Repression of CHROMOMETHYLASE 3 Prevents Epigenetic Collateral Damage in Arabidopsis

Ranjith K. Papareddy[1,4], Katalin Páldi[1,4], Anna D. Smolka[1], Patrick Hüther[1,2], Claude Becker[1,2], Michael D. Nodine[1,3,*]

[1]Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna Biocenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria; [2]Genetics, LMU Biocenter, Ludwig-Maximilians University, Martinsried, 82152, Germany; [3]Laboratory of Molecular Biology, Wageningen University, Wageningen, 6708 PB, the Netherlands. [4]These authors contributed equally to this work.
[*]Corresponding author : michael.nodine@wur.nl

ABSTRACT

DNA methylation has evolved to silence mutagenic transposable elements (TEs) while typically avoiding the targeting of endogenous genes. Mechanisms that prevent DNA methyltransferases from ectopically methylating genes are expected to be of prime importance during periods of dynamic cell cycle activities including plant embryogenesis. However, virtually nothing is known regarding how DNA methyltransferase activities are precisely regulated during embryogenesis to prevent the induction of potentially deleterious and mitotically stable genic epimutations. Here, we report that microRNA-mediated repression of CHROMOMETHYLASE 3 (CMT3) and the chromatin features that CMT3 prefers help prevent ectopic methylation of thousands of genes during embryogenesis that can persist for weeks afterwards. Moreover, CMT3-induced ectopic methylation of genes undergoing transcriptional activation can reduce their corresponding transcript levels. Therefore, the repression of CMT3 prevents epigenetic collateral damage on endogenous genes. We also provide a model that may help reconcile conflicting viewpoints regarding the functions of gene-body methylation that occurs in nearly all flowering plants.

Key words: DNA methylation, chromatin, microRNA, gene regulation, epimutation

INTRODUCTION

Methylation of DNA encoding transposable elements (TE) is required to silence their expression and consequently prevent them from mobilizing and mutagenizing genomes (Kato et al., 2003; Law and Jacobsen, 2010). Complex mechanisms have evolved to balance the high degree of sensitivity needed to direct methylation and silencing of TEs with the precision required to prevent ectopic methylation of endogenous genes (Antunez-Sanchez et al., 2020; Ito et al., 2015; Papareddy et al., 2020; Saze and Kakutani, 2011; Williams et al., 2015; Zhang et al., 2020). However, little is known about the mechanisms of epigenome homeostasis during embryogenesis when organisms are particularly vulnerable to TE-induced mutagenesis, as well as the establishment of potentially deleterious epimutations that can persist through many cell divisions and even across generations (Henderson and Jacobsen, 2007; Mathieu et al., 2007; Probst et al., 2009; Saze et al., 2003; Scheid et al., 1998).

In *Arabidopsis thaliana* (Arabidopsis), most TEs are found in pericentromeric regions of the genome to which RNA polymerases have limited access (Arabidopsis Genome Initiative, 2000; Lippman et al., 2004; 2000; Zhang et al., 2006). These TE-enriched constitutive heterochromatic regions are characterized by high densities of cytosine methylation in symmetric (CG or CHG; H $\neq$ G) and asymmetric (CHH) contexts, as well as histone H3 lysine dimethylation (H3K9me2) and other transcriptionally repressive chromatin marks (Cokus et al., 2008; Lister et al., 2008; Stroud et al., 2014). Symmetric DNA

methylation and H3K9me2 also facilitate the stable propagation of silenced states through cell divisions (Jackson et al., 2002; Lindroth et al., 2001; Stroud et al., 2013, 2014). METHYLTRANSFERASE 1 (MET1) maintains CG methylation through mitotic and meiotic cell divisions with high fidelity due to VARIANT IN METHYLATION 1/2/3 (VIM1/2/3) proteins that recognize hemi-methylated CG and recruit MET1 to methylate daughter strands (Feng et al., 2010; Finnegan and Dennis, 1993; Ning et al., 2020; Woo et al., 2008). CG methylation can also recruit RNA Polymerase IV complexes required to produce 24-nt small interfering RNAs (siRNAs) that are then loaded onto ARGONAUTE proteins and guide them to target loci by base-pairing with nucleic acids (Blevins et al., 2015; Herr et al., 2005; Papareddy et al., 2020; Zhai et al., 2015; Zilberman et al., 2003). This leads to the recruitment of DOMAINS REARRANGED METHYLTRANSFERASES 1/2 (DRM1/2) and results in de novo methylation of cytosines in all sequence contexts, including CHH, which is a hallmark feature of RNA-directed DNA methylation (RdDM) (Cao and Jacobsen, 2002; Stroud et al., 2013; Wierzbicki et al., 2008). However, RdDM is typically restricted from constitutive heterochromatin because it is inaccessible to DNA-dependent RNA polymerase IV and methyltransferases required for RdDM (Papareddy et al., 2020; Zemach et al., 2013). Instead, CHH methylation of constitutive heterochromatin is mediated by CHROMOMETHYLASE 2 (CMT2) that binds to H3K9me2 deposited by KRYPTONITE (KYP) and closely related SUPPRESSOR OF VARIATION 3-9 HOMOLOGUE PROTEIN 5/6 (SUVH5/6) methyltransferases (Stroud et al., 2014; Zemach et al., 2013). CHROMOMETHYLASE 3 (CMT3) also forms interlocking positive feedback loops with H3K9 methyltransferases (Du et al., 2012; Jackson et al., 2002; Lindroth et al., 2001), but is more closely associated with the cell cycle and mediates CHG methylation (Ning et al., 2020).

CMT3-mediated CHG methylation is largely deposited on TEs. However, CMT3 can also induce the ectopic methylation of protein-coding genes (Wendte et al., 2019). Moreover, the introduction of Arabidopsis CMT3 transgenes into *Eutrema salsugineum*, which lost CMT3 millions of years ago (Bewick et al., 2016), could occasionally reconstitute CG methylation on genes (Wendte et al., 2019). The resulting gene-body methylation (gbM) could be stably maintained independent of the CMT3 transgene for several generations (Wendte et al., 2019). However, it remains largely unknown how CMT3 is restricted to targeting heterochromatin, as well as the consequences of CMT3-induced hypermethylation of genes. Moreover, the functional significance of gbM in animals and plants has been intensely debated. Because methylated cytosines are mutagenic due to associated cytosine deamination (Shen et al., 1992; Sved and Bird, 1990), features associated with gbM have been interpreted as evidence that gbM provides selective advantages that counterbalance this mutagenesis-imposed fitness penalty. For instance, gbM is depleted from transcription start and end sites (Tran et al., 2005; Zhang et al., 2006; Zilberman et al., 2007), and it has recently been reported that gbM helps prevent transcription initiation from cryptic promoters located in gene bodies as initially proposed (Choi et al., 2019; Zilberman et al., 2007). Moreover, gbM tends to be enriched on constitutively expressed genes (Lister et al., 2008; Niederhuth et al., 2016; Takuno et al., 2017; Zhang et al., 2006), which would be consistent with gbM stabilizing gene expression by excluding certain histone variants (i.e H2A.Z) from genes (Coleman-Derr and Zilberman, 2012) and generally enhancing gene expression (Muyle and Gaut, 2019; Shahzad et al., 2021). Nevertheless, accumulating evidence is also consistent with gbM being a heritable by-product of CMT3-induced epimutations (Bewick et al., 2016, 2019; Wendte et al., 2019).

Consistent with the need to fine-tune the amount of CMT3 activities required to both silence TEs and prevent epimutations on genes, mechanisms exist that transcriptionally (Ning et al., 2020) and post-translationally (Deng et al., 2016) regulate CMT3, as well as remove H3K9me2 specifically from expressed genes (Inagaki et al., 2010; Saze et al., 2008). These and additional mechanisms are likely of utmost importance during embryo development when a proliferative morphogenesis phase produces the most fundamental cell lineages of the plant, including those that will eventually generate the gametes. Yet, how DNA methylation pathways are regulated during this phase of dynamic cell division to exquisitely balance the need for TE methylation with the prevention of potentially deleterious and stably inherited epimutations is virtually unknown.

RESULTS

Cell division is linked with CG and CHG methylation through distinct mechanisms

MET1 and VIM1/2/3 are required for the faithful transmission of mCG across cell cycles (Feng et al., 2010; Finnegan and Dennis, 1993; Ning et al., 2020; Woo et al., 2008) and accordingly had increased transcript levels in rapidly dividing early embryos that also correlated well with transcripts encoding cell-cycle activators throughout embryogenesis (Fig. 1A) (Hofmann et al., 2019; Papareddy et al., 2020). More specifically, MET1 and VIM1/2/3 transcript levels peaked at the early heart stage and were reduced afterwards before plummeting at the mature green stage. These transcript developmental dynamics were also characteristic of transcripts encoding proteins involved in licensing DNA replication (e.g. Cyclins A2/B1, CDKB1-1, MINICHROMOSOME MAINTENANCE2), heterochromatin maintenance (e.g. DDM1) and DNA methylation (e.g. CMT3), but not randomized controls (Figs. 1B,S1A). Therefore, genes required for DNA methylation and heterochromatin maintenance are tightly correlated with cell-cycle activity during embryogenesis.

To test whether the patterns observed for transcripts regulating DNA methylation reflect DNA methylation dynamics, we computed differentially methylated cytosines (DMCs) across flowers, embryos and leaves (see Methods). Similar to previous observations (Bouyer et al., 2017; Lin et al., 2017; Papareddy et al., 2020), 70% of DMCs occurred in the CHH context (Fig. 1C). Consistent with dynamic expression patterns of MET1 and CMT3, substantial fractions of DMCs respectively occurred in CG (20%) or CHG (10%) contexts. Therefore, DNA methylation is dynamically reconfigured in all sequence contexts during embryogenesis. In total, these symmetric DMCs represented 1,185 CG (Table S1) and 1,398 CHG (Table S2) differentially methylated regions (DMRs) covering 201 kb and 185.8 kb, respectively (Fig. 1D, Table S1; see Methods). Although a significant fraction of CG and CHG DMRs overlapped (n = 183; 7.1% of total), the vast majority of CG and CHG DMRs were located in non-overlapping genomic regions corresponding to euchromatic gene-rich and heterochromatic TE-rich regions of the genome, respectively (Fig. 1D). Because CHG methylation can require CG methylation (Stroud et al., 2013), we tested whether the 15.1% of CHG DMRs overlapping CG DMRs require CG methylation. Leaves deficient in CG methylation did not have reduced CHG methylation in CHG DMRs regardless of whether or not they overlapped with CG DMRs (Fig. S2B; data from Stroud et al. 2013). This indicates that CHG DMRs occur in distinct genomic regions and are largely independent of CG methylation (Fig. 1E,F).

Relative to floral bud samples, CG DMRs have slightly reduced methylation in preglobular embryos, followed by increased methylation until after the torpedo stage, when levels dramatically reduce in mature green embryos and recover in leaves (Fig. 1G). By contrast, methylation levels of CHG DMRs are relatively stable between floral buds and early embryos, then decrease in late embryos, reaching a minimum in leaves (Fig. 1H). Accordingly, changes in CG and CHG DMR methylation levels during development were significantly correlated with MET1 (Pearson's R = 0.8; *P* value = 0.03) and CMT3 (Pearson's R= 0.74; *P* value = 0.05) transcript levels, respectively (Fig. S1B,C). Therefore, although cell division rates are correlated with symmetric DNA methylation dynamics, distinct mechanisms reconfigure CG or CHG methylation genome-wide during embryogenesis.

Genome-wide coordination of symmetric DNA methylation

Because DNA methylation is concentrated on TEs (Stroud et al. 2013; Zhang et al. 2006), we next investigated global developmental dynamics of TE methylation. CG methylation on both euchromatic and heterochromatic TEs was slightly reduced in pregobular embryos and then restored to the levels found in floral buds by the early heart stage (Fig. 2A,B). Whereas CG methylation of euchromatic TEs was relatively constant for the remainder of embryogenesis, heterochromatic TEs had significantly increased methylation during late embryogenesis compared to post-embryonic tissues. Consistent with heterochromatin becoming highly condensed during embryo maturation (van Zanten et al., 2011), we found that CG hypermethylation in mature green compared to bent cotyledon embryos predominantly occurred in pericentromeric genomic regions rather than gene-rich chromosomal arms (Fig. 2C). CG methylation was required for

