

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Maintenance and Inheritance of DNA Methylation in Arabidopsis

Permalink

<https://escholarship.org/uc/item/9r85v5k7>

Author

Hsieh, Ping-Hung

Publication Date

2016

Peer reviewed|Thesis/dissertation

Maintenance and Inheritance of DNA Methylation in *Arabidopsis*

by

Ping-Hung Hsieh

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Daniel Zilberman, Co-chair

Professor Robert Fischer, Co-chair

Professor Jasper Rine

Fall 2016

Maintenance and Inheritance of DNA Methylation in *Arabidopsis*

Copyright 2016

by

Ping-Hung Hsieh

Abstract

Maintenance and Inheritance of DNA Methylation in *Arabidopsis*

By

Ping-Hung Hsieh

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Associate Professor Daniel Zilberman, Co-chair

Professor Robert Fischer, Co-chair

DNA methylation is a conserved epigenetic modification, usually of the 5th carbon of cytosine in eukaryotes, and in plants is known to regulate gene expression and silence transposable elements. Posttranslational histone tail modifications are also conserved epigenetic regulators, known in the plant kingdom to interact with DNA methylation and regulate chromatin structure. Both DNA methylation and histone modifications are reversible and, collectively, play an important role in the orchestration of dynamic transcriptional profiles during the entire life cycle of the plant. In plants, DNA methylation is found in the symmetric CG and CHG contexts, and the asymmetric CHH context (where H is A, T or C). Several studies have shown that CG methylation is catalyzed by the mammalian DNMT1 methyltransferase ortholog, MET1, and CHG methylation is maintained by the plant specific chromomethylase 3 (CMT3), through a self-reinforcing loop between CMT3 and the heterochromatic dimethylation of lysine 9 of the histone H3 subunit (H3K9me2). Due to its asymmetry, CHH methylation was thought to be maintained solely by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2, an ortholog of mammalian DNMT3-type methyltransferases) through the plant-specific RNA-directed DNA methylation (RdDM) pathway. However, DRM2-mediated CHH methylation only accounts for about 35% of total CHH methylation in *Arabidopsis*, suggesting that other proteins contribute to the maintenance of CHH methylation.

To identify which proteins are involved in the maintenance of CHH methylation in addition to DRM2, I performed a reverse genetic screen on *Arabidopsis* mutants to identify genes that can potentially affect plant DNA methylation by performing whole-genome bisulfite sequencing of each individual homozygous mutant seedling. Eventually, I found that CMT2, a homolog of CMT3, is an active methyltransferase that maintains ~70% of total CHH methylation in *Arabidopsis*, independently of RdDM. Furthermore, I determined that the

DRM2-mediated RdDM pathway mainly carries out CHH methylation in more euchromatic regions, including short TEs in chromosome arms and the edges of long TEs in pericentromeric regions, whereas CHH methylation at more heterochromatic sites, such as the bodies of long TEs, is mediated by CMT2.

This maintenance activity is thought to allow DNA methylation to carry epigenetic information through cell division and reproduction, influencing gene expression and phenotype across generations. Trans-generational inheritance is mediated by a small group of cells that includes gametes and their progenitors in flowering plants. However, methylation is usually analyzed in somatic tissues that do not contribute to the next generation, and the mechanisms of trans-generational inheritance are inferred from such studies. The male gametophyte pollen, consisting of two sperm cells and a vegetative cell, was reported to undergo DNA reprogramming during sexual reproduction: a gain of heterochromatic CHH methylation in vegetative cells and a loss of heterochromatic CHH methylation in sperm cells were attributed to the activation of DRM2 and the silencing of DRM2, respectively.

To gain further insight into how DNA methylation is inherited and reprogrammed during sexual reproduction, I analyzed *Arabidopsis thaliana* sperm and vegetative cells purified from pollen with mutations in the DRM (both DRM1 and DRM2. DRM1 is specifically expressed during early seed development and functions redundantly with DRM2), CMT2, and CMT3 methyltransferases. I found that although the gain of heterochromatic CHH methylation in the vegetative cell is primarily mediated by CMT2 instead of DRM, the effect of the *cmt2* mutation in the vegetative cell was weaker than expected, and a small but significant fraction of new CHH methylation targets in heterochromatic TEs was mediated by DRM through RdDM pathway, which supports the observation that many heterochromatic TEs are derepressed in the vegetative cell. I also showed that lack of histone H1, which elevates heterochromatic DNA methylation in somatic tissues, does not have this effect in pollen. Instead, levels of CG methylation in wild-type sperm and vegetative cells, as well as in wild-type microspores from which both pollen cell types originate, are substantially higher than in wild-type somatic tissues and are similar to those of H1-depleted roots.

In summary, I discovered a novel DNA methyltransferase in plants, CMT2, which maintains CHH methylation independently from RdDM. I showed that the DNA methylation of germlines are maintained similarly to somatic tissues, which further explains why the reprogramming of CHH methylation in pollen is most likely dependent on CMT2. Although the mechanisms of methylation maintenance are similar between pollen and somatic cells, I demonstrated that the efficiency of CG methylation is higher in pollen, allowing methylation patterns to be accurately inherited across generations. The biological significance of the

reduced CHH methylation found in sperm remains unknown. Overall, my findings expand the previous incomplete DNA methylation maintenance model with the discovery of CMT2-mediated CHH methylation. My work on the pollen methylome contributes to our further understanding of the regulation and reprogramming of DNA methylation during sexual reproduction, providing insights into transgenerational epigenetic inheritance in plants.

Contents

Table of Contents.....	<i>i</i>
List of Figures	<i>iii</i>
List of Tables.....	<i>v</i>
Acknowledgements	<i>vi</i>
Chapter I	<i>1</i>
Introduction.....	<i>1</i>
Epigenetic regulation	<i>1</i>
Small RNA-mediated gene silencing in plants.....	<i>1</i>
Histone modifications and chromatin structure.....	<i>3</i>
Landscape of DNA methylation in plants	<i>4</i>
Maintenance of DNA methylation	<i>5</i>
Initiation of de novo DNA methylation via the RDR6-RdDM pathway	<i>7</i>
Active DNA demethylation.....	<i>8</i>
Epigenetic regulation and plant development	<i>10</i>
Epigenetic regulation during sexual reproduction.....	<i>10</i>
Summary	<i>12</i>
Chapter II	<i>14</i>
The nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin	<i>14</i>
Summary	<i>14</i>
Introduction.....	<i>15</i>
Results.....	<i>16</i>
Discussion	<i>33</i>
Materials and methods	<i>35</i>
Chapter III	<i>39</i>
The Arabidopsis male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues	<i>39</i>
Summary	<i>39</i>
Introduction.....	<i>41</i>
Results.....	<i>43</i>
Discussion	<i>53</i>
Materials and methods	<i>54</i>

Chapter IV	56
Concluding remarks and discussion.....	56
References.....	61

List of Figures

Figure 1.1 Post-transcriptional gene silencing (PTGS)	2
Figure 1.2 RNA-directed DNA methylation (RDR2-RdDM).....	3
Figure 1.3 The reinforcing loop between non-CG DNA methylation and H3K9me2.	7
Figure 1.4 The transition from posttranscriptional gene silencing to transcriptional gene silencing	8
Figure 2.1 Genomic hypomethylation in mutant lines shown as the percentage of wild-type methylation.	17
Figure 2.2 DDM1 and CMT2 mediate RdDM-independent CHH methylation of long TEs.	18
Figure 2.3. Heterochromatin requires DDM1 for DNA methylation and inhibits RdDM.....	21
Figure 2.4. Correlations between DNA methylation dependence and chromatin features	23
Figure 2.5. Lack of H1 ameliorates the loss of methylation in <i>ddm1</i> plants.	25
Figure 2.6. Methylation of TE families depends on sRNA abundance and chromatin features.	26
Figure 2.7. DDM1 and RdDM collaborate to repress TE expression and transposition.	28
Figure 2.8. DDM1 mediates genic DNA methylation.	30
Figure 2.9. Patterns of genic DNA methylation (CG, CHG and CHH) in wild-type and <i>ibm1drd1</i>	31

Figure 2.10. DDM1 and RdDM synergistically regulate gene expression.	32
Figure 3.1. DNA methylation in <i>Arabidopsis</i> leaf, sperm (Spm), and vegetative cell (Veg)..	44
Figure 3.2. DNA methylation in <i>Arabidopsis</i> leaf, sperm (Spm) and vegetative cell (Veg) TEs.	45
Figure 3.3. RdDM extends into heterochromatin in the vegetative cell.	47
Figure 3.4. Heterochromatic TEs targeted by RdDM in the vegetative cell are transcriptionally competent.....	48
Figure 3.5. H1 expression in pollen.	49
Figure 3.6. Lack of H1 does not increase heterochromatic methylation in pollen.	50
Figure 3.7. More robust maintenance of CG methylation in pollen.	51
Figure 3.8. CG methylation of <i>drm</i> , <i>cmt2</i> , and <i>cmt3</i> mutant sperm remains higher than in leaves in TEs and genes.	52
Figure 3.9 Grouping of TEs based on the level of H3K9me2 in two replicates.	55

List of Tables

Table 2.1 Mean genomic coverage and DNA methylation for wild-type and mutant samples, related to Chapter II.	16
Table 2.2. Pearson correlation coefficients between CHH methylation in mutants and sRNA.	19
Table 3.1. Mean genomic coverage and DNA methylation for wild type (WT) and mutants sequenced in this chapter III	43

Acknowledgements

This dissertation represents not only what I have done in the past five years, but a collection of the genius, effort, patience, kindness and love of many people. This work would not have been possible if these people did not come into my life and selflessly supported me. I therefore want to offer my sincere thanks to all of them.

First and foremost, I would like to thank my two respected advisors, Daniel Zilberman and Robert Fischer. They are the smartest people I have ever met, and their insightful guidance has helped me accomplish the tremendous work that greatly advances our knowledge of plant epigenetics. Their patience and warm encouragement have helped me discover my true talents and gain confidence as a scientist. Their critical comments have helped me clarify my thoughts and greatly improve the quality of my papers, thesis and presentations.

I also owe current members of the Zilberman and Fischer Labs a debt of gratitude, especially David Lyons and Jaemyung Choi. They not only provided me invaluable advice during numerous discussions, but also enriched my life with music, sports, art, cultures, and laughter. I'm also indebted to the former members of Zilberman and Fischer Labs. Assaf Zemach and Tzung-Fu Hsieh taught me key skills to study epigenetics and seed biology, helping me step into this field without being lost in the first place. Jason Huff and Devin Coleman-Derr are so experienced and knowledgeable that I can always seek advice from them. I also want to thank Kyunghyuk Park and Jennifer Frost for the assistance in plant crossing. I will never forget the days we spent together in the greenhouse. Nevertheless, I would like to thank our collaborator Dr. Xiaoqi Feng and members from her lab for collecting precious samples and generating high quality data. We would not have done the beautiful work without them. I also have to thank several undergraduate students, Sukhranjan Nijjar, Nicholas Nguyen and Cindy Ye, who worked with me and we together finish numerous experiments.

My deep appreciation specially goes out to three ladies: Anne Runkel, Jennifer Frost and Jessica Rodrigues. As a non-native speaker, language has always been a major obstacle to learning and communication. I would not survive my first year of graduate school if Anne did not generously share her class notes with me and thoughtfully invite me into her study group. Jenny and Jessica have been helping polish my writings, from my proposals of qualifying exam and fellowship application to this final dissertation. I am also forever thankful that Jennifer and Jessica always remembered every important moment in my Ph.D. journey and held celebrations for me. All these memories will never fade.

A healthy social life is important for surviving graduate school. I was very fortunate to

have two amazing roommates, Wen-Chin Huang and Yi-Pei Li. I would like to thank them for making our apartment like a home away from home, and I will always remember the days we lived together and travel together. I'm very grateful to have my board game friends, Jason Huff, Marlisa Pillsbury. Jessica Rodrigues, Devin Coleman-Derr, Toshiro Nishimura, Michael Pan, Jaemyung Choi and Allison Schwartz. Those joyful nights we spent together are the highlight of my Ph.D. journey.

Finally, I would like to thank my parents and my younger sister. Because of them, I know I always have someone to count on when times are rough. Their love and support picked me up during my falls. I also want to thank them for always believing in me when I was depressed and did not have faith in myself. I would not have had the courage to embark on this journey and pursue the best I can be without them. I have to especially thank my sister for accompanying my parents when I have been away from home these years. My every achievement would not have been possible without her. My family always say they are proud of me, but in fact, they should be proud of themselves because their unconditional love and constant support helped me survive through the rough road to complete this dissertation.

謹將此論文，獻給我摯愛的家人。
謝謝你們一直相信我，讓我自由地追求我的目標。
與你們分享我的喜悅，永遠都是我努力向前的動力。
今日若我有任何榮耀與成就，都歸功於你們無私的付出。

Chapter I

Introduction

Epigenetic regulation

Over many decades, classical genetics has shown that phenotypic variations derive largely from allelic differences that result from DNA sequence changes. However, even with advances in technology that have, for example, allowed comparison of entire genomes, identifying multiple sequence changes that together contribute to complex traits, there remain many phenotypic differences that cannot be explained by DNA sequence alone (1). Such phenotypic changes, without associated DNA sequence differences, can often be instead attributed to heritable epigenetic modifications such as DNA methylation and histone tail modifications (2-4). Epigenetic modification provides an additional way to store hereditary information, and has important influences on gene expression, transposable element (TE) silencing, chromatin structure, and development. Recently, many studies have suggested that non-protein-coding RNAs (ncRNAs) are also heritable and extensively involved in epigenetic regulation at the transcriptional and posttranscriptional levels (4-6).

Small RNA-mediated gene silencing in plants

In plants, small RNAs can trigger gene silencing at both transcriptional and posttranscriptional levels. Post-transcriptional gene silencing (PTGS) represents a defense mechanism in response to RNA viruses and TE mobilization in plants (Figure 1.1). PTGS is thought to be initialized by the presence of aberrant double-stranded RNAs (dsRNAs) transcribed by RNA polymerase II (Pol II), which can originate from both endogenous and exogenous sequences (7). For example, the inverted repeats or palindromic sequences within TEs are capable of generating long double-stranded hairpin RNA structures when the TEs are transcriptionally active (8-10). Alternatively, dsRNA structures can also appear during the genome replication of RNA viruses (11). Once aberrant dsRNAs are formed, they are recognized and cleaved by two plant ribonucleases DICER-LIKE 2 and 4 (DCL2 and DCL4) into 21-22-nt primary small interference RNAs (siRNAs). One strand of these 21-22-nt primary siRNAs is loaded into Argonaute protein 1 (AGO1, possessing endonuclease activity) and specifically guide AGO1 to cleave mRNAs of TEs or RNA viruses by sequence complementarity (6).

microRNAs (miRNAs) are another small RNA molecule involved in PTGS and are produced from endogenous miRNA genes (MIR genes) (12). MIR genes are transcribed by Pol II and the transcripts also form hairpin RNA structures as the precursors of siRNAs (13, 14).

Compared to the precursors of siRNAs, the transcripts of MIR genes are shorter and the hairpin structure contains bulges due to imperfect complementarity (15). The transcripts of MIR genes are subsequently cleaved by DCL1 into 21-22-nt miRNAs. One strand of the miRNAs is also loaded into AGO1 and downregulates endogenous gene expression by cleaving mRNAs through sequence complementarity (16). miRNAs are critical for many plant developmental processes and the mutation of DCL1 results in defects in plant growth and development (17-20).

PTGS in plants can be further amplified when the cleaved transcripts are subsequently converted into dsRNAs by RNA-dependent RNA polymerase 6 (RDR6) and then cleaved by DCL2 and DCL4 to generate more 21-22-nt siRNAs. These RDR6-dependent 21-22-nt siRNAs are called secondary siRNAs and can be further loaded into AGO1 to cleave more target transcripts, consequentially producing more secondary siRNAs (21-24). The amplification of siRNAs allows PTGS to suppress TEs or viruses more efficiently.

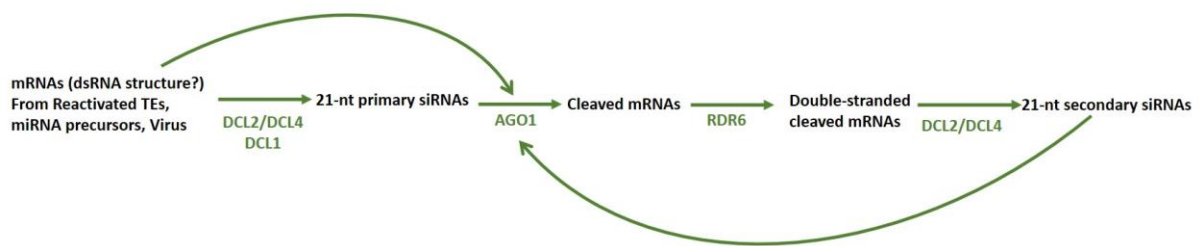


Figure 1.1 Post-transcriptional gene silencing (PTGS)

However, for the plant it would be even safer, in terms of protecting genomic integrity, to silence TEs and viruses by preventing their transcription. Thus, plants also utilize small RNAs to mediate gene silencing at the transcriptional level. Transcriptional gene silencing (TGS) can be established by RNA-directed DNA methylation (RdDM) in plants (4). RdDM begins with 24-nt small siRNA biosynthesis. Through this process, single stranded RNAs, transcribed by the plant specific RNA Polymerase IV (Pol IV), are converted into double stranded RNAs by RDR2 and then cleaved into 24-nt siRNAs by DCL3 (25-28). One strand of the siRNA is loaded into AGO4 to form an AGO4-guide-RNA complex. The AGO4-guide-RNA complex interacts with the scaffold ncRNAs of target loci, which are transcribed by a plant specific RNA polymerase V (PolV), leading to the recruitment of DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2, an ortholog of mammalian DNMT3-type methyltransferases) that catalyzes the methylation of target loci (RDR2-RdDM, Figure 1.2) (6, 26, 29-33).

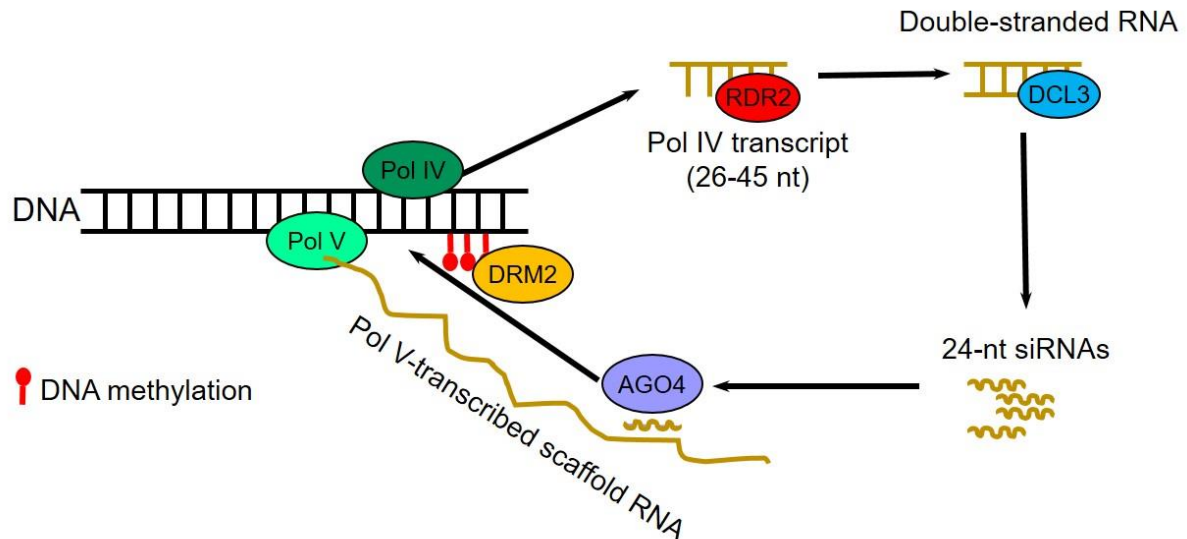


Figure 1.2 RNA-directed DNA methylation (RDR2-RdDM)

In brief, transcriptional gene silencing and posttranscriptional gene silencing can be summarized as follows: transcriptional gene silencing uses a Pol IV-RDR2-DCL3 pathway to generate 24-nt siRNAs and posttranscriptional gene silencing uses a Pol II-RDR6-DCL2/4 pathway to generate 21-22-nt siRNAs.

Histone modifications and chromatin structure

The nucleosome is the basic structural unit of chromatin, which consists of ~147 bp DNA wrapped around a histone octamer, containing two copies each of canonical histones H2A, H2B, H3 and H4 and often accompanied by an additional H1 linker histone protein connecting individual nucleosomes to each other (34-36). The linker H1 histone is thought to be associated with condensed, less accessible, and transcriptionally silent chromatin with reduced nucleosome mobility (37-39). In addition to the canonical histone proteins, the plant genome also encodes several histone variants, whose incorporation into the nucleosome leads to alteration of chromatin structure and nucleosome stability (40). For example, the incorporation of CENH3 instead of canonical H3 into a nucleosome is associated with the formation of the centromere and pericentric heterochromatin (41, 42).

Chromatin structure is dynamic and highly variable between cell cycle states, in different cell types and at different genomic regions within any given cell. However, it can be more generally classified into either condensed/inaccessible heterochromatin, or open/accessible euchromatin. In the heterochromatic state, transcription factors or other DNA binding proteins have limited access to the DNA, resulting in transcriptional silencing of genes and TEs. In contrast, chromatin is accessible and usually transcriptionally active in the euchromatic state (43).

Histone protein posttranslational modification is an important epigenetic regulation that greatly affects chromatin structure, and also interplays with DNA methylation and small RNAs. Modifications of histone proteins come in various forms (e.g. acetylation, phosphorylation, and methylation) and on different amino acids (5). Histone modifications can directly influence the inter-nucleosomal interactions between histone proteins and DNA, therefore altering chromatin structure. For example, acetylation of lysine residues of histones neutralizes the positive charge of lysine and therefore weakens the electrical attraction between positively charged histone tails and negatively charged DNA, leading to a more open/accessible chromatin state (44, 45). Moreover, histone modifications can also influence chromatin structure without a direct change of inter-nucleosomal interactions. For example, although methylation of lysine residues of histone proteins does not alter the charge of the histone tails, it can recruit histone-modification readers which recognize histone modifications and are capable of changing chromatin structure (45). Histone H3 lysine methylation is one of the most well-understood histone modifications in plants. When looking at transcription, methylation of histone H3 lysine 4 and lysine 36 (H3K4 and H3K36) generally plays an activating role, whereas methylation of H3K9 and H3K27 plays a repressive role (45, 46). In addition, H3K9me2 is enriched in heterochromatic regions containing TEs and other repetitive sequences in plants, suggesting H3K9me2 is possibly associated with transposon repression. Indeed, in *Arabidopsis* H3K9 methyltransferase mutants, *kryptonite* (*kyp* or *suvh4*), a reduced H3K9me2 level is accompanied by reactivation of TEs (47).

Landscape of DNA methylation in plants

DNA methylation in plants exists in three sequence contexts: the symmetrical CG and CHG contexts, and the asymmetrical CHH context, where H is A, T or G. In *Arabidopsis*, average DNA methylation levels are approximately 24%, 6.7% and 1.7% for CG, CHG and CHH contexts, respectively (48, 49). Data from genome-wide analyses show that CG methylation is present in both genes and TEs, whereas CHG and CHH (non-CG) methylation mainly occur in TEs in *Arabidopsis* (48, 49). Unlike in plants, most DNA methylation in human somatic cells occurs exclusively in the CG context, whereas non-CG methylation transiently exists in embryonic stem cells and was found to accumulate in neurons and glial cells of human and mouse brains (50).

DNA methylation in TEs serves as a silencing marker and together with PTGS and chromatin modifications suppresses TEs, maintaining genome integrity. The function of gene body DNA methylation is not clear. In *Arabidopsis*, genes with moderate expression levels tend to have gene body methylation (51). One possible function of gene body methylation is to prevent the deposition of the histone variant H2A.Z, which is associated with gene responsiveness, from gene body chromatin, thus reducing the variability in expression levels

(52). Although gene body methylation is usually not associated with expression silencing, a DNA methylated transcriptional start site is strongly correlated with gene silencing (51). Gene expression is negatively correlated with the density of methylated TEs near to genes, implying that methylated TEs have a deleterious impact on gene expression. As a result of natural selection, methylated TEs are typically located far away from most genes in the genome of *Arabidopsis* (53).

Maintenance of DNA methylation

DNA methylation is conserved between plants and animals, with clearly orthologous maintenance and *de novo* DNA methyltransferases and accessory proteins (54-56). In plants, the process of DNA methylation maintenance has been extensively studied. In the current model, hemimethylated CG sites generated during the process of semiconservative DNA replication are bound by chromatin binding proteins VARIANT IN METHYLATION (VIM), which are orthologs of mammalian ubiquitin-like plant homeodomain and RING finger domain 1 (UHRF1) (55-58). VIM proteins are thought to work in concert with METHYLTRANSFERASE 1 (MET1, an ortholog of mammalian DNMT1) and thus maintain CG methylation during DNA replication (49, 59). The nucleosome remodeler DECREASE in DNA METHYLATION 1 (DDM1) is also required for MET1 to maintain the CG methylation of heterochromatic TEs and the gene body methylation of highly methylated genes (60). The removal of H1 reduces the dependency on DDM1 to maintain DNA methylation, suggesting that DDM1 might provide DNA methyltransferases access to H1-containing heterochromatin (60).

CHG methylation in higher plants is mostly maintained by a self-reinforcing loop between the plant specific chromomethylase 3 (CMT3) and the heterochromatic histone mark H3K9 dimethylation (H3K9me2) (4, 61, 62). In this loop, H3K9 methyltransferase KRYPTONITE (KYP) is recruited by CHG methylation through its SRA domain and catalyzes methylation of H3K9 through its histone methyltransferase domain (63). Likewise, CMT3 is recruited by H3K9 dimethylation through its BAH- and chromo- domains, and maintains CHG methylation through its DNA methyltransferase domain (62, 64). In support of this model, in addition to other *in vitro* evidence, is a genome-wide map of H3K9me2 methylation in *Arabidopsis* suggested a strong correlation between H3K9me2 and CHG DNA methylation (65). Loss of CHG methylation in the H3K9me2 deficient mutant *kyp/suvh5/suvh6* resembles the loss of CHG methylation in *cmt3* mutants (66). In turn, CMT3 binding sites are enriched in heterochromatic regions and are highly correlated with H3K9me2 (64). A similar link between DNA methylation and histone H3 lysine 9 methylation is also observed in *Neurospora crassa* where H3K9me3 deposited by DIM-5 is recognized by Heterochromatin protein 1 (HP1) which further recruits DIM-2 to mediate DNA methylation (67).

Due to its asymmetry, during semiconservative DNA replication, one strand of methylated CHH sites will become completely unmethylated, in contrast to symmetric CG and CHG sites which become only hemimethylated. Without the hemimethylated sites as an informative locator to guide methyltransferases, CHH methylation was thought to be maintained solely by DRM2 through the RDR2-RdDM pathway (4). In addition to DRM2, *Arabidopsis* has two other DRM homologs: DRM1 and DRM3 (68-70). DRM1 is only active during early seed development and probably functions redundantly with DRM2 (68). DRM3 is reported to be catalytically inactive due to the loss of key amino acids. However, just as the mammalian catalytically inactive DNMT3L is required for methylation, DRM3 has reported to be necessary for the full levels of methylation mediated by DRM2 (69). Unlike DNMT3L, which directly interacts with DNMT3 to facilitate DNA methylation catalysis, DRM3 appears to act indirectly, through interaction with Pol V, facilitating DRM2-mediated DNA methylation, probably by promoting Pol V transcriptional elongation or stabilizing Pol V transcripts (69).

