101 specifically from expressed genes (Inagaki et al., 2010; Saze et al., 2008). These and additional
102 mechanisms are likely of utmost importance during embryo development when a proliferative
103 morphogenesis phase produces the most fundamental cell lineages of the plant, including those
104 that will eventually generate the gametes. Yet, how DNA methylation pathways are regulated
105 during this phase of dynamic cell division to exquisitely balance the need for TE methylation
106 with the prevention of potentially deleterious and stably inherited epimutations is virtually
107 unknown.

Results~~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 (Figure~~.~~ 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 (Figure~~s~~. 1B, Figure 1-figure supplement~~,~~ ~~S~~ 1A). 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 (Figure~~.~~ 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 (Supplementary File ~~Table S~~1) and 1,398 CHG (Supplementary File ~~Table S~~2) differentially methylated regions (DMRs) covering 201 kb and 185.8 kb, respectively (Figure~~.~~ 1D, Supplementary File 1~~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 (Figure~~.~~ 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 (Figure 2-figure supplement 2B~~.~~ ~~S2B~~; data from Stroud et al. 2013). This indicates that CHG DMRs occur in distinct genomic regions and are largely independent of CG methylation (Figure~~.~~ 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 (Figure 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 (Figure 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 (Figure 1-figures supplement 1B,C). Therefore, although cell division rates are correlated with symmetric DNA methylation dynamics, distinct mechanisms reconfigure CG or CHG methylation genome-wide during embryogenesis.

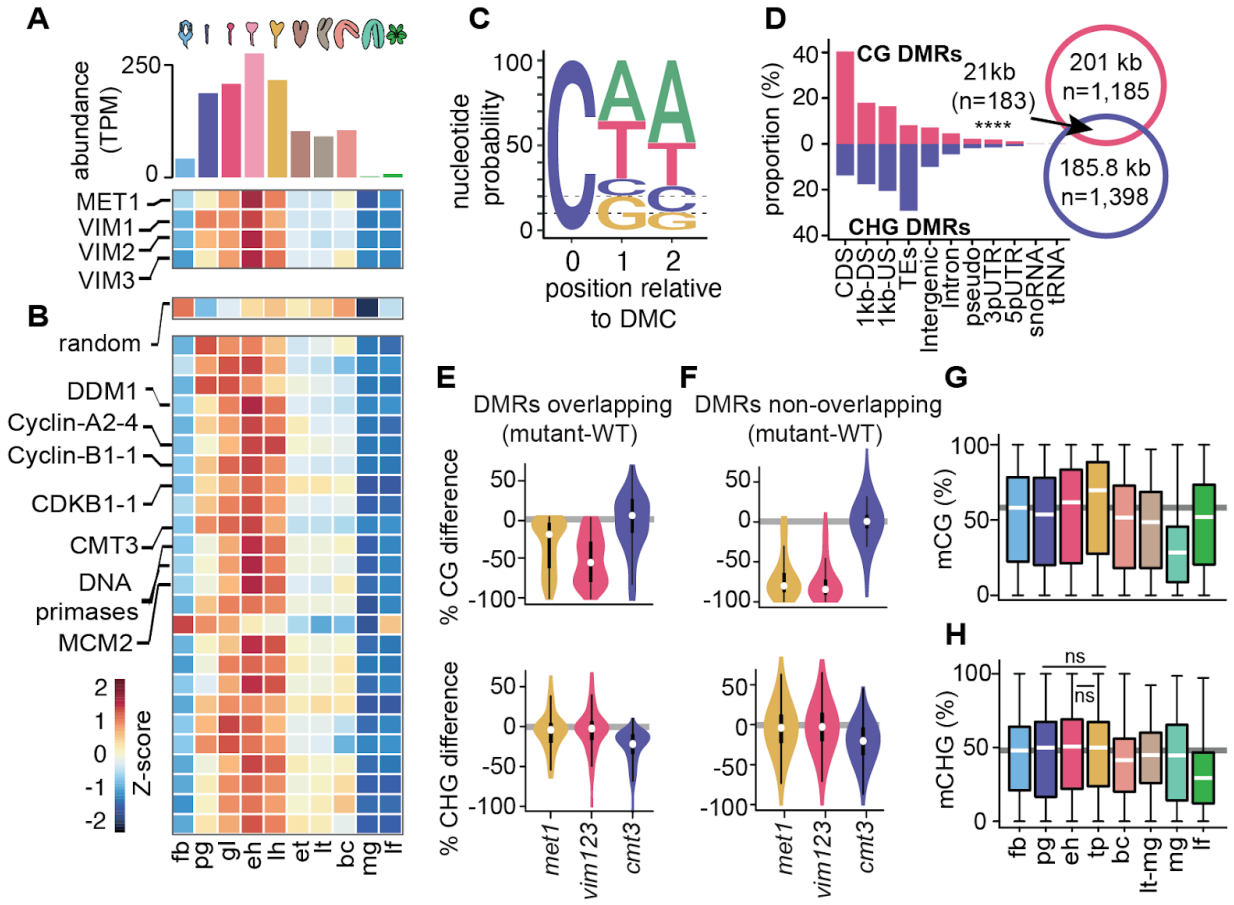

Figure 1. Cell division is linked with CG and CHG methylation through distinct mechanisms.

A) Bar chart depicting total abundance (*top*) and heat map of individual relative transcript levels (*bottom*) of genes involved in CG methylation in three biological replicates each of flowers, embryos and leaves (Hofmann et al. 2019). 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 WT~~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) and **G)** CHG DMRs (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 ((Bouyer et al., 2017; Hsieh et al., 2009; Papareddy et al., 2020; Pignatta et al., 2015). 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.

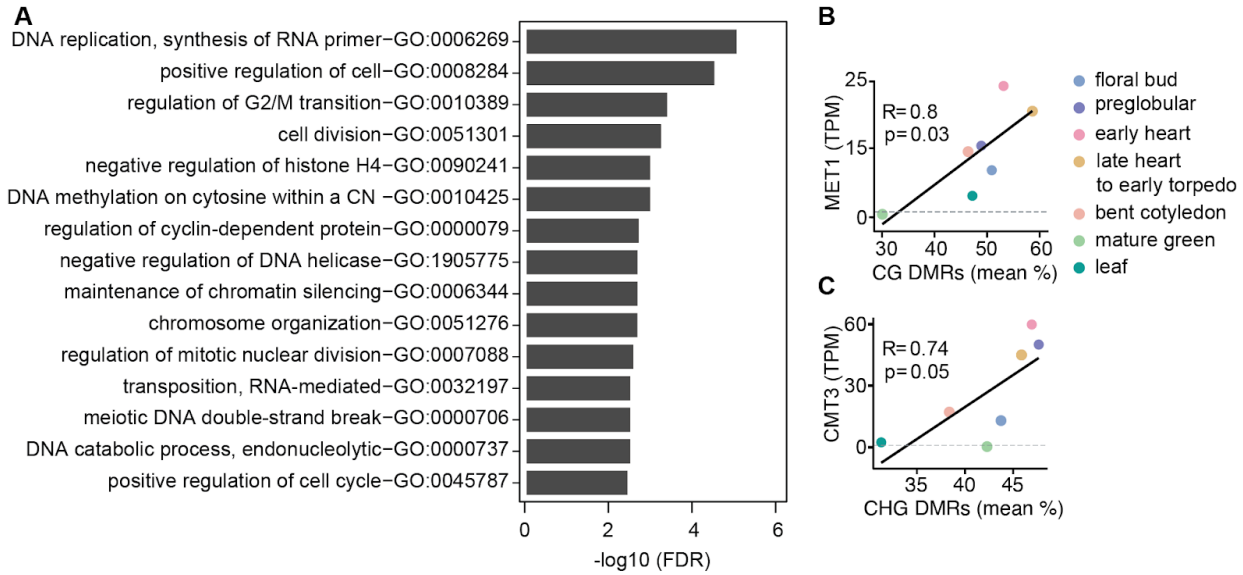

Figure 1-figure supplement ~~S~~1. Characteristics of genes and differentially methylated regions co-expressed with symmetric methyltransferases ~~(Related to Fig. 1)~~. **A)** Gene ontology enrichment of top-50 genes co-expressed with MET1 and VIM1/2/3 with false discovery rates < 0.05. **B** and **C**) 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.

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 (Figure 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 (Figure 2C). CG methylation was required for the production of 24-nt siRNAs from euchromatic TEs, but only marginally for heterochromatic TEs (Figure 2-figure supplement 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 (Figure 2-figure supplement 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 (Figure 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 (Figure 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 (Figure 2H). Therefore, CMT3-dependent CHG and CMT2-dependent CHH methylation of heterochromatic TEs are respectively positively and negatively correlated with cell division rates.

