS
 DDPLSPAEEKLP HIDEYWSDSEHIFLDANIGGVAIAPAHGSVLI ECARRELHATTPVEHPNR
 NHPTRL SLV FYQHKNLNKPQHGFELNKIKFEAKEAKNKKMKASEQKDQAANEGPEQSSEVNE
 LNQIPSHKALT LTHDNVTVSPYALTHVAGPYNHVW (SEC ID n.° 4)

FIG. 5C

FIG. 6A

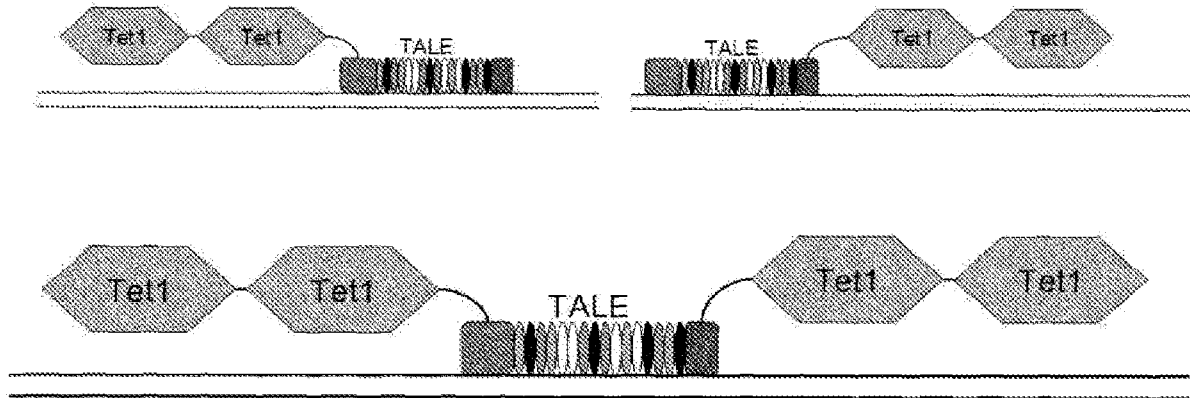


FIG. 6B.

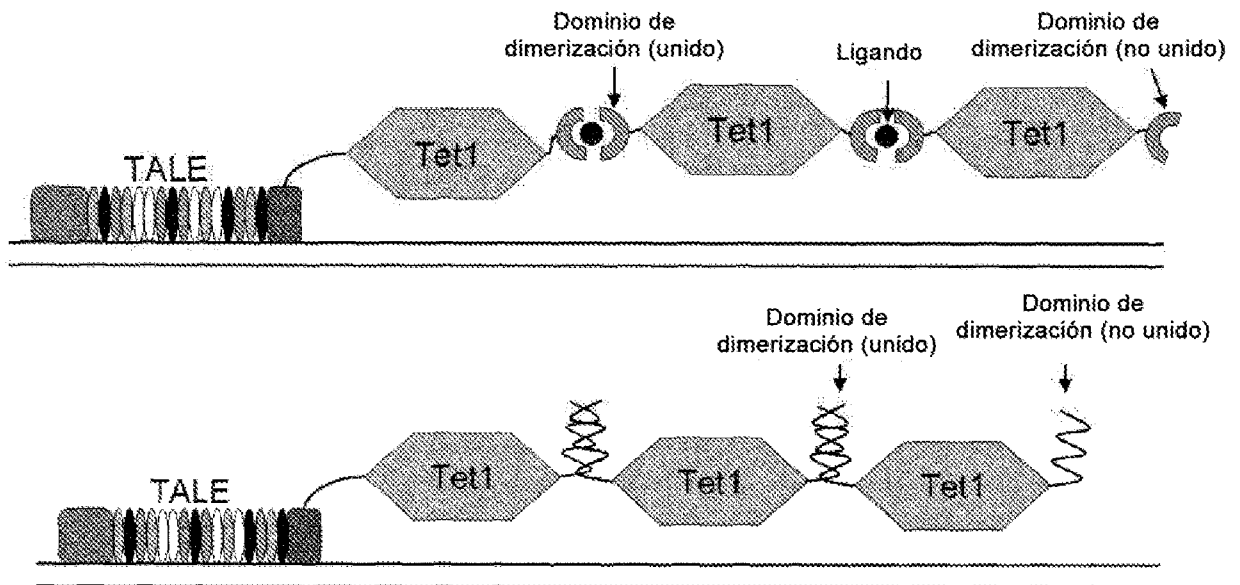


FIG. 6C.

