

# Epigenetic features drastically impact CRISPR–Cas9 efficacy in plants

Trevor Weiss  <sup>1,2,3,4</sup>, Peter A. Crisp  <sup>1,2,5</sup>, Krishan M. Rai  <sup>1</sup>, Meredith Song  <sup>6</sup>, Nathan M. Springer  <sup>1,2</sup> and Feng Zhang  <sup>1,2,3,4,\*†</sup>

<sup>1</sup> Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, Minnesota 55108, USA

<sup>2</sup> Center for Precision Plant Genomics, University of Minnesota, Saint Paul, Minnesota 55108, USA

<sup>3</sup> Microbial and Plant Genomics Institute, University of Minnesota, Minneapolis, Minnesota 55108, USA

<sup>4</sup> Center for Genome Engineering, University of Minnesota, Minneapolis, Minnesota 55108, USA

<sup>5</sup> School of Agriculture and Food Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia

<sup>6</sup> Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, Minnesota 55108, USA

\*Author for correspondence: zhangumn@umn.edu

†Senior author

F.Z. conceived, designed, and supervised the project. T.W., P.A.C., K.M.R., M.S., and N.M.S. designed and performed the experiments. T.W., K.M.R., and F.Z. analyzed the data. T.W., N.M.S., and F.Z. wrote the article.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (<https://academic.oup.com/plphys/pages/General-Instructions>) is: Feng Zhang (zhangumn@umn.edu).

## Abstract

CRISPR–Cas9-mediated genome editing has been widely adopted for basic and applied biological research in eukaryotic systems. While many studies consider DNA sequences of CRISPR target sites as the primary determinant for CRISPR mutagenesis efficiency and mutation profiles, increasing evidence reveals the substantial role of chromatin context. Nonetheless, most prior studies are limited by the lack of sufficient epigenetic resources and/or by only transiently expressing CRISPR–Cas9 in a short time window. In this study, we leveraged the wealth of high-resolution epigenomic resources in *Arabidopsis* (*Arabidopsis thaliana*) to address the impact of chromatin features on CRISPR–Cas9 mutagenesis using stable transgenic plants. Our results indicated that DNA methylation and chromatin features could lead to substantial variations in mutagenesis efficiency by up to 250-fold. Low mutagenesis efficiencies were mostly associated with repressive heterochromatic features. This repressive effect appeared to persist through cell divisions but could be alleviated through substantial reduction of DNA methylation at CRISPR target sites. Moreover, specific chromatin features, such as H3K4me1, H3.3, and H3.1, appear to be associated with significant variation in CRISPR–Cas9 mutation profiles mediated by the non-homologous end joining repair pathway. Our findings provide strong evidence that specific chromatin features could have substantial and lasting impacts on both CRISPR–Cas9 mutagenesis efficiency and DNA double-strand break repair outcomes.

## Introduction

CRISPR–Cas-based genome editing technologies have greatly advanced both basic and applied biological research. Among them, CRISPR–Cas9, the effector nuclease from a class 2 bacterial CRISPR system, has been the most widely adopted

in eukaryotes (Jinek et al., 2012). The key steps in CRISPR–Cas9-mediated genome editing involve searching, binding and then cleaving a 20-nucleotide target site directed by a guide RNA (gRNA). The resulting cleavage product with double-strand breaks (DSBs) can then be repaired by either

error-prone DNA repair pathways, such as classical nonhomologous end-joining (NHE) or micro-homology end-joining (MME), or by a template DNA-dependent pathway, that is homology-directed repair (Chen et al., 2019). Thus, specific mutations, including insertions, deletions, or point mutations, can be introduced by employing distinct DNA repair machineries (Chen et al., 2019).

Previous studies indicated that the CRISPR target sequence is the primary determinant for mutagenesis efficiency and mutation profile (Allen et al., 2018; Lazzarotto et al., 2020). Several tools have been developed to predict the efficiency and mutation outcomes solely based on CRISPR targeted sequences (Allen et al., 2018; Concordet and Haeussler, 2018; Xiang et al., 2021). However, the predictability of these tools, primarily based on data from human cells, often varies and appears to translate poorly to other species, such as plants (Naim et al., 2020). This observation suggested that nonsequence features could influence CRISPR–Cas9 mutagenesis.

Increasing evidence has revealed negative associations between CRISPR–Cas9 mutagenesis rates and heterochromatic signatures or low chromatin accessibility in multiple systems, such as yeast (*Saccharomyces cerevisiae*), zebrafish (*Danio rerio*), mouse (*Mus musculus*), human (*Homo sapiens*), and rice (*Oryza sativa*) (Wu et al., 2014; Daer et al., 2017; Uusi-Mäkelä et al., 2018; Yarrington et al., 2018; Liu et al., 2019). However, most of these studies were conducted at various genomic locations, making it difficult to separate the effect of chromatin context from those of DNA sequences. Recently, two studies investigated the impact of chromatin features by randomly inserting reporter constructs at multiple different genomic locations and then targeting an identical site within the reporter for cleavage (Gisler et al., 2019; Schep et al., 2021). Their finding confirmed previous observations that heterochromatin has a negative impact on CRISPR–Cas9 mutagenesis efficacy. Notably, specific chromatin features were also identified to impact both efficiency and mutation outcomes (Gisler et al., 2019; Schep et al., 2021). Nevertheless, these studies were limited by two factors: (1) bias in the genomic location of reporter constructs and (2) the ambiguity from whether the integrated sequences can faithfully adopt the local chromatin context. Furthermore, most of these previous studies were conducted in cell lines within a short time window (usually less than 72 h) using transiently expressed CRISPR–Cas9. It is still unclear whether the heterochromatic features have a long-lasting effect on CRISPR–Cas9 mutagenesis, or only delay it (Kallimasioti-Pazi et al., 2018).

In this study, we leveraged the high resolution epigenomic resources in the model plant species, *Arabidopsis* (*Arabidopsis thaliana*), to address the impact of chromatin features on both CRISPR–Cas9 mutagenesis efficiency and mutation outcomes. To fix the sequence variable, the *Arabidopsis* genome was scanned to identify identical CRISPR target sites located in various chromatin contexts. By using stable CRISPR–Cas9 transgenic plants targeting

multiple chromosomal regions, we systematically characterized mutagenesis efficiency and mutation outcomes with 23 distinct DNA methylation and chromatin features using a next-generation sequencing (NGS) approach. Our findings provide insight into the influences of chromatin features on CRISPR–Cas9 mutagenesis and DNA repair outcomes. This could help develop better technologies for more efficient and precise genome editing.