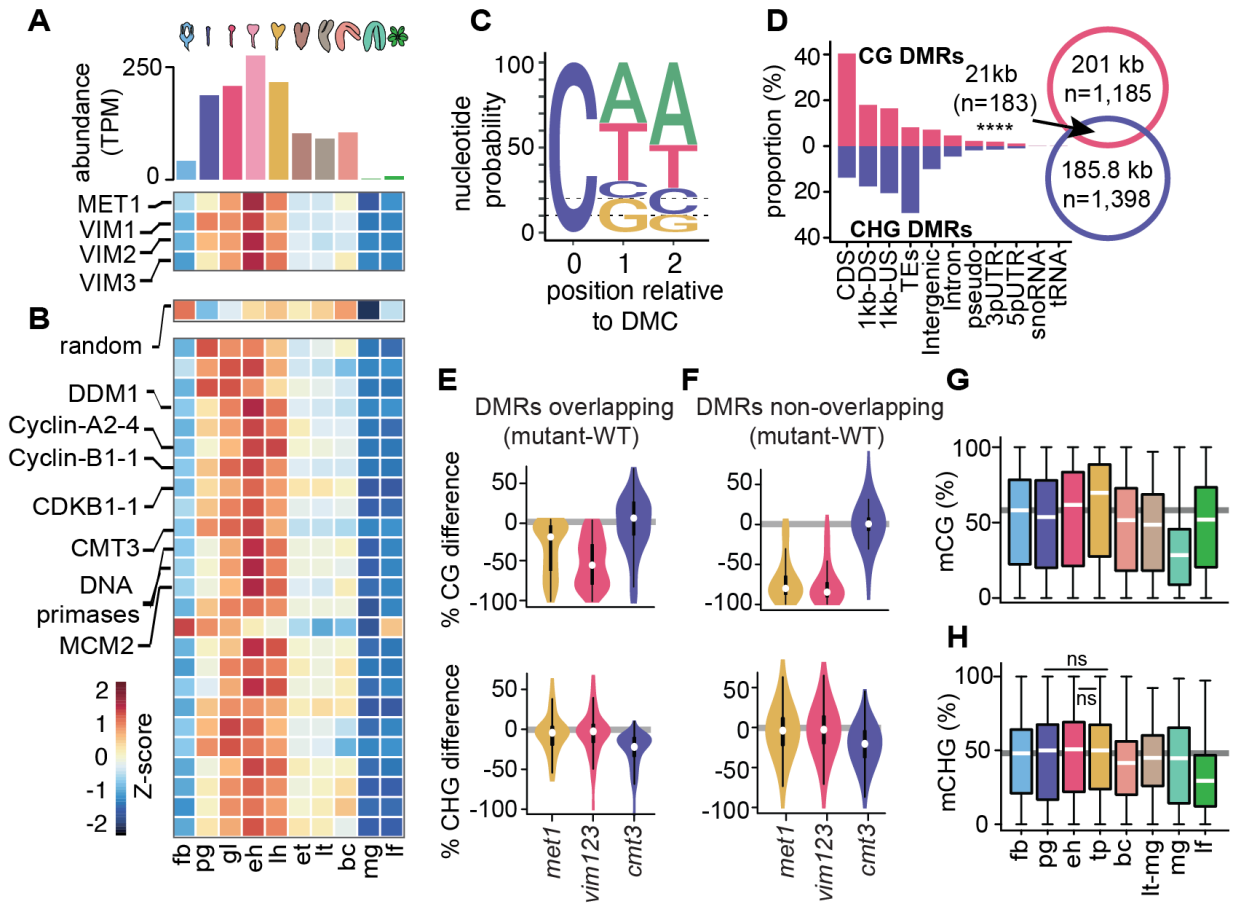

Figure 1: Cell division is linked with CG and CHG methylation through distinct mechanisms. A) Bar chart depicting total abundance (*top*) and heat map of individual relative transcript levels (*bottom*) of genes involved in CG methylation in flowers, embryos and leaves. fb, floral buds; pg, preglobular; gl, globular; eh, early heart; lh, late heart; et, early torpedo; lt, late torpedo; bc, bent cotyledon; mg, mature green; lf, leaf. **B)** Heatmap showing developmental dynamics of permuted gene set (*top*) median values (i.e. 1000 iterations of random sampling of 25 genes) and top-25 genes co-varying with MET1, VIM1, VIM2 and VIM3 obtained by employing nearest neighbour algorithm calculated based on Euclidean distance between genes and centroid expression of MET1, VIM1, VIM2 and VIM3 (*bottom*). **C)** Sequence logo representing nucleotide probability relative to differentially methylated cytosines (DMC). **D)** Proportion of CG and CHG differentially methylated regions (DMRs) overlapping genomic features. Venn diagram showing overlap between CG and CHG DMRs. Significance overlap of DMRs determined by Fisher's exact test *P* value < 0.0001 is indicated by ****. **E** and **F)** Violin plot showing CG (*top*) and CHG (*bottom*) methylation differences between mutant and wild type leaves for CHG DMRs overlapping **(E)** or not overlapping **(F)** with CG DMRs (Stroud et al., 2013). **G** and **H)** Box plots of average weighted methylation of CG DMRs (n = 1,185) **(G)** and CHG DMRs **(H)** (n = 1,398) during development. fb, floral buds; pg, preglobular; eh, early heart; tp, torpedo (6 DAP) (Pignatta et al., 2015); bc, bent cotyledon; lt-mg, late torpedo-to-early mature green (Hsieh et al., 2009); mg, mature green (Bouyer et al., 2017); lf, leaf. fb, pg, eh, bc and lf were from (Papareddy et al., 2020). Unless stated as not significant (ns), all combinations are significant with *P* values < 0.001 obtained by Mann-Whitney U test. Shaded horizontal line in the background represents the median methylation value of floral buds.

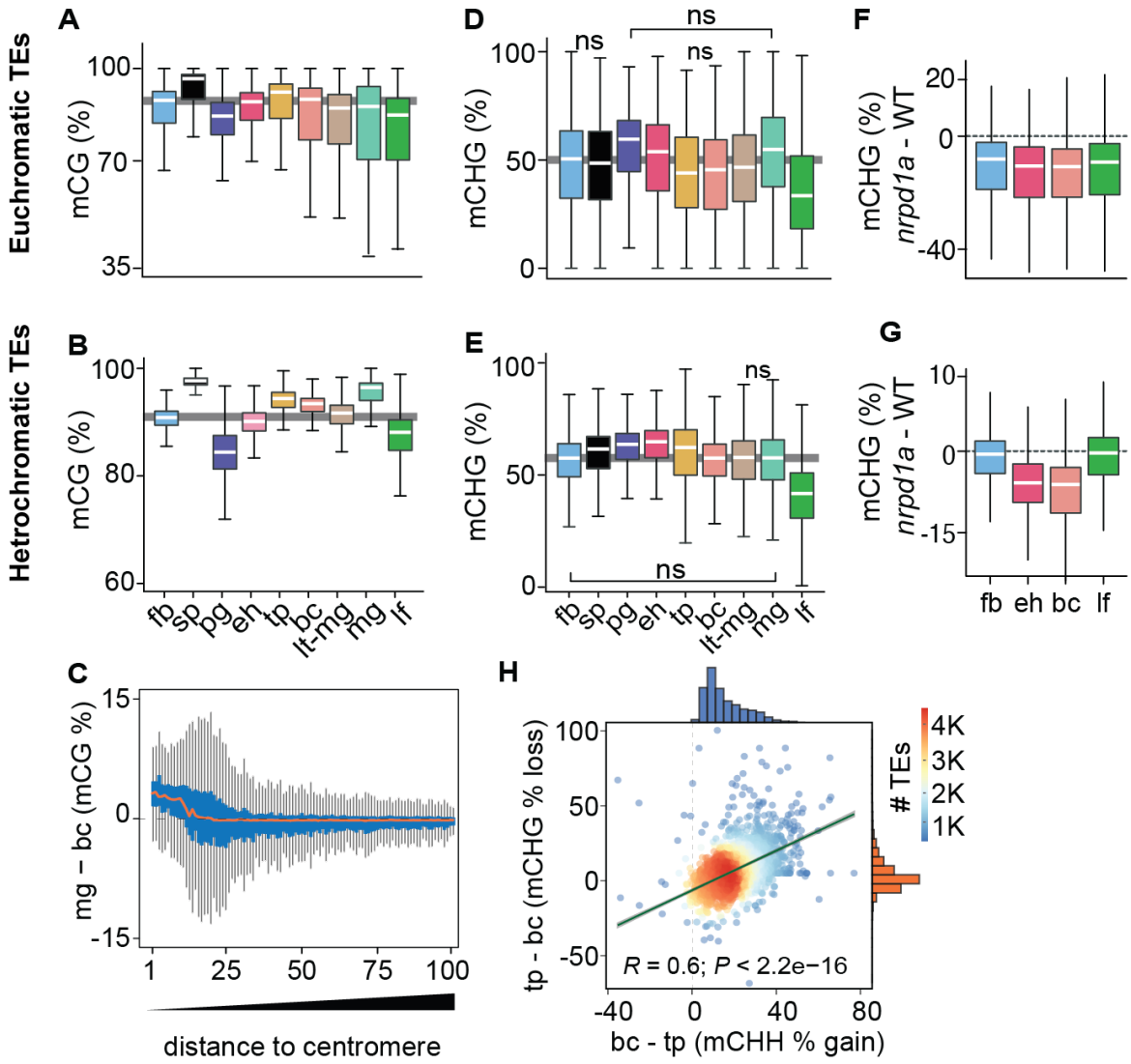

Figure 2: Genome-wide coordination of symmetric DNA methylation. A) and **B)** Boxplots of CG methylation percentages on euchromatic **(A)** and heterochromatic **(B)** TEs during development. fb, floral buds; sp, sperm (Ibarra et al., 2012); pg, preglobular; eh, early heart; tp, torpedo (6 DAP); bc, bent cotyledon; lt-mg, late torpedo-to-early mature green; mg, mature green; lf, leaf. Thick horizontal bars indicate medians, and the top and bottom edges of boxes represent the 75th and 25th percentiles, respectively. Shaded horizontal line in the background represents the median methylation value of floral buds. **C)** Difference in CG methylation between mature green (mg) and bent cotyledon (bc) embryos were calculated in 1-kb genomic bins, which were divided into percentiles and sorted based on their distance to centromeres (1 and 100 being the tile closest and furthest from centromeres, respectively). Red color line indicates the median and the top and bottom edges of the blue colored boxes represent 75th and 25th percentiles, respectively. Vertical grey bars indicate 1.5X the interquartile range. **D)** and **E)** Boxplots of CHG methylation on euchromatic **(D)** and heterochromatic **(E)** TEs during development (key as in **A**). **F** and **G)** Boxplots of CHG methylation differences between *nrpd1a* and wild type (Col-0) tissues for euchromatic **(F)** and heterochromatic **(G)** TEs. **(H)** Scatterplot showing Pearson's correlation coefficients (*R*). Differences in mCHH and mCHG between bent cotyledon (bc) and torpedo stage (tp) embryos are shown on x- and y-axes, respectively. Histograms show the number of TEs in thousands (K).

the production of 24-nt siRNAs from euchromatic TEs, but only marginally for heterochromatic TEs (Fig. S2A, data from Lister et al., 2008). Conversely, the loss of 24-nt siRNAs in *nrpd1a* mutants only had negligible effects on CG methylation of both heterochromatic and euchromatic TEs (Fig. S2B,C). Therefore, siRNA production from euchromatic regions of the genome requires CG methylation, but not vice versa.

Global CHG methylation of euchromatic and heterochromatic TEs was higher in embryos compared to leaves (Fig. 2D,E). Similar to previous observations for CHH methylation (Papareddy et al., 2020), siRNA-deficient *nrpd1a* mutant tissues had reduced CHG methylation on euchromatic or heterochromatic TEs in all or only embryonic samples, respectively (Fig. 2F,G). Intriguingly, increased CHH methylation on heterochromatic TEs was significantly correlated with decreased CHG methylation during late stages of embryogenesis when cell division rates are reduced (Fig. 2H). Therefore, CMT3-dependent CHG and CMT2-dependent CHH methylation of heterochromatic TEs are respectively positively and negatively correlated with cell division rates.

Repression of CMT3 during embryogenesis regulates methylome dynamics

CMT3 is recruited to loci by binding to H3K9me2 deposited by SUVH4/5/6 histone methyltransferases (Du et al., 2012; Jackson et al., 2002; Lindroth et al., 2001; Stroud et al., 2014). CMT3 and KYP, which is the major SUVH4 H3K9 methyltransferase, were dynamically expressed according to patterns characteristic of other cell-cycle regulated genes and CHG methylation dynamics (Figs. 1A,H,3A). More specifically, CMT3 and KYP were highly expressed in rapidly dividing early embryos and had reduced expression in late embryos until the mature stage, where they were barely detectable. Altogether, our results are consistent with the idea that the more rapid cell divisions in early embryos demand higher levels of CMT3 and KYP to maintain mCHG through the cell cycle. Moreover, IBM1, which encodes an H3K9me2 demethylase and prevents CMT3 recruitment to gene bodies (Miura et al., 2009; Saze et al., 2008), is dynamically expressed during embryogenesis in a pattern that strongly resembles CMT3 and KYP (Fig. 3A). Therefore, co-expression of IBM1 with CMT3 and KYP likely helps limit ectopic H3K9me2 and methylated CHG on gene bodies during embryogenesis as has been demonstrated during post-embryonic development (Inagaki et al., 2017).

We previously found that miR823-directed cleavage of CMT3 transcripts is highly enriched in embryos directly after morphogenesis (Plotnikova et al., 2019). In contrast to CMT3 transcript dynamics, miR823 accumulates during embryogenesis, and miR823:CMT3 cleavage products were enriched and significantly detected specifically at late heart and early torpedo stages precisely when CMT3 transcript levels were sharply decreasing (Fig. 3B). Based on these observations, we hypothesized that miR823-mediated repression of CMT3 contributes to the reduced CHG methylation levels observed during late embryogenesis.

To test if miR823-directed repression of CMT3 transcripts reduces CHG methylation levels during embryogenesis, we generated deletions in the region of the *MIR823* locus encoding the mature miRNA (Fig. S3A) and examined CMT3 transcript and CHG methylation levels. Both independently generated *mir823-1* and *mir823-2* mutants were confirmed as nulls (Fig. S3B) and had significantly increased CMT3 levels relative to wild type in bent cotyledon stage embryos at which stage CMT3 levels are normally reduced (Fig. 3C). Consistent with miR823-directed cleavage of CMT3 being highly enriched in embryos, we did not observe increased CMT3 transcripts in either leaves or floral buds of *mir823* mutants (Fig. S3D). Moreover, CHG, but not CG or CHH, methylation was increased on TEs in bent cotyledon embryos of both *mir823-1* and *mir823-2* mutants relative to wild type (Figs. 3D,S3E).

As an independent approach, we used site-directed mutagenesis to introduce synonymous mutations in the miR823 target site within CMT3 transgene constructs that included 1.41 kb upstream and 0.73 kb downstream intergenic regions, and associated cis-regulatory elements (Fig. S3C; see Methods). As controls, we also generated CMT3 constructs without mutations, and introduced these miR823-cleavable CMT3 (cCMT3), as well as the miR823-resistant (rCMT3), constructs into *cmt3-11* mutant plants (Henderson and Jacobsen, 2008). CMT3 transcript levels were increased in rCMT3 relative to cCMT3

