Histone H1 prevents non-CG methylation-mediated small RNA biogenesis in
***Arabidopsis* heterochromatin**

Jaemyung Choi^{1,#}, David B. Lyons^{1,#}, and Daniel Zilberman^{1,2,*}

¹Department of Cell and Developmental Biology, John Innes Centre, Norwich, UK

²Institute of Science and Technology, Klosterneuburg, Austria

[#]These authors contributed equally to this work

*Correspondence: daniel.zilberman@ist.ac.at

8 **Abstract**

9 Flowering plants utilize small RNA molecules to guide DNA methyltransferases to genomic
10 sequences. This RNA-directed DNA methylation (RdDM) pathway preferentially targets
11 euchromatic transposable elements. However, RdDM is thought to be recruited by
12 methylation of histone H3 at lysine 9 (H3K9me), a hallmark of heterochromatin. How RdDM
13 is targeted to euchromatin despite an affinity for H3K9me is unclear. Here we show that loss
14 of histone H1 enhances heterochromatic RdDM, preferentially at nucleosome linker DNA.
15 Surprisingly, this does not require SHH1, the RdDM component that binds H3K9me.
16 Furthermore, H3K9me is dispensable for RdDM, as is CG DNA methylation. Instead, we
17 find that non-CG methylation is specifically associated with small RNA biogenesis, and
18 without H1 small RNA production quantitatively expands to non-CG methylated loci. Our
19 results demonstrate that H1 enforces the separation of euchromatic and heterochromatic DNA
20 methylation pathways by excluding the small RNA-generating branch of RdDM from non-
21 CG methylated heterochromatin.

Introduction

Transposable elements (TEs) and their remnants comprise a substantial fraction of eukaryotic genomes and generally must be kept silent to ensure genome integrity and function (Bourque et al., 2018). TE silencing is achieved despite the disruption caused by each cell division, whereby half of the genome and histone proteins are made anew. Robust cellular memory of the inactive state is achieved by feedback interactions that reinforce and concentrate chromatin features and factors that contribute to transcriptional silencing and exclude activating factors (Allshire and Madhani, 2018; H. Zhang et al., 2018). However, silent chromatin domains are not homogenous. Flowering plants have two major types of TE-associated silent chromatin: GC-rich coding regions of autonomous TEs, and AT-rich chromatin comprised of gene-proximal TE remnants, short non-autonomous TEs and edges of autonomous TEs (Sequeira-Mendes et al., 2014; To et al., 2020; Zemach et al., 2013; Zhong et al., 2012). Although both are comprised of TEs, these chromatin types have distinct features (Sequeira-Mendes et al., 2014; Zemach et al., 2013). How two types of silent TE chromatin are distinguished and kept separate within the nucleus is a major open question.

Both types of TE chromatin feature extensive cytosine methylation in the CG context catalyzed by MET1 (plant homolog of Dnmt1) (Cokus et al., 2008; Lister et al., 2008; Zemach et al., 2013), and are also methylated at non-CG (CHG and CHH, where H is A, T or C) cytosines (Stroud et al., 2014; Zemach et al., 2013). GC-rich TE sequences have high levels of histone modifications associated with heterochromatin, including methylation of lysine nine of histone H3 (H3K9me), and are therefore known as heterochromatic TEs (Sequeira-Mendes et al., 2014; Zemach et al., 2013). Non-CG methylation at heterochromatic TEs is catalyzed primarily by chromomethylases (CMTs; CMT3 for CHG methylation and CMT2 for CHH), which are recruited to H3K9 dimethylated (H3K9me₂) nucleosomes by histone-tail-interacting domains (Du et al., 2012; Stoddard et al., 2019; Stroud et al., 2014;

Zemach et al., 2013). SUVH family H3K9me methyltransferases are in turn recruited to methylated DNA via SRA domains, forming a self-reinforcing loop (Du et al., 2014; Johnson et al., 2007; Rajakumara et al., 2011). *Arabidopsis thaliana* plants lacking functional chromomethylases (*cmt2cmt3* mutants) almost completely lack non-CG methylation at heterochromatic TEs, and their H3K9 methylation is greatly reduced (Stroud et al., 2014).

AT-rich TE sequences are low in H3K9me and other heterochromatic histone modifications, and are therefore known as euchromatic TEs (Sequeira-Mendes et al., 2014; Zemach et al., 2013). In contrast to the SUVH/CMT feedback loop that predominates in heterochromatin, RNA-directed DNA methylation (RdDM) catalyzes cytosine methylation within euchromatic TEs (Zemach et al., 2013; Zhong et al., 2012). RdDM loci are transcribed by a methylation-tolerant RNA polymerase II derivative (Pol IV) that couples co-transcriptionally with RNA-dependent RNA polymerase 2 (RDR2) to make double stranded RNA, which is processed into 23/24-nt fragments by Dicer-like 3 (DCL3) (Singh and Pikaard, 2019). These 24-nt small RNAs (sRNA) are subsequently denatured and loaded into Argonaute (AGO) protein complexes. AGO-sRNA complexes associate with another Pol II family enzyme, Pol V, to recruit DRM DNA methyltransferases (primarily DRM2 in *Arabidopsis*) (Erdmann and Picard, 2020; Matzke and Mosher, 2014; Raju et al., 2019; Wendte and Pikaard, 2017).

Like the SUVH/CMT pathway, RdDM comprises positive feedback loops. Pol V is recruited to methylated DNA, effectively seeking its own product (Liu et al., 2014; Wongpalee et al., 2019; Zhong et al., 2012). A more paradoxical feedback loop is thought to involve recruitment of Pol IV to H3K9me (Erdmann and Picard, 2020; Matzke and Mosher, 2014; Raju et al., 2019; Wendte and Pikaard, 2017). This hypothesis emerged from the observation that Pol IV-mediated sRNA production at many loci requires SHH1/DTF1, a protein that binds H3K9me₂ and monomethylated H3K9me (H3K9me₁) *in vitro* (Law et al.,

2013; Zhang et al., 2013). This model of Pol IV recruitment necessitates explaining how RdDM in general, and Pol IV specifically, is excluded from heterochromatic TEs with high H3K9me and targeted to euchromatic TEs with low H3K9me. Reliance of Pol IV on H3K9me also poses two theoretical questions. First, why would RdDM depend on a core component of the SUVH/CMT feedback loop (H3K9me₂), when the two DNA methylation systems have largely non-overlapping primary targets (Stroud et al., 2014), and RdDM targets are H3K9me-depleted? Second, the euchromatic TEs targeted by RdDM are often comprised of just one or two nucleosomes (Zemach et al., 2013). Maintenance of histone modifications is expected to be unstable at such short sequences due to the random partitioning of nucleosomes to sister chromatids following DNA replication (Angel et al., 2011; Berry and Dean, 2015; Lövkvist and Howard, 2021; Ramachandran and Henikoff, 2015; Zilberman and Henikoff, 2004). Why would RdDM, a pathway capable of almost nucleotide-level resolution (Blevins et al., 2015; Zhai et al., 2015) and specialized for silencing short TEs, be tied to a histone modification that requires longer sequences for stable propagation?

Here, we show that Pol IV activity is recruited to sequences with non-CG DNA methylation regardless of H3K9me, so that both the Pol IV and Pol V branches form positive feedback loops with the ultimate product of RdDM. We also show that linker histone H1 impedes RdDM activity in GC-rich heterochromatin, thereby restricting RdDM to AT-rich euchromatic TE. We propose that without H1, RdDM would be diluted into and effectively incapacitated by the vast stretches of non-CG methylated heterochromatin common in plant genomes (Feng et al., 2010; Niederhuth et al., 2016; Ritter and Niederhuth, 2021; Zemach et al., 2010). The affinity of H1 for GC-rich heterochromatin (Choi et al., 2020) focuses RdDM activity on short, AT-rich euchromatic TEs that RdDM is uniquely suited to silence.

Results

Histone H1 levels predict the global bifurcation of non-CG DNA methylating systems

To understand how the CMT and RdDM pathways are separated, we categorized *Arabidopsis* TEs by the dependence of their CHH methylation (mCHH) either on CMT2 (CMT TEs) or DRM2 (DRM TEs). Among 18784 TEs with more than 2% mCHH in wild type (*wt*) plants, 4486 TEs were demethylated in *cmt2* plants and 3039 TEs lost mCHH in *drm2* (mCHH in the mutants <0.02 , Fisher's exact test $p < 0.01$, TEs longer than 200 bp; Figure 1-figure supplement 1A and Figure 1-source data 1). Only 80 TEs had mCHH diminished below 2% in both mutants (Figure 1-source data 1), consistent with the largely separate sets of primary DRM and CMT targets (Sigman and Slotkin, 2016; Stroud et al., 2014).

Next, we used random forest classification (Breiman, 2001; Ishwaran et al., 2010) to identify predictors of DRM or CMT targets (Figure 1A). We included genetic and epigenetic features known to be associated with RdDM or CMT activity, as well as linker histone H1. H1 is specifically enriched in heterochromatic TEs, and its loss leads to increased DNA methylation at heterochromatic TEs and decreased methylation at euchromatic ones (Bourguet et al., 2021; Lyons and Zilberman, 2017; Papareddy et al., 2020; Rutowicz et al., 2015; Zemach et al., 2013). As expected, sRNA abundance can distinguish CMT and DRM TEs (Figure 1A). H3K9me1 is also a good classifier (Figure 1A). However, the best classifier turned out to be H1 (Figure 1A). Using all variables in Figure 1A, we could predict CMT and DRM TEs with an error rate of 2.15% (Figure 1B). With just H3K9me1 and H1, the prediction is almost as accurate (5.42% error; Figure 1B). Remarkably, H1 alone successfully identifies CMT and DRM TEs (12.17% error; Figure 1B), suggesting that H1 is fundamental to separating these silencing pathways.

RdDM activity relocates to heterochromatin without H1

To understand how H1 regulates the CMT and DRM pathways, we analyzed 24-nt sRNA expression, DNA methylation, and H3K9me2 in *h1* plants that have inactivating mutations in both of the canonical *Arabidopsis* H1 genes (Zemach et al., 2013). Consistent with published results (Bourguet et al., 2021; Lyons and Zilberman, 2017; Papareddy et al., 2020; Rutowicz et al., 2015; Zemach et al., 2013), we found an elevation of CHG methylation (mCHG), H3K9me2 and mCHH at CMT TEs (Figure 1C-D). CMT TEs are depleted of sRNAs in *wt* leaves, but sRNA expression increases 5.6-fold in *h1* plants (Figure 1D, Figure 1-figure supplement 1B-C). sRNA expression in *h1* positively correlates with that in *wt* (Figure 1-figure supplement 1B), indicating that loss of H1 amplifies sRNA expression at RdDM-capable CMT TEs rather than creating *de novo* RdDM targets.

