(19) DANMARK

(10) **DK/EP 3744835 T5**

Patent- og Varemærkestyrelsen

(12) Rettet oversættelse af europæisk patentskrift

(51) Int.Cl.: *C 12 N 9/02 (2006.01)* *A 61 K 38/17 (2006.01)* *A 61 K 48/00 (2006.01)* *A 61 P 9/00 (2006.01)* *A 61 P 25/18 (2006.01)* *A 61 P 35/00 (2006.01)* *C 07 K 14/435 (2006.01)* *C 07 K 19/00 (2006.01)* *C 12 N 15/62 (2006.01)*

(45) Oversættelsen bekendtgjort den: **2024-08-26**

(80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2024-01-31**

(86) Europæisk ansøgning nr.: **20183740.8**

(86) Europæisk indleveringsdag: **2013-05-29**

(87) Den europæiske ansøgnings publiceringsdag: **2020-12-02**

(30) Prioritet: **2012-05-29 US 201261652711 P** **2013-03-15 US 201313838520**

(62) Stamansøgningsnr: **13797024.0**

(84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**

(73) Patenthaver: **The General Hospital Corporation, 55 Fruit Street, Boston, MA 02114, USA**

(72) Opfinder: **JOUNG, Keith, J, 1 Magnolia Way, Winchester, MA 01890, USA**
MAEDER, Morgan, 872 Chestnut St., Waban, MA 02468, USA
ANGSTMAN, James, 25 Peterborough Street, Apt 15, Boston, MA, 02215, USA

(74) Fuldmægtig i Danmark: **AWA Denmark A/S, Strandgade 56, 1401 København K, Danmark**

(54) Benævnelse: **DNA MODIFICERENDE FUSIONSPROTEINER OG METODER TIL ANVENDELSE DERAF**

(56) Fremdragne publikationer:

WO-A2-2007/128982
WO-A2-2009/134409
WO-A2-2010/037001
DAVID J. GREGORY ET AL: "Selective DNA demethylation by fusion of TDG with a sequence-specific DNA-binding domain", EPIGENETICS, vol. 7, no. 4, 1 April 2012 (2012-04-01), pages 344-349, XP055250785, US ISSN: 1559-2294, DOI: 10.4161/epi.19509
CARINA FRAUER ET AL: "Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1", PLOS ONE, vol. 6, no. 2, 2 February 2011 (2011-02-02), page e16627, XP055389355, DOI: 10.1371/journal.pone.0016627
MIGUEL R. BRANCO ET AL: "Uncovering the role of 5-hydroxymethylcytosine in the epigenome", NATURE REVIEWS GENETICS, 15 November 2011 (2011-11-15), XP055057585, ISSN: 1471-0056, DOI: 10.1038/nrg3080
D. RODENHISER: "Epigenetics and human disease: translating basic biology into clinical applications", CMAJ. CANADIAN MEDICAL ASSOCIATION JOURNAL, vol. 174, no. 3, 30 January 2006 (2006-01-30), pages 341-348, XP055389418, CA ISSN: 0820-3946, DOI: 10.1503/cmaj.050774
FOLEY JONATHAN E ET AL: "Targeted mutagenesis in zebrafish using customized zinc-finger nucleases", NATURE PROTOCOLS, NATURE PUBLISHING GROUP, GB, vol. 4, no. 12, 3 December 2009 (2009-12-03), pages 1855-1868, XP009163787, ISSN: 1750-2799

Fortsættes ...

SANJANA NEVILLE E ET AL: "A transcription activator-like effector toolbox for genome engineering", NATURE PROTOCOLS, NATURE PUBLISHING GROUP, GB, vol. 7, no. 1, 5 January 2012 (2012-01-05) , pages 171-192, XP009170390, ISSN: 1750-2799
DAVID J. GREGORY ET AL: "Selective DNA demethylation by fusion of TDG with a sequence-specific DNA-binding domain", EPIGENETICS, vol. 7, no. 4, 1 April 2012 (2012-04-01), US, pages 344 - 349, XP055250785, ISSN: 1559-2294, DOI: 10.4161/epi.19509
CARINA FRAUER ET AL: "Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1", PLOS ONE, vol. 6, no. 2, 2 February 2011 (2011-02-02), pages e16627, XP055389355, DOI: 10.1371/journal.pone.0016627
D. RODENHISER: "Epigenetics and human disease: translating basic biology into clinical applications", CMAJ. CANADIAN MEDICAL ASSOCIATION JOURNAL, vol. 174, no. 3, 30 January 2006 (2006-01-30), CA, pages 341 - 348, XP055389418, ISSN: 0820-3946, DOI: 10.1503/cmaj.050774
FOLEY JONATHAN E ET AL: "Targeted mutagenesis in zebrafish using customized zinc-finger nucleases", NATURE PROTOCOLS, NATURE PUBLISHING GROUP, GB, vol. 4, no. 12, 3 December 2009 (2009-12-03), pages 1855 - 1868, XP009163787, ISSN: 1750-2799
SANJANA NEVILLE E ET AL: "A transcription activator-like effector toolbox for genome engineering", NATURE PROTOCOLS, NATURE PUBLISHING GROUP, GB, vol. 7, no. 1, 5 January 2012 (2012-01-05), pages 171 - 192, XP009170390, ISSN: 1750-2799

DESCRIPTION

Description

TECHNICAL FIELD

[0001] This invention relates to fusion proteins comprising a DNA binding domain, that binds specifically to a preselected target sequence that is within 250 bp of a CpG diresidue in a mammalian cell, wherein the DNA-binding domain is or comprises an engineered zinc finger, and a catalytic domain comprising a sequence that catalyzes demethylation of cytosines in DNA, wherein the sequence comprises amino acids 1580-2052 of Tet1 (SEQ ID NO: 2), wherein the fusion protein binds to the target sequence and reduces methylation of the CpG diresidue, *ex vivo* methods of use and such fusion proteins for use in treating cancer.

BACKGROUND

[0002] TET1 is an enzyme that catalyzes the conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) (Tahiliani, M. et al. Science 324, 930-935 (2009)). While the role of 5hmC is not entirely clear, it has been proposed that it may be an intermediate in the process of demethylation of 5-methylcytosine to cytosine. This is supported by evidence that overexpression of TET1 in cultured cells leads to an overall decrease in levels of 5mC (Tahiliani, M. et al. Science 324, 930-935 (2009)). Several mechanisms of demethylation have been proposed. There is evidence that 5hmC can be deaminated and that the resulting mismatched base is recognized by DNA glycosylases and subsequently repaired to cytosine via the base excision repair pathway (Guo et el., Cell 145, 423-434 (2011)). Alternatively, there is also evidence that iterative oxidation of 5hmC by TET1 yields 5-formylcytosine (fC) and 5-carboxylcytosine (caC), which can then be recognized by thymine DNA glycosylase and reverted to cytosine through base excision repair (He et al., Science 333, 1303-1307 (2011)). In either case, all evidence highlights TET1 as a primary catalyst for DNA demethylation.

[0003] WO2009/134409 discloses a fusion protein comprising a DNA-binding domain and a DNA cytosine demethylation domain.

[0004] WO2007/128982 discloses a fusion protein comprising a DNA-binding domain and a domain that exhibits a cytidine deaminase activity.

[0005] Gregory et al., Epigenetics, vol. 7, 20125, pages 344-349 discloses a fusion protein comprising a DNA-binding domain and a thymine-DNA-glycosylase (TDG) fused by a short glycine-rich linker.

[0006] WO2010/037001 discloses that TET1, TET2, TET3, and CXXC4 have hydrolase activity which converts the cytosine nucleotide 5-methylcytosine into 5-hydroxymethylcytosine.

[0007] Foley et al., Nature protocols, vol. 4, No. 12 pages 1855-1868 discloses zinc-finger expression vectors.

[0008] Neville, et al., Nature protocols, vol. 7, 2012, pages 171-192 discloses the design and use of TALE array fusion proteins, including TALE-transcription factors and TALENS.

[0009] Frauer, et al., PloS ONE, Volume 6, Issue 2, pages e16627 discloses the DNA binding activity of CXXC domains of mouse Dnmt1 and Tet1 proteins.

[0010] Rodenhiser et al,. CMAJ, 174(3), pages 341-348 discloses epigenetic mechanisms and their contribution to human health.

[0011] Branco et al., Nature Reviews | Genetics | Volume 13, pages 7-13 discloses the role of TET 1 in catalyzing the oxidation of 5-methylcytosine.

SUMMARY

[0012] The invention is defined in the claims. At least in part, the present invention is based on the discovery of fusion proteins comprised of an engineered custom DNA-binding domain fused to a protein capable of catalyzing hydroxylation of methylated cytosines in DNA in vitro, in cells, or in whole organisms. Such a hybrid protein would lead to formation of hydroxymethylated cytosines from methylated cytosines in DNA and subsequent demethylation of the cytosines in living cells or organisms.

[0013] Thus provided herein are fusion proteins comprising an engineered DNA-binding domain that binds specifically to a preselected target sequence that is within 250 bp of a CpG diresidue in a mammalian cell, wherein the DNA-binding domain is or comprises an engineered zinc finger, and a catalytic domain comprising a sequence that catalyzes demethylation of cytosines in DNA, i.e., that catalyzes hydroxylation or demethylation of methylated cytosines, wherein the sequence comprises amino acids 1580-2052 of Tet1 (SEQ ID NO: 2), wherein the fusion protein binds to the target sequence and reduces methylation of the CpG diresidue. In some embodiments, the fusion proteins further comprise a linker between the DNA binding domain and the catalytic domain.

[0014] In some embodiments, the linker is selected from the group consisting of Gly3Ser (SEQ ID NO:5), Gly3SerGly4SerArgSer (SEQ ID NO:6), Gly3Ser(Gly4Ser)2ArgSer (SEQ ID NO:7), Gly3Ser(Gly4Ser)3ArgSer (SEQ ID NO:8), Gly3Ser(Gly4Ser)4ArgSer (SEQ ID NO:9) and Gly3Ser(Gly4Ser)5ArgSer (SEQ ID NO: 10).

[0015] The DNA-binding domain may be or comprise an engineered transcription activator-like (TAL) effector repeat array or a zinc finger, the latter alternative forming part of the claimed invention.

[0016] In some embodiments, the fusion proteins comprise a plurality of catalytic domains.

[0017] In some embodiments, the fusion protein further comprises a nuclear localization domain/sequence.

[0018] In a further aspect, the invention provides *ex vivo* methods for reducing methylation of a selected DNA sequence in a mammalian cell, the method comprising contacting the cell with a fusion protein, as recited in the claims, comprising an engineered DNA-binding domain that binds specifically to a target sequence, wherein the target sequence is within 250 bp of the selected DNA sequence, and a catalytic domain comprising a sequence that catalyzes hydroxylation of methylated cytosines in DNA.

[0019] In another aspect, the invention provides *ex vivo* methods for reducing methylation of a selected DNA sequence in a mammalian cell, the method comprising contacting the cell with a nucleic acid, as recited in the claims, encoding a fusion protein comprising an engineered DNA-binding domain that binds specifically to a target sequence, wherein the target sequence is within 250 bp of the selected DNA sequence and a catalytic domain comprising a sequence that catalyzes hydroxylation of methylated cytosines in DNA.

[0020] In another aspect, the fusion protein or a nucleic acid encoding the fusion protein, as defined in the claims, for use in a method of reducing methylation of a CpG diresidue in a selected DNA sequence in a mammalian cell to treat cancer, the fusion protein comprising an engineered zinc finger that binds specifically to a target sequence that is within 250 bp of the CpG diresidue in the selected DNA sequence in the mammalian cell, and a catalytic domain comprising a sequence that catalyzes hydroxylation of methylated cytosine in DNA, the method comprising contacting the cell with the fusion protein or nucleic acid encoding the fusion protein,

[0021] In some embodiments, the catalytic domain comprises amino acids 1418-2136 of TET1.

[0022] In some embodiments, the cell is a human cell.

[0023] In some embodiments, the cell is in a living animal, e.g., a mammal or fish, e.g., zebrafish.

[0024] In some embodiments, the selected DNA sequence is a sequence of a gene listed in Figure 7.

[0025] In some embodiments, the DNA sequence is selected from a sequence of the RASSF1A, MGMT, BRCA1, p16, CDH1, pf6, TIMP-3, p15, p73, DAPK, APC, BRCA2, PAX3, hMLHI and RELN genes.

[0026] In some embodiments, the cell is present in an individual diagnosed with a disease listed in Fig. 7, and the selected DNA sequence is a gene associated with that disease as listed in Fig. 7.

[0027] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials are described herein for use in the present invention; other, suitable methods and materials known in the art can also be used. The materials, methods, and examples are illustrative only and not intended to be limiting. In case of conflict, the present specification, including definitions, will control.

[0028] Other features and advantages of the invention will be apparent from the following detailed description and figures, and from the claims.

DESCRIPTION OF DRAWINGS

[0029]

Figures 1A-C. Structure of TET1 and TAL-TET1 fusion proteins.

1. A) Schematic shows the predicted domain architecture of the TET1 protein (TET1FL), the catalytic domain of the TET1 protein (TET1CD) and the truncated version of TET1 in which the CXXC domain has been removed (TET1NZ). Features shown include the CXXC-type zinc-binding domain (CXXC), three nuclear localization signals (NLS), the cysteine-rich region (cys-rich) and the double-stranded β helix domain (DSBH) (Tahiliani et al., Science 2009). B) Schematic shows an exemplary TAL-TET1 fusion protein, in which TET1 is fused to the C-terminal end of the TALE protein via a Gly_3Ser linker. In some embodiments, the TET1 is fused to the N-terminal end. The engineered TALE protein binds DNA target sequence and thus localizes TET1 protein to this region. C) Demethylase activities of TAL-TET1 fusion proteins bearing various length inter-domain linkers. Graphs show the fraction of CpGs methylated (y-axis) for different positions along the length of the second KLF4 intron (x-axis, numbered relative to the start of the intron) in K562 cells transfected with plasmids expressing KL1 TALE-TET1 proteins bearing standard or variant linkers between the TALE and TET1 domains or a control plasmid expressing GFP. The arrow along the x-axis indicates the location and direction (5' to 3') of TALE target binding site. Each point represents data from a single transfection experiment. Methylation status was assessed using high-throughput bisulfite sequencing.

Figure 2. TAL Arrays Targeted to Human Klf4. Sequence of human Klf4 intron 1-2. Intron

sequences are shown in lowercase letters with neighboring exon sequences shown in uppercase letters. The four sites for which engineered customized TAL effector repeat array domains were generated are underlined and numbered. The forward sequence is SEQ ID NO: 1.

Figure 3. Targeted demethylation at the human KLF4 gene by TAL-TET1CD fusion proteins harboring a Gly_3Ser linker. Graphs show demethylation observed upon transfection of 4 different TAL-TET1CD fusion proteins into human K562 cells. Demethylation is calculated for each CpG by subtracting the %unmethylated CpGs in mock-transfected or GFP-transfected cells from the %unmethylated CpGs in cells in which TAL-TET1CD protein was expressed. Position of the TAL binding site is depicted with an arrow and direction of the arrow indicates N-terminal to C-terminal orientation of the protein. The specific TAL effector binding site targeted is indicated by a subscript, with TALi referring to binding site #1, etc. and corresponding to the numbering of sites shown in Figure 2. Where possible, error bars were calculated by determining the standard error from the means of the experimental and control samples.

Figure 4A. Targeted demethylation at the human KLF4 gene using TAL-TET1CD fusion proteins harboring more extended linker variants. Graphs show demethylation observed upon transfection of 5 different TAL-TET1CD fusion proteins into human K562 cells. Each fusion uses the same TAL DNA binding domain and TET1CD, but differs in the length and amino acid sequence used as a linker between the two domains. Demethylation is calculated for each CpG by subtracting the %unmethylated CpGs in mock-transfected or GFP-transfected cells from the %unmethylated CpGs in cells in which TAL-TET1CD protein was expressed. Position of the TAL binding site is depicted with an arrow and direction of the arrow indicates N-terminal to C-terminal orientation of the protein.

Figure 4B. Demethylation activities of TALE-TET1 CD (top row) and TALE-TET1 FL (bottom row) fusion proteins targeting first (left column) and second (right column) sites in human K562 cells. Graphs show the fraction of CpGs methylated (y-axis) for different positions along the length of the second *KLF4* intron (x-axis, numbered relative to the start of the intron). Horizontal arrow indicates the location and direction (5' to 3') of TALE target binding site. Each point represents the mean of three independent transfection experiments with bars indicating standard errors of the mean. Methylation status for each experiment was assessed using high-throughput bi-sulfite sequencing. Off-target TALE fusions bind to an unrelated sequence in the *EGFP* reporter gene. Note that the GFP control datapoints shown are the same for both graphs and are depicted twice for ease of comparison.

Figures 5A-C. Amino acid sequences of full length TET1 (TET1FL; 5A (SEQ ID NO:2)), TET1 catalytic domain (TET1CD; 5B (SEQ ID NO:3)) and TET1 protein with CXXC domain removed (TET1NZ; 5C (SEQ ID NO:4)).

Figures 6A-D. A set of schematic illustrations showing Strategies for TET1 Multimerization. A) Simple fusion of multiple TET1 domains to one TALE DNA binding domain in either amino-terminal orientation, carboxy-terminal orientation, or both. B) Non-covalent DNA interaction schemes, showing either ligand-dependent binding (inducible multimerization) or direct binding through protein-protein interacting domains. C) Covalent DNA interaction scheme showing split-inteins, here designated as Int-N on the amino-terminal protein and Int-C on the carboxy-terminal protein, before and after splicing. D) An example of the use of several multimerization approaches at once showing the use of both split-inteins as well as multimerization of TET1 domains in either amino- or carboxy-terminal orientation with respect to the TALE DNA binding protein.

Figure 7. List of human diseases and conditions associated with hypermethylation. Adapted from Table 1 of Rodenhiser and Mann, CMAJ 174(3):341-348 (2006). The references noted in the table are as follows:

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

Figures 8A-D. Targeted demethylation and activation of the human HBB gene by TALE-TET1CD Fusion Proteins. A) Schematic illustrating the human HBB locus with CpGs indicated with black arrows. Numbering indicates position on the DNA relative to the start site of transcription (right-angle arrow). Arrows indicate the location and direction (5' to 3') of 10 TALE-TET1CD fusion protein binding sites. B) Demethylation activities of HBB-1, -2, -3, -4, -5, and -6 TALE-TET1CD fusion proteins in K562 cells. Graphs show the fraction of CpGs methylated (y-axis) for different positions along the length of the HBB promoter (x-axis, numbered relative to the transcription start site) in K562 cells transfected with plasmids expressing one of the six HBB TALE-TET1CD fusion proteins (circles), an off-target TALE-TET1CD fusion to a site in the KLF4 gene (black triangles), or GFP (grey triangles). Each data point represents the mean of three independent transfection experiments with bars indicating standard errors of the mean. Methylation status for each experiment was assessed using high-throughput bi-sulfite sequencing. Note that the GFP and off-target TALE-TET1CD control data points shown are the same in both panels and are depicted twice for ease of comparison with experimental samples. Numbered arrows indicate the location and direction (5' to 3') of the various HB TALE-TET1 CD binding sites. C) Demethylation activities of HB-7, -8, -9, and -10 TALE-TET1CD fusion proteins in K562 cells. Data represented as in (B). D) Expression levels of HBB mRNA (normalized to β-actin mRNA levels) in K562 cells transfected with indicated TALE-TET1CD fusion protein expression plasmids or a control GFP expression plasmid. Means of three independent samples assayed by quantitative RT-PCR are shown with bars representing standard errors of the mean. Asterisks indicate samples with values significantly greater than those obtained with an off-target control TALE-TET1CD fusion protein (targeted to a site in the human KLF4 gene) as determined by a one-sided T-test ($P<0.05$).