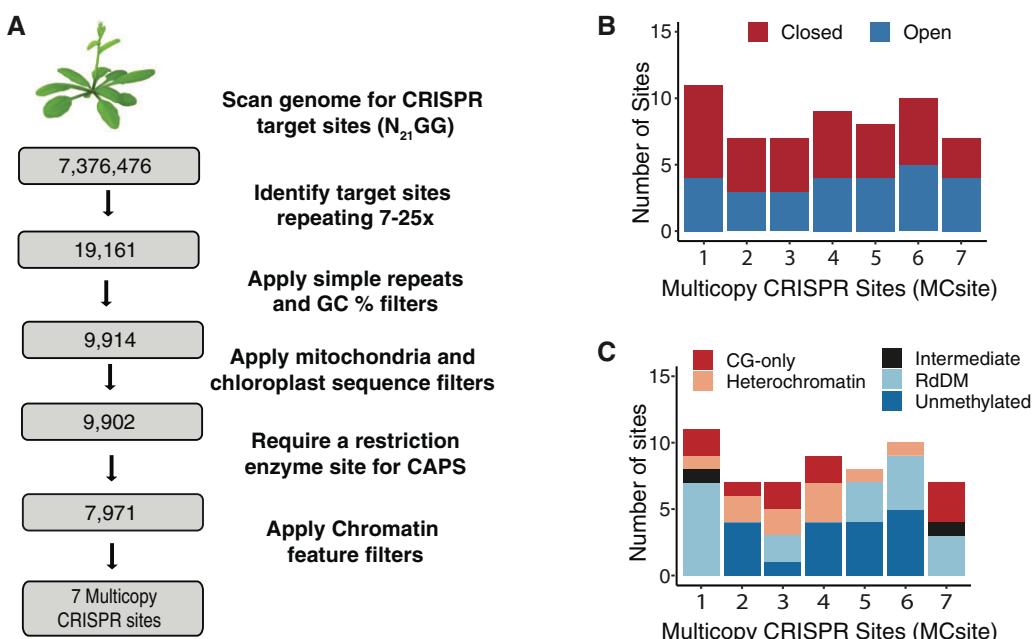
## Results

### Identification of identical CRISPR–Cas9 sites in diverse chromatin contexts

To identify identical CRISPR target sites in various chromatin contexts, the *Arabidopsis* Col-0 reference genome was scanned for 20-bp CRISPR–Cas9 recognition sequences with 3-bp NGG (the PAM sequence) at the 3'-end. Out of 7,376,476 distinct target sites, 19,161 were identified as repeating 7–25 times across the genome (Figure 1A). A series of filters were then applied to remove the sites with one of the following features: simple repeats, GC content outside the range of 40%–60%, matching sequences in mitochondria or chloroplast genomes, or containing no overlapping restriction enzyme site for subsequent mutation genotyping. The remaining 7,971 sequences, representing 92,117 total genomic sites, were assessed for three key chromatin features using 100-bp windows: chromatin accessibility indicated by ATAC-seq scores (Lu et al., 2016), DNA methylation patterns categorized as a DNA methylation domain (RdDM, heterochromatin, CG-only, unmethylated, and intermediate) (Crisp et al., 2017; Springer and Schmitz, 2017), and nine chromatin states (Supplemental Data Set S1; Sequeira-Mendes et al., 2014). Seven multicopy CRISPR sites (MCsites) were identified with individual sequences in each family having highly diverse chromatin contexts, including both open and closed chromatin, at least three different DNA methylation domains, and at least two distinct chromatin states (Figure 1, B and C and Supplemental Data Set S2).

### Differential CRISPR–Cas9 efficiencies were associated with distinct chromatin features

Next, we evaluated CRISPR–Cas9 efficacy for each of the seven MCsite families. T-DNA constructs, containing a CRISPR–Cas9 expression cassette, a firefly luciferase reporter and the bialaphos resistance (BAR) selection marker gene, were made to target each MCsite (Supplemental Figure S1A). Each construct contained two gRNA expression cassettes, one targeting the MCsite and the other as the CRISPR mutagenesis control targeting a single-copy endogenous gene located in unmethylated and accessible chromatin, the Chelatase I2 (*CHL2*) gene, as reported previously (Figure 2A; Mao et al., 2013). It is worth noting that we intentionally chose the CaMV 35S promoter to drive expression of the Cas9 and gRNA sequences because this promoter has much lower activity in *Arabidopsis* embryos than leaves (Wang et al., 2015; Yan et al., 2015). By reducing the mutagenesis potential in early embryo development



**Figure 1** Identification of MCsites in various chromatin contexts. **A**, Bioinformatic pipeline used to identify MCsites. **B**, Characterization of the chromatin accessibility as closed or open at each MCsite using ATAC-Seq data. **C**, Characterization of DNA methylation domains at each MCsite. CG-only sites contain greater than 40% mCG; Heterochromatin sites contain greater than 40% mCG and mCHG methylation; RdDM sites contain mCG, mCHG, and mCHH, with at least 15% mCHH; unmethylated sites contain less than 10% mCG, mCHG, and mCHH; and intermediate sites are everything else with data that did not meet any of the above criteria.

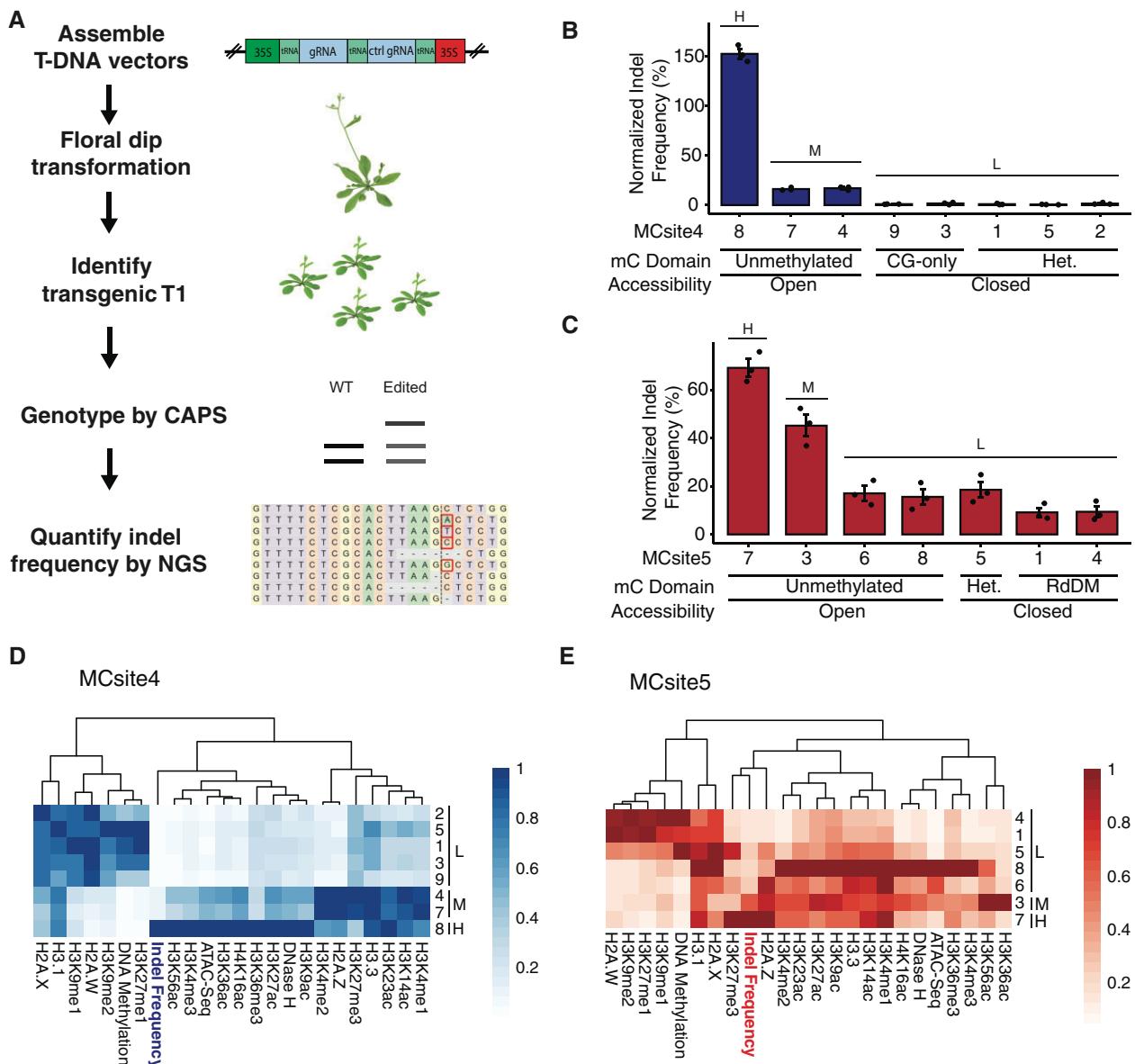
stages, we were able to capture more independent mutation events in somatic cells during leaf development.

After transformation of each T-DNA construct, the resulting CRISPR–Cas9 transgenic plants (T1) were first assessed for mutagenesis efficiency at the *CHL2* control site using the Cleaved Amplified Polymorphic Sequences (CAPS) method. Individual plants with detectable mutagenesis at the control site were then analyzed at each of the seven MCsites using a CAPS or NGS assay. We were able to identify two MCsites, MCsite4 and MCsite5, that produced notable mutagenesis for at least one of the CRISPR target sites (Supplemental Figure S1B). MCsite4 and 5, totaling 15 individual sites, were found spanning all five chromosomes (Supplemental Figure S1C). Among the 15 sites, four seemed to overlap with *Arabidopsis* genes, and only one site, MCsite5.8, was detected in transcribed RNA sequences (Supplemental Data Set S2).

