

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7207665号
(P7207665)

(45)発行日 令和5年1月18日(2023.1.18)

(24)登録日 令和5年1月10日(2023.1.10)

(51)国際特許分類

F I

C 1 2 Q 1/02 (2006.01)

C 1 2 Q 1/02

C 1 2 Q 1/6811(2018.01)

C 1 2 Q 1/6811

Z Z N A

C 1 2 N 15/09 (2006.01)

C 1 2 N 15/09 1 0 0

C 1 2 N 15/31 (2006.01)

C 1 2 N 15/09 1 1 0

C 1 2 N 15/31

請求項の数 4 (全37頁)

(21)出願番号 特願2020-528500(P2020-528500)

(86)(22)出願日 平成29年8月4日(2017.8.4)

(65)公表番号 特表2020-531572(P2020-531572
A)

(43)公表日 令和2年11月5日(2020.11.5)

(86)国際出願番号 PCT/CN2017/095988

(87)国際公開番号 WO2019/024081

(87)国際公開日 平成31年2月7日(2019.2.7)

審査請求日 令和2年4月7日(2020.4.7)

(73)特許権者 507232478

北京大学

PEKING UNIVERSITY

中華人民共和国北京市▲海▼淀区▲頤▼
和▲園▼路5号No. 5, Yiheyuan Road
, Haidian District,
Beijing 100871, Chi
na

(73)特許権者 520042939

博雅▲輯▼因(北京)生物科技有限公司
EDIGENE INC.中華人民共和国 102206 ペキン
チャンビン ディストリクト, 22 ケシ
ュエユアン ロード, ビルディング 2,
最終頁に続く

(54)【発明の名称】 メチル化により修飾されたDNA塩基を特異的に認識するT A L E R V Dおよびその使用

(57)【特許請求の範囲】

【請求項1】

細胞ゲノム中の標的配列の特定の部位に5 m Cが存在するか否かを検出する方法であって、

(1) T A L Eを含むタンパク質を細胞に導入し(但し、ヒト生体内の細胞にT A L Eを含むタンパク質を導入するのを除く)、前記T A L Eは標的配列を標的とし、前記T A L E内の、前記特定の部位を認識するR V DがH Aである；

(2) 次に、ヌクレアーゼを細胞に導入し(但し、ヒト生体内の細胞にヌクレアーゼを導入するのを除く)、前記ヌクレアーゼの標的切断部位をT A L E標的配列に位置する；

(3) 標的配列が切断されているかどうかを検出し、それによって標的配列の特定の部位に5 m Cが存在するかどうかを判断する；標的配列が切断されていない場合、前記T A L Eが前記標的配列に結合し、前記ヌクレアーゼが当該標的配列に結合して切断することができず、前記特定の部位に5 m Cが存在しているとされ；標的配列が切断されている場合、前記T A L Eが前記標的配列に結合しておらず、前記ヌクレアーゼが当該標的配列に結合して切断されており、前記特定の部位に5 m Cが存在しないとされている、ことを含む方法。

【請求項2】

前記ヌクレアーゼがエンドヌクレアーゼである、請求項1に記載の方法。

【請求項3】

前記ヌクレアーゼがC a s 9ヌクレアーゼである、請求項1に記載の方法。

【請求項 4】

ステップ（１）において Cas 9 ヌクレアーゼと sgRNA を共に細胞に導入する、請求項 3 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、DNA 結合タンパク質を用いて遺伝子を調節、編集および検出する技術に関する。

【背景技術】

【0002】

転写アクチベーター様エフェクター（transcription activator-like effector, TALE）は、植物病原菌キサントモナス属（Xanthomonas）に由来する毒性因子であり、真核生物ゲノムを再プログラム（reprogram）することができる（１，２）。TALE は、可変数のタンDEM 繰返し単位で構成される DNA 結合ドメインを含む（３）。各反復は、１２位と１３位に二つの可変性の高いアミノ酸（反復可変二重残基、repeat-variable diresidues または RVD）を有する他、３３～３５のアミノ酸残基のコンセンサス配列（consensus sequence）を含む（４，５）。TALE タンパク質の DNA 認識は、タンDEM 繰返し単位によって媒介される。タンDEM 繰返し単位は、RVD を介してヌクレオチドを標的とし、配列特異的な方法で DNA に結合し、RVD はヌクレオチドの特異性を決定する（４，６）。RVDs は DNA 塩基と直接的、配列特異的に接触し、モジュール化された DNA 認識特性により、TALE は転写活性化因子（７，８）、転写抑制因子（９，１０）またはエンドヌクレアーゼ（１１，１２）のような機能ドメインと融合することができ、これはプログラマブル遺伝子編集ツールと呼ばれる。既存の研究では、実験方法と計算方法を使用して RVD-DNA 認識コードを部分的にデコードした結果（４，６）、もっともよく使用される四つの RVD（NI、NG、HD および NN）がそれぞれ A、T、C および G/A に優先的に結合することは分かった（４，６）。

【0003】

四つの通常のデオキシリボヌクレオチドの他に、哺乳動物ゲノムはさらに、修飾された DNA 塩基を含む。例えば、５番目の DNA 塩基と呼ばれる 5-メチルシトシン（5mC）は、遺伝子発現を調節する重要なエピジェネティックマーカーである（図 1A）（１５，１６）。5mC は、１０～１１転座酵素（TET）ファミリーのタンパク質によって相次いで酸化されて 5-ヒドロキシメチルシトシン（5hmC）、5-ホルミルシトシン（5fC）および 5-カルボキシシトシン（5caC）を生成することができ、5-ホルミルシトシン（5fC）と 5-カルボキシシトシン（5caC）はチミン DNA グリコシラーゼの基質であり、最終的に修飾されていないシトシンに戻る（１７，２２）。5hmC は、～１％～１０％の修飾されたシトシンを構成し、安定したエピジェネティックマーカーと見なされ、5hmC の調節異常は癌で頻繁に観察される。

【0004】

シトシンのメチル化修飾の他に、もう一つのよく見られる DNA メチル化修飾 N⁶-メチルアデニン（N⁶-methyladenine, 6mA）は、DNA のアデニン上の共有結合修飾として原核細胞において重要な役割を果たし、複数の生物学的経路の調節に参与し、制限修飾系（restriction-modification, RM）の一部として外来 DNA の侵入を抵抗し、DNA 複製、ミスマッチ修復、遺伝子転写および転座などの過程において調節制御の役割などを果たす（４１，４７）。真核生物における 6mA に関する研究は比較的に少なく、エピジェネティクスにおける 6mA の役割は明確ではない（４６）。

【0005】

２０１５年に Cell で発表された三つの文章には、クラミドモナス、線虫、キイロシヨウジョウバエなどの真核生物のゲノムでの 6mA が報告された（４２，４３，４８）。

10

20

30

40

50

クラミドモナス (*Chlamydomonas reinhardtii*) 中の 6-メチルシトシンの位置を測定した結果、クラミドモナスのほとんどの遺伝子には 6mA の分布があり、しかもそれらのほとんどが A p T 二重塩基モードで現れており、それと同時に、6mA は転写開始部位で濃縮され、アクティブな遺伝子発現に関連している (42)。キイロショウジョウバエ (*Drosophila melanogaster*) と線虫 (*Caenorhabditis elegans*) に対する研究では、6mA が分化と発達の過程において重要な調節制御の役割を果たす可能性が高いことが分かった (48)。6mA のメチル化および脱メチル化に関連する酵素は進化において保存的であることが発見されたため、6mA は他の真核生物にも分布している可能性が高い (43)。2016 年まで、Koziołらは、アフリカツメガエル (*Xenopus laevis*) の異なる組織およびマウスとヒトの組織または細胞株を含む、脊椎動物のゲノムに 6mA が存在することを実証した。脊椎動物における 6mA 修飾の豊富さは非常に低く、クラミドモナスやキイロショウジョウバエとは違って、アフリカツメガエルとマウスゲノムにおいて 6mA はエクソン以外の領域に広く分布しているとともに、一定の配列モチーフを示していることは研究により明らかになり、これは、異なる真核生物において 6mA 修飾が異なる機能を持つ可能性があることを示している (44)。6mA の後成的修飾が高等生物における分布と、細胞及び個体の発達などの過程における役割とメカニズムをさらに研究する必要がある。

【0006】

TAL E タンパク質は、修飾された DNA 塩基を認識することができることは報告されている (24~26)。例えば、NG または N* (アスタリスクは、13 位のアミノ酸の欠失を表す) が相同 DNA 中の 5mC を認識でき (25, 27~31)、NG/N* と HD との組み合わせがインビトロ検出で 5mC/5hmC と C との区別づけに使用されていることは報告された (32)。最近の研究では、切られた反復ループ (G*, S*, および T*) を持つ TAL E タンパク質が、類似した親和性で C、5mC、5hmC、5fC および 5caC に結合することができることは報告されている (33, 34)。TAL E-DNA 複合体の結晶構造では、RVD ループは DNA の二重らせんの主溝 (duplex major groove) と接触し、1 番目の残基は適切なループ構成を安定化にし、2 番目の残基は直接的な塩基特異的な接触をする (35, 36)。現在、RVD が 5mC、5hmC および 6mA を認識する潜在能力についてさらに研究する必要がある。

【先行技術文献】

【非特許文献】

【0007】

【文献】Kay S & Bonas U (2009) How Xanthomonas type III effectors manipulate the host plant. *Curr Opin Microbiol* 12 (1):37-43.
 Kay S, Hahn S, Marois E, Hause G, & Bonas U (2007) A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318 (5850):648-651.
 Boch J & Bonas U (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* 48:419-436.
 Boch J, et al. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326 (5959):1509-1512.
 Gurlbeck D, Thieme F, & Bonas U (2006) Type III effector proteins from the plant pathogen Xanthomonas and their role in the interaction with the host plant. *J Plant Ph*

ysiol 163 (3) : 233–255.

Moscou MJ & Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326 (5959) : 1501.

Bogdanove AJ & Voytas DF (2011) TAL effectors: customizable proteins for DNA targeting. Science 333 (6051) : 1843–1846.

Morbitzer R, Romer P, Boch J, & Lahaye T (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. Proc Natl Acad Sci U S A 107 (50) : 21617–21622.

Cong L, Zhou R, Kuo YC, Cunniff M, & Zhang F (2012) Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nat Commun 3 : 968.

Garg A, Lohmueller JJ, Silver PA, & Armel TZ (2012) Engineering synthetic TAL effectors with orthogonal target sites. Nucleic Acids Res 40 (15) : 7584–7595.

Christian M, et al. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186 (2) : 757–761.

Miller JC, et al. (2011) A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29 (2) : 143–148.

Yang J, et al. (2014) Complete decoding of TAL effectors for DNA recognition. Cell research 24 (5) : 628–631.

Miller JC, et al. (2015) Improved specificity of TALE-based genome editing using an expanded RVD repertoire. Nat Methods 12 (5) : 465–471.

Kohli RM & Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. Nature 502 (7472) : 472–479.

Pastor WA, Aravind L, & Rao A (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol 14 (6) : 341–356.

Kriaucionis S & Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324 (5929) : 929–930.

Tahiliani M, et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324 (5929) : 930–935.

10

20

30

40

50

Ito S, et al. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466 (7310) :1129-1133.

He YF, et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333 (6047) :1303-1307.

Maiti A & Drohat AC (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* 286 (41) :35334-35338.

Pfaffeneder T, et al. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl* 50 (31) :7008-7012.

Huang Y & Rao A (2014) Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet* 30 (10) :464-474.

Bultmann S, et al. (2012) Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res* 40 (12) :5368-5377.

Valton J, et al. (2012) Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J Biol Chem* 287 (46) :38427-38432.

Kim Y, et al. (2013) A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol* 31 (3) :251-258.

Deng D, et al. (2012) Recognition of methylated DNA by TAL effectors. *Cell research* 22 (10) :1502-1504.

Dupuy A, et al. (2013) Targeted gene therapy of xeroderma pigmentosum cells using meganuclease and TALEN. *PLoS One* 8 (11) :e78678.

Hu J, et al. (2014) Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic Acids Res* 42 (7) :4375-4390.

Kubik G, Schmidt MJ, Penner JE, & Summerer D (2014) Programmable and highly resolved in vitro detection of 5-methylcytosine by TALEs. *Angew Chem Int Ed Engl* 53 (23) :6002-6006.

Kubik G & Summerer D (2015) Achieving single-nucleotide resolution of 5-methylcytosine detection with TALEs. *Chembiochem* 16 (

10

20

30

40

50

2) : 228–231.

Kubik G, Batke S, & Summerer D (2015) Programmable sensors of 5-hydroxymethylcytosine. *J Am Chem Soc* 137 (1) : 2–5.

Maurer S, Giess M, Koch O, & Summerer D (2016) Interrogating Key Positions of Size-Reduced TALE Repeats Reveals a Programmable Sensor of 5-Carboxylcytosine. *ACS Chem Biol* 11 (12) : 3294–3299.

Rathi P, Maurer S, Kubik G, & Summerer D (2016) Isolation of Human Genomic DNA Sequences with Expanded Nucleobase Selectivity. *J Am Chem Soc* 138 (31) : 9910–9918.

Deng D, et al. (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335 (6069) : 720–723.

Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, & Stoddard BL (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335 (6069) : 716–719.

Yang J, et al. (2013) ULTIMATE system for rapid assembly of customized TAL effectors. *PLoS One* 8 (9) : e75649.

Wu H & Zhang Y (2015) Charting oxidized methylcytosines at base resolution. *Nat Struct Mol Biol* 22 (9) : 656–661.

Yu M, et al. (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149 (6) : 1368–1380.

Hsu PD, Lander ES, & Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157 (6) : 1262–1278.

Mussolino C, et al. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* 39 (21) : 9283–9293.

Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M.C., Jabado, O.J., et al. (2012). Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nature biotechnology* 30, 1232–1239.

Fu, Y., Luo, G.Z., Chen, K., Deng, X., Yu, M., Han, D., Hao, Z., Liu, J., Lu, X., Dore, L.C., et al. (2015). N6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell* 161, 879–892.

Greer, E.L., Blanco, M.A., Gu, L., Sendinc

10

20

30

40

50

, E., Liu, J., Aristizabal-Corrales, D., Hsu, C. H., Aravind, L., He, C., and Shi, Y. (2015). DNA Methylation on N6-Adenine in *C. elegans*. *Cell* 161, 868-878.

Kozioł, M. J., Bradshaw, C. R., Allen, G. E., Costa, A. S., Frezza, C., and Gurdon, J. B. (2016). Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nature structural & molecular biology* 23, 24-30.

