nature biotechnology

Targeted DNA demethylation *in vivo* using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions

Sumiyo Morita[1], Hirofumi Noguchi[2,5], Takuro Horii[1,5], Kazuhiko Nakabayashi[3], Mika Kimura[1], Kohji Okamura[4], Atsuhiko Sakai[2], Hideyuki Nakashima[2], Kenichiro Hata[3], Kinichi Nakashima[2] & Izuho Hatada[1]

Despite the importance of DNA methylation in health and disease, technologies to readily manipulate methylation of specific sequences for functional analysis and therapeutic purposes are lacking. Here we adapt the previously described dCas9–SunTag for efficient, targeted demethylation of specific DNA loci. The original SunTag consists of ten copies of the GCN4 peptide separated by 5-amino-acid linkers. To achieve efficient recruitment of an anti-GCN4 scFv fused to the ten-eleven (TET) 1 hydroxylase, which induces demethylation, we changed the linker length to 22 amino acids. The system attains demethylation efficiencies >50% in seven out of nine loci tested. Four of these seven loci showed demethylation of >90%. We demonstrate targeted demethylation of CpGs in regulatory regions and demethylation-dependent 1.7- to 50-fold upregulation of associated genes both in cell culture (embryonic stem cells, cancer cell lines, primary neural precursor cells) and *in vivo* in mouse fetuses.

In mammalian genomes, ~70% of cytosine residues in the sequence 5′-CpG-3′ are methylated[1]. DNA methylation regulates many biological processes[2,3], and deregulation of DNA methylation has been implicated in the etiology of several diseases such as cancer and imprinting diseases[4]. Methylation of cytosines is catalyzed by DNA methyltransferases, whereas the TET family of proteins catalyzes oxidation of methylated cytosine to 5-hydroxymethylcytosine, the initial step of the demethylation pathway[5]. Although DNA methylation is thought to play important roles in several processes, in many cases its causative effects are unclear because of a lack of widely applicable techniques for adding and removing DNA methylation at specific loci. In principle, such technologies could find application in targeted epigenetic therapy. Nonspecific methods for erasure of methylation by inhibitors of DNA methyltransferases such as 5-aza-2′ -deoxycytidine have been commonly used to study the effects of demethylation on specific gene promoters[6]. However, as these reagents demethylate genomes globally, it is difficult to study the effect of specific DNA methylation events, and there is the risk of side effects in therapeutic use.

Recently, genome editing technologies, such as zinc-finger nucleases (ZFNs)[7], transcription activator-like effector nucleases (TALENs)[8], and CRISPR–Cas9 (refs. 9–11) have been adapted to recruit catalytic activity to specific loci by fusion to catalytically inactive endonucleases. Zinc fingers (ZFs)[12] and transcription activator-like effectors (TALEs)[13] fused to TET family fusion proteins were reported to hydroxylate specific loci to activate demethylation in cultured cells, but the extent of demethylation was limited in this system. In addition, in ZFN and TALEN systems, the design and protein engineering of endonucleases are required for each locus, which is time-consuming. On the other hand, CRISPR RNA-guided Cas9 nucleases use small base-pairing guide RNAs (gRNAs) to target and cleave foreign DNA elements in a sequence-specific manner. Therefore, only alteration of the target sequence in small gRNAs is required to generate new endonucleases for new loci in the CRISPR–Cas9 system. Here, we show that a catalytically inactive Cas9 (dCas9) fused to the catalytic domain of TET1 (TET1CD) hydroxylates specific loci and activates site-specific demethylation. TET1 is a dioxygenase that catalyzes the hydroxylation of 5-methylcytosine to 5-hydroxymethylcytosine and the subsequent generation of 5-formylcytosine and 5-carboxylcytosine. These modified bases are either diluted by replication or removed by thymine DNA glycosylase and base-excision repair. This activity is greatly enhanced by fusion of dCas9 to a peptide repeat sequence to recruit multiple copies of an antibody-fused TET1 hydroxylase. We also show that our system is applicable to *in vivo* manipulation of methylation of specific loci in mouse fetuses.

First, we used a simple design to manipulate methylation, a direct fusion protein of an inactive Cas9 nuclease (dCas9) and TET1. TET1 has a conserved catalytic domain at the C terminus, which has a higher catalytic activity than the full-length protein[14]. Therefore, we fused TET1CD to the catalytically inactive dCas9 (system 1, **Fig. 1a** and **Supplementary Fig. 1**). A cytosine residue within the STAT3-binding site upstream of the gene (*Gfap*) encoding the astrocyte-specific marker glial fibrillary acidic protein (GFAP) was used as a target[15]. This site is methylated in most cell types, except for astrocytes, and demethylation of this site plays a critical role in the differentiation of neural precursor cells (NPCs) into astrocytes. We designed three targets around the STAT3-binding site (**Fig. 1b**), generated gRNA vectors for them, and transiently introduced these gRNA vectors into embryonic stem cells (ESCs) with the dCas9-TET1CD fusion

[1]Institute for Molecular and Cellular Regulation, Gunma University, Maebashi, Japan. [2]Department of Stem Cell Biology and Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [3]Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, Tokyo, Japan. [4]Department of Systems BioMedicine, National Research Institute for Child Health and Development, Tokyo, Japan. [5]These authors contributed equally to this work. Correspondence should be addressed to I.H. (hatada@gunma-u.ac.jp).

Received 23 September 2015; accepted 28 July 2016; published online 29 August 2016; doi:10.1038/nbt.3658

