Electronic Supplementary Material (ESI) for Analyst.

Sequence-specific 5mC detection in live cells based on the TALE-split luciferase complementation system

*Shogo Tsuji, Kouki Shinoda, Shiroh Futaki, Miki Imanishi**

SUPPORTING INFORMATION

Material and methods

Plasmid construction

The mammalian cell expression vector L-TALE$_{LINE1}$(HD)-Sm that targets the LINE1 region with a nuclear localization signal (NLS) driven by a CMV promoter was constructed as follows. First, according to the method described by Sakuma *et al*., 16.5 tandem repeats were assembled to correspond to the L-TALE$_{LINE1}$(HD) RVDs and inserted into ptCMV-136-63-VR-HD (Addgene plasmid50699), creating L-TALE$_{LINE1}$(HD).[1] Then, the Sm-BiT coding sequence from the N197 pBiT2.1-C [TK/SmBiT] Vector (Promega) was inserted into the 3'-terminal end of the L-TALE$_{LINE1}$(HD) coding sequence, creating L-TALE$_{LINE1}$(HD)-Sm/CMV. L-TALE$_{LINE1}$(NG)-Sm/CMV was constructed similarly, except that RVD "NG" was placed at positions corresponding to CpG cytosine in the target site.

The mammalian cell expression vector R-TALE$_{LINE1}$ with an NLS driven by the CMV promoter was constructed using the platinum gate TALEN construction method.[1] The Lg-BiT coding sequence from the N196 pBiT1.1-C TK.LgBiT Vector (Promega) was inserted into the 3'-terminal end of the R-TALE$_{LINE1}$-Lg coding sequence, creating R-TALE$_{LINE1}$-Lg/CMV.

The bacterial expression vector L-TALE$_{LINE1}$(HD)-Sm was constructed as follows. The coding region of L-TALE$_{LINE1}$(HD)-Sm was inserted into pET42b (Novagen), and the His-tag coding sequence was inserted into 5'-terminal end of the L-TALE$_{LINE1}$(HD)-Sm coding sequence, creating L-TALE$_{LINE1}$(HD)-Sm/pET42b. The bacterial expression vectors L-TALE$_{LINE1}$(NG)-Sm L-TALE$_{art}$(HD)-Sm, L-TALE$_{art}$(NG)-Sm, and R-TALE$_{LINE1}$-Lg were constructed similarly.

The sequences of all these plasmids were confirmed. The amino acid sequences of these proteins are described below.

Protein purification and electrophoretic mobility shift assay (EMSA)

The L-TALE-Sm and R-TALE-Lg proteins were expressed in *E. coli* BL21 (DE3) cells by induction with 0.2 mM IPTG for 12 h at 18°C. The proteins were purified, and EMSA was performed as previously described.[2] The oligo DNAs used in EMSA are shown in **Fig. S1, S5**.

In vitro TALE-NanoBiT fragment complementation assay

L-TALE-Sm, R-TALE-Lg, and oligo DNA were mixed in TALE-EMSA buffer (12 mM Tris-HCl, pH 7.5, 60 mM KCl, 2 mM DTT, 0.05% NP-40, 50 ng/μL double-stranded poly (deoxyinosine-deoxycytosine)n (dIdC), 0.1 mg/ml bovine serum albumin, 5% glycerol, 5 mM $MgCl_2$, 0.2 mM EDTA). After incubation for 30 min at 25°C, the reaction mixture was kept at 4°C for 15 min. Then Nano-Glo Luciferase Assay Substrate (Promega) was added to the reaction mixture, and luminescence intensity was measured with the GloMax-Multi Detection System (Promega). For the experiments using oligo DNAs having different methylation percentages, C and 5mC oligo DNAs were mixed in the indicated ratios. Evaluation of NanoBiT complementation on gDNA in vitro was performed similarly except that gDNA was added instead of oligo DNA.

Genomic DNA (gDNA) preparation and methylation

gDNA was extracted from HCT116 and DKO cells (gifts from H. Suzuki) using a NucleoSpin Tissue kit (TAKARA). A portion of the gDNA from HCT116 was enzymatically methylated using M.SssI CpG methyltransferase (NEB) according to the manufacturer's instructions. Methylated DNA was purified with a NucleoSpin gDNA Clean-up kit (TAKARA).

Bisulfite sequencing and combined bisulfite restriction analysis (COBRA) of LINE1

The extracted gDNAs and methylated gDNA were bisulfite converted using an EZ DNA Methylation-Gold Kit (ZYMO research) according to the manufacturer's instructions. The converted DNA was PCR amplified by EpiTaq™ HS (TAKARA) and cloned into a pMD20T-Vector (TAKARA). Then, sequencing was performed. The sequencing results were analyzed using QUMA software.[3] COBRA was performed based on the method reported by Pobsook *et al.*, except that *Mlu*CI, an isoschizomer of *Tas*I, was used instead of *Tas*I.[4] Restriction enzyme treatments were performed according to the manufacturer's instructions. In the polyacrylamide electrophoresis, 60 and 175 bp ds DNA were used as makers.

