

## RESEARCH

# Genome-wide determination of on-target and off-target characteristics for RNA-guided DNA methylation by dCas9 methyltransferases

Lin Lin<sup>1,2,†</sup>, Yong Liu<sup>1,†</sup>, Fengping Xu<sup>3,4,5,†</sup>, Jinrong Huang<sup>3,4,5,†</sup>, Tina Fuglsang Daugaard<sup>1</sup>, Trine Skov Petersen<sup>1</sup>, Bettina Hansen<sup>1</sup>, Lingfei Ye<sup>3</sup>, Qing Zhou<sup>3,4</sup>, Fang Fang<sup>3,4</sup>, Ling Yang<sup>3,4</sup>, Shengting Li<sup>1,3</sup>, Lasse Fløe<sup>1</sup>, Kristopher Torp Jensen<sup>1</sup>, Ellen Shrock<sup>6</sup>, Fang Chen<sup>3,4,5</sup>, Huanming Yang<sup>3,4,7</sup>, Jian Wang<sup>3,4</sup>, Xin Liu<sup>3,4</sup>, Xun Xu<sup>3,4,\*</sup>, Lars Bolund<sup>1,3,4,2,8,9</sup>, Anders Lade Nielsen<sup>1</sup> and Yonglun Luo<sup>1,3,4,2,8,9,10,\*</sup>

<sup>1</sup>Department of Biomedicine, Aarhus University, Aarhus, Denmark, <sup>2</sup>Danish Regenerative Engineering Alliance for Medicine, Department of Biomedicine, Aarhus University, Aarhus, Denmark, <sup>3</sup>BGI-Shenzhen, Shenzhen 518083, China, <sup>4</sup>China National GeneBank-Shenzhen, BGI-Research, Shenzhen 518083, China, <sup>5</sup>Department of Biology, University of Copenhagen, Copenhagen, Denmark, <sup>6</sup>Department of Genetics, Harvard Medical School, Boston, MA, USA, <sup>7</sup>James D. Watson Institute of Genome Sciences, Hangzhou 310058, China, <sup>8</sup>BGI-Qingdao, 2877 Tuanjie Road, Sino-German Ecopark, Qingdao, 266000, China, <sup>9</sup>Lars Bolund Institute of Regenerative Medicine, BGI-Qingdao, China and <sup>10</sup>BrainStem - Stem Cell Center of Excellence in Neurology, Copenhagen, Denmark

\*Correspondence address. Xun Xu. China National GeneBank-Shenzhen, BGI-Research, Shenzhen 518083, China. E-mail: [xuxun@genomics.cn](mailto:xuxun@genomics.cn) and Yonglun Luo. Department of Biomedicine, Aarhus university, Batholins alle 6, DK-8000, Aarhus C, Denmark. E-mail: [alun@biomed.au.dk](mailto:alun@biomed.au.dk)

†These authors contributed equally to the study.

## Abstract

**Background:** Fusion of DNA methyltransferase domains to the nuclease-deficient clustered regularly interspaced short palindromic repeat (CRISPR) associated protein 9 (dCas9) has been used for epigenome editing, but the specificities of these dCas9 methyltransferases have not been fully investigated. **Findings:** We generated CRISPR-guided DNA methyltransferases by fusing the catalytic domain of DNMT3A or DNMT3B to the C terminus of the dCas9 protein from *Streptococcus pyogenes* and validated its on-target and global off-target characteristics. Using targeted quantitative bisulfite pyrosequencing, we prove that dCas9-BFP-DNMT3A and dCas9-BFP-DNMT3B can efficiently methylate the CpG dinucleotides flanking its target sites at different genomic loci (*uPA* and *TGFBR3*) in human embryonic kidney cells (HEK293T). Furthermore, we conducted whole genome bisulfite sequencing (WGBS) to address the specificity of our dCas9 methyltransferases. WGBS revealed that

Received: 24 February 2017; Revised: 29 June 2017; Accepted: 12 February 2018

© The Author(s) 2018. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

although dCas9-BFP-DNMT3A and dCas9-BFP-DNMT3B did not cause global methylation changes, a substantial number (more than 1000) of the off-target differentially methylated regions (DMRs) were identified. The off-target DMRs, which were hypermethylated in cells expressing dCas9 methyltransferase and guide RNAs, were predominantly found in promoter regions, 5' untranslated regions, CpG islands, and DNase I hypersensitivity sites, whereas unexpected hypomethylated off-target DMRs were significantly enriched in repeated sequences. Through chromatin immunoprecipitation with massive parallel DNA sequencing analysis, we further revealed that these off-target DMRs were weakly correlated with dCas9 off-target binding sites. Using quantitative polymerase chain reaction, RNA sequencing, and fluorescence reporter cells, we also found that dCas9-BFP-DNMT3A and dCas9-BFP-DNMT3B can mediate transient inhibition of gene expression, which might be caused by dCas9-mediated *de novo* DNA methylation as well as interference with transcription. **Conclusions:** Our results prove that dCas9 methyltransferases cause efficient RNA-guided methylation of specific endogenous CpGs. However, there is significant off-target methylation indicating that further improvements of the specificity of CRISPR-dCas9 based DNA methylation modifiers are required.

**Keywords:** DNA methylation; CRISPR; Cas9; DNMT3A; DNMT3B; dCas9; specificity; off-targets; epigenome editing

## Background

Owing to its simplicity, efficiency, and potential for multiplicity, the type II clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 (Cas9) with engineered variants have been widely used for genome and epigenome editing in many species [1–5]. The Cas9 protein is guided to a specific genomic locus containing a protospacer adjacent motif (PAM) by a small single guide RNA (gRNA), which contains a conserved scaffold sequence and a programmable guide sequence (typically 20 nt) for base pairing with the target [1]. By introducing double mutations (D10A and H840A) in the *Streptococcus pyogenes* Cas9 protein (dCas9) to inactivate its catalytic activity and fusing functional effectors to the C terminus of the dCas9, the applications of CRISPR/Cas9 are expanded to regulation of gene expression (CRISPRa and CRISPRi) [6–8], targeted DNA purification [9], visualization of specific gene regions [10], and acetylation or methylation of chromatin components [11, 12].

Genome-wide studies have revealed fundamental functional roles of DNA methylation as well as associations between aberrant DNA methylation and human diseases including cancer [13, 14]. Methylation of cytosine residues (5mC) in the mammalian genome mainly occurs at CpG dinucleotides. In promoter regions, CpG methylation is normally associated with repression of gene expression. Currently, insights into DNA methylation-associated biological processes are largely based on correlative data. Methods have been developed to methylate desired gene loci selectively by fusing programmable DNA binding proteins (zinc finger proteins [ZFs] or transcription-activator-like effectors [TALEs]) to DNA methyltransferases [3–9]. However, the laborious generation of ZFs and TALEs hampers their broader applications. Engineered dCas9 has been harnessed for targeted DNA methylation by fusing dCas9 to the catalytic domain of mammalian DNA methyltransferases, thus providing an alternative tool for more easily programmable DNA methylation [15, 16].

Currently, genome-wide characterization of the specificity of dCas9-based epigenetic modifiers is lacking. To gain more insights into the efficiency and specificity of targeted DNA methylation by CRISPR gRNA-guided dCas9 methyltransferases, we used quantitative bisulfite pyrosequencing, whole genome bisulfite sequencing, and ChIP-seq to investigate the characteristics of dCas9 methyltransferase-mediated DNA methylation in human cells.

## Methods

### Cell culture

Human embryonic kidney HEK293T cells (American Type Culture Collection) were cultured in Dulbecco's modified Eagle's medium (DMEM, Life Technologies), 10% fetal bovine serum (Sigma), 1% penicillin-streptomycin (Sigma), and 1X GlutaMAX (Life Technologies) at 37°C, 5% CO<sub>2</sub>.

### dCas9 methyltransferases plasmids

The dCas9 coding sequence was derived from pHR-SFFV-dCas9-BFP-KRAB (Addgene ID 46911) (a gift from Stanley Qi and Jonathan Weissman). The catalytic domains of DNMT1, DNMT3A, and DNMT3B were polymerase chain reaction (PCR)-amplified from pcDNA3/Myc-DNMT1 (Addgene ID 36939), pcDNA3/Myc-DNMT3A (Addgene ID 35521), and pcDNA3/Myc-DNMT3B1 (Addgene ID 35522) (a gift from Arthur Riggs), respectively. The DNMT3A (E752A) and DNMT3B (E697A) catalytically inactivating mutations were introduced by site-directed mutagenesis. All plasmids described in this study have been validated by Sanger sequencing and will be publically available through Addgene [17] (Supplementary Table S1).

### CRISPR gRNA design

Based on the observation that dCas9 methyltransferases could efficiently methylate the CpGs flanking the target sites, a web-based gRNA designing tool (dCas9 methyltransferases gRNA finder, [18]) was developed to facilitate dCas9 methyltransferase-based gRNA design. All updates regarding the dCas9 methyltransferase protocol are available on the web site [19]. All gRNA sequences are listed in Supplementary Table S1.

### Transfection and enrichment transfected cells

Unless stated elsewhere, cells were transfected with gRNAs (total 500 ng) and a dCas9 methyltransferase expression vector (500 ng) in 6-well plates using X-tremeGENE 9 DNA transfection reagent (Roche). For single gRNA or pUC19 control transfections, the amount of plasmid added was equivalent to the total amount of plasmid added for multiple gRNA transfections. For BFP-based enrichment, cells were harvested 48 hours after transfection, and dCas9 methyltransferase-expressing cells were sorted by fluorescence activated cell sorting (FACS). Briefly,

transfected cells were harvested by trypsinization, washed twice with 2% fetal bovine serum – phosphate-buffered saline (FBS-PBS), and resuspended in 500  $\mu$ L 2% FBS-PBS. Cells were stained with propidium iodide (PI) before sorting. PI-negative and BFP-positive or -negative cells were sorted with a 4 Laser BD FACS Aria III instrument. All transfections were performed in at least 2 independent experiments.

### Quantitative PCR

Total RNA was extracted from cells with the RNeasy Plus Mini kit (Qiagen, 74136) according to the manufacturer's instructions and quantified using a Nanodrop 1000 spectrophotometer. The first strand cDNA was synthesized from 100–500 ng total RNA with the iScript cDNA synthesis kit (Bio-Rad, 170–8891) following the manufacturer's instructions. Quantitative PCR (qPCR) was performed in triplicate for each sample, using the Light Cycler 480 SYBR Green I Master mix (Roche Life Science, 04887352001) and a Light Cycler 480 qPCR machine. Each qPCR reaction contained 1  $\mu$ L cDNA template (5 times diluted), 7.5  $\mu$ L qPCR Master mix (2X), and 5 pmol of each qPCR primer in a total volume of 15  $\mu$ L. The following qPCR program was used for *uPA*, *TGFBR3*, and *GAPDH*: 1 cycle at 95°C for 5 minutes; 45 cycles at 95°C for 10 seconds, 57°C for 10 seconds, and 72°C for 10 seconds during which the fluorescence signal was measured. The final product was subjected to melting curve analysis. Primers for qPCR are listed in Supplementary Table S1. Relative gene expression was calculated using the  $2^{-\Delta\Delta CT}$  method by first normalizing to the internal control *GAPDH* ( $\Delta CT$ ) and then calibrating to the transfection control *pUC19* ( $\Delta\Delta CT$ ) [20].

### DNA methylation analysis by bisulfite pyrosequencing with PyroMark Q24

Genomic DNA was extracted using the DNeasy Blood & Tissue kit (Qiagen, 69506) according to the manufacturer's instructions. A total of 200 ng of genomic DNA was bisulfite treated using the EpiTect Bisulfite kit (Qiagen, 59104) according to the manufacturer's instructions. This converts unmethylated cytosines to uracils. The bisulfite converted DNA was eluted with 20  $\mu$ L elution buffer provided in the kit. Bisulfite PCR reactions for all genes described in this study were performed in a 25  $\mu$ L volume containing 0.15  $\mu$ L Hotstar Taq polymerase (5 U/ $\mu$ L) (New England Biolabs, M0495L), 2.5  $\mu$ L 10 $\times$  Standard buffer, 0.5  $\mu$ L of 10 mM dNTPs, 1.0  $\mu$ L of each primer (10  $\mu$ M), and 1.5  $\mu$ L bisulfite converted genomic DNA. PCR was performed under the following conditions: 95°C for 5 minutes followed by 45 cycles at 94°C for 30 seconds, 58°C for 1 minute, and 72°C for 45 seconds, and, finally, at 72°C for 7 minutes. The 4  $\mu$ L PCR product was checked using gel electrophoresis. Pyrosequencing was performed with the PyroMark Q24 Advanced Reagents (Qiagen, 970922) using 20  $\mu$ L PCR product from the bisulfite treated DNA and 20  $\mu$ L sequencing primer (0.375  $\mu$ M) according to the PyroMark Q24 CpG protocol. The general degree of cytosine methylation was determined by pyrosequencing of the bisulfite converted genomic DNA using the PyroMark Q24 Advanced System (Qiagen).