In contrast to the hypermethylation of CMT TEs, DRM TEs lose H3K9me2, mCHG, mCHH and sRNA expression in *h1* plants (Figure 1C-D and Figure 1-figure supplement 1D). Despite the loss of sRNA at DRM TEs, global 24-nt sRNA abundance is not altered in *h1* plants (Figure 1-figure supplement 1E), indicating the reallocation of RdDM activity from DRM to CMT TEs. This phenomenon can be observed within individual TEs, with sRNA biogenesis and mCHH relocating from the AT-rich edges in *wt* to the GC-rich internal sequences in *h1* (Figure 1E). The relocation of sRNA production and mCHH into TE interiors in *h1* plants is also apparent in aggregate at TEs that retain substantial mCHH in *drm2* and *cmt2* mutants (intermediate TEs that are not classed either as DRM or CMT TEs; Figure 1-figure supplement 1A, F-G). CMT TE mCHH increases to the same relative extent in *h1* plants devoid of CMT2 (*h1c2*; Figure 1F and Figure 1-figure supplement 1H), indicating that mCHH hypermethylation at CMT TEs in *h1* mutants is caused by RdDM. These results indicate that RdDM relocates into heterochromatin in the absence of H1 and are consistent with recently published work (Bourguet et al., 2021; Papareddy et al., 2020).

Lack of H1 promotes sRNA biogenesis in linker DNA

Absence of H1 in *Arabidopsis* causes a preferential increase of heterochromatic TE DNA methylation within linker DNA, the regions between nucleosomes (Lyons and Zilberman, 2017). The average distance between heterochromatic nucleosomes is also reduced from ~180 to 167 bp (Choi et al., 2020). Given the relative promiscuity of RNA Pol IV initiation (Zhai et al., 2015) and the increased sRNA abundance at CMT TEs in *h1* (Figure 1D, Figure 1-figure supplement 1B-C), we asked whether patterns of sRNA production with respect to nucleosomes are altered in *h1*. As expected, overall levels of sRNA are increased around nucleosomes of CMT TEs and decreased at DRM TEs (Figure 2A-C and Figure 2-figure supplement 1). An overt sRNA linker bias is apparent in both *h1* and *wt* around the best-positioned nucleosomes (Figure 2A-C and Figure 2-figure supplement 1). This pattern becomes less obvious at less-well-positioned loci until it disappears completely (Figure 2B and Figure 2-figure supplement 1), as illustrated by measuring the correlation of the sRNA signal to itself (Figure 2D). The shortening *h1* sRNA autocorrelation around better positioned nucleosomes (Figure 2D) demonstrates how the linker histone dictates sites of sRNA production directly through linker occlusion and indirectly through nucleosome positioning.

Small RNA biogenesis is associated with H3K9me and non-CG methylation

Because H3K9me is thought to recruit Pol IV activity (Law et al., 2013; Zhang et al., 2013), we investigated how sRNA distribution changes in relation to H3K9me1/2 in *h1* plants. In *wt*, sRNA expression increases as H3K9me1 and H3K9me2 levels rise, but this trend reverses at TEs with more H3K9me and H1 (Figure 3A-B). In contrast, sRNA expression shows a relatively simple, direct relationship with H3K9me1 and H3K9me2 in *h1* plants (Figure 3A-B), suggesting that H1 prevents Pol IV from following the H3K9me gradient.

Unlike TEs, gene bodies normally have low levels of H3K9me, non-CG methylation and sRNA (Figure 3C) (H. Zhang et al., 2018). However, many genes gain H3K9me and non-CG methylation (especially mCHG) in plants lacking the H3K9 demethylase IBM1 (Miura et al., 2009). Although this hypermethylation does not require RdDM (Inagaki et al., 2010; Saze et al., 2008), recruitment of Pol IV by H3K9me would predict sRNA biogenesis in *ibm1* genes. Indeed, we find increased sRNA and mCHH levels in *ibm1* genes associated with high H3K9me₂ and mCHG (Figure 3C-D). Hence, the presence of H3K9me or non-CG methylation may be sufficient to trigger 24-nt sRNA production.

RdDM is recruited to CMT TEs independently of SHH1

The only H3K9me binding factor implicated in Pol IV recruitment is SHH1 (Law et al., 2013; Zhang et al., 2013; Zhou et al., 2018). Therefore, we tested whether CMT TE hypermethylation in *h1* plants requires SHH1. CMT TEs remain hypermethylated in *h1cmt2shh1* plants to about the same extent as in *h1cmt2* plants (Figure 4A), demonstrating that in the absence of H1, Pol IV is recruited to CMT TEs independently of SHH1.

Pol IV activity depends on a family of four CLSY putative chromatin remodeling proteins (Greenberg et al., 2013; Smith et al., 2007; Zhou et al., 2018). Simultaneous loss of CLSY1 and CLSY2 has the same effect as loss of SHH1, whereas CLSY3 and CLSY4 mediate RdDM at a largely distinct set of loci (Yang et al., 2018; Zhou et al., 2018). Mutations of SHH1 and CLSY1/2 preferentially reduce mCHH and sRNA at DRM TEs and increase mCHH at CMT TEs (Figure 4A). In contrast, *clsy3/4* mutant plants have reduced mCHH and sRNA at CMT TEs and increased mCHH and sRNA at DRM TEs (Figure 4A), suggesting that SHH1 and CLSY1/2 preferentially mediate RdDM at DRM TEs, whereas CLSY3/4 preferentially recruit Pol IV to CMT TEs. Consistently, TEs hypermethylated in *h1cmt2* and *h1cmt2shh1* show a strong overlap with published CLSY3/4-dependent sRNA

clusters and little overlap with CLSY1/2-dependent clusters (Figure 4B and Figure 4-figure supplement 1), suggesting that Pol IV relocation into heterochromatin involves CLSY3/4. However, our results do not rule out the possibility that some of the RdDM expansion in *hl* plants is mediated by CLSY1/2 or is independent of CLSY activity. Also, please note that the *wt* sRNA patterns in Figures 1D and 4A are distinct because the former is from leaves and the latter from inflorescences. Leaf sRNA levels are lower at CMT TEs and CLSY3/4 clusters compared to flowers (Figure 4-figure supplement 2), presumably due to higher expression of CLSY3/4 in reproductive tissues (Long et al., 2021; Zhou et al., 2021).

Overall, our results indicate that SHH1 is relatively unimportant for RdDM activity at H3K9me-rich CMT TEs with or without H1. The entry of Pol IV into H1-depleted heterochromatin must either involve a different H3K9me interacting factor, or a chromatin feature other than H3K9me.

RdDM expansion does not require CG methylation

Our results suggest that sRNA biogenesis at CMT TEs in *hl* mutants is mediated by CLSY3/4 Pol IV complexes. Recruitment of these complexes has been proposed to involve CG methylation (mCG) (Zhou et al., 2018). Therefore, we examined sRNA levels and DNA methylation in *hlmet1* mutants (Choi et al., 2020). Although MET1 is a CG methyltransferase, its loss also perturbs non-CG methylation and H3K9me₂ at some CMT TEs (Figure 5A and Figure 5-figure supplement 1A) (Choi et al., 2020; Deleris et al., 2012; Zabet et al., 2017; Y. Zhang et al., 2018). To understand how these changes impact sRNA production, we differentiate between two groups of CMT TEs in *met1* plants. MET1-independent CMT TEs keep non-CG methylation and H3K9me₂ in *met1* (Figure 5-figure supplement 1A) (Choi et al., 2020) and accordingly maintain sRNA expression (Figure 5B). These CMT TEs gain sRNA expression and mCHH in *hlmet1* relative to *met1* and *wt* (Figure

5B), demonstrating that mCG is not required for RdDM expansion into heterochromatin. In contrast, MET1-dependent CMT TEs, which lose non-CG methylation and H3K9me in *met1* (Figure 5-figure supplement 1A) (Choi et al., 2020), lose sRNA in *met1* and do not recover sRNA expression or mCHH in *h1met1* (Figure 5C), suggesting that non-CG methylation or H3K9me is necessary for sRNA biogenesis.

To test the above hypothesis, we grouped CLSY3/4 targets by mCHH level in *met1* ($\text{mCHH} \geq 0.05$ in *wt* and *met1*; $\text{mCHH} \geq 0.05$ in *wt* and < 0.05 in *met1*). Even though all CLSY3/4 targets lose mCG in *met1*, sRNA expression is reduced only when non-CG methylation and H3K9me2 are reduced (Figure 5D and Figure 5-figure supplement 1B), implying that the presence of non-CG methylation and/or H3K9me is sufficient to maintain CLSY3/4-dependent sRNA biogenesis. In *h1met1*, sRNA levels increase at CLSY3/4 targets where non-CG methylation is maintained: among 1565 CLSY3/4 clusters with *wt* non-CG methylation ($> 0.01\%$), 72% keep non-CG methylation in *met1* and gain sRNA expression in *h1met1* (*met1* $\text{mCH} > 0.01$), whereas 15% effectively lose all non-CG methylation in *met1* and have similarly low sRNA levels in *met1* and *h1met1* (*met1* $\text{mCH} < 0.005$, Figure 5E and Figure 5-figure supplement 1C). These results indicate that neither CLSY3/4 Pol IV activity, nor the RdDM expansion triggered by loss of H1, depend on mCG.

Severe H3K9me reduction does not perturb RdDM expansion into heterochromatin

Our results so far indicate that H1 prevents RdDM from following a gradient of either H3K9me or non-CG methylation into heterochromatin. However, heterochromatin is structurally complex and contains many factors (Feng and Michaels, 2015). To understand the overall importance of heterochromatin integrity, we tested the effects of H1 on sRNA distribution in plants with a mutation in the Swi/Snf2 chromatin remodeler DDM1, which have severely compromised heterochromatin (Kim and Zilberman, 2014; Sigman and Slotkin,

2016). The *ddm1* mutation greatly reduces heterochromatic DNA and H3K9 methylation (Ito et al., 2015; Lyons and Zilberman, 2017; Osakabe et al., 2021; Teixeira et al., 2009; Zemach et al., 2013), activates TE expression (Lippman et al., 2004; Osakabe et al., 2021; Panda et al., 2016; Panda and Slotkin, 2020; Rougée et al., 2020) and disperses nuclear heterochromatic foci (Rougée et al., 2020; Soppe et al., 2002) (Figure 6A-B and Figure 6-figure supplement 1A). However, 24-nt sRNA expression in *ddm1* is broadly similar to *wt* (Figure 6C-D and Figure 6-figure supplement 1B). Simultaneous lack of H1 and DDM1 in *h1ddm1* mutants (Lyons and Zilberman, 2017; Zemach et al., 2013) causes relocation of sRNA biogenesis into CMT and intermediate TEs that mirrors that in *h1* plants (Figure 6C-D and Figure 6-figure supplement 1B), indicating that overall heterochromatin integrity is not required for this process. Furthermore, RdDM expansion into heterochromatin occurs in *h1ddm1* despite strong H3K9me reduction compared to *wt* and *h1* (Figure 6A-B and Figure 6-figure supplement 1A). This does not rule out the possibility that H3K9me promotes Pol IV activity, because the H3K9me remaining in *h1ddm1* may be sufficient. However, the observation that sRNA production at CMT TEs is largely unaffected by a bulk H3K9me reduction argues against a primary role for H3K9me in Pol IV recruitment.