TALE-NanoBiT fragment complementation assay in live cells

HCT116 and DKO cells were seeded into a 96-well Cliniplate (Thermo Fisher Scientific) and cultured until they were about 60-70% confluent. TALE-NanoBiT fragment

expression vectors were transfected into HCT116 and DKO cells using Lipofectamine 3000 (Thermo Fisher Scientific). After 24 h, the culture medium was changed to Opti-MEM (Thermo Fisher Scientific). Then, NanoBiT fragment complementation was measured using the Nano-Glo Live Cell Assay System (Promega) according to the manufacturer's instructions. The luminescence intensity was measured with the GloMax-Multi Detection System (Promega).

Western blotting

Cells were lysed with RIPA buffer, and cell lysates were subjected to SDS-PAGE. The proteins were detected using mouse anti-FLAG (Sigma-Aldrich; F3165) or mouse anti-β-actin (Sigma-Aldrich; AC-74) antibodies as primary antibodies and anti-mouse IgG (GE Healthcare; NA931) antibodies as HRP-labeled secondary antibodies. Bands were visualized using Amersham ECL Prime (GE Healthcare).

Protein sequences

L-TALE$_{LINE1}$(HD)-Sm and L-TALE$_{LINE1}$(NG)-Sm

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAVDLRTLGYSQQQQEKIKP
KVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGV
GKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LN

LTPDQVVAIASNNGGKQALETVQRLLPVLCQDHG
LTPEQVVAIASXXGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASNNGGKQALETVQRLLPVLCQAHG
LTPAQVVAIASHDGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASXXGGKQALETVQRLLPVLCQAHG
LTPAQVVAIASNNGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNGGGKQALETVQRLLPVLCQDHG
LTPEQVVAIASNNGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASXXGGKQALETVQRLLPVLCQAHG
LTPAQVVAIASNNGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASXXGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNIGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNNGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASHDGGRPALE

SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERT

SHRVAGSGSSGGGGSGGGGSSGVTGYRLFEEIL*

= HD or NG

= RVDs

= NLS

= Sm-BiT

= 3 x FLAG tag

R-TALE$_{LINE1}$-Lg

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAVDLRTLGYSQQQQEKIKP

KVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGV

GKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP

LN

LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASHDGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNIGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASNNGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNGGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNNGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASHDGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNGGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNNGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASHDGGRPALE

SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERT

SHRVAGSGSSGGGGSGGGGSSGVFTLEDFVGDWEQTAAYNLDQVLEQGGVSSLLQNLA

VSVTPIQRIVRSGENALKIDIHVIIPYEGLSADQMAQIEEVFKVVYPVDDHHFKVILP

YGTLVIDGVTPNMLNYFGRPYEGIAVFDGKKITVTGTLWNGNKIIDERLITPDGSMLF

RVTINS*

= RVDs

= NLS

= Lg-BiT

= 3 x FLAG tag

His-tag fused L-TALE$_{LINE1}$(HD)-Sm and L-TALE$_{LINE1}$(NG)-Sm

MHHHHHHPGDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAVDLRTLGYSQ

QQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA

THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAW

RNALTGAPLN

LTPDQVVAIASNNGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASXXGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASHDGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASXXGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNNGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASNGGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASXXGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNNGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASXXGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNIGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNNGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASHDGGRPALE

SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERT

SHRVAGSGSSGGGGSGGGGSSGVTGYRLFEEIL*

= HD or NG

= RVDs

= NLS

= Sm-BiT

= 3 x FLAG tag

= His tag

His-tag fused R-TALE$_{LINE1}$-Lg

MHHHHHHPGDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAVDLRTLGYSQ

QQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA

THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAW

RNALTGAPLN

LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASHDGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNIGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASNNGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNGGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNNGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASHDGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNGGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNNGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASHDGGRPALE

SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERT

SHRVAGSGSSGGGGSGGGGSSGVFTLEDFVGDWEQTAAYNLDQVLEQGGVSSLLQNLA

VSVTPIQRIVRSGENALKIDIHVIIPYEGLSADQMAQIEEVFKVVYPVDDHHFKVILP

YGTLVIDGVTPNMLNYFGRPYEGIAVFDGKKITVTGTLWNGNKIIDERLITPDGSMLF

RVTINS*

= RVDs

= NLS

= Lg-BiT

= 3 x FLAG tag

= His tag

His-tag fused L-TALE$_{art}$(HD)-Sm and L-TALE$_{art}$(NG)-Sm

MHHHHHHPGDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAVDLRTLGYSQ

QQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA

THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAW

RNALTGAPLN

LTPDQVVAIASNGGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASXXGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASXXGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASNNGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNIGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNIGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASXXGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNIGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNIGGKQALETVQRLLPVLCQDHG

LTPDQVVAIASXXGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASNNGGKQALETVQRLLPVLCQAHG

LTPDQVVAIASNGGGKQALETVQRLLPVLCQAHG

LTPAQVVAIASNIGGKQALETVQRLLPVLCQDHG

LTPEQVVAIASHDGGRPALE

SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERT

SHRVAGSGSSGGGGSGGGGSSGVTGYRLFEEIL*

= HD or NG

= RVDs

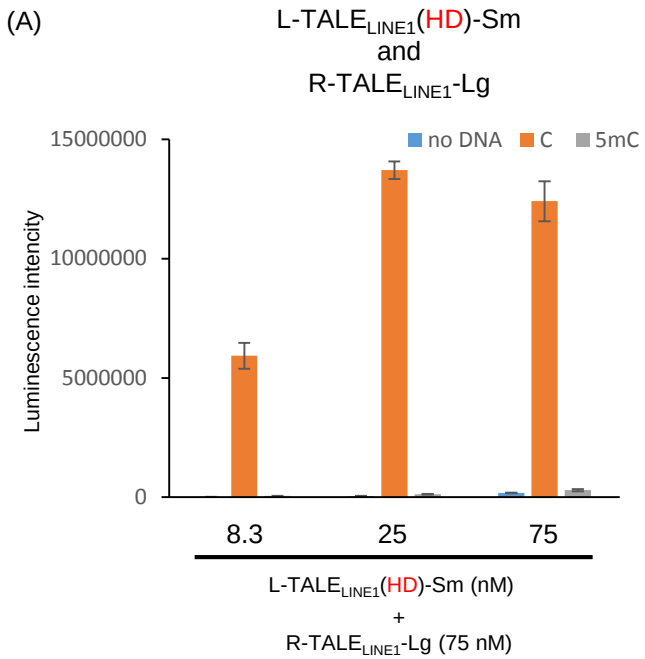
= NLS

= Sm-BiT

= 3 x FLAG tag

= His tag

Fig. S1 Oligo DNA design used in EMSA and in vitro NanoBiT complementation assay. Both oligo DNAs were fluorescent-labeled with 6-carboxyfluorescein (6-FAM). Blue boxes shown the L-TALE$_{LINE1}$ and R-TALE$_{LINE1}$ binding sites. In 5mC oligo DNA, methylated cytosines were shown in red.



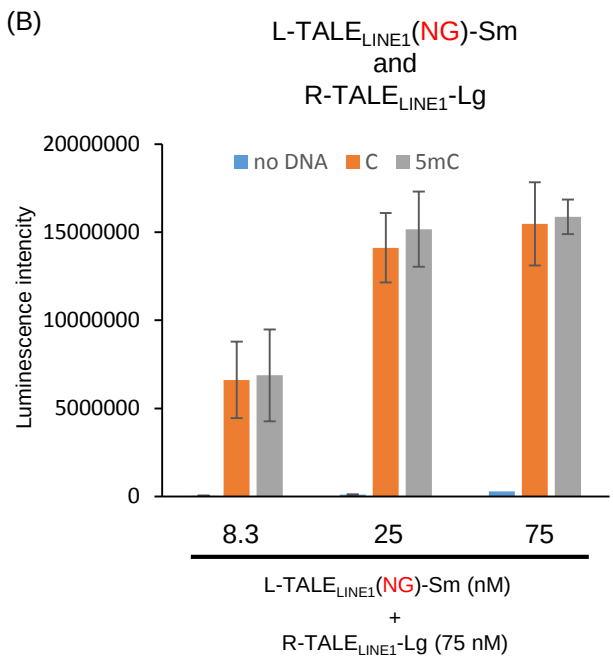
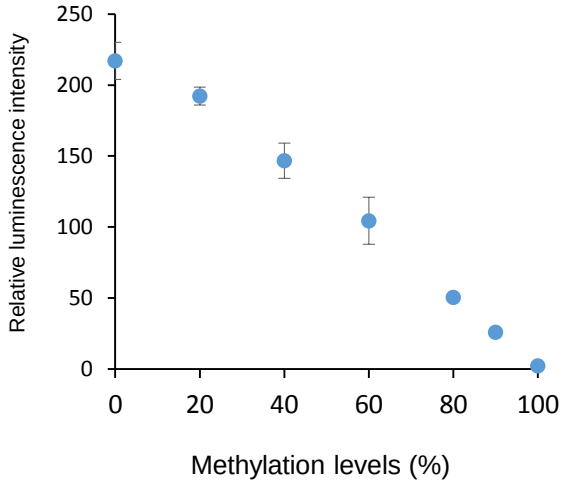


Fig. S2 (A) Luminescence intensities of the mixture containing L-TALE$_{LINE1}$(HD)-Sm (8.3, 25, or 75 nM), R-TALE$_{LINE1}$-Lg (75 nM) and C or 5mC oligo DNA (5 nM). (B) Luminescence intensities of the mixture containing L-TALE$_{LINE1}$(NG)-Sm (8.3, 25, or 75 nM) and R-TALE$_{LINE1}$-Lg (75 nM) and C or 5mC oligo DNA (5 nM). In all experiments, data are expressed as means ± SD (n = 3).