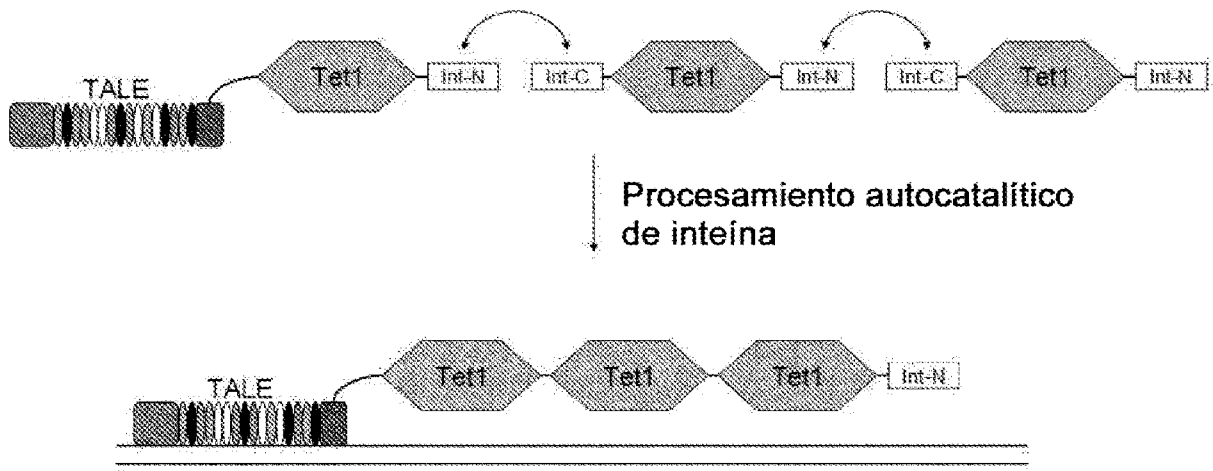


FIG. 6D.

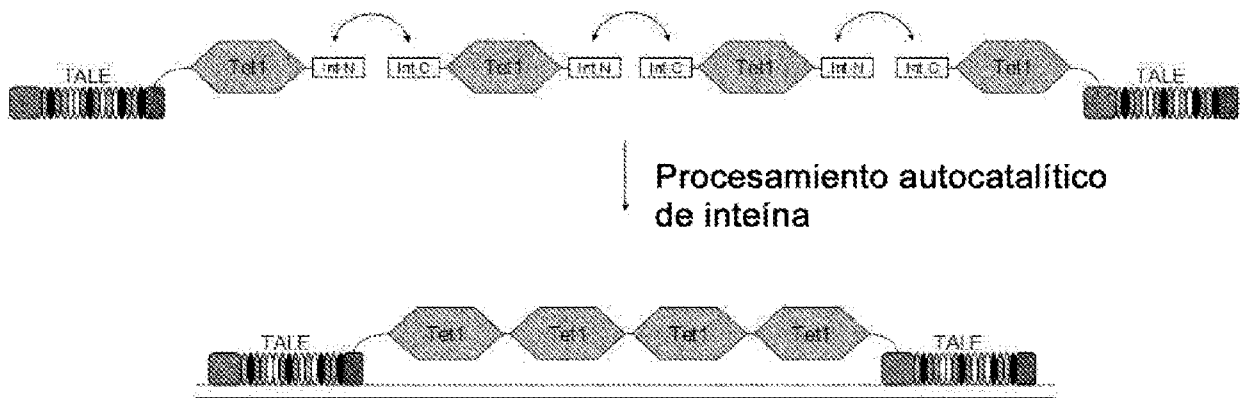


FIG. 7

Enfermedad/Afección	Gen	Referencia
Cáncer		
Vejiga	Genes múltiples	1
Cerebro (glioma)	RASSF1A	2, 3
Cerebro (glioblastoma)	MGMT	4
Mama	BRCA1	5
Mama	Genes múltiples	6, 7
Cuello uterino	P16	8
Colon	Genes múltiples	1
Esófago	CDH1	1
Cabeza y cuello	pf6, MGMT	1
Riñón	TIMP-3	1
Leucemia	p15	1
Hígado	Genes múltiples	9
Pulmón	p16, p73	1
Linfoma	DAPK	1
Mieloma	DAPK	10
Ovario	BRCA1	11
Páncreas	APC	1
Próstata	BRCA2	1, 12
Rabdomiosarcoma	PAX3	13
Útero	hMLH1	1
Neurológica		
Esquizofrenia	RELN	14, 15
Cardiovascular		
Ateroesclerosis	Genes múltiples	16