### DNA methylation analysis by bisulfite Sanger sequencing

Bisulfite converted DNA was used as the template for PCR amplifications with the BS-specific PCR primers listed in Supplementary Table S1, using the DreamTaq DNA Polymerase (Life Technologies, EP0701). PCR products were gel purified, and cloned in a TA-cloning vector (Life Technologies, 450030), and

transformed into chemically competent *Escherichia coli* cells. Cell clones were manually picked, subcultured in 250  $\mu$ L LB medium overnight, lysed, subjected to Sanger sequencing, and analyzed using BISMA [21].

### Fluorescence reporter cell assay

Five stable fluorescence reporter cell clones were established by randomly inserting various copies of the CMV promoter-driven mCherry expression cassette into HEK293T (pLV-mCherry was a gift from Pantelis Tsoulfas, Addgene ID 36084). Cells were transfected separately with each dCas9 methyltransferase expression vector (50 ng) and gRNAs (total 50 ng) in 24-well plates. One-third of the transfected cells were seeded to a new plate every 2–3 days, and the remainder was used for flow cytometry analysis. Median mCherry intensity was measured with the BD LSRFortessa cell analyzer (FACS CORE facility, Aarhus University). Identical instrument settings and control beads were applied during the time course experiment to ensure valid comparison across different time points. A total of 20 000 events were recorded for each sample. Flow cytometry data were analyzed using Flowjo software.

### Immunostaining

Forty-eight hours after transfection, cells were fixed with freshly made 4% paraformaldehyde for 15 minutes at room temperature, followed by 3 washes with Dulbecco's phosphate-buffered saline (DPBS). Cells were permeabilized in 0.3% Triton X-100 DPBS for 10 minutes and blocked in 5% goat serum–DPBS for 30 minutes. Cells were incubated with a primary rabbit anti-HA-tag antibody (C29F4, Cell Signaling 3724, 1:1000) overnight, followed by secondary antibody staining with Alexa Fluor 555 donkey anti-rabbit immunoglobulin G (A-31572, Life Technologies) at room temperature for 2 hours. Images were obtained with a confocal microscope (LSM710, Carl Zeiss).

### Southern blot analysis

Genomic DNA (15  $\mu$ g) was digested with *EcoRI* restriction enzyme overnight and then analyzed using gel electrophoresis with vacuum blotting. Primers for generating the mCherry probe are listed in Supplementary Table S1. Probe labeling was performed using the Prime-It II Random Primer Labeling kit according to the manufacturer's instructions. Prehybridization and hybridization steps were carried out at 42°C. Excess probe was washed from the membrane with saline-sodium citrate buffer, and the hybridization pattern was visualized on X-ray film by autoradiography.

### RNA sequencing

Integrity and quantity of extracted RNA was evaluated with an Agilent 2100 Bioanalyzer according to the manufacturer's instructions. After DNase I treatment, mRNA was isolated with Oligo (dT) magnetic beads. Fragmentation buffer was added to generate short fragments of mRNA. cDNA was synthesized using the mRNA fragments as templates, resolved with EB buffer for end repair and ligated with adaptors. After size selection and purification by agarose gel electrophoresis, cDNA with sizes of approximately 240 bp were used for PCR amplification (12 cycles) and library construction. Libraries were sequenced on an Ion Proton platform (>30 million reads per sample). Sequencing reads that contained low-quality, adaptor, and/or short (<30 nt) read sequences were filtered out before mapping. Then, tmap



was used to align the clean reads to the hg19 UCSC RefSeq (RNA sequences, GRCh37). No more than 3 mismatches were allowed in the alignment. Gene expression levels were calculated by transforming uniquely mapped transcript reads to transcript per million [22]. Differentially expressed genes were defined as genes with a Benjamini-Hochberg-adjusted *P* value false discovery rate (FDR)  $\leq 0.001$  and fold change  $\geq 2$  compared to pUC19 control.

### ChIP-seq

HEK293T cells were transfected with dCas9 methyltransferase and % uPA gRNAs (triplicates). Forty-eight hours after transfection, transfected cells were subjected to ChIP with a commercially available kit ChIP-IT Express Enzymatic (53009-AF, ActivMotif, distributed by Nordic Biolabs) and an anti-HA tag antibody (C29F4, Cell Signaling) according to the manufacturer's instructions. Next-generation sequencing libraries were prepared for Chip and input samples. SE50 sequencing was performed on Illumina HiSeq2500. Clean reads were mapped to human genome hg19 using SOAP2 with the parameter "-p 4 -v 2 -s 35." Unique mapping reads were sampled randomly and equally (62 723 057 reads). Peaks were called using Model-based Analysis of ChIP-Seq with *P* value  $1e-3$  compared to the input samples. Common peaks found in the triplicates were selected. Furthermore, ChIP peaks located in repeat sequences and rDNA were removed. Sequence motifs enriched within 70 bp of peak summits were identified using MEME-ChIP.

### Whole genome bisulfite sequencing library preparation and sequencing

Genomic DNA was fragmented by sonication to a mean size of 250 bp using a Bioruptor (Diagenode, Belgium), followed by the blunt-ending, dA addition to 3'-end, and adaptor ligation using the TruSeq Sample Preparation kit (Illumina Inc.) according to the manufacturer's instructions. Then, bisulfite conversion was conducted with the EZ DNA Methylation-Gold kit (ZYMO). The fragments with different insert size were excised from the same lane of a 2% (Tris-Acetate-EDTA) agarose gel. Products were purified using the QIAquick Gel Extraction kit (Qiagen) and amplified by 18 PCR cycles. The library quality was monitored using the Agilent 2100 BioAnalyzer (Agilent), and the concentration of the library was determined by quantitative PCR. Finally, the whole genome bisulfite sequencing (WGBS) libraries were paired-end sequenced on Illumina HiSeq X Ten.

After filtering out adaptor and low-quality reads, 953.7 Gb 150 bp paired-end clean data were generated. An average of 106 Gb clean data were obtained for each sample. Clean reads were aligned to the human reference genome (hg19) by BSMAP (v2.74) with the parameter "-u -v 5 -z 33 -p 6 -n 0 -w 20 -s 16 -r 0 -f 10 -L 140" [23]. Only the CpG sites with read depths  $\geq 4$  were taken into consideration for DNA methylation level calculation. The 48 502 bp lambda DNA genome was used as an extra reference for calculating the bisulfite conversion rate. Nearly complete (>99%) bisulfite conversion was documented in all libraries. For the repeat WGBS experiment, HEK293T cells were transfected with pUC19 as controls or transfected with dCas9-BFP-DNMT3A and uPA gRNAs. Transfections were conducted in triplicates. Genomic DNA was purified from all cells 48 hours after transfection without BFP-based FACS enrichment of transfected cells. WGBS library construction and sequencing were conducted as above but sequenced with less depth of 10–15X coverage.

### Identification of DMRs and attempts to exclude stochastic DMRs unrelated to the dCas9 methyltransferase treatment

The bioconductor package DSS was used to identify DMRs with the parameter "delta > = 0.1, *P* value  $\leq 0.01$ , CpG sites  $\geq 3$ , DMR length  $\geq 10$  bp, smoothing window 100 bp." Since expressing a high amount of dCas9-BFP-DNMT3A and either uPA or TGFBR3 gRNAs caused the highest *de novo* on-target methylation, we reasoned that the authentic off-target DMRs should be detected in these 2 comparisons. We first compared group 1 (dCas9-BFP-DNMT3A [500 ng] + uPA gRNAs [500 ng]) or group 3 (dCas9-BFP-DNMT3A [500 ng] + TGFBR3 gRNAs [500 ng]) to group 9 (pUC19 control).

Based on the observation of (1) dose- and gRNA-dependent *de novo* methylation of uPA, TGFBR3, and GAPDH by dCas9 methyltransferases and (2) dCas9-BFP-DNMT3A being more efficient than dCas9-BFP-DNMT3B, we reasoned that the authentic DMRs caused by dCas9 methyltransferases and uPA gRNAs should have a methylation pattern as described below:

**Hypermethylated DMRs by dCas9 methyltransferases and uPA gRNAs should meet:**

% mCpG:

- (1) **group 9** (pUC19) = < **group 5** (dCas9-DNMT-3A only [500 ng]) = < **group 7** (dCas9-DNMT-3A [50 ng] + uPA gRNAs [50 ng]) = < **group 1** (dCas9-DNMT-3A [500 ng] + uPA gRNAs [500 ng]).
- (2) **group 2** (dCas9-DNMT-3B [500 ng] + uPA gRNAs [500 ng]) = < **group 1** (dCas9-DNMT-3A [500 ng] + uPA gRNAs [500 ng])
- (3) **group 6** (dCas9-DNMT-3B [500 ng]) = < **group 2** (dCas9-DNMT-3B [500 ng] + uPA gRNAs [500 ng])

**Hypomethylated DMRs by dCas9 methyltransferases and uPA gRNAs should meet:**

% mCpG:

- (1) **group 9** (pUC19) > = **group 5** (dCas9-DNMT-3A only [500 ng]) > = **group 7** (dCas9-DNMT-3A [50 ng] + uPA gRNAs [50 ng]) > = **group 1** (dCas9-DNMT-3A [500 ng] + uPA gRNAs [500 ng]).
- (2) **group 2** (dCas9-DNMT-3B [500 ng] + uPA gRNAs [500 ng]) > = **group 1** (dCas9-DNMT-3A [500 ng] + uPA gRNAs [500 ng])
- (3) **group 6** (dCas9-DNMT-3B [500 ng]) > = **group 2** (dCas9-DNMT-3B [500 ng] + uPA gRNAs [500 ng])

Similarly, the authentic DMRs caused by dCas9 methyltransferases and TGFBR3 gRNAs should have a methylation pattern as described below:

**Hypermethylated DMRs by dCas9 methyltransferases and TGFBR3 gRNAs should meet:**

% mCpG:

- (1) **group 9** (pUC19) = < **group 5** (dCas9-DNMT-3A only [500 ng]) = < **group 8** (dCas9-DNMT-3A [50 ng] + TGFBR3 gRNAs [50 ng]) = < **group 3** (dCas9-DNMT-3A [500 ng] + TGFBR3 gRNAs [500 ng]).
- (2) **group 4** (dCas9-DNMT-3B [500 ng] + TGFBR3 gRNAs [500 ng]) = < **group 3** (dCas9-DNMT-3A [500 ng] + TGFBR3 gRNAs [500 ng])
- (3) **group 6** (dCas9-DNMT-3B [500 ng]) = < **group 4** (dCas9-DNMT-3B [500 ng] + TGFBR3 gRNAs [500 ng])

**Hypomethylated DMRs by dCas9 methyltransferases and TGFBR3 gRNAs should meet:**

% mCpG:

- (1) **group 9** (pUC19) > = **group 5** (dCas9-DNMT-3A only [500 ng]) > = **group 8** (dCas9-DNMT-3A [50 ng] + TGFBR3 gRNAs [50 ng])

- [500 ng]) > = **group 3** (dCas9-DNMT-3A [500 ng] + TGFBR3 gRNAs [500 ng]).
- (2) **group 4** (dCas9-DNMT-3B [500 ng] + TGFBR3 gRNAs [500 ng]) > = **group 3** (dCas9-DNMT-3A [500 ng] + TGFBR3 gRNAs [500 ng])
- (3) **group 6** (dCas9-DNMT-3B [500 ng]) > = **group 4** (dCas9-DNMT-3B [500 ng] + TGFBR3 gRNAs [500 ng])

We applied this methylation level-based filtering criteria to further remove potential stochastic DMRs. The remaining DMRs were subjected to all analyses as described in this study.

### Analysis of 5nt-SEED-NGG motif density

The 5nt-SEED-NGG density was calculated by counting the frequency of the sequence containing the 5nt SEED sequences preceding a NGG site on either DNA strand. The PAM density was calculated by counting the frequency of PAM sites (NGG) on either DNA strand. The median density with standard deviation is shown in the plots. Fisher exact test was conducted to compare densities between different sequence datasets.

### Statistics

All values in this study were presented as mean  $\pm$  standard deviation. The 1-way analysis of variance (ANOVA) with Bonferroni multiple testing, linear regression, Wilcoxon matched-pairs signed-rank test, Fisher exact test, and Benjamini-Hochberg-adjusted *P* value were used for statistical analysis. A *P* value < 0.05 was considered statistically significant.

## Results

### On-target DNA methylation by dCas9 methyltransferases: dCas9-BFP-DNMT3A and dCas9-BFP-DNMT3B

In mammalian cells, DNA methylation is established by *de novo* DNA methyltransferases (DNMT3A and DNMT3B) and maintained upon replication by DNMT1 [24]. Using a similar approach as Vojta et al. [15] and McDonald et al. [16], we fused DNMT1 catalytic domain, DNMT3A catalytic domain, DNMT3B catalytic domain, or EGFP to the C-terminal end of dCas9 with a blue fluorescent protein (BFP) and a triple tandem repeated flexible linker (3XG4S, Gly-Gly-Gly-Gly-Ser) (Fig. 1a and Supplementary Fig. S1a). Enrichment of cells expressing the fusion dCas9 methyltransferases were validated by BFP-based fluorescence activated cell sorting (FACS) (Supplementary Fig. S1b) and immunofluorescence staining using anti-HA tag antibody (Supplementary Fig. S1c).