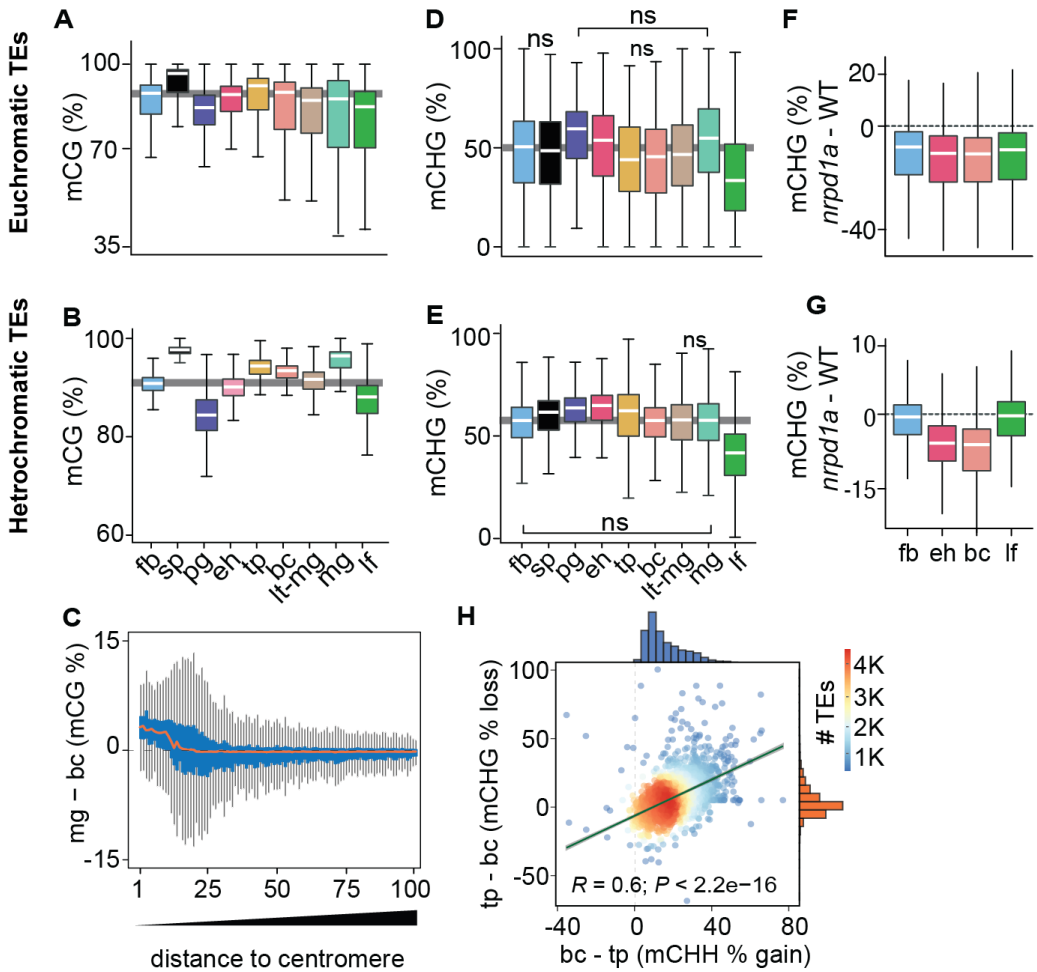

Figure 2. Genome-wide coordination of symmetric DNA methylation.

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 WT~~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).

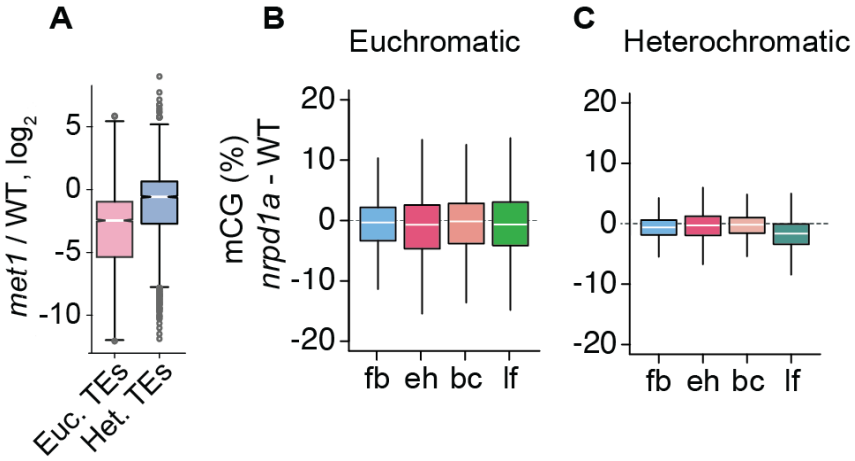

Figure ~~S~~2-figure supplement 1. Relationships between MET1 and 24-nt siRNAs ~~(Related to Fig. 2)~~. A) Boxplot illustrating relative levels of 24-nt siRNAs in *met1* relative to ~~wild-type (~~WT~~)~~ (Lister et al., 2008); Euchromatic and heterochromatic TEs are abbreviated as Euc. TEs and Het. 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.

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 (Figure~~s~~. 1A,H, Figure 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 (Figure~~s~~ 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 (Figure~~.~~ 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 (Figure 3-figure supplement 1~~. S3~~A) and examined CMT3 transcript and CHG methylation levels. Both independently generated *mir823-1* and *mir823-2* mutants were confirmed as nulls (Figure 3-figure supplement 1B~~. S3B~~) and had significantly increased CMT3 levels relative to wild type (WT) in embryos at the bent cotyledon stage when~~embryos at which stage~~ CMT3 levels are normally reduced (Figure~~.~~ 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 (Figure 3-figure supplement 1~~. S3~~D). 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 WT~~wild type~~ (Figure~~s.~~ 3D, Figure 3-supplemental figure 1~~S3~~E).

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 (Figure~~.~~ 3-figure supplement 1C~~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 lines at the bent cotyledon stage (Figure~~.~~ 3C), but not in leaves or floral buds (Figure 3-figure supplement 1~~. S3~~D), 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 (Figure~~s.~~ 3C, Figure 3-figure supplement 1~~S3~~D) 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 (Figure 3-figure supplement 1~~. S~~G). 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 WT~~wild type~~ (Figure 4-figure supplement 1~~. S4~~G,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 (Figure~~.~~ 3D) but did not globally influence CG or CHH methylation (Figure~~.~~-figure supplement 1~~ S3~~E,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 (Figure~~.~~ 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. We chose to focus on rCMT3 plants because of the large amount of hypermethylation observed in these lines during embryogenesis, and used cCMT3 plants as controls. Although TEs had increased CHG methylation levels in both cCMT3 lines relative to WT~~wild type~~, protein-coding genes were not affected (Figure~~.~~ 3F,G). In stark contrast, TEs and genes were hypermethylated in both rCMT3 lines compared to cCMT3 or WT~~wild type~~ plants, and only slightly reduced relative to the levels observed in rCMT3 bent cotyledon embryos (Figure~~.~~ 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.