Figures 9A-B. Engineered Zinc Finger-TET1 Fusion Proteins. A) ZF-TET1CD fusion protein-induced demethylation of the human KLF4 gene. Graphs show the fraction of CpGs methylated (y-axis) for different positions along the length of the second KLF4 intron (x-axis, numbered relative to the start of the intron) in K562 cells transfected with plasmids expressing one of two KLF4-targeted ZF-TET1CD fusions, an off-target ZF-TET1CD (targeted to an unrelated site in the human VEGFA gene), or GFP. Horizontal arrows indicate the location and direction (3' to 5') of the ZF target binding sites. Each point represents the mean of three independent transfection experiments with bars indicating standard errors of the mean. Methylation status for each experiment was assessed using high-throughput bi-sulfite sequencing. Note that the off-target ZF-TET1CD fusion and GFP control data shown are the same for both graphs and are depicted twice for ease of comparison. B) (top) Schematic illustrating the human HBB locus with CpGs indicated with black arrows. Numbering indicates position on the DNA relative to the start site of transcription (right-angle arrow). Colored arrows indicate the location and direction (3' to 5') of six ZF-TET1CD fusion protein binding sites. (bottom) Graphs show the fraction of CpGs methylated (y-axis) for different positions along the length of HBB gene (x-axis, numbered relative to HBB promoter transcription start site) in K562 cells transfected with plasmids expressing one of six HBB-targeted ZF-TET1CD fusions (circles correspond to arrows in (top)), an off-target ZF-TET1CD (targeted to an unrelated site in the human VEGFA gene) (black triangles), or GFP (grey triangles). Arrows indicate the location and direction (3' to 5') of the ZF target binding sites. Position of ZF-TET1CD binding sites are depicted with arrows corresponding to the circles on the graph. Data for ZF - TET 1 CD proteins targeted to the -266 and +83/+110 CpGs are shown in the left and right panels, respectively. Each point represents the mean of three independent transfection experiments with bars indicating standard errors of the mean. Methylation status for each experiment was assessed using high-throughput bi-sulfite sequencing.

Figures 10A-B. Extended time-course experiment following promoter CpG methylation and gene expression of human HBB in cells transfected with plasmids expressing TALE-TET1 fusion proteins. A) Expression levels of HBB mRNA (normalized to β-actin mRNA levels) in K562 cells transfected with plasmids expressing HB-4, HB-5, HB-6, or an off-target TALE-TET1 fusion protein or with a control GFP expression plasmid were measured 4, 7, 14, and 30 days post-transfection. (The off-target TALE-TET1 fusion is targeted to a site in the human KLF4 gene.) Means of three independent samples assayed by quantitative RT-PCR are shown with bars representing standard errors of the mean. B) Graphs showing the fraction of CpGs methylated (y-axis) for different positions along the length of the HBB promoter (x-axis, numbered relative to the transcription start site) for the same transfected cells described in (A). Each data point represents the mean of the three independent transfection experiments with bars indicating standard errors of the mean. Methylation status for each experiment was assessed using high-throughput bi-sulfite sequencing. Diamonds, day 4; squares, day 7; triangles, day 14; circles, day 30.

DETAILED DESCRIPTION

[0030] The invention is defined in the claims. Described herein are fusion proteins comprising a DNA-binding domain (i.e., an engineered custom DNA-binding domain), and one or more copies of a catalytic domain (from a different protein) comprising a sequence that catalyzes hydroxylation of methylated cytosines in DNA, with an optional linker between the domains, e.g., between the DNA-binding and catalytic domains, and/or between catalytic domains in fusion proteins that include multiple catalytic domains. As described herein, expression of a TAL effector repeat array-TET1 fusion protein in human K562 cells results in demethylation of CpGs in close proximity to the target site bound by the TAL effector repeat array part of the protein.

[0031] In an aspect of the disclosure is a hybrid protein consisting of an engineered transcription activator-like (TAL) effector repeat array fused to either the full length TET1 protein (TET1FL), the catalytic domain of TET1 (TET1CD), or a truncated version of the TET1FL protein in which the putative native DNA-binding domain has been removed (TET1NZ). DNA-binding specificity is defined by the engineered TAL effector repeat array. These DNA-binding proteins can be engineered to bind to essentially any DNA sequence and published work from various labs, as well as published and unpublished work from our lab, has demonstrated that these customizable domains can efficiently target a variety of fused domains to specific genomic locations (Reyon et al., FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol (2012).doi:10.1038/nbt.2170; Moscou and Bogdanove, Science 326, 1501-1501 (2009); Boch et al., Science 326, 1509-1512 (2009); Miller et al., Nat Biotechnol 29, 143-148 (2010)). For example, engineered TAL effector repeat arrays have been fused to the cleavage domain of the FokI endonuclease as well as activators and repressors and act to target these domains to a user-defined sequence within the context of the genome.

DNA-Binding Domains

[0032] The fusion proteins described herein can include any DNA Binding Domain (DBD) known in the art or engineered for a specific binding site. Exemplary DBDs include engineered or native TAL effector repeat arrays, engineered or native zinc fingers, homing meganucleases, chemical nucleases, and native DBDs.

TAL Effector Repeat Arrays

[0033] In aspects of the disclosure, TAL effectors of plant pathogenic bacteria in the genus Xanthomonas play important roles in disease, or trigger defense, by binding host DNA and activating effector-specific host genes. Specificity depends on an effector-variable number of imperfect, typically -33-35 amino acid repeats. Polymorphisms are present primarily at repeat positions 12 and 13, which are referred to herein as the repeat variable-diresidue (RVD). The RVDs of TAL effectors correspond to the nucleotides in their target sites in a direct, linear fashion, one RVD to one nucleotide, with some degeneracy and no apparent context dependence. In some embodiments, the polymorphic region that grants nucleotide specificity may be expressed as a triresidue or triplet (representing positions 11, 12, and 13 in the repeat).

[0034] Each DNA binding repeat can include a RVD that determines recognition of a base pair in the target DNA sequence, wherein each DNA binding repeat is responsible for recognizing one base pair in the target DNA sequence. In some embodiments, the RVD can comprise one or more of: HA for recognizing C; ND for recognizing C; HI for recognizing C; HN for recognizing G; NA for recognizing G; NH for recognizing G; SN for recognizing G or A; YG for recognizing T; and NK for recognizing G, and one or more of: HD for recognizing C; NG for recognizing T; NI for recognizing A; NN for recognizing G or A; NS for recognizing A or C or G or T; N* for recognizing C or T, wherein * represents a gap in the second position of the RVD; HG for recognizing T; H* for recognizing T, wherein * represents a gap in the second position of the RVD; and IG for recognizing T.

[0035] TALE proteins may be useful in research and biotechnology as targeted chimeric nucleases that can facilitate homologous recombination, homology-directed repair, or error-prone non-homologous end-joining repair for genome engineering (e.g., to add or enhance traits useful for biofuels or biorenewables in plants). These proteins also may be useful as, for example, transcription factors, and especially for therapeutic applications requiring a very high level of specificity such as therapeutics against pathogens (e.g., viruses) as non-limiting examples.

[0036] Methods for generating engineered TALE arrays are known in the art, see, e.g., the fast ligation-based automatable solid-phase high-throughput (FLASH) system described in USSN 61/610,212, and Reyon et al., Nature Biotechnology 30,460-465 (2012); as well as the methods described in Bogdanove & Voytas, Science 333, 1843-1846 (2011); Bogdanove et al., Curr Opin Plant Biol 13, 394-401 (2010); Scholze & Boch, J. Curr Opin Microbiol (2011); Boch et al., Science 326, 1509-1512 (2009); Moscou & Bogdanove, Science 326, 1501 (2009); Miller et al., Nat Biotechnol 29, 143-148 (2011); Morbitzer et al., T. Proc Natl Acad Sci U S A 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);

Hockemeye 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); and Zhang et al., Nat Biotechnol 29, 149-153 (2011).

Zinc Fingers

[0037] Zinc finger proteins are DNA-binding proteins that contain one or more zinc fingers, independently folded zinc-containing mini-domains, the structure of which is well known in the art and defined in, for example, Miller et al., 1985, EMBO J., 4:1609; Berg, 1988, Proc. Natl. Acad. Sci. USA, 85:99; Lee et al., 1989, Science. 245:635; and Klug, 1993, Gene, 135:83. Crystal structures of the zinc finger protein Zif268 and its variants bound to DNA show a semi-conserved pattern of interactions, in which typically three amino acids from the alpha-helix of the zinc finger contact three adjacent base pairs or a "subsite" in the DNA (Pavletich et al., 1991, Science, 252:809; Elrod-Erickson et al., 1998, Structure, 6:451). Thus, the crystal structure of Zif268 suggested that zinc finger DNA-binding domains might function in a modular manner with a one-to-one interaction between a zinc finger and a three-base-pair "subsite" in the DNA sequence. In naturally occurring zinc finger transcription factors, multiple zinc fingers are typically linked together in a tandem array to achieve sequence-specific recognition of a contiguous DNA sequence (Klug, 1993, Gene 135:83).

[0038] Multiple studies have shown that it is possible to artificially engineer the DNA binding characteristics of individual zinc fingers by randomizing the amino acids at the alpha-helical positions involved in DNA binding and using selection methodologies such as phage display to identify desired variants capable of binding to DNA target sites of interest (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). Such recombinant zinc finger proteins can be fused to functional domains, such as transcriptional activators, transcriptional repressors, methylation domains, and nucleases to regulate gene expression, alter DNA methylation, and introduce targeted alterations into genomes of model organisms, plants, and human cells (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).

[0039] Widespread adoption and large-scale use of zinc finger protein technology have been hindered by the continued lack of a robust, easy-to-use, and publicly available method for engineering zinc finger arrays. One existing approach, known as "modular assembly," advocates the simple joining together of pre-selected zinc finger modules into arrays (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). Although straightforward enough to be practiced by any researcher, recent reports have demonstrated a high failure rate for this method, particularly in the context of zinc finger nucleases (Ramirez et al., 2008, Nat. Methods, 5:374-375; Kim et al., 2009, Genome Res. 19:1279-88), a limitation that typically necessitates the construction and cell-based testing of very large numbers of zinc finger proteins for any given target gene (Kim et al., 2009, Genome Res. 19:1279-88).

[0040] Combinatorial selection-based methods that identify zinc finger arrays from randomized libraries have been shown to have higher success rates than modular assembly (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). In preferred embodiments, the zinc finger arrays are described in, or are generated as described in, WO 2011/017293 and WO 2004/099366. Additional suitable zinc finger DBDs are described in U.S. Pat. Nos. 6,511,808, 6,013,453, 6,007,988, and 6,503,717 and U.S. patent application 2002/0160940.

Native DBDs

[0041] In some aspects of the disclosure, a native DBD (e.g., a portion of a wild-type, non-engineered DNA binding protein that binds to a specific target sequence) can be used. For example, the DBD from a transcription factor, nuclease, histone, telomerase, or other DNA binding protein can be used. Typically DBDs include a structure that facilitates specific interaction with a target nucleic acid sequence; common DBD structures include helix-turn-helix; zinc finger; leucine zipper; winged helix; winged helix turn helix; helix-loop-helix; and hmg-box. The native DBD can be from any organism. See, e.g., Kummerfeld & Teichmann, Nucleic Acids Res. 34 (Database issue): D74-81 (2006). The residues in a DNA binding protein that contact DNA, and thus form part of the DBD, can be determined empirically or predicted computationally, e.g., as described in Tjong and Zhou, Nucl. Acids Res. 35:1465-1477 (2007). A database of DNA binding proteins can be used to identify DNA binding proteins and DBDs for use in the present compositions and methods; see, e.g., Harrison, Nature, 353, 715-719 (1991); Karmirantzou and 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).

[0042] Where a native DBD is used in a fusion protein described herein, the catalytic domain is from a different protein.

Homing Meganucleases

[0043] Meganucleases are sequence-specific endonucleases originating from a variety of organisms such as bacteria, yeast, algae and plant organelles. Endogenous meganucleases have recognition sites of 12 to 30 base pairs; customized DNA binding sites with 18bp and 24bp-long meganuclease recognition sites have been described, and either can be used in the present methods and constructs. See, e.g., 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); and Stoddard, Q. Rev. Biophys. 38, 49

(2005); Grizot et al., Nucleic Acids Research, 38:2006-18 (2010).

Chemical Nucleases

[0044] DNA binding domains from the so-called "chemical nucleases,"(Pingoud and Silva, Nat Biotechnol. 25:743-4 (2007)), e.g., triplex-forming oligonucleotides or peptide nucleic acids can also be utilized in the present compositions and methods; see, e.g., 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; or Eisenschmidt et al., Nucleic Acids Res. 2005; 33(22):7039-47.

Catalytic Domains

[0045] The fusion proteins include a catalytic domain comprising a sequence that catalyzes hydroxylation of methylated cytosines in DNA. Exemplary proteins include the Ten-Eleven-Translocation (TET)1-3 family, enzymes that converts 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC) in DNA.

[0046] Sequences for human TET1-3 are known in the art and are shown in the following table:

	GenBank Accession Nos.	
Gene	Amino Acid	Nucleic Acid
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
* Variant (1) represents the longer transcript and encodes the longer isoform (a). Variant (2) differs in the 5' UTR and in the 3' UTR and coding sequence compared to variant 1. The resulting isoform (b) is shorter and has a distinct C-terminus compared to isoform a.		

[0047] In some embodiments, all or part of the full-length sequence of the catalytic domain can be included, e.g., a catalytic module comprising the cysteine-rich extension and the 2OGFeDO domain encoded by 7 highly conserved exons, e.g., the TET1 catalytic domain comprising amino acids 1580-2052, TET2 comprising amino acids 1290-1905 and TET3 comprising amino acids 966-1678. See, e.g., Fig. 1 of Iyer et al., Cell Cycle. 2009 Jun 1;8(11):1698-710. Epub 2009 Jun 27, for an alignment illustrating the key catalytic residues in all three Tet proteins, and the supplementary materials thereof (available at ftp site ftp.ncbi.nih.gov/pub/aravind/DONS/supplementary_material_DONS.html) for full length sequences (see, e.g., seq 2c); in some embodiments, the sequence includes amino acids 1418-2136 of TET1 or the corresponding region in TET2/3.

[0048] Other catalytic modules can be from the proteins identified in Iyer et al., 2009.

Construction of Fusion Proteins

[0049] To generate a functional recombinant protein, the DNA binding domain is fused to at least one catalytic domain. Fusing catalytic domains to DBD to form functional fusion proteins involves only routine molecular biology techniques that are commonly practiced by those of skill in the art, see for example, U.S. Pat. Nos. 6,511,808, 6,013,453, 6,007,988, 6,503,717 and U.S. patent application 2002/0160940). Catalytic domains can be associated with the DBD domain at any suitable position, including the C- or N-terminus of the DBD. In some embodiments, the fusion proteins include multiple catalytic domains, e.g., on one or both ends of the DBD, e.g., concatenated together with an optional intervening linker; thus there can be one or more catalytic domains on each end of the DBD.

[0050] Alternatively, the catalytic domains, e.g., TET1 units, could be multimerized through specific constitutive or inducible protein-protein interaction domains (such as leucine zipper domains or ClonTech's iDimerize system, homodimerization and heterodimerization systems and ligands (e.g. AP20187, AP21967) which were previously provided by ARIAD under the brand name ARGENT. The B/B Homodimerizer (AP20187) induces dimerization of two proteins that each contain the DmrB homodimerization domain (see table below explaining domain nomenclature). The A/C Heterodimerizer (AP21967) induces dimerization of a protein possessing the DmrA domain and a second protein containing the DmrC domain. The D/D Solubilizer (alternative to AP21998) induces dissociation/disaggregation of proteins possessing DmrD domains. DmrD causes automatic self-association of proteins fused to it; see, e.g., 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 and 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)), Alternatively, multimerization could be achieved through the use of split-inteins, a class of autocatyltic intein peptides that allow for the seamless covalent splicing of two separate proteins in a predictable and efficient manner (d'Avignon, et al., Biopolymers. 2006 Oct 15;83(3):255-67; Zitzewitz, et al., Biochemistry. 1995 Oct 3;34(39):12812-9; Li et al., Hum Gene Ther. 2008 Sep;19(9):958-64). Both the protein-protein interaction and intein approaches could be optimized to produce very long multimerized strings of catalytic domains. Figures6A-D show exemplary schemes for multimerization.

[0051] Linkers that can be used in these fusion proteins (or between fusion proteins in a concatenated structure) can include any sequence that does not interfere with the function of the fusion proteins. In preferred embodiments, the linkers are short, e.g., 2-20 amino acids, and are typically flexible (i.e., comprising amino acids with a high degree of freedom such as glycine, alanine, and serine). In some embodiments, the linker comprises one or more units consisting of GGGS (SEQ ID NO:17) or GGGGS (SEQ ID NO:18), e.g., two, three, four, or more repeats of the GGGS (SEQ ID NO:17) or GGGGS (SEQ ID NO:18) unit. Other linker sequences can also be used.

Use of the Fusion Proteins

[0052] The fusion proteins designed and produced using the methods described herein can be used to perform any function where it is desired to demethylate a specific DNA sequence. The fusion proteins can be useful for the treatment of disease; for example, the fusion proteins can be targeted to a hypermethylated region of a gene listed in Table 1 of Rodenhiser and Mann, CMAJ 174(3):341-348 (2006) as hypermethylated in a disease, which is reproduced as FIG. 7 hereto, or a reference set forth therein, or for otherwise altering the structure or function of a given gene in vivo. The fusion proteins of the present invention are also useful as research tools, for example, in performing either in vivo or in vitro functional genomics studies (see, for example, U.S. Pat. No. 6,503,717 and U.S. patent application 2002/0164575).

Polypeptide Expression Systems

[0053] In order to use the fusion proteins described, it may be desirable to express the engineered proteins from a nucleic acid that encodes them. This can be performed in a variety of ways. For example, the nucleic acid encoding the fusion protein can be cloned into an intermediate vector for transformation into prokaryotic or eukaryotic cells for replication and/or expression. Intermediate vectors are typically prokaryote vectors, e.g., plasmids, or shuttle vectors, or insect vectors, for storage or manipulation of the nucleic acid encoding the fusion protein or for production of the fusion protein. The nucleic acid encoding the fusion protein can also be cloned into an expression vector, for administration to a plant cell, animal cell, preferably a mammalian cell or a human cell, fungal cell, bacterial cell, or protozoan cell.

[0054] To obtain expression, the fusion protein is typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, e.g., in Sambrook et al., Molecular Cloning, A Laboratory Manual (3d ed. 2001); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); and Current Protocols in Molecular Biology (Ausubel et al., eds., 2010). Bacterial expression systems for expressing the engineered TALE repeat protein are available in, e.g., *E. coli, Bacillus* sp., and *Salmonella* (Palva et al., 1983, Gene 22:229-235). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available.

[0055] The promoter used to direct expression of the fusion protein nucleic acid depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification of fusion proteins. In contrast, when the fusion protein is to be administered in vivo for gene regulation, either a constitutive or an inducible promoter can be used, depending on the particular use of the fusion protein. In addition, a preferred promoter for administration of the fusion protein can be a weak promoter, such as HSV TK or a promoter having similar activity. The promoter can also include elements that are responsive to transactivation, e.g., hypoxia response elements, Gal4 response elements, lac repressor response element, and small molecule control systems such as tetracycline-regulated systems and the RU-486 system (see, e.g., Gossen & 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; and Rendahl et al., 1998, Nat. Biotechnol., 16:757-761).

[0056] In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter operably linked, e.g., to the nucleic acid sequence encoding the fusion protein, and any signals required, e.g., for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding sites, or translation termination. Additional elements of the cassette may include, e.g., enhancers, and heterologous spliced intronic signals.

[0057] The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the fusion protein, e.g., expression in plants, animals, bacteria, fungus, protozoa, etc. Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and commercially available tag-fusion expression systems such as GST and LacZ. A preferred tag-fusion protein is the maltose binding protein, "MBP." Such tag-fusion proteins can be used for purification of the engineered TALE repeat protein. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, for monitoring expression, and for monitoring cellular and subcellular localization, e.g., c-myc or FLAG.

[0058] Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

[0059] Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High yield expression systems are also suitable, such as using a baculovirus vector in insect cells, with the fusion protein encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

[0060] The elements that are typically included in expression vectors also include a replicon that functions in *E. coli,* a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of recombinant sequences.

[0061] Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of protein, which are then purified using standard techniques (see, e.g., Colley et al., 1989, J. Biol. Chem., 264:17619-22; Guide to Protein Purification, in Methods in Enzymology, vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques (see, e.g., Morrison, 1977, J. Bacteriol. 132:349-351; Clark-Curtiss & Curtiss, Methods in Enzymology 101:347-362 (Wu et al., eds, 1983).