To validate that dCas9 methyltransferases can methylate endogenous CpGs, the dCas9 methyltransferases were first targeted by 5 gRNAs (uPA gRNA T1 to T5, Fig. 1b) to the uPA promoter, which contains a dense CpG island (CGI) that is hypomethylated in human cancer cells [25]. HEK293T cells were transfected with uPA gRNAs and individual dCas9 fusion expression vectors. Following BFP-based FACS enrichment of transfected cells, the percentage of methylated CpGs (mCpGs) at individual CpG sites in the uPA promoter (uPA-MR1 and uPA-MR2 genomic regions) was quantified by bisulfite pyrosequencing (Fig. 1c). Compared to the pUC19 control, cells expressing uPA gRNAs and dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B, but not dCas9-BFP-DNMT1 or dCas9-BFP-EGFP, had significantly higher mCpG levels (*P* value < 0.01, ANOVA test). This is consistent with previous reports showing that the C-terminal catalytic domains

of DNMT3A and DNMT3B, but not DNMT1, are active [26, 27]. The CpGs most efficiently *de novo* methylated were located 10–50 bp upstream and downstream of the gRNA target sites. CpGs located in the gRNA binding sites were not methylated by the dCas9 methyltransferases, most likely because CRISPR/dCas9 binding blocks the interaction of the methyltransferase domain with the CpGs (Fig. 1c). *De novo* methylation by dCas9-BFP-DNMT3A and gRNAs was further validated by bisulfite Sanger sequencing (Supplementary Fig. S1d).

To investigate dCas9 methyltransferase-mediated methylation of another genomic locus, we generated 3 gRNAs targeting the transforming growth factor beta receptor 3 (TGFBR3) promoter. Similar *de novo* methylation effects were observed for dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B with TGFBR3 gRNAs (Fig. 1d–g; Supplementary Fig. S2). Our results collectively reveal that fusion of dCas9 to the catalytic domain of DNMT3A/3B can mediate targeted *de novo* DNA methylation.

### Off-target methylation by dCas9 methyltransferases

Since high-frequency off-target mutagenesis has been observed in previous applications of CRISPR-Cas9 [28], we investigated the specificity of dCas9 methyltransferases. For this purpose, we repeated the experiment with 2 additional controls: (1) cells expressing dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B only and (2) cells expressing dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B and 3 scrambled gRNAs (gRNAs targeting the CMV promoter). We found that expression of dCas9 methyltransferases and scrambled gRNAs could cause some unspecific *de novo* methylation of the uPA promoter, but at much lower levels compared to that obtained for uPA gRNAs (Supplementary Fig. S3). A slightly increased uPA promoter methylation, although not significant, was also observed in cells expressing dCas9 methyltransferase only (Supplementary Fig. S3).

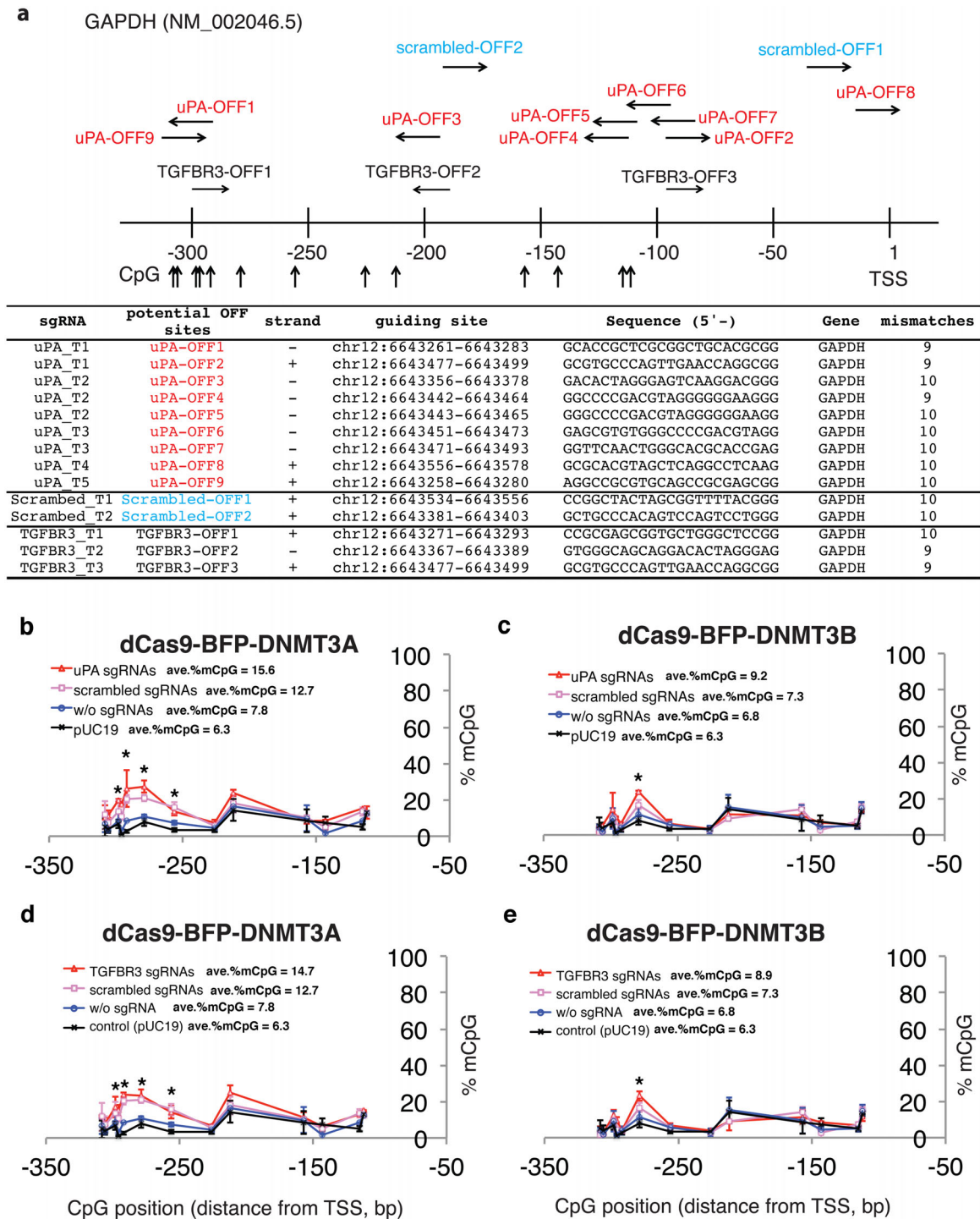
To further assess the off-target methylation, we investigated 3 genomic regions with various sequence similarities to the uPA gRNA target sites: SH2D3C (3 mismatches, Supplementary Fig. S4a), FAM221A (3 mismatches, Supplementary Fig. S4b), and GAPDH promoter (9 mismatches, Fig. 2a). We did not observe significant changes in CpG methylation at SH2D3C and FAM221A genomic sites. Surprisingly, several CpG sites in the GAPDH promoter were significantly methylated in cells expressing dCas9-BFP-DNMT3A and uPA, TGFBR3, or scrambled (CMV) gRNAs (Fig. 2b–c). The same was observed, but to a lesser extent, in cells expressing dCas9-BFP-DNMT3B (Fig. 2d–e). This effect was less prominent in cells expressing dCas9 methyltransferase only, indicating that unspecific methylation of the GAPDH promoter is RNA guided. Our results collectively reveal the existence of site-dependent off-target methylation by dCas9 methyltransferases.

### Effects of DNMT3A/3B catalytic activity and dCas9 methyltransferase expression level on on-target and off-target DNA methylation

*De novo* methylation by dCas9 methyltransferases could be mediated either by the catalytic activity of DNMT3A and DNMT3B or by the recruitment of additional DNA methylation enzymes to the binding sites facilitated by protein interactions. To elucidate the mechanism of on-target and off-target DNA methylation, we introduced the E752A and E697A catalytically inactivating mutations [29] in the DNMT3A and DNMT3B catalytic domains, respectively. To investigate the effect of dCas9 methyltransferase expression levels on on-target and off-target DNA



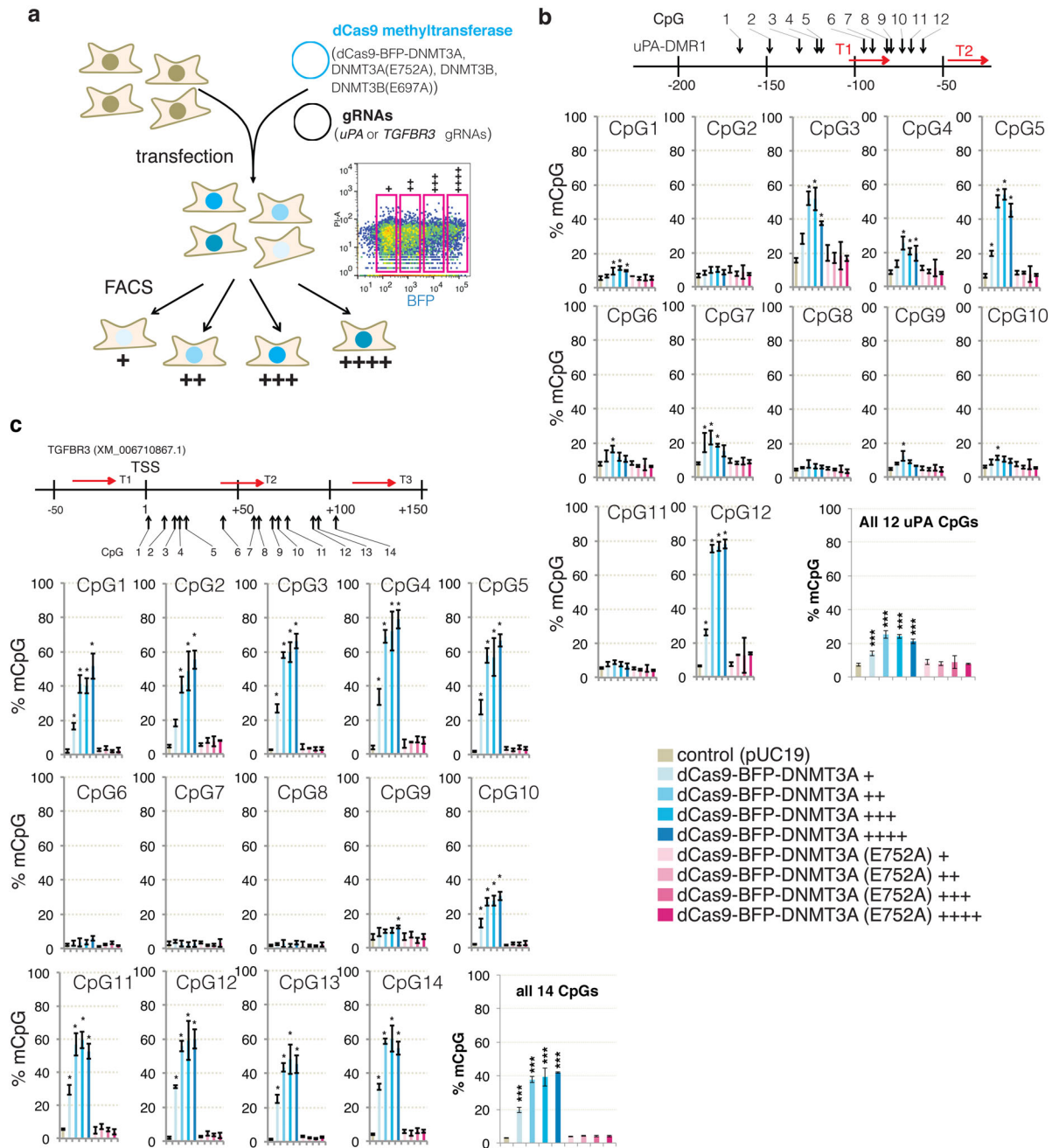




**Figure 2:** Off-target methylation of *GAPDH* promoter by dCas9 methyltransferases and gRNAs. (a) Schematic illustration of the *GAPDH* promoter. Potential off-target sites and CpGs analyzed by bisulfite pyrosequencing are indicated. Sequences of potential off-target binding sites by uPA, TGFBR3, and scrambled gRNAs with maximum 10 mismatches are listed. (b-d) Line plots of *GAPDH* promoter methylation in FACS-sorted HEK293T cells 48 hours after transfection with dCas9 methyltransferases and gRNAs. The methylation profiles from the pUC19-transfected samples were replotted as reference. Each data point in the graph represents the mean  $\pm$  SD ( $n = 2$  independent transfections). Average methylation levels for all CpGs analyzed are presented next to line legends. Asterisks (\*) represent  $P$  value  $< 0.05$  compared to pUC19 (ANOVA).

methylation, cells were sorted into 4 populations based on BFP signal intensity, a marker of dCas9 methyltransferase expression level: 1, very low: +; 2, low: ++; 3, medium: +++; and 4, high: ++++ (Fig. 3a). Bisulfite pyrosequencing analysis of the uPA (Fig. 3b) and TGFBR3 (Fig. 3c, Supplementary Fig. S5) promot-

ers revealed that only dCas9 methyltransferases, but not dCas9 methyltransferase catalytic mutants, cause dose-dependent *de novo* methylation, suggesting that *de novo* on-target methylation by dCas9 methyltransferases is mediated by the catalytic activity of DNMT3A and DNMT3B.