Mak, A. N., Bradley, P., Cernadas, R. A., Bogdanove, A. J., and Stoddard, B. L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335, 716-719. Ratel, D., Ravanat, J. L., Berger, F., and Wion, D. (2006). N6-methyladenine: the other methylated base of DNA. *BioEssays: news and reviews in molecular, cellular and developmental biology* 28, 309-315.

Wion, D., and Casadesus, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nature reviews Microbiology* 4, 183-192.

Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J., et al. (2015). N6-methyladenine DNA modification in *Drosophila*. *Cell* 161, 893-906.

【発明の概要】

【0008】

本発明は、5mC、5hmCおよび6mAに対して認識選好、およびこれらの後成的修飾に対して異なる結合特性を有するRVDを同定した。これらのRVDは、メチル化依存遺伝子活性化、効率的なゲノム編集、および5hmCに対する標的検出などにおいて使用することができる。

【0009】

本発明は一態様において、以下から選択される一つ以上のRVDを含むTALEを含む、単離されたDNA結合ポリペプチドを提供する：

5mCを特異的に認識できるHAまたはNA；

5hmCを特異的に認識できるFS；

Cと5mCの両方を認識できるN*、NGまたはKP；

Cと5hmCの両方を認識できるHVまたはKV；

5mCと5hmCの両方を認識できるK*またはRG；

C、5mCおよび5hmCの三つを認識できるG*、H*、R*、またはY*；

6mAを特異的に認識できるNP、FT、CVまたはCP；または

Aと6mAの両方を認識であるRI、NI、KIまたはHI；

ここで、*はその位置でのアミノ酸が欠失していることを表す。

【0010】

本発明はもう一態様において、以下から選択される一つ以上のRVDを含むTALEと、機能ドメインとを含む、融合タンパク質を提供する：

5mCを特異的に認識できるHAまたはNA；

5 h m C を特異的に認識できる F S ;
C と 5 m C の両方を認識できる N *、N G または K P ;
C と 5 h m C の両方を認識できる H V または K V ;
5 m C と 5 h m C の両方を認識できる K * または R G ;
C、5 m C および 5 h m C の三つを認識できる G *、H *、R *、または Y * ;
6 m A を特異的に認識できる N P、F T、C V または C P ; または
A と 6 m A の両方を認識である R I、N I、K I または H I ;
ここで、* はその位置でのアミノ酸が欠失していることを表す。

【0011】

一部の実施形態では、前記機能ドメインが、遺伝子発現を調節する機能ドメイン、後成的修飾機能ドメイン、遺伝子編集機能ドメイン、または蛍光タンパク質である。

10

【0012】

一部の実施形態では、前記遺伝子発現を調節する機能ドメインは、転写活性化因子、転写抑制因子またはその機能的フラグメントであり、前記後成的修飾機能ドメインは、メルトランスフェラーゼ、デメチラーゼまたはその機能的フラグメントであり、前記遺伝子編集機能ドメインは、ヌクレアーゼまたはその機能的フラグメントである。

【0013】

一部の実施形態では、前記遺伝子編集機能ドメインはエンドヌクレアーゼ、好ましくは F o k I エンドヌクレアーゼ、より好ましくは F o k I エンドヌクレアーゼの DNA 切断ドメインである。

20

【0014】

本発明はもう一態様において、上記 DNA 結合ポリペプチドまたは上記いずれか一つの融合タンパク質をコードするポリヌクレオチドを提供する。

【0015】

本発明はもう一態様において、上記ポリヌクレオチドを含むベクターを提供する。

【0016】

本発明はもう一態様において、上記ポリヌクレオチドまたは上記ベクターを含む宿主細胞を提供する。

【0017】

本発明はもう一態様において、

30

(1) T A L E 反復ドメインの一つ以上の R V D が H A または N A である、標的遺伝子の標的配列中のメチル化された塩基 5 m C を検出するための試薬の調製における、T A L E 反復ドメインを含むタンパク質の使用；

(2) T A L E 反復ドメインの一つ以上の R V D が F S である、標的遺伝子の標的配列中のメチル化された塩基 5 h m C を検出するための試薬の調製における、T A L E 反復ドメインを含むタンパク質の使用；または

(3) T A L E 反復ドメインの一つ以上の R V D が N P、F T、C V または C P である、標的遺伝子の標的配列中のメチル化された塩基 6 m C を検出するための試薬の調製における、T A L E 反復ドメインを含むタンパク質の使用；

を含む、標的遺伝子の標的配列中のメチル化された塩基を検出するための試薬の調製における、T A L E 反復ドメインを含むタンパク質の使用を提供する。

40

【0018】

本発明はもう一態様において、細胞における標的遺伝子の標的配列を標的結合するための試薬の調製における、上記 DNA 結合ポリペプチド、上記いずれか一つの融合タンパク質、上記ポリヌクレオチド、上記ベクターまたは上記宿主細胞の使用を提供する。

【0019】

本発明はもう一態様において、細胞における標的遺伝子発現を調節するための試薬の調製における、上記いずれか一つの融合タンパク質またはその融合タンパク質をコードするポリヌクレオチドの使用を提供し、前記融合タンパク質に含まれる機能ドメインは、遺伝子発現を調節する機能ドメインである。

50

【0020】

一部の実施形態では、前記遺伝子発現を調節する機能ドメインは、転写活性化因子またはその機能的フラグメント、若しくは転写抑制因子またはその機能的フラグメントである。

【0021】

本発明はもう一態様において、細胞における標的遺伝子に対して遺伝子編集をするための試薬の調製における、上記いずれか一つの融合タンパク質またはその融合タンパク質をコードするポリヌクレオチドの使用を提供し、前記融合タンパク質に含まれる機能ドメインは遺伝子編集機能ドメインである。

【0022】

一部の実施形態では、前記遺伝子編集は核酸切断であり、前記遺伝子編集機能ドメインはヌクレアーゼまたはその機能的フラグメント、好ましくはエンドヌクレアーゼまたはその機能的フラグメント、より好ましくはFokIエンドヌクレアーゼまたはそのDNA切断ドメインである。

【0023】

本発明はもう一態様において、細胞における標的遺伝子に対して後成的修飾をするための試薬の調製における、上記いずれか一つの融合タンパク質またはその融合タンパク質をコードするポリヌクレオチドの使用を提供し、前記融合タンパク質に含まれる機能ドメインは、後成的修飾機能ドメインである。

【0024】

一部の実施形態では、前記後成的修飾機能ドメインはメチルトランスフェラーゼ、デメチラーゼまたはその機能的フラグメントである。

【0025】

本発明はもう一態様において、上記DNA結合ポリペプチド、上記いずれか一つの融合タンパク質、または上記ポリヌクレオチドを細胞に導入して前記DNA結合ポリペプチドまたは融合タンパク質中のTALEを標的遺伝子の標的配列に結合させることを含む、細胞における標的遺伝子の標的配列を標的結合する方法を提供する。

【0026】

一部の実施形態では、上記方法において、

前記DNA結合ポリペプチドまたは融合タンパク質におけるTALEは、HAまたはNAから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列が5mCである場合しか、前記DNA結合ポリペプチドまたは融合タンパク質におけるTALEは標的遺伝子の標的配列に結合しない；

前記DNA結合ポリペプチドまたは融合タンパク質におけるTALEは、FSから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列が5hmCである場合しか、前記DNA結合ポリペプチドまたは融合タンパク質におけるTALEは標的遺伝子の標的配列に結合しない；

前記DNA結合ポリペプチドまたは融合タンパク質におけるTALEは、NP、FT、CVまたはCPから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列が6mAである場合しか、前記DNA結合ポリペプチドまたは融合タンパク質におけるTALEは標的遺伝子の標的配列に結合しない；

前記DNA結合ポリペプチドまたは融合タンパク質におけるTALEは、N*、NGまたはKPから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、Cまたは5mCであり得る；

前記DNA結合ポリペプチドまたは融合タンパク質におけるTALEは、HVまたはKVから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、Cまたは5hmCであり得る；

前記DNA結合ポリペプチドまたは融合タンパク質におけるTALEは、K*またはRGから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、5mCまたは5hmCであり得る；

前記DNA結合ポリペプチドまたは融合タンパク質におけるTALEは、G*、H*、

10

20

30

40

50

R*またはY*から選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、C、5mCまたは5hmCであり得る；または

前記DNA結合ポリペプチドまたは融合タンパク質におけるTALEは、RI、NI、KIまたはHIから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、Aまたは6mAであり得る；

ここで、*はその位置でのアミノ酸が欠失していることを表す。

【0027】

本発明はもう一態様において、上記いずれか一つの融合タンパク質、または前記融合タンパク質をコードするポリヌクレオチドを細胞に導入して前記融合タンパク質中のTALEを標的遺伝子の標的配列に結合させ、これによって、融合タンパク質中の、遺伝子発現を調節する機能ドメインである機能ドメインによって標的遺伝子の発現が調節されることを含む、細胞における標的遺伝子の発現を調節する方法を提供する。

【0028】

一部の実施形態では、上記方法において、

前記融合タンパク質におけるTALEは、HAまたはNAから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列が5mCである場合しか、前記融合タンパク質におけるTALEは標的遺伝子の標的配列に結合しない；

前記融合タンパク質におけるTALEは、FSから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列が5hmCである場合しか、前記融合タンパク質におけるTALEは標的遺伝子の標的配列に結合しない；

前記融合タンパク質におけるTALEは、NP、FT、CVまたはCPから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列が6mAである場合しか、前記融合タンパク質におけるTALEは標的遺伝子の標的配列に結合しない；

前記融合タンパク質におけるTALEは、N*、NGまたはKPから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、Cまたは5mCであり得る；

前記融合タンパク質におけるTALEは、HVまたはKVから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、Cまたは5hmCであり得る；

前記融合タンパク質におけるTALEは、K*またはRGから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、5mCまたは5hmCであり得る；

前記融合タンパク質におけるTALEは、G*、H*、R*またはY*から選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、C、5mCまたは5hmCであり得る；または

前記融合タンパク質におけるTALEは、RI、NI、KIまたはHIから選択されるRVDを含み、前記RVDの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、Aまたは6mAであり得る；

ここで、*はその位置でのアミノ酸が欠失していることを表す。

【0029】

一部の実施形態では、上記方法において、前記遺伝子発現を調節する機能ドメインは、転写活性化因子またはその機能的フラグメント、若しくは転写抑制因子またはその機能的フラグメントである。

【0030】

本発明はもう一態様において、上記いずれか一つの融合タンパク質、または前記融合タンパク質をコードするポリヌクレオチドを細胞に導入して前記融合タンパク質中のTALEを標的遺伝子の標的配列に結合させ、これによって、融合タンパク質中の、遺伝子編集機能ドメインである機能ドメインによって標的遺伝子が編集されることを含む、細胞における標的遺伝子に対して遺伝子編集をする方法を提供する。

【0031】

一部の実施形態では、上記方法において、

前記融合タンパク質におけるT A L Eは、H AまたはN Aから選択されるR V Dを含み、前記R V Dの認識部位での前記標的遺伝子の標的配列が5 m Cである場合しか、前記融合タンパク質におけるT A L Eは標的遺伝子の標的配列に結合しない；

前記融合タンパク質におけるT A L Eは、F Sから選択されるR V Dを含み、前記R V Dの認識部位での前記標的遺伝子の標的配列が5 h m Cである場合しか、前記融合タンパク質におけるT A L Eは標的遺伝子の標的配列に結合しない；

前記融合タンパク質におけるT A L Eは、N P、F T、C VまたはC Pから選択されるR V Dを含み、前記R V Dの認識部位での前記標的遺伝子の標的配列が6 m Aである場合しか、前記融合タンパク質におけるT A L Eは標的遺伝子の標的配列に結合しない；

前記融合タンパク質におけるT A L Eは、N *，N GまたはK Pから選択されるR V Dを含み、前記R V Dの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、Cまたは5 m Cであり得る；

前記融合タンパク質におけるT A L Eは、H VまたはK Vから選択されるR V Dを含み、前記R V Dの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、Cまたは5 h m Cであり得る；

前記融合タンパク質におけるT A L Eは、K *またはR Gから選択されるR V Dを含み、前記R V Dの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、5 m Cまたは5 h m Cであり得る；

前記融合タンパク質におけるT A L Eは、G *、H *、R *またはY *から選択されるR V Dを含み、前記R V Dの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、C、5 m Cまたは5 h m Cであり得る；または

前記融合タンパク質におけるT A L Eは、R I、N I、K IまたはH Iから選択されるR V Dを含み、前記R V Dの認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、Aまたは6 m Aであり得る；

ここで、*はその位置でのアミノ酸が欠失していることを表す。

【0032】

一部の実施形態では、上記方法において、前記遺伝子編集は核酸切断であり、前記遺伝子編集機能ドメインはヌクレアーゼまたはその機能的フラグメント、好ましくはエンドヌクレアーゼまたはその機能的フラグメント、より好ましくはF o k IエンドヌクレアーゼまたはそのDNA切断ドメインである。

【0033】

本発明はもう一態様において、上記いずれか一つの融合タンパク質、または前記融合タンパク質をコードするポリヌクレオチドを細胞に導入して前記融合タンパク質中のT A L Eを標的遺伝子の標的配列に結合させ、これによって、融合タンパク質中の、後成的修飾機能ドメインである機能ドメインによって標的遺伝子が後成的修飾されることを含む、細胞における標的遺伝子に対して後成的修飾をする方法を提供する。

【0034】

一部の実施形態では、上記方法において、

前記融合タンパク質におけるT A L Eは、H AまたはN Aから選択されるR V Dを含み、前記R V Dの認識部位での前記標的遺伝子の標的配列が5 m Cである場合しか、前記融合タンパク質におけるT A L Eは標的遺伝子の標的配列に結合しない；

前記融合タンパク質におけるT A L Eは、F Sから選択されるR V Dを含み、前記R V Dの認識部位での前記標的遺伝子の標的配列が5 h m Cである場合しか、前記融合タンパク質におけるT A L Eは標的遺伝子の標的配列に結合しない；

前記融合タンパク質におけるT A L Eは、N P、F T、C VまたはC Pから選択されるR V Dを含み、前記R V Dの認識部位での前記標的遺伝子の標的配列が6 m Aである場合しか、前記融合タンパク質におけるT A L Eは標的遺伝子の標的配列に結合しない；

前記融合タンパク質におけるT A L Eは、N *，N GまたはK Pから選択されるR V D

10

20

30

40

50

を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、C または 5 m C であり得る；

前記融合タンパク質における T A L E は、H V または K V から選択される R V D を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、C または 5 h m C であり得る；

