Control of Chromatin by RNA-mediated Transcriptional Silencing

by

Mohd Hafiz Bin Mohd Rothi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular, Cellular, and Developmental Biology)
in The University of Michigan
2021

Doctoral Committee:

Professor Andrzej T. Wierzbicki, Chair
Associate Professor Alan Boyle
Professor Steven Clark
Professor Gyorgyi Csankovski
Assistant Professor Kaushik Ragunathan

Mohd Hafiz Bin Mohd Rothi

mrmhafiz@umich.edu

ORCID ID: 0000-0001-7301-4719

To my wife, Pusparanee and our daughters, Iris and Irene.

ACKNOWLEDGEMENTS

I would like to thank Andrzej for all the guidance and help throughout my Ph.D. All the words of wisdom and experiences shared with me has made me think more critically about science and will help me be better prepared to face future challenges in research. I would like to acknowledge and express my utmost gratitude to all my lab mates, past and current members, for inspiring and motivating me through the work we did together and all the exciting conversations we had about science.

I would also like to thank my thesis committee for all the helpful advice and critical review of my dissertation, which improved my skills and ability to engage in effective scientific discourse as well as giving me the confidence to complete my Ph.D. I want to thank all the departmental and university administrators that made my time in the graduate program as seamless as possible by helping me resolve issues or completing applications that I had to deal with.

Finally, I would like to express my deepest gratitude and appreciation to all my family and friends that helped me get through the ups and downs of life during my Ph.D. A very special thanks to my wife, Pusparanee Hakim and our two daughters, Iris Sophia and Irene Nadya, who always brightened up my days with joy and made all the sacrifices and effort worth it. May God bless everyone and repay all the kindness that you have shown me.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

In multicellular eukaryotes, transposable elements (TE) make up a large part of the genomic content. These “selfish” genetic elements can propagate and expand the genome through their ability to replicate. However, this poses a significant risk to the stability of gene structure and overall genomic integrity which can lead to aberrant gene expression or products. To combat this threat, highly specific gene silencing mechanisms are utilized by the host to keep TEs in a constantly repressed state. In plants, transcriptional gene silencing is conducted through the RNA-directed DNA methylation (RdDM) pathway. The combined action of short interfering RNA (siRNA) and long non-coding RNA (lncRNA) direct deposition of DNA methylation at target TE regions. Subsequently, DNA methylation together with other repressive chromatin modifications turns the TE regions into a silenced state and prevent them from becoming active again. Although our understanding of the processes that lead up to the deposition of DNA methylation has been well-studied, the specific role and function of DNA methylation in gene silencing remains poorly understood. It is unclear how the presence of DNA methylation can affect the ability of transcription machinery from working at silenced regions. In contrast, DNA methylation does not hinder the transcriptional gene silencing machinery which suggests that DNA methylation plays a central role for distinguishing between different types of transcriptional activity in the DNA.

In the first story, we determined how DNA methylation interacts with nucleosomes in the context of transcriptional silencing. DNA compaction and packaging in the nucleus entirely revolves around its interaction with nucleosomes. This interaction has numerous implications for regulation of gene expression through changes in accessibility of DNA to factors involved in transcription. Here we show that RdDM can direct both DNA methylation and nucleosome positioning. Nucleosomes established by RdDM have no detectable impact on DNA methylation. Instead, DNA methylation affects nucleosome positioning. This applies not only to CHH methylation established by RdDM but also to DNA methylation in CG and CHG contexts, which is maintained by MET1 and CMT3. We propose a model where DNA methylation serves as one of the determinants of nucleosome positioning.

In the second story, we wanted to investigate the relationship between Pol V transcription and DNA methylation as a potential feedback mechanism where DNA methylation reinforces recruitment of Pol V transcription at silenced regions. Pol V transcribes in a pervasive manner throughout the genome implying that it does not require pre-existing chromatin marks such as DNA methylation to initiate transcription. However, previous studies have claimed that factors upstream of Pol V transcription that are able to bind to DNA methylation are required for the recruitment and initiation of Pol V transcription. Hence, the impact of DNA methylation on Pol V transcription remained unresolved. We found that loss of DNA methylation leads to a strong reduction of Pol V transcription. This occurs when DNA methylation is lost in all sequence contexts, which may happen not only in mutants defective in RdDM but also in mutants lacking maintenance DNA methyltransferases. Our results support a model where RdDM is maintained by a mutual reinforcement of DNA methylation and Pol V transcription with a strong crosstalk with other silencing pathways.

CHAPTER I

Introduction

1.0.1 Transposable Elements

Genetic information is encoded in deoxyribonucleic acid (DNA) where the order of nitrogenous bases determines the building blocks of life itself [1]. Genes form the basic unit of heredity in living organisms and the complete genetic sequence of an organism is known as a genome [2]. Since the advent of high-throughput DNA sequencing techniques, increasing amounts of information about genomes have been uncovered. Remarkably, only a small portion of a genome encodes for protein-coding genes while the majority are non-coding regions [3], [4]. For example, about 1% of the human genome is protein-coding while the rest of the genome is non-coding and sometimes perceived as non-essential to the point of being labeled as "junk DNA" [3], [4].

Some of the most prevalent components in eukaryotic genomes are transposable elements (TE). In multicellular eukaryotes they can comprise as much as 27% - 66% of the genome in mammals and up to 89% of the genome in plants [5]. Transposons were initially discovered in maize when Barbara McClintock observed that kernel color would be affected whenever a mobile DNA element, the Ac/Ds transposon, inserted itself into the gene responsible for pigmentation [6]. Transposons have been found to be present in every organism that has been studied from

Prokaryotes to Eukaryotes, making it a source for variation in genomes which helped shape the evolutionary process in every species [7].

Transposons can be categorized into two major classes, class I and class II transposons [8], [9]. Class I transposons are RNA transposons or retrotransposons which when transcribed will produced RNA transcripts that are reverse transcribed back into DNA and later reinserted into the genome in a random region. This method of transposition is described as the "copy-and-paste" method where the original retrotransposon does not move. Instead, new copies that are made from the original retrotransposon will be inserted into the genome, leading to a significant increase in the population of the retrotransposon [9], [10]. Within class I, retrotransposons are further divided into long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons [long interspersed elements (LINE) and short interspersed elements (SINE)].

Class II transposons are DNA transposons that can excise themselves out of the genome and reinsert into the genome again in a new yet random position. This method of transposition is described as the "cut-and-paste" method where the transposon does not typically make new copies of itself and moves around randomly in the genome [9], [10]. In class II, transposons are divided into terminal inverted repeat (TIR) transposons and Helitron/Rolling circle (RC) transposons. Both classes of transposons contain multiple families of transposons which are further categorized by their sequence structure and ability to transpose autonomously or by utilizing components from autonomous transposons [9], [10].

The impact of transposons on genomes have been associated with both positive and negative effects [11]. As genomes evolved through time with the presence of transposons, there have been numerous points where exaptation or domestication of wild transposons have occurred.

A common way that transposons affect gene expression is by insertion into genes which breaks the structure of the gene [12]. This can lead to null mutations and has been utilized as a tool to create mutants for genetic studies [13], [14]. Another way transposons affect gene expression is by rewiring the way genes are expressed. This can occur when a transposon inserts itself into regulatory elements such as enhancers and repressors. In maize, *Vgt1* is a conserved non-coding sequence (CNS) that regulates flowering time and is found approximately 70kb upstream of the AP2 gene, a transcription factor that functions as a negative regulator for flowering time [15]. Interestingly, insertion of a type of transposon known as miniature inverted repeat transposable element (MITE) in *Vgt1* has been strongly associated with flowering time [15]. In a different a study examining the potential origins of miRNA genes, *Made1* which is also a MITE, likely contributed to the unique structure of the miRNA gene *has-mir-548* in humans [16]. The structure of *Made1* contains inverted repeats with a palindromic feature and has been suggested to contribute to the formation of the unique miRNA stem-loop structure formed by *has-mir-548* [16].

Another example of transposon exaptation is the evolution of light-sensing in higher plants. To detect changes in the level of light in the environment, plants evolved to develop special photoreceptors [17]. The main photoreceptors in plants are known as phytochromes, where once they become photoactivated, leads to a cascade of events that are involved in light response [17]. In the example of phytochrome A (*phyA)*, the active form is imported from the cytoplasm into the nucleus where it activates transcription factors responsible for light-mediated responses [18]. Two proteins, FHY1 and FHL were shown to be required for *phyA* to accumulate in the nucleus and these proteins are regulated by two other transcription factors, FHY3 and FAR1 [18], [19]. Evolutionary studies on FHY3 and FAR1 revealed that these genes were likely derived from a type of DNA transposon known as *Mutator*-like elements (MULEs) [20]–[22]. Coding-sequence

exaptation of the transposase proteins from the transposons formed the binding sites used by the transcription factors to carry out their functions [20]. This also indicates that the domestication of the MULE transposon probably occurred near the rise of angiosperm evolution [19]. Hence, through numerous serendipitous events, transposons have contributed to evolutionary developments in organisms.

Transposition events also pose a risk for genomes as it could lead to disadvantageous mutations or significant aberrations to genomic structure and integrity. In the case of humans for example, this can lead to the development of genetic diseases and cancer [23]. A common way that rampant transposition events can occur is when the transposons become deregulated or reactivated within the cell and quickly spread throughout the genome. In one type of colorectal cancer, it was discovered that an active human LINE-1 (L1) retrotransposon had inserted itself into the tumor suppressor gene *APC*, leading to the disruption of its function [24]. Another study that looked at the prevalence of L1 insertions within human populations, suggested that active somatic L1 retrotransposons can lead to a higher risk of cancer initiation within affected individuals [25]. Insertions by transposons can also lead to the formation of cryptic splice site within genes. In a study examining retinoblastoma in a patient, they discovered L1 insertions in the *RB1* gene [21]. This lead to the formation of various aberrant transcripts that range from exon skipping to the introduction of new cryptic exons between exon 13 and exon 17 [26]. Despite the many examples of transposon insertions causing serious problems in the genome, these are rare events and most of the population do not suffer from it. This alludes to the fact that many organisms have evolved efficient silencing mechanisms that reduce or completely abolish the negative effects of transposable elements (TE) extant in the genome.

1.0.2 Gene Silencing

In the early 90s, Napoli *et al.* carried out experiments to enhance the colors of the petunia flower by introducing a chimeric *Chalcone synthetase (Chs)* gene into the plant genome [27]. Surprisingly, they observed the opposite effect where the color appeared to be faded or absent in certain parts of the flower which led them to propose the phenomenon as co-suppression [27], [28]. In an earlier study, it was demonstrated that introducing antisense RNA transcripts that are complementary to a gene in plants could lead to the inhibition of gene expression [29]. These initial findings instigated the notion that another layer of gene regulation potentially existed post-transcriptionally in living organisms [30].

Subsequent studies focused on the aspect of antisense RNA working as a suppression factor by binding to complementary mRNA transcripts [31]–[33]. However, a surprising finding by Fire *et al.* revealed the true mechanism of gene silencing which lead them to coin the term RNA interference (RNAi) [34]. In their work using *Caenorhabditis elegans*, they observed that the strongest down-regulation of the target RNA was by a negative control used in their experiments, a double-stranded RNA (dsRNA) that had both sense and antisense strands, while single strands of either sense or antisense had modest effects [34]. Further studies showed that RNA degradation is mediated by short interfering RNAs (siRNAs) which range from 21-23 nucleotides (nt) that are generated by the cleavage of long double-stranded RNA by ribonuclease III or otherwise known as *Dicer* [35]–[37]. The siRNAs are subsequently taken up by another nuclease known as *Argonaute* (*Ago*), which forms the RNA-induced silencing complex (RISC) that finds complementary mRNA targets to degrade [38].

With the elucidation of the RNAi mechanism, the function and roles of other types of small RNAs became better understood as well. An example of another type of small RNA that is involved in RNAi are microRNAs [39], [40]. Early work on a gene known as *lin-4* in *C. elegans*

showed that it was strongly associated with developmental timing defects which lead to severe phenotypes in the morphology of the animal [41], [42]. Unexpectedly, it was later discovered that *lin-4* did not code for any protein and seemed to exist as a noncoding RNA, which indicated that the main mechanism may come from a RNA-based interaction [43]. When the suppressor mutation to *lin-4* was found, known as *lin-14*, researchers quickly connected that the *lin-4* RNA strand had a strong complementarity to the 3' untranslated region (UTR) of the *lin-14* mRNA strand [44]. In the same study, they demonstrated that negative regulation occurred at the posttranscriptional level where LIN-14 protein levels were drastically reduced while RNA levels remained fairly similar between wildtype and *lin-4* mutant [44]. These foundational studies established miRNA as a subset of RNAi mechanisms that that are involved in gene regulation.

A major difference between siRNA and miRNAs is that in most cases for miRNAs, the target mRNA is not sliced by the RISC complex [45]. Instead, miRNAs promote the repression of translational machinery as well as the removal of the polyadenylated tail (poly-A) from the mRNA, which later leads to the degradation of the RNA transcript by exonucleases [46]–[48]. This difference in mechanism has been suggested to be due to the level of complementarity between miRNA and its target sequence [49], [50]. In animals, miRNAs have a seed sequence which provides most of the complementarity and flanking sequences that partially match to their target sequence [51], [52]. However, for AGOs to carry out slicing, it has been shown that high complementarity is needed where the active site of the binding pocket can catalyze the cutting activity [53]–[55]. In plants, however, miRNA functions similarly to siRNA because most plant miRNAs target coding regions and have high complementarity to their targets [56], [57].

Another variable aspect in posttranscriptional silencing are the biological roles that they serve in certain organisms. RNAi carried out by siRNA has been shown to be driven by an

immunological purpose where foreign RNA that originate from sources such as viruses and TEs, are identified by the structure of a dsRNA and processed by DICER and AGO for immediate degradation by the RISC machinery [58]–[60]. This defense mechanism has been shown to be conserved in both plants and animals to prevent unwanted viral infections and aberrations caused by TEs [59]. siRNAs have also been artificially synthesized for the purposes of research and potential therapies [61]. miRNAs are encoded in the genome and are involved in regulation of endogenous genes. miRNAs have been shown to be critical factors that help coordinate changes in gene regulation of development and differentiation of cells and tissues in many organisms [39]. However, in the case of TEs, they are usually integrated into the host genome and will be constantly inherited from generation to generation. Therefore, due to the immense importance of genome defense, it is performed by multiple parallel pathways at the posttranscriptional and chromatin level.

1.0.3 Chromatin modifications

DNA strands are tightly packaged inside the nuclei by wrapping around spool-like structures known as nucleosomes [62]. Nucleosomes are formed from a histone octamer structure composed of 2 copies each of histone H2A, H2B, H3, and H4 [63]. This structure which is composed of DNA and proteins form the molecule known as chromatin, a term initially coined by Walther Flemming in 1879 due to its ability to take up aniline dyes [64], [65]. Our understanding of chromatin has seen an astronomical leap since then due to improved biochemical techniques as well as the invention of high throughput sequencing technology which have uncovered its highly complex characteristic and biological functions that extend beyond packaging [66]–[69]. A major feature that has been a focus of study are the chemical modifications that can exist in the chromatin that play crucial roles in chromatin accessibility, gene regulation and chromatin looping events [69],

[70]. These modifications can be present on the DNA strand or on histones, specifically posttranslational modifications on the tail which extrudes from the nucleosome structure [71], [72].

There have been a large and diverse number of histone tail modifications identified in eukaryotes and it used to be proposed that the information stored in these modifications could be recognized as the histone code [73]. The most commonly studied posttranslational modifications are the addition of either methylation or acetylation on specific amino acids present on the histone tail of H3 and H4 [72]. These histone modifications are placed by proteins known as writers such as histone methyltranferases or removed by proteins known as erasers such as histone demethylases. Furthermore, each of these modification is strongly correlated with either a repressive or an active chromatin state in regards to gene expression [72]. For example, histone methylation such as H3K9me2 or H3K27me3 are associated with gene silencing and usually present in pericentromeric and heterochromatic regions in the chromosome [74]–[76]. Alternatively, histone acetylation such as H4K16Ac and H3K23Ac are strongly correlated with active gene expression and highly enriched at transcription start sites (TSS) of genes [77]. The key to regulation by histone modifications are the histone modification readers that are able to affect accessibility for transcription or actively recruit transcription factors [70]. In addition, histone modifications have also been shown to be involved in DNA methylation pathways which adds to the complexity of chromatin regulation [78].

DNA modifications were identified soon after the discovery of its role as the genetic material when methylated cytosine (5-meC) was found in calf thymus DNA [79]. Initially, there was not an immediately obvious biological function to this modification but the level of ubiquity in the genome suggested that it had the potential to have one [80], [81]. The addition of DNA methylation is catalyzed by writers which are DNA methyltransferases and can be actively

removed by DNA demethylases through a process involving glycosylation or passively after multiple rounds of DNA replication [82, p. 2], [83], [84]. Later studies started to report a strong correlation between cytosine DNA methylation and gene silencing in mammals and plants [85], [86]. In addition, DNA methylation in plants serves as an epigenetic mark as it could be passed on from generation to generation [87]. In mammals however, the genome undergoes a global demethylation step in germ cells which suggests that DNA methylation may not be a heritable chromatin mark in this case [88]. Besides writers and erasers for DNA methylation and histone modifications, there are many readers of these marks which process the information in different ways to elicit a specific effect to the chromatin.

Proteins that can recognize and bind to DNA methylation through protein binding domains such as SET- AND RING-ASSOCIATED (SRA) domains [89], [90]. In plants, for example, SUVH4, SUVH5 and SUVH6 which have SRA domains, can bind DNA methylation which leads to the placement of H3K9me2 within the same region [91], [92]. There are also readers for histone modifications such as CMT3 who possess BAH and chromo domains, allowing it to bind to H3K9me2 marks leading to the deposition of DNA methylation, forming a positive feedback loop for the pathway [91], [93]. This crosstalk between different chromatin modifications facilitate the formation of stable heterochromatin, a repressed chromatin state, particularly in TEs which will be passed on trans generationally [78]. However, for *de novo* silencing to occur on newly inserted TEs in the genome, another pathway known as transcriptional gene silencing performs this role at the chromatin level.

1.0.4 Transcriptional Gene Silencing

A study by Verdel *et al.* reported a variant of the RNAi mechanism in *Schizosaccaromyces pombe* that could lead to the formation of heterochromatin at regions targeted for silencing [94]. Utilizing similar components of the PTGS pathway such as DICER and AGOs, siRNAs generated from the pathway could target specific genomic regions and direct the RNA-induced transcriptional silencing (RITS) machinery to silence them through chromatin modification like histone H3 lysine 9 methylation (H3K9me) [94]. This mechanism was described as transcriptional gene silencing (TGS) and broadened the repertoire of pathways driven by small RNAs involved in gene regulation.

The first evidence for RNA affecting the chromatin state was shown in plants where potato spindle viroid genes that were artificially introduced into the genome were quickly silenced and marked with a DNA modification known as DNA methylation [95]. In another study by Mette *et al.*, they carried out an experiment to determine if expressing the promoter of a gene would lead to the silencing of an unlinked gene that shared sequence homology [96]. In one of their transgenic plants, an unexpected error led to the formation of aberrant transcripts that had inverted repeats, allowing a stem-loop structure to form [96]. This lead to the deposition of DNA methylation at the transgene and the homologous gene causing silencing, which suggested that a dsRNA intermediate structure was crucial for the process [96].

In *Drosophila*, introduction of multiple tandem copies of transgenes was shown to cause the silencing of the transgenes as well as the endogenous gene [97]. Interestingly, the silencing effect was shown to require the Polycomb Complex, which is known to be involved the formation of heterochromatin and regulation of homeotic genes [98]. Further studies showed that repetitive sequences in fly and mouse germline cells can induce silencing through the Piwi protein which associates with piwi-interacting RNAs (piRNAs) [99]–[101]. In addition, piRNAs have been

proposed to function as a defense mechanism against aberrant TE activity and has been shown to function upstream of DNA methyltransferases at transposons targeted for silencing via DNA methylation [102], [103].

However, the best studied TGS pathway in terms of mechanistic insights has been in plants due to its genetic robustness where key mutants in silencing pathways remain viable in comparison to other model organisms [104]. Current models of TGS pathways involve the presence of long noncoding RNA (lncRNA) which is produced by DNA-dependent RNA polymerase II (Pol II) in many organisms [94], [105]. Plants evolved to have two specialized RNA polymerases, DNA-dependent RNA polymerase IV (Pol IV) and DNA-dependent RNA polymerase V (Pol V) which specifically produce noncoding RNA in TGS [106], [107]. In addition, null mutations in these specialized polymerases do not affect the viability of the organism unlike Pol II which typically leads to severe defects or even lethality [108]–[111]. This gives plants a significant advantage as a genetic model for TGS and allows interrogation of the pathway at a genome-wide scale.

1.0.5 RNA-directed DNA methylation

In plants, TGS is established through a pathway known as the RNA-directed DNA Methylation (RdDM) pathway [104]. It mostly targets TEs and repetitive elements for silencing through the placement of repressive chromatin modifications such as DNA methylation, H3K9me2 and nucleosome positioning [Figure 1.1] [104], [112], [113]. This is thought to help prevent TE mobility from occurring which could harm the integrity of the genome and lead to aberrations in gene products or expression levels. DNA methylation in plants can occur in three different sequence contexts, CG, CHG, and CHH (H = A/T/G) [87]. In addition, there are three major pathways that are responsible for DNA methylation in plants which are mainly distinguished by

methylation sequence context and whether it is involved in maintenance of DNA methylation or initiating *de novo* DNA methylation [87], [112], [114]. In the first pathway, symmetrical CG methylation is maintained through the action of DNA METHYLTRANSFERASE 1 (MET1) during DNA replication, where in coordination with VARIANT IN METHYLATION (VIM1), immediately deposits DNA methylation onto the newly synthesized DNA strand, preserving the epigenetic mark between generations [115], [116].

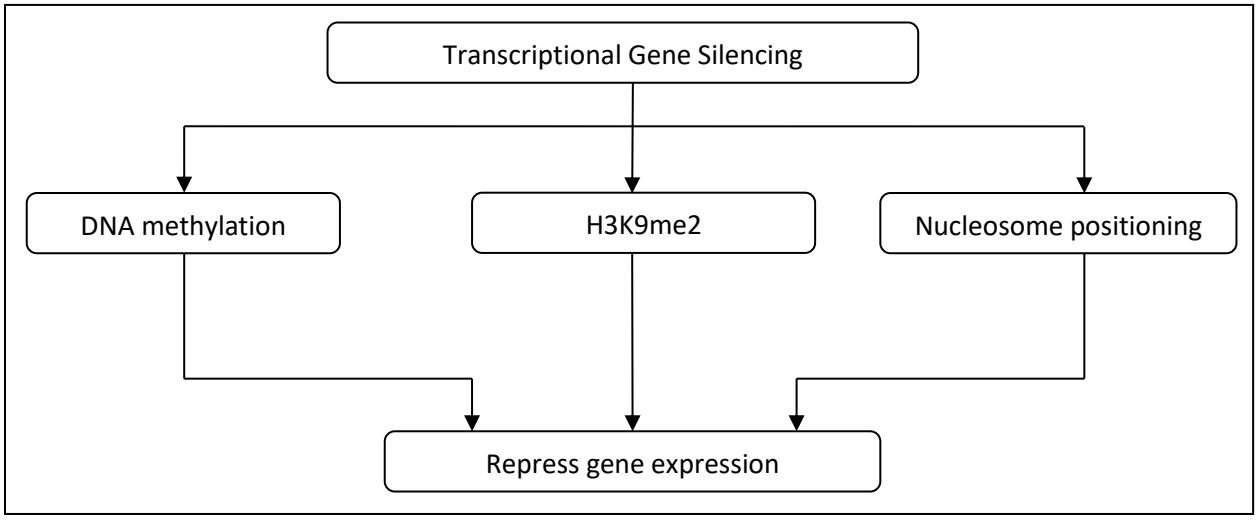

Figure 1.1: **Simplified workflow for transcriptional gene silencing mechanism in plants.**

In the second pathway, symmetrical CHG methylation is maintained by a different DNA methyltransferase known as CHROMOMETHYLASE 3 (CMT3). CMT3 possesses a H3K9me2 binding domain and will bind to regions that are enriched with H3K9me2 such as heterochromatic and centromeric regions [91]. When CMT is bound to the H3K9me2 marks and is within proximity to the DNA, it will deposit DNA methylation within the bound region [91]. Additionally, another DNA methyltransferase protein known as CHROMOMETHYLASE 2 (CMT2) can also bind to H3K9me2 marks, however, it will deposit DNA methylation asymmetrically in the CHH sequence context [112], [117]. The CMT pathway has been shown to be closely associated with the activity

of histone methyltransferases such as SU(VAR)3-9 HOMOLOGUES 4 (SUVH4), SU(VAR)3-9 HOMOLOGUES 5 (SUVH5), and SU(VAR)3-9 HOMOLOGUES 6 (SUVH6) [92], [118], [119]. These histone methyltransferases possess SRA domains which can bind to DNA methylation marks [92], [93], [119], [120]. This leads to a positive feedback loop situation where once either DNA methylation or H3K9me2 marks are present in a region, the actions of CMT proteins and SUVH proteins will reinforce each other until the regions turns into highly repressive state [78].

The final pathway is RdDM which methylates DNA in any sequence context (CG/CHG/CHH) in asymmetrical fashion due to its ability to carry out *de novo* methylation [117]. It is thought to be the pathway that can target new TE insertions for silencing and involved in most of the transgene silencing events observed previously [121]–[123]. RdDM occurs in two major steps, siRNA biogenesis and *de novo* DNA methylation [104], [124]. In the first step, Pol IV is recruited to target regions through SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1) which can bind to H3K9me2 marks [83], [84], [85]. Next, POL IV will transcribe a noncoding RNA that is subsequently converted into a dsRNA by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) [107], [127]. The dsRNA is then taken up by DICER-LIKE3 (DCL3) which cleaves into 24 nt siRNA [127], [128]. These siRNA are modified at the 3’ ends by HUA ENCHANCER 1 (HEN1), subsequently loaded into ARGONAUTE 4 (AGO4) and transported back into the nucleus [129]–[131].

