Decorating chromatin for enhanced genome editing using CRISPR-Cas9

Evelyn Chen[a,b,1], Enrique Lin-Shiao[a,b,1,2], Mohammad Saffari Doost[a,b], and Jennifer A. Doudna[a-h,3]

[a]Department of Molecular and Cell Biology, University of California, Berkeley, CA USA 94720

[b]Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA USA 94720

[c]Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA USA 94720

[d]Department of Chemistry, University of California, Berkeley, Berkeley, CA USA 94720

[e]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA USA 94720

[f]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, Berkeley, CA USA 94720

[g]Gladstone Institutes, University of California, San Francisco, CA USA 94158

[h]Gladstone-UCSF Institute of Genomic Immunology, San Francisco, CA USA 94158

[1]E.C. and E.L.-S. contributed equally to this work.

[2]Present address: BridgeBio Pharma, Palo Alto, CA USA 94301

[3]Correspondence to: doudna@berkeley.edu

Classification: Biological Sciences - Genetics

Keywords: CRISPR-Cas9, genome editing, chromatin, epigenetics

Abstract

CRISPR-associated (Cas) enzymes have revolutionized biology by enabling RNA-guided genome editing. Homology-directed repair (HDR) in the presence of donor templates is currently the most versatile way to introduce precise edits following CRISPR-Cas-induced double-stranded DNA cuts, but HDR efficiency is generally low relative to end-joining pathways that lead to insertions and deletions (indels). We tested the hypothesis that HDR could be increased using a Cas9 construct fused to PRDM9, a chromatin remodeling factor that deposits histone methylations H3K4me3 and H3K36me3 shown to mediate homologous recombination in human cells. Our results show that the fusion protein contacts chromatin specifically at the Cas9 cut site in DNA to double the observed HDR efficiency and increase the HDR:indel ratio by 3-fold compared to that induced by Cas9 alone. HDR enhancement occurred in multiple cell lines with no increase in off-target genome editing. These findings underscore the importance of chromatin structure for the choice of DNA repair pathway during CRISPR-Cas genome editing and provide a new strategy to increase the efficiency of HDR.

Significance Statement

CRISPR-Cas-mediated homology-directed repair (HDR) enables precision genome editing for diverse research and clinical applications, but HDR efficiency is often low due to competing end-joining pathways. Here, we describe a simple strategy to influence DNA repair pathway choice and improve HDR efficiency by engineering CRISPR-Cas9-methyltransferase fusion proteins. This strategy highlights the impact of histone modifications on DNA repair following CRISPR-Cas-induced double-stranded breaks and adds to the CRISPR genome editing toolbox.

Introduction

CRISPR-Cas9 genome editing provides a transformative opportunity to study and treat a wide range of diseases (1–4). CRISPR enzymes generate RNA-guided double-stranded breaks (DSBs) in DNA that trigger targeted DNA repair by mechanisms including non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), and homology-directed repair (HDR) (5). NHEJ and MMEJ pathways lead to a heterogeneous mixture of insertions and deletions (indels), which have been harnessed to facilitate gene disruption. HDR in the presence of donor templates can produce precise sequence insertions, but its low efficiency relative to NHEJ and MMEJ repair remains a limitation of genome editing applications (6). Cas9-based methods including base editing (7, 8) and prime editing (9) enable targeted substitutions and small insertions without requiring a DSB. However, these approaches are limited to single nucleotide substitutions or insertions smaller than 50 bases and have also been associated with off-target RNA editing (10). Therefore, HDR continues to be the most versatile method for targeted substitutions and insertions.

Previous studies of the effects of chromatin on DNA repair outcomes following CRISPR-Cas9-induced DSBs have focused primarily on the differences between euchromatin (more accessible chromatin) and heterochromatin (tightly packaged and inaccessible chromatin) (11, 12). Heterochromatin temporarily inhibits targeting by CRISPR-Cas9 *in vivo*, but DNA repair pathway choice does not differ significantly between heterochromatin and euchromatin after the initial delay (13–15). To date, the role of site-specific histone modifications on DNA repair pathway choice following CRISPR-Cas9-induced DSBs has remained largely unexplored.

Trimethylation of H3K4 (H3K4me3) and H3K36 (H3K36me3) are histone modifications found to mediate homologous recombination in human cells (16–19). During meiotic recombination, PRDM9 deposits both H3K4me3 and H3K36me3 for ZCWPW1 recognition, enabling Spo11-catalyzed DSBs and DMC1-mediated homologous strand invasion and crossovers (Fig. S1A) (18–21). In somatic cells, SETD2 deposits H3K36me3 to recruit CtIP to DSB sites, promoting DNA end resection and RPA and RAD51 foci formation that stimulates accurate genome repair (Fig. S1A) (17). SETD2-mediated H3K36me3 deposition also enables V(D)J recombination in the adaptive immune system and broadly decorates transcriptionally active regions to ensure transcriptional fidelity (17, 22). While these histone marks have been thoroughly characterized, it remains unclear how H3K4me3 and H3K36me3 influence DNA repair pathway choice following DSBs induced by CRISPR-Cas9.