前記融合タンパク質における T A L E は、K * または R G から選択される R V D を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、5 m C または 5 h m C であり得る；

前記融合タンパク質における T A L E は、G *、H *、R * または Y * から選択される R V D を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、C、5 m C または 5 h m C であり得る；または

前記融合タンパク質における T A L E は、R I、N I、K I または H I から選択される R V D を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、A または 6 m A であり得る；

ここで、* はその位置でのアミノ酸が欠失していることを表す。

【0035】

一部の実施形態では、上記方法において、前記後成的修飾機能ドメインはメチルトランスフェラーゼ、デメチラーゼまたはその機能的フラグメントである。

【0036】

本発明はもう一態様において、上記いずれか一つの融合タンパク質、または前記融合タンパク質をコードするポリヌクレオチドを細胞に導入して前記融合タンパク質中の T A L E を標的遺伝子の標的配列に結合させることを含む生細胞染色体の標識方法を提供し、前記機能ドメインが蛍光タンパク質であり、前記融合タンパク質中の T A L E を標的遺伝子の標的配列に結合させることによって標的配列に対する蛍光標識を実現する。

【0037】

一部の実施形態では、上記方法において、

前記融合タンパク質における T A L E は、H A または N A から選択される R V D を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列が 5 m C である場合しか、前記融合タンパク質における T A L E は標的遺伝子の標的配列に結合しない；

前記融合タンパク質における T A L E は、F S から選択される R V D を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列が 5 h m C である場合しか、前記融合タンパク質における T A L E は標的遺伝子の標的配列に結合しない；

前記融合タンパク質における T A L E は、N P、F T、C V または C P から選択される R V D を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列が 6 m A である場合しか、前記融合タンパク質における T A L E は標的遺伝子の標的配列に結合しない；

前記融合タンパク質における T A L E は、N *、N G または K P から選択される R V D を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、C または 5 m C であり得る；

前記融合タンパク質における T A L E は、H V または K V から選択される R V D を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、C または 5 h m C であり得る；

前記融合タンパク質における T A L E は、K * または R G から選択される R V D を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、5 m C または 5 h m C であり得る；

前記融合タンパク質における T A L E は、G *、H *、R * または Y * から選択される R V D を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、C、5 m C または 5 h m C であり得る；または

前記融合タンパク質における T A L E は、R I、N I、K I または H I から選択される R V D を含み、前記 R V D の認識部位での前記標的遺伝子の標的配列の特定の塩基はそのメチル化状態が不確定であり、A または 6 m A であり得る；

10

20

30

40

50

ここで、*はその位置でのアミノ酸が欠失していることを表す。

【0038】

本発明はもう一態様において、以下を含む、細胞ゲノム中の標的配列の特定の部位に 5 mC が存在するか否かを検出する方法を提供する：

(1) TALE を含むタンパク質を細胞に導入し、前記 TALE は標的配列を標的とし、前記 TALE 内の、前記特定の部位を認識する RVD が HA または NA である；

(2) 次に、ヌクレアーゼを細胞に導入し、前記ヌクレアーゼの標的切断部位を TALE 標的配列に位置する；

(3) 標的配列が切断されているかどうかを検出し、それによって標的配列の特定の部位に 5 mC が存在するかどうかを判断する；標的配列が切断されていない場合、前記 TALE が前記標的配列に結合し、前記ヌクレアーゼが当該標的配列に結合して切断することができず、前記特定の部位に 5 mC が存在するとされ；標的配列が切断されている場合、前記 TALE が前記標的配列に結合しておらず、前記ヌクレアーゼが当該標的配列に結合してそれを切断しており、前記特定の部位に 5 mC が存在しないとされている。

10

【0039】

本発明はもう一態様において、以下のステップを含む、細胞ゲノム中の標的配列の特定の部位に 5 hmC が存在するか否かを検出する方法を提供する：

(1) TALE を含むタンパク質を細胞に導入し、前記 TALE は標的配列を標的とし、前記 TALE 内の、前記特定の部位を認識する RVD が FS である；

(2) 次に、ヌクレアーゼを細胞に導入し、前記ヌクレアーゼの標的切断部位を TALE 標的配列に位置する；

(3) 標的配列が切断されているかどうかを検出し、それによって標的配列の特定の部位に 5 hmC が存在するかどうかを判断する；標的配列が切断されていない場合、前記 TALE が前記標的配列に結合し、前記ヌクレアーゼが当該標的配列に結合して切断することができず、前記特定の部位に 5 hmC が存在するとされ；標的配列が切断されている場合、前記 TALE が前記標的配列に結合しておらず、前記ヌクレアーゼが当該標的配列に結合してそれを切断しており、前記特定の部位に 5 hmC が存在しないとされている。

20

【0040】

本発明はもう一態様において、以下を含む、細胞ゲノム中の標的配列の特定の部位に 6 mA が存在するか否かを検出する方法を提供する：

(1) TALE を含むタンパク質を細胞に導入し、前記 TALE は標的配列を標的とし、前記 TALE 内の、前記特定の部位を認識する RVD が NP、FT、CV または CP である；

(2) 次に、ヌクレアーゼを細胞に導入し、前記ヌクレアーゼの標的切断部位を TALE 標的配列に位置する；

(3) 標的配列が切断されているかどうかを検出し、それによって標的配列の特定の部位に 6 mA が存在するかどうかを判断する；標的配列が切断されていない場合、前記 TALE が前記標的配列に結合し、前記ヌクレアーゼが当該標的配列に結合して切断することができず、前記特定の部位に 6 mA が存在するとされ；標的配列が切断されている場合、前記 TALE が前記標的配列に結合しておらず、前記ヌクレアーゼが当該標的配列に結合してそれを切断しており、前記特定の部位に 6 mA が存在しないとされている。

30

40

【0041】

一部の実施形態では、前記ヌクレアーゼがエンドヌクレアーゼである。

【0042】

一部の実施形態では、前記ヌクレアーゼが Cas9 ヌクレアーゼであり、しかもステップ (1) において前記 Cas9 ヌクレアーゼと sgRNA を共に細胞に導入する。

【図面の簡単な説明】

【0043】

【図1】修飾されたシトシンを認識するすべての潜在的な TALE RVDs を評価するためのスクリーニング概略図である。(a) は C、5 mC および 5 hmC の化学構造であ

50

る。(b)は、T A L E アクチベーターとG F P 発現レポーターDNAフラグメントで構成される、修飾されたシトシンに対して新しいR V D をスクリーニングするためのシステムの概略図である。(c)では、T A L E - (E *)₃が5 m C レポーターDNAフラグメントに対する場合のように、構築されたT A L E がレポーターDNAフラグメントに結合しない場合(左図)、G F P 発現はベースラインレベルにある(右図);逆に、T A L E - (G *)₃が5 m C レポーターDNAフラグメントに対する場合のように、T A L E がレポーターDNAフラグメントに密接に結合する場合(左図)、G F P 発現は上昇する(右図)。m C h e r r y 強度は、T A L E - (X X ')₃ プラスミドのトランスフェクション効率を示す。

【図2】5 m C と5 h m C を含むレポーターDNAフラグメントの調製を示す図である。5 m C と5 h m C は、5 m C と5 h m C を含むレポーターDNAフラグメントを生成するためのプライマーに組み込まれた。H P L C クロマトグラフィーは、(a)5 m C および(b)5 h m C の組み込みを示しており、拡大画像から、5 h m C のピークは明確に観察された。(c)は、5 m C と5 h m C を含むレポーターDNAフラグメントのP C R 増幅の概略図である。

【図3】5 m C と5 h m C に対するT A L E R V D の効率および特異性に対する完全な評価を示した図である。(a)5 m C と5 h m C のスクリーニングデータは、ヒートマップを使用してまとめられた。比較のために、従来のCおよびTレポーターDNAフラグメントの結果も示している。レポーターDNAフラグメントのアイデンティティを異なる色で示し、異なるレポーターDNAフラグメントのE G F P 活性を番号付けし、色の明るさは、ベースラインレベルに対して標準化された、レポーターDNAフラグメントに対するT A L E 構造体の誘導倍数を示す。ここで、アミノ酸の1文字の略語が使用されている。

(b)図aの予備スクリーニング結果から、5 m C と5 h m C に対して認識能力を持っているいくつかのR V D を選択して実験し、具体的には、5 m C と5 h m C のレポーターシステムに対してE G F P 活性化倍数が大きいいくつかのR V D を選択して3回重複した確認実験をし、この図は、修飾されたシトシンに対するR V D の選好を示している。塩基選好によってR V D を群分けし、各群において13位の塩基によってグループ化した。データは、平均値±S D、n=3;*P<0.05、**P<0.005である。

【図4】修飾されたシトシンへの420個のT A L E R V D の結合選好を示す図である。このデータはヒートマップ(図3a)に対応している。Y軸はE G F P レポーターの誘導倍数であり、X軸はR V D である。R V D の1番目の残基によってヒストグラムをソートし、2番目の残基のアルファベット順にデータをリストした。

【図5】インビトロ保護アッセイによるT A L E R V D s のDNA認識の定量的測定を示す図である。(a)インビトロ保護アッセイの原理を示す図である。簡単に言えば、T A L E タンパク質(すなわち、図中のT A L E フエクタータンパク質)は特定の配列のDNAフラグメントに結合すると、M s p I 制限エンドヌクレアーゼ部位をブロックし、それによってエンドヌクレアーゼの切断を阻害し、変性P A G E 分析において保護された完全長バンドと切断されたDNAバンドが生成する。DNAの保護効率は、T A L E タンパク質のDNAへの結合効率を反映している。(b)切断されていないまたは保護されたDNAフラグメントを測定することにより標準化された保護効率を取得し、それを異なるT A L E R V D の保護曲線に適合させた。H i l l s l o p e (G r a p h P a d) を用いてその曲線を特異的結合曲線に適合させた。すべての実験を3回繰り返した。(c)(b)から計算された阻害定数であり、同じR V D の最低阻害定数に対する各定数の比率を括弧内に示している。R V D の阻害定数は、異なるR V D を含むT A L E タンパク質がC、5 m C および5 h m C に対する切断保護実験を通して保護効率を得て、G r a p h P a d P r i s m 6 ソフトウェアを使用して保護効率曲線を適合させて阻害定数を計算して得られたものであり、この阻害定数は、C、5 m C および5 h m C に対する異なるR V D の結合効率を示し、阻害定数の値が小さければ小さいほど、R V D の保護効率が強く、対応するDNAフラグメントへの結合が強いことを表す。ここでいう同じR V D の最低阻害定数とは、当該R V D とC、5 m C および5 h m C との結合効率が最も高いグループの阻

10

20

30

40

50

害定数値を指す。

【図6】インビトロ保護アッセイでの、異なるT A L E R V Dのエピジェネティックマーカーとしてのシトシンへの特異的結合を示す図である。(a)精製されたT A L Eタンパク質の代表的なサイズ排除クロマトグラフィーである。(b)S D S - P A G E分析は、精製されたT A L Eタンパク質の分子量が計算された分子量とよく相関することを示している。(c)インビトロ保護アッセイの代表的なゲル画像である。図6から分かるように、M A P K 6 - H DはCを最高の効率で保護することができ、H Aは、修飾されていないCより高い効率で5 m Cと5 h m Cを保護し、F Sは最高の効率で5 h m Cを保護する。

【図7】メチル化依存性遺伝子の発現活性化と遺伝子編集を示す図である。(a)T A L E T E T 1は、T E T 1遺伝子の転写開始部位(T S S)の上流約80 b pでの16 b p D N A配列を標的としている。この領域での三つのC p G(Cは黒色で表示)はすべて、H e L a細胞では高度にメチル化されているが、H E K 2 9 3 T細胞ではメチル化されていない。(b)異なるR V Dを含むT A L E T E T 1でトランスフェクションしたH e L aとH E K 2 9 3 T細胞におけるT E T 1の相対的m R N Aレベルを示す図である。(c)T A L E L R P 2は、L R P 2遺伝子のT S Sの上流100 b pでの16 b p配列を標的としている。これら二つの領域での二つのC p GはH e L a細胞では中程度のレベルのメチル化を含むが、H E K 2 9 3 T細胞ではメチル化されていない。(d)異なるR V Dを含むT A L E L R P 2でトランスフェクションしたH e L aとH E K 2 9 3 T細胞におけるL R P 2の相対的m R N Aレベルを示す図である。(e)T A L E N (T r a n s c r i p t i o n a c t i v a t o r - l i k e e f f e c t o r n u c l e a s e、すなわちF o k Iエンドヌクレアーゼと融合したT A L Eエフェクター)の標的配列の位置を示す図である。メチル化されたC p Gは赤色で表示される。(f)異なるR V Dを使用したT A L E Nの遺伝子編集効率を示す図である。データは、平均値±S D、n=3；*P<0.05、**P<0.005である。

【図8】メチル化依存性遺伝子の発現活性化と遺伝子編集を示す図である。(A)R V D N A、G*、およびY*を有するT A L E T E T 1でトランスフェクションしたH e L aとH E K 2 9 3 T細胞におけるT E T 1の相対的m R N Aレベルを示す図である。(B)R V D N A、G*、およびY*を有するT A L E L R P 2でトランスフェクションしたH e L aとH E K 2 9 3 T細胞におけるL R P 2の相対的m R N Aレベルを示す図である。(C)R V D N A、G*、およびY*を有するT A L E Nのゲノム編集効率を示す図である。データは、平均値±S D、n=3；*P<0.05、**P<0.005である。

【図9】ゲノムDNAにおける単一塩基解像度での5 h m Cの検出を示す図である。(a)新しく同定されたR V Dを用いて塩基レベルの解像度で5 h m Cを検出するワークフローである。簡単に言えば、標的ゲノム領域は、C a s 9を介したDNAに切断されないように、T A L Eによって保護されている。(b)m E S Cゲノム内の単一の5 h m C部位を標的とするT A L E - F S (黒色)とT A L E - H D (灰色)の保護効率を示す図である。(c)T A L E - F Sがm E S C、R A W 2 6 4 . 7、L - M (T K -) およびL 9 2 9細胞のゲノムDNAにおける単一5 h m C部位に対する保護効率を示す図である。当該特定の部位では、すべての細胞株の中で、m E S Cゲノムは最高の5 h m C修飾レベルを含む。