Recent studies have shown that although Pol IV-transcribed transcripts display similar transcriptional start sites to Pol II-transcribed transcripts, the length of Pol IV transcripts is surprisingly short (26-45 nt) (71, 72). The short Pol IV-transcribed transcripts suggest that each 24-nt siRNA is likely processed from one Pol IV transcript and Pol IV can likely only transcribe into the edges of long TEs, as Pol IV uses similar transcriptional start sites to Pol II (71, 72). Bohmdorfer *et al.* further found that although the length of Pol V-transcribed scaffold transcripts is longer (median length of 205 nt), Pol V-transcribed transcripts still tend to be enriched for the edges of TEs as Pol IV-transcribed transcripts (72, 73). These results explain why DRM2 preferentially targets short TEs and the edges of long TEs (60). It also suggests that although Pol IV and Pol V are capable of transcribing methylated DNA, they still cannot transcribe into the more heterochromatic TE body of long TEs. In addition to DRM2, CMT3 was also found to partially contribute to the maintenance of CHH methylation (64, 66).

Several studies have suggested the existence of a reinforcing loop between RdDM-mediated non-CG methylation and H3K9 methylation (61, 74-76). A Pol IV-interacting protein SAWADEE HOMEODOMAIN HOMOLOGUE (SHH1) was found to bind to methylated H3K9, indicating that this histone modification can subsequently recruit small RNA producing machinery, thus bridging the gap between RdDM-mediated non-CG methylation and H3K9 methylation (74, 76). Moreover, removing active histone marker H3K4 methylation was shown to be important for the binding of SHH1 (74).

Furthermore, non-CG methylation also forms a reinforcing loop with Pol V-transcribed scaffold RNAs. For the recruitment of Pol V, two catalytically inactive SUVH proteins (SUVH2 and 9) were reported to bind to methylated DNA and then interact with Pol V, resulting

in DNA methylation (75). Support for this mechanism was demonstrated by targeting SUVH2 to an unmethylated locus, which led to the methylation of the unmethylated locus (75). These reinforcing loops together ensure the robust silencing of TEs (Figure 1.3).

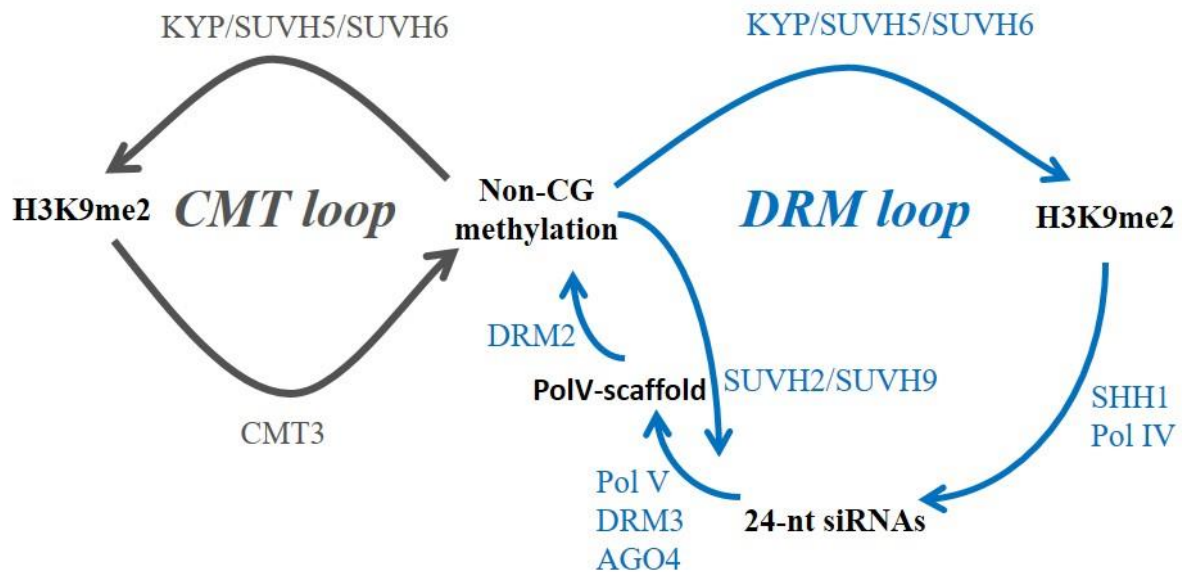


Figure 1.3 The reinforcing loop between non-CG DNA methylation and H3K9me2.

Initiation of *de novo* DNA methylation via the RDR6-RdDM pathway

How is DNA methylation in TEs established in the first place? For example, when a new TE-like element is introduced into a plant genome, e.g. through horizontal transfer or upon virus infection; or when silenced TEs are derepressed and thus leading to transposition under certain circumstances, how does a plant recognize and silence them epigenetically? Due to mutual dependency between non-CG methylation and H3K9 methylation, the RDR2-RdDM pathway is more likely to be in charge of maintenance DNA methylation rather than the *de novo* DNA methylation of such a transcriptionally active and unmethylated element. However, several important studies have shown that RDR6-dependent RdDM (RDR6-RdDM) initializes *de novo* DNA methylation and leads to transcriptional gene silencing (Figure 4) (77-84).

In the RDR6-RdDM pathway, the invading sequences first initialize PTGS as described above. Secondary siRNAs appear to be funneled into AGO6 in addition to PTGS, and then act to guide interactions of the AGO6 complex with scaffold RNA of target loci transcribed by Pol V, leading to the recruitment of DRM2 for target loci methylation (6, 78). An intriguing question arises from this model: how does Pol V transcribe the scaffold RNA from an unmethylated TEs, since it has been shown that the recruitment of Pol V requires SUVH2/9 in order to bind to methylated DNA? One possible explanation is that Pol II can also transcribe the scaffold RNA for RdDM as previously reported (32, 85). The key to initialize RDR6-RdDM in this model was reported to depend on whether the secondary siRNAs can be produced, on what kind of primary siRNAs are produced, and where the cleavage site of TE transcripts

occurs (81). RDR6-dependent RdDM has been shown to target many full-length autonomous TEs when they are transcriptionally activated in *Arabidopsis* (81). In *Arabidopsis*, the nascent miRNA-mediated PTGS can also initialize RDR6-RdDM by cleaving the transcripts of transcriptionally-activated heterochromatic TEs, thus generating epigenetically activated small interfering RNAs (easiRNAs) as secondary RNAs to initialize RDR6-RdDM (82). Virus infection was also shown to induce RDR6-RdDM as supported by a virus-induced gene silencing model in *Arabidopsis* (84). One can expect that once RDR6-RdDM establishes the very first DNA methylation event on the transcribed TEs or virus genome, that RDR2-RdDM, KYP and CMT3 will further be recruited and reinforce the DNA methylation and H3K9me2 to achieve TGS.

An alternative model of this kind of transition from PTGS to TGS was proposed by studying the silencing of an active *EVADE* element in *Arabidopsis*. In this model, the overwhelming mRNAs from proliferated *EVADE* copies saturate DCL2 and DCL4 in the PTGS defense system. DCL3 therefore competes for *EVADE* mRNAs with DCL2 and DCL4 and producing 24-nt siRNAs that directly enter the RDR2-RdDM pathway (Figure 1.4) (86).

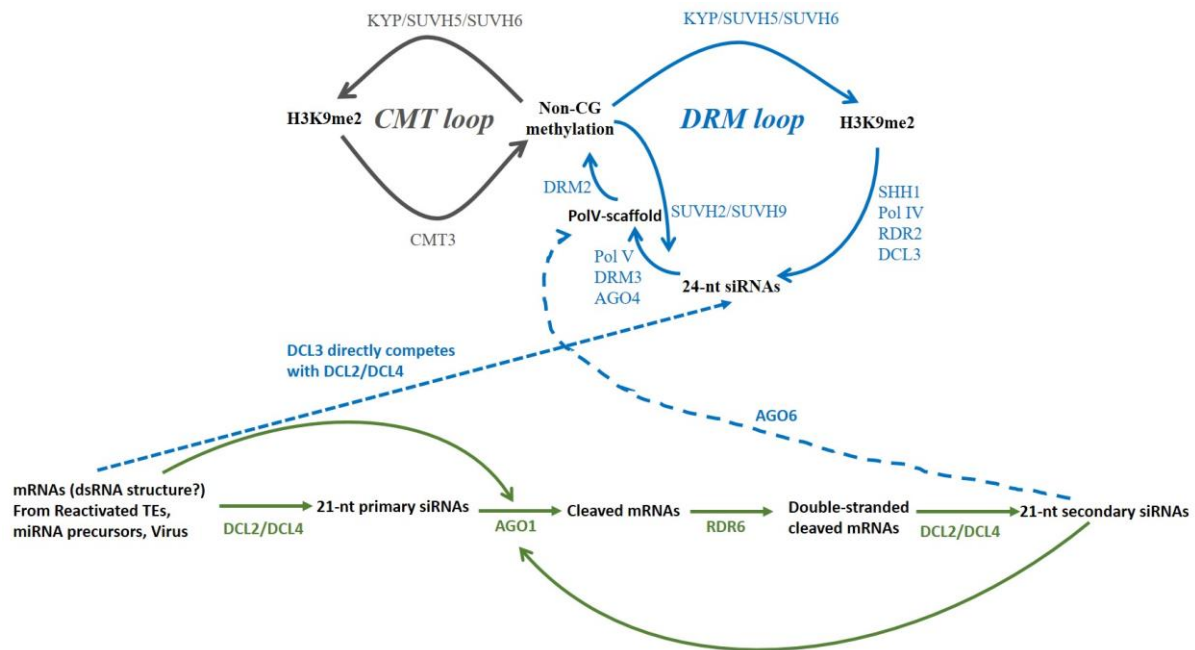


Figure 1.4 The transition from posttranscriptional gene silencing to transcriptional gene silencing

Active DNA demethylation

Given the suppressive nature of DNA methylation, plants need to counteract this and remove DNA methylation at transcriptional start sites or from TEs close to genes, in order to regulate gene expression. In *Arabidopsis*, DNA demethylation occurs by the base excision-repair pathway. A group of DNA glycosylases comprised of REPRESSOR OF SILENCING1 (ROS1), DEMETER (DME), DEMETER-LIKE2 (DML2) and DEMETER-LIKE3 (DML3)

initiate the pathway by excising the 5-methylcytosine base, creating an abasic site. Downstream enzymes, AP endonuclease, DNA polymerase and ligase, repair the abasic site, insert unmethylated cytosine, and seal the gap in the DNA sugar phosphate backbone, respectively. Among them, DME is specifically expressed in the companion cells during sexual reproduction, the central cell and vegetative cell in the female and male gametophytes, respectively (87, 88). Whereas ROS1, DML2 and DML3 are widely expressed in sporophyte leaf and root tissues (88, 89). The glycosylase-mediated active DNA demethylation pathway appears to be conserved in widely divergent plant species including rice and tomato (90, 91).

DME is a required protein for reproduction in *Arabidopsis*, and *dme* mutations cause silencing of the maternally expressed MEDEA (MEA) and FERTILIZATION INDEPENDENT SEED2 (FIS2) proteins in endosperm, resulting in the loss of endosperm gene imprinting, endosperm overproliferation, arrested embryo development and seed abortion (87). Conversely, the expression of most genes is unaffected in the *ros1dml2dml3* (*rdd*) triple mutant, although after several in-bred generations, the *rdd* mutant exhibits pleiotropic phenotypic changes, indicative of gene expression alteration at regions where DNA methylation is no longer adjusted appropriately in the continued absence of ROS1/DML2/DML3 (89). Moreover, ROS1-mediated active DNA demethylation was found to be required for the expression of ERF2, a negative regulator of stomatal formation (92).

How do plants know when to remove excess DNA sporophyte DNA methylation? ROS1 was found to be silenced when RdDM machinery is lost. Thus, a model was proposed whereby ROS1 expression is turned on upon detection that particular RdDM targeting sites become methylated (93, 94). Recently, this RdDM targeting sensor locus was found to be located in the ROS1 promoter and partially overlapped with a Helitron TE (93, 94). Unlike the canonical silencing effect of DNA methylation, ROS1 is activated when the sensor locus is methylated by RdDM and is suppressed when the sensor locus is demethylated by ROS1 itself. Therefore, when the levels of DNA methylation are high and the sensor locus is methylated in a cell, ROS1 is turned on to perform its demethylase activity. After that, ROS1 demethylates its targeting loci including the sensor locus of its promoter, turning off its expression when the high levels of DNA methylation are removed. These properties allow ROS1 to be a rheostat that monitors and adjusts DNA methylation in *Arabidopsis* (93, 94).

It is unclear how the DML glycosylases determine where to remove DNA methylation, although ROS1 was found to preferentially target gene-adjacent TEs, suggesting that ROS1 might protect genes from the suppressive impact of nearby methylated TEs. Furthermore, ROS1-targeted TEs feature high levels of H3K18ac, H3K23ac and H3K27me3, and low levels of H3K27me and H3K9me2 (95, 96). The active acetylated markers of ROS1 targeting sites

require the protein INCREASED DNA METHYLATION 1 (IDM1) (96). Coincidentally, DME-targeted TEs also tend to be small, AT-rich, and nucleosome-depleted euchromatic TEs (97). The histone chaperone SSRP1 is also required for DME-mediated DNA demethylation (98), indicating that the chromatin environment is critical for the targeting of active demethylation.

Epigenetic regulation and plant development

DNA methylation and histone modifications play an important role in the orchestration of varying expression profiles for the genome during the entire plant life cycle, i.e. during seed development, shoot/root apical meristem (SAM/RAM) regulation, floral transition (from vegetative phase to reproductive phase), gametogenesis, and embryogenesis (97, 99-102). For example, the induction of stem cell fate determinant gene WUSCHEL (WUS) in *Arabidopsis* callus correlates with increased levels of activating marks H3K4me3 and H3K9ac and decreased levels of repressive H3K9me2 in the regulatory regions of WUS (103). The expression of WUS is also higher in *met1* callus, suggesting DNA methylation is also involved in the regulation of WUS expression (103).

Unlike the mammalian germline, which is established during embryogenesis, plant germlines differentiate from the SAM in later stages of the plant life cycle (99, 104). FLOWERING LOCUS C (FLC) plays a key role in regulating the floral transition. In the vegetative phase, FLC is expressed, repressing the floral pathway integrators including SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), LEAFY (LFY), FLOWERING LOCUS T (FT) and FLOWERING LOCUS D (FD) (99, 104). When the floral transition is induced by external signals, FLC is repressed via histone modifications and thus the floral pathway integrators are derepressed, triggering floral transition (105). During this process, the SAM develops into the inflorescence meristem (IM), and is followed by the formation of floral meristem (FM) in the flank of IM, whereby the FM goes on to produce the reproductive organs (99, 104).

Epigenetic regulation during sexual reproduction

In *Arabidopsis*, the germline generation process consists of two parts: sporogenesis and gametogenesis (100). During sporogenesis, certain cells in the reproductive tissues differentiate into meiocytes, which undergo meiosis to produce haploid microspores (1N) and macrospores (1N) in the male and female reproductive tissues, respectively. Following gametogenesis, the microspore undergoes two rounds of mitoses to generate sperm cells. In the first mitosis, the microspore divides asymmetrically to generate a vegetative cell (1N), which stops dividing at this point, and a generative cell (1N). In the second mitosis, the generative cell gives rise to two sperm cells and the sperm cells are then engulfed into the cytoplasm of the vegetative cell,

forming the mature pollen grain – the male gametophyte (106).

In contrast, the megaspore undergoes three rounds of mitoses leading to a syncytial premature female gametophyte with eight nuclei. The premature female gametophyte further develops into a seven-cell mature gametophyte, including two synergid cells (1N), three antipodal cells (1N), one egg cell (1N), and one central cell (2N) which is formed by the fusion of two haploid polar nuclei (107). Flowering plants produce their seeds through a special double fertilization method (108, 109). After pollination, the pollen grain rehydrates and the vegetative cell germinates to produce a pollen tube that delivers two sperm cells to the female gametophyte. Then, one sperm cell fertilizes the egg cell to generate the diploid embryo, and the other sperm cell fertilizes the central cell to form the triploid endosperm (108, 109).

One would expect a dynamic transcriptional change accompanied by a large scale epigenetic change during these sophisticated processes. Indeed, cytological studies have shown the chromatin states of these sexual reproductive cells experience dramatic changes during sexual reproduction. For the meiocytes, chromatin decondensation was observed cytologically in the male meiocyte, accompanied by reduced linker histone H1-GFP. Consistent with the decondensed chromatin state, an increased cytological signal of active histone modification H3K4me3 and a decreased cytological signal of repressive histone modifications H3K27me1 and H3K27me3 were observed in the male meiocyte (110). Similarly, chromatin decondensation state and reduced H1-GFP were also observed in the female meiocyte (111).

In addition to the meiocytes, gametophytes also show dynamic chromatin changes. For example, the sperm cell displays highly condensed chromatin whereas the vegetative cell shows decondensed chromatin (112). Consistent with chromatin decondensation in the vegetative cell, the loss of histone variant CenH3 and H3K9me2 foci were also observed in vegetative cells. The mutation of a chaperone protein Cdc48/p97 was found to block the loss of CenH3 and H3K9me2 foci and prevent centromeric heterochromatin decondensation in vegetative cell nuclei, suggesting Cdc48/p97 might actively remove H3K9me2 and switch CenH3 in the vegetative cell (113). However, these studies were all done at the cytological level. More detailed molecular examination of the chromatin changes will be needed for further understanding how chromatin dynamics can regulate sexual reproduction.

In addition to the change of chromatin structure, two studies suggest that siRNAs from somatic cells are critical for gamete development during plant sexual reproduction. AGO9 was found to be specifically expressed in ovule companion somatic cells surrounding female meiocytes during sexual reproduction in *Arabidopsis* and maize (114, 115). The function of AGO9 is thought to be in generating 24-nt siRNA from ovule companion somatic cells to help

silence TEs in egg cells in a non-cell-autonomous manner during sexual reproduction. Additionally, AGO9 was also found to be critical for the cell fate specification of megaspore and meiosis during sexual reproduction (114, 115).

The changes in DNA methylation during sexual reproduction, especially male gametogenesis, have been extensively examined by single-base resolution bisulfite sequencing. Studies have shown that plant germlines undergo much less DNA methylation reprogramming compared to mammalian germlines (116). However, a gain of heterochromatic CHH methylation and a loss of heterochromatic CHH methylation were observed in vegetative cells and sperm cells, respectively (102). The gain of heterochromatic CHH methylation in the vegetative cell was attributed to active RdDM-mediated CHH methylation because DRM2 was shown to be active in the vegetative cell (102). In the vegetative cell, derepressed heterochromatic TEs were suggested to generate abundant 21-nt easiRNAs and these easiRNAs were thought to contribute to RdDM-mediated CHH methylation, leading to a gain of CHH methylation in the vegetative cell (102).

In contrast to this, the loss of CHH methylation in the sperm cell was thought to be caused by DRM2 silencing (102). In addition to gaining CHH methylation, vegetative cells also exhibit local CG hypomethylation in short and euchromatic TEs, which was shown to be mediated by DME (97, 102). It was proposed that DME-mediated DNA demethylation activates TEs in the vegetative cell, and that DME-activated TEs generate siRNAs that are exported to the sperm cell and reinforce DNA methylation of corresponding TEs (97). A similar model was also proposed for the female gametophyte where siRNAs from the central cell reinforce DNA methylation of corresponding TEs in the egg cell (97). After fertilization, DRM2 starts to be expressed in the zygote, and thus CHH methylation regained in the embryo was thought to be mediated by RdDM guided by the siRNAs imported from the companion cells (i.e. vegetative cells and central cells) (68).

Summary

Since the first single-base resolution methylome experiments in *Arabidopsis* published in 2008 (48, 49), our understanding of plant DNA methylation has grown rapidly. Mechanisms for maintaining the three types of DNA were represented by the model: CG methylation is mediated by MET1, CHG is mediated by CMT3 and CHH methylation is primarily maintained by DRM2 through RdDM (4, 48, 49). There are also many fundamental studies expanding the network of this model through their discovery of interacting partners and accessory proteins (66). Many studies have shown how important DNA methylation is during plant development and in their response to environmental cues. However, the following questions about the maintenance and inheritance of DNA methylation were still unanswered.

First, the maintenance model for CHH methylation seemed to be incomplete because a substantial amount of CHH methylation remains in the *drm1drm2* double mutant and even in the *drm1drm2cmt3* triple mutant. This suggested that CHH methylation is maintained by enzymes other than DRM1/2 and CMT3 (48, 49).

Second, although the decondensed chromatin in the vegetative cell might be more accessible and favorable for RdDM, it was unknown how RdDM can dramatically change its preference to target these heterochromatic TEs. This was because the model that RdDM is responsible for the reprogramming of CHH methylation during sexual reproduction was built upon the presumption that DRM2 is the primary DNA methyltransferase maintaining CHH methylation, which was incomplete.

Third, the semiconservative model of CG methylation maintenance implies that even a modest reduction in maintenance efficiency below 100% should lead to dilution of methylation with each cell division that eventually causes complete loss of DNA methylation. Although, CG methylation levels measured in somatic cells appear to be below what would be required for stable maintenance, transgenerational losses of CG methylation in wild type *Arabidopsis* plants are not observed (48, 49), implying the existence of an unknown transgenerational DNA methylation inheritance mechanism.

In this thesis, I will address these questions by using genomic and genetic approaches as well as a fluorescence activated cell-sorting technique.

Chapter II

The nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin

Most of the following chapter has been published as part of a peer reviewed article in the *Cell*, USA:

Assaf Zemach, M. Yvonne Kim, Ping-Hung Hsieh, Devin Coleman-Derr, Leor Eshed-Williams, Ka Thao, Stacey L. Harmer, and Daniel Zilberman (2013). The nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*, USA, 153(1):193-205. doi: 10.1016/j.cell.2013.02.033.

Copyright is retained by the authors.

Contributions:

Daniel Zilberman conceived and designed the experiment and provided discussion and advice at various stages. Devin Coleman-Derr, Leor Eshed-Williams, Ka Thao and Stacey L. Harmer provided laboratory and data analysis assistance. Assaf Zemach, M. Yvonne Kim and Ping-Hung Hsieh contributed equally to experimental design, data analysis, molecular biology experiments.

Sequencing data are deposited in GEO with accession number GSE41302.

Summary

Nucleosome remodelers of the DDM1/Lsh family are required for DNA methylation of transposable elements, but the reason for this is unknown. How DDM1 interacts with other methylation pathways, such as small RNA-directed DNA methylation (RdDM), which is thought to mediate asymmetric methylation through DRM enzymes, is also unclear. Here, we show that most asymmetric methylation is facilitated by DDM1 and mediated by the methyltransferase CMT2 separately from RdDM. We find that heterochromatic sequences preferentially require DDM1 for DNA methylation, and that this preference depends on linker histone H1. RdDM is instead inhibited by heterochromatin and absolutely requires the nucleosome remodeler DRD1. Together, DDM1 and RdDM mediate nearly all transposon methylation, and collaborate to repress transposition and regulate the methylation and expression of genes. Our results indicate that DDM1 provides DNA methyltransferases access to H1-containing heterochromatin to allow stable silencing of transposable elements in cooperation with the RdDM pathway.

Introduction

DNA methylation in flowering plants occurs in three sequence contexts: CG, CHG and CHH, where H is any nucleotide except G. Methylation in each context is believed to be primarily catalyzed by a specific family of DNA methyltransferases: MET1 (homologous to animal Dnmt1) for CG, chromomethylases (CMT) for CHG, and DRM (homologous to animal Dnmt3) for CHH (117). The majority of plant methylation is found in transposable elements (TEs), where methylation occurs in all sequence contexts and is crucial for the repression of TE expression and transposition (117). Substantial methylation is also found in the bodies of active genes, where methylation is generally restricted to the CG context (51, 117).

The establishment of plant DNA methylation in all sequence contexts, and the maintenance of CHH methylation is mediated by a specialized branch of the RNA interference pathway referred to as RNA-directed DNA methylation (RdDM) (117). RdDM relies on two plant-specific homologs of RNA polymerase II – pol IV and pol V. The pol IV branch of RdDM is thought to synthesize the long RNA molecules that are made double-stranded by RNA-dependent RNA polymerase 2 (RDR2) and processed by Dicer-like nucleases into small interfering RNA (sRNA). RNA pol V and associated factors are believed to produce nascent transcripts from target loci that are recognized by sRNA-containing AGO4 complexes that target DNA methylation via DRM enzymes (118).

DNA methylation is also influenced by chromatin factors: CHG methylation by CMT3 requires dimethylation of histone H3 at lysine 9 (H3K9me₂), to which CMT3 binds via its chromo and bromo adjacent homology domains (64, 117), and CHG methylation is kept out of genes by a histone demethylase, IBM1, which removes H3K9me₂ from gene bodies (119). A more enigmatic chromatin factor that is essential for normal DNA methylation is the Snf2 family nucleosome remodeler DDM1 (120, 121). Snf2 remodelers hydrolyze ATP to move along DNA, altering nucleosome composition and placement and allowing other proteins to access the DNA (122). DDM1 can shift nucleosomes *in vitro* (123), and its mutation has been reported to cause a profound loss of methylation from some TEs and repeats (120), but not from genes (121). Mutation of Lsh, the mouse homolog of DDM1, causes a similar methylation phenotype (124), indicating that DDM1 remodelers are ancient components of the DNA methylation pathway. The loss of DDM1 leads to strong transcriptional activation of TEs (121), and inbred *ddm1* mutant lines have increased rates of transposition (125). sRNAs correspond to the TEs hypomethylated in *ddm1* mutant plants, leading to the suggestion that DDM1 participates in RdDM (121). However, DDM1 and RdDM synergize to silence rDNA loci (126), indicating that RdDM can function without DDM1, and DDM1 can mediate CHH methylation independently of RdDM at some TEs (127). Thus, the related questions of how DDM1 interacts with the methyltransferase pathways, including RdDM, and why some sequences require

DDM1 for methylation and others do not, remain largely unanswered.

Results

DDM1 and RdDM separately mediate nearly all DNA methylation in TEs

To understand how DDM1 mediates DNA methylation, we quantified genomic methylation of *ddm1* mutant *Arabidopsis thaliana* plants (Table 2.1).