H3K9me and non-CG methylation can be decoupled in heterochromatin

H3K9me and non-CG DNA methylation are closely associated in heterochromatin due to the feedback loop between CMT2/3 and the SUVH4/5/6 H3K9 methyltransferases (Du et al., 2012; Stoddard et al., 2019; Stroud et al., 2014). To isolate the effects of these features on sRNA biogenesis, we examined DNA methylation, H3K9me and sRNA levels in *c2c3* and *h1c2c3* plants. While CG methylation is largely unaffected, non-CG methylation is specifically abolished at CMT TEs in these plants (Figure 7-figure supplement 1A), consistent with previously published *c2c3* results (Stroud et al., 2014). As expected, H3K9me

is also greatly reduced (Figure 7-figure supplement 1A), but some H3K9me1 and H3K9me2 remains in heterochromatin. Specifically, 875 CMT TEs maintain H3K9me1 and 1126 maintain H3K9me2 in *c2c3*, while in *h1c2c3* we identified 2434 H3K9me1-enriched CMT TEs and 1443 H3K9me2-enriched CMT TEs (Figure 7A and 7B). Principal component analysis shows that H3K9me in these mutants associates with mCG, followed by CG and CCG density (which contributes to mCG density; Figure 7C and Figure 7-figure supplement 1B), suggesting that SUVH4/5/6 are recruited to mCG in the absence of non-CG methylation.

This conclusion is supported by a complementary pattern of H3K9 methylation changes in *h1c2c3* vs. *met1*. TEs that lose H3K9me2 in *met1*, suggesting H3K9me dependence on mCG, maintain H3K9me in the absence of mCHG/mCHH in *h1c2c3* (Figure 7D). Conversely, TEs that lose H3K9me in *h1c2c3*, suggesting H3K9me dependence on mCHG/mCHH, retain H3K9me2 in *met1* (Figure 7E). This indicates that H3K9me at mCG-dense CMT TEs is partially dependent on mCG, leading to considerable H3K9me retention in *c2c3*, and especially *h1c2c3* plants. The ability of mCG to recruit H3K9me is consistent with published work, including studies that show RdDM-independent initiation of the CMT-SUVH feedback loop specifically at CG-methylated sequences (Miura et al., 2009; To et al., 2020; Zabet et al., 2017) and the observed affinity of SUVH histone methyltransferase SRA domains for mCG *in vitro* (Johnson et al., 2007; Li et al., 2018; Rajakumara et al., 2011).

24-nt sRNA production specifically correlates with non-CG methylation

The decoupling of H3K9me and non-CG methylation in *h1c2c3* plants allowed us to determine how each feature is associated with sRNA biogenesis. In *h1* plants, H3K9me2, DNA methylation in every context, and sRNA expression together increase in direct relation to *wt* H1 prevalence, as loss of H1 increases accessibility of previously H1-rich TEs (Figure 8A and Figure 8-figure supplement 1A) (Bourguet et al., 2021; Lyons and Zilberman, 2017;

Papareddy et al., 2020; Zemach et al., 2013). H3K9me1/2, DNA methylation, and sRNA levels are also all positively correlated in *hl* plants, though the correlation between H3K9me2 and sRNA is weak (Figure 8B and Figure 8-figure supplement 1B). In contrast, the coupling of H3K9me with DNA methylation and sRNA levels nearly disappears when comparing *hlc2c3* to *c2c3* (Figure 8C-D and Figure 8-figure supplement 1C-D). Relative H3K9me1/2 abundance increases with *wt* H1 levels, whereas DNA methylation and sRNA changes show at best a very weak relationship with *wt* H1 enrichment (Figure 8C and Figure 8-figure supplement 1C).

Two correlated groups remain in *hlc2c3*: H3K9me1/2 with mCG, and sRNA with mCHG/mCHH (Figure 8D and Figure 8-figure supplement 1D). The linear correlations between sRNA and either H3K9me1 or mCG observed in *hl* (Figure 8E) become kinked in *hlc2c3* (Figure 8F), resembling the association between sRNA and H3K9me1 in *wt* (Figure 3B). The overall pattern of *hlc2c3* sRNA at CMT and intermediate TEs resembles *wt* far more than *hl* (Figure 8G and Figure 8-figure supplement 1E). The patterns and levels of sRNA and mCHH at DRM TEs are also similar between *hlc2c3* and *wt* (Figure 8-figure supplement 1F-G). Only the association between non-CG methylation and sRNA remains linear in *hlc2c3* (Figure 8E and 8F). This dynamic can be observed at an individual array of CMT TEs (Figure 8H). 24-nt sRNA expression is confined to the edges of the CMT TE array in *wt*, but follows H3K9me and DNA methylation throughout the array in *hl* plants (Figure 8H). In *hlc2c3*, non-CG methylation within the array is strongly reduced, but H3K9me is maintained, and sRNA expression exhibits a broadly *wt* pattern associated with remaining mCHH but not with H3K9me (Figure 8H).

It is important to note that in plants lacking CMT2/3, all mCHH should be catalyzed by RdDM, and a correlation between sRNA (product of the Pol IV pathway) and mCHH (product of the Pol V pathway) is therefore expected regardless of how Pol IV is recruited.

The key observations are that loss of CMT2/3 in *h1c2c3* plants (and the associated loss of mCHG/mCHH) largely abrogates the relocation of Pol IV activity into heterochromatin (Figure 8G-H and Figure 8-figure supplement 1E-G), and the remaining heterochromatic sRNA biogenesis is not associated with H3K9me or mCG (Figure 8D-F). These results do not support the hypothesis that Pol IV is recruited by H3K9me, and offer non-CG methylation as the most likely alternative. Our data suggest the hypothesis that without H1, non-CG methylation catalyzed by CMT2/3 pulls Pol IV into heterochromatin, and loss of CMT2/3 allows Pol IV to return to its mostly euchromatic *wt* targets.

CLSY1/2 RdDM activity specifically associates with non-CG methylation

24-nt sRNA expression is globally associated with non-CG methylation rather than H3K9me in *h1c2c3*, but these correlations are primarily driven by heterochromatic regions with low *wt* RdDM. To determine if this trend translates to euchromatic TEs where SHH1 is required for RdDM, we analyzed associations between H3K9me, DNA methylation, and sRNA expression in published CLSY1/2 sRNA clusters in *wt* plants (Figure 9A) (Zhou et al., 2018). In clusters grouped by H3K9me and mCHH, sRNA expression is associated with high mCHH, but not with high H3K9me (Figure 9A), supporting the idea that non-CG methylation dictates Pol IV localization (with the caveat that non-CG methylation is a product of RdDM).

As a further test of our hypothesis, we analyzed published data from plants lacking the three H3K9 methyltransferases implicated in the CMT/SUVH positive feedback loop. In these *suvh4/5/6* mutants, H3K9me₂ and non-CG methylation are strongly diminished and sRNA expression of CLSY1/2 clusters is decreased (Stroud et al., 2014; Zhou et al., 2018). If H3K9me₂ recruits Pol IV via SHH1, the limited remaining H3K9me would be expected to correlate with sRNA. Instead, we find sRNA expression in *suvh4/5/6* follows mCHH but not H3K9me₂ (Figure 9B-C, compare left and right elements in 9C), consistent with our

observations in heterochromatin. 24-nt sRNA correlates much more strongly with non-CG methylation than with H3K9me2 in *suvh4/5/6* plants (Figure 9D), highlighting the limited importance of H3K9me for sRNA biogenesis.

Finally, we assayed CLSY1/2 clusters with low *wt* H3K9me2 but high *wt* sRNA and mCHH (LH CLSY1/2 clusters) in *polv* mutants to determine whether non-CG methylation is required to maintain sRNA expression. RNA Pol V is not directly involved in sRNA production, but is an essential RdDM component required for DNA methylation because it recruits DRM2 (Erdmann and Picard, 2020; Matzke and Mosher, 2014; Raju et al., 2019; Wendte and Pikaard, 2017). Therefore, *polv* mutants allow us to differentiate non-CG methylation as a cause vs. a consequence of Pol IV activity. 90% of the 662 LH CLSY1/2 clusters lose mCHH in *polv* plants (mCHH<0.05, Figure 9E), and the overall non-CG methylation of LH CLSY1/2 clusters is greatly reduced without Pol V (Figure 9F). In *suv4/5/6* mutants, LH CLSY1/2 clusters maintain sRNA expression, whereas sRNA expression in *polv* mutants is greatly reduced (Figure 9G). Furthermore, mCG at LH CLSY1/2 clusters is higher in *polv* than in *suvh4/5/6* plants (Figure 9H). Therefore, sRNA biogenesis is not sensitive to the loss of either H3K9me2 or mCG and specifically requires non-CG methylation.

Discussion

We have examined intertwined chromatin features – sRNA production, DNA methylation, and H3K9 methylation – to understand how the genomic sites of Pol IV activity are specified. We find that two main factors are involved. First, linker histone H1 prevents sRNA production in heterochromatin (Figure 10). Without H1, RdDM relocates from its usual euchromatic targets into heterochromatic TEs (Figure 1 and Figure 1-figure supplement 1), as has been recently observed by an independent study (Papareddy et al., 2020). Another

heterochromatic protein, the histone variant H2A.W, may also contribute to the exclusion of RdDM from heterochromatin, but this effect is modest and only observed when H1 is absent (Bourguet et al., 2021). In the presence of H1, lack of H2A.W instead strengthens the exclusion of RdDM from heterochromatin, potentially due to enhanced heterochromatic H1 accumulation (Bourguet et al., 2021). Overall, the available evidence indicates that H1 is the major factor excluding Pol IV from heterochromatin.

Second, we find that non-CG methylation promotes Pol IV activity (Figure 10), contrary to the well-established view that Pol IV is recruited by H3K9me (Erdmann and Picard, 2020; Law et al., 2013; Raju et al., 2019; Wendte and Pikaard, 2017; Zhang et al., 2013), and the more recent proposal that mCG may be involved (Zhou et al., 2018). The hypothesis that non-CG methylation recruits Pol IV has a long history (Herr et al., 2005; Li et al., 2020; Zemach et al., 2013), but testing it has been challenging because non-CG methylation is associated with other epigenetic and chromatin features, including mCG and H3K9me (Law and Jacobsen, 2010; Xu and Jiang, 2020; H. Zhang et al., 2018). The link with H3K9me has been particularly difficult to break because of the CMT-SUVH feedback loop (Du et al., 2012; Johnson et al., 2007; Li et al., 2018; Stoddard et al., 2019).

However, we have used *h1c2c3*, *suvh4/5/6* and *polv* mutants to disentangle H3K9me and non-CG methylation. In all three backgrounds, sRNA biogenesis follows non-CG methylation instead of H3K9me (Figure 8, 9 and Figure 8-figure supplement 1). The *h1c2c3* line has been particularly informative due to the many TEs that maintain H3K9me but lack non-CG methylation (Figure 8 and Figure 8-figure supplement 1). H3K9me may be substantially retained in *h1c2c3* heterochromatin because lack of H1 allows SUVH methyltransferases easier access, so that the weak affinity of their SRA domains for mCG suffices for effective recruitment (Johnson et al., 2007; Li et al., 2018; Rajakumara et al., 2011). Whatever the mechanism, the strong linear association between sRNA biogenesis and

non-CG methylation, and the lack of such an association with H3K9me and mCG (Figure 8 and 9), provide strong support for the hypothesis that non-CG methylation recruits Pol IV (Figure 10).