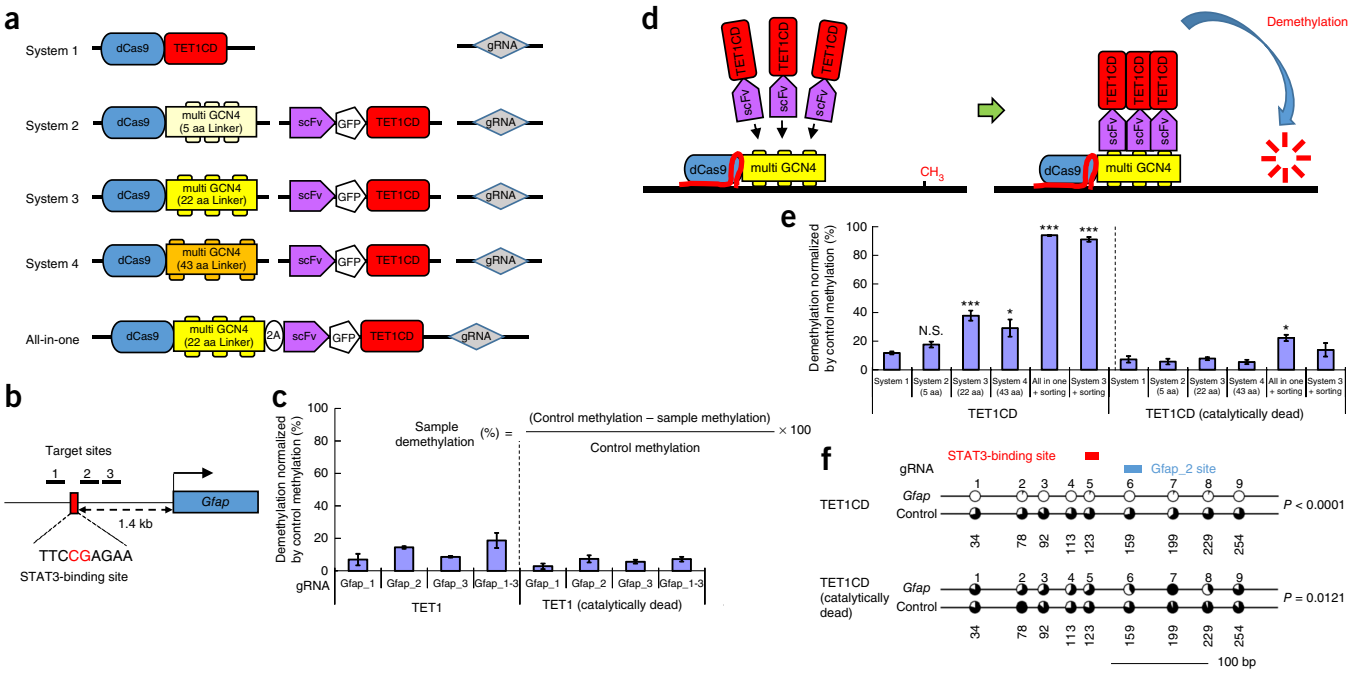

Figure 1 Targeted demethylation in ESCs using CRISPR–Cas9 and a peptide-repeat-based amplification. (**a**) Structure of vector components used in the experiments. (**b**) The mouse *Gfap* locus. STAT3-binding site is indicated by a red box; it has a methylation-sensitive CpG site (red characters). The locations of the targets (1–3) for gRNAs (Gfap_1–3) are indicated by black bars. (**c**) Demethylation activities of dCas9 directly bound to TET1CD (system 1) using gRNA targets in **b** (Gfap_1–3), and of dCas9 directly bound to catalytically dead TET1. Methylation of the STAT3-binding site was analyzed by COBRA. Error bars, mean ± s.e.m. ($n = 2$ from two independent experiments). (**d**) A scheme for CRISPR–Cas9 and a peptide-repeat-based amplification of demethylation. dCas9 fused to a peptide repeat can recruit multiple copies of antibody (scFv)-fused TET1CD. Thus, multiple copies of TET1CD can demethylate the target more efficiently. (**e**) Demethylation activities of CRISPR–Cas9 and peptide-repeat-based amplification systems with various linker lengths (systems 2–4) compared to the system that does not use amplification (system 1), for both active and catalytically dead TET1. + Sorting, ESCs transfected with the all-in-one vector or system 3 and sorted by FACS to select GFP-expressing cells. Demethylation of the STAT3-binding site was analyzed by COBRA as in **c**. The gRNA used was target 2 of *Gfap*. Error bars, mean ± s.e.m. ($n = 3$ from two independent experiments). N.S., not significant; $^{*}P < 0.05$; $^{***}P < 0.005$ versus system 1 (two-sided Student's *t*-test). (**f**) Methylation for both active and catalytically dead TET1 surrounding the Gfap target site in ESCs transfected with *Gfap* gRNA (Gfap_2) or a control using system 3 and expressing GFP. Black portion of circles indicates the percentage of methylation in each CpG site. Each number beneath the circles indicates the position. Scale indicates distance in bp. For each group, at least 14 randomly selected clones were sequenced and analyzed. The statistical significance of the difference between the two groups of the entire set of CpG sites was evaluated with the Mann-Whitney *U*-test (the Wilcoxon rank-sum test).

protein expression vector (pCAG-dCas9TET1CD). Methylation of the STAT3-binding site was analyzed by combined bisulfite restriction analysis (COBRA)[16]. In each assay, methylation in cells transfected with the control vector (empty gRNA vector) was defined as 0% demethylation (100% methylation) and demethylation in each sample was normalized by the control. These three gRNAs, Gfap_1, Gfap_2, and Gfap_3 (**Fig. 1c**), showed 7%, 14%, and 9% demethylation at the STAT3-binding site, respectively (**Fig. 1c**). On the other hand, dCas9 with catalytically dead TET1CD showed lower levels of demethylation, and unrelated gRNAs (UR_1, UR_2, and UR_3) did not show any demethylation (**Fig. 1c**, data not shown). Thus, this simple system induces gRNA-dependent specific demethylation; however, the extent of demethylation is at most 14% and significant improvement was not observed even when the three gRNAs were used simultaneously (**Fig. 1c**).

Next, we attempted to amplify the demethylation activity using dCas9 fused to a SunTag[17] to recruit multiple copies of antibody-fused TET1CD (**Fig. 1d**). SunTag is a tandem repeat of ten copies of the 19 amino-acid GCN4 peptide separated by amino acid linkers of 5 amino acid residues. It was initially used to amplify gene expression by fusing it to dCas9 to recruit multiple copies of the herpes virus transcriptional activation domain (VP64) fused to a single-chain variable fragment (scFv) of the anti-GCN4 antibody[17]. We applied this system to demethylate the STAT3-binding site of *Gfap* using an expression vector for Gfap_2 gRNA, dCas9 with ten copies of the GCN4 peptide, and an anti-GCN4 peptide antibody (scFv)-superfolder green fluorescent protein (sfGFP)[18]-TET1CD fusion protein in ESCs (system 2, **Fig. 1a** and **Supplementary Fig. 2**).

As the extent of demethylation was not improved using the original SunTag (system 2) (**Fig. 1e**), we optimized the length of the linker separating each 19-amino-acid (aa) GCN4 peptide unit of the array fused to dCas9. We tested linkers with 5 aa (system 2, **Fig. 1a** and **Supplementary Fig. 2**), 22 aa (system 3, **Fig. 1a** and **Supplementary Fig. 3**), and 43 aa (system 4, **Fig. 1a** and **Supplementary Fig. 4**), and compared the demethylation activity. Because of the technical limitations of synthetic gene technology, we reduced the copy number of the GCN4 peptide in the vectors with a linker of 22 and 43 aa to 5 and 4, respectively. Demethylation was best with a linker of 22 aa, reaching as high as 38%, although the copy number of the GCN4 peptide was reduced, second highest with a linker of 43 aa, and worst with a linker of 5 aa (**Fig. 1e**). The improvement in performance with longer linkers may be explained by the reduction in steric hindrance between the scFV-GFP-TET1CD fusion proteins that retain catalytic activity (**Supplementary Fig. 5a**, possibility 2), rather than an increase in peptide array binding (**Supplementary Fig. 5a**, possibility 1), because the Cas9 peptide array of system 2 bound amounts of antibody-sfGFP-

