Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation

James I. McDonald1,*, Hamza Celik2,*, Lisa E. Rois3, Gregory Fishberger3, Tolison Fowler1, Ryan Rees3, Ashley Kramer2, Andrew Martens2, John R. Edwards1,† and Grant A. Challen2,4,‡

ABSTRACT

Advances in sequencing technology allow researchers to map genome-wide changes in DNA methylation in development and disease. However, there is a lack of experimental tools to site-specifically manipulate DNA methylation to discern the functional consequences. We developed a CRISPR/Cas9 DNA methyltransferase 3A (DNMT3A) fusion to induce DNA methylation at specific loci in the genome. We induced DNA methylation at up to 50% of alleles for targeted CpG dinucleotides. DNA methylation levels peaked within 50 bp of the short guide RNA (sgRNA) binding site and between pairs of sgRNAs. We used our approach to target methylation across the entire CpG island at the Cdkn2a promoter, three CpG dinucleotides at the Arf promoter, and the CpG island within the Cdkn1a promoter to decrease expression of the target gene. These tools permit mechanistic studies of DNA methylation and its role in guiding molecular processes that determine cellular fate.

KEY WORDS: CRISPR/Cas9-based system, CpG dinucleotides, DNA methylation

INTRODUCTION

DNA methylation of CpG dinucleotides is a prominent epigenetic modification of the mammalian genome that can influence gene expression, and aberrant distribution of DNA methylation is associated with a spectrum of human disorders including cancers (Egger et al., 2004). Despite intensive study, it remains unclear which CpG dinucleotides must change methylation state in order to alter transcription. Genome-wide analyses have found associations between DNA methylation and reduced gene expression (Hinshelwood et al., 2009). From a literature search, we also been suggested that DNA methylation occurs after the loss of expression, and that expression changes can alter DNA methylation patterns (Bestor et al., 2015; Busslinger et al., 1983). Here, we sought to develop tools for locus-specific epigenetic remodeling to directly address the role of DNA methylation in regulating gene expression.

Targeted DNA methylation approaches have been attempted by fusing DNA methyltransferase enzymes (DNMTs) to DNA-binding proteins such as zinc finger proteins (ZFPs) (Siddique et al., 2013), and transcriptional activator-like effector (TALE) (Bernstein et al., 2015). However, engineering custom proteins for each targeted sequence is laborious and requires specialized expertise. Moreover, in these studies, induced DNA methylation of the targeted loci was relatively poor, with substantial off-target activity. An engineered form of the clustered, regularly interspaced, short palindromic repeat (CRISPR) system has emerged as an alternative for achieving site-specific DNA targeting (Jinek et al., 2012). Here, the Cas9 endonuclease is directed to genomic targets by engineered short guide RNAs (sgRNAs) (Jinek et al., 2012). Because the sgRNA is the DNA sequence-specific component of the system, it allows for efficient targeting of multiple regions due to the ease of design and synthesis of new sgRNAs (relative to engineering new custom proteins for each target site). A Cas9 mutant (D10A and H840A); henceforth referred to as dCas9) that lacks endonuclease activity but can still be recruited by sgRNA(s) (Jinek et al., 2012) has recently been used to target genes in mammalian cells for transcriptional activation (Perez-Pinera et al., 2013a,b; Maeder et al., 2013; Mali et al., 2013). Here, we demonstrate an easily reprogrammable CRISPR/dCas9 DNMT fusion capable of inducing site-specific DNA methylation.

RESULTS

To design a flexible system to target DNA methylation, we fused dCas9 to the catalytic domain of the de novo DNA methyltransferase DNMT3A (Fig. 1A). To test this system we targeted DNA methylation to the tumor suppressor gene Cdkn2a (cyclin dependent kinase 2A), which inhibits progression through the cell cycle (Liggett and Sidransky, 1998). Cdkn2a is one of the most frequently hypermethylated genes in The Cancer Genome Atlas (Ciriello et al., 2013), and numerous clinical studies show a negative correlation between Cdkn2a methylation and expression in colorectal cancer (Shima et al., 2011). While it is generally assumed that Cdkn2a methylation induces gene silencing, it has also been suggested that DNA methylation occurs after the loss of expression (Hinshelwood et al., 2009). From a literature search, we identified 17 publications that associate Cdkn2a methylation with expression and/or cancer (Fig. S1.1). Overwhelmingly, these papers studied the differentially methylated region (cancer DMR, cDMR) on the 3′ end of the CpG island that overlapped the first exon of Cdkn2a (Fig. 1B).
We computationally designed three sgRNAs (g1a, g7a and g33a) to target this region and test whether DNA methylation was sufficient to induce gene silencing. We validated the ability of each sgRNA to target the CDKN2A locus by transfecting them with active CRISPR/Cas9 and measuring the ability of Cas9 to cleave the locus (Fig. S1.2, Table S1). We then transfected HEK293T cells with the pool of three sgRNAs along with either a normal dCas9-DNMT3A catalytic domain (CD) fusion (dCas9-D3A) or one with a DNMT3AE756A mutation (dCas9-mD3A, Fig. 1A), which abolishes DNA methyltransferase activity (Reither et al., 2003). Transfection efficiencies were >80-90% for all experiments as measured by co-transfection with GFP-containing plasmids. We analyzed DNA methylation levels for 20 days post-transfection using Illumina sequencing of two amplicon regions (amplicon bisulfite sequencing, ABS). ABS results were validated using Sanger bisulfite sequencing (R2=0.83; Fig. S1.3), and DNA methylation levels at CpGs analyzed in both of two independent ABS amplicons showed strong correspondence (R2=0.98; Fig. 1C). All CpGs had >100× sequencing coverage, with a median coverage of 15,200.

