nature plants

ARTICLES

https://doi.org/10.1038/s41477-020-00810-z

Check for updates

RNA interference-independent reprogramming of DNA methylation in *Arabidopsis*

Taiko Kim To[1,6]✉, Yuichiro Nishizawa[1,6], Soichi Inagaki[1,2,3,6], Yoshiaki Tarutani[2,6], Sayaka Tominaga[1], Atsushi Toyoda[4], Asao Fujiyama[4], Frédéric Berger[5] and Tetsuji Kakutani[1,2]✉

DNA methylation is important for silencing transposable elements (TEs) in diverse eukaryotes, including plants. In plant genomes, TEs are silenced by methylation of histone H3 lysine 9 (H3K9) and cytosines in both CG and non-CG contexts. The role of RNA interference (RNAi) in establishing TE-specific silent marks has been extensively studied, but the importance of RNAi-independent pathways remains largely unexplored. Here, we directly investigated transgenerational de novo DNA methylation of TEs after the loss of silent marks. Our analyses uncovered potent and precise RNAi-independent pathways for recovering non-CG methylation and H3K9 methylation in most TE genes (that is, coding regions within TEs). Characterization of a subset of TE genes without the recovery revealed the effects of H3K9 demethylation, replacement of histone H2A variants and their interaction with CG methylation, together with feedback from transcription. These chromatin components are conserved among eukaryotes and may contribute to chromatin reprogramming in a conserved manner.

Cytosine methylation is important for silencing transposable elements (TEs) in diverse eukaryotes including plants[1–3]. In plant genomes, most TEs are silenced by methylation at cytosines of both CG and non-CG (or CH, where H can be A, T or C) contexts. In *Arabidopsis*, CG methylation is maintained by a conserved DNA methyltransferase (MTases) called MET1 (METHYLTRANSFERASE 1)[4,5]. CH methylation includes methylation at CHG and CHH sites, which are mostly catalysed by another class of DNA MTases, CHROMOMETHYLASE 3 and 2 (CMT3 and CMT2), respectively[6–9]. These CMTs are recruited to regions with histone H3 at lysine 9 (H3K9me). H3K9me, in turn, is recruited to regions with CH methylation[7,10,11], generating a self-reinforcing positive feedback loop[12]. Thus, methylation at CG and CH sites, as well as H3K9me, can be epigenetically inherited during cell divisions. An important question is, how are these heritable silencing marks specifically targeted to TEs or excluded from active genes?

Factors affecting differential modifications between TEs and active genes have been genetically investigated in *Arabidopsis*. Mutations in the chromatin-remodeller gene *DECREASE IN DNA METHYLATION 1* (*DDM1*) compromise DNA methylation at both CG and CH sites in TEs, although actively transcribed genes are largely unaffected[9,13,14]. Actively transcribed genes are affected by mutations in a Jumonji domain-containing histone demethylase gene, *INCREASE IN BONSAI METHYLATION 1* (*IBM1*). In *ibm1* mutants, CH methylation accumulates in expressed genes, although silent TEs are unaffected[15–18]. A remaining question is how these factors interact with others to establish the differential modifications between TEs and genes.

To understand how TE-specific DNA methylation is established, a powerful complementary approach is to investigate recovery of modifications after their loss. The *ddm1*-induced loss of DNA methylation is inherited to progeny even after the restoration of *DDM1* gene function[14,19]. However, slow remethylation is detected in TEs with matching 24-nucleotide (nt) small-interfering RNAs (siRNAs)[20], and the remethylation depends on RNA interference (RNAi)-based pathways or RNA-directed DNA methylation (RdDM). RdDM is an activity that directs de novo DNA methylation by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) through pathways triggered by double-stranded RNAs[21–26]. The observation suggests the importance of siRNA-based mechanisms in establishing DNA methylation at TEs. However, mutations in the RdDM machinery mainly affect TEs in euchromatic regions[9,27], suggesting that the RdDM machinery is dispensable at least for the maintenance of DNA methylation in more heterochromatic regions of TEs. In addition, the *ddm1* mutation abolishes methylation of CG, CH and H3K9 together. Interactions among modifications during the recovery remain largely unexplored, especially by genetic approaches.

Here, we directly investigated transgenerational de novo DNA methylation of TEs after the loss of CG and CH methylation separately. Our analyses uncovered potent and precise RNAi-independent pathways for recovering CH methylation and H3K9 methylation, specifically functioning in most TE genes (that is, coding regions within TEs). Characterization of a subset of TE genes without recovery revealed the impact of H3K9 demethylation, replacement of histone H2A variants and their interaction with CG methylation, together with feedback from transcription. Our analyses uncovered the importance of these unexpected pathways in epigenome differentiation.

Results

Recovery of CG methylation in TE genes is associated with siRNA. In plant genomes, TEs are enriched in both CG and CH methylation. To understand the dynamics of each modification, we examined the recovery of CG methylation and CH methylation separately, with methylation in one context lost and the other remaining. We first examined the recovery of CG methylation in the presence of CH methylation, using mutants of CG MTase gene *MET1*

[1]Department of Biological Sciences, The University of Tokyo, Tokyo, Japan. [2]Department of Integrated Genetics, National Institute of Genetics (NIG), Mishima, Shizuoka, Japan. [3]PREST, Japan Science and Technology Agency, Kawaguchi, Japan. [4]Center for Genetic Resource Information, National Institute of Genetics, Mishima, Shizuoka, Japan. [5]Gregor Mendel Institute, Austrian Academy of Sciences, Vienna, Austria. [6]These authors contributed equally: Taiko Kim To, Yuichiro Nishizawa, Soichi Inagaki, Yoshiaki Tarutani. ✉e-mail: tkt@bs.s.u-tokyo.ac.jp; tkak@bs.s.u-tokyo.ac.jp

and its three cofactor genes *VIM1*, *VIM2* and *VIM3* (*VARIANT IN METHYLATION*[28]) (Extended Data Fig. 1a). After genetically crossing the *met1* mutant and the *vim1 vim2 vim3* triple mutant, we examined de novo CG methylation in their F1 progeny (Extended Data Fig. 1a). The F1 progeny inherited CG-hypomethylated genomes from both parents in a background where wild-type (WT) *MET1* and *VIM* gene products are available (Extended Data Fig. 1a). Difference in dynamics was noted between TE genes (that is, coding regions within TEs) and normal protein coding genes (red and black dots in Extended Data Fig. 1b–g, respectively). Consistent with previous reports[29,30], remethylation occurred in TE genes, but not in protein coding genes (Extended Data Fig. 1b–g), and the remethylation efficiency of TE genes correlated with the abundance of matching siRNAs (Extended Data Fig. 1h–j); the remethylation was markedly less efficient in TE genes without high levels of 24-nt siRNAs, suggesting that RdDM is involved.

CH methylation in TE genes recovers efficiently. To examine the dynamics of CH methylation in the presence of CG methylation, we used *Arabidopsis* CH MTase and H3K9 MTase mutants (Fig. 1a). CMT2 and CMT3 methylate CH sites. These CH methylases are recruited to regions with H3K9me, which is a mark of silent chromatin[7,8]. Thus, CH methylation was abolished in the *cmt2 cmt3* double mutant (hereafter referred to as *cc*) and also in the triple mutant of three redundant H3K9 MTase genes *SUVH4*, *SUVH5* and *SUVH6* (hereafter referred to as *sss*), for both CHG (Fig. 1b,c) and CHH sites (Fig. 1f,g), while CG methylation was largely unaffected (Fig. 1j,k). We examined de novo CH methylation in F1 progeny resulting from the genetic crosses between *cc* and *sss*. The F1 progeny inherited CH hypomethylated genomes from both parents in a background where the WT *CMT* and *SUVH* gene products are available. In the F1 (*cc*×*sss*) progeny, most TE genes were CH remethylated substantially, approaching the level of the WT plants, for both CHG sites (Fig. 1d,n,o) and CHH sites (Fig. 1h,p,q). CH remethylation was consistently seen in the two individual F1 plants of reciprocal crosses in almost identical patterns (Fig. 1e,i). The consistent recovery of CH methylation was also evident at the single-nucleotide level (Fig. 1r–y), reflecting that a recovered CH methylation profile was very similar to the original WT pattern (Fig. 1n–q). Such consistent recovery of CH methylation is in contrast to ectopic and/or stochastic CH methylation found in *met1* or *ddm1* mutants, as well as their progeny outcrossed to WT[29,31–35]. While CG methylation is lost in *met1* or *ddm1* mutants, it is largely unaffected in *cc* and *sss* mutants and their F1 (Fig. 1). The robust and precise control of CH methylation may be related to the remaining CG methylation (further analyses later in Fig. 4).

CH methylation recovery in TE genes does not depend on RdDM. The roles of siRNAs in the de novo establishment of DNA methylation have been well investigated. Transgenerational remethylation from the *ddm1*-induced loss of DNA methylation is mainly found in TEs with matching 24-nt siRNAs[20] (Extended Data Fig. 3a,b, *x* axis). Similarly, CG remethylation in the presence of CH methylation also occurred for TEs with matching 24-nt siRNAs[29,30] (Extended Data Fig. 1h–j). In contrast, in F1 hybrids between *cc* and *sss*, CH remethylation occurred efficiently in a substantial number of TE genes with few siRNA detected in the *sss* mutant (Extended Data Fig. 3a,b, *y* axis), suggesting that siRNA-based de novo DNA methylation may be dispensable for the recovery of CH methylation.