(A)



(B)

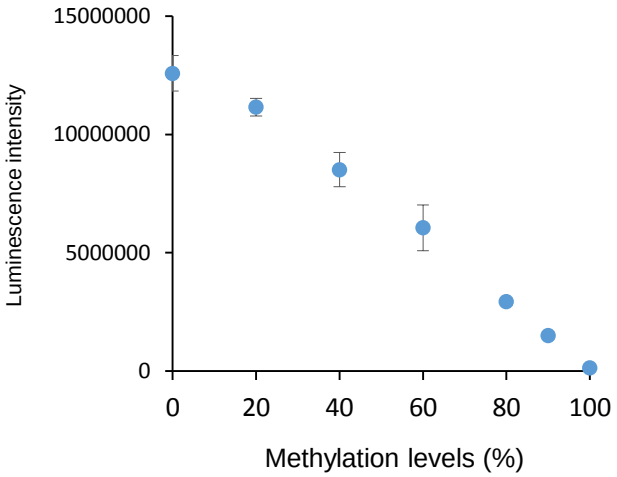


Fig. S3 (A) Luminescene intensities of the mixture containing L-TALE$_{LINE1}$(HD)-Sm (25 nM) and R-TALE$_{LINE1}$-Lg (75 nM) and oligo DNAs having different methylation levels (5 nM). The luminescence intensity was normalized to that of no DNA condition. Raw data are shown in (B). Data are expressed as means ± SD (n = 3).

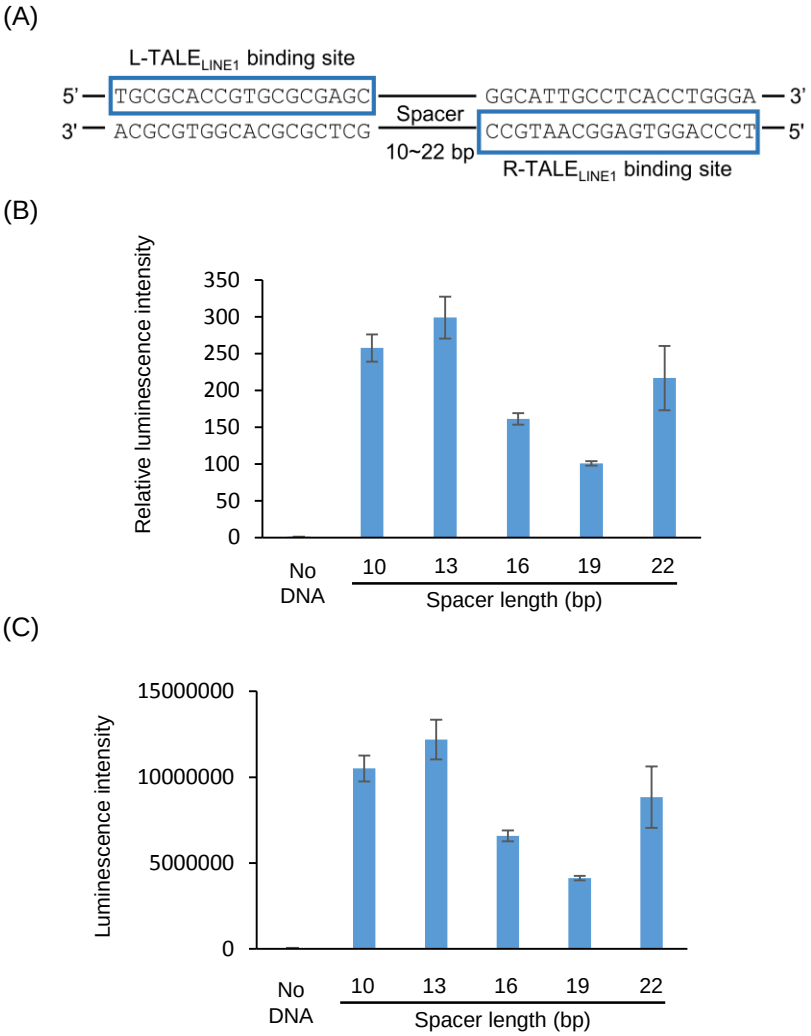
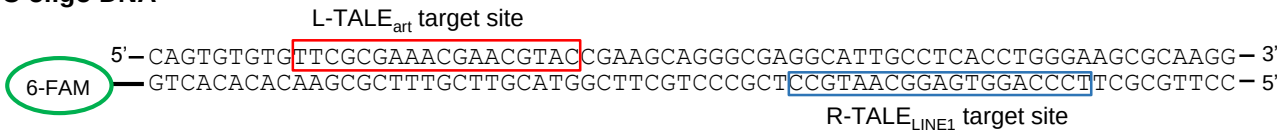