In this study, we investigated how pre-existing and newly deposited H3K4me3 and H3K36me3 affect the efficiency of CRISPR-Cas9-mediated HDR by engineering and testing PRDM9-Cas9 fusion proteins. PRDM9 is expressed exclusively in germ cells during meiosis, and previous work in mice demonstrated that it does not influence transcription, making it an attractive fusion partner for Cas9 editing tool development (23). We show that both endogenous and newly decorated chromatin influence DNA repair pathway choice. In particular, the presence of H3K4me3 and H3K36me3 favors DSB repair via HDR following CRISPR-Cas9 cuts. We explored the efficacy of PRDM9-Cas9-catalyzed genome editing across different cell types, demonstrating a doubling in HDR rate and a tripling in the HDR:indel ratio relative to that observed using Cas9 alone. These findings underscore the importance of chromatin modification state for DNA repair pathway choice during CRISPR-Cas9-mediated genome editing and provide a new strategy to enhance HDR efficiency.

Results

Endogenous chromatin architecture modulates DNA repair pathway choice

To assess whether endogenous chromatin architecture affects Cas9 activity and DNA repair pathway choice, we identified genomic loci with varying chromatin modifications based on publicly available ENCODE ChIP-seq data from human embryonic kidney 293 cells (HEK293T) (24). We selected nine target sites with differing levels of H3K4me3 and H3K36me3 enrichment, including disease-relevant sites *HBB* and *LDLR* that encode the beta-globin component of hemoglobin and the low-density lipoprotein receptor, respectively (Fig. 1A). We examined HDR frequencies at the selected sites by transfecting HEK293T cells with Cas9 and single guide RNA (sgRNA) expression plasmids along with a single-stranded oligodeoxynucleotide (ssODN) donor template encoding a point mutation upstream of the PAM motif adjacent to the sgRNA binding site at each locus (25). While absolute HDR frequencies did not correlate with endogenous histone modifications, HDR:indel ratios were significantly higher at sites enriched with both H3K4me3 and H3K36me3 histone methylations compared to unenriched sites (Fig. 1B). Of note, the HDR:indel ratio (0.77 ± 0.04) at site 5, which is decorated extensively with histone marks, was four-fold higher compared to sites 6-9, which lack H3K4me3 and H3K36me3 enrichment. These results suggest that endogenous H3K4me3 and H3K36me3 marks favor HDR over error-prone end-joining pathways following Cas9-induced DSBs compared to sites devoid of these marks.

CRISPR-Cas9-methyltransferase fusion proteins modify histones site-specifically

Based on the above findings and previous reports (14, 26), we hypothesized that newly deposited histone modifications might influence DNA repair pathway choice following Cas9-induced DNA cutting (Fig. 1C). To test this idea, we constructed genes encoding four chimeric proteins in which histone methyltransferases are fused at the N-terminus of Cas9 in the pCAGGS expression vector (Fig. 1D). The PRDM9-Cas9 fusion comprises the KRAB domain which recruits additional proteins to facilitate recombination, the PR/SET domain which catalyzes methyltransferase activity, and a post-SET single zinc finger (ZnF) (23). The N-terminal domains of PRDM9 may be important for mediating interactions with chromatin remodeling factor HELLS and forming a pioneer complex to open chromatin (27). We excluded the C-terminal ZnF array of PRDM9 to avoid its endogenous DNA binding activity (28). PRDM9dC-Cas9 is a truncated version lacking the post-SET ZnF involved in negative autoregulation of methyltransferase activity and is thus predicted to show higher methylation activity (Fig. 1D) (23). SETD2-Cas9 includes the SET and post-SET domains of SETD2 which deposit H3K36me3 marks (Fig. 1D). Additionally, we generated a fusion of Cas9 to SETMAR, a histone methyltransferase that deposits H3K36me2 marks shown to be important for NHEJ repair (29, 30) (Fig. 1D; Fig. S1A).

We performed Western blot analysis to determine the level of expression of each fusion compared to unmodified Cas9 in HEK293T cells. When we delivered equal amounts (500ng) of plasmids expressing either Cas9 or the fusion proteins, we observed slightly lower expression of each fusion compared to Cas9 (Fig. S1B). Notably, scaling the amounts of plasmids expressing the fusion proteins based on their size compared to unmodified Cas9 did not lead to increased

expression (Fig. S1B). We reasoned that the lower expression of fusion constructs could be due to the larger protein sizes or decreased stability.

We then investigated whether the Cas9 fusion proteins deposit histone modifications in a site-specific manner. We selected sites 7 (intergenic) and 9 (exon 5 of *LDLR*) and determined H3K4me3 and H3K36me3 enrichment by chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR). To eliminate the potential effects of DSB remodeling on chromatin modifications, we engineered fusion proteins of nuclease-dead Cas9 (dCas9) to each histone methyltransferase described above. We then performed ChIP using HEK293T cells transfected with either dCas9 or one of the dCas9 fusion proteins and sgRNAs targeting site 7 or 9. By qPCR with primers flanking each target site or respective control genomic sites ~20 kilobases downstream, we determined the presence of H3K4me3 and H3K36me3 three days post-transfection.