From this point, the pathway moves into the second step. At the same target regions, Pol V will transcribe lncRNAs with the help of upstream factors DEFECTIVE IN RNA-DIRECTED DNA METHYLATION (DRD1), DEFECTIVE IN MERISTEM SILENCING 3 (DMS3) and RNA-DIRECTED DNA METHYLATION 1 (RDM1) [132], [133]. Pol V transcripts are then believed to act as a scaffold for other downstream factors to bind for *de novo* DNA methylation

process. Pol V and SUPRESSOR OF TY INSERTION 5 - LIKE (SPT5L) / KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (KTF1), a transcription elongation factor, will recruit AGO4 to the regions using their AGO-hook domain and subsequently the lncRNA is thought to base-pair with siRNA [111], [134]–[136]. Next, IDN2, an RNA binding protein, will interact with Pol V transcript and finally DRM2, a *de novo* DNA methyltransferase will be recruited to deposit DNA methylation marks to the target sites which are recognized as RdDM regions [101], [102], [103].

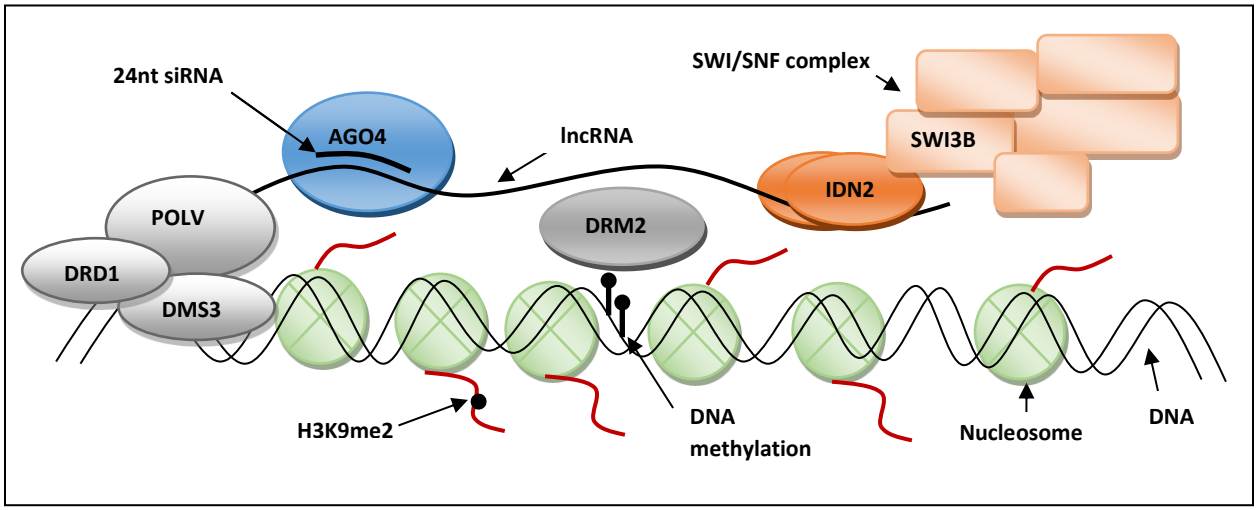

Figure 1.2: **Current model for the POLV-mediated *de novo* DNA methylation pathway.**

Currently, most studies in the *de novo* DNA methylation step have focused on the mechanism by which the pathway deposits DNA methylation at regions targeted for silencing. A major gap still exists in our understanding of how silencing is achieved once DNA methylation has been placed. Genomic regions where gene expression is strongly repressed are characterized by the presence of repressive chromatin marks and compact packaging of DNA that causes the region to become inaccessible to transcription factors [139]. However, in the context of RdDM, this becomes more complicated as most of the regions that it targets are not located in the

pericentromeric or large heterochromatic domains. Instead, RdDM factors such as Pol V and AGO4 are strongly enriched in TEs that are present in the intergenic regions in euchromatic domains [133], [135]. This observation implies that RdDM must use a mechanism that allows it to silence regions in a precise manner without causing inadvertent silencing of adjacent genes.

In addition to DNA methylation, other types of chromatin modifications have been reported to be strongly associated with RdDM regions. Histone methyltransferases such as SUVH4 possess SRA domains that allow them to bind to methylated DNA and RdDM regions are typically enriched for H3K9me2 marks [78], [135]. Recent work has also shown that SILENZIO, a molecular chaperone J-domain protein is recruited to methylated regions through its interaction with MBD5 and MBD6, a class of methyl readers, and acts as transcriptional repressor [140]. These findings suggests that DNA methylation serves as a point of reference for repressive factors to localize and exert their effect once the regions are marked with DNA methylation. Another type of chromatin modification that was previously reported is nucleosome positioning [113], [141]. This finding came from the discovery that IDN2 was able to interact with a subunit of the SWI/SNF chromatin remodeling complex called SWI3B [113]. In the same study, it was shown that SWI/SNF had a strong correlation with gene silencing effects in RdDM regions which suggested that the pathway may be using nucleosome positioning as a silencing mechanism [113].

Nucleosomes present a physical barrier to transcription factors from accessing sequence motifs that are important for activation of gene expression [142]. Work by Chodavarapu *et al*. found that at the genome-wide level, DNA methylation is enriched within the nucleosomal region [143]. This finding implies that DNA methylation is strongly correlated with the presence of nucleosomes and that there could be a potential function for this relationship. In another study using a genome-wide approach in mammalian cells, they showed that highly methylated CpG

regions are enriched with nucleosomes adding more evidence for a potential relationship between nucleosomes and DNA methylation [144]. However, *in vitro* studies looking at the biochemical effect of DNA methylation in relation to nucleosome binding demonstrated that the DNA modification led to reduced flexibility of the DNA strand making it less favorable to binding nucleosomes [145]. Indeed, work by Lyons and Zilberman looking at the mechanistic relationship between DNA methylation and nucleosome reported that DNA methylation is generally enriched in linker regions between nucleosomes [146]. They determine that nucleosome bound DNA is inaccessible to DNA methyltransferases and that chromatin remodelers such as DECREASED DNA METHYLATION 1 (DDM1) and Lsh are required for DNA methylation to be enriched within nucleosomal regions [146]. These findings show that in *in vivo* conditions, through the action of chromatin remodelers, conditions which make DNA methylation unfavorable for nucleosome binding *in vitro* can be overcome and cause nucleosomal DNA to be enriched with DNA methylation. In addition, this provides a potential clue for the presence of chromatin remodelers downstream of Pol V transcription which might be to facilitate DNA methylation deposition.

In the current model for RdDM, the exact process for nucleosome positioning to occur downstream of Pol V transcription is still poorly understood [Figure 1.2]. Previous work did not conclusively determine how DNA methylation and nucleosome positioning function in relation to each other to establish gene silencing [113]. Hence, this opens two possible scenarios for the presence of the SWI/SNF chromatin remodeling complex in RdDM. Either it is needed to facilitate DNA methylation deposition by remodeling the chromatin to provide DRM2 access to nucleosomal DNA, which implies an indirect role in terms of gene silencing, or that the deposition of DNA methylation via RdDM affects where nucleosomes are positioned by SWI/SNF to

potentially inhibit active transcription by Pol II, which suggests an active role for nucleosome positioning in relation to gene silencing. In Chapter II, we investigated the relationship between DNA methylation and nucleosome positioning in the context of transcriptional gene silencing.

Another aspect of the *de novo* DNA methylation step that has not been fully explored are the events occurring upstream of Pol V transcription. We still do not fully understand how Pol V initiates transcription, especially in the scenario where a new transposon has inserted into the genome [121]. Recent work has shown that Pol V is able to transcribe pervasively throughout the genome as a surveillance mechanism [147]. This provides some insight that Pol V can transcribe in most regions but does not explain how Pol V is able to initiate transcription in a highly non-specific manner. Upstream factors required for Pol V transcription have been identified such as the DDR complex, but these factors do not provide any mechanistic insight besides opening up the DNA ahead of Pol V to facilitate transcription [132].

In a study by Johnson *et al.* which focused on factors upstream of Pol V transcription, they identified SU(VAR)3-9 HOMOLOGUES 2 (SUVH2) and SU(VAR)3-9 HOMOLOGUES 9 (SUVH9), two catalytically dead histone methyltransferase with SRA domains that are required for Pol V transcription [90]. This finding led them to propose that Pol V can be recruited to target regions by pre-existing DNA methylation marks and feed back into the pathway to reinforce silencing [90]. An important implication of the positive feedback hypothesis is that it predicts that Pol V transcript levels would be substantially reduced in mutants downstream of Pol V transcription. Instead, other studies that tested Pol V transcription levels in mutants downstream of Pol V do not observe a reduction in these mutants which contradicts the positive feedback mechanism implied by the function of SUVH2 and SUVH9 [111], [136], [148], [149].

However, findings from the Pol V surveillance study indicates that Pol V transcribes regions in varying levels from very low at the surveillance stage to high transcription at established RdDM regions [147]. This suggests that Pol V can transition from surveillance to maintenance in a yet to be identified mechanism and that the positive feedback hypothesis may still be plausible at the maintenance stage where Pol V transcription is higher. Hence, the mechanism for the maintenance stage needs to be better understood and the potential positive feedback mechanism conclusively established to gain more insight into the dynamic nature of Pol V transcription initiation. In Chapter III, we investigated the role of DNA methylation in reinforcing noncoding RNA transcription through a positive feedback mechanism to establish maintenance of gene silencing.

CHAPTER II

DNA Methylation Affects Nucleosome Positioning in Transcriptional Silencing

The contents of this chapter have been submitted for publication and is currently in the review process. Shriya Sethuraman conducted data analysis and generated plots shown in Figure 2.5B, 2.5C, 2.6C, 2.6D, 2.8A, 2.8B, 2.9 and 2.10. Jakob Dolata prepared MNase H3-ChIP and whole genome bisulfite samples for sequencing. Alan Boyle provided guidance and consultation in bioinformatic analysis. I performed all other experiments and data analysis shown in this chapter.

2.1 Abstract

Repressive chromatin modifications are instrumental in regulation of gene expression and transposon silencing. In *Arabidopsis thaliana*, transcriptional silencing is performed by the RNA-directed DNA methylation (RdDM) pathway. In this process, two specialized RNA polymerases, Pol IV and Pol V, produce non-coding RNAs, which recruit several RNA-binding proteins and lead to the establishment of repressive chromatin marks. An important feature of chromatin is nucleosome positioning, which has also been implicated in RdDM. We show that RdDM is able to direct both DNA methylation and nucleosome positioning. Nucleosomes established by RdDM have no detectable impact on DNA methylation. Instead, DNA methylation affects nucleosome

positioning. This applies not only to CHH methylation established by RdDM but also to DNA methylation in CG and CHG contexts, which is maintained by MET1 and CMT3. We propose a model where DNA methylation serves as one of the determinants of nucleosome positioning.

2.2 Introduction

Transcriptional gene silencing (TGS) pathways play an important role in maintaining genomic integrity in eukaryotes. They rely on repressive chromatin modifications, which are specifically targeted to silence transposable elements (TE) present in the genome. TGS pathways are conserved in fungi, animal and plant kingdoms, denoting their importance in the control of genome stability [78]. In plants, TGS is established and partially maintained through RNA-directed DNA methylation (RdDM), which consists of two major steps, biogenesis of short interfering RNA (siRNA) and *de novo* DNA methylation [104].

In the first step, RNA polymerase IV (Pol IV) binds to loci targeted for silencing and produces noncoding RNA, which is then converted into a double-stranded form (dsRNA) by RNA-dependent RNA polymerase 2 (RDR2) and cleaved into 24-nucleotide siRNA by DICER-LIKE 3 (DCL3) [96], [126]–[128], [150]. siRNAs are then incorporated into ARGONAUTE 4 (AGO4) and other related AGOs, forming AGO-siRNA complexes [131]. In the second step, RNA polymerase V (Pol V) produces long noncoding RNA (lncRNA) that acts as a scaffold or otherwise helps recruit downstream effectors [111], [136], [151]. The AGO4-siRNA complex is recruited to Pol V-transcribed loci leading to stepwise binding of INVOLVED IN *DE NOVO* 2 (IDN2) and DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) which deposits DNA methylation [82], [136], [137], [148], [152]. However, the mechanisms by which DNA methylation and other repressive features of chromatin contribute to transcriptional gene silencing are not fully understood.

RNA-directed DNA methylation is functionally intertwined with nucleosome modifications and positioning. This includes the involvement of pre-existing histone modifications and putative chromatin remodelers in recruitment of both Pol IV and Pol V [90], [126], as well as the establishment of repressive histone modifications and nucleosome positioning in the second step of RdDM [111], [113], [117], [141], [153]. The involvement of active chromatin remodeling in transcriptional silencing by RdDM was suggested by an interaction of IDN2 with SWITCH 3B (SWI3B), a subunit of the Switch/Sucrose Non Fermenting (SWI/SNF) chromatin remodeling complex [113]. Subunits of this complex also interact with other silencing factors, including HISTONE DEACETYLASE 6 (HDA6) and MICRORCHIDIA 6 (MORC6), which indicates that SWI/SNF may be involved in various aspects of gene silencing [141], [153, p. 6]. This is consistent with SWI/SNF being multifunctional and affecting not only gene silencing but also various other aspects of plant gene regulation [154]–[160].

There are several indications that nucleosome positioning and DNA methylation are somehow connected throughout plant genomes [143], [146], [161]. However, the exact nature of this connection varies depending on species and genomic regions tested [146], [161]. In Arabidopsis, nucleosomes determined by MNase digestion protections have been reported to generally correlate with DNA methylation [143]. However, the opposite correlation exists on a subset of Arabidopsis nucleosomes and throughout genomes of certain other species [146], [161]. This difference may be explained by the DNA binding of linker histones, which prevent methylation of linker DNA, and by the activity of DDM1, which facilitates methylation of nucleosomal DNA [146], [162]. In Arabidopsis these two proteins counteract the general preference to methylate linker DNA [146].

The involvement of linker histones, DDM1 and SWI/SNF in determining the pattern of DNA methylation indicates that the observed connection between nucleosomes and DNA methylation is primarily determined by nucleosomes being inaccessible to DNA methyltransferases. This is supported by *in vitro* data indicating preferential methylation of linker DNA [163]. However, the opposite relationship has been observed on a few individual loci, where nucleosomes were affected by the *drm2* mutation [113]. This indicates that DNA methylation may affect nucleosome positioning. This alternative causality is also supported by some *in vitro* data [164]. Therefore, the relationship between nucleosomes and DNA methylation remains only partially resolved.

Here, we explore the mechanism by which RdDM affects nucleosome positioning in *Arabidopsis thaliana*. We demonstrate that Pol V and more broadly RdDM establish both DNA methylation and nucleosome positioning. The SWI/SNF complex is not required for DNA methylation on positioned nucleosomes. Instead, DNA methylation is needed for nucleosome positioning on differentially methylated regions. We propose a model where DNA methylation serves as one of the determinants of nucleosome positioning.

2.3 Results

2.3.1 Pol V affects a subset of nucleosomes

Pol V has been previously shown to affect protection to MNase digestion of certain genomic regions [113]. To conclusively attribute these protections to nucleosome positioning, we expanded this experiment by combining MNase digestion with immunoprecipitation using an anti-H3 antibody (MNase H3 ChIP-seq) in two biological replicates of Col-0 wildtype and *nrpe1*, a mutant of the largest subunit of Pol V [Figure 2.1A]. We identified 690 nucleosomes stabilized by Pol V, where signal was at least 2-fold higher in Col-0 compared to *nrpe1* with a false discovery rate

(FDR) of less than 0.05 [Figure 2.1B]. We also identified 3082 Pol V destabilized nucleosomes, where signal was at least 2-fold higher in *nrpe1* compared to Col-0 with an FDR of less than 0.05 [Figure 2.2A].

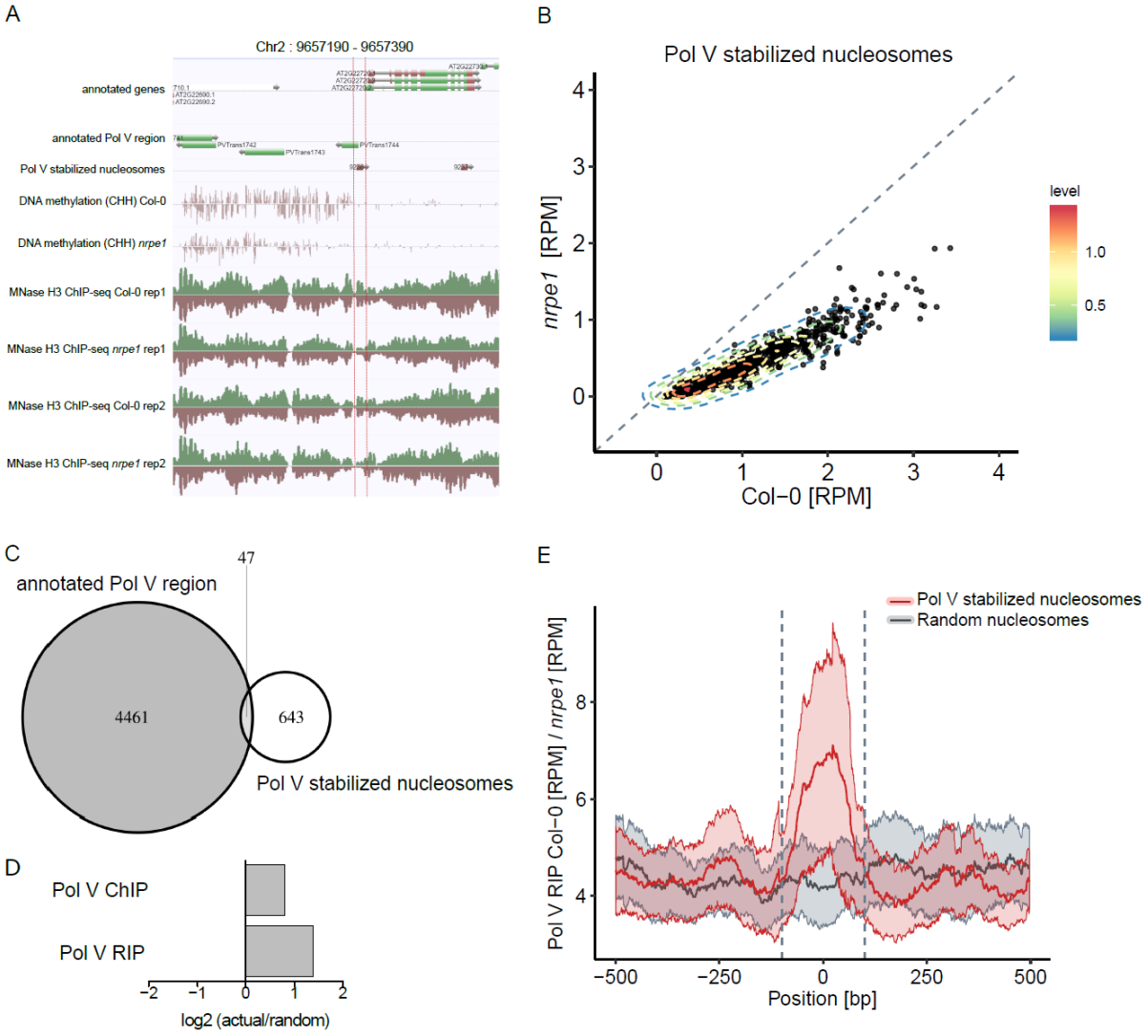

Figure 2.1: **Pol V affects a subset of nucleosomes:** A. Genome browser screenshot showing a Pol V stabilized nucleosome (indicated by red vertical lines). Shown data include TAIR10 genome annotations, previously identified annotated Pol V transcripts [165], Pol V stabilized nucleosomes, CHH methylation [112] and MNase H3 ChIP datasets from two biological replicates. B. Comparison of MNase H3 ChIP-seq signal in Col-0 and *nrpe1* on Pol V stabilized nucleosomes.

Average signal levels from two biological replicates are shown. C. Overlap between Pol V stabilized nucleosomes and previously published annotated Pol V transcribed regions [165]. D. Enrichment of Pol V stabilized nucleosomes on previously published Pol V bound regions identified using ChIP [166] and Pol V transcribed regions identified using RIP [165]. $p < 0.001$, random permutation test, 1000 iterations. E. Pol V RNA immunoprecipitation signal [165] on Pol V stabilized nucleosomes and random nucleosomes. The nucleosomal regions are indicated with vertical dashed lines. Ribbons indicate bootstrap confidence intervals with $p < 0.05$.

To test if Pol V stabilized and destabilized nucleosomes are located within genomic regions with high levels of Pol V transcription, we overlapped identified nucleosomes with previously published Pol V-transcribed regions [165] that represent RdDM Pol V transcription [167]. Pol V stabilized nucleosomes showed a limited overlap with annotated Pol V-transcribed regions [Figure 2.1C], which was still significantly more than expected by chance [Figure 2.1D]. Consistently, the average level of Pol V transcription on Pol V stabilized nucleosomes was strongly enriched compared to adjacent regions or random sequences [Figure 2.1E]. Furthermore, like Pol V transcription, Pol V stabilized nucleosomes are enriched in intergenic and promoter regions [Figure 2.2B-C]. On the other hand, overlaps between Pol V destabilized nucleosomes and annotated Pol V-transcribed regions were less likely than expected by chance [Figure 2.2D-E]. This indicates that Pol V stabilized nucleosomes may be present within regions subject to high levels of RdDM Pol V transcription but also within regions with low Pol V transcription. We conclude that previously demonstrated impact of Pol V on MNase protections [113] may be conclusively attributed to Pol V affecting a pool of nucleosomes.

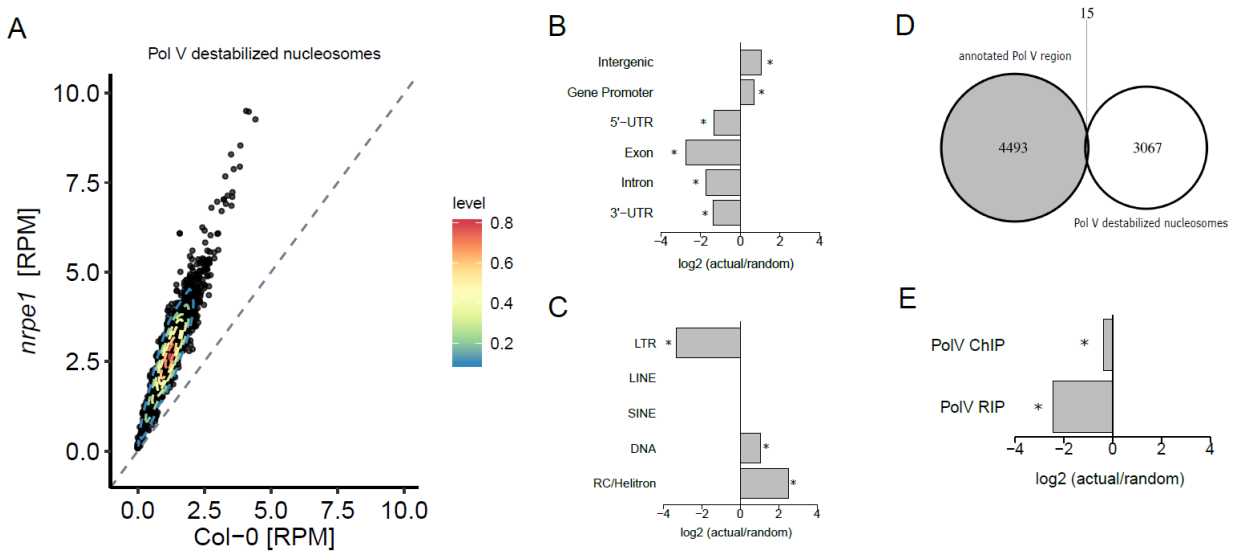

Figure 2.2: **Pol V affects a subset of nucleosomes.** A. MNase H3-ChIP seq signal on Pol V destabilized nucleosomes. B. Enrichment of Pol V stabilized nucleosomes on various genomic regions (random permutation test; 1000 iterations; stars indicate p-values < 0.001). C. Enrichment of Pol V stabilized nucleosomes on annotated transposable element regions (random permutation test; 1000 iterations; stars indicate p-value < 0.001). D. Overlap between Pol V destabilized nucleosomes and annotated Pol V transcribed regions. E. Enrichment of Pol V destabilized nucleosomes on annotated Pol V transcribed or bound regions (random permutation test; 1000 iterations; stars indicate p-value < 0.001).

2.3.2 AGO4 and IDN2 affect nucleosome positioning

Impact of Pol V on nucleosomes may be explained by the interaction between a lncRNA-binding protein IDN2 and a subunit of the SWI/SNF chromatin remodeling complex SWI3B [113]. Because IDN2 is recruited to Pol V transcripts downstream of AGO4 [148], this predicts that events occurring downstream of Pol V transcription should also affect nucleosome positioning. To test this prediction, we performed MNase-H3 ChIP followed by qPCR in Col-0 wild type, *nrpe1*, *ago4-1* and *idn2-1* mutants. We detected a substantial decrease of the nucleosome signals in all three tested mutants compared to wildtype at Pol V stabilized nucleosomes [Figure 2.3A-G]. While

nrpe1, as expected, strongly affected all tested nucleosomes, *ago4* and *idn2* had more variable effects [Figure 2.3A-G]. These findings indicate that AGO4 and IDN2 both contribute to Pol V-mediated nucleosome positioning. This suggests that events occurring downstream of Pol V transcription are involved in Pol V-mediated nucleosome positioning.