Over the 20-day time course we observed an increase in DNA methylation at the CDKN2A target locus that ranged from 20-43% at its peak on day three. Background methylation from transfection with dCas9-mD3A was consistently less than 1.5%, while background methylation from an off-target sgRNA, which controls for DNMT3A-CD overexpression, was less than 14% on day three. Induced DNA methylation levels were highest over a set of eight CpGs directly between the g33a and g7a sgRNA target sites (Fig. 1C). Increases in CHG and CHH methylation were minimal (Fig. S1.4). DNA methylation decreased rapidly after passaging the cells on day four, but stabilized 20 days post-transfection at 6-10% (Fig. 1D; Fig. S1.5). Despite the literature support for a negative
correlation between expression and DNA methylation in this region, we did not observe a measurable effect on CDKN2A gene expression by RT-qPCR (Fig. 1F). This suggests that a limited increase in methylation in the region 100-400 bp downstream of the CDKN2A TSS is insufficient to trigger gene silencing.

Spatially, the induced DNA methylation spiked near the sgRNA target sites and dropped quickly toward background levels at surrounding CpG sites. Analysis of DNA methylation induced by single sgRNAs indicates that methylation occurs primarily within 50 bp of the sgRNA binding site (Fig. S1.6). Higher DNA methylation levels were often observed 3′ of the sgRNA binding site (Fig. S1.6). Our initial data from the three pooled sgRNAs suggested that CpG methylation was higher between pairs of sgRNAs. To investigate this effect, we transfected pairs of sgRNA with varying intervening distances and monitored methylation of six clustered CpGs between the sgRNA pairs using ABS (Fig. 1E). DNA methylation of the three CpGs (cluster 1) within 20 bp of the fixed sgRNA (g33a) did not change with addition of a second sgRNA 77 bp or further away. However, the methylation level increased from 30-40% to 42-53% when sgRNAs were paired within 80 bp and both sgRNAs were within 50 bp of CpG cluster 2 (Fig. 1E). This suggests the DNMT3A-CD activity at the target locus is additive.

We next tested whether we could use our approach to methylate an entire CpG-island (CGI). We designed 17 sgRNAs (Fig. 2A; Table S1) to target DNA methylation across the CDKN2A CGI, which spans the TSS. We applied three combinations of sgRNAs (Set 1, 2, All) to test whether inducing DNA methylation of the entire CGI could decrease gene expression (Fig. 2A). ABS analysis of eight amplicons (minimum per CpG sequencing depth of 100) showed that the DNA methylation level increased to an average of 22% across the entire region with a peak of 54% (Fig. 2B). As an off-target negative control, we used three sgRNAs targeted to the ARF promoter located ∼20 kb away. The average background methylation at CDKN2A after treatment with off-target ARF sgRNAs was 9% (Fig. 2B). The other two sgRNA sets (Set 2 and All) induced similar increases of methylation across the CGI overlapping the TSS of CDKN2A (Fig. S2).

Analysis of CDKN2A expression by RT-qPCR in all three sgRNA targeting experiments (Set 1, 2, All) indicated an average 39% decrease in CDKN2A mRNA expression after targeting with dCas9-D3A (Fig. 2C). Cells transfected with dCas9-mD3A showed a 16-26% reduction in CDKN2A expression, likely due to CRISPR inhibition (Fig. 2C). Across the three replicates (Set 1, 2, All) expression decreased by an average of 17% in dCas9-D3A relative to dCas9-mD3A (P<0.01 paired-t-test). This indicated that DNA methylation decreases gene expression at the CDKN2A promoter in a context dependent manner.

(A) Locations of the 17 sgRNAs used in Sets 1 and 2 and regions sequenced in B. sgRNA coordinates are in Table S1. (B) Percent methylation is plotted for regions designated in A. Set 1 sgRNA target sites are indicated above the graphs. Data for All sgRNA, Set 2 sgRNA, and off-target SgRNA is presented in Fig. S2. (C) Methylation induced by CDKN2A-targeted sgRNA Set 1, Set 2, and All sgRNA, decreases gene expression. Relative expression of CDKN2A is normalized to a mock-treated control. Error bars=mean±s.e.m. (All, n=2; Set 1, n=1; Set 2, n=1; all performed in technical triplicate, paired t-test).
methylation directly decreased CDKN2A expression, but targeting of the entire CGI was required to trigger this effect. Our results are consistent with other studies that find a similar reduction in gene expression after inducing methylation at the CDKN2A promoter using ZFP- and TALE-based systems (Bernstein et al., 2015; Cui et al., 2015).

To verify the effect of our system with a separate locus, we designed three inwardly-directed (5' to 3') sgRNAs that bracketed three CpG sites in the ARF promoter located 150-170 bp downstream of the TSS (Fig. 3A). ARF-targeted sgRNAs increased the DNA methylation level to 27-30% at these three CpG sites with less than 15% methylation induced in adjacent sites (Fig. 3B). Induced methylation of the ARF promoter was associated with a 19% decrease in its expression (Fig. 3C).

To further validate our design rules, we attempted to induce DNA methylation in a different system in a regulated fashion with a minimalistic combination of sgRNAs. dCas9-D3A and dCas9-mD3A were cloned into a doxycycline-inducible lentivirus where expression of the fusion protein is linked to mCherry via internal ribosome entry site (IRES), and targeted to the Cdkn1a locus with two rationally-designed sgRNAs. We designed two sgRNAs targeting the CpG island located in the Cdkn1a promoter (Fig. 4A) and cloned them into lentiviral vectors (in which sgRNAs and GFP are expressed constitutively from two different promoters) (Fig. S4.1) and transduced 32D cells (mouse IL3-dependent myeloid progenitor cell line with an unmethylated Cdkn1a promoter and active Cdkn1a expression). Additional negative controls were generated by transducing 32D cells with sgRNAs only, dCas9-D3A only, or sgRNAs plus dCas9-mD3A. Cells were then FACS-sorted by GFP and/or mCherry to establish stable cell lines. Analysis of the stable cell lines induced by doxycycline showed successful expression of dCas9-fusions and sgRNAs via mCherry and GFP signal (Fig. S4.2). In the absence of doxycycline, the expression of fusion constructs was silenced (Fig. S4.2).