	Nuclear					Chloroplast
Sample	Mean coverage	C	CG	CHG	CHH	CHH
WT roots	13.69	6.68%	25.82%	8.05%	2.70%	0.16%
WT shoots	3.08	6.48%	25.99%	8.76%	2.49%	0.05%
<i>drd1-7</i>	22.61	5.78%	25.51%	7.19%	1.82%	0.04%
<i>rdr2-1</i>	39.76	5.61%	25.16%	7.16%	1.80%	0.20%
<i>ddm1-2</i>	13.50	3.32%	10.85%	3.48%	1.84%	0.20%
<i>ddm1drd1</i>	10.42	2.08%	10.79%	1.65%	0.45%	0.02%
<i>ddm1rdr2</i>	85.37	2.01%	10.95%	1.74%	0.46%	0.11%
<i>met1-6</i>	15.50	2.28%	1.02%	6.78%	1.59%	0.05%
<i>cmt3-12</i>	6.11	4.96%	26.30%	1.12%	1.87%	0.18%
<i>cmt2-3</i>	3.44	5.20%	25.94%	7.76%	0.75%	0.16%
<i>cmt2-4</i>	9.15	5.42%	26.67%	8.56%	0.991%	0.34%
<i>cmt2-5</i>	7.61	5.45%	26.18%	8.82%	1.05%	0.35%
<i>cmt2-6</i>	8.15	5.41%	26.49%	8.88%	1.01%	0.35%
<i>drm2-3</i>	4.65	5.66%	26.09%	7.55%	1.61%	0.17%
<i>ibm1drd1</i>	75.49	6.1%	25.4%	11.2%	1.6%	0.11%
H1;DDM1	26.85	6.0%	23.1%	8.2%	2.6%	0.6%
H1; <i>ddm1</i>	22.09	3.2%	12.6%	3.1%	1.6%	0.4%
<i>h1;ddm1</i> replicate 1	17.27	4.5%	18.3%	6.1%	1.8%	0.1%
<i>h1;ddm1</i> replicate 2	18.47	4.5%	18.2%	5.9%	1.8%	0.3%
<i>h1</i> ;DDM1	26.29	5.5%	22.3%	8.6%	1.9%	0.1%
H1 roots	11.82	6.54%	26.50%	8.39%	2.61%	0.22%
<i>h1</i> roots	8.89	6.89%	27.12%	10.69%	2.54%	0.13%

Table 2.1 Mean genomic coverage and DNA methylation for wild-type and mutant samples, related to Chapter II. *drd1*, *rdr2*, *ddm1* and *met1* samples are from roots; *ddm1drd1*, *ddm1rdr2*, *cmt3*, *cmt2*, *drm2*, *ibm1*, *ibm1drd1*, *H1DDM1*, *H1ddm1*, *h1ddm1* (both replicates), and *h1DDM1* samples are from 2-3 week shoots. Chloroplast CHH methylation is a measure of cytosine non-conversion and other errors.

Lack of DDM1 caused a 58%, 57% and 32% overall reduction of CG, CHG and CHH methylation, respectively (Figure 2.1), reflected in much lower TE methylation in all sequence contexts (Figure 2.2A). The strong loss of CHH methylation might be caused by a dependence of RdDM on methylation in other contexts, as proposed to explain the loss of non-CG methylation in *met1* mutant plants (Figure 2.1) (49), or may indicate that a large fraction of plant CHH methylation is mediated by DDM1 independently of RdDM (127).

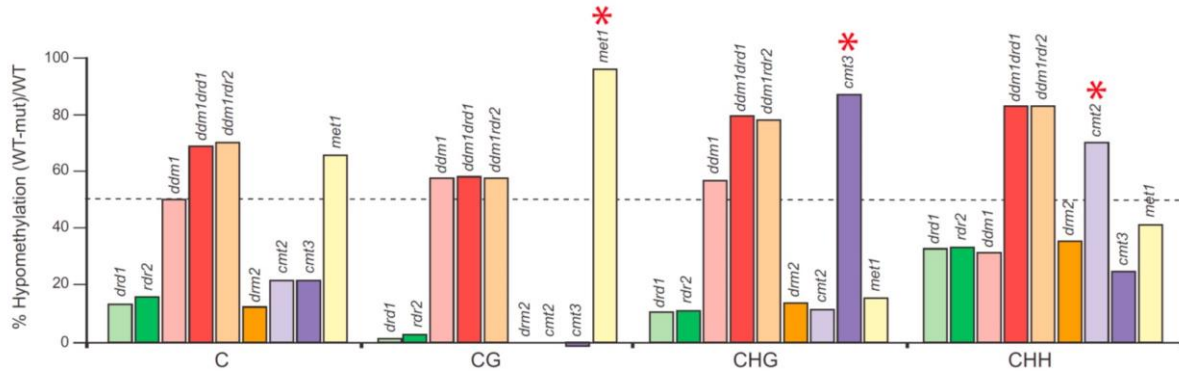


Figure 2.1 Genomic hypomethylation in mutant lines shown as the percentage of wild-type methylation. Methyltransferase mutant lines with the strongest methylation loss in a given context are denoted by asterisks. Mean genomic methylation levels can be found in Table2.1.

To answer this question, we determined DNA methylation in plants with a mutation in RdDM pathway gene *RDR2*, which is required for the production of all endogenous sRNA molecules (128). We also analyzed methylation in plants lacking *DRD1*, a Snf2 remodeler that positively regulates RdDM at a number of loci (129, 130) and forms a complex with RdDM pathway proteins that is thought to facilitate pol V activity (130, 131). Lack of *RDR2* and *DRD1* caused a relatively modest loss of CHH methylation (Figures 2.1 and 2.2A), demonstrating that the majority of CHH methylation does not require RdDM (132). Combining the *ddm1* mutation with either *rdr2* or *drd1* caused a nearly complete loss of CHH methylation (Figures 2.1 and 2.2A), as well as of CG and CHG methylation in TEs (Figure 2.2A), demonstrating that a great deal of *Arabidopsis* CHH methylation is mediated by DDM1 separately from RdDM, and that the two pathways together are responsible for almost all DNA methylation of TEs. The methylation phenotypes of *drd1* and *ddm1drd1* mutants are virtually indistinguishable from those of *rdr2* and *ddm1rdr2*, respectively (Figure 2.2A), indicating that *DRD1* is absolutely required for RdDM.

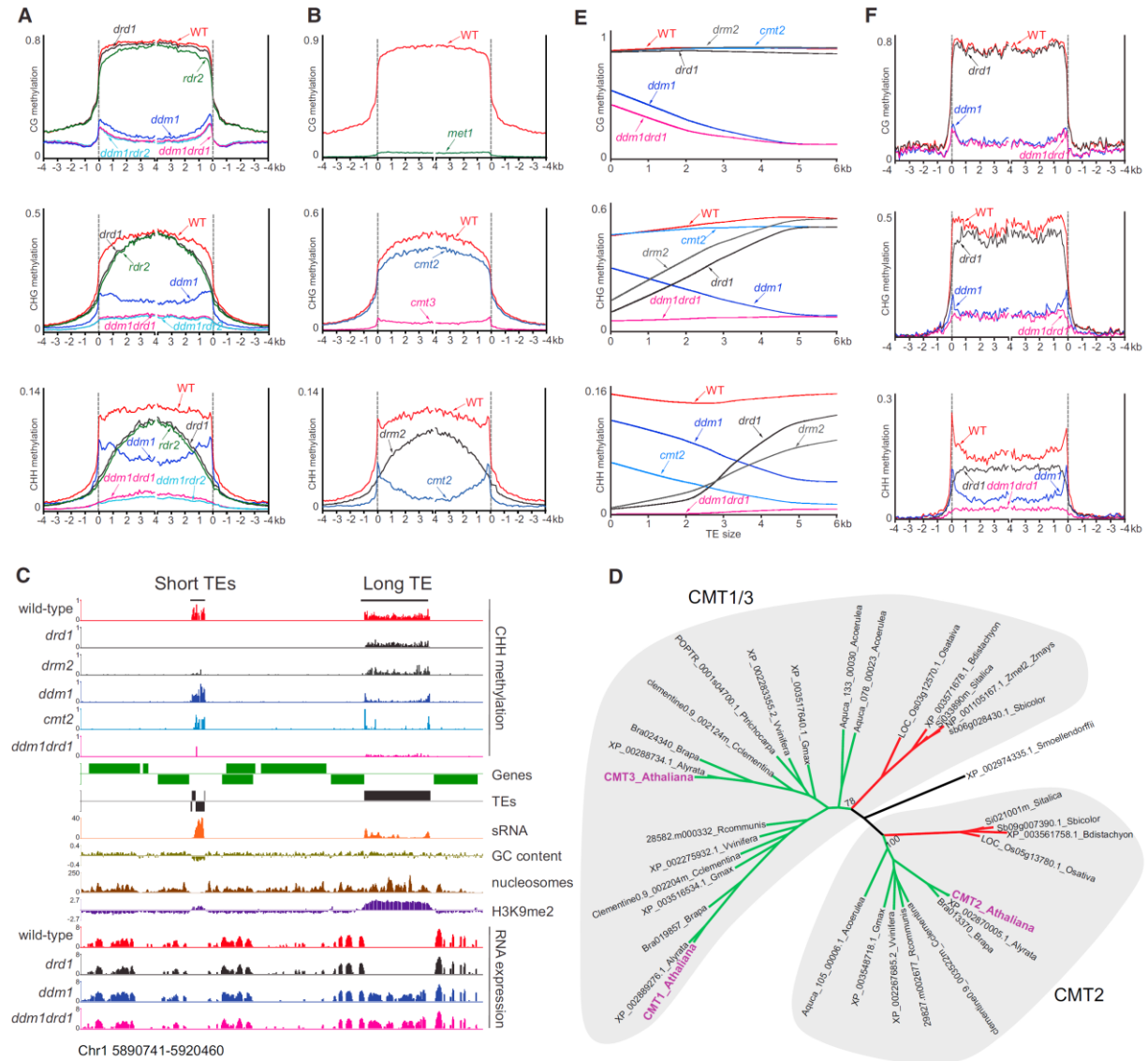


Figure 2.2 DDM1 and CMT2 mediate RdDM-independent CHH methylation of long TEs.

(A) Patterns of TE DNA methylation (CG, CHG and CHH) in wild-type and indicated mutants. *Arabidopsis* TEs were aligned at the 5' end or the 3' end, and average methylation for all cytosines within each 100-bp interval is plotted. The dashed lines represent the points of alignment. (B) Patterns of TE methylation in *met1*, *cmt3*, *cmt2* and *drm2* plants. (C) CHH methylation, sRNA, GC content, nucleosomes, H3K9me2 and RNA levels of a representative region. Genes and TEs oriented 5' to 3' and 3' to 5' are shown above and below the line, respectively. (D) Phylogenetic tree of angiosperm chromomethylases, with *Selaginella moellendorffii* (black) as an outgroup. Dicots are shown in green and monocots in red. (E) LOWESS fit of CG, CHG and CHH methylation levels in wild-type and indicated mutants calculated in 50-bp windows and plotted against TE size. (F) DNA methylation in wild-type and indicated mutants was averaged specifically in long TEs (> 4 kb) as described in (A).

DDM1-dependent CHH methylation is catalyzed by CMT2

The presence of extensive RdDM-independent CHH methylation raises the question of which DNA methyltransferase is responsible. Previous studies have demonstrated that MET1 and CMT3 mediate virtually all *Arabidopsis* CG and CHG methylation, respectively (49, 133), and our data confirm these results (Figures 2.1 and 2.2B). However, even plants lacking DRM1 (which appears to be specifically active during early seed development (68)), DRM2, and CMT3 have substantial residual CHH methylation (49, 133), indicating that another methyltransferase must be involved. Mutation of *DRM2* causes CHH methylation loss that closely resembles that in RdDM mutants (Figures 2.2A, 2.2B and 2.2C), consistent with the established link between DRM2 and RdDM (117).

We thus decided to perform a reverse genetic screen on *Arabidopsis* mutants to identify genes that can potentially affect plant DNA methylation by performing whole-genome bisulfite sequencing of each individual homozygous mutant seedling. We sequenced more than 250 mutants for genes which were predicted in the Chromatin Database website (<http://www.chromdb.org/>) to be associated with chromatin in some way, such as DNA methyltransferases, RdDM components, chromatin remodeling proteins, and histone modification enzymes. Unexpectedly, lack of CMT2, a homolog of CMT3 (Figure 2.2D), has little impact on CHG methylation but causes a major loss of CHH methylation (Figures 2.2B). The observation that DRM2 accounts for RdDM (Figures 2.2A, 2.2B and 2.2C) indicates that CMT2 is responsible for the DDM1-mediated, sRNA-independent CHH methylation (Figure 2.2A). In support of this conclusion, residual methylation in *cmt2* plants is correlated with that in *ddm1* and anticorrelated with that in RdDM mutants (Table 2.2), whereas residual methylation in *drm2* plants (mediated by CMT2) is uncorrelated with sRNA abundance (Table 2.2). CMT2 appears to have evolved prior to the radiation of angiosperms (Figure 2.2D) to methylate a different sequence context than canonical chromomethylases (51, 64, 133).

	<i>cmt2</i>	sRNA	<i>drd1</i>	<i>rdr2</i>	<i>drm2</i>
sRNA	0.115				
<i>drd1</i>	-0.2496	-0.0571			
<i>rdr2</i>	-0.2286	-0.0419	0.7128		
<i>drm2</i>	-0.1407	-0.0029	0.5181	0.493	
<i>ddm1</i>	0.4022	0.1053	-0.1109	-0.136	-0.0178

Table 2.2. Pearson correlation coefficients between CHH methylation in mutants and sRNA.

Pearson correlation coefficients were calculated between CHH methylation in the indicated mutants and sRNA in 50-bp windows that correspond to TEs.

DDM1 and RdDM mediate methylation of distinct TE sizes and domains

The *ddm1* and *drd1* mutant lines exhibit a similar absolute level but very different patterns of CHH methylation loss (Figures 2.1 and 2.2A). The *drd1* mutation strongly reduces TE CHH methylation near the points of alignment, whereas *ddm1* has a larger effect away from the points of alignment (Figure 2.2A). The same distinction is evident for CG and CHG methylation (Figure 2.2A), and for the CHH methylation phenotypes of *drm2* and *cmt2* (Figure 2.2B). Lack of DDM1 and RdDM thus affects TEs very differently.

More than 80% of annotated *Arabidopsis* TEs are shorter than 1000 bp (134), and such TEs would only contribute to the patterns of TE methylation shown in Figures 2.2A and 2.2B close to the points of alignment, suggesting that the differences between *ddm1* and *drd1*, as well as between *cmt2* and *drm2*, may be caused by differential hypomethylation of short and long TEs. Indeed, the *ddm1* and *cmt2* hypomethylation effects are positively correlated with TE size, whereas *drd1* and *drm2* hypomethylation is negatively correlated with TE size (Figures 2.2C and 2.2E). Pericentric heterochromatin and chromosome arms are enriched for long and short TEs, respectively (135), and consequently DDM1 preferentially mediates DNA methylation near the centromeres, whereas RdDM predominantly functions along the chromosome arms.

The methylation patterns in Figure 2.2A are also consistent with DDM1 and DRD1 mediating methylation differently at the edges and inside the bodies of TEs. To examine this issue, we averaged methylation across TEs longer than 4 kb, so that short TEs would not influence the methylation level near the points of alignment. The hypomethylation induced by *drd1* is strongest at TE edges, whereas *ddm1* hypomethylation is greater within TE bodies (Figures 2.2C and 2.2F). The little remaining non-CG methylation in the *ddm1drd1* double mutant is evenly distributed across the entire TE sequence (Figure 2.2F). Similarly, *cmt2* nearly eliminates CHH methylation of TEs longer than 4 kb except at the edges, where CHH methylation is mediated by DRM2 (Figures 2.2C). Taken together, our results indicate that DDM1 is preferentially required for DNA methylation within the bodies of long TEs, which is catalyzed by MET1 (CG), CMT3 (CHG) and CMT2 (CHH), whereas RdDM mostly targets short TEs and TE edges through DRM2 (Figures 2.2C, 2.2E and 2.2F), consistent with the observed enrichment of pol V at such sequences (131).

Heterochromatin requires DDM1 for DNA methylation and inhibits RdDM

Because DDM1 can remodel nucleosomes, we asked whether chromatin features are responsible for the differential requirement of DDM1 for DNA methylation. Sequence composition is thought to be a major determinant of the nucleosome landscape (136), and indeed *ddm1* TE demethylation in all sequence contexts is strongly correlated with nucleosome

occupancy and GC content (Figure 2.3A). Short TEs and TE edges are relatively AT-rich and nucleosome-depleted (Figures 2.3B, 2.3C and 2.4A), consistent with the preferential requirement of DDM1 to maintain DNA methylation in the bodies of long TEs (Figures 2.2E, 2.2F and 2.3A).

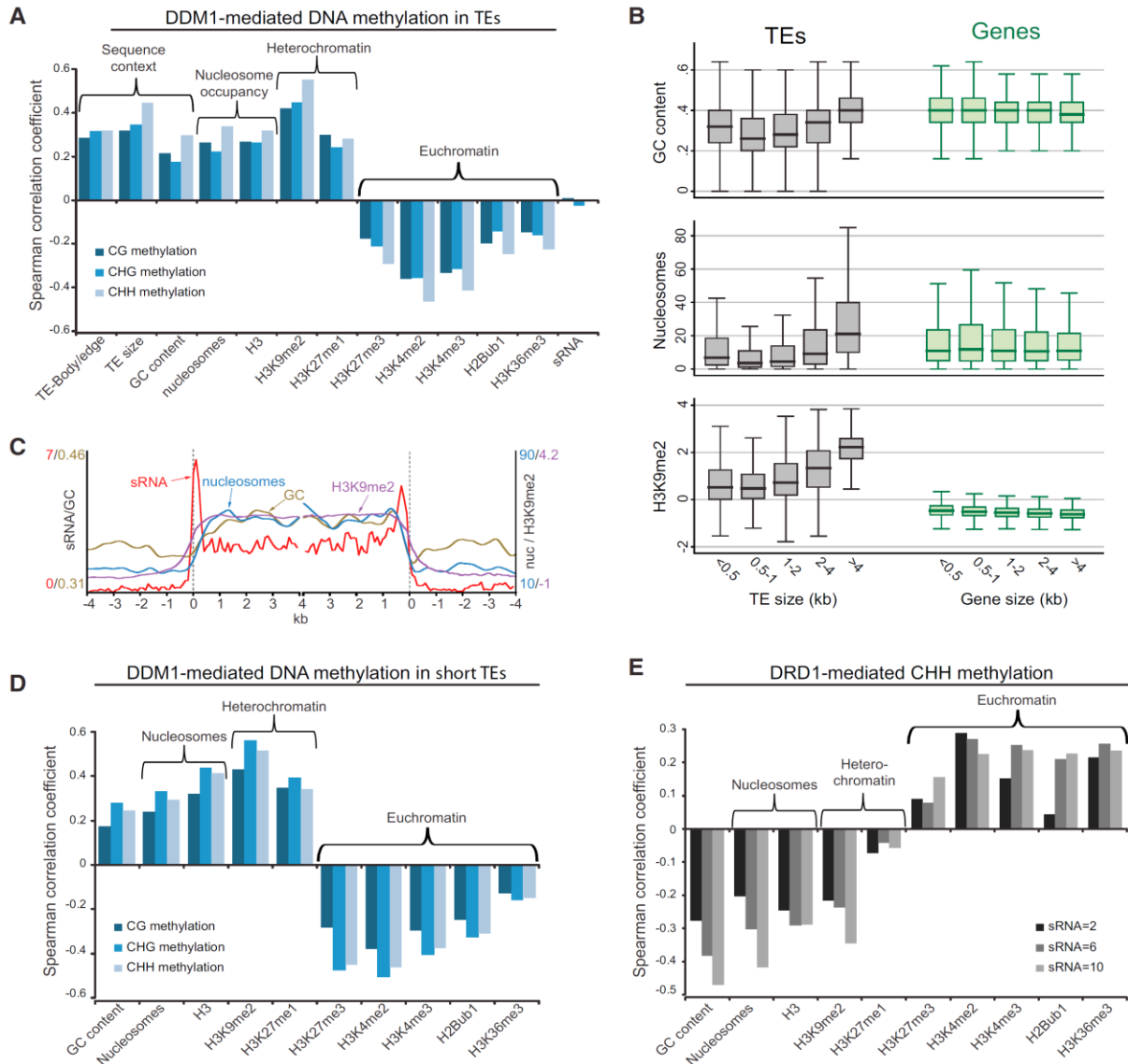


Figure 2.3. Heterochromatin requires DDM1 for DNA methylation and inhibits RdDM.

(A) Spearman correlation coefficients between DDM1-mediated methylation (drd1 methylation minus ddm1drd1 methylation) and DNA sequence and chromatin features of TEs in 50-bp windows. (B) Box plots showing GC content, nucleosome occupancy and H3K9me2 levels in 50-bp windows within TEs and genes of the indicated size. (C) sRNA, GC content, nucleosome occupancy and H3K9me2 levels were averaged in long TEs (> 4 kb) as described in Figure 1. (D) Spearman correlation coefficients between DDM1-mediated methylation in short TEs (<500 bp) and chromatin features. (E) Spearman correlation coefficients between DRD1-mediated CHH methylation (ddm1 methylation minus ddm1drd1 methylation) and chromatin features, calculated for 50-bp windows with three different levels of sRNA.

However, nucleosome occupancy alone is unlikely to determine the dependence of DNA methylation on DDM1 because genic GC content is similar to that of long TEs and nucleosome occupancy in genes is higher than in most TEs (Figure 2.3B), presumably because genes and long TEs are composed largely of protein-coding sequences. The properties of nucleosomes in euchromatic genes do differ from those of heterochromatic TEs, as exemplified by different sets of posttranslational histone modifications (137). Somewhat unexpectedly, we find that H3K9me2, the most well-studied histone modification associated with plant heterochromatin, is enriched in the bodies of long TEs compared to short TEs (Figures 2.3B and 2.3C). More generally, DDM1-mediated TE DNA methylation – the residual methylation in the *drd1* mutant – is correlated with heterochromatic histone modifications, anticorrelated with euchromatic ones, and not correlated with sRNA abundance (Figures 2.3A and 2.4B). Furthermore, the dependence of short TE methylation on DDM1 also correlates with GC content, nucleosome occupancy and histone modifications (Figure 2.3D) just like methylation of all TEs (Figure 2.3A), indicating that the chromatin environment, rather than TE size, ultimately determines the extent to which DDM1 is required for maintenance of DNA methylation.

RdDM would be expected to occur at sRNA-associated loci. Indeed, DRD1-mediated methylation – the residual methylation in the *ddm1* mutant – is positively correlated with sRNA abundance (Figure 2.4C), and TE edges, which are hypermethylated at CHH sites in comparison to internal sequences in wild-type plants (Figure 2.2F), are preferentially targeted by sRNA (Figure 2.3C) (138). However, sRNA molecules are also abundantly derived from TE bodies (Figure 2.3C). To understand how chromatin structure affects RdDM, we analyzed DRD1-mediated DNA methylation at sequences with similar levels of sRNA. With sRNA levels held constant, DRD1-mediated methylation is negatively correlated with GC content, nucleosome occupancy and heterochromatic histone modifications (Figures 2.3E, 2.4B and 2.4D). Thus, RdDM is inhibited by heterochromatin, as has been suggested by (112).

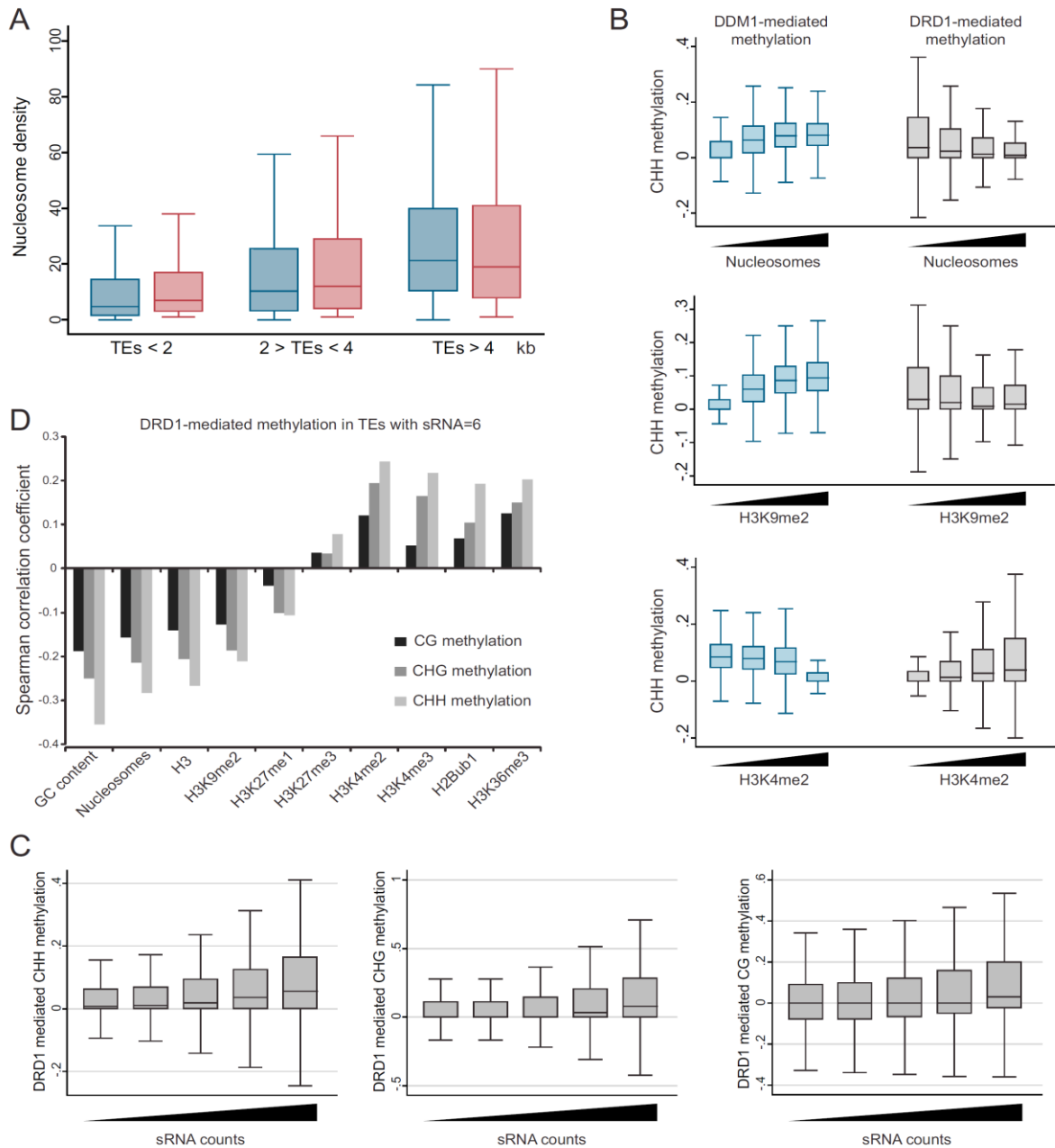


Figure 2.4. Correlations between DNA methylation dependence and chromatin features

(A) Box plots of nucleosome occupancy from this study (red) and from (139) (blue) of 50-bp windows within TEs of different sizes. (B) Box plots of DDM1- and DRD1-mediated TE CHH methylation within 50-bp windows with sRNA level of 6 counts and different levels of nucleosome counts (left to right: 0 to 6, 7 to 16, 17 to 34 and 34 to 76), H3K9me2 (left to right: log2 of -1 to 1.2, 1.2 to 1.9, 1.9 to 2.5, and 2.5 to 3.9), and H3K4me2 (left to right: log2 of -4 to -2, -2 to -1.5, -1.5 to -0.7, and -0.7 to 1). (C) Box plots of DRD1-mediated TE methylation within 50-bp windows with different levels of sRNA (left to right: 0, 1 to 2, 3 to 6, 7 to 15, and over 15). (D) Spearman correlation coefficients between DRD1-mediated TE DNA methylation (*ddm1* DNA methylation minus *ddm1drd1* DNA methylation) and chromatin features within 50-bp windows with sRNA level of 6 counts.