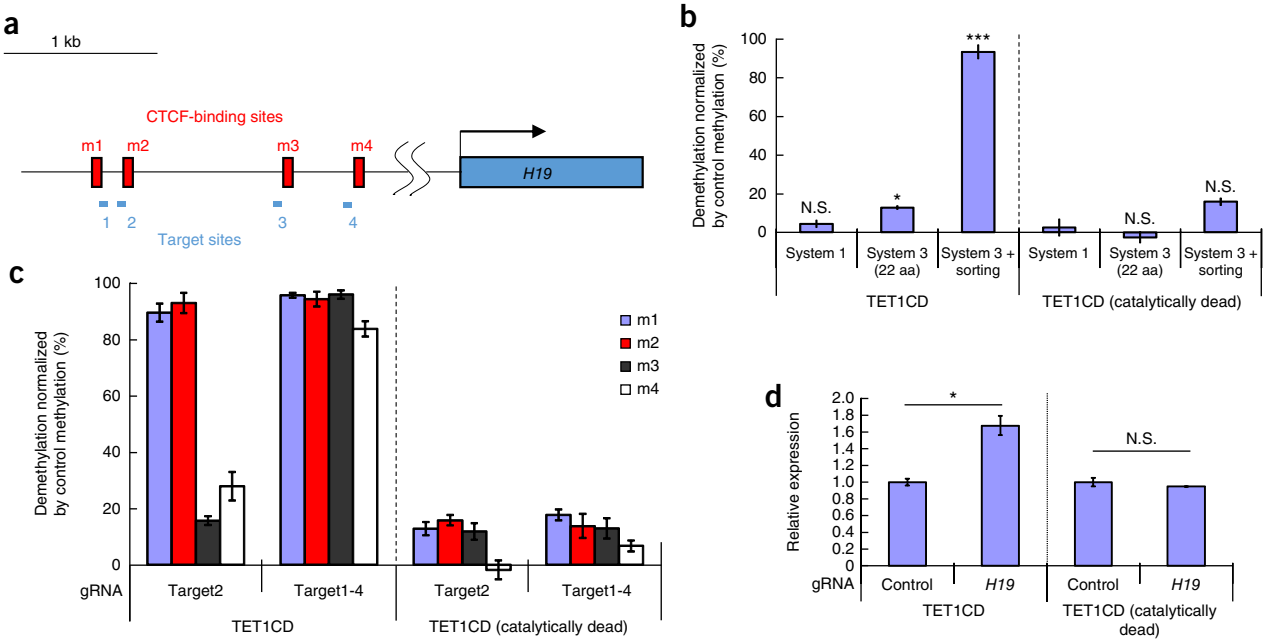

Figure 2 Targeted demethylation of the imprinted gene *H19* in ESCs using CRISPR–Cas9 and a peptide-repeat-based amplification. (**a**) The mouse *H19* locus is shown, with CTCF-binding sites indicated by red boxes. These CTCF-binding sites have methylation-sensitive CpG sites. Location of the targets (H19DMR_1–4) used for gRNAs are indicated by blue bars. Scale indicates distance in kb. (**b**) Demethylation of the m2 CTCF-binding site of *H19* using system 1, system 3, and system 3 with sorting. Demethylation activities for both active and catalytically dead TET1 are shown. Demethylation was analyzed as in **Figure 1c**. Error bars, mean ± s.e.m. ($n = 3$ from two independent experiments). $^*P < 0.05$; $^{***}P < 0.005$ versus system 1 (two-sided Student's *t*-test). (**c**) Demethylation of CTCF-binding sites (m1–m4) of *H19* using system 3 (with sorting) with active or catalytically dead TET1. Demethylation was analyzed as in **Figure 1c**. Data are shown as the mean ± s.e.m. ($n = 3$ from two independent experiments). (**d**) Quantitative PCR analysis of *H19*. Data are shown as the mean ± s.e.m. ($n = 3$ from two independent experiments). N.S., not significant; $^*P < 0.05$ (two-sided Student's *t*-test).

TET1CD that were at least comparable to the amounts of fusion protein bound by the other systems (**Supplementary Fig. 5b**).

To rule out the possibility that demethylation is due to overexpression of TET1CD, we performed experiments using only dCas9-GCN4 or scFV-GFP-TET1CD, in addition to an experiment using catalytically dead TET1CD (H1671Y and D1673A mutations) (**Fig. 1e** and **Supplementary Fig. 5c**). In contrast to the strong demethylation in the presence of both dCas9-GCN4 and scFV-GFP-TET1CD (**Supplementary Fig. 5c**, 1), no demethylation was observed in the presence of only scFV-GFP-TET1CD (**Supplementary Fig. 5c**, 4; $P < 0.005$), indicating that demethylation is not due to overexpression of TET1CD. In the experiments in which only dCas9-GCN4 was expressed (**Supplementary Fig. 5c**, 3) and in which dCas9-GCN4 and catalytically dead TET1CD were expressed (**Supplementary Fig. 5c**, 2), slight demethylation was observed. However, the extent of demethylation was considerably lower than the strong demethylation in the presence of both dCas9-GCN4 and scFV-GFP-TET1 ($P < 0.005$). This slight demethylation can probably be explained by the physical hindrance of surrounding DNA from the maintenance methylation system (Dnmt1).

In conclusion, the expression of both dCas9-GCN4 and scFV-GFP-TET1 is required for effective demethylation of the system. The low solubility of scFV could lead to protein aggregation, resulting in increased cell death[17]. Therefore, we measured the viability of ESCs into which systems 2, 3, and 4 had been introduced by propidium iodide (PI) staining and did not find any differences between the three systems (**Supplementary Fig. 6**).

To further improve the demethylation efficiency, we used fluorescence-activated cell sorting (FACS) to select GFP-expressing transfected cells. For this purpose, we generated an all-in-one vector including the gRNA, dCas9 with the GCN4 array of system 3, and the antibody-sfGFP-TET1CD fusion protein (**Fig. 1a** and **Supplementary Fig. 7**). Demethylation was almost complete in GFP-sorted all-in-one vector-transfected ESCs (**Fig. 1e**). GFP-sorted ESCs transfected with three separate vectors also showed almost complete demethylation (**Fig. 1e**). Cells incorporating one vector harboring DNA encoding GFP may be competent for the incorporation of other vectors. Thus, we achieved complete demethylation of the targeted locus by enhancing the sensitivity of the read-out by sorting cells.

We validated the range of demethylation at the target site in a sorted sample by bisulfite sequencing[19]. We found that demethylation occurred at sites at least 200 bp from the target (**Fig. 1f**, **Supplementary Figs. 8** and **Fig. 9c**) and also confirmed that systems 3 and 4 are better than system 2. We also validated the off-target activity in the same sample by bisulfite sequencing of off-target sites of the Gfap_2 gRNA. No significant off-target activity was observed (**Supplementary Fig. 9a**). In addition, we performed whole-genome bisulfite sequencing and RNA sequencing (RNA-seq) analyses. Demethylation of the *Gfap* locus was confirmed (**Supplementary Fig. 9b**,**c**) and low off-target activity was also confirmed (**Supplementary Fig. 9b**, 1 vs. 2), judging from the variation of data (a vs. b). The difference in methylation between active and catalytically dead TET1 was also small (1 vs. 3), indicating low non-specific TET1 activity. RNA-seq analysis confirmed these results (**Supplementary Fig. 9d**).

