Journal Name

COMMUNICATION

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

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

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

We established a method for converting TALE-DNA binding to luminescence, by combining TALE and a split luciferase system. Furthermore, using a methylation-sensitive TALE, sequence-specific 5mC detection of genomic DNA was achieved in live cells. This study provides a new strategy for exploring the biological functions of 5mC.

DNA methylation, which occurs at cytosine bases of CpG dinucleotides in mammals, is an important epigenetic marker that regulates gene expression, chromatin remodeling, and genome stability.[1] DNA methylation status changes dynamically during biological events and tumorigenic processes.[2] Owing to the importance of methylated cytosine (5mC), numerous 5mC detection methods have been developed.[3] However, methods for detecting 5mC in live cells are limited, and if available, such methods would allow the investigation of the biological roles of the dynamics of methylation.[4] One promising way to detect DNA-methylation in live cells is to use proteins that directly recognize 5mC, such as the methyl-CpG-binding domain (MBD). Changes in DNA methylation status have been visualized by live-cell imaging of a fluorescence protein-fused MBD.[4a-d] However, because the MBD binds to all mCpG sites without sequence selectivity, this method is not applicable to investigate the biological roles of the methylation state at individual cytosine residues.

Transcription activator-like effector (TALE),[5] a programmable DNA-binding protein, is another candidate to detect DNA methylation. The DNA-binding specificity of TALE is determined by a series of tandem repeats. Each repeat recognizes one target base. The base preference of a repeat is defined by two amino acids, called repeat variable diresidues (RVDs). Therefore, TALEs can be readily designed to target specific DNA sequences by simply modifying the RVDs. A recent study revealed that the RVD "HD" could differentiate C from 5mC.[6] This 5mC sensitivity of RVD "HD" allows us to design TALEs that bind to the target sequence only when the target sequence is not methylated. Based on this property of TALE, Kubik *et al*. reported sequence-specific 5mC detection of extracted genomic DNA (gDNA).[6a] In their study, they employed the following principle: TALE binds to a target site of gDNA and inhibits primer extension by KF(exo-) DNA polymerase. Quantification of the extension products led to the quantitative detection of mC levels at the target site. Although this study showed the potential of TALE to discriminate methylation states of a specific site of gDNA, the method was not applicable to in-cell detection. To trace the DNA methylation state over time, a method that converts TALE-DNA binding to a readily detectable signal in live cells is required.

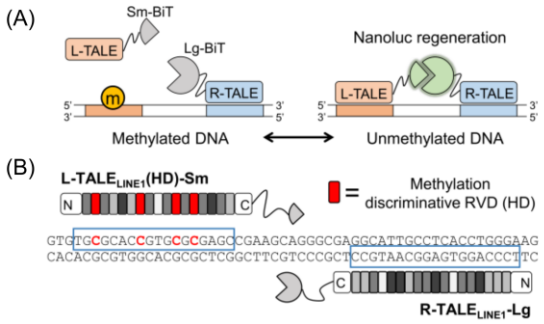

Fig. 1 Schematic representation of sequence-specific DNA methylation detection by TALE-NanoBiT. (A) L-TALE and R-TALE are fused to Sm-BiT and Lg-BiT, respectively. When the L-TALE target sequence is unmethylated, two TALEs come into close proximity, resulting in NanoBiT fragment complementation. (B) L-TALE$_{LINE1}$(HD)-Sm and R-TALE$_{LINE1}$-Lg, targeting LINE1 regions, were designed. RVD "HD" was used for methylation discriminative RVD. CpG cytosines in the target sequence are shown in red. Blue boxes indicate target DNA sequences in the LINE1 region.

To visualize TALE-DNA binding in live cells, bimolecular fluorescence complementation (BiFC) technology has been used. [7] BiFC is based on structural complementation between two non-fluorescent N-terminal and C-terminal fragments of a fluorescent protein.[8] TALEs attached to N- or C-halves of the fluorescent proteins localize on adjacent sites by binding to the target DNA region, resulting in the complementation of the original fluorescent protein. Using BiFC-TALEs, telomeres and centromeres have been visualized in live cells.[7] However, BiFC is irreversible and requires time for fluorophore maturation.[8] The irreversibility and slow-acting property

Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan.
†Electronic Supplementary Information (ESI) available.