When CRISPR–Cas9 mutagenesis efficiency was assessed at the individual sequences of MCsite4 and 5 using NGS, substantial variations were observed across individual target sites, consistent with the CAPS data (Supplemental Figure S2A). To control for variation in CRISPR–Cas9 expression levels between plants, the mutation rate at each targeted site was normalized to the *CHL2* control. As a result, the normalized frequencies at MCsite4 and MCsite5 sites ranged from an average 0.61%–152.28% and 9.17%–69.17%, respectively (Figure 2, B and C). Pairwise comparisons indicated mutagenesis frequencies of MCsite4 sites can be categorized into three distinct groups ( $P < 0.001$ ): the high editing group (group H: MCsite4.8, 152.28%), the moderate editing

frequency group (group M: MCsite4.4 with 17.04% and 4.7 with 16.20%), and the low editing frequency group (group L: MCsite4.1, 4.2, 4.3, 4.5, and 4.9, ranging from 0.61% to 1.71%) (Figure 2B). Comparison between the highest and lowest MCsite4 edited sites revealed a 249.64-fold difference. Similarly, mutagenesis frequencies at the MCsite5 sequences can also be grouped into the high, moderate, and low editing groups ( $P < 0.01$ ) with 7.54-fold differences between the highest and lowest edited sites (Figure 2C). We then characterized the local sequence context surrounding each target site (25 bp from each side). High sequence similarities were found in these extended sequences for both MCsite4 and 5 families (Supplemental Figure S2, B and C). Phylogenetic analyses within each target site family found no evident associations between sequence similarity and the mutagenesis frequency groups (Supplemental Figure S2, B and C). Thus, the local sequence context could not explain differential mutagenesis frequencies observed from individual target sites.

The initial association analysis of indel frequencies with DNA methylation domains and chromatin accessibility indicated unmethylated and accessible sites generally had higher mutagenesis levels than the methylated and inaccessible sites (Figure 2, B and C). Notably, negative associations between mutagenesis frequency and DNA methylation levels at both MCsite4 and 5 sites could also be observed when the cytosine methylation status was examined at the single nucleotide level of individual CRISPR target sites (Supplemental Figure S3). To systematically investigate the relationship between mutagenesis efficiency and chromatin



**Figure 2** Characterization of CRISPR–Cas9 mutation frequencies and chromatin features at MCsites. **A**, The NGS-based workflow to analyze CRISPR–Cas9 mutagenesis efficiency and repair profile at MCsites. **B** and **C**, Normalized mutation frequency for individual MCsite4 and MCsite5 CRISPR targets. The standard error (SEM) was calculated for each target site with three replicates. A one-way analysis of variance and Tukey's multiple comparison test were performed to group the target sites into H (high), M (moderate), and L (low) mutagenesis levels ( $P < 0.05$ ). The DNA methylation domain (mC domain) and accessibility features were listed below each target site. Het stands for Heterochromatin DNA methylation domain. RdDM stand for RNA-directed DNA methylation. **D** and **E**, Hierarchical clustering heatmap of distinct chromatin features (x-axis) for individual MCsite4 or MCsite5 sites (y-axis). H (high), M (moderate), and L (low) indicated the mutagenesis group for each site. Each feature was normalized to a scale of 0–1, with 1 indicating the highest level (the darkest color) and 0 indicating the lowest level (white). The dendrogram indicated similarity clustering of chromatin features.

features, we further characterized individual MCsites using all 23 chromatin features, including distinct histone modifications and histone variants (Supplemental Data Set S3; Liu et al., 2018). When hierarchical cluster analysis was performed, the lowly edited sites (group L) tended to cluster with the heterochromatic features, such as H2A.W, H3K9me1, H3K9me2, H3K27me1, and hyper DNA methylation, for the MCsite4 and 5 targets with exceptions observed for MCsite5.6 and MCsite5.8 (Figure 2, D and E). While they

are in the lowly edited group, these two sites appeared to be associated with unmethylated and accessible chromatin features (Figure 2, C and E). On the contrary, the highly and moderately edited groups (groups H and M) appeared to be associated with accessible, active chromatin features such as histone acetylation, H3K36me3, and H3K4 methylation (Figure 2, D and E; Roudier et al., 2011). To examine the impact of individual features, we performed correlation analyses by plotting mutagenesis efficiency at all 15 target sites

with each chromatin and DNA methylation feature (Supplemental Figure S4). Consistent with the hierarchical cluster analysis, strong positive correlations were observed between mutagenesis frequency and euchromatin-related features such as H3K56ac ( $R = 0.85$ ,  $P = 1.8e-13$ ), H3K9ac ( $R = 0.82$ ,  $P = 8.6e-12$ ), H3K36ac ( $R = 0.75$ ,  $P = 4.2e-9$ ), H3K27ac ( $R = 0.71$ ,  $P = 4.1e-8$ ), H3K36me3 ( $R = 0.71$ ,  $P = 3.5e-8$ ), H3K4me3 ( $R = 0.65$ ,  $P = 1.6e-6$ ), and accessibility ( $R = 0.51$ ,  $P = 0.00038$ ) measured with ATAC-Seq data (Supplemental Figure S4). Moreover, strong negative correlations were found between mutagenesis frequencies and the heterochromatin-related features H2A.W ( $R = -0.57$ ,  $P = 4.1e-5$ ), H3K9me1 ( $R = -0.55$ ,  $P = 0.00011$ ), H3K9me2 ( $R = -0.55$ ,  $P = 1e-4$ ), and cytosine DNA methylation ( $R = -0.45$ ,  $P = 0.0022$ ) (Supplemental Figure S4).

### Improving CRISPR–Cas9 mutagenesis efficiency through combined reduction of DNA methylation

Because of the strong negative association observed between the lowly edited sites and heterochromatic features, we hypothesized that altering these chromatin states could improve mutagenesis efficiency at the refractory sites. In this study, we chose to perturb DNA methylation at the lowly edited sites because high levels of DNA methylation are often associated with heterochromatic features (Zemach and Grafi, 2003; Johnson et al., 2007; Crisp et al., 2020). We first sought to test the possible impact of CHG methylation on mutagenesis efficiencies because CHG methylation is often associated with H3K9me2 and heterochromatin (Springer and Schmitz, 2017). To this end, the *chromomethylase3* (*cmt3-11t*) mutant was chosen due to the well-documented genome-wide reduction in CHG methylation (Stroud et al., 2013). Analysis of the single-base resolution DNA methylation profiles indicated that a substantial reduction in CHG methylation was confirmed at individual MCsite4 and 5 CRISPR targets in this mutant (Supplemental Figure S5). The CRISPR–Cas9 T-DNA constructs targeting MCsite4 and 5 sequences were transformed into the *cmt3-11t* mutant using the same procedure described above. As a result, transgenic *cmt3-11t* plants (T1) were identified with detectable mutagenesis activities at both MCsite4 and 5 target sites (Supplemental Figure S6). After normalizing to the CHLI2 control target site, we observed a nearly identical pattern in mutagenesis efficiency between WT and *cmt3-11t* plants for both MCsite4 and MCsite5 (Figure 3, A and B). These results suggested that a reduction of CHG methylation alone is not sufficient to improve CRISPR–Cas9 mutagenesis at the refractory sites.