【図10】5 h m Cを含むDNAに対するT A L E - F Sの選択的保護を示す図である。5 m C、5 h m C、および修飾されていないCを含むDNA(M A P K 6遺伝子と同じ配列を持つ)を異なる比率でペアして混合した。5 m C(灰色の円)の割合が増加すると、保護効率は僅かに上昇し、5 h m Cの割合が増加すると(Cおよび5 m Cと混合し、黒丸、黒色三角形で表示)、保護効率は大幅に向上し、これは、5 h m Cに対するR V D F Sの選択的保護を示している。

【図11】6 m AおよびAに対するT A L E - (X X ')₃の結合特定を示す図である。

【図12】6 m AおよびAに対する一部のT A L E - (X X ')₃の結合特性を示す図である。R V Dの2位のアミノ酸によって群分けし、各群のR V Dについて6 m Aレポーターシステムの活性化効率に従って低から高へソートした。縦軸はレポーターシステムE G F

10

20

30

40

50

Pの活性化倍数であり、灰色はAレポーターシステムに対応し、黒色は6mAレポーターシステムに対応し、横軸はRVDであり、反復した後の6mAの平均値が5より大きいデータ群のみが示されている。データは、 $\text{means} \pm \text{s.d.}$ 、 $n=3$ である。

【図13】A、T、CおよびGレポーターシステムに対する異なるRVDの認識効率を示す図である。

【発明を実施するための形態】

【0044】

本発明は、DNAへのTALEタンパク質の結合がDNA塩基修飾により影響されることを示している。本発明は、420のRVDについて研究した結果、5mC、5hmCおよび／または6mAに対して独特の特異性を有するRVDを同定した。5mC、5hmC、及び6mAは、高等真核生物の重要なエピジェネティックマーカーである。メチル化およびヒドロキシメチル化群は、塩基対形成を妨げないが、DNA二本鎖の主溝に存在しているので、TALEタンパク質との相互作用に影響を与える。

【0045】

TALE-DNA複合体の構造は、13位のアミノ酸がセンス鎖のDNA塩基と直接に相互作用する唯一の残基であり、12位の残基の役割が、塩基対認識プロセスで適切なループ構造を安定化にすることであることを示している(35, 36)。13位に小さいアミノ酸(GlyとAla)である場合、あるいはアミノ酸が欠失している場合、5mCに対する親和性が増加できることは本発明により証明された。この観察結果は、N*およびNG(自然にTを認識する)が5mCに結合できるという前の発見と一致している。13位に大きな側鎖が欠失することは、5mCのメチル基を収容するのに十分なスペースを提供できる可能性がある。ただし、この一般的な傾向には例外がある。例えば、5mCに対するHGの親和性が非常に弱く、HDと比べてHGが13位において小さい残基を含み、HDがCの天然の結合体であることは本発明において観察された。興味深いことに、12位のHisがArgに置き換えられた場合(したがってRGになる)、5mCへの強い結合が観察された。実は、RGは5hmCをも認識する。これらの観察結果は、二重残基が修飾に対する認識にはより複雑なモードが存在するかもしれないことを示している。

【0046】

本発明は、TALEに媒介された、いくつかの高度にメチル化されたゲノム領域のメチル化依存性遺伝子活性化とゲノム編集を証明した。重要な対照として、同じ領域にシトシンのメチル化が存在しない場合(異なる細胞内)、遺伝子の活性化はほとんど観察されていない。したがって、本発明によって発見されたRVDは、インビボでの標的遺伝子の修飾状態によって標的遺伝子を操作するという可能性を提供することができる。ゲノム刷り込みと疾患などを含む、多くの重要な生物学的イベントに係るメチル化可変領域(differentially methylated region、DMR)が多く存在することは知られている。したがって、エピジェネティックマーカーを読み取るTALEタンパク質の独特な能力により、将来的にはインビボでのエピゲノム依存性の応用が可能になる。

【0047】

本文で使用される「ポリヌクレオチド」という用語は、線状または環状の立体構造及び一本鎖または二本鎖形態のデオキシリボヌクレオチドまたはリボヌクレオチドポリマーを指す。

【0048】

本発明において、用語の「ポリペプチド」、「ペプチド」及び「タンパク質」は互いに置き換えて使用でき、アミノ酸の重合体を指し、その中で、一つ以上のアミノ酸は天然アミノ酸、または化学的類似体若しくは修飾誘導体であってもよい。

【0049】

本文の「結合」とは、高分子間(例えば、タンパク質と核酸との間)の配列特異的な非共有相互作用を指す。本文で使用される用語の「結合ポリペプチド」は、別の分子と非共有結合することができるポリペプチドまたはタンパク質である。この別の分子は、DNA分子、RNA分子及び／またはタンパク質分子であってもよい。

【0050】

本発明で使用される用語の「T A L E」とは、DNA結合ドメイン（T A L E反復ドメインまたはT A L E繰り返し単位とも呼ばれる）およびその両側のN-末端とC-末端非反復配列を含み、DNA配列を特異的に認識できる転写活性化因子様エフェクター（T r a n s c r i p t i o n A c t i v a t o r - l i k e E f f e c t o r s）を指す。前記DNAドメインは、タンデムの「繰り返し単位」で構成される。各「繰り返し単位」に33～35のアミノ酸が含まれ、そのうち12位と13位の残基は標的認識の肝心な部位であり、反復可変二重残基（r e p e a t - v a r i a b l e d i r e s i d u e s またはR V D）と呼ばれ、各R V Dは一つの塩基しか認識できない。T A L EまたはそのDNAドメインは、R V Dを介してそのR V Dに順に対応するDNA標的配列を認識する。

10

【0051】

天然T A L Eは一般的に1.5～33.5の繰り返し単位を含むが、DNAに対する効果的な認識と結合には一般的には6.5の繰り返し単位が必要であり、より強い活性を示すには10.5以上の繰り返し単位が必要であることは研究によって明らかになった（B o c h, J e n s, a n d U l l a B o n a s. “X a n t h o m o n a s A v r B s 3 f a m i l y - t y p e I I I e f f e c t o r s : d i s c o v e r y a n d f u n c t i o n . ” A n n u a l r e v i e w o f p h y t o p a t h o l o g y 48 (2010) : 419-436. ; B o c h, J e n s, e t a l. “B r e a k i n g t h e c o d e o f D N A b i n d i n g s p e c i f i c i t y o f T A L - t y p e I I I e f f e c t o r s . ” S c i e n c e 326. 5959 (2009) : 1509-1512.) 。

20

【0052】

T A L E繰り返し単位は、切られた繰り返し単位であることができ、半繰り返し単位とも呼ばれ、つまり、完全は繰り返し単位のN-末端の一部であり、切られた繰り返し単位はR V Dを含む。一般的には、天然T A L E反復ドメインのカルボキシ末端の最後の繰り返し単位は、切られた繰り返し単位である。繰り返し単位は一般的には17～20のアミノ酸を含む。

【0053】

本発明では、一部の実施形態では、T A L Eの繰り返し単位は、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、または35であってもよい。T A L Eの繰り返し単位は、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、または34個の完全な繰り返し単位と一つの半繰り返し単位を含んでもよい。

30

【0054】

好ましい実施形態では、T A L Eは、14個の完全な繰り返し単位と一つの半繰り返し単位とを含み、半繰り返し単位は、T A L E繰り返し単位全体のカルボキシ末端に位置する。

【0055】

好ましい実施形態では、T A L E中の単一の「繰り返し単位」は、L T P E Q V V A I A S X X ' G G K Q A L E T V Q R L L P V L C Q A H Gであってもよい。一部の実施形態では、T A L E中の半繰り返し単位配列は、L T P E Q V V A I A S X X ' G G K Qである。その中で、X X ' はR V Dである。

40

【0056】

本発明の実施例で使用されるT A L E繰り返し単位配列は、X a n t h o m o n a s (キサントモナス属) のA v r B s 3タンパク質アミノ酸配列である。この配列の他に、本発明のR V Dはその他の繰り返し単位配列を含むT A L Eにも適用可能である。A v r B s 3は、異なるキサントモナス属の亜種において異なるホモログを持っており、その具体的配列について、「B o c h, J e n s, a n d U l l a B o n a s. “X a n t h o

50

monas AvrBs3 family-type III effectors: discovery and function. "Annual review of phytopathology 48 (2010): 419-436."を参照できる。

【0057】

本発明において、ポリペプチド配列中のアミノ酸は一文字略語で示され、本発明に係るアミノ酸およびその一文字略語は下記通りである。

【0058】

【表1】

グリシン	Glycine	Gly	G
アラニン	Alanine	Ala	A
バリン	Valine	Val	V
ロイシン	Leucine	Leu	L
イソロイシン	Isoleucine	Ile	I
プロリン	Proline	Pro	P
フェニルアラニン	Phenylalanine	Phe	F
チロシン	Tyrosine	Tyr	Y
トリプトファン	Tryptophan	Trp	W
セリン	Serine	Ser	S
スレオニン	Threonine	Thr	T
システイン	Cystine	Cys	C
メチオニン	Methionine	Met	M
アスパラギン	Asparagine	Asn	N
グルタミン	Glutamine	Gln	Q
アスパラギン酸	Aspartic acid	Asp	D
グルタミン酸	Glutamic acid	Glu	E
リジン	Lysine	Lys	K
アルギニン	Arginine	Arg	R
ヒスチジン	Histidine	His	H

【0059】

本発明において、RVDを説明する時、*はその位置でのアミノ酸が欠失していることを表す。

【0060】

本発明において、「塩基」と「ヌクレオチド」は互いに置き換えて使用でき、プリン塩基またはピリミジン塩基、リボースまたはデオキシリボース、及びリン酸の三つの物質からなる化合物を指し、DNA配列とRNA配列の主な組成成分である。一般的なデオキシヌクレオチドには、シトシン（C）、チミン（T）、アデニン（A）、及びグアニン（G）が含まれる。

【0061】

上記四つの従来のデオキシリボヌクレオチドの他に、ほ乳類のゲノムには、修飾されたDNA塩基も含まれている。例えば、5番目のDNA塩基と呼ばれる5-メチルシトシン（5mC）は、遺伝子発現を調節する重要なエピジェネティックマーカーである。5mC

は、10～11転座酵素（TET）ファミリーのタンパク質によって相次いで酸化されて5-ヒドロキシメチルシトシン（5hmC）、5-ホルミルシトシン（5fC）および5-カルボキシシトシン（5caC）を生成することができる。シトシンのメチル化修飾の他に、もう一つのよく見られるDNAメチル化修飾N6-メチルアデニン（N⁶-methyladenine, 6mA）は、DNAのアデニン上の共有結合修飾として原核細胞において重要な役割を果たす。

【0062】

本発明でいう「メチル化修飾塩基」とは、メチル化修飾を有する塩基をいい、5-メチル化シトシン（5mC）、5-ヒドロキシメチル化シトシン（5hmC）、及び6-メチル化アデニン（6mA）を含む。

【0063】

本発明は、5mC、5hmC、6mAを特異的に認識するRVD、及びこれらのメチル化修飾塩基および対応する非修飾塩基を認識できる縮退RVDを発見した。詳細は下記表を参照できる。

【0064】

【表2】

認識塩基	RVD
5mC	HA、NA
5hmC	FS
C、5mC	N*、NG、KP
C、5hmC	HV、KV
5mC、5hmC	K*、RG
C、5mC、5hmC	G*、H*、R*、Y*
6mA	NP、FT、CV、CP
6mA、A	RI、NI、KI、HI

【0065】

上記表によれば、RVD HAまたはNAは5mCを特異的に認識でき、すなわち、5mCを5hmCとCとから区別付けることができる；RVD FSは5hmCを特異的に認識でき、すなわち、5hmCを5mCとCとから区別付けることができる；RVD NP、FT、CVまたはCPは6mAを特異的に認識でき、すなわち、6mAとAを区別付けることができる；縮退RVD N*、NGまたはKPはCと5mCを認識できる。

【0066】

本発明において、特に断らない限り、RVDにより認識される塩基を説明する時、「C」を用いる場合、メチル化修飾のないシトシンを意味し、「A」を用いる場合、メチル化修飾のないアデニンを意味し、「5mC」を用いる場合、5-メチルシトシンを指し、「5hmC」を用いる場合、5-ヒドロキシメチル化シトシンを指し、「6mA」を用いる場合、6-メチル化アデニンを指す。

【0067】

本発明によれば、ある特定のメチル化修飾塩基に対する「特異的認識」とは、この特定のメチル化修飾塩基に対するRVDの結合親和性が、その他の形態の修飾を有する同じ塩基よりも有意に大きい、または修飾のない同じ塩基より大きい、またはその他の異なる塩基より大きいことをいう。

【0068】

結合親和性は、当業者によく知られている様々な方法で測定することができる。例えば、下記の参考文献に記載されているようにTALE-VP64-mCherry構造体を

構築し、異なる修飾塩基と蛍光タンパク質遺伝子を含むレポーターDNAフラグメントを構築し、T A L E - V P 6 4 - m C h e r r y 構造体が細胞内で発現したT A L E - V P 6 4 タンパク質がレポーターDNAフラグメントを結合かつ活性化して引き起こした蛍光タンパク質シグナルの上昇倍数を利用して、T A L E に含まれるR V D とレポーターDNAフラグメントに含まれる異なる修飾塩基との結合親和性を確定する。ある特定の修飾塩基に対するR V D のE G F P 活性化倍数は、その他の形態の塩基に対するE G F P 活性化倍数より有意に高い場合、R V D が当該特定の修飾塩基を特異的に認識できると考えられる。結合親和性は、本発明の実施例4に記載されているインビトロ保護アッセイによっても確定できる。

【0069】

上記表によれば、R V D H A またはN A が5 m C を特異的に認識できることを本発明は発見した。R V D H A またはN A の5 m C への結合親和性は、5 h m C とC への結合親和性より有意に高く、このR V D を利用すれば、5 m C を5 h m C とC とを区別付けることができ、T A L E と5 m C との特異的な結合を実現することができ、5 m C に依存する様々な具体的な使用を実現することができる。

【0070】

5 m C に依存する様々な具体的な使用は、遺伝子中の5 m C の検出、5 m C に依存する、遺伝子発現調節、遺伝子編集、後成的修飾など（すなわち、標的配列に5 m C が存在する場合にのみ遺伝子発現調節、遺伝子編集または後成的修飾をせずに、対応する位置にC または5 h m C である場合には遺伝子発現調節、遺伝子編集または後成的修飾をする）、5 m C に依存する生細胞染色体標識（すなわち、染色体中の対応する位置に5 m C を有する遺伝子のみを標識し、対応する位置にはC または5 h m C である場合は標識せずに、これによって生細胞における遺伝子のシトシンのメチル化状況を観察することができる）、5 m C を含む配列に特異的に結合できるタンパク質の調製を含むが、これらに限られていない。