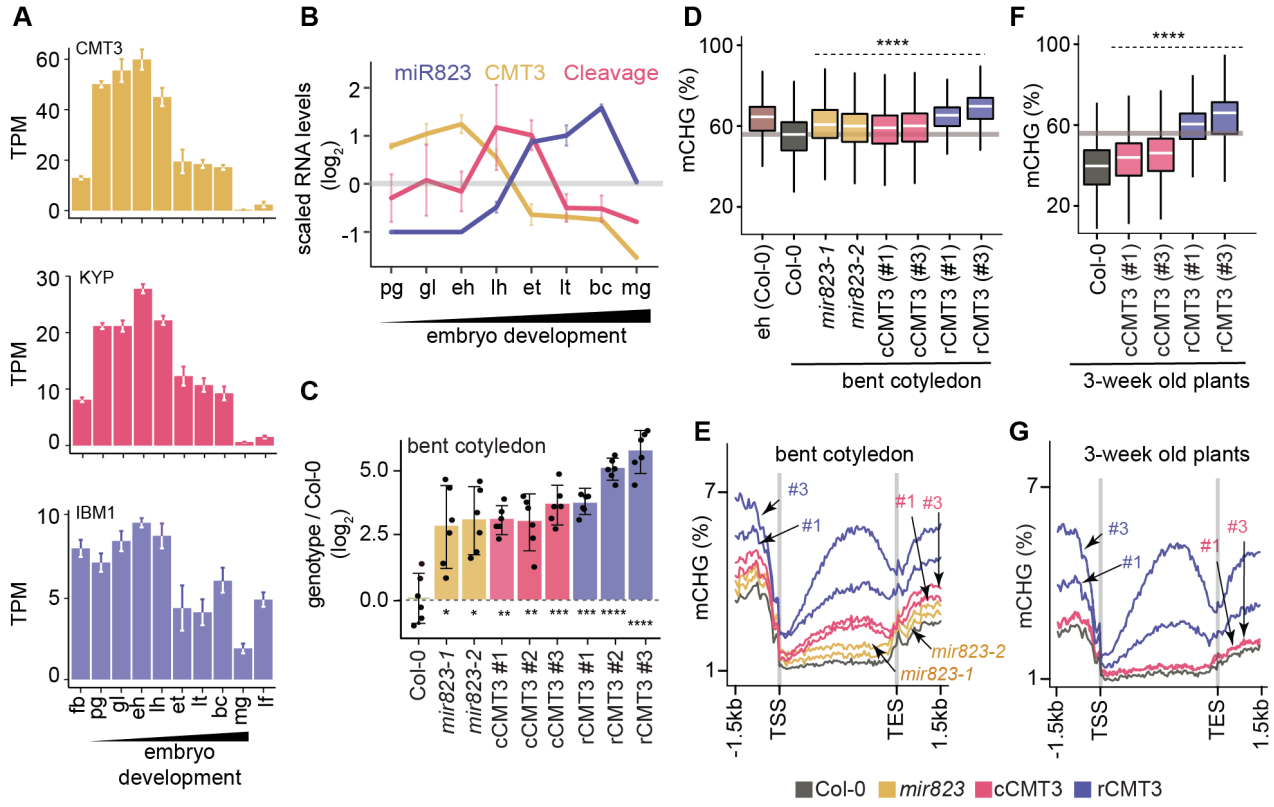

Figure 3: Repression of CMT3 during embryogenesis regulates methylome dynamics. A) Barplots illustrating transcript levels of CMT3 (*top*), KYP (*middle*) and IBM1 (*bottom*) in flowers, embryos and leaves. fb, floral buds; pg, preglobular; gl, globular; eh, early heart; lh, late heart; et, early torpedo; lt, late torpedo; bc, bent cotyledon; mg, mature green; lf, leaf. Error bars represent standard errors. **B)** Line graphs showing the relative RNA abundance of miR823 (*blue*), CMT3 RNA (*yellow*) and miR823:CMT3 cleavage products (*pink*). **C)** Log_2-transformed relative CMT3 target transcript levels in bent cotyledon embryos (8 DAP; day after pollination) from wild type plants (Col-0), or *cmt3-11* plants expressing either miR823-cleavable CMT3 (cCMT3) or miR823-resistant CMT3 (rCMT3) versions. Each dot represents the mean of two technical replicates of embryos. Bars represent mean values and error bars indicate standard error. Asterisks indicate whether the transcript levels observed in *mir823* mutant, cCMT3 and rCMT3 embryos were significantly different compared to wild type (Two-tailed Student's t tests; ****, ***, **, and * represent *P* values < 0.0001, < 0.001, < 0.01, and < 0.05, respectively). Color-coded according to the key. **D)** Boxplots of CHG methylation on transposons with ≥5 informative cytosines covered by ≥4 reads and classified as either euchromatic or heterochromatic in Papareddy et al., 2020. *P values* < 0.0001 based on Mann-Whitney U tests of methylation differences between wild type and either mutant or transgenic bent cotyledon embryos are represented by ****. **E)** Metaplots of average CHG methylation percentages across genes bodies from transcription start sites (TSS) to transcription end sites (TES), 1.5-kb upstream and 1.5-kb downstream of genes in bent cotyledon embryos. Color-coded according to the key. **F** and **G)** Boxplots of CHG methylation on transposons **(F)** and metaplots of CHG methylation on genes **(G)** in three-week old plants as described in **D** and **E**, respectively.

lines at the bent cotyledon stage (Fig. 3C), but not in leaves or floral buds (Fig. S3D), which further indicates that miR823-directed cleavage and repression of CMT3 is highly enriched in embryos transitioning between morphogenesis and maturation. CMT3 levels were also increased in cCMT3 and rCMT3 lines compared to Col-0 in embryos, leaves and floral buds (Figs. 3C,S3D) suggesting that miR823 is not sufficient to repress transgenic CMT3 to the same extent as endogenous CMT3 transcripts. Although we cannot rule out that this is due to missing cis-regulatory repressive elements in the transgenes, increased gene dosage and positional effects of the transgenes seems more likely. Upstream and downstream intergenic regions were included in the CMT3 constructs (Fig. SG). Moreover, although relative transgene copy numbers were not significantly different across the independently generated cCMT3 and rCMT3 transgenic lines, they were higher than endogenous CMT3 in wild type (Fig. S4G,H). Nevertheless, it is clear that CMT3 levels are finely tuned during embryogenesis. Together with the analysis of *mir823* mutants and miR823-mediated CMT3 transcript cleavage products (Plotnikova et al., 2019), these results strongly indicate that miR823 cleaves and represses CMT3 levels during mid-embryogenesis. Consistent with what we observed in *mir823* mutants, increased CMT3 transcript levels in cCMT3 and rCMT3 embryos resulted in CHG hypermethylation of TEs (Fig. 3D) but did not globally influence CG or CHH methylation (Fig. S3E,F). Remarkably, increased CMT3 transcript levels in *mir823* mutants, cCMT3 and most strikingly rCMT3 embryos were associated with ectopic CHG methylation on protein-coding gene bodies and flanking regions in bent cotyledon embryos (Fig. 3E). Therefore, both TEs and genes are hypermethylated when CMT3 levels are not properly down-regulated upon the morphogenesis-to-maturation transition during embryogenesis.

To test whether miR823-directed repression of CMT3 and prevention of CHG methylation of genes that we observed in embryos persists after embryogenesis, we next profiled methylomes of cCMT3 and rCMT3 plants three weeks after germination. Although TEs had increased CHG methylation levels in both cCMT3 lines relative to wild type, protein-coding genes were not affected (Fig. 3F-G). In stark contrast, TEs and genes were hypermethylated in both rCMT3 lines compared to cCMT3 or wild type plants, and only slightly reduced relative to the levels observed in rCMT3 bent cotyledon embryos (Fig. 3F,G). Together with miR823-independent processes (e.g. IBM1 removal of H3K9me2), miR823-directed repression of CMT3 is therefore required to prevent the hypermethylation of protein-coding genes that can be maintained weeks after the completion of embryogenesis.

Chromatin features associated with CMT3-induced gene methylation

To yield insights into how genes are hypermethylated upon the derepression of CMT3, we determined whether certain genomic features were associated with CMT3-induced genic methylation. Towards this end, we first selected 22,637 nuclear-encoded protein-coding genes that had $\geq$5 methylC-seq reads overlapping CHG sites in rCMT3 line #3 and that were expressed (i.e. $\geq$1 TPM in any tissue based on Hofmann et al., 2019). We chose rCMT3 line #3 because it had the strongest genome-wide CHG hypermethylation and focused on expressed genes to exclude those that may have TE-like features, which could confound analysis. We then used k-means clustering of the differences between rCMT3 line #3 and wild type bent cotyledon embryos to partition this set of genes into four clusters (Fig. S4A). These clusters were comprised of 1,439 to 7,882 genes (6.4% to 34.8% of total) and ranged from groups of genes that had no methylation changes (cluster 1) to those that were strongly hypermethylated (cluster 4) in rCMT3 compared to wild type embryos (Figs. 4A,B,S4B,C). The same patterns were observed across these clusters in embryos from an independently generated rCMT3 transgenic (line #1), which indicates that CMT3-induced hypermethylation is not stochastic (Fig. S4B,C). As expected, CMT3-induced hypermethylation predominantly occurred in the CMT3-preferred CWG context (Gouil and Baulcombe, 2016; Li et al., 2018) although hypermethylation was also found in CCG and slightly but significantly in non-CHG contexts, including CG characteristic of gbM similar to previous observations (Fig. S4D) (Wendte et al., 2019).

Consistent with methyltransferases preferring nucleosome-rich DNA as substrates (Chodavarapu et al., 2010), CMT3- induced hypermethylation was proportional to patterns of nucleosome occupancy and biased towards

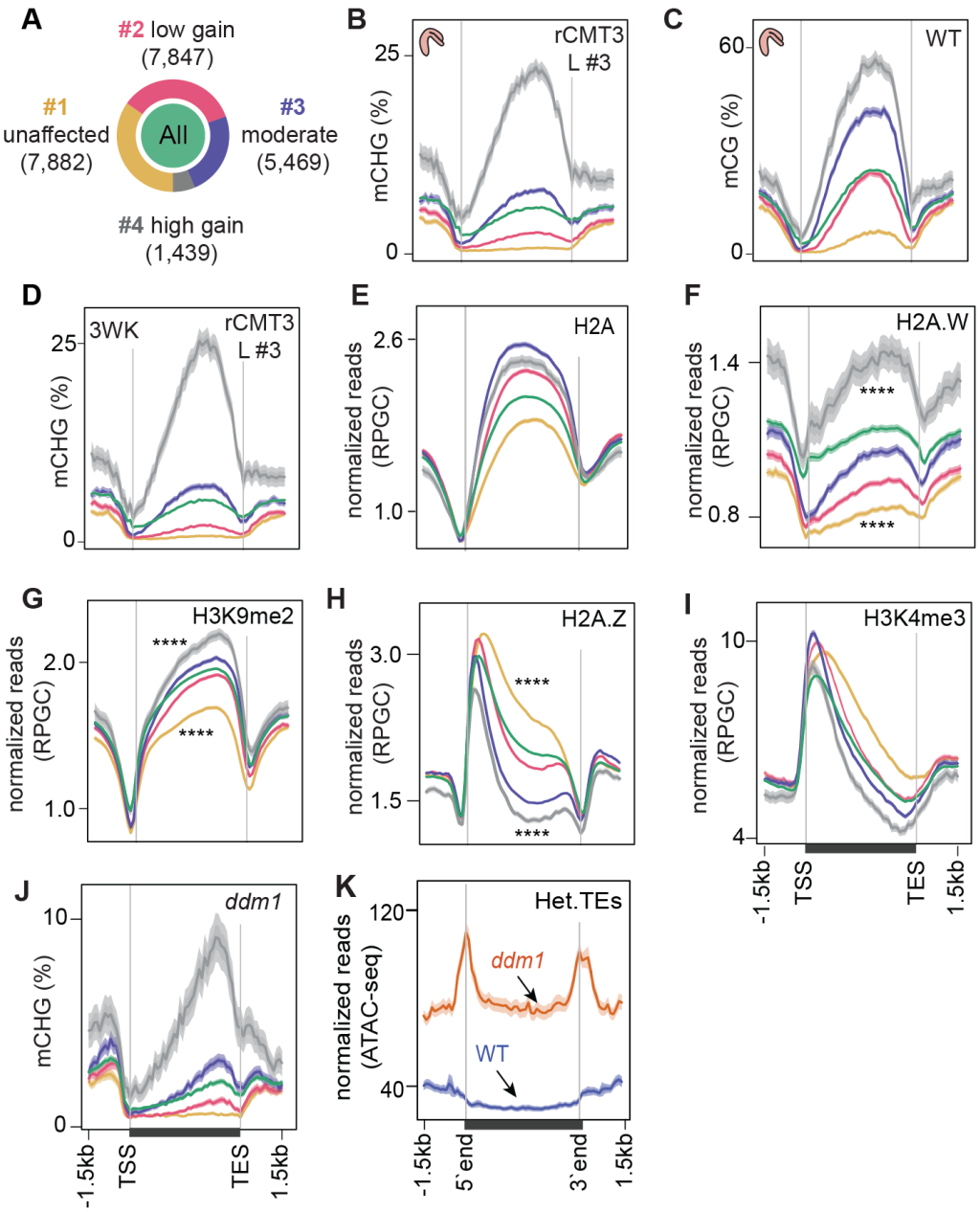

Figure 4: Chromatin features associated with CMT3-induced gene methylation. A) Proportion of genes in each cluster partitioned using k-means clustering algorithm based on differences in mCHG between rCMT3 (line #3) and wild type embryos. Unaffected genes (*yellow*), low mCHG gain genes (*red*), moderate mCHG gain genes (*blue*) and high mCHG gain genes (*grey*). Green inner circle represents all expressed genes. **B-D)** Metaplots showing mCHG on gene clusters in bent cotyledon embryos from rCMT3 line #3 (L #3) **(B)**, mCG on gene clusters in wild type bent cotyledon embryos **(C)** and mCHG in rCMT3 (L #3) three-week old plants (3WK) **(D)**. Shaded ribbons in metaplots represent standard deviations. **E-I)** Metaplots showing normalized reads per genomic content (RPGC) average values of histone variant H2A **(E)**, H2A.W **(F)** (Yelagandula et al., 2014), H3K9me2 **(G)** (Stroud et al., 2014), H2A.Z **(H)** (Yelagandula et al., 2014) and H3K4me2 **(I)** (Maher, 2020). *P* values < 0.0001 obtained by Mann-Whitney U test based on differences between genes in cluster 1 or 4 compared to all genes is represented by ****. **J)** Metaplots showing mCHG on gene clusters in seventh generation *ddm1* mutants (Stroud et al., 2013). **K)** Normalized ATAC-seq reads (Zhong et al., 2021) representing accessibility of heterochromatic TEs (Het.TEs) in WT and *ddm1* mutants as defined in Papareddy et al., 2020.

the 3′ ends of gene bodies, which was highly similar to CG methylation (Figs. 4C,S4E). Nucleosome spacing is promoted by linker histone 1 (H1) (Choi et al., 2019; Fan et al., 2003) and CMT3-induced CHG hypermethylation was proportional to H1 levels across gene bodies (Fig. S4F). Because nucleosome occupancy was not as readily distinguishable between clusters of affected genes (i.e. clusters 2-4) (Fig. S4E,F), we hypothesized that histone variants conferring differential nucleosome stabilities and chromatin accessibility may influence ectopic CMT3-induced hypermethylation (Osakabe et al., 2018). Indeed, CHG hypermethylation across the four groups was positively correlated with levels of the stable histone variants H2A, H2A.X and most notably H2A.W that was recently shown to be required for CHG methylation (Figs. 4E,F,S4G) (Bourguet et al., 2021; Yelagandula et al., 2014). CMT3-induced CHG hypermethylation was also tightly associated with transcriptionally repressive H3K9me2 marks, which are required for interdependent feedback loops with CMT3 (Fig. 4G). It was inversely related to H2A.Z (Fig. 4H) and marks indicative of active transcription including, H3K4me3 and H3K9ac (Figs. 4I,S4H). Further suggesting that deregulated CMT3 prefers features typically associated with inaccessible chromatin, genes with CHG hypermethylation had reduced chromatin accessibility (Fig. S4I) and were generally closer to heterochromatic centromeres (Fig. S4J). Moreover, the most CMT3-induced hypermethylated genes (i.e. cluster 4) were also substantially hypermethylated in *ddm1* mutants (Fig. 4J) that have increased heterochromatic accessibility (Fig. 4K) and decreased stability (Mathieu et al., 2003; Soppe et al., 2002; Zhong et al., 2021). Although CMT3-induced CHG hypermethylation was strongly associated with CG gene-body methylation (gbM), both the independence of developmental mCHG DMRs (Fig. 1E,F) and the gain of mCHG being associated with proportional loss of mCG over genes in ddm1 mutants (Fig. S4K) (Ito et al., 2015; Stroud et al., 2013; Zemach et al., 2013) indicate that mCG is not strictly required for ectopic CHG hypermethylation of genes. Instead, the associations between chromatin features of genes and their propensity for CMT3-induced hypermethylation altogether suggest that excessive CMT3 is ectopically recruited to genic chromatin characterized by nucleosome stability and inaccessibility.