Stable cells were induced with doxycycline for eight days. Cdkn1a promoter DNA methylation and gene expression were then evaluated by bisulfite sequencing and RT-qPCR respectively. Cells transduced with dCas9-D3A together with sgRNAs showed an increase in Cdkn1a promoter DNA methylation by over 25% across the entire target region (and by as much as 40-85% at individual CpGs around sgRNAs) compared to all control groups (Fig. 4B,C), and a reduction in Cdkn1a expression by approximately 40% (Fig. 4D). There was no difference in Cdkn1a promoter DNA methylation or expression between any control groups (Fig. 4B-D). Cells transduced with dCas9-D3A together with sgRNAs showed a growth advantage, consistent with an increase in proliferation due to repression of a cell cycle inhibitor (Fig. 4E).

We reverse cloned viral integrants to identify clones that contained either one of the two guides, or both sgRNAs together with dCas9-D3A. No induced promoter DNA methylation or reduction in gene expression was observed in clones expressing only a single sgRNA (Fig. 5A,B). In contrast, Cdkn1a promoter methylation was increased and its expression decreased in clones that expressed both sgRNAs (Fig. 5A,B). To confirm the presence of both sgRNAs were required to execute induced DNA methylation, clones that contained only a single sgRNA were transduced with their complementary sgRNA. Following this, after eight-days of induction, Cdkn1a DNA methylation increased across the promoter by up to 33% compared to the parental clone that contained only a single sgRNA (Fig. 5C). Cdkn1a expression was reduced by 40-50% in the complemented clones (Fig. 5D).
data indicate that targeting with multiple sgRNAs results in a synergism and induces robust DNA methylation. We also examined the stability of the induced DNA methylation by withdrawing doxycycline for eight days. Cdkn1a promoter DNA methylation and expression profiles remained largely unchanged (Fig. 5E,F) even though expression of dCas9-D3A was silenced (Fig. S4.2).

To assess non-specific activity in this system, we initially attempted to analyze off-target genomic areas with high homology to predicted binding regions of both Cdkn1a sgRNAs. However, no regions with high homology were found as these sgRNAs were designed based on their minimal off-target binding and the presence of a PAM-site. As an alternative, we measured the global DNA methylation levels between cells that contained fusion proteins of a PAM-site. As an alternative, we measured the global DNA methylation levels between cells that contained fusion proteins of a PAM-site. As an alternative, we measured the global DNA methylation levels between cells that contained fusion proteins of a PAM-site. Though modest, the expression decreases caused by induced methylation were sufficient to decrease expression of all three genes. Methylation increases and changes in expression were highly significant and reproducible either by using multiple distinct sgRNA combinations in the case of CDKN2A or at the clonal levels as observed for Cdkn1a. Moreover, the reduction in Cdkn1a expression clearly had functional consequences (increased proliferation) for the transduced cells. Though modest, the expression decreases caused by induced methylation were consistent with previously published results using ZFP and TALE fusions (Bernstein et al., 2015; Cui et al., 2015).

The effects of the induced methylation also appeared to be context dependent. While methylation of the entire CGI at the CDKN2A promoter repressed gene expression, inducing DNA methylation of a region 100-400 bp downstream of the CDKN2A TSS alone was insufficient to affect expression despite the frequent observation of a negative correlation between methylation and expression in this region. This indicates the importance of the flexibility to target multiple regions offered by our CRISPR/dCas9 DNMT fusion system.

Although approaches for targeted DNA methylation have been previously described, our method is advantageous for several reasons including: (1) the ease of designing new sgRNAs for targeting, (2)
higher levels of induced DNA methylation, (3) little off-target activity. This approach can be used to interrogate the effects of DNA methylation on only a few CpG sites by bracketing them with sgRNAs, or can be used to test the effects of broader increases in DNA methylation by using many sgRNAs simultaneously. Further, the method described here provides a robust, reprogrammable approach to allow researchers to easily and thoroughly explore the functional roles of DNA methylation changes in development and disease.

MATERIALS AND METHODS
dCas9 fusion protein design and construction
The catalytic domain (CD) of human DNMT3A (amino acids 598 to 912 of NP_783328.1) both with and without the E756A mutation was cloned between the NheI and AgeI sites of pCMV_dCas9_VP64 (Addgene plasmid #49015, Cambridge, MA) with a NLS and FLAG tag linker. For lentivirus plasmids, TetO-FUW-OSKMD (Addgene plasmid #20321), was digested with EcoRI and a multiple cloning site containing EcoRI, XbaI, NheI, AgeI, PspXI, Ascl and EcoRI was cloned into this plasmid to generate TetO-FUW plasmid. An IRES-mCherry insert was amplified by PCR and cloned into AgeI- and AscI-digested TetO-FUW to generate TetO-FUW-IRES-mCherry. dCas9-D3A or dCas9-mD3A was cloned into this plasmid using XbaI and AgeI sites to generate TetO-dCas9-D3A or TetO-dCas9-mD3A. In the final plasmid, expression of both fusion proteins is controlled from the tetracycline operator (tetOP) and a minimal CMV promoter. The full amino acid sequence of the dCas9-3a or dCas9-3aΔ protein is shown in Fig. S4.1. All plasmids are available in Addgene, and detailed protocol information is available at http://epigenomics.wustl.edu/epigenomeEditing and http://www.challennlab.com.

sgRNA design
Target sequences were entered into the MIT sgRNA design software (http://crispr.mit.edu/), the BROAD sgRNA design tool (http://www.broadinstitute.org/rna公开/analysis-tools/sgrna-design-v1) (Doench et al., 2014), and the sgRNAcas9 tool (version 2.0.10) (Xie et al., 2014). The intersection of sgRNA target sites produced by all tools was taken for further analysis. sgRNA sequences that failed the BROAD test (score>0.2) were excluded. sgRNA were selected based on high BROAD scores and location relative to other sgRNAs. sgRNA coordinates and sequences are in Tables S1 and S2. Oligonucleotides corresponding to the target sites were annealed and cloned into MLM3636 (Addgene plasmid #43860). For lentivirus plasmids, U6-sgRNA coding sequences were amplified from pMLM3636 by PCR and cloned into a EcoRI and Clal linearized plVTTHM lentiviral backbone (Addgene plasmid #12247) generating pl-pLsgRNA1 and pl-pLsgRNA2 in which GFP is expressed from EF1a promoter.