[0062] Any of the known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, nucleofection, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the protein of choice.

[0063] In some embodiments, the fusion protein includes a nuclear localization domain which provides for the protein to be translocated to the nucleus. Several nuclear localization sequences (NLS) are known, and any suitable NLS can be used. For example, many NLSs have a plurality of basic amino acids, referred to as a bipartite basic repeats (reviewed in Garcia-Bustos et al, 1991, Biochim. Biophys. Acta, 1071:83-101). An NLS containing bipartite basic repeats can be placed in any portion of chimeric protein and results in the chimeric protein being localized inside the nucleus. In preferred embodiments a nuclear localization domain is incorporated into the final fusion protein, as the ultimate functions of the fusion proteins described herein will typically require the proteins to be localized in the nucleus. However, it may not be necessary to add a separate nuclear localization domain in cases where the DBD domain itself, or another functional domain within the final chimeric protein, has intrinsic nuclear translocation function.

Use of Fusion Proteins in Gene Therapy

[0064] The fusion proteins described herein can be used to regulate gene expression or alter gene sequence in gene therapy applications in the same. See for example U.S. Pat. No. 6,511,808, U.S. Pat. No. 6,013,453, U.S. Pat. No. 6,007,988, U.S. Pat. No. 6,503,717, U.S. patent application 2002/0164575, and U.S. patent application 2002/0160940. The methods can include administering one or more of the fusion proteins described herein targeted to one or more genes. Since methylation of many CpGs across hundreds of basepairs of DNA in promoters or imprinted regions can influence gene expression (Han et al., Mol Cell Biol. 2008 Feb;28(3):1124-35, Mabaera, et al., Blood. 2007 Aug 15;110(4):1343-52), it may be desirable to reduce methylation across longer sequences. If a larger region of a gene is desired to be demethylated, a plurality of fusion proteins that all target positions on the same gene, e.g., positions within about 1000, 500, 300, 250, 150, 100, 50, 40, 30, or 20 bp of the site to be demethylated, can be administered. Alternatively or in addition, one or a plurality of fusion proteins that are multimerized as described herein (see, e.g., Figs.6A-D) can be administered.

[0065] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding the fusion protein into mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding fusion proteins to cells in vitro. Preferably, the nucleic acids encoding the fusion proteins are administered for in vivo or ex vivo gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, 1992, Science, 256:808-813; Nabel & Felgner, 1993, TIBTECH, 11:211-217; Mitani & 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 & Perricaudet, 1995, Br. Med. Bull., 51:31-44; Haddada et al., in Current Topics in Microbiology and Immunology Doerfler and Bohm (eds) (1995); and Yu et al., 1994, Gene Ther., 1:13-26.

[0066] Methods of non-viral delivery of nucleic acids encoding the fusion proteins include lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA or RNA, artificial virions, and agent-enhanced uptake of DNA or RNA. Lipofection is described in e.g., U.S. Pat. No. 5,049,386, No. 4,946,787; and No.4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024. Delivery can be to cells (ex vivo administration) or target tissues (in vivo administration).

[0067] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., 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; U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[0068] The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding the fusion proteins takes advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro and the modified cells are administered to patients (ex vivo). Conventional viral based systems for the delivery of fusion proteins could include retroviral, lentivirus, adenoviral, adeno-associated, Sendai, and herpes simplex virus vectors for gene transfer. Viral vectors are currently the most efficient and versatile method of gene transfer in target cells and tissues. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0069] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., 1992, J. Virol., 66:2731-39; Johann et al., 1992, J. Virol., 66:1635-40; Sommerfelt et al., 1990, Virololgy, 176:58-59; Wilson et al., 1989, J. Virol., 63:2374-78; Miller et al., 1991, J. Virol., 65:2220-24; WO 94/26877).

[0070] In applications where transient expression of the fusion protein is preferred, adenoviral based systems can be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., 1987, Virology 160:38-47; U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, 1994, Hum. Gene Ther., 5:793-801; Muzyczka, 1994, J. Clin. Invest., 94:1351). Construction of recombinantAAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., 1985, Mol. Cell. Biol. 5:3251-60; Tratschin et al.,1984, Mol. Cell. Biol., 4:2072-81; Hermonat & Muzyczka, 1984, Proc. Natl. Acad. Sci. USA, 81:6466-70; and Samulski et al., 1989, J. Virol., 63:3822-28.

[0071] In particular, at least six viral vector approaches are currently available for gene transfer in clinical trials, with retroviral vectors by far the most frequently used system. All of these viral vectors utilize approaches that involve complementation of defective vectors by genes inserted into helper cell lines to generate the transducing agent.

[0072] pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials (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 was the first therapeutic vector used in a gene therapy trial. (Blaese et al., 1995, Science, 270:475-480). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors (Ellem et al., 1997, Immunol Immunother., 44:10-20; Dranoffet al., 1997, Hum. Gene Ther., 1:111-112).

[0073] Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. Typically, the vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system (Wagner et al., 1998, Lancet, 351:1702-1703; Kearns et al., 1996, Gene Ther., 9:748-55).

[0074] Replication-deficient recombinant adenoviral vectors (Ad) are predominantly used for colon cancer gene therapy, because they can be produced at high titer and they readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the Ad E1a, E1b, and E3 genes; subsequently the replication defector vector is propagated in human 293 cells that supply deleted gene function in trans. Ad vectors can transduce multiple types of tissues in vivo, including nondividing, differentiated cells such as those found in the liver, kidney and muscle system tissues. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection (Sterman et al., 1998, Hum. Gene Ther. 7:1083-89). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker et al., 1996, Infection, 24:15-10; Sterman 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; Sterman et al., 1998, Hum. Gene Ther., 7:1083-89.

[0075] Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and Ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

[0076] In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. A viral vector is typically modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the viruses outer surface. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han et al., 1995, Proc. Natl. Acad. Sci. USA, 92:9747-51, reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, filamentous phage can be engineered to display antibody fragments (e.g., Fab or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

[0077] Gene therapy vectors can be delivered in vivo by administration to an individual patient, typically by systemic administration (e.g., intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells ex vivo, such as cells explanted from an individual patient (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or stem cells (e.g., universal donor hematopoietic stem cells, embryonic stem cells (ES), partially differentiated stem cells, non-pluripotent stem cells, pluripotent stem cells, induced pluripotent stem cells (iPS cells) (see e.g., Sipione et al., Diabetologia, 47:499-508, 2004)), followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

[0078] Ex vivo cell transfection for diagnostics, research, or for gene therapy (e.g., via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In a preferred embodiment, cells are isolated from the subject organism, transfected with nucleic acid (gene or cDNA), encoding the fusion protein, and re-infused back into the subject organism (e.g., patient). Various cell types suitable for ex vivo transfection are well known to those of skill in the art (see, e.g., Freshney et al., Culture of Animal Cells, A Manual of Basic Technique (5th ed. 2005)) and the references cited therein for a discussion of how to isolate and culture cells from patients).

[0079] In one embodiment, stem cells (e.g., universal donor hematopoietic stem cells, embryonic stem cells (ES), partially differentiated stem cells, non-pluripotent stem cells, pluripotent stem cells, induced pluripotent stem cells (iPS cells) (see e.g., Sipione et al., Diabetologia, 47:499-508, 2004)) are used in ex vivo procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types in vitro, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ cells in vitro into clinically important immune cell types using cytokines such a GM-CSF, IFN-gamma and TNF-alpha are known (see Inaba et al., 1992, J. Exp. Med., 176:1693-1702).

[0080] Stem cells can be isolated for transduction and differentiation using known methods. For example, stem cells can be isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panB cells), GR-1 (granulocytes), and 1ad (differentiated antigen presenting cells) (see Inaba et al., 1992, J. Exp. Med., 176:1693-1702).

[0081] Vectors (e.g., retroviruses, adenoviruses, liposomes, etc.) containing nucleic acids encoding the fusion protein can be also administered directly to the organism for transduction of cells in vivo. Alternatively, naked DNA or RNA can be administered. The DNA or RNA can be modified to improve its properties, e.g., to increase half-life or to reduce immunologic response (see, e.g., US20130123481; 20130115272; and 20120251618). Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route. Alternatively, stable formulations of the fusion protein can also be administered.

[0082] Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions available, as described below (see, e.g., Remington: The Science and Practice of Pharmacy, 21st ed., 2005).

Delivery Vehicles

[0083] An important factor in the administration of polypeptide compounds, such as the fusion proteins of the present invention, is ensuring that the polypeptide has the ability to traverse the plasma membrane of a cell, or the membrane of an intra-cellular compartment such as the nucleus. Cellular membranes are composed of lipid-protein bilayers that are freely permeable to small, nonionic lipophilic compounds and are inherently impermeable to polar compounds, macromolecules, and therapeutic or diagnostic agents. However, proteins and other compounds such as liposomes have been described, which have the ability to translocate polypeptides such as fusion protein across a cell membrane.

[0084] For example, "membrane translocation polypeptides" have amphiphilic or hydrophobic amino acid subsequences that have the ability to act as membrane-translocating carriers. In one embodiment, homeodomain proteins have the ability to translocate across cell membranes. The shortest internalizable peptide of a homeodomain protein, Antennapedia, was found to be the third helix of the protein, from amino acid position 43 to 58 (see, e.g., Prochiantz, 1996, Curr. Opin. Neurobiol., 6:629-634). Another subsequence, the h (hydrophobic) domain of signal peptides, was found to have similar cell membrane translocation characteristics (see, e.g., Lin et al., 1995, J. Biol. Chem., 270:14255-58).

[0085] Examples of peptide sequences that can be linked to a protein, for facilitating uptake of the protein into cells, include, but are not limited to: peptide fragments of the tat protein of HIV (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); a 20 residue peptide sequence which corresponds to amino acids 84-103 of the p16 protein (see Fahraeus et al., 1996, Curr. Biol., 6:84); the third helix of the 60-amino acid long homeodomain of Antennapedia (Derossi et al., 1994, J. Biol. Chem., 269:10444); the h region of a signal peptide, such as the Kaposi fibroblast growth factor (K-FGF) h region (Lin et al., supra); the VP22 translocation domain from HSV (Elliot & O'Hare, 1997, Cell, 88:223-233); or supercharged proteins or intraphilins, e.g., as described in US20120100569; US20110112040; 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; and Lawrence et al (2007) J. Am. Chem. Soc. 129, 10110. See also, e.g., 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; and Deshayes et al., 2005, Cell. Mol. Life Sci., 62:1839-49. Other suitable chemical moieties that provide enhanced cellular uptake may also be chemically linked to the Fusion proteins described herein.

[0086] Toxin molecules also have the ability to transport polypeptides across cell membranes. Often, such molecules are composed of at least two parts (called "binary toxins"): a translocation or binding domain or polypeptide and a separate toxin domain or polypeptide. Typically, the translocation domain or polypeptide binds to a cellular receptor, and then the toxin is transported into the cell. Several bacterial toxins, including Clostridium perfringens iota toxin, diphtheria toxin (DT), Pseudomonas exotoxin A (PE), pertussis toxin (PT), Bacillus anthracis toxin, and pertussis adenylate cyclase (CYA), have been used in attempts to deliver peptides to the cell cytosol as internal or amino-terminal fusions (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; and Novak et al., 1992, J. Biol. Chem., 267:17186-93).

[0087] Such subsequences can be used to translocate fusion proteins across a cell membrane. The fusion proteins can be conveniently fused to or derivatized with such sequences. Typically, the translocation sequence is provided as part of a fusion protein. Optionally, a linker can be used to link the fusion protein and the translocation sequence. Any suitable linker can be used, e.g., a peptide linker.

[0088] The fusion protein can also be introduced into an animal cell, preferably a mammalian cell, via liposomes and liposome derivatives such as immunoliposomes. The term "liposome" refers to vesicles comprised of one or more concentrically ordered lipid bilayers, which encapsulate an aqueous phase. The aqueous phase typically contains the compound to be delivered to the cell, i.e., the fusion protein.

[0089] The liposome fuses with the plasma membrane, thereby releasing the compound into the cytosol. Alternatively, the liposome is phagocytosed or taken up by the cell in a transport vesicle. Once in the endosome or phagosome, the liposome either degrades or fuses with the membrane of the transport vesicle and releases its contents.

[0090] In current methods of drug delivery via liposomes, the liposome ultimately becomes permeable and releases the encapsulated compound (e.g., the fusion protein or a nucleic acid encoding the same) at the target tissue or cell. For systemic or tissue specific delivery, this can be accomplished, for example, in a passive manner wherein the liposome bilayer degrades over time through the action of various agents in the body. Alternatively, active compound release involves using an agent to induce a permeability change in the liposome vesicle. Liposome membranes can be constructed so that they become destabilized when the environment becomes acidic near the liposome membrane (see, e.g., Proc. Natl. Acad. Sci. USA, 84:7851 (1987); Biochemistry, 28:908 (1989)). When liposomes are endocytosed by a target cell, for example, they become destabilized and release their contents. This destabilization is termed fusogenesis. Dioleoylphosphatidylethanolamine (DOPE) is the basis of many "fusogenic" systems.

[0091] Such liposomes typically comprise the fusion protein and a lipid component, e.g., a neutral and/or cationic lipid, optionally including a receptor-recognition molecule such as an antibody that binds to a predetermined cell surface receptor or ligand (e.g., an antigen). A variety of methods are available for preparing liposomes as described in, e.g., Szoka et al., 1980, Annu. Rev. Biophys. Bioeng., 9:467, U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,946,787, PCT Publication. No. WO 91/17424, Deamer & 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, Chapter 1); Hope et al., 1986, Chem. Phys. Lip., 40:89; Gregoriadis, Liposome Technology (1984) and Lasic, Liposomes: from Physics to Applications (1993)). Suitable methods include, for example, sonication, extrusion, high pressure/homogenization, microfluidization, detergent dialysis, calcium-induced fusion of small liposome vesicles and ether-fusion methods, all of which are well known in the art.

[0092] In certain embodiments, it is desirable to target liposomes using targeting moieties that are specific to a particular cell type, tissue, and the like. Targeting of liposomes using a variety of targeting moieties (e.g., ligands, receptors, and monoclonal antibodies) has been previously described (see, e.g., U.S. Pat. Nos. 4,957,773 and 4,603,044).

[0093] Examples of targeting moieties include monoclonal antibodies specific to antigens associated with neoplasms, such as prostate cancer specific antigen and MAGE. Tumors can also be diagnosed by detecting gene products resulting from the activation or over-expression of oncogenes, such as ras or c-erbB2. In addition, many tumors express antigens normally expressed by fetal tissue, such as the alphafetoprotein (AFP) and carcinoembryonic antigen (CEA). Sites of viral infection can be diagnosed using various viral antigens such as hepatitis B core and surface antigens (HBVc, HBVs) hepatitis C antigens, Epstein-Barr virus antigens, human immunodeficiency type-1 virus (HIV1) and papilloma virus antigens. Inflammation can be detected using molecules specifically recognized by surface molecules which are expressed at sites of inflammation such as integrins (e.g., VCAM-1), selectin receptors (e.g., ELAM-1) and the like.

[0094] Standard methods for coupling targeting agents to liposomes can be used. These methods generally involve incorporation into liposomes lipid components, e.g., phosphatidylethanolamine, which can be activated for attachment of targeting agents, or derivatized lipophilic compounds, such as lipid derivatized bleomycin. Antibody targeted liposomes can be constructed using, for instance, liposomes which incorporate protein A (see Renneisen et al., 1990, J. Biol. Chem., 265:16337-42 and Leonetti et al., 1990, Proc. Natl. Acad. Sci. USA, 87:2448-51).

Dosages

[0095] For therapeutic applications, the dose of the fusion protein to be administered to a patient can be calculated in a similar way as has been described for zinc finger proteins, see for example U.S. Pat. No. 6,511,808, U.S. Pat. No. 6,492,117, U.S. Pat. No. 6,453,242, U.S. patent application 2002/0164575, and U.S. patent application 2002/0160940. In the context of the present disclosure, the dose should be sufficient to effect a beneficial therapeutic response in the patient over time. In addition, particular dosage regimens can be useful for determining phenotypic changes in an experimental setting, e.g., in functional genomics studies, and in cell or animal models. The dose will be determined by the efficacy, specificity, and K_D of the particular fusion protein employed, the nuclear volume of the target cell, and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular patient.

Pharmaceutical Compositions and Administration

[0096] Appropriate pharmaceutical compositions for administration of the fusion proteins of the present invention can be determined as described for zinc finger proteins, see for example U.S. Pat. No. 6,511,808, U.S. Pat. No. 6,492,117, U.S. Pat. No. 6,453,242, U.S. patent application 2002/0164575, and U.S. patent application 2002/0160940. Fusion proteins, and expression vectors encoding fusion proteins, can be administered directly to the patient for modulation of methylation patterns, e.g., and gene expression, and for therapeutic or prophylactic applications, for example, for treatment of diseases listed in Fig. 7 as associated with hypermethylation, including cancer (e.g., bladder, brain (e.g., glioma, or glioblastoma), breast, cervical, colon, colorectal, esophagus, head/neck, kidney, leukemia, liver, lung, lymphoma, myeloma, ovary, pancreas, prostate, rhabdomyosarcoma, and uterus cancer); schizophrenia; memory formation; and atherosclerosis. Thus the methods can include identifying a subject who has a disease associated with hypermethylation (e.g., optionally including obtaining a sample and detecting methylation of DNA, e.g., of the associated gene listed in Fig. 7, and selecting the subject if their DNA is hypermethylated), and administering a therapeutically effective amount of a fusion protein, or a nucleic acid encoding a fusion protein, as described herein, to the subject.

[0097] Administration of therapeutically effective amounts is by any of the routes normally used for introducing Fusion proteins into ultimate contact with the tissue to be treated. The fusion proteins are administered in any suitable manner, preferably with pharmaceutically acceptable carriers. Suitable methods of administering such modulators are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0098] Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions that are available (see, e.g., Remington: The Science and Practice of Pharmacy, 21st ed., 2005).

[0099] The fusion proteins, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

[0100] Formulations suitable for parenteral administration, such as, for example, by intravenous, intramuscular, intradermal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. The disclosed compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials. Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described.

EXAMPLES

[0101] The invention is further described in the following examples, which do not limit the scope of the invention described in the claims. Any examples not falling within then scope of the claims are provided for comparison purposes only.

Example 1. Generation of TAL-TET1 Fusion Proteins to Specific Genomic Loci in the Endogenous Human KLF4 Gene

[0102] Plasmids expressing three different versions of the human TET1 protein were constructed based on the predicted domain architecture described by Tahiliani et al. (Science 324, 930-935 (2009)) (Figure 1A).

[0103] The full TET1 coding sequence was synthesized as gBlocks (Integrated DNA Technologies) and assembled by standard restriction enzyme digest and ligation to construct TET1FL and TET1CD expression vectors. All TALEs were assembled using the FLASH method and were cloned into TALE-TET1 expression vectors (pJA344C7, pJA345D4, pJA344E9 and pJA247) containing one of four 0.5 C-terminal TALE repeats, an N-terminal nuclear localization signal, the Δ152 TALE N-terminal domain and the +95 TALE C-terminal domain as previously described (Miller et al., Nat Biotechnol 29, 143-148 (2010); Reyon et al., Nature Biotechnology 30:460-465 (2012)). In all expression plasmids used, either TET1FL or TET1CD was fused to the C-terminal end of the TALE-derived DNA-binding domain via a GlyGlyGlySer (SEQ ID NO:18) linker and expression of this fusion was driven by an EF1alpha promoter. For the Western blot experiments, expression vectors had an additional triple-flag tag cloned upstream of the nuclear localization signal.

[0104] TET1FL is the full length TET1 protein, while TET1CD comprises only the catalytic domain of the protein, including the Cys-rich region and the double-stranded β-helix domain (Figure 5). TET1NZ is an intermediate truncation variant that lacks a CXXC-type zinc-binding domain, which may be responsible for directing DNA-binding of the endogenous protein (Zhang et al., Cell Res 20, 1390-1393 (2010); Xu et al., Mol. Cell 42, 451-464 (2011); Figure 5). Note that TET1NZ maintains more of the native protein than TET1CD alone. Briefly, DNA fragments encoding full-length TET1 (TET1FL; amino acids 1-2136) and TET1 catalytic domain (TET1CD; amino acids 1418-2136) were synthesized in 8 pieces (Integrated DNA Technologies) and iteratively cloned together using standard restriction digest/ligation based cloning.