Our data linking 24-nt biogenesis with non-CG methylation do not mean that such methylation is absolutely required for Pol IV recruitment. Indeed, there is residual 24-nt biogenesis in *ddcc* mutants that lack non-CG methylation (Stroud et al., 2014). One possibility is that the factor or factors recruiting Pol IV to non-CG methylation have weak affinity for mCG, which could recruit Pol IV in the absence of non-CG methylation, analogous to our proposed mode of SUVH4/5/6 recruitment in plants lacking CMT2/3. Other chromatin features may also recruit or facilitate Pol IV activity. However, our results indicate that non-CG methylation is the major Pol IV recruiting genomic feature under normal conditions.

The linking of Pol IV activity to non-CG methylation instead of H3K9me resolves several thorny issues. First, the observation that SHH1 – the proposed H3K9me reader – is preferentially required for RdDM where H3K9me is low (Zhou et al., 2018) can be easily accommodated if H3K9me is not directly involved in RdDM. Similarly, the finding that severe loss of H3K9me in *suvh4/5/6* mutants is accompanied by only a modest reduction of sRNA levels (Zhou et al., 2018) is no longer mysterious. At a more fundamental level, this hypothesis ties RdDM in a feedback loop with its product and unties it from a histone modification produced by the distinct CMT-SUVH pathway and depleted from RdDM target sequences. Breaking RdDM from dependence on any histone modification is also conceptually important because a core theoretical strength of RdDM is the ability to maintain methylation at much shorter sequences than those where stable histone-based epigenetic inheritance is possible (Angel et al., 2011; Lövkvist and Howard, 2021; Ramachandran and Henikoff, 2015; Zilberman and Henikoff, 2004).

410 Long TEs that can be effectively silenced by the histone-dependent CMT-SUVH
411 pathway tend to be relatively GC-rich because they contain coding sequences (Sequeira-
412 Mendes et al., 2014; To et al., 2020; Zemach et al., 2013). In contrast, short non-autonomous
413 TEs and TE remnants tend to lack coding sequences and are thus AT-rich. In this context, the
414 GC sequence preference of *Arabidopsis* H1 (Choi et al., 2020) may be key. GC bias is far
415 from a H1 universal, with most animal H1 variants preferring AT-rich DNA (Cao et al., 2013;
416 Izaurralde et al., 1989; Tomaszewski and Jerzmanowski, 1997). The preferences of plant H1
417 may have evolved, at least in part, to target it to coding sequences, including those of
418 autonomous heterochromatic TEs. This would allow H1 to exclude RdDM from such
419 sequences, which can cover vast tracts of plant genomes (Michael, 2014; Suzuki and Bird,
420 2008), and focus RdDM on the short TEs it is specialized to silence. The interplay of H1 and
421 non-CG methylation can thus produce the preferential activity of RdDM at short, AT-rich
422 TEs observed throughout flowering plants (Gouil and Baulcombe, 2016; Numa et al., 2015;
423 Tan et al., 2018).

Key Resources Table				
Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
antibody	anti-H3K9me1 (Rabbit polyclonal)	Millipore	07-450 RRID:AB_310625	1:200
antibody	anti-H3K9me2 (Mouse monoclonal)	Abcam	ab1220 RRID:AB_449854	1:200
commercial assay or kit	Library construction (Native ChIP)	Tecan	3460-24	
commercial assay or kit	Bisulfite conversion	QIAGEN	59104	
commercial assay or kit	Library construction (bisulfite sequencing)	New England Biolabs	E7645 and E7335S	
commercial assay or kit	Library construction (small RNA)	Illumina	RS-200-0012 and RS-200-0024	
software, algorithm	cutadapt	DOI:10.14806/ej.17.1.200	RRID:SCR_011841	
software, algorithm	bowtie	DOI: 10.1186/gb-2009-10-3-r25	RRID:SCR_005476	
software, algorithm	deepTools2	DOI:10.1093/nar/gkw257		
software, algorithm	dzlabtools	DOI:10.1126/science.1172417		https://zilbermanlab.net/tools/

software, algorithm	RandomFore stExplainer	DOI:10.1198/jas a.2009.tm08622		
software, algorithm	IGV	DOI:10.1038/nbt .1754	RRID:SCR_0 11793	
software, algorithm	Gene Cluster 3.0	DOI:10.1093/bio informatics/bth0 78		
software, algorithm	corrplot	doi:10.1198/000 313002533 doi:10.1080/000 31305.1996.104 74371		
software, algorithm	Treeview	DOI:10.1093/bio informatics/bth0 78	RRID:SCR_0 16916	

Biological materials

cmt2 and *cmt2cmt3* (Stroud et al., 2014; Zemach et al., 2013) plants were crossed to *h1.1h1.2* (Zemach et al., 2013) plants to generate *h1cmt2* and *h1cmt2cmt3* plants. To establish the *h1cmt2shh1* mutant line, we crossed *h1+/-cmt2* plants with *shh1* (SALK_074540C) plants, then isolated *h1cmt2shh1* homozygous siblings. *met1*, *h1met1*, *ddm1*, *h1ddm1* plants were described previously (Choi et al., 2020; Lyons and Zilberman, 2017). *Arabidopsis thaliana* seedlings were germinated and grown for 4-5 weeks on soil at 20-25°C in growth chambers (16 hr day/8 hr night) for all the experiments performed except for *met1*, *h1met1*, and corresponding *wt* seedling sRNA libraries. These seedlings were germinated and grown for two weeks in half-strength Gamborg's B-5 liquid media (Caisson Labs, cat. No. GBP07) at 22-25°C under continuous light with shaking at 125 RPM.

Bisulfite Sequencing (BS-seq) library preparation

BS-seq libraries were constructed using genomic DNA (gDNA) extracted from rosette leaves of 4-5-week-old plants. 500 ng total gDNA was sheared to 100-1000bp using Bioruptor Pico (Diagenode), then purified with 1.2X volume of SPRI beads (Beckman Coulter, cat. No. A63881). Fragmented gDNA was ligated to NEBNext Adaptor for Illumina using NEBNext Ultra II DNA library prep kit for Illumina (New England Biolabs, cat. No. E7645). We performed bisulfite conversion twice with ligated libraries (QIAGEN, cat. No. 59104) to prevent incomplete conversion (<99% conversion) of unmethylated cytosines. Converted libraries were subjected to SPRI bead purification with 0.8X volume of beads. We amplified bisulfite-converted libraries with NEB next indexing primers (New England Biolabs Inc. cat. No. E7335S).

Small RNA sequencing library preparation

To isolate small RNA, we extracted total RNA from rosette leaves of 4-5-week-old plants using Trizol (Invitrogen, cat. No. 15596026) according to manufacturer's manual. To remove DNA from samples, 5 µg of RNA was treated with DNA-free DNA removal kit (Thermo, cat. No. AM1907). 1 µg of DNA-free total RNA was subjected to sRNA library construction according to manufacturer's protocol (Illumina, cat. No. RS-200-0012 and RS-200-0024).

Native Chromatin Immunoprecipitation and sequencing library preparation

MNase digestion of native chromatin was carried out on 0.5g of 4 week old *Arabidopsis* rosette leaves as described previously (Lyons and Zilberman, 2017). Digestion was stopped with EGTA and chromatin was rotated at 4°C for 30 minutes. The preparation was then centrifuged for 10 minutes at 2000 RPM and solubilized chromatin fragments were isolated by aspirating supernatant immediately. Chromatin was then diluted to 1ml in wash buffer A (50mM Tris-HCl pH 7.5, 50mM NaCl, 10mM EDTA) and antibody added at 1µl per 0.1 g of

total starting material (Millipore, cat. No. 07-450 for H3K9me1, Abcam, cat. No. ab1220 for H3K9me2). Dilute Tween-20 was added to a final concentration of 0.1%, and the mixture was rotated overnight at 4°C. All buffers were supplemented with PMSF and protease inhibitor (Roche (Merck), cat. No. 11873580001). A standard immunoprecipitation procedure was used the following day. Briefly, pre-blocked Protein-A and -G dynabeads (Invitrogen, cat. No. 10001D and 10003D) were incubated with the chromatin preparation for 3 hrs. rotating at 4°C, and the beads/chromatin mixture was then washed on ice in Tris-EDTA buffer with increasing concentrations of NaCl, starting at 50mM and ending at 150mM. DNA was eluted from beads by shaking in 1% SDS and 1% NaHCO₃ for 10 minutes at 55°C, and DNA was purified with phenol-chloroform extraction. Input and ChIP DNA was converted into sequencing libraries using Celero DNA reagents (Tecan, cat. No. 3460-24) following manufacturer's instructions.

Sequencing

Sequencing was performed at the John Innes Centre with the NextSeq 500 (Illumina), except for sRNA libraries from seedlings (*wt*, *met1*, and *hlmet1*). These seedling libraries were sequenced at the Vincent J. Coates Genomic Sequencing Laboratory at the University of California, Berkeley with the HiSeq 4000 (Illumina).

Sequence Alignment and data preparation

For sRNA-seq libraries, adapter sequences were removed from reads using cutadapt (Martin, 2011). 18-28 bp, 21nt, and 24 nt fragments were isolated using the following cutadapt options: -m 18 -M 28, -m 21 -M 21, -m 24 -M 24. Reads were mapped with Bowtie (Langmead et al., 2009) allowing up to 1 mismatch and up to 10 multi-mapped reads. Aligned 21-nt or 24-nt read counts were normalized by reads per kilobase per million mapped reads (RPKM) of 18-

28 bp fragments. ChIP-seq libraries were mapped with Bowtie (Langmead et al., 2009) allowing up to 2 mismatches and up to 10 multi-mapped reads. To calculate enrichment, ChIP samples were divided by input samples and transformed into $\log_2$ ratio values using deepTools2 bamCompare (Ramírez et al., 2016). For H3K9me1 and H3K9me2 from WT, *hl*, *ddm1*, *h1ddm1*, *c2c3*, and *h1c2c3*, we used a random subset of input reads equivalent to 25% of the total uniquely mapped reads of the corresponding IP for input into bamCompare. For BS-seq libraries, reads were mapped with the bs-seq pipeline (<https://zilbermanlab.net/tools/>).

Description of *Arabidopsis* genome features

‘Transposable elements’ include transposon annotation from (Panda and Slotkin, 2020). Araport11 TE genes and pseudogenes, and genomic regions with TE-like DNA methylation (Cheng et al., 2017; Choi et al., 2020; Panda and Slotkin, 2020; Shahzad et al., 2021). We filtered out elements shorter than 250 bp. Previously, we merged overlapping TE annotations into single TE unit, then defined heterochromatic TEs and euchromatic TEs as transposons that have more than 0 or less than 0 H3K9me2 ($\log_2$ ChIP/Input) in *wt* plants (Choi et al., 2020). Both CMT and DRMs target these merged, long TEs, as the edges of TEs are methylated by DRMs and the bodies of TEs are methylated by CMTs. Therefore, to isolate TEs with non-CG methylation dependent on CMTs or DRMs, we did not merge TE annotations here. Among TEs with mCHH methylation ($mCHH > 0.02$), CMT-dependent TEs were defined as the TEs that lost mCHH methylation in *cmt2* plants ($mCHH < 0.02$ in *cmt2*). DRM-dependent TEs were defined as the TEs that lost mCHH methylation in *drm2* plants ($mCHH < 0.02$ in *drm2*). sRNA cluster annotation is from (Zhou et al., 2018).