Histone H1 mediates the dependence of heterochromatic DNA methylation on DDM1

Our results indicate that DDM1 is preferentially required for methylation of heterochromatic sequences in all contexts, most likely by allowing methyltransferases access to the DNA. To determine which component of heterochromatin blocks enzyme access, we examined histone H1, which binds to the nucleosome core and the linker DNA that separates nucleosomes (140), condenses chromatin and inhibits nucleosome mobility and transcription *in vitro* (141, 142), and is associated with more compact, less accessible and transcriptionally silent chromatin *in vivo* (143-146). Loss of H1 has been reported to cause disparate changes in genomic DNA methylation: mice with reduced H1 specifically lose DNA methylation at the regulatory regions of several imprinted genes (145), whereas loss of H1 leads to extensive hypermethylation in the fungus *Ascobolus immersus* (144) and apparently stochastic methylation gains and losses in *Arabidopsis* (147).

Plants with mutant alleles in the two canonical *Arabidopsis* *H1* genes (147, 148) exhibit a complex DNA methylation phenotype. Euchromatic TEs (those with low H3K9me2) lose DNA methylation in *h1* (Figure 2.5A), whereas H3K9me2-rich heterochromatic TEs exhibit a global increase of DNA methylation (Figures 2.5A and 2.5B), supporting our hypothesis that H1 impedes access of DNA methyltransferases to heterochromatin. Loss of H1 almost completely suppresses the reduction of TE CHH methylation in *ddm1*, and greatly ameliorates the reduction of TE CG and CHG methylation (Figures 2.5B and 2.5C). Most strikingly, H3K9me2-rich heterochromatic TEs are not preferentially hypomethylated in *h1ddm1* plants, as they are in *ddm1* (Figures 2.5C). Instead, *h1ddm1* causes heterochromatic TEs to lose less DNA methylation than more euchromatic TEs (black traces in Figure 2.5C), similar to *h1* (green traces in Figure 2.5C), indicating that the loss of DDM1 affects euchromatic and heterochromatic TEs similarly when H1 is not present (Figure 2.5D). Lack of H1 still destabilizes the methylation of euchromatic TEs when combined with *ddm1*, but heterochromatic TEs are methylated rather efficiently when both DDM1 and H1 are absent (Figures 2.5C). Our results indicate that the differential importance of DDM1 for the maintenance of DNA methylation in heterochromatic versus euchromatic TEs is governed by H1.

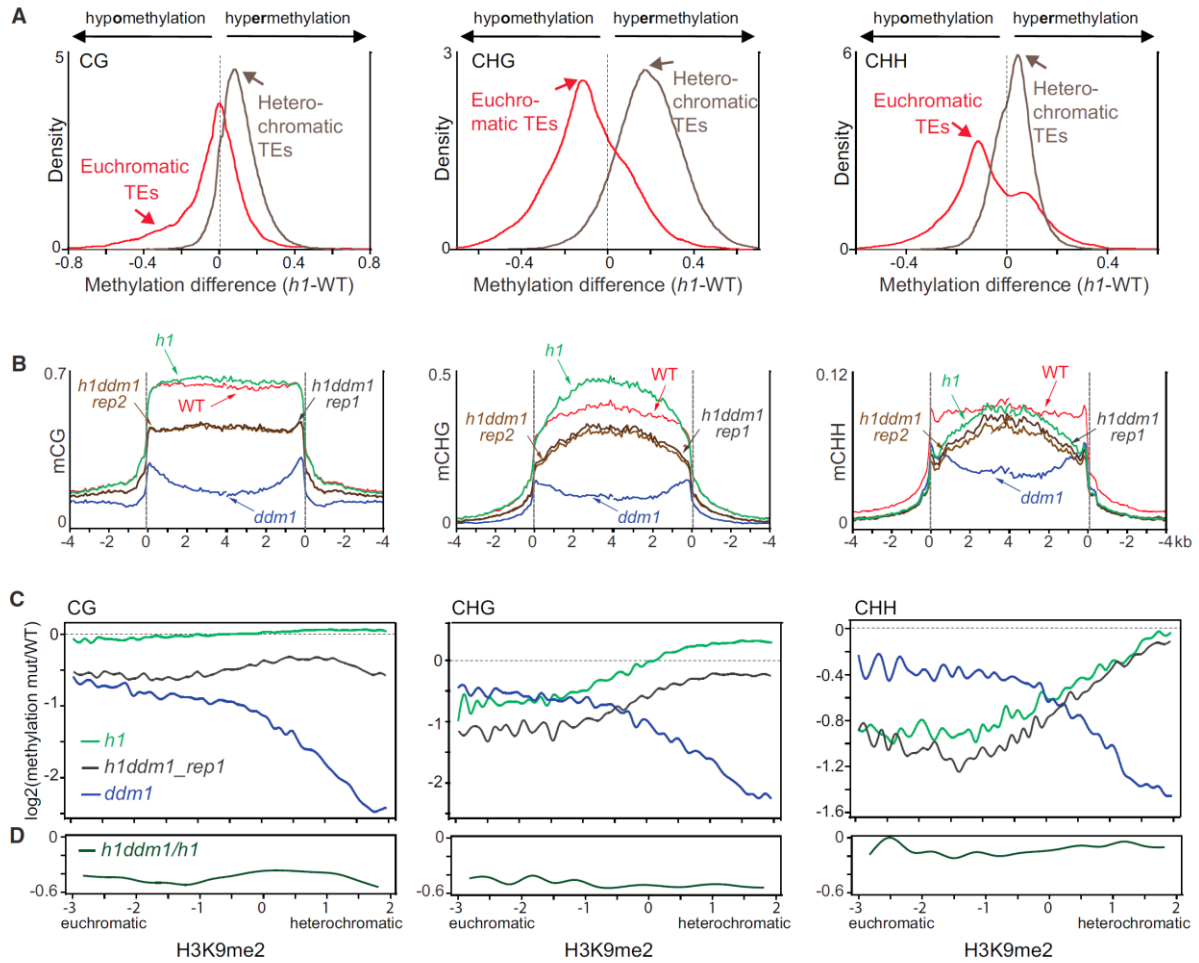


Figure 2.5. Lack of H1 ameliorates the loss of methylation in *ddm1* plants.

(A) Kernel density plots of methylation differences between h1 and wild-type (positive numbers indicate greater methylation in h1). TEs with H3K9me2 log2 scores lower than -1 and higher than 1 are considered euchromatic and heterochromatic, respectively. The arrows emphasize global differences (a shifted peak) or extensive local differences (a broad shoulder). (B) Average methylation of TEs in sibling wild-type (WT), h1, ddm1, and h1ddm1 (two biological replicates) seedlings is plotted as in Figure 1. (C-D) M-spline curve fits of log2 DNA methylation ratios in 50-bp windows plotted against H3K9me2 level.

Methylation of TE families depends on sRNA abundance and chromatin features

The *Arabidopsis* genome contains a variety of TE families that have different mechanisms of transposition, internal structure, and chromosomal localization (134, 135). We chose four such families to examine DNA methylation mediated by DDM1 and RdDM in more detail: *Gypsy*, *Copia*, MuDR and LINE elements. *Gypsy* elements are long terminal repeat (LTR)-flanked retrotransposons that are concentrated in pericentric heterochromatin. *Copia* LTR elements are more evenly dispersed, as are LINE non-LTR retrotransposons and the terminal inverted repeat (TIR)-flanked MuDR DNA transposons.

The four TE families have distinct sRNA distributions: *Gypsy* elements have high levels of sRNA across the entire sequence, *Copia* and MuDR elements preferentially accumulate sRNA at their 5' and 3' terminal repeats, and LINEs, which lack 5' repeats, have a spike in sRNA abundance at the 3' end (Figure 2.6A). DDM1-mediated DNA methylation – the residual methylation in the *drd1* mutant – does not appear to be influenced by sRNA, as exemplified by efficient methylation of sRNA-poor *Copia* and LINE TE bodies as well as sRNA-rich *Gypsy* elements in *drd1* (compare the black traces in Figures 2.6B with Figure 2.6A), supporting our family-independent TE analysis (Figure 2.3A). CHH RdDM – the residual methylation in the *ddm1* mutant – resembles the distribution of sRNA in *Copia*, MuDR and LINE elements (compare the blue trace in Figure 2.6B with Figure 2.6A). Although *Gypsy* elements are evenly covered by sRNA (Figure 2.6A), RdDM is still preferentially localized at their edges (Figure 4B), supporting our conclusion that the heterochromatic environment of internal *Gypsy* sequences inhibits RdDM (Figure 2.3E).

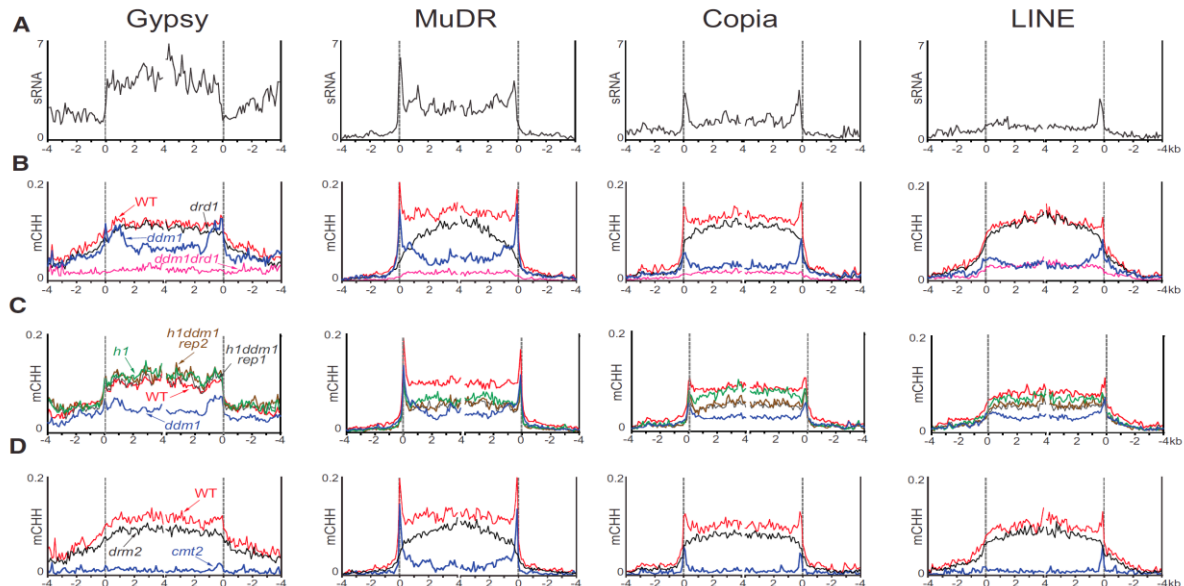


Figure 2.6. Methylation of TE families depends on sRNA abundance and chromatin features. (A–D) Averaged sRNA abundance (A) and CHH methylation levels (B–D) are plotted as in Figure 1 for TEs belonging to four distinct families. The *ddm1* trace in (C) represents siblings of the wild-type (WT), *h1*, and *h1ddm1* seedlings analyzed in this panel and is independent of the *ddm1* roots analyzed in (B).

Lack of H1 ameliorates the CG, CHG and CHH methylation losses caused by *ddm1* in all TE families (Figures 2.6C). In particular, CHH methylation in *h1ddm1* is similar to that in *h1* at *Gypsy* and MuDR elements, and to a lesser extent at *Copia* and LINE TEs (Figure 2.6C).

CHH methylation in *drm2* plants closely resembles that in *drd1* mutants at all four TE families (compare the black traces in Figures 2.6B and 2.6D), further substantiating the link between RdDM and DRM2. The *cmt2* CHH methylation phenotype is similar to but stronger than that of *ddm1* at MuDR, *Copia*, and LINE elements (compare the blue traces in Figures 2.6B and 2.6D) – virtually all CHH methylation is lost in *cmt2* plants except at sRNA-targeted terminal repeat sequences (Figure 2.6D), supporting our conclusion that CMT2 mediates CHH methylation independently of sRNA. Surprisingly, lack of CMT2 essentially eliminates CHH methylation at *Gypsy* elements (Figure 2.6D), suggesting that the CHH methylation at *Gypsy* elements requires the assistance of DDM1. However, an unexpected substantial amount of CHH methylation remains at *Gypsy* elements in *ddm1* mutants (Figure 2.6B), indicating that RdDM can maintain CHH methylation at *Gypsy* elements in the absence of DDM1, but not in wild type plants. This kind of transition from CMT2-mediated CHH methylation to RdDM-mediated CHH methylation at *Gypsy* elements was found to be caused by the derepression of *Gypsy* elements in *ddm1* mutants, triggering *de novo* RDR6-RdDM and RDR2-RdDM (81).

DDM1 and RdDM collaborate to repress TE transcription and transposition

Consistent with the importance of DDM1 for the maintenance of DNA methylation in the bodies of long TEs, where TE-encoded genes required for transposition are located, our RNA-seq analysis revealed that many TEs (2294) are reactivated in *ddm1* mutant plants (Figure 2.7A). In contrast, lack of DRD1, which primarily affects RdDM of non-coding short TEs and TE edges, caused the reactivation of just 44 TEs (Figure 2.7A). In agreement with our methylation analyses, TEs reactivated in *ddm1* are longer and more heterochromatic than those reactivated in *drd1* (Figure 2.7B). In both mutants, TE reactivation is associated with DNA hypomethylation (Figures 2.7C). The *ddm1drd1* double mutant, which showed additive to synergistic hypomethylation (Figures 2.7C), led to stronger TE transcriptional reactivation than either of the single mutants (Figure 2.7D). This is particularly exemplified by *Gypsy* elements that require both DDM1 and RdDM for full methylation (Figures 2.6B), which are synergistically hyperactivated in *ddm1drd1* (Figure 2.7E).

Mutations in *DDM1* and *MET1* have been shown to cause transposition of a few TEs, including *CACTA* and *EVADE*, but only after several generations of inbreeding (125, 149). We found that *CACTA* and *EVADE* transpose within the first homozygous generation of *ddm1drd1* and *ddm1rdr2* mutants (Figures 2.7F). This result emphasizes that the DDM1 and RdDM pathways collaborate to prevent TE expression and mobilization.

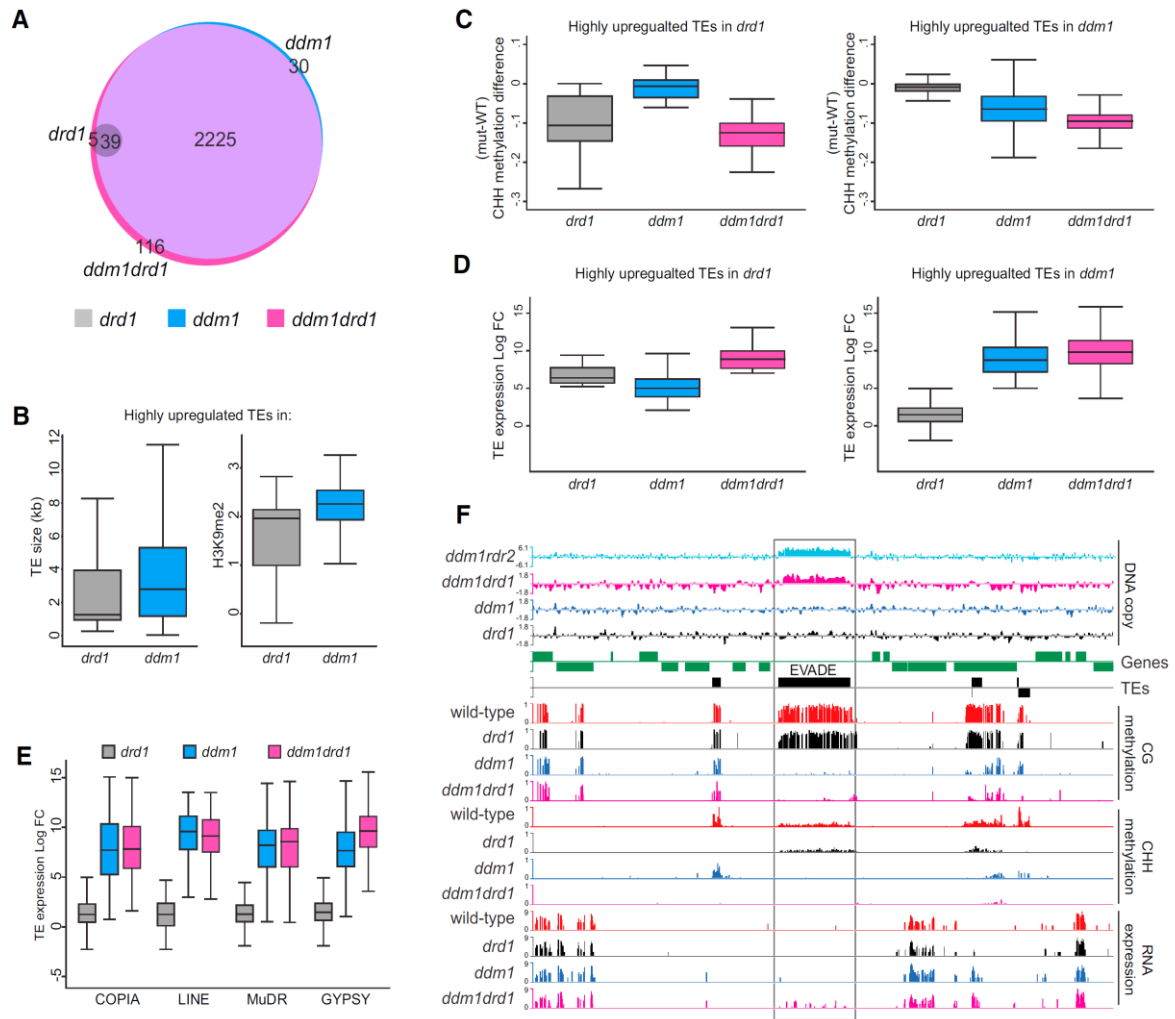


Figure 2.7. DDM1 and RdDM collaborate to repress TE expression and transposition.

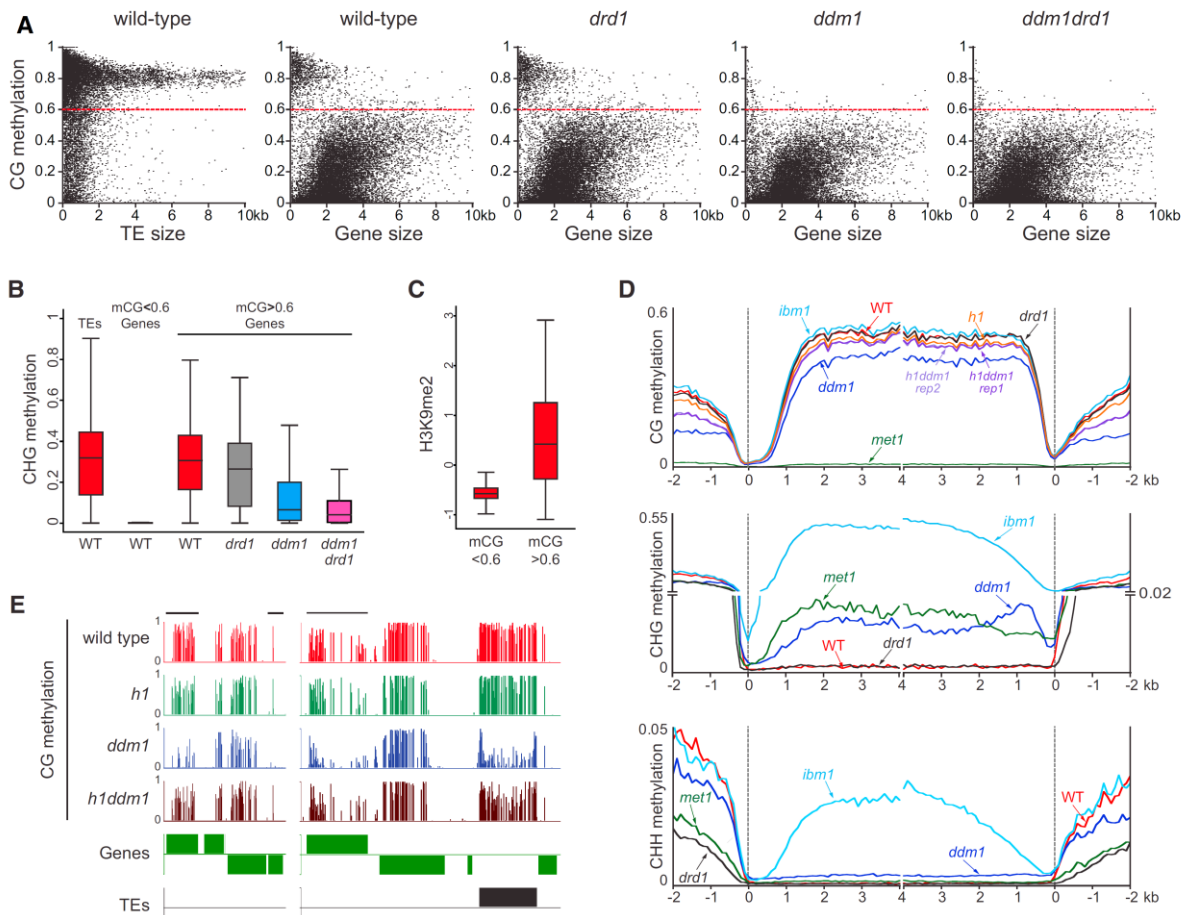
(A) Venn diagram of significantly upregulated TEs in *drd1*, *ddm1* and *ddm1drd1* mutants. (B-D) Box plots of the sizes and H3K9me2 levels (B), absolute fractional CHH demethylation of 50-bp windows (C), and expression compared to wild-type (D) of TEs that are at least 32-fold overexpressed either in *drd1* or in *ddm1*. (E) Box plots of TE family expression in the indicated mutants with respect to wild-type. (F) DNA sequencing coverage (log2(reads in mutant/read in wild-type)), DNA methylation and RNA levels near the LTR retrotransposon *EVADE* (*AT5TE20395*). The sequence coverage is indicative of *EVADE* copy number relative to wild type (125).

DDM1 mediates gene body DNA methylation

Plant genes, including those with presumed or demonstrated biological functions, can have methylation patterns that resemble TEs (150, 151), and may therefore be regulated more like TEs than conventional genes. TEs are generally highly methylated at CG sites (Figure 2.8A, left panel), whereas most *Arabidopsis* genes have lower methylation levels, except for a distinct group that resembles TEs (Figure 2.8A). A cutoff of 60% overall CG methylation separates the

two genic groups rather cleanly (Figure 2.8A). The 1284 highly methylated genes resemble TEs in many aspects – they are concentrated in pericentric heterochromatin, and enriched for non-CG methylation (Figures 2.8B) and H3K9me2 (Figure 2.8C). CG methylation of such genes is lost in *ddm1* but not *drd1* mutants (Figure 2.8A), and non-CG methylation is lost partially in both mutants and synergistically in *ddm1drd1* (Figures 2.8B). Because of these characteristics, we will refer to the highly methylated TE-like genes as heterochromatic genes, and to the rest as euchromatic genes.

Despite the reported requirement of DDM1 and its mouse homolog Lsh for maintenance of DNA methylation in TEs but not genes (121, 124), we find that CG methylation of at least 5,348 euchromatic gene bodies (50% of all methylated euchromatic genes) is significantly reduced in *ddm1* plants (Fisher's exact test p-value < 0.0005), whereas only 85 genes are significantly hypermethylated, leading to an overall methylation loss of 20% (Figures 2.8D). Nevertheless, *h1* suppresses the genic hypomethylation caused by *ddm1* (Figures 2.8D and 2.8E), with numbers of significantly hypomethylated (3,878) and hypermethylated (1,488) genes in *h1ddm1* resembling *h1*. These results demonstrate that DDM1 is important for DNA methyltransferase access to all types of sequences, and that the DDM1 requirement depends on the extent of heterochromatin, from the most heterochromatic long TEs, to the less heterochromatic short TEs, to the least heterochromatic genes (Figure 2.8F).



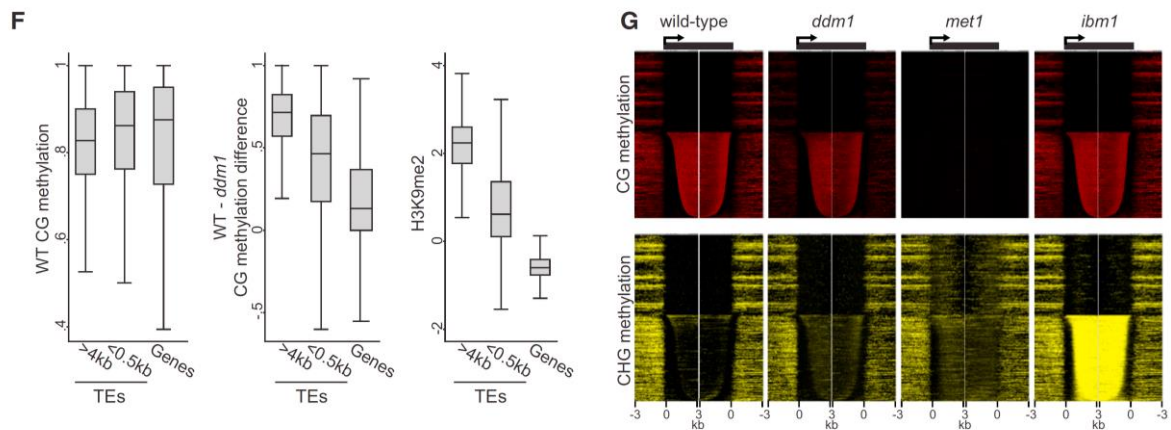


Figure 2.8. DDM1 mediates genic DNA methylation.

(A) CG methylation was averaged in TEs (left panel) and genes for the indicated genotypes (right panels), and plotted against TE or gene size. Note the group of relatively short genes above the red line that are highly methylated in wild-type and significantly hypomethylated in *ddm1* and *ddm1drd1* mutants, similarly to TEs. (B) Box plots of averaged CHG methylation in TEs, euchromatic genes (mCG < 0.6), and heterochromatic genes (mCG > 0.6). (C) Box plots of H3K9me2 in euchromatic and heterochromatic genes. (D) Genes with average CG methylation between 20% and 60% (euchromatic genes) were aligned as described in Figure 1A. The Y-axis of the CHG plot was broken at 0.02 to improve visualization. (E) Distribution of CG methylation in representative genes *AT1G04700*, *AT1G04750* and *AT1G67220* (emphasized by horizontal black bars) that lose methylation in *ddm1* but not in *h1ddm1*. (F) Box plots of wild-type CG methylation (left), absolute fractional CG demethylation in *ddm1* (middle), and H3K9me2 (right) of 50-bp windows within long TEs (> 4 kb), short TEs (< 500 bp) and euchromatic genes. (G) Heat maps of CG (red) and CHG (yellow) DNA methylation in genes aligned at the 5' end (left half of each panel) and the 3' end (right half of each panel). More intense color indicates greater methylation. Genes without wild-type CG methylation (shown in the top half of each panel) were stacked from the top of chromosome 1 to the bottom of chromosome 5; genes containing CG methylation islands (shown at the bottom of each panel) were sorted based on the starting position (for 5' panels) or ending position (for 3' panels) of the wild-type CG methylation island.