[0105] These fragments encoding TET1FL and TET1CD were then cloned into a plasmid vector that enables their expression as fusions to the carboxy-terminal end of four different engineered TAL repeat array protein domains designed to bind different 18 bp target sites within the second intron of the human Klf4 gene (see Figure 2). The linker sequences used to join the TET1 protein or TET1 protein fragments to the TAL effector repeat arrays were composed of one of the following linkers: Gly_3Ser (SEQ ID NO:5), $Gly_3SerGly_4SerArgSer$ (SEQ

ID NO:6), $Gly_3Ser(Gly_4Ser)_2ArgSer$ (SEQ ID NO:7), $Gly_3Ser(Gly_4Ser)_3ArgSer$ (SEQ ID NO:8), $Gly_3Ser(Gly_4Ser)_4ArgSer$ (SEQ ID NO:9), or $Gly_3Ser(Gly_4Ser)_5ArgSer$ (SEQ ID NO:10). Lengthening the linker between the TALE repeat array and the TET1 CD did not appreciably alter demethylation efficiencies observed (Figure 1C). Based on these results, all additional fusion proteins described in this report were constructed using the TALE-TET1CD architecture with a short GGGS linker.

[0106] The final expression plasmid placed the fusion proteins under the transcriptional control of a constitutively active EF1α promoter. TET1NZ was constructed by removing the CXXC domain located at amino acids 584-624 of the full length protein. This was achieved by using tailed PCR of a small segment of TET1FL that placed a BamHI site immediately after the CXXC domain and included a PmlI site upstream of the CXXC domain that is unique in the TET1FL expression plasmid. This PCR product was then cloned into the TET1FL expression plasmid using a BamHI/PmlI double digest of both the PCR product and the TET1FL plasmid followed by ligation and transformation.

[0107] The four genomic loci chosen to target for demethylation (and for which TAL effector repeat arrays were designed) were selected based on several criteria. All four loci lie within intron 1-2 of the human KLF4 gene, which exhibits marks of open chromatin based on data from the ENCODE project (as accessed through the UCSC Genome Browser), suggesting its likely accessibility for binding by TAL effector DNA binding domains. This intron was also attractive because it harbors multiple CpG diresidues that are methylated in K562 cells but that are unmethylated in human embyronic stem cells, suggesting that methylation of these cytosines is not necessary for cell survival or viability.

[0108] TAL effector repeat arrays were engineered using the previously described FLASH assembly method (Reyon et al., Nature Biotechnology 30:460-465(2012)). DNA fragments encoding assembled arrays were cloned into expression constructs containing amino acids 153 to 288 from the TALE13 protein and the +95 C-terminal domain, comprised of amino acids 715 to 809 from the C-terminal end of TALE13 (Miller et al., Nat Biotechnol 29, 143-148 (2010)).

Example 2. Site-Specific Demethylation of CpGs in the Human KLF4 Locus Mediated by TAL-TET1 Fusion Proteins

[0109] The abilities of these proteins to demethylate adjacent CpGs were then tested in human K562 cells. The K562 cells were cultured in RPMI 1640 medium (Gibco) with 10% HI-FBS (Gibco), 1% Glutamax (Gibco) and 1% Pen/Strep (Gibco), and were transfected with plasmids encoding various TAL-TET1CD fusion proteins harboring the Gly_3Ser linker by nucleofecting 1×10^6 dividing cells with 10 µg of DNA in 100 µl of Kit V solution (Lonza) using program T-016 on the Nucleofector 2b Device (Lonza). Nucleofected cells were incubated in 6-well plates at 37°C for 4 days following nucleofection. Genomic DNA was harvested 4 days post-transfection.

Bisulfite DNA sequencing of KLF4 intron 1-2 from these transfected cell populations was performed as follows. Genomic DNA was isolated from transfected cells using the Qiagen Blood Mini kit. 200-1000 ng of genomic DNA was bisulfite treated using either the EZ DNA Methylation Kit (Zymo), EZ DNA Methylation-Lightning Kit (Zymo), or Cells-to-CpG Bisulfite Conversion Kit (Applied Biosystems) following recommended protocols. PCR amplification of Bis-DNA was performed using Accuprime HiFi Taq DNA polymerase system (Invitrogen). PCR products were TA-cloned using the Original TA Cloning protocol (Invitrogen) or the pGEM-T Easy Vector System (Promega) and transformed into either Top10 or XL1-Blue competent cells. Individual clones were miniprepped and sequenced and sequences were analyzed using the BISMA software available online at (Rohde et al., BMC Bioinformatics 2010, 11:230), using default parameters.

[0110] The results revealed robust demethylation at CpGs positioned adjacent to the site bound by each of the TAL-TET1CD fusion proteins. Figure 3 shows the rates of demethylation for CpGs near the four different TAL binding sites. TAL-TET1CD protein targeted to site1 showed the highest levels of demethylation with 32.5% and 45.4% demethylation of CpGs located 9 and 15 bp downstream of the target binding site, respectively. TAL-TET1CD protein targeted to site 2 showed no demethylation of CpGs located 1 and 3bp downstream of the binding site but did show 36.5% and 28.2% demethylation at CpGs located 17 and 27bp downstream, respectively. TAL-TET1 protein targeted to site 3 showed several greater than 25% demethylation at CpGs located 22, 36, 116 and 143bp downstream, and a TAL-TET1CD protein targeted to site 4 showed a 37% demethylation at a CpG located 13bp downstream but failed to show significant demethylation at any other CpGs. In additional experiments, the KL1 TALE-TET1CD fusion protein reduced the methylation of CpGs located 10 and 16 bp from the 3' boundary of the TALE binding site by 54% and 66%, respectively. Interestingly, this KL-1 fusion also induced demethylation of CpGs located ~150 bp away from 3' edge of its binding site, albeit to a lesser degree than the more proximal CpGs.

[0111] Control fusion proteins bearing a TALE repeat array targeted to an unrelated EGFP reporter gene sequence did not demethylate CpGs in the KLF4 intron, demonstrating that demethylation requires specific binding to the target locus by the TALE repeats and is not due simply to overexpression of proteins harboring TET1 hydroxymethylase activity.

[0112] Overall, this demonstrates the ability of engineered TAL-TET1CD fusion proteins harboring a Gly3Ser linker to induce 30-50% demethylation of CpGs located approximately 10-30bp downstream of the TAL binding site. (Note for all of these cases, "downstream" refers to bases located 3' to the last base of the binding site contacted by the carboxy-terminal TAL effector repeat in the fusion protein.) TAL-TET1CD fusion proteins harboring more extended linkers ($Gly_3SerGly_4SerArgSer$ (SEQ ID NO:6), $Gly_3Ser(Gly_4Ser)_2ArgSer$ (SEQ ID NO:7), $Gly_3Ser(Gly_4Ser)_3ArgSer$ (SEQ ID NO:8), $Gly_3Ser(Gly_4Ser)_4ArgSer$ (SEQ ID NO:9), or $Gly_3Ser(Gly_4Ser)_5ArgSer$ (SEQ ID NO:10) also showed similar levels of demethylation (Figure 4A).

[0113] In comparison experiments, TALE fusions bearing the TET1 CD domain induced significantly greater decreases in methylation of CpGs proximal to the TALE binding site than those bearing the full-length TET1 protein as judged by deep bisulfite sequencing (Figure 4B). Because the TALE-full-lengthTET1 fusions were less active than TALE-TET1CD fusions, further experiments focused on TALE-TET1CD fusions.

Example 3. Engineered TALE-TET1CD Proteins Targeted to the Human HBB Locus

[0114] To extend and further generalize the finding that TALE-TET1CD fusion proteins can be used to identify CpG methylation marks that influence endogenous gene expression, a TALE-TET1CD fusion protein was designed to demethylate CpGs in an additional locus, the human beta-globin (HBB) gene promoter.

[0115] Published work has identified a small number of CpGs (about four) in the human β-globin and γ-globin promoters which are differentially methylated in fetal liver and adult bone marrow cells and has suggested that methylation of these CpGs may dictate expression of these genes during development (Mabaera et al., Blood. 2007 Aug 15;110(4):1343-52). Ten TALE arrays targeted to sequences near these differentially methylated CpGs at the human β-globin locus were engineered: four TALEs targeted to a set of two CpGs ~100bp downstream of the transcription start site and six TALEs targeted to a set of two CpGs ~300bp upstream of the start site (Figure 8A).

[0116] Human K562 cells were cultured in RPMI 1640 medium (Gibco) with 10% HI-FBS (Gibco), 1% Glutamax (Gibco) and 1% Pen/Strep (Gibco), and were transfected with plasmids encoding various TALE-TET1CD fusion proteins harboring the Gly_3Ser linker by nucleofecting 1×10^6 dividing cells with 10 μg of DNA in 100 μl of Kit V solution (Lonza) using program T-016 on the Nucleofector 2b Device (Lonza). Nucleofected cells were incubated in 6-well plates at 37°C for 4 days following nucleofection. Genomic DNA was harvested 4 days post-transfection. Bisulfite DNA sequencing of the HBB locus from these transfected cell populations was performed as follows. Genomic DNA was isolated from transfected cells using the Qiagen Blood Mini kit. 200-1000 ng of genomic DNA was bisulfite treated using either the EZ DNA Methylation Kit (Zymo), EZ DNA Methylation-Lightning Kit (Zymo), or Cells-to-CpG Bisulfite Conversion Kit (Applied Biosystems) following recommended protocols. PCR amplification of Bis-DNA was performed using Pyromark PCR kit (Qiagen). Illumina adapters and barcodes were added by PCR with Phusion High-Fidelity PCR enzyme (NEB) and amplicons were sequenced on an Illumina MiSeq system. Total RNA was isolated from the same cells with the PureLink RNA mini kit (Ambion) according to manufacturer's instructions. Reverse transcription was performed with the SuperscriptIII RT kit (Invitrogen) and Taqman assays were run on an Applied Biosystems 7500Fast Real Time PCR machine.

[0117] HBB gene Expression levels were assayed as follows. RNA was reverse transcribed using the SuperScript III First-Strand Synthesis SuperMix and oligo-dT (Life Technologies). Quantitative PCR was performed with Taqman Universal PCR Mastermix (Applied Biosystems) on an ABI 7500 Fast Real-Time PCR system with the following primer/probe sets: Forward HBB primer 5'-CAAGGGCACCTTTGCCACAC-3' (SEQ ID NO:11); Reverse HBB Primer 5'-TTTGCCAAAGTGATGGGCCA-3' (SEQ ID NO: 12); HBB Taqman Probe 5'-/56-FAM/CCTGGGCAA/ZEN/CGTGCTGGTCTGTGT/3IABkFQ/-3' (SEQ ID NO: 13); Forward β-actin (ACTB) Primer: 5'- GGCACCCAGCACAATGAAG-3' (SEQ ID NO:14); Reverse ACTB primer 5'- GCCGATCCACACGGAGTACT-3' (SEQ ID NO: 15); ACTB Taqman Probe 5'-/SMAX550-Y/TCAAGATCA/ZEN/TTGCTCCTCCTGAGCGC/3IABlk_FQ/-3' (SEQ ID NO:16).

[0118] Although all ten TALE-TET1CD fusions (termed HB-1 through HB-10) induced significant demethylation of CpGs near their respective binding sites in human K562 cells (Fig. 8B and 8C), significant increases in HBB gene expression as measured by quantitative RT-PCR were observed with only four of these proteins (HB-3, HB-4, HB-5, and HB-6) (Fig. 8D). Of note, the three proteins (HB-4, HB-5, and HB-6) that induced the greatest fold-activation of the promoter were the only fusions that induced significant demethylation of the CpG at position -266 (numbered relative to the transcription start site; Fig. 8D); in addition, the magnitude of HBB gene activation observed with these three TALE-TET1CD fusions correlated with the extent of demethylation (Fig. 8D). Control experiments showed that HB-4, HB-5, and HB-6 proteins bearing the H1671Y/D1673A mutations (which inactivate TET1 catalytic domain activity) failed to demethylate the -266 CpG and also failed to efficiently activate HBB gene expression in K562 cells.

[0119] Western Blotting was performed as follows. Transfected cells were lysed in radioimmunoprecipitation assay protein extraction buffer (Cell Signaling) and centrifuged for 5 min at 4°C to harvest the supernatant. The concentration of the extracted protein was determined using a Thermo Scientific protein assay kit. Ten micrograms of the extracted protein were loaded per lane onto 4% to 12% Tris-glycine gels (Life Technologies) for electrophoresis, and electro-transferred to nitrocellulose membrane (Life Technologies). This membrane was blocked with 5% nonfat dry milk in Tris-buffered Saline-Tween buffer [25 mM Tris (pH 7.4), 137 mM NaCl, 2.7 mM KCl, 0.1% Tween 20] for 1 hour at room temperature, followed by incubation with FLAG monoclonal antibody (Sigma) and β-actin (C4) HRP monoclonal antibody (Santa Cruz Biotechnology) in blocking buffer for 2 hours at room temperature. After washing with TBS-T buffer, the membrane was incubated with goat anti-mouse IgG (H+L) - HRP conjugate antibody (Bio-Rad Laboratories) at a dilution of 1 :7500 in blocking buffer for 1 hour at room temperature. The membrane was developed using ImmunoCruz chemiluminescence system (Santa Cruz Biotechnology) and exposed to BioMax XAR film (Kodak). These Western blot experiments confirm that the loss of demethylation and gene activation activities observed with the catalytically inactive mutants of HB-5 and HB-6 was not due to decreased protein expression in K562 cells.

[0120] To test whether the -266 CpG demethylation and increased HBB gene expression induced by HB-4, HB-5-, and HB-6 are stable over time, these variables were assayed in a more extended time-course experiment. K562 cells transfected with expression plasmids encoding HB-4, HB-5, HB-6, or a control TALE-TET1CD fusion targeted to the KLF4 locus were assessed for -266 CpG demethylation and HBB gene expression at 4, 7, 14, and 30 days post-transfection. HB-4, HB-5, and HB-6 fusions all showed their highest levels of fold-activation at post-transfection day 4 with HBB expression steadily decreasing by day 30 to the same level observed with the control KLF4-targeted TALE-TET1CD protein. The results are shown in Figures10A-B. Strikingly, for all three HBB-targeted TALE-TET1CD proteins, the extent of -266 CpG methylation over time in the population of transfected cells directly paralleled the fold-activation of HBB gene expression while the methylation status of another CpG at position -306 did not change significantly during the course of the experiment.

[0121] Taken together, these findings strongly suggest that demethylation of the -266 CpG may require continued expression of a targeted TALE-TET1CD protein and that demethylation of this particular CpG is required for the observed activation of HBB gene expression. These results demonstrate that the TALE-TET1CD fusion proteins can be used to demethylate specific CpG residues in a gene promoter and thereby upregulate expression of that gene in living cells.

Example 4. Engineered ZF-TET1 Fusion Proteins

[0122] In addition to generating TALE-TET1 fusions, Zinc Finger (ZF)-TET1 fusion proteins were also engineered. The zinc finger arrays were constructed targeted to two sites in the KLF4 gene and to six sites in the HBB gene. These zinc finger arrays each consisted of six fingers and were assembled using two-finger units according to the method originally described by Moore et al., Proc Natl Acad Sci USA 98, 1432-1436 (2001), and similar to the ZF engineering approach used by Sangamo Biosciences (Doyon et al., Nat Biotechnol 26, 702-708 (2008)). The six-finger zinc finger arrays were assembled using two-finger units derived from three-finger proteins selected by the OPEN method (Maeder et al., Mol. Cell 31, 294-301 (2008)). Assembled six-finger proteins were fused to the catalytic domain of TET1 using the linker sequence Leu-Arg-Gly-Ser (SEQ ID NO:19) and to an N-terminal nuclear localization signal with expression of the ZF-TET1CD fusion protein driven by an EF1alpha promoter.

[0123] Of the two ZF-TET1CD fusions targeted to KLF4, one failed to demethylate nearby CpGs while the other was able to induce demethylation of CpGs located 6 and 24bp downstream of the ZF binding site by 61% and 18%, respectively (Figure 9A). One ZF-TET1CD fusion targeted to the -266 CpG in the HBB promoter failed to induce demethylation but all five fusions targeted to the region approximately 100 bp downstream of the transcription start site were able to demethylate target CpGs with efficiencies comparable to those observed with TALE-TET1CD proteins (Figure 9B).

[0124] This data demonstrates that the catalytic domain of TET1 can be fused to alternate DNA-binding domains in order to generate site-specific DNA demethylation enzymes, resulting upregulation of expression of targeted genes in living cells.

OTHER EMBODIMENTS

[0125] It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

REFERENCES CITED IN THE DESCRIPTION

Cited references

Patent documents cited in the description

- WO2009134409A [0003]
- WO2007128982A [0004]
- WO2010037001A [0006]
- US61610212 [0036]
- WO2011017293A [0040]
- WO2004099366A [0040]
- US6511808B [0040] [0049] [0064] [0095] [0096]
- US6013453A [0040] [0049] [0064]
- US6007988A [0040] [0049] [0064]
- US6503717B [0040] [0049] [0052] [0064]
- US20020160940A [0040] [0049] [0064] [0095] [0096]
- US20020164575A [0052] [0064] [0095] [0096]
- US5049386A [0066]
- US4946787A [0066] [0067] [0091]
- US4897355A [0066]
- WO9117424A [0066] [0091]
- WO9116024A [0066]
- US4186183A [0067] [0091]
- US4217344A [0067] [0091]
- US4235871A [0067] [0091] [0091]
- US4261975A [0067] [0091] [0091]

- US4485054A [0067] [0091] [0091]
- US4501728A [0067] [0091] [0091]
- US4774085A [0067] [0091] [0091]
- US4837028A [0067] [0091] [0091]
- WO9426877A [0069]
- US4797368A [0070]
- WO9324641A [0070]
- US5173414A [0070]
- US20130123481A [0081]
- US20130115272A [0081]
- US20120251618A [0081]
- US20120100569A [0085]
- US20110112040A [0085]
- US4957773A [0092]
- US4603044A [0092]
- US6492117B [0095] [0096]
- US6453242B [0095] [0096]

Non-patent literature cited in the description

- **TAHILIANI, M. et al.**Science, 2009, vol. 324, 930-935 [0002] [0002]
- **GUO**Cell, 2011, vol. 145, 423-434 [0002]
- **HE et al.**Science, 2011, vol. 333, 1303-1307 [0002]
- **GREGORY et al.**Epigenetics, vol. 7, 20125344-349 [0005]
- **FOLEY et al.**Nature protocols, vol. 4, 121855-1868 [0007]
- **NEVILLE et al.**Nature protocols, 2012, vol. 7, 171-192 [0008]
- **FRAUER et al.**PloS ONE, vol. 6, 2e16627- [0009]
- **RODENHISER et al.**CMAJ, vol. 174, 3341-348 [0010]
- **BRANCO et al.**Nature Reviews | Genetics |, vol. 13, 7-13 [0011]
- **TAHILIANI et al.**Science, 2009, [0029]
- **RODENHISERMANN**CMAJ, 2006, vol. 174, 3341-348 [0029] [0052]
- **ESTELLER MCORN PGBAYLIN SB et al.**Cancer Res, 2001, vol. 61, 3225-9 [0029]
- **UHLMANN KROHDE KZELLER C et al.**Int J Cancer, 2003, vol. 106, 52-9 [0029]
- **GAO YGUAN MSU B et al.**Clin Chim Acta, 2004, vol. 349, 173-9 [0029]
- **BELLO MJALONSO MEAMINOSO C et al.**Mutat Res, 2004, vol. 554, 23-32 [0029]
- **MANCINI DNRODENHISER DIAINSWORTH PJ et al.**Oncogene, 1998, vol. 16, 1161-9 [0029]
- **SZYF MPAKNESHAN PRABBANI SA.**Biochem Pharmacol, 2004, vol. 68, 1187-97 [0029]
- **WIDSCHWENDTER MJONES PA**Oncogene, 2002, vol. 21, 5462-82 [0029]