To generalize the results, we applied our technology to another locus, the differential methylation region (DMR) of the paternally methylated imprinted gene *H19*. There are four methylation-sensitive CTCF-binding sites (m1–m4)[20], which are important for regulation of *H19* imprinting, in the DMR (**Fig. 2a**). We introduced a gRNA targeting m2 (H19DMR_2) into ESCs with dCas9-TET1CD (**Fig. 1a**), system 3 (**Fig. 1a**), or system 3 followed by cell sorting.

We observed a notable improvement of demethylation using system 3 compared to dCas9-TET1CD, and complete demethylation at the m2 site in GFP-sorted cells (**Fig. 2b**). GFP-sorted cells were further analyzed for methylation in the surrounding area. The m1 site, which is 200 bp from the target, showed complete demethylation (**Fig. 2c** and **Supplementary Fig. 10**). On the other hand, the m3 and m4 sites, which are located more than 1 kb from the target, were only partially demethylated, suggesting that effective demethylation does not extend more than 1 kb. To test the possibility of multiple targeting, we designed and introduced gRNAs for m1–m4 with system 3 (H19DMR_1–4). In GFP-sorted cells, all four sites (m1–m4) were nearly completely demethylated (**Fig. 2c**), indicating that multiple sites can be demethylated using multiple gRNAs. In addition, no significant off-target activity was observed in *H19* by bisulfite sequencing of off-target sites of gRNA for m1-m4 (**Supplementary Fig. 11**).

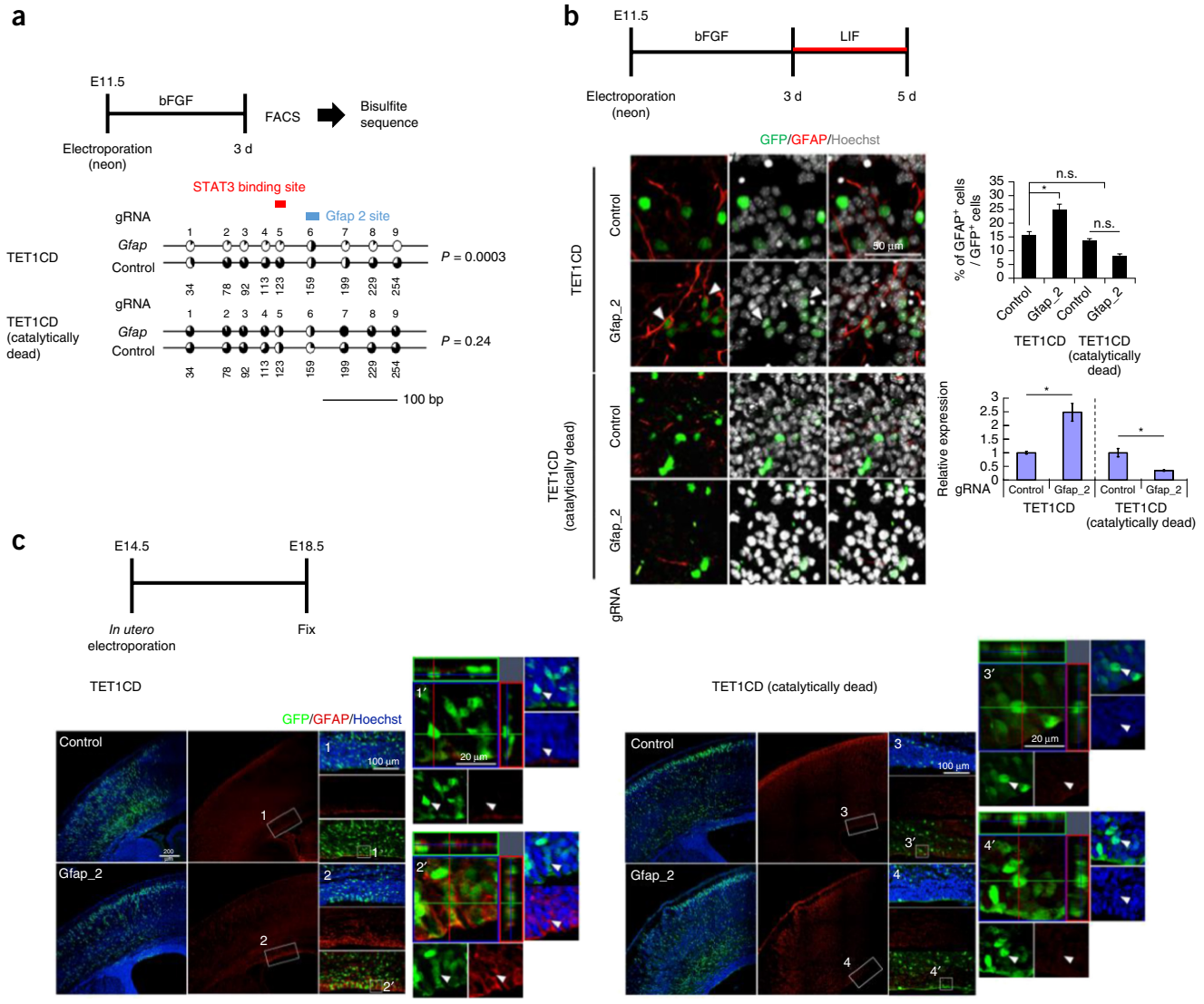

Figure 3 *In vitro* and *in vivo* targeted demethylation using CRISPR–Cas9 and a peptide-repeat-based amplification. (**a**) Methylation surrounding the *Gfap* target site in NPCs expressing GFP. Black in circles indicates the percentage of methylation in each CpG site. Each number beneath the circles indicates the position. Scale indicates distance in bp. For each group, at least 14 randomly selected clones were sequenced and analyzed. The statistical significance of the difference between the two groups of the entire set of CpG sites was evaluated with the Mann-Whitney *U*-test (also called the Wilcoxon rank-sum test). (**b**) Differentiation ability of methylation-edited cells with active or catalytically dead TET1. The all-in-one vector targeting *Gfap* (Gfap_2) or control was introduced into isolated NPCs, which were subsequently differentiated using LIF. After induction, the percentage of GFAP-positive cells among GFP-positive cells was significantly increased compared to the control (left). GFAP (red) and GFP (green) were mainly stained in processes and nuclei, respectively. DNA was stained with Hoechst (gray). To assess the astrocyte differentiation, over 300 GFP-positive cells from one sample ($n = 3$ per group) were counted (right top). Experiments were independently replicated at least three times. Error bars, mean ± s.e.m. $^*P < 0.05$ (ANOVA with Tukey's *post-hoc* test). Scale bar, 50 μm. Quantitative PCR analysis of NPCs 6 h after induction of differentiation (right bottom). n.s., not significant; $^*P < 0.05$ (two-sided Student's *t*-test). (**c**) E18.5 brain sections that were electroporated at E14.5 with the all-in-one vector harboring either active or catalytically dead TET1 and the *Gfap*-targeting gRNA or the control vector. Green, red, and blue indicate GFP, GFAP, and Hoechst, respectively. Magnified images of the boxed areas indicated (ventricular zone, VZ) are also shown in 1, 2, 3, and 4, and confocal orthogonal images of the boxed areas indicated in 1, 2, 3, and 4 are shown in 1′, 2′, 3′, and 4′, respectively. The white arrowheads in 1′, 2′, 3′, and 4′ indicate representative electroporated cells. Scale bars, 1–4(100 μm), 1′–4′(20 μm).