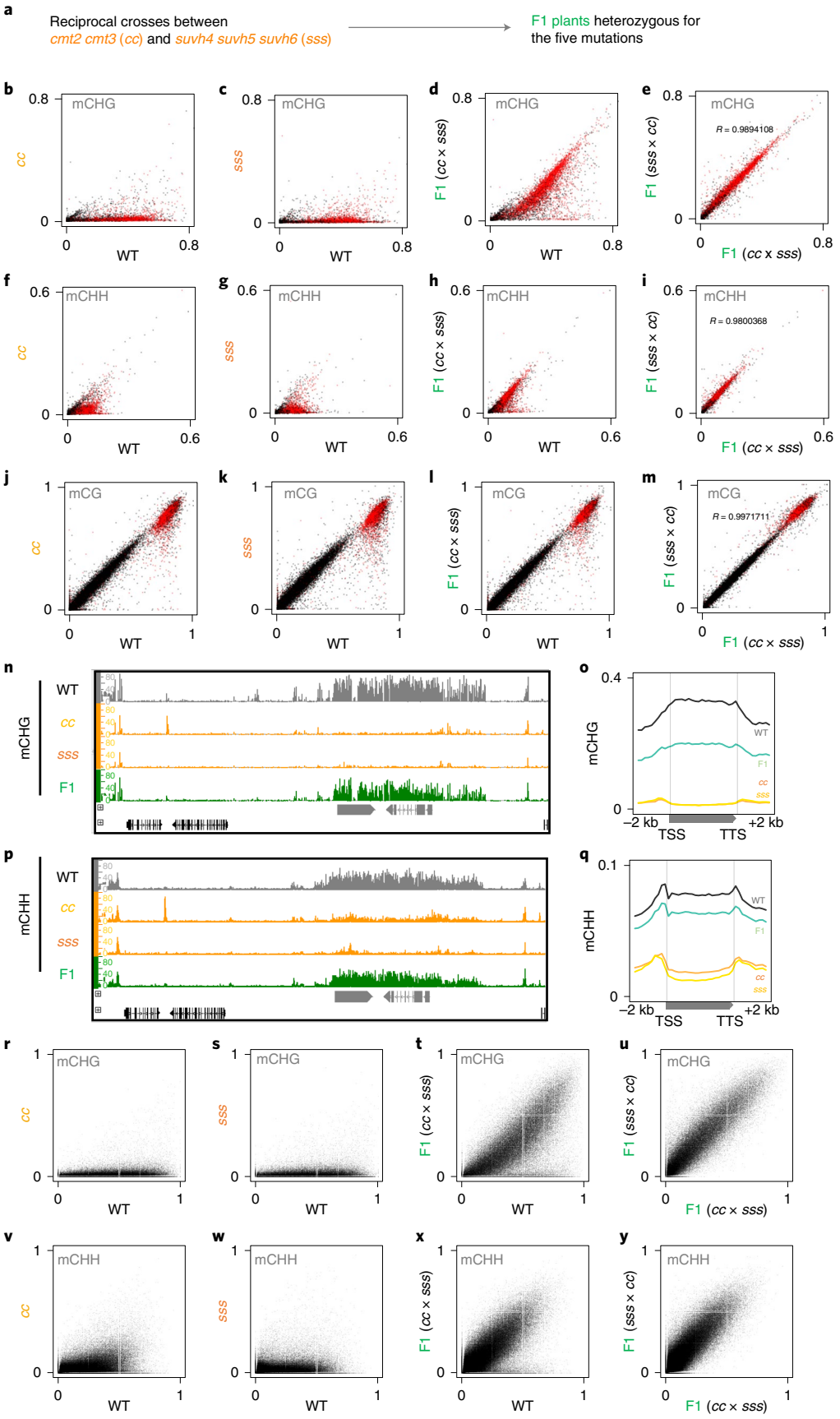
We therefore genetically tested whether this de novo CH methylation depends on RdDM, by repeating the *cc*×*sss* cross in a mutant *drm2* background. DRM2 is a de novo DNA MTase that is directed by RNAi[21–23,25]. We crossed the *drm2 cc* and *drm2 sss* mutants and examined the F1 hybrids (Fig. 2a). In the F1 hybrids, CH remethylation occurred efficiently for most TE genes even in the absence of a functional *DRM2* gene (Fig. 2c,f and Extended Data Fig. 3f). This efficient recovery of CH methylation in the *drm2* background was also evident at the single-nucleotide level (Extended Data Fig. 4a,b). In F1 hybrids with *DRM2* and *drm2* backgrounds, we also detected efficient recovery of H3K9me2 (Fig. 2h and Extended Data Fig. 3e), which directs the deposition of CH methylation. Consistent results were obtained when we tested the effects of mutations in other RdDM components, that is, RNA-dependent RNA polymerases (RDR1, RDR2 and RDR6), which are responsible for siRNA production[25]. We detected efficient CH methylation recovery in the F1 (*cc*×*sss*) hybrid in a background of triple mutations in *RDR1*, *RDR2* and *RDR6* (*rdr126*, Fig. 2a,d,g and Extended Data Fig. 3f), further supporting the conclusion that RdDM is dispensable for the efficient recovery of CH methylation and H3K9me2 in TE genes.

RdDM plays a central role in the recovery of CH methylation in non-coding regions of TEs. The analyses above were conducted on TE genes, which represent only the coding regions of TEs. We also examined impact of RdDM on TEs that do not contain coding regions, and observed very different DNA methylation dynamics (Extended Data Fig. 4c,d). In the F0 generation, the *drm2 cc* and *drm2 sss* mutants showed more extensive losses of CH methylation than *DRM2 cc* and *DRM2 sss*, respectively, suggesting that the remaining CH methylation of non-coding TEs in *cc* and *sss* depends on *DRM2* function. In the F1 hybrids, the remethylation efficiency was affected strongly by the *drm2* mutation in the non-coding TEs (Extended Data Fig. 4c,d), although it was not affected much in the TE genes (Extended Data Fig. 4a,b). Thus, although *DRM2* function was largely dispensable for CH remethylation in TE genes (coding regions of TEs), *DRM2* was involved in remethylation more in the non-coding TEs. Our analyses of these results and reanalysis of the published data[9] also indicate that, in addition to the non-coding TEs (shaded area in Extended Data Fig. 5, left), non-coding regions within TEs that contain TE genes are also targets of RdDM (shaded areas in Extended Data Fig. 5, middle and right). Together, these results suggest that RdDM affects non-coding regions of TEs and that coding regions can be remethylated in an RdDM-independent manner.

CH methylation recovery is less efficient in TE genes targeted by H3K9 demethylase, IBM1. Our results revealed that CH methylation generally recovered efficiently in the coding regions of most TE genes. In a subset of TE genes, however, the recovery was extremely inefficient, and that was reproducibly seen in multiple F1 plants (Fig. 3a). To identify factors affecting the RNAi-independent

Fig. 1 | CH methylation in TE genes recovers efficiently. a, Materials examined. **b**–**d**, CHG methylation levels of TE genes in the parental mutants, *cmt2 cmt3* (*cc*) (**b**), *suvh4 suvh5 suvh6* (*sss*) (**c**) or a F1 hybrid (**d**), compared to a WT plant. **e**, Comparison of two individual F1 plants of reciprocal crosses (*sss*×*cc* and *cc*×*sss*). *R* represents the Pearson's correlation coefficient between two methylation values in the TE genes. **f**–**i**, CHH methylation levels, as shown in **b**–**e**. **j**–**m**, CG methylation levels, as shown in **b**–**e**. **n**, Genome browser view of CHG methylation for a genomic region with a TE remethylated in F1 progeny from a cross between *cc* and *sss*. The coding regions of cellular genes (black) and TE genes (grey) are shown. The region shown corresponds to nucleotides 15888700–15927330 in chromosome 1. **o**, Averaged CHG methylation profiles around TE genes in a WT plant, two mutant parents and the F1 progeny. The grey arrow below represents the transcription unit, from the TSS to the TTS. In each genotype, mean values are shown for the TE genes. **p**,**q**, CHH methylation profile presented in the format used in **n** and **o**. **r**–**y**, CHG and CHH methylation levels were plotted for each cytosine within TE genes on chromosome 1. In each of these genotypes, metaplots of DNA methylation are also shown in Extended Data Fig. 2 on TE genes (**a**) and genes (**b**).

reprogramming, we selected 73 of these atypical TE genes with especially inefficient CHG methylation recovery (<10% of WT CHG methylation in both of reciprocal F1 hybrids, as represented by the red dots in Fig. 3a and listed in the Supplementary Table; although many additional TE genes showed intermediate level of the recovery, we used conservative criteria here). These TE genes also showed



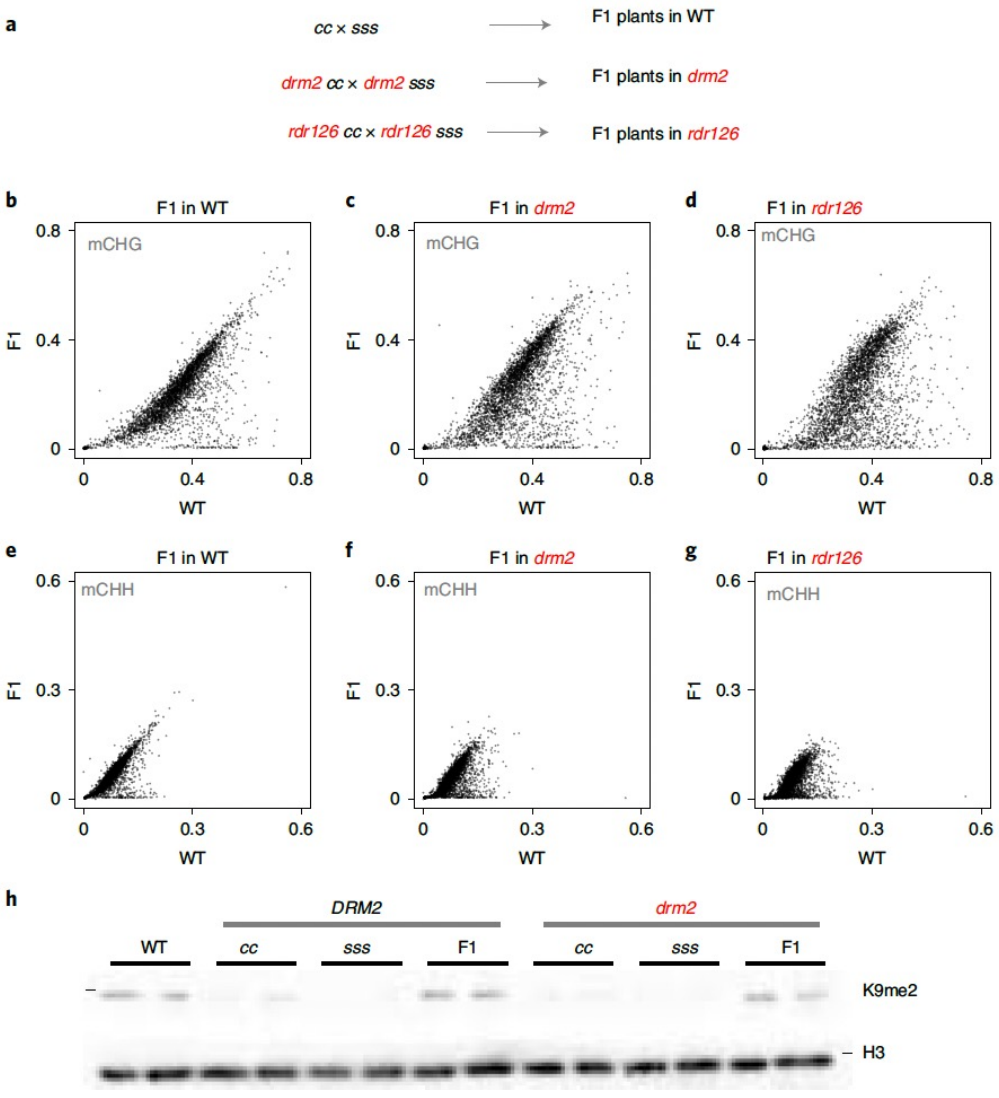


Fig. 2 | CH methylation recovery in TE genes does not depend on RdDM. a, Materials examined. **b–d**, CHG methylation of TE genes compared between WT and *cc* × *sss* individual F1 plants with backgrounds of WT (**b**), *drm2* mutation (**c**) or *rdr1 rdr2 rdr6* triple mutation (*rdr126*) (**d**). **e–g**, CHH methylation of TE genes compared between WT and *cc* × *sss* individual F1 plants with backgrounds of WT (**e**), *drm2* mutation (**f**) or *rdr1 rdr2 rdr6* triple mutation (*rdr126*) (**g**). **h**, Western blot analysis of H3K9me2 levels in the F0 mutants and F1 hybrid plants in *DRM2* and *drm2* backgrounds. The blot of H3 is shown below for the control. Black bars show positions of molecular weight marker of the size 17 kDa. Two biological replicates were examined for each genotype. Another biological replicate is shown in Extended Data Fig. 3e. Source data are provided.

defects in the recovery of CHH methylation and H3K9me2 (red dots in Extended Data Fig. 6a,b).