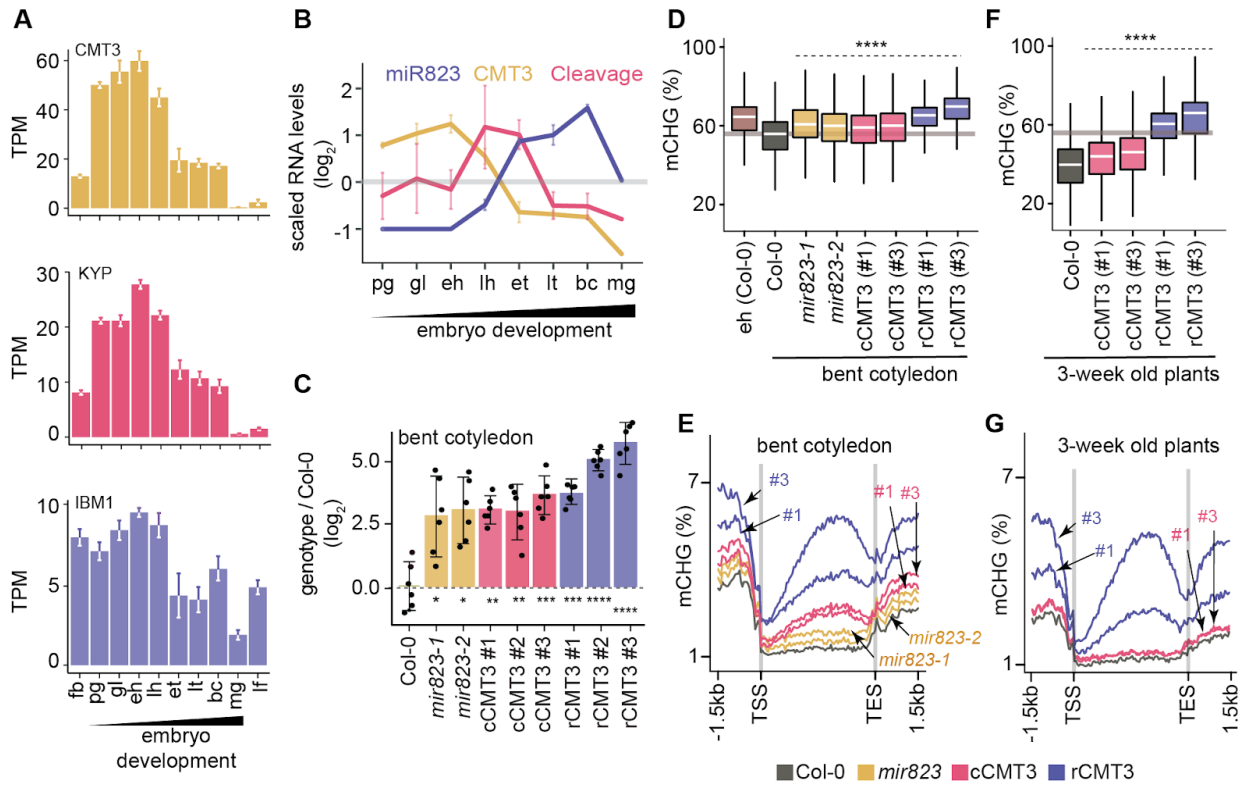

Figure 3. Repression of CMT3 during embryogenesis regulates methylome dynamics. A) Barplots illustrating transcript levels of CMT3 (*top*), KYP (*middle*) and IBM1 (*bottom*) in flowers, embryos and leaves. fb, floral buds; pg, preglobular; gl, globular; eh, early heart; lh, late heart; et, early torpedo; lt, late torpedo; bc, bent cotyledon; mg, mature green; lf, leaf. ~~Error bars represent standard errors of the means of three biological replicates.~~ **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 WT~~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 and~~.~~ b~~B~~ars represent mean values. E~~and~~ error bars in A-C represent ~~indicate~~ standard errors of the means of three biological replicates. Asterisks indicate whether the transcript levels observed in *mir823* mutant, cCMT3 and rCMT3 embryos were significantly different compared to WT~~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 WT~~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.

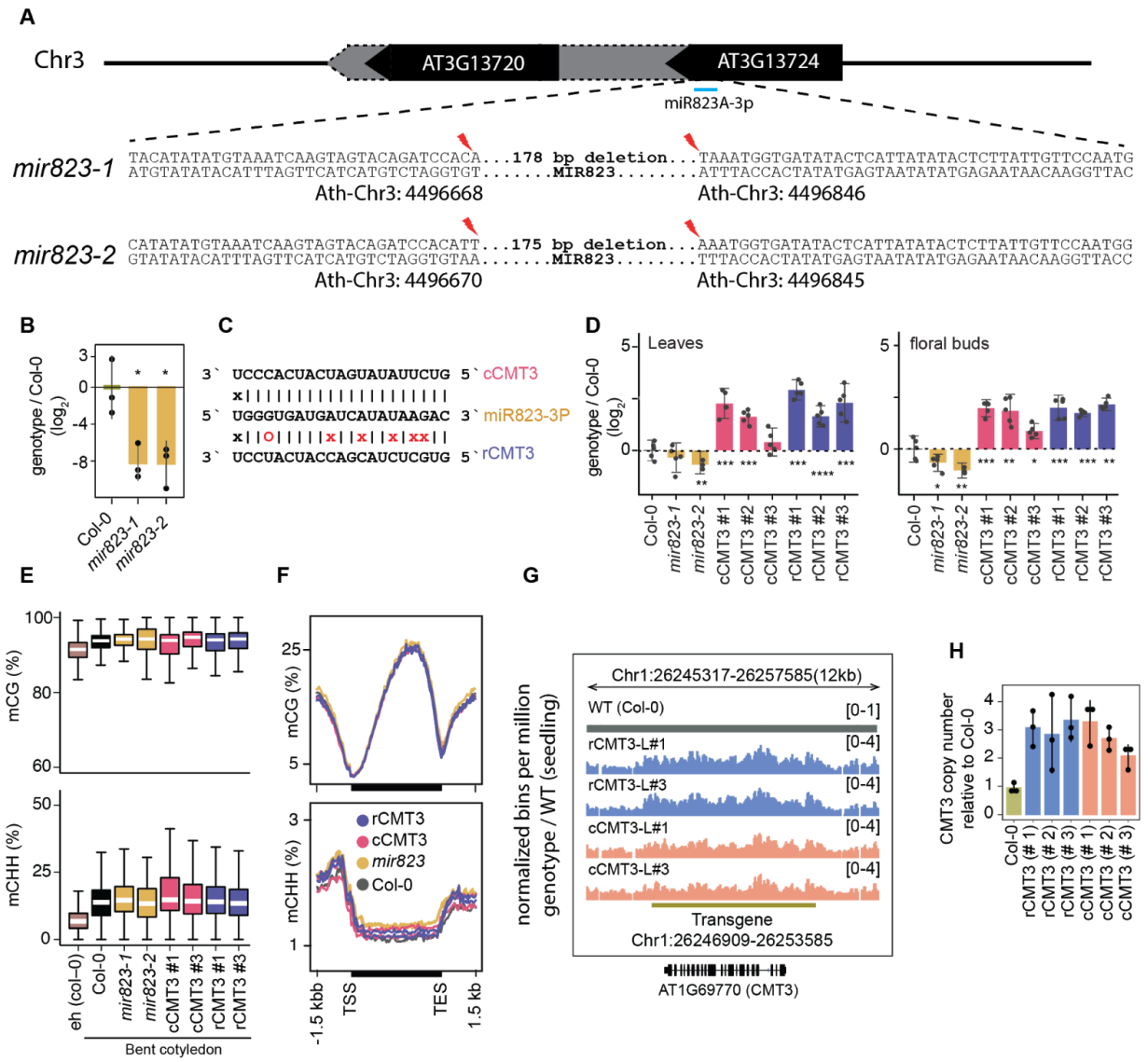

Figure 3-figure supplement 1~~Figure S3~~. *mir823* mutants and effects of miR823-directed repression of CMT3~~(Related to Fig. 3)~~. **A**) Diagram of sequences deleted by CRISPR/Cas9 approach from MIR823A locus in *miR823-1* and *miR823-2*. **B**) Relative levels of miR823 in WT~~wild-type~~ (Col-0), *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 WT~~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 WT~~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 WT~~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 WT~~wild-type~~ (see Methods). **H**) CMT3 copy number quantification based on qPCR in three-week old leaves from WT~~wild-type~~ (Col-0), or independently generated rCMT3 or cCMT3 transgenics in the *cmt3-11* background.

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 ≥5 methylC-seq reads overlapping CHG sites in rCMT3 line #3 and that were expressed (i.e. ≥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 focussed 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 WT~~wild-type~~ bent cotyledon embryos to partition this set of genes into four clusters (Figure 4-figure supplement 1~~-S4~~A). 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 with 3' biases (cluster 4) in rCMT3 compared to WT~~wild-type~~ embryos (Figure ~~s.~~ 4A,B, Figure 4-figure supplement 1~~S4~~B,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 (Figure 4-figure supplement 1~~-S4~~B,C).

TE-like methylated (teM) genes generally have non-CG methylation on their gene bodies without strong 3' biases (Kawakatsu et al. 2016; Bewick et al. 2016). To check whether rCMT3 induced genic CHG methylation is affected by teMs, we intersected our gene clusters with previously defined teMs (Bewick et al. 2016) and found that only 272 of 22,637 (0.012%) expressed genes overlapped teMs (Figure 4-figure supplement 1D). Hypermethylated clusters (cluster 3 and 4) contained more teMs compared to unmethylated or lowly methylated gene

clusters (Figure 4-figure supplement 1E). However, rCMT3 embryos still had genic CHG hypermethylation and 3' biases after excluding teM genes; whereas, WT embryos remained devoid of CHG methylation (Figure 4-figure supplement 1F,G). Therefore, our analysis is not confounded by either TEs or teM genes. 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 (Figure 4-figure supplement 1H) (Wendte et al., 2019).

Consistent with methyltransferases preferring nucleosome-rich DNA as substrates (Chodavarapu et al. 2010; Du et al. 2012), CMT3-induced hypermethylation was proportional to patterns of nucleosome occupancy and biased towards the 3' ends of gene bodies, which was highly similar to CG methylation (Figure ~~s.~~ 4C, Figure 4-figure supplement 1~~S4I~~E). 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 (Figure 4-figure supplement 1~~. S4J~~F). Because nucleosome occupancy was not as readily distinguishable between clusters of affected genes (i.e. clusters 2-4) (Figure 4-figure supplement 1~~. S4~~I,JE,~~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 (Figure ~~s.~~ 4E,F, Figure 4-figure supplement 1~~S4~~K~~G~~) (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 (Figure~~.~~ 4G). It was inversely related to H2A.Z (Figure~~.~~ 4H) and marks indicative of active transcription including~~,~~ H3K4me3 and H3K9ac (Figure~~s.~~ 4I, Figure 4-figure supplement 1~~S4~~L~~.H~~). Further suggesting that deregulated CMT3 prefers features typically associated with inaccessible chromatin, genes with CHG hypermethylation had reduced chromatin accessibility (Figure 4-figure supplement 1~~. S4M.~~I) and were generally closer to heterochromatic centromeres (Figure~~.~~ 4-figure supplement 1~~S4N.~~J). Moreover, the most CMT3-induced hypermethylated genes (i.e. cluster 4) were also substantially hypermethylated in *ddm1* mutants (Figure~~.~~ 4J) that have increased heterochromatic accessibility (Figure~~.~~ 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 (Figure~~s~~ 1E,F) and the gain of mCHG being associated with proportional loss of mCG over genes in *ddm1* mutants (Figure 4-figure supplement~~s~~ 1~~S4O~~K) (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.