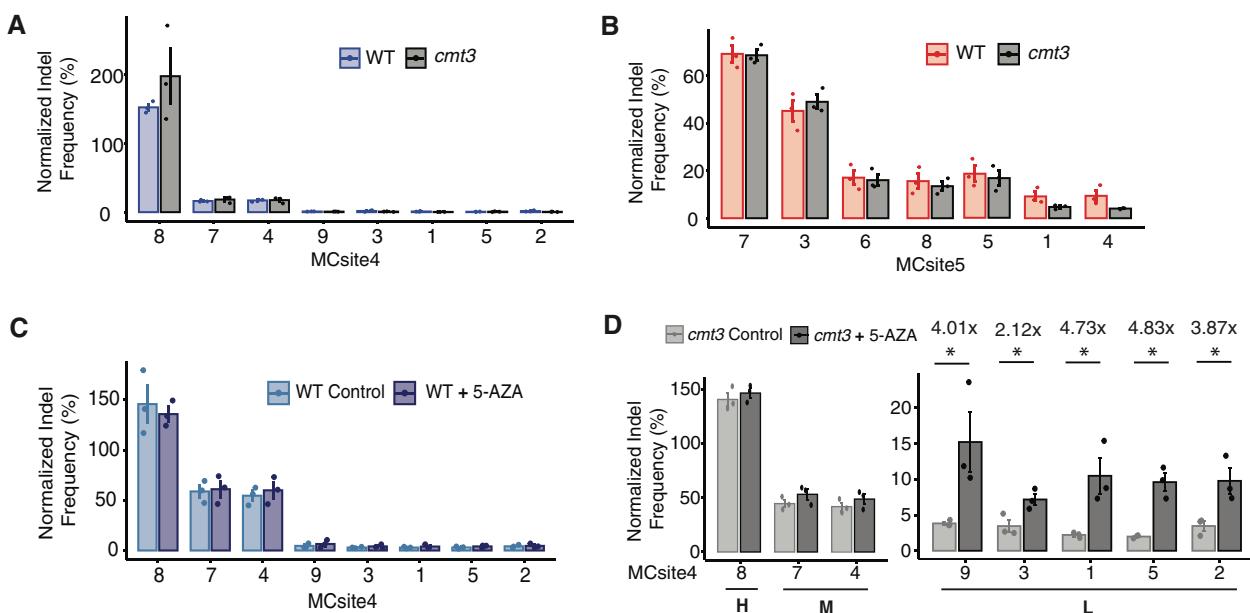
Although additional genetic mutant lines with perturbed DNA methylation, such as Methyltransferase 1 (*met1*), were available (Stroud et al., 2013), we have not been able to obtain transgenic CRISPR–Cas9 T-DNA events in these mutant lines. Therefore, we chose a chemical approach to reduce genome-wide DNA methylation in all 5mC contexts using 5-azacytidine (Griffin et al., 2016). Analysis of the DNA methylation for 5-azacytidine treated plants revealed substantial

reductions in DNA methylation at the highly methylated sequences at both MCsite4 and 5 (Supplemental Figure S7A; Griffin et al., 2016). We focused on MCsite4 in the subsequent 5-azacytidine experiment due to the lower mutagenesis frequencies at the lowly edited sites. The T2 siblings with the CRISPR–Cas9 T-DNA targeting MCsite4 in the wild-type background were obtained from a self-pollinated T1 plant used above. After 2 weeks of growth with or without 100- $\mu$ M 5-azacytidine, individual plants were subjected to mutagenesis analyses using the NGS assay (Supplemental Figure S7B). No significant differences were observed for the normalized mutagenesis frequency at each MCsite4 site between the untreated and treated samples (Figure 3C). Thus, partial reductions in all DNA methylation contexts using 5-azacytidine did not reveal changes in mutagenesis frequencies.

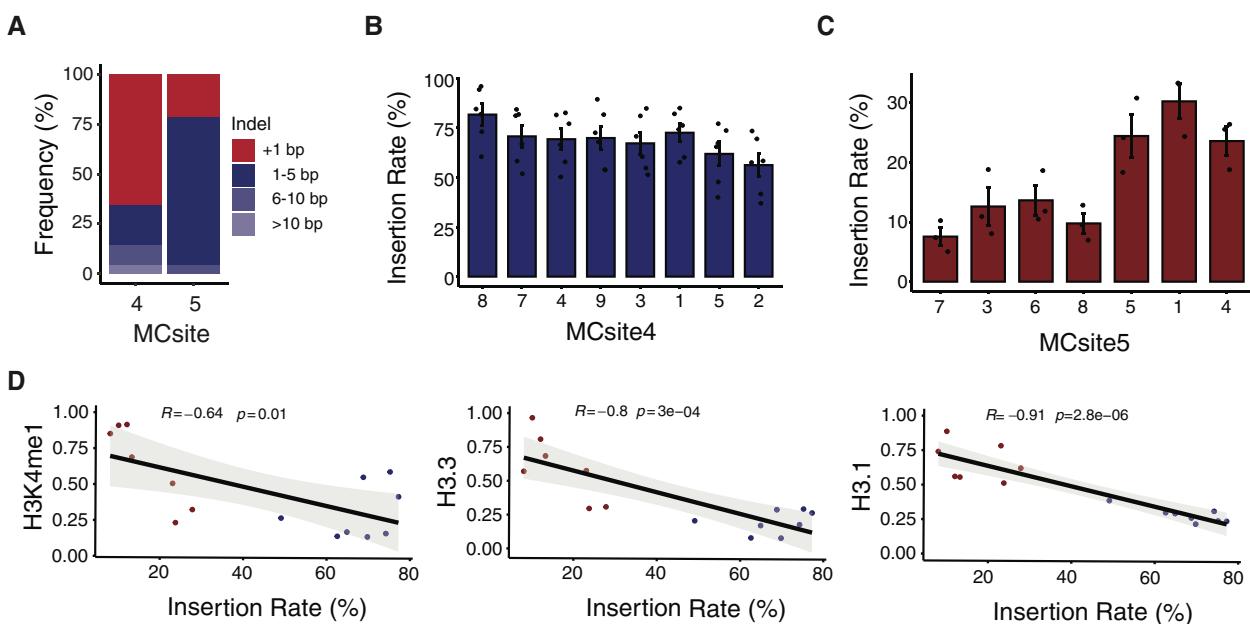
Lastly, we tested the impact of combined reductions in DNA methylation using both the *cmt3-11t* mutant and 5-azacytidine treatments. The T2 siblings with the MCsite4-targeting T-DNA were obtained from a self-pollinated T1 plant in the *cmt3-11t* mutant background, grown for 2 weeks with or without 100- $\mu$ M 5-azacytidine, and subjected to NGS analysis using the same procedure (Supplemental Figure S7C). After normalizing to the CHLI2 control, we observed two distinct patterns for mutagenesis efficiency across individual MCsite4 CRISPR targets. At the unmethylated and accessible sites (groups H and M), no significant differences in mutagenesis frequency were found between the 5-azacytidine treated mutants and the control group (Figure 3D). On the contrary, significant increases in mutagenesis frequency were observed at the inaccessible and hypermethylated sites (group L) when 5-azacytidine treatment was combined with the *cmt3-11t* mutant, that is MCsite4.9 (4.01-fold), MCsite4.3 (2.12-fold), MCsite4.1 (4.73-fold), MCsite4.5 (4.83-fold), and MCsite4.2 (3.87-fold) (Figure 3D). Together, these data indicated that strong reduction in DNA methylation in multiple contexts could result in significant improvement of CRISPR–Cas9 mutagenesis efficiency at the hypermethylated refractory sites.

### Differential CRISPR–Cas9 mutational profiles are associated with distinct chromatin features

CRISPR–Cas9-induced mutations are typically composed of either small deletions or 1 bp insertions (Allen et al., 2018). Small deletions are mainly derived from the NHEJ or MMEJ pathway through exonuclease-mediated end processing and ligation, while the 1-bp insertions were mostly generated from blunt-end or 1-bp staggered cleavage by Cas9 followed by DNA polymerase-mediated end filling (Gisler et al., 2019). To investigate the potential impact of chromatin context on CRISPR–Cas9 mutation outcomes, we examined the insertion and deletion profiles for both MCsite4 and 5. As expected, most mutations contained 1-bp insertions or small deletions (<10 bp) for both MCsite4 and 5 (Figure 4A). MCsite4 was preferentially repaired as 1-bp insertions, while MCsite5 showed a strong bias towards



**Figure 3** Comparisons of CRISPR–Cas9 mutagenesis frequencies using genetic and chemical approaches. **A** and **B**, Normalized mutation frequency was compared for MCsite 4 and 5 in wild-type and *cmt3* mutant background. **C**, Normalized mutation frequencies for MCsite4 sites in wild-type plants were compared with or without 5-azacytidine treatment. **D**, Normalized mutagenesis frequencies were compared for the MCsite4 sites in the *cmt3* mutant plants with or without 5-azacytidine treatment. The MCsite4 sites were categorized in High (H; site 8), Moderate (M; sites 7 and 4), and Low (L; sites 9, 3, 1, 5, and 2) mutagenesis groups. The standard error (SEM) was calculated for each target site with three replicates in (A)–(D). The fold changes in mutagenesis were calculated for the sites in group L in (D) (shown above bars). Mann–Whitney *U* test was performed in each panel, with asterisks indicating significant *P*-values (*P* < 0.05). If no asterisk is present, *P*-values are not significant.