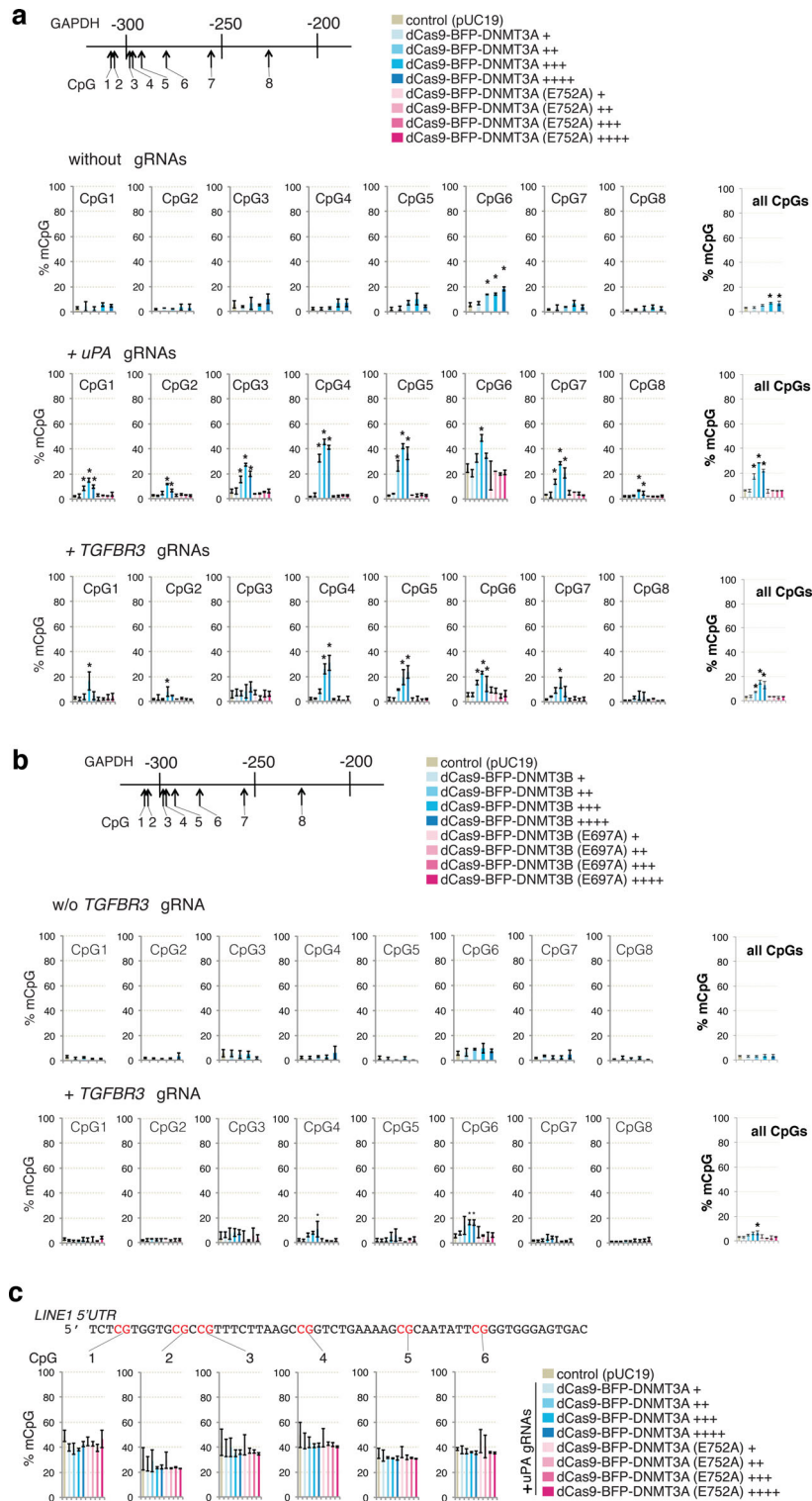
**Figure 3:** On-target methylation by dCas9 methyltransferases. (a) Schematic illustration of the experiment. dCas9 methyltransferase-expressing cells were enriched by FACS 48 hours after transfection and sorted according to the BFP signal: +, ++, +++, +++++. Right: Representative FACS plot and gating. (b) and (c) Bar charts indicating % mCpG for individual CpG and average values of all CpG sites in the uPA (b) and TGFB3 (c) target regions. The schematic illustrations above the bar graphs show gRNA binding sites and CpG sites analyzed. Value represents mean  $\pm$  SD (n = 3). Asterisk (\*) indicates statistical significance ( $P < 0.05$ , ANOVA) compared to the control after Bonferroni correction. Figure legend for bar graphs in (b) and (c) is presented at bottom right.

Next, we investigated the effect of dCas9 methyltransferase expression level on off-target methylation by analyzing the GAPDH promoter methylation in the FACS-sorted cells with different BFP signal intensity (+, ++, +++, and +++++). Consistent with previous results, co-expression of dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B (Fig. 4a, b) with either uPA or TGFB3 gRNAs significantly increased *de novo* methylation of GAPDH promoter CpGs compared to cells expressing dCas9 methyltransferase without gRNAs or pUC19. Furthermore, titrating dCas9 methyltransferase expression levels decreased unspecific methylation

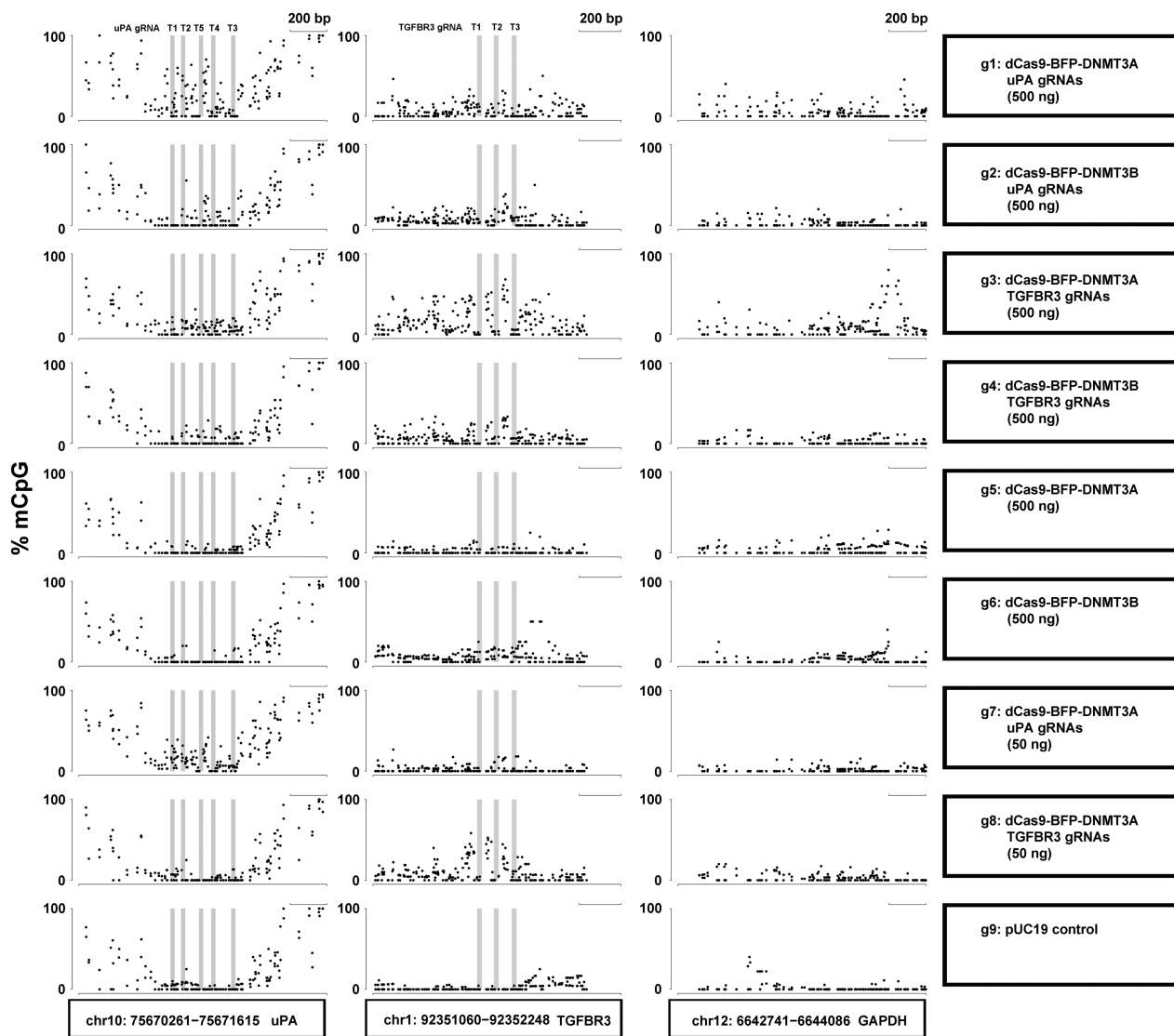
of the GAPDH promoter (Fig. 4a, b). Similarly, methyltransferase catalytic mutants do not cause *de novo* methylation of GAPDH. Since *de novo* methylation of gRNA-targeted genes was also decreased by dCas9 methyltransferase titration (Fig. 3), our results collectively suggest that altering dCas9 methyltransferase expression levels cannot efficiently reduce unspecific methylation relative to targeted methylation.

To investigate global methylation levels, repetitive LINE1 elements were investigated as they represent a surrogate marker for global DNA methylation [30]. We measured the LINE1 5'UTR





**Figure 4:** Off-target methylation by dCas9 methyltransferases. **(a)** Bar charts indicating % mCpG at individual CpGs and total % mCpG (8 CpG sites) for the GAPDH promoter in cells expressing different levels (BFP signal: +, ++, +++, +++) of dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3A(E752A) alone or together with either uPA or TGFBR3 gRNAs. **(b)** Bar charts indicating % mCpG in the GAPDH promoter in cells expressing different levels (BFP signal: +, ++, +++, +++) of dCas9-BFP-DNMT3B or dCas9-BFP-DNMT3B(E697A) alone or with TGFBR3 gRNAs. **(c)** LINE1 5'UTR methylation in cells expressing uPA gRNAs with different levels of either dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3A(E752A). Cells transfected with pUC19 were used as controls. Values represent mean  $\pm$  SD ( $n = 3$ ). Asterisks (\*) represent  $P$  value  $< 0.05$  (ANOVA) compared to pUC19.



**Figure 5:** *De novo* methylation of *uPA*, *TGFBR3*, and *GAPDH* promoters by dCas9 methyltransferase measured with WGBS. Dot plots of % mCpG for individual CpG sites in the *uPA*, *TGFBR3*, and *GAPDH* promoter regions. Each dot represents 1 CpG site. Right panel indicates the transfected plasmids. mCpG levels were quantified by WGBS. Scale bar, 200 bp.

methylation by bisulfite pyrosequencing, which revealed that expression of dCas9-BFP-DNMT3A and *uPA* gRNAs did not result in significant *LINE1* methylation changes (Fig. 4c).

### Genome-wide bisulfite sequencing revealed off-target methylation by dCas9 methyltransferases

Prompted by the unspecific methylation of *GAPDH* promoter by dCas9 methyltransferases, we investigated the genome-wide off-target methylation characteristics by CRISPR dCas9 methyltransferases using WGBS. WGBS was conducted in HEK293T cells transfected with (i) pUC19 (control), (ii) dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B alone, and (iii) dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B with either *uPA* or *TGFBR3* gRNAs with 2 different doses (50 ng or 500 ng) (Supplementary Fig. S6a). Using the Illumina HiSeq X platform, we generated more than 100 Gb of clean data for each sample (more than 30 $\times$  coverage with a 99.5% bisulfite conversion rate). This allowed us to analyze the methylation pattern at single base-pair resolution. Since mainly CpG

dinucleotides are subject to methylation in HEK293T cells (Supplementary Fig. S6b), all following analyses are based on CpG methylation in the entire genome (approximately 40 000 000 CpG sites). First, we examined *uPA*, *TGFBR3*, and *GAPDH* promoter methylation as revealed by WGBS in all 9 groups. WGBS confirmed that the *uPA* and *TGFBR3* gRNAs could target dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B to the *uPA* and *TGFBR3* loci and methylate CpGs flanking the gRNA binding sites in a dose- and gRNA-dependent manner (Fig. 5). Furthermore, our WGBS data revealed that some dCas9 methyltransferase-mediated *de novo* methylation of *uPA*, *TGFBR3*, and *GAPDH* (off-target) promoters occurred in a broad region surrounding the gRNA binding site.