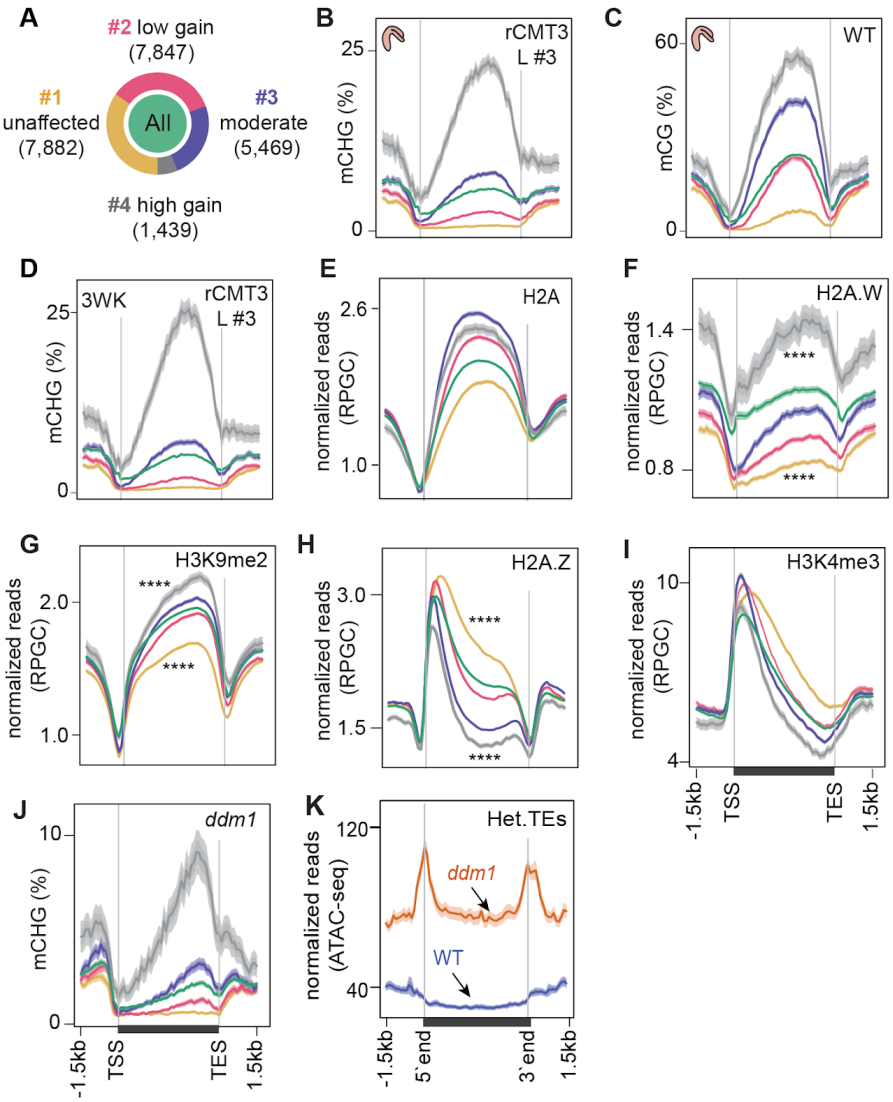

Figure 4. Chromatin features associated with CMT3-induced gene methylation. A) Proportion of genes in each cluster partitioned using k-means clustering algorithm based on differences in mCHG between rCMT3 (line #3) and WT~~wild type~~ embryos. Unaffected genes (*yellow*), low mCHG gain genes (*red*), 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

WT~~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.

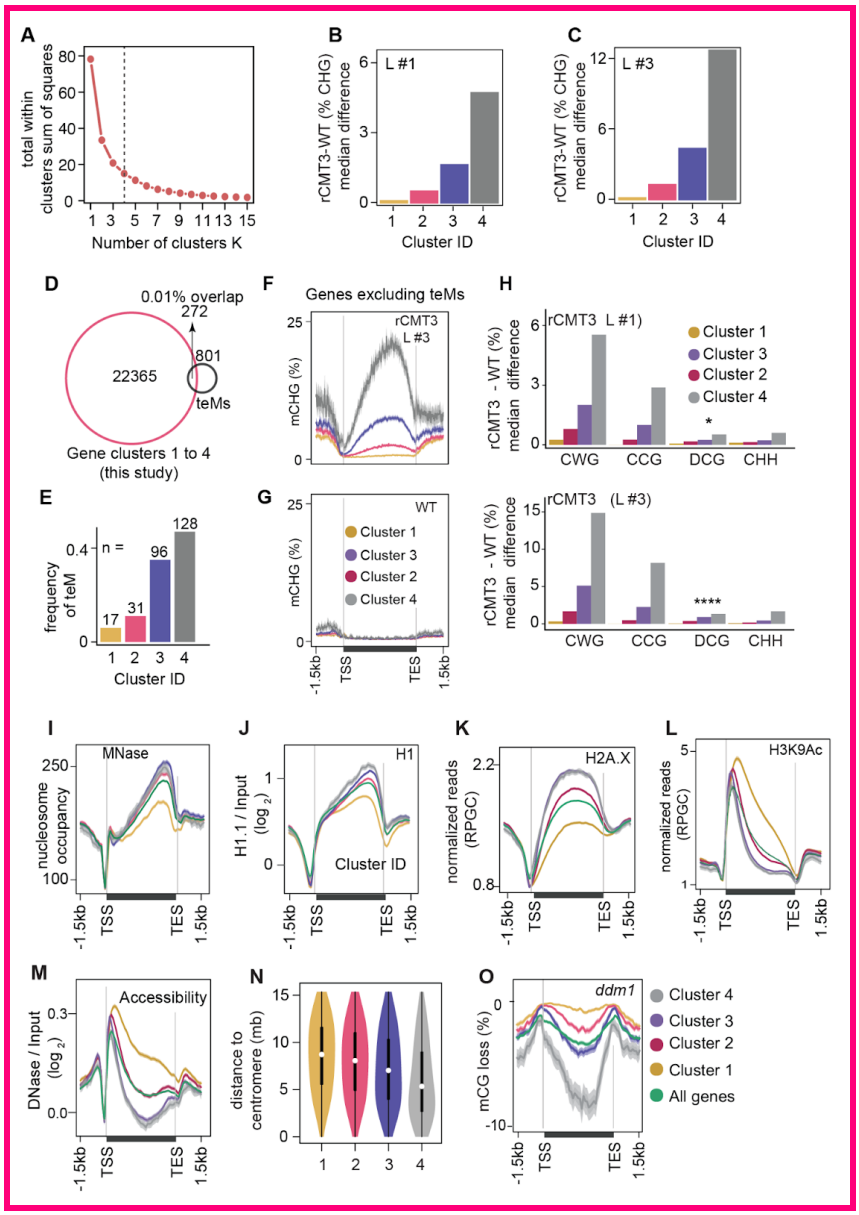

Figure 4-figure supplement 1~~S4~~. Partitioning of CMT3-induced hypermethylated genes and associated chromatin features~~ (Related to Fig. 4)~~. **A)** Determining the optimal K-value based difference in genic mCHG between rCMT3 (Line #3) and wild-type bent cotyledon embryos using the elbow method. 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**) Overlap between expressed genes used for K-means clustering in main figure 4A and all genes classified as teM in Bewick et al 2016. **E**) Number of total intersected teM in each gene cluster. **F-G**) Metaplot showing average weighted methylation rate in rCMT3 Line #3 (**F**) and ~~wild-type (~~WT~~)~~ (**G**) after excluding 272 teM intersecting genes. **H~~D~~**) Barplots illustrating median gain of methylation in rCMT3 (Line #1) (*top*) and rCMT3 (Line #3) (*bottom*) 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. **I-M~~E-I~~**) Metaplots showing nucleosome occupancy obtained from MNase-seq data (Rutowicz et al., 2019) (**I~~E~~**), enrichment of linker histone 1 (H1) on gene clusters (Choi et al.) (**J~~F~~**), normalized reads per genomic content (RPGC) average value of histone variant H2A.X (Yelagandula et al., 2014) (**K~~G~~**), H3K9Ac (Wang et al., 2019) (**L~~H~~**) and DNase-Seq signal on gene clusters representing accessibility (Choi et al.) (**M~~I~~**). **N~~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. **O~~K~~**) Metaplot of CG methylation in differences in *ddm1* compared to ~~wild-type (~~WT~~)~~ (Stroud et al., 2013).