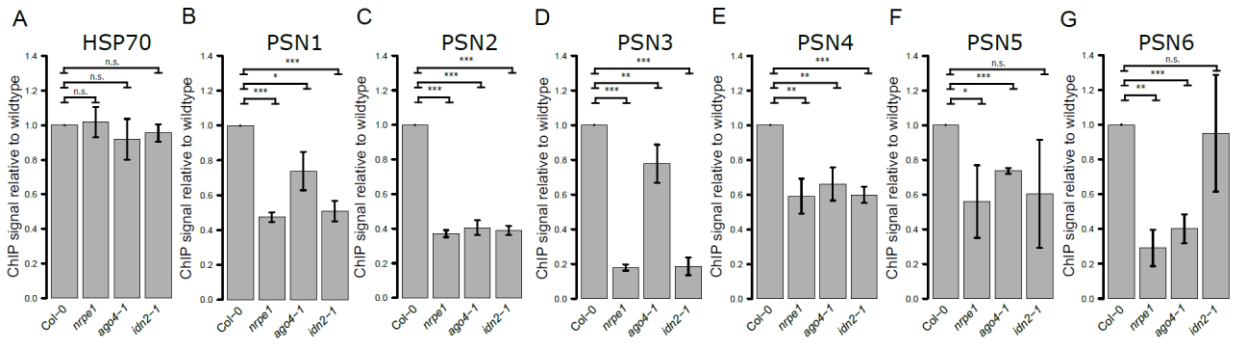

Figure 2.3: **AGO4 and IDN2 affect nucleosome positioning:** A.–G. Locus-specific analysis of MNase H3-ChIP qPCR levels on Pol V stabilized nucleosomes in Col-0, *nrpe1*, *ago4-1* and *idn2-1*. Significance tested using t-test (n.s. = not significant, ** = p-value < 0.01,*** = p-value < 0.001). ChIP signal values were normalized to *ACTIN2* and Col-0 wild-type. Error bars show standard deviations from three biological replicates.

2.3.3 Pol V stabilized nucleosomes are enriched in DNA methylation

The impact of the RdDM pathway on nucleosome positioning suggests that RdDM may establish both DNA methylation and nucleosome positioning on a pool of the same loci. To test this prediction, we determined genome-wide levels of DNA methylation using whole-genome bisulfite sequencing in Col-0 wildtype and *nrpe1* in two biological replicates. We plotted DNA methylation in the CHH context at Pol V stabilized nucleosomes and 500 bp adjacent regions [Figure 2.4A]. CHH DNA methylation was significantly enriched on Pol V stabilized nucleosomes compared to both the adjacent regions and the *nrpe1* mutant [Figure 2.4A]. To test if this enrichment is also

dependent on AGO4 and IDN2, we used previously published whole-genome bisulfite sequencing datasets [112]. Likewise, we detected a reduction in the average DNA methylation levels in both *ago4-1* and *idn2-1* [Figure 2.5A]. These findings indicate that at a subset of its targets, Pol V affects nucleosome positioning in parallel with establishing DNA methylation and therefore, nucleosome positioning is linked to RdDM.

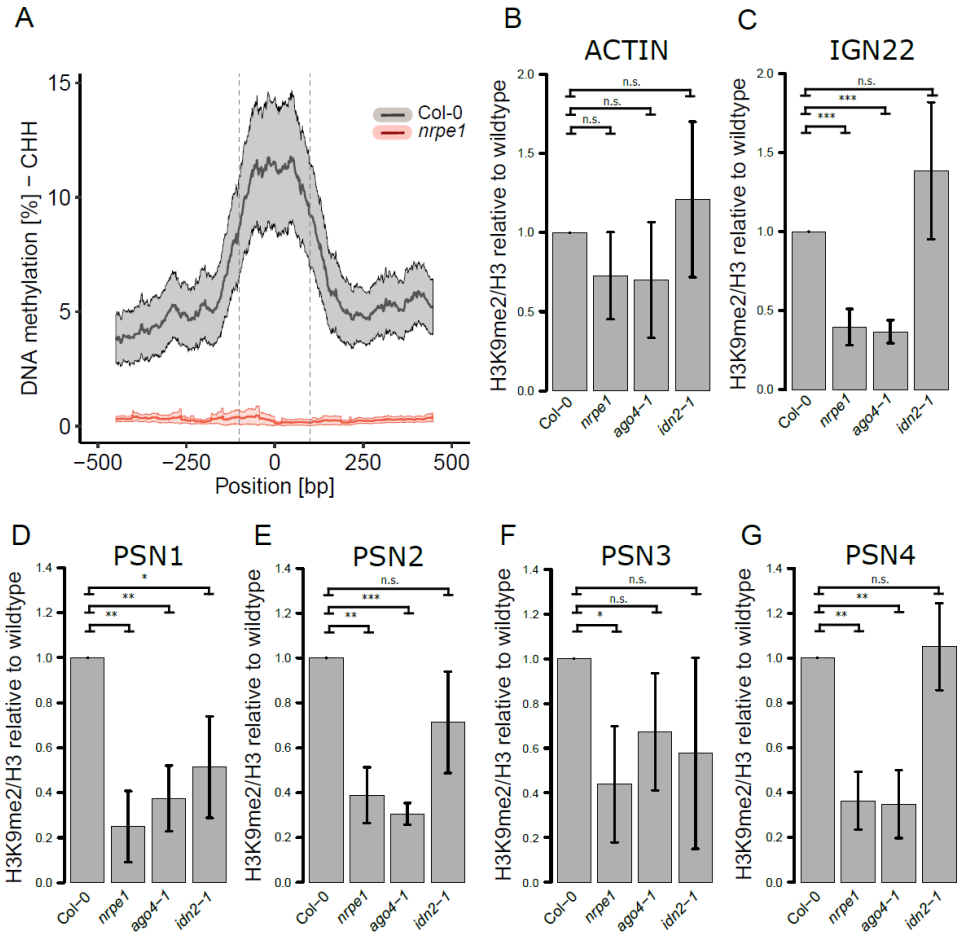

Figure 2.4: **Pol V stabilized nucleosomes are enriched with DNA methylation:** A. Average profile of DNA methylation levels (CHH context) on Pol V stabilized nucleosomes. B–G. Locus-specific analysis of H3K9me2 levels on *ACTIN2*, *IGN22* and Pol V stabilized nucleosomes in Col-0, *nrpe1*, *ago4* and *idn2-1*. Significance tested using t-test (n.s. = not significant, * = $p < 0.05$, ** = $p< 0.01$,*** = $p< 0.001$). H3K9me2 ChIP signal values were normalized to H3 and Col-0 wild-type. Error bars show standard deviations from three biological replicates.

Enrichment of DNA methylation on Pol V stabilized nucleosomes is inconsistent with the reported preference of DNA methylation towards linker regions [146]. To determine if this inconsistency is a general property of our datasets, we identified all nucleosome positions genome-wide (n=650,610) and measured the average CHH methylation levels at nucleosomes and 500 bp adjacent regions. We observed that CHH methylation was enriched on linker regions and depleted on nucleosomes [Figure 2.5BC]. Although this enrichment was significant on the genome-wide scale, the difference in average DNA methylation levels between nucleosomes and linkers was still relatively small [Figure 2.5B]. We conclude that nucleosome positioning by RdDM overcomes the weak general preference to methylate linker DNA.

We further tested if Pol V-dependent nucleosome positioning is linked to the establishment of repressive chromatin marks by RdDM by assaying the levels of H3K9me2 on Pol V stabilized nucleosomes. MNase ChIP-qPCR using anti-H3K9me2 antibody in wildtype, *nrpe1*, *ago4-1* and *idn2-1* revealed that the levels of H3K9me2 were significantly reduced on the tested Pol V stabilized nucleosomes in *nrpe1* and *ago4* [Figure 2.4B-G]. At the same time, H3K9me2 was unchanged on a negative control locus [Figure 2.4B] and reduced on a positive control RdDM locus [Figure 2.4C]. The *idn2* mutant showed a locus-specific effect, which is consistent with partial redundancy of IDN2 and its paralogs [168], [169]. This indicates that at least at the tested loci, Pol V affects nucleosome positioning in parallel with establishing H3K9me2. Together, these results indicate that Pol V stabilized nucleosomes are enriched in CHH methylation and H3K9me2. This suggests that nucleosome positioning is linked to the entire RdDM pathway and repressive chromatin marks established by this process.

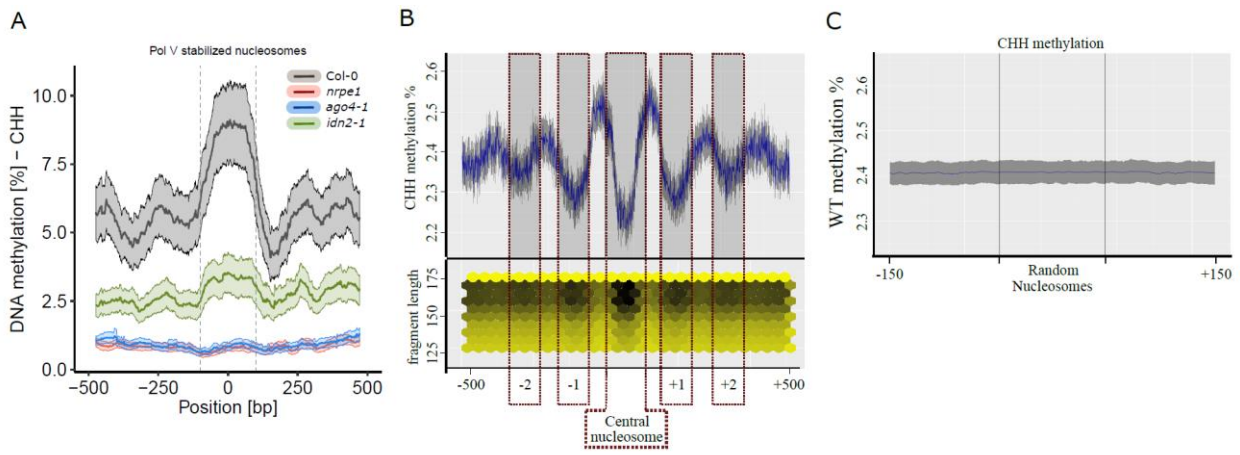

Figure 2.5: **Pol V stabilized nucleosomes are enriched in DNA methylation:** A. Average levels of CHH methylation on and around Pol V stabilized nucleosomes in Col-0, *nrpe1*, *ago4* and *idn2* using datasets from Stroud et al (2013). B. Average levels of CHH methylation on and around all annotated nucleosomes genome wide. Dark grey shading indicates the annotated nucleosome and four neighboring nucleosomes. Ribbon indicates bootstrap confidence intervals with $p < 0.05$. Heatmap on the bottom shows average MNase H3 ChIP signal levels at and around annotated nucleosomes (X axis) plotted by sequenced fragment length (y axis). C. Average levels of CHH methylation at random nucleosome-sized regions. Ribbon indicates bootstrap confidence intervals with $p < 0.05$.

2.3.4 SWI/SNF complex is not required for DNA methylation on positioned nucleosomes

Establishment of both nucleosome positioning and DNA methylation by the RdDM pathway suggests that nucleosome positioning and DNA methylation may be established in parallel. Alternatively, DNA methylation may affect nucleosome positioning or nucleosomes may affect the pattern of DNA methylation. We tested the latter possibility by manipulating nucleosome positioning and testing the levels of CHH methylation. Nucleosome positioning was partially disrupted by mutating *SWI3B*, which has previously been shown to be involved in Pol V-mediated nucleosome positioning [113]. Although *swi3b* null mutants are embryo lethal [156], we took

advantage of the well documented observation that *SWI3B* is haploinsufficient [113], [156], [170] and used the *swi3b/+* heterozygous plants. As expected, the MNase-H3 ChIP signal on Pol V stabilized nucleosomes was slightly but significantly reduced in *swi3b/+* [Figure 2.6A], which was confirmed using locus-specific assays [Figure 2.7A]. Despite this impact on nucleosomes, CHH methylation was not significantly changed in *swi3b/+* [Figure 2.6B]. This indicates that a minor disruption of nucleosome positioning has no detectable direct impact on DNA methylation in the CHH context.

To further test if nucleosomes have an impact on CHH methylation, we looked genome-wide and identified SWI3B stabilized nucleosomes, which are defined as nucleosomes that have a higher MNase H3 ChIP-seq signal level in wildtype compared to *swi3b/+* with FDR of less than 0.05. In total, we identified 4089 SWI3B stabilized nucleosomes, where the average nucleosome signal was significantly and reproducibly decreased in *swi3b/+* [Figure 2.6C]. CHH methylation was unaffected in *swi3b/+* on SWI3B stabilized nucleosomes [Figure 2.6D]. This further supports our conclusion that reduction of nucleosome positioning in *swi3b/+* had no detectable direct impact on DNA methylation in the CHH context. Although the impacts of more substantial disruptions of nucleosomes on DNA methylation remain unknown, our findings are consistent with DNA methylation and nucleosome positioning being established in parallel or DNA methylation working upstream of nucleosome positioning.

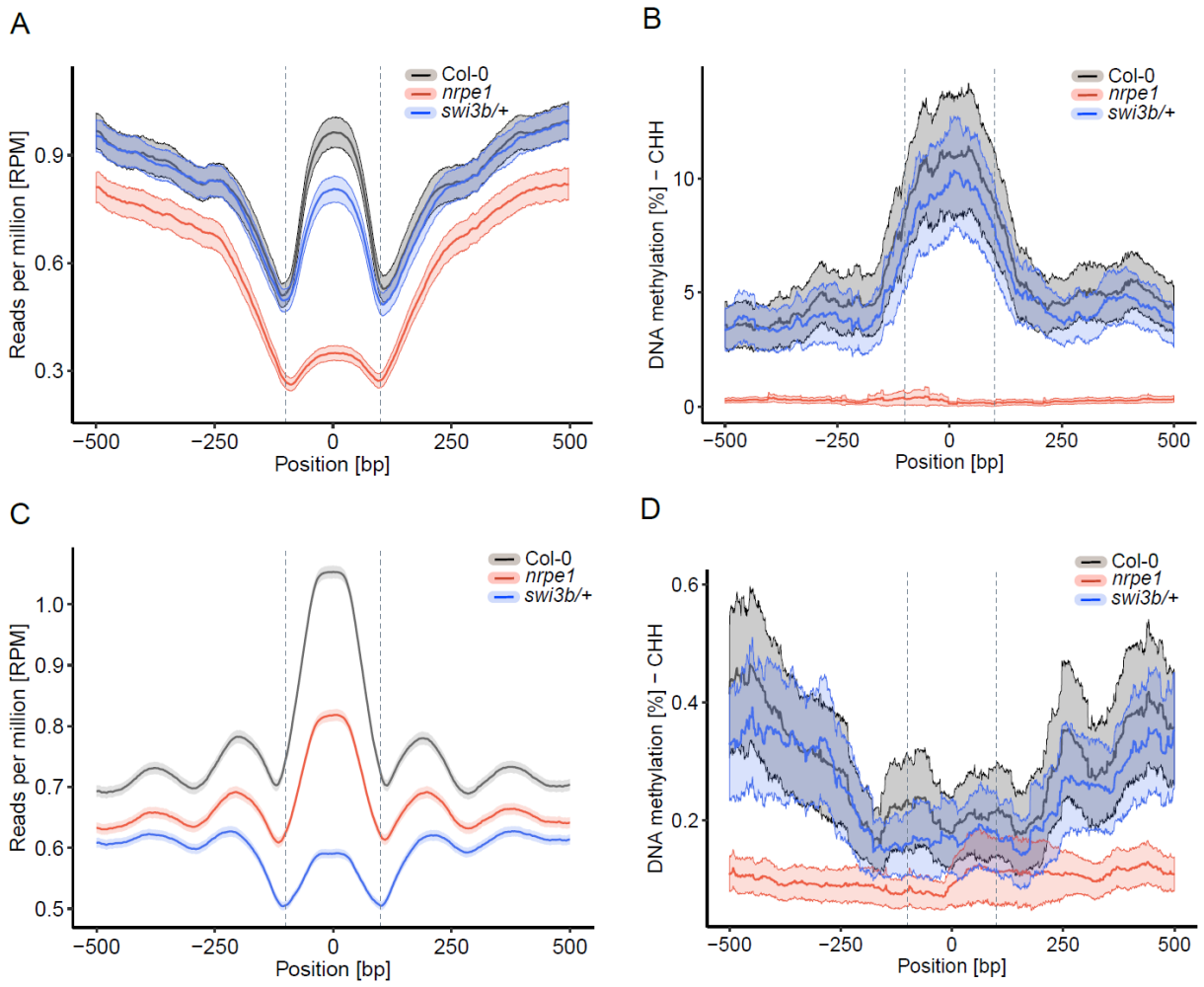

Figure 2.6: **SWI/SNF complex is not required for DNA methylation on positioned nucleosomes:** A. Average levels of MNase H3 ChIP-seq signal on Pol V stabilized nucleosomes in Col-0, *nrpe1* and *swi3b/+*. Ribbons indicate bootstrap confidence intervals with $p < 0.05$. B. Average levels of CHH methylation on and around Pol V stabilized nucleosomes. X axis indicates position (bp). Ribbons indicate bootstrap confidence intervals with $p < 0.05$. C. Average levels of MNase H3 ChIP of Col-0, *nrpe1* and *swi3b/+* at and around SWI3B stabilized nucleosomes. Ribbons indicate bootstrap confidence intervals with $p < 0.05$. D. Average levels of CHH methylation on and around SWI3B stabilized nucleosomes. X axis indicates position (bp). Ribbons indicate bootstrap confidence intervals with $p < 0.05$.

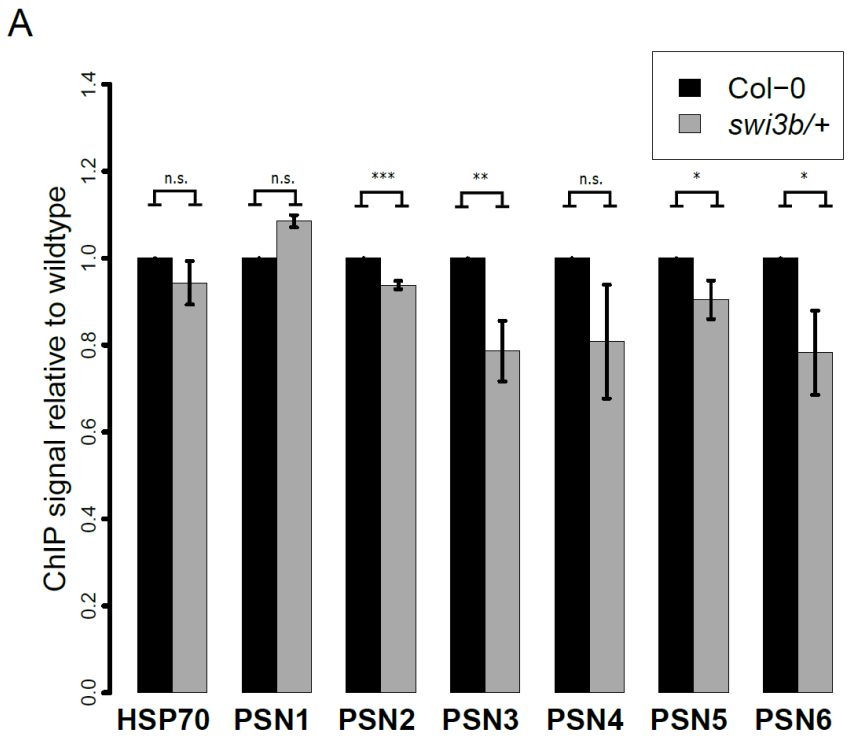

Figure 2.7: **SWI/SNF complex is not required for DNA methylation on positioned nucleosomes**: A. Locus-specific validation of Pol V stabilized nucleosomes by MNase H3 ChIP followed by qPCR. Significance tested using t-test (n.s. = not significant, * = p-value < 0.05, ** = p-value < 0.01,*** = p-value < 0.001). ChIP signal values were normalized to ACTIN2 and Col-0 wild-type. Error bars indicate standard deviations from three biological replicates.

2.3.5 CHH methylation affects nucleosome positioning

To determine if CHH methylation and nucleosome positioning are established in parallel or CHH methylation affects the pattern of nucleosome positioning, we manipulated the levels of CHH methylation and determined the pattern of nucleosome positioning. We first used previously published datasets [112] to identify differentially methylated regions (DMRs), where CHH methylation is reduced in the *drm2* mutant. We then assayed nucleosome positioning by MNase H3 ChIP-seq in two biological replicates of Col-0 wildtype and *drm2* mutant. At DRM2 DMRs

[Figure 2.8A], the nucleosome signal was generally enriched in Col-0 wild type relative to neighboring regions [Figure 2.8B]. This is consistent with Pol V stabilized nucleosomes being enriched in CHH methylation [Figure 3A]. In the *drm2* mutant, the nucleosome signal was significantly reduced [Figure 2.8B]. This indicates that CHH methylation established by DRM2 affects nucleosome positioning.

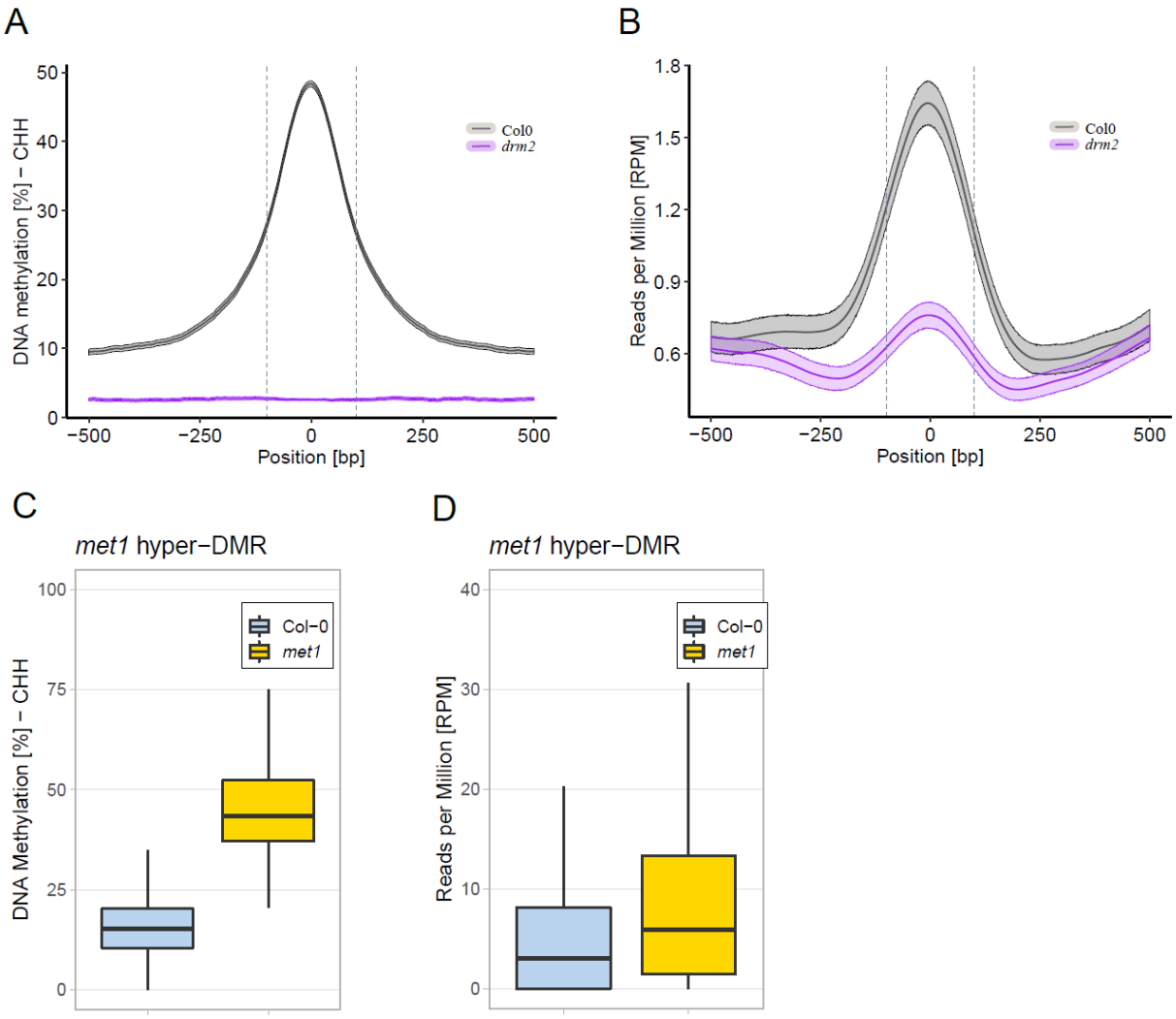

Figure 2.8: **CHH methylation affects nucleosome positioning:** A. Average levels of CHH methylation at and around regions that lose CHH methylation in the *drm2* mutant (DRM2 DMRs). Ribbons indicate bootstrap confidence intervals with $p < 0.05$. B. Average levels of MNase H3 ChIP signal at and around DRM2 DMRs. Ribbons indicate bootstrap confidence intervals with p

< 0.05. C. CHH DNA methylation levels on regions where CHH methylation was increased in the *met1* mutant (MET1 hyper-DMRs). D. MNase H3 ChIP signal levels on MET1 hyper-DMRs.

DMRs are not expected to exactly match lengths and positions of nucleosomes genome-wide. To increase the resolution of this analysis we took the list of all nucleosomes identified by MNase-H3 ChIP and looked only at nucleosomes that overlap DRM2 DMRs. These nucleosomes had the expected loss of CHH methylation in *drm2* [Figure 2.9A]. They also had a significant reduction of the nucleosome signal in *drm2* [Figure 2.9B]. This further confirms that CHH methylation established by DRM2 affects nucleosome positioning.

The levels of CHH methylation are greatly variable between silenced loci. To determine if the level of CHH methylation is associated with the loss of nucleosome signal, we split DRM2 DMRs into five groups based on the difference in CHH methylation between Col-0 wild type and *drm2*. Loci with the highest loss of CHH methylation in *drm2* also had the greatest reduction of the nucleosome signal in *drm2* [Figure 2.9C]. This indicates that there is a quantitative relationship between reduction of CHH methylation and reduction of nucleosome positioning.

