Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

*Nat Struct Mol Biol*. Author manuscript; available in PMC 2014 July 18.

Published in final edited form as:

*Nat Struct Mol Biol*. 2014 January ; 21(1): 64–72. doi:10.1038/nsmb.2735.

# The roles of non-CG methylation in Arabidopsis

**Hume Stroud**[1], **Truman Do**[1], **Jiamu Du**[2], **Xuehua Zhong**[1], **Suhua Feng**[1,3,4], **Lianna Johnson**[1], **Dinshaw J. Patel**[2], and **Steven E. Jacobsen**[1,3,4]

[1] Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, California, USA.

[2] Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York, USA.

[3] Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, Los Angeles, California, USA.

[4] Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, California, USA.

## Abstract

DNA methylation occurs in CG and non-CG sequence contexts. Non-CG methylation is abundant in plants, and is mediated by *CHROMOMETHYLASE* (CMT) and *DOMAINS REARRANGED METHYLTRANSFERASE* (DRM) proteins; however its roles remain poorly understood. Here we characterize the roles of non-CG methylation in *Arabidopsis thaliana*. We show that a poorly characterized methyltransferase, CMT2, is a functional methyltransferase *in vitro* and *in vivo*. CMT2 preferentially binds histone H3 lysine 9 (H3K9) dimethylation and methylates non-CG cytosines that are regulated by H3K9 methylation. We revealed the contributions and redundancies between each non-CG methyltransferase in DNA methylation patterning and in regulating transcription. We also demonstrate extensive dependencies of small RNA accumulation and H3K9 methylation patterning on non-CG methylation, suggesting self-reinforcing mechanisms between these epigenetic factors. The results suggest that non-CG methylation patterns are critical in shaping the histone modification and small non-coding RNA landscapes.

## INTRODUCTION

DNA methylation plays roles in different biological processes such as gene regulation and imprinting. In *Arabidopsis thaliana*, DNA is methylated in three cytosine contexts: CG,

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence should be addressed to S.E.J. (jacobsen@ucla.edu).

AUTHOR CONTRIBUTIONS

H.S., T.D., J.D., X.Z., S.F. and L.J. performed the experiments. S.E.J. and D.J.P. oversaw the study. H.S. designed the study, analyzed data, and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Accession codes.

All sequencing data have been deposited in GEO with accession GSE51304.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

CHG, and CHH (where H=A, T, or C)[1]. In mammals, DNA is primarily methylated in CG contexts, however, studies have uncovered the presence of non-CG methylation in certain cell types such as embryonic stem cells and brains cells[2-7]. In Arabidopsis, CG methylation is maintained by MET1, the plant homolog of DNMT1. CHG and CHH methylation are site-specifically methylated by CMT3 and DRM2[8,9]. CMT3 is controlled by histone H3 lysine 9 (H3K9) methylation[10-12]. DRM2 is targeted to certain loci through an RNA-directed DNA methylation (RdDM) pathway involving 24-nucleotide small interfering RNAs (24nt-siRNAs)[1]. Heterochromatin in Arabidopsis is enriched in both CG and non-CG methylations as well as H3K9 methylation and 24nt-siRNAs, however the relationships between each of these marks remain poorly understood.

The abundant non-CG methylation in plants compared to mammals may in part be explained by the presence of plant specific CMT genes. In addition to CMT3, the Arabidopsis genome encodes two other CMT genes: CMT1 and CMT2. CMT1 is expressed at low levels and is truncated in many Arabidopsis ecotypes[13]. CMT2 is expressed and is a putative DNA methyltransferase. A recent study performed whole-genome methylation profiling in *cmt2* mutants and found loss of CHH methylation predominantly at large TEs that were heterochromatic[9]. Genetic evidence suggested that the chromatin remodeler DDM1 in part allows access for MET1, CMT3, and CMT2 to heterochromatin[9]. However, the mechanism of CMT2 targeting to heterochromatin, the roles it plays, and its relationship with other DNA methyltransferases is not understood.

Here, we set out to characterize the roles of non-CG methylation. We first show that CMT2 is a functional non-CG methyltransferase. CMT2 preferentially methylates unmethylated DNA *in vitro*, and methylates both CHG and CHH sites *in vitro* and *in vivo*. We find that CMT2 binds H3K9 methylation *in vitro* and that H3K9 methylation controls non-CG methylation through CMT2. We also uncover that the number of methyl groups on H3K9 may influence CMT2 and CMT3 targeting. Given the identification of CMT2 as a functional methyltransferase, we generated all possible combinations of non-CG methyltransferase mutants, and examined the contributions and redundancies between each non-CG methyltransferase in DNA methylation patterning and gene silencing. While it is clear that 24nt-siRNAs and H3K9 methylation guide non-CG methylation, we reveal extensive dependencies of both 24nt-siRNAs and H3K9 methylation patterning on non-CG methylation. This suggests that non-CG methylation plays a critical role in regulating these marks. Furthermore, we find elevated histone acetylation levels throughout sites that lose non-CG methylation. Our results provide insights into non-CG methylation targeting and will help to guide further studies of the biology of DNA methylation.

## RESULTS

### CMT2 strongly methylates both CHG and CHH sites *in vitro*

To examine whether CMT2 plays a role in methylating the genome, we performed whole genome bisulfite sequencing (BS-seq) in two different CMT2 T-DNA insertion mutants, *cmt2-7* and *cmt2-3*[8]. We found that global CHH methylation is substantially reduced, whereas CG and CHG methylation were largely undisturbed (Fig. 1a), consistent with a recent study[9]. For the rest of the study we focused in *cmt2-7*, which we confirmed to be a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

null mutant by RT-PCR (Supplementary Fig. 1a). In contrast to *cmt2* mutants, *cmt3* mutants lost CHG methylation globally but only affected CHH methylation at limited sites in the genome[8]. Thus CMT2 and CMT3 appear to have different sequence preferences.

To understand the difference between the sequence specificity between CMT2 and CMT3 we sought to examine CMT2 methyltransferase activity *in vitro*. To test if CMT2 could methylate DNA *in vitro*, we assayed whether CMT2 can methylate oligonucleotides of different methylation status. We used oligos that were unmethylated, oligos that were methylated in all sequences contexts on only one strand (hemimethylated), and as a negative control, oligos that were methylated in all sequence contexts in both strands (fully-methylated) (see Online Methods)[10]. We found that CMT2 preferentially methylated unmethylated oligos compared to hemimethylated oligos *in vitro* (Fig. 1b). This was in contrast to CMT3, which preferentially methylated hemimethylated oligos.[10] We further assayed sequence specificity of methylation by CMT2 and found that it did not methylate CG sites (Supplementary Fig. 1c). Rather, CMT2 strongly methylated both CHG and CHH sites (Fig. 1c). This was in contrast to CMT3, which substantially preferred to methylate CHG sites compared to CHH sites[10] (Supplementary Fig. 1b). Hence the methyltransferase activity of CMT2 is distinct from that of CMT3 such that it preferentially methylates unmethylated DNA and effectively methylates both CHG sites and CHH sites *in vitro*. These findings are consistent with our *in vivo* studies (see below) showing that CMT2 not only mediates CHH methylation but also mediates CHG methylation.

## CMT2 activity is mediated by H3K9 methylation

KRYPTONITE (KYP or SUVH4), SUVH5, and SUVH6 are the major H3K9 methyltransferases in Arabidopsis[11,12]. We previously showed that loss of CHG methylation in *kyp suvh5 suvh6* triple mutants mimicked the loss of CHG methylation in *cmt3* mutants genome-wide[8]. However, extensive loss of CHH methylation was also observed in *kyp suvh5 suvh6* but not in *cmt3*, suggesting that there must be another methyltransferase(s) methylating CHH sites[8]. About 86% of *kyp suvh5 suvh6* CHH hypomethylated sites overlapped with *cmt2* CHH hypomethylated sites, suggesting that H3K9 methylation regulates bulk CHH methylation through CMT2 (Fig. 2a and b). A smaller fraction of KYP SUVH5 SUVH6 regulated CHH sites overlapped with DRM2 target sites (Fig. 2a), which likely is explained by the dependency of Pol IV recruitment on H3K9 methylation through the histone binding protein SHH1[14,15]. We performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) on H3K9me2 in wild type and *kyp suvh5 suvh6* mutants, and confirmed that loss of CHH methylation in *kyp suvh5 suvh6* was associated with loss of H3K9me2 (Fig. 2b).

Structural and functional work has suggested that the BAH and chromo domains of CMT3 bind H3K9 methylation[10]. Because CMT2 and CMT3 proteins have very similar domain configurations (Supplementary Fig. 2a), we hypothesized that CMT2 may also recognize H3K9 methylation. To test this, we assayed binding of recombinant CMT2 protein to different histone modifications on a peptide array. Interestingly, we found preferential binding of CMT2 to H3K9 di- and trimethylated peptides (H3K9me2, H3K9me3), but less binding to H3K9 monomethylated (H3K9me1) peptides (Fig. 2c and Supplementary Fig.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2b), which was further confirmed by our ITC binding data (Fig. 2d). This data was in contrast to CMT3, which bound H3K9me1, -me2, and -me3 equally well (Fig. 2e)[10]. In addition, all the ITC binding curves yielded N values around 2, indicating that two histone tail peptides bind to each CMT molecule and that the dual recognition of methylated H3K9 tails is therefore likely to be a general feature of chromomethylase family DNA methyltransferases.