Next, we analyzed the global DNA methylation profile. Consistent with the *LINE1* assay (Fig. 3c), expression of dCas9 methyltransferase alone or together with gRNAs was not associated with global methylation changes (Supplementary Fig. S6c, d). Since we have only 1 replicate per group and stochastic methylations frequently occur in cancer cells during cultivation [31], we analyzed the data with DSS-single (a method

developed by Wu et al. for detecting DMRs from WGBS data without replicates [32]) to identify DMRs caused by dCas9 methyltransferase and gRNAs. We compared cells transfected with dCas9 methyltransferases with or without gRNAs to control cells (transfected with pUC19 control plasmid). More than 10 000 hyper or hypo DMRs were identified by DSS-single (Supplementary Fig. S7). Next, based on the observation that (1) there is dose- and gRNA-dependency of uPA, TGFBR3, and GAPDH methylation by dCas9 methyltransferase and (2) dCas9-BFP-DNMT3A is more efficient than dCas9-BFP-DNMT3B, we applied a stringent filtering step to remove potentially stochastic DMRs. Following this filtering, we identified more than 1000 DMRs resulting from dCas9 methyltransferase together with either uPA gRNAs (hypermethylated DMRs [hyper-DMRs] = 3671; hypomethylated DMRs [hypo-DMRs] = 1807) or TGFBR3 gRNAs (hyper-DMRs = 2267; hypo-DMRs = 1662) (Supplementary Tables S2–S5). These DMRs were on average 63–81 bp and contained an average of 5–9 CpGs (Supplementary Fig. S8). The average methylation levels of these hyper-/hypo-DMRs differ significantly between pUC19 control cells, cells expressing dCas9 methyltransferase only, cells expressing low amounts of dCas9 methyltransferase and gRNAs, and cells expressing high amounts of dCas9 methyltransferase and gRNAs (Fig. 6a, Supplementary Fig. S9a). Only a very small portion of the DMRs (hyper-DMRs = 192; hypo-DMRs = 81) were commonly found among DMRs caused by dCas9 methyltransferase and uPA compared to TGFBR3 gRNAs (Fig. 6b), suggesting that the majority of the off-target DMRs are RNA guided. Taken together, our WGBS result revealed that expression of dCas9 methyltransferases together with gRNAs can cause substantial off-target methylation.

### Characteristics of dCas9 methyltransferase off-targets

To better describe the characteristics of dCas9 methyltransferase off-targets, we stratified hyper- and hypo-DMRs according to their localization in particular types of genomic regions, including promoters, coding sequences, introns, 5' untranslated regions (5'-UTR), 3'-UTR, CGIs, CGI shores, Alu sequences, LINE1 (L1) sequences, and LINE2 (L2) sequences. Our results showed that hyper-DMRs were predominantly enriched in promoters, 5'-UTR, and CGIs, whereas hypo-DMRs were enriched in repeated sequences Alu and LINE1 (Fig. 6c–d, Supplementary Fig. S9b–c). Consistent with this finding, a metaplot of average methylation levels for all genes before the DSS-single call also showed that transcription start site flanking regions (overlapping with promoters and 5' UTR) were hypermethylated in cells expressing dCas9 methyltransferase and gRNAs (Supplementary Fig. S10).

Since dCas9 preferentially binds open chromatin regions [33], we further analyzed DNase I hypersensitivity regions based on ENCODE data from HEK293T cells (GEO#: GSM1008573) and quantified the average methylation level in DNase I hypersensitivity sites (DHS) (as an indication of sites with an open chromatin state). The DHS flanking regions (1 kb upstream and downstream) were used as a control. Compared to cells transfected with pUC19, cells expressing dCas9 methyltransferase and gRNAs had significantly higher methylation levels in the DHS sites ( $P$  value < 0.05; Wilcoxon matched-pairs signed-rank test) (Fig. 6e, Supplementary Fig. S9d). Furthermore, only hyper-DMRs, but not hypo-DMRs, were significantly enriched in DHS ( $P$  value <  $1e-300$ , Fisher exact test, Fig. 6f–g and Supplementary Fig. S9e–f), which collectively suggests that open chromatin regions are prone to unspecific methylation by dCas9 methyltransferase and gRNAs.

Previous studies have discovered that complementary base pairing between gRNA guide sequences and the PAM-proximal 5nt region (5ntSEED-PAM) is crucial for off-target binding [33, 34]. We also assessed the density of individual gRNA 5ntSEED-PAM sequence (5'-NNNNNNGG-3') in the hyper- and hypo-DMRs. For each DMR, we included the 100-bp flanking sequences when calculating the presence of 5ntSEED-PAM sequence density. This is based on the previous observation that dCas9 methyltransferases methylate CpGs flanking the gRNA binding site. We consistently observed significant enrichment of 5ntSEED-PAM sequences for all gRNAs in the hyper-DMRs, but not hypo-DMRs (Fig. 6h, Supplementary Fig. S9g). Taken together, this shows that if guided by gRNAs, dCas9 methyltransferases can cause substantial off-target methylation of genomic regions with open chromatin accessibility, i.e., promoters and 5' UTR, as well as CGI. Our finding between the off-target methylation and the chromatin accessibility is also consistent with our recent discovery that CRISPR/Cas9 cleaves more efficiently in euchromatin than heterochromatin regions [35].

### dCas9 methyltransferase-mediated hypermethylated DMRs are weakly correlated with off-target binding

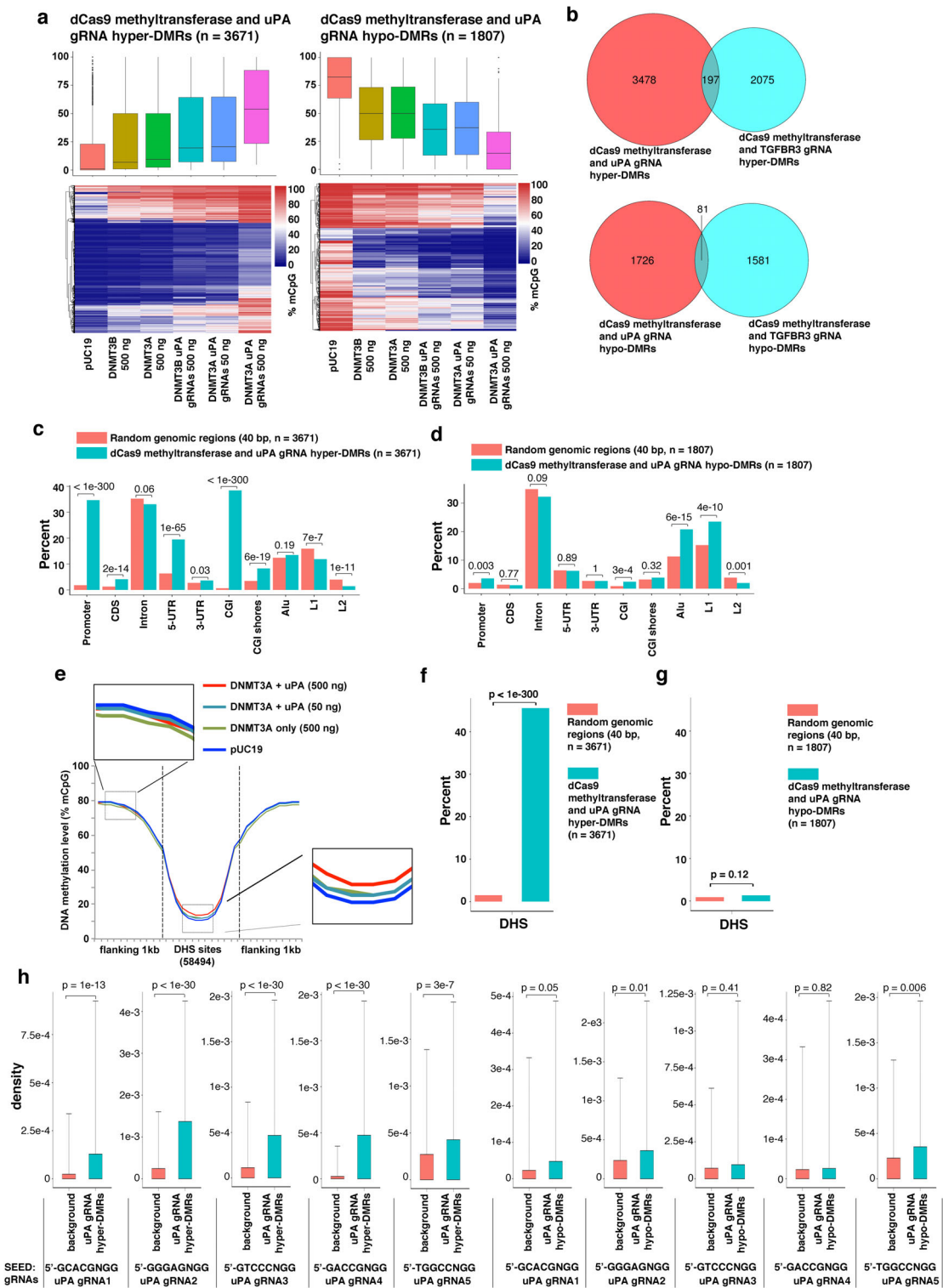
To further investigate the association between dCas9 methyltransferase off-target methylation and dCas9 off-target DNA binding, we studied off-target binding sites in HEK293T cells expressing dCas9 methyltransferase and uPA gRNAs using ChIP-seq. Using pair-wise comparison as the previous approach for dCas9 [34], 805 enriched peaks ( $P$  value < 0.001, Supplementary Table S6) were identified. These ChIP peaks were scattered throughout the genome and significantly enriched in DHS genomic regions (Fig. 7a, b). Using MEME motif scanning of ChIP peaks [36], we identified the most significant motif GGGAGAGGGAGNGG ( $P = 1.0e-593$ ). This motif is identical to the 11-bp seed sequences of uPA gRNA T2 (GAGCCGGCGGGAGAGGGAG(GGG)) and the PAM (NGG) site (Fig. 7c), suggesting that T2 is dominant compared to other uPA gRNAs in mediating off-target binding. Analysis of 5ntSEED-PAM sequence density further confirmed that uPA T2 binding sites were overrepresented in the ChIP peaks (Fig. 7d). A previous study has shown that the choice of gRNAs has a great effect on dCas9 off-target binding [34]. The uPA gRNA T2 is highly G-rich or AG-rich in the seed region. This can potentially be the cause of most of the off-target activities. This could be the explanation of why we have found 40 times more off-target binding sites compared to those reported in the study by Liu et al. [37].

Next, we analyzed the correlation between the ChIP peaks and the uPA DMRs (including the flanking 100 bp of each DMR). There is a significantly increased overlap between ChIP peaks and uPA hyper-DMRs ( $P = 0.006$ , Fisher exact test), but not uPA hypo-DMRs ( $P = 1$ , Fisher exact test) (Fig. 7e). However, the percentage of uPA hyper-DMRs overlaps with ChIP peaks is still very low (11 of 3671 hyper-DMRs, 0.3%). Since the average methylation level of all ChIP peak regions exceeds 60% (Supplementary Fig. S11), this may partially explain why there is a low correlation between ChIP peaks and DMRs given potential functional difficulty in further increasing the methylation level. Furthermore, ChIP-seq only identified sites to which the dCas9 methyltransferase binds strongly.