We have previously shown that the absence of heterochromatin marks (for example, H3K9me and CH methylation) from active genes depends on the H3K9 demethylase, IBM1 (refs. [15–17]). Thus, the inefficient CH methylation recovery in these TE genes could be due to IBM1 activity. To test this possibility, we examined the effects of *ibm1* mutation on their CH methylation recovery in the F1 hybrids between the *sss* and *cc* mutants (Fig. 3b). Indeed, many of these TE genes showed substantial CH methylation recovery in the *ibm1* mutant background for both CHG sites (Fig. 3c,d,g) and CHH sites (Fig. 3e,f,h), suggesting that H3K9 removal by IBM1 contributes to the difference between TE genes with and without the efficient recovery of CH methylation.

The TE genes affected by IBM1 did not gain CH methylation even in the presence of functional CMTs and SUVHs in the F1 progeny, and the CH methylation pattern of these TE genes in the F1 hybrids (Extended Data Fig. 2c, right) was almost indistinguishable from that of normal protein coding genes (Extended Data Fig. 2b), which lack CG and CH methylation at the transcription start site (TSS), even though the internal regions (bodies) are often CG methylated[36–38]. Thus, we designated these TE genes as gene-like TE genes (GLTs). GLTs showed gene-like epigenetic modification patterns not only in CH methylation but also in CG methylation and other epigenetic marks, as described below.

TE genes without efficient CH methylation recovery show reduced CG methylation. GLTs (that is, TE genes without efficient CH methylation recovery) showed a loss of CG methylation in the F0 mutants (*cc* and *sss*) and their F1 progeny (Fig. 4a,b). When TE genes were sorted according to the levels of CHG methylation recovery (left part in Fig. 4b), the TE genes with the least efficient CHG methylation recovery showed a strong reduction of CG methylation, especially around TSSs and transcription termination sites (TTSs) (Fig. 4b). Thus, GLTs showed not only CH but also CG methylation pattern similar to cellular genes, which

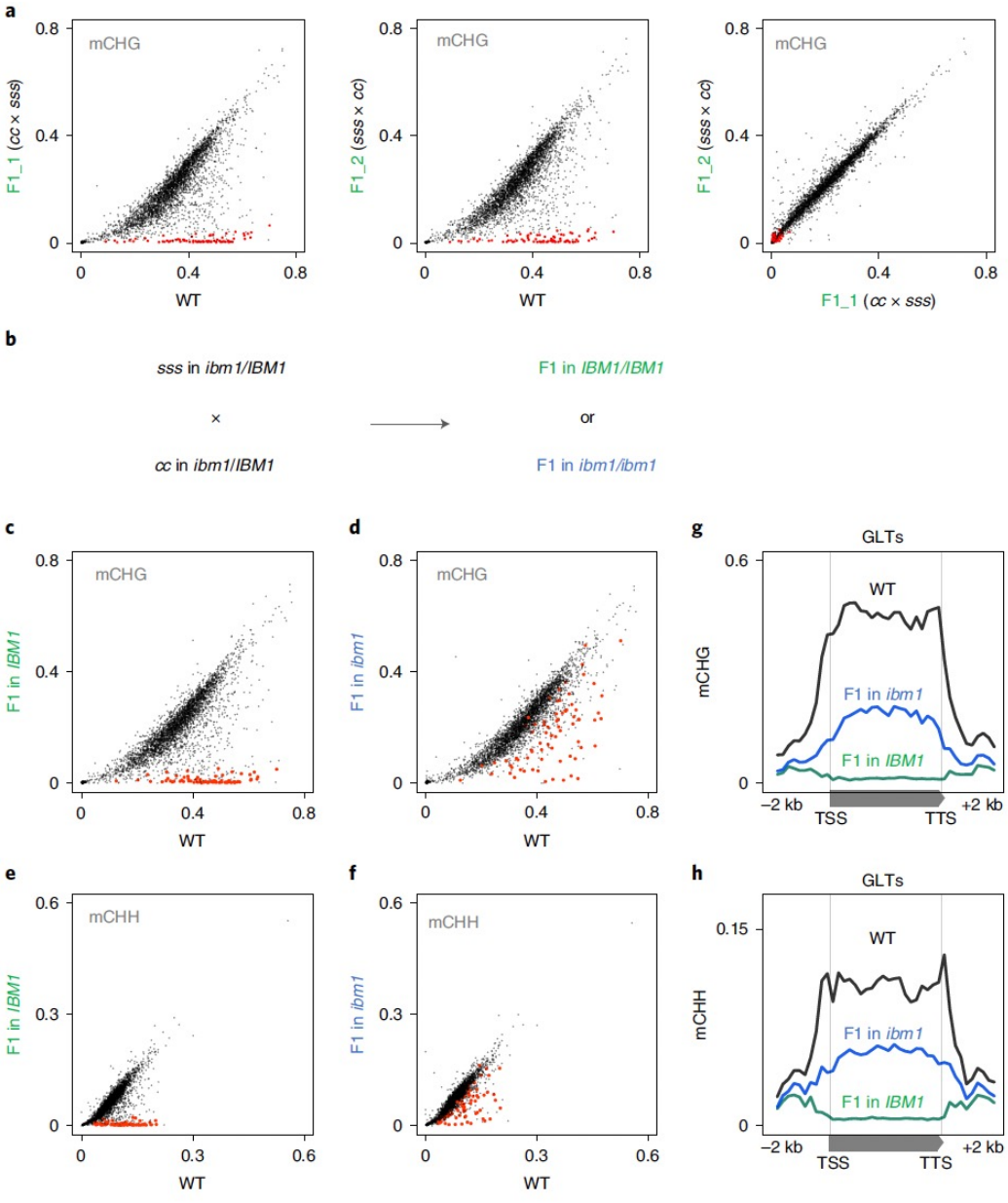
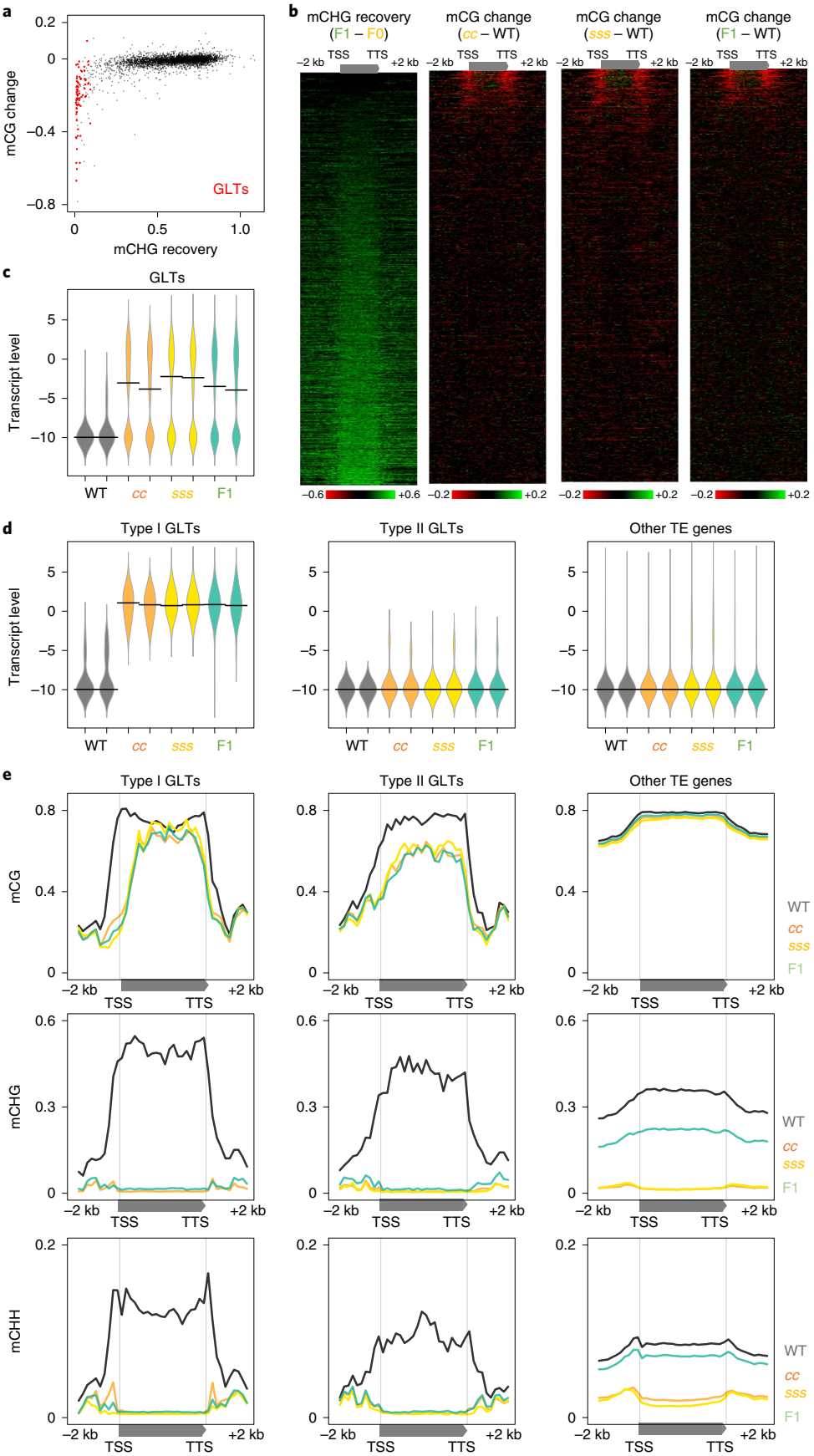


Fig. 3 | CH methylation recovery is less efficient in TE genes targeted by H3K9 demethylase, IBM1. a, A subset of TE genes do not recover CHG methylation. The original data for panels on the left and right are the same as those shown in Fig. 1d,e, respectively. TE genes with <10% of WT CHG methylation levels in both F1 hybrids of reciprocal crosses ($n=73$) are highlighted in red. We designated these TE genes GLTs. **b**, Genetic scheme used to examine effect of *ibm1* mutation on CH methylation recovery. **c**–**f**, CHG and CHH methylation recovery in F1 progenies of *IBM1/IBM1* (**c**,**e**) and *ibm1/ibm1* (**d**,**f**) backgrounds. GLTs ($n=73$) and other TE genes ($n=3,381$) are represented by red and black dots, respectively. GLTs are highlighted with a bigger dot size. **g**,**h**, CHG (**g**) and CHH (**h**) methylation profiles of GLTs in the F1s in *IBM1* and *ibm1* backgrounds, with WT data shown as controls.

do not have CG methylation in TSSs and TTSs (Extended Data Fig. 2b,c)[36–38].