The sensitivity of CMT2 to number of methyl groups on H3K9 *in vitro* led us to investigate whether this property influences the sites that CMT2 and CMT3 are targeted. To test this, we performed ChIP-seq on H3K9me1 and compared to H3K9me2. We did not analyze H3K9me3 since this mark is present at extremely low levels[16] and is associated with active genes[17], which are devoid of non-CG methylation. We compared sites that are regulated by both CMT2 and CMT3 to sites that where regulated by CMT3 but not CMT2 (see Online Methods). At sites regulated by both CMT2 and CMT3, there were higher levels of H3K9me2 compared to sites methylated by CMT3 but not CMT2 (Fig. 2f). Hence CMT2 is preferentially associated with H3K9me2 whereas CMT3 does not show such preference. This supports our finding that CMT2 binds H3K9me2 with a substantial preference over H3K9me1, whereas CMT3 can bind both H3K9me1 and H3K9me2 almost equally (Fig. 2c-e)[10]. Our results indicate that the number of methyl groups on H3K9 may influence CMT protein targeting to the genome.

## Interplays between non-CG methyltransferases in methylation

The finding that CMT2 plays an important role in maintaining CHH methylation levels in the genome led us to generate mutants containing all possible combinations of non-CG methyltransferase mutants. We crossed *cmt2* to *cmt3* and to *drm1 drm2* double mutants (DRM1 is expressed only in female gametes[18]). We generated single nucleotide resolution maps of DNA methylation in the mutants by performing BS-seq. We first looked at non-CG methylation patterns over all TEs and chromosomes. We found that non-CG methylation in the genome was eliminated in *drm1 drm2 cmt2 cmt3* quadruple mutants (Fig. 3a, b and Supplementary Fig. 3a, b). This indicated that DRM1 DRM2, CMT2, and CMT3 are collectively responsible for all non-CG methylation in the Arabidopsis genome. This finding enabled us to determine the contributions of each non-CG methyltransferases in DNA methylation patterning. We observed that both CHG and CHH methylation are redundantly regulated by all non-CG methyltransferases to a certain extent (Fig. 3a-d). This suggests that different pathways cooperate to regulate non-CG methylation patterning.

## CMT2 and CMT3 methylate CHG sites in a redundant manner

CMT3 tends to methylate large TEs and sites distal to genes[8,9]. In *cmt3* mutants, a strong but partial loss of CHG methylation occurs[8,9] (Fig. 3a-d). We found that in *cmt2 cmt3* (*cmt2 cmt3*) double mutants there was stronger loss of CHG methylation than in *cmt3* mutants (Fig. 3c, e-g and Supplementary Fig. 3b). These sites were non-overlapping with DRM2 regulated sites (Fig. 3c). This suggests that while CMT2 preferentially methylates CHH sites, it also methylates CHG sites. This result is consistent with our finding that CMT2 can also methylate CHG sites *in vitro* (Fig. 1c). Hence while the main role of CMT2 is to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

methylate CHH sites, CMT2 and CMT3 function partially redundantly to methylate CHG sites.

### DRM2 target sites are methylated by both DRM2 and CMT3

DRM2 tends to methylate the edges of large TEs as well as small TEs that are proximal to genes [8,9]. In *drm1 drm2* mutants, loss of DNA methylation occurs in CHH contexts and to a lesser extent in CHG contexts[8] (Fig. 3f). This suggests that a different methyltransferase is methylating CHG at DRM2 target sites. In *cmt3* mutants, CHG methylation was partially reduced at DRM2 target sites, and in *drm1 drm2 cmt3* mutants CHG methylation was nearly completely lost (Fig. 3f). Hence CMT3 also methylates DRM2 sites. There was almost complete loss of non-CG methylation at DRM2 sites in *drm1 drm2 cmt3* mutants in the presence of a functional CMT2 (Fig. 3f). This suggests that CMT2 plays a very minor role at DRM2 target sites. Thus generally at DRM2 target sites, CMT3 and DRM2 methylate cytosines in CHG contexts, and DRM2 methylates cytosines in CHH contexts.

### CMT2 and DRM2 mediate all CHH methylation in the genome

Mutations in CMT2 or DRM2 alone are not sufficient to eliminate CHH methylation in the genome (Fig. 3a-g). However, we found that *drm1 drm2 cmt2* mutants essentially eliminated all CHH methylation in the genome (Fig. 3b, d-g). In fact, 99% of *drm1 drm2 cmt2 cmt3* CHH hypomethylated differentially methylated regions (DMRs) overlapped with *drm1 drm2 cmt2* CHH DMRs (Supplementary Fig. 3c). DRM2 and CMT2 methylate almost completely non-overlapping sites in the genome (Supplementary Fig. 3d). Hence a large proportion of heterochromatin can be divided into regions that are CMT2 targeted and those that are DRM2 targeted.

### Relative H3K9me1 and 2 levels at CMT2 and DRM2 target sites

Our finding of CMT2 binding preferentially to H3K9me2 led us to compare H3K9me1 and H3K9me2 levels at CMT2 target sites and DRM2 target sites. We found that the relative levels of H3K9me2 to H3K9me1 were higher at CMT2 target sites compared to DRM2 target sites (Fig. 3h). Furthermore, because DRM2 targets the edges of TEs[8,9], we sought to examine the distributions of H3K9me1 and H3K9me2 over TEs. We found that H3K9me1 was especially enriched at boundaries of TEs whereas H3K9me2 was enriched over the body of TEs (Fig. 3i). This H3K9me1/2 distribution was consistent with the distribution of sites methylated by DRM2 and CMT2 (Fig. 3i). These results are consistent with the fact that SHH1, a factor involved in recruiting RNA polymerase IV (Pol IV) to promote DRM2 targeting, exhibits similar *in vitro* binding to H3K9me1, -me2 and –me3 as observed for CMT3 (Fig. 2e)[10,14,15], whereas CMT2 preferably binds H3K9me2 (Fig. 2c, d). These results further suggest that the number of methyl groups on H3K9 may influence non-CG methyltransferase targeting.

### CMT2, CMT3, and DRM2 cooperatively regulate TE expression

DNA methylation is implicated in transcriptional regulation. Because for the first time we possessed a mutant with largely normal levels of CG methylation but a complete lack of non-CG methylation, we were able to test the extent to which non-CG methylation regulates

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

expression of TEs and genes. We performed mRNA sequencing (mRNA-seq) on the different combinations of non-CG methylation mutants (Supplementary Fig. 4a). We defined TE derepression by using stringent cutoffs (see Online Methods), and only selected TEs that showed significant misregulation in two biological replicates. TE derepression was most prominent in mutants containing *cmt3* mutations, suggesting that CMT3 plays the strongest role in transcriptional silencing of TEs among non-CG methyltransferases (Fig. 4a). We found relatively minor upregulation of TEs in *cmt2* mutants despite CMT2 methylating a substantial proportion of the genome (Fig. 1a). This together with the fact that *drm1 drm2* mutants alone or *drm1 drm2 cmt2* mutants showed modest TE derepression defects (Fig. 4a) suggest that CHH methylation itself may not play a major role in TE silencing. However, when combining *cmt3* mutations with *cmt2* or *drm1 drm2* mutations, we observed an increased number of TEs upregulated, suggesting that CHH and CHG methylation redundantly silence TEs (Fig. 4a). Notably, upon loss of all non-CG methylation in *drm1 drm2 cmt2 cmt3* mutants, there was a large increase in the number of TEs upregulated (Fig. 4a and Supplementary Fig. 4a). In fact, there was a global increase in RNA-seq reads in heterochromatic regions in *drm1 drm2 cmt2 cmt3* relative to wild type (Fig. 4b). Although both DNA type and retrotransposons were regulated by non-CG methylation, there was over-representation of DNA/Mariner, LINE/L1, LTR/Copia, and LTR/Gypsy transposons (Supplementary Fig. 4b and Supplementary Table 1). Hence, different non-CG methyltransferases cooperate to silence TEs in the genome. We next measured the changes in non-CG methylation levels associated with changes in TE expression. The degree of TE upregulation correlated with the degree of loss of non-CG methylation in the mutants, indicating that these TEs are indeed regulated by non-CG methylation (Fig. 4c). Hence non-CG methylation plays important roles in silencing TEs.