We examined the expression of the *H19* gene. *H19* is an imprinted gene and one of the alleles is already expressed. Therefore, complete demethylation should result in, at most, a twofold increase in expression. We observed a 1.7-fold increase in *H19* expression (**Fig. 2d**).

In addition, we applied our technology to the *RHOXF2B*, *CARD9*, *SH3BP2*, and *CNKSR1* genes in human cells, and found significant ($P < 0.005$) demethylation and induction of expression for these genes (**Supplementary Fig. 12a**,**b**).

We next applied our technology *in vivo* in brain of mouse fetuses. Before direct application to the brain, we tested isolated NPCs. The all-in-one vector with a gRNA targeting the STAT3-binding site of the GFAP-encoding gene (Gfap_2) was introduced into NPCs isolated from mouse embryos by electroporation. GFP-positive cells were isolated by FACS, and methylation around the STAT3-binding site was analyzed. Induction of active TET1, but not catalytically dead TET1, in the all-in-one system decreased methylation of the STAT3-binding site and its surrounding area in the *Gfap* promoter (**Fig. 3a**) and increased *Gfap* expression (**Fig. 3b**). To examine the differentiation ability of methylation-edited cells, electroporated NPCs were differentiated using leukemia inhibitory factor (LIF), a well-known astrocyte-inducing cytokine[15]. After induction, the percentage of cells positive for the astrocyte-specific marker GFAP was significantly ($P < 0.05$) increased compared to when the control gRNA or catalytically dead TET1 was used (**Fig. 3b**).

Next, we introduced the all-in-one vector into the ventricular zone (VZ) of mouse fetal brain at E14 by *in utero* electroporation[21] and samples were fixed at E18. To determine whether this targeted manipulation induces expression of GFAP, recovered embryonic brains were stained for GFAP. In GFP-positive cells in the VZ of brain transduced with the all-in-one vector and the gRNA targeting the STAT3-binding site of the GFAP-encoding gene (Gfap_2), expression of GFAP was detected in all embryonic brains examined (**Fig. 3c** and **Supplementary Fig. 13a**). By contrast, in the control, in which the all-in-one vector with control gRNA or catalytically dead TET1 was introduced, the GFAP signal was not induced in GFP-positive cells (**Fig. 3c** and **Supplementary Fig. 13a**). We also examined a gRNA (Gfap_0) located further away from the STAT3-binding site than Gfap_2 and also found a similar induction of expression *in vivo* (**Supplementary Fig. 13a**,**b**) and demethylation (**Supplementary Fig. 13c**).

A experiment using a luciferase reporter construct in adult rat hippocampus cells showed that the promoter activity of *Gfap* was affected by Gfap_2, but not by Gfap_0 (**Supplementary Fig. 14**). This suggests that the bulky Cas9-effector at the locus could access the STAT3-binding site *in vitro* when Gfap_2 was used as a gRNA, but not when Gfap_0 was used. Nevertheless, in *in vivo* experiments, the Gfap_2 gRNA effectively induced demethylation and expression but not when catalytically dead TET1 was used. Thus, our method can be used for *in vivo* applications. Future studies will be needed to determine whether similar interventions could have therapeutic applications.

In summary, we describe a method for targeted demethylation of endogenous genes. Seven of nine gRNAs showed demethylation of >50% and four of these seven gRNAs showed demethylation of >90% (**Figs. 1e**, **2c** and **Supplementary Fig. 12b**), whereas a previously published TALE system shows variable demethylation of 15–84% only in a small fraction of targets[13]. There are two explanations for this. Our system recruits multiple copies of TET1CD, whereas the TALE system recruits only one. Therefore, the 5-methylcytosine hydroxylase activity recruited is higher in our system. The other explanation is a difference in methylation sensitivity. Cas9 can cleave methylated DNA[22], whereas TALEs cannot[23]. Compared to the TALE method[13], our system showed long-range demethylation. There was >90% demethylation of CpG sites 100 bp from the target sites (**Figs. 1f** and **2c**), whereas the highest degree of demethylation is only within 30 bp using the TALE system[13]. This difference could also be due to the recruitment of multiple copies of TET1CD. The ZF system in combination with TET2 shows, at most, 10% demethylation[12]. ZFNs seem not to be sensitive to methylation[24]; therefore, the difference could be explained by the multiple recruitment of TET1CD in our system or differences between TET1 and TET2. Recently, a system, dCas9 with gRNA inserting two copies of bacteriophage MS2 RNA elements, which could recruit multiple MS2 coat proteins fused to TET1CD has been reported for the demethylation of specific genes. However, this system only induced minimal demethylation at loci tested[25]. Finally, although there are some previous reports regarding *in vivo* manipulation of the epigenome[26–29], to our knowledge CRISPR-based manipulation of the epigenome *in vivo* has not been reported previously. The successful *in vivo* application suggests that the investigation of potential therapeutic uses is warranted.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Sequence Read Archive: BS-seq (SRP075309), RNA-seq (SRP075305). Accession numbers: pCAG-dCas9TET1CD (LC169507), pCAG-dCas9-10xGCN4_v4 (LC169508), pCAG-dCas9-5xPlat2AflD (LC169509), pCAG-dCas9-3.5xSuper (LC169510), pCAG-scFvGCN4sfGFPTET1CD (LC169511), pPlatTET-gRNA2 (LC169512).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by the Basic Science and Platform Technology Program for Innovative Biological Medicine from the Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT); The Japan Agency for Medical Research and Development (AMED) to I.H. and AMED-CREST and AMED to K.N.; and a Grant-in-Aid for Challenging Exploratory Research (grant number 15K14452) to K.N. The authors plan to make the reagents widely available to the academic community through Addgene (http://www.addgene.org/?gclid=CKvf88_a2ccCFQNwvAodSbUGiQ). We appreciate the technical assistance provided by The Research Support Center, Research Center for Human Disease Modeling, Kyushu University Graduate School of Medical Sciences.

AUTHOR CONTRIBUTIONS

I.H. conceived the project. I.H., S.M., T.H., K.H., and K.N. designed the experiments. S.M., H.N., T.H., K.N., M.K., K.O., and A.S. performed the experiments and analyzed the data. I.H., S.M., H.N., and K.N. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Razin, A. *et al.* Variations in DNA methylation during mouse cell differentiation in vivo and in vitro. *Proc. Natl. Acad. Sci. USA* **81**, 2275–2279 (1984).
2. Reik, W. & Dean, W. DNA methylation and mammalian epigenetics. *Electrophoresis* **22**, 2838–2843 (2001).
3. Bird, A.P. & Wolffe, A.P. Methylation-induced repression--belts, braces, and chromatin. *Cell* **99**, 451–454 (1999).
4. Feinberg, A.P. & Tycko, B. The history of cancer epigenetics. *Nat. Rev. Cancer* **4**, 143–153 (2004).
5. Carey, N., Marques, C.J. & Reik, W. DNA demethylases: a new epigenetic frontier in drug discovery. *Drug Discov. Today* **16**, 683–690 (2011).
6. Oki, Y. & Issa, J.P. Review: recent clinical trials in epigenetic therapy. *Rev. Recent Clin. Trials* **1**, 169–182 (2006).
7. Porteus, M.H. & Carroll, D. Gene targeting using zinc finger nucleases. *Nat. Biotechnol.* **23**, 967–973 (2005).
8. Miller, J.C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–148 (2011).

9. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
10. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
11. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
12. Chen, H. *et al.* Induced DNA demethylation by targeting Ten-Eleven Translocation 2 to the human ICAM-1 promoter. *Nucleic Acids Res.* **42**, 1563–1574 (2014).
13. Maeder, M.L. *et al.* Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat. Biotechnol.* **31**, 1137–1142 (2013).
14. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
15. Takizawa, T. *et al.* DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain. *Dev. Cell* **1**, 749–758 (2001).
16. Xiong, Z. & Laird, P.W. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* **25**, 2532–2534 (1997).
17. Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S. & Vale, R.D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
18. Pédelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. & Waldo, G.S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
19. Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* **89**, 1827–1831 (1992).
20. Bell, A.C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482–485 (2000).
21. Tabata, H. & Nakajima, K. Neurons tend to stop migration and differentiate along the cortical internal plexiform zones in the Reelin signal-deficient mice. *J. Neurosci. Res.* **69**, 723–730 (2002).
22. Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
23. Bultmann, S. *et al.* Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res.* **40**, 5368–5377 (2012).
24. Chen, S. *et al.* A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucleic Acids Res.* **41**, 2769–2778 (2013).
25. Xu, X. *et al.* A CRISPR-based approach for targeted DNA demethylation. *Cell Discov.* **2**, 16009 (2016).
26. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).
27. Heller, E.A. *et al.* Locus-specific epigenetic remodeling controls addiction- and depression-related behaviors. *Nat. Neurosci.* **17**, 1720–1727 (2014).
28. Stolzenburg, S. *et al.* Stable oncogenic silencing in vivo by programmable and targeted de novo DNA methylation in breast cancer. *Oncogene* **34**, 5427–5435 (2015).
29. Li, K. *et al.* Manipulation of prostate cancer metastasis by locus-specific modification of the CRMP4 promoter region using chimeric TALE DNA methyltransferase and demethylase. *Oncotarget* **6**, 10030–10044 (2015).

ONLINE METHODS

Ethics statement. All animal experiments were approved by the Animal Care and Experimentation Committee of Gunma University. All mice were treated according to Fundamental Guidelines for Proper Conduct of Animal Experiment and Related Activities in Academic Research Institutions under the jurisdiction of the Ministry of Education, Culture, Sports, Science and Technology of Japan. All efforts were made to minimize animal suffering and to reduce the number of animals used.

Animals. All (10- to 11-week-old females) pregnant mice (ICR background) were obtained from SLC Inc. (Shizuoka, Japan).

Construction of plasmids for targeted demethylation. The dCas9-TET1CD fusion protein expression vector (pCAG-dCas9TET1CD) was generated by fusing cDNA encoding the catalytically inactive nuclease codon-optimized *S. pyogenes* Cas9 (dCas9) to the N-terminal of human TET1CD in an expression vector using the CAG promoter. The dCas9 fragment was amplified from Addgene plasmid 48240 by PCR. The TET1CD fragment was amplified from human cDNA by PCR. The dCas9 fragment of systems 1–4 was amplified from Addgene plasmid 60903 by PCR. The aa sequence of GCN4 used was EELLSKNYHLENEVARLKK. The sequences of the linker between each GCN4 peptide unit were GSGSG (systems 2), GSGSGGSGSGSGGSGSGGSGSG (system 3), and GSGSGGSGSGG SGSGGSGSGGSGGSGSGGSGSGGSGSGGSGSG (system 4). The GFP fragment was amplified from Addgene plasmid 60904. The scFv fragment was also amplified from Addgene plasmid 60904. All the fusion proteins were expressed under the control of the CAG promoter. The all-in-one vector was generated by fusing components 1 and 2 of system 3 using 2A peptide (GSGATNFSLLKQAGDVEENPGP). The vector sequences are described in **Supplementary Figures 1–4**, **7**. The authors plan to make the plasmids widely available to the academic community through Addgene (http://www.addgene.org/?gclid=CKvf88_a2ccCFQN-wvAodSbUGiQ).

Construction of gRNAs. The gRNA vectors for *Gfap* and *H19* were generated by inserting the target sequences into Addgene plasmid 41824. Cloning was performed by linearization of an *Afl* II site and Gibson assembly-mediated incorporation of the gRNA insert fragment. The target sequences are described in **Supplementary Sequences**.

Cell culture. ESCs were made from mouse (C57BL/6J) blastocysts and were validated by chimera formation. ESCs were cultured at 37 °C under 5% CO_2 in Dulbecco's Modified Eagle's Medium (DMEM)-high glucose (D6429-500ML, Sigma) supplemented with 1% Fetal Bovine Serum (FBS), 17.5% KSR100 (10828028, Gibco), 0.2% 2-mercaptoethanol (21985-023, Gibco), and 1×10^3 units/mL ESGRO mouse LIF (ESG1107, Millipore). A549 (RIKEN BRC) cells were cultured in Minimum Essential Medium with 10% FBS, non-essential amino acids. 1–87 (Cell Resource Center for Biomedical Research, Tohoku University) cells were cultured in RPMI1640 with 10% FBS. HEK293T (RIKEN BRC) cells were cultured in DMEM with 10% FBS. Cells were transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocols, harvested 48 h later, and used directly for assays or first sorted using FACSAriaII (BD Biosciences). The molar ratio of the dCas9-peptide array fusion vector, scFv-GFP-TET1CD vector, and gRNA vector in the transfection was 1:2:4, respectively. Cell lines have not been tested routinely for *Mycoplasma* contamination.

DNA methylation analysis. Genomic DNA was treated with the Epitect Plus DNA Bisulfite Kit (QIAGEN) according to the manufacturer's instruction. The modified DNA was amplified with the PCR primers described in **Supplementary Sequences**. The percentages of demethylation at the STAT3-binding site of *Gfap* and at the m1–m4 sites of *H19* were determined by COBRA. Briefly, amplified fragments were cleaved with restriction enzymes (described in **Supplementary Sequences**) whose recognition sites were located in these sites and subjected to PAGE. Methylation was calculated as the percentage of cleaved DNA by densitometric analysis of gels stained with ethidium bromide. In each assay, methylation in cells transfected with the control vector (empty gRNA vector) was defined as 100% methylation (0% demethylation) and demethylation of each sample was normalized by the control. Sample demethylation (%) = (control methylation − sample methylation)/control methylation × 100. For methylation analysis of the surrounding area and off-targets, bisulfite sequencing analysis was performed. Briefly, amplified fragments were ligated into the TOPO vector (Invitrogen) and at least 14 clones were sequenced. Sequences were analyzed by the methylation analysis tool QUantification tool for Methylation Analysis (QUMA). The statistical significance of the difference between two groups of the entire set of CpG sites was evaluated with the Mann-Whitney *U*-test (also called the Wilcoxon rank-sum test).