GLTs can be divided into two groups of TE genes with high and low levels of transcriptional derepression in *cc* and *sss* mutants, and their F1 progeny (Fig. 4c,d). We classify them as type I and type II GLTs, respectively. Type I GLTs were enriched in *COPIA* family retrotransposons, which consistently tend to be long, while type II GLTs were enriched in *COPIA* and *LINE* families of retrotransposons (Extended Data Fig. 6c,d and Supplementary Table). In WT plants, both types of GLT are heavily methylated at both CG and CH sites, to a level comparable to the other TE genes (Fig. 4e and Extended Data Fig. 6e,f). However, once methylation is lost in the *cc* and *sss* mutants, the methylation does not recover in the F1 hybrids and the plants show DNA methylation profiles similar to those of normal genes (Fig. 4e and Extended Data Fig. 2b,c). In other words, GLTs show epigenetic profiles of normal TE genes in the WT plants, but after passing through one generation with the *cc* or *sss* mutations, they acquire the epigenetic features of normal genes, even in the presence of functional CMTs and SUVHs in the F1 progeny.

One possible explanation for the CG methylation loss observed after the loss of CH methylation could be through transcriptional derepression. However, both GLTs with and without transcriptional derepression (types I and II, respectively) showed comparable CG



methylation loss in the *cc* and *sss* mutants, and their F1 progeny (Fig. 4e, top), suggesting that transcriptional derepression alone does not account for the observed CG methylation loss.

Dynamics of H2A variants in GLTs are distinct from those in the other TE genes. In general, CG methylation correlates negatively with the histone variant, H2A.Z[39,40]. Mutations affecting

Fig. 4 | TE genes without efficient CH methylation recovery show reduced CG methylation. a, TE genes without efficient CHG methylation recovery (that is, GLTs, $n=73$, red dots) show reduced CG methylation in the F1 hybrid. In each TE gene, *x* and *y* axes represent the ratio of mCHG levels in F1 against WT (F1/WT) and the difference in mCG levels (F1–WT), respectively. TE genes with low CH methylation levels (mCHG <0.1 or mCHH <0.03) were excluded ($n=444$) to avoid division by values near zero, and the remaining TE genes were analysed ($n=3,459$). **b**, TE genes are aligned on the basis of the efficiency of CHG methylation recovery in a *cc* × *sss* F1 hybrid (left, TE genes with the least efficient recovery are shown on top, making GLTs concentrated there). GLTs, that is, TE genes with less efficient CHG methylation recovery, show a loss of CG methylation, especially around TSS and TTS in the parental mutants (*cc* and *sss*) and their F1 progeny. Genome browser views of genomic regions with GLTs are shown in Extended Data Fig. 6f. TE genes with low CH methylation (mCHG <0.1 or mCHH <0.03) and short TE genes (<2 kb) were excluded, and the remaining TE genes ($n=1,872$) were analysed. **c**, GLTs ($n=73$) include two populations of TE genes, with and without transcriptional derepression in mutants and F1. Beanplot shows GLT expression levels in WT, mutant (*cc* and *sss*) and F1 plants, shown by $\log_2$ RPKM. Each horizontal black line represents the median. For each genotype, RNA expression levels were examined in two independent replicates. **d**, GLTs are classified into two populations: that is, those with high (RPKM > 0.1 in both *cc* and *sss*, $n=34$) and low (RPKM < 0.1 in both *cc* and *sss*, $n=31$) levels, shown as type I GLTs and type II GLTs, respectively. Beanplots showing transcription levels for each type of GLTs as well as other TE genes ($n=3,381$) are in the same format shown in **c**. **e**, Metaplots of CG, CHG and CHH methylation in type I and type II GLTs, and the other TE genes ($n=34$, 31 and 3,381, respectively).

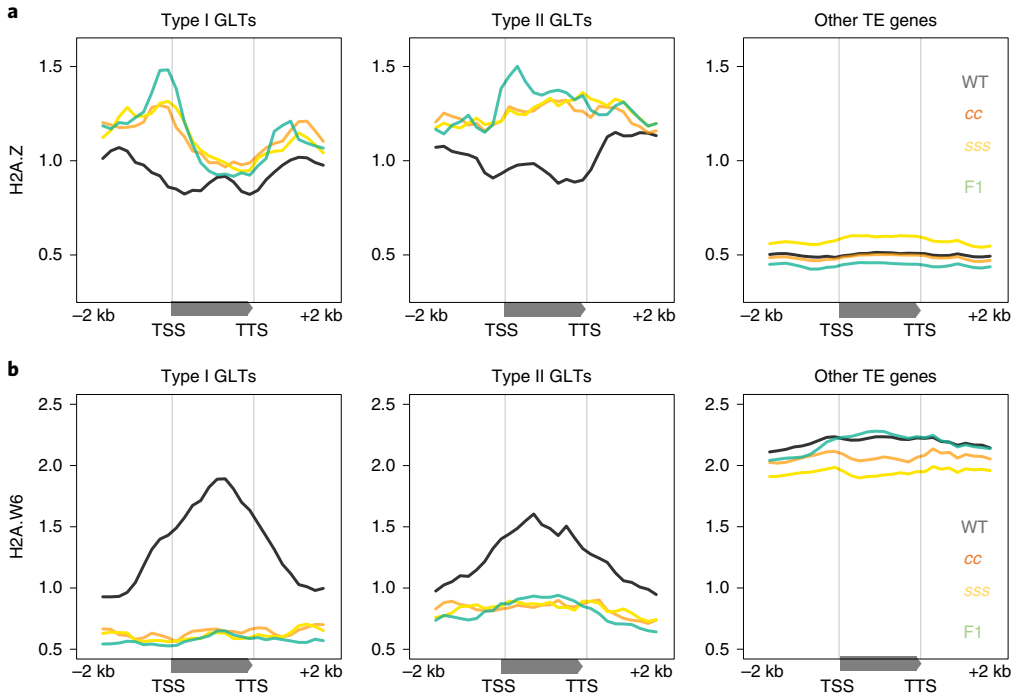


Fig. 5 | Dynamics of H2A variants in GLTs are distinct from those in the other TE genes. a,**b**, Results of ChIP–seq experiments performed to detect H2A.Z (**a**) and H2A.W6 (**b**). The mean of RPKM value for each segment was normalized to that of input DNA. Type I and II GLTs and the other TE genes are shown ($n=34$, 31 and 3,381, respectively). The results for H2A.W7, another H2A.W protein[43], show a pattern similar to that of H2A.W6 (Extended Data Fig. 7a). The ChIP–seq results for biological replicates are shown in Extended Data Fig. 7b.

CG methylation or H2A.Z result in ectopic accumulation of the other epigenetic mark, at least partially[39,41,42], suggesting mutual inhibitions after their establishment. However, the effects of these H2A variants on dynamics of other epigenetic marks are less well characterized. Consistent with the previous reports, our chromatin-immunoprecipitation coupled with high-throughput sequencing (ChIP–seq) results revealed that the loss of CG methylation in GLTs was associated with a gain of H2A.Z in *cc* and *sss* mutants, as well as in their hybrid (Fig. 5a and Extended Data Fig. 7b). The gain of H2A.Z in GLTs was associated with loss of H2A.W, the H2A variant that marks constitutive heterochromatin[43] (Fig. 5b and Extended Data Fig. 7a,b). This replacement of H2A variants occurred specifically at GLTs; TE genes other than GLTs did not show comparable loss of H2A.W or gain of H2A.Z (Fig. 5a,b and Extended Data Fig. 7a,b).

Crosstalk among CG methylation, H2A.Z and transcription. The correlations between CG methylation losses and H2A.Z gains in GLTs are consistent, suggesting a possible link between the changes in these two marks[39,41,42]. We then genetically tested whether the gain of H2A.Z can act upstream of the loss of CG methylation in GLTs, using mutants of three redundant H2A.Z genes (*HTA8*, *HTA9* and *HTA11*)[41] in *Arabidopsis*. As CG methylation changes can often be inherited over multiple plant generations, we separately examined the effects of H2A.Z on the induction and maintenance of CG methylation changes induced by the *cc* mutation. In the genetic scheme shown in Fig. 6a, the triple mutations of three H2A.Z genes (hereafter referred to as *zzz*) were fixed to be homozygous before or after fixation of the *cc* mutations to be homozygous. The effect of H2A.Z mutations differed depending on the order of fixation and depending on the states of transcriptional derepression, as described below.