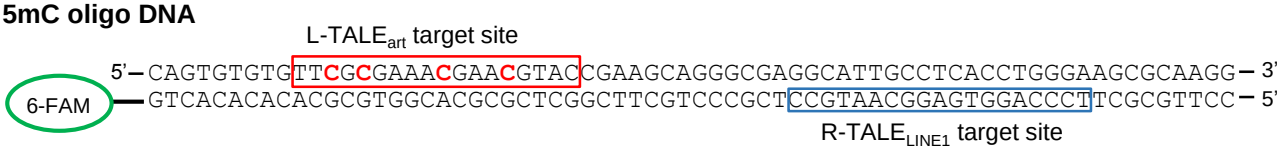
Fig. S4 (A) The design of oligo DNA fragments having different spacer lengths between two TALE binding sites. (B) Luminescene intensities of the mixture containing L-TALE$_{LINE1}$(NG)-Sm (25 nM) and R-TALE$_{LINE1}$-Lg (75 nM) and oligo DNAs having different spacer lengths (5 nM). The luminescence intensity was normalized to that of no DNA condition. Raw data are shown in (C). Data are expressed as means ± SD (n = 3).

(A)

C oligo DNA



5mC oligo DNA



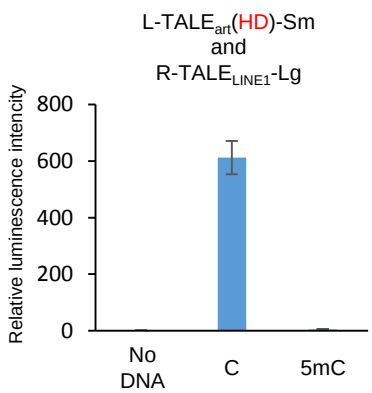
(B)

K_d values of TALE-NanoBiT fragments

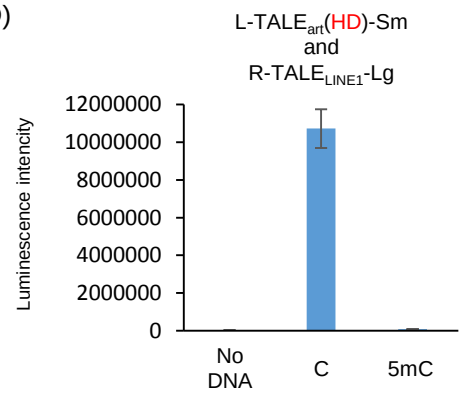
Proteins	K_d (nM)[a]	
	C oligo DNA	mC oligo DNA
L-TALE$_{art}$(HD)-Sm	74	5.4
L-TALE$_{art}$(NG)-Sm	3.0	88

[a] determined by EMSA

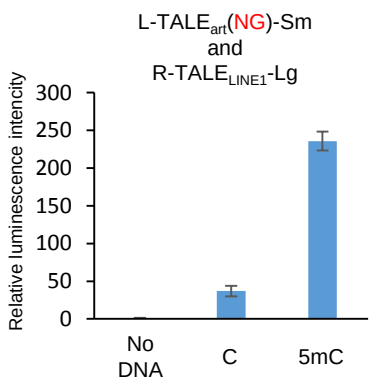
(C)



(D)



(E)



(F)