Cell culture, lentiviral production and doxycycline induction
HEK293T cells were acquired from ATCC (CRL-3216, Manassas, VA) and grown in DMEM supplemented with 10% FBS (Gibco, Waltham, MA), 1× Penicillin/Streptomycin (Gibco), and 2 mM GlutaMax (Gibco). For transfection experiments, 3×10^5 HEK293T cells were plated in a 60 mm dish. The next day, the cells were transfected with Lipofectamine LT (Thermo Fisher Scientific, Waltham, MA). The Lipofectamine-DNA ratio was 3.5, with a total of 5.5 μg of plasmid DNA. The mass of Cas9-DNMT3A-CD fusion plasmid was equal to the total mass of the CRISPR/RNA plasmids. Since HEK293T cells incorporate either all plasmids or none, 0.5 or 0.7 μg of pMaxGFP was co-transfected in order to indicate the transfection efficiency. The plasmid DNA was first diluted in Gibco OptiMEM, then Lipofectamine:DNA ratio P<0.05; ****P<0.0001, error bars=mean±s.e.m.
MiniPrep kit and quantified with the Qubit dsDNA broad range assay (Thermo Fisher Scientific, Waltham, MA). gDNA was bisulfite converted with the Zymo Research EZ DNA methylation kit or the Epitech Bisulfite Kit (Qiagen) according to the manufacturer’s instructions. All samples underwent bisulfite conversion with a high efficiency of at least 98% as determined by conversion of unmethylated, non-CpG cytosines. For **CDKN2A**, the target regions were amplified with the Qiagen PyroMark PCR kit and CDKN2A_B primers in Table S3. PCR products were cloned into the Promega pGEM-T Easy plasmid and transformed into NEB 10β competent cells. PCR products from individual colonies were sequenced by Sanger.

A nested PCR strategy was used to amplify **Cdkn1a** target genomic DNA site. In primary PCR reaction, 20 ng of total bisulfite converted DNA was used as a template with the primer pair F1 (GTTTAGATTTTTAGAGGGAGGG) and R1 (CAAAAACTAAAAATAACTACCATCC). PCR was performed using Taq DNA polymerase (Invitrogen) using following cycling conditions; (1) 94°C for 3 min, (2) 95°C for 45 s, (3) 58°C for 58°C, (4) 72°C for 45 s and (5) 72°C for 10 min. Steps 2-4 were repeated for 40 cycles. For the second part of the nested PCR, 1.0 µl of the primary PCR reaction was used as a template and the same PCR conditions were used with primer pairs F1 and R2 (TCTCTAATTCTACCTATATACC). The PCR amplicons were then gel extracted and cloned into pCR4-TOPO TA vector (Invitrogen). The PCR amplicons were purified over a 1.5% agarose II gel in 1× TBE in order to remove incompletely ligated DNA. Custom 13 bp barcode index sequences were added via PCR using NEB Phusion (NEB, Ipswich, MA). The standard Phusion PCR protocol was followed but with the following primer concentrations (for a 50 µl reaction): 1 µl of Illumina Primer 1.0 (25 µM), 1 µl of Illumina PCR Primer 2.0 (0.5 µM), and 1 µl of the index primer (25 µM). The thermocycler protocol was the following: (1) 98°C, 30 s; (2) 98°C, 10 s; (3) 64°C, 30 s; (4) 72°C, 30 s; (5) Return to step two 11 times; (6) 72°C 5 min; (7) 4°C hold. The PCR products were again purified over a 1.5% agarose II gel in 1× TBE, and their concentration was measured with the Qubit dsDNA high sensitivity kit. Individually indexed samples were pooled and submitted for sequencing.

Amplicon bisulfite sequencing data were checked for quality using fastQC, adaptor and poor quality sequence (quality less than 20) was trimmed using fqtrim, and the trimmed sequences were mapped to the target sequences using Bismark (Krueger and Andrews, 2011).

**ELISA methylation analysis**

For quantification of total 5-mC level, the ELISA-based MethylFlash Methylated DNA Quantification Kit (Colorimetric; all from Epigentek, distributed by BioCat GmbH, Heidelberg, Germany) were utilized to quantify the amount of 5-mC in the DNA of transduced 32D cell lines. Quantification was performed in technical and biological duplicates according to the manufacturer’s instructions. 100 ng of total DNA isolated from each transfected cells was used for 5-mC quantification. For absolute 5-mC quantifications, a standard curve was generated by plotting the various concentrations of the positive control provided with the kit against the optical density (OD) at 450 nm.

**Expression analysis**

For human cells, RNA was extracted with the Zymo Research Quick mRNA Miniprep kit. RNA concentration was measured with the Qubit dsDNA broad range assay (Thermo Fisher Scientific, Waltham, MA). gDNA was bisulfite converted with the Zymo Research EZ DNA methylation kit or the Epitech Bisulfite Kit (Qiagen) according to the manufacturer’s instructions. All samples underwent bisulfite conversion with a high efficiency of at least 98% as determined by conversion of unmethylated, non-CpG cytosines. For **CDKN2A**, the target regions were amplified with the Qiagen PyroMark PCR kit and CDKN2A_B primers in Table S3. PCR products were cloned into the Promega pGEM-T Easy plasmid and transformed into NEB 10β competent cells. PCR products from individual colonies were sequenced by Sanger.