Strikingly, PRDM9-dCas9 and PRDM9dC-dCas9 improved H3K4me3 and H3K36me3 enrichment at site 7 by at least 88-fold compared to dCas9 without modifying the downstream site, indicating that the fusion proteins specifically deposited histone marks at the Cas9-targeted site (Fig. 1E). The PRDM9 fusion proteins also effectively deposited H3K4me3 marks at site 9 (100-fold enrichment by PRDM9-dCas9 and 35-fold enrichment by PRDM9dC-dCas9), although they only increased H3K36me3 by up to 4.8-fold (Fig. S1C). SETD2-dCas9 increased H3K36me3 by 30-fold at site 7 and 16-fold at site 9 (Fig. 1E; Fig. S1C), while SETMAR-dCas9 only achieved a 1.7-fold increase in H3K36me2 at site 7 and a 1.3-fold increase at site 9 compared to dCas9 (Fig. 1E; Fig. S1C). It is possible that the truncated versions of SETD2 and SETMAR in the fusion designs lack domains that play additional roles in catalyzing methylation.

CRISPR-Cas9-methyltransferase fusion proteins display higher HDR and HDR:indel ratios

We next evaluated the editing outcomes of the fusion constructs via a BFP-to-GFP conversion assay in HEK293T cells stably expressing a BFP reporter (31). Cells were transfected with plasmids encoding either Cas9 or one of the four fusion proteins and a sgRNA targeting the BFP gene, as well as a ssODN template encoding a three-nucleotide (nt) change. Cells express GFP if the Cas9-induced DSB is repaired via HDR; on the other hand, cells lose BFP expression if the cut is repaired via end-joining pathways leading to indel formation (Fig. 2A). BFP-/GFP+ (HDR) and BFP-/GFP- (indel) cells were gated relative to cells transfected with a non-targeting control (NTC) sgRNA (Fig. 2B).

Using flow cytometry seven days post-transfection, we observed that in the absence of HDR templates, the Cas9 fusion proteins induced modestly reduced genome edits relative to those induced by unmodified Cas9, except for PRDM9dC-Cas9 which displayed a 50% decrease in editing efficiency (Fig. 2C). This could result from reduced nuclease activity of the Cas9 fusion proteins or increased deposition of histone modifications that improved the rate of perfect DNA repair following Cas9-catalyzed cutting. When ssODN templates were co-introduced, PRDM9 and SETD2 fusion proteins greatly increased HDR frequency and HDR:indel ratios (Fig. 2D-E). Strikingly, PRDM9-Cas9 achieved a 15.6 ± 0.8% HDR knock-in efficiency compared to 9.6 ± 0.5% with unmodified Cas9 (Fig. 2D). Additionally, PRDM9-Cas9 displayed a nearly three-fold higher HDR:indel ratio (0.3 ± 0.02) compared to Cas9 (0.1 ± 0.01)

(Fig. 2E). Together, these data provide evidence that the Cas9-methyltransferase fusion proteins improve HDR-mediated DNA repair while decreasing indel formation.

PRDM9-Cas9 fusion proteins enhance HDR:indel ratios at endogenous genomic loci and in different cell types

We examined the ability of the Cas9 fusion proteins to drive HDR at endogenous genomic sites. To this end, we transfected HEK293T cells with Cas9 and sgRNA expression plasmids and ssODN templates and measured the extent of genome modification by next-generation sequencing (NGS). PRDM9-Cas9, PRDM9dC-Cas9, and SETD2-Cas9 all produced higher HDR levels than did unmodified Cas9 at both site 7 and 9 (Fig. 3A-B). PRDM9-Cas9 displayed HDR frequencies of 11 ± 0.2% and 9.2 ± 0.6% at sites 7 and 9, compared to 5.8 ± 0.5% and 5.6 ± 0.2% with Cas9 at the respective sites (Fig. 3A-B). Moreover, PRDM9-Cas9 and SETD2-Cas9 improved HDR:indel ratios by 3-fold or more compared to unmodified Cas9 (Fig. 3A-B).

Since PRDM9-Cas9 generated the highest HDR frequency at both sites, we tested this fusion construct at additional endogenous genomic loci using suitable sgRNAs. PRDM9-Cas9-induced HDR frequencies were up to two-fold higher than those induced by Cas9 at sites 6-9, which lacked endogenous histone modifications (Fig. 3C). However, the fusion did not enhance HDR efficiencies at two additional sites 10 and 11 (exon in *SERPINA1*) (Fig. 3C). Additionally, although PRDM9-Cas9 showed higher HDR:indel ratios compared to Cas9 at sites 6-9, only modest improvements in HDR:indel ratios were observed at sites 10 and 11 (Fig. 3C). Interestingly, these two sites did not show endogenous H3K4me3 or H3K36me3 enrichment, suggesting that editing efficiency may be influenced by other histone modifications and factors beyond chromatin architecture. In particular, a previous study reported that genomic sites with high endogenous H4K16 acetylation enrichment are associated with increased HDR efficiency (32).

Given that the Cas9 fusion constructs can modify adjacent chromatin, we then examined whether this could lead to higher off-target editing by evaluating editing activity at seven potential off-target sites associated with two well-characterized sgRNA sequences (33). Importantly, PRDM9-Cas9 does not lead to higher off-target editing compared to Cas9 (Fig. 3D). Together, these data highlight the ability of PRDM9-Cas9 to improve Cas9-mediated HDR efficiency and the HDR:indel ratio via *de novo* modifications of chromatin architecture without increasing off-target effects.