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 WT~~wild-type~~ and rCMT3 (line #s 1 and 3) bent cotyledon embryos. Principal component analysis revealed that WT~~wild-type~~ and rCMT3 biological replicates clustered according to genotype and in similar positions along the dominant principal component axis corresponding to developmental time (Figure~~s~~ 5A). This indicates that our mRNA-seq datasets captured gene expression variation inherent to WT~~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 (Figure~~s~~ 5B, Figure 5-figure supplement 1~~S5~~A). We then identified 916 genes that were differentially expressed between rCMT3 and WT embryos (i.e. ≥2-fold differences and adj. *P* values ≤ 0.01; see Methods) (Figure 5-figure supplement 1~~ S5~~B,C and Supplementary File ~~Table S~~3). Differentially expressed genes (DEGs), defined by comparing either rCMT3 line #1 or rCMT3 line #3 with WT~~wild-type~~, were commonly detected in both independently generated lines with 87.5% of genes overlapping (Figure 5-figure supplement~~ 1S5~~C). 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 were~~as~~ not directly due to their hypermethylation (Figure 5-figure supplement 1~~. S5~~D). 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 (Supplementary File ~~Table S~~4; 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 = 1.29e-05) (Figure 5-figure supplement 1~~. S5~~E). Consistent with CMT3-induced hypermethylation repressing their expression, the DMRs overlapping these 21 down-regulated genes were significantly CHG hypermethylated compared to genomic bins (Figure~~.~~ 5C). Moreover, the stronger-expressing rCMT3 line #3 had significantly higher CHG methylation compared to rCMT3 line #1 (Figure~~.~~ 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 (Figure 5-figure supplement 1~~. S5~~F). Strikingly, transcripts corresponding to these 21 CMT3-induced hypermethylated and down-regulated genes were rapidly increasing when embryos were transitioning to the maturation phase (Figure~~.~~ 5D). Moreover, 10 of these 21 genes (Fisher's exact test, *P* value = 1.49e-13), were among a group of 381 genes previously identified to also be rapidly activated at these time points (Figure 5-figure supplement 1~~. S5~~G) (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 WT~~wild type~~. Altogether, our expression and methylation analyses suggest that when CMT3 is not properly repressed it can induce ectopic hypermethylation of genes. Furthermore, we suggest that CMT3-induced hypermethylation of gene promoters or bodies can reduce the steady state levels of transcripts from genes that are in the process of switching from silent to active transcriptional states. However, additional experiments are required to directly test whether gene-body CHG methylation can repress gene expression.

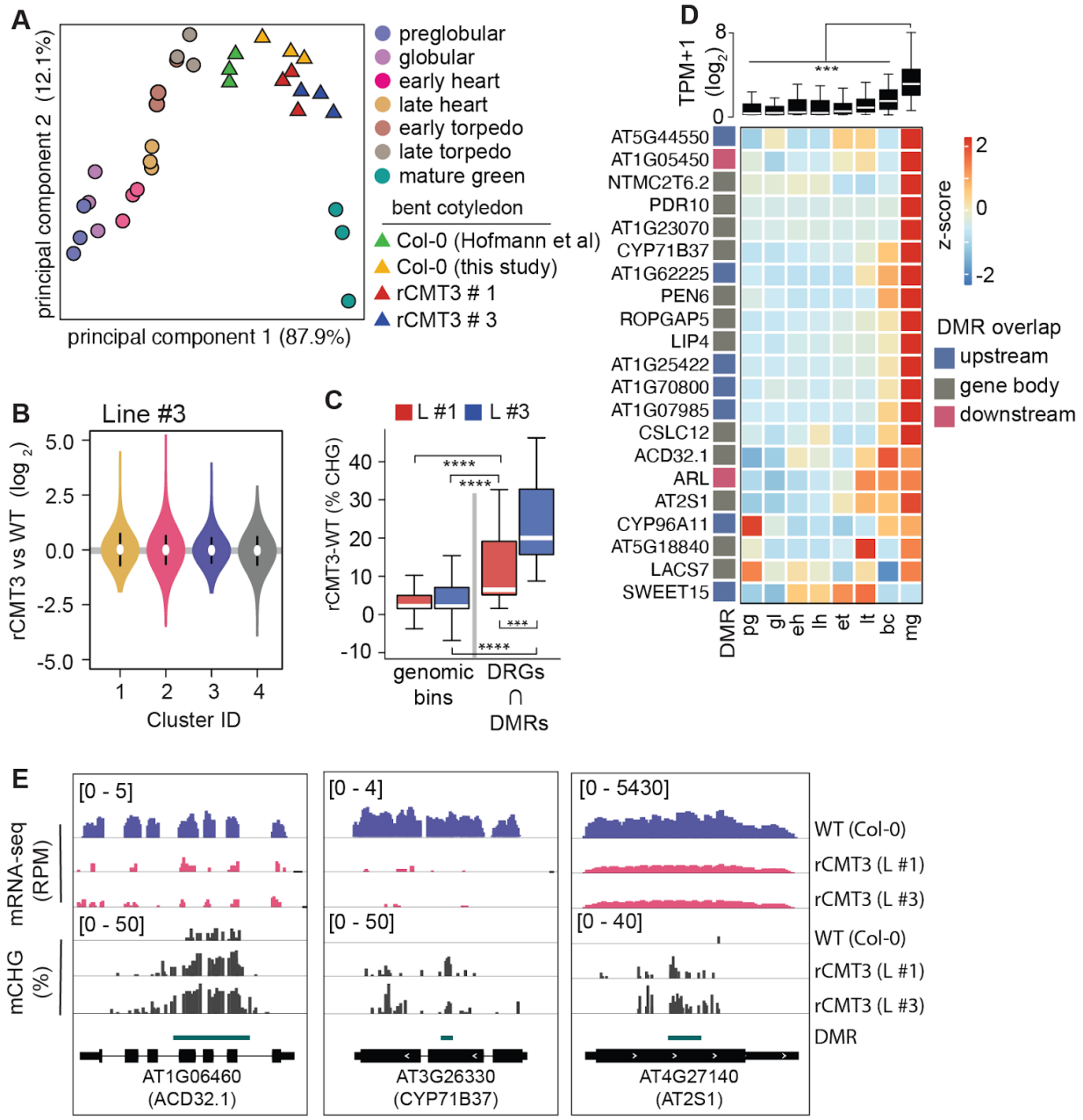

Figure 5. Impact of CMT3-induced hypermethylation on gene expression. A) Principal component analysis of mRNA-seq from three biological replicates of rCMT3 and WT~~wild-type~~ (~~WT,~~ Col-0) bent cotyledon embryos generated in this study along with floral buds, embryos, leaves, from (Hofmann et al., 2019) and color-coded according to the key. **B**) Violin plot showing transcript fold changes in rCMT3 (line #3) compared to WT~~wild-type~~ (~~WT,~~ Col-0) bent cotyledon embryos per cluster as defined in Figure~~:~~ 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.

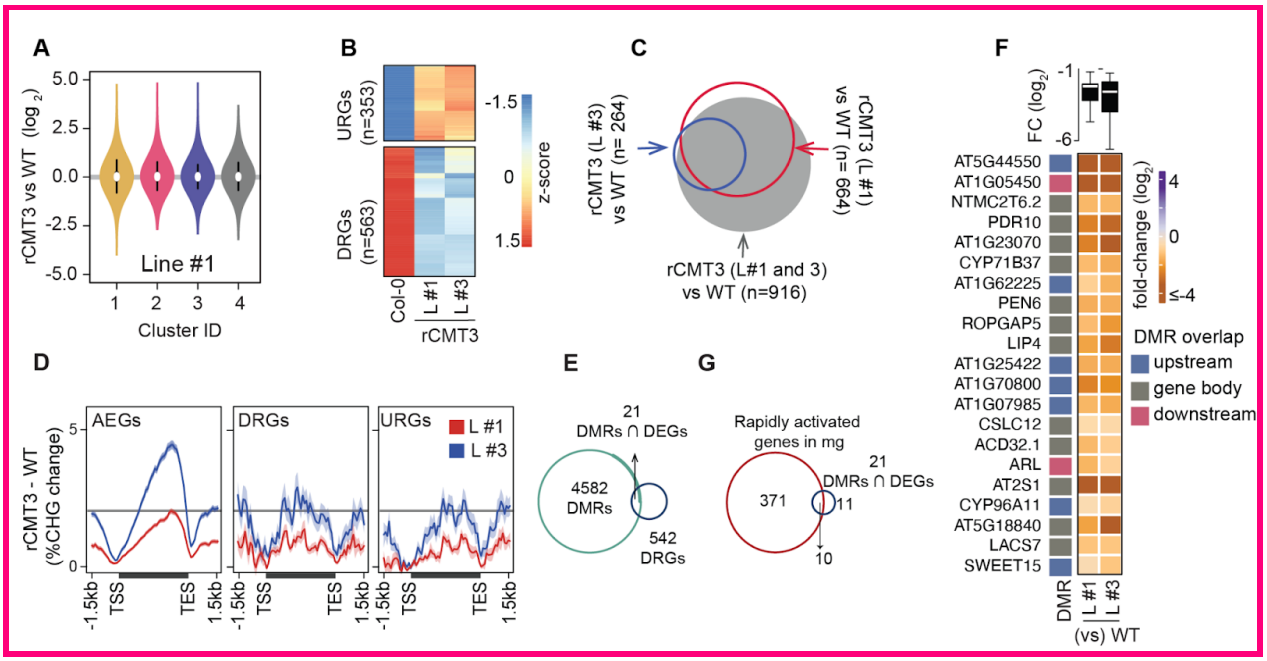

Figure 5-figure supplement 1~~S5~~. Additional information regarding the influence of CMT3-induced hypermethylation on gene expression~~ (Related to Fig. 5)~~. **A**) Violin plot showing expression fold change in rCMT3 (line #3) compared to WT~~wild type~~ (Col-0) bent cotyledon embryos per cluster as defined in Figure~~s~~ 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 WT~~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 WT~~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.