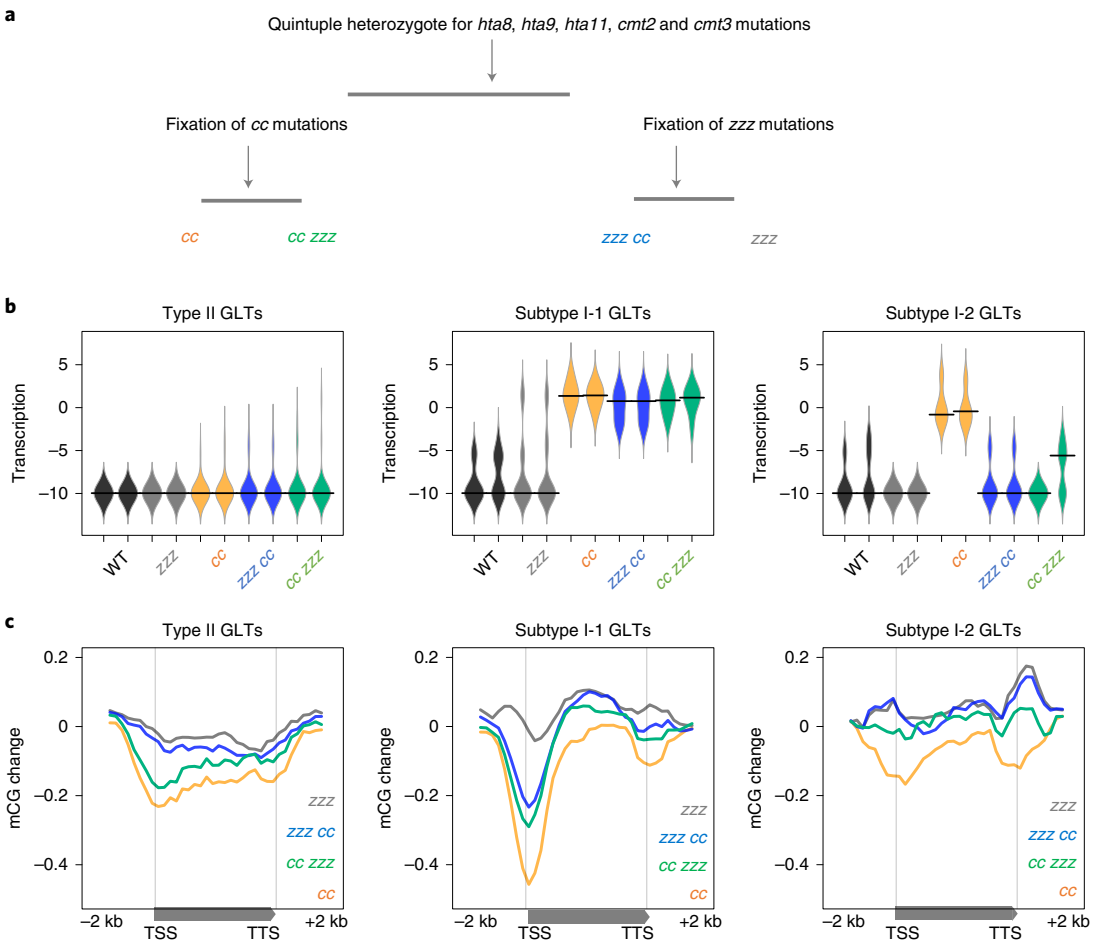


Fig. 6 | Crosstalk among CG methylation, H2A.Z and transcription. a, Genetic scheme used to study the effects of mutations in three redundant *H2A.Z* genes (*HTA8*, *HTA9* and *HTA11*) on the *cc*-induced loss of CG methylation. From self-pollinated progeny of a quintuple heterozygote (top), *H2A.Z* triple mutations (*hta8 hta9 hta11* (*zzz*)) or CMT double mutations (*cc*) were fixed as homozygous first (middle), and the other mutations were fixed as homozygous in the next generation (bottom). Siblings with mutations only in *cc* or *zzz* were examined in parallel as controls. Since the plants contain a genome fraction of Ws ecotype that originates from the *hta8* mutant[41], TE genes that were not mapped in at least one sample were excluded from the analysis. **b**, Subtypes of GLT with different transcription patterns. Type I, which is transcribed in *cc*, are further divided to subtype I-1 ($n=13$) and I-2 ($n=5$), depending on the transcription in the quintuple mutants. Beanplot showing transcription levels for each subtype of GLTs, as well as Type II GLTs ($n=27$) in the same format shown in Fig. 4c. **c**, Effect of *zzz* mutations on the *cc*-induced loss of CG methylation in each subtype of GLTs. Metaplots showing the CG methylation difference for each mutant compared to the WT plants. Metaplots and beeswarm plots of the original CG methylation patterns for each genotype are shown in Extended Data Fig. 6b,c.

In the type II GLTs, which remain silent (Fig. 6b, left), the CG methylation loss was suppressed almost completely if *zzz* mutations were fixed before the fixation of *cc* mutations ('*zzz cc*' in Fig. 6c, left). In contrast, if order of fixation was reversed ('*cc zzz*'), the CG methylation loss largely remained, indicating that H2A.Z is dispensable for inheritance of the CG methylation loss. Thus, H2A.Z is necessary for the induction, but not inheritance, of the *cc*-induced CG methylation loss in the type II GLTs.

A contrasting pattern was observed in the type I GLTs. Type I GLTs showed transcriptional derepression in *cc*, but a few of type I GLTs were silenced in the background of additional *zzz* mutations (Extended Data Fig. 8a, left). We therefore separately analysed DNA methylation in type I GLTs with and without transcriptional repression in the quintuple mutants (subtype I-2 and I-1 GLTs, respectively, Fig. 6b–d). In the subtype I-1 GLTs, which were transcribed in the quintuple mutants, the *cc*-induced loss of CG methylation was suppressed only partially by *zzz* mutation, irrespective of the timing of the fixation of the *cc* mutations and *zzz* mutations (*zzz cc* and *cc zzz* in Fig. 6c,d, centre). We suspected that the differences in the effects of *zzz* mutation on subtype I-1 and type II GLTs reflect their transcriptional states. Consistent with this idea, the subtype I-2 GLTs, which showed transcriptional repression in the quintuple mutants, showed almost complete suppression of *cc*-induced CG methylation loss in *zzz cc* (Fig. 6c,d, right), as was the case in type II GLTs. These results suggest that H2A.Z and transcription additively induce the CG methylation loss.

GLTs showed properties different from those of TE genes affected by RNAi. Our analyses above on GLTs revealed association of the CH methylation recovery with CG methylation and H2A variants. GLTs are TE genes that do not recover CH methylation even in the presence of RdDM machinery. On the other hand, a subset of TE genes did not recover CH methylation only when RdDM machinery are non-functional (Fig. 2b–g and Extended Data Fig. 9a). We therefore designated them as RNAi-dependent TE genes (RDTs), and separately characterized three categories of TE genes: GLTs (no CH methylation recovery, $n=73$), RDTs (no CH methylation recovery only when RdDM is non-functional, $n=134$) and the other TE

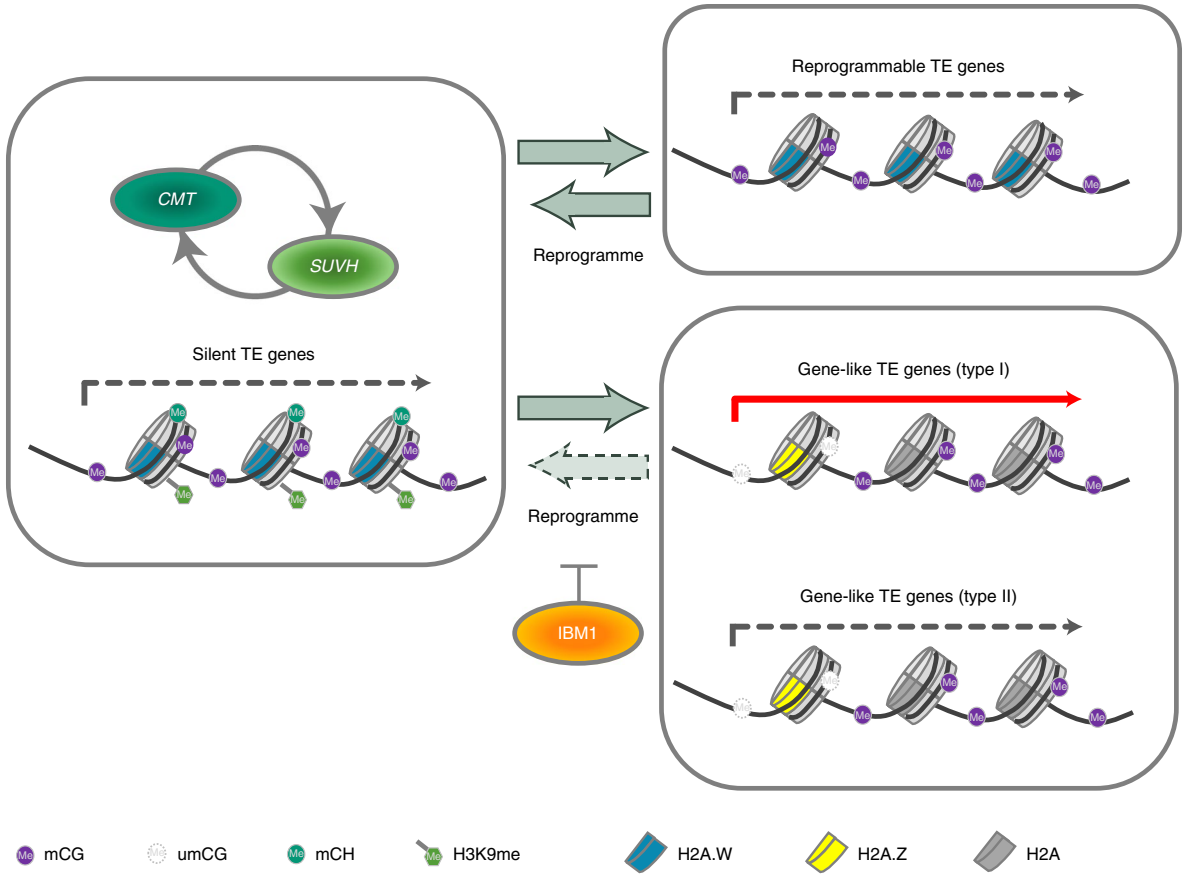


Fig. 7 | Model of the modification dynamics of TE genes during the loss and gain of heterochromatin machinery. Left, in WT plants, TE genes are silenced by H3K9me and DNA methylation of CG and CH sites (shown as mCG and mCH, respectively) with the histone variant, H2A.W. CH methylases, CMTs and H3K9 methylases, SUVHs, generate a positive feedback loop to maintain mCH and H3K9me. Right, in plants with mutations in the *CMT* or *SUVH* genes, both H3K9me and mCH were lost. Most TE genes (top right) retain H2A.W and mCG, and they acquire H3K9me and mCH efficiently when SUVHs and CMTs are available (Reprogram, arrow pointing to the left). A subset of TE genes, referred to here as GLTs, show distinct modification dynamics (bottom right). The GLTs lose H2A.W and gain H2A.Z (Fig. 5). The gain of H2A.Z induces loss of mCG (Fig. 6). This type of modification pattern does not recover efficiently, even in conditions where both SUVH and CMT proteins are available (Figs. 3 and 4). This lack of remethylation in GLTs depends on the H3K9 demethylase, IBM1 (Fig. 3).

genes (RNAi-independent CH methylation recovery, $n = 3{,}247$). With regards to the CH methylation recovery and RNAi, the question is, how do the properties of GLTs and RDTs overlap?