**Figure 4** Analyses of CRISPR–Cas9 mutation profiles at MCsites. **A**, Distribution of mutation outcomes at MCsite4 and MCsite5. The mutations were categorized as 1-bp insertions (+ 1 bp, red), 1–5 bp deletions (Δ 1–5 bp, dark blue), 6–10 bp deletions (Δ 6–10 bp, blue), and deletions of more than 10 bp (Δ >10 bp, light blue). **B** and **C**, The 1-bp insertion frequencies at individual MCsite4 and MCsite5 targets. The standard error (SEM) was calculated for each target site with replicates (*n* = 6 for MCsite4; *n* = 3 for MCsite5). Kruskal–Wallis analysis indicated the *P*-value for MCsite4 and MCsite5 are 0.17 and 0.016, respectively. **D**, Correlation plots between the insertion rate and three chromatin markers, H3K4me1, H3.3, and H3.1, with the blue dots for MCsite4 and the red dots for MCsite5. The trendline was plotted with gray indicating the standard error. The *R*- and *P*-values were calculated using Spearman's rank correlation coefficient.

small deletion outcomes (Figure 4A). While the major mutation types at individual sites within each family seemed to be highly similar (Supplemental Figure S8, A–D), further analysis revealed substantial variations for the insertion rate between individual sites. The rate of insertion outcomes ranged from 56.25% to 81.73% and 7.57% to 30.27% in MCsite4 and MCsite5 sites, respectively (Figure 4, B and C). We then conducted a correlation analysis using the 23 epigenetic features with the insertion rates at all 15 sites. Three histone H3-related features, H3K4me1 ( $R = -0.64$ ,  $P = 0.01$ ), H3.3 ( $R = -0.83$ ,  $P = 3e-4$ ), and H3.1 ( $R = -0.91$ ,  $P = 2.8e-6$ ), were identified with significant negative correlations with 1-bp insertional mutations (Figure 4D and Supplemental Figure S9). Thus, these findings strongly suggested that chromatin features could not only affect CRISPR–Cas9 mutagenesis efficiency but also influence mutation outcomes.

## Discussion

In recent studies, chromatin contexts have been demonstrated to have significant impacts on CRISPR–Cas9-mediated genome editing (Wu et al., 2014; Daer et al., 2017; Yarrington et al., 2018; Liu et al., 2019). Most of these findings indicated that heterochromatic features at the CRISPR target regions could impede CRISPR–Cas9 mutagenesis efficiency in multiple systems, such as yeast, rice, mouse, and human cell lines (Wu et al., 2014; Daer et al., 2017; Yarrington et al., 2018; Liu et al., 2019). However, it remained unclear whether heterochromatic features could only temporarily delay CRISPR–Cas9 mutagenesis (Kallimasioti-Pazi et al., 2018). In this study, we systematically characterized the impact of 23 distinct DNA methylation and chromatin features on CRISPR–Cas9 mutagenesis efficiency and mutation outcomes by investigating CRISPR–Cas9 transgenic Arabidopsis plants. Consistent with the previous studies, our results demonstrated that inaccessible and heterochromatic features were associated with low mutagenesis efficiency. Such repressive effects can be long-lasting through plant development leading up to a 250-fold difference between identical CRISPR target sites. However, it was worth noting that, although this observation broadly holds for both MCsite4 and 5, two lowly edited sites in MCsite5, MCsite5.6 and 5.8, appeared to be associated with open and active chromatin features. The weaker associations observed here could be due to the overall higher mutagenesis frequencies at the lowly edited MCsite5 targets than those at the MCsite4 targets. Furthermore, mutagenesis efficiency at the target sequences in accessible chromatin regions could also have substantial variations ranging from 1.53-fold to 10-fold at MCsite4 and MCsite5, respectively (Figure 2, B and C). These observations suggested that specific chromatin features other than just open or closed chromatin should be considered to account for CRISPR–Cas9 mutagenesis efficiency. Close examination of individual chromatin features identified several euchromatic marks, such as H3K9ac, H3K56ac, H3K36ac, H3K27ac, H3K4 methylation, and

H3K36m3, that were positively correlated with mutagenesis efficiency. Further investigation with a larger data set will be needed to dissect their impacts on CRISPR–Cas9 mutagenesis in greater detail.

In this study, we observed strongly negative correlations between CRISPR–Cas9 mutagenesis efficiency and repressive chromatin features, such as DNA hypermethylation, low DNA accessibility, and H3K9 methylation. To test the hypothesis that modulating some of these features could improve mutagenesis frequency at the lowly edited sites, we sought to reduce DNA methylation by using both genetic and chemical approaches. Our results indicated that partial reduction of DNA methylation using either a mutant affecting CHG methylation or 5-azacytidine treatment alone was not sufficient to improve mutagenesis efficiency at any of the tested sites. When combining the CHG deficiency mutant with 5-azacytidine chemical treatment, 2.1- to 4.8-fold improvements in mutagenesis efficiency were found at the lowly edited sites but not at the highly and moderately edited sites. Combined reduction of DNA methylation in multiple contexts has been demonstrated to increase chromatin accessibility and even alter the higher order 3D chromatin organization in Arabidopsis (Zhong et al., 2021). Thus, mutagenesis efficiency improvement at lowly edited sites observed here could have resulted from increasing chromatin accessibility and/or changing 3D chromatin organization due to the substantial reduction of DNA methylation in multiple contexts; but this would need to be experimentally verified, for example using ATAC-Seq or DNase I hypersensitivity assays. Additionally, as discussed above, open/close chromatin structure alone could not explain all the variations in editing efficiencies. A systematic approach would be required to further dissect the relationship between epigenetic features and targeted mutagenesis efficiency. Furthermore, as reported previously, repressive chromatin features could act as the barriers to hinder CRISPR–Cas9 binding and cleavage (Wu et al., 2014). However, it is noted that the primary readout in this study, targeted mutation frequency, is an indirect measurement of CRISPR–Cas9 target binding and cleavage dependent on DSB repair. The mutation efficiency differences observed here could reflect differences in either CRISPR–Cas9 binding, cleaving, DNA repair, or a combination of these. Further investigation would be required to differentiate these possibilities.

In addition to the impacts on mutagenesis efficiency, chromatin features have been suggested to influence CRISPR–Cas9 mutation outcomes. For example, a recent study with more than 1,000 copies of identical insertion sites indicated the 1-bp insertions were found more prevalent in euchromatin than in heterochromatin, likely through recruiting different DNA repair machinery (Schep et al., 2021). However, conflicting results were also reported showing little impact of chromatin features on mutation outcomes (Kallimasioti-Pazi et al., 2018; Gisler et al., 2019). In this study, we observed significant variations for the 1-bp insertion rate in different chromatin contexts. Of the 23

chromatin features analyzed, three distinct histone H3 features, H3K4me1, H3.3, and H3.1, exhibited significantly strong negative correlations with the 1-bp insertion rate. Interestingly, H3K4me1 was also identified by Schep et al. (2021) as a marker to correlate with distinct mutation outcomes, while H3.1 and H3.3 were not included in their study. This suggested that the balance between 1-bp insertions and small deletions could be influenced by specific chromatin markers. Distinct chromatin features have been reported to recruit different DNA repair machinery and result in different repair outcomes in mammalian systems (Fnu et al., 2011; Jacquet et al., 2016; Luijsterburg et al., 2016). Further investigation is needed to address the potential roles of specific chromatin markers in determining DNA repair outcomes.