Classification of MET1-dependent and -independent CMT TEs

We previously defined MET1-dependent TEs as the TEs that lost H3K9me2 in *met1* plants (Choi et al., 2020). In this study, to evaluate how DNA methylation affects CLSY3/4-dependent sRNA expression, we defined MET1-dependent TEs as the TEs that lost mCHH methylation in *met1* ($\text{mCHH in } wt \geq 0.05$, $\text{mCHH in } met1 < 0.02$), and MET1-independent TEs as ones that keep mCHH methylation in *met1* ($\text{mCHH in } wt \geq 0.05$, $\text{mCHH in } met1 \geq 0.05$).

Random forest classification and prediction

To measure the importance of each genetic and epigenetic marker to classify DRM and CMT TEs, we first calculated average enrichment of various histone modifications, histone H1, average sRNA expression and DNA methylation level at each TE using `window_by_annotation.pl` Perl script (<https://zilbermanlab.net/tools/>). We also included density of various cytosine sequence contexts. The importance of each variable was evaluated using ‘`randomForest`’ and ‘`measure_importance`’ function in `RandomForestExplainer` R package (Ishwaran et al., 2010). The importance matrices were visualized by ‘`plot_multi_way_importance`’ function of the same package.

To evaluate the predictive power of each variable, we randomly divided TEs into training and validation sets. The random forest classifier was built using TEs in the training set with indicated variables and the classification of each TE (DRM or CMT). The trained model was used to predict the category of TEs in the validation set, and the error rate was calculated by comparing the predicted classification and its actual classification. We used ‘`randomforest`’ and ‘`predict`’ function in `randomForest` R Package.

Data visualization

Enrichment scores of various genomic and epigenomic features were generated by window_by_annotation.pl Perl scripts (<https://zilbermanlab.net/tools/>). For scatter plots and heatscatter plots in Figure 1, the enrichment scores were imported to R (Davey et al., 1997) and visualized by ggplot2 R package (Wickham, 2009) or ‘heatscatter’ function in LSD R package (Venables and Ripley, 2002). For scatter plots and heatscatter plots in other figures, TEs were sorted by their GC content, then average feature enrichments of 100 TEs were calculated to reduce the variability of data. DNA methylation, H3K9 methylation and sRNA distribution around TEs were generated with ends_analysis.pl and average_ends_new.pl Perl scripts (<https://zilbermanlab.net/tools/>). For sRNA distribution, we removed bins with higher than 200 rpkm to prevent outliers skewing the average. For proportional Venn diagram, TE ID lists in each group were uploaded to BioVenn (Hulsen et al., 2008). To visualize the relationship among genetic, epigenetic features and sRNA expression in *c2c3* and *h1c2c3* plants, principal component analysis was applied to arrays of features using Gene Cluster 3.0 (de Hoon et al., 2004; Figure 6C). For Pearson’s correlation coefficient plots, the DNA methylation, H3K9 methylation, and sRNA expression level matrices were imported to R and visualized using corrplot R package (Friendly, 2002; Murdoch and Chow, 1996; Figure 7 and 8). Screenshots of *Arabidopsis* genomic loci were taken in IGV (Robinson et al., 2011; Thorvaldsdottir et al., 2013). Treeview was used to generate heatmaps (de Hoon et al., 2004). For sRNA plots around nucleosomes (Figure 2), previously published nucleosome dyad coordinates were used (Lyons and Zilberman 2017) as anchors around which 10-bp bins of 24-nt sRNA were averaged and plotted. Autocorrelation estimates were generated on these averages using the built-in R ‘acf’ function.

Use of Previously Published Data

551 DNA methylation data of *wt*, *drm2*, *c2c3*, *ddcc* and *ibm1* plants (Stroud et al., 2014; Zemach
552 et al., 2013), DNA methylation and sRNA data of *clsy1/2*, *clsy3/4*, and *shh1* plants (Zhou et
553 al., 2018), DNA methylation, MNase, well-positioned nucleosome loci data of *wt* and *hl*
554 plants (Lyons and Zilberman, 2017), DNA methylation, H1 and H3K9me data of *wt*, *met1*
555 and *hlmet1* plants (Choi et al., 2020), H3K9me2 and sRNA expression data of *wt* and *ibm1*
556 plants (Fan et al., 2012; Lai et al., 2020), DNA methylation, H3K9me2, and sRNA
557 expression data of *suvh4/5/6* plants (Papareddy et al., 2020; Stroud et al., 2014), and DNA
558 methylation and sRNA data of *polv* plants (Johnson et al., 2014; Zhong et al., 2012) were
559 obtained through GEO (GEO accessions: GSE51304, GSE41302, GSE99694, GSE122394,
560 GSE108487, GSE32284, GSE152971, GSE52041 and GSE39247).

561 **Data Availability**

562 The accession number for the NGS data reported in this paper is GEO: GSE179796.

563 **Additional Files**

564 Figure 1-source data 1

565 Lists of CMT- and DRM-dependent transposons and intermediate transposons in *Arabidopsis*

566 **Acknowledgements**

567 We thank X. Feng for helpful comments on the manuscript. This work was supported by a
568 European Research Council grant MaintainMeth (725746) to D.Z.

569

570 Data generated from this study is available under GEO accession GSE179796. The authors
571 declare no competing financial interests.

572 **Figure Legends**

573 **Figure 1** Histone H1 prohibits RNA-directed DNA methylation of CMT-dependent
574 heterochromatic transposons.

575 **(A)** The importance of DNA methylation, histone H3 modifications, sRNA, H1, and cytosine
576 sequence context to predict CMT TE or DRM TE classes by random forest classification. **(B)**
577 Prediction of CMT or DRM TE classes by random forest classification with all variables, H1
578 and H3K9me1, or only H1. **(C and D)** Heatmaps of H3K9me2 and mCHG levels **(C)** and
579 mCHH and sRNA levels **(D)** at CMT and DRM TEs in *wt* and *h1* plants. TEs were sorted by
580 mCHH level in *wt*. **(E)** Example of DNA methylation and sRNA expression at a CMT TE in
581 *wt* and *h1* (AT1TE58075). **(F)** mCHH difference between *wt* and *h1* (x-axis) versus *h1cmt2*
582 (*h1c2*) and *cmt2* (*c2*; y-axis) at CMT TEs.

583 **Figure 1-figure supplement 1** CMT TEs gain non-CG DNA methylation and sRNA
584 expression in *h1* plants.

585 **(A)** Heatmaps of H3K9me, H1, GC content, mCHG, and mCHH at CMT, DRM and
586 intermediate TEs in *wt*, *cmt2* (*c2*), *cmt2cmt3* (*c2c3*), *drm2* (*d2*), and *drm1drm2cmt2cmt3*
587 (*ddcc*) plants. 3807 intermediate TEs had more than 2% mCHH in *cmt2* and *drm2* plants
588 (TEs longer than 200 bp). TEs were sorted by mCHH level in *wt*. **(B)** Average sRNA
589 expression level of CMT TEs in *wt* and *h1* plants. **(C)** Average sRNA distribution around
590 CMT TEs in *wt* and *h1* plants. **(D)** Average sRNA expression level of DRM TEs in *wt* and *h1*
591 plants. **(E)** Normalized number of mapped reads of 21-nt and 24-nt length sRNAs in *wt* and
592 *h1* plants. The number of mapped reads of 21-nt sRNA was normalized to 1. **(F and G)**
593 Average sRNA distribution **(F)** and mCHH level **(G)** around intermediate TEs in *wt* and *h1*
594 plants. **(H)** Heatmaps of mCHH in *c2* and *h1cmt2* (*h1c2*) plants. TEs were sorted by mCHH

595 level in *wt*. (**B** and **D**) rpkm indicates reads per kilobase of transcript, per million mapped
596 reads. *r* indicates Pearson's correlation. (**C**) rpm indicates reads per million mapped reads.

597 **Figure 2** RdDM is preferentially active in linker DNA in *hl* plants.

598 (**A** and **C**) Example of a CMT TE (Chr2:3,944,600-3,946,400) (**A**) and DRM TE
599 (Chr2:6,389,500-6,392,500) (**C**) with well-positioned nucleosomes (yellow boxes).
600 Smoothed MNase-seq (apricot), sRNA expression (gray) and mCHH (indigo) are plotted. (**B**)
601 Average sRNA expression around well positioned or poorly positioned nucleosomes at CMT
602 TEs. (**D**) Autocorrelation estimates of average sRNA values shown in (**B**) to illustrate
603 shortened sRNA phasing in *hl* mutants corresponding to shortened nucleosome repeat length.
604 Nucleosome positioning data and designations are from (Lyons and Zilberman, 2017).

605 **Figure 2-figure supplement 1** RdDM is enriched at linker DNA around well positioned
606 nucleosomes.

607 Average sRNA expression around well positioned (left) to increasingly poorly positioned
608 nucleosomes at DRM TEs.

609 **Figure 3** Small RNA biogenesis is associated with H3K9me and non-CG methylation.

610 (**A** and **B**) Average H3K9me2 (**A**) or H3K9me1 (**B**) (x-axis) and sRNA expression level (y-
611 axis) in *wt* and *hl*. Each dot represents the average of 100 TEs sorted by GC content. (**C**)
612 Distribution of H3K9me2, non-CG methylation and sRNA expression around 5' ends of
613 genes in *wt* and *ibm1* plants. (**D**) A boxplot shows sRNA expression level at genes in *wt* and
614 *ibm1* plants. Genes that have more than 5% mCHH in *ibm1* or less than 5% mCHH in *ibm1*
615 are plotted separately.

616 **Figure 4** SHH1 is not required for non-CG hypermethylation in *hl*.

617 (A) Heatmaps of mCHH and sRNA expression at CMT and DRM TEs in plants with *shh1* or

618 *clsy* mutations. (B) Venn diagram of TEs in indicated categories.

619 **Figure 4-figure supplement 1** Loss of *hl* causes CMT2-independent hypermethylation at

620 TEs with CLSY3/4 sRNA clusters. Boxplots show mCHH at TEs within indicated CLSY

621 clusters.

622 **Figure 4-figure supplement 2** Heatmaps of *wt* sRNA expression at CMT, DRM TEs and

623 CLSY1/2, CLSY3/4 sRNA clusters in leaves and flowers (Zhou et al., 2018).

624 **Figure 5** sRNA expression at CLSY3/4 clusters is independent of mCG.