Table 1 K_d values of TALE-NanoBiT fragments

Proteins	K_d (nM)[a]	
	C oigo DNA	5mC oligo DNA
L-TALE$_{LINE1}$(HD)-Sm	7.9 ± 3.2	39 ± 9.1
L-TALE$_{LINE1}$(NG)-Sm	14.0 ± 0.9	11.5 ± 1.4
R-TALE$_{LINE1}$-Lg	27 ± 19	29 ± 20

[a] determined by EMSA

of BiFC may obscure the analysis of dynamic changes in DNA methylation. Split luciferase is another class of complementation reporter, and it has been used for investigating protein interactions and epigenetic states of gDNA.[4e, 9] The main advantage of using split luciferase is that the complementation is a reversible reaction.[9a, 10] Among the split luciferases, the recently developed split NanoLuc luciferase system, called NanoLuc Binary technology (NanoBiT), which showed bright luminescence and rapid and reversible reconstitution, seems to be the best complementation reporter for investigating DNA methylation.[11]

In this study, we established a sequence-specific 5mC detection method applicable to cellular assays by combining a methylation-sensitive DNA binding protein, TALE, and a state-of-the-art split luciferase system, NanoBiT. In our method, the small fragment of NanoBiT (Sm-BiT) and the large fragment of NanoBiT (Lg-BiT) are fused to left TALE (L-TALE) and right TALE (R-TALE), creating L-TALE-Sm and R-TALE-Lg, respectively (Fig. 1A). L-TALE is designed to bind to a specific sequence containing a CpG site using RVD "HD" as a methylation discriminative RVD. L-TALE-Sm is expected to bind only to unmethylated target sequences because of the unmethylated C-selectivity of RVD "HD." On the contrary, R-TALE is designed to bind to a sequence that is adjacent to the L-TALE target sequence. R-TALE binds to the target site in a methylation-independent manner because the R-TALE target site has no CpG. Only when the target site is not methylated, both TALEs bind at adjacent sites. The bindings of TALEs bring the NanoBiT fragments fused to TALEs into close proximity, resulting in the complementation of NanoBiT fragments, which generates luminescence.

Here, we targeted long interspersed nuclear elements 1 (LINE1), which is a highly repetitive element. Methylation levels of the LINE1 region are significantly associated with global DNA methylation and have been widely used as a surrogate of global DNA methylation.[12] We designed L-TALE$_{LINE1}$(HD)-Sm, with RVD "HD" as a methylation-discriminative RVD, to target an 18 nt sequence of the LINE1 region containing 4 CpG sites (Fig. 1B). We also designed R-TALE$_{LINE1}$-Lg to target the adjacent locus of the L-TALE$_{LINE1}$(HD)-Sm target site (Fig. 1B). In addition, RVD "NG," which has a slight 5mC-preference, was used instead of RVD "HD" at the positions opposite to CpG cytosine bases to design L-TALE$_{LINE1}$(NG)-Sm as a control.

First, the DNA binding abilities of the TALE-NanoBiT proteins were evaluated by EMSA (Table 1, Fig. S1, ESI†). In EMSA, L-TALE$_{LINE1}$(HD)-Sm showed about 5-fold lower K_d value for the C target than 5mC, while L-TALE$_{LINE1}$(NG)-Sm showed similar K_d values between the C and 5mC targets (Table 1). These data suggested that L-TALE$_{LINE1}$(HD)-Sm discriminated the methylation-states of the target sequence because of the unmethylated-C selectivity of RVD "HD." As expected, R-TALE$_{LINE1}$-Lg, which targeted the sequence with no CpG, showed similar K_d values between the C and 5mC oligo DNAs (Table 1).