## CMT3 and DRM2, but not CMT2, regulate protein-coding genes

DNA methylation also regulates expression of protein-coding genes. By applying the same stringent cutoffs as we did for TEs, we defined 166 protein-coding genes significantly upregulated and 117 genes down-regulated in *drm1 drm2 cmt2 cmt3* mutants. Genes that became upregulated in *drm1 drm2 cmt2 cmt3* mutants were substantially associated with high levels of non-CG methylation in wild type (Supplementary Fig. 4c), as well as non-CG DMRs in *drm1 drm2 cmt2 cmt3* mutants (Fig. 4d) indicating that these genes are regulated by non-CG methylation. In contrast, genes down-regulated in *drm1 drm2 cmt2 cmt3* mutants did not show association with non-CG methylation, suggesting that down-regulation of these genes is likely an indirect effect (Fig. 4d and Supplementary Fig. 4c). This result indicates that non-CG methylation primary acts as a repressor of transcription. Gene ontology analysis of genes upregulated in *drm1 drm2 cmt2 cmt3* mutants indicated some association with response genes (Supplementary Fig. 4d); however, the list contained a variety of genes with different functions (Supplementary Table 2).

The fact that DRM2 targets sites proximal to genes suggests that it may function to regulate gene expression[8,9]. These sites are methylated by CMT3 and DRM2, but not CMT2 (Fig. 3f). Consistent with this fact, gene upregulation was most prominent in *drm1 drm2 cmt3* mutants compared to any other combinations of mutants (Fig. 4e). In fact, *drm1 drm2 cmt2 cmt3* mutants did not show substantial increase in gene expression levels compared to *drm1*

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

*drm2 cmt3* mutants (Fig. 4e). This is in contrast to our analysis of TEs (Fig. 4a). *SUPPRESSOR OF drm1 drm2 cmt3* (SDC) is a gene redundantly regulated by DRM2 and CMT3, and is responsible for the developmental phenotypes of *drm1 drm2 cmt3* mutants[19]. SDC was not more expressed in *drm1 drm2 cmt2 cmt3* compared to *drm1 drm2 cmt3* mutants (Supplementary Fig. 4e), consistent with the morphological defects the plants exhibited (Supplementary Fig. 4f). Hence while TEs are cooperatively silenced by DRM2, CMT2, and CMT3, protein-coding genes are largely cooperatively regulated by CMT3 and DRM2 but not CMT2.

## 24-nt siRNAs and non-CG methylation at DRM2 target sites

DRM2 is guided by 24nt-siRNAs to target loci[1]. The biogenesis of 24nt-siRNA depends on Pol IV. However, at certain loci siRNA accumulation has also been shown to depend on downstream RdDM factors such as Pol V and DRM2[14,20-22]. We sought to examine the extent to which siRNA accumulation depends on non-CG methylation by performing small RNA sequencing (smRNA-seq). We found that in *drm1 drm2 cmt3* mutants there was strong loss of 24nt-siRNAs (Fig. 5a). This suggests that loss of non-CG methylation at these sites causes loss of 24nt-siRNAs. Loss of 24nt-siRNA at these sites was not observed in *cmt2* mutants, nor was the degree of loss substantially enhanced in *drm1 drm2 cmt2 cmt3* mutants compared to *drm1 drm2 cmt3* mutants (Fig. 5a), consistent with the finding that CMT2 generally does not act at DRM2 target sites. Our results uncover an almost complete dependency of 24nt-siRNA accumulation on non-CG methylation at DRM2 target sites, suggesting a strong self-reinforcing loop mechanism.

## 24-nt siRNAs and non-CG methylation at CMT2 target sites

Upstream RdDM factors such as Pol IV are responsible for most 24nt-siRNA produced in the genome[23-25]. By analyzing ChIP-seq data on Pol IV[14] we confirmed that Pol IV protein was physically enriched at CMT2 target sites (Supplementary Fig. 5a). Known upstream RdDM mutants such as *dms4*, *pol iv*, and *rdr2*, which strongly reduce 24nt-siRNA across the genome[23-26], did not substantially reduce CHH methylation at CMT2 dependent sites (Supplementary Fig. 5b). In contrast, we observed that both *drm1 drm2 cmt3* mutants and *cmt2* single mutants had partial but consistent loss of 24nt-siRNA accumulation at CMT2 target sites (Fig. 5b). There was substantially more loss of 24nt-siRNAs upon loss of all non-CG methylation in *drm1 drm2 cmt2 cmt3* quadruple mutants (Fig. 5b). This suggests that non-CG methylation partially regulates 24nt-siRNAs at these sites. While these 24nt-siRNAs do not control non-CG methylation in cis, one possibility is that they target other elements in trans[27], such as newly inserted TEs[28]. Our results suggest that there is an almost complete dependency of 24nt-siRNA on non-CG methylation at DRM2 target sites, and partial dependency of 24nt-siRNA on non-CG methylation at CMT2 target sites. As explored below, a possible mechanism for this dependency may be through H3K9 methylation.

## Non-CG methylation globally controls H3K9 methylation

Most non-CG methylation in the genome is regulated by H3K9 methylation (Fig. 2)[8,10,14,15]. Although partially, H3K9 methylation has also been suggested to be dependent

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

on DNA methylation at certain loci, suggesting a self-reinforcing loop between DNA methylation and H3K9 methylation[29-31]. This self-reinforcing loop is likely mediated at least in part by the SRA domains of the H3K9 methyltransferases KYP, SUVH5, and SUVH6 that preferentially bind methylated DNA[29]. However, the extent of this dependency remains poorly understood. We performed ChIP-seq on H3K9me2 in wild type, *drm1 drm2 cmt2 cmt3* mutants, and the *kyp suvh5 suvh6* triple H3K9 methyltransferase mutant. Strikingly, by analyzing the distribution of H3K9me2 across chromosomes we found strong loss of H3K9me2 in *drm1 drm2 cmt2 cmt3* mutants (Fig. 6a). Inspection of the data on the genome browser confirmed loss of H3K9me2 in *drm1 drm2 cmt2 cmt3* mutants (Fig. 6b and Supplementary Fig. 6a). In fact, the degree of loss of H3K9me2 in *drm1 drm2 cmt2 cmt3* mutants was as strong as in *kyp suvh5 suvh6* mutants (Fig. 6a, b and Supplementary Fig. 6a). Loss of H3K9me2 in *drm1 drm2 cmt2 cmt3* mutants occurred at both CMT2 targeted sites and DRM2 targeted sites, although the loss appeared stronger at CMT2 dependent sites (Fig. 6c). Strong loss of 24nt-siRNA in *drm1 drm2 cmt2 cmt3* mutants at DRM2 target sites (Fig. 5a) is likely in part explained by loss of H3K9 methylation, since 24nt-siRNA accumulation is dependent on the H3K9 methylation binding protein SHH1[14,15]. Our results indicate that non-CG methylation mediates genome-wide H3K9 methylation patterning.

## 24nt-siRNA accumulation is mediated by H3K9 methylation

Our finding of extensive self-reinforcing loops between H3K9 methylation and non-CG methylation in part provides an explanation for the self-reinforcing loop between 24nt-siRNA accumulation and non-CG methylation. At DRM2 target sites, non-CG methylation is required for H3K9 methylation (Fig. 6c), which then regulates 24nt-siRNAs through SHH1 binding to H3K9 methylation. Consistent with this model, in *kyp suvh5 suvh6* mutants, there was a strong loss of 24nt-siRNAs at DRM2 target sites (Fig. 5a). At CMT2 target sites, non-CG methylation is almost completely required for H3K9 methylation (Fig. 6c). Consistent with the fact that H3K9me2 is lost to a similar extent in *drm1 drm2 cmt2 cmt3* and *kyp suvh5 suvh6* mutants (Fig. 6a), we found similar degrees of loss of 24nt-siRNA in *drm1 drm2 cmt2 cmt3* and *kyp suvh5 suvh6* mutants compared to wild type (Fig. 5b). Hence it is likely that non-CG methylation controls H3K9 methylation which then regulates the biogenesis of 24nt-siRNA.

## CG methylation and heterochromatic H3K9 methylation

Genome-wide elimination of CG methylation by mutation of the CG methyltransferase, MET1, resulted in loss of H3K9me2 at certain sites[32,33], although the mechanism is not understood. We analyzed H3K9me2 ChIP data in wild type and *met1* mutants[34]. As expected, we observed loss of H3K9me2 at certain sites in *met1* mutants (Fig. 6d). However, we found that these were sites that also lost non-CG methylation in *met1* mutants (Supplementary Fig. 6b). On the other hand, we did not observe genome-wide loss of H3K9me2 in *met1* mutants as we found in *drm1 drm2 cmt2 cmt3* mutants (Fig. 6d and Supplementary Fig. 6c). This suggests that H3K9 methylation is much more dependent on non-CG methylation than it is on CG methylation. While we cannot rule out the possibility that loss of H3K9me2 at certain sites in *met1* mutants is directly due to loss of CG methylation, it seems likely that loss of H3K9me2 in *met1* mutants is due to loss of non-CG

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

methylation at these sites. Our results suggest that non-CG methylation plays a dominant role in regulating H3K9 methylation patterning throughout the genome.