625 (A) The number of CMT TEs ($mCHH \geq 0.05$ in *wt*) that maintain mCHH in *met1*

626 ($mCHH \geq 0.05$ in *met1*; 2498) or lose mCHH in *met1* ($mCHH < 0.02$ in *met1*; 798). (B and C)

627 Averaged sRNA distribution and mCHH levels around CMT TEs in *wt*, *met1*, and *hlmet1*

628 (*hml*) plants that maintain mCHH in *met1* ($mCHH \geq 0.05$ in *met1*; B) and lose mCHH in *met1*

629 ($mCHH < 0.02$ in *met1*; C). (D) Boxplots of H3K9me2, DNA methylation, and sRNA

630 expression at CLSY3/4 sRNA clusters in *wt* and *met1* (*m1*). CLSY3/4 clusters that maintain

631 more than 5% mCHH in *met1* or less than 5% mCHH in *met1* are plotted separately. (E)

632 sRNA expression level at CLSY3/4 sRNA clusters that maintain non-CG methylation

633 ($mCH > 0.01$) in *met1* or lose non-CG methylation ($mCH < 0.005$) in *met1*. Non-CG

634 methylation (mCH) density equals number of mCH sites per base pair.

635 **Figure 5-figure supplement 1** DNA methylation and sRNA expression changes in *met1* and

636 *hlmet1*.

(A) H3K9me2 levels around CMT TEs that maintain mCHH ($mCHH \geq 0.05$) in *met1* and lose mCHH ($mCHH < 0.02$) in *met1*. (B) Kernel density distribution of sRNA expression changes at CLSY3/4 sRNA clusters in *met1* (vs. *wt*) that maintain more than 5% mCHH in *met1* or less than 5% mCHH in *met1*. (C) Heatmaps of DNA methylation and sRNA expression at CLSY3/4 clusters in *wt*, *met1* (*m1*), and *h1met1* (*hm1*). TEs were sorted by non-CG methylation level (mCH) in *met1*.

Figure 6 Severely reduced H3K9 methylation does not prevent sRNA expansion.

(A-C) Distribution of H3K9 methylation (A and B) and sRNA expression (C) around 5' ends of CMT and DRM TEs in *wt*, *h1*, *ddm1*, and *h1ddm1* plants. (D) Averaged sRNA distribution around CMT and intermediate TEs in *wt*, *h1*, *ddm1*, and *h1ddm1* plants.

Figure 6-figure supplement 1 sRNA expression and H3K9 methylation changes in *h1*, *ddm1* and *h1ddm1*.

(A and B) Boxplots of H3K9 methylation levels (A) and sRNA expression (B) at CMT TEs in *wt*, *h1*, *ddm1*, and *h1ddm1* plants.

Figure 7 Non-CG DNA methylation and H3K9 methylation are decoupled in *h1c2c3*.

(A) Number of CMT TEs with H3K9 methylation (average H3K9me1 (K9me1) or H3K9me2 (K9me2) > 0.5) in *cmt2cmt3* (*c2c3*) or *h1cmt2cmt3* (*h1c2c3*) plants. (B) Example of DNA methylation, CG and CCG density, H1 level, H3K9 methylation and sRNA expression around CMT TEs in *c2c3* and *h1c2c3* plants (Chr3:14,495,000-14,520,000). (C) Principal component analysis of H3K9me, cytosine content (total GC content, CG, CCG, CHG, CHH, CWG, CWA, and nonCWA (W=A and T)), DNA methylation, and sRNA expression in *c2c3*

658 and *hlc2c3* plants. **(D and E)** H3K9me levels at CMT TEs with high H3K9me1 ($H3K9me1 \geq$
659 0.5; **D**) or low H3K9me1 ($H3K9me1 < 0.5$; **E**) in *hlc2c3* plants.

660 **Figure 7-figure supplement 1** Analysis of chromatin features at CMT TEs.

661 **(A)** Boxplots of DNA methylation and H3K9 methylation levels at CMT TEs in *wt*,
662 *cmt2cmt3* (*c2c3*), and *hlcmt2cmt3* (*hlc2c3*) plants. **(B)** Cytosine context density (total GC
663 content, CG, CCG, CWG, CWA, and nonCWA (W=A and T)) at CMT TEs with high
664 H3K9me1 ($H3K9me1 \geq 0.5$; H) and low H3K9me1 ($H3K9me1 < 0.5$; L) in *hlc2c3* plants.

665 **Figure 8** sRNA expression specifically correlates with non-CG methylation.

666 **(A and C)** Boxplots of H3K9me2, mCHG, and sRNA expression changes in *hl* vs. *wt* **(A)**
667 and *hlc2c3* vs. *c2c3* **(C)**. **(B and D)** Correlation among H3K9 methylation, DNA methylation,
668 and sRNA expression in *hl* plants **(B)** and *hlc2c3* plants **(D)**. **(E and F)** sRNA expression
669 relation to H3K9me1, CG and non-CG methylation density in *hl* plants **(E)** and *hlc2c3*
670 plants **(F)**. Each dot represents the average of 100 TEs sorted by GC content. DNA
671 methylation density equals number of methylated sites per base pair. **(G)** Average sRNA
672 expression level of CMT TEs in *wt*, *hl*, *c2c3* and *hlc2c3* plants. **(H)** Example of DNA
673 methylation, sRNA expression, H3K9 methylation (K9me1 and K9me2), and H1.1
674 distribution at CMT TEs in *wt*, *hl*, and *hlc2c3* plants (Chr2:6,548,000-6,559,000).

675 **Figure 8-figure supplement 1** H3K9 methylation, DNA methylation and sRNA expression
676 in *hl* and *hlc2c3* plants. **(A and C)** Boxplots of H3K9me1, mCG and mCHH change in *hl* vs.
677 *wt* **(A)** and *hlc2c3* vs. *c2c3* **(C)**. **(B and D)** Boxplots of H3K9 methylation, DNA methylation,
678 and sRNA expression levels at CMT and DRM TEs in *hl* plants **(B)** and *hlc2c3* plants **(D)**.
679 **(E and F)** Average sRNA expression level of intermediate TEs **(E)** and DRM TEs **(F)** in *wt*,

680 *h1*, *c2c3* and *h1c2c3* plants. **(G)** Boxplots of sRNA expression and mCHH levels at DRM
681 TEs in *wt*, *h1*, *c2c3*, and *h1c2c3* plants.

682 **Figure 9** CLSY1/2-dependent sRNA expression is associated with non-CG methylation.
683 **(A)** Boxplots of H3K9me1, H3K9me2, mCHH and sRNA expression levels at CLSY1/2-
684 dependent sRNA clusters in *h1c2c3* plants. sRNA clusters were classified by H3K9
685 methylation level (H3K9me1>0.5, H3K9me2>0 as high H3K9me (H), and the rest as low
686 H3K9me (L)) and mCHH level (mCHH>0.05 as high mCHH (H) and the rest as low mCHH
687 (L)). **(B)** Boxplots of H3K9me2, mCHH and sRNA expression levels at CLSY1/2-dependent
688 sRNA clusters in *suvh4/5/6* plants. sRNA clusters were classified by H3K9me2 level
689 (H3K9me2>0 as high H3K9me2 (H) and the rest as low H3K9me2 (L)) and mCHH level as
690 in (A). **(C)** Examples of CLSY1/2 sRNA clusters with high H3K9me2 in *suvh4/5/6* but
691 different non-CG methylation levels (Chr1:17,520,000-17,538,000). **(D)** Correlation among
692 H3K9me2, DNA methylation, and sRNA expression in *suvh456* plants. **(E)** Overlap between
693 H3K9me2 low/mCHH high CLSY1/2 clusters (LH) in *suvh4/5/6* plants and mCHH
694 demethylated CLSY1/2 clusters in *polv* plants. **(F-H)** Boxplots of non-CG methylation levels
695 **(F)**, sRNA expression **(G)**, and mCG levels **(H)** at 597 CLSY1/2 clusters that lose mCHH in
696 *polv* (blue in panel E).

697 **Figure 10** Histone H1 prevents non-CG methylation-mediated small RNA biogenesis in
698 *Arabidopsis* heterochromatin.

699 **(A)** In *wt* plants, H1 binds to GC-rich CMT TEs to restrict access of RNA polymerase IV
700 (pol IV). Pol IV binds to DRM TEs and produces sRNA. **(B)** In *h1* plants, RNA pol IV can
701 transcribe non-CG methylated CMT TEs to produce 24-nt sRNA, which leads to DNA
702 methylation of CMT TEs and reduced activity at DRM TEs.

References

- Allshire RC, Madhani HD. 2018. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* **19**:229–244. doi:10.1038/nrm.2017.119
- Angel A, Song J, Dean C, Howard M. 2011. A Polycomb-based switch underlying quantitative epigenetic memory. *Nature* **476**:105–108. doi:10.1038/nature10241
- Berry S, Dean C. 2015. Environmental perception and epigenetic memory: Mechanistic insight through FLC. *Plant J* **83**:133–148. doi:10.1111/tpj.12869
- Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, Tang H, Pikaard CS. 2015. Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. *Elife* **4**:e09591. doi:10.7554/eLife.09591
- Bourguet P, Picard CL, Yelagandula R, Pélissier T, Lorković ZJ, Feng S, Pouch-Pélissier M-N, Schmücker A, Jacobsen SE, Berger F, Mathieu O. 2021. The histone variant H2A.W and linker histone H1 co-regulate heterochromatin accessibility and DNA methylation. *Nat Commun* **12**:2683. doi:10.1038/s41467-021-22993-5
- Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, Mager DL, Feschotte C. 2018. Ten things you should know about transposable elements. *Genome Biol* **19**:199. doi:10.1186/s13059-018-1577-z
- Breiman L. 2001. Random forests. *Mach Learn* **45**:5–32. doi:10.1023/A:1010933404324
- Cao K, Lailler N, Zhang Y, Kumar A, Uppal K, Liu Z, Lee EK, Wu H, Medrzycki M, Pan C, Ho P-Y, Cooper GP, Dong X, Bock C, Bouhassira EE, Fan Y. 2013. High-resolution mapping of H1 linker histone variants in embryonic stem cells. *PLoS Genet* **9**:e1003417. doi:10.1371/journal.pgen.1003417
- Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. 2017. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant*

728 *J* **89**:789–804. doi:10.1111/tpj.13415

729 Choi J, Lyons DB, Kim MY, Moore JD, Zilberman D. 2020. DNA methylation and histone
730 H1 jointly repress transposable elements and aberrant intragenic transcripts. *Mol Cell*
731 **77**:310–323.e7. doi:10.1016/j.molcel.2019.10.011

732 Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF,
733 Pellegrini M, Jacobsen SE. 2008. Shotgun bisulphite sequencing of the Arabidopsis
734 genome reveals DNA methylation patterning. *Nature* **452**:215–219.
735 doi:10.1038/nature06745

736 Davey C, Pennings S, Allan J. 1997. CpG methylation remodels chromatin structure in vitro.
737 *J Mol Biol* **267**:276–288. doi:10.1006/JMBI.1997.0899

738 de Hoon MJL, Imoto S, Nolan J, Miyano S. 2004. Open source clustering software.
739 *Bioinformatics* **20**:1453–1454. doi:10.1093/bioinformatics/bth078

740 Deleris A, Stroud H, Bernatavichute Y, Johnson E, Klein G, Schubert D, Jacobsen SE. 2012.
741 Loss of the DNA methyltransferase MET1 induces H3K9 hypermethylation at PcG
742 target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis
743 thaliana. *PLoS Genet* **8**:e1003062. doi:10.1371/journal.pgen.1003062