Quantitative RT-PCR analysis. Total RNA was prepared from isolated tissues using the Allprep DNA/RNA mini kit (QIAGEN). Gene expression levels were measured with LightCycler 480 (Roche) using SYBR Premix Ex Taq (Takara) according to the manufacturer's instructions. Expression levels were normalized against the level of 18S ribosomal RNA (ESCs) or actin (HEK293 cells). Primer sequences are described in **Supplementary Sequences**.

Primary cell culture. E11.5 mouse forebrains were dissected and triturated in calcium- and magnesium-free Hanks' balanced salts solution (Sigma, 369 H2387). The cells were electroporated with 2.5 μg of the all-in-one vector (expressing gRNA for the control or the *Gfap* locus) prepared in 10 μl of R buffer using the Neon Transfection System (Invitrogen, 1,200 V with one 40 ms pulse). Electroporated cells were cultured on a poly-ornithine/fibronectin-coated 6 cm dish or an 8-well chamber slide in N2-supplemented DMEM/F-12 (Invitrogen, 11320-033) containing 10 ng/ml basic fibroblast growth factor (PeproTech, 100-18B) for 3 d under 5% CO_2 at 37 °C. The cells cultured on 6 cm dishes were sorted using a FACS Aria flow cytometer (BD Bioscience) to isolate GFP-positive cells. The cells cultured on 8-well chamber slides were further cultured with 50 ng/ml LIF (Millipore, ESG1107) for 2 d to induce astrocyte differentiation.

***In utero* electroporation.** *In utero* electroporation was performed at E14.5 as previously described. DNA was dissolved in phosphate-buffered saline (PBS; pH 7.4) to a concentration of 1.5 μg/μl with 0.05% Fast Green to monitor the injection. Pregnant mice were deeply anesthetized, and the uterine horns were exposed and placed on a sterile gauze. Then, 1–2 μl of DNA solution was injected into the lateral ventricle of the embryonic brain through the uterine with a glass micropipette. The uterus was kept wet by dropping saline (prewarmed at 37 °C) between the electrodes (CUY650P5). Thereafter, 50 ms electronic pulses of 45 V were charged five times at 950 ms intervals using a square-pulse electroporator (CUY21EDIT; Nepa Gene Company). The uterine horns were placed back in the abdominal cavity, and the abdominal wall and skin were sewed with surgical sutures. Embryos were harvested at E15.5 for bisulfite sequencing or E18.5 for immunohistochemistry.

Tissue preparation. E18.5 embryos were deeply anesthetized with ice before being perfused with 4% paraformaldehyde (PFA) prepared in PBS. Brains were dissected and fixed with 4% PFA prepared in PBS overnight at 4 °C. For cryoprotection, fixed brains were stored in 15% sucrose prepared in PBS overnight at 4 °C and then in 30% sucrose prepared in PBS overnight at 4 °C. The brain was embedded in optimal cutting temperature (OCT) compound (Tissue Tek, Sakura Finetek, 25608-930) and frozen at −80 °C for cryosectioning. Frozen brains were serially sectioned in the coronal plane at a thickness of 20 μm using a Leica CM 1900 cryostat (Leica Microsystems). Cryostat sections were then affixed to Matsunami adhesive-coated glass slides (Matsunami Glass, S9441) and subjected to immunohistochemistry. For bisulfite sequencing, E15.5 mouse forebrains were dissected and digested in S-MEM (Gibco) containing 0.1% papain (Sigma), and triturated with 60 mg/ml DNase I (Sigma) and 10% FBS. Dissociated cells were then sorted using a FACS Aria flow cytometer to isolate GFP-positive cells and subjected to bisulfite treatment.

Immunohistochemistry and immunostaining. Cryosections were washed with PBS, blocked for 1 h at room temperature with blocking solution (3% FBS, 0.5% Triton X-100, and 0.25% Tween-20), and incubated overnight at 4 °C with

doi:10.1038/nbt.3658

primary antibodies diluted in blocking solution. The following primary antibodies were used: rabbit anti-GFAP (1:500; Sigma, G9269) and chick anti-GFP (1:500; Aves Labs, GFP-1020). After three washes in PBS, sections were incubated for 2 h with the corresponding secondary antibodies as follows: CF488 donkey anti-chick IgY (IgG) (H+L), highly cross-adsorbed (1:500; Biotium, 20166); and CF543 donkey anti-rabbit IgG (H+L), highly cross-adsorbed (1:500; Biotium, 20038). Hoechst 33258 (1:500; Nacalai Tesque) was used for nuclear staining. For immunostaining, cells were fixed with 4% PFA prepared in PBS for 20 min, incubated for 3 h at 4 °C with primary antibodies diluted in blocking solution, and then incubated with the corresponding secondary antibodies. After a final rinse with PBS, samples were mounted on glass slides with Immu-Mount (Thermo Scientific, 9990412). Immunofluorescence images of cultured cells were taken with a fluorescence microscope (Leica AF600) and a confocal microscope (LSM700, Carl Zeiss), and combined for analysis using Adobe Photoshop Elements 10. For quantification of tissue sections, cortical sections at the same anatomical level were analyzed, and confocal images were taken with a confocal microscope (LSM700, Carl Zeiss). Stitching of 20 × objective images was performed using Zen 2011 software (blue edition, Carl Zeiss) to cover electroporated regions in the cortex of each coronal section and combined for analysis using Adobe Photoshop Elements 10. To assess astrocyte differentiation, at least 300 GFP-positive cells were counted per sample ($n = 3$ per group for the *in vitro* assay and 4–6 brains per group for *in vivo* analysis). GFAP-positive cells among GFP-positive cells were counted in high-magnification images, and each GFAP-positive cell was identified by GFAP staining around the nucleus, as indicated by both GFP and Hoechst. All experiments were independently replicated at least three times.

Quantification of cell viability. Two days after transfection, ESCs were resuspended in PBS (-) containing PI (2 μg/ml) and then incubated for 5 min at room temperature in the dark. The cells were analyzed using a FACSVerse flow cytometer (BD Biosciences) with a 488-nm blue laser.

Co-immunoprecipitation. ESCs were transfected using systems 2–4 with Gfap_2 gRNA, harvested, and lysed in lysis buffer (25 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% NP-40, and 5% glycerol). The cell lysate was centrifuged at 13,000*g* for 10 min at 4 °C. The supernatant was incubated with anti-HA magnetic beads (Thermo Scientific) for 30 min at room temperature. The beads were washed three times with TBST buffer (25 mM Tris-HCl pH 7.4, 0.15 M NaCl, and 0.05% Tween 20) and once with water. The HA-tagged GCN4 peptide array was eluted with sodium dodecyl sulfate (SDS) sample buffer (60 mM Tris-HCl, 1% SDS, 10% glycerol, and 50 mM dithiothreitol) at 95 °C for 10 min. Eluents were subjected to western blot analysis with an anti-GFP polyclonal antibody (MBL, cat. #598) or an anti-HA tag (C29F4) rabbit monoclonal antibody (Cell Signaling Technology, cat. #3724).