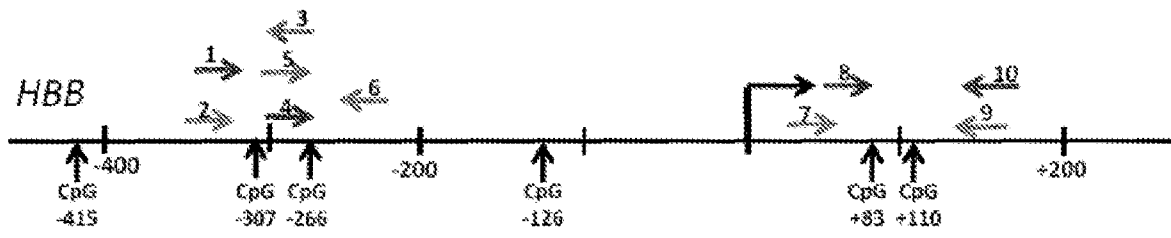


FIG. 8A

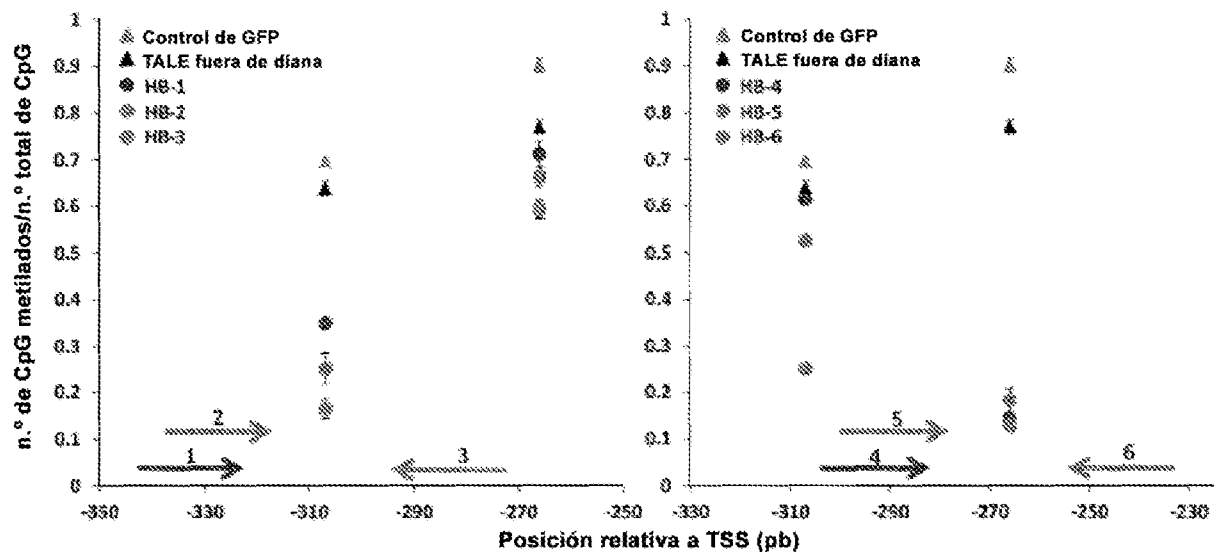