- **LEA JSCOLEMAN RKURIEN A et al.**Am J Obstet Gynecol, 2004, vol. 190, 674-9 [0029]
- **DE ZHU J.**Cell Res, 2005, vol. 15, 272-80 [0029]
- **CHIM CSKWONG YLFUNG TK et al.**Leuk Res, 2004, vol. 28, 379-85 [0029]
- **ESTELLER MSILVA JMDOMINQUEZ G et al.**J Natl Cancer Inst, 2000, vol. 92, 564-9 [0029]
- **LI LCOKINO STDAHIYA R**Biochim Biophys Acta, 2004, vol. 1704, 87-102 [0029]
- **KURMASHEVA et al.**Pediatr Blood Cancer, 2005, vol. 44, 328-37 [0029]
- **SHARMA RP**Schizophr Res, 2005, vol. 72, 79-90 [0029]
- **COSTA ECHEN YDAVIS J et al.**Mol Interv, 2002, vol. 2, 47-57 [0029]
- **LUND GANDERSON LLAURIA M et al.**J Biol Chem, 2004, vol. 279, 29147-54 [0029]
- **REYON et al.**FLASH assembly of TALENs for high-throughput genome editing.Nat Biotechnol, 2012, [0031]
- **MOSCOUBOGDANOVE**Science, 2009, vol. 326, 1501-1501 [0031]
- **BOCH et al.**Science, 2009, vol. 326, 1509-1512 [0031] [0036]
- **MILLER et al.**Nat Biotechnol, 2010, vol. 29, 143-148 [0031] [0103] [0108]
- **REYON et al.**Nature Biotechnology, 2012, vol. 30, 460-465 [0036] [0103] [0108]
- **BOGDANOVEVOYTAS**Science, 2011, vol. 333, 1843-1846 [0036]
- **BOGDANOVE et al.**Curr Opin Plant Biol, 2010, vol. 13, 394-401 [0036]
- **SCHOLZEBOCH**J. Curr Opin Microbiol, 2011, [0036]
- **MOSCOUBOGDANOVE**Science, 2009, vol. 326, 1501- [0036]
- **MILLER et al.**Nat Biotechnol, 2011, vol. 29, 143-148 [0036]
- **MORBITZER et al.**T. Proc Natl Acad Sci U S A, 2010, vol. 107, 21617-21622 [0036]
- **MORBITZER et al.**Nucleic Acids Res, 2011, vol. 39, 5790-5799 [0036]
- **ZHANG et al.**Nat Biotechnol, 2011, vol. 29, 149-153 [0036] [0036]
- **GEISSLER et al.**PLoS ONE, 2011, vol. 6, e19509- [0036]
- **WEBER et al.**PLoS ONE, 2011, vol. 6, e19722- [0036]
- **CHRISTIAN et al.**Genetics, 2010, vol. 186, 757-761 [0036]
- **LI et al.**Nucleic Acids Res, 2011, vol. 39, 359-372 [0036]
- **MAHFOUZ et al.**Proc Natl Acad Sci USA, 2011, vol. 108, 2623-2628 [0036]
- **MUSSOLINO et al.**Nucleic Acids Res, 2011, [0036]
- **LI et al.**Nucleic Acids Res, 2011, vol. 39, 6315-6325 [0036]
- **CERMAK et al.**Nucleic Acids Res, 2011, vol. 39, e82- [0036]
- **WOOD et al.**Science, 2011, vol. 333, 307- [0036]
- **HOCKEMEYE et al.**Nat Biotechnol, 2011, vol. 29, 731-734 [0036]
- **TESSON et al.**Nat Biotechnol, 2011, vol. 29, 695-696 [0036]
- **SANDER et al.**Nat Biotechnol, 2011, vol. 29, 697-698 [0036]
- **HUANG et al.**Nat Biotechnol, 2011, vol. 29, 699-700 [0036]
- **MILLER et al.**EMBO J., 1985, vol. 4, 1609- [0037]
- **BERG**Proc. Natl. Acad. Sci. USA, 1988, vol. 85, 99- [0037]
- **LEE et al.**Science, 1989, vol. 245, 635- [0037]
- **KLUG**Gene, 1993, vol. 135, 83- [0037] [0037]
- **PAVLETICH et al.**Science, 1991, vol. 252, 809- [0037]
- **ELROD-ERICKSON et al.**Structure, 1998, vol. 6, 451- [0037]
- **REBAR et al.**Science, 1994, vol. 263, 671- [0038]

- **CHOO et al.**Proc. Natl. Acad. Sci. USA, 1994, vol. 91, 11163- [0038]
- **JAMIESON et al.**Biochemistry, 1994, vol. 33, 5689- [0038]
- **WU et al.**Proc. Natl. Acad. Sci. USA, 1995, vol. 92, 344- [0038]
- **CARROLL**Gene Ther., 2008, vol. 15, 1463-68 [0038]
- **CATHOMEN**Mol. Ther., 2008, vol. 16, 1200-07 [0038]
- **WU et al.**Cell. Mol. Life Sci., 2007, vol. 64, 2933-44 [0038]
- **SEGAL et al.**Biochemistry, 2003, vol. 42, 2137-48 [0039]
- **BEERLI et al.**Nat. Biotechnol., 2002, vol. 20, 135-141 [0039]
- **MANDELL et al.**Nucleic Acids Res., 2006, vol. 34, W516-523 [0039]
- **CARROLL et al.**Nat. Protoc., 2006, vol. 1, 1329-41 [0039]
- **LIU et al.**J. Biol. Chem., 2002, vol. 277, 3850-56 [0039]
- **BAE et al.**Nat. Biotechnol., 2003, vol. 21, 275-280 [0039]
- **WRIGHT et al.**Nat. Protoc., 2006, vol. 1, 1637-52 [0039]
- **RAMIREZ et al.**Nat. Methods, 2008, vol. 5, 374-375 [0039]
- **KIM et al.**Genome Res., 2009, vol. 19, 1279-88 [0039] [0039]
- **MAEDER et al.**Mol. Cell, 2008, vol. 31, 294-301 [0040] [0122]
- **JOUNG et al.**Nat. Methods, 2010, vol. 7, 91-92 [0040]
- **ISALAN et al.**Nat. Biotechnol., 2001, vol. 19, 656-660 [0040]
- **KUMMERFELDTEICHMANN**Nucleic Acids Res., 2006, vol. 34, D74-81 [0041]
- **TJONGZHOU**Nucl. Acids Res., 2007, vol. 35, 1465-1477 [0041]
- **HARRISON**Nature, 1991, vol. 353, 715-719 [0041]
- **KARMIRANTZOUHAMODRAKAS**Protein Eng., 2001, vol. 14, 7465-472 [0041]
- **KUMAR et al.**BMC Bioinformatics., 2007, vol. 8, 463- [0041]
- **KUMAR et al.**J Biomol Struct Dyn., 2009, vol. 26, 6679-86 [0041]
- **LIN et al.**PLoS One, 2011, vol. 6, 9e24756- [0041]
- **SILVA, G et al.**Current Gene Therapy, 2011, vol. 11, 11-27 [0043]
- **ARNOULD et al.**Journal of Molecular Biology, 2006, vol. 355, 443-58 [0043]
- **ARNOULD et al.**Protein Engineering Design & Selection, 2011, vol. 24, 27-31 [0043]
- **STODDARD, Q.**Rev. Biophys., 2005, vol. 38, 49- [0043]
- **GRIZOT et al.**Nucleic Acids Research, 2010, vol. 38, 2006-18 [0043]
- **PINGOUDSILVA**Nat Biotechnol., 2007, vol. 25, 743-4 [0044]
- **SCHLEIFMAN et al.**Methods Mol Biol., 2008, vol. 435, 175-90 [0044]
- **ARIMONDO et al.**Mol Cell Biol., 2006, vol. 26, 1324-33 [0044]
- **MAJUMDAR et al.**J Biol Chem., 2008, vol. 283, 1711244-52 [0044]
- **SIMON et al.**Nucleic Acids Res., 2008, vol. 36, 113531-8 [0044]
- **EISENSCHMIDT et al.**Nucleic Acids Res., 2005, vol. 33, 227039-47 [0044]
- **IYER et al.**Cell Cycle., 2009, vol. 8, 111698-710 [0047]
- **BURNETT et al.**J. Leukoc. Biol., 2004, vol. 75, 4612-623 [0050]
- **FREEMAN et al.**Cancer Res., 2003, vol. 63, 238256-8563 [0050]
- **CASTELLANO et al.**Curr. Biol., 1999, vol. 9, 7351-360 [0050]
- **CRABTREESCHREIBER**Trends Biochem. Sci., 1996, vol. 21, 11418-422 [0050]
- **GRAEF et al.**Embo. J., 1997, vol. 16, 185618-5628 [0050]
- **MUTHUSWAMY et al.**Mol. Cell. Biol., 1999, vol. 19, 106845-6857 [0050]
- **D'AVIGNON et al.**Biopolymers, 2006, vol. 83, 3255-67 [0050]

- **ZITZEWITZ et al.**Biochemistry, 1995, vol. 34, 3912812-9 [0050]
- **LI et al.**Hum Gene Ther., 2008, vol. 19, 9958-64 [0050]
- **SAMBROOK et al.**Molecular Cloning, A Laboratory Manual20010000 [0054]
- **KRIEGLER**Gene Transfer and Expression: A Laboratory Manual19900000 [0054]
- Current Protocols in Molecular Biology20100000 [0054]
- **PALVA et al.**Gene, 1983, vol. 22, 229-235 [0054]
- **GOSSENBUJARD**Proc. Natl. Acad. Sci. USA, 1992, vol. 89, 5547- [0055]
- **OLIGINO et al.**Gene Ther., 1998, vol. 5, 491-496 [0055]
- **WANG et al.**Gene Ther., 1997, vol. 4, 432-441 [0055]
- **NEERING et al.**Blood, 1996, vol. 88, 1147-55 [0055]
- **RENDAHL et al.**Nat. Biotechnol., 1998, vol. 16, 757-761 [0055]
- **COLLEY et al.**J. Biol. Chem., 1989, vol. 264, 17619-22 [0061]
- Guide to Protein PurificationMethods in Enzymology19900000vol. 182, [0061]
- **MORRISON**J. Bacteriol., 1977, vol. 132, 349-351 [0061]
- **CLARK-CURTISSCURTISS et al.**Methods in Enzymology19830000vol. 101, 347-362 [0061]
- **GARCIA-BUSTOS et al.**Biochim. Biophys. Acta, 1991, vol. 1071, 83-101 [0063]
- **HAN et al.**Mol Cell Biol., 2008, vol. 28, 31124-35 [0064]
- **MABAERA et al.**Blood., 2007, vol. 110, 41343-52 [0064] [0115]
- **ANDERSON**Science, 1992, vol. 256, 808-813 [0065]
- **NABELFELGNER**TIBTECH, 1993, vol. 11, 211-217 [0065]
- **MITANICASKEY**TIBTECH, 1993, vol. 11, 162-166 [0065]
- **DILLON**TIBTECH, 1993, vol. 11, 167-175 [0065]
- **MILLER**Nature, 1992, vol. 357, 455-460 [0065]
- **VAN BRUNT**Biotechnology, 1988, vol. 6, 1149-54 [0065]
- **VIGNE**Restorat. Neurol. Neurosci., 1995, vol. 8, 35-36 [0065]
- **KREMERPERRICAUDET**Br. Med. Bull, 1995, vol. 51, 31-44 [0065]
- **HADDADA et al.**Current Topics in Microbiology and Immunology19950000 [0065]
- **YU et al.**Gene Ther., 1994, vol. 1, 13-26 [0065]
- **CRYSTAL**Science, 1995, vol. 270, 404-410 [0067]
- **BLAESE et al.**Cancer Gene Ther., 1995, vol. 2, 291-297 [0067]
- **BEHR et al.**Bioconjugate Chem., 1994, vol. 5, 382-389 [0067]
- **REMY et al.**Bioconjugate Chem., 1994, vol. 5, 647-654 [0067]
- **GAO et al.**Gene Ther., vol. 2, 710-722 [0067]
- **AHMAD et al.**Cancer Res., 1992, vol. 52, 4817-20 [0067]
- **BUCHSCHER et al.**J. Virol., 1992, vol. 66, 2731-39 [0069]
- **JOHANN et al.**J. Virol., 1992, vol. 66, 1635-40 [0069]
- **SOMMERFELT et al.**Virololgy, 1990, vol. 176, 58-59 [0069]
- **WILSON et al.**J. Virol., 1989, vol. 63, 2374-78 [0069]
- **MILLER et al.**J. Virol., 1991, vol. 65, 2220-24 [0069]
- **WEST et al.**Virology, 1987, vol. 160, 38-47 [0070]
- **KOTIN**Hum. Gene Ther., 1994, vol. 5, 793-801 [0070]
- **MUZYCZKA**J. Clin. Invest., 1994, vol. 94, 1351- [0070]
- **TRATSCHIN et al.**Mol. Cell. Biol., 1985, vol. 5, 3251-60 [0070]

- **TRATSCHIN et al.**Mol. Cell. Biol., 1984, vol. 4, 2072-81 [0070]
- **HERMONATMUZYCZKA**Proc. Natl. Acad. Sci. USA, 1984, vol. 81, 6466-70 [0070]
- **SAMULSKI et al.**J. Virol., 1989, vol. 63, 3822-28 [0070]
- **DUNBAR et al.**Blood, 1995, vol. 85, 3048- [0072]
- **KOHN et al.**Nat. Med., 1995, vol. 1, 1017- [0072]
- **MALECH et al.**Proc. Natl. Acad. Sci. USA, 1997, vol. 94, 12133-38 [0072]
- **BLAESE et al.**Science, 1995, vol. 270, 475-480 [0072]
- **ELLEM et al.**Immunol Immunother., 1997, vol. 44, 10-20 [0072]
- **DRANOFFET**Hum. Gene Ther., 1997, vol. 1, 111-112 [0072]
- **WAGNER et al.**Lancet, 1998, vol. 351, 1702-1703 [0073]
- **KEARNS et al.**Gene Ther., 1996, vol. 9, 748-55 [0073]
- **STERMAN et al.**Hum. Gene Ther., 1998, vol. 7, 1083-89 [0074] [0074]
- **ROSENECKER et al.**Infection, 1996, vol. 24, 15-10 [0074]
- **STERMAN et al.**Hum. Gene Ther, 1998, vol. 9, 71083-89 [0074]
- **WELSH et al.**Hum. Gene Ther., 1995, vol. 2, 205-218 [0074]
- **ALVAREZ et al.**Hum. Gene Ther., 1997, vol. 5, 597-613 [0074]
- **TOPF et al.**Gene Ther., 1998, vol. 5, 507-513 [0074]
- **HAN et al.**Proc. Natl. Acad. Sci. USA, 1995, vol. 92, 9747-51 [0076]
- **SIPIONE et al.**Diabetologia, 2004, vol. 47, 499-508 [0077] [0079]
- **FRESHNEY et al.**Culture of Animal Cells, A Manual of Basic Technique20050000 [0078]
- **INABA et al.**J. Exp. Med., 1992, vol. 176, 1693-1702 [0079] [0080]
- Remington: The Science and Practice of Pharmacy20050000 [0082] [0098]
- **PROCHIANTZ**Curr. Opin. Neurobiol, 1996, vol. 6, 629-634 [0084]
- **LIN et al.**J. Biol. Chem., 1995, vol. 270, 14255-58 [0084]
- **ENDOH et al.**Methods Mol. Biol., 2010, vol. 623, 271-281 [0085]
- **SCHMIDT et al.**FEBS Lett, 2010, vol. 584, 1806-13 [0085]
- **FUTAKI**Biopolymers, 2006, vol. 84, 241-249 [0085]
- **FAHRAEUS et al.**Curr. Biol., 1996, vol. 6, 84- [0085]
- **DEROSSI et al.**J. Biol. Chem., 1994, vol. 269, 10444- [0085]
- **ELLIOTO'HARE**Cell, 1997, vol. 88, 223-233 [0085]
- **CRONICAN et al.**Chem Biol., 2011, vol. 18, 833- [0085]
- **CRONICAN et al.**ACS Chem. Biol., 2010, vol. 5, 747- [0085]
- **MCNAUGHTON et al.**Proc. Natl. Acad. Sci. USA, vol. 106, 6111- [0085]
- **LAWRENCE et al.**J. Am. Chem. Soc., 2007, vol. 129, 10110- [0085]
- **CARON et al.**Mol Ther., 2001, vol. 3, 310-318 [0085]
- **LANGEL**Cell-Penetrating Peptides: Processes and ApplicationsCRC Press20020000 [0085]
- **EL-ANDALOUSSI et al.**Curr. Pharm. Des., 2005, vol. 11, 3597-3611 [0085]
- **DESHAYES et al.**Cell. Mol. Life Sci., 2005, vol. 62, 1839-49 [0085]
- **ARORA et al.**J. Biol. Chem., 1993, vol. 268, 3334-41 [0086]
- **PERELLE et al.**Infect. Immun., 1993, vol. 61, 5147-56 [0086]
- **STENMARK et al.**J. Cell Biol., 1991, vol. 113, 1025-32 [0086]
- **DONNELLY et al.**Proc. Natl. Acad. Sci. USA, 1993, vol. 90, 3530-34 [0086]
- **CARBONETTI et al.**Abstr. Annu. Meet. Am. Soc. Microbiol., 1995, vol. 95, 295- [0086]

- **SEBO et al.**Infect. Immun., 1995, vol. 63, 3851-57 [0086]
- **KLIMPEL et al.**Proc. Natl. Acad. Sci. USA, 1992, vol. 89, 10277-81 [0086]
- **NOVAK et al.**J. Biol. Chem., 1992, vol. 267, 17186-93 [0086]
- Proc. Natl. Acad. Sci. USA, 1987, vol. 84, 7851- [0090]
- Biochemistry, 1989, vol. 28, 908- [0090]
- **SZOKA et al.**Annu. Rev. Biophys. Bioeng, 1980, vol. 9, 467- [0091]
- **DEAMERBANGHAM**Biochim. Biophys. Acta, 1976, vol. 443, 629-634 [0091]
- **FRALEY et al.**Proc. Natl. Acad. Sci. USA, 1979, vol. 76, 3348-52 [0091]
- **HOPE et al.**Biochim. Biophys. Acta, 1985, vol. 812, 55-65 [0091]
- **MAYER et al.**Biochim. Biophys. Acta, 1986, vol. 858, 161-168 [0091]
- **WILLIAMS et al.**Proc. Natl. Acad. Sci. USA, 1988, vol. 85, 242-246 [0091]
- Liposomes19830000 [0091]
- **HOPE et al.**Chem. Phys. Lip., 1986, vol. 40, 89- [0091]
- **GREGORIADIS**Liposome Technology, 1984, [0091]
- **LASIC**Liposomes: from Physics to Applications, 1993, [0091]
- **RENNEISEN et al.**J. Biol. Chem., 1990, vol. 265, 16337-42 [0094]
- **LEONETTI et al.**Proc. Natl. Acad. Sci. USA, 1990, vol. 87, 2448-51 [0094]
- **TAHILIANI et al.**Science, 2009, vol. 324, 930-935 [0102]
- **ZHANG et al.**Cell Res, 2010, vol. 20, 1390-1393 [0104]
- **XU et al.**Mol. Cell, 2011, vol. 42, 451-464 [0104]
- **ROHDE et al.**BMC Bioinformatics, 2010, vol. 11, 230- [0109]
- **MOORE et al.**Proc Natl Acad Sci USA, 2001, vol. 98, 1432-1436 [0122]
- **DOYON et al.**Nat Biotechnol, 2008, vol. 26, 702-708 [0122]

P A T E N T K R A V

1. Fusionsprotein omfattende

et DNA-bindende domæne som binder specifikt til en forvalgt målsekvens der er inden for 250 bp af en CpG direst i en mammal celle, hvor det DNA-bindende domæne er eller omfatter en konstrueret zink finger; og

et katalytisk domæne omfattende en sekvens der katalyserer demethylering af cytosiner i DNA, hvor sekvensen omfatter aminosyre 1580-2052 af Tet1 (SEQ ID NO:2), hvor fusionsproteinet binder til målsekvensen og reducerer methylering af CpG direst.

2. Fusionsprotein ifølge krav 1, yderligere omfattende en linker mellem det DNA-bindende domæne of det katalytiske domæne.