744 Du J, Johnson LM, Groth M, Feng S, Hale CJ, Li S, Vashisht AA, Gallego-Bartolome J,
745 Wohlschlegel JA, Patel DJ, Jacobsen SE. 2014. Mechanism of DNA methylation-
746 directed histone methylation by KRYPTONITE. *Mol Cell* **55**:495–504.
747 doi:10.1016/j.molcel.2014.06.009

748 Du J, Zhong X, Bernatavichute Y V., Stroud H, Feng S, Caro E, Vashisht AA, Terragni J,
749 Chin HG, Tu A, Hetzel J, Wohlschlegel JA, Pradhan S, Patel DJ, Jacobsen SE. 2012.
750 Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes
751 directs DNA methylation in plants. *Cell* **151**:167–180. doi:10.1016/j.cell.2012.07.034

752 Erdmann RM, Picard CL. 2020. RNA-directed DNA methylation, *PLoS Genet.* **16**, e1009034.

doi:10.1371/journal.pgen.1009034

Fan D, Dai Y, Wang X, Wang Z, He H, Yang H, Cao Y, Deng XW, Ma L. 2012. IBM1, a JmjC domain-containing histone demethylase, is involved in the regulation of RNA-directed DNA methylation through the epigenetic control of RDR2 and DCL3 expression in Arabidopsis. *Nucleic Acids Res* **40**:8905–8916. doi:10.1093/nar/gks647

Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, Ukomadu C, Sadler KC, Pradhan S, Pellegrini M, Jacobsen SE. 2010. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci USA* **107**:8689–8694. doi:10.1073/pnas.1002720107

Feng W, Michaels SD. 2015. Accessing the inaccessible: The organization, transcription, replication, and repair of heterochromatin in plants. *Annu Rev Genet* **49**:439–459. doi:10.1146/annurev-genet-112414-055048

Friendly M. 2002. Corrgrams: Exploratory displays for correlation matrices. *Am Stat* **56**:316–324. doi:10.1198/000313002533

Gouil Q, Baulcombe DC. 2016. DNA methylation signatures of the plant chromomethyltransferases. *PLoS Genet* **12**:e1006526. doi:10.1371/journal.pgen.1006526

Greenberg MVC, Deleris A, Hale CJ, Liu A, Feng S, Jacobsen SE. 2013. Interplay between active chromatin marks and RNA-directed DNA methylation in Arabidopsis thaliana. *PLoS Genet* **9**:e1003946. doi:10.1371/journal.pgen.1003946

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**:118–120. doi:10.1126/science.1106910

Hulsen T, de Vlieg J, Alkema W. 2008. BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**:488. doi:10.1186/1471-2164-9-488

778 Inagaki S, Miura-Kamio A, Nakamura Y, Lu F, Cui X, Cao X, Kimura H, Saze H, Kakutani
779 T. 2010. Autocatalytic differentiation of epigenetic modifications within the Arabidopsis
780 genome. *EMBO J* **29**:3496–3506. doi:<https://doi.org/10.1038/emboj.2010.227>

781 Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. 2010. High-dimensional
782 variable selection for survival data. *J Am Stat Assoc* **105**:205–217.
783 doi:[10.1198/jasa.2009.tm08622](https://doi.org/10.1198/jasa.2009.tm08622)

784 Ito T, Tarutani Y, To TK, Kassam M, Duvernois-Berthet E, Cortijo S, Takashima K, Saze H,
785 Toyoda A, Fujiyama A, Colot V, Kakutani T. 2015. Genome-wide negative feedback
786 drives transgenerational DNA methylation dynamics in Arabidopsis. *PLoS Genet*
787 **11**:e1005154. doi:[10.1371/journal.pgen.1005154](https://doi.org/10.1371/journal.pgen.1005154)

788 Izaurrealde E, Käs E, Laemmli UK. 1989. Highly preferential nucleation of histone H1
789 assembly on scaffold-associated regions. *J Mol Biol* **210**:573–585.
790 doi:[https://doi.org/10.1016/0022-2836\(89\)90133-2](https://doi.org/10.1016/0022-2836(89)90133-2)

791 Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE. 2007. The
792 SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr Biol*
793 **17**:379–384. doi:<https://doi.org/10.1016/j.cub.2007.01.009>

794 Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, Zhong X, Marson G,
795 Pellegrini M, Segal DJ, Patel DJ, Jacobsen SE. 2014. SRA- and SET-domain-containing
796 proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507**:124–128.
797 doi:[10.1038/nature12931](https://doi.org/10.1038/nature12931)

798 Kim MY, Zilberman D. 2014. DNA methylation as a system of plant genomic immunity.
799 *Trends Plant Sci* **19**:320–326. doi:[10.1016/j.tplants.2014.01.014](https://doi.org/10.1016/j.tplants.2014.01.014)

800 Lai Y, Lu XM, Daron J, Pan S, Wang J, Wang W, Tsuchiya T, Holub E, McDowell JM,
801 Keith Slotkin R, Le Roch KG, Eulgem T. 2020. The arabidopsis PHD-finger protein
802 EDM2 has multiple roles in balancing NLR immune receptor gene expression, *PLoS*

803 *Genet* **16**:e1008993. doi:10.1371/journal.pgen.1008993

804 Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient
805 alignment of short DNA sequences to the human genome. *Genome Biol* **10**:R25.
806 doi:10.1186/gb-2009-10-3-r25

807 Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, Strahl BD, Patel DJ, Jacobsen
808 SE. 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires
809 SHH1. *Nature* **498**:385–389. doi:10.1038/nature12178

810 Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation
811 patterns in plants and animals. *Nat Rev Genet* **11**:204–220. doi:10.1038/nrg2719

812 Li J, Yang D-L, Huang H, Zhang G, He L, Pang J, Lozano-Durán R, Lang Z, Zhu J-K. 2020.
813 Epigenetic memory marks determine epiallele stability at loci targeted by de novo DNA
814 methylation. *Nat Plants* **6**:661–674. doi:10.1038/s41477-020-0671-x

815 Li X, Harris CJ, Zhong Z, Chen W, Liu R, Jia B, Wang Z, Li S, Jacobsen SE, Du J. 2018.
816 Mechanistic insights into plant SUVH family H3K9 methyltransferases and their
817 binding to context-biased non-CG DNA methylation. *Proc Natl Acad Sci USA*
818 **115**:E8793–E8802. doi:10.1073/pnas.1809841115

819 Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, Richard McCombie W, Lavine
820 K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R.
821 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature*
822 **430**:471–476. doi:10.1038/nature02651

823 Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008.
824 Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*
825 **133**:523–536. doi:10.1016/J.CELL.2008.03.029

826 Liu Z-W, Shao C-R, Zhang C-J, Zhou J-X, Zhang S-W, Li L, Chen S, Huang H-W, Cai T, He
827 X-J. 2014. The SET domain proteins SUVH2 and SUVH9 are required for Pol V

occupancy at RNA-directed DNA methylation loci. *PLoS Genet* **10**:e1003948.
doi:10.1371/journal.pgen.1003948

Long J, Walker J, She W, Aldridge B, Gao H, Deans S, Vickers M, Feng X. 2021. Nurse
cell-derived small RNAs define paternal epigenetic inheritance in Arabidopsis. *Science*
373:eabh0556. doi:10.1126/science.abh0556

Lövkvist C, Howard M. 2021. Using computational modelling to reveal mechanisms of
epigenetic Polycomb control. *Biochem Soc Trans* **49**:71–77. doi:10.1042/BST20190955

Lyons DB, Zilberman D. 2017. DDM1 and Lsh remodelers allow methylation of DNA
wrapped in nucleosomes. *Elife* **6**:e30674. doi:10.7554/eLife.30674

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
reads. *EMBnet.journal* **17**:10–12. doi:10.14806/ej.17.1.200

Matzke MA, Mosher RA. 2014. RNA-directed DNA methylation: an epigenetic pathway of
increasing complexity. *Nat Rev Genet* **15**:394–408. doi:10.1038/nrg3683

Michael TP. 2014. Plant genome size variation: bloating and purging DNA. *Brief Funct
Genomics* **13**:308–317. doi:10.1093/bfpg/elu005

Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, Kakutani T. 2009. An Arabidopsis
jmjC domain protein protects transcribed genes from DNA methylation at CHG sites.
EMBO J **28**:1078–1086. doi:10.1038/emboj.2009.59

Murdoch DJ, Chow ED. 1996. A Graphical Display of Large Correlation Matrices. *Am Stat*
50:178–180. doi:10.1080/00031305.1996.10474371

Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim K Do, Li Q, Rohr NA, Rambani A,
Burke JM, Udall JA, Egesi C, Schmutz J, Grimwood J, Jackson SA, Springer NM,
Schmitz RJ. 2016. Widespread natural variation of DNA methylation within
angiosperms. *Genome Biol* **17**:194. doi:10.1186/s13059-016-1059-0

Numa H, Yamaguchi K, Shigenobu S, Habu Y. 2015. Gene body CG and CHG methylation

853 and suppression of centromeric CHH methylation are mediated by DECREASE IN
854 DNA METHYLATION¹ in rice. *Mol Plant* **8**:1560–1562.
855 doi:10.1016/j.molp.2015.08.002

856 Osakabe A, Jamge B, Axelsson E, Montgomery SA, Akimcheva S, Kuehn AL, Pisupati R,
857 Lorković ZJ, Yelagandula R, Kakutani T, Berger F. 2021. The chromatin remodeler
858 DDM1 prevents transposon mobility through deposition of histone variant H2A.W. *Nat*
859 *Cell Biol* **23**:391–400. doi:10.1038/s41556-021-00658-1

860 Panda K, Ji L, Neumann DA, Daron J, Schmitz RJ, Slotkin RK. 2016. Full-length
861 autonomous transposable elements are preferentially targeted by expression-dependent
862 forms of RNA-directed DNA methylation. *Genome Biol* **17**:170. doi:10.1186/s13059-
863 016-1032-y

864 Panda K, Slotkin RK. 2020. Long-read cDNA sequencing enables a “gene-like” transcript
865 annotation of transposable elements. *Plant Cell* **32**:2687–2698.
866 doi:10.1105/tpc.20.00115

867 Papareddy RK, Páldi K, Paulraj S, Kao P, Lutzmayer S, Nodine MD. 2020. Chromatin
868 regulates expression of small RNAs to help maintain transposon methylome
869 homeostasis in Arabidopsis. *Genome Biol* **21**:251. doi:10.1186/s13059-020-02163-4

870 Rajakumara E, Law JA, Simanshu DK, Voigt P, Johnson LM, Reinberg D, Patel DJ,
871 Jacobsen SE. 2011. A dual flip-out mechanism for 5mC recognition by the Arabidopsis
872 SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in
873 vivo. *Genes Dev* **25**:137–152. doi:10.1101/gad.1980311

874 Raju SKK, Ritter EJ, Niederhuth CE. 2019. Establishment, maintenance, and biological roles
875 of non-CG methylation in plants. *Essays Biochem* **63**:743–755.
876 doi:10.1042/EBC20190032