Reduction of the nucleosome signal upon reduction of CHH methylation in *drm2* suggests increase of CHH methylation should cause an increase in the nucleosome signal. To test this possibility, we used the *met1* mutant, which is defective in the maintenance of CG methylation and has compensatory hypermethylation in other contexts [171]. Using previously published datasets [112] we identified loci, where CHH methylation was increased in the *met1* mutant [Figure 2.8C]. We then assayed nucleosome positioning by MNase H3 ChIP-seq in two biological replicates of Col-0 wildtype and *met1* mutant. Despite the complexity of interactions between DNA methylation in various sequence contexts, the nucleosome signal was increased on loci with

CHH hypermethylation [Figure 2.8D]. This indicates that the increase of CHH methylation leads to the increase of nucleosome positioning.

Overall, these results are inconsistent with CHH methylation and nucleosome positioning being established in parallel. Instead, they suggest that DNA methylation established by RdDM may be a direct or indirect determinant of nucleosome positioning.

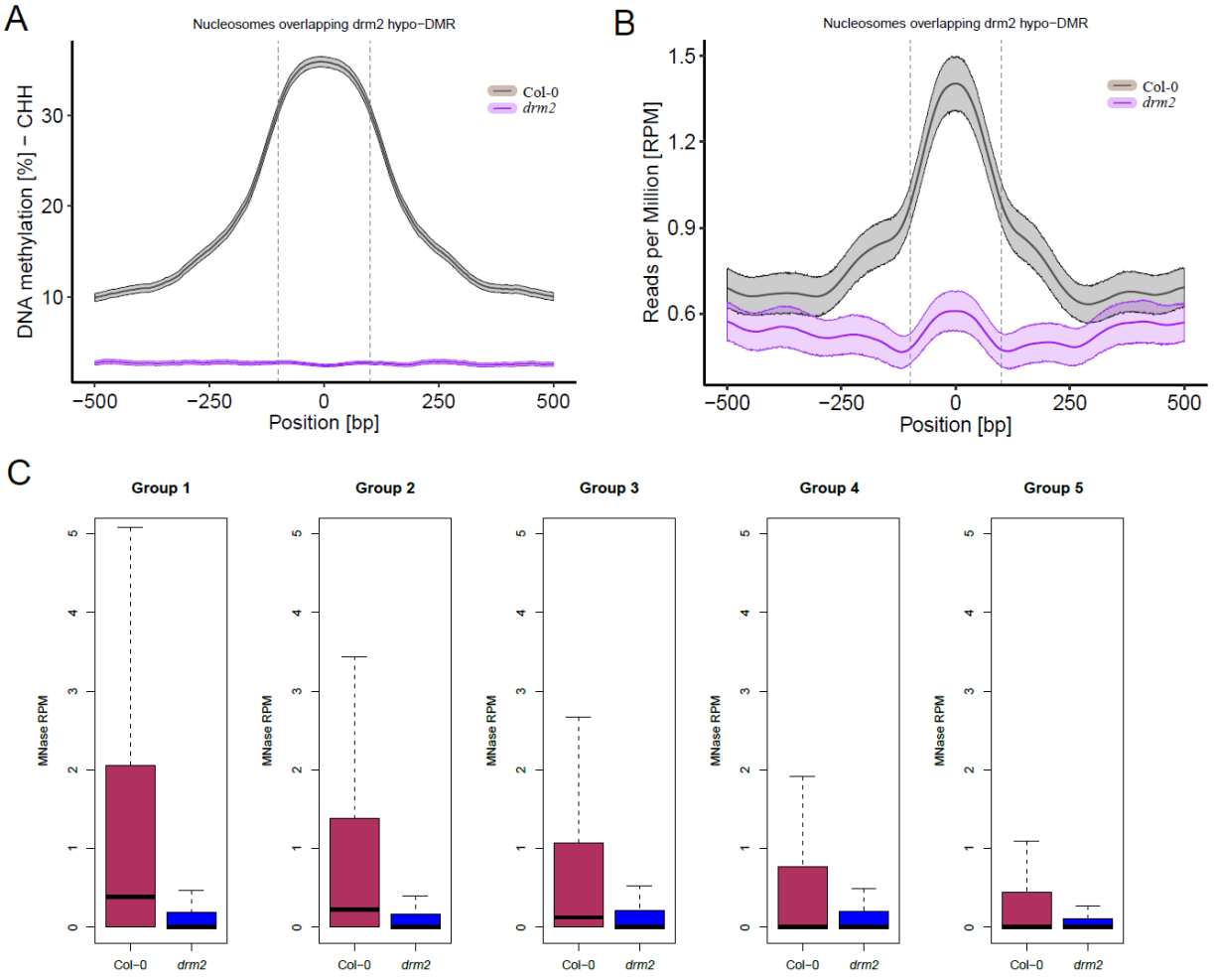

Figure 2.9: **CHH methylation affects nucleosome positioning**: A. Average levels of CHH methylation at and around annotated nucleosomes that overlap DRM2 DMRs. Ribbons indicate bootstrap confidence intervals with $p < 0.05$. B. Average levels of MNase H3 ChIP signal at nucleosomes overlapping DRM2 DMRs. Ribbons indicate bootstrap confidence intervals with $p <$

0.05. C. Levels of MNase H3 ChIP signal at DRM2 DMRs split into five groups based on the difference in CHH methylation between Col-0 wild type and *drm2* with Group 1 having the highest difference and Group 5 having the lowest difference.

2.3.6 CG and CHG methylation affect nucleosome positioning

Impact of DNA methylation on the pattern of nucleosome positioning may be a specific property of the RdDM pathway. Alternatively, it may be a more general property of DNA methylation established by one or more other silencing pathways. To distinguish between these possibilities, we used the *met1* mutant, which loses most CG methylation and is at least partially independent of RdDM [112], [171]. MET1 CG DMRs had the expected strong reduction of CG methylation [Figure 2.10A]. MNase-H3 ChIP performed in Col-0 wild type revealed a significant enrichment of the nucleosome signal on the DMRs [Figure 2.10B], which suggests that CG methylation maintained by MET1 also counteracts the general weak enrichment of DNA methylation on linkers. MNase-H3 ChIP in the *met1* mutant showed a significant reduction of the nucleosome signal on MET1 DMRs [Figure 2.10B]. This indicates that CG methylation maintained by MET1 affects nucleosome positioning.

To determine if CHG methylation also affects nucleosome positioning we used the *cmt3* mutant. CMT3 is responsible for maintaining CHG methylation and is also at least partially independent of RdDM [172]. CMT3 DMRs identified using previously published datasets in the CHG context [112] had the expected strong reduction of CHG methylation [Figure 2.10C]. MNase-H3 ChIP performed in the *cmt3* mutant revealed a significant reduction of the nucleosome signal on CMT3 DRMs [Figure 2.10D]. This indicates that CHG methylation maintained by CMT3 also affects nucleosome positioning.

Overall, these results indicate that the impact of DNA methylation on the pattern of nucleosome positioning is not a unique property of RdDM. Instead, DNA methylation maintained by MET1 and CMT3 pathways also has an impact on nucleosome positioning.

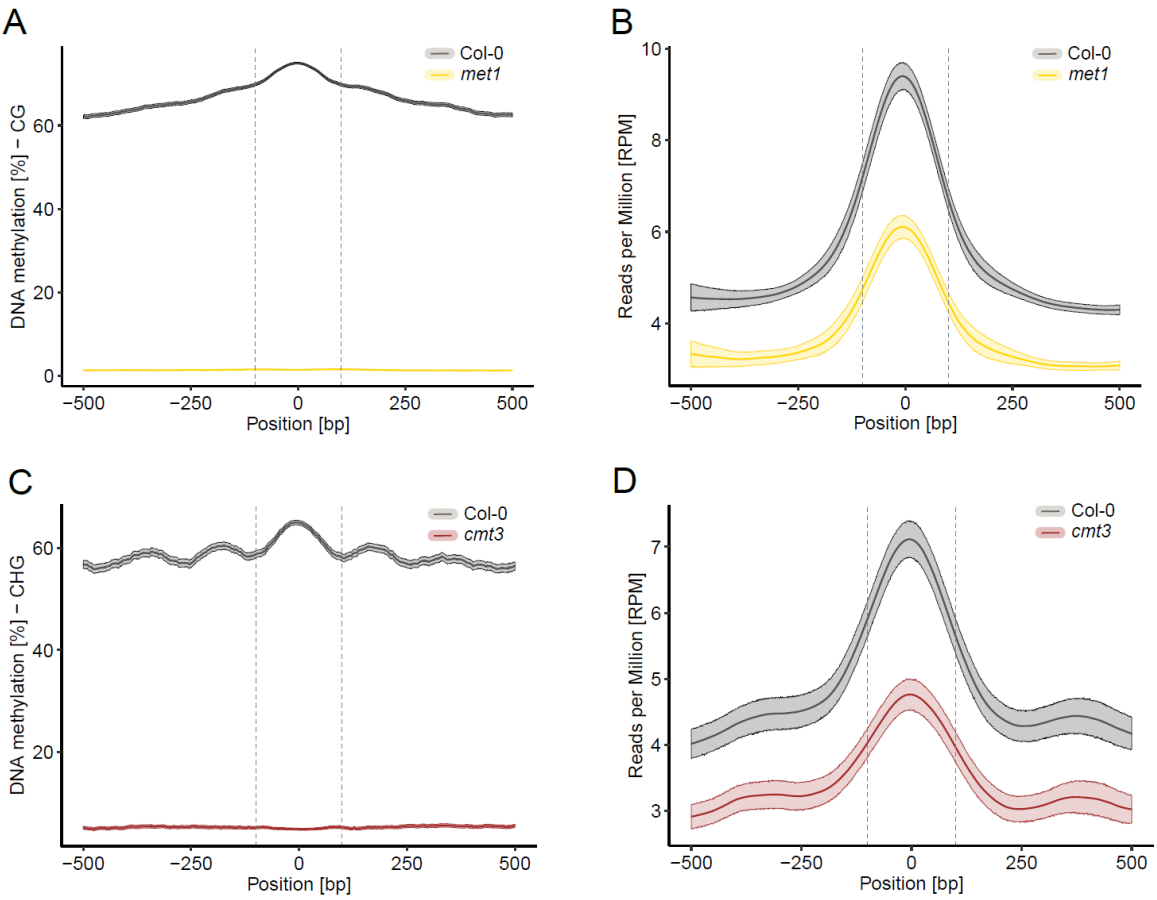

Figure 2.10: **SWI/SNF complex is not required for DNA methylation on SWI3B stabilized nucleosomes**: A. Average levels of CG methylation at and around regions that lose CG methylation in the *met1* mutant (MET1 DMRs). Ribbons indicate bootstrap confidence intervals with $p < 0.05$. B. Average levels of MNase H3 ChIP signal at and around MET1 DMRs. Ribbons indicate bootstrap confidence intervals with $p < 0.05$. C. Average levels of CHG methylation at and around regions that lose CHG methylation in the *cmt3* mutant (CMT3 DMRs). Ribbons

indicate bootstrap confidence intervals with $p < 0.05$. D. Average levels of MNase H3 ChIP signal at and around CMT3 DMRs. Ribbons indicate bootstrap confidence intervals with $p < 0.05$.

2.4 Discussion

We propose a model where DNA methylation is a determinant of nucleosome positioning in RdDM. In this model, non-coding transcription by both Pol IV and Pol V leads to the recruitment of AGO4 and IDN2. IDN2 interacts with a subunit of SWI/SNF but there is little effect on nucleosome positioning from this interaction alone. Instead, the subsequent recruitment of DRM2 and establishment of DNA methylation activates chromatin remodelers and leads to changes in nucleosome positioning. Coordinated establishment of various chromatin marks leads to repression of Pol II promoters within the silenced region of the genome.

There are several possibilities as for how DNA methylation affects nucleosome positioning. First, this effect may be explained by distinct intrinsic properties of DNA containing 5-methylcytosines, as suggested by [164]. Alternatively, DNA methylation may facilitate the recruitment or activation of SWI/SNF, either directly or by the involvement of other proteins that are sensitive to the presence of 5-methylcytosines. Yet another possibility is that DNA methylation may affect nucleosome positioning by changing the pattern of posttranslational histone modifications. This includes H3K9me2, which may recruit proteins that modulate the activity of chromatin remodelers. This also includes histone deacetylation, which may affect physical properties of the nucleosomes [153], [173]. Finally, it also remains possible that the impact of DNA methylation on nucleosome positioning may be much more indirect with loss of DNA methylation causing transcriptional activation or other major functional changes within chromatin, which then affect the placement of nucleosomes.

The importance of DNA methylation for nucleosome positioning has a significant impact on our understanding of the RdDM pathway. It argues against the pathway being branched after IDN2 recruitment [113]. Instead, it supports the notion that events occurring co-transcriptionally at the sites of Pol V transcription are organized in a stepwise genetic pathway [148]. Although when studied genetically, this pathway appears linear, various steps of the pathway are likely to rely on the cooperative recruitment or activation of subsequent factors. One example of such a connection is the requirement of both IDN2-SWI3B interaction and DNA methylation for nucleosome positioning. Other examples include the recruitment of AGO4, which has been proposed to rely on the interaction of AGO4 with the NRPE1 C-terminal domain and with Pol V transcripts [136], [174]. Similarly, there is evidence of DRM2 being recruited by interactions with AGO4 and other RdDM factors [175], [176].

Our model is consistent with the notion that events in the late stages of RdDM lead to a concerted establishment of DNA methylation, posttranslational histone modifications and nucleosome positioning, which together form a repressive chromatin structure. This explains the robustness of transcriptional silencing, where coordinated establishment of various repressive chromatin marks leads to efficient repression of Pol II transcription. It is also consistent with the general difficulty to experimentally tease apart various repressive chromatin modifications established by this pathway.

SWI/SNF and nucleosome positioning act in maintenance of RdDM, where transcription of heterochromatic regions by Pol IV and Pol V may involve the removal or repositioning of previously positioned nucleosomes. This is supported by the involvement of putative chromatin remodelers in initiation and/or elongation of transcription by both of those polymerases [111], [177, p. 1], [178], [179]. Nucleosome positioning established as an outcome of RdDM may serve

to re-create the pattern of nucleosomes disrupted by Pol IV and Pol V. *De novo* RdDM in newly inserted TEs is a distinct scenario, since Pol V is expected to transcribe at very low surveillance levels [167] and no pre-existing repressive chromatin modifications are expected to exist. The role of nucleosome positioning in this *de novo* process remains unexplored.

The involvement of DNA methylation in determining the pattern of nucleosomes extends beyond RdDM targets. The impact of CG and CHG methylation maintained by MET1 and CMT3 silencing pathways on stabilizing nucleosomes indicates that DNA methylation may affect nucleosome patterns beyond RdDM. This is consistent with findings in other eukaryotes [144], [164]. Such an effect of DNA methylation on nucleosome positioning counteracts the overall preference to methylate linkers and contributes to local correlations between nucleosomes and DNA methylation. This property of nucleosomes is consistent with previous reports [143] and may involve the activity of DDM1 [146]. It may also be explained by an indirect mechanism, where the loss of DNA methylation leads to activation of transcription, which then affects the pattern of nucleosomes.

Existing evidence does not support the view that DNA methylation is the primary determinant of the nucleosome pattern. This role remains reserved for a combination of intrinsic factors and active chromatin remodeling. The role of DNA methylation is more limited and probabilistic, clearly visible in meta-analysis of large pools of sequences. Therefore, opposite behaviors of individual loci are expected. Moreover, global losses of DNA methylation in RdDM and DNA methyltransferase mutants may affect the patterns of nucleosomes by a combination of cis- and trans-acting factors, which could only be distinguished using tools targeting DNA methylation to specific loci.

2.5 Materials and Methods

2.5.1 Plant material

Col-0 ecotype wildtype, *nrpe1*/*nrpd1b-11* [110], *ago4-1* (introgressed into the Col-0 background [136]), *idn2-1* [137], *drm2-2* (SAIL_70_E12, [111], *swi3b-2* (GABI_302G08, [156]), *cmt3-11* (SALK_148381) and *met1-3* [180] were grown at 22ºC under white LED light in 16h/8h day/night cycle.

2.5.2 Antibodies

Rabbit polyclonal anti-histone H3 antibody (ab1791) and mouse monoclonal anti-H3K9me2 antibody (ab1220) were obtained from Abcam.

2.5.3 MNase H3 ChIP-seq

2g of approximately 3.5-week old *Arabidopsis thaliana* mature leaf tissue, which was crosslinked with 0.5% formaldehyde, was ground in liquid nitrogen. MNase H3 ChIP of Col-0, *met1*, *cmt3* and *drm2* was carried out as described previously [113]. MNase H3 ChIP of Col-0, *nrpe1* and *swi3b* was carried out using the following protocol. Cold nuclei isolation buffer I (10 mM Tris HCl pH8, 10mM $MgCl_2$, 0.4 M sucrose, 0.035% β-mercaptoethanol, 1mM phenylmethylsulfonyl fluoride (PMSF)) was added. Tissue was resuspended by vigorous vortexing and shaking. Sample was filtered using Miracloth into new 50 ml tube on ice. Miracloth was washed with 10 ml of nuclei isolation buffer I. Sample was centrifuged 15 min, 4000 g, 4˚C.

Supernatant was discarded and nuclei pellet was resuspended using 1 ml of cold nuclei isolation buffer II (10 mM Tris HCl pH8, 10 mM $MgCl_2$, 0.4 M sucrose, 1% Triton X-100, 0.035% β-mercaptoethanol, 1mM phenylmethylsulfonyl fluoride (PMSF), 0.02 tab/ml cOmplete EDTA-free, 0.004 mg/ml Pepstatin A). Sample was transferred to 1.5 ml tube and centrifuged for 5 min,

2000 g, 4˚C. This step was repeated two more times. Pellet was resuspended using 300 µl of cold Nuclei isolation buffer II and layered on top of cold 900 ml Nuclei isolation buffer III (10 mM Tris HCl pH8, 2 mM $MgCl_2$, 1.7 M sucrose, 0.15% Triton X-100, 0.035% β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.02 tab/ml cOmplete EDTA-free, 0.004 mg/ml Pepstatin A) in 1.5 ml tube. Sample was centrifuged for 30 min, 16000 g, 4˚C and supernatant was discarded.

Isolated nuclei were washed twice with Micrococcal Nuclease (MNase) reaction buffer (10 mM Tris HCl pH8, 15 mM NaCl, 60 µM KCl, 1mM $CaCl_2$) and resuspended in the same buffer. MNase enzyme (NEB; 200 Kunitz unit/µl) was added and samples were mixed by vortexing. Samples were digested for 10 minutes at 30℃. 1 volume of MNase stop buffer (30 mM Tris HCl pH8, 225 mM NaCl, 10 mM ethylenediaminetetraacetic acid (EDTA), 10 mM egtazic acid (EGTA), 0.2% sodium dodecyl sulphate (SDS), 2% Tween 20) was then added to stop the reaction. To release the chromatin from the nuclei, the sample was vortexed vigorously 5 times and centrifuged for 10 min, 14000 g. The supernatant was then transferred to a new tube. Samples for H3 ChIP were then diluted in 1 volume ChIP dilution buffer (16.7 mM Tris HCl pH8, 1.2 mM ethylenediaminetetraacetic acid (EDTA), 167 mM NaCl, 1.1% Triton X-100, 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.02 tab/ml cOmplete EDTA-free, 0.004 mg/ml Pepstatin A). H3 antibody was added and sample was incubated 12-16 hours, 4˚C with rotation.

Protein A magnetic beads (Pierce™) were washed three times with IP buffer (50 mM HEPES pH7.5, 150 mM NaCl, 10 µM $ZnSO_4$, 1% Triton X-100, 0.05% sodium dodecyl sulphate (SDS), 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.02 tab/ml cOmplete EDTA-free, 0.004 mg/ml Pepstatin A) and resuspended in 50 µl IP buffer. Beads were added to IP sample and incubated for 1 hour, 4˚C with rotation. Immunoprecipitated chromatin bound to magnetic beads

was collected using magnetic separator. Beads were washed 5 min with cold buffers: two times with low salt buffer (20 mM Tris HCl pH8, 2 mM ethylenediaminetetraacetic acid (EDTA), 150 mM NaCl, 1% Triton X-100, 0.1% sodium dodecyl sulphate (SDS)), once with high salt buffer (20 mM Tris HCl pH8, 2 mM ethylenediaminetetraacetic acid (EDTA), 0.5 M NaCl, 1% Triton X-100, 0.1% sodium dodecyl sulphate (SDS)), once with LiCl buffer (20 mM Tris HCl pH8, 2 mM ethylenediaminetetraacetic acid (EDTA), 250 mM LiCl, 1% NP-100, 1% sodium deoxycholate)) and twice with TE buffer (10 mM Tris HCl pH8, 1 mM ethylenediaminetetraacetic acid (EDTA)). After the last wash, samples were transferred into new a tube and beads were collected using a magnetic separator.

For library preparation, magnetic beads were incubated with 100 µl Elution buffer (10 mM Tris HCl pH8, 1 mM ethylenediaminetetraacetic acid (EDTA), 1% sodium dodecyl sulphate (SDS)) in a thermomixer (65˚C, 1400 rpm, 30 min). Beads were collected using magnetic separator and supernatant was transferred into new tube. Step was repeated and both supernatants combined. IP samples were de-crosslinked by Proteinase K treatment (5 µl, 65˚C, 12 h). Samples were purified using QIAquick® PCR Purification Kit (35 µl of EB buffer were used). Library for Illumina sequencing was prepared using either MicroPlex Library Preparation^{TM} Kit (Diagenode) according manufacturer instruction, using in-house library preparation based on Bowman *et al* [181], or prepared by the University of Michigan Advanced Genomics Core. MNase ChIP-seq experiments were performed in two biological replicates and sequenced by either 50 bp or 150 bp paired-end sequencing at the University of Michigan Advanced Genomics Core.

2.5.4 MNase H3 & H3K9me2 ChIP-qPCR

Nuclei were extracted from 2g of approximately 3.5-week old *Arabidopsis thaliana* mature leaf tissue which was crosslinked with formaldehyde [0.5%] as described previously[113] and were

digested with Micrococcal Nuclease (MNase ; NEB) for 10 minutes at 30℃. MNase-digested chromatin was immunoprecipitated with anti-histone H3 antibody or anti-H3K9me2 antibody. DNA was purified and used for qPCR analysis. MNase ChIP-qPCR experiments were performed in three biological replicates with region-specific primers listed in Table 2.1.

2.5.5 Whole genome bisulfite-seq

Genomic DNA was isolated from approximately 3.5-week old *Arabidopsis thaliana* mature leaf tissue of Col-0 wild type, *swi3b*/+ and *nrpe1* using DNeasy Plant Mini Kit (QIAGEN). DNA was processed for bisulfite treatment and library generation at the University of Michigan Advanced Genomics Core.

2.5.6 Bioinformatic data analysis

MNase H3 ChIP-seq paired-end reads from two independent biological replicates were aligned and processed to the *Arabidopsis* TAIR10 genome with Bowtie2 [182]. Mapped reads were deduplicated using PICARD tools (http://broadinstitute.github.io/picard) and filtered by fragment length between 120-170 bp and MAPQ value of >=2. Differential nucleosomes were identified using DANPOS2 [183] by filtering nucleosomes with more than 2 fold enrichment in either in Col-0 for PolV stabilized nucleosomes or in *nrpe1* for PolV destabilized nucleosomes and FDR< 0.05. Nucleosomes were then filtered using the negative binomial test with reads from biological replicates using the NBPseq R package [184]. For subsequent analysis we selected nucleosomes which showed more than 2 fold-change and FDR < 0.05. We further refined the nucleosome positions for well-positioned nucleosomes by filtering for main peak nucleosomes using iNPS [185]. Nucleosome data was (RPM) normalized and visualized on heatmaps and profiles by calculating the number of reads using BEDTools 2.15.0 at nucleosome dyads [186]. Overlap analyses with nucleosomes were performed with 1000 permutated genomic regions to obtain

expected numbers and p values. SWI3B-stabilized nucleosomes were filtered for higher read counts in Col-0 than the *swi3b* mutant and an FDR<0.05. These nucleosomes were then further filtered using the negative binomial test with reads from biological duplicates using NBPseq and the nucleosomes with FDR<0.01 were selected for further analysis.

The sequencing reads from whole genome bisulfite-seq datasets were mapped to the TAIR10 genome using the Bismark software allowing no mismatches [187]. DNA methylation levels were calculated by the ratio of #C/(#C+#T) after selecting for Cs with at least 5 sequenced reads. Differentially Methylated Regions (DMRs) were identified using methylKit package in R [188]. The bin sizes used were 100bp bins with a step-size of 50bps. 10 minimum bases were required in each tile. A 25% (hypo-DMRs) or 20% (hyper-DMRs) minimum methylation differences were selected for in each of the tiles and an FDR value of 0.01 was used. The number of MNase-H3 ChIP-seq reads overlapping these DMRs were then plotted as a profile.