A nested PCR strategy was used to amplify **Cdkn1a** target genomic DNA site. In primary PCR reaction, 20 ng of total bisulfite converted DNA was used as a template with the primer pair F1 (GTTTAGATTTTTAGAGGGAGGG) and R1 (CAAAAACTAAAAATAACTACCATCC). PCR was performed using Taq DNA polymerase (Invitrogen) using following cycling conditions; (1) 94°C for 3 min, (2) 95°C for 45 s, (3) 58°C for 58°C, (4) 72°C for 45 s and (5) 72°C for 10 min. Steps 2-4 were repeated for 40 cycles. For the second part of the nested PCR, 1.0 µl of the primary PCR reaction was used as a template and the same PCR conditions were used with primer pairs F1 and R2 (TCTCTAATTCTACCTATATACC). The PCR amplicons were then gel extracted and cloned into pCR4-TOPO TA vector (Invitrogen). The PCR amplicons were purified over a 1.5% agarose II gel in 1× TBE in order to remove incompletely ligated DNA. Custom 13 bp barcode index sequences were added via PCR using NEB Phusion (NEB, Ipswich, MA). The standard Phusion PCR protocol was followed but with the following primer concentrations (for a 50 µl reaction): 1 µl of Illumina Primer 1.0 (25 µM), 1 µl of Illumina PCR Primer 2.0 (0.5 µM), and 1 µl of the index primer (25 µM). The thermocycler protocol was the following: (1) 98°C, 30 s; (2) 98°C, 10 s; (3) 64°C, 30 s; (4) 72°C, 30 s; (5) Return to step two 11 times; (6) 72°C 5 min; (7) 4°C hold. The PCR products were again purified over a 1.5% agarose II gel in 1× TBE, and their concentration was measured with the Qubit dsDNA high sensitivity kit. Individually indexed samples were pooled and submitted for sequencing.

Amplicon bisulfite sequencing data were checked for quality using fastQC, adaptor and poor quality sequence (quality less than 20) was trimmed using fqtrim, and the trimmed sequences were mapped to the target sequences using Bismark (Krueger and Andrews, 2011).

**ELISA methylation analysis**

For quantification of total 5-mC level, the ELISA-based MethylFlash Methylated DNA Quantification Kit (Colorimetric; all from Epigentek, distributed by BioCat GmbH, Heidelberg, Germany) were utilized to quantify the amount of 5-mC in the DNA of transduced 32D cell lines. Quantification was performed in technical and biological duplicates according to the manufacturer’s instructions. 100 ng of total DNA isolated from each transfected cells was used for 5-mC quantification. For absolute 5-mC quantifications, a standard curve was generated by plotting the various concentrations of the positive control provided with the kit against the optical density (OD) at 450 nm.

**Expression analysis**

For human cells, RNA was extracted with the Zymo Research Quick mRNA Miniprep kit. RNA concentration was measured with the Qubit dsDNA broad range assay (Thermo Fisher Scientific, Waltham, MA). gDNA was bisulfite converted with the Zymo Research EZ DNA methylation kit or the Epitech Bisulfite Kit (Qiagen) according to the manufacturer’s instructions. All samples underwent bisulfite conversion with a high efficiency of at least 98% as determined by conversion of unmethylated, non-CpG cytosines. For **CDKN2A**, the target regions were amplified with the Qiagen PyroMark PCR kit and CDKN2A_B primers in Table S3. PCR products were cloned into the Promega pGEM-T Easy plasmid and transformed into NEB 10β competent cells. PCR products from individual colonies were sequenced by Sanger.

A nested PCR strategy was used to amplify **Cdkn1a** target genomic DNA site. In primary PCR reaction, 20 ng of total bisulfite converted DNA was used as a template with the primer pair F1 (GTTTAGATTTTTAGAGGGAGGG) and R1 (CAAAAACTAAAAATAACTACCATCC). PCR was performed using Taq DNA polymerase (Invitrogen) using following cycling conditions; (1) 94°C for 3 min, (2) 95°C for 45 s, (3) 58°C for 58°C, (4) 72°C for 45 s and (5) 72°C for 10 min. Steps 2-4 were repeated for 40 cycles. For the second part of the nested PCR, 1.0 µl of the primary PCR reaction was used as a template and the same PCR conditions were used with primer pairs F1 and R2 (TCTCTAATTCTACCTATATACC). The PCR amplicons were then gel extracted and cloned into pCR4-TOPO TA vector (Invitrogen) and resulting plasmids were transformed into TOP10 cells (Invitrogen). The target region was PCR amplified from different clones and the PCR products were sequenced by the Beckman Coulter Genomics (Danvers, MA, USA). Sanger bisulfite sequencing analysis was performed using BIQ Analyzer (Bock et al., 2005).

**Amplicon bisulfite sequencing**

gDNA was extracted, bisulfite converted and PCR amplified as above. A tenfold molar excess of Illumina sequencing Y-adapter was then annealed to 100 ng of PyroMark PCR product (total Y-adapter mass varies with PCR product length) with NEB Quick Ligase for 15 min. The ligations were purified over a 1.5% agarose II gel in 1× TBE in order to remove incompletely ligated DNA. Custom 13 bp barcode index sequences were added via PCR using NEB Phusion (NEB, Ipswich, MA). The standard Phusion PCR protocol was followed but with the following primer concentrations (for a 50 µl reaction): 1 µl of Illumina Primer 1.0 (25 µM), 1 µl of Illumina PCR Primer 2.0 (0.5 µM), and 1 µl of the index primer (25 µM). The thermocycler protocol was the following: (1) 98°C, 30 s; (2) 98°C, 10 s; (3) 64°C, 30 s; (4) 72°C, 30 s; (5) Return to step two 11 times; (6) 72°C 5 min; (7) 4°C hold. The PCR products were again purified over a 1.5% agarose II gel in 1× TBE, and their concentration was measured with the Qubit dsDNA high sensitivity kit. Individually indexed samples were pooled and submitted for sequencing.