FIG. 8B

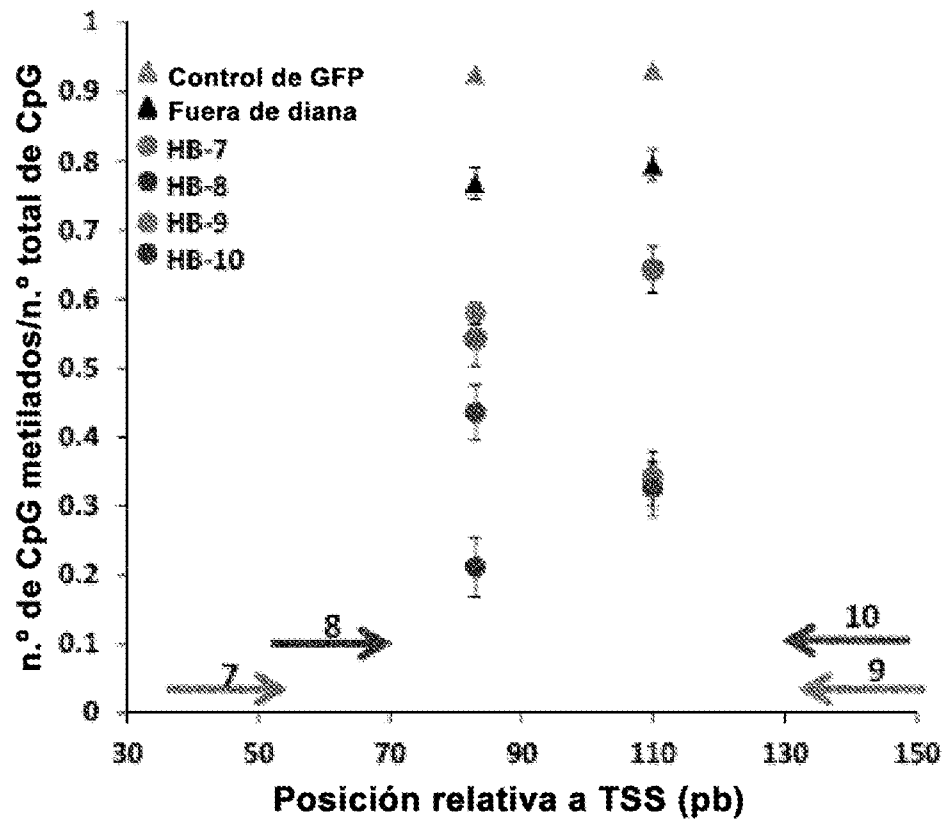


FIG. 8C