Next, we compared PRDM9-Cas9- and Cas9-induced HDR efficiency using ssODN templates that either include or lack a mutation at the PAM site. Previous reports have shown that ssODNs that introduce a blocking mutation at the PAM significantly increase HDR efficiency by preventing the retargeting of the edited site (34). We designed ssODN templates that either disrupt or retain the PAM sequence at sites 7, 9, and 10. Consistent with previous reports, unmodified Cas9 showed two-fold higher HDR efficiency at sites 7 and 9 using ssODNs with PAM mutations (10.6% ± 0.3% and 12.3% ± 0.5%) compared to ssODNs without PAM mutations (5.8% ± 0.5% and 5.6% ± 0.2%) although no significant difference was observed at site 10 (Fig. 4A) (34). Of note, our PRDM9-Cas9 fusion approach alone achieved similar improvements in HDR efficiency compared to unmodified Cas9 using ssODN templates with PAM mutations (Fig. 4A). PRDM9-Cas9 increased the HDR:indel ratio more significantly than the PAM mutation strategy (Fig. 4B). Taken together, our findings suggest that PRDM9-Cas9

can be utilized to improve HDR efficiency effectively without the need to introduce PAM mutations.

It has been previously reported that individual Cas9 gRNAs lead to distinct and predictable indel patterns resulting from NHEJ and MMEJ repair of Cas9-catalyzed DSBs (35–38). NHEJ primarily leads to small indels, whereas MMEJ leads to deletions associated with microhomologies of at least 2bp flanking the DSB site (6, 39, 40). More specifically, the identity of the position 4 nucleotide before the PAM has been reported to influence indel patterns, with a guanine (G) at position 4 preferentially favoring the formation of primarily deletions (37). Furthermore, the frequency of MMEJ-mediated deletions has been shown to predict HDR efficiency at a given Cas9-targeted site (38). We therefore investigated whether position 4 nucleotide identity predicts indel patterns at sites 6-11 in HEK293T cells transfected with either Cas9 or PRDM9-Cas9 without a donor template. Based on previous work, we defined NHEJ as deletions of less than 3bp and insertions, and MMEJ as deletions of greater than or equal to 3bp (38). Consistent with previous reports, the frequency of MMEJ-mediated deletions was higher than NHEJ repair at all three target sites with a G at position 4 (Fig. S1D) (36, 37). Importantly, PRDM9-Cas9-mediated editing in the absence of a donor template did not alter the indel patterns at each target site compared to Cas9.

To investigate whether PRDM9-Cas9 can improve HDR efficiency in other cell types, we then tested the fusion construct at sites 7-9 in HeLa and U2OS cells. We only observed a modest increase in HDR frequency by PRDM9-Cas9 in these cell types, which may be due to differences in histone modifications and DNA repair processes that are cell type-specific (Fig. 4C-D). Nonetheless, PRDM9-Cas9 enhanced HDR:indel ratios by up to two-fold across all three target sites in both HeLa and U2OS cells (Fig. 4C-D). Overall, PRDM9-Cas9 improves the frequency of HDR relative to indels at multiple endogenous sites across different mammalian cell lines.

Discussion

HDR-mediated precision genome editing holds great potential for research and clinical applications, but HDR efficiency is often low and considerably variable across different genomic loci and cell types. Here, we report a simple strategy to influence DNA repair pathway choice and improve HDR efficiency by engineering CRISPR-Cas9-methyltransferase fusion proteins. These fusion proteins specifically decorated chromatin at target sites with H3K4me3 and H3K36me3, which have been reported previously to mediate homologous recombination during meiosis and somatic DNA repair (16–19). Among the four fusion constructs tested, PRDM9-Cas9 significantly improved HDR efficiency and HDR:indel ratio compared to Cas9 at multiple genomic loci without increasing off-target editing. A key advantage of PRDM9-Cas9 is that it can enhance HDR without influencing transcription, as previous work demonstrated that PRDM9 is not involved in transcriptional regulation (23).

While PRDM9-Cas9 achieved a clear improvement in HDR efficiency at target sites lacking endogenous H3K4me3 and H3K36me3 enrichment, the effect was not observed across all sites investigated. This observation may be due to other site-specific histone modifications and factors beyond chromatin architecture that affect Cas9-mediated editing efficiency at different loci and in different cell types. Importantly, recent studies reported on the importance of sgRNA design and CRISPR-Cas cut (blunt vs. staggered) on subsequent DNA repair pathway

choice (38). Our Cas9 fusion strategy can be combined with these findings and other approaches, such as asymmetric donor templates and tiling sgRNAs, to further improve HDR efficiency (31). Nonetheless, PRDM9-Cas9 consistently improved HDR:indel ratios compared to unmodified Cas9, making it a valuable tool for precise genome editing applications that demand a high HDR frequency relative to indel formation, such as gene correction in sickle cell disease (41). While the lower observed indel formation may be explained by increased DSB repair via sister chromatid or interhomolog homologous recombination, further work is required to elucidate the mechanisms involved.