Amplicon bisulfite sequencing data were checked for quality using fastQC, adaptor and poor quality sequence (quality less than 20) was trimmed using fqtrim, and the trimmed sequences were mapped to the target sequences using Bismark (Krueger and Andrews, 2011).
RNA BR kit. RNA integrity was determined by visualizing rRNA bands using agarose gel electrophoresis. Reverse transcription was performed using the Bio-Rad iScript Reverse Transcriptase kit. Quantitative reverse-transcription polymerase chain reaction (RT-qPCR) was performed with the Bio-Rad iCycler Universal with SYBR Green reagent on an Applied Biosystems ViiA7 instrument. The thermocycler protocol was the following: (1) 95°C, 20 s; (2) 95°C, 3 s; (3) 60°C, 20 s; for 40 cycles. qPCR primers are listed in Table S4. A melt curve was performed to indicate there was no off-target amplification. Data was analyzed as described by Hellelums et al. (2007) using the geometric mean of ACTB, GAPDH, and RPL0 as an internal control. The All sgRNA sample represents data from two independent transfection experiments. The data for all remaining samples derives from technical replication using the same RNA sample. P-values were calculated with paired sample t-tests on the normalized levels of gene expression.

For mouse cells, total RNA was isolated from 5×10^6 FACSC-sorted cells using NucleoSpin RNA XS (Macherey-Nagel) and reverse transcribed with the SuperScript VILO kit (Life Technologies). cDNA samples were used for qPCR.

ΔΔCt

Hematology, Alex

Funding

analysis. A.K. helped with cell culture and off-target analysis. A.M. designed the methylation analyses. G.F. and R.R. performed cloning of target loci for sequencing J.I.M and H.C were equally responsible for cell culture, methylation and expression analyses. The authors declare no competing or financial interests.

Acknowledgements

The authors thank the Genome Engineering and iPSC Center at Washington University School of Medicine for production and validation of sgRNAs. We thank Jessica Hoslington-Lopez for sequencing assistance, and G.F. for critical reading of this manuscript. We thank the Alvin J. Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital for the use of the Siteman Flow Cytometry Core, which provided cell sorting and analysis services. The Siteman Cancer Center is supported in part by NCI Cancer Center Support Grant #P30 CA91842.

Competing interests

The authors declare no competing or financial interests.

Author contributions

J.L.M and H.C. were equally responsible for cell culture, methylation and expression experiments, data analysis, and writing the manuscript. L.E.R. performed methylation analyses and assisted with writing the manuscript. T.F. performed methylation analyses at E.F. and R.R. performed cloning of target loci for sequencing analysis. A.K. helped with cell culture and off-target analysis. A.M. designed the cloning of dCas9-D3A and dCas9-mD3A fusion proteins. J.R.E and G.A.C supervised the work, edited and approved the manuscript.

Funding

We acknowledge the following funding sources: National Institutes of Health [NIHDK R01DK102428], the Edward Mallinckrodt Jr Foundation, the American Society of Hematology, Alex’s Lemonade Stand Foundation for Childhood Cancer, the Children’s Discovery Institute, the Sidney Kimmel Foundation for Cancer Research and the V Foundation for Cancer Research (to G.A.C.), the Siteman Cancer Center, U.S. Department of Defense Congressionally Directed Medical Research Program for Breast Cancer [W81XWH-11-1-0041], and the National Institute of General Medicine Sciences [2R01GM108611, NLM R21LM11198] (to J.R.E.), and National Institutes of Health T32 CBM Training Grant [2T32GM007067-37] for pre-doctoral support to J.M.

Supplementary information

Supplementary information available online at http://bio.biologists.org/lookup/suppl? vid=10.1242/bio.019067/-/DC1

References


873
(p16) promoter methylation and loss of expression in 902 colorectal cancers: cohort study and literature review. *Int. J. Cancer* 128, 1080-1094.


Fig. S1.1. (a) UCSC genome browser view showing the regions of CDKN2A with either a reported relationship between methylation and expression or reported to be cancer-specific changes. The number beside each region corresponds to the reference in (b). (b) Table containing the hg19 coordinates of each sequencing region as well as the relevant reference.
References:


Merlo, Adrian, James G. Herman, Li Mao, Daniel J. Lee, Edward Gabrielson, Peter C. Burger, Stephen B. Baylin, and David Sidransky. 1995. “5′ CpG Island Methylation Is