Impact of CMT3-induced hypermethylation on gene expression

Because CHG methylation of TEs contributes to their repression (Stroud et al., 2014), we tested whether CMT3-induced ectopic CHG hypermethylation of protein-coding genes also represses their expression levels. Namely, we performed mRNA-seq on three biological replicates of wild type and rCMT3 (line #s 1 and 3) bent cotyledon embryos. Principal component analysis revealed that wild type and rCMT3 biological replicates clustered according to genotype and in similar positions along the dominant principal component axis corresponding to developmental time (Fig. 5A). This indicates that our mRNA-seq datasets captured gene expression variation inherent to wild type and rCMT3 genotypes, as well as that our staging was accurate. Differences in global transcript levels were not observed across the four clusters with increasing levels of CMT3-induced CHG methylation suggesting that ectopic CHG methylation alone was not sufficient to globally repress gene expression (Figs. 5B,S5A). We then identified 916 genes that were differentially expressed between rCMT3 and WT embryos (i.e. $\geq$2-fold differences and adj. *P* values $\leq$0.01; see Methods) (Fig. S5B,C and Table S3). Differentially expressed genes (DEGs), defined by comparing either rCMT3 line #1 or rCMT3 line #3 with wild type, were commonly detected in both independently generated lines with 87.5% of genes overlapping (Fig. S5C). In both rCMT3 lines, DEGs were less hypermethylated compared to all expressed genes, which indicates that the vast majority of changes in gene expression observed upon up-regulation of CMT3 was not directly due to their hypermethylation (Fig. S5D). We then examined whether hypermethylation affects a subset of genes by computing DMRs in rCMT3 compared to WT bent cotyledon embryos and identified 4,603 (97% of total) and 127 (3% of total) CHG hypermethylated and hypomethylated DMRs, respectively (Table S4; see Methods). Further suggesting that CHG hypermethylation has minimal direct consequences on the expression of most genes under the conditions examined, we found that only a small but significant number of the down-regulated genes (including 1.5 kb regions flanking their transcriptional units) overlapped DMRs (21 of 542, 3.8% of total; Fisher's exact test, *P* value = 1.29^{e-05}) (Fig. S5E). Consistent with CMT3-induced hypermethylation

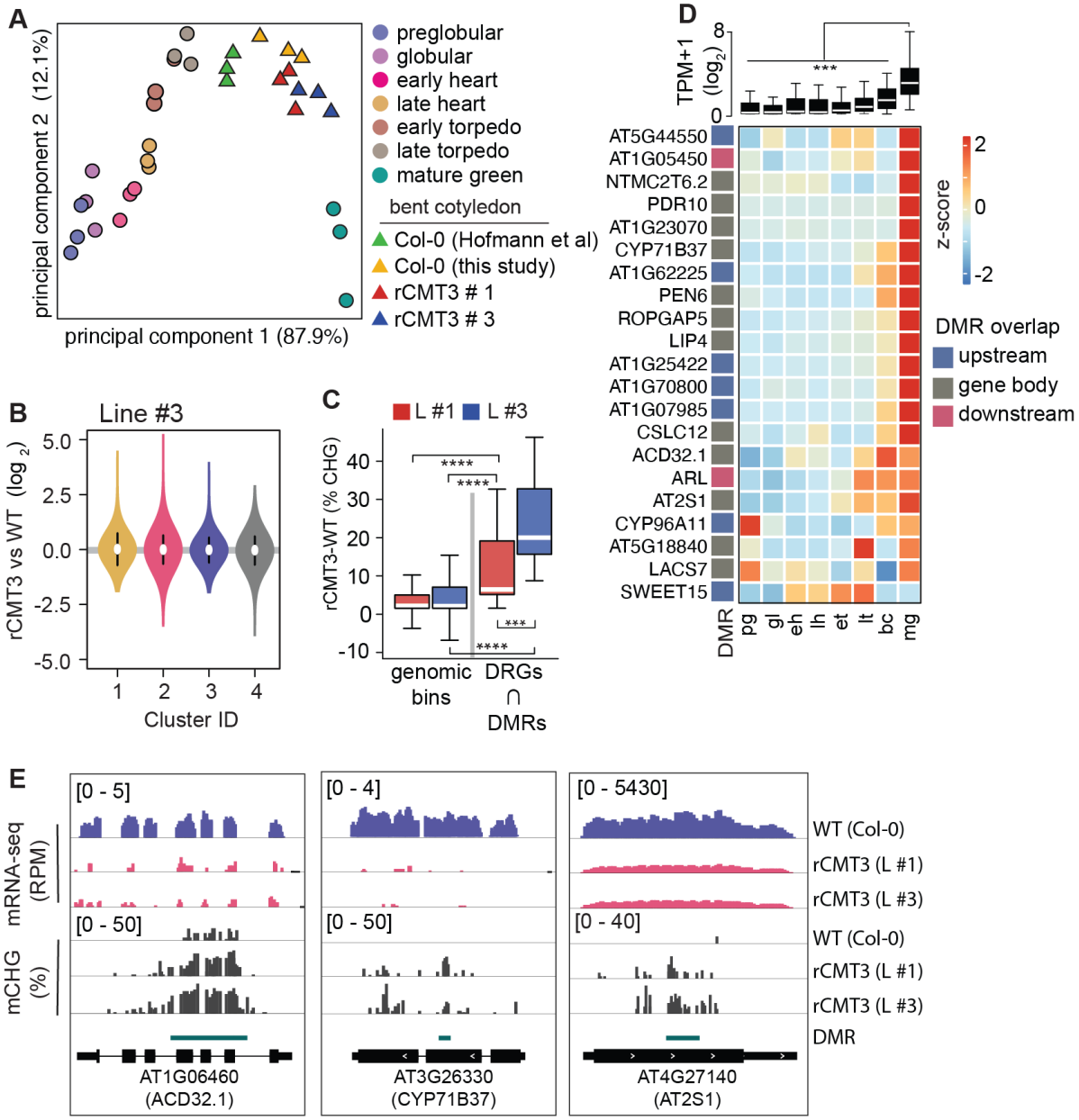

Figure 5: Impact of CMT3-induced hypermethylation on gene expression. A) Principal component analysis of mRNA-seq from rCMT3 and wild type (WT; Col-0) bent cotyledon embryos generated in this study along with floral buds, embryos, leaves, from (Hofmann et al., 2019) and color-coded according to the key. **B)** Violin plot showing transcript fold changes in rCMT3 (line #3) compared to wild type (WT; Col-0) bent cotyledon embryos per cluster as defined in Fig. 4A. **C)** Boxplot showing difference in methylation comparing rCMT3 to WT in down-regulated genes (DRG) intersecting with DMRs and similarly sized genomic bins of 213-bp as controls. *P* values < 0.001 and < 0.0001 based on Mann-Whitney U tests are represented by *** and **** respectively. **D)** Boxplot (*top*) and heatmap (*bottom*) of transcript levels of DRGs intersecting DMRs during embryogenesis. *P* values < 0.001 based on differences in transcript levels between mature green (mg) and all other stages of embryogenesis based on Mann-Whitney U test are represented by ***. **E)** Integrative genome viewer (IGV) screenshot of representative down regulated genes associated with DMRs.

repressing their expression, the DMRs overlapping these 21 down-regulated genes were significantly CHG hypermethylated compared to genomic bins (Fig. 5C). Moreover, the stronger-expressing rCMT3 line #3 had significantly higher CHG methylation compared to rCMT3 line #1 (Fig. 5C). This further supports that increased CMT3 levels lead to more ectopic CHG methylation (Inagaki et al., 2010, 2017). However, the transcript levels of these 21 genes were only moderately reduced in rCMT3 line #3 compared to rCMT3 line #1, suggesting non-linear relationships between gene hypermethylation and transcript levels (Fig. S5F). Strikingly, transcripts corresponding to these 21 CMT3-induced hypermethylated and down-regulated genes were rapidly increasing when embryos were transitioning to the maturation phase (Fig. 5D). Moreover, 10 of these 21 genes (Fisher's exact test, *P* value = 1.49^{e-13}), were among a group of 381 genes previously identified to also be rapidly activated at these time points (Fig. S5G) (Hofmann et al., 2019). Nearly half of these 381 genes (n = 183, 48%) were also among the 563 significantly down-regulated genes in rCMT3 embryos compared to wild type. Altogether, our expression and methylation analyses suggest that when CMT3 is not properly repressed it can induce ectopic hypermethylation of genes. Furthermore, CMT3-induced hypermethylation can reduce the steady state levels of transcripts from genes that are in the process of switching from silent to active transcriptional states.

DISCUSSION

DNA methylation is faithfully propagated across cell cycles by methyltransferases to ensure robust silencing of TEs (Borges et al., 2021; Law and Jacobsen, 2010; Mathieu et al., 2007; Ning et al., 2020; Probst et al., 2009; Saze et al., 2003). However, it is not well understood how DNA methyltransferases are regulated following periods of rapid division to prevent off-targeting of genes and their consequential repression. Cell division rates are highly dynamic during Arabidopsis embryogenesis. We found that the expression of MET1 and CMT3 methyltransferases and corresponding CG and CHG methylation are intricately linked to mitotic indices through distinct mechanisms (Fig. 1). Moreover, miR823-mediated cleavage and repression of CMT3 following the proliferative early phase of embryogenesis helps prevent excess CMT3 from ectopically methylating protein-coding genes that can persist for weeks afterwards (Fig. 3). CMT3-induced hypermethylation of genes was highly associated with features conferring nucleosome stability (Fig. 4) and resulted in the repression of genes that are transcriptionally activated (Fig. 5). Repression of CMT3 following a period when it is needed in high quantity to keep pace with TE methylation therefore prevents CMT3 from ectopically targeting protein-coding genes for methylation. This resulting epigenetic collateral damage on protein-coding genes can negatively affect gene expression. Our results are consistent with the model that CMT3-induced epimutations give rise to CG gene-body methylation (gbM) that can be maintained by MET1 across many generations (Wendte et al., 2019).

Complex mechanisms are required to specifically silence mutagenic TEs rather than endogenous genes (Antunez-Sanchez et al., 2020; Deng et al., 2016; Lee et al., 2021; Lister et al., 2008; Papareddy et al., 2020; Saze and Kakutani, 2011; Williams et al., 2015; Zhang et al., 2020). Mechanisms regulating epigenome homeostasis are of paramount importance during Arabidopsis embryogenesis due to highly dynamic cell cycle and transcriptional activities, as well as the establishment of cell lineages that will produce all future cell types including the gametes. MET1 and CMT3 methyltransferases are required for TE methylation (Kato et al., 2003; Stroud et al., 2014) and are expressed at high levels during early embryogenesis likely because this is a period of rapid cell division. CHG and CHH methylation exhibit opposite developmental dynamics depending on the tissue's mitotic index (Figs. 1,2,6) (Papareddy et al., 2020). When embryos are transitioning to stages with reduced cell division, decreased CMT3-mediated CHG methylation is correlated with increased CMT2-mediated CHH methylation (Fig. 2H). Unlike CMT2, CMT3 can also target protein-coding genes for CHG methylation (Stroud et al., 2014) and lead to the recruitment of transcriptionally repressive H3K9me2 methyltransferases such as KYP (Du et al., 2014; Jackson et al., 2002; Lindroth et al., 2001). Therefore, handing over TE silencing to CMT2-dependent CHH methylation in cells with reduced division rates likely reduces ectopic methylation of protein-coding genes. In addition to what we observed during embryogenesis, varying degrees of mitotic indices across devel-

opment can readily explain the genome-wide patterns of non-CG methylation reported thus far (Borges et al., 2021; Calarco et al., 2012; Gutzat et al., 2020; Ji et al., 2019; Kawakatsu et al., 2016, 2017; Lin et al., 2017; Narsai et al., 2017; Papareddy and Nodine, 2021).

CMT3, KYP and their corresponding DNA and histone methylation marks form interdependent feedback loops that perpetuate silencing through cell divisions (Du et al., 2015; Ning et al., 2020). Consistent with the transcription-coupled H3K9me2 demethylase IBM1 breaking these loops and preventing ectopic CHG hypermethylation of genes, we found that CMT3, KYP and IBM1 were highly expressed during early embryogenesis (Fig. 3). After this rapidly dividing morphogenesis phase, transcripts from CMT3, KYP and IBM1 decrease, and miR823 directs the cleavage and repression of excess CMT3 to help prevent hypermethylation of protein-coding genes (Fig. 3). Excess CMT3 induces CHG methylation on distinct regions of protein-coding genes that are characteristic of stable nucleosomes including transcriptionally repressive H3K9me2 marks that bind to CMT3. Although the distribution of CMT3-induced CHG hypermethylation is strikingly similar to CG gene-body methylation of genes (Fig. 4B,C), this appears to be due to common targeting mechanisms by CMT3 and MET1 rather than a strict prerequisite of CG. In fact, mutants with reduced CG methylation (Fig. S4K) (Jacobsen and Meyerowitz, 1997; Lister et al., 2008; Saze and Kakutani, 2007; Stroud et al., 2013) or species largely devoid of genic CG methylation (Wendte et al., 2019) can still recruit CHG on genes. CMT3-induced CHG methylation of genes that we observed in rCMT3 transgenic plants was similar to ectopic gain of genic mCHG in *ddm1* mutants (Fig. 4). Notably, heterochromatin becomes destabilized in *ddm1* mutants (Fig. 4K) (Mathieu et al., 2003; Soppe et al., 2002) and CMT3 prefers features associated with stable (Figs. 4E,F,S4G) (Bourguet et al., 2021; Osakabe et al., 2018; Yelagandula et al., 2014) over unstable nucleosomes such as H2A.Z with active marks (Figs. 4H,I,S4H). Although destabilization of heterochromatin has been inversely correlated with genic CHG methylation (Ito et al., 2015; Zhang et al., 2020), chromatin features underlying this mechanism are unclear. Therefore, we propose that destabilization of heterochromatin in *ddm1* mutants redirects CMT3 to genic regions with stable nucleosomes. Therefore, factors such as DDM1 that stabilize heterochromatin may be yet another mechanism required to regulate CMT3 activity in order to achieve proper epigenome homeostasis (Fig. 6).

CMT3-induced CHG hypermethylation of genes did not globally affect steady state transcript levels (Fig. 5). However, we observed exceptional association between CHG hypermethylation and repression of genes that switch from transcriptionally inactive to active states. Because IBM1-mediated removal of H3K9me2 marks is coupled to transcription (Inagaki et al., 2017), it is possible that CMT3-induced methylation can form feedback loops with H3K9me2 methyltransferases when genes are transcriptionally inert. However, when genes are switched on, H3K9me2 could repress initial rounds of transcription before it is removed by IBM1. Accordingly, it may be difficult to detect the effects of ectopic CHG methylation on gene expression when quantifying transcripts at steady state with standard mRNA-seq. Importantly, CMT3-induced CHG hypermethylation due at least partially to loss of miR823 repression in embryos is largely maintained for weeks after detectable miRNA activity (Fig. 3). Therefore, epigenetic collateral damage occurring in embryos may also negatively impact gene expression later in life.