We propose a model to account for the impacts of chromatin features on CRISPR–Cas9 mutagenesis efficiency and mutation outcomes (Figure 5). In the first step, chromatin features are the key determinants for CRISPR–Cas9 recognition and binding efficiency (Figure 5). In general, heterochromatic features, such as H3K9 methylations, H3K27me1, H2A.W, and DNA hypermethylation, could substantially reduce chromatin accessibility and thus reduce the recognition and binding efficiency of CRISPR–Cas9. Such repressive effects could persist through cell division and development. After CRISPR–Cas9 locates and binds the genomic target site, it can introduce a DSB with either a 1-bp 5' overhang (staggered cut), or blunt ends (Gisler et al., 2019). The cleaved product can be repaired to yield three outcomes: wild-type sequence by perfect ligation, 1-bp insertions, or small deletions (less than 10 bp) (Figure 5). It has been proposed that the staggered cut primarily leads to 1-bp insertion via template-dependent repair, facilitated by a DNA polymerase, while the blunt cut mainly results in small deletions through end resection (Gisler et al., 2019; Schmid-Burgk et al., 2020). The blunt ends could occasionally be repaired by DNA polymerase, likely the members from DNA polymerase family X, without DSB end resection, resulting in template-independent 1-bp insertions (Gisler et al., 2019). In this study, we observed both templated 1-bp insertion and template-independent insertions, as exemplified in MCsite5 and MCsite4, respectively (Supplemental Figure S8, A and B). Nevertheless, the balance between 1-bp insertion and small deletion products is primarily dependent on the repair choices between short-range DSB end resection and DNA polymerase end filling (Figure 5; Lemos et al., 2018; Schmid-Burgk et al., 2020). While sequence features are a key determinant in mutation profile, our data indicated that chromatin features, such as H3K4me1, H3.3, and H3.1, could also significantly impact the balance between 1-bp insertion and small deletion outcomes. These specific chromatin features may exert their influences on the balance between the staggered and blunt cut, or through modulating the balance between DNA polymerase end filling and short-range DSB end resection during NHEJ. In fact, previous studies have

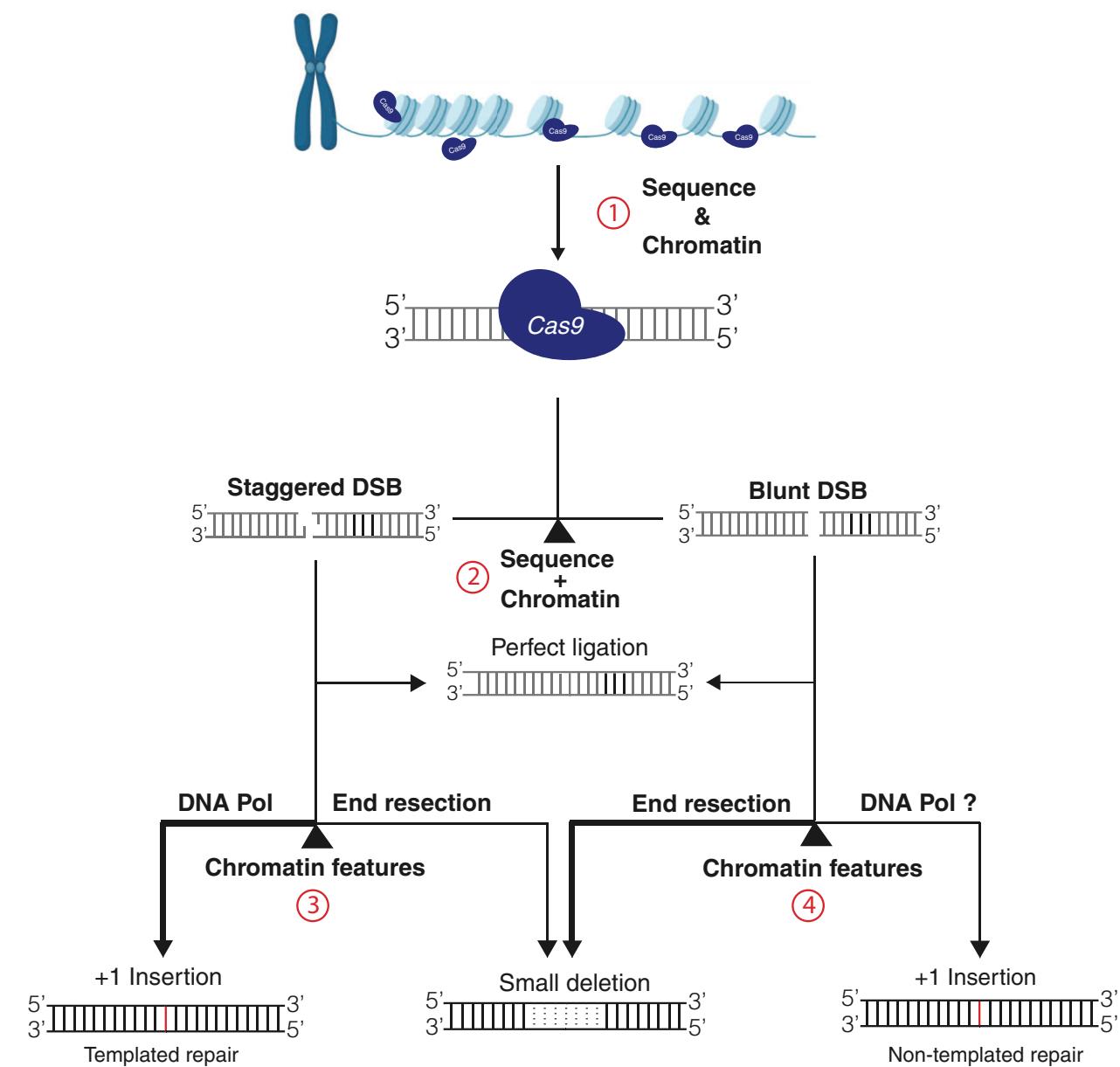
demonstrated that distinct chromatin features, such as the H3.3 variant, could impact DSB end resection (Luijsterburg et al., 2016).

Recently much effort has been made to develop computational tools to predict CRISPR–Cas9 editing efficiency and/or repair outcomes. To our knowledge, these tools primarily relied on sequence features (Allen et al., 2018; Concordet and Haeussler, 2018; Xiang et al., 2021). Yet, our results indicated that nonsequence features should also be taken into consideration for the prediction of both CRISPR–Cas9 efficiency and mutation outcomes. One implication from this study is that heterochromatin-related features, such as low accessibility, H3K9me1, H3K9me2, H3K27me1, H2A.W, and DNA hypermethylation, should be avoided in order to design gRNAs with high efficiency. When it comes to predicting CRISPR–Cas9 mutation profiles, chromatin features such as H3K4me1, H3.3, and H3.1 should also be considered. Future studies using genetic mutants that result in altered states of histone modifications will be interesting to further dissect the impact of distinct chromatin features on CRISPR–Cas9 mutagenesis. A better understanding of the interplay between chromatin dynamics and CRISPR–Cas9 will enable the development of more precise and efficient genome engineering technologies.

## Materials and methods

### Identification of multicopy CRISPR sites

To identify gRNAs that matched to multiple places in the genome, the *Arabidopsis* (*A. thaliana*) Col-0 reference genome (TAIR10) was parsed to identify every NGG PAM site using *seqkit locate* (Shen et al., 2016) with the search motif NNNNNNNNNNNNNNNNNNNNNNNNNNGG. The number of occurrences of each nonredundant sequence in the genome was summarized using *csvtk*. The resulting distinct target sites were then filtered in R (v4.1.2) to retain gRNAs that had between 7 and 25 distinct matches. Sites were then eliminated if they had simple sequence motifs consisting of 5 As, Ts, Gs, or Cs in a row; CG content outside 40%–60% or if the sequence was found in the chloroplast or mitochondria genome; to give a final list of 9,902 candidate gRNAs. The gRNA identification script is available at [https://github.com/pedrocrisp/Weiss\\_et\\_al\\_gRNA\\_chromatin](https://github.com/pedrocrisp/Weiss_et_al_gRNA_chromatin). The potential gRNAs were also screened to identify those with restriction enzyme recognition motifs (from a list commercially available enzymes from NEB) that overlapped position 17–18 of the gRNA (between position 3 and 4 bp from the PAM) such that an indel mutation in this position would disrupt restriction enzyme recognition and cleave for efficient screening of edited transgenic lines. A final list of 7,971 gRNAs was then annotated with chromatin state information by overlapping the coordinates of the gRNA target sites with chromatin annotation files using *bedtools* (Quinlan and Hall, 2010). Chromatin data included nine histone states (Sequeira-Mendes et al., 2014); chromatin accessibility (Lu et al., 2016); and DNA methylation (Crisp et al., 2017;



Downloaded from <https://academic.oup.com/plphys/article/190/2/1153/6605859> by guest on 22 September 2023

**Figure 5** Model depicting the factors that influence CRISPR–Cas9 mutagenesis efficiency and DNA DSB mutation outcomes. The four key check points were indicated in the numerical order with red circles. At Point 1, both sequence and chromatin features are the key determinants for CRISPR–Cas9 recognition and binding efficiency. At Point 2, sequence and chromatin features are crucial to determine staggered or blunt cleavage by CRISPR–Cas9. The cleaved products can be ligated back perfectly or go to Point 3 or 4, where chromatin features likely influence occurrence of end resection. With end resection, DSB repair typically generates small deletions. When the end resection is limited, the broken ends will be directly filled by DNA polymerase (DNA Pol) to generate 1-bp insertions. One-bp insertions can be the products from template-dependent repair using the 1-bp 5' overhang by staggered cleavage, or from template-independent repair by filling in blunt ends. The staggered cleavage primarily leads to 1-bp insertions, while the blunt cut mainly results in small deletions as indicated by line thickness.