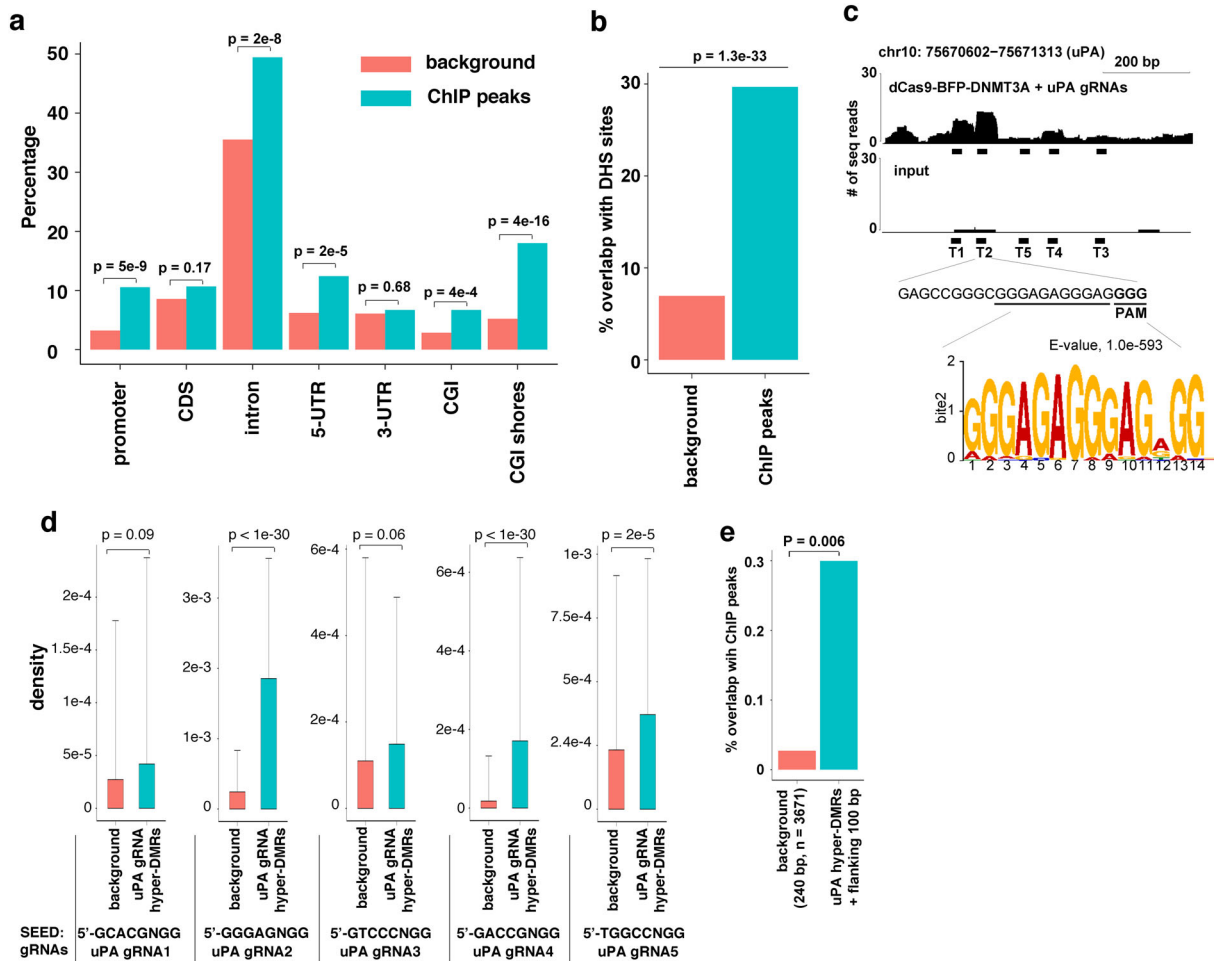
### Effects of dCas9 methyltransferases on gene expression

Methylation of promoter DNA can be correlated with inhibition of gene transcription. To determine whether the dCas9





**Figure 6:** Genomic characteristics of off-target DMRs caused by dCas9 methyltransferases and uPA gRNAs. (a) Box plot (top) and heat map clustering (bottom) of the hypermethylated (left) and hypomethylated (right) DMRs resulting from dCas9 methyltransferases and uPA gRNAs. (b) Venn diagram presentation of hypermethylated (top) or hypomethylated (bottom) DMRs caused by dCas9 methyltransferases and uPA gRNAs compared to TGFBR3 gRNAs. (c and d) Bar chart illustrating the percentage of the identified uPA hypermethylated (c) or hypomethylated (d) DMRs that fall into the different types of genomic regions indicated. Background represents a random sample of the same number of similar sized genomic windows that fall into the categories indicated. Values above bars are P values between background and uPA-DMRs (Fisher exact test). (e) Metaplot of average CpG methylation levels in 58 494 DNase I hypersensitive sites (DHS) and 1 kb upstream and downstream flanking regions. (f and g) Bar chart of % uPA hypermethylated (f) or hypomethylated (g) DMRs falling into DHS core regions. (h) Density of 5nt-SEED-NGG for uPA gRNAs (T1 to T5) in background genomic windows and uPA DMRs + flanking 100 bp. Values represent median density with 1 standard deviation. P values (t test) are given above the bar charts.

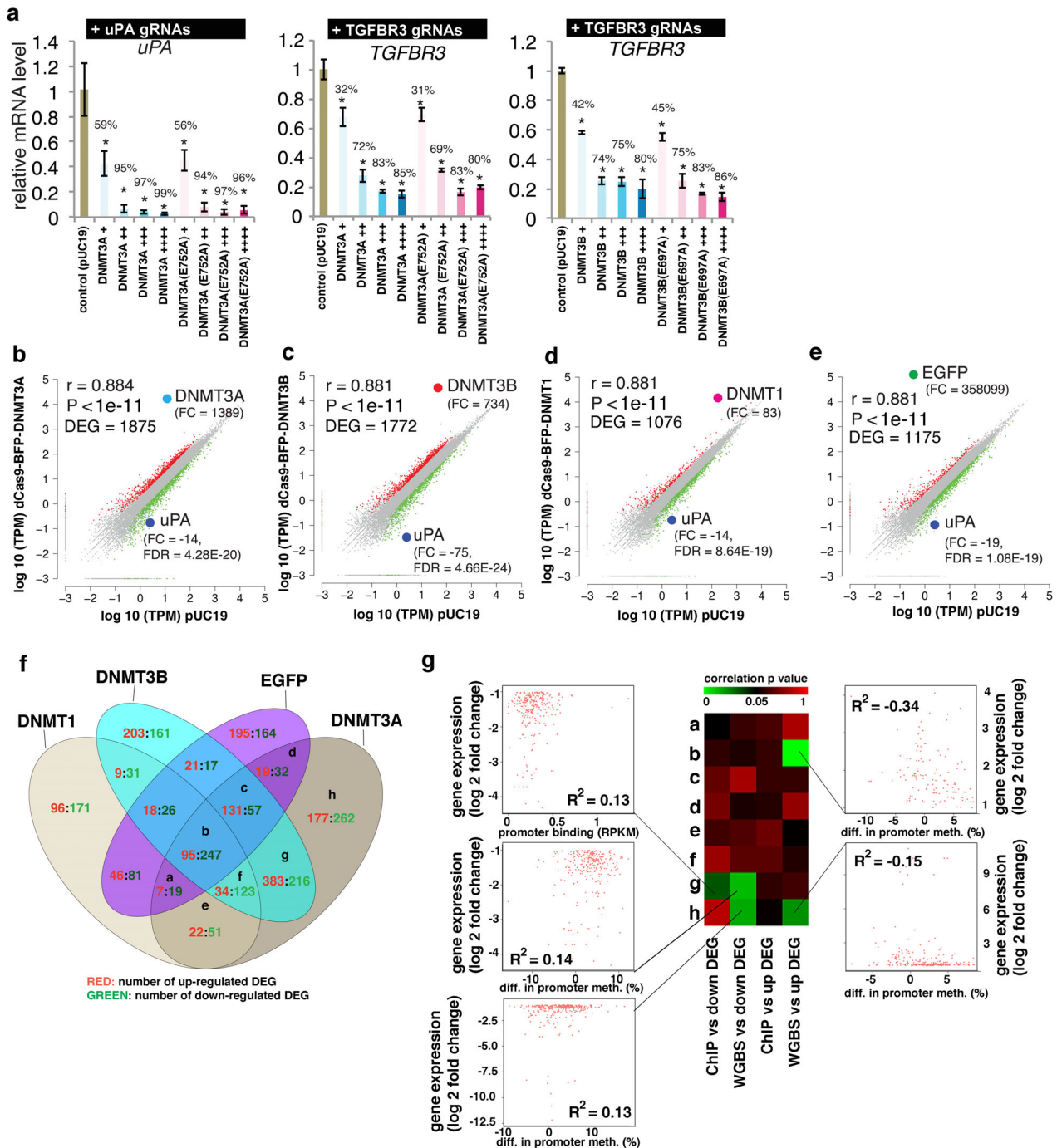


**Figure 7:** Correlation between dCas9-BFP-DNMT3A off-target binding and off-target methylation. (a) Bar chart illustrating the percentage of ChIP peaks from cells expressing dCas9-BFP-DNMT3A and uPA gRNAs or background control regions (random sampling of the same number of similar sized genomic windows as the ChIP peaks) falling into the different types of genomic regions indicated. P values between background and ChIP peaks indicated above bars, Fisher exact test. (b) Bar chart of % ChIP peaks falling into DHS core regions. (c) Representative plot of ChIP-seq reads in the uPA promoter, uPA gRNA T2 sequences, and the top motif identified by MEME-ChIP. (d) Density of 5nt-SEED-NGG for uPA gRNAs (T1 to T5) ChIP peaks. Background is a random sample of the same number of similar sized genomic windows as ChIP peaks. Values represent median density with 1 standard deviation. P values are given for the indicated comparisons (t test). (e) Bar plot of % ChIP peaks overlapping with hypermethylated DMRs caused by dCas9 methyltransferase and uPA gRNAs. Background is a random sample of the same number of similar sized genomic windows as DMRs.

methyltransferase-mediated uPA and *TGFBR3* promoter methylation could inhibit gene expression, we measured uPA and *TGFBR3* mRNA levels by qPCR in HEK293T cells. Compared to the pUC19 transfection control, both uPA and *TGFBR3* expressions were significantly decreased in cells expressing dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B and either uPA or *TGFBR3* gRNAs (Fig. 8a). However, the reduced uPA and *TGFBR3* expression does not appear to be only associated with the *de novo* DNA methylation by dCas9 methyltransferases (Fig. 8a), as inactivating dCas9 methyltransferase mutants dCas9-BFP-DNMT3A(E752A) and dCas9-BFP-DNMT3B(E697A) also cause similar degrees of expression inhibition despite their lack of *de novo* DNA methylation activity.

To investigate whether the inhibition of gene expression is specific to the gRNA targeted genes, we conducted RNA sequencing in HEK293T cells expressing dCas9 methyltransferase and uPA gRNAs. A large number (>1000) of differentially expressed genes (DEGs) significantly (FDR  $P$  value < 0.001, fold change > 2) were found in cells expressing uPA gRNAs and either dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B (Fig. 8b-c). However, similar

effects on the global transcription profile were observed in cells expressing uPA gRNAs with dCas9-BFP-DNMT1 or with dCas9-BFP-EGFP lacking *de novo* DNA methylation activity (Fig. 8d-e). We cross-compared DEGs among the 4 groups, and 342 (18–32%) genes were commonly identified (Fig. 8f). For DEGs found in cells expressing dCas9-BFP-DNMT3A and uPA gRNAs, we also performed integrative analyses of the expression change, promoter methylation, and promoter binding intensity (Fig. 8g). Very weak but significant correlation was identified for a few clusters of DEGs. Taken together, these results suggest that the nonspecific alteration of transcription is not merely caused by promoter methylation or binding of dCas9 methyltransferase. Since uPA is an important factor in regulating cell proliferation and inhibition of cell growth was found in cells expressing dCas9 methyltransferases and uPA gRNAs (Supplementary Fig. S12), the large number of differentially expressed genes might be a result of altered cellular functions. Taken together, our results clearly indicate that inhibition of uPA and *TGFBR3* expression by dCas9 methyltransferase and corresponding gRNAs is not merely due to *de novo* DNA methylation of their promoters.



**Figure 8:** Effect of dCas9 methyltransferases on gene expression. (a) Relative gene expression levels of uPA and TGFB3 in cells expressing different levels of dCas9-BFP-DNMT3A, dCas9-BFP-DNMT3B, dCas9-BFP-DNMT3A(E752A), or dCas9-BFP-DNMT3B(E697A). mRNA expression was measured by qPCR and quantified as fold change compared to control cells transfected with pUC19. Bar charts depict mean change in mRNA level compared to pUC19 controls. Data represent mean  $\pm$  SD ( $n = 3$  independent transfections). Mean percentage decrease in mRNA level compared to pUC19 is presented on top of bars. Asterisks (\*) represent P value  $< 0.05$  compared to pUC19. (b-e) Dot plots of  $\log_{10}$  (TPM) for all genes expressed in the BFP positive (BFPp) cells expressing uPA gRNAs (T1-T5) and dCas9-BFP-DNMT3A (b), dCas9-BFP-DNMT3B (c), dCas9-BFP-DNMT1 (d), or dCas9-BFP-EGFP (e) plotted against  $\log_{10}$  (TPM) in a pUC19 control group. DEGs are marked in red (upregulated) and green (downregulated) (fold change  $\geq 2$ , FDR  $< 0.001$ ). Fold changes compared to pUC19 and FDR P values for DNMT1, DNMT3A, DNMT3B, EGFP, and uPA are shown. (f) Venn diagram representation of cross-comparison of DEGs. (g) Integrative analysis of gene expression change, promoter methylation, and promoter binding caused by dCas9-BFP-DNMT3A and uPA gRNAs for the different clusters of DEGs. Heat map represents linear regression P values. Dot plots were given for significant correlations ( $P < 0.05$ ).