### Loss of non-CG methylation induces histone hyperacetylation

Histone acetylation is associated with open chromatin and actively transcribed genes. Given the strong loss of the repressive histone mark H3K9me2 in *drm1 drm2 cmt2 cmt3*, we sought to examine the effects on genome-wide histone acetylation patterns. We performed ChIP-seq on H3K23 acetylation (H3K23ac) and H3 on wild type, *drm1 drm2 cmt2 cmt3*, and *kyp suvh5 suvh6* mutants. As expected, H3K23ac was enriched in promoter regions of active genes in wild type (Supplementary Fig. 6d). We observed genome-wide increases of histone acetylation in *drm1 drm2 cmt2 cmt3* mutants and *kyp suvh5 suvh6* mutants at sites that lost DNA methylation (Fig. 6c, e and Supplementary Fig. 6e). Elevation of histone acetylation levels were not restricted to transcriptionally upregulated TEs and genes (Fig. 6e and Supplementary Fig. 6f), suggesting that this phenomenon cannot simply be explained by more transcription in the mutants. Consistent with the elevation in histone acetylation, we found substantial chromocenter decondensation in *drm1 drm2 cmt2 cmt3* mutants (Supplementary Fig. 6g). Hence non-CG methylation is required to keep heterochromatin in a deacetylated and compacted state.

## DISCUSSION

In this study we characterized a series of mutants affecting non-CG methylation including the poorly understood methyltransferase CMT2. This analysis has uncovered the roles of each non-CG methyltransferase in DNA methylation patterning and gene silencing. Furthermore, our finding of extensive cross talks between non-CG methylation and H3K9 methylation provide insights into the mechanisms of cross talk between different silencing pathways. All data generated in this study can be visualized in a modified UCSC browser (http://genomes.mcdb.ucla.edu/AthBSseq/) along with other epigenomic datasets.

At DRM2 target sites, there is a self-reinforcing loop between non-CG methylation, H3K9 methylation and 24nt-siRNAs (Fig. 7a). H3K9 methylation is required for CMT3 targeting to methylate CHG sites at a subset of DRM2 sites, as well as for DRM2 targeting through binding of SHH1[14], which methylates the remaining non-CG sites. SHH1 binding to H3K9 methylation is required for 24nt-siRNA accumulation at a subset of DRM2 sites[14]. The 24nt-siRNAs then directs DRM2[35]. Our data suggest that DRM2 and CMT3 mediated non-CG methylation is required for H3K9 methylation, which is in large part is mediated by KYP SUVH5 SUVH6. The H3K9 methylation then directs CMT3 and DRM2 pathways for non-CG methylation.

At CMT2 target sites, there is also a self-reinforcing loop between non-CG methylation, H3K9 methylation and 24nt-siRNAs (Fig. 7b). Our results suggest that both CMT2 and CMT3 mediate CHG methylation and CMT2 mediates CHH methylation at these sites through binding to H3K9 methylation. Non-CG methylation mediated by CMT2 and CMT3 regulates H3K9 methylation mediated by KYP SUVH5 SUVH6. H3K9 methylation may then partially regulate 24nt-siRNAs produced at these sites through a similar mechanism that occurs at DRM2 target sites. Because these 24nt-siRNAs are also dependent on Pol

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

IV[25,26], a speculation is that there may be H3K9 methylation readers other than SHH1 that recruit Pol IV to CMT2 sites. While these 24nt-siRNAs do not appear to play a major role in guiding DRM2 in cis, they might function to silence TEs in trans[27,28].

In summary, our data demonstrate that the CMT2, CMT3, and DRM2 methyltransferases collaborate to control non-CG methylation, and participate in self-reinforcing loop mechanisms with H3K9 methylation and small RNAs to control gene silencing throughout the genome.

## ONLINE METHODS

### Plant Material

All mutant lines used in this study were in the Columbia ecotype background. *drm1 drm2 cmt3* and *kyp suvh5 suvh6* mutants were previously described[11,36]. The *cmt2* T-DNA allele used in this study was *cmt2-7* (WISCDSLOX7E02) and *cmt2-3* (SALK_012874). *cmt2-7* was used for subsequent crosses. Plants were grown under continuous light, and three-week-old leaves were used for all experiments, except for small RNA sequencing (see below).

### RT-PCR

Total RNA was extracted from leaves with Trizol, and treated with DNase I (Roche). cDNA was synthesized with oligo-dTs using Superscript II (Invitrogen). PCR was performed on CMT2 (JP10697: GAGAAATCCTAAAACGTCCG and JP10698: CAGCCCATTTCGTCACGAC) and ACTIN (JP2452: TCGTGGTGGTGAGTTTGTTAC and JP2453: CAGCATCATCACAAGCATCC).

### Recombinant Protein Expression and Purification

The N-terminal fragment of *Arabidopsis* CMT2 (residues 1-503) did not show homology to any known domain, nor did it BLAST to any other plant species, and was not included. The N-terminal truncated CMT2 (residues 504 - 1295), including all the functional domains (the BAH domain, the chromodomain, and the DNA methyltransferase domain), was cloned into a self-modified vector which fuses an N-terminal hexa-histidine plus yeast sumo tag to the target protein. The recombinant plasmid was transformed into *E. coli* strain BL21(DE3) RIL (Stratagene). The cells were cultured in LB media at 37°C until OD600 reached 0.6. The media was subsequently cooled to 20°C and 0.25 mM IPTG was added to induce the protein expression overnight. The recombinant expressed protein was purified using a HisTrap FF column (GE Healthcare) followed by a Q FF column (GE Healthcare) and a Hiload Superdex G200 16/60 column (GE Healthcare). The purified protein was concentrated to 15 mg/ml and was stocked in – 80°C for further using. The N-terminal truncated *Arabidopsis* CMT3 (residues 46-839), including all the functional domains (the BAH domain, the chromodomain, and the DNA methyltransferase domain), was cloned, expressed, and purified using the same protocol as CMT2.

### Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC)-based binding experiments were conducted using a MicroCalorimeter iTC 200 instrument at 4 °C. Purified protein samples were dialyzed

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

overnight against a buffer of 100 mM NaCl, 2 mM β-mercaptoethanol, and 20 mM HEPES, pH 7.5 at 4 °C. Then, the protein samples were diluted and the lyophilized peptides were dissolved with the same buffer. The titration was conducted according to standard protocol and the data was fitted using the program Origin 7.0.

### DNA Methyltransferase Activity Assay

DNA methyltransferase assay was performed as previously described[10] except that 2 μg of recombinant CMT2 protein was used. Oligos used for the assays are shown in Supplementary Table 3.

### Histone Peptide Array

Thirty μg of recombinant CMT2 protein was screened on a MODified histone array slide following manufacturer instructions (Active Motif) using His antibody and was developed with Enhanced chemiluminescence (GE Healthcare). All analyses were performed using the manufacturer software (Active Motif).

### Whole Genome Bisulfite Sequencing (BS-seq)

500 ng of genomic DNA was used to generate BS-seq libraries as previously described[8,37]. 50-mer sequencing reads were analyzed. Identical reads were collapsed into single reads, and reads were mapped to the TAIR10 genome using BS-seeker by allowing up to 2 mismatches. Fractional DNA methylation levels were computed by #C/(#C+#T). DMRs were defined exactly as previously described [8].

### mRNA Sequencing

RNA was extracted from 0.1 g tissue using Trizol (Invitrogen). We performed mRNA-seq experiments on two biological replicates for each genotype tested. Libraries were generated and sequenced following manufacturer instructions (Illumina). Data were analyzed as previously described[38]. Reads were mapped to the TAIR10 genome using Bowtie[39] by allowing up to two mismatches and only keeping reads that uniquely map to the genome. Genes and TEs were defined as deregulated in a mutant using a four-fold cutoff and a corrected $p<0.01$. Only genes and TEs that showed consistent deregulation in two independent experiments were defined as significantly deregulated. To avoid divisions with zero, elements with zero reads were assigned the lowest non-zero gene or TE expression values within each library.

### smRNA Sequencing

Total RNA was extracted from 0.2 g of flowers using Trizol (Invitrogen). siRNAs were purified as previously described[40] with the following modifications. To precipitate high molecular weight RNAs, 25% PEG was added to a final concentration of 12.5% instead of 5% PEG. For small RNA purification from LMW RNA, SYBR® Gold was used to stain the gel. The gel was crushed using Gel Breaker Tubes (IST Engineering Inc), and the debris was filtered using 5 μm Filter tubes (IST Engineering Inc). The final elution of the RNA was done in 5 μL of nuclease-free $H_2O$ for subsequent generation of libraries for high throughput sequencing. Libraries were generated and sequenced following manufacturer instructions

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Illumina TruSeq Small RNA Sample Preparation Kits). Adapter sequences were clipped off before mapping. Reads were mapped to the TAIR10 genome using Bowtie[39] by allowing no mismatches and only keeping reads that uniquely map to the genome. For the analyses, the smRNA counts were normalized to the size of each smRNA library by dividing to the number of reads to the number of total uniquely mapping reads of 21 bp in size.