【0071】

R V D F S が5 h m C を特異的に認識できることを本発明はさらに発見した。R V D F S の5 h m C への結合親和性は、5 m C とC への結合親和性より有意に高く、このR V D を利用すれば、5 h m C を5 m C とC とを区別付けることができ、T A L E と5 h m C との特異的な結合を実現することができ、5 h m C に依存する様々な具体的な使用を実現することができる。

【0072】

5 h m C に依存する様々な具体的な使用は、遺伝子中の5 h m C の検出、5 h m C に依存する、遺伝子発現調節、遺伝子編集、後成的修飾など（すなわち、標的配列に5 h m C が存在する場合にのみ遺伝子発現調節、遺伝子編集または後成的修飾をせずに、対応する位置にC または5 m C である場合には遺伝子発現調節、遺伝子編集または後成的修飾をする）、5 h m C に依存する生細胞染色体標識（すなわち、染色体中の対応する位置に5 h m C を有する遺伝子のみを標識し、対応する位置にはC または5 m C である場合は標識せずに、これによって生細胞における遺伝子のシトシンのメチル化状況を観察することができる）、5 h m C を含む配列に特異的に結合できるタンパク質の調製を含むが、これらに限られていない。

【0073】

また、R V D N P 、F T 、C V またはC P が6 m A を特異的に認識できることを本発明は発見した。これらのR V D の6 m A への結合親和性はA への結合親和性より有意に高く、これらのR V D を利用すれば、6 m A をA とを区別付けることができ、T A L E と6 m A との特異的な結合を実現することができ、6 m A に依存する様々な具体的な使用を実現することができる。

【0074】

6 m A に依存する様々な具体的な使用は、遺伝子中の6 m A の検出、6 m A に依存する、遺伝子発現調節、遺伝子編集、後成的修飾など（すなわち、標的配列に6 m A が存在す

10

20

30

40

50

る場合にのみ遺伝子発現調節、遺伝子編集または後成的修飾をせずに、対応する位置にAである場合には遺伝子発現調節、遺伝子編集または後成的修飾をする）、6mAに依存する生細胞染色体標識（すなわち、染色体中の対応する位置に6mAを有する遺伝子のみを標識し、対応する位置にはAである場合は標識せずに、これによって生細胞における遺伝子のシトシンのメチル化状況を観察することができる）、6mAを含む配列に特異的に結合できるタンパク質の調製を含むが、これらに限られていない。

【0075】

また、縮退RVD N*、NGまたはKPがCと5mCを認識できることを本発明は発見した。これらの縮退RVDは類似した結合親和性でCと5mCに結合し、しかもこれらの縮退RVDのCと5mCへの結合親和性は5hmCへの結合親和性より有意に高い。

【0076】

また、縮退RVD HVまたはKVがCと5hmCを認識できることを本発明は発見した。これらの縮退RVDは類似した結合親和性でCと5hmCに結合し、しかもこれらの縮退RVDのCと5hmCへの結合親和性は5mCへの結合親和性より有意に高い。

【0077】

また、縮退RVD K*またはRGが5mCと5hmCを認識できることを本発明は発見した。これらの縮退RVDは類似した結合親和性で5mCと5hmCに結合し、しかもこれらの縮退RVDの5mCと5hmCへの結合親和性は5mCへの結合親和性より有意に高い。

【0078】

また、縮退RVD G*、H*、R*またはY*がC、5mC、及び5hmCを認識できることを本発明は発見した。これらの縮退RVDは類似した結合親和性でC、5mC、及び5hmCに結合する。

【0079】

また、縮退RVD RI、NI、KIまたはHIが6mAとAを認識できることを本発明は発見した。これらの縮退RVDは類似した結合親和性でAと6mAに結合する。

【0080】

これらの縮退RVDは、二つまたは三つの異なるメチル化修飾されたまたはメチル化修飾されていない塩基を同時に認識することができ、塩基のメチル化修飾状況を知らなくても使用することができ、TALEの標的結合効率を向上させ、TALEと標的配列との結合に対するメチル化修飾の影響を低減することができる。例えば、細胞ゲノム中の5mCは、TETファミリータンパク質の触媒作用下で5hmCに酸化されることができ、5mCと5hmCを同時に認識できる縮退RVDを使用すると、シトシンメチル化の違いによる結合効率の低下などの問題を避けることができる。そのため、具体的な実験では、実験の目的によって、具体的な実験のニーズを満たすように、ある一つのメチル化修飾塩基を特異的に認識するRVD、二つのメチル化形態の塩基を認識する縮退RVD、及び三つのメチル化形態の塩基を認識する縮退RVDを組み合わせて使用することができる。

【0081】

本発明のRVDは、特定のメチル化形態の塩基に結合する必要がある任意の応用に使用することができ、これらの応用は、インビトロまたはインビボであってもよく、これらの応用は非治療的であってもよい。

【0082】

本発明のRVDを含むTALEは、特定のメチル化形態を有する塩基に結合するためのDNA結合ポリペプチドとして発現されることができる。場合によって、このDNA結合ポリペプチドは「抗体」として機能して、その「抗原」（すなわち、特定のメチル化形態を有する塩基の標的配列）に結合することができる。場合によって、このDNA結合ポリペプチドは、特定のメチル化形態を有する塩基の標的配列に結合して、ヌクレアーゼによって切断されないように、またはその他のDNA結合ポリペプチド（例えば、転写調節因子など）と相互作用をしないようにそれを保護することができる。

【0083】

10

20

30

40

50

本発明の R V D を含む T A L E はまた、蛍光タンパク質とカップリングして融合タンパク質を形成することができる。この融合タンパク質を用いて、生細胞の染色体上の、特定のメチル化形態を有する塩基の標的配列に結合させ、これによって生細胞中の染色体の動的变化を観察することができる。

【0084】

蛍光タンパク質は当業者に周知であり、緑色蛍光タンパク質 (G F P)、強化緑色蛍光タンパク質 (E G F P)、赤色蛍光タンパク質 (R F P) または青色蛍光タンパク質 (B F P) などを含むが、これらに限られていない。

【0085】

本発明の R V D を含む T A L E はまた、機能ドメインにカップリングして融合タンパク質を形成することができ、この融合タンパク質を使用して、特定のメチル化形態の塩基を有する標的配列を操作することができる。この操作は、遺伝子編集、遺伝子発現の調節、または後成的修飾などであってもよく、前記機能ドメインは、遺伝子編集機能ドメイン、遺伝子発現を調節するドメイン、または後成的修飾ドメインであってもよい。

【0086】

用語の「遺伝子編集」とは、遺伝子の挿入、削除または置換を含む、標的部位での遺伝子配列を変更することを指す。例えば、前記遺伝子編集は、ヌクレアーゼを使用して標的部位に DNA 二本鎖切断して DNA 一本鎖ギャップなどを形成し、続いて DNA 配列の非同相末端結合 (N H E J) 修復中に DNA の挿入または削除 (i n s e r t i o n a n d d e l e t i o n , i n d e l) が生じてフレームシフト突然変異を引き起こし、これによって遺伝子ノックアウトの目的を達成する。遺伝子編集機能ドメインとは、遺伝子編集機能を実現可能なアミノ酸配列を指す。

【0087】

本発明の R V D を含む T A L E と遺伝子編集機能ドメインとを含む融合タンパク質を使用して遺伝子編集をする時、遺伝子編集機能ドメインはヌクレアーゼであってもよい。ヌクレアーゼは、エンドヌクレアーゼ、ジンクフィンガーヌクレアーゼ (Z F N)、C a s 9 ヌクレアーゼを含むが、これらに限られていない。C a s 9 ヌクレアーゼの使用は本分野で周知であり、その使用は通常、標的配列の切断を実現するように C a s 9 ヌクレアーゼと s g R N A を共に細胞に導入することである。

【0088】

本発明において、遺伝子編集を行う場合、前記融合タンパク質を T A L E N の形で提供可能であることが好ましく、この場合、遺伝子編集機能ドメインは F o k I エンドヌクレアーゼの DNA 切断ドメインである。

【0089】

用語の「遺伝子発現の調節」とは、非コード RNA、及び一つ以上のタンパク質またはタンパク質サブユニットをコードする RNA を含む、遺伝子の発現または RNA 分子のレベルを変更することを指す。「遺伝子発現の調節」はさらに、一つ以上の遺伝子産物、タンパク質またはタンパク質サブユニットの活性を変更することを含む。遺伝子発現を調節する機能ドメインとは、標的遺伝子の発現を調節できるアミノ酸配列をいう。

【0090】

前記遺伝子発現を調節する機能ドメインは、転写活性化因子またはその機能的フラグメント、若しくは転写抑制因子またはその機能的フラグメントであってもよい。

【0091】

用語の「後成的修飾」とは、標的遺伝子の DNA 配列を変更せずに、DNA メチル化修飾、DNA 脱メチル化などを含む DNA の修飾を指す。後成的修飾機能ドメインとは、標的遺伝子に対して後成的修飾が可能なアミノ酸配列を指す。

【0092】

後成的修飾機能ドメインは、メチルトランスフェラーゼ、デメチラーゼであってもよい。

【0093】

用語の「機能的フラグメント」とは、その配列が完全長のタンパク質またはポリペプチ

10

20

30

40

50

ドの一部であるが、完全長のタンパク質またはポリペプチドと同じ機能を有するタンパク質またはポリペプチド、例えば、ヌクレアーゼの切断機能ドメインのような、特定の実験条件下で対応する機能を発揮可能なタンパク質ドメインである。

【0094】

本文に記載の細胞は、任意の細胞または細胞株であってもよく、植物、動物（例えば、マウス、ラット、霊長類、家畜、ウサギなどの哺乳動物）、魚などの細胞であってもよく、真核生物（例えば、酵母、植物、真菌、魚、及びネコ、イヌ、ねずみ、ウシ、ヒツジ、ブタなどのような哺乳動物細胞）細胞であってもよい。

【0095】

本文に記載の細胞は、卵母細胞、K562細胞、CHO（チャイニーズハムスター卵巣）細胞、HEP-G2細胞、BaF-3細胞、シュナイダー（Schneider）細胞、COS細胞（SV40T-抗原を発現するサル腎細胞）、CV-1細胞、HuTu80細胞、NTERA2細胞、NB4細胞、HL-60細胞、及びHeLa細胞、HEK293T細胞などであってもよい。

【0096】

本発明のいずれか一項の技術方案の方法は、インビトロまたはインビボで実施することができる。

【0097】

本発明のいずれか一項の技術方案の方法は、非治療的であってもよい。

【実施例】

【0098】

実施例1 材料と方法

【0099】

1、DNAの合成と精製

オリゴ（Oligo）DNAプライマーを、5mCと5hmCホスホルアミダイトを含む標準試薬（Glen Research）を使用してExpedite 8909 DNA/RNAシンセサイザーで合成した。Glen Research Corp. が推奨した標準法でOligo DNAを脱保護し、Glen-Pak DNA精製カートリッジ（purification cartridge）を使用してOligo DNAを精製した。

【0100】

合成されたDNAを高速液体クロマトグラフィー（HPLC）で検証した。要するに、DNAをリボザイムP1（Sigma、N8630）およびアルカリホスファターゼ（Sigma、P4252）でヌクレオシドに消化させた。SB-Aq C18カラム（Agilent）で5%～50%アセトニトリルを使用して30分間内にヌクレオシドを分離した。

【0101】

2、細胞培養、トランスフェクション、フローサイトメトリー

HEK293T細胞（スタンフォード大学のStanley Cohen Labから）、及びHeLa細胞（本Labで保存）をDMEMで培養し、10%FBSおよび1%ペイシリン-ストレプトマイシンを添加し、37℃、5%CO₂の条件下で培養した。トランスフェクションの24時間前に、1ウェル当たり 7×10^4 細胞の播種密度で細胞を24ウェルプレートに播種した。0.15 µgのTALE-（XX'）₃プラスミドと0.15 µgのレポーターDNAを用いてポリエチレンイミン（PEI）で各ウェルの細胞を同時トランスフェクションした。トランスフェクションの48時間後、細胞を収集し、BD LSR Fortessaフローサイトメーター（BD Biosciences）で分析した。488nmと561nmの波長のレーザーを使用してそれぞれEGFPとmCherryタンパク質発現を定量化した。各サンプルから少なくとも10000イベントを収集し、分析に十分なデータを取得した。mCherry蛍光密度が $5 \times 10^3 \sim 5 \times 10^4$ である細胞は、分析に用いられた。

10

20

30

40

50

【0102】

3、TALENの構築

TALENプラスミドの骨格には、CMVプロモーター、核局在化シグナル、TALEアミノ末端とカルボキシル末端の非反復配列、及びエンドヌクレアーゼのFokIモノマーが含まれており、その具体的な配列について、参考文献37を参照できる。

使用中、異なるRVDを含むTALE繰り返し単位がTALEN骨格ベクターに挿入され、異なるRVDの効果が検証された。その構築方法について、Yang, Junjiao, et al. "Assembly of Customized TALE Effectors Through Advanced ULTIMATE System." TALENs: Methods and Protocols (2016): 49-60を参照できる。

10

【0103】

4、TALEタンパク質の発現と精製

発現及び精製されたTALEタンパク質は、インビトロ保護アッセイに使用された。

前述のように、ULTIMATEシステムを使用して標準RVD（すなわち、NI、NG、HDおよびNN）を有するTALE繰り返し単位を構築した（37）。新しいRVDを使用したTALE繰り返し単位の場合、新しいRVDを含むTALE繰り返し単位モノマーは個別に合成された。前述のように、同じULTIMATEプロトコルを使用してこれらのTALE構造体の最終組み立てを実行した（37）。

【0104】

TALE発現プラスミドの構築は、TALE繰り返し単位をTALEN骨格に組み換えることにより実現された。TALENのN-およびC-末端配列を含み、しかも中間繰り返し単位を持つフラグメントを、対応するTALENプラスミドから増幅し、pET-28a(+)のNheIおよびHindIII部位にクローニングした。