Discussion~~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 (Figure~~.~~ 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 (Figure~~.~~ 3). CMT3-induced hypermethylation of genes was highly associated with features conferring nucleosome stability (Figure~~.~~ 4) and resulted in the repression of genes that are transcriptionally activated (Figure~~.~~ 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 appears to~~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 (Figures~~.~~ 1, Figure 2, Figure 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 (Figure~~.~~ 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 development 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 (Figure~~s~~ 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 (Figure~~s~~ 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 (Figure~~s~~ 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 (Figure 4-figure supplement ~~s~~4O~~K~~) (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 (Figure~~s~~ 4). Notably, heterochromatin becomes destabilized in *ddm1* mutants (Figure~~s~~ 4K) (Mathieu et al., 2003; Soppe et al., 2002) and CMT3 prefers features associated with stable (Figure~~s~~ 4E,F, Figure 4-figure supplement 1~~S4~~K~~G~~) (Bourguet et al., 2021; Osakabe et al., 2018; Yelagandula et al., 2014) over unstable nucleosomes such as H2A.Z with active marks (Figure~~s~~ 4H,I, Figure 4-figure supplement 1~~S4~~L~~H~~). 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

596 required to regulate CMT3 activity in order to achieve proper epigenome homeostasis (Figure
597 6).

598 CMT3-induced CHG hypermethylation of genes did not globally affect steady state transcript
599 levels (Figure 5). However, we observed exceptional association between CHG
600 hypermethylation and repression of genes that switch from transcriptionally inactive to active
601 states. Because IBM1-mediated removal of H3K9me2 marks is coupled to transcription (Inagaki
602 et al., 2017), it is possible that CMT3-induced methylation can form feedback loops with
603 H3K9me2 methyltransferases when genes are transcriptionally inert. However, when genes are
604 switched on, H3K9me2 could repress initial rounds of transcription before it is removed by
605 IBM1. Accordingly, it may be difficult to detect the effects of ectopic CHG methylation on gene
606 expression when quantifying transcripts at steady state with standard mRNA-seq. It is possible
607 that we observed a repressive effect of CHG hypermethylation on a subset of genes because
608 we profiled a developmental stage in which hundreds of genes become transcriptionally
609 activated at the onset of embryo maturation. Nevertheless, we cannot completely exclude that
610 the repression of hypermethylated genes undergoing transcriptional activation is due to
611 secondary effects of other genes influenced by CMT3-induced hypermethylation. Importantly,
612 CMT3-induced CHG hypermethylation due at least partially to loss of miR823 repression in
613 embryos is largely maintained for weeks after detectable miRNA activity (Figure 3). Therefore,
614 epigenetic collateral damage occurring in embryos may also negatively impact gene expression
615 later in life. However, future experiments are required to directly test the relationship between
616 CMT3-induced hypermethylation and gene expression.

617 Transcriptional (Ning et al., 2020), post-transcriptional (Figure 3), post-translational (Deng
618 et al., 2016), post-hoc (Saze et al., 2008) and perhaps substrate-related (Figure 4) mechanisms
619 fine-tune CMT3 activities to levels required to specifically silence mutagenic TEs but not genes.
620 However, errors in restricting CMT3 to heterochromatin are inevitable on an evolutionary
621 timescale (Zhang et al., 2020) and recent studies indicate that CMT3-induced methylation of
622 genes precedes gbM (Wendte et al., 2019). Because gbM can be stably maintained over many
623 generations by MET1 and its functional significance is debatable (Bewick et al., 2016, 2019;
624 Choi et al., 2019; Coleman-Derr and Zilberman, 2012; Le et al., 2020; Picard and Gehring,
625 2017; Shahzad et al., 2021; Takuno and Gaut, 2013; Wendte et al., 2019; Williams et al., 2021;
626 Zilberman, 2017), it cannot be excluded that gbM is merely an evolutionary record of epigenetic
627 collateral damage events that occurred in the past (Bewick and Schmitz, 2017; Bewick et al.,
628 2017). Our results suggest that derepressed CMT3 and MET1 both prefer genic regions
629 characterized by increased nucleosome stability (Figure 4). Accordingly, CMT3-induced CHG

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

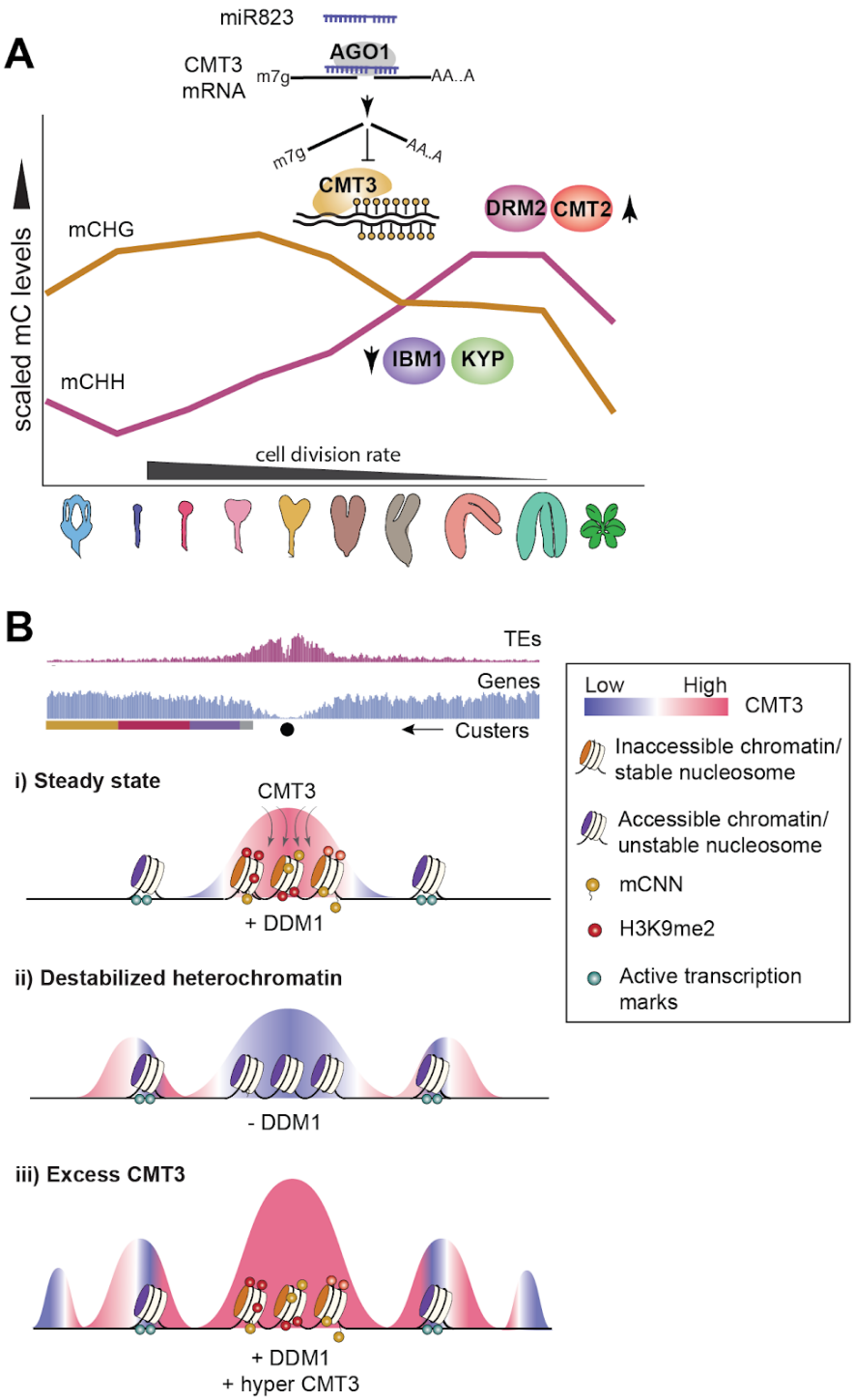

Figure 6. Models for CMT3 regulation during periods of fluctuating cell division rates and destabilized chromatin. A) Model of non-CG methylation dynamics during embryo development and corresponding regulatory mechanisms. **B**) Model for how CMT3 equilibrium is maintained to restrict its activity to 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 Figure 4A. Black dot represents the centromere. (i) In steady state,