### Chromatin Immunoprecipitation (ChIP) Sequencing

One gram of tissue was ground in liquid nitrogen, and ChIP was performed as previously described[14] using the following antibodies: H3K9me2 (Abcam 1220), H3 (Abcam 1791), H3K9me1 (Upstate 07-450), and H3K23ac (Millipore 07-355). Libraries were generated and sequenced following manufacturer instructions (Illumina). Reads were mapped to the TAIR10 genome using Bowtie[39] by allowing up to two mismatches and only keeping reads that uniquely map to the genome. Reads mapping to identical locations were collapsed into one read. Two independent ChIP-seq experiments on biological replicates were performed on H3K9me2 and H3K23ac on wild type, *drm1 drm2 cmt2 cmt3* and *kyp suvh5 suvh6* mutants, which led to the similar conclusions.

### Chromocenter Compaction Assay

Chromocenter compaction assays were performed as previously described[41] with the following modifications. Following post-fix, the slides were washed three times in PBS for 5 minutes each. The nuclei were then stained and mounted in Vectashield mounting media with DAPI (Vector H-1200). At least 100 nuclei were analyzed for each genotype.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We thank M. Akhavan for Illumina sequencing. Sequencing was performed at the UCLA BSCRC BioSequencing Core Facility. We thank W. Yang and M. Pellegrini for help with the UCSC browser. We thank L. Tao and D. Chen for technical help with experiments. We thank F. Berger for helpful comments. This work was supported by US National Institutes of Health grant GM60398 and National Science Foundation grant 0701745 (S.E.J.). H.S. was supported by a UCLA Dissertation Year Fellowship. X.Z. is a research fellow of Ruth L. Kirschstein National Research Service Award (F32GM096483-01). S.F. is a Special Fellow of the Leukemia & Lymphoma Society. S.E.J. is supported as an investigator of the Howard Hughes Medical Institute.

## REFERENCES

1. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010; 11:204–220. [PubMed: 20142834]
2. Lister R, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009; 462:315–322. [PubMed: 19829295]
3. Xie W, et al. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell. 2012; 148:816–831. [PubMed: 22341451]
4. Laurent L, et al. Dynamic changes in the human methylome during differentiation. Genome Res. 2010; 20:320–331. [PubMed: 20133333]
5. Meissner A, et al. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 2005; 33:5868–5877. [PubMed: 16224102]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

6. Ramsahoye BH, et al. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A. 2000; 97:5237–5242. [PubMed: 10805783]
7. Varley KE, et al. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. 2013; 23:555–567. [PubMed: 23325432]
8. Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. Cell. 2013; 152:352–364. [PubMed: 23313553]
9. Zemach A, et al. The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. Cell. 2013; 153:193–205. [PubMed: 23540698]
10. Du J, et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell. 2012; 151:167–180. [PubMed: 23021223]
11. Ebbs ML, Bender J. Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. Plant Cell. 2006; 18:1166–1176. [PubMed: 16582009]
12. Jackson JP, Lindroth AM, Cao X, Jacobsen SE. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature. 2002; 416:556–560. [PubMed: 11898023]
13. Henikoff S, Comai L. A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in Arabidopsis. Genetics. 1998; 149:307–318. [PubMed: 9584105]
14. Law JA, et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature. 2013; 498:385–389. [PubMed: 23636332]
15. Zhang H, et al. DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proc Natl Acad Sci U S A. 2013; 110:8290–8295. [PubMed: 23637343]
16. Johnson L, et al. Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res. 2004; 32:6511–6518. [PubMed: 15598823]
17. Roudier F, et al. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. Embo J. 2011; 30:1928–1938. [PubMed: 21487388]
18. Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F. DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. Curr Biol. 2012; 22:1825–1830. [PubMed: 22940470]
19. Henderson IR, Jacobsen SE. Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. Genes Dev. 2008; 22:1597–1606. [PubMed: 18559476]
20. Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. RNA polymerase IV directs silencing of endogenous DNA. Science. 2005; 308:118–120. [PubMed: 15692015]
21. Pontier D, et al. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes Dev. 2005; 19:2030–2040. [PubMed: 16140984]
22. Zilberman D, Cao X, Jacobsen SE. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science. 2003; 299:716–719. [PubMed: 12522258]
23. Kasschau KD, et al. Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol. 2007; 5:e57. [PubMed: 17298187]
24. Mosher RA, Schwach F, Studholme D, Baulcombe DC. PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. Proc Natl Acad Sci U S A. 2008; 105:3145–3150. [PubMed: 18287047]
25. Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE. Role of RNA polymerase IV in plant small RNA metabolism. Proc Natl Acad Sci U S A. 2007; 104:4536–4541. [PubMed: 17360559]
26. Wierzbicki AT, et al. Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. Genes Dev. 2012; 26:1825–1836. [PubMed: 22855789]
27. McCue AD, Nuthikattu S, Slotkin RK. Genome-wide identification of genes regulated in trans by transposable element small interfering RNAs. RNA Biol. 2013; 10

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

28. Zhong X, et al. DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. Nat Struct Mol Biol. 2012; 19:870–875. [PubMed: 22864289]
29. Johnson LM, et al. The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol. 2007; 17:379–384. [PubMed: 17239600]
30. Inagaki S, et al. Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. Embo J. 2010; 29:3496–3506. [PubMed: 20834229]
31. Mathieu O, Probst AV, Paszkowski J. Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. Embo J. 2005; 24:2783–2791. [PubMed: 16001083]
32. Soppe WJ, et al. DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. Embo J. 2002; 21:6549–6559. [PubMed: 12456661]
33. Johnson L, Cao X, Jacobsen S. Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. Curr Biol. 2002; 12:1360–1367. [PubMed: 12194816]
34. Deleris A, et al. Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis thaliana. PLoS Genet. 2012; 8:e1003062. [PubMed: 23209430]
35. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet. 2009; 41:630–634. [PubMed: 19377477]
36. Chan SW, et al. RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in arabidopsis. PLoS Genet. 2006; 2:e83. [PubMed: 16741558]
37. Feng S, Rubbi L, Jacobsen SE, Pellegrini M. Determining DNA methylation profiles using sequencing. Methods Mol Biol. 2011; 733:223–238. [PubMed: 21431774]
38. Stroud H, et al. DNA methyltransferases are required to induce heterochromatic re-replication in Arabidopsis. PLoS Genet. 2012; 8:e1002808. [PubMed: 22792077]
39. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]
40. Lu C, Meyers BC, Green PJ. Construction of small RNA cDNA libraries for deep sequencing. Methods. 2007; 43:110–117. [PubMed: 17889797]
41. Moissiard G, et al. MORC family ATPases required for heterochromatin condensation and gene silencing. Science. 2012; 336:1448–1451. [PubMed: 22555433]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