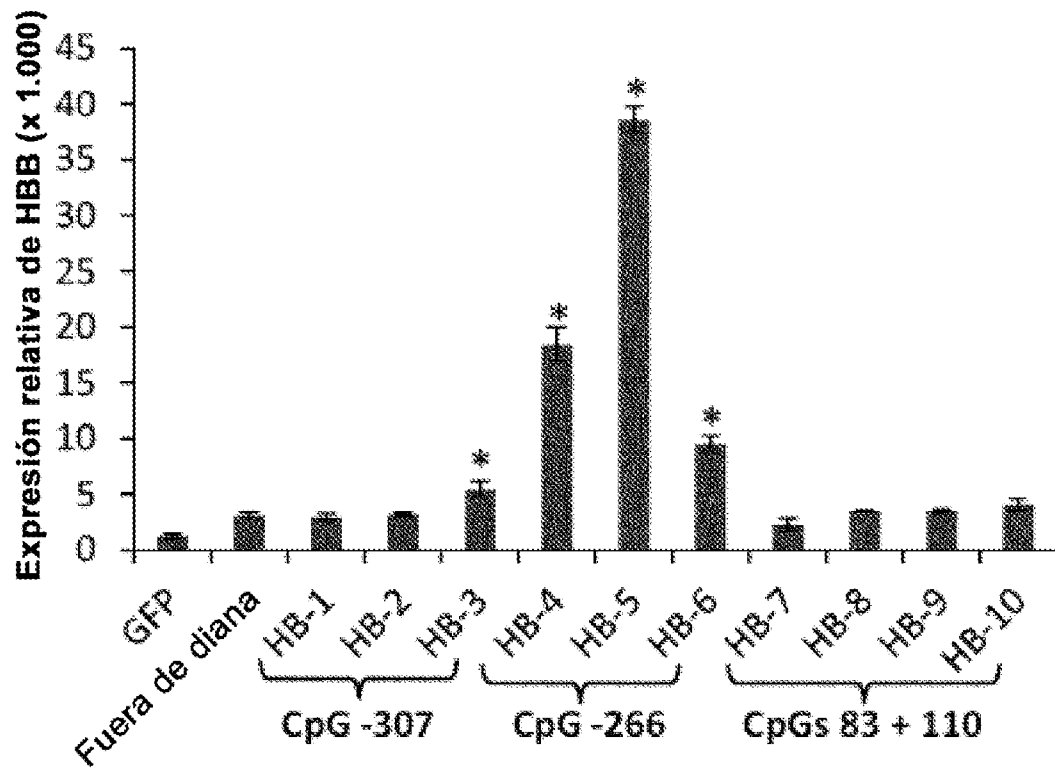


FIG. 8D

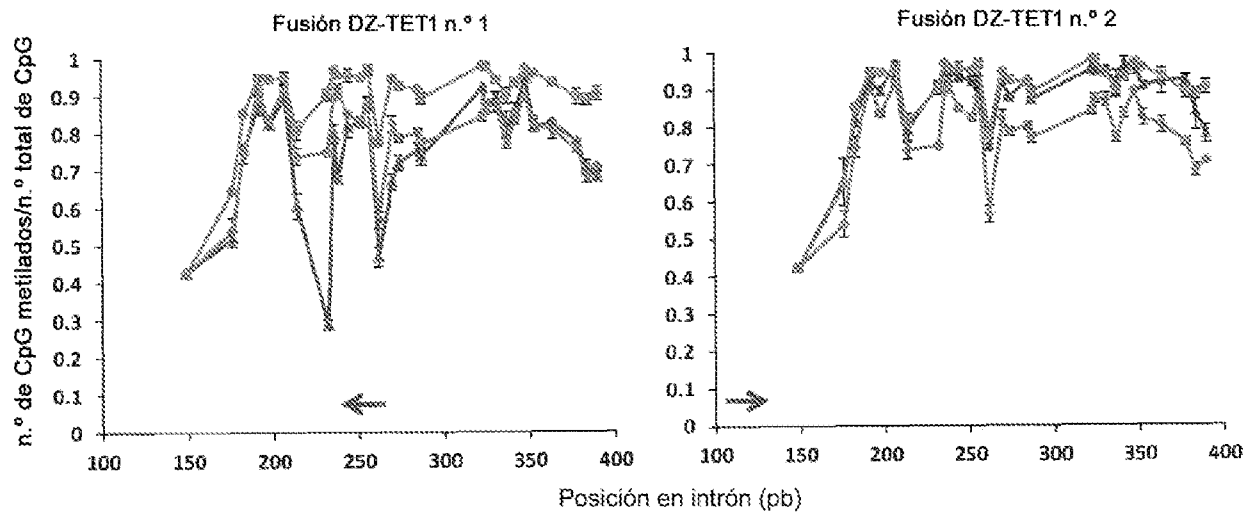