3. Fusionsprotein ifølge krav 2, hvor linkeren er valgt fra gruppen bestående af Gly3Ser (SEQ ID NO:5), Gly3SerGly4SerArgSer (SEQ ID NO:6), Gly3Ser(Gly4Ser)2ArgSer (SEQ ID NO:7), Gly3Ser(Gly4Ser)3ArgSer (SEQ ID NO:8), Gly3Ser(Gly4Ser)4ArgSer (SEQ ID NO:9) og Gly3Ser(Gly4Ser)5ArgSer (SEQ ID NO:10).

4. Fusionsprotein ifølge ét hvilket som helst af kravene 1-3, omfattende en flerhed af katalytiske domæner.

5. Fusionsprotein ifølge ét hvilket som helst af de foregående krav, yderligere omfattende et/en nuklear lokalisations-domæne/sekvens.

6. Ex vivo fremgangsmåde til reducering af methylering af en valgt DNA sekvens i en mamalsk celle, fremgangsmåden omfattende at bringe cellerne i kontakt med et fusionsprotein ifølge ét hvilket som helst af kravene 1-5 eller en nukleinsyre der koder for fusionsproteinet ifølge ét hvilket som helst af kravene 1-5.

7. Fusionsprotein ifølge et hvilket som helst af kravene 1-5, eller nukleinsyre der koder fusionsproteiner ifølge et hvilket som helst af kravene 1-5, til anvendelse i en fremgangsmåde til reducering af methylering af en CpG direst i en valgt DNA sekvens i en mamalsk celle til behandling af cancer, fremgangsmåden omfattende at bringe cellerne i kontakt med fusionsproteinet eller en nukleinsyre der koder fusionsproteinet.

8. Fremgangsmåde ifølge krav 6 eller fusionsprotein eller nukleinsyre der koder fusionsproteinet til anvendelse ifølge krav 7, hvor det katalytiske domæne omfatter aminosyre 1418-2136 af Tet1.

9. Fremgangsmåde ifølge krav 6, eller fusionsprotein eller nukleinsyre der koder fu-

sionsproteinet til anvendelse ifølge krav 7, hvor cellen er en human celle.

10. Fremgangsmåde ifølge krav 6, hvor den valgte DNA sekvens er valgt fra en sekvens af RASSF1A, MGMT, BRCA1, p16, CDH1, TIMP-3, p15, p73, DAPK, APC, og RELN gener.

11. Fremgangsmåde ifølge krav 7, hvor cellen er tilstede i et individ:

(i) diagnosticeret med hjernecancer (glioma), hvor den valgte DNA sekvens er RASSFIA;

(ii) diagnosticeret med hjerne cancer (glioblastoma), hvor den valgte DNA sekvens er MGMT;

(iii) diagnosticeret med brystkræft, hvor den valgte DNA sekvens er BRCA1;

(iv) diagnosticeret med, livmoderhalskræft, hvor den valgte DNS sekvens er p16;

(v) diagnosticeret med oesophagus kræft, hvor den valgte DNA sekvens er CDH1;

(vi) diagnosticeret med hoved- og nakkekræft, hvor de valgte DNA sekvenser er pf6 og MGMT:

(vii) diagnosticeret med nyrekræft, hvor den valgte DNA sekvens er TIMP-3;

(viii) diagnosticeret med lungekræft, hvor de valgte DNA sekvenser er p16 og p73;

(ix) diagnosticeret med ovariekræft, hvor den valgt DNA sekvens er BRCA1;

(x) diagnosticeret med bugspytkirtelkræft, hvor den valgte DNA sekvens er APC;

(xi) diagnosticeret med prostatakræft, hvor den valgte DNA sekvens er BRCA2;

(xii) diagnosticeret med livmoderkræft, hvor den valgte DNA sekvens er hMLHI;

(xiii) diagnosticeret med leukæmi, hvor den valgte DNA sekvens er p15;

(xiv) diagnosticeret med lymfeknudekræft, hvor den valgte DNA sekvens er DAPK;

(xv) diagnosticeret med myelomatose, hvor den valgte Dna sekvens er DAPK;

(xvi) diagnosticeret med rabdomyosarkom, hvor den valgte DNA sekvens er PAX3; eller,

(xvii) diagnosticeret med skizofreni, hvor den valgte DNA sekvens er RELN.

DRAWINGS

Drawing

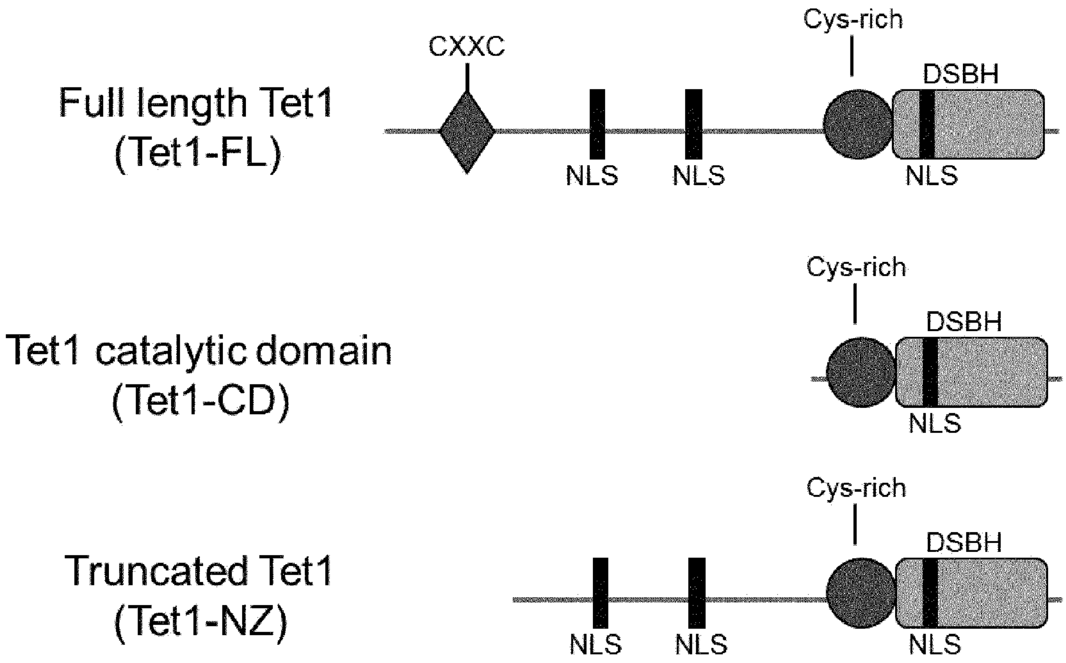


FIG. 1A

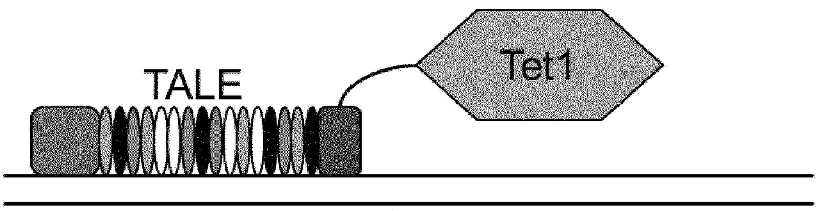


FIG. 1B

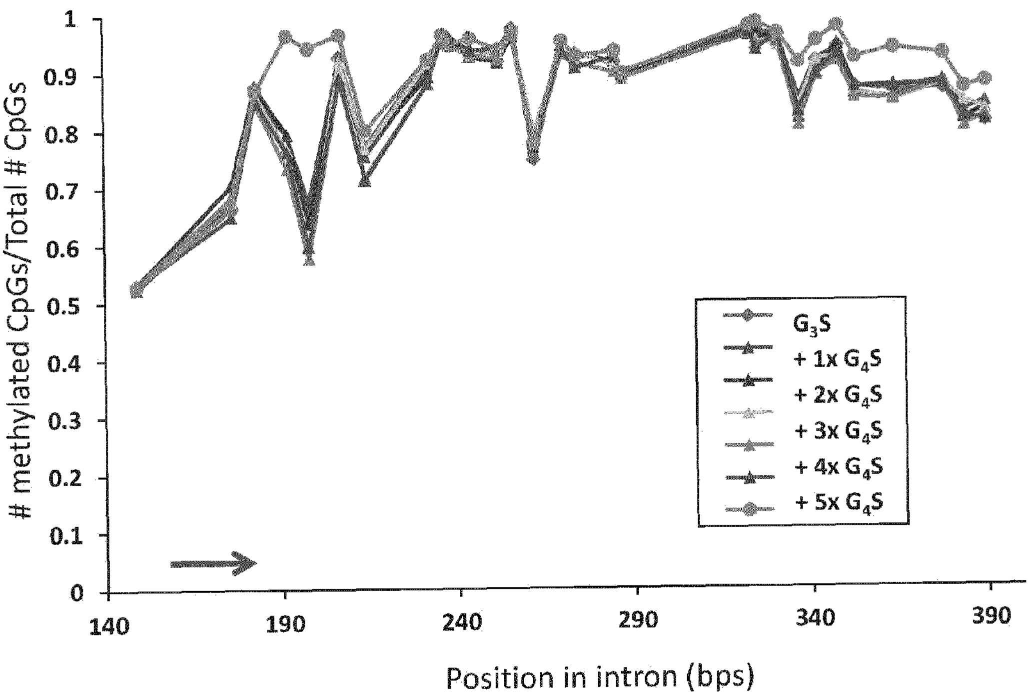

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
methylated CpGs/Total # CpGs
140
190
240
290
340
390
Position in intron (bps)
G_3S
+ 1x G_4S
+ 2x G_4S
+ 3x G_4S
+ 4x G_4S
+ 5x G_4S

FIG. 1C

5'-GGGGGCGGGCGGAGCCTTGgggtcaccaagagccactgaa
3'-CCCCCGCCCGCCTCGGAACcccagtggttctcggtgactt

cgagggtaaggagcggccaccgcgcggactccgcaacctt
gctcccattcctcgccggtggcgcgcctgaggcgttggaa

1
cccgggatagcttccgtgtccttggctcagctgtgtatgc
gggccctatcgaaggcacaggaaccgagtcgacacatacg

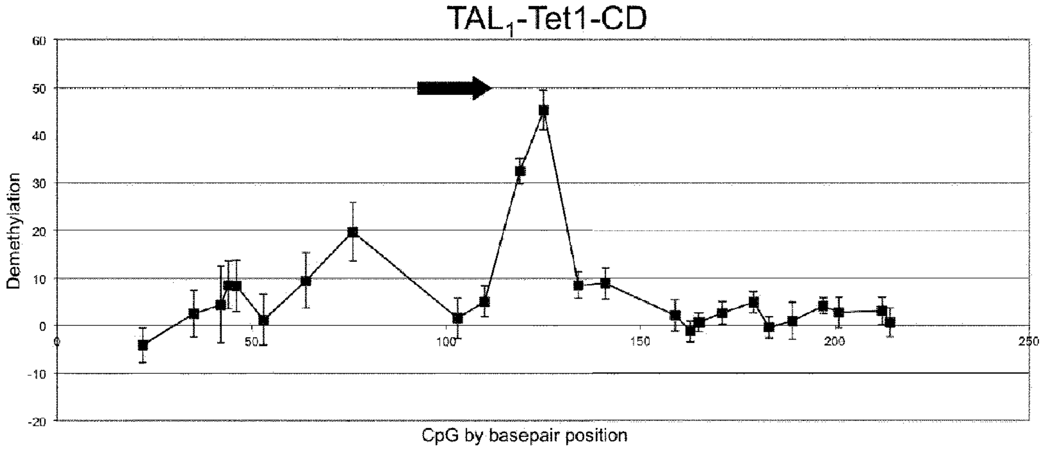
ccgtggtgcgagcactgcggagccgcctgggcgcaagcc
ggcaccacgctcgtgacgcctcggcggacccgcgttcgg

2
gggtcttcaccagtccccgggcgcgctgtcggtagggcgg
cccagaagtggtcaggggcccgcgcgacagccatcccgcc

gcgctctcgggccaccggcgggaggtggccgcgtcccca
cgcgagagcccggtggcccgccctccaccggcgcaggggt

gacactggctgggccagagtgagtttagcGCGTAGGCGGC-3
ctgtgaccgacccggtctcactcaaatcgCGCATCCGCCG-5
3
4

FIG. 2



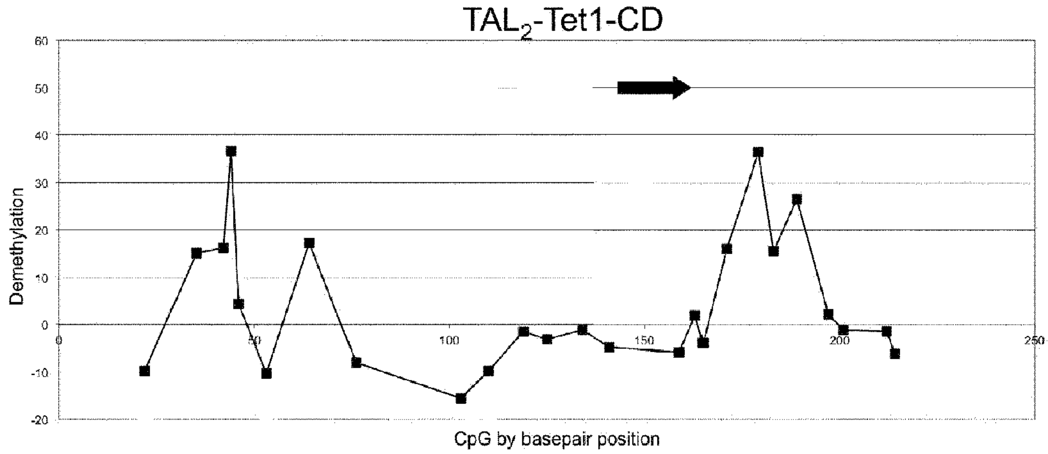
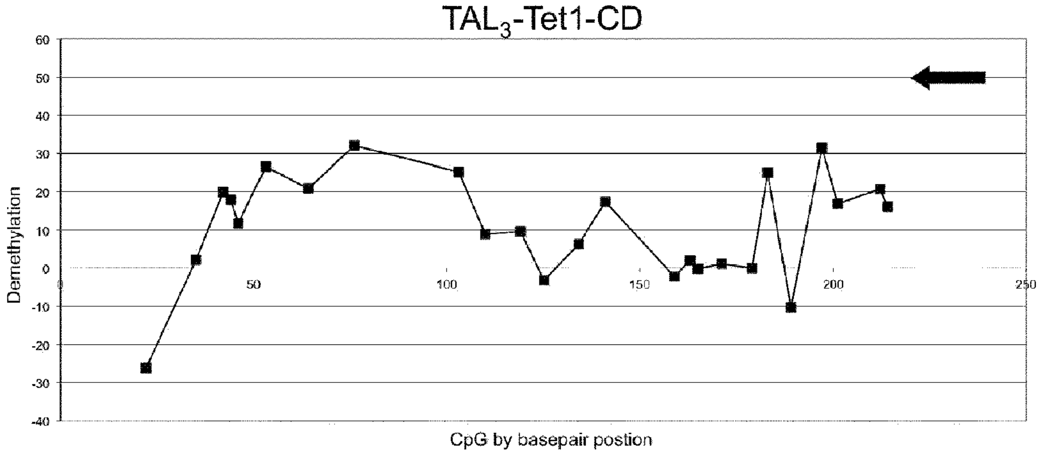


FIG. 3, page 1 of 2



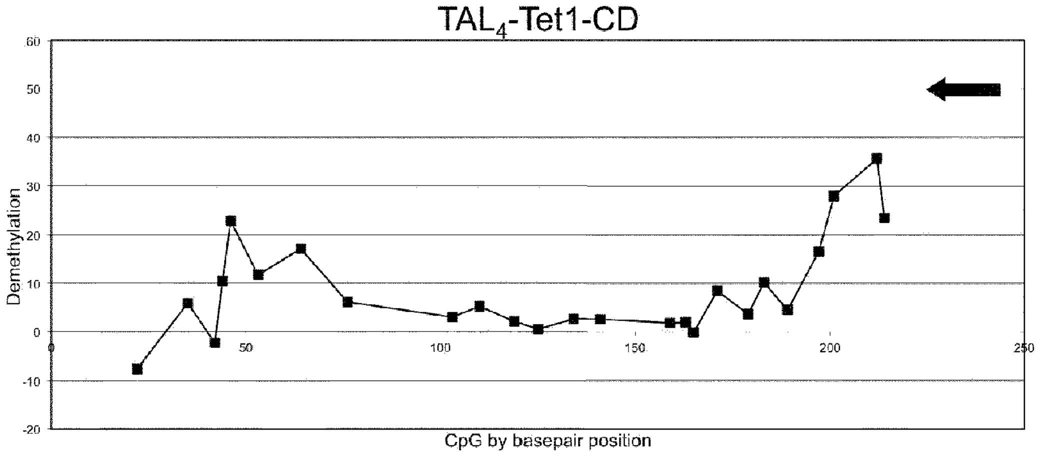


FIG. 3, page 2 of 2

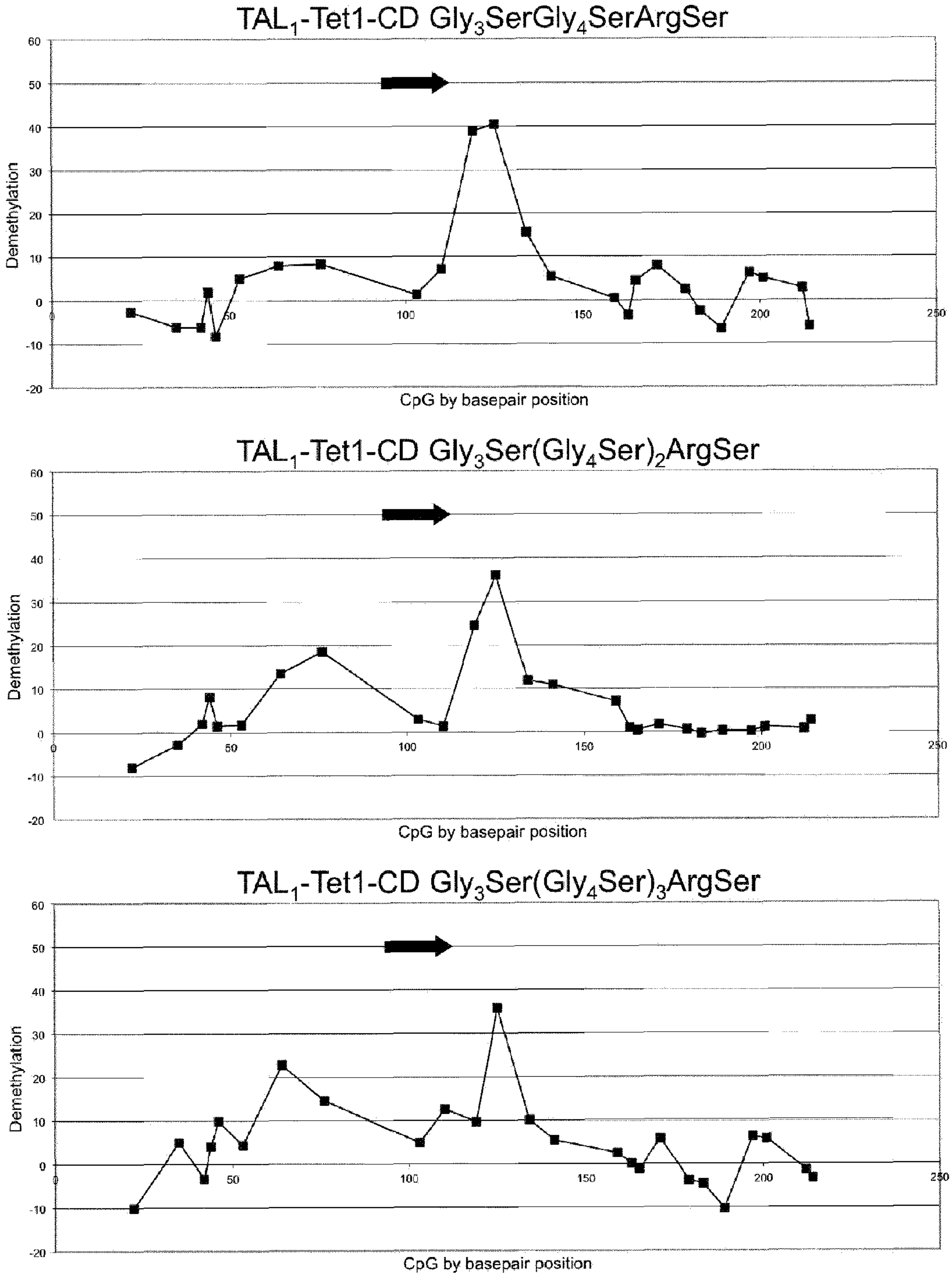
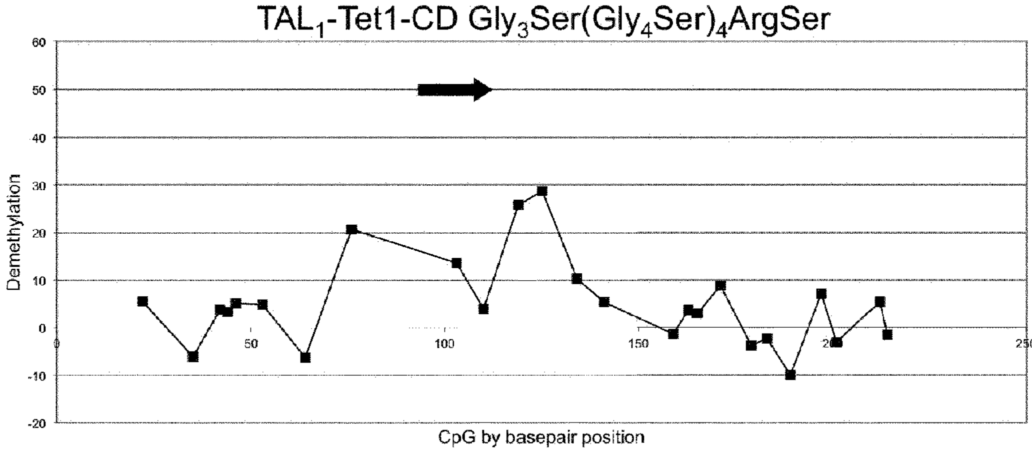