Whole-genome bisulfite sequencing. Whole-genome bisulfite sequencing libraries were prepared using the post-bisulfite adaptor-tagging (PBAT) protocol[32]. Genomic DNA samples (100 ng of each spiked with 0.5 ng of lambda phage DNA (Takara)) were subjected to bisulfite treatment using an EZ DNA Methylation Gold Kit (Zymo Research) followed by conversion to double-stranded DNA using Klenow fragments (3′–5′ exo(-); New England BioLabs) using BioPEA2N4 (5′-biotin-ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN-3′) (first strand). The biotinylated strand was captured using Dynabeads M-280 Streptavidin (Invitrogen), followed by conversion to double-stranded DNA using the Klenow fragment (3′→5′ exo(-)), with PE-reverse-N4 (5′-CAAGCAGAAGACGGCATACGAGATNNNN-3′) (second strand). After removing the biotinylated first strand DNA, the second strand was subjected to primer extension with primer-3 (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′) using Phusion Hot Start High-Fidelity DNA Polymerase (Finnzymes) to obtain double-stranded DNA products. The concentrations of the resultant PBAT libraries were measured by quantitative PCR using PE-forward and PE-reverse primers (Illumina). To obtain a sufficient amount of libraries for subsequent sequencing, the libraries were subjected to six cycles of PCR amplification using PE-forward and PE-reverse primers, followed by DNA purification using Agencourt AMPure XP (Beckman Coulter). The PBAT libraries were sequenced on a HiSeq2500 system (Illumina). Cluster generation was performed using the HiSeq PE Cluster Kit v4-cBot HS kit. Single read (225 bp) or paired-end read (125 bp × 2) sequencing was performed using the HiSeq SBS v4-HS kit. Real-time analysis and base calling were performed using HiSeq Control Software Version 2.2.68 (Illumina). At least 400 million reads were obtained for each sample. The conversion of bcl files to fastq files was done using bcl-2fast2 version 1.8.3 software (Illumina). After removing sequences from random primers, the remaining sequences were aligned to the reference mouse genome (mm10) using Bismark v.0.13.1 (http://www.bioinformatics.babraham.ac.uk/projects/bismark/). The resultant BAM files were sorted and converted to SAM files using Samtools (http://samtools.sourceforge.net/). The SAM files were analyzed by methylKit (http://code.google.com/p/methylkit), an R package for DNA methylation analysis, to determine the methylation level of individual CpG sites.

RNA sequencing. RNA quality was assessed with an Agilent RNA 6000 Nano Kit (Agilent, cat. #5067-1511) on an Agilent 2100 Bioanalyzer. RNA was prepared using the TruSeq Stranded mRNA LT Sample Prep Kit-Set A (Illumina, cat. #RS-122-2101). Each library was prepared from 750 ng of total RNA. The cDNA libraries were quantified using the KAPA Library Quantification kit (KAPA Biosynthesis, cat. #KK4835), and their quality and size were checked using the Agilent High Sensitivity DNA Kit (Agilent, cat. #5067-4626). The samples were loaded onto a cBot (Illumina) for cluster generation in a flow cell using the HiSeq PE Cluster Kit v4-cBot HS kit (PE-402-4002, Illumina), and the flow cell was subjected to sequencing using a HiSeq 2500 system (Illumina). A paired-end run of 125 bp × 2 was performed using the HiSeq SBS v4-HS kit (FC-401-4003, Illumina). Real-time analysis and base calling were performed using HiSeq Control Software Version 2.2.68 (Illumina). At least 69 million read pairs were obtained for each sample. The conversion of bcl files to fastq files was performed using bcl2fast2 version 1.8.3 software (Illumina). Alignment of obtained sequences to the mouse genome (mm10) was performed using TopHat2 (http://ccb.jhu.edu/software/tophat/index.shtml). The resultant BAM files were analyzed by Cufflinks 2.2.1 (http://cole-trapnell-lab.github.io/cufflinks/), which assembles transcripts, estimates their abundances, and tests for differential expression among samples. The estimated expression levels of genes were calculated as the numbers of fragments per kilobase of transcript per million mapped reads (FPKM). The expression levels of 10045 RefSeq genes with the minimum FPKM value among four samples (ESsort1–4) higher than 0.125 were used for scatter plot and correlation analyses.

Luciferase reporter assay. For the reporter assay of the *Gfap* promoter, a neural progenitor cell line from adult rat hippocampus (HCN cells) was used as previously described[30]. HCN cells were co-transfected with the all-in-one vector (expressing gRNA for the control or the *Gfap* locus) and a GFAP promoter-reporter plasmid, which expresses firefly luciferase under the regulation of the 2.6 kb *Gfap* promoter (GF1L)[30]. As an internal control, the sea pansy luciferase-expressing vector under the control of the human elongation factor-1a promoter was also co-transfected[31]. One day after transfection, cells were stimulated with LIF (50 ng/ml) for 8 h and lysed with Reporter Lysis Buffer (Promega). The luciferase activity of the lysates was measured with the Dual-Glo Luciferase Assay System (Promega) according to the manufacturer's protocol. A Wallac 1420 ARVO/Light (PerkinElmer Life and Analytical Science) luminometer was used for detection. Firefly luciferase activities were determined by three independent transfections and normalized by comparison with Renilla luciferase activities as the internal control.

Reproducibility and statistical analysis. At least three mice were analyzed per group. Statistical analyses were performed using either a two-sided Student's *t*-test for comparisons between two groups or a one-way ANOVA with Tukey's multiple comparison test for comparisons among multiple groups. The Mann-Whitney *U*-test was used to compare DNA methylation levels among samples in bisulfite sequencing analysis. No sample size estimates were performed. No blinding was used.

Selection of off-target sites for bisulfite sequencing analysis. A flowchart for the selection of sites for off-target analysis is presented in **Supplementary Figure 15**. We searched for off-targets using a web tool called CRISPR direct

(http://crispr.dbcls.jp/). By using this tool, the 12 bases in the 3′ region of the target sequence adjacent to the PAM were searched against the genome because this region contains critical residues determining target specificity. Next, we removed the sites unsuitable for analysis (sequences containing repeats and those giving no PCR primers by Meth Primer (http://www.urogene.org/methprimer/) in the default condition except for the product size). As for the off-target analysis of *Gfap*, all these sites were subjected to off-target analysis. As for the off-targets of *H19*, we selected all the off-targets in which more than 16 of 20 bases match. If there were fewer than three selected sites, sites of lower homology were selected.

All information regarding the off-targets is presented in **Supplementary Sequences**.

30. Hsieh, J., Nakashima, K., Kuwabara, T., Mejia, E. & Gage, F.H. Histone deacetylase inhibition-mediated neuronal differentiation of multipotent adult neural progenitor cells. *Proc. Natl. Acad. Sci. USA* **101**, 16659–16664 (2004).
31. Nakashima, K. *et al*. Synergistic signaling in fetal brain by STAT3-Smad1 complex bridged by p300. *Science* **284**, 479–482 (1999).
32. Miura, F., Enomoto, Y., Dairiki, R. & Ito, T. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. Nucleic Acids Res. 40, e136 (2012).

doi:10.1038/nbt.3658