To determine whether longer-term inhibition of gene expression can be facilitated by dCas9 methyltransferases, 5 HEK293T fluorescent reporter cell clones carrying different copies of a CMV-mCherry expression cassette (Supplementary Fig. S13a, b) were generated. We quantified mCherry level by FACS for 2 weeks after transfection. We observed that the number of dCas9 methyltransferase-expressing cells peaked on day 2 and decreased gradually (Supplementary Fig. S13c). Maximal inhibition of mCherry levels were observed on day 5 after transfection (Supplementary Fig. S13d-h). Compared to other dCas9 fusion proteins, the dCas9-BFP-DNMT3A fusion resulted in the highest and longest inhibition of mCherry expression in the reporter cells (4 of 5 clones) (Supplementary Fig. S13d-h). The transient and prolonged inhibition efficacy varied among the 5 cell clones. For example, clone 2, which has the lowest copy number of transgene, showed the highest transient and longest inhibition by dCas9-BFP-DNMT3A (Supplementary Fig. S13e). However, expression of mCherry was, in all clones, not significantly different from the pUC19 control after 2 weeks, suggesting that inhibition of gene expression by dCas9 methyltransferases is not stably maintained.

## Discussion

Since dCas9 methyltransferases are targeted to a specific genomic locus simply by a small gRNA, this system is more convenient than ZF- or TALE-based methyltransferases [29, 38, 39]. Recently, Vojta et al. [15] and McDonald et al. [16] reported that directly fusing DNMT3A to dCas9 could be used to induce DNA methylation at specific loci in HEK293T cells. Consistent with that, we show that dCas9-BFP-DNMT3A can methylate CpGs flanking the gRNA binding sites in genomic loci, further proving the general applicability of dCas9 methyltransferases for targeted DNA methylation in mammalian cells. In addition, our study shows for the first time that the fusion of dCas9 to DNMT3B is also capable of inducing specific DNA methylation, although the efficiency is lower than that of DNMT3A. Additionally, Stepper et al. showed that the dCas9-DNMT3A-DNMT3L fusion can further improve *de novo* methylation efficiency compared to dCas9-DNMT3A [40]. Together with the reported systems, the dCas9 methyltransferases system reported in this study further broadens the availability and applicability of CRISPR-based reprogramming of DNA methylation. Based on the observation that dCas9 methyltransferases can efficiently methylate the flanking CpG sites from the gRNA binding site, we have developed an open-source web-based gRNA designing tool for dCas9 methyltransferase gRNAs [18].

On the basis of extensive gene-specific bisulfite pyrosequencing and WGBS, we identified novel off-target methylation characteristics that appear to be predominantly enriched in promoter, 5'UTR, CGI, and open chromatin regions. Since most of these genomic regions are hypomethylated in HEK293T cells, it was expected that the off-target DMRs would be enriched in such regions. In other genomic regions, which already have a high level of methylation, a further methylation by dCas9 methyltransferase is not achievable. We discovered that open chromatin regions are highly prone to off-target methylation by dCas9-methyltransferase. The location of the GAPDH promoter in a DHS region explains why this region is subjected to highly off-target methylation. To further confirm our finding, we repeated the WGBS experiment in triplicates (Supplementary Fig. S14). Our result confirmed that the hyper-methylated DMRs identified in a previous WGBS experiment (Fig. 6a) are

significantly increased in cells overexpressing dCas9-BFP-DNMT3A and uPA in the repeated experiments (Supplementary Fig. S14a). Consistently, DHS regions were significantly methylated in cells expressing dCas9-BFP-DNMT3A and uPA (Supplementary Fig. S14b).

Our study also revealed the gRNA dependency of off-target methylation. This is consistent with the observations of McDonald et al. [16] and Vojta et al. [15]. Additionally, we discovered that even in the absence of gRNAs, expression of the dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B alone can cause some unspecific DNA methylation. This gRNA-independent off-target methylation effect is even more pronounced when too many dCas9 methyltransferases, or the DNMT3A catalytic domain, enter the nucleus. For example, increasing dCas9 methyltransferase expression level, fusing the catalytic domain of DNMT3A or DNMT3B directly to Cas9 without the BFP linker, or overexpressing the DNMT3A catalytic domain will cause increased gRNA-independent off-target methylation (see extended data and description in Supplementary File 1).

In this study, we found that expressing dCas9 methyltransferases and gRNAs could also cause significant demethylation of genomic regions enriched in repeated sequences. Repeated sequences, which make up more than half of the human genome, are generally highly methylated, and their dynamics, to some extent, are associated with normal development and tumorigenesis. A previous study of methylation in repeated sequences has shown that with increasing age from adulthood, there is a global decrease in DNA methylation in repeated sequences and intergenic genome sequences [41]. We also observed that expression of dCas9 methyltransferase alone or together with gRNA can inhibit HEK293T cell growth (Supplementary Fig. S12). The hypomethylated DMRs could potentially be the result of inhibited cell proliferation by dCas9 methyltransferase and gRNAs. This should be investigated in future studies.

Improvement of dCas9 methyltransferase specificity, to minimize the gRNA-dependent and gRNA-independent off-target activity, is crucial for future applications of the technology. McDonald et al. [16] have observed significant reduction in off-target methylation using DOX inducible dCas9-DNMT3A. Consistent with these findings, we found that reducing the dCas9 methyltransferase and gRNA expression levels, as well as lowering the dCas9 methyltransferase nuclear entry efficiency (Supplementary File 1), can reduce off-target methylation. However, this approach also reduced on-target methylation levels accordingly. Thus, this may not represent a plausible way of increasing the specificity of the system. New approaches should be developed to reduce off-target methylation while maintaining sufficient on-target methylation efficiencies. The results presented in this study highlight the importance of inclusion of extensive controls in subsequent experiments, such as catalytically inactive dCas9 methyltransferase mutants, scrambled gRNAs, and gRNA free settings. This is necessary for reliable interpretations of correlations between specific DNA methylation events by dCas9 methyltransferase, gene expression regulation, and phenotypic effects.

Off-target effect is a major concern in CRISPR/Cas9-based DNA manipulation technologies and applications. Unlike the original CRISPR/Cas9 technology, of which the endonuclease activity of Cas9 depends heavily on the base-pairing between the guide sequences and the target site (proto-spacer), the dead Cas9 (dCas9)-derived CRISPR technologies and applications are more dependent on the physical interaction between the dCas9/gRNA complex and the DNA loci and is more tolerant to mismatches.

The dCas9 in such CRISPR-derived systems is acting as a cargo protein, bringing whatever proteins/domains to a specific genomic locus guided by the small gRNA. As demonstrated in Fig. 2 and Supplementary Fig. S4 of this study and several previous investigations by ChIP-seq [42–44], the criteria of defining off-target sites (simply based on mismatches) from wild-type Cas9 is not suitable for the dCas9-derived effector proteins, such as the dCas9 methyltransferases. Although we only evaluated the dCas9 methyltransferases, we speculate that these off-target effects are most likely to be the same for other kinds of dCas9-based effectors.

In this study, we also observed that dCas9 methyltransferases can efficiently inhibit expression of genes in human cells. However, the transient inhibition of gene expression could have resulted from both promoter methylation and blockage of transcription by dCas9 methyltransferases. A previous study reported that targeted DNA methylation by a zinc finger-based methyltransferase is not stably maintained [45]. Our time-course experiments to study the inhibition of gene expression gradually decreased during *in vitro* expansion of the transfected cells. This could be the result of removal of the *de novo* established epigenetic marks, dilution of the dCas9 methyltransferase expression plasmids, and/or negative selection of the cells expressing dCas9 methyltransferases. We also realize that DNA methylation and gene expression analyses were conducted in cells transiently transfected with dCas9 methyltransferase expression plasmids, which might lead to severe overexpression of the dCas9 methyltransferases. Thus, future studies could benefit from being conducted in cells stably or conditionally expressing low copy numbers of dCas9 methyltransferase to minimize off-target methylation. Our study is the first to reveal novel characteristics of the on-target and off-target DNA methylation by dCas9 methyltransferases on a genome-wide scale with single-base resolution and it highlights the need for development of CRISPR DNA methylation editing systems with higher specificity.

## Conclusions

The dCas9 methyltransferases presented here, and other dCas9 fusion protein systems described previously [11, 12, 15, 16], provide useful tools for targeted epigenome editing. Continued improvement of the specificities of these systems and combining tools to enable simultaneous modification of multiple histones and DNA loci will enable more precise and stable regulation of gene structure and function. Such CRISPR gRNA-guided programmable epigenetic modification tools will hopefully have broad research applications to delineate the association between specific epigenetic changes, gene-expression regulation, and phenotypes.

## Availability of supporting data

RNA sequencing, WGBS, and ChIP-seq data are available from the publicly available repository (GEO) and in GigaDB [46].

RNA-seq: GSE74935

WGBS: GSE92310, GSE92311

ChIP-seq: GSE92261

## Additional files

Supplementary Fig. S1. Validation of dCas9 methyltransferase expression and uPA promoter methylation. (a) Schematic

overview of the human DNA methyltransferases (DNMT1, DNMT3A, and DNMT3B), with the N-terminal regulatory region, a C-terminal catalytic domain (CD), and the cytosine C5-DNA methyltransferase motifs highlighted. The first amino acid (a.a) residue of the C-terminal catalytic domain, which was fused to the dCas9, is indicated by an arrow. (b) Representative FACS sorting and re-analysis of HEK293T cells 48 hours after transfection. Gating for BFP positive (BFPp) and negative (BFPn) cells are indicated. (c) Laser scanning microscopy of dCas9 methyltransferase expression in HEK293T cells, 48 hours after transfection. The BFP signal from the dCas9-BFP-DNMT1 transfected cells was enhanced since the BFP signal from the dCas9-BFP-DNMT1 fusion was initially weaker compared to that from the other three fusion proteins. Scale bar: 20  $\mu\text{m}$ . (d) Validation of RNA-guided uPA methylation (uPA-MR1) by dCas9-BFP-DNMT3A using bisulfite Sanger sequencing.

Supplementary Fig. S2. Validation of dCas9 methyltransferase-mediated TGFB3 methylation in HEK293T cells by bisulfite Sanger sequencing. TGFB3 methylation by dCas9 methyltransferase and gRNAs was validated by bisulfite Sanger sequencing. CpG methylation status is indicated according to the absolute nucleotide position and color-coded as red, methylated; blue, unmethylated; or white, unknown methylation state based on the sequencing signal.

Supplementary Fig. S3. Validation of *de novo* methylation of uPA by dCas9 methyltransferase and uPA gRNAs. Line plots of uPA-MR2 methylation in cells transfected with pUC19 (control), dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B only, and dCas9-BFP-DNMT3A or dCas9-BFP-DNMT3B together with either uPA gRNAs or scrambled gRNAs.

Supplementary Fig. S4. Effect of dCas9 methyltransferases on two potential off-target sites (SH2D3C and FAM221A). (a-b) Schematic illustration of the SH2D3C (b) and FAM221A (c) off-target loci, with off-target sites indicated by red arrows. Sequences of uPA gRNA (T2), SH2D3C, and FAM221A off-target sites are given above, with the PAM (red letters) and mismatches (green letters) indicated. CpGs analyzed are indicated by black arrows; numbers indicate distances (in bp) from the transcription start site (TSS) of the gene (SH2D3C, NM.001252334.1) or (FAM221A, XM.011515369.1). Y-axis represents % mCpG level for each CpG site and X-axis represents distance (in bp) from TSS. The CpG methylation level from the control samples (pUC19 transfection) was re-plotted as a reference. Each data point in the graph represents the mean percentage of CpGs methylated  $\pm$  SD ( $n = 2$ , independent transfections).

Supplementary Fig. S5. Effects of DNMT3B catalytic activity and expression level on *de novo* TGFB3 methylation. Bar charts of % mCpG level for individual CpG sites of the TGFB3 targeted regions in dCas9 methyltransferase-expressing cells. Cells were enriched by FACS 48 hours after transfection and sorted according to the BFP signal: +, ++, +++, +++++. The schematic illustrations above the bar charts show gRNA binding sites and CpG sites analyzed. Asterisk (\*) indicates statistical significance ( $P < 0.05$ , ANOVA) compared to the pUC19 control group after Bonferroni correction. Percentage values represent % decrease of TGFB3 expression compared to pUC19.

Supplementary Fig. S6. WGBS analysis of cells expressing dCas9 methyltransferase and gRNAs. (a) Summary of WGBS including clean data, clean reads, clean rate, mapped reads, uniquely mapped reads and rate, and bisulfite conversion rate, for each experimental group and control (pUC19). (b) Average percentage of methylated cytosine (% mC) for whole-genome CpG sites, CHG sites, and CHH sites. “H” represents A, C, and T. (c-d) Average mCpG level (percentage) stratified according to

individual chromosome or whole genome for all samples measured by WGBS.

Supplementary Fig. S7. Differentially methylated regions (DMRs) identified by DSS-single method. DMRs were categorized as hypermethylated or hypomethylated compared to control sample (pUC19 transfection).