Transcriptional (Ning et al., 2020), post-transcriptional (Fig. 3), post-translational (Deng et al., 2016), post-hoc (Saze et al., 2008) and perhaps substrate-related (Fig. 4) mechanisms fine-tune CMT3 activities to levels required to specifically silence mutagenic TEs but not genes. However, errors in restricting CMT3 to heterochromatin are inevitable on an evolutionary timescale (Zhang et al., 2020) and recent studies indicate that CMT3-induced methylation of genes precedes gbM (Wendte et al., 2019). Because gbM can be stably maintained over many generations by MET1 and its functional significance is debatable (Bewick et al., 2016, 2019; Choi et al., 2019; Coleman-Derr and Zilberman, 2012; Le et al., 2020; Picard and Gehring, 2017; Shahzad et al., 2021; Takuno and Gaut, 2013; Wendte et al., 2019; Williams et al., 2021; Zilberman, 2017), it cannot be excluded that gbM is merely an evolutionary record of epigenetic collateral damage events that occurred in the past (Bewick and Schmitz, 2017;Bewick et al., 2017). Our results suggest that

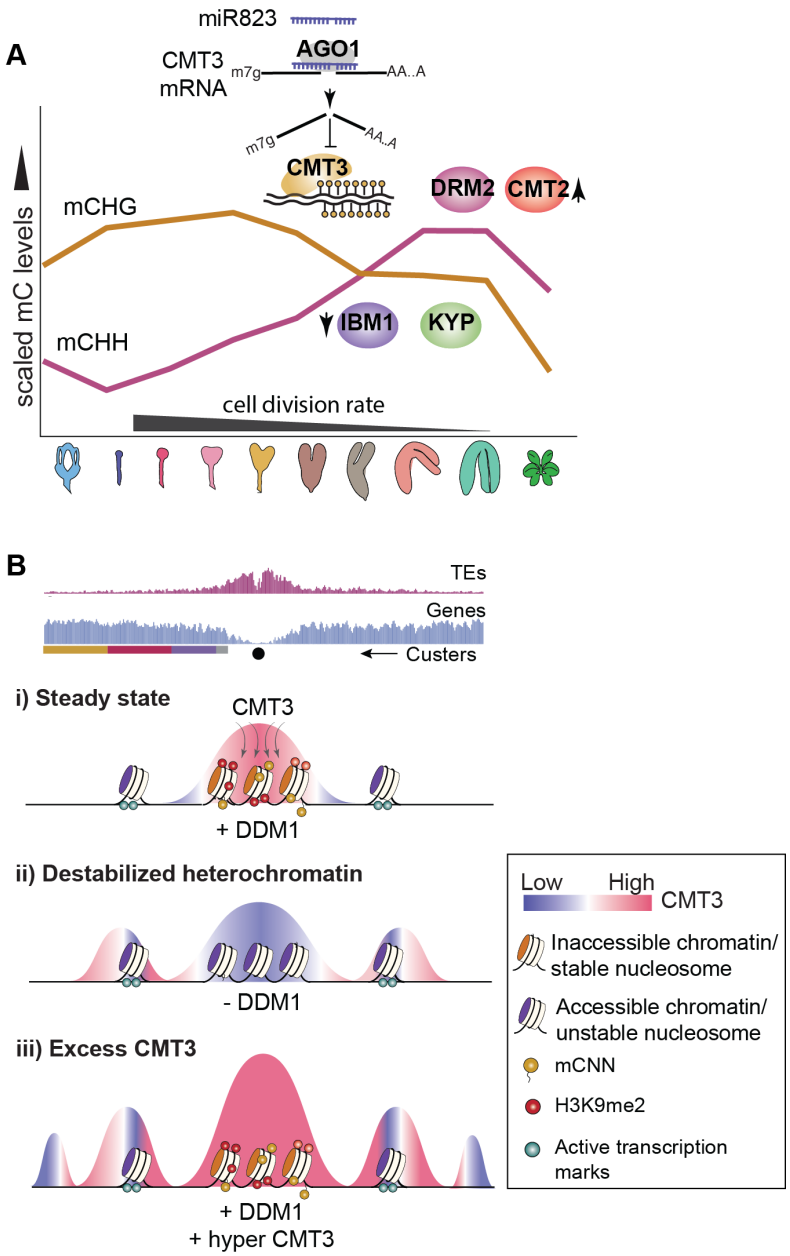

Figure 6: Models for CMT3 regulation during periods of fluctuating cell division rates and destabilized chromatin. A) Model of non-CG methylation dynamics during embryo development and corresponding regulatory mechanisms. **B)** Model for how CMT3 equilibrium is maintained to restrict its activity to heterochromatin. Density of transposable elements (TEs) (*top; red*) and genes (*middle; blue*) on chromosome 1. Cartoon illustration of gene cluster location (*bottom*) according to key in Fig. 4A. Black dot represents the centromere. (i) In steady state, stable nucleosomes along with H3K9me2 and DNA methylation provides positive reinforcement to sequester CMT3 to constitutive heterochromatin. (ii) Loss of DDM1 results in destabilized and accessible heterochromatin (Zhong et al., 2021), characterized by loss of H3K9me2 and stable nucleosomes (Osakabe et al., 2021). Accessible chromatin or DNA without stable nucleosomes is no longer a preferable substrate for CMT3 and results in CHG hypomethylation of TEs. CMT3 will now be readily available and redirected to genic regions where it induces ectopic CHG methylation in proportion to the levels of stable nucleosomes and chromatin marks. (iii) Excess levels of CMT3 causes genome-wide CHG hypermethylation with a preference for stable nucleosomes associated with repressive marks that tend to be in regions closer to centromeres compared to chromosomal arms.

derepressed CMT3 and MET1 both prefer genic regions characterized by increased nucleosome stability (Fig. 4). Accordingly, CMT3-induced CHG hypermethylation tends to occur away from transcription start and end sites of genes in a nearly identical pattern as observed for gbM (Fig. 4). We propose that CHG methylation is more tolerated in central/3' biased regions because they are relatively inaccessible to trans-acting factors that regulate transcription. Moreover, our results suggest that CMT3-induced hypermethylation can repress genes that are transcriptionally activated (Fig. 5). Perhaps genes that are consistently expressed can accumulate CHG methylation without having a large effect on steady state transcript levels and resulting fitness penalties, and thus be more likely to accumulate gbM over evolutionary time. In other words, miR823-mediated repression is one of several ways to prevent CMT3 from ectopically methylating protein-coding genes. However, CMT3 off-targeting on genes may still occur despite these complex regulatory mechanisms and the resulting epigenetic collateral damage can be recorded as heritable gbM. The characteristic features of gbM may not pertain to its current functions, but rather the consequences of transient CHG methylation that occurred in the past and were selected on during evolution.

MATERIALS AND METHODS

Plant material and growth conditions

Arabidopsis thaliana accession Columbia-0 (Col-0) were grown in controlled growth chambers at 20-22°C under a 16-h light/8-h dark cycle with incandescent lights (130 to 150 µmol/m^2/s).

Generation of transgenic lines

The control genomic CMT3 construct (miR823-cleavable; cCMT3) was generated by PCR amplification of the CMT3 locus including 1,408 bp upstream and 730 bp downstream of the TAIR10-annotated transcription start and end sites, respectively. PCR primers included overhangs for subsequent Gibson assembly into MultiSite-Gateway destination vector pAlligatorR43 (Kawashima et al., 2013). The miR823-resistant CMT3 construct (rCMT3) was generated by PCR site-directed mutagenesis (Q5 Site-Directed Mutagenesis Kit, New England Biolabs) using the cCMT3 construct as a template to introduce six silent mutations as shown in Fig. S3C. Both cCMT3 and rCMT3 construct sequences were analyzed for mutations using Sanger sequencing. All primers used are listed in the Table S5. The constructs were transformed into *cmt3-11T* (SALK_148381) using the Agrobacterium floral dip method (Clough and Bent, 1998), and transformants were selected based on seed-coat RFP signal under fluorescent light (Zeiss SteREO DiscoveryV.8). Multiple independent first-generation transgenic (T1) lines were identified for cCMT3 and rCMT3, and three and four were respectively characterized in bent cotyledon embryos for each.

Generation of CRISPR/Cas9 knockout mutants for *MIR823*

CRISPR/Cas9 knockout mutants in MIR823 were created by using a modified pHSE401 binary vector (Addgene #62201) according to the protocol detailed by (Xing et al., 2014). Primers containing the sequences for the two guide RNAs targeting the *MIR823* locus flanking the miR823 sequence (Fig. S3A and Table S5) were amplified together with the pCBCD-T1T2 plasmid (Addgene #50590), and the resulting PCR product was subsequently assembled into the pHSE401 binary vector using GoldenGate cloning method (Xing et al., 2014). Plants were transformed with the floral dip method as described above; and Cas9-positiveseeds were selected based on the presence of seed coat RFP signal. Deletion lines were identified with PCR using primers flanking gRNA-targeted sites (Fig. S3A and Table S5). Deletion mutants were confirmed and mapped by Sanger sequencing.

qRT-PCR analysis

Leaves (two-week old rosettes), floral clusters (five weeks) and bent cotyledon embryos (eight DAP) were homogenized in 500 µl TRIzol reagent (Invitrogen) and total RNA was isolated and purified according to manufacturer's recommendations. For mRNA, 200 ng of total RNA was used for cDNA synthesis with SuperScript III Reverse Transcriptase (Thermo Fisher Scientific). The cDNA was diluted two-fold for embryos or ten-fold for leaves and floral buds with nuclease-free water. Two µL of diluted cDNA was used as a template for the qRT-PCR with Fast SYBR Green Master Mix (Roche) on a LightCycler 96 instrument (Roche) with two technical replicates for each biorep. For miRNA823 quantification, cor-

responding stem-loop primers were added to the RT reaction (adapted from Yang et al., 2014) and miR823 levels were measured using Fast SYBR Green Master Mix (Roche) with miRNA823 specific forward primer and a stem-loop specific universal reverse primer. U6 snRNA was used as the reference RNA (adapted from Shen et al., 2010). Primers used for qRT-PCR are listed in Table S5.

Embryo isolation and nucleic acid extraction

Bent cotyledon embryos were dissected from seeds eight days after pollination and also selected based on morphology to ensure accurate staging. Embryos were serially washed 4× with nuclease-free water under an inverted microscope. Approximately 50 embryos per replicate were isolated and stored at -80°C until further use. RNA was isolated as previously described (Lutzmayer et al., 2017; Plotnikova et al., 2019). Genomic DNA was extracted from embryos and three-week old plants using Quick-DNA^{TM} Micro prep Kit (Zymo D3020) according to the recommendations of the manufacturer.

DNA methylation profiling and analysis

MethylC-Seq libraries were generated as described previously (Papareddy et al., 2020) and sequenced in single-read mode on an Illumina HiSeq 2500 or Nextseq 550 instrument. Adapters and the first six bases corresponding to random hexamers used during the pre-amplification step were trimmed from MethylC-seq reads using *Trim Galore*. Bisulfite-converted reads were aligned against the TAIR10 genome (Lamesch et al., 2012) in non-directional mode using Bismark (*bismark –non_directional -q –score-min L,0,-0.4*) (Krueger and Andrews, 2011). *Methylpy* software was used to extract weighted methylation rates for each available cytosine from BAM files containing only deduplicated and uniquely mapped reads (Schultz et al., 2015). Reads mapping to the unmethylated chloroplast genome were used to calculate bisulfite conversion rates. FASTQ files obtained from publicly available methylomes generated from sperm (Ibarra et al., 2012), early torpedo (Pignatta et al., 2015), mid-torpedo to early maturation (Hsieh et al., 2009), mature green embryos (Bouyer et al., 2017) and DNA methylation mutant leaves (Stroud et al., 2013) were also processed in a similar manner except that alignments were performed in directional mode and only 5′ end nucleotides of the reads with m-bias were removed. Differentially methylated regions (DMRs) were identified using *Methylpy* (Schultz et al., 2015).

Briefly, biological replicates were pooled and differentially methylated cytosines (DMCs) were identified by root mean squared tests with false discovery rates $\leq$0.01. DMRs were defined by collapsing DMCs with $\geq$4 reads within 500 bps to single units requiring $\geq$8 and $\geq$4 DMCs for CG and CHN sites, respectively (N = A,T,C,G; H $\neq$ G). Using these parameters, DMRs were identified across floral bud, early heart, early torpedo, bent cotyledon, mature green and leaf samples, and merged into a single bedFile using the BEDtools *merge* function (Quinlan and Hall, 2010). Resulting DMRs were then used to calculate the methylation rate on all analyzed tissues and genotypes. We assigned that a gene and a DMR are associated if the DMR is overlapping within 1.5kb upstream or downstream of TAIR10 annotated gene bodies using BEDtools *closest* function. For down-regulated genes overlapping with DMRs with above criteria, significance was tested using BEDtools *fisher* function with nuclear genome as a background control.

mRNA profiling and analysis

Smart-seq2 mRNA libraries were generated from 1 µl of the 7 µl bent cotyledon embryo total RNA as previously described (Hofmann et al., 2019; Picelli et al., 2014). Both amplified cDNA and final libraries were inspected using Agilent HS NGS Fragment Kit (DNF-474) to control for library quality and proper length distributions. Libraries were sequenced in single-read mode on an Illumina HiSeq 2500 or NextSeq 550 machine. Raw FASTQ files from technical replicates were merged, quality filtered and trimmed for adapter sequences with *Trim Galore* using default parameters. Trimmed reads were aligned using STAR (Dobin et al., 2013) against a genome index generated using the TAIR10 genome fasta file and all transcripts in the GTF of Ensembl build TAIR10 annotation set (release version 44). Aligned transcriptome bam files were used to quantify read counts per gene and transcript abundance using RSEM (Li and Dewey, 2011). Along with the transcriptomes generated in this study, publicly available embryonic transcriptomes (Hofman et al., 2019) used for PCA were analyzed in the same fashion as described above (Table S6). Prior to PCA (Fig. 5A), read

counts derived from nuclear protein-coding genes were subjected to variance stabilizing transformations using DESeq2 (Love et al., 2014). Differential gene expression analysis was performed using DESeq2 for genes with at least five aligned reads. Genes with $\geq$2-fold differences and adjusted p-value $\leq$ 0.01 were classified as differentially expressed genes (DEGs). Nearest-neighbour genes in Fig. 1A,B were classified based on Euclidean distance. First, the centroid expression of MET1 and VIM1/2/3 was calculated for all tissue types represented in the developmental time series. This centroid value was then used to calculate Euclidean distance of all TAIR10-annotated protein-coding genes and sorted based on their distances.