**Fig. S1.2.** sgRNA target validation for g1a and g33a (a) and for g7a (b) using a mismatch detection assay. Cells were transfected with active CRISPR/Cas9 and the indicated sgRNA. The parental band is 655 for (a) and 755 for (b). Cleavage products are 384 and 326 for g7a (blue arrows), 384 and 271 for g1a (blue arrows), and 472 and 183 for g33a (red arrows). NHEJ frequencies are estimated at 28% for both g1a and g33a, and at 24% for g7a.
Fig. S1.3. Validation of the Amplicon Bisulfite Sequencing (ABS) method. ABS (a) and Sanger bisulfite sequencing (b) data for the red region in Figure 1c. For Sanger data, each line represents one read. Filled circles represent a methylated CpG; empty circles represent
an unmethylated CpG. The percent methylation of all CpGs in the green boxes is included on top of each box. (c) Scatter plot of the Sanger sequencing versus ABS data showing high correspondence between each method. (d) ABS sequencing of the same amplicon in (a) for genomic DNA artificially methylated using M.SssI at 0%, 50% and 100% levels.
Fig. S1.4. CHG methylation (a) and CHH methylation (b) after transfection of CDKN2A locus with three pooled sgRNA (g1a, g7a, and g33a). Methylation data is from the red region in Figure 1a. Peak methylation is observed at the same six CpG cluster where we observe peak CpG methylation (Figure 1c, near position 300 bp).
**Fig. S1.5.** Percent methylation for additional CpGs over the time course in Figure 1d. CpG locations indicated in each panel are relative to CDKN2A’s TSS as in Figure 1c.
Fig. S1.6. sgRNA directed methylation is strand specific and decreases after 50 bp. Arrows indicate the 5’ to 3’ direction of the sgRNA. (a) Percent methylation in each direction on either side of a sgRNA site for cells transfected with a single sgRNA. (b) Percent methylation relative to the sgRNA target site. Data comes from cells transfected with both g33a and the indicated sgRNA (Figure 1d). CpGs that fall within 50 bp of the g33a sgRNA are removed to avoid confounding effects from the second sgRNA. Methylation levels are background-subtracted using methylation levels from off-target sgRNAs as the background (Figure 2b).
Fig. S2. (a) Methylation at the CDKN2A promoter after treatment with the dCas9-D3A and either All or Set 2 CDKN2A sgRNA. (b) Methylation induced at the CDKN2A promoter after treatment with catalytically inactive dCas9-mD3A and Set 2 sgRNA. (c) Methylation induced at the CDKN2A promoter after treatment with sgRNA to the ARF or RASSF1A promoters as an off-target control. Sequencing of the RASSF1A locus is not shown, since it was highly methylated in all samples, including those with catalytically inactive DNMT3A.
**Fig. S4.1.** (a) Schematic representation of TetO-dCas9-3A or TetO-dCas9-mD3A vector. In this lentiviral plasmid, expression of both fusion proteins is controlled from the tetOP and a minimal CMV promoter. Schematic representation of lentiviral pL-sgRNA vector (bottom) in which a sgRNA is expressed from human U6 promoter and GFP is expressed from EF1a promoter. (b) Amino acid sequence of the dCas9-3A fusion protein used in this study. Region highlighted in yellow represents the amino acid sequence of dCas9, whereas the region shown in green encodes FLAG tag. Region shown in light blue encodes DNMT3A catalytic domain, in which the amino acid shown in red is the location of the point mutation (E765A) to inactivate the catalytic activity of DNMT3A. Purple is NLS. (c) Cdkn1a target DNA region (5’ to 3’) with CpGs are shown in red color. Regions highlighted in yellow show two sgRNA binding sites with CRISPR/Cas9 PAM recognition sites underlined.
Fig. S4.2. (a) FACS plots showing that 32D cells transduced with sgRNAs constitutively express GFP. dCas9 fusion proteins whose expression is linked to mCherry under the control of doxycycline inducible promoter are also successfully expressed in the presence of doxycycline. (b) Western blot showing dCas9 fusions are expressed in the presence of doxycycline.
doxycycline, and their expression is silenced upon its withdrawal. Fusion proteins were detected with an antibody specifically recognizing Cas9 protein. (c) Bisulfite sequencing analysis of Cdkn1a target locus transduced with either one of the two sgRNAs (top two rows) or both (bottom two rows). (n=12-15 biological replicates representative of two independent experiments). (d) DNA methylation analysis of Cdkn1a target locus in 32D cells with one sgRNA which were then transduced with their complementary sgRNA. (n=12-15 biological replicates representative of two independent experiments). (e) DNA methylation analysis of Cdkn1a locus followed deinduction for eight days (n=12-16 biological replicates representative of two independent experiments). Source data for bisulfite sequencing in panels c, d, e in Table S5.
Fig. S6. Schematic representation of off-target CpG islands (indicated with green boxes) for the genes (a) Cdkn1a (off-target), (b) Srsf3, (c) Cdkn1b, and (d) Cdkn2d. Black boxes indicate each exon of indicated transcript (not to scale). Number of CpGs which are assayed for their methylation profile is indicated in each triangle for a given gene.
## Table S1. CDKN2A targeted sgRNAs.

<table>
<thead>
<tr>
<th>sgRNA</th>
<th>sgRNA Sequence</th>
<th>Genomic Coordinates (hg19)</th>
<th>3</th>
<th>sgRNA Pool</th>
<th>Subset 1</th>
<th>Subset 2</th>
<th>Distance Series</th>
</tr>
</thead>
<tbody>
<tr>
<td>g33A</td>
<td>GTGGCCAGCCAGTCGCGCGAAGG</td>
<td>chr9:21974774-21974796</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g1A</td>
<td>TGGGGCGGACCGCTGGCTCGG</td>
<td>chr9:21974873-21974895</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g7A</td>
<td>CTTTGCGTGAAGATACCGCGG</td>
<td>chr9:21974988-21975010</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>TCGAGAGATCTGTACCGCGG</td>
<td>chr9:21974449-21974471</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>CTGGGCTTTCCAGCCTCCAGGG</td>
<td>chr9:21974488-21974510</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>ACCCCTTCTGAAAACCTCCAGG</td>
<td>chr9:21974540-21974562</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>GTGGGGTAGGGTCTGCACGG</td>
<td>chr9:21974655-21974677</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>CCAACCGCAGCATAAGTTACGG</td>
<td>chr9:21974693-21974715</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>GGCACCCCTGAAAGTCCAGGG</td>
<td>chr9:21975070-21975092</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A7</td>
<td>CTGGGAGCAGGGGAGGGCCAGG</td>
<td>chr9:21975193-21975215</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A8</td>
<td>CGGAGGAAGGAAAGGGGAGGGG</td>
<td>chr9:21975282-21975304</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A9</td>
<td>TCGGGGTGTGTTGTCATAGGG</td>
<td>chr9:21975335-21975357</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A10</td>
<td>GTGTATATTCGATTTCGCTGG</td>
<td>chr9:21975421-21975443</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A11</td>
<td>CTGAAAATACAGGCTGAGGGG</td>
<td>chr9:21975469-21975491</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A12</td>
<td>GAGCAGTCCGACTTCCAAGGG</td>
<td>chr9:21975535-21975557</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A13</td>
<td>GCGTGTAAACTCCACGGTCG</td>
<td>chr9:21975584-21975606</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A14</td>
<td>GAGCAGACTCTGAAAAGATGAGG</td>
<td>chr9:21975643-21975665</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A15</td>
<td>GGAGGCCTGAGGCAGTGCTCGG</td>
<td>chr9:21975011-21975033</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A16</td>
<td>TCCAGAGGATTTGAGGGACAGG</td>
<td>chr9:21974961-21974983</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A17</td>
<td>GGGGCTTGCTGCTGCAGAGGG</td>
<td>chr9:21974896-21974918</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table S2. ARF and RASSF1A targeted sgRNA sequences.