FIG. 9A

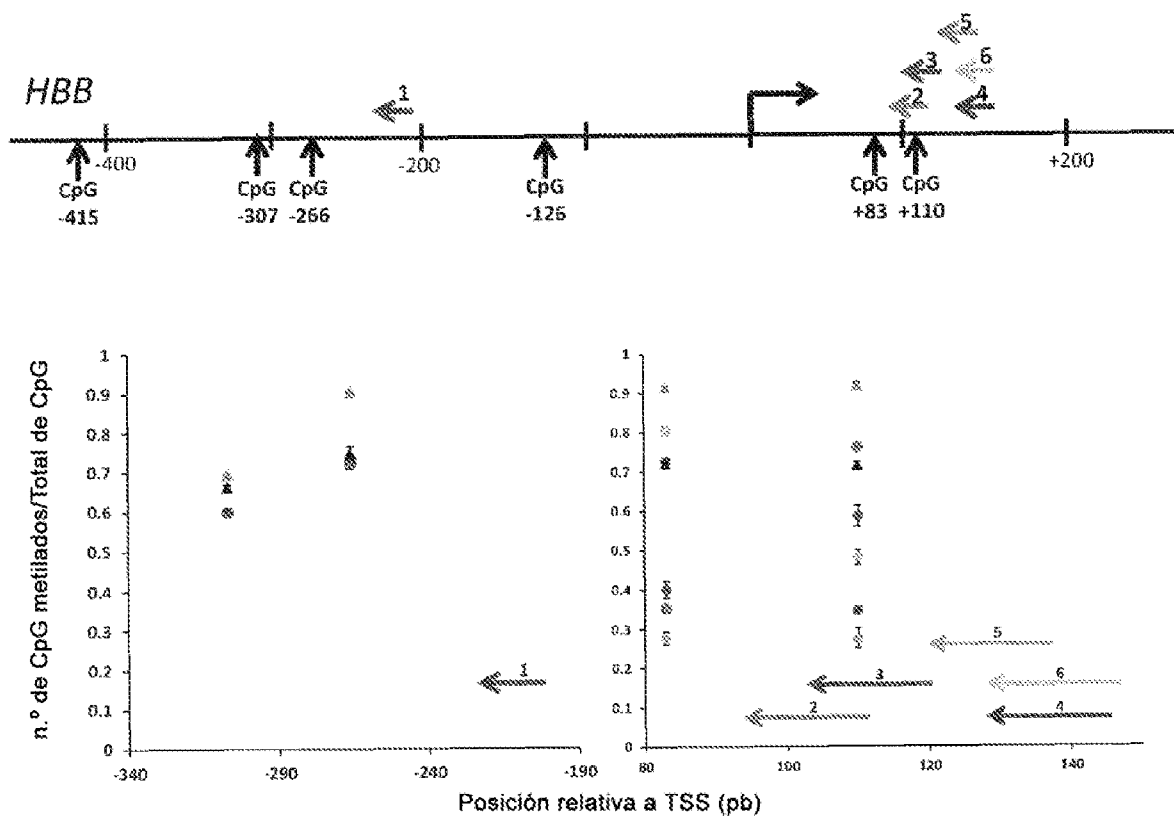


FIG. 9B

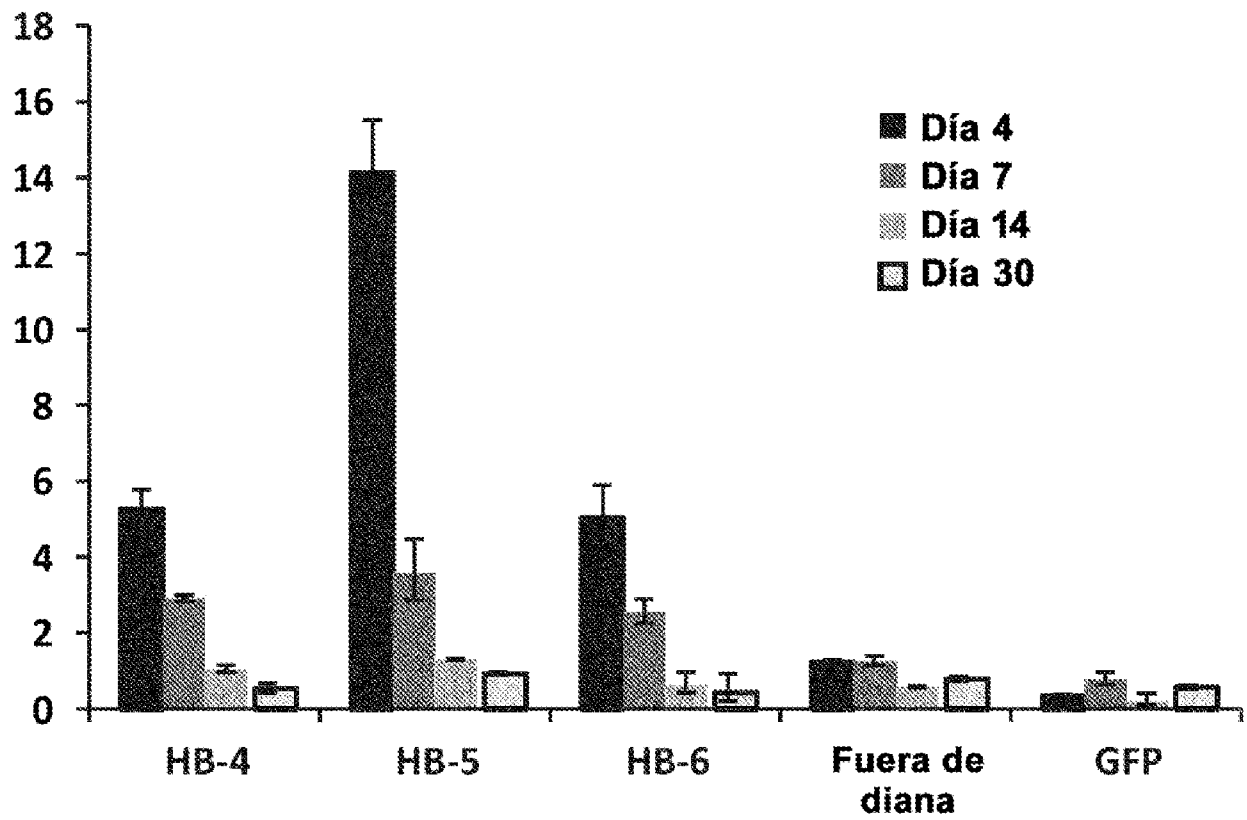