Depositing specific chromatin modifications at the DNA cut site might also reduce chromothripsis, a phenomenon observed during DNA cleavage including CRISPR-Cas9-induced genome editing that leads to genome instability (42). Furthermore, the methylation activity of histone methyltransferase SETMAR has been shown to suppress chromosomal translocations, suggesting that the SETMAR-Cas9 fusion construct could be exploited for multiplex editing (43).

Overall, our findings suggest that both endogenous and newly deposited histone modifications influence DNA repair pathway choice during CRISPR-Cas9-mediated genome editing. The engineered fusion proteins described here provide a transient strategy to enhance HDR efficiency by decorating adjacent nucleosomes with H3K4me3 and H3K36me3, allowing the development of better editing tools for research and therapeutic approaches.

Materials and Methods

Dataset

We selected target sites based on narrow-peak calls for H3K4me3 and H3K36me3 in HEK293 cells. ChIP-seq datasets were obtained from the ENCODE portal with the following identifier: ENCSR372WXC (**https://www.encodeproject.org/reference-epigenomes/ENCSR372WXC/**). Aligned, non-duplicated pairs were used for peak calling using MACS2 (q value cutoff set to be 0.01). Only overlapping peaks shared by biological replicates were used to identify sites enriched for H3K4me3 and/or H3K36me3.

Plasmid construction

The Cas9 expression plasmid pCAGGS was a gift from Jennifer R. Hamilton. Cas9 was amplified from the expression plasmid using CloneAmp HiFi PCR Premix (Takara Bio) for 35 cycles (98□ for 10 s, 55□ for 15 s, and 72□ for 10 s; then 72□ for 1 min). PRDM9 and PRDM9dC were amplified from human cDNA purchased from GenScript (ORF Clone ID OHu03253). SETD2 was amplified from SETD2-GFP (Addgene plasmid # 80653; http://n2t.net/addgene:80653), and SETMAR was amplified from SETMAR (3BO5) (Addgene plasmid # 25250; http://n2t.net/addgene:25250). In-Fusion Cloning (Takara Bio) was used to clone PRDM9-Cas9, PRDM9dC-Cas9, SETD2-Cas9, and SETMAR-Cas9 into the pCAGGS expression vector. All Cas9 variants were confirmed by Sanger sequencing.

DNA oligonucleotides (IDT) for Cas9 sgRNA, including a non-targeting negative control sgRNA, were cloned into U6-sgRNA expression vectors (Table S1). Oligos were resuspended in nuclease-free water, and forward and reverse oligos were mixed with T4 polynucleotide kinase and 10X T4 DNA ligase buffer and placed in a thermal cycler for phosphorylation and annealing (37□ for 30 min; 95□ for 5 min; decrease temperature down to 25□ at 5□/min).

Annealed oligos were mixed with the linearized vector, T4 DNA Ligase, and 10X T4 DNA ligase buffer and incubated at 16°C overnight for ligation.

Mammalian cell line culture and lipofection

All cell lines (HEK293T, HeLa, U2OS, IMR90) were cultured in Dulbecco's Modified Eagle Medium (DMEM; Thermo Fisher) supplemented with 10% fetal bovine serum (FBS, VWR), and 1% Penicillin/Streptomycin (P/S, Gibco). All cells were cultured at 37ºC in a 5% CO2 air incubator. Routine checks for mycoplasma contamination were performed using the MycoAlert mycoplasma detection kit (Lonza). Lipofection was performed using Lipofectamine 3000 (Thermo Fisher Scientific) according to the manufacturer's instructions. 50,000 cells per well were seeded in 24-well plates 24 hours prior to lipofection. Cells were transfected with 500 ng Cas nuclease expression plasmid, 150 ng sgRNA expression plasmid, and 1.5 pmol ssODN templates (IDT, Table S2) per well.

Flow cytometry analysis

Cells were resuspended in FACS buffer (1% BSA in PBS) and analyzed by flow cytometry for BFP-negative cells (end-joining pathways) and GFP-positive cells (homology-directed repair pathway) seven days post-transfection. Flow cytometry was performed on an Attune NxT flow cytometer with a 96-well autosampler (Thermo Fisher Scientific). Data analysis was performed using the FlowJo v10 software.

Illumina deep sequencing analysis

DNA was extracted 72 hours post-transfection using QuickExtract DNA Extraction Solution (Lucigen) and heated at 65□ for 20 min followed by 95□ for 20 min. DNA samples were then amplified with PrimeSTAR GXL DNA Polymerase (Takara Bio) with PCR forward/reverse primers containing Illumina adapter sequences (Table S3) for 30 cycles (98□ for 10 s, 55□ for 15 s, and 68□ for 1 min).

The resulting amplicons were cleaned by adding 25 uL of amplicon to 45 uL of magnetic beads (UC Berkeley Sequencing Core). The samples were placed on a 96-well magnetic plate for 5 min, and the supernatant was removed. The samples were washed twice with 200 uL of 70% ethanol and eluted in 40 uL of Tris-EDTA Buffer (Corning).

The purified samples were sequenced on an Illumina iSeq by QB3 Genomics at UC Berkeley. NGS sequencing reads were analyzed for HDR-mediated modifications and indels using CRISPResso2 (https://crispresso.pinellolab.partners.org) in batch mode using default parameters.