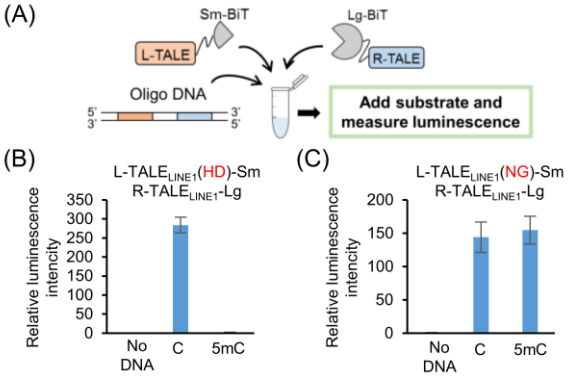

Fig. 2 Evaluation of TALE-NanoBiT fragment complementation on oligo DNA. (A) Schematic representation of the experiment. Luminescence intensities of the mixture containing C or 5mC oligo DNA (5 nM), R-TALE$_{LINE}$-Lg (75 nM), and L-TALE$_{LINE1}$(HD)-Sm (25 nM) (B) or L-TALE$_{LINE1}$(NG)-Sm (25 nM) (C). In all experiments, the luminescence intensity was normalized to that of the no DNA condition. Data are expressed as the mean ± SD (n = 3).

The TALE-NanoBiT fragment complementation was successfully induced in the presence of the target DNA in a manner dependent on the CpG methylation of the target sequence (Fig. 2). In the experiment, L-TALE$_{LINE1}$ (HD)-Sm (25 nM), R-TALE $_{LINE1}$-Lg (75 nM), and either C or 5mC oligo DNA (5 nM) were mixed so that almost all oligo DNA molecules could be bound by R-TALE$_{LINE1}$-Lg and the fraction bound by L-TALE$_{LINE1}$(HD)-Sm would be different between C and 5mC oligo DNAs based on their K_d values (Table 1). Furimazine, a substrate of NanoBiT, was added to the mixtures, and then the luminescence intensity was measured (Fig. 2A). The luminescence intensity observed in the mixture containing C oligo DNA was about 280-fold higher than that in the no DNA condition, while a slight increase (2.6-fold) was observed in the mixture containing 5mC oligo DNA (Fig. 2B, S2A). In addition, the luminescence intensity decreased in proportion to the methylation percentage of the oligo DNA (Fig. S3). We also examined the luminescence intensity at higher and lower L-TALE$_{LINE1}$(HD)-Sm (Fig. S2A). Under all examined conditions, L-TALE$_{LINE1}$(HD)-Sm in combination with R-TALE$_{LINE1}$-Lg showed significantly higher luminescence signals for the C oligo DNA compared with 5mC oligo DNA. On the other hand, the L-TALE $_{LINE1}$ (NG)-Sm and R-TALE $_{LINE1}$-Lg pair showed similar luminescence intensities on C and 5mC oligo DNAs, consistent with the similar K_d values of L-TALE $_{LINE1}$ (NG)-Sm to C and 5mC oligo DNAs obtained from EMSA (Fig. 2C, S2B). These data suggested that the luminescence intensity reflected the TALE-DNA binding and that the low luminescence intensity of the L-TALE$_{LINE1}$(HD)-Sm and R-TALE $_{LINE1}$-Lg pair on 5mC oligo DNA was due to the low preference of RVD "HD" for 5mC. Interestingly, the luminescence intensity of the L-TALE $_{LINE1}$ (HD)-Sm and R-TALE $_{LINE1}$-Lg pair on C oligo DNA was over 100-fold higher than that on 5mC oligo DNA. The difference in the luminescence intensity was unexpectedly large considering that L-TALE$_{LINE1}$(HD)-Sm showed only a five-fold lower K_d value for the C target than for 5mC. Although the mechanism is unclear, it is possible that L-TALE$_{LINE1}$(HD) dissociates from 5mC oligo DNA too rapidly to lead to complementation of NanoBiT fragments.