Springer and Schmitz, 2017). DNA methylation data were converted to methylation domains for each 100-bp non-overlapping window of the TAIR10 genome using the method detailed in Crisp et al. (2020). Chromatin accessibility at the CRISPR target site was called either open or closed based on the presence of an overlapping ATAC-Seq peak or lack thereof, respectively, using the accessibility profiles in

(Lu et al., 2016). Gene and transposable element annotations were downloaded from Araport v11.

#### T-DNA vector construction

The CRISPR–Cas9 constructs were created using the Golden Gate assembly method as outlined previously (Čermák et al., 2017). The gRNA sequences were first assembled into the

pMOD\_B2301 vector containing the gRNA targeting the multicopy CRISPR site (MCsite) and the CHIL2 positive control using the oligos listed in *Supplemental Table S3*. T-DNA constructs were assembled by combining pMOD\_A0101 (the Cas9 component), pMOD\_B2301 containing the gRNA array, pMZ105 (the luciferase reporter), and pTRANS230d (the T-DNA backbone) with the Golden Gate method as described previously (Čermák et al., 2017). The modular components used to build the T-DNA plasmids can be found at <https://www.addgene.org/browse/article/28189956/> (Čermák et al., 2017). The T-DNA constructs described in this study are available through Addgene.

### Plant materials and growth conditions

The *A. thaliana* Columbia ecotype (Col-0) was used in these experiments. The *cmt3-11t* (stock CS16392) genotype was acquired from the Arabidopsis Biological Resource Center. Floral dip transformation was performed according to the protocol as previously outlined (Zhang et al., 2006). Transgenic T1 seeds were sown on soil and exposed to BASTA selection to recover transgenic plants. Plants were grown in a growth chamber with the following conditions: 16-/8-h light/dark cycle, 22°C, and 55% humidity.

### Characterization of chromatin features

Previously analyzed datasets (BigWig files) for the 23 chromatin features were downloaded from Plant Chromatin State Database (Liu et al., 2018). The data sets for wild-type *Arabidopsis* with and without 100-μM 5-azacytidine treatment were downloaded from Griffin et al. (2016). For each data set, the values for the 1-kb window (500-bp upstream and 500-bp downstream from the center of gRNA) were calculated using the DeepTools2 computeMatrix in the reference-point mode. Similarly, the values for each nucleotide of target sites were calculated using the scale-region mode at single base resolution. Each data set was then normalized to allow for comparison on a scale of 0–1 with 1 indicating the highest level of that feature.

### 5-Azacytidine treatment and luciferase screening

T2 seedlings from self-pollinated T1 *Arabidopsis* plants were grown on 1% (w/v) agar plates containing 0.5 Murashige and Skoog (PhytoTech Labs) and 100-μM 5-azacytidine (Griffin et al., 2016). After 2 weeks, seedlings were screened for the presence of the transgene using a luciferase reporter. The luciferase assay procedure was performed using the Bio-Glo Luciferase Assay System (Promega Corp., Madison, Wisconsin, USA) in accordance with the manufacturer's instructions.

### Mutation genotyping and characterization of mutation profiles

Genotyping was performed using two methods: genomic PCR followed by restriction enzyme digestion (CAPS) and the NGS assay using Illumina paired-end read amplicon sequencing. All tissues for genotyping were collected at 2 weeks post germination for the CTAB-base genomic DNA

extraction. PCR was performed using GoTaq Green Mastermix (Promega Corp., Madison, Wisconsin, USA) according to the manufacturer's instructions, with an annealing temperature of 55°C (CHIL2 and MCsite4) or 60°C (MCsite5) with an extension time of 1 min. Primers to amplify CHIL2, MCsite4, and MCsite5 target sites can be found in *Supplemental Table S1*. Amplicons were then subjected to restriction enzyme digestion using *BsmAI* (CHIL2), *AluI* (MCsite4), or *DrdI* (MCsite5) according to the manufacturer's instructions. PCR amplicons generated with the corresponding primers were subjected to Illumina paired-end read sequencing (Genewiz Inc., South Plainfield, New Jersey, USA). The raw NGS reads were analyzed using CRISPResso2 to estimate indel mutation rates (Clement et al., 2019). To analyze the mutation profiles for each sample, the NGS reads with indel mutations were extracted from the CRISPResso2 output files with a 2% threshold. The resulting output files were then loaded into R studio (version 4.1.0) for data visualization using ggplot (41). The total read counts for CRISPR–Cas9 editing frequency and repair profiles can be found in *Supplemental Table S2*. Normalized indel frequencies were calculated by dividing the indel frequency of each MCsite by the CHIL2 positive control indel frequency within each replicate.

### Data availability

All sequencing data analyzed in this article will be available at the National Center for Biotechnology Information under BioProject Accession PRJNA795172 (*Supplemental Data Set S4*).

### Accession numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession number PRJNA795172.

### Supplemental data

The following materials are available in the online version of this article.

**Supplemental Figure S1.** Characterization of multicopy CRISPR sites (MCsites) for CRISPR–Cas9 mutagenesis.

**Supplemental Figure S2.** Non-normalized mutagenesis efficiency and sequence comparison for individual target sites in MCsite4 and MCsite5.

**Supplemental Figure S3.** Single nucleotide heatmap of DNA methylation levels at multicopy CRISPR site (MCsite) 4 (blue) and 5 (red) protospacer and PAM (bold) sequences from 0 (unmethylated) to 100 (fully methylated).

**Supplemental Figure S4.** Correlation analysis for CRISPR–Cas9 mutagenesis frequencies and chromatin features.

**Supplemental Figure S5.** Characterization of the single-based DNA methylation status at multicopy CRISPR site (MCsite) 4 and 5 in the wild-type and *cmt3* mutant plants.

**Supplemental Figure S6.** Unnormalized mutagenesis frequencies for multicopy CRISPR site (MCsite) 4 (blue and gray) and 5 (red and gray) in the *cmt3* mutant plants.

**Supplemental Figure S7.** 5-Azacytidine treatment of the wild-type and *cmt3* T2 seedlings.

**Supplemental Figure S8.** Characterization of mutation outcomes for multicopy CRISPR site (MCsite) 4 (blue) and 5 (red).

**Supplemental Figure S9.** Correlation analysis for 1-bp insertion rate and chromatin features.

**Supplemental Table S1.** Primer sequences to amplify each CRISPR target site analyzed in these experiments.

**Supplemental Table S2.** Summary of NGS reads count for each tested target site.

**Supplemental Table S3.** Oligos for cloning MCsite and CHL2 gRNAs.

**Supplemental Data Set S1.** Characterization of the sequences, DNA methylation, chromatin accessibility, and chromatin states for the 7,971 candidate CRISPR target sites identified.

**Supplemental Data Set S2.** Annotations of MCsite 4 and 5.

**Supplemental Data Set S3.** Plant Chromatin State Database (PCSD) files for MCsite 4 and 5.

**Supplemental Data Set S4.** PRJNA795172 accession key.

## Acknowledgments

We thank all members of the Zhang Lab for helpful discussions.