異なるRVDを有するTALEの配列（精製用Hisタグ、TALEのN-およびC-末端配列、及び、DNAを特異的に認識できるTALE繰り返し単位を含む）は、pET28ベクター（Novagen）にクローニングされた。細胞密度がOD600 8.0に達すると、1.0mMイソプロピルβ-D-チオガラクトシド（IPTG）を用いて大腸菌BL21(DE3)においてTALEタンパク質の過剰発現を誘導した。20℃で16時間増殖させた後、細胞を回収し、25mM Tris-HCl pH8.0および150mM NaClを含むバッファーに再懸濁し、超音波処理により細胞を破壊した。Ni²⁺-ニトリロトリアセテート親和性樹脂（Ni-NTA、GE healthcare）（Buffer A: 10mM Tris-HCl pH8.0、300mM NaCl、Buffer B: 10mM Tris-HCl pH8.0、300mM NaClおよび500mMイミダゾール）、およびHiLoad superdax PG200（GE Healthcare）（Buffer GF: 10mM Tris-HCl、pH8.0、100mM NaCl）を順次に用いて組み換えタンパク質を精製した。

20

30

【0105】

5、TALEN繰り返し単位

以下の実施例で使用されたTALE繰り返し単位は、連続した14の繰り返し単位と一つの半繰り返し単位とを含み、各単一の繰り返し単位は34のアミノ酸残基を含み、その配列は、LTPEQVVVAIASXX'GGKQALETVQRLLPVLCQAHG(SEQ ID NO. 1)であり、半繰り返し単位は単一の繰り返し単位の最初の17のアミノ酸残基を含み、その配列は、LTPEQVVVAIASXX'GGKQ(SEQ ID NO. 2)である。ここで、XX'はRVDである。

40

本実施例で説明された材料と方法は、下記の実施例2～7に用いられている。

【0106】

実施例2 人工スクリーニングシステムの構築

【0107】

人工スクリーニングシステムは、レポーターDNAエレメントとTALE-VP64発

50

現ライブラリーで構成されている。

T A L E - V P 6 4 発現ライブラリーは、400のT A L E - V P 6 4 - m C h e r r y 構造体を含み、各T A L E - V P 6 4 - m C h e r r y 構築体は、細胞にトランスフェクションされるとT A L E - V P 6 4 融合タンパク質を発現する環状プラスミドである（具体的な構築方法について、後文の参考文献37を参照できる）。図1Bに示すように、各構築体は人工T A L E アレイを含み、この人工T A L E アレイは、V P 6 4 に融合された14.5の反復を含み、1～6、10～14番目の反復、および最後の半反復（図1Bで14.5番目の反復として示されている）にあって、異なる構築体の間では同じであるが、7～9番目の反復は、異なる構築体の間で異なる。各構築体について、人工T A L E アレイの7～9番目の反復の三つの連続したモノマーのR V Dは、同じ6つのランダムに合成されたヌクレオチドによりコードされ、すなわち、7～9番目の繰り返し単位は、三つの同一のR V Dをタンデムで発現し、これらの人工T A L E アレイはT A L E - (X X')₃と呼ばれ、これによって、異なるR V Dの5mCおよび5hmCに対する認識を検出するために、異なるテストR V D X X'を有する400のT A L Eを形成した。ここで、XおよびX'はそれぞれ、反復の12と13番目の残基（すなわちR V D）を表す。また、N*が5mCを認識することができることは発見されていたため、20のT A L E - (X*)₃が別途に組み立てられており、13番目の残基は欠失している。後文で、上記のT A L E - (X X')₃とT A L E - (X*)₃をまとめてT A L E - (X X')₃と呼ぶ。したがって、使用されるT A L E - V P 6 4 発現ライブラリーは合計、420種類の異なるT A L E - (X X')₃を含む420種類のT A L E - V P 6 4 - m C h e r r y 構築体を含み、後文でまとめてT A L E 構築体と呼ばれ、T A L E - (X X')₃プラスミドまたはT A L E - (X X')₃発現プラスミドとも呼ばれる。

【0108】

T A L E - V P 6 4 発現ライブラリーを生成した。具体的には、420種類のT A L E - (X X')₃を2種類に分けて、そのうち400種類のT A L E - (X X')₃プラスミドの7～9番目の繰り返し単位のR V Dの12と13番目のアミノ酸残基は、20種類の天然アミノ酸残基の組み合わせであり、このようなT A L E - (X X')₃プラスミドの構築方法は、後文の参考文献13に記載されている。

【0109】

他の20種類のT A L E - (X X')₃の7～9番目の繰り返し単位に発現されたR V Dは、13位のアミノ酸残基が欠失しているR V Dであり、すなわち、A*、C*、D*、E*、F*、G*、H*、I*、K*、L*、M*、N*、P*、Q*、R*、S*、T*、V*、W*、Y*である。これらの20種類のT A L E - (X X')₃発現プラスミドの構築について、後文の参考文献13に報告されているT A L E - (X X')₃の分別構築法が使用される。すなわち、特定のR V Dをコードするフォワードプライマー5' - t C G T C T C a G A A C A G G T T G T A G C C A T A G C T T C T N N N N N N G G A G G T A A G C A G G C A C T G G A A - 3' (SEQ ID NO. 3) (N N N N N N Nは、特定のR V Dをコードする塩基配列を表す)と、同じのリバースプライマー5' - a a C G T C T C a G T T C G G G G G T C A A C C C A T G A G C C T G A C A C A G T A C T G G G A G C A G G C G C T G C A C G G T T T C C A G T G C C T G C T T - 3' (SEQ ID NO. 4)とを用いて、B s m B I 制限性エンドヌクレアーゼ部位を両端に有する、102bp長のT A L E モノマーフラグメントをアニーリングとPCR伸長により生成した。その後、6つのG o l d e n - G a t e 消化ライゲーションサイクルを通してT A L E モノマーフラグメントを連結し、プライマーG - l i b - F とG - l i b - R を使用してT A L E マルチマーを増幅した。最後に、ゲル回収法を使用して、3つのT A L E モノマーしか含有しないフラグメントを選択的に回収し、予め構築したライブラリー発現ベクターに連結し、T r a n s 1 - T 1 コンピテント細胞に形質転換した。対応するR V Dを正確に発現するT A L E - (X X')₃プラスミドをS a n g e r シーケンスにより得た。その中で、

G - l i b - F : 5' - T A G C T A T A C G T C T C A T T G A C C C C C G A A

CAGGTTGTAGCC-3' (SEQ ID NO. 5)

G-lib-R: 5'-TAGCTATACGTCTCACCCATGAGCCTG
ACACAGTACTGGGAGCA-3' (SEQ ID NO. 6) である。

【0110】

レポーターDNAエレメントは、T A L E - (X X ')₃ 認識配列、miniCMVプロモーター、EGFPタンパク質コーディング配列、およびpoly Aシグナルを含む直線状DNAフラグメントである(図1b)。レポーターDNAエレメントのT A L E - (X X ')₃ 認識配列は15塩基長であり、そのうち1~6、10~15番目の塩基は、それぞれライブラリーT A L E 構築体の1~6、10~14.5番目の反復に含まれるRVDにより認識されることができる。レポーターDNAエレメントのT A L E - (X X ')₃ 認識配列の7~9番目の塩基は、対応するメチル化修飾塩基に対する異なるRVDの結合能力をテストするための、連続した3つの5mC、5hmCまたは6mAであってもよく、それぞれ5mCレポーターDNAエレメント、5hmCレポーターDNAエレメント、または5mAレポーターDNAエレメントと呼ばれる。スクリーニングされるメチル化修飾塩基に基づいて、一つまたはそれ以上のレポーターDNAエレメントの使用を決定する。レポーターDNAエレメントは、化学合成された特定のメチル化修飾塩基を含むフォワードプライマーReport-Fと、同じのリバースプライマーReport-Rとを使用してPCR増幅によって得られたものであり、サイズは約1450bpである。

【0111】

プライマーの配列の下記通りである。

Report-F:

5'-G*C*C*AGATATACGCGTTACTGGAGCCATCTGGCCN
NNTACGTAGGCGTGTAC-3' (SEQ ID NO. 7)、ここで、Nは5
mC、5hmC、または6mAを表す；

Report-R: 5'-A*G*C*GTCTCCCGTAAAGCACTAAATC
GGAACCCCTAAAGGGAGC-3' (SEQ ID NO. 8)

(*は、主にレポーターDNAエレメントを細胞内でヌクレアーゼによる分解から保護する役割を果たすチオ修飾塩基を表し、下線は、T A L E - (X X ')₃ 認識配列、すなわち、T A L E 結合配列を表す。)

【0112】

レポーターDNAエレメントの構築過程は次の通りである。まず、大腸菌からレポータープラスミドpcDNA6_3Aを増幅し(図2c)、このレポータープラスミドは、T A L E - (X X ')₃ の結合部位のテンプレート配列CTGGCCAAATACGTA (SEQ ID NO. 9)を含む。次に、5mC、5hmCまたは6mAを含む上記のプライマーを化学合成し、T A L E 結合配列に5mC、5hmCまたは6mAを含む直線状レポーターDNAエレメントをPCRによって生成し(図2c)、フォワードプライマーはT A L E - (X X ')₃ 結合配列CTGGCCNNNTACGTA (SEQ ID NO. 10)を含み、その中で、Nは5mC、5hmCまたは6mAを表し、最小CMVプロモーター(pminiCMV)とその下流のEGFP遺伝子のすぐ上流に位置する。その中で、T A L E 結合配列の7~9番目の反復塩基に対応するのは、三つの連続した5mC、5hmC、または6mAである。

また、人工スクリーニングシステムは、下記の参考文献13に記載の方法に従って構築された環状DNAである、CおよびTのレポーターDNAエレメントを含んでもよい。その中に含まれるT A L E - (X X ')₃ 認識配列は、NNNがCCCまたはTTTであることを除き、その他は上記と同じである。

上記の人工スクリーニングシステムを使用して、レポーターDNAエレメントのT A L E 結合配列へのT A L E - (X X ')₃ の結合の特異性を、EGFPの蛍光レベルを測定することにより検出した。これによって、T A L E RVD認識スクリーニングを体系的に評価するためのプラットフォームが構築されて取得された。

【0113】

10

20

30

40

50

実施例3 修飾されたシトシンを認識するT A L E R V Dのスクリーニング**【0114】**

420個のRVDの5mCおよび5hmCへの結合親和性を測定するために、420個のT A L E構築体のそれぞれを、3種類のEGFPレポーターDNAエレメント（それぞれ、三つの連続したC、5mCまたは5hmCを含む）のうちの一つと一緒にHEK293T細胞に導入した。EGFPおよびmCherry蛍光レベルを、FACS分析を使用して測定した（図1c）。対応する塩基のベースラインレベルに対する各RVDのEGFP発現の倍数変化を比較することにより、T A L E構築体の420個のRVDのそれぞれがC、5mCおよび5hmCへの結合特異性を確定した。C、5mCおよび5hmCの1260のデータポイント、および前の研究（以下の参考文献13を参照）の、Tに対する420のデータポイントをヒートマップ（heat map）にまとめた（図3aおよび図4）。

10

図3aの予備スクリーニング結果から、5mCと5hmCへの結合力のより高いRVDをいくつか選択して3回の反復の確認実験をした。そのうち、5mC、5hmCレポーターDNAフラグメントについてEGFP活性化倍数が4以上であるRVDは、これらの2つのヌクレオチドに対してそれぞれより高い結合力を有する者であると認定され、その結果は図3bに示される。

【0115】

その結果から分かるように、スクリーニングにより、5mCおよび5hmCを効果的に認識する特異性の縮退RVDが得られた。5mCへの高い結合力を持つ複数の結合剤が認識され、それらは、13位のアミノ酸残基によって、Glyを含むRVDs（NG、KGおよびRG）、Alaを含むRVDs（HAおよびNA）、および欠失を含むRVDs（N*、K*、H*、R*、Y*、およびG*）の3種類に分けられた。Glyを含むRVDsと欠失を含むRVDsの中には、一般のRVDs（5mC、5hmC、および通常のCを認識する）と縮退RVDs（5mCおよび5hmCを認識する）の両方がある。興味深いことに、Alaを含む2つのRVDs（HAおよびNA）は、5mCに対して選択性を有する。これまでの研究では、NG（Tの天然結合剤）とN*が使用されて5mCが認識された。スクリーニングの過程において、これら2つのRVDsをも認識したが、本研究中で報告された多くの新しいRVDsの5mCへの結合親和性は、前記2つのRVDsを上回った。例えば、HA、NA、およびX*（Xは、K、H、Y、およびGを指す）は、5mCに対してより高い結合親和性を示している。これら3種類のRVDsが通常のTにも結合することは未だに発見していないが、これは驚くことではない。それらは、小さい側鎖が付いているアミノ酸残基を有するものであるか、13位に当該残基が欠失しているものである。

20

30

【0116】

5hmCに選択的に結合するRVDは今まで報告されていない。上述のように、5hmCによく結合する縮退RVDsと一般のRVDsを特定した。それらの中で、これらの5hmC結合剤について～15倍の誘導が観察され、これは、5hmCに強い親和性を有することは証明された。また、13位の残基にセリン（FS、YSおよびWS）を有する一組の新しい5hmC-結合RVDsも観察された。それらの5hmCへの親和性は、一般の結合剤より低いが、5mCよりも5hmCに優先的に結合するため、フォワードおよび選択的な5hmC認識の可能性を提供する。総じていえば、5mCおよび5hmCの一般の結合剤と縮退結合剤は13位にグリシンまたは欠失を含む傾向が有り、5mCおよび5hmCの特異性結合剤はそれぞれ13位にアラニンまたはセリンを有することを発見した。

40

【0117】

実施例4 5mC、5hmC、および通常のCに対するRVDの結合親和性と特異性の定量測定

【0118】

実施例3で得られた新しいRVDによるDNAの認識は、インビトロ保護アッセイによって検証された（反応原理は図5aに示されている）。この実験では、MAPK6遺伝子

50

における遺伝子配列、5' - TTCAGCTGGAT [CCGGAGGA] GCGGATATAACCAAGG - 3' (SEQ ID NO. 11) が化学合成によって合成された。この配列に基づいて設計されたTAL E認識配列は、角括弧で示される配列であり、この配列には、一つのMsp I制限性エンドヌクレアーゼ認識部位（下線に示される）が含まれている。化学合成時に、所定の位置（Msp I認識部位の2番目のC）には、C、5mCまたは5hmCのDNAオリゴヌクレオチドを含有させる。様々な濃度のTAL Eタンパク質の存在下で、エンドヌクレアーゼMsp IをDNAプローブに加えた。TAL Eタンパク質の、それによって認識されたシトシン塩基への結合は、エンドヌクレアーゼによるDNAの切断を阻害し、その結果、保護された完全長バンドと切断されたDNAバンドが変性PAGE分析で生じた。次に、各RVDについて保護効率を計算して阻害定数（ K_i 、結合親和性の逆数で測定した）の形で表示した。RVDの阻害定数は、異なるRVDを含むTAL Eタンパク質がC、5mCおよび5hmCに対する切断保護実験を通して保護効率を得て、GraphPad Prism 6ソフトウェアを使用して保護効率曲線を適合させて阻害定数を計算して得られたものであり、この阻害定数は、C、5mCおよび5hmCに対する異なるRVDの結合効率を示し、阻害定数の値が小さければ小さいほど、RVDの保護効率が強く、対応するDNAフラグメントへの結合が強いことを表す。