Supplementary Fig. S8. Histogram charts of the distribution of DMR length (bp), and number of CpGs per DMR. DMRs included in this figure are those remaining after the stringent filtering step (see methods). Mean DMRs length (in bp) and mean number of CpG per DMR were given for each chart.

Supplementary Fig. S9. Genomic characteristics of off-target DMRs caused by dCas9 methyltransferases and TGFBR3 gRNAs. (a) Box plot (top) and heatmap clustering (bottom) of the hypermethylated (left) and hypomethylated (right) DMRs caused by dCas9 methyltransferases and TGFBR3 gRNAs. (b-c) Bar chart illustrating the percentage of the identified TGFBR3 hypermethylated (c) or hypomethylated (d) DMRs that fall into the different types of genomic regions indicated. Background represents of a random sample of the same number of similar sized genomic windows that fall into the categories indicated. Values above bars are P values between background and TGFBR3 DMRs, Fisher's exact test. (d) Metaplot of average CpG methylation levels in 58 494 DNase I hypersensitive sites (DHS) and 1 kb upstream and downstream flanking regions. (e-f) Bar chart of % TGFBR3 hypermethylated (f) or hypomethylated (g) DMRs falling into DHS core regions. (h) Density of 5nt-SEED-NGG for TGFBR3 gRNAs (T1 to T3) in background genomic windows and TGFBR3 DMRs + flanking 100 bp. Values represent median density with one standard deviation. P values (t-test) are given above the bar charts.

Supplementary Fig. S10. Average methylation levels of seven genomic regions in all annotated genes (hg19). (a-d) Each line indicates the genome-wide average methylation levels across seven genomic regions: upstream 2kb of the transcription start site, first exon, first intron, internal exons, internal introns, last exon, and downstream 2kb of the last exon.

Supplementary Fig. S11. The average methylation level in ChIP-peaks and flanking regions. Bar chart presents the average methylation level of all dCas9-BFP-DNMT3A and uPA gRNA off-target binding sites ( $n = 7754$ ) found by ChIPseq, as well as the 2kb upstream and downstream region.

Supplementary Fig. S12. Effect of dCas9 methyltransferases and uPA gRNAs on cell growth. Cell growth was determined by counting the number of cell clones derived from 1000 BFP positive cells after transfection. Values represent mean and one standard deviation from 6 experimental repeats. Asterisks represent a P value  $< 0.05$  (ANOVA) compared to pUC19 transfection control.

Supplementary Fig. S13. Effects of dCas9 methyltransferases on mCherry expression in fluorescence reporter cell lines. (a) Schematic illustration of the mCherry fluorescence transgene expression cassette. The target sites of the gRNAs within the CMV promoter are indicated by red arrows (5'-3', targeting sense or antisense strands). (b) Southern blot analysis of five cell clones with the transgene cassette randomly and stably integrated into the genome. (c) Flow cytometry-based analysis of the percentage of BFP positive cells in the fluorescence reporter cells at 2, 5, 8 and 14 days after transient transfection with CMV gRNAs (T1-T3) and dCas9-BFP-DNMT1, dCas9-BFP-DNMT3A, dCas9-BFP-DNMT3B, or dCas9-BFP-EGFP. (d-h) % mCherry fluorescence median intensity in these five clones at day 2, 5, 8, and 14 days following transient transfection with CMV gRNAs (T1-T3) and dCas9-BFP-DNMT1, dCas9-BFP-DNMT3A, dCas9-BFP-DNMT3B,

or dCas9-BFP-EGFP. Control cells were transfected with pUC19. Percent inhibition of mCherry expression was calculated by normalizing the median mCherry fluorescence intensity to that from the pUC19 transfected cells at each time point. Figures are plotted using the mean % mCherry median  $\pm$  SD ( $n = 3$ , independent transfections). ANOVA with Bonferroni comparison was performed for cell clone 2. "a," "b," "c," and "d," indicates a P-value  $< 0.05$  compared to the pUC19 control for the corresponding transfection group.

Supplementary Fig. S14. Validation of hypermethylated DMRs and DHS methylation caused by dCas9-BFP-DNMT3A and uPA gRNAs by WGBS. (a) Box plot of % methylation level of hypermethylated DMRs dCas9 found in previous WGBS experiment. The WGBS in the repeat experiment was conducted as described in the methods sections, but cells were not FACS enriched and sequenced in lower depth than previous experiment. (b) Metaplot of average CpG methylation levels in 58 494 DNase I hypersensitive sites (DHS) and 1 kb upstream and downstream flanking regions. P value represents Wilcoxon matched pairs signed rank test between treated and control groups.

Supplementary Table S1. List of plasmids deposited to Addgene, qPCR primers, gRNA sequences, bisulfite PCR primers, bisulfite pyrosequencing primers, and DNA regions analyzed for methylation.

Supplementary Table S2. List of hypermethylated DMRs caused by dCas9 methyltransferases and uPA gRNAs

Supplementary Table S3. List of hypomethylated DMRs caused by dCas9 methyltransferases and uPA gRNAs

Supplementary Table S4. List of hypermethylated DMRs caused by dCas9 methyltransferases and TGFBR3 gRNAs

Supplementary Table S5. List of hypomethylated DMRs caused by dCas9 methyltransferases and TGFBR3 gRNAs

Supplementary Table S6. List of binding peaks caused by dCas9-BFP-DNMT3A and uPA gRNAs

Supplementary File 1. Extended discussion and results.

## Competing interests

All authors report no competing interests.

## Abbreviations

BFP: blue fluorescent protein; Cas9: CRISPR-associated protein 9; CGI: CpG island; CRISPR: clustered regularly interspaced short palindromic repeats; dCas9: nuclease deficient Cas9 or dead Cas9; DEG: differentially expressed gene; DHS: DNase I hypersensitivity sites; DMR: differentially methylated region; FACS: fluorescence-activated cell sorting; FBS: fetal bovine serum; gRNA: guide RNA; PAM: protospacer adjacent motif; PBS: phosphate-buffered saline; qPCR: quantitative polymerase chain reaction; PI: propidium iodide; TALE: transcription-activator-like effectors; TGFBR3: transforming growth factor beta receptor 3; UTR: untranslated region; WGBS: whole-genome bisulfite sequencing; ZF: zinc finger protein.

## Author contributions

L.L., L.B., and Y.L. conceived the idea. H.Y., J.W., L.B., X.L., X.X., A.L.N., and Y.L. planned and oversaw the study. L.L., Y. Liu, F.X., J.H., T.F.D., T.S.P., B.H., L.Y., Q.Z., F.F., L.Y., S.L., K.T.J., L.F., E.S., and Y.L. performed experiments and analyzed the data. L.L., J.H., and Y.L. prepared the figures. L.L. and Y.L. drafted the manuscript, and all authors revised the manuscript.



## Acknowledgements

This work was partially supported by grants from the Danish Research Council for Independent Research (DFF-1337-00128 to Y.L.), the Sapere Aude Young Research Talent Prize (DFF-1335-00763A to Y.L.), the Innovation Fund Denmark (BrainStem, to Y.L.), and the Lundbeck Foundation (R173-2014-1105 to Y.L., R151-2013-14439 to L.B., and R219-2016-1375 to L.L.). A.L.N. was supported by the Toyota-Foundation and the Lundbeck Foundation. FACS was performed with help from Charlotte Christie Petersen and Anni Skovbo at the FACS Core Facility, Aarhus University, Denmark.

## References

- Jinek M, Chylinski K, Fonfara I et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;**337**:816–21.
- Mali P, Yang L, Esvelt KM et al. RNA-guided human genome engineering via Cas9. *Science* 2013;**339**:823–6.
- Jinek M, East A, Cheng A et al. RNA-programmed genome editing in human cells. *elife* 2013;**2**:e00471.
- Cong L, Ran FA, Cox D et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013;**339**:819–23.
- Vad-Nielsen Lin L, Bolund L, Nielsen AL, Luo Y. Golden-gate assembly of CRISPR gRNA expression array for simultaneously targeting multiple genes. *Cell Mol Life Sci* 2016;**73**:4315–25.
- Qi LS, Larson MH, Gilbert LA et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 2013;**152**:1173–83.
- Cheng AW, Wang H, Yang H et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res* 2013;**23**:1163–71.
- Gilbert LA, Larson MH, Morsut L et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 2013;**154**:442–51.
- Fujita T, Fujii H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem Biophys Res Commun* 2013;**439**:132–6.
- Chen B, Gilbert LA, Cimini BA et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 2013;**155**:1479–91.
- Kearns NA, Pham H, Tabak B et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* 2015;**12**:401–3.
- Hilton IB, D'Ippolito AM, Vockley CM et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 2015;**33**:510–7.
- Jones PA, Baylín SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 2002;**3**:415–28.
- Brena RM, Costello JF. Genome-epigenome interactions in cancer. *Hum Mol Genet* 2007;**16** Spec No 1:R96–R105.
- Vojta A, Dobrinic P, Tadic V et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* 2016;**44**(12):5615–28.
- McDonald JI, Celik H, Rois LE et al. Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. *Biology Open* 2016;**5**:866–74.
- Addgene Yonglun Luo Lab Plasmid page <https://www.addgene.org/Yonglun Luo/>. Accessed 1st January 2018
- Luo Lab CRISPR - Search sgRNA binding sites (U6) page <http://luolab.au.dk/views/gRNA.cgi>. Accessed 1st January 2018
- Luo Lab Website <http://luolab.au.dk/>. Accessed 1st January 2018
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>(-delta delta C(T))</sup> method. *Methods* 2001;**25**:402–8.
- Rohde C, Zhang Y, Reinhardt R et al. BISMAR - Fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics* 2010;**11**:230.
- Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* 2012;**131**:281–5.
- Xi Y, Li W. BSMAP: whole genome bisulfite sequence Mapping program. *BMC Bioinformatics* 2009;**10**:232.
- Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 2010;**11**:204–20.
- Pakneshan P, Szyf M, Farias-Eisner R et al. Reversal of the hypomethylation status of urokinase (uPA) promoter blocks breast cancer growth and metastasis. *J Biol Chem* 2004;**279**:31735–44.
- Gowher H, Jeltsch A. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. *J Biol Chem* 2002;**277**:20409–14.
- Margot JB, Aguirre-Arteta AM, Di Giacco BV et al. Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. *J Mol Biol* 2000;**297**:293–300.
- Fu Y, Foden JA, Khayter C et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 2013;**31**(9):822–6.
- Rivenbark AG, Stolzenburg S, Beltran AS et al. Epigenetic reprogramming of cancer cells via targeted DNA methylation. *Epigenetics* 2012;**7**:350–60.
- Yang AS, Estecio MR, Doshi K et al. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* 2004;**32**:38e–38.
- Landan G, Cohen NM, Mukamel Z et al. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat Genet* 2012;**44**:1207–14.
- Wu H, Xu T, Feng H et al. Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Res* 2015;**43**:e141.
- Kuscu C, Arslan S, Singh R et al. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* 2014;**32**:677–83.
- Wu X, Scott DA, Kriz AJ et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* 2014;**32**:670–6.
- Jensen KT, Floe L, Petersen TS et al. Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Lett* 2017;**591**(13):1892–1901.
- Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 1994;**2**:28–36.
- Liu XS, Wu H, Ji X et al. Editing DNA methylation in the mammalian genome. *Cell* 2016;**167**:233–247.e17.
- Bernstein DL, Le Lay JE, Ruano EG et al. TALE-mediated epigenetic suppression of CDKN2A increases replication

- in human fibroblasts. *J Clin Invest* 2015;125:1998–2006.
39. Meister GE, Chandrasegaran S, Ostermeier M. Heterodimeric DNA methyltransferases as a platform for creating designer zinc finger methyltransferases for targeted DNA methylation in cells. *Nucleic Acids Res* 2010;38:1749–59.
40. Stepper P, Kungulovski G, Jurkowska RZ et al. Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. *Nucleic Acids Res* 2016;45(4):1703–1713.
41. Suelves M, Carrio E, Nunez-Alvarez Y et al. DNA methylation dynamics in cellular commitment and differentiation. *Brief Funct Genomics* 2016;15:443–53.
42. O'Geen H, Henry IM, Bhakta MS et al. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res* 2015;43:3389–404.
43. Zhang XH, Tee LY, Wang XG et al. Off-target effects in CRISPR/Cas9-mediated genome engineering. *Molecular Therapy - Nucleic Acids* 2015;4:e264.
44. Duan J, Lu G, Xie Z et al. Genome-wide identification of CRISPR/Cas9 off-targets in human genome. *Cell Res* 2014;24:1009–12.
45. Kungulovski G, Nunna S, Thomas M et al. 2014:Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. *Epigenetics & Chromatin* 2015;8:12.
46. Lin L, Liu Y, Xu F et al. Supporting data for “Genome-wide determination of on-target and off-target characteristics for RNA-guided DNA Methylation by dCas9 methyltransferases” GigaScience Database 2018. <http://dx.doi.org/10.5524/100406>.