Loss of DDM1 causes extensive genic CHG hypermethylation

Loss of DDM1 has been reported to cause CHG hypermethylation of a few genes (121, 152), and genic hypermethylation was also reported in mouse cells lacking Lsh (124), but whether *ddm1* causes extensive genic CHG hypermethylation, as has been reported for *Arabidopsis met1* and *ibm1* mutants (49, 119), is unknown. We find that lack of DDM1 leads to substantial CHG hypermethylation in euchromatic gene bodies (Figure 2.8D), but this hypermethylation differs from that caused by *met1*. Hypermethylation in *ddm1* plants is higher toward the 3' end, whereas *met1*-mediated CHG hypermethylation is more prevalent near the 5' end (Figures 2.8D). CHG hypermethylation in *ddm1* is restricted to genes that exhibit CG

methylation, and is excluded from 5' and 3' genic sequences like wild-type CG methylation (Figure 2.8G). In contrast, *met1* causes CHG methylation of genes that lack CG methylation in wild-type, and the hypermethylation extends to the 3', and to a lesser extent the 5' regions of genes that are normally not CG methylated (Figure 2.8G).

Like the *ddm1* phenotype, CHG hypermethylation in plants lacking the H3K9me2 demethylase IBM1 (Figure 2.8D) is confined almost exclusively to genic regions that bear CG methylation (Figure 2.8G). Such genes also exhibit some CHG methylation in wild-type plants (Figure 2.8G), which may be due to imperfect activity of IBM1. The *ibm1* mutation also causes substantial CHH hypermethylation of genes (Figure 2.8D) (52) that is insensitive to *drd1* (Figure 2.9), and is thus likely mediated by CMT2. However, hypermethylation in *ddm1* plants is unlikely to be primarily caused by IBM1 malfunction (153), because genic CHG methylation is 5'-biased in *ibm1*, unlike the 3'-biased methylation in *ddm1* (Figures 2.8D). This interpretation is also supported by the normal expression of the *IBM1* transcript in *ddm1*, which is disrupted in *met1* (154). Thus, lack of MET1, IBM1 and DDM1 leads to distinct genic hypermethylation phenotypes likely mediated by different, though potentially overlapping mechanisms.

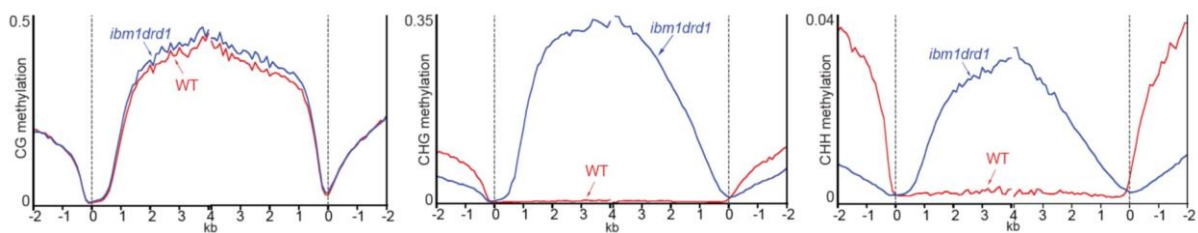


Figure 2.9. Patterns of genic DNA methylation (CG, CHG and CHH) in wild-type and *ibm1drd1*, plotted as in Figure 2.8D.

DDM1 and RdDM synergistically regulate gene expression

The importance of TE silencing by DNA methylation is not restricted to preventing transposition – demethylation and activation of TEs can also alter the expression of nearby genes (155, 156). We found 179 genes upregulated in *ddm1* plants (Figure 2.10A), all but one of which are heterochromatic (Figure 2.10B), consistent with the importance of DDM1 for the methylation of heterochromatin generally and heterochromatic genes specifically (Figures 2.8A, 2.8B and 2.10C). Only ten genes are upregulated in *drd1* plants (Figure 2.10A), six of which are euchromatic (Figure 2.10B), and only two of which overlap with genes upregulated by *ddm1*. One of the heterochromatic genes upregulated in *drd1* is *AT1G59930* (Figure 2.10C), a maternally-expressed imprinted gene that is activated by DNA demethylation (156). All six *drd1*-upregulated euchromatic genes (*AT1G21940*, *AT1G35730*, *AT4G01985*, *AT4G09350*, *AT4G16460*, *AT5G41830*) have short TEs or TE edges that are hypomethylated in *drd1* but not

in *ddm1* in close proximity to the transcriptional start sites, such as the 150 bp *Copia* fragment (*MET1*) upstream of *AT1G35730*.

More genes (214) are overexpressed in the *ddm1drd1* double mutant, including 23 euchromatic genes and five of the six euchromatic genes upregulated in *drd1* (Figures 2.10A and 2.10B). One such gene is *SDC* (Figure 2.10D), the overexpression of which confers the characteristic phenotype of the *drm1drm2cmt3* methyltransferase mutant line (155) that is also exhibited by *ddm1drd1* plants (Figure 2.10E). Similar to *Gypsy* elements, the repetitive *SDC* promoter is targeted by sRNA and H3K9me2 (Figure 2.10D), and its methylation and silencing is mediated by DDM1 and RdDM (Figure 7D). *SDC* is also a maternally-expressed imprinted gene regulated by DNA demethylation (156), as is *FWA*, which is modestly upregulated in *ddm1drd1* plants due to demethylation of its SINE-related promoter. Overall, our results demonstrate that DDM1 and RdDM collaborate to methylate gene-adjacent repetitive elements, thereby maintaining appropriate patterns of gene expression.

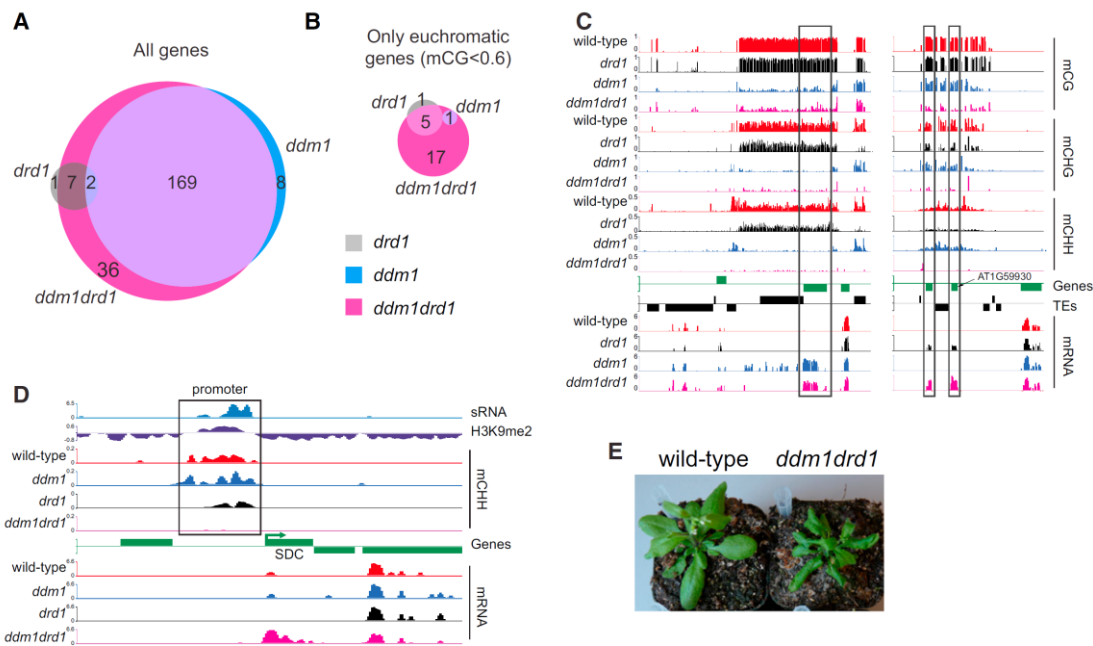


Figure 2.10. DDM1 and RdDM synergistically regulate gene expression.

(A-B) Venn diagram of significantly ($p < 0.05$) upregulated genes in *drd1*, *ddm1* and *ddm1drd1* mutants. (C) DNA methylation and RNA levels near *AT1G46696*, and the linked genes *AT1G59920* and *AT1G59930*. (D) sRNA, H3K9me2, CHH methylation and RNA levels near *SDC* (*AT2G17690*). (E) Phenotypes of wild-type (flat leaves) and *ddm1drd1* (leaves curled downward) plants.

Discussion

The targeting of plant DNA methylation has been carefully dissected at individual loci, and the methyltransferases that catalyze CG and CHG methylation throughout the genome are known (117), but the identity of the pathways that mediate the bulk of genomic methylation has remained a mystery. Here, we find that DDM1 and RdDM separately mediate nearly all DNA methylation in *Arabidopsis* TEs. DDM1 is required for methylation in all sequence contexts of highly heterochromatic TEs (Figures 2.2A and 2.3A). This requirement is reduced at less heterochromatic elements, is least in euchromatic genes (Figure 2.8F), and depends on histone H1 (Figures 2.5B, 2.5C and 2.5D). Together with the preferential demethylation of short euchromatic TEs during plant sexual development (151, 157), our results demonstrate that a division of the genome into genes and TEs can only explain the biology of DNA methylation if chromatin configuration is also considered.

Lack of access to DNA is postulated to be a core property of heterochromatin that enforces gene silencing by preventing binding of transcription factors and RNA polymerase (158). At the same time, stable maintenance of inaccessible heterochromatin requires DNA methylation and histone modifications like H3K9me2 that are catalyzed by enzymes that need to access chromatin. This conundrum is exemplified by the RdDM pathway that silences TE expression, yet requires TE transcripts to function (118). Our results indicate that H1 restricts access to nucleosomal DNA, and that DDM1 overcomes this restriction to enable the maintenance of DNA methylation and silencing of diverse TEs. Without DDM1, DNA methyltransferases cannot efficiently methylate inaccessible heterochromatic TEs (Figures 2.2A and 2.3A), leading to derepression and transposition (Figures 2.7). Without H1, less heterochromatic sequences lose methylation (Figures 2.5A), presumably because they become more accessible to enzymes that catalyze euchromatic histone modifications and antagonize DNA methylation (96, 157, 159). The balance between exclusion and access is thus essential for the stable propagation of chromatin states.

The influence of the chromatin environment on DNA methylation has important implications for how different classes of TEs are silenced. Short TEs, which are preferentially found near active genes (151, 157), are generally relatively euchromatic (Figure 2.3B), and rely on RdDM for silencing (Figure 2.7B), whereas silencing of the more heterochromatic longer TEs (Figure 2.3B) that are usually found away from genes (151, 157) relies primarily on DDM1 (Figure 2.7B). Despite these differences, DDM1 and RdDM contribute to the methylation and silencing of most TEs, leading to a synergistic loss of methylation (Figures 2.2A, 2.6B and 2.7C) and repression (Figures 2.7D and 2.7E), as well as to enhanced transposition (Figures 2.7F), when both pathways are mutated.

Our data suggest that RdDM operates primarily through DRM2, and is thus responsible for a relatively minor fraction of genomic methylation (Figures 2.2A and 2.2B) (132). We show that the majority of CHH methylation is mediated by CMT2 independently of RdDM (Figures 2.2A, 2.2B, 2.2C, 2.6D and Table 2.2), presumably by binding to H3K9me2 like its CMT3 homolog (64). CMT2 forms a distinct family in monocots and dicots (Figure 2.2D), indicating that this enzyme catalyzes CHH methylation throughout flowering plants, including important crops such as rice. Curiously, we have not been able to identify a CMT2 homolog in maize, suggesting that CHH methylation may be entirely dependent on RdDM in this species.

RdDM, by definition, only targets loci that generate sRNA, and therefore affects TEs, which are at least somewhat heterochromatic compared to genes, almost exclusively (Figures 2.2, 2.6B and 2.8D) (131). Production of sRNA can be influenced by many factors, including TE structure and copy number (160). Thus, the abundant *Gypsy* LTR retrotransposons that make up the bulk of pericentric heterochromatin generate sRNA differently from the more dispersed *Copia*, LINE and MuDR elements (Figure 2.6A), and RdDM is crucial for *Gypsy* silencing (Figure 2.7E) despite the highly heterochromatic nature of these elements. Nonetheless, RdDM is more efficient at less heterochromatic sRNA-producing loci (Figure 2.3E) (112). This is a curious observation because RdDM, which functions to methylate and silence TEs, might be expected to work well in heterochromatin. Furthermore, why would DDM1, a protein apparently adapted to remodel heterochromatic nucleosomes, not facilitate RdDM?

Our observation that DRD1 is required for RdDM offers a potential explanation. DRD1 is associated with RNA polymerase V, a derivative of RNA polymerase II that shares most pol II subunits (118). Pol II has evolved to transcribe genes, and thus the machinery associated with pol V is likely adapted to function in euchromatin. Unlike DDM1, DRD1 may remodel heterochromatic nucleosomes inefficiently. Because RdDM is a branch of the RNA interference pathway that functions to cleave aberrant and viral mRNA, it likely evolved independently of heterochromatic DDM1-associated pathways. Thus, plants possess two separate mechanisms for methylating and silencing TEs that rely on distinct remodelers with differing nucleosome preferences.

Materials and methods

Biological materials.

The *ddm1-2* (120), *met1-6* (161), *rdr2-1* (128) and *ibm1-6* (52) mutant lines were described previously. The *drd1-7* (GABI_503F06) and *h1.2-1* (*AT2G30620*; GABI_406H11) mutants in the Col-0 ecotype were obtained from the GABI-KAT collection (www.gabi-kat.de/). T-DNA insertion lines for *drm2-3* (SALK_150863), *cmt3-12* (SALK_148381), *cmt2-3* (SALK_012874), *cmt2-4* (SALK_201637), *cmt2-5* (SAIL_906_G03; CS863642), *cmt2-6* (SAIL_1236_D12; CS863007) and *h1.1-1* (*AT1G06760*, SALK_128430C) alleles in the Col-0 ecotype were obtained from TAIR (www.arabidopsis.org). T-DNA insertions were confirmed by PCR-based genotyping. Roots were grown for three to four weeks under continuous light in a sterile Gamborg's B5 medium and seedlings were grown on MS medium under continuous light following two days of cold treatment.

Phylogenetic analysis.

The catalytic domains of the indicated CMT methyltransferase proteins were aligned using MUSCLE v3.7. All alignment files were checked, modified and converted to NEXUS files using MacClade v4.06. Phylogenetic trees were inferred using MrBayes v3.1.2. Two independent Markov chain Monte Carlo (MCMC) runs of four chains using the default fixed Poisson model were started from independent random trees, and were carried through four million generations, with trees sampled every 100th generation. Convergence was confirmed by checking that the standard deviations of split frequencies were <0.01, that the log probabilities of the data given the parameter values fluctuated within narrow limits, that the PSRF (potential scale reduction factor) = 1.000, and by examining the plot of the generation versus the log probability of the data. The first 7,500 stored trees from each run were discarded and the remaining 32,501 trees were used to construct the consensus tree. Final trees were checked and graphically presented in FigTree v1.2.2 (<http://tree.bio.ed.ac.uk/software/figtree>). Trees were made from the highly conserved amino acid blocks by excluding weakly conserved N- and C-termini along with unique inserts, yielding a final dataset of 487 characters.

BS-seq.

About 500 ng of genomic DNA was isolated from roots or seedlings, fragmented by sonication, end repaired and ligated to custom-synthesized methylated adapters (Eurofins MWG Operon) according to the manufacturer's (Illumina) instructions for gDNA library construction. Adaptor-ligated libraries were subjected to two successive treatments of sodium bisulfite conversion using the EpiTect Bisulfite kit (Qiagen) as outlined in the manufacturer's instructions. The bisulfite-converted libraries were then amplified by PCR using the following conditions: 2.5 U of ExTaq DNA polymerase (Takara Bio), 5 µl of 10X Extaq reaction buffer,

25 μ M dNTPs, 1 μ l Primer 1.1, 1 μ l Primer 2.1 (50 μ l final). PCR reactions were carried out as follows: 95 °C 3 min, then 12-14 cycles of 95 °C 30 sec, 65 °C 30 sec and 72 °C 60 sec. The enriched libraries were either gel-purified (~300 bp band) or purified with solid phase reversible immobilization (SPRI) method using AM-Pure beads (Beckman Coulter) prior to quantification with a Bioanalyzer (Agilent). Sequencing on the Illumina platform was performed at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley.

RNA-seq.

Total RNA samples from 3-week-old roots were isolated using the RNeasy mini kit (Qiagen #74106) including on-column DNase treatment. mRNA was purified from 10-50 μ g of total RNA by either (i) two cycles of poly-A enrichment using the Oligotex kit (Qiagen #72022), followed by a rRNA removal step using the RiboMinus Eukaryote Kit for RNA-Seq (Invitrogen #A1083702), or by (ii) using the Ribo-Zero rRNA Removal Kit (Plant Leaf). Precipitated mRNA samples were eluted with 9 μ l of RNase free water and fragmented with 1 μ l of 10X fragmentation buffer (Ambion, #AM8740) at 70 °C. Reactions were stopped after 5 minutes by adding 1 μ l Stop buffer and RNA was purified by ethanol precipitation. cDNA was synthesized from 100-300 ng of mRNA using SuperScript III reverse transcriptase (Invitrogen #18080-051). Double-stranded DNA was synthesized according to the instructions for the SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen). DNA was cleaned with a QIAquick PCR spin column (Qiagen, #28106), sequencing adapters were ligated according to the Illumina protocol and library was amplified by 18 cycles of PCR using Phusion DNA polymerase (NEB, #F-530). Bands around 300 bp were gel-purified and cloned for validation. Traditional sequencing confirmed that the libraries were properly constructed, showing high percentage of mRNA over rRNA. The libraries were sequenced at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley to generate single ends (SE) 36-50 base reads.

MNase-seq.

Arabidopsis roots (1 g) were ground in liquid nitrogen, resuspended in 20 ml of HBM buffer (25mM Tris, pH 7.6, 0.44M Sucrose, 10 mM MgCl₂, 2 mM spermine and 0.1% Triton X-100), homogenized, filtered through miracloth, transferred to a 30 ml round bottom glass tube, centrifuged at 2000g (4 °C) for 10 min and resuspended in 1 ml HBB buffer (25mM Tris, pH 7.6, 0.44M Sucrose, 10 mM MgCl₂ and 0.1% Triton X-100). Nuclei were further spun down at 200g, 4 °C for 2 min and resuspended in 1 ml of TNE buffer (10 mM Tris, pH 8.0, 100 mM NaCl and 1 mM EDTA). MNase digestion was done with 4 μ l of 1M CaCl₂ and 1 μ l of diluted (1/20) MNase (200ul/ml; Sigma #N-3755) per 100 μ l of pellet nuclei. Nuclei were then divided to several tubes and digestion was stopped at 45 sec intervals with 10mM EDTA. Digested nuclei were spun down at maximum speed for 5 min at 4 °C, and released soluble nucleosomes were collected from the supernatant. Following RNase A and proteinase K digestion, DNA was

purified using phenol/chloroform. Purified DNA samples were run on a 2% agarose gel, digested samples with most enriched intact mononucleosomes were chosen and bands corresponding to ~150 bp were cut and purified with a Gel Purification kit (Qiagen). Illumina libraries were constructed and sequenced at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley to generate paired ends (PE) 36 base reads.

Genomic data.

Data for sRNA and CHH methylation in the *rdd* triple mutant were derived from (49), for *ibm1-6* DNA methylation from (52), for H3K9me2 from (65) and (162), for H3K27me3, EMF1 ChIP-chip, and *emf1* and *emf2* RNA microarray expression from (163), and for all other histone modifications from (137).

Gene and TE methylation and chromatin pattern analyses.

Arabidopsis genes or transposons were aligned at the 5' or the 3' end. To avoid averaging the edges of short genes or TEs with the middles of long sequences, we did not use data within 1.5 kb of the opposite end for genes or with 250 bp of the opposite end for TEs. After the alignment, genes with CG methylation over 60% were excluded from the analysis because they behave like transposons. Genes with low CG methylation (<20%) were also excluded from methylation analyses because they decrease the dynamic range without substantively contributing to the analysis. To avoid complications in calculating TE size caused by serial TE insertions near the centromeres, only TEs located on the chromosome arms were included in analyses where TEs were filtered by size. GC content was calculated by averaging the GC content in 50-bp windows.

Kernel density plots.

Density plots in Figure 3A were generated with the difference between *h1* and wild-type root methylation in 50-bp windows. We plotted windows with at least 10 informative sequenced cytosines and fractional methylation of at least 0.3 for CG and 0.1 for CHG and CHH in *h1* or wild-type. Genes with over 60% and under 20% CG methylation were excluded from the analysis. We used data from sibling *h1* and wild-type roots because the sibling *h1* and wild-type seedlings used in Figure 3B-D and other *h1* analyses were segregated from a cross to heterozygous *ddm1* plants. Even though the *ddm1* line had been extensively backcrossed to wild type, this line contained some TEs that had been demethylated when *ddm1* was homozygous, and which did not recover methylation in backcrossed progeny (127). These segregating demethylated TEs complicated the density analysis by mimicking sequences that are specifically differentially methylated in *h1* plants.

Expression analysis.

To identify genes and TEs differentially expressed in the mutants relative to the wild-type control samples, RNA-seq datasets were mapped to the TAIR cDNA and TE annotations and analyzed using the Bioconductor package edgeR (164). This package uses an empirical Bayesian approach based upon the negative binomial distribution to model digital expression data. Before comparing expression between genotypes, we first imposed an expression value threshold, excluding genes that did not have at least one read per million reads in at least three libraries. When all annotated genes were considered, 70% passed this expression threshold, whereas when highly methylated genes (CG methylation greater than 0.6) were excluded, 74% of the genes passed this expression threshold. A negative binomial was fitted to the data, using parameters estimated from the RNA-seq data. The dispersion of individual genes and TEs was estimated using the quantile-adjusted conditional maximum likelihood method and an empirical Bayes strategy. Genes and TEs differentially expressed between genotypes were determined using the exact test for the negative binomial distribution (which has strong parallels to Fisher's exact test). The resulting p-values were adjusted using Benjamini and Hochberg's approach (165) to control the false discovery rate to below 5%.

Chapter III

The *Arabidopsis* male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues

Most of the following chapter has been published as part of a peer reviewed article in the *Proceedings of the National Academy of Sciences, USA*:

Ping-Hung Hsieh, Shengbo He, Toby Buttress, Hongbo Gao, Matthew Couchman, Robert L. Fischer, Daniel Zilberman, and Xiaoqi Feng (2016). *Arabidopsis* male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues. *Proceedings of the National Academy of Sciences, USA*. doi: 10.1073/pnas.1619074114

Copyright is retained by the authors.

Contributions:

Robert L. Fischer, Daniel Zilberman, and Xiaoqi Feng designed research; Shengbo He, Toby Buttress, Hongbo Gao and Xiaoqi Feng performed experiments; Ping-Hung Hsieh, Shengbo He, Matthew Couchman and Xiaoqi Feng analyzed data; and Ping-Hung Hsieh, Robert L. Fischer, Daniel Zilberman, and Xiaoqi Feng wrote the paper.

Sequencing data are deposited in GEO with accession number GSE87170.

Summary

Cytosine DNA methylation regulates the expression of eukaryotic genes and transposons. Methylation is copied by methyltransferases after DNA replication, which results in faithful transmission of methylation patterns during cell division and, at least in flowering plants, across generations. Trans-generational inheritance is mediated by a small group of cells that includes gametes and their progenitors. However, methylation is usually analyzed in somatic tissues that do not contribute to the next generation, and the mechanisms of trans-generational inheritance are inferred from such studies. To gain a better understanding of how DNA methylation is inherited, we analyzed purified *Arabidopsis thaliana* sperm and vegetative cells – the cell types that comprise pollen – with mutations in the *DRM*, *CMT2* and *CMT3* methyltransferases. We find that DNA methylation dependency on these enzymes is similar in sperm, vegetative cells and somatic tissues, although *DRM* activity extends into heterochromatin in vegetative cells, likely reflecting transcription of heterochromatic transposons in this cell type. We also show that lack of histone H1, which elevates heterochromatic DNA methylation in somatic tissues,

does not have this effect in pollen. Instead, levels of CG methylation in wild-type sperm and vegetative cells, as well as in wild-type microspores from which both pollen cell types originate, are substantially higher than in wild-type somatic tissues and similar to those of H1-depleted roots. Our results demonstrate that the mechanisms of methylation maintenance are similar between pollen and somatic cells, but the efficiency of CG methylation is higher in pollen, allowing methylation patterns to be accurately inherited across generations.

Introduction

Cytosine methylation is a covalent DNA modification that regulates transcription in eukaryotes (166). The highest levels of methylation in plant and animal genomes are typically located within symmetric CG dinucleotides (166). Methylation in this sequence context is virtually ubiquitous in plant transposable elements (TEs), which are transcriptionally silenced by methylation, but also occurs within many genes without disrupting their expression (166, 167). CG methylation is catalyzed by the Dnmt1 methyltransferase family, called MET1 in plants (166, 167). MET1 restores full methylation of hemimethylated CG dinucleotides generated by DNA replication, thereby perpetuating methylation patterns after cell division (166, 167). This maintenance activity is thought to allow DNA methylation to carry epigenetic information – and influence gene expression and phenotype – across generations (101, 168). The nature of this mechanism predicts that imperfect maintenance of CG methylation should lead to complete loss as methylation is diluted during each cell division, so that the only stable methylation states for a CG site in a population of cells should be fully methylated or fully unmethylated. However, the methylation levels measured at *Arabidopsis thaliana* CG sites appear to be too low for stable maintenance (169, 170). Therefore, exactly how CG methylation is so robustly inherited in flowering plants is not entirely clear.

In addition to MET1, plants possess the chromomethylase (CMT) and DRM methyltransferase families. In *Arabidopsis*, CMT3 catalyzes methylation of semi-symmetric CNG sites, which is typically analyzed as CHG (H stands for A, T or G) to avoid overlap with CG (166, 167). A related enzyme, CMT2, catalyzes asymmetric (CHH) methylation, primarily in heterochromatic TEs (60, 61). Both enzymes rely on dimethylation of lysine 9 of histone H3 (H3K9me₂), a histone modification characteristic of plant heterochromatin (61, 64). DRM enzymes (DRM1 and DRM2 in *Arabidopsis*), which are guided by the small RNA-directed DNA methylation (RdDM) pathway (171), catalyze CHH methylation of more euchromatic TEs (60, 172, 173). Methylation mediated by CMT and DRM enzymes, collectively referred to as non-CG methylation, functions to repress TE expression and is almost completely absent from genes (166, 167). Non-CG methylation varies substantially between plant cell types and tissues (97, 102, 151, 174, 175) for reasons that remain largely unexplained.

Trans-generational inheritance of DNA methylation patterns is carried out by gametes and the cellular lineages from which they differentiate. The shoot apical meristem, a small group of stem cells that develops early during *Arabidopsis* embryogenesis, gives rise to all above-ground tissues, including the floral meristems that produce the sexual organs (176). In these, certain cells differentiate into meiocytes, which undergo meiosis to produce haploid spores (177, 178). The spores go on to divide by mitosis to create the multicellular male and female gametophytes. The male gametophyte, pollen, consists of two sperm cells and a vegetative cell,

which forms the pollen tube that delivers the sperm into the female gametophyte (177, 178). As this developmental sequence illustrates, plants specify dedicated sexual lineages much later than animals, which set aside the germline during embryogenesis (179). Nonetheless, only a very small fraction of plant cells can give rise to gametes. Despite their importance, these cells are rarely directly examined in studies of DNA methylation, so that most of our knowledge about the mechanisms of epigenetic methylation inheritance is inferred from analyses of differentiated tissues that do not contribute to the next generation.