649 stable nucleosomes along with H3K9me2 and DNA methylation provides positive reinforcement to sequester CMT3
650 to constitutive heterochromatin. (ii) Loss of DDM1 results in destabilized and accessible heterochromatin (Zhong et
651 al., 2021), characterized by loss of H3K9me2 and stable nucleosomes (Osakabe et al., 2021). Accessible chromatin
652 or DNA without stable nucleosomes is no longer a preferable substrate for CMT3 and results in CHG
653 hypomethylation of TEs. CMT3 will now be readily available and redirected to genic regions where it induces
654 ectopic CHG methylation in proportion to the levels of stable nucleosomes and chromatin marks. (iii) Excess levels
655 of CMT3 causes genome-wide CHG hypermethylation with a preference for stable nucleosomes associated with
656 repressive marks that tend to be in regions closer to centromeres compared to chromosomal arms.
657

Materials and methods~~ATERIALS AND METHODS~~

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
gene (*Arabidopsis thaliana*)	CHROMOMETHYLASE 3 (CMT3)	TAIR	AT1G69770	
gene (*Arabidopsis thaliana*)	MICRORNA 823A (MIR823A)	TAIR	AT3G13724	
genetic reagent (*Arabidopsis thaliana*)	miR823-cleavable CMT3 (cCMT3)	this paper		pAlligatorR43/promoterCMT3::genomic CMT3
genetic reagent (*Arabidopsis thaliana*)	miR823-resistant CMT3 (rCMT3)	this paper		pAlligatorR43/promoterCMT3::resistant CMT3 (generated from cCMT3 with site-directed mutagenesis)
recombinant DNA reagent	pAlligatorR43 (plasmid)	DOI: 10.7554/eLife.04501		mCherry selection marker
recombinant DNA reagent	pHSE401 (plasmid)	Addgene	#62201	CRISPR/Cas9 plasmid

recombinant DNA reagent	pCBCD-T1T2 (plasmid)	Addgene	#50590	CRISPR/Cas9 plasmid
strain (*Arabidopsis thaliana*)	*cmt3-11T*	NASC	SALK_148381	T-DNA insertion mutant of *CMT3*
strain (*Arabidopsis thaliana*)	*mir823-1*	this paper		miR823 knockout mutant
strain (*Arabidopsis thaliana*)	*mir823-2*	this paper		miR823 knockout mutant
commercial kit	Q5 Site-Directed Mutagenesis Kit	New England Biolabs	#E0554S	
commercial kit	Fast SYBR Green Master Mix	Roche	#06402712001	
commercial kit	SuperScript III Reverse Transcriptase	Thermo Fisher Scientific	#18080093	
commercial kit	TRIzol	Invitrogen	#15596026	
Software	Lightcycler 96®	Roche Diagnostics	Version 1.1.0.1320	

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 Figure 3-figure supplement 1~~. S3~~C. Both cCMT3 and rCMT3 construct sequences were analyzed for mutations using Sanger sequencing. All primers used are listed in the Supplementary File ~~Table S~~5. 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 (Figure 3-figure supplement 1~~. S3~~A and Supplementary File ~~Table S~~5) 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-positive seeds were selected based on the presence of seed coat RFP signal. Deletion lines were identified with PCR using primers flanking gRNA-targeted sites (Figure 3-figure supplement 1~~. S3~~A and Supplementary File ~~Table S~~5). 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, corresponding 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 Supplementary File ~~Table S~~5.

Sample size estimation, e~~E~~mbryo isolation and nucleic acid extraction

Sample sizes were determined based on a combination of the required statistical power, ability to acquire samples and cost of the experiments. 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, two biological replicates were pooled and differentially methylated cytosines (DMCs) were identified by root mean squared tests with false discovery rates ≤ 0.01. DMRs

were defined by collapsing DMCs with ≥4 reads within 500 bps to single units requiring ≥8 and ≥4 DMCs for CG and CHN sites, respectively (N = A,T,C,G; H ≠ 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 1.5 kb 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 (Supplementary File ~~Table S~~6). Prior to PCA (Figure~~.~~ 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 ≥2-fold differences and adjusted p-value ≤ 0.01 were classified as differentially expressed genes (DEGs). Nearest-neighbour genes in Figure~~.~~ 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 WT~~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 WT~~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 WT~~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; G.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~~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.

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

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

876 Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method forAgrobacterium-mediated
877 transformation ofArabidopsis thaliana: Floral dip transformation of Arabidopsis. Plant J. *16*,
878 735–743.

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

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

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

886 Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., and Notredame, C.
887 (2017). Nextflow enables reproducible computational workflows. Nat. Biotechnol. *35*, 316–319.

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

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

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

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

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

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

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

907 Finnegan, E.J., and Dennis, E.S. (1993). Isolation and identification by sequence homology of a
908 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 transposon 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

947 non-CG methylation in immobilization of transposons in Arabidopsis. Curr. Biol. *13*, 421–426.

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

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

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

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

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

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

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

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

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

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

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

976 Lin, J.-Y., Le, B.H., Chen, M., Henry, K.F., Hur, J., Hsieh, T.-F., Chen, P.-Y., Pelletier, J.M.,
977 Pellegrini, M., Fischer, R.L., et al. (2017). Similarity between soybean and Arabidopsis seed
978 methylomes and loss of non-CG methylation does not affect seed development. Proc. Natl.
979 Acad. Sci. U. S. A. *114*, E9730–E9739.

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

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

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

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

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

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

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

998 Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C., and Paszkowski, J. (2007).
999 Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. Cell
1000 *130*, 851–862.

1001 Miura, A., Nakamura, M., Inagaki, S., Kobayashi, A., Saze, H., and Kakutani, T. (2009). An
1002 Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG
1003 sites. EMBO J. *28*, 1078–1086.

1004 Muyle, A., and Gaut, B.S. (2019). Loss of Gene Body Methylation in Eutrema salsugineum Is
1005 Associated with Reduced Gene Expression. Mol. Biol. Evol. *36*, 155–158.

1006 Narsai, R., Gouil, Q., Secco, D., Srivastava, A., Karpievitch, Y.V., Liew, L.C., Lister, R., Lewsey,
1007 M.G., and Whelan, J. (2017). Extensive transcriptomic and epigenomic remodelling occurs
1008 during Arabidopsis thaliana germination. Genome Biol. *18*, 172.

1009 Niederhuth, C.E., Bewick, A.J., Ji, L., Alabady, M.S., Kim, K.D., Li, Q., Rohr, N.A., Rambani, A.,
1010 Burke, J.M., Udall, J.A., et al. (2016). Widespread natural variation of DNA methylation within
1011 angiosperms. Genome Biol. *17*, 194.

1012 Ning, Y.-Q., Liu, N., Lan, K.-K., Su, Y.-N., Li, L., Chen, S., and He, X.-J. (2020). DREAM
1013 complex suppresses DNA methylation maintenance genes and precludes DNA
1014 hypermethylation. Nat Plants *6*, 942–956.

1015 Osakabe, A., Lorkovic, Z.J., Kobayashi, W., Tachiwana, H., Yelagandula, R., Kurumizaka, H.,
1016 and Berger, F. (2018). Histone H2A variants confer specific properties to nucleosomes and
1017 impact on chromatin accessibility. Nucleic Acids Res. *46*, 7675–7685.

1018 Osakabe, A., Jamge, B., Axelsson, E., Montgomery, S.A., Akimcheva, S., Kuehn, A.L., Pisupati,
1019 R., Lorković, Z.J., Yelagandula, R., Kakutani, T., et al. (2021). The chromatin remodeler DDM1
1020 prevents transposon mobility through deposition of histone variant H2A.W. Nat. Cell Biol. *23*,
1021 391–400.

1022 Papareddy, R.K., and Nodine, M.D. (2021). Plant Epigenetics: Propelling DNA Methylation
1023 Variation across the Cell Cycle. Curr. Biol. *31*, R129–R131.

1024 Papareddy, R.K., Páldi, K., Paulraj, S., Kao, P., Lutzmayer, S., and Nodine, M.D. (2020).

1025 Chromatin regulates expression of small RNAs to help maintain transposon methylome
1026 homeostasis in Arabidopsis. Genome Biol. *21*, 251.

1027 Picard, C.L., and Gehring, M. (2017). Proximal methylation features associated with nonrandom
1028 changes in gene body methylation. Genome Biol. *18*, 73.

1029 Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014).
1030 Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

1031 Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W., and Gehring, M. (2015).
1032 Correction: Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene
1033 imprinting. Elife *4*.

1034 Plotnikova, A., Kellner, M.J., Schon, M.A., Mosiolek, M., and Nodine, M.D. (2019). MicroRNA
1035 Dynamics and Functions During Arabidopsis Embryogenesis. Plant Cell *31*, 2929–2946.

1036 Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle.
1037 Nat. Rev. Mol. Cell Biol. *10*, 192–206.

1038 Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic
1039 features. Bioinformatics *26*, 841–842.

1040 Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible
1041 platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

1042 Rutowicz, K., Lirski, M., Mermaz, B., Teano, G., Schubert, J., Mestiri, I., Kroteń, M.A., Fabrice,
1043 T.N., Fritz, S., Grob, S., et al. (2019). Linker histones are fine-scale chromatin architects
1044 modulating developmental decisions in Arabidopsis. Genome Biol. *20*, 157.