The effect of the spacer length within two binding sites of TALEs was evaluated. We designed five oligo DNA fragments, which had the binding sites of L-TALE$_{LINE1}$ and R-TALE$_{LINE1}$ separated by different lengths of the spacer (Fig. S4A), and TALE-NanoBiT

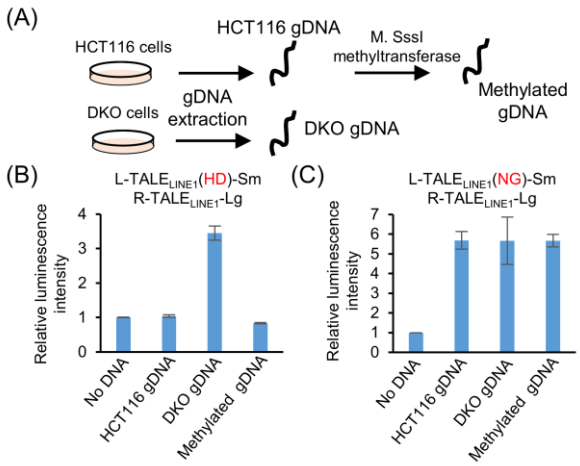

Fig. 3 Methylation detection of gDNA. (A) Schematic representation of the preparation of gDNAs containing different methylation states. Luminescence intensities of the mixture containing indicated gDNA (45 ng), R-TALE$_{LINE1}$-Lg (10 nM) and L-TALE$_{LINE1}$(HD)-Sm (10 nM) (B) or L-TALE$_{LINE1}$(NG)-Sm (10 nM) (C). The luminescence intensity was normalized to that of no DNA condition. Data are expressed as means ± SD (n = 3).

fragment complementation on these oligo DNAs was evaluated (Fig. S4B, S4C). The luminescence intensity decreased as the spacer length became longer (10-19 bp), although the reduction was within only three-fold. However, oligo DNA with a 22 bp spacer induced higher luminescence intensity than with 16 and 19 bp spacers. These results suggest that the orientation of TALEs also affects the complementation efficiency. The importance of the spacer length and the orientation of TALEs is common to other TALE technologies, such as BiFC-TALE and TALE nuclease (TALEN).[7, 13]

To evaluate the universality of our system, we designed another L-TALE, L-TALE$_{art}$, targeting an artificial sequence containing 4 CpG sites (Fig. S5). L-TALE$_{art}$(HD)-Sm and R-TALE$_{LINE1}$-Lg were successfully complemented on the target DNA in a manner dependent on the CpG methylation of the target sequence, indicating the applicability of our TALE-NanoBiT system to other DNA sequences. In addition, in this context, L-TALE$_{art}$(NG)-Sm showed 14-fold lower K_d value for the 5mC target than C, and induced stronger luminescence on 5mC oligo DNA than C (Fig. S5). These results suggested that our method can adopt other RVDs apart from RVD "HD" as a base-discriminating RVD.

The methylation states of gDNA were clearly detected using TALE-NanoBiT system (Fig. 3). We extracted gDNA from a colon cancer cell line, HCT116 cells (Fig. 3A). A portion of the gDNA was enzymatically methylated using M.SssI CpG methyltransferase to prepare fully methylated gDNA. Meanwhile, highly unmethylated gDNA was extracted from DNA (cytosine-5)-methyltransferase (DNMT) 1/DNMT3B double-knockout HCT116 (DKO) cells.[14] The LINE1 methylation levels of these gDNAs were determined by combined bisulfite restriction analysis (COBRA) and bisulfite sequencing.[3a, 3e] In COBRA, HCT116 gDNA had 80.4% LINE1 methylation and 88.1% after treatment with M.SssI methyltransferase, whereas DKO gDNA had 4.0% LINE1 methylation (Fig. S6). Similar results were obtained using bisulfite sequencing (Fig. S7). To evaluate NanoBiT fragment complementation on these gDNAs that had different methylation states, L-TALE$_{LINE1}$(HD)-Sm, R-TALE$_{LINE1}$-Lg, and each gDNA were mixed, and luminescence intensity was measured. Compared to the no DNA condition, a significant increase in luminescence intensity was observed only on the gDNA from DKO cells that was highly unmethylated (Fig. 3B, S8A). On the other hand, when using L-TALE$_{LINE1}$(NG)-Sm instead of L-TALE$_{LINE1}$(HD)-Sm, an increase in luminescence intensity was observed on all three types of gDNA (Fig. 3C, S8B). These results demonstrate that L-TALE$_{LINE1}$(HD)-Sm and R-TALE$_{LINE1}$-Lg detected the differences in the methylation states of gDNAs.