To help address this deficiency, we analyzed DNA methylation in purified *Arabidopsis* sperm and vegetative cells with mutations in *CMT3*, *CMT2*, and both *DRM1* and *DRM2*, respectively. We also analyzed sperm and vegetative cells with mutation of both genes encoding canonical histone H1, a chromatin protein that globally reduces heterochromatic DNA methylation in all sequence contexts (60). Despite the reported absence of H3K9me2 from the vegetative cell nucleus (112), we find that methyltransferase dependencies of non-CG methylation in the sperm and vegetative cells are similar to those of leaves and other examined tissues, although RdDM partially extends into vegetative cell heterochromatin. Unlike in somatic tissues, mutation of H1 does not substantially increase heterochromatic methylation in either sperm or vegetative cell. Instead, methylation of CG sites is elevated in wild-type pollen, resembling *h1* mutant somatic tissues. The higher CG methylation levels in pollen are easier to reconcile with stable trans-generational maintenance, indicating that CG methylation efficiency is reduced in somatic cells with limited division potential. This in turn suggests that small DNA methylation differences between somatic cells or tissues may be caused by variance in maintenance efficiency instead of active developmental reprogramming.

Results

Mechanisms of DNA methylation maintenance are similar between pollen and leaves.

We isolated sperm and vegetative cell nuclei from *cmt3*, *cmt2*, *drm1drm2* (*drm*), and control wild-type (WT) plants by fluorescence-activated cell sorting (FACS) (97). Genome-wide analysis of DNA methylation by bisulfite sequencing (Table 3.1).

Sample	Nuclear				Chloroplast
	Coverage	CG, %	CHG, %	CHH, %	Overall CHH, %
WT sperm	115.46	30.34	11.25	1.46	0.55
<i>cmt2</i> sperm	30.08	29.35	10.06	0.68	0.22
<i>cmt3</i> sperm	16.56	28.64	0.99	0.69	0.20
<i>drm</i> sperm	94.81	28.79	10.37	1.13	0.62
<i>h1</i> sperm	34.90	30.34	11.05	1.23	0.43
WT vegetative nuclei	40.66	26.68	12.35	3.40	0.56
<i>cmt2</i> vegetative nuclei	28.83	24.77	11.30	1.38	0.22
<i>cmt3</i> vegetative nuclei	10.54	22.76	1.48	1.86	0.34
<i>drm</i> vegetative nuclei	23.04	25.78	11.38	2.25	0.67
<i>h1</i> vegetative nuclei	38.47	26.69	12.45	3.14	0.51

Table 3.1. Mean genomic coverage and DNA methylation for wild type (WT) and mutants sequenced in chapter III. Chloroplast CHH methylation is an estimate of cytosine nonconversion rate and other errors (contaminating chloroplast DNA is present in the purified vegetative nuclei samples). Mean methylation is calculated by averaging methylation of individual cytosines in each context.

The results revealed that, as expected, none of the mutations have a major effect on CG DNA methylation in either genes or TEs (Figure 3.1A and 3.1B), just as in leaves (Figure 3.1A and 3.1B) and other somatic tissues (60, 61, 66, 169, 170). The *cmt3* mutation has a similarly strong effect on TE CHG methylation in sperm, vegetative cells and leaves (Figure 3.1C), indicating that CHG methylation in both pollen cell types is maintained primarily by CMT3, despite the reported lack of H3K9me2 in the vegetative nucleus (112). Likewise, mutation of either *cmt2* or *drm* affects the patterns of CHH methylation similarly in sperm, vegetative cells and leaves (Figure 3.1D), even though CHH methylation is about three-fold higher in vegetative cells compared to sperm (Figure 3.1D) (97, 102). As in leaves and other tissues (60, 61), sperm and vegetative cell CHH methylation in euchromatic TEs – here defined as those with a low level of H3K9me2 in leaves (60) – is primarily dependent on *drm* (Figure 3.1E and 3.2A).

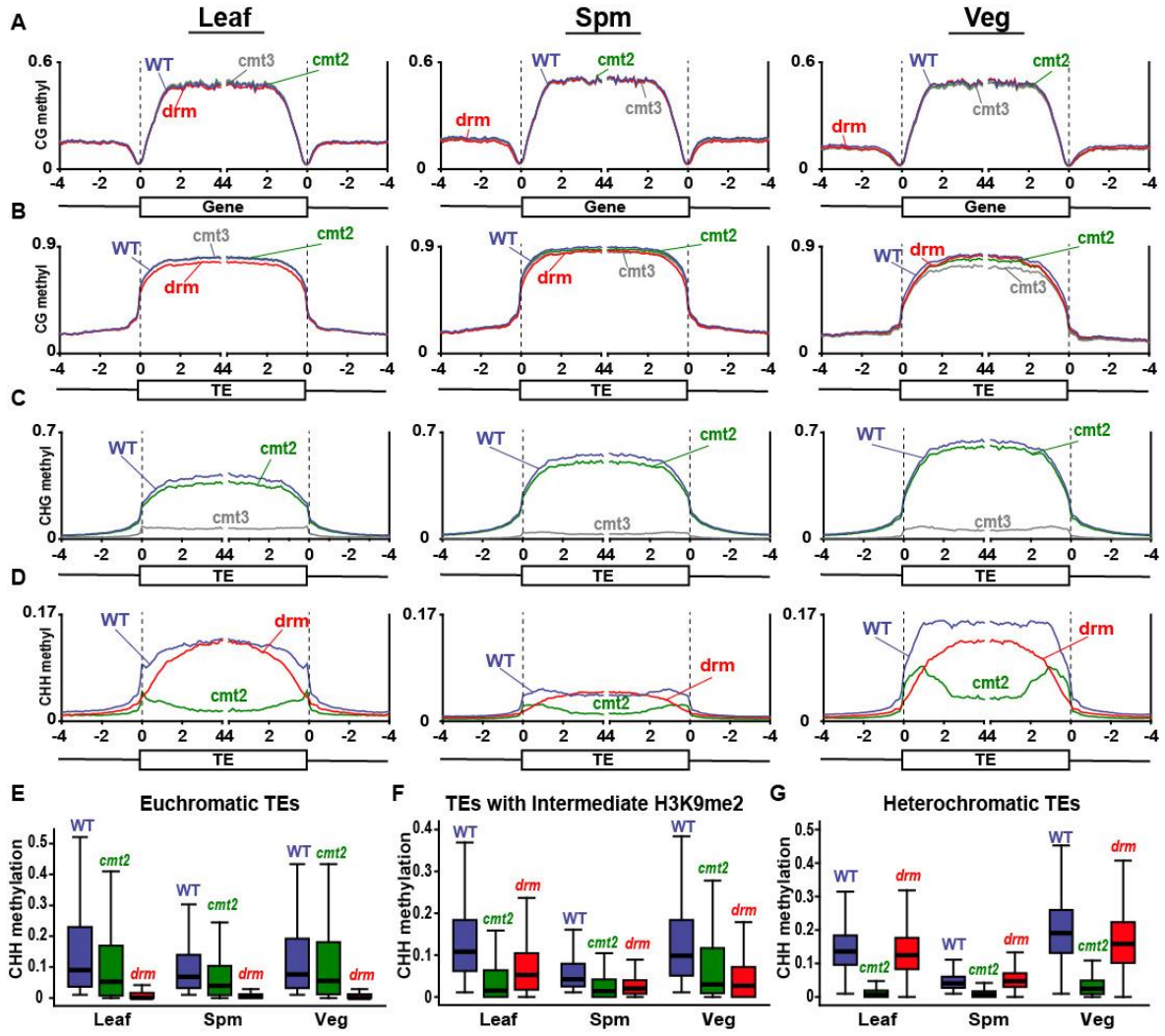


Figure 3.1. DNA methylation in *Arabidopsis* leaf, sperm (Spm), and vegetative cell (Veg).

Genes (A) and TEs (B–D) were aligned at the 5' and 3' ends. Methylation within each 100-bp interval was averaged and plotted from 4 kb away from the annotated gene or TE (negative numbers) to 4 kb into the annotated region (positive numbers). The dashed lines represent the points of alignment. (E–G) Box plots show CHH methylation levels within 50-bp windows in WT, *cmt2*, and *drm* mutants in TEs with low H3K9me2 in leaves (euchromatic TEs; E), TEs with intermediate H3K9me2 (F), and TEs with high H3K9me2 (heterochromatic TEs; G). Each box encloses the middle 50% of the distribution, with the horizontal line marking the median and vertical lines marking the minimum and maximum values that fall within 1.5 times the height of the box. Only windows with methylation greater than 1% in WT tissues and with at least 20 informative sequenced cytosines across all samples are included.

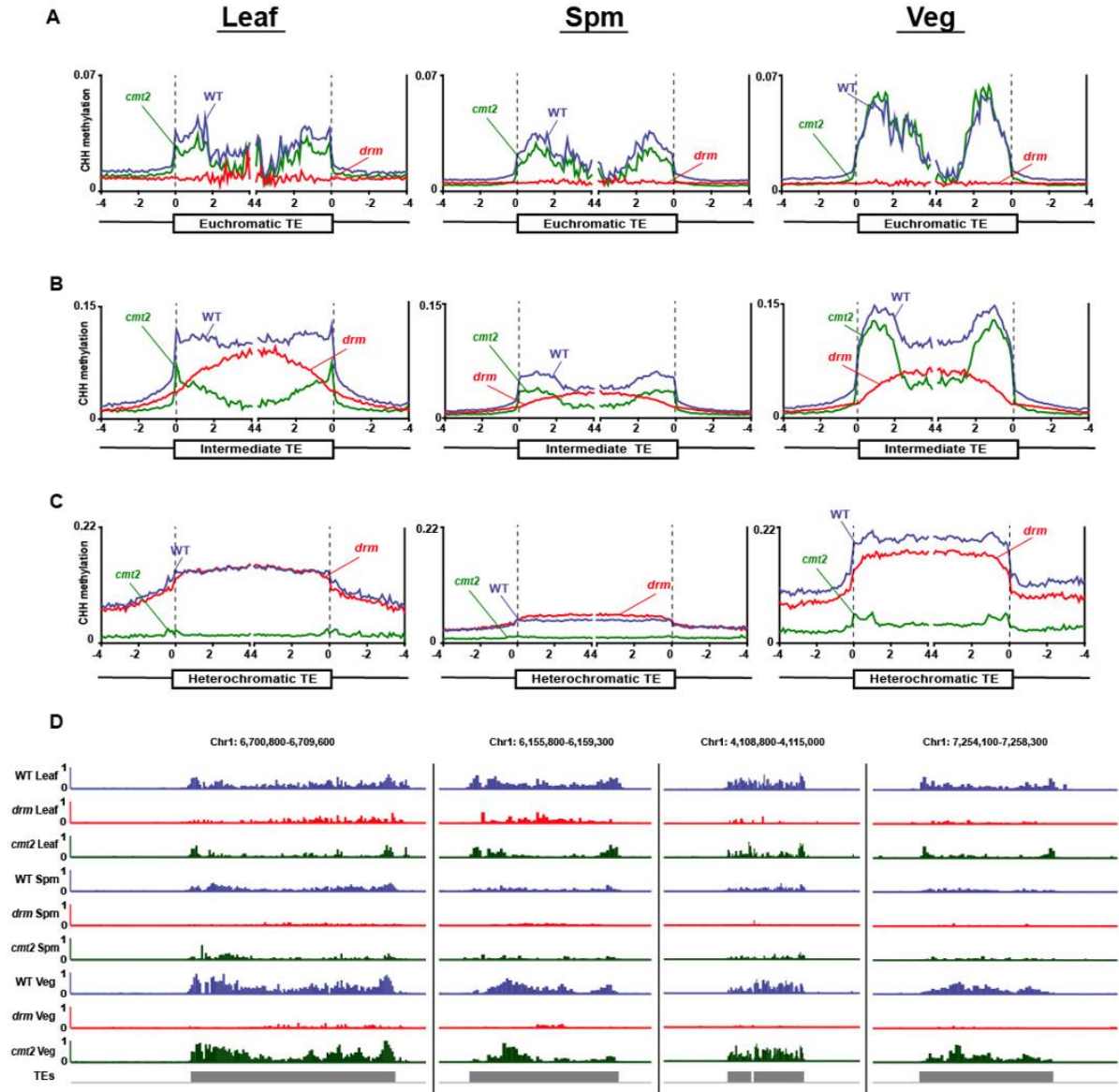


Figure 3.2. DNA methylation in *Arabidopsis* leaf, sperm (Spm) and vegetative cell (Veg) TEs.

(A-C) Euchromatic TEs, TEs with intermediate H3K9me2, and heterochromatic TEs were aligned at the 5' and 3' ends. Methylation within each 100-bp interval was averaged and plotted from 4 kb away from the annotated gene or TE (negative numbers) to 4 kb into the annotated region (positive numbers). The dashed lines represent the points of alignment. (D) Snapshots of CHH methylation of TEs with intermediate H3K9me2 in leaf, Spm and Veg of WT, *cmt2* and *drm* mutants.

CHH methylation of TEs with intermediate H3K9me2 is dependent on *drm* and *cmt2* (Figure 3.1F), with *cmt2* causing a stronger reduction in TE bodies and *drm* in TE edges (Figure 3.2B) (60), and methylation of the most heterochromatic TEs is performed primarily by CMT2 in sperm, vegetative cells and leaves (Figure 3.1G and 3.2C). An interesting feature of the data is that vegetative cells have only modestly higher CHH methylation of euchromatic TEs compared to sperm (Figure 3.1E), whereas heterochromatic TEs are far more methylated

in the vegetative cell (Figure 3.1G). The elevated CHH methylation in the vegetative cell is thus largely caused by increased activity of H3K9me2-dependent CMT2 in heterochromatin (Figure 3.1G). Overall, our results demonstrate that the roles of the major *Arabidopsis* DNA methyltransferases in pollen are remarkably similar to those in somatic tissues.

New heterochromatic RdDM targets in the vegetative cell.

Although pollen and somatic heterochromatic CHH methylation is largely dependent on CMT2, the effect of the *cmt2* mutation is weaker in the vegetative cell compared to leaf and sperm (Figure 3.1G and 3.2C). The *cmt2* mutation also doesn't reduce CHH methylation in TEs with intermediate H3K9me2 as much in the vegetative cell as it does in leaf (Figure 3.1F and 3.2B). This is easily noticeable in the methylation data as heterochromatic TEs that retain substantial CHH methylation in *cmt2* vegetative cells (Figure 3.3A) and as intermediate TEs that retain essentially WT methylation levels in *cmt2* vegetative cells (Figure 3.2D). To systematically identify heterochromatic loci with CMT2-independent CHH methylation, we compared methylation levels between all 50-bp windows in heterochromatic TEs that retain CHH methylation in either *cmt2* vegetative cell or *cmt2* leaf (Figure 3.3B). The vast majority of windows have more methylation in the vegetative cell (Figure 3.3B). As expected, CHH methylation of these windows is dependent on CMT2 in leaf and sperm (Figure 3.3C). WT vegetative cells have higher CHH methylation of these loci than leaf or sperm, which is depended on both CMT2 and DRM (Figure 3.3C), resembling the pattern observed in TEs with intermediate H3K9me2 (Figure 3.1F). TEs containing such loci also exhibit higher CHH and CHG methylation than other heterochromatic TEs in the vegetative cell (Figure 3.3D). Thus, in addition to the usual maintenance of heterochromatic CHH methylation by CMT2, some parts of vegetative cell heterochromatin are also targeted by DRM via RdDM, suggesting that these loci are less heterochromatic in this cell type. However, the 466,450 bp covered by such loci in our analysis represent only 7.9% of CMT2-dependent heterochromatic loci in the vegetative cell (5,905,450 bp), so that the vast majority of vegetative cell heterochromatic CHH methylation requires CMT2 (Figure 3.1G).

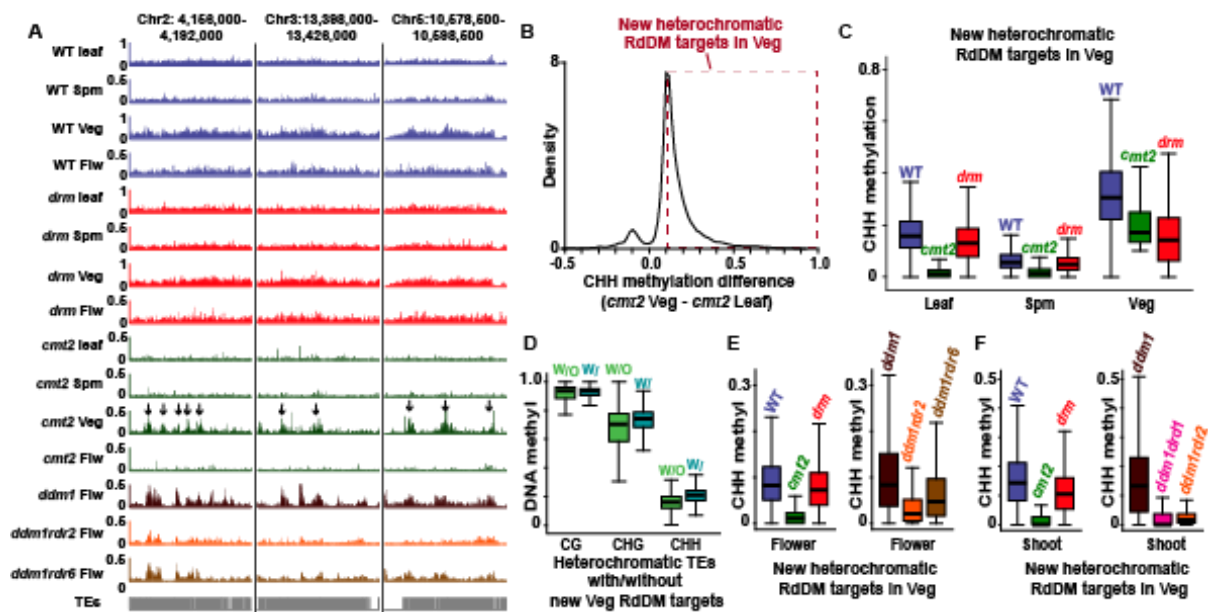


Figure 3.3. RdDM extends into heterochromatin in the vegetative cell.

(A) Snapshots of CHH methylation in leaf, sperm (Spm), vegetative cell (Veg), and flower (Flw) of WT, *cmt2*, and *drm* mutants (the arrows point to loci that retain CHH methylation in *cmt2* mutant vegetative cell, but not in leaf, sperm, and flower). (B) A density plot showing the frequency distribution of CHH DNA methylation differences between 50-bp windows in heterochromatic TEs that retain at least 10% CHH methylation in either *cmt2* vegetative cell or *cmt2* leaf. Only windows with at least 20 informative sequenced cytosines in both samples are included. The dashed box indicates windows in which CHH methylation is at least 10% greater in *cmt2* vegetative cell than in *cmt2* leaf. These windows are defined as new heterochromatic RdDM targets in the vegetative cell. (C) Box plots show CHH methylation levels of 50-bp windows marked by the dashed box in B (heterochromatic RdDM targets in the vegetative cell) in WT, *cmt2*, and *drm* mutants. Only windows with at least 20 informative sequenced cytosines in each sample are included. (D) Box plots show vegetative cell methylation levels of heterochromatic TEs with RdDM targets in the vegetative cell (W/) and other heterochromatic TEs (W/O). (E and F) Box plots show CHH methylation levels of 50-bp windows marked by the dashed box in B in the indicated genotypes for flowers (E) and shoots (F).

RdDM-targeted heterochromatic TEs are likely transcribed in the vegetative cell.

Arabidopsis TEs are generally transcriptionally silenced, but activation of heterochromatic TEs has been reported in the vegetative cell (180). Furthermore, transcriptional activation of such TE in the *ddm1* mutant background has been associated with RdDM targeting (81), suggesting that DRM-targeted heterochromatic TEs (Figure 3.3A, 3.3B and 3.3C) may be transcribed in the vegetative cell. To test this hypothesis, we examined how TEs targeted by RdDM in the vegetative cell are methylated in *ddm1* mutant flowers (81) and shoots (60). As expected, CHH methylation of these loci is maintained by CMT2 in WT flowers and shoots (Figure 3.3A, 3.3E and 3.3F). However, CHH methylation becomes RdDM-

dependent in both tissues in *ddm1* plants, as it is greatly reduced by additional mutation of either *RDR2* or *DRD1* (Figure 3.3E and 3.3F), genes that are required for RdDM (171). CHH methylation in *ddm1* flowers is also reduced by mutation of *RDR6*, which contributes to RdDM of transcriptionally active loci (Figure 3.3E) (81). This can be seen at individual TEs, where the *RDR2*- and *RDR6*-dependent methylation patterns of *ddm1* flowers resemble those of *cmt2* vegetative cells (Figure 3.3A).

Heterochromatic TEs targeted by RdDM in the vegetative cell are also more likely to be transcribed in *ddm1* mutants than other heterochromatic TEs (Figure 3.4A), and they tend to be much longer (Figure 3.4B), suggesting they are mostly full-length, transcriptionally competent elements (81). Overall, our results support the interpretation that RdDM targets transcriptionally activated heterochromatic TEs in the vegetative cell.

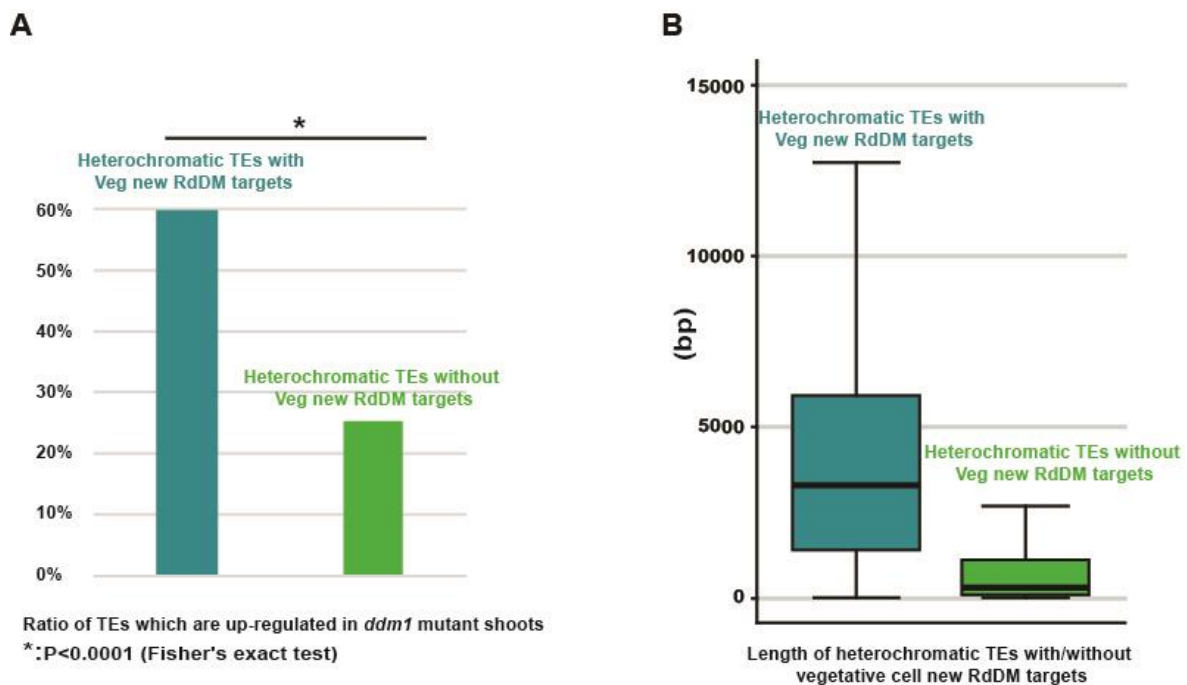


Figure 3.4. Heterochromatic TEs targeted by RdDM in the vegetative cell are transcriptionally competent. (A) RdDM-targeted TEs are more likely to be transcribed in *ddm1* mutant shoots (7). 804 out of the 1349 heterochromatic TEs targeted by RdDM in the vegetative cell are up-regulated in *ddm1* mutant shoots, whereas only 375 out of the 1479 heterochromatic TEs not targeted by RdDM in the vegetative cell are up-regulated. (B) Heterochromatic TEs targeted by RdDM in the vegetative cell (1349) are much longer than heterochromatic TEs not targeted by RdDM (1479).

Histone H1 is present in sperm but absent from the vegetative cell.

The targeting of heterochromatic loci by RdDM in the vegetative cell (Figure 3.3), as well as much higher levels of heterochromatic CHH methylation compared to sperm (Figure 3.1G), suggest that heterochromatin may be more accessible to DNA methyltransferases in the

vegetative nucleus. This is consistent with the decondensation of chromatin and lack of heterochromatic foci reported in the vegetative cell (112). Histone H1 is a chromatin protein that reduces the efficiency of heterochromatic DNA methylation (60). We therefore wondered if increased heterochromatic methylation may be caused by reduced abundance of H1 in the vegetative cell. The *Arabidopsis* genome encodes two canonical, widely expressed *H1* genes, *H1.1* and *H1.2*, as well as a truncated, stress-induced *H1.3* gene generally expressed at a much lower level (181).

We find that *H1.1-GFP* and *H1.2-GFP* reporter proteins (111) are present in sperm but undetectable in vegetative nuclei (Figure 3.5A and 3.5B). *H1.3-GFP* is not detectable in pollen (Figure 3.5C). Quantitative RT-PCR experiments show that *H1.1* and *H1.2* transcript levels are much lower in pollen than in leaf (Figure 3.5D and 3.5E), whereas *H1.3* is undetectable in pollen (Figure 3.5F). Vegetative cell chromatin thus appears to be depleted of histone H1, and sperm chromatin may contain less H1 than leaves and other tissues.

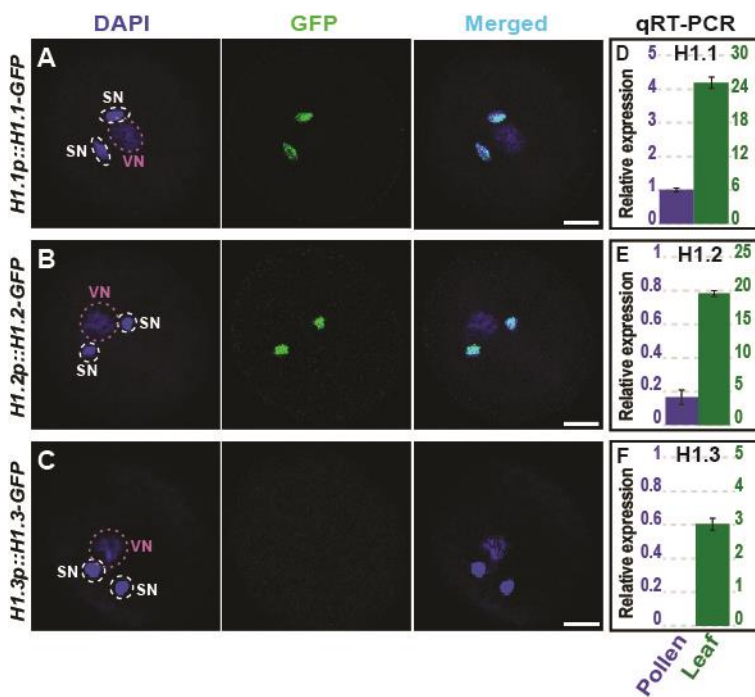


Figure 3.5. H1 expression in pollen. (A–C) H1.1-GFP (A) and H1.2-GFP (B) reporter constructs (27) are expressed in the sperm cells but not in the vegetative cell of pollen, whereas H1.3-GFP is not expressed in pollen (C). DAPI stains DNA in the pollen nuclei. SN, sperm nucleus; VN, vegetative nucleus. (Scale bar, 5 μm.) (D–F) Quantitative RT-PCR of H1.1 (D), H1.2 (E), and H1.3 (F) in pollen and leaf. The y axis scales show relative expression in pollen (left) and leaf (right).