877 Ramachandran S, Henikoff S. 2015. Replicating nucleosomes. *Sci Adv* **1**:e1500587.

doi:10.1126/sciadv.1500587

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**:W160–W165. doi:10.1093/nar/gkw257

Ritter EJ, Niederhuth CE. 2021. Intertwined evolution of plant epigenomes and genomes. *Curr Opin Plant Biol* **61**:101990. doi:https://doi.org/10.1016/j.pbi.2020.101990

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**:24–26. doi:10.1038/nbt.1754

Rougée M, Quadrana L, Zervudacki J, Hure V, Colot V, Navarro L, Deleris A. 2020. Polycomb mutant partially suppresses DNA hypomethylation-associated phenotypes in *Arabidopsis*. *Life Sci Alliance* **4**:e202000848. doi:10.26508/lsa.202000848

Rutowicz K, Puzio M, Halibart-Puzio J, Lirski M, Kotliński M, Kroteń MA, Knizewski L, Lange B, Muszewska A, Śniegowska-Świerk K, Kościelniak J, Iwanicka-Nowicka R, Buza K, Janowiak F, Żmuda K, Jöesaar I, Laskowska-Kaszub K, Fogtman A, Kollist H, Zielenkiewicz P, Tiuryn J, Siedlecki P, Swiezewski S, Ginalski K, Kobłowska M, Archacki R, Wilczynski B, Rapacz M, Jerzmanowski A. 2015. A specialized histone H1 variant is required for adaptive responses to complex abiotic stress and related DNA methylation in *Arabidopsis*. *Plant Physiol* **169**:2080–2101. doi:10.1104/pp.15.00493

Saze H, Shiraishi A, Miura A, Kakutani T. 2008. Control of Genic DNA Methylation by a jmjC Domain-Containing Protein in *Arabidopsis thaliana*. *Science* **319**:462–465. doi:10.1126/science.1150987

Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U, Gutierrez C. 2014. The functional topography of the *Arabidopsis* genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* **26**:2351–2366. doi:10.1105/tpc.114.124578

903 Shahzad Z, Moore JD, Zilberman D. 2021. Gene body methylation mediates epigenetic
 904 inheritance of plant traits. *bioRxiv* 2021.03.15.435374. doi:10.1101/2021.03.15.435374

905 Sigman MJ, Slotkin RK. 2016. The first rule of plant transposable element silencing:
 906 Location, location, location. *Plant Cell* **28**:304–313. doi:10.1105/tpc.15.00869

907 Singh J, Pikaard CS. 2019. Reconstitution of siRNA biogenesis in vitro: Novel reaction
 908 mechanisms and RNA channeling in the RNA-directed DNA methylation pathway. *Cold*
 909 *Spring Harb Symp Quant Biol* **84**:195–201. doi:10.1101/sqb.2019.84.039842

910 Smith LM, Pontes O, Searle I, Yelina N, Yousafzai FK, Herr AJ, Pikaard CS, Baulcombe DC.
 911 2007. An SNF2 protein associated with nuclear RNA silencing and the spread of a
 912 silencing signal between cells in Arabidopsis. *Plant Cell* **19**:1507–1521.
 913 doi:10.1105/tpc.107.051540

914 Soppe WJJ, Jasencakova Z, Houben A, Kakutani T, Meister A, Huang MS, Jacobsen SE,
 915 Schubert I, Fransz PF. 2002. DNA methylation controls histone H3 lysine 9 methylation
 916 and heterochromatin assembly in Arabidopsis. *EMBO J* **21**:6549–6559.
 917 doi:10.1093/emboj/cdf657

918 Stoddard CI, Feng S, Campbell MG, Liu W, Wang H, Zhong X, Bernatavichute Y, Cheng Y,
 919 Jacobsen SE, Narlikar GJ. 2019. A nucleosome bridging mechanism for activation of a
 920 maintenance DNA methyltransferase. *Mol Cell* **73**:73–83.e6.
 921 doi:10.1016/J.MOLCEL.2018.10.006

922 Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE. 2014. Non-CG
 923 methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol*
 924 **21**:64–72. doi:10.1038/nsmb.2735

925 Suzuki MM, Bird A. 2008. DNA methylation landscapes: Provocative insights from
 926 epigenomics. *Nat Rev Genet* **9**:465–476. doi:10.1038/nrg2341

927 Tan F, Lu Y, Jiang W, Wu T, Zhang R, Zhao Y, Zhou D-X. 2018. DDM1 represses

928 noncoding RNA expression and RNA-directed DNA methylation in heterochromatin.
 929 *Plant Physiol* **177**:1187–1197. doi:10.1104/pp.18.00352
 930 Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, Cruaud C, Poulain J,
 931 Berdasco M, Fraga MF, Voinnet O, Wincker P, Esteller M, Colot V. 2009. A role for
 932 RNAi in the selective correction of DNA methylation defects. *Science* **323**:1600–1604.
 933 doi:10.1126/science.1165313
 934 Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV):
 935 High-performance genomics data visualization and exploration. *Brief Bioinform* **14**:178–
 936 192. doi:10.1093/bib/bbs017
 937 To TK, Nishizawa Y, Inagaki S, Tarutani Y, Tominaga S, Toyoda A, Fujiyama A, Berger F,
 938 Kakutani T. 2020. RNA interference-independent reprogramming of DNA methylation
 939 in Arabidopsis. *Nat Plants* **6**:1455–1467. doi:10.1038/s41477-020-00810-z
 940 Tomaszewski R, Jerzmanowski A. 1997. The AT-rich flanks of the oocyte-type 5S RNA
 941 gene of *Xenopus laevis* act as a strong local signal for histone H1-mediated chromatin
 942 reorganization in vitro. *Nucleic Acids Res* **25**:458–465. doi:10.1093/nar/25.3.458
 943 Venables WN, Ripley BD. 2002. Modern applied statistics with S. Springer.
 944 Wendte JM, Pikaard CS. 2017. The RNAs of RNA-directed DNA methylation. *Biochim*
 945 *Biophys Acta - Gene Regul Mech* **1860**:140–148.
 946 doi:https://doi.org/10.1016/j.bbagrm.2016.08.004
 947 Wickham H. 2009. Ggplot2 : Elegant graphics for data analysis. New York: Springer-Verlag.
 948 Wongpalee SP, Liu S, Gallego-Bartolomé J, Leitner A, Aebersold R, Liu W, Yen L, Nohales
 949 MA, Kuo PH, Vashisht AA, Wohlschlegel JA, Feng S, Kay SA, Zhou ZH, Jacobsen SE.
 950 2019. CryoEM structures of Arabidopsis DDR complexes involved in RNA-directed
 951 DNA methylation. *Nat Commun* **10**:3916. doi:10.1038/s41467-019-11759-9
 952 Xu L, Jiang H. 2020. Writing and reading histone H3 lysine 9 methylation in Arabidopsis.

953 *Front. Plant Sci.* **11**:452. doi: 10.3389/fpls.2020.00452

954 Yang D-L, Zhang G, Wang L, Li J, Xu D, Di C, Tang K, Yang L, Zeng L, Miki D, Duan C-G,
 955 Zhang H, Zhu J-K. 2018. Four putative SWI2/SNF2 chromatin remodelers have dual
 956 roles in regulating DNA methylation in Arabidopsis. *Cell Discov* **4**:55.
 957 doi:10.1038/s41421-018-0056-8

958 Zabet NR, Catoni M, Prischi F, Paszkowski J. 2017. Cytosine methylation at CpCpG sites
 959 triggers accumulation of non-CpG methylation in gene bodies. *Nucleic Acids Res*
 960 **45**:3777–3784. doi:10.1093/nar/gkw1330

961 Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL,
 962 Zilberman D. 2013. The Arabidopsis nucleosome remodeler DDM1 allows DNA
 963 methyltransferases to access H1-containing heterochromatin. *Cell* **153**:193–205.
 964 doi:10.1016/j.cell.2013.02.033

965 Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of
 966 eukaryotic DNA methylation. *Science* **328**:916–919. doi:10.1126/science.1159151

967 Zhai J, Bischof S, Wang H, Feng S, Lee T-F, Teng C, Chen X, Park SY, Liu L, Gallego-
 968 Bartolome J, Liu W, Henderson IR, Meyers BC, Ausin I, Jacobsen SE. 2015. A one
 969 precursor one siRNA model for pol IV-dependent siRNA biogenesis. *Cell* **163**:445–55.
 970 doi:10.1016/j.cell.2015.09.032

971 Zhang H, Lang Z, Zhu J-K. 2018. Dynamics and function of DNA methylation in plants. *Nat*
 972 *Rev Mol Cell Biol* **19**:489–506. doi:10.1038/s41580-018-0016-z

973 Zhang H, Ma ZY, Zeng L, Tanaka K, Zhang CJ, Ma J, Bai G, Wang P, Zhang SW, Liu ZW,
 974 Cai T, Tang K, Liu R, Shi X, He XJ, Zhu JK. 2013. DTF1 is a core component of RNA-
 975 directed DNA methylation and may assist in the recruitment of Pol IV. *Proc Natl Acad*
 976 *Sci USA* **110**:8290–8295. doi:10.1073/pnas.1300585110

977 Zhang Y, Harris CJ, Liu Q, Liu W, Ausin I, Long Y, Xiao L, Feng L, Chen Xu, Xie Y, Chen

978 Xinyuan, Zhan L, Feng S, Li JJ, Wang H, Zhai J, Jacobsen SE. 2018. Large-scale
 979 comparative epigenomics reveals hierarchical regulation of non-CG methylation in
 980 Arabidopsis. *Proc Natl Acad Sci USA* E1069–E1074. doi:10.1073/pnas.1716300115
 981 Zhong X, Hale CJ, Law JA, Johnson LM, Feng S, Tu A, Jacobsen SE. 2012. DDR complex
 982 facilitates global association of RNA polymerase V to promoters and evolutionarily
 983 young transposons. *Nat Struct Mol Biol* **19**:870–875. doi:10.1038/nsmb.2354
 984 Zhou M, Coruh C, Xu G, Bourbousse C, Lambolez A, Law JA. 2021. The CLASSY family
 985 controls tissue-specific DNA methylation patterns in Arabidopsis. *bioRxiv*
 986 2021.01.23.427869. doi:10.1101/2021.01.23.427869
 987 Zhou M, Palanca AMS, Law JA. 2018. Locus-specific control of the de novo DNA
 988 methylation pathway in Arabidopsis by the CLASSY family. *Nat Genet* **50**:865–873.
 989 doi:10.1038/s41588-018-0115-y
 990 Zilberman D, Henikoff S. 2004. Silencing of transposons in plant genomes: Kick them when
 991 they're down. *Genome Biol* **5**:249. doi:10.1186/gb-2004-5-12-249

Figure 1

Figure 2

Figure 3

A

Figure 5

Figure 6

Figure 7

Figure 8

Figure 9

Figure 10

Figure 1-figure supplement 1

Figure 2-figure supplement 1

Figure 4-figure supplement 1

Figure 4-figure supplement 2

Figure 5-figure supplement 1

Figure 6-figure supplement 1

Figure 7-figure supplement 1

Figure 8-figure supplement 1