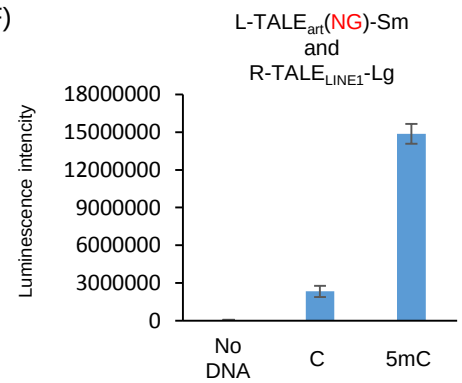
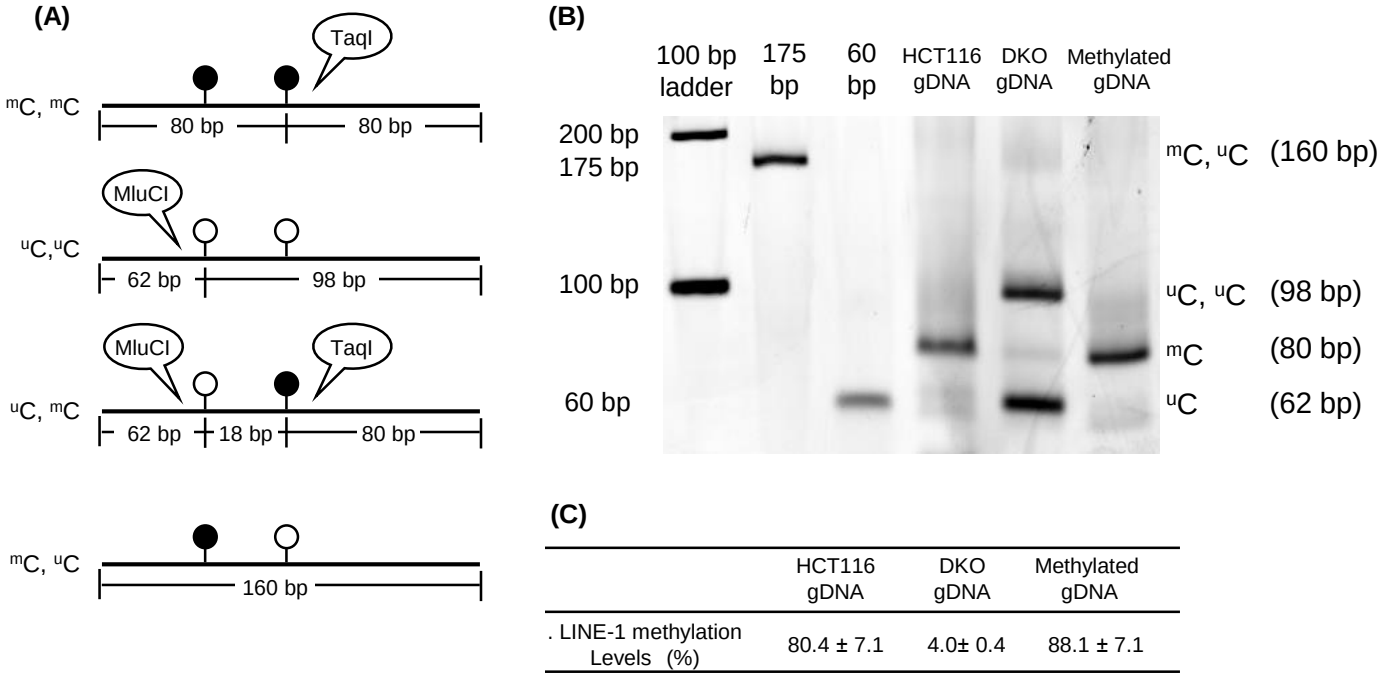


Fig. S5 Methylation detection of oligo DNA by L-TALE$_{art}$-Sm (A) Oligo DNA design used in EMSA and in vitro NanoBiT complementation assay of L-TALE$_{art}$-Sm. Both oligo DNAs were fluorescent-labeled with 6-carboxyfluorescein (6-FAM). Red boxes shown the L-TALE$_{art}$ binding sites. Blue boxes shown the R-TALE$_{LINE1}$ binding sites. In 5mC oligo DNA, methylated cytosines were shown in red. (B) Apparent dissociation constant K_d (nM) of a L-TALE$_{art}$-Sm to the C or 5mC oligo DNA. (n=1) (C) Luminescence intensities of the mixture containing L-TALE$_{art}$(HD)-Sm (25 nM), R-TALE$_{LINE1}$-Lg (75 nM) and C or 5mC oligo DNA (5 nM). The luminescence intensity was normalized to that of no DNA condition. Raw data are shown in (D). (E) Luminescence intensities of the mixture containing L-TALE$_{art}$(NG)-Sm (25 nM) and R-TALE$_{art}$-Lg (75 nM) and C or 5mC oligo DNA (5 nM). The luminescence intensity was normalized to that of no DNA condition. Raw data are shown in (F). In all experiments without (B), data are expressed as means ± SD (n = 3).



(C)

	HCT116 gDNA	DKO gDNA	Methylated gDNA
. LINE-1 methylation Levels (%)	80.4 ± 7.1	4.0± 0.4	88.1 ± 7.1

Fig. S6 LINE1-methylation status confirmation of the gDNAs by COBRA. (A) Positive restriction sites of possible methylation patterns of LINE1 region. Methylated and unmethylated CpGs are ^{m}C and ^{u}C, respectively. The dark circles represent methylated cytosines and the hollow circles represent unmethylated cytosine. (B) Electrophoresis of the gDNAs after MluCI and TaqI treatment. One hundred and sixty, 98, 80, and 62 bp are $^{m}C^{u}C$, $^{u}C^{u}C$, ^{m}C and ^{u}C, respectively. (C) LINE-1 methylation levels were calculated as $^{m}C/(^{u}C^{u}C+^{m}C+^{u}C)\%$. $^{m}C^{u}C$ or $^{u}C^{u}C$ or ^{m}C or ^{u}C was determined as (mCuC or uCuC or mC or uC)/($^{u}C^{u}C+^{m}C^{u}C+^{m}C+^{u}C$)%, respectively. Data are expressed as means ± SD. n=3.