Lack of H1 does not increase heterochromatic methylation in pollen.

To determine how histone H1 influences DNA methylation in pollen, we analyzed methylation of FACS-purified sperm and vegetative cells with mutations in *H1.1* and *H1.2* (*h1* mutants (60); Table 3.1). Unlike somatic tissues (60), vegetative cells do not show substantially increased heterochromatic methylation in any sequence context (Figure 3.6A, 3.6B and 3.6C), consistent with undetectable expression of H1 in the vegetative cell (Figure 3.5).

Heterochromatic methylation is also largely unaffected in sperm cells (Figure 3.6A, 3.6B and 3.6C), perhaps due to lower levels of H1 in this cell type (Figure 3.5). Overall, heterochromatic CG methylation in WT sperm and vegetative cells is similar to *h1* roots and substantially higher than in WT roots (Figure 3.6A), which is readily apparent even at individual loci (Figure 3.6D). Pollen heterochromatic CG methylation is therefore substantially higher than in somatic tissues, potentially due to reduced levels of H1.

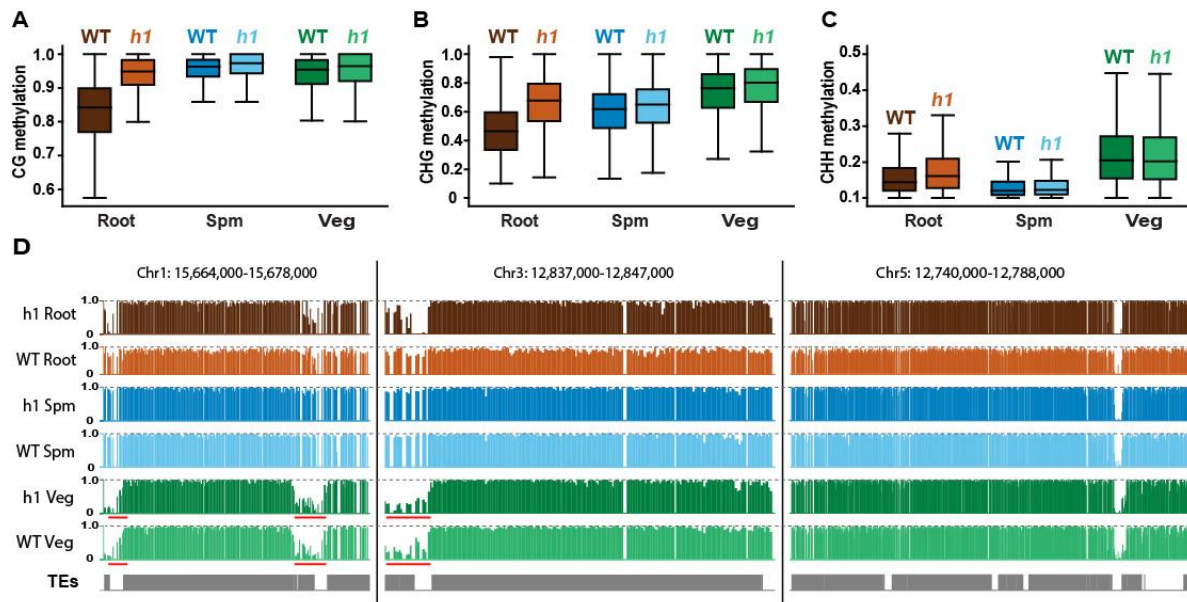


Figure 3.6. Lack of H1 does not increase heterochromatic methylation in pollen.

(A–C) Box plots show DNA methylation levels in 50-bp windows within heterochromatic TEs in WT and *h1* mutant root, sperm (Spm), and vegetative cell (Veg). Only windows with at least 10 informative sequenced cytosines and methylation of at least 30% for CG and 10% for CHG and CHH are included. (D) Snapshots of CG methylation in root, sperm, and vegetative cell of WT and *h1* mutants. Note how DNA methylation levels compare with the horizontal dashed line that denotes 100% methylation. Underlined regions bordering heterochromatic TEs are subject to active DNA demethylation in the vegetative cell (15).

CG methylation is more robustly maintained in pollen than in leaf or root.

We were intrigued by the higher levels of heterochromatic CG methylation in sperm and vegetative cells (Figure 3.6A and 3.6D) because they suggested an explanation for a longstanding mystery. The semi-conservative model of CG methylation maintenance (166) implies that even a modest reduction in maintenance efficiency below 100% should lead to dilution of methylation with each cell division that eventually causes complete loss (182). However, CG methylation levels measured in somatic cells (Figure 3.6A) (169, 170) appear to be below what would be required for stable maintenance (182). Somatic tissues do not contribute to the next generation, so higher methylation efficiency in cells that do, including gametes, would solve the problem. Indeed, methylation of individual CG sites in sperm and

vegetative cells, as well as in the microspores from which they arise, is much higher than in leaves and roots, both in heterochromatic TEs and genes (Figure 3.7A and 3.7B). CG methylation of euchromatic TEs is also higher in sperm and microspores than in somatic tissues (Figure 3.7C). The low methylation of euchromatic TEs in vegetative cells (Figure 3.7C) is attributable to extensive active DNA demethylation of such sequences in this cell type (Figure 4D) (97, 102), which may also explain the somewhat lower CG methylation of genes compared to sperm and microspores (Figure 3.7A).

A potential concern regarding our analysis is that we are comparing methylation between pure male reproductive cells and complex somatic tissues. Methylation heterogeneity between cell types within leaves and roots could, when averaged, create the impression of an overall lower methylation efficiency even though methylation within each cell type is as robust as in pollen. To circumvent this issue, we analyzed published methylation data from multiple purified root cell types (175). These cells show CG methylation levels comparable to those of whole roots and well below those of male reproductive cells (Figure 3.7D, 3.7E and 3.7F). Therefore, the lower CG methylation levels we observe in somatic cells and tissues are caused by reduced methylation efficiency rather than tissue heterogeneity.

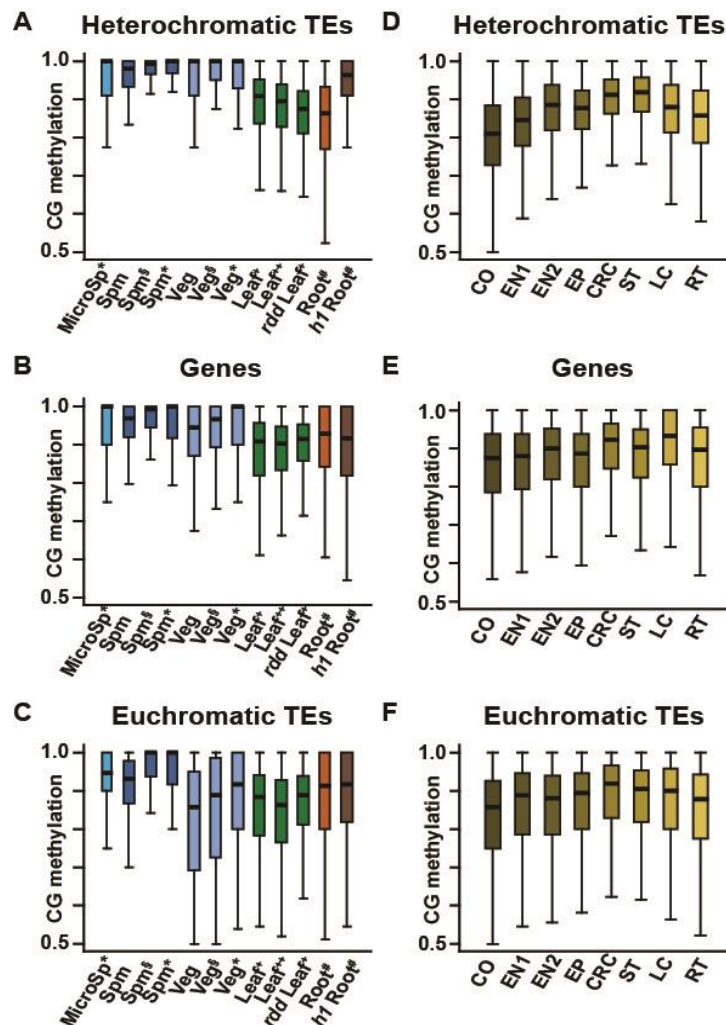


Figure 3.7. More robust maintenance of CG methylation in pollen. (A–F)

Box plots show CG methylation for individual CG sites with methylation greater than 50% and at least 10 informative sequenced cytosines. Published data in A–C are from the following: *, ref. 16; †, ref. 15; ‡, ref. 23; ‡‡, ref. 8; #, ref. 7. Published data in D–F are from ref. 17. CO, cortex; CRC, columella root cap; EN, endodermis; EP, epidermis; LC, lower columella; MicroSp, microspore; RT, root tip; Spm, sperm; ST, stele; Veg, vegetative cell.

To understand the mechanism of increased CG methylation efficiency in pollen, we tested several key known pathways. First, we asked if active DNA demethylation contributes to reduced efficiency in somatic tissues by analyzing methylation of leaves lacking ROS1, DML2 and DML3, the *Arabidopsis* DNA demethylases expressed in somatic tissues (183). The *ros1dml2dml3* (*rdd*) triple mutant did not substantially affect CG methylation in leaves compared with sperm and vegetative cells (Figure 3.7A, 3.7B and 3.7C), consistent with the limited genomic hypermethylation observed in this mutant (170). We also found that CG methylation of *drm*, *cmt2* and *cmt3* mutant sperm remained higher than in leaves in TEs and genes (Figure 3.8). Our data indicate that – with the exception of MET1 – no single known pathway can explain the increased CG methylation efficiency in pollen.

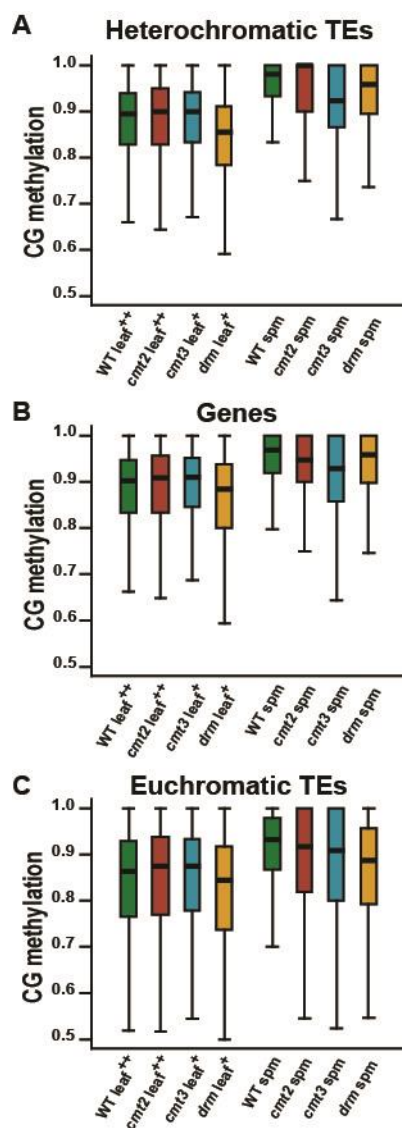


Figure 3.8. CG methylation of *drm*, *cmt2*, and *cmt3* mutant sperm remains higher than in leaves in TEs and genes.

(A–C) Box plots show CG methylation for individual CG sites with methylation greater than 50% and at least 10 informative sequenced cytosines.

Discussion

The nuclei of sperm and vegetative cells are drastically different. The sperm nucleus is small, with densely packed chromatin (Figure 3.5) and obvious H3K9me₂-containing heterochromatic foci (112). The vegetative nucleus is larger (Figure 3.5), lacks heterochromatic foci and cytologically-detectable H3K9me₂ (112), and has much higher levels of CHH methylation (Figure 3.1D) (97, 102). Chromomethylases are dependent on H3K9me₂, so it is reasonable to hypothesize that CHH and even CHG methylation in the vegetative nucleus may be largely dependent on RdDM. The columella cells in the *Arabidopsis* root cap also have much higher CHH methylation than neighboring cells, with elevated RdDM proposed as the cause (175). Our data do show that RdDM extends somewhat into heterochromatic TEs in vegetative cells (Figure 3.3), which is consistent with published results (112) and may reflect TE activation in the vegetative nucleus (180). However, CHG methylation is still dependent on CMT3 (Figure 3.1C) and heterochromatic CHH methylation primarily on CMT2 (Figure 3.1G). In general, maintenance of non-CG methylation is remarkably similar between vegetative, sperm and leaf cells (Figure 3.1C, 3.1D, 3.1F, and 3.1G). It is likely that the active removal of H3K9me₂ in the vegetative nucleus (113) occurs after DNA methylation is deposited, allowing CMTs to work. Overall, our results demonstrate that large, global CHH methylation changes can occur with only minor alterations of pathway specificity.

Although DNA methylation pathways function similarly in pollen and somatic tissues, the efficiency of CG methylation is substantially higher in pollen (Figure 3.7). The semi-conservative model of CG methylation inheritance (166), and the considerable genetic evidence in support of this model (116, 184, 185), is more compatible with the higher efficiency of CG methylation in the male germline (microspores and sperm), and with the similarly high efficiency observed in female plant gametes (186), than with that observed in somatic tissues (182). In heterochromatic TEs, much or all of the increased efficiency might be accounted by reduced levels of histone H1 (Figure 3.5), but the mechanism must be different in genes and euchromatic TEs, where loss of H1 does not facilitate CG methylation (Figure 3.7B and 3.7C). Instead of a unifying mechanism, the similar pollen-soma methylation efficiency differences in genes and TEs (Figure 3.7) may be an unavoidable consequence of distinct selection pressures. Gametes – and cells that might give rise to gametes – have the potential to undergo an essentially unlimited number of divisions that span generations, and should therefore be under strong selection for very efficient methylation maintenance. In comparison, somatic cells will divide very few times, and need a methylation maintenance activity that is just sufficient to keep TE silencing and other methylation functions from degenerating. Therefore, mutations that reduce methylation efficiency in somatic cells but keep it above this threshold would not be counter-selected. Under these conditions, somatic methylation efficiency would be expected to settle at this equilibrium threshold. An important consequence of less-efficient somatic

methylation is that small methylation differences between somatic tissues or cell types (175) may be caused by maintenance fluctuations rather than developmental reprogramming.

Materials and methods

Isolation of *A. thaliana* sperm and vegetative cell nuclei.

A. thaliana plants were grown under 16h light/ 8h dark in a growth chamber (20°C, 80% humidity). Sperm and vegetative cell nuclei were isolated by FACS based on via SYBR Green staining, as previously described (97).

Whole-genome bisulfite sequencing.

Bisulfite sequencing libraries for sperm and vegetative cells were constructed using the Ovation Ultralow Methyl-Seq Library Systems (Nugen, #0336) and EpiTect Fast Bisulfite Conversion (Qiagen, #59802) kits according to the kit protocols, except the incorporation of two rounds of bisulfite conversion. Illumina sequencing was performed at the UC Berkeley Vincent J. Coates Genomics Sequencing Laboratory, the DNA sequencing facility of the University of Cambridge Department of Biochemistry, Novogene Ltd., and the Bauer Core Facility at Harvard University. Sequenced reads (75 or 100 base single end) were mapped to the TAIR10 reference genome and DNA methylation of each cytosine was ascertained as previously described (97).

Published genomic data.

DNA methylation data for wild-type, *cmt3*, *cmt2*, and *drm1drm2* (*drm*) leaf tissue are from (61) and (66). DNA methylation data for wild-type and *h1* root is from (60). DNA methylation data for sperm, vegetative cell, and microspore are from (97) and (102). DNA methylation data for wild-type, *cmt2*, *drm2*, *ddm1*, *ddm1rdr2*, *ddm1rdr6* and *ddm1drd1* flower and shoot tissues is from (81) and (60). Leaf H3K9me2 and histone H3 ChIP-seq data are from (61).

Definition of genomic features.

Only genes with CG methylation between 20% and 60% were used for analysis in this paper. Overlapping TE annotations were merged.

Grouping of TEs based on the level of H3K9me2.

TEs were classified into three groups based on density plots: TEs with log₂ (H3K9me2 IP/H3 IP) below the lower peak in both replicates were classified as euchromatic; TEs with log₂ (H3K9me2 IP/H3 IP) between the peaks in both replicates were classified as intermediate; TEs with log₂ (H3K9me2 IP/H3 IP) above the higher peak in both replicates were classified as heterochromatic (Figure 3.9).

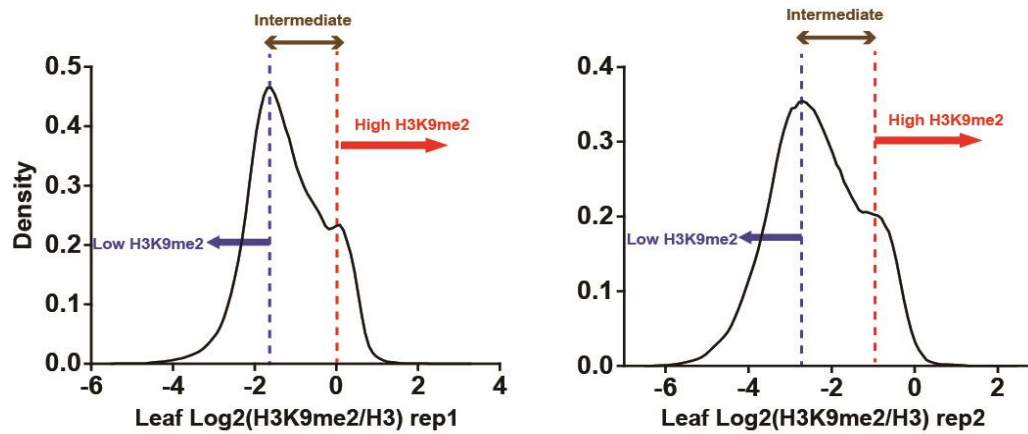


Figure 3.9 Grouping of TEs based on the level of H3K9me2 in two replicates.

Confocal microscopy.

Pollen grains were isolated by vortexing open flowers in PBS with 1% of Triton-X-100 and 0.1 $\mu\text{g/ml}$ of DAPI and spread on slides for confocal microscopy under DAPI and GFP channels, respectively.

Quantitative RT-PCR.

Total RNA was extracted from mature pollen and rosette leaves, respectively, and treated with DNase I. Equal amounts of total RNA from the two samples were used for reverse transcription and quantitative PCR. The *ACT8* gene was used as an internal control.

Chapter IV

Concluding remarks and discussion

There has been a substantial amount of data generated in the past ten years that delineates the profiles, regulation and function of epigenetic modifications in plants, including whole-genome bisulfite sequencing for methylomes (48, 49, 51), chromatin immunoprecipitation sequencing for histone modifications (61, 187), Hi-C sequencing for chromatin interactions (162, 188), MNase sequencing for nucleosome positioning (139, 189) and DNase I sequencing for DNA accessibility (190, 191). Among these types of epigenetic regulation, DNA methylation can be viewed as a fundamental component that helps determine the higher orders of chromatin structure. Therefore, to understand epigenetic regulation in development, it is important to know how DNA methylation is regulated within a generation, inherited through generations, how it interacts between parental epigenomes, and in subsequent generations of offspring. At the time when I started my project, the mechanism of CHH methylation maintenance in plants was a mystery. Many studies had generated several DNA methyltransferase compound mutants to determine whether CHH methylation is lost completely, but all the attempts either failed to remove all CHH methylation or the mutant physiological phenotype was too severe to be analyzed (48, 49, 192, 193).

We thus decided to perform a reverse genetic screen on *Arabidopsis* mutants to identify genes that can potentially affect plant DNA methylation by performing whole-genome bisulfite sequencing of each individual homozygous mutant seedling. We sequenced more than 250 mutants for genes which were predicted in the Chromatin Database website (<http://www.chromdb.org/>) to be associated with chromatin in some way, such as DNA methyltransferases, RdDM components, chromatin remodeling proteins, and histone modification enzymes. We discovered that CMT2 is responsible for more than 70% of CHH methylation in *Arabidopsis*, and works with DRM2 to mediate full levels of CHH methylation (Figure 2.1). Unlike DRM2-targeted TEs, we found that CMT2 preferentially targets long and H3K9me2-enriched heterochromatic TEs (Figure 2.2B, 2.3A, 2.3B and 2.3C). Moreover, the siRNA levels of CMT2-targeted sites are also relatively lower compared to DRM2-targeted sites and CMT2 also has canonical BAH- and chromodomains, which are predicted to enable H3K9me2 binding ability (Figure 2.3C and 2.6A). Together with these findings, we proposed a model for CMT2-mediated CHH methylation: CMT2 requires H3K9me2 to mediate CHH methylation, and CMT2-mediated CHH methylation is also required for H3K9me2 deposition in TEs. A study carried out by Stroud *et al.* further supports our model. Stroud *et al.* showed that CMT2 can bind to H3K9me2 and H3K9me1 using *in vitro* binding assays (61). The loss of non-CG methylation in the *drm1drm2cmt2cmt3* quadruple mutant is accompanied with the

complete loss of H3K9me2 (61), which is not observed in *cmt3* or *drm1drm2* mutants (61, 194), demonstrating that CMT2-mediated CHH methylation participates in the reinforcing loop between non-CG DNA methylation and H3K9me2. Together with our work, from these data we can propose a comprehensive three-part model for the maintenance of DNA methylation in *Arabidopsis*: first, MET1 is responsible for CG methylation of gene bodies and TEs; second, CMT2 and CMT3 together maintain non-CG methylation in TEs through a positive feedback loop with H3K9me2; and lastly, DRM2 maintains non-CG methylation of short and euchromatic TEs, which are not targeted by CMT2 and CMT3.

Plants such as *Arabidopsis* reprogram their DNA less than mammals and most DNA methylation states are faithfully inherited (4). We wondered how the current DNA methylation maintenance model can be applied to explain high-fidelity transgenerational DNA methylation inheritance. Because of the difficulties in collecting germline cells, most of our knowledge concerning transgenerational DNA methylation is based on studies performed in somatic tissues, which do not contribute to the DNA methylation pattern in offspring (48, 49). We therefore decided to study how DNA methylation is maintained in reproductive tissues by sequencing sperm cells and vegetative cells with mutations in non-CG methyltransferase genes. We confirmed that CG and CHG methylation in these cell types is well-maintained, as reported previously (Figure 3.1B and 3.1C) (97, 102).

The gain of heterochromatic CHH methylation in vegetative cells, and the loss of heterochromatic CHH methylation in sperm cells, was previously attributed to the activation of DRM (DRM1 and DRM2) and the silencing of DRM, respectively (102). The gain of DRM-mediated CHH methylation in vegetative cells can be further explained by the RdDM-favorable decondensed chromatin environment and the abundant 21-nt easiRNAs from derepressed heterochromatic TEs, proposed to be caused by the downregulation of DDM1 in the vegetative cell (102, 112). But what we found in our study was that CMT2 is the actual player involved in the gain of heterochromatic CHH methylation in the vegetative cell, instead of DRM (Figure 3.1D, 3.1G and 3.2C). Although we did find DRM activity slightly extends into heterochromatin in vegetative cells, likely reflecting the transcription of heterochromatin (Figure 3.3), heterochromatic CHH methylation is mostly dependent on CMT2 in sperm, vegetative cells, and somatic tissues (Figure 3.1D and 3.2C). It is likely that the loss of heterochromatic CHH methylation in sperm is also due to the silencing of CMT2 instead of DRM, given the similar enzyme dependency.

I would further argue that CHH remethylation of heterochromatic TEs in the zygote during embryogenesis is not mainly mediated by the reactivation of DRM, as has been previously proposed: The current model is that the easiRNAs from the vegetative cell are imported and

retained in the sperm cell, and they can guide RdDM to reestablish heterochromatic CHH methylation in the zygote after fertilization. However, if DRM is responsible for the remethylation of heterochromatic TEs in the zygote, we should expect to see a dramatic loss of heterochromatic CHH methylation in the *drm* mutant. However, many studies have shown that the loss of DRM causes reduced CHH methylation in short and euchromatic TEs rather than heterochromatic TEs in somatic tissues and likely also in embryos (Figure 2.1 and 2.2B) (48, 49). In contrast, as discussed above, several studies have shown that the CHH methylation of heterochromatic TEs is lost in *cmt2* mutant leaf and seedling, suggesting that heterochromatic CHH hypomethylation in the zygote cannot be remethylated without CMT2 (Figure 2.1 and 2.2B) (48, 49). Thus, it is plausible that CMT2 remethylates the zygotic CHH sites of heterochromatic TEs during embryogenesis. However, it is still possible that DRM does remethylate zygotic CHH sites of heterochromatic TEs during embryogenesis, but that the heterochromatic CHH methylation cannot be efficiently maintained during plant development without CMT2, leading to the loss of heterochromatic CHH methylation observed in the *cmt2* mutant leaf and seedling. To answer this question, one can simply sequence the methylomes of *drm* and *cmt2* mutant embryos to see whether the heterochromatic CHH methylation is regained. Furthermore, using CMT2 to remethylate heterochromatic TEs would also be more energetically efficient than using DRM because CMT3-mediated CHG is well-maintained in sperm cells and possibly in egg cells (Figure 3.1C). Thus, the reinforcing feedback loop between CMT2-mediated CHH methylation and H3K9me2 should be reestablished efficiently in the zygote (61).

For CG methylation of pollen, we found that pollen displays high fidelity CG methylation maintenance compared to somatic tissues, allowing methylation patterns to be accurately inherited across generations (Figure 3.7). The high efficiency of CG methylation maintenance in pollen can be partially explained by increased accessibility due to the lower levels of H1 linker histone protein in heterochromatic TEs, which would in principle allow methyltransferases greater access, as shown for somatic tissues (Figure 3.5). However, the reason why euchromatic sites are also maintained with equally high efficiency in the pollen remains unknown.

The assumption that the downregulation of DDM1 in the vegetative cell leads to the derepression of heterochromatic TEs (102) has become arguable, following our new analysis on the methylome of the germline. Reactivation of TEs in *ddm1* mutant somatic tissues is most likely due to the dramatic loss of DNA methylation in all three contexts, and a decondensed chromatin state can result from losing DNA methylation (Figure 2.2A) (195, 196). In support of this, the DNA methyltransferase mutant *met1* also exhibits dramatic loss of DNA methylation and decondensed chromatin (49, 196). Since DNA methylation in the vegetative

cell is well-maintained, however, whereby levels are even higher than that of somatic tissues in all three contexts (Figure 3.1B, 3.1C and 3.1D), the downregulation of DDM1 is unlikely to be the main reason why chromatin is decondensed in the vegetative cell, which subsequently causes derepression of heterochromatic TEs. Instead, chromatin decondensation in the vegetative cell is likely to be due to downregulation of as-yet unidentified chromatin regulatory proteins, e.g. those that maintain the superstructure of chromosomes. Indeed, AtMORC1 and AtMORC6 in *Arabidopsis* are predicted to catalyze alterations in chromosome superstructure, and the loss of them leads to decondensation of pericentromeric heterochromatin accompanied by derepression of TEs without concomitant DNA methylation loss, reinforcing the importance of chromosome superstructure on TE silencing (162).