FIG. 4A, page 1 of 2



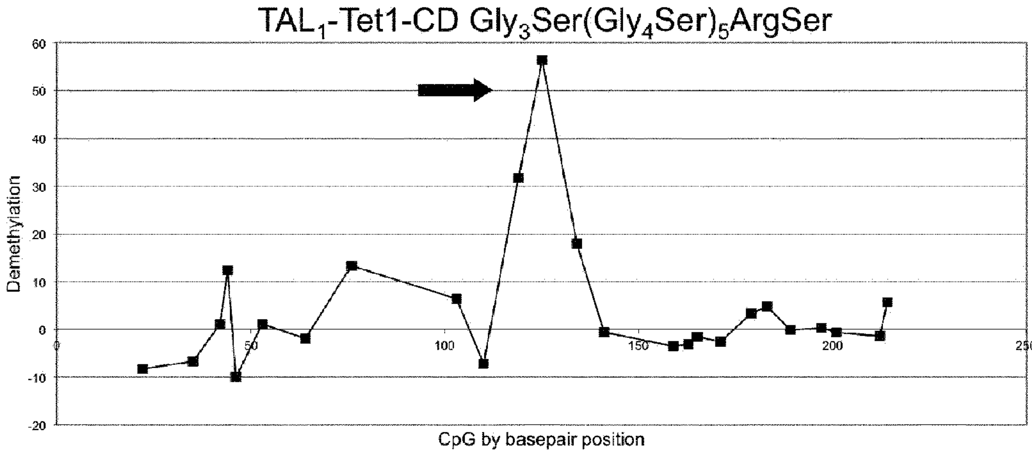


FIG. 4A, page 2 of 2

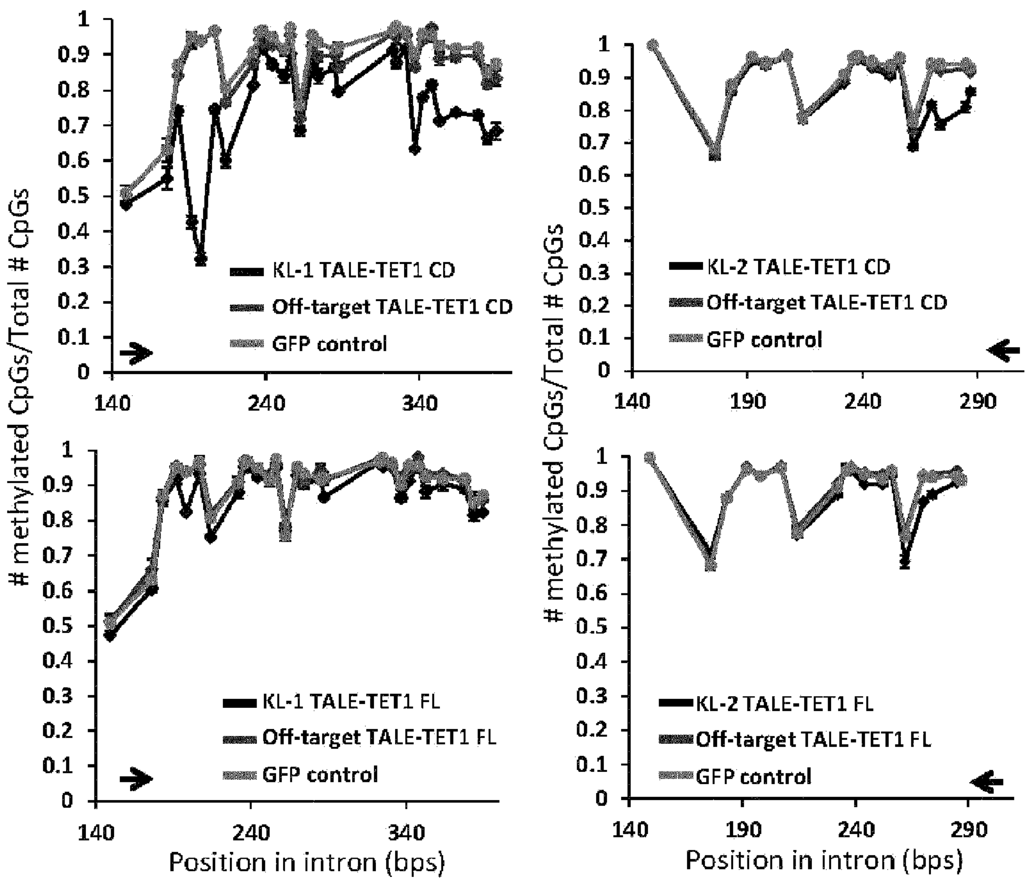

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
140
240
340
KL-1 TALE-TET1 CD
Off-target TALE-TET1 CD
GFP control
140
190
240
290
KL-2 TALE-TET1 CD
Off-target TALE-TET1 CD
GFP control
methylated CpGs/Total # CpGs
KL-1 TALE-TET1 FL
Off-target TALE-TET1 FL
GFP control
KL-2 TALE-TET1 FL
Off-target TALE-TET1 FL
GFP control
Position in intron (bps)

FIG. 4B

Tet1-FL

MSRSRHARPSRLVRKEDVNKKKKNSQLRKTTKGANKNVASVKTLSPGKLKQLIQERDVKKKT
EPKPPVPVRSLLTRAGAARMNLDRTEVLFQNPESLTCNGFTMALRSTSLSRRLSQPPLVVAK
SKKVPLSKGLEKQHDCDYKILPALGVKHSENDSVPMQDTQVLPDIETLIGVQNPSLLKGKSQ
ETTQFWSQRVEDSKINIPTHSGPAAEILPGPLEGTRCGEGLFSEETLNDTSGSPKMFAQDTV
CAPFPQRATPKVTSQGNPSIQLEELGSRVESLKLSDSYLDPIKSEHDCYPTSSLNKVIPDLN
LRNCLALGGSTSPTSVIKFLLAGSKQATLGAKPDHQEAFEATANQQEVSDTTSFLGQAFGAI
PHQWELPGADPVHGEALGETPDLPEIPGAIPVQGEVFGTILDQQETLGMSGSVVPDLPVFLP
VPPNPIATFNAPSKWPEPQSTVSYGLAVQGAIQILPLGSGHTPQSSSNSEKNSLPPVMAISN
VENEKQVHISFLPANTQGFPLAPERGLFHASLGIAQLSQAGPSKSDRGSSQVSVTSTVHVVN
TTVVTMPVPMVSTSSSSYTTLLPTLEKKKRKRCGVCEPCQQKTNCGECTYCKNRKNSHQICK
KRKCEELKKKPSVVVPLEVIKENKRPQREKKPKVLKADFDNKPVNGPKSESMDYSRCGHGEE
QKLELNPHTVENVTKNEDSMTGIEVEKWTQNKKSQLTDHVKGDFSANVPEAEKSKNSEVDKK
RTKSPKLFVQTVRNGIKHVHCLPAETNVSFKKFNIEEFGKTLENNSYKFLKDTANHKNAMSS
VATDMSCDHLKGRSNVLVFQQPGFNCSSIPHSSHSIINHHASIHNEGDQPKTPENIPSKEPK
DGSPVQPSLLSLMKDRRLTLEQVVAIEALTQLSEAPSENSSPSKSEKDEESEQRTASLLNSC
KAILYTVRKDLQDPNLQGEPPKLNHCPSLEKQSSCNTVVFNGQTTTLSNSHINSATNQASTK
SHEYSKVTNSLSLFIPKSNSSKIDTNKSIAQGIITLDNCSNDLHQLPPRNNEVEYCNQLLDS
SKKLDSDDLSCQDATHTQIEEDVATQLTQLASIIKINYIKPEDKKVESTPTSLVTCNVQQKY
NQEKGTMQQKPPSSVHNNHGSSLTKQKNPTQKKTKSTPSRDRRKKKPTVVSYQENDRQKWEK
LSYMYGTICDIWIASKFQNFGQFCPHDFPTVFGKISSSTKIWKPLAQTRSIMQPKTVFPPLT
QIKLQRYPESAEEKVKVEPLDSLSLFHLKTESNGKAFTDKAYNSQVQLTVNANQKAHPLTQP
SSPPNQCANVMAGDDQIRFQQVVKEQLMHQRLPTLPGISHETPLPESALTLRNVNVVCSGGI
TVVSTKSEEEVCSSSFGTSEFSTVDSAQKNFNDYAMNFFTNPTKNLVSITKDSELPTCSCLD
RVIQKDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSHGCPIAKWV
LRRSSDEEKVLCLVRQRTGHHCPTAVMVVLIMVWDGIPLPMADRLYTELTENLKSYNGHPTD
RRCTLNENRTCTCQGIDPETCGASFSFGCSWSMYFNGCKFGRSPSPRRFRIDPSSPLHEKNL
EDNLQSLATRLAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACLDFCAHPHRD
IHNMNNGSTVVCTLTREDNRSLGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEAKIKSGAIEV
LAPRRKKRTCFTQPVPRSGKKRAAMMTEVLAHKIRAVEKKPIPRIKRKNNSTTTNNSKPSSL
PTLGSNTETVQPEVKSETEPHFILKSSDNTKTYSLMPSAPHPVKEASPGFSWSPKTASATPA
PLKNDATASCGFSERSSTPHCTMPSGRLSGANAAAADGPGISQLGEVAPLPTLSAPVMEPLI
NSEPSTGVTEPLTPHQPNHQPSFLTSPQDLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAE
EKLPHIDEYWSDSEHIFLDANIGGVAIAPAHGSVLIECARRELHATTPVEHPNRNHPTRLSL
VFYQHKNLNKPQHGFELNKIKFEAKEAKNKKMKASEQKDQAANEGPEQSSEVNELNQIPSHK
ALTLTHDNVVTVSPYALTHVAGPYNHWV (SEQ ID NO:2)

FIG. 5A

Tet1-NZ

EELKKKPSVVVPLEVIKENKRPQREKKPKVLKADFDNKPVNGPKSESMDYSRCGHGEEQKLE
LNPHTVENVTKNEDSMTGIEVEKWTQNKKSQLTDHVKGDFSANVPEAEKSKNSEVDKKRTKS
PKLFVQTVRNGIKHVHCLPAETNVSFKKFNIEEFGKTLENNSYKFLKDTANHKNAMSSVATD
MSCDHLKGRSNVLVFQQPGFNCSSIPHSSHSIINHHASIHNEGDQPKTPENIPSKEPKDGSP
VQPSLLSLMKDRRLTLEQVVAIEALTQLSEAPSENSSPSKSEKDEESEQRTASLLNSCKAIL
YTVRKDLQDPNLQGEPPKLNHCPSLEKQSSCNTVVFNGQTTTLSNSHINSATNQASTKSHEY
SKVTNSLSLFIPKSNSSKIDTNKSIAQGIITLDNCSNDLHQLPPRNNEVEYCNQLLDSSKKL
DSDDLSCQDATHTQIEEDVATQLTQLASIIKINYIKPEDKKVESTPTSLVTCNVQQKYNQEK
GTMQQKPPSSVHNNHGSSLTKQKNPTQKKTKSTPSRDRRKKKPTVVSYQENDRQKWEKLSYM
YGTICDIWIASKFQNFGQFCPHDFPTVFGKISSSTKIWKPLAQTRSIMQPKTVFPPLTQIKL
QRYPESAEEKVKVEPLDSLSLFHLKTESNGKAFTDKAYNSQVQLTVNANQKAHPLTQPSSPP
NQCANVMAGDDQIRFQQVVKEQLMHQRLPTLPGISHETPLPESALTLRNVNVVCSGGITVVS
TKSEEEVCSSSFGTSEFSTVDSAQKNFNDYAMNFFTNPTKNLVSITKDSELPTCSCLDRVIQ
KDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSHGCPIAKWVLRRS
SDEEKVLCLVRQRTGHHCPTAVMVVLIMVWDGIPLPMADRLYTELTENLKSYNGHPTDRRCT
LNENRTCTCQGIDPETCGASFSFGCSWSMYFNGCKFGRSPSPRRFRIDPSSPLHEKNLEDNL
QSLATRLAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACLDFCAHPHRDIHNM
NNGSTVVCTLTREDNRSLGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEAKIKSGAIEVLAPR
RKKRTCFTQPVPRSGKKRAAMMTEVLAHKIRAVEKKPIPRIKRKNNSTTTNNSKPSSLPTLG
SNTETVQPEVKSETEPHFILKSSDNTKTYSLMPSAPHPVKEASPGFSWSPKTASATPAPLKN
DATASCGFSERSSTPHCTMPSGRLSGANAAAADGPGISQLGEVAPLPTLSAPVMEPLINSEP
STGVTEPLTPHQPNHQPSFLTSPQDLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLP
HIDEYWSDSEHIFLDANIGGVAIAPAHGSVLIECARRELHATTPVEHPNRNHPTRLSLVFYQ
HKNLNKPQHGFELNKIKFEAKEAKNKKMKASEQKDQAANEGPEQSSEVNELNQIPSHKALTL
THDNVVTVSPYALTHVAGPYNHWV (SEQ ID NO:3)

FIG. 5B

Tet1-CD

LPTCSCLDRVIQKDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSH
GCPIAKWVLRRSSDEEKVLCLVRQRTGHHCPTAVMVVLIMVWDGIPLPMADRLYTELTENLK
SYNGHPTDRRCTLNENRTCTCQGIDPETCGASFSFGCSWSMYFNGCKFGRSPSPRRFRIDPS
SPLHEKNLEDNLQSLATRLAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACLD
FCAHPHRDIHNMNNGSTVVCTLTREDNRSLGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEAK
IKSGAIEVLAPRRKKRTCFTQPVPRSGKKRAAMMTEVLAHKIRAVEKKPIPRIKRKNNSTTT
NNSKPSSLPTLGSNTETVQPEVKSETEPHFILKSSDNTKTYSLMPSAPHPVKEASPGFSWSP
KTASATPAPLKNDATASCGFSERSSTPHCTMPSGRLSGANAAAADGPGISQLGEVAPLPTLS
APVMEPLINSEPSTGVTEPLTPHQPNHQPSFLTSPQDLASSPMEEDEQHSEADEPPSDEPLS
DDPLSPAEEKLPHIDEYWSDSEHIFLDANIGGVAIAPAHGSVLIECARRELHATTPVEHPNR
NHPTRLSLVFYQHKNLNKPQHGFELNKIKFEAKEAKNKKMKASEQKDQAANEGPEQSSEVNE
LNQIPSHKALTLTHDNVVTVSPYALTHVAGPYNHWV (SEQ ID NO:4)

FIG. 5C

FIG. 6A

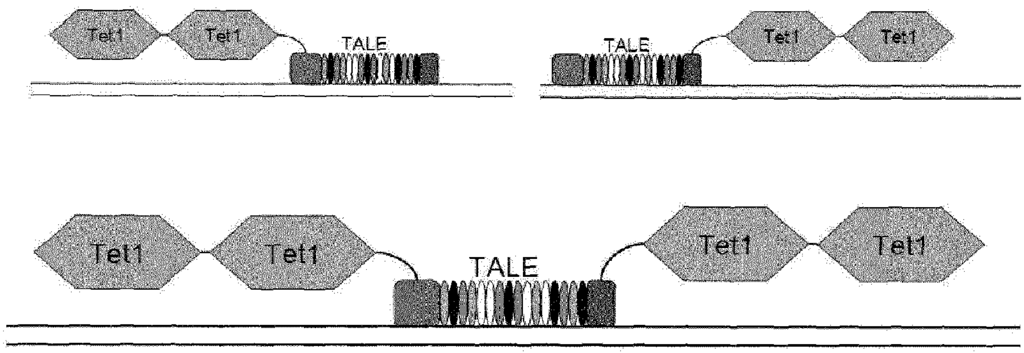

Tet1
Tet1
TALE
TALE
Tet1
Tet1
Tet1
Tet1
TALE
Tet1
Tet1

FIG. 6B.

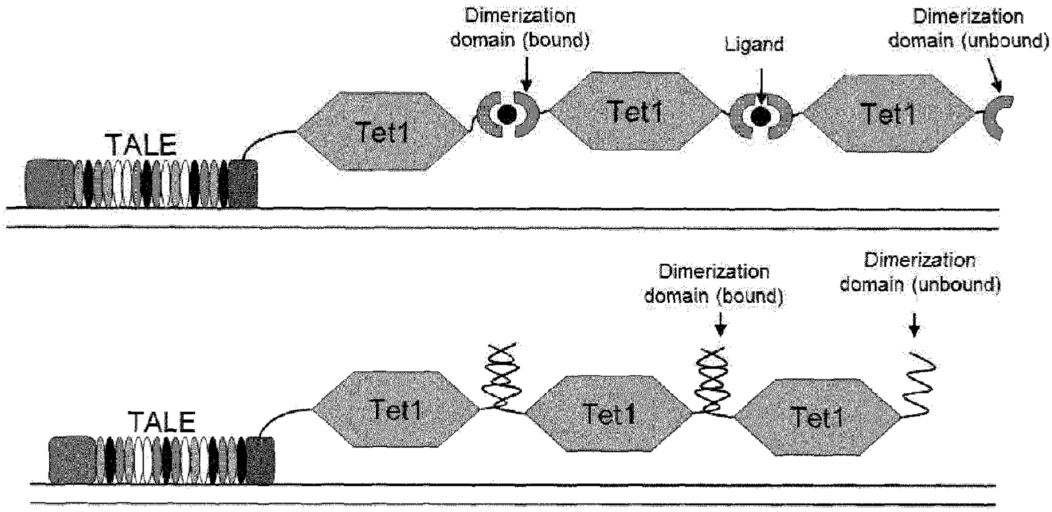

Dimerization domain (bound)
Ligand
Dimerization domain (unbound)
TALE
Tet1
Tet1
Tet1
Dimerization domain (bound)
Dimerization domain (unbound)
TALE
Tet1
Tet1
Tet1

FIG. 6C.