【0119】

エンドヌクレアーゼMsp Iを使用してインビトロ保護アッセイを行った（その原理は図5aに示される）。各10 μ lの反応系には、1 nMの標記されたDNA、1 μ lの10X CutSmart Buffer (NEB)、および100 nMのNaClが含まれている。TAL Eタンパク質を加え、最終濃度は10 nM～8 μ Mであった。結合システムを25℃で30分間インキュベートした。そして、0.4 U Msp Iを加えてインキュベーションを15分間続けた。10 μ lのホルムアミドを添加して反応を停止させ、その後95℃で5分間加熱した。保護されたおよび切断されたDNAをUrea-PAGEで分離し、化学発光核酸検出モジュールキット (Chemiluminescent Nucleic Acid Detection Module Kit (Thermo)) でイメージングした。

【0120】

アッセイを、RVD HDを用いて最適化した。HDは、通常のシトシンに対して高い親和性を有する天然結合剤である。HDは、Cに対して低い K_i が観察され、5mCおよび5hmCに対する K_i が少なくとも30倍高いことが観察された（図5bおよびc、図6c）。これは、結合親和性に対する定量的評価における保護アッセイの能力を証明した。このインビトロアッセイでは、NGとN*は5hmCに結合せず、5mCのみに結合することが示されている（図5bおよびc）。スクリーニング結果（図3b）から代表的なRVDを選択して更なる評価をした。5mC特異性RVD HAは5mCに対して最も低い K_i を示し、このインビトロアッセイでは、5mCに対する選択性は、Cおよび5hmCに対する選択性の約5～7倍である。5hmC特異性RVD FSは、5hmCへの結合親和性が、5mCへのHAの親和性ほど強くないように見えるが、5hmCに対する選択性がCおよび5mCに対する選択性の約5～6倍であることを示している。また、縮退RVD RGは5mCおよび5hmCに対してかなりの保護を示しているが、C、5mCおよび5hmCに結合する一般のRVD R*は、前記三つのすべてに対して類似した親和性を示している（図5bおよび図5cを参照）。

【0121】

実施例5 新しいRVDがメチル化依存的に遺伝子発現を活性化した

【0122】

生体内でシトシンのメチル化を認識する新しく同定されたRVDの潜在能力を研究するために、ヒト細胞での標的遺伝子の活性化におけるそれらの性能について研究した。特定の遺伝子の活性化を実現するために以前に開発されたTAL E-VP64を使用してTAL E-activatorを設計構築した（37）。TAL E-activatorプラスミドの骨格には、CMVプロモーター、核局在化シグナル、TAL Eアミノ末端とカル

ボキシル末端の非反復配列、及び活性化因子VP64が含まれており、その具体的な配列について、後文の参考文献37を参照できる。

使用中、異なるRVDを含むTALE繰り返し単位がTALE-activator骨格に挿入され、異なるRVDの効果が検証された。その構築方法について、Yang, Junjiao, et al. "Assembly of Customized TALE Effectors Through Advanced ULTIMATE System." TALENs: Methods and Protocols (2016): 49-60を参照できる。

【0123】

まず、USCSデータベースからの既存のメチル化データを使用して、TET1遺伝子を選択し、そのプロモーターはHeLa細胞では高いメチル化レベルを有しているが、HEK293T細胞では脱メチル化されている(図7a)。標的TET1遺伝子を含むTALE繰り返し単位のTALE-activatorを構築した。HeLa細胞では、5mC特異的HA、縮退RG、および一般のR*はいずれもTET1の発現を有意に活性化し(有意な活性化の基準は、対照群と比べてTET1の発現量を高めることができ、しかも発現量が対照群と比べて有意に向上したことであり、*、 $P < 0.05$; **, $P < 0.005$)、その中で、RGは約10倍の活性化に達し(図7b)、同定された3つの新しいRVD(HA、RG、R*)はいずれも、NGとN*よりも優れた性能を発揮したことは証明された。また、HDはTET1の発現を有意に向上させることはない。HEK293T細胞では、HDは、脱メチル化されたTET1プロモーターによく結合し、その発現をさらに強め(すでに高い発現レベルを有しているが)、HAとRGは遺伝子発現に影響せず、一般のR*は通常のCに対する親和性は、HDより低く、遺伝子発現をわずかにアップレギュレートした。NGとN*は、未修飾Cを区別付けることが難しいため、TET1遺伝子発現をわずかに活性化する。

【0124】

次に、LRP2遺伝子を標的とするTALE繰り返し単位を含むTALE-activatorを構築した。それらはLRP2遺伝子のプロモーター領域を標的しおり、このプロモーター領域は、HeLa細胞では中程度にメチル化されているが、HEK293T細胞では脱メチル化されている(図7c)。また、この領域には2つのCpGしか含まれていないため、RVDを介した区別にとってさらに困難である。

HEK293TとHeLa細胞を6ウェルプレートに播種し、60%の密度まで増殖させた。各ウェルについて、2 μ gのTALE-アクチベータープラスミドをLipofectamine(R)2000(Invitrogen)でトランスフェクションした。mCherry陽性細胞をフローサイトメトリーにより選別する前に、トランスフェクションされた細胞を3日間培養した。総RNAをmCherry陽性細胞から単離して逆転写をし、ViiATM7 Real-Time PCR System(Applied Biosystems)でSYBR Green 2X premix II(Takara)を使用して標準反応条件下でリアルタイムPCR分析を行った。

5mCに結合したRVD(HA、RG)がHeLa細胞において遺伝子を有意に活性化したことは観察された。HEK293T細胞では、HDと一般のRVD R*のみがLRP2遺伝子の発現を活性化し、5mCに結合したRVDはLRP2遺伝子の発現を活性化しなかった(図7d)。したがって、同定された新しいRVD(HA、RG)は、インビボで中程度メチル化された部位と、メチル化されていない部位とを区別付けることができる。

【0125】

実施例6 新しいRVDによるメチル化依存性ゲノム編集

【0126】

メチル化依存性ゲノム編集の可能性を確認するために、異なるRVDを含むTALEN構築体(TALE繰り返し単位をTALEN発現ベクターに挿入して得られ、TALEN発現ベクター(すなわち、TALENプラスミドの骨格)には、CMVプロモーター、核

局在化シグナル、T A L E アミノ末端とカルボキシル末端の非反復配列、及びエンドヌクレアーゼの F o k I モノマーが含まれており、その具体的な配列について、以下の参考文献 37 を参照でき、その構築方法について、Yang, Junjiao, et al. “Assembly of Customized TAL Effectors Through Advanced ULtimate System.” TALENs: Methods and Protocols (2016): 49–60 を参照できる) を使用してヒト PLXNB2 遺伝子を標的として DNA 切断を行った (図 7 e)。HeLa 細胞で高度にメチル化されている (UCSC によるデータ)、PLXNB2 の 2 番目のエクソンを選択し、TALEN を介した DNA 切断を indel 比 (すなわち、挿入欠失比) を使用して評価した。

【0127】

HeLa 細胞を 6 ウェルプレートに播種し、60% の密度まで増殖させた。各ウェルについて、Xtreme Gene HP (Roche) を使用して、一対の TALEN プラスミドと pmaxGFP (Lonza Group Ltd.) を 9:9:2 の比率 (0.9 μg:0.9 μg:0.2 μg) でコトランスフェクションした。GFP 陽性細胞をフローサイトメトリーにより選別する前に、トランスフェクションされた細胞を 3 日間培養した。TALEN-標的領域は、単離された GFP 陽性細胞のゲノム DNA から PCR により増幅されたものである。前述のように、ミスマッチ感受性 T7 エンドヌクレアーゼ (T7E1; New England Biolabs) を使用して、TALEN を介した indel を分析した (41)。

【0128】

その結果、TALEN-HD は、無視されてもよい編集効率を示し (図 7 f)、これは、この領域には 3 つの 5mC 修飾があり、その結合を効果的に阻止したことを示した。HD を含む 3 つの RVD が、5mC に結合した RVD (HA、R*、NG、および N* が検出された) で置き換えられると、高い indel 比が観察された (図 7 f および 図 8 c)。これらの結果は、これらの RVD がヒト細胞でメチル化依存性ゲノム編集を可能にしたことを示している。

【0129】

実施例 7 単一塩基解像度での哺乳動物ゲノム中の 5hmC の RVD 媒介検出

【0130】

シトシンのメチル化比率は、バイサルファイトシーケンスで確定できるが、従来のバイサルファイトシーケンスでは 5hmC と 5mC を区別付けることができない (38)。C と 5mC に結合する TALE タンパク質を使用した間接的な 5hmC 検出は以前にも報告されている (32)。5hmC 認識 RVD を含む TALE タンパク質を使用して直接 5hmC 検出を行う可能性を研究するために、まず、5hmC、5mC、および C を特定の部位に取り込んだモデル DNA 配列を合成し、5hmC 検出に対する RVD FS の選択性を検出した。インビトロ保護アッセイでは、5hmC の割合が増加するにつれて、保護された完全長 DNA が直線的に増加した (図 10)。逆に、5mC と C の割合が変化すると、保護率はわずかに変化した。この実験では、それぞれ C、5mC、および 5hmC を含みかつ同じ配列を持つ DNA フラグメントを図に示す割合で混合した。黒丸は、5mC と 5hmC の混合物における 5hmC の割合が 0% から 100% に増加した時の保護程度の変化を示す。黒色三角は、C と 5hmC の混合物における 5hmC の割合が 0% から 100% に増加した時の保護程度の変化を示す。灰色の円は、C と 5mC の混合物における 5mC の割合が 0% から 100% に増加した時の保護程度の変化を示す。図 10 から分かるように、C と 5mC の混合物における 5mC の割合が増加するにつれて、TALE-FS が DNA に対する保護程度は僅かにしか増加していない。これと比べて、C と 5hmC の混合物、および 5mC と 5hmC の混合物における 5hmC の割合が増加するにつれて、TALE-FS が DNA に対する保護程度も大幅に増加し、これは、TALE-FS が、5hmC を含む DNA フラグメントに対して選択的な保護作用を有することを示している。これらの観察結果は、5hmC 特異的な RVD FS がゲノム DNA サンプルに使用さ

10

20

30

40

50

れて、複雑な修飾状況（興味のあるヌクレオチドには、少なくともC、5mC、および5hmC修飾が同時に存在する）下で5hmC修飾を検出することができることを示している。

【0131】

そして、FSを含むTALEタンパク質（すなわち、図9aのTALEフェクタータンパク質）を使用してゲノムDNAで部位特異性5hmC検出を行った。ゲノムDNAの複雑性を考慮にして、制限性酵素の代わりにCRISPR/Cas9システムを使用し、この保護アッセイでDNA切断を生成した（図9a）。マウスslc9a9遺伝子のイントロンの10bp配列を選択した。その中の最初のシトシンがmES細胞で高度にヒドロキシメチル化されていることは報告された（39）。

10

【0132】

反応条件は次の通りである。各10μlの反応系には、50ngのゲノムDNA、1μlの10X Cas9ヌクレアーゼ反応緩衝液（NEB）、および1nMのDTTが含まれている。TALEタンパク質を20nM~500nMの最終濃度まで添加した。結合反応を25℃で30分間インキュベーションした。5μlのプレインキュベーションしたCas9とsgRNAを加え、37℃で1時間インキュベーションを続けた。95℃で5分間加熱することにより反応を停止させた。Ampure BeadsでDNAを精製し、SYBR Green 2X premix II（Takara）on LightCycler(R)96（Roche）でqPCRを分析した。

【0133】

20

その結果、TALE-FSの保護効率はTALE-HDよりはるかに高く（図9b）、これは、TALE-FSがゲノムDNAの複雑な環境での一つの単一の5hmC部位を検出できることを示している。5hmC検出におけるこの方法の能力をさらに研究するために、この方法を、同じ部位でのヒドロキシメチル化レベルが不明な他の細胞株のゲノムDNAに適用した。mESCサンプルと比べて、比較的に低い濃度のTALEタンパク質（RAW264.7、L-M（TK-）、およびL929細胞）が存在する場合、これらの細胞のゲノムDNAに対する保護ははるかに弱く（図9c）、これは、これらの細胞でのこの特定の部位での5hmCのレベルが低いことを示している。上記結果は、新しく同定されたRVDを含むTALEタンパク質を、塩基レベルの解像度でゲノムDNAにおけるヒドロキシメチル基の状態の検出に使用することができることを示している。

30

【0134】

実施例8 6mAを認識するTALEタンパク質RVDの同定

【0135】

実施例2に記載したのと同じスクリーニングシステム、すなわち、TALE-(XX')₃独立RVDライブラリーと6mA修飾を含む直線状DNAレポーターシステムを使用して、それぞれHEK293T細胞にコトランスフェクションした後、6mAレポーターシステムに対するTALE-(XX')₃のEGFP発現活性化倍数をフローサイトメトリーにより検出した。図11は、6mAに対する420のRVDのスクリーニング結果のヒートマップである。

【0136】

40

6mAのスクリーニング結果のヒートマップから分かるように、6mAレポーターシステムに対して活性化効率を有するTALE-(XX')₃は比較的に多く、その1位のアミノ酸は主にHis（H）、Lys（K）、Asn（N）、およびArg（R）であり、これらの高効率なRVDの2位のアミノ酸のほとんどは、Ile（I）、Pro（P）、Ser（S）、Thr（T）またはVal（V）である。重ね合わせたヒートマップ（図11c）から分かるように、上記の6mAに対して高い認識能力を持つRVDの中で、XI、XS、XT、XVなどのシリーズの多くのRVDは、非修飾アデニンに対してもよい認識能力を持っている。また、たとえばXPシリーズのRVDのような一部は6mAに対する特異性が良い。図12は、予備スクリーニング結果から、6mAに対して認識能力を持っているいくつかのRVDを選択して実験し、具体的には、6mAのレポーターシステ