Western blotting

Cells were resuspended in lysis buffer (20mM Tris pH 7.5, 1mM MgCl2, 1mM CaCl2, 137mM NaCl, 10% Glycerol, 1% NP-40) containing Benzonase (12.5 u/mL) and Protease Inhibitor Cocktail (Roche) and incubated at 4□ for 1 h with rotation. Proteins from whole cell lysates were separated by 4–20% TGX gel (Bio-Rad) and transferred to a nitrocellulose membrane (0.2µm pore size) (Invitrogen). Membranes were blocked and incubated with primary antibodies at 4°C overnight followed by secondary antibodies at room temperature for 1 h. Primary antibodies

used for Western blot analysis were FLAG (Sigma) and GAPDH (Cell Signaling Technology). Secondary Alexa Fluor antibodies were purchased from Thermo Fisher.
Imaging was performed using the Odyssey imaging system (LI-COR). Double bands were caused by incomplete cleavage at the self-cleaving P2A peptide, leading to an uncleaved byproduct (44)**.**

Chromatin immunoprecipitation-qPCR

Cells were crosslinked for 5 min in 1% formaldehyde, and the reaction was quenched by the addition of glycine to 125 mM and incubation for 5 min. Cells were washed twice with lysis buffer 1 (50mM HEPES-KOH pH 7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1x protease inhibitor) and lysis buffer 2 (10mM Tris-HCl pH 8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA, 1x protease inhibitors), respectively. Then, cells were resuspended in lysis buffer 3 (10mM Tris-HCl pH 8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine, 1x protease inhibitors) and sheared by sonication (peak power 140, duty factor 5%, 200 cycles per burst, 900 s treatment time) using a Covaris S2 (UC Berkeley Functional Genomics Laboratory). 500 ug lysate was incubated at 4□ overnight with 2 ug of appropriate antibodies (H3K4me3 (Abcam), H3K36me3 (Abcam), H3K36me2 (Diagonde), and IgG (Abcam)) pre-bound to 30 uL of Protein G Magnetic Beads (Thermo Fisher). The IP/bead mixture was washed 5 times with RIPA wash buffer (50mM HEPES-KOH pH 7.5, 500mM LiCl, 1mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate) and eluted from the beads for 30 min at 65 □. Both input and IP samples were reverse crosslinked by incubating at 65□ overnight with shaking. Then, samples were incubated at 37□ for 2 h with RNase A (0.2 mg/mL) and at 55□ for 2 h with Proteinase K (0.2 mg/mL). DNA samples were purified using 400 uL UltraPure™ Phenol:Chloroform:Isoamyl Alcohol (25:24:1) (Thermo Fisher).

Quantitative PCR was performed using QuantStudio 3 real-time PCR system (Thermo Fisher). Reactions were prepared in triplicates containing 2.5 uL of DNA sample, 100 nM of each primer (Table S4), 5 uL of Power SYBR Green Master Mix (Thermo Fisher). qPCR settings were: 50℃ for 2 min, 95℃ for 2 min, followed by 40 cycles of 95℃ for 1 min and 60℃ for 30 sec. The % of input DNA was calculated as ((Ct[Input] - log2(Input Dilution Factor)) − Ct[ChIP]) × 100%.

Author Contributions

This project was conceived by E.L.-S. and J.A.D. E.L.-S and E.C. initiated the study and designed and conducted all of the experiments. M.S.D. aided in molecular biology experiments. E.C., E.L.-S and J.A.D. wrote the manuscript. All authors reviewed and commented on the manuscript.

Competing Interest Statements

The Regents of the University of California have patents issued and pending for CRISPR technologies on which J.A.D. is an inventor. J.A.D. is a cofounder of Caribou Biosciences, Editas Medicine, Scribe Therapeutics, Intellia Therapeutics, and Mammoth Biosciences. J.A.D. is a scientific advisory board member of Vertex, Caribou Biosciences, Intellia Therapeutics, Scribe Therapeutics, Mammoth Biosciences, Algen Biotechnologies, Felix Biosciences, The Column Group, and Inari. J.A.D. is Chief Science Advisor to Sixth Street, a Director at Johnson

& Johnson, Altos and Tempus, and has research projects sponsored by Apple Tree Partners and Roche.

Research reported in this publication was supported by the Centers for Excellence in Genomic Science of the National Institutes of Health under award number RM1HG009490. E.L.-S was supported by the National Institute of General Medical Sciences under award number F32GM142146-01. J.A.D. is a Howard Hughes Medical Institute investigator.

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Information. Additional data related to this paper may be requested from the authors.