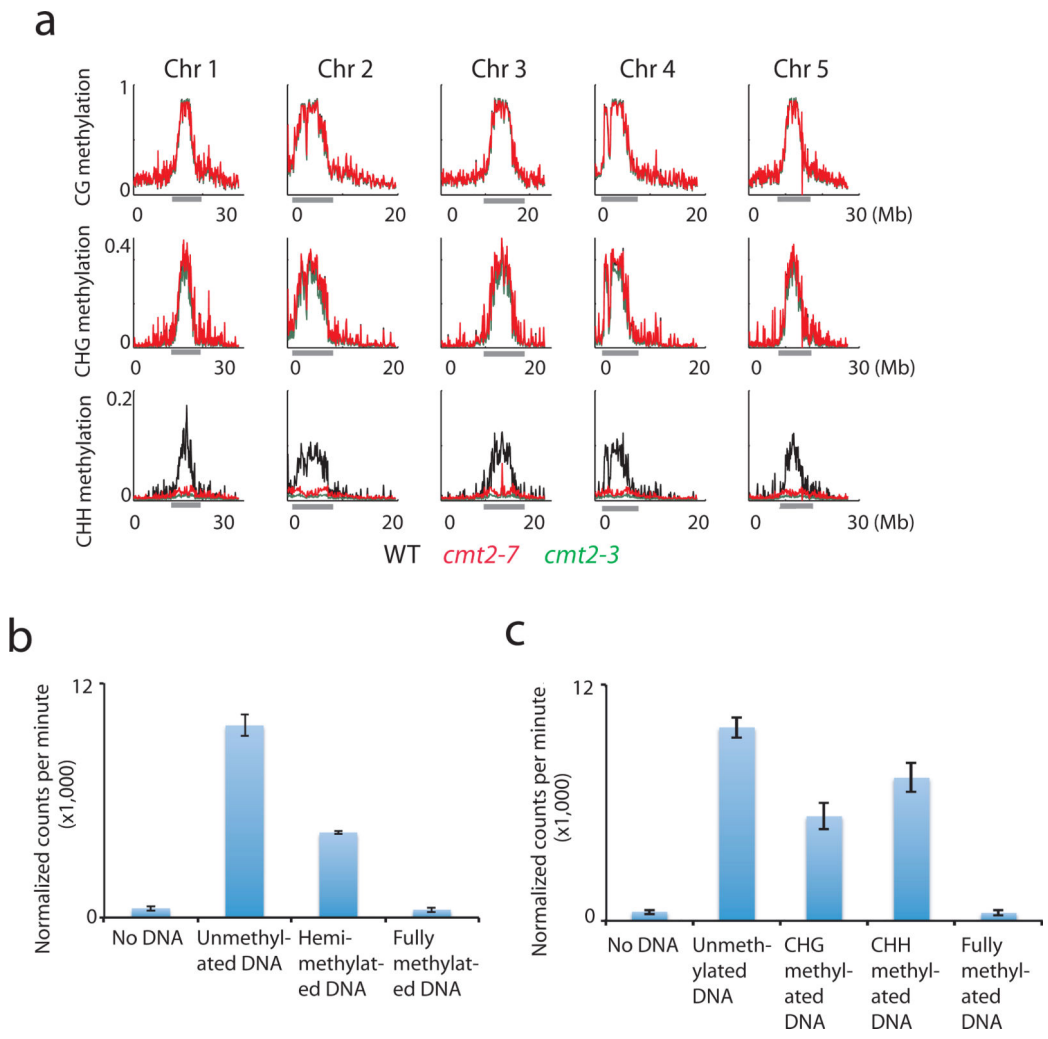


**Figure 1.**

*In vitro* activity of CMT2. (**a**) Fractional DNA methylation levels of cytosines in CG, CHG, and CHH contexts across chromosomes. Grey bars indicate pericentromeric heterochromatin. (**b**) CMT2 *in vitro* methylation activity on DNA of different methylation status. The values for unmethylated and hemimethylated DNA were normalized according to the number of available (i.e. unmethylated) cytosines. Error bars represent SD for two technical replicates. (**c**) CMT2 *in vitro* methylation activity on DNA of different methylation status. Sequence specificities of CMT2 were assessed. Error bars represent SD for two technical replicates.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

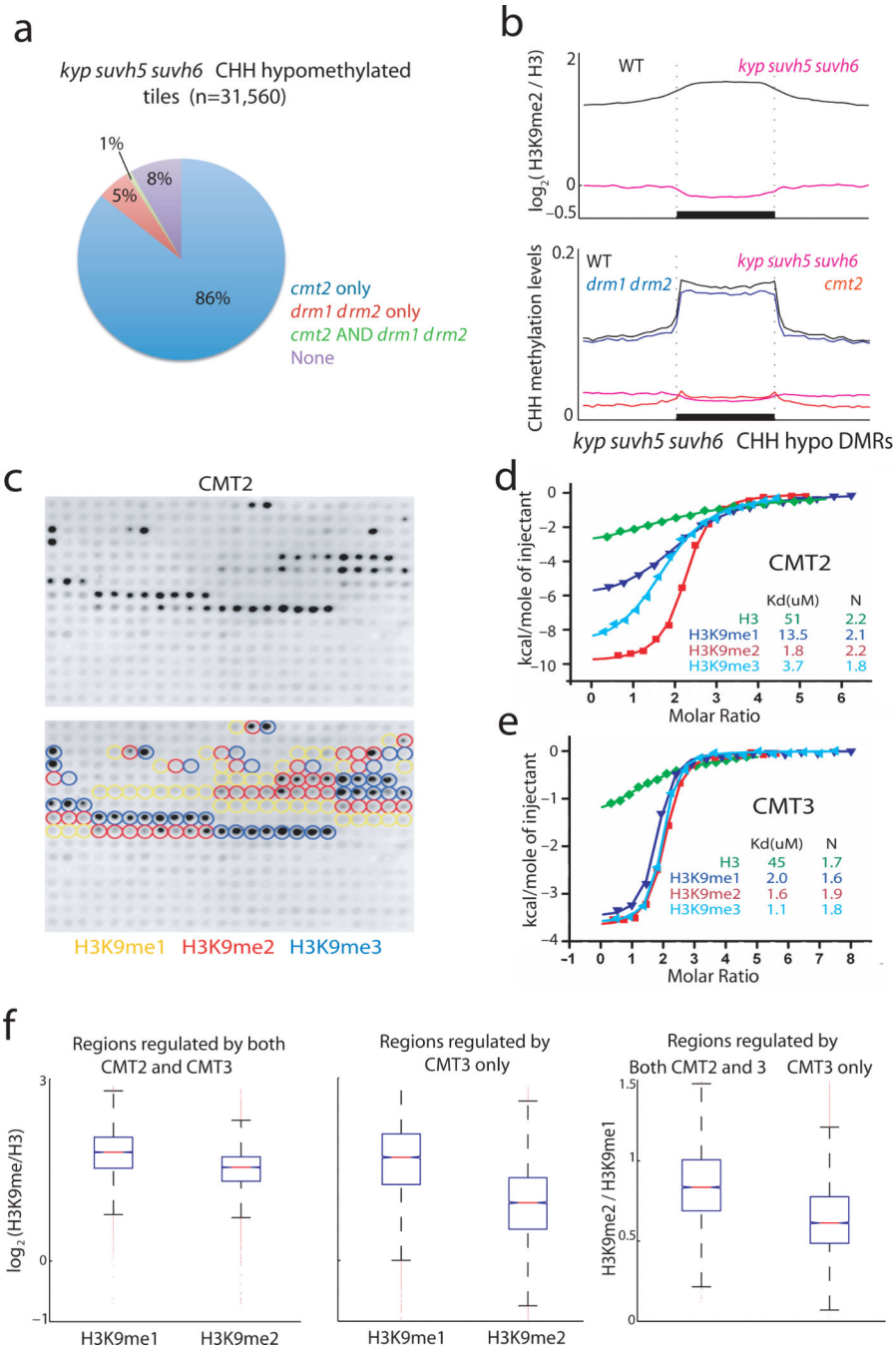


**Figure 2.**

CMT2 is mediated by H3K9 methylation. (**a**) Percentages of *kyp suvh5 suvh6* CHH hypomethylated 100 bp tiles overlapping with *cmt2* and *drm1 drm2* CHH hypomethylated tiles. (**b**) Average distribution of H3K9me2 and CHH methylation over previously defined *kyp suvh5 suvh6* CHH hypomethylation DMRs. Middle region represents the DMR and the flanking regions were scaled such that they are the same lengths as the middle region. (**c**) CMT2 binding assay to different histone modifications on a peptide array. The yellow, red, and blue circles indicate peptides containing mono-, di-, and trimethylated H3K9me2

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

peptides, respectively. (**d**) ITC binding curves for complex formation between CMT2 protein and H3K9me3, H3K9me2, H3K9me1, and unmodified H3 peptides. Kd values and the N values are listed as insert. (**e**) ITC binding curves for CMT3 protein. (**f**) Normalized H3K9me1 and H3K9me2 ChIP-seq reads in indicated regions are shown. Here and throughout, red lines, median; edges of boxes, 25$^{th}$ (bottom) and 75$^{th}$ (top) percentiles; error bars, minimum and maximum points within 1.5 × interquartile range; red dots, outliers.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