## Funding

M.S. was supported by the Undergraduate Research Opportunities Program (UROP) from the University of Minnesota. Funding from NSF IOS-1934384 to N.M.S. partially supported this work. P.A.C. was supported by an ARC Discovery Early Career Award (DE200101748). F.Z. was supported by the startup fund from the Department of Plant and Microbial Biology, University of Minnesota.

*Conflict of interest statement.* None declared.

## References

Allen F, Crepaldi L, Alsinet C, Strong AJ, Kleshchevnikov V, De Angeli P, Páleníková P, Khodak A, Kiselev V, Kosicki M et al. (2018) Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol* **37**: 64–72

Cermák T, Curtin SJ, Gil-Humanez J, Čegan R, Kono TJY, Konečná E, Belanto JJ, Starker CG, Mathre JW, Greenstein RL et al. (2017) A multipurpose toolkit to enable advanced genome engineering in plants. *Plant Cell* **29**: 1196–1217

Chen K, Wang Y, Zhang R, Zhang H, Gao C (2019) CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu Rev Plant Biol* **70**: 667–697

Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE et al. (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* **37**: 224–226

Concordet J-P, Haeussler M (2018) CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* **46**: W242–W245

Crisp PA, Ganguly DR, Smith AB, Murray KD, Estavillo GM, Searle I, Ford E, Bogdanović O, Lister R, Borevitz JO et al. (2017) Rapid recovery gene downregulation during excess-light stress and recovery in *Arabidopsis*. *Plant Cell* **29**: 1836–1863

Crisp PA, Marand AP, Noshay JM, Zhou P, Lu Z, Schmitz RJ, Springer NM (2020) Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. *Proc Natl Acad Sci USA* **117**: 23991–23990

Daer RM, Cutts JP, Brafman DA, Haynes KA (2017) The impact of chromatin dynamics on Cas9-mediated genome editing in human cells. *ACS Synth Biol* **6**: 428–438

Fnu S, Williamson EA, De Haro LP, Brenneman M, Wray J, Shaheen M, Radhakrishnan K, Lee S-H, Nickoloff JA, Hromas R (2011) Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. *Proc Natl Acad Sci USA* **108**: 540–545

Gisler S, Gonçalves JP, Akhtar W, de Jong J, Pindyurin AV, Wessels LFA, van Lohuizen M (2019) Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. *Nat Commun* **10**: 1598

Griffin PT, Niederhuth CE, Schmitz RJ (2016) A comparative analysis of 5-azacytidine- and zebularine-induced DNA demethylation. *G3* **6**: 2773–2780

Jacquet K, Fradet-Turcotte A, Avvakumov N, Lambert J-P, Roques C, Pandita RK, Paquet E, Herst P, Gingras AC, Pandita TK et al. (2016) The TIP60 complex regulates bivalent chromatin recognition by 53BP1 through direct H4K20me binding and H2AK15 acetylation. *Mol Cell* **62**: 409–421

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816–821

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr Biol* **17**: 379–384

Kallimasioti-Pazi EM, Thelakkad Chathoth K, Taylor GC, Meynert A, Ballinger T, Kelder MJE, Lalevée S, Sanli I, Feil R, Wood AJ (2018) Heterochromatin delays CRISPR–Cas9 mutagenesis but does not influence the outcome of mutagenic DNA repair. *PLoS Biol* **16**: e2005595

Lazzarotto CR, Malinin NL, Li Y, Zhang R, Yang Y, Lee G, Cowley E, et al. (2020) CHANGE-Seq reveals genetic and epigenetic effects on CRISPR–Cas9 genome-wide activity. *Nat Biotechnol* **38**: 1317–1327

Lemos BR, Kaplan AC, Bae JE, Ferrazzoli AE, Kuo J, Anand RP, Waterman DP, Haber JE (2018) CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proc Natl Acad Sci USA* **115**: E2040–E2047

Liu G, Yin K, Zhang Q, Gao C, Qiu J-L (2019) Modulating chromatin accessibility by transactivation and targeting proximal dsgRNAs enhances Cas9 editing efficiency *in vivo*. *Genome Biol* **20**: 145

Liu Y, Tian T, Zhang K, You Q, Yan H, Zhao N, Yi X, Xu W, Su Z (2018) PCSD: a plant chromatin state database. *Nucleic Acids Res* **46**: D1157–D1167

Luijsterburg MS, de Krijger I, Wiegant WW, Shah RG, Smeenk G, de Groot AJL, Pines A, Vertegaal ACO, Jacobs JJL, Shah GM et al. (2016) PARP1 links CHD2-mediated chromatin expansion and H3.3 deposition to DNA repair by non-homologous end-joining. *Mol Cell* **61**: 547–562

Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ (2016) Combining ATAC-Seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res* **45**: e41–e41

Mao Y, Zhang H, Xu N, Zhang B, Gou F, Zhu J-K (2013) Application of the CRISPR–Cas system for efficient genome engineering in plants. *Mol Plant* **6**: 2008–2011

Naim F, Shand K, Hayashi S, O'Brien M, McGree J, Johnson AAT, Dugdale B, Waterhouse PM (2020) Are the current gRNA ranking prediction algorithms useful for genome editing in plants? *PLoS ONE* **15**: e0227994

**Quinlan AR, Hall IM** (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842

**Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L et al.** (2011) Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J* **30**: 1928–1938

**Schep R, Brinkman EK, Leemans C, Vergara X, van der Weide RH, Morris B, van Schaik T, Manzo SG, Peric-Hupkes D, van den Berg J et al.** (2021) Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol Cell* **81**: 2216–2230.e10

**Schmid-Burgk JL, Gao L, Li D, Gardner Z, Strecker J, Lash B, Zhang F** (2020) Highly parallel profiling of Cas9 variant specificity. *Mol Cell* **78**: 794–800.e8

**Sequeira-Mendes J, Aragüez I, Peiró R, Méndez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U, Gutierrez C** (2014) The functional topography of the *Arabidopsis* genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* **26**: 2351–2366

**Shen W, Le S, Li Y, Hu F** (2016) SeqKit: a cross-platform and ultra-fast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**: e0163962

**Springer NM, Schmitz RJ** (2017) Exploiting induced and natural epigenetic variation for crop improvement. *Nat Rev Genet* **18**: 563–575

**Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE** (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**: 352–364

**Usi-Mäkelä MIE, Barker HR, Bäuerlein CA, Häkkinen T, Nykter M, Rämet M** (2018) Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the Zebrafish (*Danio rerio*). *PLoS ONE* **13**: e0196238

**Wang Z-P, Xing H-L, Dong L, Zhang H-Y, Han C-Y, Wang X-C, Chen Q-J** (2015) Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in *Arabidopsis* in a single generation. *Genome Biol* **16**: 144

**Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, Cheng AW, Trevino AE, Konermann S, Chen S, et al.** (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* **32**: 670–676

**Xiang X, Corsi GI, Anthon C, Qu K, Pan X, Liang X, Han P, Dong Z, Liu L, Zhong J, et al.** (2021) Enhancing CRISPR-Cas9 gRNA efficiency prediction by data integration and deep learning. *Nat Commun* **12**: 3238

**Yan L, Wei S, Wu Y, Hu R, Li H, Yang W, Xie Q** (2015) High-efficiency genome editing in *Arabidopsis* using YAO promoter-driven CRISPR/Cas9 system. *Mol Plant* **8**: 1820–1823

**Yarrington RM, Verma S, Schwartz S, Trautman JK, Carroll D** (2018) Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *Proc Natl Acad Sci USA* **115**: 9351–9358

**Zemach A, Grafi G** (2003) Characterization of *Arabidopsis thaliana* methyl-CpG-binding domain (MBD) proteins. *Plant J* **34**: 565–572

**Zhang X, Henriques R, Lin S-S, Niu Q-W, Chua N-H** (2006) Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat Protocols* **1**: 641–646

**Zhong Z, Feng S, Duttke SH, Potok ME, Zhang Y, Gallego-Bartolomé J, Liu W, Jacobsen SE** (2021) DNA methylation-linked chromatin accessibility affects genomic architecture in *Arabidopsis*. *Proc Natl Acad Sci USA* **118**: e2023347118