50

ムに対してEGFP活性化倍数が5より大きいいくつかのRVDを選択して3回重複した確認実験をした結果である。

【0137】

6mAに対する認識能力および選好は全体から見ると、依然としてRVDの2位のアミノ酸と密接に関連している。本研究では、XPシリーズRVDとNA、CV、FTなどのRVDは、6mAに対する明確な選好を示すが、XI、XCおよび一部のXTシリーズは、非修飾アデニンおよびN6-メチルアデニンに対する認識に明確な選好がない。Ile(I)とA塩基の接触は、その側鎖とアデニンC8およびN7との間に形成されたファンデルワールス相互作用である(45)ため、6位アミノ基上のメチル基の増加による影響を受けない可能性がある。6mAに対して高い特異性(6mA/A>5)を持つRVDの中で、FT、CV、CPおよびNPは、メチル化修飾のないその他の塩基に対する認識のバックグラウンド値が低く(図13)、その中で、NPが6mAに対する認識能力は最も高く、FTが低く、CVおよびCPがより低くなり、これらは、6mAに対して最適な選好を有するRVDの選択肢と見なすことができる。

【0138】

以上によって、一般に、13位に小さいアミノ酸(GlyとAla)である場合、あるいはアミノ酸が欠失している場合、5mCに対する親和性が増加できることは本研究により発見された。この観察結果は、N*およびNG(自然にTを認識する)が5mCに結合できるという前の発見と一致している。13位に大きな側鎖が欠失することは、5mCのメチル基を収容するのに十分なスペースを提供できる可能性がある。ただし、この一般的な傾向には例外がある。例えば、5mCに対するHGの親和性が非常に弱く、HDと比べてHGが13位において小さい残基を含み、HDがCの天然の結合体であることは観察された。興味深いことに、12位のHisがArgに置き換えられた場合(したがってRGになる)、5mCへの強い結合が観察された。実は、RGは5hmCをも認識する。これらの観察結果は、二重残基が修飾に対する認識にはより複雑なモードが存在するかもしれないことを示している。TALEsの修飾認識メカニズムを十分に解明するには、これらの新しいRVDsと修飾シトシンとから形成された複合体におけるこれらの新しいRVDの結晶構造が必要である。

【0139】

本発明はさらに、TALEに媒介された、いくつかの高度にメチル化されたゲノム領域のメチル化依存性遺伝子活性化とゲノム編集を証明した。重要な対照として、同じ領域にシトシンのメチル化が存在しない場合(異なる細胞内)、遺伝子の活性化はほとんど観察されていない。したがって、本研究で報告された新しいRVDは、インビボでの標的遺伝子の修飾状態によって標的遺伝子を操作するという可能性を提供することができる。ゲノム刷り込みと疾患などを含む、多くの重要な生物学的イベントに係るメチル化可変領域(differentially methylated region、DMR)が多く存在することは知られている。したがって、エピジェネティックマーカーを読み取るTALEタンパク質の独特な能力により、将来的にはインビボでのエピゲノム依存性の応用が可能になる。

【0140】

また、本研究は、CV、FT、NPなど、N6-メチルアデニンに対して選好度の高いRVDを、ハイスループットスクリーニング法により見つけた。これらのRVDは、配列特異的N6-メチルアデニン結合TALEタンパク質の構築に使用でき、抗体に類似した役割を果たすことができ、さらに、未修飾のA塩基のみを認識するRVDと組み合わせて使用して、6mAの定量的または定性的検出を行うこともできる。NIなどの、6mAおよびA塩基に対して選好のないRVDは、潜在的なアデニンメチル化修飾を含む配列を偏差なくターゲティングすることに使用でき、これによって、メチル化修飾による遺伝子編集の効率低下という問題を克服することができる。

10

20

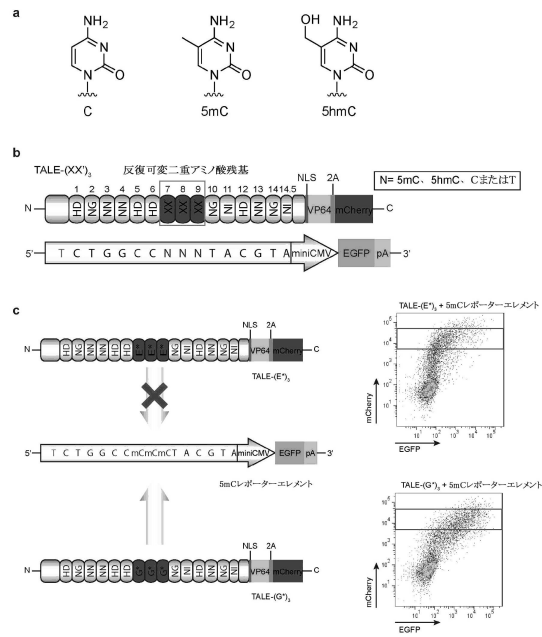
30

40

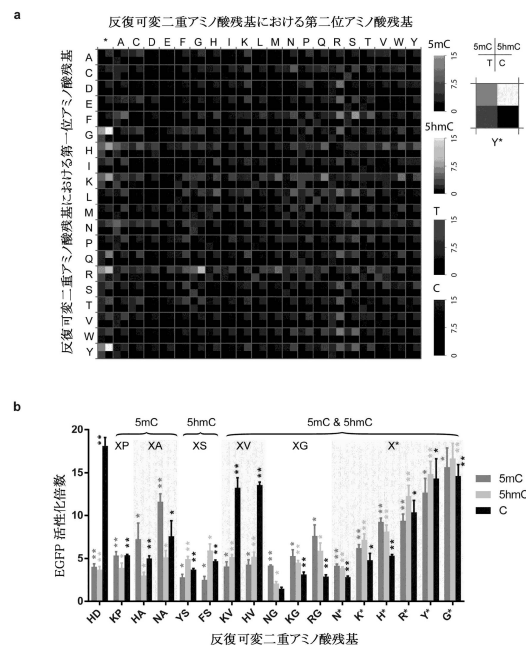
50

【図面】

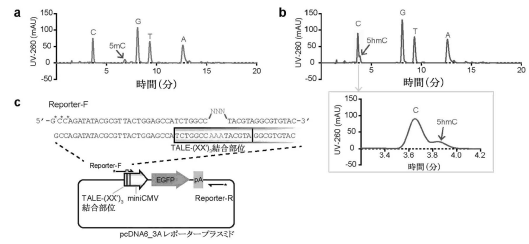
【図 1】



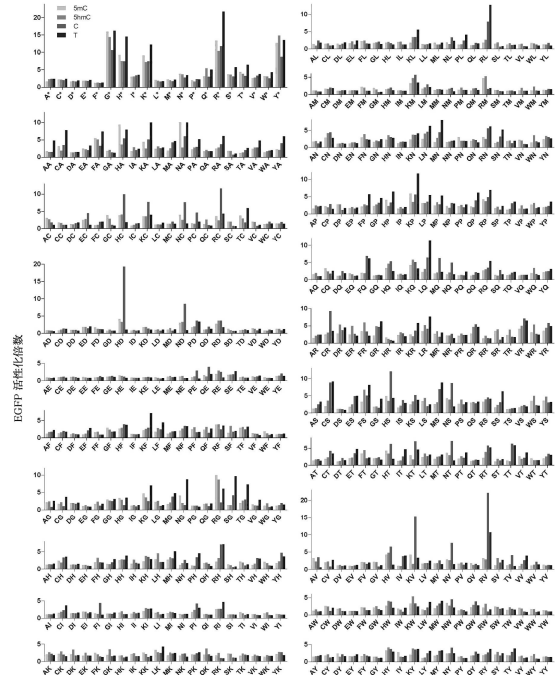
【図 3】



【図 2】



【図 4】



10

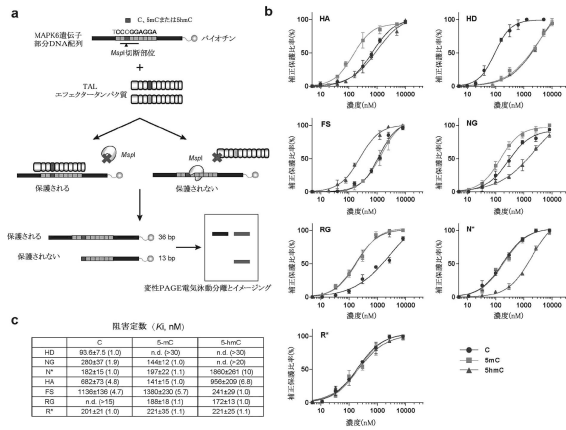
20

30

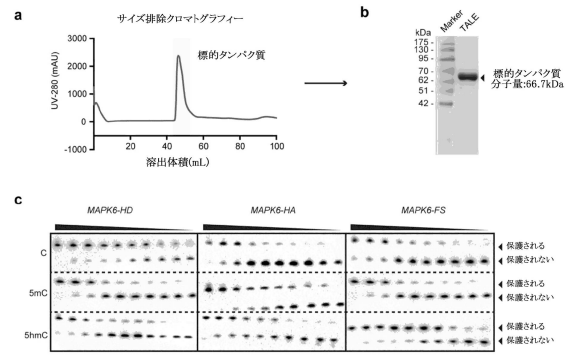
40

50

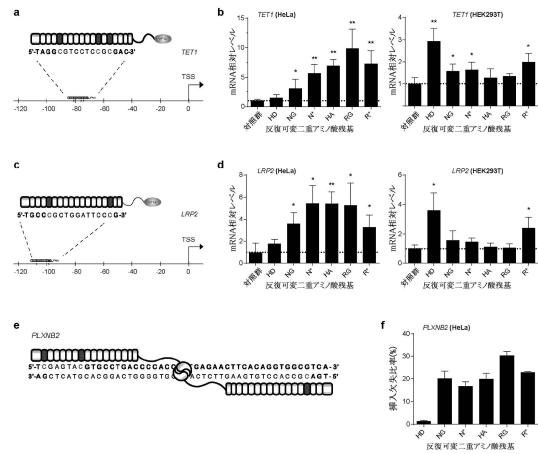
【図 5】



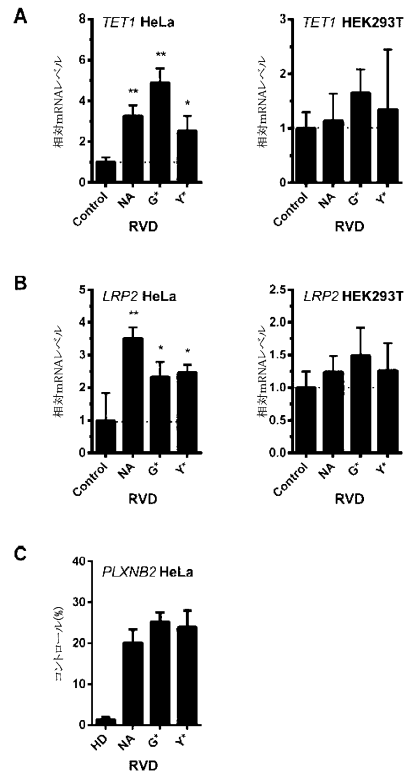
【図 6】



【図 7】



【図 8】



10

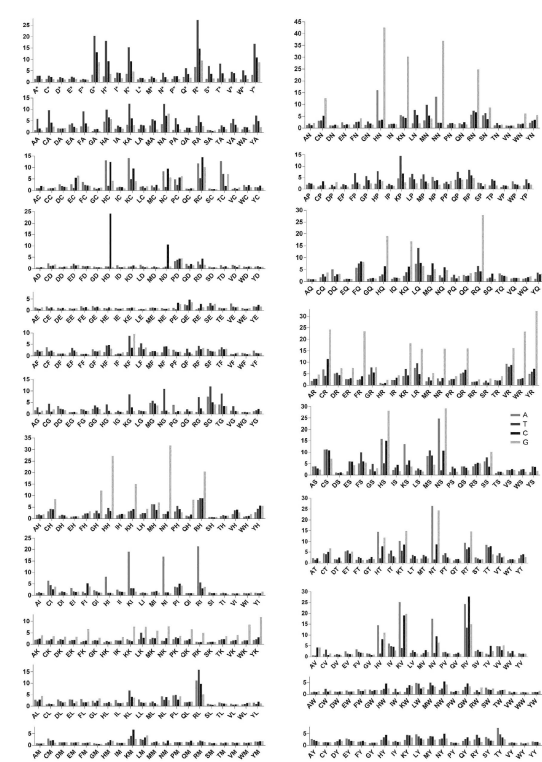
20

30

40

50

【図 13】



10

20

【配列表】
0007207665000001.app

30

40

50

フロントページの続き

フロア 2, ライフ サイエンス パーク

Life Science Park, Floor 2, Building 2, 22 KeXueYuan Road, Changping District, Beijing 102206 China

(74)代理人 110002000

弁理士法人栄光事務所

(72)発明者 魏文▲勝▼

中華人民共和国 100871 ペキン ハイディアンのディストリクト, 5 ユイヒユアン ロード, ペキン ユニバーシティ内

(72)発明者 伊成器

中華人民共和国 100871 ペキン ハイディアンのディストリクト, 5 ユイヒユアン ロード, ペキン ユニバーシティ内

(72)発明者 ▲張▼媛

中華人民共和国 100871 ペキン ハイディアンのディストリクト, 5 ユイヒユアン ロード, ペキン ユニバーシティ内

(72)発明者 郭生杰

中華人民共和国 100871 ペキン ハイディアンのディストリクト, 5 ユイヒユアン ロード, ペキン ユニバーシティ内

(72)発明者 朱晨旭

中華人民共和国 100871 ペキン ハイディアンのディストリクト, 5 ユイヒユアン ロード, ペキン ユニバーシティ内

(72)発明者 ▲劉▼▲路▼▲路▼

中華人民共和国 100871 ペキン ハイディアンのディストリクト, 5 ユイヒユアン ロード, ペキン ユニバーシティ内

審査官 中野 あい

(56)参考文献 特表2013-529083 (JP, A)

米国特許出願公開第2016/0138094 (US, A1)

特表2015-510772 (JP, A)

Cell Res, 2012 Sep 4, vol. 22, no. 10, pp. 1502-1504

Chembiochem, 2015 (Epub 2014 Dec 17), vol. 16, no. 2, pp. 228-231

Chem Commun, 2016, vol. 52, no. 99, pp. 14238-14241

Nat Commun, 2017 Oct 12, vol. 8, no. 1, article no. 901 (pp. 1-9)

(58)調査した分野 (Int.Cl., DB名)

C12N 15/00-15/90

JSTPlus/JMEDPlus/JST7580 (JDreamIII)

Caplus/MEDLINE/EMBASE/BIOSIS (STN)