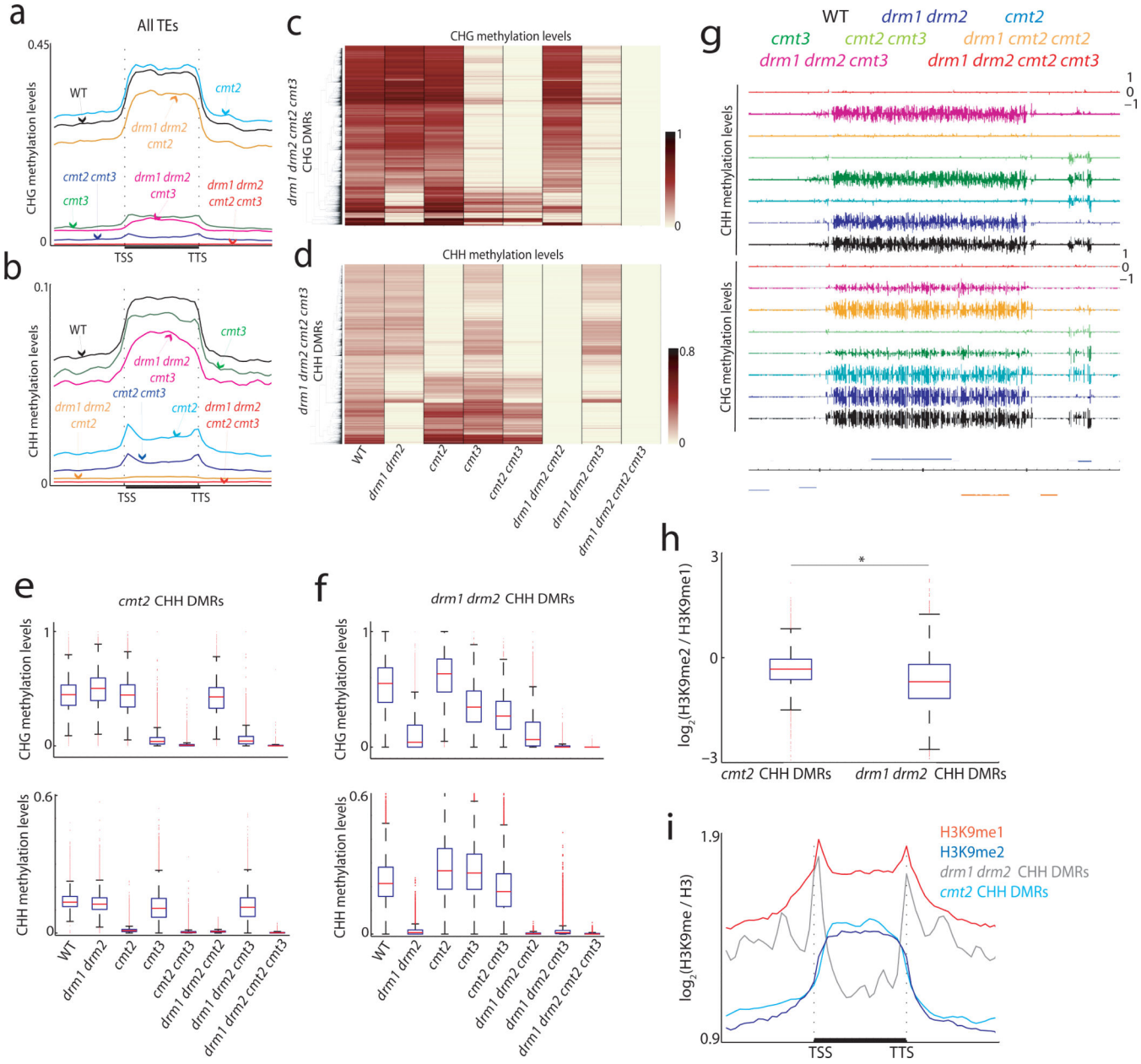


**Figure 3.**

Dissecting contributions of non-CG methyltransferases in DNA methylation patterning. (**a**) Average distribution of CHG methylation in indicated genotypes over all TEs. TSS= transcription start site; TTS= transcription termination site. (**b**) Average distribution of CHH methylation in indicated genotypes over all TEs. (**c**) Heatmaps of CHG methylation levels within *drm1 drm2 cmt2 cmt3* CHG hypomethylation DMRs. The columns represent the indicated genotypes, and the rows represent the DMRs. Rows were sorted by complete linkage hierarchical clustering with Euclidean distance as a distance measure. (**d**) Heatmaps of CHH methylation levels within *drm1 drm2 cmt2 cmt3* CHH hypomethylation DMRs. (**e**) Boxplots of CHG and CHH methylation levels in *cmt2* CHH DMRs. (**f**) Boxplots of CHG and CHH methylation levels in *drm1 drm2* CHH DMRs. (**g**) Genome browser views of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

CHG and CHH methylation in chromosome 1. Blue bars, TEs; Yellow bars, genes. (**h**) Boxplots of H3K9me2 levels relative to H3K9me1 in CMT2 target sites and DRM2 target sites. *$P$=6.5 × $10^{-224}$ by two-tailed Wilcoxon rank sum test. (**i**) Average distributions of H3K9me1 and H3K9me2 levels over long TEs. The log2 ratios of H3K9me1 and H3K9me2 to H3 were plotted over TEs of greater than 2 kilobases in size. Distribution of *drm1 drm2* and *cmt2* CHH hypomethylation DMRs are also shown for comparison (arbitrary scales).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

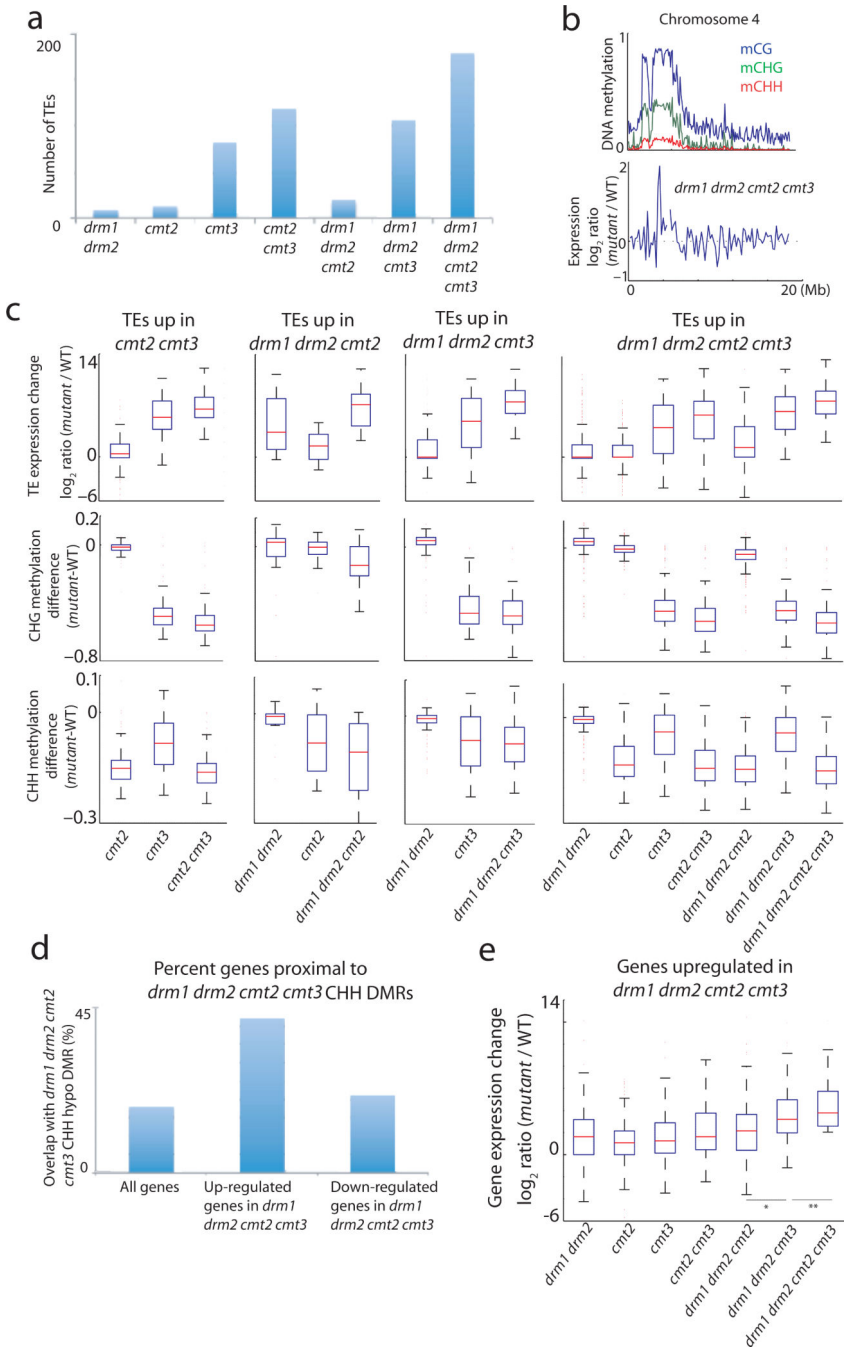


**Figure 4.**

Non-CG methyltransferases cooperatively silence TEs and genes. (**a**) Number of TEs defined to be significantly upregulated in indicated genotypes. (**b**) Distribution of RNA-seq reads in *drm1 drm2 cmt2 cmt3* relative to wild type. Wild-type DNA methylation levels are plotted in the top panel to indicate heterochromatic regions. (**c**) TE expression change in mutant relative to wild type and associated changes in CHG and CHH methylation levels in defined upregulated TEs are plotted. (**d**) Percentage of genes within one kilobase of *drm1 drm2 cmt2 cmt3* CHH hypomethylation DMRs. (**e**) Protein-coding gene expression levels of

genes defined to be upregulated in *drm1 drm2 cmt2 cmt3* mutants. *Medians significantly different at a 95% confidence interval. **Medians not different at a 95% confidence interval.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

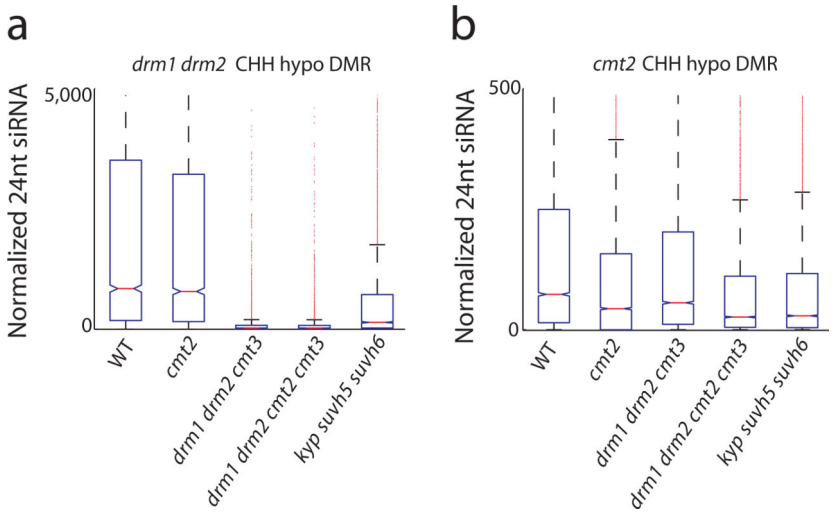


**Figure 5.**

Relationship between non-CG methylation and 24nt-siRNA accumulation. (**a**) 24ntsiRNA levels in DRM2 target sites. 24nt-siRNA levels were normalized by the counts of 21nt-siRNA levels for each genotype. (**b**) 24nt-siRNA levels in CMT2 target sites. 24nt-siRNA levels were normalized by the counts of 21nt-siRNA levels for each genotype.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