Locus	**Name**	**Sequence (5'-3')**	**Application**
		Nucleosome validation	
PSN1	MH487	caggttgtgagttcgaatcgt	ChIP-qPCR
	MH488	catctccgttagccaccttt	
PSN2	MH489	tgagattttaccgggtccac	ChIP-qPCR
	MH490	cccttatacgtaatttccatcaca	
PSN3	MH491	ggagtgggatgtagactcgaa	ChIP-qPCR
	MH492	ctagtggtaccgcagggttt	
PSN4	MH493	cgatcggttcgatctcctta	ChIP-qPCR
	MH494	taacggttcaacccgagaaa	
PSN5	MH495	tctcccccacaatttctgtc	ChIP-qPCR
	MH496	aaatggacccctcattgtca	
PSN6	MH501	acagatagcgctgtacagatttta	ChIP-qPCR
	MH502	tcatttgatatgcgttttgttt	
ACTIN2	Actin2-A118	gagagattcagatgcccagaagtc	ChIP-qPCR
	Actin2-A119	tggattccagcagcttcca	[111]
HSP70	A512	ctcttcctcacacaatcataaaca	ChIP-qPCR
	A513	cagaattgttcgccggaaag	[189]

		H3K9me2 validation	
IGN22	MH537	cgggtccttggactcctgat	ChIP-qPCR [135]
	MH538	tcgtgaccggaataattaaatgg	
ACTIN2	Actin2-A118	gagagattcagatgcccagaagtc	ChIP-qPCR [111]
	Actin2-A119	tggattccagcagcttcca	
PSN1	MH487	caggttgtgagttcgaatcgt	ChIP-qPCR
	MH488	catctccgttagccaccttt	
PSN3	MH491	ggagtgggatgtagactcgaa	ChIP-qPCR
	MH492	ctagtggtaccgcagggttt	

Table 2.1: Oligonucleotides used in this study, Related to the Experimental Procedures

2.5.7 Other datasets used in this study

Arabidopsis genome annotations (TAIR10) were obtained from TAIR (www.arabidopsis.org). Pol V ChIP-seq data (SRA054962) and peak list and Pol V RIP-seq data (GSE70290) and annotated regions were published previously [165], [190]. DNA methylation data from *idn2*, *ago4*, *drm2*, *met1* and *cmt3* mutants as well as corresponding Col-0 and *nrpe1* controls were obtained from GSE39901 [112].

2.6 Acknowledgements

2.6.1 Author Contributions

M Hafiz Rothi*: Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, United States

Contributions: MHR, Conception and design, Analysis and interpretation of data, Data acquisition, Drafting or revising the article.

Shriya Sethuraman*: Bioinformatics Graduate Program, University of Michigan, Ann Arbor, United States

Contributions: SS, Conception and design, Analysis and interpretation of data, Data acquisition, Drafting or revising the article.

Jakub Dolata: Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, United States

Contributions: JD, Conception and design, Data acquisition, Revising the article.

Alan P Boyle: Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, United States

Contributions: APB, Analysis and interpretation of data, Revising the article.

Andrzej T Wierzbicki: Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, United States

Contributions: ATW, Conception and design, Analysis and interpretation of data, Drafting or revising the article.

2.6.2 Funding

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM108722 to ATW. JD was supported by the National Science Center grants Etiuda UMO-2014/12/T/NZ2/00246 and Opus UMO- 2017/25/B/NZ1/00603.

2.6.3 Data and materials availability

High throughput sequencing datasets obtained in this study have been deposited in Gene Expression Omnibus under accession GSE148173.

CHAPTER III

Reinforcement of Transcriptional Silencing by a Positive Feedback Between DNA Methylation and Non-Coding Transcription

The contents of this chapter have been submitted for publication and is currently in the review process. Masayuki Tsuzuki prepared Pol V IPARE samples for sequencing. Shriya Sethuraman optimized the sequencing read mapping protocol. Andrzej T. Wierzbicki conducted data analysis and generated plots shown in Figure 3.3C, 3.3D, 3.4C, 3.5, 3.6, 3.7D, 3.7E, 3.9 and 3.10. I performed all other experiments and data analysis shown in this chapter.

3.1 Abstract

Non-coding transcription is an important determinant of heterochromatin formation. In *Arabidopsis thaliana* a specialized RNA polymerase V (Pol V) transcribes pervasively and produces long non-coding RNAs. These transcripts work with small interfering RNA to facilitate locus-specific establishment of RNA-directed DNA methylation (RdDM). Subsequent maintenance of RdDM is associated with elevated levels of Pol V transcription. However, the impact of DNA methylation on Pol V transcription remained unresolved. We found that DNA

methylation strongly enhances Pol V transcription. The level of Pol V transcription is reduced in mutants defective in RdDM components working downstream of Pol V, indicating that RdDM is maintained by a mutual reinforcement of DNA methylation and Pol V transcription. Pol V transcription is affected only on loci that lose DNA methylation in all sequence contexts in a particular mutant, including mutants lacking maintenance DNA methyltransferases, which suggests that RdDM works in a complex crosstalk with other silencing pathways.

3.2 Introduction

RNA-directed DNA methylation (RdDM) is a plant transcriptional silencing pathway which targets transposable elements (TE), transgenes and repetitive sequences [191]. These loci are then turned off by the establishment of repressive chromatin marks, including posttranslational histone modifications, nucleosome positioning and DNA methylation [192], [193]. RdDM is determined by two classes of non-coding RNA [194], [195]. The first is small interfering RNA (siRNA), which is produced by the activities of RNA-dependent RNA polymerases and Dicer-like proteins [196]. siRNA incorporates into Argonaute proteins and gives them sequence-specificity towards loci complementary to siRNA [197], [198]. The second class of non-coding RNA involved in RdDM is produced by a specialized RNA polymerase, Pol V [166], [167], [199]–[201]. Pol V transcribes long non-coding RNA (lncRNA) and lncRNA is required for recognition of target loci by siRNA-Argonaute complexes, which has been proposed to occur via siRNA-lncRNA base-pairing [198], [200], [202], [203]. The consequence of this recognition is recruitment of chromatin modifiers and heterochromatin formation [204]–[206].

The most important property of RdDM is its locus specificity, which assures that TEs are recognized and silenced, but essential protein-coding genes are not targeted. This specificity is

achieved when a TE is newly integrated or activated. As a TE becomes transcribed by Pol II, it produces aberrant transcripts, which are the preferred substrates for RNA-dependent RNA polymerases and give rise to primary siRNAs [196], [207]. Pol V has been recently shown to transcribe broadly and surveil the genome to make it competent to receive the silencing signal from primary siRNA [167]. Therefore, Pol V contributes little or no sequence-specificity to the initiation of RdDM.

Once initiated, silencing is often not maintained epigenetically and has to be reinforced by a continuous activity of the RdDM pathway. This process involves another specialized RNA polymerase, Pol IV, which produces substrates for RDR2 and DCL3 and leads to relatively high accumulation of 24nt siRNA [208], [209]. It also involves Pol V, which transitions from a very low level of surveillance transcription to a more efficient production of lncRNAs on silenced loci [167]. Both events are caused by the presence of repressive chromatin marks. H3K9me2 is recognized by the SHH1 protein, which recruits Pol IV [210], [211]. Methylated DNA is bound by SUVH2 and SUVH9 proteins, which facilitate Pol V transcription [212], [213]. Consistently, Pol V association with chromatin is often reduced in the *met1* mutant [212]. This strongly suggests that RdDM is a self-reinforcing mechanism, where DNA methylation and H3K9me2 enhance Pol IV and Pol V transcription, which leads to further establishment of repressive chromatin marks.

The presence of a self-reinforcing feedback loop between elevated Pol V transcription and DNA methylation has one important implication. It suggests that disruption of RdDM factors that work downstream of Pol V should lead to loss of DNA methylation and subsequently reduction of Pol V transcription. Surprisingly, it is not the case and Pol V transcripts still accumulate in those mutants, including *spt5l*, *ago4* and *drm2* [167], [198], [200], [201], [204], [214]. This

inconsistency indicates that the relationship between Pol V transcription and DNA methylation remains unresolved.

One possible explanation for the inability to disrupt the RdDM feedback loop is the presence of multiple overlapping silencing pathways [172], [212], [215]–[217]. In this scenario, maintenance of silencing on a subset of RdDM loci may be performed not only by RdDM but also by MET1 and/or CMT3. In this study we tested this possibility by performing genome-wide identification of Pol V transcription in mutants defective in downstream RdDM components and DNA methyltransferases. We found that loci transcribed by Pol V are indeed targeted by multiple overlapping and partially redundant silencing pathways. This confounds the ability to detect the self-reinforcing properties of RdDM. When effects of other pathways are eliminated, the positive feedback of Pol V transcription and DNA methylation becomes clearly detectable.

3.3 Results

3.3.1 RdDM loci are targeted by multiple silencing pathways

RdDM has been proposed to work as a self-reinforcing feedback loop [212], which predicts that mutants in components acting downstream of Pol V should affect the accumulation of Pol V transcripts. Several studies indicated that this is not the case and Pol V transcripts accumulate in *spt5l*, *ago4* and *drm2* mutants [167], [198], [200], [201], [204], [214]. To resolve these conflicting results, we performed Pol V IPARE in the *drm2* mutant, and reanalyzed previously published comparable Pol V IPARE datasets in Col-0, *ago4* and *spt5l* [167]. As expected, the overall accumulation of Pol V transcripts on all known RdDM Pol V-transcribed regions [167] was only slightly reduced in *spt5l*, *ago4* or *drm2* mutants. This reduction was much smaller than observed in *nrpe1*, a mutant in the largest subunit of Pol V (Fig. 3.1A, Fig. 3.2AB).

One potential explanation of this observation is that not all DNA methylation is lost in the studied mutants [216]. To test this hypothesis, we reanalyzed previously published whole genome bisulfite sequencing datasets [218] and determined the levels of DNA methylation in all three contexts on the same known RdDM Pol V-transcribed regions [167]. We found that while CHH methylation was substantially reduced, the levels of CG methylation remained high in *spt5l*, *ago4* and *drm1/2* mutants (Fig. 3.1B). The remaining CG methylation may explain why these mutants only have minor effects on Pol V transcription.

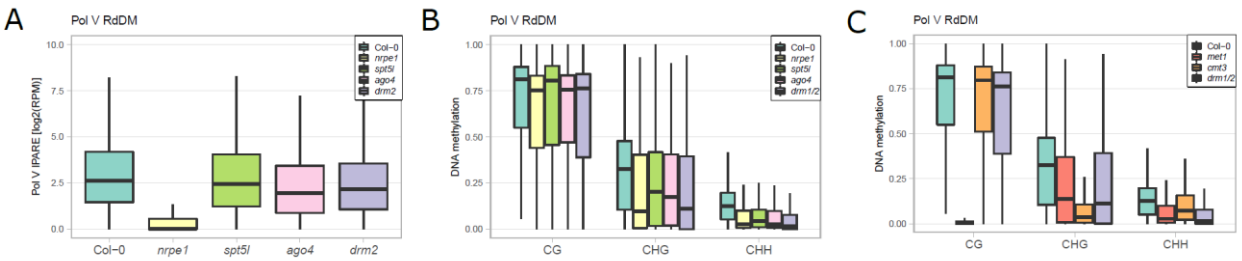

Figure 3.1: **RdDM loci are targeted by multiple silencing pathways**: A. Small effects of mutants in downstream components of RdDM on Pol V transcription throughout the genome. Pol V IPARE signal levels were plotted on previously identified Pol V RdDM regions [167] in Col-0, *nrpe1*, *spt5l*, *ago4*, and *drm2*. Individual biological replicates are shown in Fig. 3.2AB. B. Presence of symmetric DNA methylation in RdDM mutants. DNA methylation levels [218] were plotted on previously identified Pol V RdDM regions [167] in CG, CHG and CHH contexts in Col-0, *nrpe1*, *spt5l*, *ago4*, and *drm1/2*. C. Residual DNA methylation in DNA methyltransferase mutants. DNA methylation levels [218] were plotted on previously identified Pol V RdDM regions [167] in CG, CHG and CHH contexts in Col-0, *met1*, *cmt3*, and *drm1/2*.

High levels of residual DNA methylation in RdDM mutants are consistent with previous observations that RdDM loci are commonly targeted by several silencing pathways [172], [212], [217]. To provide further support for this conclusion, we determined the levels of DNA methylation on RdDM Pol V-transcribed regions [167] in DNA methyltransferase mutants, which disrupt various silencing pathways. The *cmt3* mutant had a strong reduction of CHG methylation

only (Fig. 3.1C). *drm1/2* had reduced levels of CHH and to a smaller extent CHG methylation but no major change in CG methylation (Fig. 3.1C). *met1* had an almost complete loss of CG methylation but only partial reductions of CHH and CHG methylation (Fig. 3.1C). This indicates that as expected, RdDM Pol V-transcribed loci are targeted not only by RdDM but also by variable contributions of CMT3 and MET1. Together, these results indicate that RdDM loci are targeted by multiple overlapping silencing pathways.

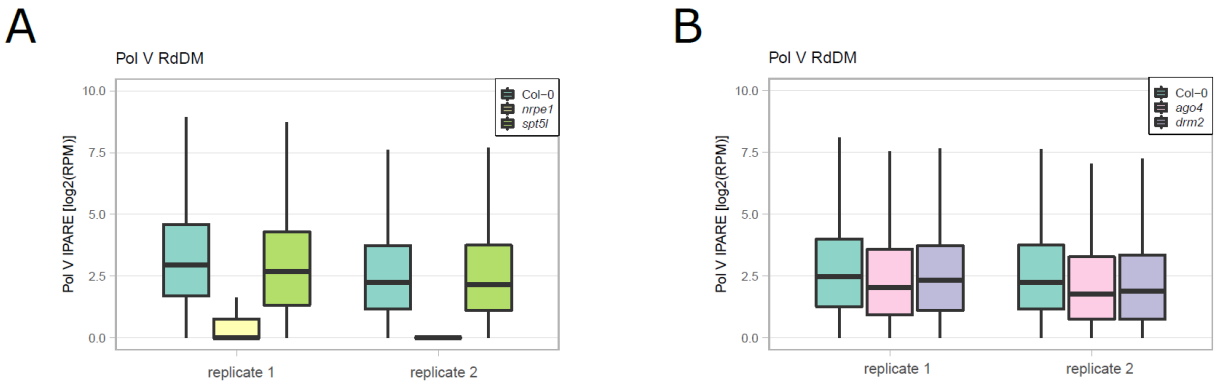

Figure 3.2: **Individual biological replicates of datasets showing minimal effects of mutants in downstream components of RdDM on Pol V transcription:** Pol V IPARE signal levels were plotted on previously identified Pol V RdDM regions [167] in Col-0, *nrpe1* and *spt5l* (A) and Col-0, *ago4*, and *drm2* (B).

3.3.2 Maintenance of RdDM requires DNA methylation by DRM2

Presence of multiple silencing pathways on RdDM loci may confound the ability to test the role of DNA methylation for Pol V transcription. To overcome this limitation, we took advantage of the fact that each particular locus may be targeted by any combination of silencing pathways and relative contributions of various pathways at least partially depend on the frequency of cytosines in particular contexts [172]. This means that some loci may be primarily silenced by just one pathway and therefore a subset of loci is expected to have no DNA methylation in *drm2* in all

contexts. To identify these loci, we found differentially methylated regions (DMRs) that lose CHH methylation in *drm1/2* (*drm1/2* DMRs) and are transcribed by Pol V. We then split these DMRs into two categories based on the presence or absence of CG and CHG methylation in *drm1/2*. The control group had CG and CHG methylation detectable in *drm1/2* (Fig. 3.3A, "Both CG and CHG present"). The second group had no CHG and no CG methylation detectable in *drm1/2* (Fig. 3.3A, "Neither CG nor CHG present"). We then calculated the abundance of Pol V transcription in those groups in Col-0 wild type and *drm2* mutant. While the control group had only a small reduction of Pol V transcription in *drm2* (Fig. 3.3B), the group with no CHG and no CG methylation had a substantially greater reduction of Pol V transcription in *drm2* (Fig. 3.3B, Fig. 3.4C). The level of Pol V transcription in *drm2* on loci with no CHG and no CG methylation in *drm2* was significantly lower than on control loci ($p < 10^{-16}$, Wilcoxon test). This indicates that loss of DNA methylation in all contexts in *drm2* leads to a substantial reduction of Pol V transcription.

To further confirm the role of all DNA methylation contexts for maintaining high levels of Pol V transcription, we performed a reciprocal analysis. We identified Pol V-transcribed genomic regions, where Pol V IPARE signal was significantly reduced in *drm2* and control loci where no difference in Pol V IPARE signal was detected in *drm2* (Fig. 3.3C, Fig. 3.4C). We then assayed DNA methylation in Col-0 wild type, *drm1/2* and *nrpe1*. Loci where Pol V transcription was DRM2-independent showed strong reductions of CHG and CHH methylation but mostly maintained relatively high levels of CG DNA methylation in *drm1/2* (Fig. 3.3D). In contrast, loci that lost Pol V transcription in *drm2* also lost DNA methylation in all sequence contexts, including CG (Fig. 3.3D). Levels of CG methylation in *drm2* on loci that lost Pol V transcription in *drm2* were significantly lower than at loci where Pol V transcription was DRM2-independent ($p < 10^{-179}$, Wilcoxon test). This indicates that residual CG methylation allows maintaining high levels of

Pol V transcription and the reduction of Pol V transcription in *drm2* is associated with the loss of DNA methylation in all sequence contexts. Levels of DNA methylation in all contexts were similar in *drm1/2* and *nrpe1* on both categories of loci (Fig. 3.4D), which is consistent with Pol V being generally required for DNA methylation by DRM2.

Together, these results indicate that RdDM Pol V transcription requires DNA methylation in at least one sequence context. This is consistent with RdDM operating as a self-reinforcing feedback loop and enhanced Pol V transcription on silenced loci playing an important role in this feedback.

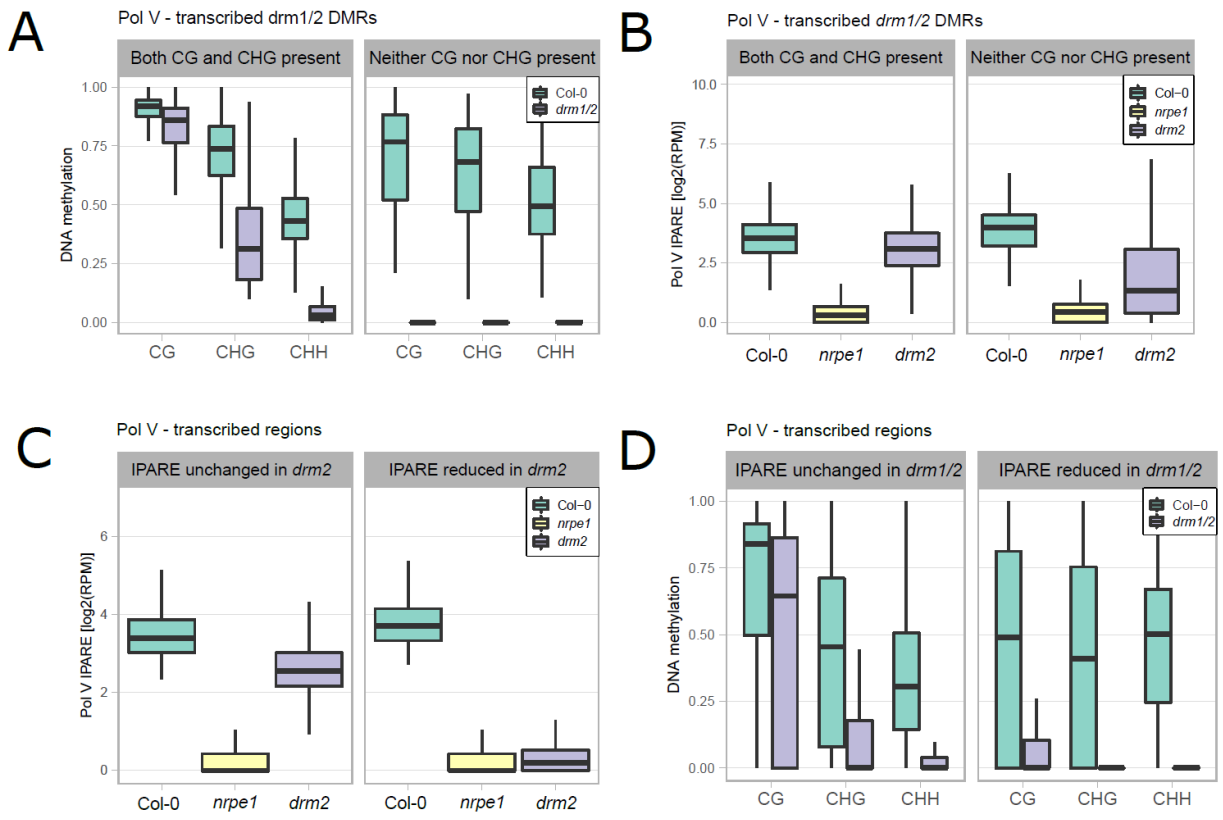

Figure 3.3: **Maintenance of RdDM requires DNA methylation by DRM2 at loci not targeted by other silencing pathways**: A. Control plot showing *drm1/2* DMRs split by the presence or absence of symmetric methylation in *drm1/2*. DNA methylation levels [218] were plotted on Pol V-transcribed *drm1/2* CHH DMRs split by the levels of CG and CHG methylation. There were

3113 DMRs with both CG and CHG present in both Col-0 and *drm2* as well as 276 DMRs with CG and CHG present in Col-0 but absent in *drm2*. DMRs were identified by difference between the whole genome bisulfite sequencing (WGBS) CHH signal of Col-0 and *drm1/2* > 0.2 and FDR < 0.01. Presence of DNA methylation was defined as WGBS signal > 0.2 (CG) or > 0.1 (CHG). Absence of DNA methylation was defined as WGBS signal of 0. B. Substantial reduction of Pol V transcription in *drm2* on loci that lose DNA methylation in all contexts. Pol V IPARE signal was plotted on two categories of Pol V-transcribed *drm1/2* DMRs in Col-0, *nrpe1*, and *drm2*. Individual biological replicates are shown in Fig. 3.4AB. C. Control plot showing genomic Pol V-transcribed bins split by the impact of DRM2 on Pol V transcription. Pol V IPARE signal was plotted on Pol V-transcribed regions with either Pol V IPARE reduced (1246 bins) or unchanged (8945 bins) in *drm2*. Bins were identified as Pol V-transcribed by IPARE signal being significantly greater in Col-0 compared to *nrpe1* (FDR < 0.05 [184]). IPARE signal was defined as reduced in *drm2* by FDR < 0.05, and as unchanged in *drm2* by FDR > 0.9 and fold change smaller than 2. Individual biological replicates are shown in Fig. 3.4C. D. Substantial reduction of DNA methylation in *drm2* in all contexts on genomic bins with DRM2-dependent Pol V transcription. DNA methylation levels [218] in CG, CHG and CHH contexts were plotted on Pol V-transcribed regions with Pol V IPARE signal reduced or unchanged in *drm2*. Corresponding data for *nrpe1* and total levels of DNA methylation in all contexts are shown in Fig. 3.4D.

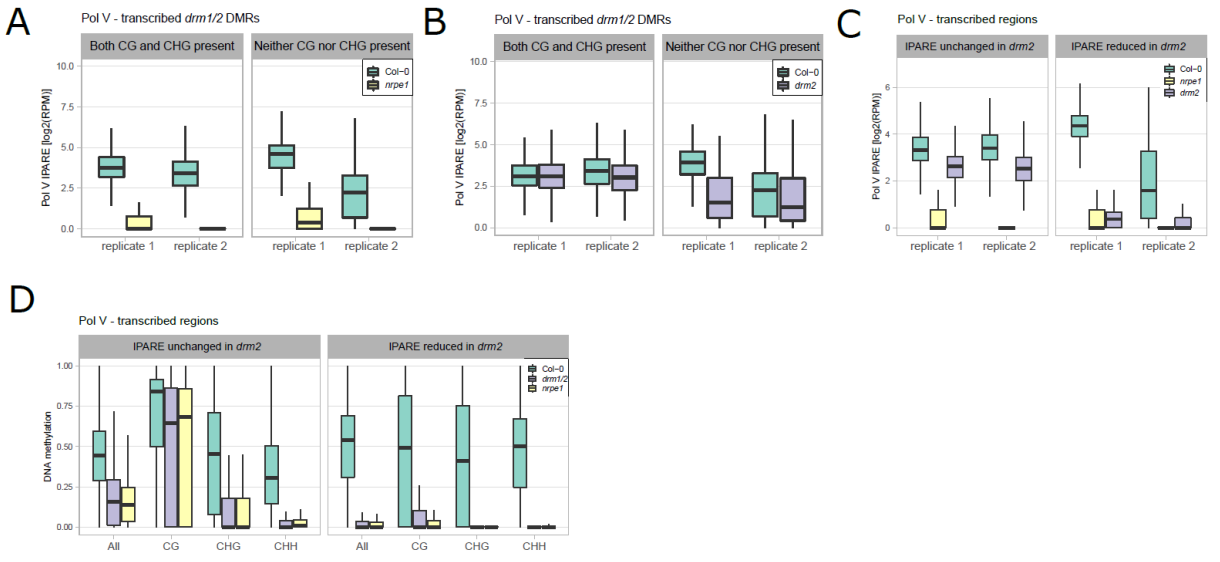

Figure 3.4: **Maintenance of RdDM requires DNA methylation by DRM2 at loci not targeted by other silencing pathways**: A. Individual biological replicates of data showing a substantial reduction of Pol V IPARE signal in *nrpe1* on loci that lose DNA methylation in all contexts shown in Fig. 3.3B. Pol V IPARE signal was plotted on two categories of Pol V-transcribed *drm1/2* DMR in Col-0 and *nrpe1*. B. Individual biological replicates of data showing a substantial reduction of Pol V IPARE in *drm2* on loci that lose DNA methylation in all contexts shown in Fig. 3.3B. Pol

V IPARE signal was plotted on two categories of Pol V-transcribed *drm1/2* DMRs in Col-0 and *drm2*. C. Individual biological replicates of control data showing genomic bins split by the presence or absence of DRM2-dependent Pol V transcription shown in Fig. 3.3C. Pol V IPARE signal was plotted on regions with either Pol V IPARE reduced or unchanged in *drm2*. D. Substantial reduction of DNA methylation in *drm1/2* and *nrpe1* in all contexts on genomic bins with DRM2-dependent Pol V transcription. DNA methylation levels [218] in CG, CHG and CHH contexts as well as total DNA methylation levels were plotted on Pol V-transcribed regions with Pol V IPARE signal reduced or unchanged in *drm2*.