In the *Arabidopsis* genome, TE genes are enriched in pericentromeric regions, but GLTs and RDTs are both localized in chromosomal arm regions (Extended Data Fig. 9b,c). They are associated with a relatively low level of H2A.W and high level of H2A.Z, compared to the other TE genes (Extended Data Fig. 9d). Despite these similarities, GLTs and RDTs are different in the H3K9me level and in response to *cc* and *sss* mutations. RDTs are short and euchromatic (low in H3K9me2 level, Extended Data Fig. 9e,f), which is consistent with previous reports about properties of RdDM targets[9,27]. In contrast, GLTs, despite their localization in chromosome arms, have high levels of H3K9me, comparable to those in the other TE genes (Extended Data Fig. 9f). In *cc* or *sss* mutants, GLTs showed reduced CG methylation, but the change was subtle in RDTs (Extended Data Fig. 9g) even in a *drm2* or *rdr126* background, where the CH methylation recovery was compromised. In addition, RDTs did not show the replacement of H2A.W to H2A.Z in *cc* or *sss* mutants as in GLTs (Extended Data Fig. 9h). These differences further support our conclusion that the links of CH methylation recovery to CG methylation and H2A are distinct from the well-characterized RNAi pathway.

Discussion

Generally, CG and CH methylation redundantly silence TEs[12,44,45]. CG methylation loss, induced by a *ddm1* or *met1* mutant, does not recover efficiently even in a WT background, and the recovery only occurs in sequences with abundant matching small RNAs[20,29,30]. Here, we report that CH methylation and H3K9me2 recover very efficiently and globally in most of coding regions of TEs when they are CG methylated. Unlike the recovery of CG methylation, the recovery of CH methylation and H3K9me2 does not depend on the well-characterized RNAi-based de novo DNA methylation pathway, suggesting the existence of previously uncharacterized pathways.

Regarding the differential modifications of active genes and TE genes, an important question arising from this study is, what mark(s), other than siRNA, does a host recognize to introduce CH methylation and H3K9me specifically to coding regions of TEs? A potential candidate mark could be DNA methylation at CG sites (Fig. 7). Generally, CH remethylation efficiency correlated with CG methylation at TSSs (Fig. 4a,b). CG methylation in TSSs is a feature specific to TE genes and does not occur in cellular genes. In addition, the recovery of DNA methylation was slow from the *ddm1* mutant[14,19,20], which loses not only CH but also CG methylation in TEs. An alternative, but not mutually exclusive factor that determines the epigenetic states of TE genes could be H2A

variants, such as H2A.W and H2A.Z. While H2A.W localizes to silent TEs, H2A.Z localizes to active genes. The genome-wide distribution of H2A.W matches well with that of H3K9me2, but, the global pattern of H2A.W does not change in *sss* and *cc* mutants[43] (Fig. 5 and Extended Data Fig. 7). Thus, H2A.W (or the lack of H2A.Z) might allow the cell to memorize where to re-introduce H3K9me2 and CH methylation. Consistent with this hypothesis, H2A.W was markedly decreased (and H2A.Z increased) in the GLTs (that is, TE genes without efficient CH methylation recovery). In addition, our genetic analyses revealed that H2A.Z functions upstream of the CG methylation loss and transcriptional derepression of GLTs. As CG methylation and H2A.Z can control each other[39,41,42], it is possible that both CG methylation and H2A variants mark the TE genes (Fig. 7).

This study revealed that while non-coding regions are targets of RNAi-induced silencing, coding transcription units can acquire silent marks in an RNAi-independent manner, indicating potential involvement of transcription for epigenome reprograming. In fact, the involvement of transcription was seen in the crosstalk among modifications mentioned above. Mutations of H2A.Z (*zzz*) suppressed the induction of CG hypomethylation by mutations in CH methylases (*cc*) almost completely if GLTs were silent (Fig. 6c), but only partially if transcribed (subtype I-1 GLTs in Fig. 6c), suggesting that the *cc*-induced loss of CG methylation is due to parallel effects of H2A.Z accumulation and transcriptional derepression. H2A.Z is necessary for the maintenance of CG methylation loss in type I GLTs, but not in type II GLTs (Fig. 6c), suggesting that H2A.Z could also antagonize CG methylation recovery that occurs in type I GLT. As this recovery occurs only in type I GLTs, but not in type II GLTs, transcription is associated with the recovery of DNA methylation. Consistent with this observation, it has been reported that TE sequences newly introduced into the *Arabidopsis* genome acquire heritable DNA methylation marks efficiently when they have intact transcription units[46].

Genomes of vertebrates and plants contain a substantial proportion of TE sequences to be silenced. RNAi should be of essential importance for silencing foreign sequences introduced into the genome[21,46–48]. On the other hand, remethylation by RNAi has been often associated with ectopic DNA methylation[29,30]. Although RNAi has a high impact on epigenome dynamics, it should be controlled in trade-off conditions to ensure appropriate expression of nearby genes[49]. The RNAi-independent reprograming pathways we uncovered here function in more robust and precise manners, and control most TE genes globally. Such mechanisms should be important for coordinated gene functions, especially in organisms with large and complex genomes, with a high proportion of TE sequences to control.

Our results also suggest that this RNAi-independent reprograming involves crosstalk among CG methylation, H2A variants and H3K9me, with feedback from transcription. These chromatin components are conserved among eukaryotes[50], and their importance for epigenome reprogramming may also be conserved.

Methods

Plant materials and growth conditions. The mutants *met1-6*, *vim123*, *cmt23* and *suvh456, rdr126* and *h2a.z* (*hta8 hta9 hta11*) are kind gifts from R. Fischer, E. Richards, J. Bender and D. Zilberman[9,28,41,51–53]. First-generation homozygous mutants were used for genetic crosses to create the F1 plants. In the experiment to see effect of mutations of *H2A.Z* genes on the *cmt2 cmt3* mutation (Fig. 6), progeny from two rounds of self-pollinations from a quintuple heterozygote plant, *HTA8/hta8 HTA9/hta9 HTA11/hta11 CMT2/cmt2 CMT3/cmt3* were used. The plant shown as *cc zzz* and *zzz cc* in Fig. 6 were fixed for the *zzz* and *cc* mutations, first by self-pollination of the quintuple heterozygote and subsequently fixed for other mutations, respectively; *cc zzz* and *cc* are from self-pollinated progeny of an *HTA8/hta8 HTA9/hta9 HTA11/hta11 cmt2/cmt2 cmt3/cmt3* plant, and *zzz cc* and *zzz* are from self-pollinated progeny of an *HTA8/hta8 HTA9/hta9 HTA11/hta11 cmt2/cmt2 cmt3/cmt3* plant. The original quintuple heterozygote was generated by crossing of double *cc* heterozygous mutant and *zzz* triple heterozygote generated by three times of backcrossing an *zzz* triple mutant[41] to a Col WT plant to reduce the possible effects of the WS background of the original *hta8-1* mutant. For F2 plants from the cross between *met1-6* and *vim123* (Extended Data Fig. 1), F1 plants of *met1-6* and *vim123* all in heterozygotes were self-pollinated, and the plants with the genotype *MET1/MET1 VIM1/VIM1 VIM2/VIM2 vim3/vim3* were used.

The annotations of genes and TEs are based on The *Arabidopsis* Information Resource (TAIR)[54]. The details of the annotation of TE genes are provided on the TAIR website (https://www.arabidopsis.org).

Whole genome bisulfite sequencing (WGBS). Genomic DNA was isolated from the rosette leaves of individual plants using the Nucleon Phytopure genomic DNA extraction kit (GE Healthcare), and WGBS was performed as described previously[55]. Two independent biological replicates for F1 plants were analysed, with the exception of F1s from an *ibm1/IBM1* background. Genomic DNA was fragmented using Covaris S220, and sizes between 300 and 450 bp were gel purified. The libraries were prepared using the TruSeq DNA LT Sample Prep Kit (Illumina) and subjected to bisulfite conversion using the MethylCode Bisulfite Conversion Kit (Life Technologies). Bisulfite-treated DNA were amplified with KAPA HiFi HotStart Uracil ReadyMix (Kapa Biosystems) and then purified with Agencourt AMPure XP (Beckman Coulter).

Raw sequence data were deposited in the Gene Expression Omnibus (GEO) (GSE148753). The raw sequence data were quality checked by fastqc (v.1.8.0). The adaptor sequences and low-quality regions were trimmed using Trimmomatic[56]. Mapping to the reference genome (TAIR10), deduplication and methylation extraction were performed using Bismark v.0.10.1 (ref. [57]). The parameters used for mapping were -n 1 –l 20 –e 90. Uniquely aligned reads were used and reads with more than one alignment with the same read quality were removed. The cytosine methylation levels of each context within a gene were calculated using the ratio of the number of methylated cytosines to that of total cytosines within the coding regions (weighted methylation level[58]). We used Perl script to count the numbers of methylated and total cytosines and R studio (v.1.1.463) to plot DNA methylation patterns. For browser views, bedgraph files were created and visualized using the Integrated Genome Browser[59]. To create metaplots around TE genes, TE genes and their neighbouring regions (2 kb) were divided into 20 and ten segments, respectively, then the values of methylated cytosines over total cytosines were calculated for each segment using Perl script. For heatmaps in Fig. 4b and Extended Data Fig. 6e, TE genes with low CH methylation (mCHG <0.1 or mCHH <0.03) and short TE (<2 kb) genes were excluded and the remaining TE genes ($n = 1{,}872$) were analysed. The processed data were visualized via TreeView 3 (ref. [60]). To calculate methylation for each cytosine in Fig. 1 and Extended Data Fig. 4, only cytosines with a read depth >9 were used. Perl script was used to select cytosines within coding regions of TEs (but not in non-coding regions of TEs), or cytosines within non-coding TEs (but not in coding regions of TEs). The 'remethylation level' of the progenies shown in Supplementary Table were defined as the methylation level of the progeny divided by that of WT. The methylation recovery shown in Extended Data Fig. 3a–d was calculated as follows. First, F1 methylation was subtracted by their parental DNA methylation (F1–F0), and then divided by the loss of methylation in the parents: (F1–F0)/(WT–F0). To calculate the remethylation efficiency, the TE genes showing a small methylation decrease in the parents (for CG and CHG (WT, mid-parental value) <0.1; for CHH (WT, mid-parental value) <0.05) were excluded to avoid division by values near zero. As for the analysis using *ddm1*-epiallele F9 (Extended Data Fig. 3), the TE genes, which were originally from *ddm1*-derived chromosomes, were used for the analysis.