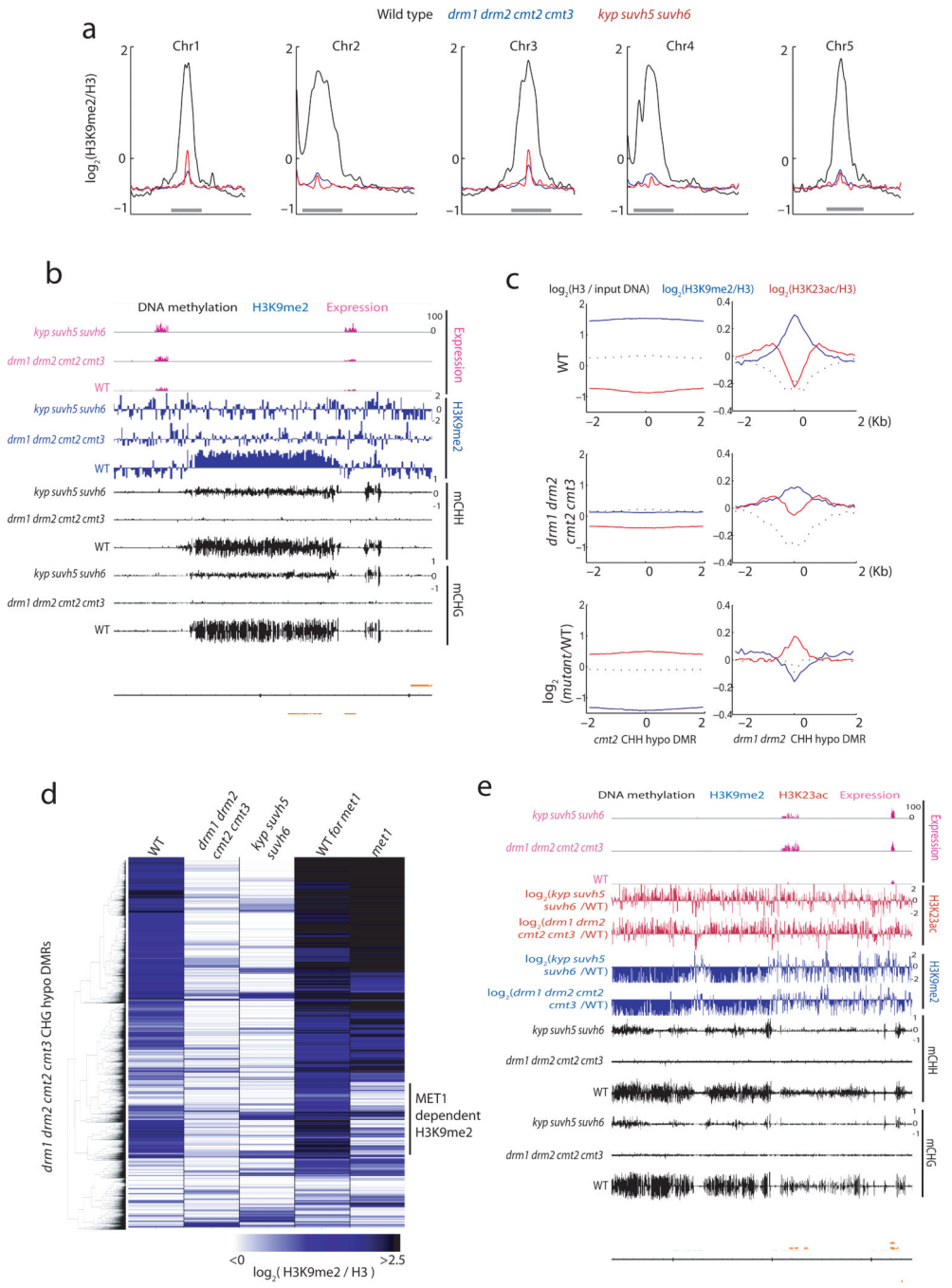


**Figure 6.**

Relationship between non-CG methylation and H3K9 methylation. (**a**) Distribution of H3K9me2 relative to H3 over chromosomes. The graphs were shifted such that all the graphs aligned on the euchromatic arms. Grey bars indicate pericentromeric heterochromatin. (**b**) Genome browser views of DNA methylation, expression levels, and H3K9me2 in wild type, *drm1 drm2 cmt2 cmt3*, and *kyp suvh5 suvh6* mutants in chromosome 1. Blue bars, TEs; Yellow bars, genes. (**c**) Average distribution of H3K9me2 and H3K23ac relative to H3 over *cmt2* and *drm1 drm2* CHH hypomethylation DMRs. (**d**) Heatmaps of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

H3K9me2 levels within *drm1 drm2 cmt2 cmt3* mutant CHG hypomethylation DMRs. H3K9me2 was normalized to H3. Two wild-type H3K9me2 data are shown since H3K9me2 data for *met1* has a separate wild-type control[34]. (**e**) Genome browser views of DNA methylation, expression levels, H3K23ac, and H3K9me2 in wild type, *drm1 drm2 cmt2 cmt3*, and *kyp suvh5 suvh6* mutants in chromosome 1. Blue bars, TEs; Yellow bars, genes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

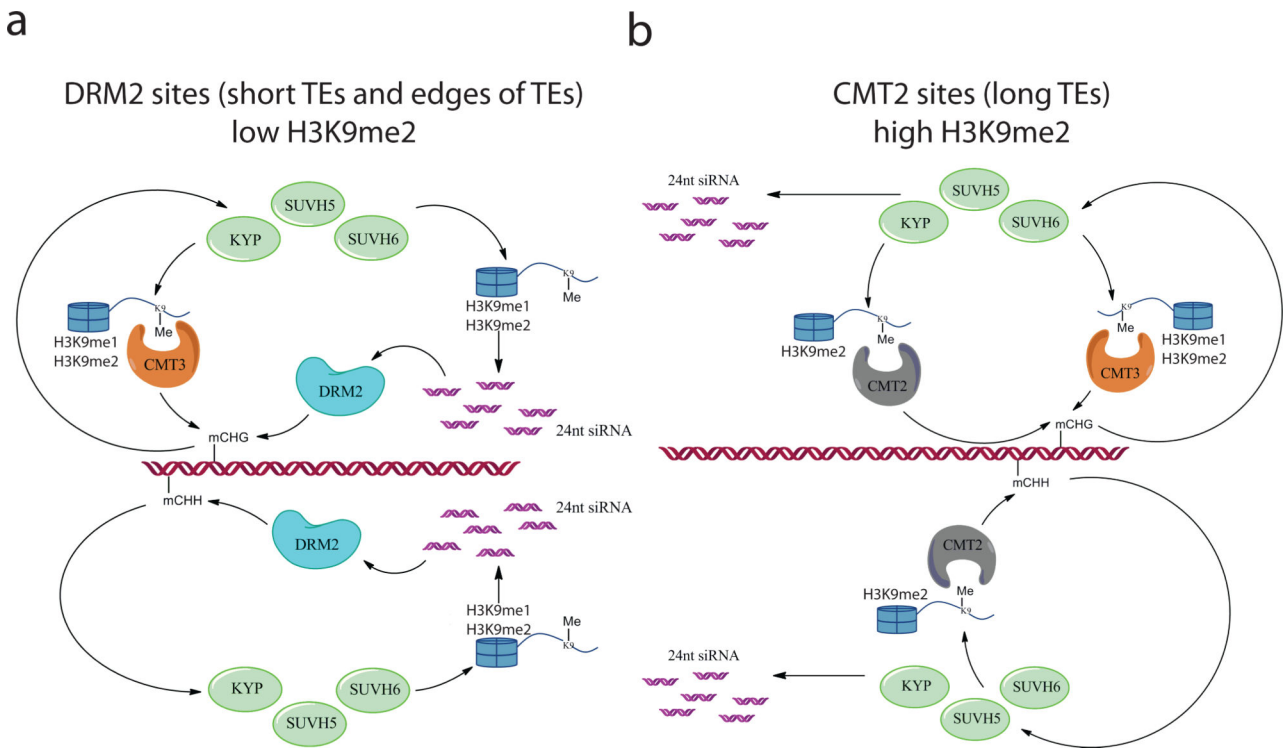


**Figure 7.**

Non-CG methylation pathways. (**a**) Non-CG methylation pathways at DRM2 target sites. See text for description. (**b**) Non-CG methylation pathways at CMT2 target sites. See text for description.