<table>
<thead>
<tr>
<th>sgRNA</th>
<th>sgRNA Sequence</th>
<th>Genomic Coordinates (hg19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARF_1</td>
<td>GCACGCAGCGCCGAGGGGAGG</td>
<td>chr9:21994286-21994308</td>
</tr>
<tr>
<td>ARF_2</td>
<td>GTGTCGCAAGGGCGGCACGCG</td>
<td>chr9:21994363-21994385</td>
</tr>
<tr>
<td>ARF_3</td>
<td>AGAAACGCTCAGCGGCGGCGG</td>
<td>chr9:21994260-21994282</td>
</tr>
<tr>
<td>RASSF1A_1</td>
<td>GCACCCAGAGTTTCATGCGCG</td>
<td>chr3:50378369-50378391</td>
</tr>
<tr>
<td>RASSF1A_2</td>
<td>GAACACCGGTTATCTCCGGTGG</td>
<td>chr3:50378524-50378546</td>
</tr>
<tr>
<td>RASSF1A_3</td>
<td>TGTGAGGGGACGAGAGGGG</td>
<td>chr3:50378461-50378483</td>
</tr>
</tbody>
</table>

### Table S3. Amplicons for bisulfite sequencing.

<table>
<thead>
<tr>
<th>Amplicon</th>
<th>Genomic Coordinate</th>
<th>Amplicon Length (bp)</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDKN2A_B</td>
<td>chr9:21974899-21975181</td>
<td>283</td>
<td>F: GTAGGTGAGGGAGGATTTAGTT&lt;br&gt;R: TCTAAACCAACCCCTCC</td>
</tr>
<tr>
<td>CDKN2A_A</td>
<td>chr9:21974594-21974985</td>
<td>392</td>
<td>F: TTTTTAGAGGTAGGAGGATAGG&lt;br&gt;R: CTACCTAATTCATCCCTACA</td>
</tr>
<tr>
<td>CDKN2A_1</td>
<td>chr9:21974140-21974555</td>
<td>215</td>
<td>F: TTGATTTTTAAAGTGAGGGATTATAAGG&lt;br&gt;R: ATTTTCAAAAAATTTATAATCTACAAAGCTTCC</td>
</tr>
<tr>
<td>CDKN2A_2</td>
<td>chr9:21974523-21974920</td>
<td>398</td>
<td>F: GAGGGTTGTAGTTATAATTTTTTTTG&lt;br&gt;R: AAAAACTAATACATACCCAAAAATAAAAAC</td>
</tr>
<tr>
<td>CDKN2A_3</td>
<td>chr9:21974890-21975191</td>
<td>301</td>
<td>F: GTTTTATTTTTTTGATGATTAGTTAGTTTTT&lt;br&gt;R: ATATAAAAAACATATAAAAAACCAAT</td>
</tr>
<tr>
<td>CDKN2A_4</td>
<td>chr9:21975160-21975525</td>
<td>365</td>
<td>F: GATTGGGTTTTTTTTTATTGTTTTTTAT&lt;br&gt;R: ACTAAAAATTTCAAATCAAATCACCCC</td>
</tr>
<tr>
<td>CDKN2A_5</td>
<td>chr9:21975512-21975862</td>
<td>351</td>
<td>F: TAGAAATTTTTGTTAAGGATTTTTTGGAGAG&lt;br&gt;R: TAAAAATATTTAAAAACAAAAACAAATTTTAAAAAC</td>
</tr>
<tr>
<td>CDKN2A_6</td>
<td>chr9:21975634-21975976</td>
<td>343</td>
<td>F: TTGGTATAGAGTAGTTTTGAAAAGATGAGG&lt;br&gt;R: ATTAATACATCTTTAAACTCAAAACAACACC</td>
</tr>
<tr>
<td>ARF</td>
<td>chr9:21994178-21994434</td>
<td>257</td>
<td>F: TTGGGTTTTTTGTTAGTTAGGG&lt;br&gt;R: CCTCAATACATCACAC</td>
</tr>
</tbody>
</table>
Table S4. qPCR primers for expression measurements.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>ACATCGCTCAGACACCATG</td>
<td>TGTTGAGGTCAATGAAAGG</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>GCTGCCCAACGCACCGAATAGTTA</td>
<td>TGCACGCGTGGGTGAGAG</td>
</tr>
<tr>
<td>ARF</td>
<td>GATGCTACTGAGGAGCCAGCTCTA</td>
<td>GCACAGCTCCGCCTCCTC</td>
</tr>
<tr>
<td>ACTB</td>
<td>CGACAGGATGCAGAAGGAG</td>
<td>TCCTGCTTGCTTGATCCACT</td>
</tr>
<tr>
<td>RPL0</td>
<td>AGACTGGAGACAAGTGGGA</td>
<td>CAGACAGACACTGGCAACA</td>
</tr>
</tbody>
</table>

Table S5.

Click here to Download Table S5