ChIP-seq analysis

ChIP-seq data for H2A variants and H3K9me2 were downloaded from GSE50942 (Yelagandula et al., 2014) and GSE51304 (Stroud et al., 2014) respectively. H3K9 acetylation marks were from GSE98214 (Wang et al., 2019). H3K4me3 marks were obtained from GSE152243 (Maher, 2020). All FASTQ files were trimmed and quality filtered using *Trim Galore* default parameters. Trimmed reads were aligned against the TAIR10 genome using BWA-MEM (Li and Durbin, 2009). Multi-mapping reads and clonal duplicates were removed using *MarkDuplicates* from the Picard Tools suite (Toolkit, 2019). The resulting BAM files containing alignments were sorted, indexed and used as input for the *bamCoverage* function of deepTools (Ramírez et al., 2014) to obtain genome normalized coverage with parameters *–normalizeUsing 'RPGC'*. Processed bigwig files for H1 Chromatin Affinity purification followed by sequencing (ChAP) and DNase-seq datasets were obtained from GSE122394 (Choi et al., 2019). MNase-Seq data was obtained from GSE113556 (Rutowicz et al., 2019). ATAC-seq processed bigwig files for wild type and *ddm1* mutants were from GSE155503 (Zhong et al., 2021).

Metaplots

ChIP, ATAC, MNase, DNase and MethylC-seq metaplots were plotted using the R library Seqplots (Stempor and Ahringer, 2016). Body, upstream, and downstream regions of TEs or genes were split into equal-sized bins, and the average levels for each bin was calculated and plotted.

CMT3 transgene copy number estimation

CMT3 transgene copy number was estimated using two methods: qPCR and coverage calculation. For the qPCR method, genomic DNA was extracted from leaves of three-week old plants using the CTAB DNA isolation method (Aboul-Maaty and Oraby, 2019). Relative transgene copy number was determined by using the qPCR-based method as described (Shepherd et al., 2009). *ACTIN2* was used as a control gene while transgene copy number was calculated based on CMT3 levels. For the coverage method, Bismark-aligned and deduplicated BAM files from wild type, cCMT3 and rCMT3 lines were processed with DeepTools to obtain normalized genome coverage as bins per million mapped reads (BPM) units with the *bamCoverage* function and following parameters: *–binsize 50 –skipNAs –normalizeUsing 'BPM' –ignoreForNormalization mitochondria chloroplast*. The resulting bigwig files were used to calculate genome-wide coverage fold-changes relative to wild type using the deepTools function *bigwigCompare –skipNAs –operation "ratio"*. CMT3 locus was displayed with the Integrative Genomics Viewer (IGV)

Availability of data and material

All sequencing data generated in this study are available at the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO, https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE171198. ChIP-Seq and mRNA-seq bioinformatic analysis pipelines were based on Nextflow (Di Tommaso et al., 2017) and the nf-core framework (Ewels et al., 2020) are available at https://github.com/Gregor-Mendel-Institute/RKP2021-CMT3.

CONTRIBUTIONS

R.K.P. and M.D.N. conceived the project; R.K.P. and K.P. developed the methodology; R.K.P., K.P. and A.D.S. conducted the experiments; R.K.P. and P.H. developed the software and performed formal analysis; C.B. supervised P.H; R.K.P. and M.D.N. wrote and edited the manuscript; M.D.N. supervised the project and acquired funding.

ACKNOWLEDGEMENTS

We thank the Vienna Biocenter Core Facilities GmbH (VBCF) Next Generation Sequencing and Plant Sciences Facilities for next-generation sequencing and plant growth chamber access, respectively, and the Institute of Molecular Pathology-Institute of Molecular Biology-Gregor Mendel Institute Molecular Biology Services for instrument access and support. R.K.P personally thanks Pierre Bourguet and Michael Borg for invaluable discussions. We also thank Zdravko Lorkovic, Bhagyshree Jamge, Robin Burns, Eriko Sasaki, Magnus Nordborg and Frédéric Berger for sharing thoughts and reagents; and members of the Nodine lab for valuable input. This work was supported by the European Research Council under the European Union's Horizon 2020 Research and Innovation Program grant 637888 to M.D.N.

REFERENCES

Aboul-Maaty, N.A.-F., and Oraby, H.A.-S. (2019). Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. Bull. Natl. Salmon Resour. Cent. 43.

Antunez-Sanchez, J., Naish, M., Ramirez-Prado, J.S., Ohno, S., Huang, Y., Dawson, A., Opassathian, K., Manza-Mianza, D., Ariel, F., Raynaud, C., et al. (2020). A new role for histone demethylases in the maintenance of plant genome integrity. Elife 9.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408, 796–815.

Bewick, A.J., and Schmitz, R.J. (2017). Gene body DNA methylation in plants. Curr. Opin. Plant Biol. 36, 103–110.

Bewick, A.J., Ji, L., Niederhuth, C.E., Willing, E.-M., Hofmeister, B.T., Shi, X., Wang, L., Lu, Z., Rohr, N.A., Hartwig, B., et al. (2016). On the origin and evolutionary consequences of gene body DNA methylation. Proc. Natl. Acad. Sci. U. S. A. 113, 9111–9116.

Bewick, A.J., Niederhuth, C.E., Ji, L., Rohr, N.A., Griffin, P.T., Leebens-Mack, J., and Schmitz, R.J. (2017). The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. Genome Biol. 18, 65.

Bewick, A.J., Zhang, Y., Wendte, J.M., Zhang, X., and Schmitz, R.J. (2019). Evolutionary and Experimental Loss of Gene Body Methylation and Its Consequence to Gene Expression. G3 9, 2441–2445.

Blevins, T., Podicheti, R., Mishra, V., Marasco, M., Wang, J., Rusch, D., Tang, H., and Pikaard, C.S. (2015). Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. Elife 4, e09591.

Borges, F., Donoghue, M.T.A., LeBlanc, C., Wear, E.E., Tanurdžić, M., Berube, B., Brooks, A., Thompson, W.F., Hanley-Bowdoin, L., and Martienssen, R.A. (2021). Loss of Small-RNA-Directed DNA Methylation in the Plant Cell Cycle Promotes Germline Reprogramming and Somaclonal Variation. Curr. Biol. 31, 591–600.e4.

Bourguet, P., Picard, C.L., Yelagandula, R., Pélissier, T., Lorković, Z.J., Feng, S., Pouch-Pélissier, M.-N., Schmücker, A., Jacobsen, S.E., Berger, F., et al. (2021). The histone variant H2A.W and linker histone H1 co-regulate heterochromatin accessibility and DNA methylation. bioRxiv doi: https://doi.org/10.1101/2020.03.19.998609

Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F., and Colot, V. (2017). DNA methylation dynamics during early plant life. Genome Biol. 18, 179.

Calarco, J.P., Borges, F., Donoghue, M.T.A., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J.A., Becker, J.D., et al. (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell 151, 194–205.

Cao, X., and Jacobsen, S.E. (2002). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc. Natl. Acad. Sci. U. S. A. 99 Suppl 4, 16491–16498.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.-Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J., et al. (2010). Relationship between nucleosome positioning and DNA methylation. Nature 466, 388–392.

Choi, J., Lyons, D.B., Kim, M.Y., Moore, J.D., and Zilberman, D. (2019). DNA Methylation and Histone H1 Jointly Repress Transposable Elements and Aberrant Intragenic Transcripts. Mol. Cell.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana: Floral dip transformation of Arabidopsis. Plant J. 16, 735–743.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452, 215–219.

Coleman-Derr, D., and Zilberman, D. (2012). Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. PLoS Genet. 8, e1002988.

Deng, S., Jang, I.-C., Su, L., Xu, J., and Chua, N.-H. (2016). JMJ24 targets CHROMOMETHYLASE3 for proteasomal degradation in Arabidopsis. Genes Dev. 30, 251–256.

Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. Nat. Biotechnol. 35, 316–319.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A., Terragni, J., Chin, H.G., Tu, A., et al. (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell 151, 167–180.

Du, J., Johnson, L.M., Groth, M., Feng, S., Hale, C.J., Li, S., Vashisht, A.A., Wohlschlegel, J.A., Patel, D.J., and Jacobsen, S.E. (2014). Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. Mol. Cell 55, 495–504.

Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J. (2015). DNA methylation pathways and their crosstalk with histone methylation. Nat. Rev. Mol. Cell Biol. 16, 519–532.

Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., and Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nat. Biotechnol. 38, 276–278.

Fan, Y., Nikitina, T., Morin-Kensicki, E.M., Zhao, J., Magnuson, T.R., Woodcock, C.L., and Skoultchi, A.I. (2003). H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. Mol. Cell. Biol. 23, 4559–4572.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. Proc. Natl. Acad. Sci. U. S. A. 107, 8689–8694.

Finnegan, E.J., and Dennis, E.S. (1993). Isolation and identification by sequence homology of a putative cytosine methyltransferase from Arabidopsis thaliana. Nucleic Acids Res. 21, 2383–2388.

Gouil, Q., and Baulcombe, D.C. (2016). DNA Methylation Signatures of the Plant Chromomethyltransferases. PLoS Genet. 12, e1006526.

Gutzat, R., Rembart, K., Nussbaumer, T., Hofmann, F., Pisupati, R., Bradamante, G., Daubel, N., Gaidora, A., Lettner, N., Donà, M., et al. (2020). Arabidopsis shoot stem cells display dynamic transcription and DNA methylation patterns. EMBO J. 39, e103667.

Henderson, I.R., and Jacobsen, S.E. (2007). Epigenetic inheritance in plants. Nature 447, 418–424.

Henderson, I.R., and Jacobsen, S.E. (2008). Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. Genes Dev. 22, 1597–1606.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science 308, 118–120.

Hofmann, F., Schon, M.A., and Nodine, M.D. (2019). The embryonic transcriptome of Arabidopsis thaliana. Plant Reprod. 32, 77–91.

Hsieh, T.-F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and Zilberman, D. (2009). Genome-wide demethylation of Arabidopsis endosperm. Science 324, 1451–1454.

Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.-F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., et al. (2012). Active DNA demethylation in plant companion cells reinforces trans-

poson methylation in gametes. Science 337, 1360–1364.

Inagaki, S., Miura-Kamio, A., Nakamura, Y., Lu, F., Cui, X., Cao, X., Kimura, H., Saze, H., and Kakutani, T. (2010). Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. EMBO J. 29, 3496–3506.

Inagaki, S., Takahashi, M., Hosaka, A., Ito, T., Toyoda, A., Fujiyama, A., Tarutani, Y., and Kakutani, T. (2017). Gene-body chromatin modification dynamics mediate epigenome differentiation in Arabidopsis. EMBO J. 36, 970–980.

Ito, T., Tarutani, Y., To, T.K., Kassam, M., Duvernois-Berthet, E., Cortijo, S., Takashima, K., Saze, H., Toyoda, A., Fujiyama, A., et al. (2015). Genome-Wide Negative Feedback Drives Transgenerational DNA Methylation Dynamics in Arabidopsis. PLOS Genetics 11, e1005154.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature 416, 556–560.

Jacobsen, S.E., and Meyerowitz, E.M. (1997). Hypermethylated SUPERMAN epigenetic alleles in arabidopsis. Science 277, 1100–1103.

Ji, L., Mathioni, S.M., Johnson, S., Tucker, D., Bewick, A.J., Do Kim, K., Daron, J., Slotkin, R.K., Jackson, S.A., Parrott, W.A., et al. (2019). Genome-Wide Reinforcement of DNA Methylation Occurs during Somatic Embryogenesis in Soybean. Plant Cell 31, 2315–2331.

Kato, M., Miura, A., Bender, J., Jacobsen, S.E., and Kakutani, T. (2003). Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. Curr. Biol. 13, 421–426.

Kawakatsu, T., Stuart, T., Valdes, M., Breakfield, N., Schmitz, R.J., Nery, J.R., Urich, M.A., Han, X., Lister, R., Benfey, P.N., et al. (2016). Unique cell-type-specific patterns of DNA methylation in the root meristem. Nat Plants 2, 16058.

Kawakatsu, T., Nery, J.R., Castanon, R., and Ecker, J.R. (2017). Dynamic DNA methylation reconfiguration during seed development and germination. Genome Biol. 18, 171.

Kawashima, T., Kitamura, K., Suzuki, K., Nonaka, M., Kamijo, S., Takemoto-Kimura, S., Kano, M., Okuno, H., Ohki, K., and Bito, H. (2013). Functional labeling of neurons and their projections using the synthetic activity–dependent promoter E-SARE. Nature Methods 10, 889–895.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571–1572.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 40, D1202–D1210.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11, 204–220.

Le, N.T., Harukawa, Y., Miura, S., Boer, D., Kawabe, A., and Saze, H. (2020). Epigenetic regulation of spurious transcription initiation in Arabidopsis. Nat. Commun. 11, 3224.

Lee, Y.-S., Maple, R., Dürr, J., Dawson, A., Tamim, S., Del Genio, C., Papareddy, R., Luo, A., Lamb, J.C., Amantia, S., et al. (2021). A transposon surveillance mechanism that safeguards plant male fertility during stress. Nat Plants 7, 34–41.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760.

Li, X., Harris, C.J., Zhong, Z., Chen, W., Liu, R., Jia, B., Wang, Z., Li, S., Jacobsen, S.E., and Du, J. (2018). Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. Proc. Natl. Acad. Sci. U. S. A. 115, E8793–E8802.

Lin, J.-Y., Le, B.H., Chen, M., Henry, K.F., Hur, J., Hsieh, T.-F., Chen, P.-Y., Pelletier, J.M., Pellegrini, M., Fischer, R.L., et al. (2017). Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. Proc. Natl. Acad.

Sci. U. S. A. 114, E9730–E9739.

Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S., and Jacobsen, S.E. (2001). Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science 292, 2077–2080.

Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature 430, 471–476.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133, 523–536.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Lutzmayer, S., Enugutti, B., and Nodine, M.D. (2017). Novel small RNA spike-in oligonucleotides enable absolute normalization of small RNA-Seq data. Sci. Rep. 7, 5913.

Maher, K.A. (2020). Applications of Next-Generation Sequencing Strategies for the Identification and Characterization of Enhancers in Plants. Emory University.

Mathieu, O., Jasencakova, Z., Vaillant, I., Gendrel, A.-V., Colot, V., Schubert, I., and Tourmente, S. (2003). Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in Arabidopsis. Plant Cell 15, 2929–2939.

Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C., and Paszkowski, J. (2007). Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. Cell 130, 851–862.

Miura, A., Nakamura, M., Inagaki, S., Kobayashi, A., Saze, H., and Kakutani, T. (2009). An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. EMBO J. 28, 1078–1086.

Muyle, A., and Gaut, B.S. (2019). Loss of Gene Body Methylation in Eutrema salsugineum Is Associated with Reduced Gene Expression. Mol. Biol. Evol. 36, 155–158.

Narsai, R., Gouil, Q., Secco, D., Srivastava, A., Karpievitch, Y.V., Liew, L.C., Lister, R., Lewsey, M.G., and Whelan, J. (2017). Extensive transcriptomic and epigenomic remodelling occurs during Arabidopsis thaliana germination. Genome Biol. 18, 172.

Niederhuth, C.E., Bewick, A.J., Ji, L., Alabady, M.S., Kim, K.D., Li, Q., Rohr, N.A., Rambani, A., Burke, J.M., Udall, J.A., et al. (2016). Widespread natural variation of DNA methylation within angiosperms. Genome Biol. 17, 194.