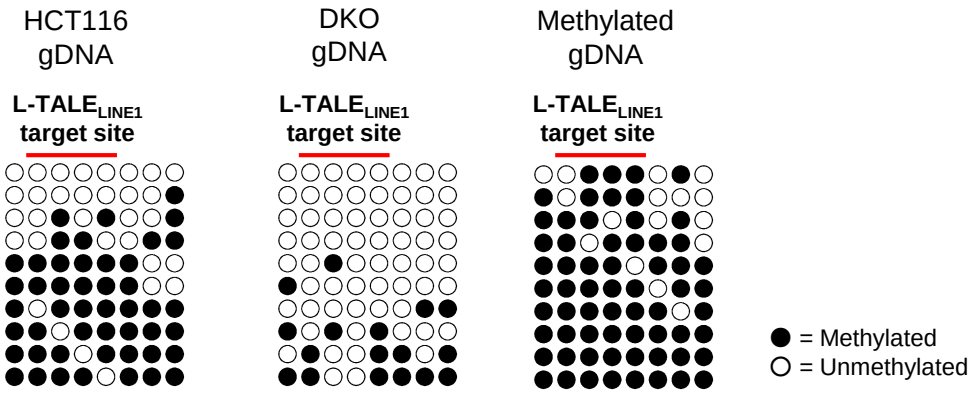


Fig. S7 LINE1-methylation status confirmation of the gDNAs by bisulfite sequencing. The dark circles represent methylated cytosines and the hollow circles represent unmethylated cytosine. Red bars indicate the CpG position targeted by L-TALE$_{LINE1}$.

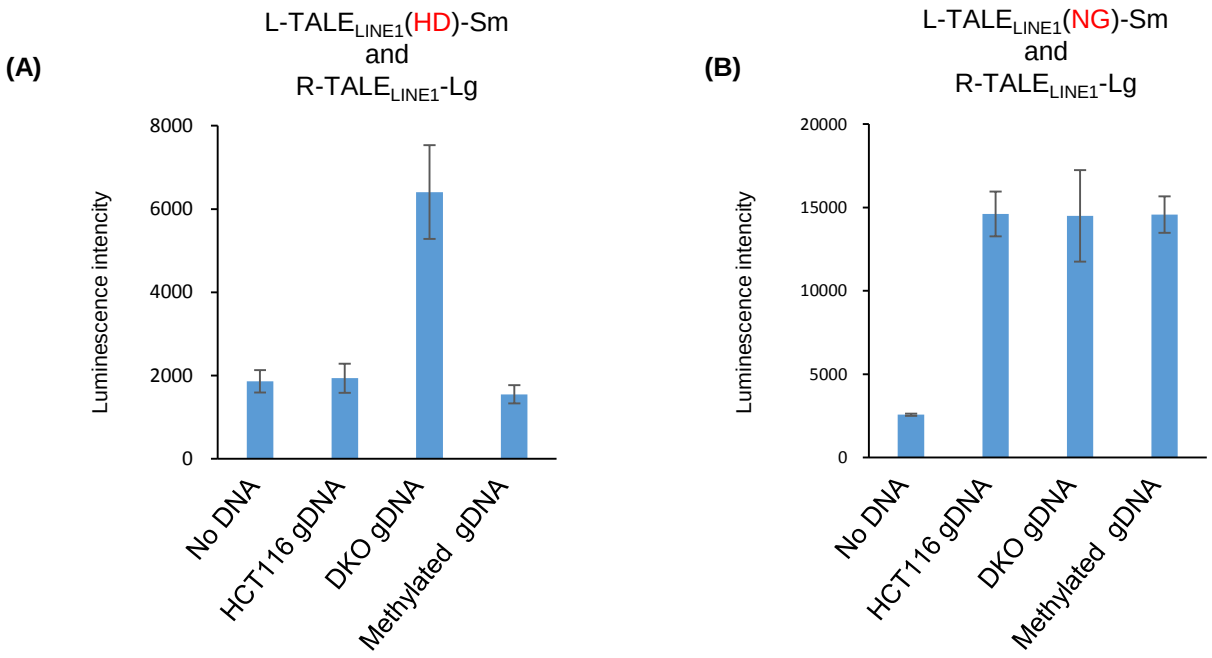


Fig. S8 (A) Luminescene intensities of the mixture containing L-TALE$_{LINE1}$(HD)-Sm (10 nM), R-TALE$_{LINE1}$-Lg (10 nM) and indicated gDNA (45 ng). (B) Luminescene intensities of the mixture containing L-TALE$_{LINE1}$(NG)-Sm (10 nM), R-TALE$_{LINE1}$-Lg (10 nM) and indicated gDNA (45 ng). Data are expressed as means ± SD (n = 3).