Furthermore, the DNA methylation state was successfully detected even in live cells (Fig. 4). We expressed the L-TALE$_{LINE1}$(HD)-Sm and R-TALE$_{LINE1}$-Lg pair in HCT116 or DKO cells (Fig. 4A). Twenty-four hours after the transfection, the culture medium was changed to reduced serum medium, furimazine was added, and the luminescence intensity of these cells was measured. We set the L-TALE$_{LINE1}$(NG)-Sm and R-TALE$_{LINE1}$-Lg pair as a standard in each cell line because the pair showed similar increases in luminescence in the presence of either methylated or unmethylated gDNA (Fig. 3C). By comparing the luminescence of the L-TALE$_{LINE1}$(HD)-Sm and R-TALE$_{LINE1}$-Lg pair to that of the control pair, we evaluated the methylation-discriminative ability of the L-TALE$_{LINE1}$(HD)-Sm and R-TALE$_{LINE1}$-Lg pair. In HCT116 cells that had highly methylated gDNA, the L-TALE$_{LINE1}$(HD)-Sm and R-TALE$_{LINE1}$-Lg pair showed a significantly lower luminescence intensity than the L-TALE$_{LINE1}$(NG)-Sm and R-TALE$_{LINE1}$-Lg pair (Fig. 4B, S9A). The difference in luminescence was not due to differences in transfection efficiency because the expression levels of both pairs were similar (Fig. S10A). On the other hand, in DKO cells that had highly unmethylated gDNA, both pairs showed similar luminescence intensities (Fig. 4C, S9B). The expression levels of both pairs were also similar in DKO cells **(**Fig. S10B). These results suggested that in DKO cells, L-TALE$_{LINE1}$(HD)-Sm bound to gDNA as much as L-TALE$_{LINE1}$(NG)-Sm, while in HCT116 cells, the DNA binding of L-TALE$_{LINE1}$(HD)-Sm was less than that of L-TALE$_{LINE1}$(NG)-Sm. The fact that L-TALE$_{LINE1}$(NG)-Sm bound to both methylated and unmethylated gDNA indicated that L-TALE$_{LINE1}$(HD)-Sm preferentially bound to unmethylated DKO gDNA. Thus, the applicability of the TALE-NanoBiT system to detect the methylation states of gDNA in live cells was demonstrated.

In conclusion, we established the TALE-NanoBiT system to convert TALE-DNA binding to an easily detectable luminescence signal. In addition, using a methylation-sensitive TALE, the methylation states of the target sequence could be distinguished. Importantly, the TALE-NanoBiT system worked well even in live cells, demonstrating a new way of detecting the methylation states of gDNA over time.

To date, the sequence-specific detection of 5mC has not been achieved with existing methods using the MBD. In this study, we detected the methylation states of the LINE1 region as a model target. Because of the programmable sequence specificity and methylation sensitivity of TALE, TALE-NanoBiT allows detection of the methylation states of other desirable specific sites. Although the protein design of the TALE-NanoBiT system could be further optimized, this study demonstrated the usability of TALE-NanoBiT for sequence-specific detection of the methylation states of gDNA in live cells. Recently reported TALE scaffolds with enhanced 5mC selectivity may also help improve the 5mC-discriminative ability of TALE-NanoBiT.[15]

Apart from 5mC, there are many types of 5mC derivatives, including 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosines (5caC).[16] Although it was suggested that