Ning, Y.-Q., Liu, N., Lan, K.-K., Su, Y.-N., Li, L., Chen, S., and He, X.-J. (2020). DREAM complex suppresses DNA methylation maintenance genes and precludes DNA hypermethylation. Nat Plants 6, 942–956.

Osakabe, A., Lorkovic, Z.J., Kobayashi, W., Tachiwana, H., Yelagandula, R., Kurumizaka, H., and Berger, F. (2018). Histone H2A variants confer specific properties to nucleosomes and impact on chromatin accessibility. Nucleic Acids Res. 46, 7675–7685.

Osakabe, A., Jamge, B., Axelsson, E., Montgomery, S.A., Akimcheva, S., Kuehn, A.L., Pisupati, R., Lorković, Z.J., Yelagandula, R., Kakutani, T., et al. (2021). The chromatin remodeler DDM1 prevents transposon mobility through deposition of histone variant H2A.W. Nat. Cell Biol. 23, 391–400.

Papareddy, R.K., and Nodine, M.D. (2021). Plant Epigenetics: Propelling DNA Methylation Variation across the Cell Cycle. Curr. Biol. 31, R129–R131.

Papareddy, R.K., Páldi, K., Paulraj, S., Kao, P., Lutzmayer, S., and Nodine, M.D. (2020). Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in Arabidopsis. Genome Biol. 21, 251.

Picard, C.L., and Gehring, M. (2017). Proximal methylation features associated with nonrandom changes in gene body methylation. Genome Biol. 18, 73.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171–181.

Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W., and Gehring, M. (2015). Correction: Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. Elife 4.

Plotnikova, A., Kellner, M.J., Schon, M.A., Mosiolek, M., and Nodine, M.D. (2019). MicroRNA Dynamics and Func-

tions During Arabidopsis Embryogenesis. Plant Cell 31, 2929–2946.

Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nat. Rev. Mol. Cell Biol. 10, 192–206.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–W191.

Rutowicz, K., Lirski, M., Mermaz, B., Teano, G., Schubert, J., Mestiri, I., Kroteń, M.A., Fabrice, T.N., Fritz, S., Grob, S., et al. (2019). Linker histones are fine-scale chromatin architects modulating developmental decisions in Arabidopsis. Genome Biol. 20, 157.

Saze, H., and Kakutani, T. (2007). Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. EMBO J.

Saze, H., and Kakutani, T. (2011). Differentiation of epigenetic modifications between transposons and genes. Curr. Opin. Plant Biol. 14, 81–87.

Saze, H., Mittelsten Scheid, O., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nat. Genet. 34, 65–69.

Saze, H., Shiraishi, A., Miura, A., and Kakutani, T. (2008). Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana. Science 319, 462–465.

Scheid, O.M., Afsar, K., and Paszkowski, J. (1998). Release of epigenetic gene silencing by trans-acting mutations in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 95, 632–637.

Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N., Nery, J.R., Urich, M.A., Chen, H., et al. (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. Nature 523, 212–216.

Shahzad, Z., Moore, J.D., and Zilberman, D. (2021). Gene body methylation mediates epigenetic inheritance of plant traits. bioRxiv doi: https://doi.org/10.1101/2021.03.15.435374

Shen, J., Xie, K., and Xiong, L. (2010). Global expression profiling of rice microRNAs by one-tube stem-loop reverse transcription quantitative PCR revealed important roles of microRNAs in abiotic stress responses. Molecular Genetics and Genomics 284, 477–488.

Shen, J.C., Rideout, W.M., 3rd, and Jones, P.A. (1992). High frequency mutagenesis by a DNA methyltransferase. Cell 71, 1073–1080.

Shepherd, C.T., Moran Lauter, A.N., and Scott, M.P. (2009). Determination of transgene copy number by real-time quantitative PCR. Methods Mol. Biol. 526, 129–134.

Soppe, W.J.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. EMBO J. 21, 6549–6559.

Stempor, P., and Ahringer, J. (2016). SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics. Wellcome Open Res 1, 14.

Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell 152, 352–364.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat. Struct. Mol. Biol. 21, 64–72.

Sved, J., and Bird, A. (1990). The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. Proc. Natl. Acad. Sci. U. S. A. 87, 4692–4696.

Takuno, S., and Gaut, B.S. (2013). Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. Proc. Natl. Acad. Sci. U. S. A. 110, 1797–1802.

Takuno, S., Seymour, D.K., and Gaut, B.S. (2017). The Evolutionary Dynamics of Orthologs That Shift in Gene Body Methylation between Arabidopsis Species. Mol.

Biol. Evol. 34, 1479–1491.

The Cold Spring Harbor Laboratory, Washington University Genome Sequencing Center, and PE Biosystems Arabidopsis Sequencing Consortium (2000). The Complete Sequence of a Heterochromatic Island from a Higher Eukaryote. Cell 100, 377–386.

Toolkit, P. (2019). Broad institute, GitHub repository. See Http://broadinstitute. Github. Io/picard.

Tran, R.K., Henikoff, J.G., Zilberman, D., Ditt, R.F., Jacobsen, S.E., and Henikoff, S. (2005). DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes. Curr. Biol. 15, 154–159.

Wang, L., Wang, C., Liu, X., Cheng, J., Li, S., Zhu, J.-K., and Gong, Z. (2019). Peroxisomal -oxidation regulates histone acetylation and DNA methylation in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 116, 10576–10585.

Wendte, J.M., Zhang, Y., Ji, L., Shi, X., Hazarika, R.R., Shahryary, Y., Johannes, F., and Schmitz, R.J. (2019). Epimutations are associated with CHROMOMETHYLASE 3-induced de novo DNA methylation. Elife 8.

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 135, 635–648.

Williams, B.P., Pignatta, D., Henikoff, S., and Gehring, M. (2015). Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. PLoS Genet. 11, e1005142.

Williams, B.P., Bechen, L.A., Pohlmann, D.A., and Gehring, M. (2021). Somatic DNA demethylation generates tissue-specific methylation states and impacts flowering time. bioRxiv doi: https://doi.org/10.1101/2021.03.29.437569

Woo, H.R., Dittmer, T.A., and Richards, E.J. (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. PLoS Genet. 4, e1000156.

Xing, H.-L., Dong, L., Wang, Z.-P., Zhang, H.-Y., Han, C.-Y., Liu, B., Wang, X.-C., and Chen, Q.-J. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biol. 14, 327.

Yang, L.-H., Wang, S.-L., Tang, L.-L., Liu, B., Ye, W.-L., Wang, L.-L., Wang, Z.-Y., Zhou, M.-T., and Chen, B.-C. (2014). Universal stem-loop primer method for screening and quantification of microRNA. PLoS One 9, e115293.

Yelagandula, R., Stroud, H., Holec, S., Zhou, K., Feng, S., Zhong, X., Muthurajan, U.M., Nie, X., Kawashima, T., Groth, M., et al. (2014). The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. Cell 158, 98–109.

van Zanten, M., Koini, M.A., Geyer, R., Liu, Y., Brambilla, V., Bartels, D., Koornneef, M., Fransz, P., and Soppe, W.J.J. (2011). Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation. Proc. Natl. Acad. Sci. U. S. A. 108, 20219–20224.

Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153, 193–205.

Zhai, J., Bischof, S., Wang, H., Feng, S., Lee, T.-F., Teng, C., Chen, X., Park, S.Y., Liu, L., Gallego-Bartolome, J., et al. (2015). A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis. Cell 163, 445–455.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W.-L., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell 126, 1189–1201.

Zhang, Y., Wendte, J.M., Ji, L., and Schmitz, R.J. (2020). Natural variation in DNA methylation homeostasis and the emergence of epialleles. Proc. Natl. Acad. Sci. U. S. A. 117, 4874–4884.

Zhong, Z., Feng, S., Duttke, S.H., Potok, M.E., Zhang, Y., Gallego-Bartolomé, J., Liu, W., and Jacobsen, S.E. (2021). DNA methylation-linked chromatin accessibility affects genomic architecture in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 118.

Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and animals. Genome Biol. 18, 87.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumu-

lation and DNA and histone methylation. Science 299, 716–719.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. 39, 61–69.

SUPPLEMENTAL INFORMATION

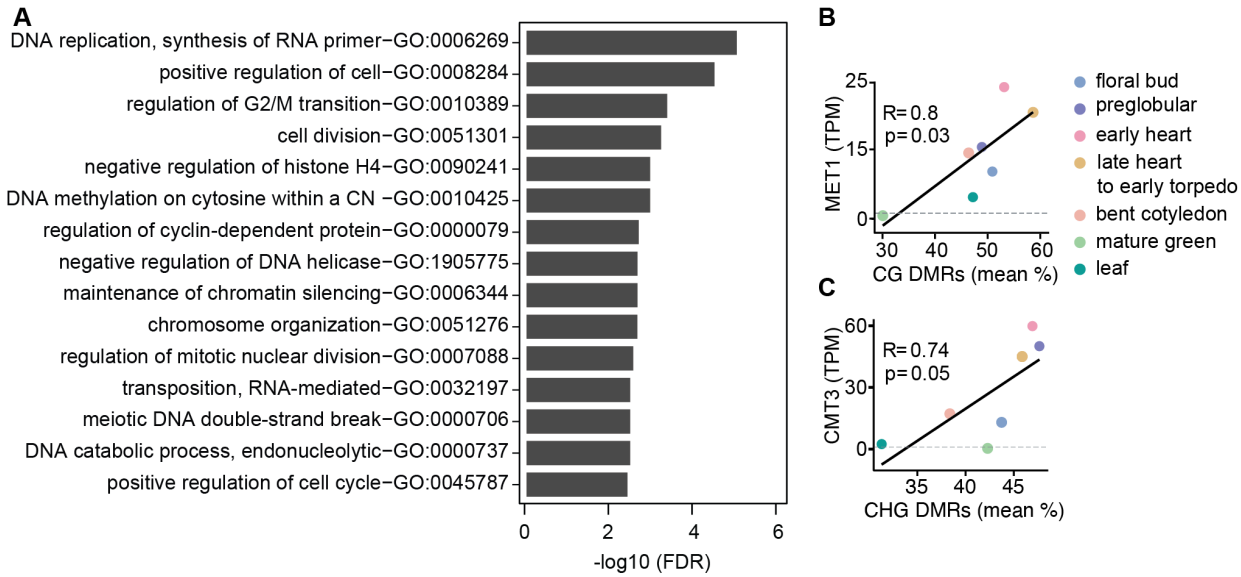

Figure S1. Characteristics of genes and differentially methylated regions co-expressed with symmetric methyltransferases (Related to Fig. 1). **A)** Gene ontology enrichment of top-50 genes co-expressed with MET1 and VIM1/2/3 with false discovery rates < 0.05. **B** and **C**) Scatterplots showing Pearson's *R* between MET1 transcript levels (TPM; transcripts per million) and mean-weighted CG methylation of developmental CG DMRs (B) or CMT3 TPM and mean-weighted CHG methylation of developmental CHG DMRs.

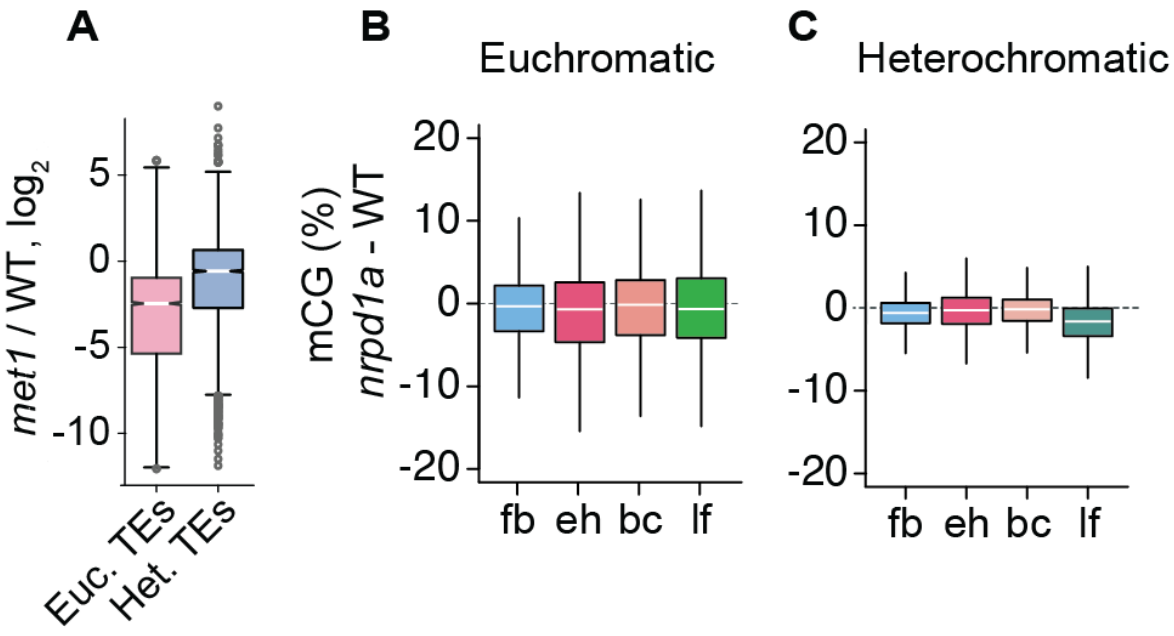

Figure S2. Relationships between MET1 and 24-nt siRNAs (Related to Fig. 2). **A)** Boxplot illustrating relative levels of 24-nt siRNAs in *met1* relative to wild type (WT) (Lister et al., 2008); Euchromatic and heterochromatic TEs are abbreviated as Euc. TEs and Het. TEs, respectively. **B** and **C**) Boxplots of CG methylation differences between *nrpd1a* and wild-type tissues for euchromatic (**B**) and heterochromatic (**C**) TEs. fb, floral buds; eh, early heart; bc, bent cotyledon; lf, leaf. Thick horizontal bars indicate medians, and the top and bottom edges of boxes indicate the 75th and 25th percentiles, respectively.