To test the differential regulation between coding and non-coding regions (Extended Data Fig. 5), TE sequences (TAIR) were classified according to the number containing TE genes: no TE gene (TEGn0, $n = 28{,}971$), one TE gene (TEGn1, $n = 1{,}772$) or two or more TE coding genes (TEGn2, $n = 446$). The coding gene (if there was one or more) of each TE sequence was divided into 20 segments, and each of the non-coding regions (edges or intergenic regions) as well as their neighbouring sequences into ten segments. If the TE contained more than two TE genes, the two outermost genes were analysed. Perl script was used to calculate the total read counts and methylation levels in each segment. The data for *ddm1*, *rdr2* and WT shown in Extended Data Fig. 5a are from reanalysis of the published data (GSE41302)[9].

Methylation-sensitive restriction enzyme–quantitative PCR (MSRE–qPCR). To analyse the methylation levels of specific regions, MSRE–qPCR was performed. Genomic DNA was digested with the restriction enzyme *Pst* I (TaKaRa Bio), which digests CTGCAG site when the first C is unmethylated. The digested DNA was quantified by real-time qPCR for specific regions using Light Cycler 480 (Roche) and KAPA SYBR Fast qPCR kit (Roche). Region without *Pst* I site (*AT3G02515*) was also analysed for the control and used to normalize the amount of DNA. The primers used were as follows; 5′-CAAGGGATTGATATGGTACTAG-3′ and 5′-CGAAACTTCTTGTACACATAACATGGC-3′ for *ATGP3-1*, 5′-CATGA CATTGGCTCTTCCACTC-3′ and 5′-GAAGCAAACACTGAGTATGACTCG-3′ for *VANDAL14*, 5′-CGCCATTCGTGTTGTTCGCTA-3′ and 5′-GAGGGCAAGC AGCATAGTCAC-3′ for *COPIA22*, 5′-TCGTTGCTTTGGTTTCCACGC-3′ and 5′-GACCGACGTTGAACAGCTACG-3′ for *AT3G02515*.

ChIP–seq analysis. Seedlings grown for 16 d on MS Agar plates were frozen in liquid nitrogen, ground and fixed in 0.5% formaldehyde. After nuclei isolation, Micrococcal nuclease (Sigma N3755) was used to digest chromatin into almost one nucleosome length. Nuclear membrane was disrupted by brief sonication (TOMY UD-201; output, 1; 2 s). After centrifugation, the supernatant was used for immunoprecipitation. Antibodies used were anti-H2A.W6 and H2A.W7 (ref. [43]), anti-H3K9me2 (MABI0307), anti-HA tag (3F10, Roche) and anti-H2A.Z. The H2A.Z antibody was raised against the H2A.Z-specific peptide KPSGSDKDKDKKKP[43] and validated by comparing with the previously published results for the H2A.Z antibody[43], as shown in Supplementary Fig. 1. Libraries were prepared using SMARTer ThruPLEX DNA-seq Kit. The 50-bp single-end sequences were obtained with a HiSeq4000 sequencer (Illumina) in The V.J. Coates Genomics Sequencing Laboratory at UC Berkeley. Two independent biological replicates were analysed for each genotype. The raw sequence data were quality checked by fastqc (v.1.8.0_60). The reads were quality filtered using Trimmomatic[56] and mapped to TAIR10 genome using Bowtie[61] with the option -M 1–best. Metaplots were created by counting the mapped reads within each segment of genes or outer regions using Perl script, reads per kilobase of transcript per million (RPKM) normalized and then further normalized by that of input DNA. Rstudio (v.1.1.463) was used to plot the normalized values.

To perform spike-in normalization, we created a *Schizosaccharomyces pombe* strain with HA-tagged histone, TTY001. A DNA fragment containing HA-tagged histone HTB1 with a Leu1 marker was made via a PCR-based method and transformed into the strain SFY1 by lithium acetate recombination. The transformant was selected on EMM–Leu agar plates (EMM agar plate supplemented with glucose, uracil, histidine, lysine and adenine), and confirmed by DNA sequencing around the HTB1 site. Yeast ChIP was performed as follows. The TTY001 strain with HA-tagged histone was grown in 50 ml of EMM–Leu liquid media to an optical density of 0.9, fixed with 1% formaldehyde and quenched with glycine. After collection, the pellet was washed with cold PBS and suspended in 0.4 ml of lysis buffer (50 mM HEPES KOH pH 7.4, 150 mM NaCl, 1 mM EDTA pH 8.0, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate and 1× cOmplete proteinase inhibitor cocktail (Roche)). Glass beads (Sigma) were added and shaken together for 45 min in a cold room. Beads were removed and the volume was adjusted to 3 ml. Sonication was performed using TOMY UD-201 (output, 3; 20 cycles of 10 s on and 50 s off). After centrifuge, supernatant was used as yeast chromatin extract. Aliquots of chromatin extracts from both yeast and plant were reverse crosslinked, DNA purified and quantified with Qubit 3.0 (Thermo). Yeast chromatin extract was mixed to plant chromatin extract at 5%. Anti-HA (3F10, Roche) was used to purify yeast chromatin.

RNA-sequencing (RNA-seq) analysis. Total RNAs were purified from 12-day-old plants grown on MS Agar plates using the RNeasy Plant Mini Kit (Qiagen), and then treated with DNase I (TaKaRa). Libraries were constructed using the KAPA Stranded RNA-seq Library Preparation Kit according to the manufacturer's instructions. RNA fragmentation was performed at 94 °C for 7 min. Libraries were pooled and sequenced as for ChIP–seq. The sequencing was performed in a 50-bp single end as for ChIP–seq analysis. The obtained data were quality filtered and reads mapped to genes were counted through the STAR algorithm[62]. After RPKM normalization, Rstudio (v.1.1.463) was used to plot expression levels. Two independent biological replicates were analysed for each genotype.

Western blot analysis. Seedings grown for 16 d were frozen, ground and suspended in Honda buffer (0.44 M sucrose, 1.25% ficoll, 2.5% dextran T40, 20 mM HEPES KOH pH 7.4, 10 mM $MgCl_2$, 0.5% Triton X-100, 5 mM DTT and cOmplete protease inhibitor cocktail (Roche)). The lysate was centrifuged at 3,000*g* for 10 min at 4 °C. Nuclei were resuspended with 1× SDS sample buffer and boiled for 5 min. After centrifugation, the supernatant was used for analysis. Nuclei proteins were separated with 15% polyacrylamide gel and transferred onto a polyvinyldifluoride membrane (Biorad). Histone proteins were detected using the antibodies H3 with dilution 1:5,000 (Abcam, Ab1791) and H3K9me2 with dilution 1:2,000 (MAB Institute Inc., MABI0307).

Classification of TE genes. We analysed all TE genes ($n = 3{,}903$) annotated in TAIR[54] unless otherwise stated. To calculate CH methylation recovery shown in Fig. 4a, we excluded TE genes with low CH methylation in WT plants ($n = 444$) and TE genes with CH methylation (for CHG > 0.1 and CHH < 0.03; $n = 3{,}459$) were analysed. 'GLTs' (that is, the TE genes without remethylation in *cc*×*sss* F1 ($n = 73$)) were categorized by low F1 methylation level (less than 10% of WT CHG methylation in two independent F1 plants among those with CH methylation in WT plants. 'Type I GLTs' are from TE genes with transcriptional derepression in both *cc* and *sss* ($n = 34$), while 'type II GLTs' are from TE genes without transcriptional derepression in both *cc* and *sss* ($n = 31$). In Fig. 6 and Extended Data Fig. 8, we further classified type I GLTs into two groups according to their transcriptional status in *zzzcc* and *cczzz*. Since the plants contain a genome fraction of the Ws ecotype that originates from the *hta8* mutant[41], TE genes without mapped sequencing reads in at least one sample were excluded from the analysis. Type I-1 GLTs are TE genes with sustained expression in both *zzzcc* and *cczzz* ($n = 13$) whereas Type I-2 GLTs are resilenced in both *zzzcc* and *cczzz* ($n = 5$).

The list of GLTs and their subtypes are in Supplementary Table. 'RDTs' are defined as the TE genes with inefficient CH methylation recovery in the RdDM mutant background ($n = 134$; F1 CHG methylation/WT CHG methylation >0.1 in a WT background but not in a *rdr126* or *drm2* background). In contrast to GLTs, 'Other TE genes' in figures, except for Extended Data Fig. 9, are from TE genes with CH methylation in WT (CHG >0.1 and CHH >0.03) except for GLTs ($n = 3{,}381$). Other TE genes in Extended Data Fig. 9 are defined as CH methylated TE genes excluding GLTs and RDTs ($n = 3{,}252$).

Calculation of genes and TE genes density. To calculate genes and TE gene densities along the genome, Perl script was used to count the number of genes and TE genes in a 100-kb window. To calculate it around the specific TE genes, the number of genes and TE genes were counted within the adjacent areas of 50 kb on both sides and averaged (mean). In both cases, the values were visualized using the Integrated Genome Browser[59].

Sources of public data used for analyses. The published data were obtained from the GEO database; GSE10967 for siRNAs in *met1* and the control[63], GSE51304 for siRNAs in *suvh456* and the control[8], GSE61028 for siRNAs in *ddm1* (ref. [64]) and GSE98553 for siRNAs in WT leaves[65]. The methylome data for *ddm1*-F9 plants (*ddm1*-epiRIL) are in GSE62206 (ref. [66]). For *ddm1*-F9 plants, the data from epiRIL98 were used, because this line has the most abundant *ddm1*-derived chromosome segments[35,66]. The methylome data for *ddm1*, *rdr2* and WT shown in Extended Data Fig. 5 are in GSE41302 (ref. [9]). The coordinates, if TAIR8 was used, were updated into TAIR10 coordinates using a script in TAIR[54]. For the analyses of siRNAs in *met1* mutant, the processed data (GSM277609 for *met1* and GSM277608 for its WT[63]) were used. The processed data contain the reads with perfect matches to the genome, including multi-aligned reads. Therefore, we divided the read counts by the numbers of alignments. For the analyses of siRNAs in *suvh456*, we used the raw data (SRR1005421.sra for *suvh456* and SRR1005417.sra for its WT) to perform comparable analyses with that of *met1*. At first, the reads were quality checked and trimmed for the adapter sequences using software Trimmomatic[56]. The quality passed and adapter trimmed reads were mapped to the genome TAIR10 using Bowtie[61] with option -a -v 0 to report only perfect matches. For the reads with multiple alignments, the read counts were divided by the numbers of matched alignments. Perl script was used to calculate the adjusted read counts for 24-nt siRNAs matching within gene and TE gene regions and RPKM normalized. TE genes with low or high siRNA levels in Extended Data Fig. 1h were defined under or over RPKM value of 0.1, respectively.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

WGBS, ChIP–seq and RNA-seq reads in this study were deposited in the GEO with the accession number GSE148753. All other reasonable requests for data and research materials are available via contacting the corresponding author. Source data are provided with this paper.