Two questions still need to be addressed in more detail. First, why do plants lower their CMT2-mediated CHH methylation during sexual reproduction? Second, why do plants generate so many 21-nt easiRNAs that are probably not used for DNA methylation? A decrease of CMT2-mediated CHH methylation was previously observed in *Arabidopsis* inflorescences (81, 102), suggesting that it might be very important to suppress CMT2 during sexual reproduction. Our observation that the CMT2 seems to be missing in maize might provide a potential explanation. The loss of CMT2 was reported to increase heat stress tolerance in *Arabidopsis* and the natural accessions of *Arabidopsis* with CMT2 mutation were found to be associated with the adaptation to variable temperatures (197, 198). Consistent with this finding, maize is known for possessing a higher temperature optimum than C3 plants (199), suggesting that it is possible that the loss of CMT2-mediated heterochromatic CHH methylation during sexual reproduction might increase the heat tolerance of the male germline cells in *Arabidopsis*. To test if the hypothesis is correct, ectopically expressing CMT2 during sexual reproduction could be expected to make germlines more susceptible to heat and heat-related stresses.

As discussed above, 21-nt easiRNAs do not account for the major gain of heterochromatic CHH methylation through RdDM in the vegetative cell, and possibly also not in the zygote. However, since it is likely that CMT2 is turned off in the sperm for unknown reasons, the existence of easiRNAs can be more important during sexual reproduction by antagonizing accidentally expressed TEs in the sperm via PTGS to compensate the loss of CHH methylation. Some small RNAs generated from retrotransposons were shown to be able to regulate gene expression and influence plant stress response (200). Therefore, in addition to genomic immunity, it is also possible that these 21-nt easiRNA from heterochromatic TEs might regulate expression of some important genes *in-trans* during sexual reproduction (82).

Overall, my findings expand the previous incomplete DNA methylation maintenance model with the discovery of CMT2-mediated CHH methylation. My work on the pollen

methylome contributes to our further understanding of the regulation and reprogramming of DNA methylation during sexual reproduction, providing insights into transgenerational epigenetic inheritance in plants.

References

1. Mosher RA & Melnyk CW (2010) siRNAs and DNA methylation: seedy epigenetics. *Trends Plant Sci* 15(4):204-210.
2. Goldberg AD, Allis CD, & Bernstein E (2007) Epigenetics: a landscape takes shape. *Cell* 128(4):635-638.
3. Slotkin RK & Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8(4):272-285.
4. Law JA & Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11(3):204-220.
5. Jenuwein T & Allis CD (2001) Translating the histone code. *Science* 293(5532):1074-1080.
6. Borges F & Martienssen RA (2015) The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* 16(12):727-741.
7. Lisch D (2009) Epigenetic Regulation of Transposable Elements in Plants. *Annu Rev Plant Biol* 60:43-66.
8. Axtell MJ (2013) Classification and Comparison of Small RNAs from Plants. *Annual Review of Plant Biology, Vol 64* 64:137-159.
9. Bousios A, *et al.* (2016) A role for palindromic structures in the cis-region of maize Sirevirus LTRs in transposable element evolution and host epigenetic response. *Genome Res* 26(2):226-237.
10. Slotkin RK, Freeling M, & Lisch D (2005) Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nat Genet* 37(6):641-644.
11. Tenllado F & Diaz-Ruiz JR (2001) Double-stranded RNA-mediated interference with plant virus infection. *J Virol* 75(24):12288-12297.
12. He L & Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 5(7):522-531.
13. Lee O & Kim VN (2004) Evidence that microRNA genes are transcribed by RNA polymerase II. *Cell Struct Funct* 29:68-68.
14. Lee Y, *et al.* (2004) MicroRNA genes are transcribed by RNA polymerase II. *Embo J* 23(20):4051-4060.
15. Valencia-Sanchez MA, Liu JD, Hannon GJ, & Parker R (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. *Gene Dev* 20(5):515-524.
16. Rogers K & Chen XM (2013) Biogenesis, Turnover, and Mode of Action of Plant MicroRNAs. *Plant Cell* 25(7):2383-2399.
17. McElver J, *et al.* (2001) Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. *Genetics* 159(4):1751-1763.
18. Schwartz BW, Yeung EC, & Meinke DW (1994) Disruption of Morphogenesis and Transformation of the Suspensor in Abnormal Suspensor Mutants of *Arabidopsis*.

- Development* 120(11):3235-3245.
19. Thompson BE, *et al.* (2014) The dicer-like1 Homolog fuzzy tassel Is Required for the Regulation of Meristem Determinacy in the Inflorescence and Vegetative Growth in Maize. *Plant Cell* 26(12):4702-4717.
 20. Liu B, *et al.* (2005) Loss of function of OsDCL1 affects microRNA accumulation and causes developmental defects in rice. *Plant Physiology* 139(1):296-305.
 21. Fei Q, Xia R, & Meyers BC (2013) Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* 25(7):2400-2415.
 22. Ronemus M, Vaughn MW, & Martienssen RA (2006) MicroRNA-targeted and small interfering RNA-mediated mRNA degradation is regulated by Argonaute, Dicer, and RNA-dependent RNA polymerase in Arabidopsis. *Plant Cell* 18(7):1559-1574.
 23. Cuperus JT, *et al.* (2010) Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nat Struct Mol Biol* 17(8):997-U111.
 24. Wang XB, *et al.* (2010) RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana. *P Natl Acad Sci USA* 107(1):484-489.
 25. Xie ZX, *et al.* (2004) Genetic and functional diversification of small RNA pathways in plants. *Plos Biol* 2(5):642-652.
 26. Wierzbicki AT, Haag JR, & Pikaard CS (2008) Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. *Cell* 135(4):635-648.
 27. Onodera Y, *et al.* (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120(5):613-622.
 28. Pikaard CS, Haag JR, Ream T, & Wierzbicki AT (2008) Roles of RNA polymerase IV in gene silencing. *Trends Plant Sci* 13(7):390-397.
 29. Zilberman D, Cao XF, & Jacobsen SE (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299(5607):716-719.
 30. Cuerda-Gil D & Slotkin RK (2016) Non-canonical RNA-directed DNA methylation. *Nat Plants* 2(11):16163.
 31. Wierzbicki AT, Ream TS, Haag JR, & Pikaard CS (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* 41(5):630-634.
 32. Zhang HM & Zhu JK (2011) RNA-directed DNA methylation. *Curr Opin Plant Biol* 14(2):142-147.
 33. Cao XF & Jacobsen SE (2002) Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* 12(13):1138-1144.
 34. Luger K, Mader AW, Richmond RK, Sargent DF, & Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature* 389(6648):251-260.

35. Luger K, Dechassa ML, & Tremethick DJ (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Bio* 13(7):436-447.
36. Thomas JO (1999) Histone H1: location and role. *Curr Opin Cell Biol* 11(3):312-317.
37. Raghuram N, Carrero G, Th'ng J, & Hendzel MJ (2009) Molecular dynamics of histone H1. *Biochem Cell Biol* 87(1):189-206.
38. Robinson PJJ & Rhodes D (2006) Structure of the '30 nm' chromatin fibre: A key role for the linker histone. *Curr Opin Struc Biol* 16(3):336-343.
39. Fan YH, *et al.* (2005) Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* 123(7):1199-1212.
40. Talbert PB & Henikoff S (2016) Histone variants on the move: substrates for chromatin dynamics. *Nat Rev Mol Cell Biol*.
41. McKinley KL & Cheeseman IM (2016) The molecular basis for centromere identity and function. *Nat Rev Mol Cell Biol* 17(1):16-29.
42. Steiner FA & Henikoff S (2015) Diversity in the organization of centromeric chromatin. *Curr Opin Genet Dev* 31:28-35.
43. Allis CD & Jenuwein T (2016) The molecular hallmarks of epigenetic control. *Nat Rev Genet* 17(8):487-500.
44. Struhl K (1998) Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev* 12(5):599-606.
45. Bannister AJ & Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* 21(3):381-395.
46. Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature* 447(7143):407-412.
47. Jackson JP, Lindroth AM, Cao X, & Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* 416(6880):556-560.
48. Cokus SJ, *et al.* (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452(7184):215-219.
49. Lister R, *et al.* (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133(3):523-536.
50. Lister R, *et al.* (2013) Global epigenomic reconfiguration during mammalian brain development. *Science* 341(6146):1237905.
51. Zemach A, McDaniel IE, Silva P, & Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328(5980):916-919.
52. Coleman-Derr D & Zilberman D (2012) Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. *PLoS Genet* 8(10):e1002988.
53. Hollister JD & Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-

- off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* 19(8):1419-1428.
54. Goll MG & Bestor TH (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74:481-514.
 55. Bostick M, *et al.* (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317(5845):1760-1764.
 56. Woo HR, Pontes O, Pikaard CS, & Richards EJ (2007) VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes Dev* 21(3):267-277.
 57. Woo HR, Dittmer TA, & Richards EJ (2008) Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. *PLoS Genet* 4(8):e1000156.
 58. Sharif J, *et al.* (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450(7171):908-912.
 59. Kankel MW, *et al.* (2003) Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics* 163(3):1109-1122.
 60. Zemach A, *et al.* (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153(1):193-205.
 61. Stroud H, *et al.* (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol* 21(1):64-72.
 62. Johnson LM, *et al.* (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr Biol* 17(4):379-384.
 63. Du J, *et al.* (2014) Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Mol Cell* 55(3):495-504.
 64. Du J, *et al.* (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* 151(1):167-180.
 65. Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, & Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. *PLoS One* 3(9):e3156.
 66. Stroud H, Greenberg MV, Feng S, Bernatavichute YV, & Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* 152(1-2):352-364.
 67. Du J, Johnson LM, Jacobsen SE, & Patel DJ (2015) DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol* 16(9):519-532.
 68. Jullien PE, Susaki D, Yelagandula R, Higashiyama T, & Berger F (2012) DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. *Curr Biol* 22(19):1825-1830.
 69. Zhong X, *et al.* (2015) Domains rearranged methyltransferase3 controls DNA methylation and regulates RNA polymerase V transcript abundance in Arabidopsis.

- Proc Natl Acad Sci U S A* 112(3):911-916.
70. Costa-Nunes P, Kim JY, Hong E, & Pontes O (2014) The cytological and molecular role of domains rearranged methyltransferase3 in RNA-dependent DNA methylation of *Arabidopsis thaliana*. *BMC Res Notes* 7:721.
 71. Blevins T, *et al.* (2015) Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in *Arabidopsis*. *Elife* 4:e09591.
 72. Zhai J, *et al.* (2015) A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis. *Cell* 163(2):445-455.
 73. Bohmdorfer G, *et al.* (2016) Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. *Elife* 5.
 74. Law JA, *et al.* (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* 498(7454):385-389.
 75. Johnson LM, *et al.* (2014) SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* 507(7490):124-128.
 76. Zhang H, *et al.* (2013) 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* 110(20):8290-8295.
 77. Nuthikattu S, *et al.* (2013) The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol* 162(1):116-131.
 78. Panda K & Slotkin RK (2013) Proposed mechanism for the initiation of transposable element silencing by the RDR6-directed DNA methylation pathway. *Plant Signal Behav* 8(8).
 79. Fultz D, Choudury SG, & Slotkin RK (2015) Silencing of active transposable elements in plants. *Curr Opin Plant Biol* 27:67-76.
 80. McCue AD, *et al.* (2015) ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *Embo J* 34(1):20-35.
 81. Panda K, *et al.* (2016) Full-length autonomous transposable elements are preferentially targeted by expression-dependent forms of RNA-directed DNA methylation. *Genome Biol* 17(1):170.
 82. Creasey KM, *et al.* (2014) miRNAs trigger widespread epigenetically activated siRNAs from transposons in *Arabidopsis*. *Nature* 508(7496):411-415.
 83. Duan CG, *et al.* (2015) Specific but interdependent functions for *Arabidopsis* AGO4 and AGO6 in RNA-directed DNA methylation. *Embo J* 34(5):581-592.
 84. Bond DM & Baulcombe DC (2015) Epigenetic transitions leading to heritable, RNA-mediated de novo silencing in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 112(3):917-922.
 85. Zheng B, *et al.* (2009) Intergenic transcription by RNA polymerase II coordinates Pol

- IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. *Genes Dev* 23(24):2850-2860.
86. Mari-Ordonez A, *et al.* (2013) Reconstructing de novo silencing of an active plant retrotransposon. *Nat Genet* 45(9):1029-1039.
 87. Choi Y, *et al.* (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. *Cell* 110(1):33-42.
 88. Schoft VK, *et al.* (2011) Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. *Proc Natl Acad Sci U S A* 108(19):8042-8047.
 89. Penterman J, *et al.* (2007) DNA demethylation in the Arabidopsis genome. *Proc Natl Acad Sci U S A* 104(16):6752-6757.
 90. Liu R, *et al.* (2015) A DEMETER-like DNA demethylase governs tomato fruit ripening. *Proc Natl Acad Sci U S A* 112(34):10804-10809.
 91. Ono A, *et al.* (2012) A null mutation of ROS1a for DNA demethylation in rice is not transmittable to progeny. *Plant J* 71(4):564-574.
 92. Yamamuro C, *et al.* (2014) Overproduction of stomatal lineage cells in Arabidopsis mutants defective in active DNA demethylation. *Nat Commun* 5:4062.
 93. Williams BP, Pignatta D, Henikoff S, & Gehring M (2015) Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. *PLoS Genet* 11(3):e1005142.
 94. Lei M, *et al.* (2015) Regulatory link between DNA methylation and active demethylation in Arabidopsis. *Proc Natl Acad Sci U S A* 112(11):3553-3557.
 95. Tang K, Lang Z, Zhang H, & Zhu JK (2016) The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. *Nat Plants* 2(11):16169.
 96. Qian W, *et al.* (2012) A histone acetyltransferase regulates active DNA demethylation in Arabidopsis. *Science* 336(6087):1445-1448.
 97. Ibarra CA, *et al.* (2012) Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* 337(6100):1360-1364.
 98. Ikeda Y, *et al.* (2011) HMG domain containing SSRP1 is required for DNA demethylation and genomic imprinting in Arabidopsis. *Dev Cell* 21(3):589-596.
 99. Kaufmann K, Pajoro A, & Angenent GC (2010) Regulation of transcription in plants: mechanisms controlling developmental switches. *Nat Rev Genet* 11(12):830-842.
 100. Kawashima T & Berger F (2014) Epigenetic reprogramming in plant sexual reproduction. *Nat Rev Genet* 15(9):613-624.
 101. Heard E & Martienssen RA (2014) Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* 157(1):95-109.
 102. Calarco JP, *et al.* (2012) Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* 151(1):194-205.
 103. Li W, *et al.* (2011) DNA methylation and histone modifications regulate de novo

- shoot regeneration in *Arabidopsis* by modulating WUSCHEL expression and auxin signaling. *PLoS Genet* 7(8):e1002243.
104. Ahmad A, Zhang Y, & Cao XF (2010) Decoding the epigenetic language of plant development. *Mol Plant* 3(4):719-728.
 105. Zhu D, Rosa S, & Dean C (2015) Nuclear organization changes and the epigenetic silencing of FLC during vernalization. *J Mol Biol* 427(3):659-669.
 106. Twell D (2011) Male gametogenesis and germline specification in flowering plants. *Sex Plant Reprod* 24(2):149-160.
 107. Yang WC, Shi DQ, & Chen YH (2010) Female gametophyte development in flowering plants. *Annu Rev Plant Biol* 61:89-108.
 108. Faure JE (2001) Double fertilization in flowering plants: discovery, study methods and mechanisms. *C R Acad Sci III* 324(6):551-558.
 109. Dresselhaus T, Sprunck S, & Wessel GM (2016) Fertilization Mechanisms in Flowering Plants. *Curr Biol* 26(3):R125-139.
 110. She W & Baroux C (2015) Chromatin dynamics in pollen mother cells underpin a common scenario at the somatic-to-reproductive fate transition of both the male and female lineages in *Arabidopsis*. *Front Plant Sci* 6:294.
 111. She W, *et al.* (2013) Chromatin reprogramming during the somatic-to-reproductive cell fate transition in plants. *Development* 140(19):4008-4019.
 112. Schoft VK, *et al.* (2009) Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. *EMBO Rep* 10(9):1015-1021.
 113. Merai Z, *et al.* (2014) The AAA-ATPase molecular chaperone Cdc48/p97 disassembles sumoylated centromeres, decondenses heterochromatin, and activates ribosomal RNA genes. *Proc Natl Acad Sci U S A* 111(45):16166-16171.
 114. Olmedo-Monfil V, *et al.* (2010) Control of female gamete formation by a small RNA pathway in *Arabidopsis*. *Nature* 464(7288):628-632.
 115. Singh M, *et al.* (2011) Production of viable gametes without meiosis in maize deficient for an ARGONAUTE protein. *Plant Cell* 23(2):443-458.
 116. Feng S, Jacobsen SE, & Reik W (2010) Epigenetic reprogramming in plant and animal development. *Science* 330(6004):622-627.
 117. Law JA & Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11(3):204-220.
 118. Haag JR & Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol* 12(8):483-492.
 119. Saze H & Kakutani T (2011) Differentiation of epigenetic modifications between transposons and genes. *Curr Opin Plant Biol* 14(1):81-87.
 120. Jeddloh JA, Stokes TL, & Richards EJ (1999) Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat Genet* 22(1):94-97.

121. Lippman Z, *et al.* (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430(6998):471-476.
122. Ryan DP & Owen-Hughes T (2011) Snf2-family proteins: chromatin remodellers for any occasion. *Curr Opin Chem Biol* 15(5):649-656.
123. Brzeski J & Jerzmanowski A (2003) Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. *J Biol Chem* 278(2):823-828.
124. Tao Y, *et al.* (2011) Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences. *Proc Natl Acad Sci U S A* 108(14):5626-5631.
125. Tsukahara S, *et al.* (2009) Bursts of retrotransposition reproduced in Arabidopsis. *Nature* 461(7262):423-426.
126. Blevins T, Pontes O, Pikaard CS, & Meins F, Jr. (2009) Heterochromatic siRNAs and DDM1 independently silence aberrant 5S rDNA transcripts in Arabidopsis. *PLoS One* 4(6):e5932.
127. Teixeira FK, *et al.* (2009) A role for RNAi in the selective correction of DNA methylation defects. *Science* 323(5921):1600-1604.
128. Xie Z, *et al.* (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* 2(5):E104.
129. Kanno T, *et al.* (2004) Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr Biol* 14(9):801-805.
130. Law JA, *et al.* (2010) A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in Arabidopsis. *Curr Biol* 20(10):951-956.
131. Zhong X, *et al.* (2012) DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat Struct Mol Biol.*
132. Wierzbicki AT, *et al.* (2012) Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. *Genes Dev* 26(16):1825-1836.
133. Cokus SJ, *et al.* (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452(7184):215-219.
134. Buisine N, Quesneville H, & Colot V (2008) Improved detection and annotation of transposable elements in sequenced genomes using multiple reference sequence sets. *Genomics* 91(5):467-475.
135. Ahmed I, Sarazin A, Bowler C, Colot V, & Quesneville H (2011) Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. *Nucleic Acids Res* 39(16):6919-6931.
136. Iyer VR (2012) Nucleosome positioning: bringing order to the eukaryotic genome. *Trends Cell Biol* 22(5):250-256.
137. Roudier F, *et al.* (2011) Integrative epigenomic mapping defines four main chromatin

- states in Arabidopsis. *EMBO J* 30(10):1928-1938.
138. Lee TF, *et al.* (2012) RNA polymerase V-dependent small RNAs in Arabidopsis originate from small, intergenic loci including most SINE repeats. *Epigenetics* 7(7):781-795.
 139. Chodavarapu RK, *et al.* (2010) Relationship between nucleosome positioning and DNA methylation. *Nature* 466(7304):388-392.
 140. Thomas JO (1999) Histone H1: location and role. *Curr Opin Cell Biol* 11(3):312-317.
 141. Pennings S, Meersseman G, & Bradbury EM (1994) Linker histones H1 and H5 prevent the mobility of positioned nucleosomes. *Proc Natl Acad Sci U S A* 91(22):10275-10279.
 142. Robinson PJ & Rhodes D (2006) Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr Opin Struct Biol* 16(3):336-343.
 143. Ascenzi R & Gantt JS (1999) Subnuclear distribution of the entire complement of linker histone variants in Arabidopsis thaliana. *Chromosoma* 108(6):345-355.
 144. Barra JL, Rhounim L, Rossignol JL, & Faugeron G (2000) Histone H1 is dispensable for methylation-associated gene silencing in Ascombolus immersus and essential for long life span. *Mol Cell Biol* 20(1):61-69.
 145. Fan Y, *et al.* (2005) Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* 123(7):1199-1212.
 146. Raghuram N, Carrero G, Th'ng J, & Hendzel MJ (2009) Molecular dynamics of histone H1. *Biochem Cell Biol* 87(1):189-206.
 147. Wierzbicki AT & Jerzmanowski A (2005) Suppression of histone H1 genes in Arabidopsis results in heritable developmental defects and stochastic changes in DNA methylation. *Genetics* 169(2):997-1008.
 148. Rea M, *et al.* (2012) Histone H1 affects gene imprinting and DNA methylation in Arabidopsis. *Plant J* 71(5):776-786.
 149. Mirouze M, *et al.* (2009) Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* 461(7262):427-430.
 150. You W, *et al.* (2012) Atypical DNA methylation of genes encoding cysteine-rich peptides in Arabidopsis thaliana. *BMC Plant Biol* 12:51.
 151. Zemach A, *et al.* (2010) Local DNA hypomethylation activates genes in rice endosperm. *Proc Natl Acad Sci U S A* 107(43):18729-18734.
 152. Saze H & Kakutani T (2007) Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. *EMBO J* 26(15):3641-3652.
 153. Miura A, *et al.* (2009) An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J* 28(8):1078-1086.
 154. Rigal M, Kevei Z, Pelissier T, & Mathieu O (2012) DNA methylation in an intron of the

- IBM1 histone demethylase gene stabilizes chromatin modification patterns. *EMBO J* 31(13):2981-2993.
155. Henderson IR & Jacobsen SE (2008) Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev* 22(12):1597-1606.
 156. Hsieh TF, *et al.* (2011) Regulation of imprinted gene expression in Arabidopsis endosperm. *Proc Natl Acad Sci U S A* 108(5):1755-1762.
 157. Ibarra CA, *et al.* (2012) Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* 337(6100):1360-1364.
 158. Grewal SI & Jia S (2007) Heterochromatin revisited. *Nat Rev Genet* 8(1):35-46.
 159. Probst AV, *et al.* (2004) Arabidopsis histone deacetylase HDA6 is required for maintenance of transcriptional gene silencing and determines nuclear organization of rDNA repeats. *Plant Cell* 16(4):1021-1034.
 160. Martienssen RA (2003) Maintenance of heterochromatin by RNA interference of tandem repeats. *Nat Genet* 35(3):213-214.
 161. Xiao W, *et al.* (2003) Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Developmental Cell* 5(6):891-901.
 162. Moissiard G, *et al.* (2012) MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* 336(6087):1448-1451.
 163. Kim SY, Lee J, Eshed-Williams L, Zilberman D, & Sung ZR (2012) EMF1 and PRC2 cooperate to repress key regulators of Arabidopsis development. *PLoS Genet* 8(3):e1002512.
 164. Robinson MD, McCarthy DJ, & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.
 165. Benjamini Y & Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* (57).
 166. Law JA & Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* 11:204-220.
 167. Kim MY & Zilberman D (2014) DNA methylation as a system of plant genomic immunity. *Trends Plant Sci* 19(5):320-326.
 168. Weigel D & Colot V (2012) Epialleles in plant evolution. *Genome Biol* 13(10):249.
 169. Cokus SJ, *et al.* (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452:215-219.
 170. Lister R, *et al.* (2008) Highly integrated single-base resolution maps of the Arabidopsis genome. *Cell* 133(3):523-536.
 171. Matzke MA & Mosher RA (2014) RNA-directed DNA methylation: an epigenetic

- pathway of increasing complexity. *Nat Rev Genet* 15(6):394-408.
172. Zhong X, *et al.* (2012) DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat Struct Mol Biol* 19(9):870-875.
 173. Huettel B, *et al.* (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *EMBO J* 25(12):2828-2836.
 174. Hsieh T-F, *et al.* (2009) Genome-wide demethylation of *Arabidopsis* endosperm. *Science* 324:1451-1454.
 175. Kawakatsu T, *et al.* (2016) Unique cell-type-specific patterns of DNA methylation in the root meristem. *Nat Plants* 2(5):16058.
 176. Irish VF (2010) The flowering of Arabidopsis flower development. *Plant J* 61(6):1014-1028.
 177. Feng X, Zilberman D, & Dickinson H (2013) A conversation across generations: soma-germ cell crosstalk in plants. *Dev Cell* 24(3):215-225.
 178. Schmidt A, Schmid MW, & Grossniklaus U (2015) Plant germline formation: common concepts and developmental flexibility in sexual and asexual reproduction. *Development* 142(2):229-241.
 179. Saffman EE & Lasko P (1999) Germline development in vertebrates and invertebrates. *Cell Mol Life Sci* 55(8-9):1141-1163.
 180. Slotkin RK, *et al.* (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136(3):461-472.
 181. Jerzmanowski A, Przewłoka M, & Grasser KD (2000) Linker histones and HMG1 proteins of higher plants. *Plant Biology* 2(6):586-597.
 182. Haerter JO, Lovkvist C, Dodd IB, & Snepken K (2014) Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic Acids Res* 42(4):2235-2244.
 183. Penterman J, *et al.* (2007) DNA demethylation in the Arabidopsis genome. *Proc Natl Acad Sci U S A* 104:6752-6757.
 184. Becker C, *et al.* (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. *Nature* 480(7376):245-249.
 185. Schmitz RJ, *et al.* (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science* 334(6054):369-373.
 186. Park K, *et al.* (2016) DNA demethylation is initiated in the central cells of Arabidopsis and rice *Proc Natl Acad Sci U S A* submitted.
 187. Sani E, Herzyk P, Perrella G, Colot V, & Amtmann A (2013) Hyperosmotic priming of Arabidopsis seedlings establishes a long-term somatic memory accompanied by specific changes of the epigenome. *Genome Biol* 14(6):R59.
 188. Feng S, *et al.* (2014) Genome-wide Hi-C analyses in wild-type and mutants reveal

- high-resolution chromatin interactions in Arabidopsis. *Mol Cell* 55(5):694-707.
189. Zhang X, *et al.* (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. *Plos Biol* 5(5):e129.
 190. Shu H, Gruissem W, & Hennig L (2013) Measuring Arabidopsis chromatin accessibility using DNase I-polymerase chain reaction and DNase I-chip assays. *Plant Physiol* 162(4):1794-1801.
 191. Zhang W, *et al.* (2012) High-resolution mapping of open chromatin in the rice genome. *Genome Res* 22(1):151-162.
 192. Chan SW, *et al.* (2006) RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in arabidopsis. *PLoS Genet* 2(6):e83.
 193. Zhang X & Jacobsen SE (2006) Genetic analyses of DNA methyltransferases in Arabidopsis thaliana. *Cold Spring Harb Symp Quant Biol* 71:439-447.
 194. Inagaki S, *et al.* (2010) Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. *Embo J* 29(20):3496-3506.
 195. Fransz P, ten Hoopen R, & Tessadori F (2006) Composition and formation of heterochromatin in Arabidopsis thaliana. *Chromosome Res* 14(1):71-82.
 196. Soppe WJ, *et al.* (2002) DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. *Embo J* 21(23):6549-6559.
 197. Dubin MJ, *et al.* (2015) DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. *Elife* 4:e05255.
 198. Shen X, *et al.* (2014) Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. *PLoS Genet* 10(12):e1004842.
 199. Crafts-Brandner SJ & Salvucci ME (2002) Sensitivity of photosynthesis in a C4 plant, maize, to heat stress. *Plant Physiol* 129(4):1773-1780.
 200. McCue AD, Nuthikattu S, Reeder SH, & Slotkin RK (2012) Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genet* 8(2):e1002474.