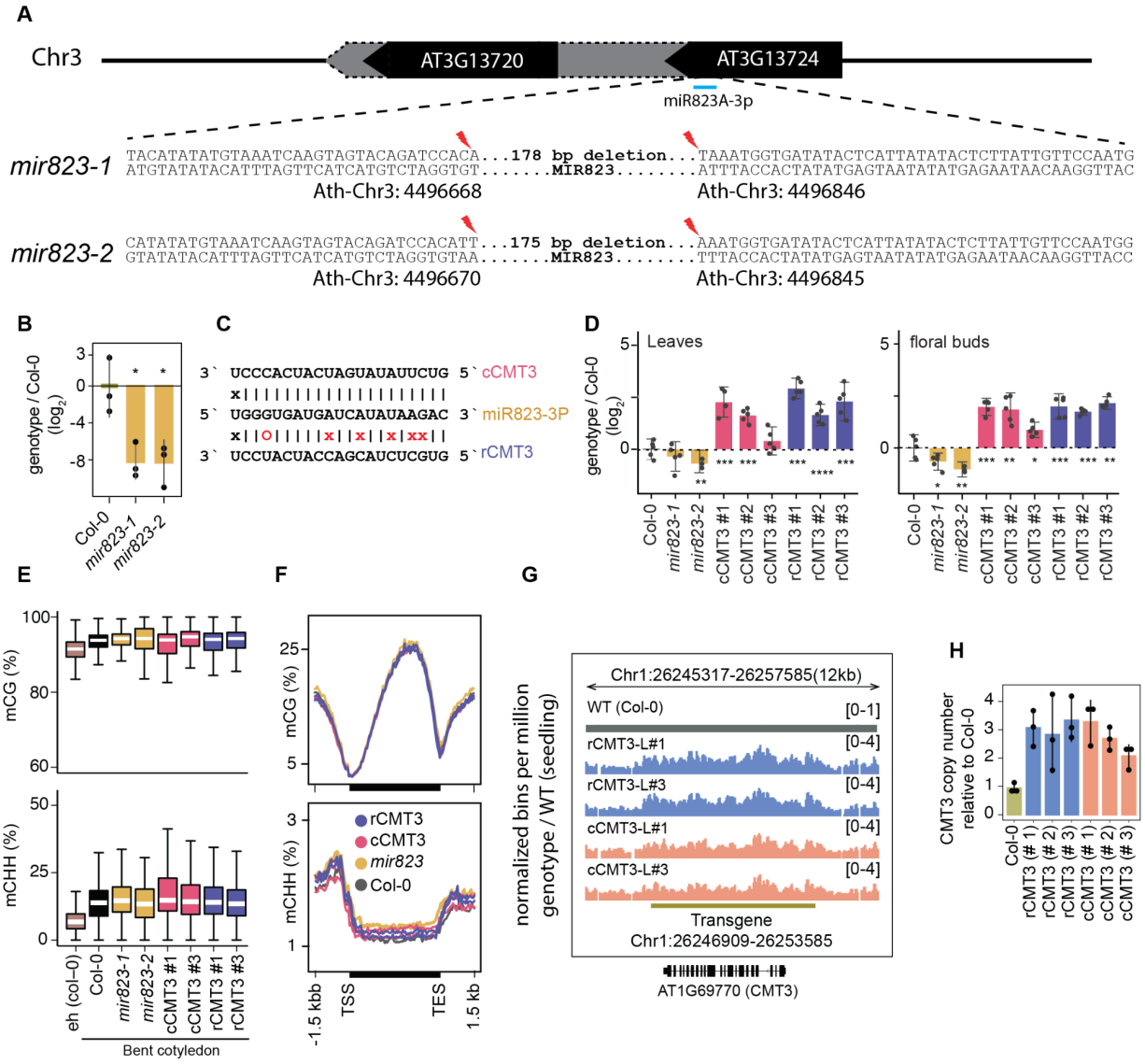

Figure S3. *mir823* mutants and effects of miR823-directed repression of CMT3 (Related to Fig. 3). **A**) Diagram of sequences deleted by CRISPR/Cas9 approach from MIR823A locus in *miR823-1* and *miR823-2*. **B**) Relative levels of miR823 in wild type (Col-0), mir823-1 and mir823-2 bent cotyledon embryos (~8 DAP; days after pollination). Stem-loop qPCR values were normalized to U6 and then divided by the levels observed in wild-type plants before $\log_2$-transformation. Each dot represents the mean of two technical replicates of embryos, and error bars indicate standard error. *P* values < 0.01 based on two-tailed Student's t-test of differences between wild type and *miR823* mutants are represented by *. **C**) Schematics of miR823 target site in CMT3 transcripts. Base-pairing interactions of miR823 with either wild-type target sites (cleavable, cCMT3) or miRNA-resistant target sites (resistant, rCMT3) are indicated above and below, respectively. Mutations introduced are labeled in red, and Watson-Crick

base-pairing (I), non-base-pairing (X), and G:U wobbles (O) for each pair are indicated. **D**) Relative CMT3 transcript levels in two-week old leaves (*left*) or floral buds (*right*) from wild type plants (Col-0), *mir823-1*, *mir823-2*, or *cmt3-11* plants carrying either miR823-cleavable CMT3 (cCMT3) or miR823-resistant CMT3 (rCMT3) transgenes. Bars represent mean values and error bars indicate standard errors. Asterisks indicate whether the transcript levels observed in *mir823* mutants, or cCMT3 and rCMT3 transgenics were significantly different compared to wild type (Two-tailed Student's t-tests; ****, ***, **, and * represent *P* values < 0.0001, < 0.001, < 0.01, and < 0.05, respectively). **E**) Boxplots of CG (*top*) and CHH (*bottom*) methylation on euchromatic and heterochromatic transposons with ≥4 informative cytosines covered by ≥5 reads. **F**) Metaplots of average CG (*top*) and CHH (*bottom*) methylation percentages across genes bodies, and 1.5 kb upstream and downstream regions. Genotypes are color-coded according to the key. **G**) Integrative Genomics Viewer screenshot of CMT3 locus showing normalized coverage of methylC-seq reads compared to wild type (see Methods). **H**) CMT3 copy number quantification based on qPCR in three-week old leaves from wild type (Col-0), or independently generated rCMT3 or cCMT3 transgenics in the *cmt3-11* background.

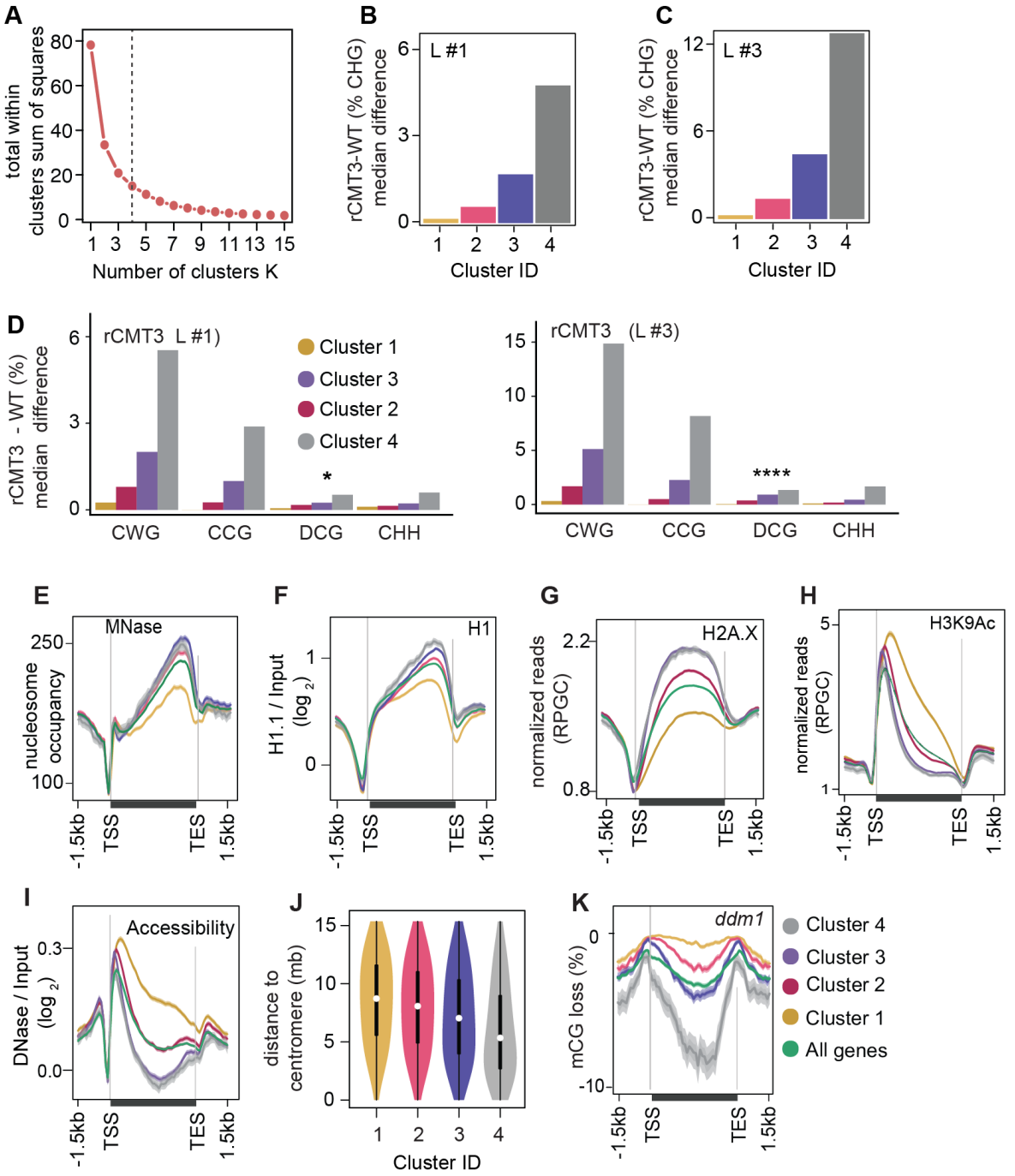

Figure S4. Partitioning of CMT3-induced hypermethylated genes and associated chromatin features (Related to Fig. 4). **A)** Determining the optimal K-value based difference in genic mCHG between rCMT3 (Line #3) and wild-type bent cotyledon embryos using the elbow method. Four clusters were selected as optimal because the total within cluster sum of squares (WSS) became marginal after a K value of four. **B** and **C**) Barplots showing median mCHG gain on annotated gene bodies in rCMT3 (Line #1) (B) and rCMT3 (Line #3) (**C**) compared to

wild-type bent cotyledon embryos. **D**) Barplots illustrating median gain of methylation in rCMT3 (Line #1) (*left*) and rCMT3 (Line #3) (*right*) compared to wild-type bent cotyledon embryos in various trinucleotide cytosine contexts where W = A or T and D ≠ C. *P* Values <0.05 and <0.0001 obtained by Mann Whitney U test based on difference in DCG methylation between rCMT3 and WT in bent cotyledon embryos were represented by * and **** respectively. **E**-**I**) Metaplots showing nucleosome occupancy obtained from MNase-seq data (Rutowicz et al., 2019) (**E**), enrichment of linker histone 1 (H1) on gene clusters (Choi et al.) (**F**), normalized reads per genomic content (RPGC) average value of histone variant H2A.X (Yelagandula et al., 2014) (**G**), H3K9Ac (Wang et al., 2019) (**H**) and DNase-Seq signal on gene clusters representing accessibility (Choi et al.) (**I**). **J**) Violin plot showing distances between genes and centromeres per cluster. White dots indicate the median and vertical black bars indicate 1.5X interquartile ranges. **K**) Metaplot of CG methylation in differences in *ddm1* compared to wild type (WT) (Stroud et al., 2013).

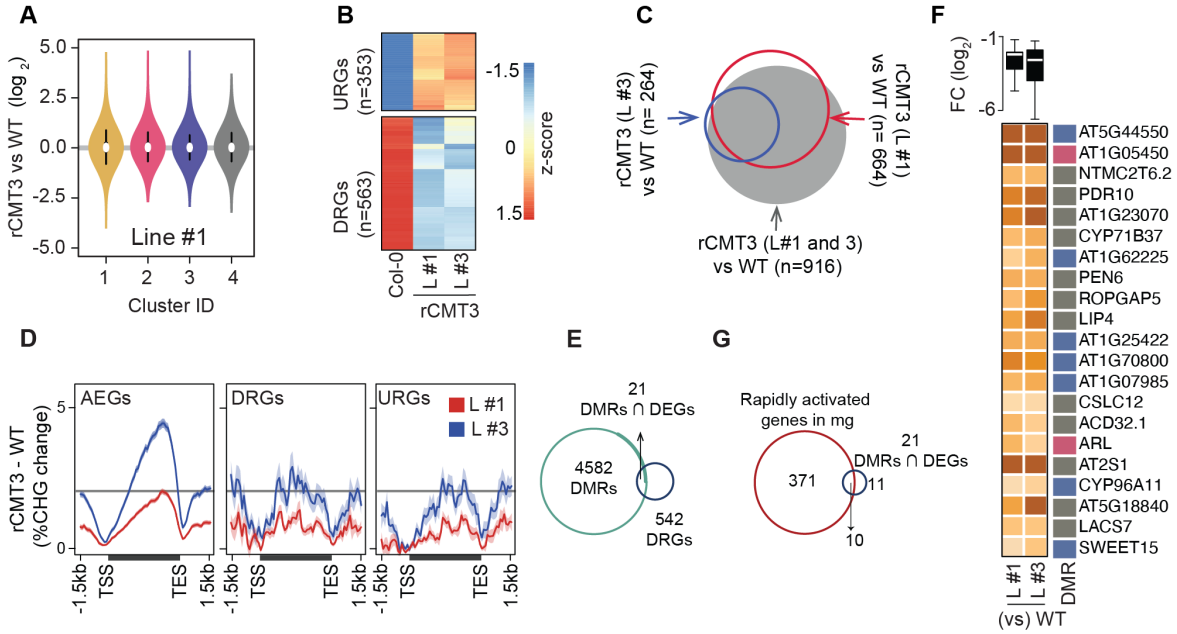

Figure S5. Additional information regarding the influence of CMT3-induced hypermethylation on gene expression (Related to Fig. 5). **A**) Violin plot showing expression fold change in rCMT3 (line #3) compared to wild type (Col-0) bent cotyledon embryos per cluster as defined in Fig. 4A. **B**) Scaled heatmap of differentially expressed genes (DEGs) when rCMT3 bent cotyledon are compared to WT with upregulated (URGs) and downregulated genes (DRGs) shown at the top and bottom, respectively. **C**) Three-way Venn diagram showing the proportion of DEGs overlapping with each labelled comparison. **D**) Metaplot of difference in CHG methylation in rCMT3 (line #1) (*red) or rCMT3 (*line #3) (*blue*) compared to wild type for all expressed genes (AEGs, *left*), downregulated genes (DRGs, *middle*) and upregulated genes (URGs, *right*). **E**) Venn diagram showing overlap between hyper differentially methylated regions (DMRs) and down regulated genes (DRGs) in rCMT3 compared to wild type bent cotyledon embryos. **F**) Boxplot showing $\log_2$ fold-change (FC) of transcript levels between rCMT3 and Col-0 (WT) for all DRGs intersecting DMRs (*top*). Heatmap showing fold-change of individual DRGs (*bottom*). **G**) Venn diagram showing overlap between genes rapidly activated at mature green stage (cluster D6 genes based on (Hofmann et al., 2019)) and 21 DRGs/DMRs.

The following supplemental tables are provided separately as Excel spreadsheets:

Table S1. Developmental CG differentially methylated regions.

Table S2. Developmental CHG differentially methylated regions.

Table S3. Transcript levels in rCMT3 compared to wild type embryos.

Table S4. CHG differentially methylated regions in rCMT3 compared to wild type.

Table S5. Oligonucleotides used in this study.

Table S6. MethylC-seq and mRNA-seq mapping statistics.