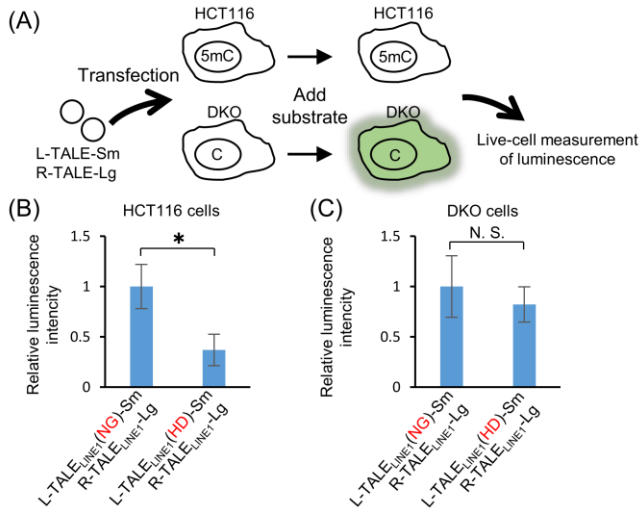

Fig. 4 Live-cell detection of LINE1 methylation. (A) Schematic representation of the experiment. Luminescence intensities of (B) HCT116 cells or (C) DKO cells. The luminescence intensity was normalized to that of L-TALE$_{LINE1}$(NG)-Sm and R-TALE$_{LINE1}$-Lg pair-transfected cells. Data are expressed as the mean ± SD (n = 3). * $P < 0.05$ (Student's *t*-test).

these 5mC derivatives have independent roles as epigenetic markers,[17] the details have not been well studied because of the lack of methods for detecting these modified nuclear bases in live cells. TALE RVDs with sensitivity to 5hmC and 5caC have also been reported.[6c, 6d, 6f] Using these RVDs, our strategy to detect epigenetic changes based on the TALE-NanoBiT system can be applied to other such modified nuclear bases. To the best of our knowledge, TALE is the only protein with a modified base sensitivity with designable sequence specificity.

Because detection of the changes in the 5mC status can be used to determine changes in biological and tumorigenic processes, our method for detecting 5mC in live cells should be of great clinical and research value.

We thank Takashi Yamamoto (Hiroshima University) for the plasmids used to construct the TALEs and Hiromu Suzuki (Sapporo Medical University) for the HCT116 and DKO cells. This work was supported by JST/CREST (JPMJCR14W3) and JSPS KAKENHI (16H03281) (M. I.).

Conflicts of interest

There are no conflicts to declare.

References

1 J. A. Law and S. E. Jacobsen, *Nat. Rev. Genet*., 2010, 11, 204.

2 (a) Z. D. Smith and A. Meissner, *Nat. Rev. Genet*., 2013, 14, 204; (b) S. E. Brown, M. F. Fraga, I. C. G. Weaver, M. Berdasco and M. Szyf, *Epigenetics*, 2007, 2, 54; (c) G. Egger, G. Liang, A. Aparicio and P. A. Jones, *Nature*, 2004, 429, 457; (d) Y. Bergman and H. Cedar, *Nat. Struct. Mol. Biol*., 2013, 20, 274.

3 (a) S. J. Cokus, S. Feng, X. Zhang, Z. Chen, B. Merriman, C. D. Haudenschild, S. Pradhan, S. F. Nelson, M. Pellegrini and S. E. Jacobsen, *Nature*, 2008, 452, 215; (b) F. Santos, B. Hendrich, W. Reik and W. Dean, *Dev. Biol.*, 2002, 241, 172; (c) S. Kobayakawa, K. Miike, M. Nakao and K. Abe, *Genes Cells*, 2007, 12, 447; (d) C. Desjobert, M. E. Maï, T. Gérard-Hirne, D. Guianvarc'h, A. Carrier, C. Pottier, P. B. Arimondo and J. Riond, *Epigenetics*, 2015, 10, 82; (e) T. Pobsook, K. Subbalekha, P. Sannikorn and A. Mutirangura, *Clin. Chim. Acta,* 2011, 412, 314.

4 (a) T. Yamazaki, K. Yamagata and T. Baba, *Dev. Biol*., 2007, 304, 409; (b) J. Ueda, K. Maehara, D. Mashiko, T. Ichinose, T. Yao, M. Hori, Y. Sato, H. Kimura, Y. Ohkawa and K. Yamagata, *Stem Cell Rep*., 2014, 2, 910; (c) R. Zhang, L. Liu, Y. Yao, F. Fei, F. Wang, Q. Yang, Y. Gui and X. Wang, *Sci. Rep*., 2017, 7, 5430; (d) Y. Hori, N. Otomura, A. Nishida, M. Nishiura, M. Umeno, I. Suetake and K. Kikuchi, J. Am. Chem. Soc., 2018, 40, 1686; (e) X. Huang, R. Narayanaswamy, K. Fenn, S. Szpakowski, C. Sasaki, J. Costa, P. Blancafort and P. M. Lizardi, *DNA Cell Biol*., 2012, Suppl 1, S-2–S-10; (f) Y. Stelzer, C. S. Shivalila, F. Soldner, S. Markoulaki and R. Jaenisch, *Cell*, 2015, 163, 218; (g) C. Lungu, S. Pinter, J. Broche, P. Rathert and A. Jeltsch, *Nat. Commun*., 2017, 8, 649.