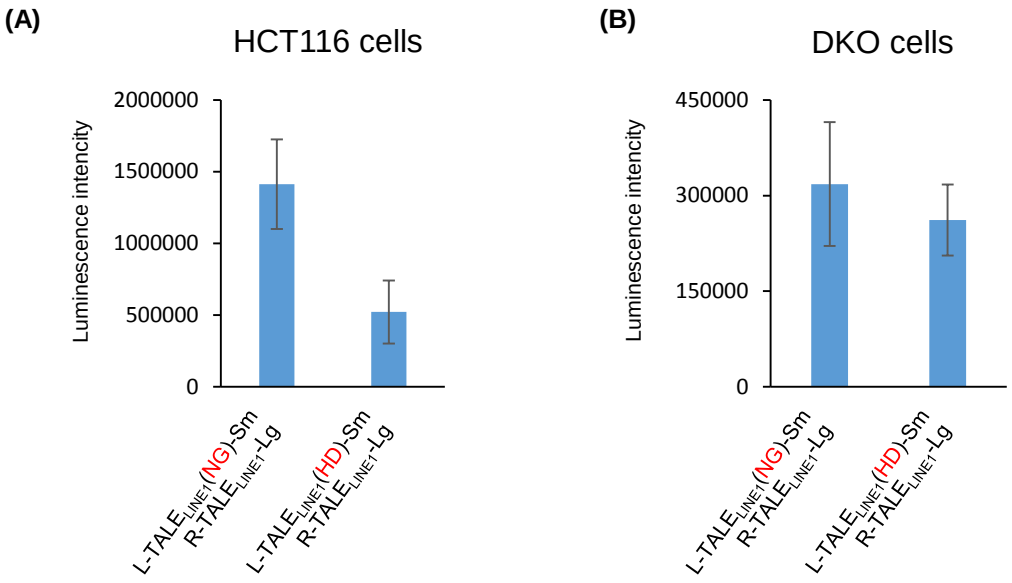


Fig. S9 Live-cell methylation detection of gDNA by TALE-NanoBiT. Luminescene intensities of (B) HCT116 cells or (C) DKO cells. Data are expressed as means ± SD (n = 3). * $P < 0.05$ (Student's *t*-test).

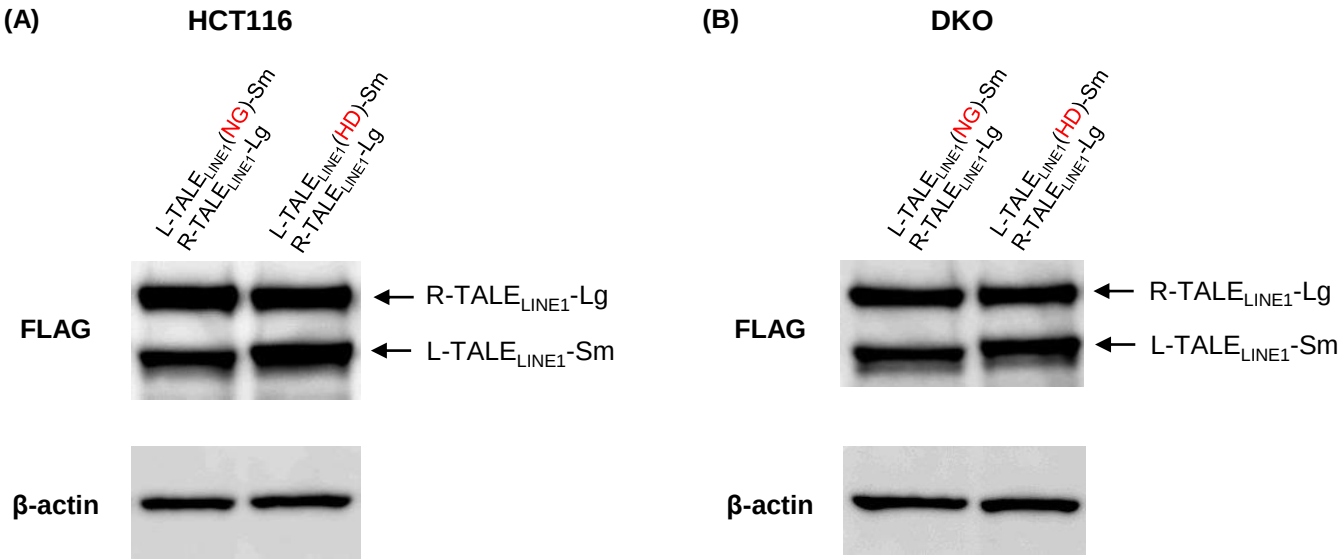


Fig. S10 Expression of TALE-NanoBiT fragments in (A) HCT116 cells and (B) DKO cells. All TALE-NanoBiT fragments contain three FLAG tags at their N-termini.

Acknowledgment

The TALEN kit used for constructing Platinum TALENs was a gift from Takashi Yamamoto (Addgene kit # 1000000043).

References

1 T. Sakuma, H. Ochiai, T. Kaneko, T. Mashimo, D. Tokumasu, Y. Sakane, K. Suzuki, T. Miyamoto, N. Sakamoto, S. Matsuura and T. Yamamoto, *Sci. Rep*., 2013, 3, 3379.

2 S. Tsuji, S. Futaki and M. Imanishi, *Chem. Commun*., 2016, 52, 14238.

3 Y. Kumaki, M. Oda and M. Okano, *Nucleic Acids Res*., 2008, 36, W170.

4 T. Pobsook, K. Subbalekha, P. Sannikorn and A. Mutirangura, *Clin. Chim. Acta*, 2011, 412, 314.