3.3.3 Downstream components are required for maintenance of RdDM

The self-reinforcing loop between Pol V transcription and DNA methylation is expected to be disrupted not only in the *drm2* mutant but also in mutants defective in other downstream RdDM components, including *spt5l* and *ago4*. To test this prediction, we analyzed Pol V IPARE from the *spt5l* mutant [167] and identified Pol V-transcribed genomic regions that had no changes of Pol V transcription in *spt5l* (Fig. 3.5A, Fig. 3.6A). These regions had strong reductions of CHG and CHH methylation but retained high levels of CG methylation in *spt5l* (Fig. 3.5B). In contrast, regions with significant reductions of Pol V transcription in *spt5l* (Fig. 3.5A, Fig. 3.6A) had substantial reductions of DNA methylation in all sequence contexts, including CG (Fig. 3.5B). Levels of CG methylation in *spt5l* at loci that lost Pol V transcription in *spt5l* were significantly lower than at loci where Pol V transcription was SPT5L-independent ($p < 10^{-250}$, Wilcoxon test). This indicates that residual CG methylation allows maintaining high levels of Pol V transcription and a subset of loci where the level of Pol V transcription is dependent on SPT5L also loses DNA methylation in all sequence contexts in *spt5l*.

We further tested the contribution of AGO4 to the self-reinforcement of RdDM by analyzing Pol V IPARE in the *ago4* mutant. Pol V-transcribed genomic regions with no reductions of Pol V transcription in *ago4* (Fig. 3.5C) had strong reductions of CHG and CHH methylation

but retained high levels of CG methylation in *ago4* (Fig. 3.5D). Regions that lost Pol V transcription in *ago4* (Fig. 3.5C, Fig. 3.6B) also showed substantial reduction of DNA methylation in *ago4* in all sequence contexts, including CG (Fig. 3.5D). Levels of CG methylation in *ago4* at loci that lost Pol V transcription in *ago4* were significantly lower than at loci where Pol V transcription was AGO4-independent ($p < 10^{-51}$, Wilcoxon test). This further demonstrates the role of residual CG methylation in maintaining Pol V transcription and shows that a subset of loci where Pol V transcription is dependent on AGO4 also loses DNA methylation in *ago4* in all sequence contexts.

Together, these results demonstrate that Pol V transcription is enhanced by DNA methylation and confirm that RdDM is controlled by a self-reinforcing feedback loop between the level of Pol V transcription and DNA methylation. This feedback loop may be disrupted by mutating *SPT5L* or *AGO4* and is only detectable on loci with no confounding activity of other silencing pathways.

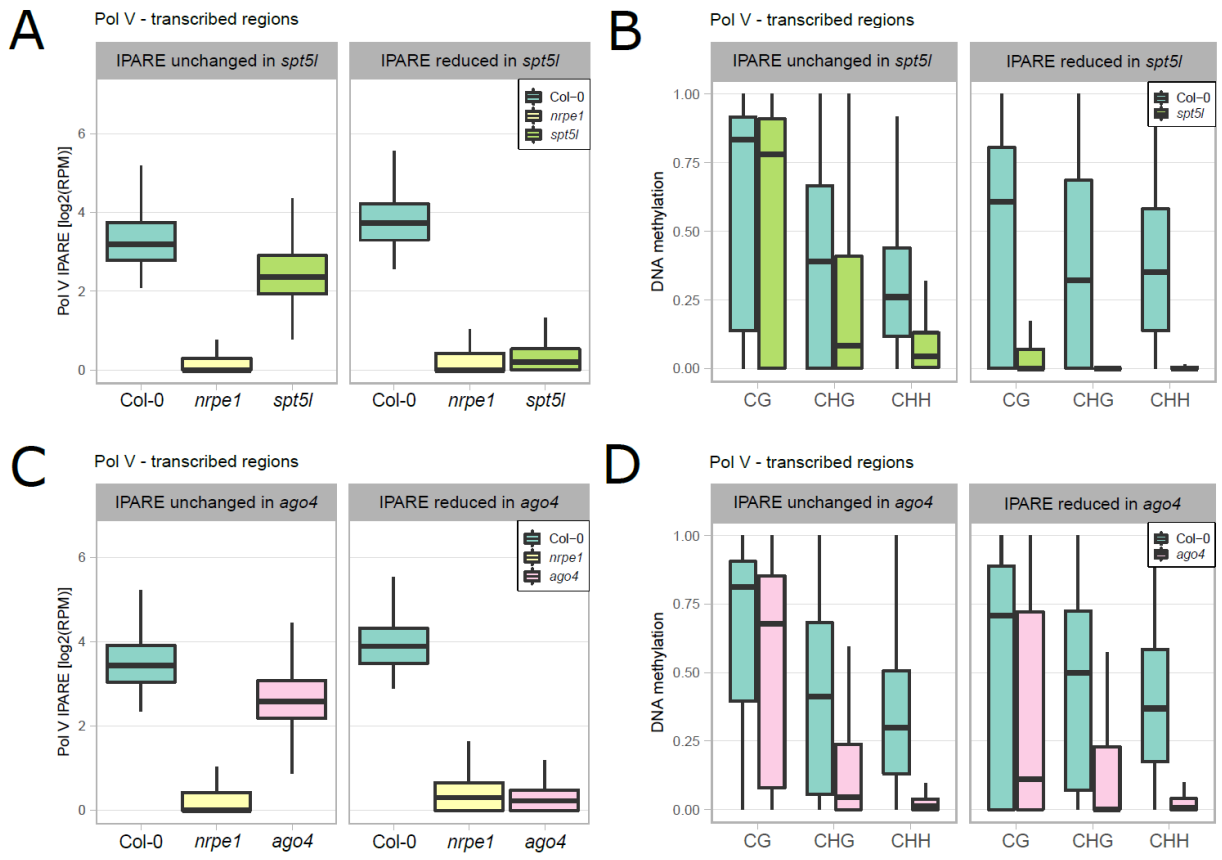

Figure 3.5: **Downstream components are required for maintenance of RdDM at loci where they are needed for DNA methylation in all contexts:** A. Control plot showing Pol V-transcribed genomic bins split by the impact of SPT5L on Pol V transcription. Pol V IPARE signal was plotted on Pol V transcribed regions with either Pol V IPARE reduced (1304 bins) or unchanged (13115 bins) in *spt5l*. Bins were identified as Pol V-transcribed by IPARE signal being significantly greater in Col-0 compared to *nrpe1* (FDR < 0.05 [184]). IPARE signal was defined as reduced in *spt5l* by FDR < 0.05, and as unchanged in *spt5l* by FDR > 0.9 and fold change smaller than 2. Individual biological replicates are shown in Fig. 3.6A. B. Substantial reduction of DNA methylation in *spt5l* in all contexts on genomic bins with SPT5L-dependent Pol V transcription. DNA methylation levels [218] in CG, CHG and CHH contexts was plotted on regions with Pol V IPARE signal reduced or unchanged in *spt5l*. C. Control plot showing genomic bins split by the presence or absence of AGO4-dependent Pol V transcription. Pol V IPARE signal was plotted on regions with either Pol V IPARE reduced (1048 bins) or unchanged (9181 bins) in *ago4*. Bins were identified as Pol V-transcribed by IPARE signal being significantly greater in Col-0 compared to *nrpe1* (FDR < 0.05 [184]). IPARE signal was defined as reduced in *ago4* by FDR < 0.05, and as unchanged in *ago4* by FDR > 0.9 and fold change smaller than 2. Individual biological replicates are shown in Fig. 3.6B. D. Substantial reduction of DNA methylation in *ago4* in all contexts on

genomic bins with AGO4-dependent Pol V transcription. DNA methylation levels [218] in CG, CHG and CHH contexts were plotted on Pol V-transcribed regions with Pol V IPARE signal reduced or unchanged in *ago4*.

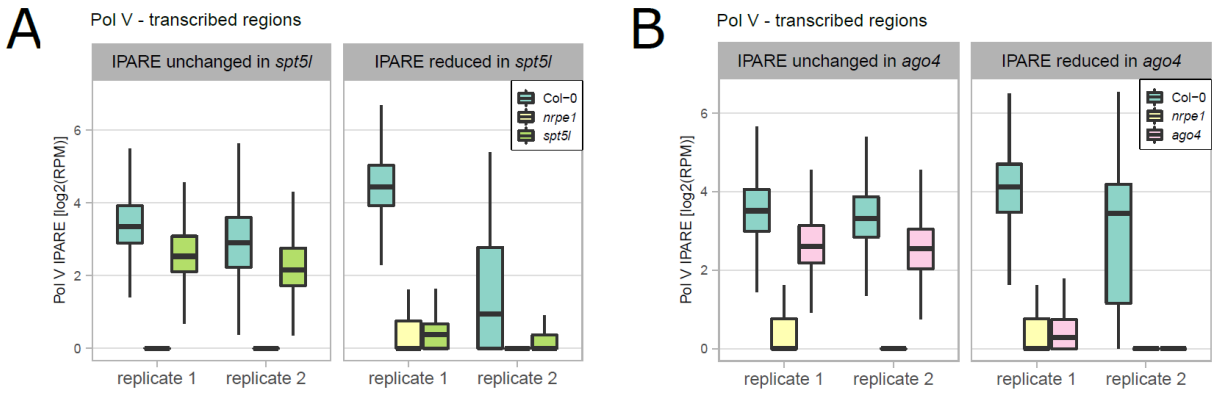

Figure 3.6: **Downstream components are required for maintenance of RdDM at loci where they are needed for DNA methylation in all contexts**: A. Individual biological replicates of control data showing genomic bins split by the presence or absence of SPT5L-dependent Pol V transcription. Pol V IPARE signal was plotted on regions with either Pol V IPARE reduced or unchanged in *spt5l*. B. Individual biological replicates of control data showing genomic bins split by the presence or absence of AGO4-dependent Pol V transcription. Pol V IPARE signal was plotted on regions with either Pol V IPARE reduced or unchanged in *ago4*.

3.3.4 MET1 is needed for maintenance of RdDM

The most prominent silencing pathway that overlaps RdDM is maintenance of CG methylation by MET1 [212]. Disruption of this process by mutating *MET1* affects the levels of CHH methylation and has an impact on Pol V binding to chromatin [212]. This pathway is likely to be responsible for high levels of CG methylation remaining in *nrpe1* and downstream mutants on RdDM Pol V-transcribed loci (Fig. 1B). To test the impact of MET1 on Pol V transcription, we performed Pol V IPARE in the *met1* mutant. The overall accumulation of Pol V transcripts on all known RdDM Pol V-transcribed regions [167] was reduced in *met1* to a greater extent than in *drm2* or *cmt3* but

was still strongly enriched over the background level observed in *nrpe1* (Fig. 1A, Fig. 3.7A). This indicates that maintenance of CG methylation by MET1 is important but not absolutely required for Pol V transcription.

Our findings that loss of DNA methylation in all contexts in downstream RdDM mutants leads to reduction of Pol V transcription suggest a similar relationship in *met1*. To test this possibility, we found DMRs that lose CG methylation in *met1* (*met1* DMRs) and are transcribed by Pol V. We then split these DMRs into categories based on the presence or absence of CHG and CHH methylation in *met1* (Fig. 3.7B) and calculated the abundance of Pol V transcription in those groups in Col-0 wild type and *met1* mutant (Fig. 3.7C). Regions with no CHG and no CHH methylation in *met1* had a substantially greater reduction of Pol V transcription in *met1* than regions that retain CHG and CHH methylation in *met1* (Fig. 3.7C). The level of Pol V transcription in *met1* on loci with no CHG and no CHH methylation in *met1* was significantly lower than on control loci ($p < 10^{-16}$, Wilcoxon test). This indicates that loss of DNA methylation in all contexts in *met1* leads to a substantial reduction of Pol V transcription.

To further confirm the role of all DNA methylation contexts for maintaining high levels of Pol V transcription, we performed a reciprocal analysis. We identified Pol V-transcribed loci where Pol V transcription was unchanged in *met1* (Fig. 3.7D). These loci lost CG methylation but retained substantial levels of CHG and CHH methylation in *met1* (Fig. 3.7E). In contrast, loci with significantly reduced Pol V transcription in *met1* (Fig. 3.7D) had strong reductions of DNA methylation in all sequence contexts, including CHG and CHH (Fig. 3.7E). Levels of CHG and CHH methylation in *met1* at loci that lost Pol V transcription in *met1* were significantly lower than at loci where Pol V transcription was MET1-independent ($p < 10^{-199}$ and $p < 10^{-291}$ respectively, Wilcoxon test). This indicates that remaining CHG and CHH methylation allows maintaining Pol

V transcription in *met1*. Reduction of Pol V transcription in *met1* at a subset of loci is associated with the loss of DNA methylation in all sequence contexts.

These results demonstrate that at a subset of loci, disruption of CG methylation maintenance in the *met1* mutant leads to loss of DNA methylation in all sequence contexts. This negatively affects the level of Pol V transcription and disrupts the maintenance of RdDM. This indicates that MET1 is involved in determining the level of Pol V transcription and therefore contributes to the maintenance of RdDM.

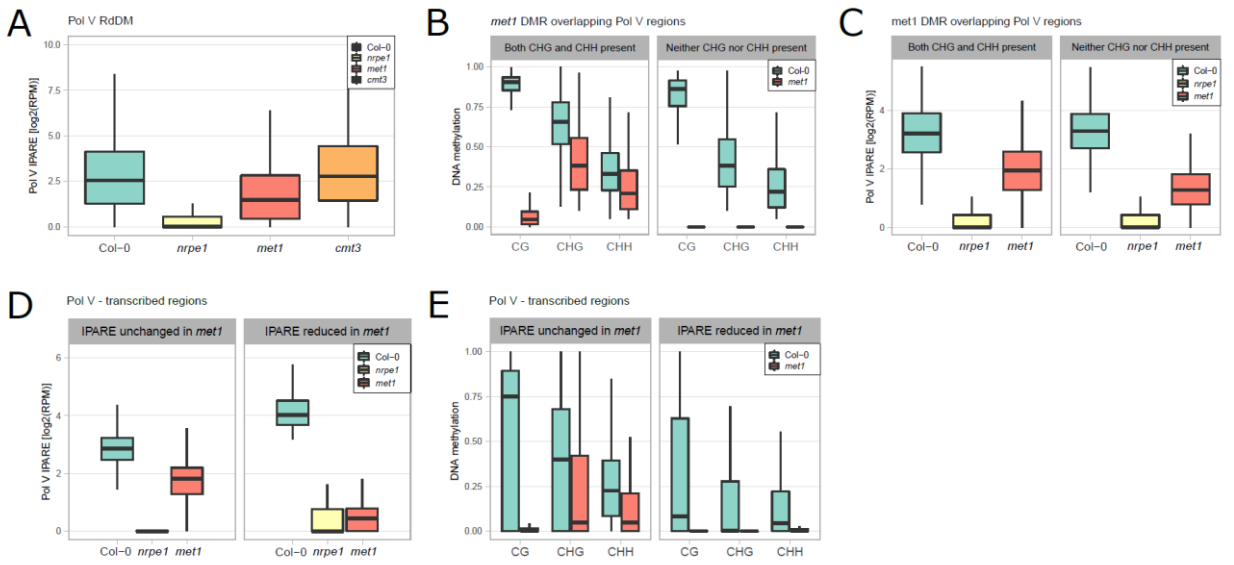

Figure 3.7: **MET1 is needed for maintenance of RdDM at loci where it affects DNA methylation in all contexts**: A. Effects of DNA methyltransferase mutants on Pol V transcription. Pol V IPARE signal levels were plotted on previously identified Pol V RdDM regions [167] in Col-0, *nrpe1*, *met1* and *cmt3*. Individual biological replicates are shown in Fig. 3.8. B. Control plot showing *met1* DMRs split by the presence or absence of non-CG methylation in *met1*. DNA methylation levels [218] were plotted on Pol V-transcribed *met1* CG DMRs split by the level of CHG and CHH methylation in *met1*. There were 1819 DMRs with both CHG and CHH present in both Col-0 and *met1* as well as 995 DMRs with CHG and CHH present in Col-0 but absent in *met1*. DMRs were identified by difference between the WGBS CG signal of Col-0 and *met1* > 0.55 and FDR < 0.01. Presence of DNA methylation was defined as WGBS signal > 0.1 (CHG) or > 0.05 (CHH). Absence of DNA methylation was defined as WGBS signal of 0. C. Substantial reduction of Pol V transcription in *met1* on loci that lose DNA methylation in all contexts in *met1*.

Pol V IPARE signal was plotted on two categories of Pol V-transcribed *met1* DMRs in Col-0, *nrpe1*, and *met1*. D. Control plot showing Pol V-transcribed genomic bins split by the presence or absence of MET1-dependent Pol V transcription. Pol V IPARE signal was plotted on regions with either Pol V IPARE reduced (2231 bins) or unchanged (5755 bins) in *met1*. Bins were identified as Pol V-transcribed by IPARE signal being significantly greater in Col-0 compared to *nrpe1* (FDR < 0.05 [184]). IPARE signal was defined as reduced in *met1* by $p < 0.01$ at 2-fold change or greater calculated using GFOLD [219], and as unchanged in *met1* by GFOLD $p < 0.01$ at 0.1-fold change or smaller and fold change smaller than 2. E. Substantial reduction of DNA methylation in *met1* in all contexts on genomic bins with MET1-dependent Pol V transcription. DNA methylation levels [218] in CG, CHG and CHH contexts were plotted on regions with Pol V IPARE signal reduced or unchanged in *met1*.

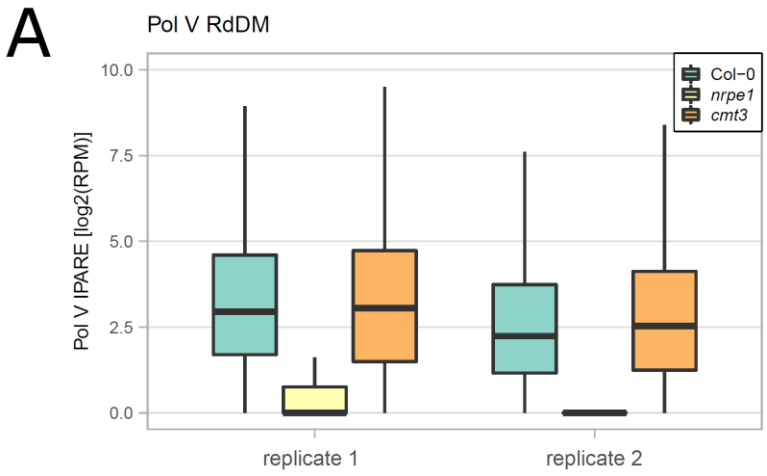

Figure 3.8: **Individual biological replicates of datasets showing minimal effects of mutants in DNA methyltransferase CMT3 on Pol V transcription**: Pol V IPARE signal levels were plotted on previously identified Pol V RdDM regions [167] in Col-0 and *cmt3*.

3.3.5 CMT3 affects RdDM maintenance at a subset of loci

Although RdDM loci are also often targeted by CMT3 ([217] and Fig. 1C), DNA methylated in CHG contexts is not preferentially bound by SUVH2 or SUVH9 *in vitro* [220]. This predicts that CMT3 should not contribute to the maintenance of RdDM and mutating *CMT3* should not lead to the loss of RdDM Pol V transcription. To test this prediction, we identified Pol V-transcribed

regions that had significant reductions of Pol V transcription in *cmt3* (Fig. 3.9A, Fig. 3.10A). These sequences only partially overlapped loci with Pol V transcription dependent on AGO4, DRM2 or MET1 (Fig. 3.10B). We then compared them to regions with no change of Pol V transcription in *cmt3* (Fig. 3.9A, Fig. 3.10A). Regions where Pol V transcription was unchanged in *cmt3* had a partial reduction of CHG methylation but retained high levels of CG and CHH methylation in *cmt3* (Fig. 3.9B), higher than in *nrpe1* (Fig. 3.10C). In contrast, regions where Pol V transcription was significantly reduced in *cmt3* also had substantial reductions of DNA methylation in *cmt3* in all sequence contexts (Fig. 3.9B), greater than in *nrpe1* (Fig. 3.10C). Levels of CG, CHG and CHH methylation in *cmt3* at loci that lost Pol V transcription in *cmt3* were significantly lower than at loci where Pol V transcription was CMT3-independent ($p < 10^{-142}$ for CG, $p < 10^{-116}$ for CHG, and $p < 10^{-234}$ for CHH, Wilcoxon test). This indicates that CMT3 contributes to the maintenance of RdDM. At a subset of loci, disruption of CHG methylation maintenance in the *cmt3* mutant leads to loss of DNA methylation in all contexts, which disrupts the maintenance of RdDM.

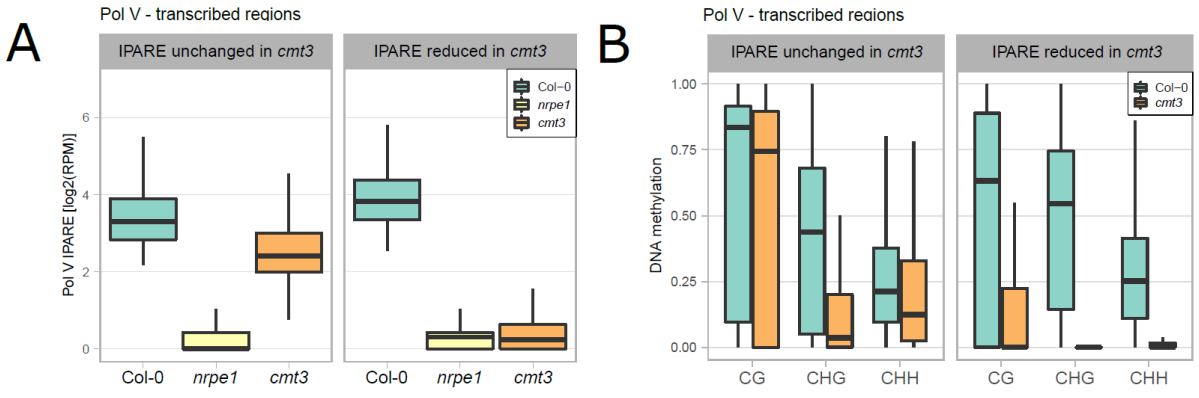

Figure 3.9: **CMT3 affects RdDM maintenance at loci where it is needed for DNA methylation in all contexts**: A. Control plot showing Pol V-transcribed genomic bins split by the presence or absence of CMT3-dependent Pol V transcription. Pol V IPARE signal was plotted on regions with either Pol V IPARE reduced (912 bins) or unchanged (8735 bins) in *cmt3*. Bins were identified as Pol V-transcribed by IPARE signal being significantly greater in Col-0 compared to *nrpe1* (FDR

< 0.05 [184]). IPARE signal was defined as reduced in *cmt3* by FDR < 0.05, and as unchanged in *cmt3* by FDR > 0.9 and fold change smaller than 2. Individual biological replicates are shown in Fig. 3.10A. B. Substantial reduction of DNA methylation in *cmt3* in all contexts on genomic bins with CMT3-dependent Pol V transcription. DNA methylation levels [218] in CG, CHG and CHH contexts were plotted on regions with Pol V IPARE signal reduced or unchanged in *cmt3*.

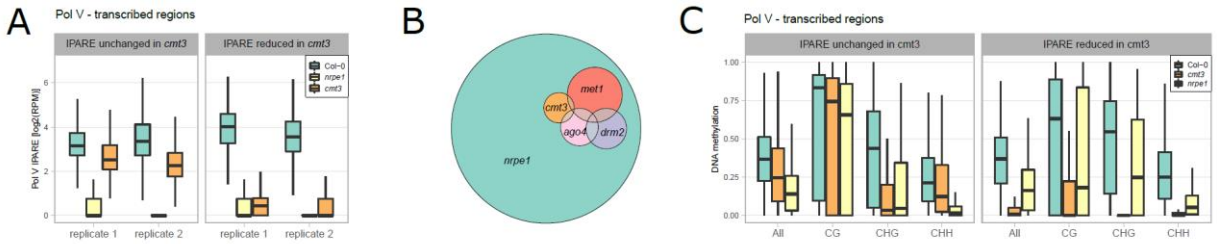

Figure 3.10: **CMT3 affects RdDM maintenance at loci where it is needed for DNA methylation in all contexts** A. Individual biological replicates showing control data with genomic bins split by the presence or absence of CMT3-dependent Pol V transcription presented in Fig. 3.9A. Pol V IPARE signal was plotted on regions with either Pol V IPARE reduced or unchanged in *cmt3*. B. Overlaps of genomic bins with Pol V transcription reduced in *nrpe1*, *met1*, *cmt3* and *ago4*. C. Substantial reduction of DNA methylation in *cmt3* and *nrpe1* in all contexts on genomic bins with CMT3-dependent Pol V transcription. DNA methylation levels [218] in CG, CHG and CHH contexts as well as total DNA methylation levels were plotted on Pol V-transcribed regions with Pol V IPARE signal reduced or unchanged in *cmt3*.

3.3.6 RdDM feedback is enriched on TE edges

Edges of long TEs are known to be preferentially targeted by DRM2-dependent CHH methylation [217], [218], [221] and Pol V transcription, which has been proposed to act as a determinant of heterochromatin/euchromatin boundaries [199]. In contrast, regions inside long TEs are primarily silenced by epigenetically maintained CHG and CG methylation [217], [218], [221]. This suggests that edges of long TEs are likely to be targeted by stable silencing by the positive feedback of RdDM. To test this prediction, we identified genomic bins, where significant reduction of Pol V transcription in the *drm2* mutant indicates the presence of positive feedback by RdDM. We then

overlapped these regions with genes and TEs. Distribution of loci with RdDM feedback resembled the overall pattern of Pol V transcription [199] in being enriched on intergenic regions and depleted on LTR TEs (Fig. 3.11A). Importantly, it was more strongly enriched on edges of long TEs than on the inner regions of long TEs (Fig. 3.11A). To further confirm that TE edges are preferential targets of the RdDM feedback, we plotted DRM2-dependent Pol V transcription on RdDM-targeted TEs [123], [222]. Average levels of DRM2-dependent Pol V transcription were enriched on edges of studied TEs (Fig. 3.11B, Fig. 3.12AB), which is consistent with relatively low amounts of DNA methylation remaining on those regions in *drm2* (Fig. 3.11C-E). This indicates that RdDM feedback is preferentially active on the edges of TEs, which is consistent with the role of RdDM in determining boundaries of heterochromatin.