5 (a) J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt and U. Bonas, *Science*, 2009, 326, 1509; (b) A. J. Bogdanove and D. F. Voytas, *Science*, 2011, 333, 1843; (c) T. Gaj, C. A. Gersbach and C. F. Barbas, *Trends Biotechnol*., 2013, 31, 397; (d) 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.

6 (a) G. Kubik, M. J. Schmidt, J. E. Penner and D. Summerer, *Angew. Chem., Int. Ed.*, 2014, 53, 6002; (b) G. Kubik and D. Summerer, *ChemBioChem*, 2015, 16, 228; (c) G. Kubik, S. Batke and D. Summerer, *J. Am. Chem. Soc*., 2015, 137, 2; (d) S. Maurer, M. Giess, O. Koch and D. Summerer, *ACS Chem. Biol.,* 2016, 11, 3294; (e) S. Tsuji, S. Futaki and M. Imanishi, *Chem. Commun*., 2016, 52, 14238; (f) Y. Zhang, L. Liu, S. Guo, J. Song, C. Zhu, Z. Yue, W. Wei and C. Yi, *Nat. Commun*., 2017, 8, 901.

7 H. Hu, H. Zhang, S. Wang, M. Ding, H. An, Y. Hou, X. Yang, W. Wei, Y. Sun and C. Tang, *Sci. Rep.,* 2017, 7, 40192.

8 (a) S. Cabantous, T. C. Terwilliger and G. S. Waldo, *Nat. Biotechnol*., 2005, 23, 102; (b) Y. J. Shyu and C. d. Hu, *Trends Biotechnol*., 2008, 26, 622.

9 (a) I. Remy and S. W. Michnick, *Nat. Methods*, 2006, 3, 977; (b) A. H. Badran, J. L. Furman, A. S. Ma, T. J. Comi, J. R. Porter and I. Ghosh, *Anal. Chem*., 2011, 83, 7151.

10 K. S. Yang, M. X. Ilagan, D. Piwnica-Worms and L. J. Pike, *J. Biol. Chem*., 2009, 284, 7474.

11 A. S. Dixon, M. K. Schwinn, M. P. Hall, K. Zimmerman, P. Otto, T. H. Lubben, B. L. Butler, B. F. Binkowski, T. Machleidt, T. A. Kirkland, M. G. Wood, C. T. Eggers, L. P. Encell and K. V. Wood, *ACS Chem. Biol*., 2016, 11, 400.

12 D. J. Weisenberger, M. Campan, T. I. Long, M. Kim, C. Woods, E. Fiala, M. Ehrlich and P. W. Laird, *Nucleic Acids Res*., 2005, 33, 6823.

13 J. C. Miller, S. Tan, G. Qiao, K. A. Barlow, J. Wang, D. F. Xia, X. Meng, D. E. Paschon, E. Leung, S. J. Hinkley , G. P. Dulay, K. L. Hua, I. Ankoudinova, G. J. Cost, F. D. Urnov, H. S. Zhang, M. C. Holmes, L. Zhang, P. D. Gregory and E. J. Rebar, *Nat. Biotechnol*., 2011, 29, 143.

14 I. Rhee, K. E. Bachman, B. H. Park, K. W. Jair, R. W. Yen, K. E. Schuebel, H. Cui, A. P. Feinberg, C. Lengauer, K. W. Kinzler, S. B. Baylin and B. Vogelstein, *Nature*, 2002, 416, 552.

15 P. Rathi, A. Witte and D. Summerer, *Sci. Rep*., 2017, 7, 15067.

16 (a) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind and A. Rao, Science, 2009, 324, 930; (b) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He and Y. Zhang, *Science*, 2011, 333, 1300.

17 L. Shen, C. X. Song, C. He and Y. Zhang, *Annu. Rev. Biochem*., 2014, 83, 585.