Code availability

The code used in this study are available via contacting the corresponding author.

Received: 25 May 2020; Accepted: 16 October 2020;
Published online: 30 November 2020

References

1. Ito, H. & Kakutani, T. Control of transposable elements in *Arabidopsis thaliana*. *Chromosome Res.* **22**, 217–323 (2014).
2. Kim, M. Y. & Zilberman, D. DNA methylation as a system of plant genomic immunity. *Trends Plant Sci.* **19**, 320–326 (2014).
3. Saze, H. & Kakutani, T. Differentiation of epigenetic modifications between transposons and genes. *Curr. Opin. Plant Biol.* **14**, 81–87 (2011).
4. Finnegan, E. J. & Dennis, E. S. Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Res.* **21**, 2383–2388 (1993).
5. Finnegan, E. J., Peacock, W. J. & Dennis, E. S. Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. *Proc. Natl Acad. Sci. USA* **93**, 8449–8454 (1996).
6. Bartee, L., Malagnac, F. & Bender, J. Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes Dev.* **15**, 1753–1758 (2001).
7. Jackson, J. P., Lindroth, A. M., Cao, X. & Jacobsen, S. E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**, 556–560 (2002).
8. Stroud, H. et al. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat. Struct. Mol. Biol.* **21**, 64–72 (2014).

9. Zemach, A. et al. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205 (2013).
10. Du, J. et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**, 167–180 (2012).
11. Malagnac, F., Bartee, L. & Bender, J. An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J.* **21**, 6842–6852 (2002).
12. Johnson, L., Cao, X. & Jacobsen, S. Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr. Biol.* **12**, 1360–1367 (2002).
13. Jeddeloh, J. A., Stokes, T. L. & Richards, E. J. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat. Genet.* **22**, 94–97 (1999).
14. Vongs, A., Kakutani, T., Martienssen, R. A. & Richards, E. J. *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**, 1926–1928 (1993).
15. Inagaki, S. et al. Autocatalytic differentiation of epigenetic modifications within the *Arabidopsis* genome. *EMBO J.* **29**, 3496–3506 (2010).
16. Inagaki, S. et al. Gene-body chromatin modification dynamics mediate epigenome differentiation in *Arabidopsis*. *EMBO J.* **36**, 970–980 (2017).
17. Saze, H., Shiraishi, A., Miura, A. & Kakutani, T. Control of genic DNA methylation by a jmjC domain-containing protein in *Arabidopsis thaliana*. *Science* **319**, 462–465 (2008).
18. Miura, A. et al. An *Arabidopsis* jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J.* **28**, 1078–1086 (2009).
19. Kakutani, T., Munakata, K., Richards, E. J. & Hirochika, H. Meiotically and mitotically stable inheritance of DNA hypomethylation induced by ddm1 mutation of *Arabidopsis thaliana*. *Genetics* **151**, 831–838 (1999).
20. Teixeira, F. K. et al. A role for RNAi in the selective correction of DNA methylation defects. *Science* **323**, 1600–1604 (2009).
21. Cao, X. & Jacobsen, S. E. Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc. Natl Acad. Sci. USA* **99**, 16491–16498 (2002).
22. Law, J. A. & Jacobsen, S. E. Molecular biology. Dynamic DNA methylation. *Science* **323**, 1568–1569 (2009).
23. Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408 (2014).
24. Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. & Matzke, A. J. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19**, 5194–5201 (2000).
25. Cuerda-Gil, D. & Slotkin, R. K. Non-canonical RNA-directed DNA methylation. *Nat. Plants* **2**, 16163 (2016).
26. Nuthikattu, S. et al. The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol.* **162**, 116–131 (2013).
27. Huettel, B. et al. Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *EMBO J.* **25**, 2828–2836 (2006).
28. Woo, H. R., Dittmer, T. A. & Richards, E. J. Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*. *PLoS Genet.* **4**, e1000156 (2008).
29. Rigal, M. et al. Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in *Arabidopsis thaliana* F1 epihybrids. *Proc. Natl Acad. Sci. USA* **113**, E2083–E2092 (2016).
30. Catoni, M. et al. DNA sequence properties that predict susceptibility to epiallelic switching. *EMBO J.* **36**, 617–628 (2017).
31. Kishimoto, N. et al. Site specificity of the *Arabidopsis* METI DNA methyltransferase demonstrated through hypermethylation of the superman locus. *Plant Mol. Biol.* **46**, 171–183 (2001).
32. Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C. & Paszkowski, J. Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. *Cell* **130**, 851–862 (2007).
33. Saze, H. & Kakutani, T. Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. *EMBO J.* **26**, 3641–3652 (2007).
34. Stroud, H., Greenberg, M. V. C., Feng, S., Bernatavichute, Y. V. & Jacobsen, S. E. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**, 352–364 (2013).
35. Ito, T. et al. Genome-wide negative feedback drives transgenerational DNA methylation dynamics in *Arabidopsis*. *PLoS Genet.* **11**, e1005154 (2015).
36. To, T. K., Saze, H. & Kakutani, T. DNA methylation within transcribed regions. *Plant Physiol.* **168**, 1219–1225 (2015).
37. Tran, R. K. et al. DNA methylation profiling identifies CG methylation clusters in *Arabidopsis* genes. *Curr. Biol.* **15**, 154–159 (2005).
38. Zhang, X. et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**, 1189–1201 (2006).
39. Zilberman, D., Coleman-Derr, D., Ballinger, T. & Henikoff, S. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**, 125–129 (2008).
40. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916–919 (2010).
41. Coleman-Derr, D. & Zilberman, D. Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. *PLoS Genet.* **8**, e1002988 (2012).
42. Nie, W. F. et al. Histone acetylation recruits the SWR1 complex to regulate active DNA demethylation in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **116**, 16641–16650 (2019).
43. Yelagandula, R. et al. The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in *Arabidopsis*. *Cell* **158**, 98–109 (2014).
44. Kato, M., Miura, A., Bender, J., Jacobsen, S. E. & Kakutani, T. Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*. *Curr. Biol.* **13**, 421–426 (2003).
45. Tsukahara, S. et al. Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature* **461**, 423–426 (2009).
46. Fultz, D. & Slotkin, R. K. Exogenous transposable elements circumvent identity-based silencing, permitting the dissection of expression-dependent silencing. *Plant Cell* **29**, 360–376 (2017).
47. Marí-Ordóñez, A. et al. Reconstructing de novo silencing of an active plant retrotransposon. *Nat. Genet.* **45**, 1029–1039 (2013).
48. Slotkin, R. K., Freeling, M. & Lisch, D. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nat. Genet.* **37**, 641–644 (2005).
49. Williams, B. P., Pignatta, D., Henikoff, S. & Gehring, M. Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. *PLoS Genet.* **11**, e1005142 (2015).
50. Kawashima, T. et al. Diversification of histone H2A variants during plant evolution. *Trends. Plant Sci.* **20**, 419–425 (2015).
51. Xiao, W. et al. Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev. Cell* **5**, 891–901 (2003).
52. Ebbs, M. L. & Bender, J. Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. *Plant Cell* **18**, 1166–1176 (2006).
53. Sasaki, T., Kobayashi, A., Saze, H. & Kakutani, T. RNAi-independent de novo DNA methylation revealed in *Arabidopsis* mutants of chromatin remodeling gene *DDM1*. *Plant J.* **70**, 750–758 (2012).
54. Lamesch, P. et al. The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210 (2012).
55. Fu, Y. et al. Mobilization of a plant transposon by expression of the transposon-encoded anti-silencing factor. *EMBO J.* **32**, 2407–2417 (2013).
56. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
57. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
58. Schultz, M. D., Schmitz, R. J. & Ecker, J. R. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends. Genet.* **28**, 583–585 (2012).
59. Nicol, J. W., Helt, G. A., Blanchard, S. G. Jr, Raja, A. & Loraine, A. E. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* **25**, 2730–2731 (2009).
60. Keil, C. et al. Treeview 3.0 (beta 1)—Visualization and analysis of large data matrices. *Zenodo* https://zenodo.org/record/160573/export/xm#.X5FehNBKhPY (2018).
61. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
62. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
63. Lister, R. et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
64. Slotkin, R. K. et al. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**, 461–472 (2009).
65. Lutzmayer, S., Enugutti, B. & Nodine, M. D. Novel small RNA spike-in oligonucleotides enable absolute normalization of small RNA-seq data. *Sci. Rep.* **7**, 5913 (2017).
66. Colomé-Tatché, M. et al. Features of the *Arabidopsis* recombination landscape resulting from the combined loss of sequence variation and DNA methylation. *Proc. Natl Acad. Sci. USA* **109**, 16240–16245 (2012).

Acknowledgements

We thank K. Kato, M. Takahashi, K. Takashima and A. Terui for technical assistance; V. Colot, E. Dennis, Y. Hiromi and L. Quadrana for comments on the manuscript and J. Bender, R. Fischer, E. Richards and D. Zilberman for sharing mutant strains. Computations were partially performed on the NIG supercomputer at NIG, Japan. This work was supported by grants from the Mitsubishi Foundation (to T.K.), Japanese Ministry of Education, Culture, Sports, Science and Technology (nos. 26221105, 15H05963 and 19H00995 to T.K.; nos. 19H05740 and 17K15059 to T.K.T.), CREST grant, Japan (no. JPMJCR15O1 to T.K.), Systems Functional Genetics Project of the

Transdisciplinary Research Integration Center, ROIS, Japan (no. to Y.T., A.T., A.F. and T.K.), PREST grant from Japan Science and Technology Agency (JPMJPR17Q1 to S.I.), Austrian Academy of Sciences (to F.B.) and FWF stand-alone programs P32054 and P33380 (to F.B.).

Author contributions

T.K.T., Y.N., S.I. Y.T., F.B. and T.K. designed the study. T.K.T., Y.N., S.I. Y.T., S.T., A.T., A.F. and T.K. performed the experiments. T.K.T., Y.N., S.I. and Y.T. analysed the data. T.K.T. and T.K. wrote the paper incorporating comments from the other authors.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41477-020-00810-z.

Supplementary information is available for this paper at https://doi.org/10.1038/s41477-020-00810-z.

Correspondence and requests for materials should be addressed to T.K.T. or T.K.

Peer review information *Nature Plants* thanks Zhaobo Lang, Jixian Zhai and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.