References

1. M. Jinek, *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
2. M. Jinek, *et al.*, RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
3. P. Mali, *et al.*, RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
4. L. Cong, *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
5. J. R. Chapman, M. R. G. Taylor, S. J. Boulton, Playing the end game: DNA double-strand break repair pathway choice. *Mol Cell* **47**, 497–510 (2012).
6. H. H. Y. Chang, N. R. Pannunzio, N. Adachi, M. R. Lieber, Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* **18**, 495–506 (2017).
7. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
8. N. M. Gaudelli, *et al.*, Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
9. A. V. Anzalone, *et al.*, Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
10. T. S. Nambiar, L. Baudrier, P. Billon, A. Ciccia, CRISPR-based genome editing through the lens of DNA repair. *Molecular Cell* **82**, 348–388 (2022).
11. X. Chen, *et al.*, Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res* **44**, 6482–6492 (2016).
12. R. M. Daer, J. P. Cutts, D. A. Brafman, K. A. Haynes, The impact of chromatin dynamics on Cas9-mediated genome editing in human cells. *ACS Synth Biol* **6**, 428–438 (2017).
13. M. A. Horlbeck, *et al.*, Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife* **5**, e12677 (2016).
14. R. M. Yarrington, S. Verma, S. Schwartz, J. K. Trautman, D. Carroll, Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *PNAS* **115**, 9351–9358 (2018).
15. E. M. Kallimasioti-Pazi, *et al.*, Heterochromatin delays CRISPR-Cas9 mutagenesis but does not influence the outcome of mutagenic DNA repair. *PLoS Biol* **16**, e2005595 (2018).
16. C.-C. Pai, *et al.*, A histone H3K36 chromatin switch coordinates DNA double-strand break repair pathway choice. *Nat Commun* **5**, 4091 (2014).
17. S. X. Pfister, *et al.*, SETD2-dependent histone H3K36 trimethylation Is required for homologous recombination repair and genome stability. *Cell Reports* **7**, 2006–2018 (2014).
18. K. Paigen, P. M. Petkov, PRDM9 and its role in genetic recombination. *Trends Genet* **34**, 291–300 (2018).
19. N. R. Powers, *et al.*, The meiotic recombination activator PRDM9 trimethylates both H3K36 and H3K4 at recombination hotspots in vivo. *PLoS Genet.* **12**, e1006146 (2016).
20. M. Mahgoub, *et al.*, Dual histone methyl reader ZCWPW1 facilitates repair of meiotic double strand breaks in male mice. *eLife* **9**, e53360 (2020).
21. D. Wells, *et al.*, ZCWPW1 is recruited to recombination hotspots by PRDM9 and is essential for meiotic double strand break repair. *eLife* **9**, e53392 (2020).
22. Z. Ji, *et al.*, The histone methyltransferase Setd2 is indispensable for V(D)J recombination. *Nat Commun* **10**, 3353 (2019).
23. S. Thibault-Sennett, *et al.*, Interrogating the functions of PRDM9 domains in meiosis. *Genetics* **209**, 475–487 (2018).
24. I. Dunham, *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
25. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Y. 2009 Almendros, Short motif

sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740.
26. I. Strohkendl, *et al.*, Inhibition of CRISPR-Cas12a DNA targeting by nucleosomes and chromatin. *Science Advances* **7**, eabd6030.
27. C. Spruce, *et al.*, HELLS and PRDM9 form a pioneer complex to open chromatin at meiotic recombination hot spots. *Genes Dev* **34**, 398–412 (2020).
28. Y. Striedner, *et al.*, The long zinc finger domain of PRDM9 forms a highly stable and long-lived complex with its DNA recognition sequence. *Chromosome Res* **25**, 155–172 (2017).
29. J. W. Edmunds, L. C. Mahadevan, A. L. Clayton, Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J* **27**, 406–420 (2008).
30. S. Fnu, *et al.*, Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. *Proc Natl Acad Sci U S A* **108**, 540–545 (2011).
31. C. D. Richardson, G. J. Ray, M. A. DeWitt, G. L. Curie, J. E. Corn, Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **34**, 339–344 (2016).
32. N. Horikoshi, *et al.*, Pre-existing H4K16ac levels in euchromatin drive DNA repair by homologous recombination in S-phase. *Commun Biol* **2**, 253 (2019).
33. S. Q. Tsai, *et al.*, GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187–197 (2015).
34. S. Okamoto, Y. Amaishi, I. Maki, T. Enoki, J. Mineno, Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs. *Sci Rep* **9**, 4811 (2019).
35. M. van Overbeek, *et al.*, DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol Cell* **63**, 633–646 (2016).
36. M. W. Shen, *et al.*, Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018).
37. A. Taheri-Ghahfarokhi, *et al.*, Decoding non-random mutational signatures at Cas9 targeted sites. *Nucleic Acids Research* **46**, 8417–8434 (2018).
38. K. J. Tatiossian, *et al.*, Rational selection of CRISPR-Cas9 guide RNAs for homology-directed genome editing. *Mol Ther* **29**, 1057–1069 (2021).
39. C. D. Yeh, C. D. Richardson, J. E. Corn, Advances in genome editing through control of DNA repair pathways. *Nat Cell Biol* **21**, 1468–1478 (2019).
40. L. Ranjha, S. M. Howard, P. Cejka, Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma* **127**, 187–214 (2018).
41. S. Pattabhi, *et al.*, In vivo outcome of homology-directed repair at the HBB gene in HSC using alternative donor template delivery methods. *Mol Ther Nucleic Acids* **17**, 277–288 (2019).
42. M. L. Leibowitz, *et al.*, Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat Genet* **53**, 895–905 (2021).
43. J. Wray, *et al.*, The transposase domain protein Metnase/SETMAR suppresses chromosomal translocations. *Cancer Genetics and Cytogenetics* **200**, 184–190 (2010).
44. J. H. Kim, *et al.*, High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLOS ONE* **6**, e18556 (2011).