FIG. 10A

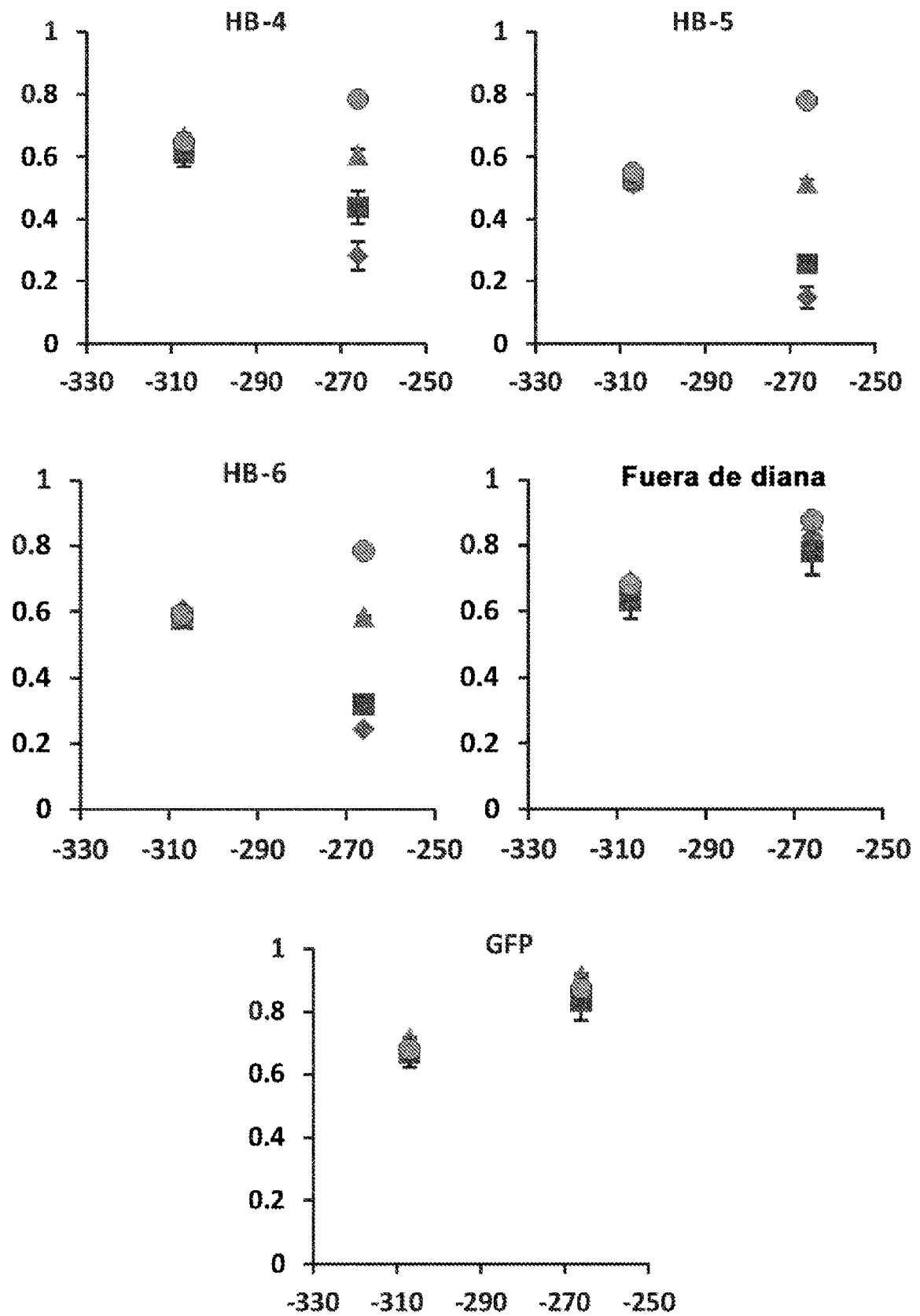


FIG. 10B