1045 Saze, H., and Kakutani, T. (2007). Heritable epigenetic mutation of a transposon−flanked
1046 Arabidopsis gene due to lack of the chromatin−remodeling factor DDM1. EMBO J.

1047 Saze, H., and Kakutani, T. (2011). Differentiation of epigenetic modifications between
1048 transposons and genes. Curr. Opin. Plant Biol. *14*, 81–87.

1049 Saze, H., Mittelsten Scheid, O., and Paszkowski, J. (2003). Maintenance of CpG methylation is
1050 essential for epigenetic inheritance during plant gametogenesis. Nat. Genet. *34*, 65–69.

1051 Saze, H., Shiraishi, A., Miura, A., and Kakutani, T. (2008). Control of genic DNA methylation by
1052 a jmjC domain-containing protein in Arabidopsis thaliana. Science *319*, 462–465.

1053 Scheid, O.M., Afsar, K., and Paszkowski, J. (1998). Release of epigenetic gene silencing by
1054 trans-acting mutations in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. *95*, 632–637.

1055 Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N.,
1056 Nery, J.R., Urich, M.A., Chen, H., et al. (2015). Human body epigenome maps reveal
1057 noncanonical DNA methylation variation. Nature *523*, 212–216.

1058 Shahzad, Z., Moore, J.D., and Zilberman, D. (2021). Gene body methylation mediates
1059 epigenetic inheritance of plant traits. bioRxiv doi: https://doi.org/10.1101/2021.03.15.435374.

1060 Shen, J., Xie, K., and Xiong, L. (2010). Global expression profiling of rice microRNAs by
1061 one-tube stem-loop reverse transcription quantitative PCR revealed important roles of

1062 microRNAs in abiotic stress responses. Molecular Genetics and Genomics *284*, 477–488.

1063 Shen, J.C., Rideout, W.M., 3rd, and Jones, P.A. (1992). High frequency mutagenesis by a DNA
1064 methyltransferase. Cell *71*, 1073–1080.

1065 Shepherd, C.T., Moran Lauter, A.N., and Scott, M.P. (2009). Determination of transgene copy
1066 number by real-time quantitative PCR. Methods Mol. Biol. *526*, 129–134.

1067 Soppe, W.J.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen,
1068 S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9
1069 methylation and heterochromatin assembly in Arabidopsis. EMBO J. *21*, 6549–6559.

1070 Stempor, P., and Ahringer, J. (2016). SeqPlots - Interactive software for exploratory data
1071 analyses, pattern discovery and visualization in genomics. Wellcome Open Res *1*, 14.

1072 Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013).
1073 Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis
1074 methylome. Cell *152*, 352–364.

1075 Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E.
1076 (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat.
1077 Struct. Mol. Biol. *21*, 64–72.

1078 Sved, J., and Bird, A. (1990). The expected equilibrium of the CpG dinucleotide in vertebrate
1079 genomes under a mutation model. Proc. Natl. Acad. Sci. U. S. A. *87*, 4692–4696.

1080 Takuno, S., and Gaut, B.S. (2013). Gene body methylation is conserved between plant
1081 orthologs and is of evolutionary consequence. Proc. Natl. Acad. Sci. U. S. A. *110*, 1797–1802.

1082 Takuno, S., Seymour, D.K., and Gaut, B.S. (2017). The Evolutionary Dynamics of Orthologs
1083 That Shift in Gene Body Methylation between Arabidopsis Species. Mol. Biol. Evol. *34*,
1084 1479–1491.

1085 The Cold Spring Harbor Laboratory, Washington University Genome Sequencing Center, and
1086 PE Biosystems Arabidopsis Sequencing Consortium (2000). The Complete Sequence of a
1087 Heterochromatic Island from a Higher Eukaryote. Cell *100*, 377–386.

1088 Toolkit, P. (2019). Broad institute, GitHub repository. See Http://broadinstitute. Github. Io/picard.

1089 Tran, R.K., Henikoff, J.G., Zilberman, D., Ditt, R.F., Jacobsen, S.E., and Henikoff, S. (2005).
1090 DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes. Curr. Biol.
1091 *15*, 154–159.

1092 Wang, L., Wang, C., Liu, X., Cheng, J., Li, S., Zhu, J.-K., and Gong, Z. (2019). Peroxisomal
1093 β-oxidation regulates histone acetylation and DNA methylation in Arabidopsis. Proc. Natl. Acad.
1094 Sci. U. S. A. *116*, 10576–10585.

1095 Wendte, J.M., Zhang, Y., Ji, L., Shi, X., Hazarika, R.R., Shahryary, Y., Johannes, F., and
1096 Schmitz, R.J. (2019). Epimutations are associated with CHROMOMETHYLASE 3-induced de
1097 novo DNA methylation. Elife *8*.

1098 Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding transcription by RNA
1099 polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes.

1100 Cell *135*, 635–648.

1101 Williams, B.P., Pignatta, D., Henikoff, S., and Gehring, M. (2015). Methylation-sensitive
1102 expression of a DNA demethylase gene serves as an epigenetic rheostat. PLoS Genet. *11*,
1103 e1005142.

1104 Williams, B.P., Bechen, L.A., Pohlmann, D.A., and Gehring, M. (2021). Somatic DNA
1105 demethylation generates tissue-specific methylation states and impacts flowering time. bioRxiv
1106 doi: https://doi.org/10.1101/2021.03.29.437569.

1107 Woo, H.R., Dittmer, T.A., and Richards, E.J. (2008). Three SRA-domain methylcytosine-binding
1108 proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis.
1109 PLoS Genet. *4*, e1000156.

1110 Xing, H.-L., Dong, L., Wang, Z.-P., Zhang, H.-Y., Han, C.-Y., Liu, B., Wang, X.-C., and Chen,
1111 Q.-J. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biol. *14*,
1112 327.

1113 Yang, L.-H., Wang, S.-L., Tang, L.-L., Liu, B., Ye, W.-L., Wang, L.-L., Wang, Z.-Y., Zhou, M.-T.,
1114 and Chen, B.-C. (2014). Universal stem-loop primer method for screening and quantification of
1115 microRNA. PLoS One *9*, e115293.

1116 Yelagandula, R., Stroud, H., Holec, S., Zhou, K., Feng, S., Zhong, X., Muthurajan, U.M., Nie, X.,
1117 Kawashima, T., Groth, M., et al. (2014). The histone variant H2A.W defines heterochromatin
1118 and promotes chromatin condensation in Arabidopsis. Cell *158*, 98–109.

1119 van Zanten, M., Koini, M.A., Geyer, R., Liu, Y., Brambilla, V., Bartels, D., Koornneef, M., Fransz,
1120 P., and Soppe, W.J.J. (2011). Seed maturation in Arabidopsis thaliana is characterized by
1121 nuclear size reduction and increased chromatin condensation. Proc. Natl. Acad. Sci. U. S. A.
1122 *108*, 20219–20224.

1123 Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer,
1124 S.L., and Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA
1125 methyltransferases to access H1-containing heterochromatin. Cell *153*, 193–205.

1126 Zhai, J., Bischof, S., Wang, H., Feng, S., Lee, T.-F., Teng, C., Chen, X., Park, S.Y., Liu, L.,
1127 Gallego-Bartolome, J., et al. (2015). A One Precursor One siRNA Model for Pol IV-Dependent
1128 siRNA Biogenesis. Cell *163*, 445–455.

1129 Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W.-L., Chen, H., Henderson, I.R.,
1130 Shinn, P., Pellegrini, M., Jacobsen, S.E., et al. (2006). Genome-wide high-resolution mapping
1131 and functional analysis of DNA methylation in arabidopsis. Cell *126*, 1189–1201.

1132 Zhang, Y., Wendte, J.M., Ji, L., and Schmitz, R.J. (2020). Natural variation in DNA methylation
1133 homeostasis and the emergence of epialleles. Proc. Natl. Acad. Sci. U. S. A. *117*, 4874–4884.

1134 Zhong, Z., Feng, S., Duttke, S.H., Potok, M.E., Zhang, Y., Gallego-Bartolomé, J., Liu, W., and
1135 Jacobsen, S.E. (2021). DNA methylation-linked chromatin accessibility affects genomic
1136 architecture in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. *118*.

1137 Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and
1138 animals. Genome Biol. *18*, 87.

1139 Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific
1140 siRNA accumulation and DNA and histone methylation. Science *299*, 716–719.

1141 Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide
1142 analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between
1143 methylation and transcription. Nat. Genet. *39*, 61–69.

1144

1145 **Additional files**

1146

1147 **Supplementary File 1.**

1148 CG differentially methylated regions during development.

1149

1150 **Supplementary File 2**.

1151 CHG differentially methylated regions during development.

1152

1153 **Supplementary File 3**.

1154 Transcript levels in rCMT3 compared to wild type embryos.

1155

1156 **Supplementary File 4**.

1157 CHG differentially methylated regions in rCMT3 compared to wild type.

1158

1159 **Supplementary File 5**.

1160 Oligonucleotides used in this study.

1161

1162 **Supplementary File 6**.

1163 MethylC-seq and mRNA-seq mapping statistics.