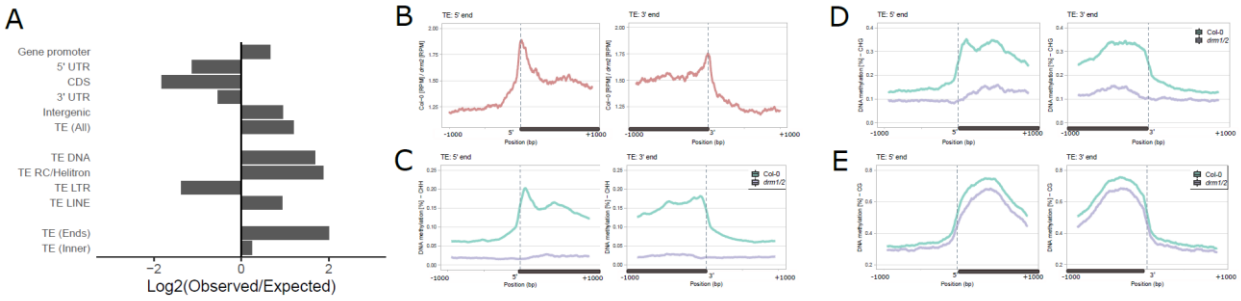

Figure 3.11: **RdDM feedback is enriched on TE edges**: A. Overlaps of genomic bins that show evidence of DRM2-dependent Pol V transcription with genome annotations, including regions within genes, TE families, ends of long TEs and inner regions of long TEs [123]. The plot shows ratios between observed overlaps and average expected overlaps calculated from 1000 permutations of random genomic bins. For all reported enrichments and depletions $p < 0.001$ (permutation test). B. Average levels of DRM2-dependent Pol V transcription on 5' and 3' ends of TEs targeted by RdDM [123]. C-E. Average levels of DNA methylation [218] in the (C) CHH, (D) CHG and (E) CG contexts on 5' and 3' ends of TEs targeted by RdDM [123].

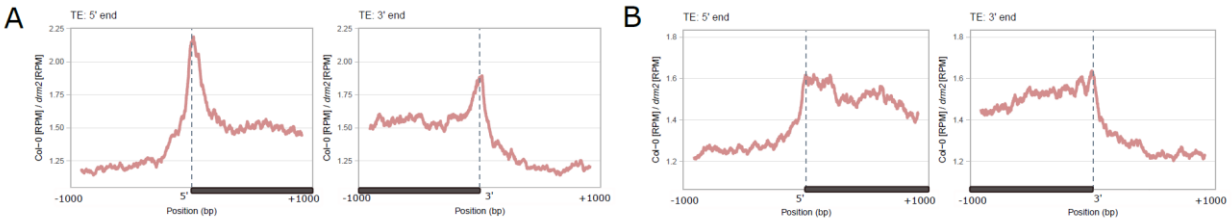

Figure 3.12: **RdDM feedback is enriched on TE edges:** A. Average levels of DRM2-dependent Pol V transcription on 5' and 3' ends of TEs targeted by RdDM [123]. Biological replicate 1. B. Average levels of DRM2-dependent Pol V transcription on 5' and 3' ends of TEs targeted by RdDM [123]. Biological replicate 2.

3.4 Discussion

Our results indicate that RdDM is a self-reinforcing process, where Pol V transcription and DNA methylation enhance each other to maintain silencing. Presence of DNA methylation in at least one sequence context positively affects Pol V transcription and DNA methylation in multiple sequence contexts allows a crosstalk with other silencing mechanisms. Therefore, maintenance of DNA methylation on particular loci by MET1 and CMT3 pathways contributes to enhanced transcription by Pol V. Locus-specific contributions of individual silencing pathways are determined by a combination of the frequency of cytosines in particular contexts [172], presence of H3K9me2 [223] and other factors.

The mechanism of Pol V transcription enhancement by DNA methylation is unlikely to be mediated exclusively by Pol V recruitment as Pol V has been shown to transcribe broadly, even in euchromatin [167]. Instead, DNA methylation may allow both Pol V recruitment and Pol V transcription at elevated rates, typical of RdDM loci [167]. This is likely to be partially mediated by methylated DNA-binding of SUVH2 and SUVH9 and the recruitment of the DDR complex [212], [213], [220]. However, these factors also contribute to the low level of non-RdDM Pol V transcription which indicates that the mechanism of Pol V transition from surveillance to RdDM

transcription is likely to be more complex [167]. More importantly, there are many loci in the genome which have high levels of DNA methylation but no evidence of RdDM Pol V transcription, such as genes with body DNA methylation [204]. This indicates that DNA methylation is not sufficient to specifically control Pol V transcription. One potential explanation of the variable levels of Pol V transcription is exclusion of Pol V by Pol II and associated chromatin modifications. Another possibility is that there is an additional, yet unknown factor, which works together with DNA methylation to control the level of Pol V transcription.

Enhancement of Pol V transcription on methylated loci allows efficient recruitment of siRNA-AGO4 complexes to silenced loci [197]–[199] and facilitates further DNA methylation by DRM2 [204]. Therefore, loss of AGO4 or SPT5L leads to the reduction of DNA methylation and consequent reduction of Pol V transcription. Enhancement of Pol V transcription on methylated loci is likely accompanied by recruitment of Pol IV and elevated production of siRNA, which explains why loss of downstream silencing factors leads to reduction of siRNA accumulation on subsets of loci [224], [225].

Self-reinforcement of RdDM is particularly important on edges of TEs, which are preferentially transcribed by Pol V [199]. This is consistent with the role of RdDM in precisely determining the boundaries between heterochromatin and euchromatin [199], [226]. The importance of RdDM self-reinforcement on TE edges may be explained by the low resolution of MET1 and CMT3 pathways, which is limited by the distribution of cytosines in symmetric contexts and/or the nucleosome size. In contrast, RdDM is enhanced by CHH methylation, which is more frequent and allows higher resolution of Pol V transcription determination [199]. Pol V has also been shown to preferentially transcribe into TEs, which indicates that Pol V transcription

may be enhanced by the proximity of euchromatin and heterochromatin, which could further contribute to precise determination of TE boundaries.

Our observations that MET1 and CMT3 are needed for elevated Pol V transcription at certain loci suggest that RdDM is efficiently maintained only if DNA methylation is above a certain threshold level. Loci where RdDM is capable of maintaining DNA methylation above this threshold may be silenced exclusively by RdDM. However, loci where RdDM cannot maintain DNA methylation above the threshold require at least one other silencing pathway for efficient silencing. The basis of this threshold mechanism remains unknown, however it is likely to integrate the level of Pol V transcription and the amount and properties of siRNA. This possibility is supported by the observation that tethering Pol V to the *FWA* locus leads to increased levels of DNA methylation [227]. The mechanism of threshold is also likely to be controlled by a balance between DNA methylation and demethylation [228]. The existence of such a threshold would be particularly important in *de novo* silencing as it would prevent inadvertent silencing of essential genes by low amounts of siRNA.

3.5 Materials and Methods

3.5.1 Reagents

The antibody against the largest subunit of Pol V (NRPE1) was described previously [167], [199], [229].

3.5.2. Biological Resources

We used the following genotypes of *Arabidopsis thaliana*: Columbia-0 ecotype (wildtype), *nrpe1* (*nrpd1b-11* [110]), *ago4-1* (introgressed into the Col-0 background [198]), *spt5l* (SALK_001254),

drm2-2 (SAIL_70_E12), *cmt3-11* (SALK_148381), and *met1-3* [180]. Plants were grown at 22ºC under white LED light in 16h/8h day/night cycle.

3.5.3 Computational resources

During data analysis we used bowtie2 2.2.9 [167], BEDTools 2.15.0 [186], the NBPseq R package [184], GFOLD [219], Bismark [187] and methylKit R package [188]. *Arabidopsis* genome annotations (TAIR10) were obtained from TAIR (www.arabidopsis.org). Previously published high throughput sequencing datasets were obtained from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/). Pol V IPARE data (GSE146913) and annotated regions were published previously [167]. TE regions annotated by RdDM categories were provided by the Slotkin lab [123], [222]. DNA methylation data (GSE39901) were obtained from previously published datasets [218].

3.5.4 Statistical Analyses

Significant differences in the levels of Pol V transcription were identified using Robinson and Smyth's exact negative binomial test implemented in the NBPseq R package [184] using data from two independent biological replicates. For the *met1* mutant significant differences in the levels of Pol V transcription were identified using generalized fold change algorithm implemented in GFOLD [219]. Levels of DNA methylation or Pol V transcription on groups of genomic bins were compared using the Wilcoxon test.

3.5.5 Pol V IPARE

Three grams of aerial tissue of 18-day old plants were used for Pol V IPARE experiments carried out as described [167]. High throughput sequencing was performed at the University of Michigan Advanced Genomics Core.

3.5.6 Bioinformatic Analysis

Pol V IPARE sequencing reads were processed and aligned to the *Arabidopsis* TAIR10 genome with bowtie2 as described previously [167]. Pol V IPARE levels were plotted as boxplots by counting the number of reads in studied genomic regions using BEDTools and normalized as number of reads per million mapped reads (RPM) [186]. Information about IPARE datasets generated and used in this study is presented in Table S1.

To identify differentially transcribed regions by Pol V, we counted the number of IPARE reads in 100 bp bins with a step-size of 50 bp across the whole genome. We then tested for differential Pol V transcription in the bins between Col-0 and specific mutants with false discover rate (FDR) < 0.04 using NBPseq [184]. Overlap analyses between Pol V IPARE reduced in *drm2* regions and specific genomic regions (Fig. 6A) were performed with 1000 permutated genomic regions using BEDTools to obtain expected numbers and p-values [186]. TE ends were defined as 150 bp at the end of TEs and TE inner are the remainders of annotated TEs. Average profiles of Pol V IPARE signal at ends of Pol V RdDM TEs with lengths of more than 500 bp, were plotted with Col-0 divided by *drm2*. Reductions of Pol V transcription in *drm2*, *spt5l*, *ago4* and *cmt3* mutants was determined by FDR < 0.05. Pol V transcription was determined to be unchanged if FDR was greater than 0.9 and fold change smaller than 2. Reduction of Pol V transcription in *met1*, which was based on one replicate of Pol V IPARE was determined using GFOLD [219] with the

$p < 0.01$ at 2-fold change or greater. Pol V transcription was determined to be unchanged in *met1* if $p < 0.01$ at 0.1-fold change or smaller and fold change smaller than 2.

Sequencing reads from whole genome bisulfite-seq datasets were mapped to the *Arabidopsis* TAIR10 genome using Bismark allowing no mismatches [187]. DNA methylation levels were calculated by the ratio of #C/(#C+#T) after selecting for Cs with at least 5 sequenced reads. Differentially Methylated Regions (DMRs) were identified using methylKit [188]. The bin sizes used were 100 bp bins with a step-size of 50 bp. A minimum of 10 bases was required in each bin. For *drm1/2* DMRs, 25% minimum difference in CHH context DNA methylation was selected for in each of the tiles with FDR < 0.01. For *met1* DMRs, 55% minimum difference in CG context DNA methylation was selected for in each of the tiles with FDR < 0.01. DNA methylation levels used as the cutoff for presence of each context in the DNA methylation categories in Figures 2-5 were 5% CHH, 10% CHG and 20% CG. DNA methylation levels used as the cutoff for absence of each context in the DNA methylation categories in Figures 2-5 were 0% CHH, 0% CHG and 0% CG in the respective mutant that was tested.

Datasets	**Exp. group**	**GEO acc.**	**Total reads**	**reads post-trimming**	**mapped reads**	**deduplicated reads**	**nuclear**
Col-0 IPARE	1	GSM4409524	Described previously in Tsuzuki *et al*. (2020)				
nrpe1 IPARE	1	GSM4409525	Described previously in Tsuzuki *et al*. (2020)				
spt5l IPARE	1	GSM4409526	Described previously in Tsuzuki *et al*. (2020)				
cmt3 IPARE	1	GSM5171710	15689618	5782750	3094943	2743344	2540409

Col-0 IPARE	2	GSM4409529	Described previously in Tsuzuki *et al*. (2020)				
ago4 IPARE	2	GSM4409530	Described previously in Tsuzuki *et al*. (2020)				
drm2 IPARE	2	GSM5171711	16778574	8329297	4508122	3475878	3114715
met1-3 IPARE	2	GSM5171712	15883353	7977408	4302897	2795337	2479618
Col-0 IPARE	3	GSM4409533	Described previously in Tsuzuki *et al*. (2020)				
nrpe1 IPARE	3	GSM4409534	Described previously in Tsuzuki *et al*. (2020)				
spt5l IPARE	3	GSM4409535	Described previously in Tsuzuki *et al*. (2020)				
ago4 IPARE	3	GSM4409536	Described previously in Tsuzuki *et al*. (2020)				
drm2 IPARE	3	GSM5171713	15851954	9500166	4382140	2946804	2514335
cmt3 IPARE	3	GSM5171714	16817100	9007012	3884136	2922802	2599268

Table 3.1: High throughput sequencing datasets obtained in this study. Experimental groups correspond to datasets generated in parallel from plants grown at the same time.

3.6 Acknowledgements

3.6.1 Author Contributions

M Hafiz Rothi: Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, United States

Contributions: MHR, Conception and design, Analysis and interpretation of data, Drafting or revising the article.

Masayuki Tsuzuki: Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, United States

Contributions: MS, Conception and design, Acquisition of data, Drafting or revising the article.

Shriya Sethuraman: Bioinformatics Graduate Program, University of Michigan, Ann Arbor, United States

Contributions: SS, Analysis and interpretation of data, Drafting or revising the article.

Andrzej T Wierzbicki: Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, United States

Contributions: ATW, Conception and design, Analysis and interpretation of data, Drafting or revising the article.

3.6.2 Funding

This work was supported by National Institute of General Medical Sciences of the National Institutes of Health [R01GM108722 to ATW] and Japan Society for the Promotion of Science [Overseas Research Fellowship 201860706 to MT and KAKENHI 19K23715 to MT].

3.6.3 Data and materials availability

High throughput sequencing datasets obtained in this study are available in Gene Expression Omnibus under accession number GSE168869.

CHAPTER IV

Conclusions

4.0.1 Introduction

Cytosine DNA methylation is a chromatin modification which has been extensively studied in respect to its function in gene silencing and the formation of a repressive chromatin state [87]. In addition, the biochemical processes involved in writing, reading, and erasing DNA methylation has been well elucidated in various systems [230]. Despite this, the mechanisms by which DNA methylation carries out its role in gene silencing is still poorly understood. This is due to the existence of many different methyl readers that can recognize and bind to the mark leading to different types of effects [231]. Moreover, in certain organisms such as plants, DNA can be methylated in different contexts which can be bound by methyl readers that possess a higher affinity towards a specific context adding another layer of complexity [232].

Plants utilize DNA methylation mainly for gene silencing and it can be deemed indispensable considering the existence of multiple DNA methylation pathways that provide a high level of redundancy and robustness [87]. This may be due the substantial amount of TEs that are still present in plant genomes which constantly poses a tremendous risk if left unchecked [5].

Consequently, plants evolved to have one of the most sophisticated systems utilizing DNA methylation as a method of gene regulation in comparison to animals where often DNA methylation is either present in low amounts or completely absent [233], [234]. In addition, plants are also capable of faithfully passing on DNA methylation between generations making it a proper model for epigenetic studies [235]. Hence, plants serve as one of the best models to study and gain insights into the mechanistic aspects of DNA methylation.

DNA methylation has been shown to serve as a marker for various types of proteins to bind to and exert a specific downstream effect [90], [92], [140]. This implies that DNA methylation in part serves as recruitment factor and facilitates localization of methyl binding proteins within the same region. At the start of this work, the most well-established downstream feature involving DNA methylation in gene silencing has been with histone methyltransferases, namely SUVH4, SUVH5, and SUVH6, which possess SRA domains that allow it to bind to methylated DNA [78], [91], [120]. Their recruitment would facilitate the deposition of H3K9me2 marks within the same region [92], [120]. Previously, it was shown that IDN2 interacts with SWI3B, a component of the SWI/SNF chromatin remodeling complex, which led to the discovery that nucleosome positioning is a feature of RdDM [113]. However, it was not determined how DNA methylation and nucleosome positioning function in relation to each other to establish gene silencing.

It was also shown that SUVH2 and SUVH9, catalytically dead histone methyltransferases with SRA domains, are required for Pol V transcription which led to the proposal that a positive feedback mechanism may exist between DNA methylation and Pol V transcription [90]. This has been largely unconfirmed due the fact that studies which tested the level of Pol V transcripts in mutants downstream of Pol V transcription reported little to no change [111], [134], [136], [148]. Hence, the existence of a positive feedback mechanism involving DNA methylation remains

unresolved. The goal of my research has been to expand our understanding of the function and mechanisms involving DNA methylation in RNA-mediated transcriptional silencing.

4.0.2 Findings

The RdDM pathway functions as a gene silencing pathway by directly modifying the chromatin [124]. This occurs through the addition of DNA methylation, H3K9me2 and nucleosome positioning events [124]. The relationship between DNA methylation and H3K9me2 pathways has been well established [78], [91], [92]. In contrast, little is known about the relationship between DNA methylation and nucleosome positioning in the context of RdDM. We added more insight into this phenomenon in Chapter II. First, we wanted to confirm that nucleosome positions were affected with the loss of Pol V because the previous study had used a non-specific protection assay with MNase to conclude that Pol V is positioning nucleosomes. Indeed, using MNase coupled with H3 ChIP in two biological replicates, we were able to identify 690 nucleosomes as Pol V stabilized nucleosomes. In addition, we found that these nucleosomes were enriched in regions similar to those by RdDM such as TEs, intergenic regions and gene promoter regions.

The only RdDM factor that SWI3B interacts with is IDN2 which means that it is possible that nucleosome positioning can occur directly through IDN2's association with Pol V transcripts and independent of other factors such as AGO4 and DRM2 [113]. However, when we measured nucleosome signals at Pol V stabilized nucleosomes in Col-0, *nrpe1*, *ago4* and *idn2*, we observed that nucleosome positioning requires AGO4 and that Pol V stabilized nucleosomes are enriched with DNA methylation and H3K9me2. This suggested that the entire RdDM pathway may be required for nucleosome positioning to occur which means that DNA methylation or nucleosome positioning could be important for the presence of the other. We tested two potential scenarios,

either nucleosome positioning is required for DNA methylation or DNA methylation is required for nucleosome positioning. We measured the DNA methylation levels at Pol V and SWIB stabilized nucleosomes in Col-0, *nrpe1* and *swi3b/+* and found that DNA methylation is largely unchanged between Col-0 and *swi3b/+*. Subsequently we measured the nucleosome signal levels in Col-0 and *drm2* at *drm2* DMRs and saw a substantial loss of signal. Taken together, we concluded that DNA methylation functions upstream of nucleosome positioning in RdDM and is required for nucleosome positioning at target regions.

Initiation of Pol V transcription is still a poorly understood aspect in RdDM [104]. With recent work proposing that Pol V has potentially up to three different modes of transcription, namely surveillance, transition, and maintenance, understanding Pol V initiation in each scenario become even more pertinent [147]. In our work in Chapter III, we successfully teased apart the factors involved in the proposed positive feedback mechanisms to demonstrate that it does exist and more importantly that it has strong crosstalk with other silencing pathways. We showed using Pol V IPARE analysis that Pol V transcript levels remain stable in downstream mutants at RdDM annotated regions. This was shown to be caused by residual DNA methylation that still remain in other contexts that were maintained by other silencing pathways. Next, we showed that if you filtered *drm2* DMRs for the presence of CG and CHG contexts, Pol V transcript levels decrease in the *drm2* mutant. This analysis revealed that if a silenced region depends entirely on RdDM for DNA methylation in all contexts, Pol V transcription levels will be affected which suggests that the DNA methylation deposited by RdDM enhances Pol V transcription, which constitutes a positive feedback loop. In addition, we showed that regions were DNA methylation was maintained by other silencing pathways including MET1 and CMT3 were also transcribed by Pol V. In both *met1* and *cmt3* DMRs, when a loss of methylation in all contexts occurs, this leads to

the significant decrease in Pol V transcription. However, this happens in only a subset of loci in *met1* and *cmt3* DMRs because RdDM has small overlaps in regions maintained by other silencing pathways.

It was shown that Pol V transcription determines the edges of heterochromatin, in particular at the ends of long transposons [165]. We wanted to test if a positive feedback mechanism could be involved in the strong enrichment of Pol V transcription at these regions. Indeed, using classification analysis, DRM2-dependent Pol V transcription regions were enriched in TEs, especially at the ends of TEs. When we measured Pol V transcript levels at the ends of TEs, we observed a substantial decrease in Pol V transcript levels in *drm2* compared to Col-0, which indicates that enhanced levels of Pol V transcription are due to a positive feedback mechanism between DNA methylation and Pol V transcription. Although, CHH and CHG methylation were significantly reduced in *drm2* in these regions, CG methylation were largely unaffected, which is consistent with the mark being maintained by MET1 and therefore Pol V transcription may still occur to a lesser extent. Taken together, we conclude that a positive feedback loop between DNA methylation and Pol V transcription reinforces transcriptional silencing leading to maintenance of silencing.

4.0.3 Implications

Our work described in Chapter II has provided more insight into the complex relationship between DNA methylation and nucleosomes. Previous studies have attempted to characterize the interaction between DNA methylation and nucleosomes in both *in vitro* and *in vivo* conditions. It was shown using MNase-seq in *Arabidopsis thaliana* that nucleosomal DNA is more highly methylated compared to linker DNA [143]. This finding raised speculation that nucleosomes can

shape the methylation landscape in genomes. Interestingly, *in vitro* studies examining how DNA modifications affect the ability of DNA to wrap around nucleosome showed that DNA methylation causes DNA to become less flexible and decreases the mechanical stability to keep it wrapped around nucleosomes [145]. This suggests that methylated DNA is not a naturally favorable region for nucleosomes bind and would require additional factors to overcome this obstacle. Our work has shown that in the case of RdDM, DNA methylation is affecting nucleosome positioning. In addition, nucleosomal DNA at positioned nucleosomes are enriched with DNA methylation instead of linker DNA. Based on the *in vitro* study, we would predict that methylated DNA positions DNA by creating pockets of unmethylated regions flanked by methylated DNA which will give nucleosomes a perfect landing spot to position. This does not seem to be the case in our finding which suggests that *in vivo* methylated DNA can stabilize nucleosomes within its region.

It has been shown in another study that in general, linker DNA tends to be methylated in comparison to nucleosomal DNA and it was proposed that nucleosome inhibit the activity of DNA methyltransferases [146]. Moreover they demonstrated that factors such as DDM1 and Lsh are required to remodel chromatin and provide DNA methyltransferases access to nucleosomes in order for nucleosomal DNA to be methylated [146]. In respect to our findings, we confirmed that the general trend is that linker regions tend to be more methylated than nucleosomal regions. However, we found that SWI/SNF was not required for wild type DNA methylation levels at Pol V stabilized nucleosomes which is inconsistent with the idea that SWI/SNF facilitates DRM2 to methylate DNA. Hence, at least in the case of RdDM, chromatin remodeling does not seem to precede DNA methylation.

In contrast, our findings show that DNA methylation seems to precede nucleosome positioning at DMRs which posits that DNA methylation potentially acts as a recruiting factor for

nucleosomes positioning to occur at. A study in the human genome showed that highly methylated CpG islands tend to have high nucleosome occupancy as well [144]. This effect was strongly associated with silencing mechanisms present in the system. Increased compaction of DNA to make it inaccessible to transcription factors is an intuitive concept where nucleosomes form a physical barrier to Pol II transcription [236], [237]. Hence, our work points towards RdDM directing nucleosomes to be positioned at target regions as a way of creating this barrier. However, this also potentially presents barrier to the transcriptional gene silencing polymerases as well. Pol V has been shown to require upstream chromatin remodelers for transcription although current evidence is still lacking in terms of their biochemical activity [132]. Thus, it may be possible that Pol V stabilized nucleosome possess a dual purpose of blocking Pol II but can be recognized by Pol V transcription factors.

In Chapter III, we addressed a long-standing issue regarding the model of a positive feedback mechanism between Pol V transcription and DNA methylation. It was shown that Pol V transcription required the presence of SUVH2 and SUVH9 [90]. In addition, SUVH2 and SUVH9 possess an SRA domain which allows it to bind to methylated DNA [238, p. 2]. Hence, it was proposed that DNA methylation that was deposited through RdDM could feed back into the pathway by recruitment of Pol V to enhance transcription within the region. However, studies that looked at Pol V transcription levels in mutants downstream of Pol V transcription did not observe a substantial decrease in Pol V transcript levels [111], [134], [136], [147], [148], [239]. Our work shows that DNA methylation enhances Pol V transcription to reinforce transcriptional gene silencing at target regions.

This finding shares parallels to the TGS system in *Schizosaccharomyces pombe* that utilizes H3K9 methylation as the repressive mark deposited by the pathway instead of DNA

methylation which is absent in yeasts [240], [241]. In *S. pombe*, RNAi driven by *cen* RNA, originating from centromeric repeat regions, initiate association of the RNA Induced Transcriptional Silencing (RITS) complex to chromatin at target regions [94]. Subsequently, a histone methyltransferase known as *clr4* is recruited to these regions where H3K9me marks will be deposited [242]. H3K9me marks are then used to recruit the RITS complex back through the chromodomain of *chp1*, a component of the complex [242], [243]. Further studies looking at the feedback mechanism found that it is important for maintaining silencing at target regions which mediates transgenerational epigenetic inheritance of silenced alleles [244]. Consistent with our finding that RdDM uses the chromatin mark that it deposits, DNA methylation, as a factor for recruiting Pol V back for transcription, this demonstrates that the mechanisms for maintaining gene silencing have evolved to work in a relatively similar manner despite the different type of chromatin modification used in organisms from different kingdoms.