Fig. 1: Endogenous histone modifications mediate DNA repair pathway choice.

A) UCSC genome browser tracks showing normalized endogenous H3K4me3 and H3K36me3 enrichment at target site ± 1.5 kilobases (kb) in HEK293T cells based on reanalyzed ENCODE ChIP-seq datasets. Each plot spans 3 kb.

B) HDR frequency and HDR:indel ratio measured by NGS at target sites in HEK293T cells transfected with plasmids expressing Cas9 and sgRNA along with a ssODN template with 50-nt homology arms. Data represents mean ± standard deviation (s.d.) (n = 2). *P < 0.05, determined by student's two-tailed t-test.

C) Schematic of four Cas9-methyltransferase fusion proteins designed to decorate chromatin to modulate the choice of DNA repair pathway following Cas9-induced DSBs. Histone methyltransferases are fused to Cas9 at the N-terminus.

D) Schematic of Cas9-histone lysine methyltransferase (KMT) fusion protein activity. Cas9 fusion protein is guided by a sgRNA (purple) to the DNA target site (dark blue), which may be embedded within a heterochromatic region. Cas9 fusion protein deposits histone marks (orange and red) at the target site, which influence the choice of DNA repair pathway following Cas9-induced DSBs.

E) H3K4me3, H3K36me3, and H3K36me2 enrichment shown as a percentage of input DNA measured by ChIP-qPCR at site 7 (intergenic) in HEK293T cells. Cells were transfected with plasmids expressing the dCas9 fusion proteins and a sgRNA targeting either site 7 or a control site ~20 kilobases downstream. Data represents mean ± s.d. (n = 2). *P < 0.05, determined by student's two-tailed t-test.

Fig. 2: Engineered CRISPR-Cas9-methyltransferase fusion proteins produce higher HDR and HDR:indel ratios relative to Cas9

A) Schematic of BFP-to-GFP reporter assay. In brief, BFP+/GFP- cells can be converted to BFP-/GFP+ cells via HDR or to BFP-/GFP- via indel formation.

B) Flow cytometry scatter plots for BFP-to-GFP reporter cells showing frequency of HDR (BFP-/GFP+) and indel (BFP-/GFP-). Cells were transfected with plasmids expressing the fusion proteins and sgRNA along with a ssODN template with a 91nt homology arm on the PAM-proximal side of the DSB and a 36nt homology arm on the PAM-distal side. A non-targeting negative control (NTC) sgRNA was included.

C) Editing activity of Cas9-methyltransferase fusion proteins in the absence of ssODN template measured by flow cytometry in BFP-to-GFP reporter cells transfected with plasmids encoding fusion protein and sgRNA.

D) HDR frequency measured by flow cytometry in BFP-to-GFP reporter cells transfected with plasmids encoding fusion protein and sgRNA, along with ssODN template.

E) HDR:indel ratio measured by flow cytometry in BFP-to-GFP reporter cells transfected with plasmids encoding fusion protein and sgRNA, along with ssODN template.

For B-E, data represents mean ± s.d. (n = 3). *P < 0.05, determined by student's two-tailed t-test.

Fig. 3: CRISPR-Cas9-methyltransferase fusion proteins display increased HDR efficiency and HDR:indel ratios across multiple endogenous sites

A) HDR frequency and HDR:indel ratio measured by NGS at site 7 (intergenic) in HEK293T cells transfected with plasmids expressing the fusion proteins and sgRNA along with a ssODN template with 50-nt homology arms.
B) HDR frequency and HDR:indel ratio measured by NGS at site 9 (exon 5 of *LDLR*) in HEK293T cells transfected with the fusion proteins and sgRNA with ssODN template.
C) HDR frequency and HDR:indel ratio measured by NGS at multiple genomic loci (sites 6-11) in HEK293T cells transfected with PRDM9-Cas9 and sgRNA with ssODN template.
D) Off-target activity of PRDM9-Cas9 measured by NGS at 7 potential off-target sites predicted from 2 sgRNAs.
For A-D, data represents mean ± s.d. (n = 3). **P* < 0.05, ***P* < 0.01, determined by student's two-tailed t-test.

Fig. 4. PRDM9-Cas9 fusion protein produces increased HDR:indel ratios in different cell types

A) HDR frequency measured by NGS at multiple genomic loci (sites 7, 9, 10) in HEK293T cells transfected with PRDM9-Cas9 and sgRNA with ssODN templates that either include (hash pattern) or lack (solid pattern) a mutation at the PAM site.
B) HDR:indel ratio measured by NGS at multiple genomic loci (sites 7, 9, 10) in HEK293T cells transfected with PRDM9-Cas9 and sgRNA with ssODN templates that either include (hash pattern) or lack (solid pattern) a mutation at the PAM site.
C) HDR frequency and HDR:indel ratio measured by NGS at multiple genomic loci (sites 7-9) in HeLa cells transfected with PRDM9-Cas9 and sgRNA with ssODN template.
D) HDR frequency and HDR:indel ratio measured by NGS at multiple genomic loci (sites 7-9) in U2OS cells transfected with PRDM9-Cas9 and sgRNA with ssODN template.
For A-D, data represents mean ± s.d. (n = 3). **P* < 0.05, ***P* < 0.01, determined by student's two-tailed t-test.