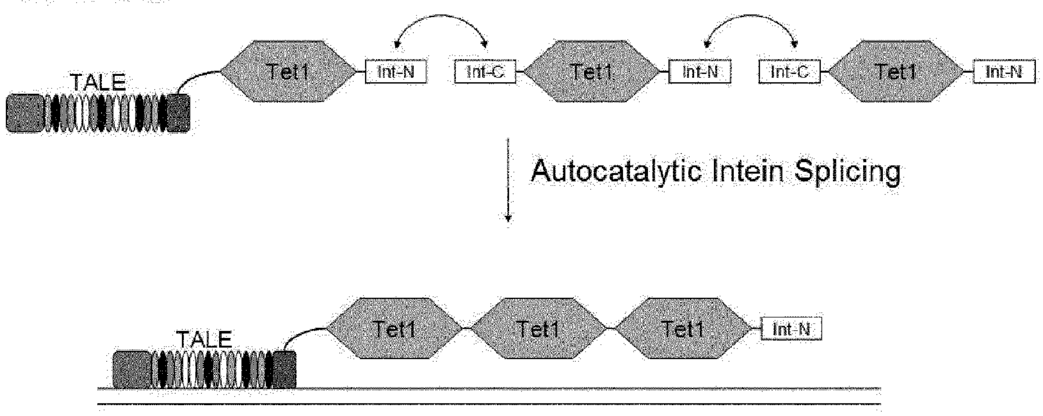

TALE
Tet1
Int-N
Int-C
Tet1
Int-N
Int-C
Tet1
Int-N
Autocatalytic Intein Splicing
TALE
Tet1
Tet1
Tet1
Int-N

FIG. 6D.

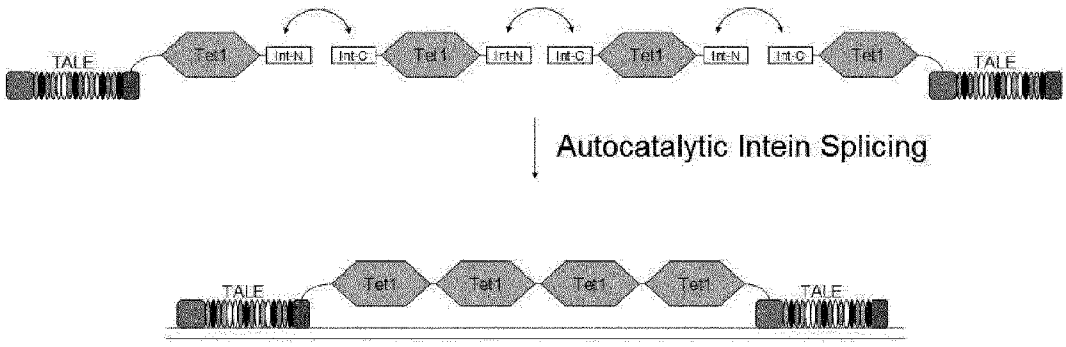
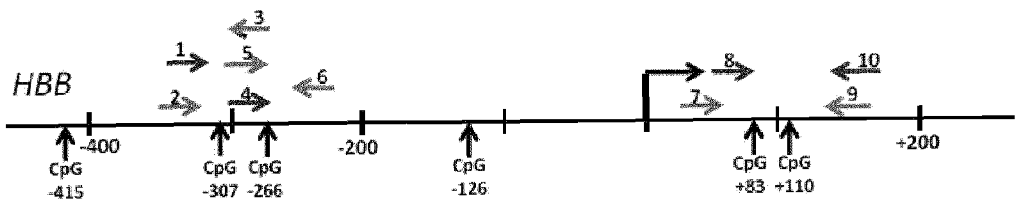

TALE
Tet1
Int-N
Int-C
Tet1
Int-N
Int-C
Tet1
Int-N
Int-C
Tet1
TALE
Autocatalytic Intein Splicing
TALE
Tet1
Tet1
Tet1
Tet1
TALE

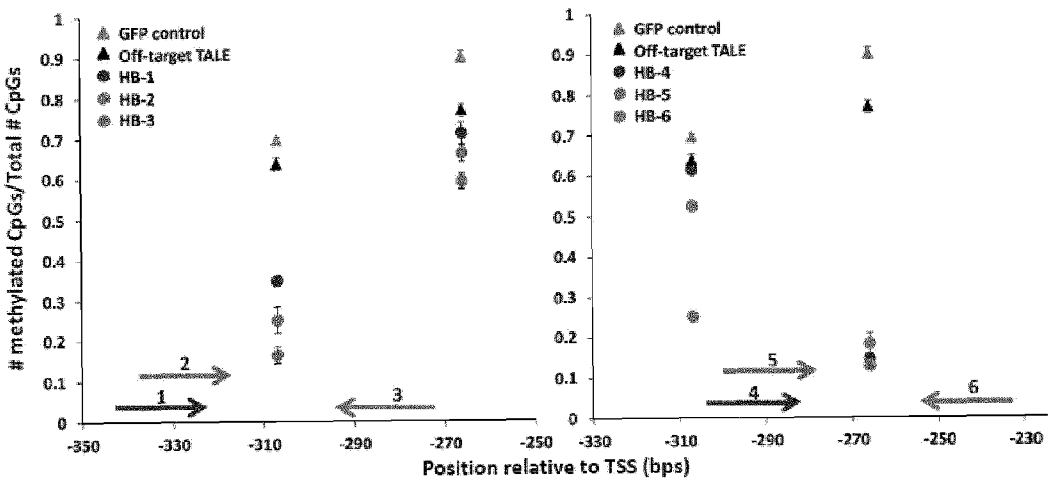
FIG. 7

Disease /condition	Gene	Reference
Cancer		
Bladder	Multiple genes	1
Brain (glioma)	RASSFIA	2, 3
Brain (glioblastoma)	MGMT	4
Breast	BRCA1	5
Breast	Multiple genes	6, 7
Cervix	P16	8
Colon	Multiple genes	1
Esophagus	CDH1	1
Head & neck	pf6, MGMT	1
Kidney	TIMP-3	1
Leukemia	p15	1
Liver	Multiple genes	9
Lung	pl6, p73	1
Lymphoma	DAPK	1
Myeloma	DAPK	10
Ovary	BRCA1	11
Pancreas	APC	1
Prostate	BRCA2	1, 12
Rhabdomyosarcoma	PAX3	13
Uterus	hMLHI	1
Neurologic		
Schizophrenia	RELN	14, 15
Cardiovascular		
Atherosclerosis	Multiple genes	16



HBB
1
2
3
4
5
6
7
8
9
10
-400
-200
+200
CpG -415
CpG -307
CpG -266
CpG -126
CpG +83
CpG +110

FIG. 8A



GFP control
Off-target TALE
HB-1
HB-2
HB-3
HB-4
HB-5
HB-6
methylated CpGs/Total # CpGs
Position relative to TSS (bps)

FIG. 8B

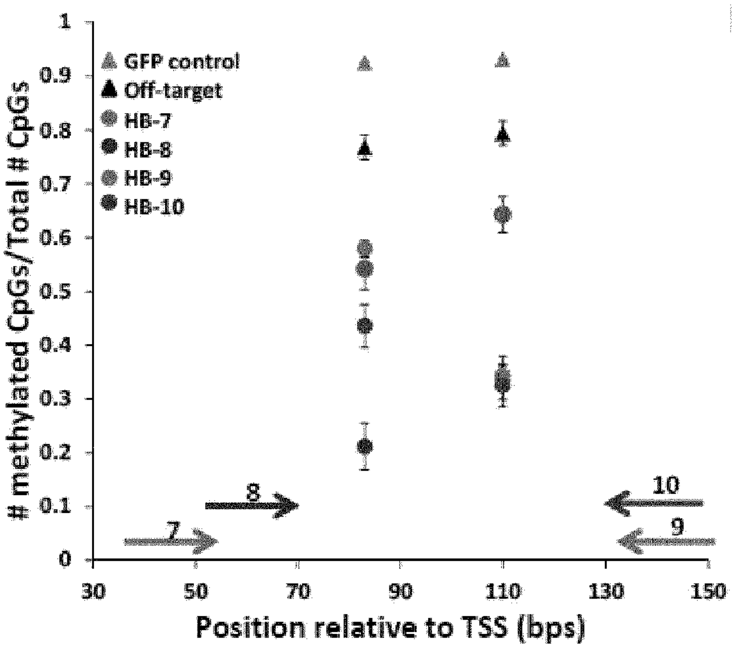

GFP control
Off-target
HB-7
HB-8
HB-9
HB-10
methylated CpGs/Total # CpGs
Position relative to TSS (bps)
8
7
10
9

FIG. 8C

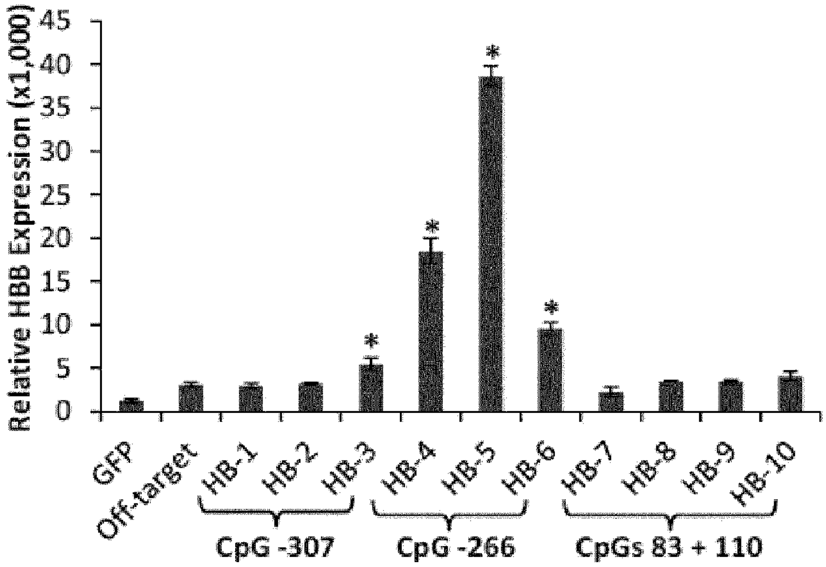

Relative HBB Expression (x1,000)
GFP
Off-target
HB-1
HB-2
HB-3
HB-4
HB-5
HB-6
HB-7
HB-8
HB-9
HB-10
CpG -307
CpG -266
CpGs 83 + 110

FIG. 8D

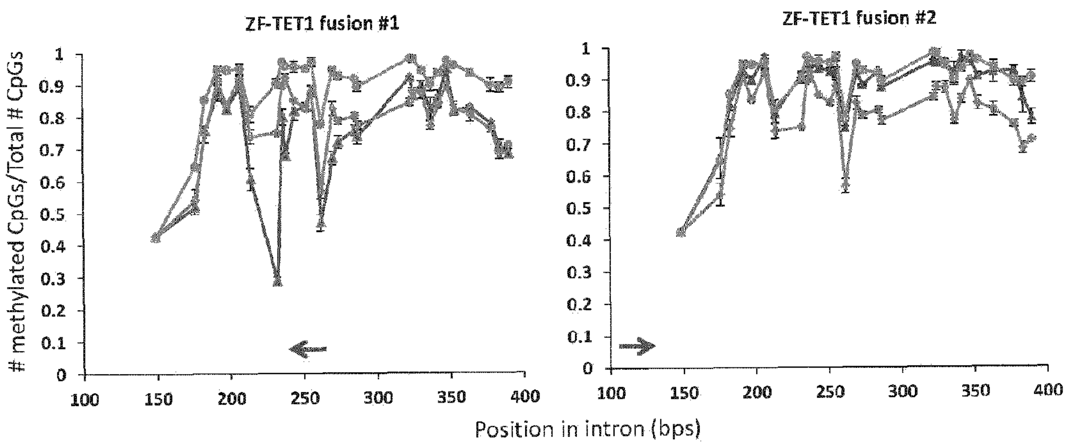

ZF-TET1 fusion #1
ZF-TET1 fusion #2
methylated CpGs/Total # CpGs
Position in intron (bps)

FIG. 9A

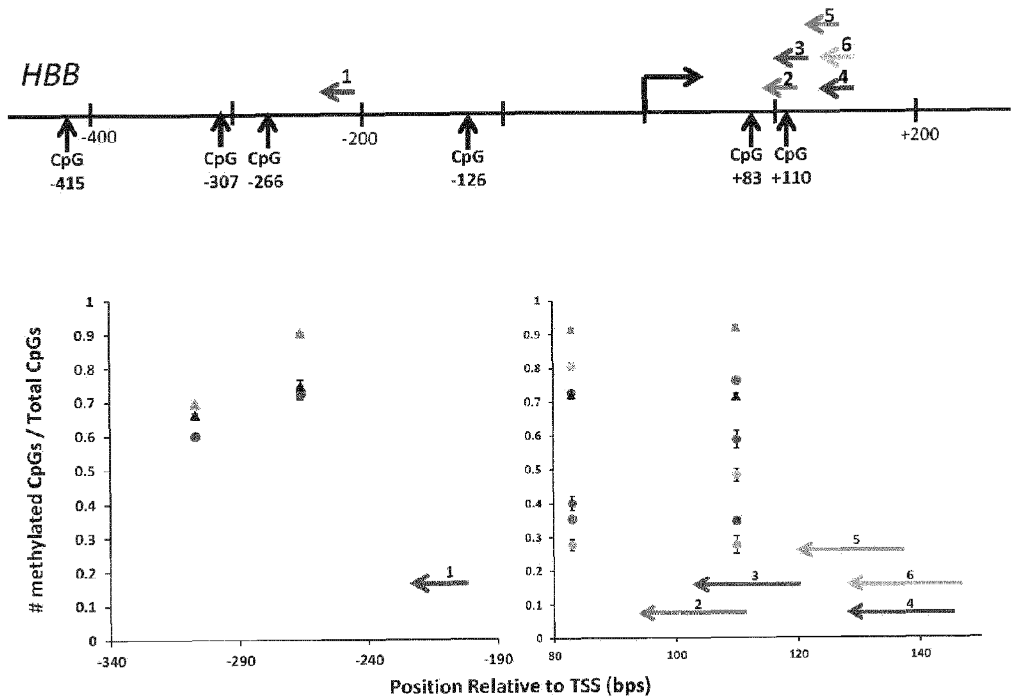

HBB
CpG -415
CpG -307
CpG -266
CpG -126
CpG +83
CpG +110
-400
-200
+200
1
2
3
4
5
6
methylated CpGs / Total CpGs
Position Relative to TSS (bps)

FIG. 9B

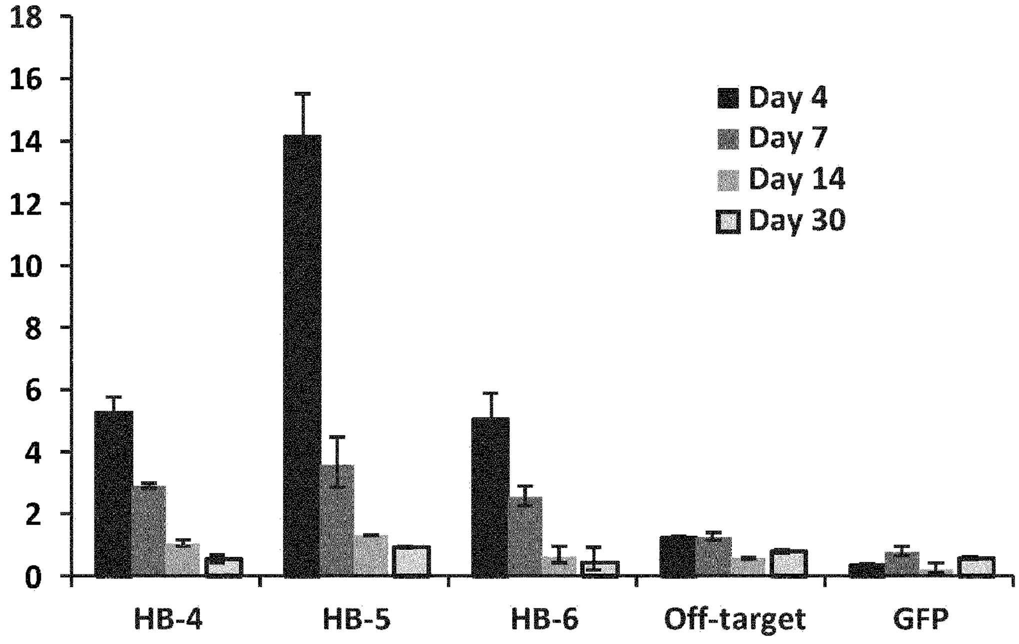

18
16
14
12
10
8
6
4
2
0
Day 4
Day 7
Day 14
Day 30
HB-4
HB-5
HB-6
Off-target
GFP

FIG. 10A

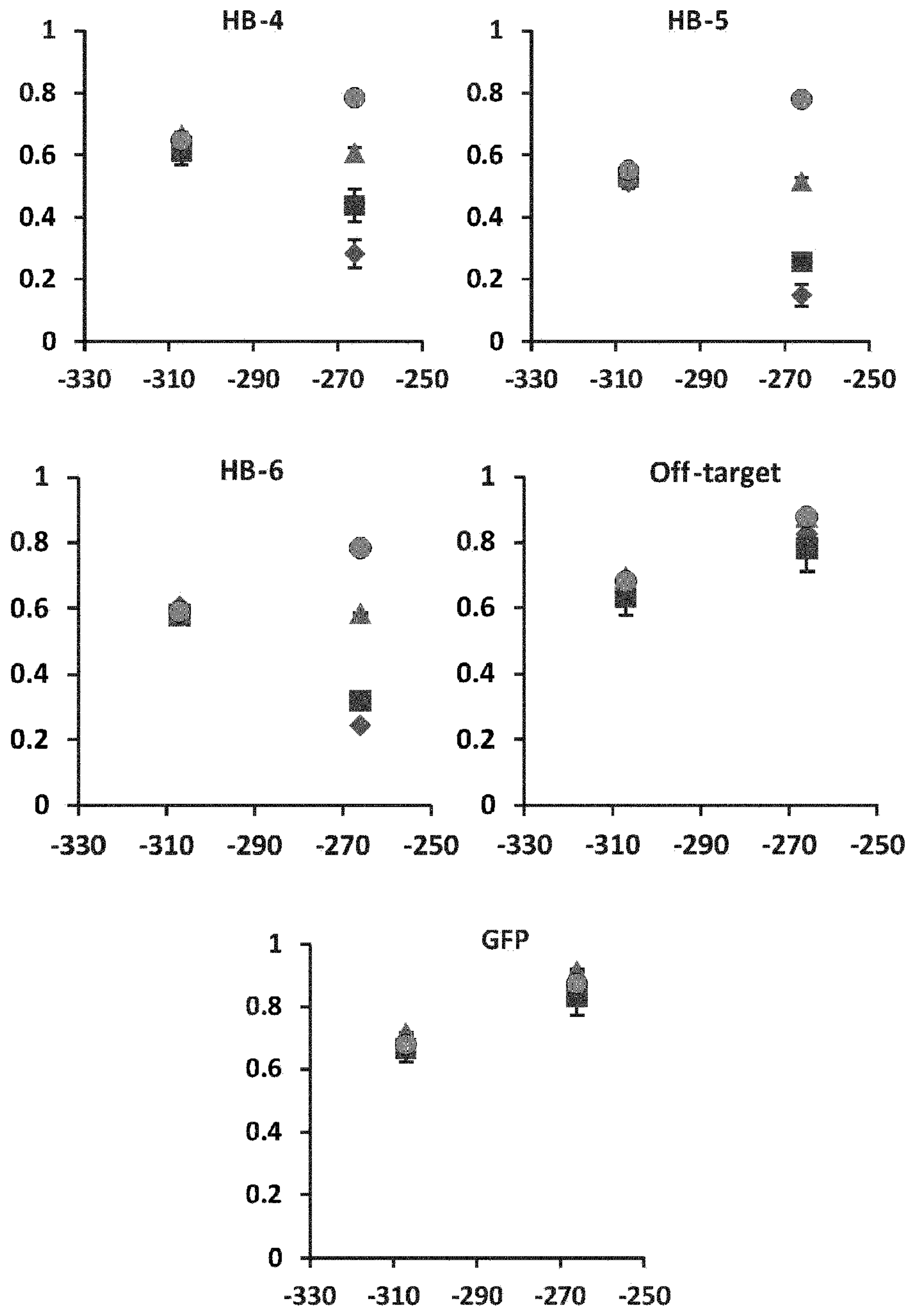


FIG. 10B

SEKVENSLISTE

Sekvenslisten er udeladt af skriftet og kan hentes fra det Europæiske Patent Register.

The Sequence Listing was omitted from the document and can be downloaded from the European Patent Register.