4.0.4 Limitations

This work is centered around a plant-specific transcriptional gene silencing pathway. However, our findings are likely applicable in other organisms that possess and utilizes DNA methylation for gene regulation. In addition, most organisms contain TEs in their genomes where gene silencing pathways are likely present to control TE activity [5]. A limitation with using *Arabidopsis thaliana* in this work is that the level of TE content is relatively low at approximately 10% in comparison to other plants such as maize where it can be up to 84% [5]. Hence, it is possible that transcriptional gene silencing may work in a different manner in other organisms and that some findings may not be applicable as well. However, it is also due to this limitation that mutants in RdDM can be studied where the phenotypes are less severe and allow us to gain more

insight into mechanisms pertaining to chromatin modifications as well as noncoding RNA in the case of Pol IV and Pol V [106].

In our work, we utilize different types of assays to measure genome-wide changes through high-throughput sequencing. For instance, we use the bisulfite conversion method to measure DNA methylation levels and antibody-based methods such as ChIP and RIP to measure enrichment of DNA or RNA associated with the protein. A limitation of these different approaches is that each assay has a certain level of sensitivity. Hence, when we cross reference different types of datasets it is possible that we are unable to observe strong overlaps between them which only allows limited interpretation of the result. In addition, genome-wide approaches mostly capture general trends that occur in the cell. Therefore, locus-specific effects can show varying levels of behavior that either match the general trend or is completely different. This limitation is taken into account by applying statistical analysis whenever genome-wide observations are made and further confirmed with locus-specific validation whenever applicable.

As a pathway that is mainly responsible for silencing new insertions of TEs, RdDM is likely to have tissue specific activity related to reactivation of TEs [121], [122]. However, our work uses whole aerial tissues in assays which largely neglects this aspect of RdDM. This limits our observations to an amalgamated picture of events from different tissues that likely have varying levels of RdDM activity. Hence, our observations may be diluted or weakened to the point where accurate interpretations become more difficult. A major limitation with working on SWI/SNF chromatin remodeling complex in plants is that the biochemical activity of this complex is predicated on knowledge established in other organisms [245], [246]. Although there have been many genetic studies done in plants involving SWI/SNF complex, we still do not know if these remodelers function similarly to their homologues in organisms such as yeast and mammals [156],

[247], [248]. Hence, interpretation of results implying direct activity such as nucleosome positioning is difficult to establish with the limited knowledge of plant chromatin remodeler biochemistry.

4.0.5 Future Directions

Work in described in Chapter II provides a novel paradigm where DNA methylation affects nucleosome positioning in RNA-mediated gene silencing. Other studies have reported that nucleosomal DNA can have highly methylated DNA, however, the mechanism by which this situation happens is not fully understood [143], [144], [249]. We still do not know if methylated DNA directly interacts with the nucleosomes for positioning or if there are methyl readers that act as intermediate factors to facilitate this. Recently, new methyl readers were identified to be associated with RdDM silencing involving molecular chaperone proteins [140]. Hence, there could still be many more unknown methyl readers that could be involved in nucleosome positioning as well. In addition, we still do not fully understand if DNA methylation needs to be present in a specific manner or if there is a threshold level before nucleosomes will be positioned. Recently, there have been multiple studies reporting clustered regularly interspaced short palindromic repeats (CRISPR) based tools that can target specific regions for DNA methylation [250], [251]. It would be interesting to directly test in the future if DNA methylation can direct nucleosome positioning using these tools.

Although nucleosome positioning by RdDM has been associated with gene silencing, we still do not have a clear idea of what elements these nucleosomes are protecting [113], [141]. Our current understanding of cis-elements and transcription factor binding sites in plants is still quite limited [252], [253]. In the context of TE silencing, nucleosome may be positioned at sequences

that can be recognized by Pol II associated transcription factors. A comprehensive sequence motif analysis could provide more insight into the critical elements where nucleosomes are positioned. Future work could also involve Pol II ChIP-seq assays to determine if Pol II occupancy levels increase in the absence of nucleosomes positioned by RdDM.

It was shown that RdDM is involved in inhibiting chromatin looping [254]. In Chapter II, we show that Pol V stabilized nucleosomes had a relatively small overlap with Pol V transcribed regions. It is possible that this is due to nucleosome positioning occurring at a region where RdDM is affecting it from a long-range distance. In mammalian systems, it was shown that DNA methylation could affect the presence of CTCF through the level of nucleosome occupancy [255]. It would be interesting to see if high resolution chromatin contact maps in DNA methyltransferase mutants show an increase of looping due to the loss of methylation.

Pol II initiates transcription by binding to promoter regions which contain specific sequences recognized by transcription factors [256]. However, this does not seem to be the case for Pol V [190], [257]. Attempts to identify potential promoter sequences for Pol V have not yielded any conclusive results and it has been speculated that Pol V uses internal promoters to initiate transcription [165], [190]. In addition, a new model has been proposed where Pol V can undergo up to three modes of transcription beginning from surveillance, then transition and finally into a maintenance state [147]. We show in this work that at least in the maintenance state, Pol V transcription can be initiated through DNA methylation marks that are either deposited by RdDM itself as a positive feedback mechanism or DNA methylation that is deposited by other silencing pathways that share a region with RdDM. It would be interesting to look at factors that are needed during the transitionary stage for Pol V initiation. DNA methylation is likely one of them however it does not seem to be enough for transition to occur [147]. It is possible that the addition or removal

of histone modifications or changes in histone variants could also contribute to transitioning from the surveillance state. Future work in this area should be focused on identifying which of these factors in addition to DNA methylation could lead to Pol V transcription transition phase.

4.0.6 Concluding Remarks

This research work has generated many datasets pertaining to DNA methylation, nucleosome occupancy, and Pol V transcripts that have been deposited into public repositories where some have been made public and some which hopefully will be made public once they are accepted for publication. There are still many interesting biological questions that we have not fully explored which can be potentially answered using these datasets. It is our hope that these datasets will be beneficial and facilitate future discoveries by researchers in the field and the entire scientific community.

We aimed to expand our understanding and knowledge of the function and mechanisms involving DNA methylation in RNA-mediated transcriptional silencing. We have shown that DNA methylation affects nucleosome positioning in RdDM and that DNA methylation can enhance Pol V transcription through a positive feedback loop. These novel mechanistic insights have opened up more possible processes by which a simple DNA modification can take part in to carry out its role in gene silencing.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] J. D. Watson and F. H. C. Crick, "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid," *Nature*, vol. 171, no. 4356, Art. no. 4356, Apr. 1953, doi: 10.1038/171737a0.

[2] G. W. Beadle and E. L. Tatum, "Genetic Control of Biochemical Reactions in Neurospora," *PNAS*, vol. 27, no. 11, pp. 499–506, Nov. 1941, doi: 10.1073/pnas.27.11.499.

[3] E. S. Lander *et al.*, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, Art. no. 6822, Feb. 2001, doi: 10.1038/35057062.

[4] J. C. Venter *et al.*, "The Sequence of the Human Genome," *Science*, vol. 291, no. 5507, pp. 1304–1351, Feb. 2001, doi: 10.1126/science.1058040.

[5] L. Guio and J. González, "New Insights on the Evolution of Genome Content: Population Dynamics of Transposable Elements in Flies and Humans," in *Evolutionary Genomics: Statistical and Computational Methods*, M. Anisimova, Ed. New York, NY: Springer, 2019, pp. 505–530. doi: 10.1007/978-1-4939-9074-0_16.

[6] B. McClintock, "The origin and behavior of mutable loci in maize," *PNAS*, vol. 36, no. 6, pp. 344–355, Jun. 1950, doi: 10.1073/pnas.36.6.344.

[7] A. Hua-Van, A. Le Rouzic, T. S. Boutin, J. Filée, and P. Capy, "The struggle for life of the genome's selfish architects," *Biology Direct*, vol. 6, no. 1, p. 19, Mar. 2011, doi: 10.1186/1745-6150-6-19.

[8] D. J. Finnegan, "Eukaryotic transposable elements and genome evolution," *Trends in Genetics*, vol. 5, pp. 103–107, Jan. 1989, doi: 10.1016/0168-9525(89)90039-5.

[9] T. Wicker *et al.*, "A unified classification system for eukaryotic transposable elements," *Nature Reviews Genetics*, vol. 8, no. 12, Art. no. 12, Dec. 2007, doi: 10.1038/nrg2165.

[10] Z. Joly-Lopez and T. E. Bureau, "Diversity and evolution of transposable elements in Arabidopsis," *Chromosome Res*, vol. 22, no. 2, pp. 203–216, Jun. 2014, doi: 10.1007/s10577-014-9418-8.

[11] D. Lisch, "How important are transposons for plant evolution?," *Nat Rev Genet*, vol. 14, no. 1, pp. 49–61, Jan. 2013, doi: 10.1038/nrg3374.

[12] H. Candela and S. Hake, "The art and design of genetic screens: maize," *Nat Rev Genet*, vol. 9, no. 3, pp. 192–203, Mar. 2008, doi: 10.1038/nrg2291.

[13] Y.-I. Hsing *et al.*, "A rice gene activation/knockout mutant resource for high throughput functional genomics," *Plant Mol Biol*, vol. 63, no. 3, pp. 351–364, Feb. 2007, doi: 10.1007/s11103-006-9093-z.

[14] A. M. Settles *et al.*, "Sequence-indexed mutations in maize using the UniformMu transposon-tagging population," *BMC Genomics*, vol. 8, no. 1, p. 116, May 2007, doi: 10.1186/1471-2164-8-116.

[15] S. Salvi *et al.*, "Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize," *PNAS*, vol. 104, no. 27, pp. 11376–11381, Jul. 2007.

[16] J. Piriyapongsa and I. K. Jordan, “A Family of Human MicroRNA Genes from Miniature Inverted-Repeat Transposable Elements,” *PLOS ONE*, vol. 2, no. 2, p. e203, Feb. 2007, doi: 10.1371/journal.pone.0000203.
[17] S. Mathews, “Phytochrome-mediated development in land plants: red light sensing evolves to meet the challenges of changing light environments,” *Molecular Ecology*, vol. 15, no. 12, pp. 3483–3503, 2006, doi: https://doi.org/10.1111/j.1365-294X.2006.03051.x.
[18] A. Hiltbrunner *et al.*, “Nuclear Accumulation of the Phytochrome A Photoreceptor Requires FHY1,” *Current Biology*, vol. 15, no. 23, pp. 2125–2130, Dec. 2005, doi: 10.1016/j.cub.2005.10.042.
[19] R. Lin, L. Ding, C. Casola, D. R. Ripoll, C. Feschotte, and H. Wang, “Transposase-Derived Transcription Factors Regulate Light Signaling in Arabidopsis,” *Science*, vol. 318, no. 5854, pp. 1302–1305, Nov. 2007, doi: 10.1126/science.1146281.
[20] M. E. Hudson, D. R. Lisch, and P. H. Quail, “The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway,” *The Plant Journal*, vol. 34, no. 4, pp. 453–471, 2003, doi: https://doi.org/10.1046/j.1365-313X.2003.01741.x.
[21] R. Lin and H. Wang, “Arabidopsis FHY3/FAR1 Gene Family and Distinct Roles of Its Members in Light Control of Arabidopsis Development,” *Plant Physiology*, vol. 136, no. 4, pp. 4010–4022, Dec. 2004, doi: 10.1104/pp.104.052191.
[22] D. Lisch, “Mutator transposons,” *Trends in Plant Science*, vol. 7, no. 11, pp. 498–504, Nov. 2002, doi: 10.1016/S1360-1385(02)02347-6.
[23] L. M. Payer and K. H. Burns, “Transposable elements in human genetic disease,” *Nature Reviews Genetics*, vol. 20, no. 12, Art. no. 12, Dec. 2019, doi: 10.1038/s41576-019-0165-8.
[24] Y. Miki *et al.*, “Disruption of the APC Gene by a Retrotransposal Insertion of L1 Sequence in a Colon Cancer,” *Cancer Res*, vol. 52, no. 3, pp. 643–645, Feb. 1992.
[25] E. C. Scott, E. J. Gardner, A. Masood, N. T. Chuang, P. M. Vertino, and S. E. Devine, “A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer,” *Genome Res.*, vol. 26, no. 6, pp. 745–755, Jun. 2016, doi: 10.1101/gr.201814.115.
[26] C. Rodríguez-Martín *et al.*, “Familial retinoblastoma due to intronic LINE-1 insertion causes aberrant and noncanonical mRNA splicing of the RB1 gene,” *Journal of Human Genetics*, vol. 61, no. 5, Art. no. 5, May 2016, doi: 10.1038/jhg.2015.173.
[27] C. Napoli, C. Lemieux, and R. Jorgensen, “Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans.,” *The Plant Cell*, vol. 2, no. 4, pp. 279–289, Apr. 1990, doi: 10.1105/tpc.2.4.279.
[28] R. A. Jorgensen, P. D. Cluster, J. English, Q. Que, and C. A. Napoli, “Chalcone synthase cosuppression phenotypes in petunia flowers: comparison of sense vs. antisense constructs and single-copy vs. complex T-DNA sequences,” *Plant Mol Biol*, vol. 31, no. 5, pp. 957–973, Aug. 1996, doi: 10.1007/BF00040715.
[29] J. R. Ecker and R. W. Davis, “Inhibition of gene expression in plant cells by expression of antisense RNA,” *PNAS*, vol. 83, no. 15, pp. 5372–5376, Aug. 1986, doi: 10.1073/pnas.83.15.5372.
[30] S. N. Covey, N. S. Al-Kaff, A. Lángara, and D. S. Turner, “Plants combat infection by gene silencing,” *Nature*, vol. 385, no. 6619, Art. no. 6619, Feb. 1997, doi: 10.1038/385781a0.

[31] M. Metzlaff, M. O'Dell, P. D. Cluster, and R. B. Flavell, "RNA-Mediated RNA Degradation and Chalcone Synthase A Silencing in Petunia," *Cell*, vol. 88, no. 6, pp. 845–854, Mar. 1997, doi: 10.1016/S0092-8674(00)81930-3.
[32] M. H. Kumagai, J. Donson, G. della-Cioppa, D. Harvey, K. Hanley, and L. K. Grill, "Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA.," *PNAS*, vol. 92, no. 5, pp. 1679–1683, Feb. 1995, doi: 10.1073/pnas.92.5.1679.
[33] A. J. Hamilton and D. C. Baulcombe, "A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants," *Science*, vol. 286, no. 5441, pp. 950–952, Oct. 1999, doi: 10.1126/science.286.5441.950.
[34] A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello, "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans," *Nature*, vol. 391, no. 6669, Art. no. 6669, Feb. 1998, doi: 10.1038/35888.
[35] E. Bernstein, A. A. Caudy, S. M. Hammond, and G. J. Hannon, "Role for a bidentate ribonuclease in the initiation step of RNA interference," *Nature*, vol. 409, no. 6818, Art. no. 6818, Jan. 2001, doi: 10.1038/35053110.
[36] P. D. Zamore, T. Tuschl, P. A. Sharp, and D. P. Bartel, "RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals," *Cell*, vol. 101, no. 1, pp. 25–33, Mar. 2000, doi: 10.1016/S0092-8674(00)80620-0.
[37] S. M. Elbashir, J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, and T. Tuschl, "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells," *Nature*, vol. 411, no. 6836, Art. no. 6836, May 2001, doi: 10.1038/35078107.
[38] S. M. Hammond, E. Bernstein, D. Beach, and G. J. Hannon, "An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells," *Nature*, vol. 404, no. 6775, Art. no. 6775, Mar. 2000, doi: 10.1038/35005107.
[39] D. P. Bartel, "MicroRNAs: Genomics, Biogenesis, Mechanism, and Function," *Cell*, vol. 116, no. 2, pp. 281–297, Jan. 2004, doi: 10.1016/S0092-8674(04)00045-5.
[40] T. Treiber, N. Treiber, and G. Meister, "Regulation of microRNA biogenesis and its crosstalk with other cellular pathways," *Nature Reviews Molecular Cell Biology*, vol. 20, no. 1, Art. no. 1, Jan. 2019, doi: 10.1038/s41580-018-0059-1.
[41] H. R. Horvitz and J. E. Sulston, "Isolation and Genetic Characterization of Cell-Lineage Mutants of the Nematode Caenorhabditis Elegans," *Genetics*, vol. 96, no. 2, pp. 435–454, Oct. 1980.
[42] M. Chalfie, H. R. Horvitz, and J. E. Sulston, "Mutations that lead to reiterations in the cell lineages of C. elegans," *Cell*, vol. 24, no. 1, pp. 59–69, Apr. 1981, doi: 10.1016/0092-8674(81)90501-8.
[43] R. C. Lee, R. L. Feinbaum, and V. Ambros, "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14," *Cell*, vol. 75, no. 5, pp. 843–854, Dec. 1993, doi: 10.1016/0092-8674(93)90529-Y.
[44] B. Wightman, I. Ha, and G. Ruvkun, "Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans," *Cell*, vol. 75, no. 5, pp. 855–862, Dec. 1993, doi: 10.1016/0092-8674(93)90530-4.
[45] M. R. Fabian, N. Sonenberg, and W. Filipowicz, "Regulation of mRNA Translation and Stability by microRNAs," *Annu. Rev. Biochem.*, vol. 79, no. 1, pp. 351–379, Jun. 2010, doi: 10.1146/annurev-biochem-060308-103103.
[46] P. H. Olsen and V. Ambros, "The lin-4 Regulatory RNA Controls Developmental Timing in Caenorhabditis elegans by Blocking LIN-14 Protein Synthesis after the Initiation of

Translation," *Developmental Biology*, vol. 216, no. 2, pp. 671–680, Dec. 1999, doi: 10.1006/dbio.1999.9523.
[47] T. H. Beilharz *et al.*, "microRNA-Mediated Messenger RNA Deadenylation Contributes to Translational Repression in Mammalian Cells," *PLOS ONE*, vol. 4, no. 8, p. e6783, Aug. 2009, doi: 10.1371/journal.pone.0006783.
[48] L. Wu, J. Fan, and J. G. Belasco, "MicroRNAs direct rapid deadenylation of mRNA," *PNAS*, vol. 103, no. 11, pp. 4034–4039, Mar. 2006, doi: 10.1073/pnas.0510928103.
[49] J. G. Doench, C. P. Petersen, and P. A. Sharp, "siRNAs can function as miRNAs," *Genes Dev.*, vol. 17, no. 4, pp. 438–442, Feb. 2003, doi: 10.1101/gad.1064703.
[50] Y. Zeng, R. Yi, and B. R. Cullen, "MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms," *PNAS*, vol. 100, no. 17, pp. 9779–9784, Aug. 2003, doi: 10.1073/pnas.1630797100.
[51] M. C. Vella, E.-Y. Choi, S.-Y. Lin, K. Reinert, and F. J. Slack, "The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3′UTR," *Genes Dev.*, vol. 18, no. 2, pp. 132–137, Jan. 2004, doi: 10.1101/gad.1165404.
[52] B. P. Lewis, C. B. Burge, and D. P. Bartel, "Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets," *Cell*, vol. 120, no. 1, pp. 15–20, Jan. 2005, doi: 10.1016/j.cell.2004.12.035.
[53] Y. Wang, S. Juranek, H. Li, G. Sheng, T. Tuschl, and D. J. Patel, "Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex," *Nature*, vol. 456, no. 7224, Art. no. 7224, Dec. 2008, doi: 10.1038/nature07666.
[54] Y. Wang *et al.*, "Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes," *Nature*, vol. 461, no. 7265, Art. no. 7265, Oct. 2009, doi: 10.1038/nature08434.
[55] J. S. Parker, "How to slice: snapshots of Argonaute in action," *Silenc*, vol. 1, p. 3, Jan. 2010, doi: 10.1186/1758-907X-1-3.
[56] A. A. Millar and P. M. Waterhouse, "Plant and animal microRNAs: similarities and differences," *Funct Integr Genomics*, vol. 5, no. 3, pp. 129–135, Jul. 2005, doi: 10.1007/s10142-005-0145-2.
[57] C. Llave, Z. Xie, K. D. Kasschau, and J. C. Carrington, "Cleavage of Scarecrow-like mRNA Targets Directed by a Class of Arabidopsis miRNA," *Science*, vol. 297, no. 5589, pp. 2053–2056, Sep. 2002, doi: 10.1126/science.1076311.
[58] Y. Qiu *et al.*, "Flavivirus induces and antagonizes antiviral RNA interference in both mammals and mosquitoes," *Science Advances*, vol. 6, no. 6, p. eaax7989, Feb. 2020, doi: 10.1126/sciadv.aax7989.
[59] S.-W. Ding and O. Voinnet, "Antiviral Immunity Directed by Small RNAs," *Cell*, vol. 130, no. 3, pp. 413–426, Aug. 2007, doi: 10.1016/j.cell.2007.07.039.
[60] P. V. Maillard, A. G. van der Veen, E. Z. Poirier, and C. Reis e Sousa, "Slicing and dicing viruses: antiviral RNA interference in mammals," *The EMBO Journal*, vol. 38, no. 8, p. e100941, Apr. 2019, doi: 10.15252/embj.2018100941.
[61] H. Dana *et al.*, "Molecular Mechanisms and Biological Functions of siRNA," *Int J Biomed Sci*, vol. 13, no. 2, pp. 48–57, Jun. 2017.
[62] A. Cutter and J. J. Hayes, "A Brief Review of Nucleosome Structure," *FEBS Lett*, vol. 589, no. 20 0 0, pp. 2914–2922, Oct. 2015, doi: 10.1016/j.febslet.2015.05.016.

[63] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, “Crystal structure of the nucleosome core particle at 2.8 Å resolution,” *Nature*, vol. 389, no. 6648, pp. 251–260, Sep. 1997, doi: 10.1038/38444.
[64] R. D. Kornberg and J. O. Thonmas, “Chromatin Structure: Oligomers of the Histones,” *Science*, vol. 184, no. 4139, pp. 865–868, May 1974, doi: 10.1126/science.184.4139.865.
[65] W. Flemming, “Neue Beiträge zur Kenntniss der Zelle. II. Theil,” *Archiv f. mikrosk. Anat.*, vol. 37, no. 1, pp. 685–751, Dec. 1891, doi: 10.1007/BF02954311.
[66] E. Li, “Chromatin modification and epigenetic reprogramming in mammalian development,” *Nat Rev Genet*, vol. 3, no. 9, Art. no. 9, Sep. 2002, doi: 10.1038/nrg887.
[67] T. Kouzarides, “Chromatin Modifications and Their Function,” *Cell*, vol. 128, no. 4, pp. 693–705, Feb. 2007, doi: 10.1016/j.cell.2007.02.005.
[68] Q. Szabo, F. Bantignies, and G. Cavalli, “Principles of genome folding into topologically associating domains,” *Science Advances*, vol. 5, no. 4, p. eaaw1668, Apr. 2019, doi: 10.1126/sciadv.aaw1668.
[69] M. J. Rowley and V. G. Corces, “Organizational principles of 3D genome architecture,” *Nat Rev Genet*, vol. 19, no. 12, pp. 789–800, Dec. 2018, doi: 10.1038/s41576-018-0060-8.
[70] S. L. Klemm, Z. Shipony, and W. J. Greenleaf, “Chromatin accessibility and the regulatory epigenome,” *Nat Rev Genet*, vol. 20, no. 4, Art. no. 4, Apr. 2019, doi: 10.1038/s41576-018-0089-8.
[71] S. Kumar, V. Chinnusamy, and T. Mohapatra, “Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond,” *Front. Genet.*, vol. 9, 2018, doi: 10.3389/fgene.2018.00640.
[72] A. J. Bannister and T. Kouzarides, “Regulation of chromatin by histone modifications,” *Cell Res*, vol. 21, no. 3, Art. no. 3, Mar. 2011, doi: 10.1038/cr.2011.22.
[73] T. Jenuwein and C. D. Allis, “Translating the Histone Code,” *Science*, vol. 293, no. 5532, pp. 1074–1080, Aug. 2001, doi: 10.1126/science.1063127.
[74] T. A. Volpe, C. Kidner, I. M. Hall, G. Teng, S. I. S. Grewal, and R. A. Martienssen, “Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi,” *Science*, vol. 297, no. 5588, pp. 1833–1837, Sep. 2002, doi: 10.1126/science.1074973.
[75] Y. Jacob *et al.*, “ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing,” *Nature Structural & Molecular Biology*, vol. 16, no. 7, p. 763, Jul. 2009, doi: 10.1038/nsmb.1611.
[76] Y. Jacob *et al.*, “Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases,” *Nature*, vol. 466, no. 7309, p. 987, Aug. 2010, doi: 10.1038/nature09290.
[77] L. Lu, X. Chen, D. Sanders, S. Qian, and X. Zhong, “High-resolution mapping of H4K16 and H3K23 acetylation reveals conserved and unique distribution patterns in Arabidopsis and rice,” *Epigenetics*, vol. 10, no. 11, pp. 1044–1053, Nov. 2015, doi: 10.1080/15592294.2015.1104446.
[78] J. Du, L. M. Johnson, S. E. Jacobsen, and D. J. Patel, “DNA methylation pathways and their crosstalk with histone methylation,” *Nature Reviews Molecular Cell Biology*, vol. 16, no. 9, p. 519, Sep. 2015, doi: 10.1038/nrm4043.
[79] R. D. Hotchkiss, “THE QUANTITATIVE SEPARATION OF PURINES, PYRIMIDINES, AND NUCLEOSIDES BY PAPER CHROMATOGRAPHY,” *Journal of Biological Chemistry*, vol. 175, no. 1, pp. 315–332, Aug. 1948, doi: 10.1016/S0021-9258(18)57261-6.

[80] A. Razin and H. Cedar, “Distribution of 5-methylcytosine in chromatin,” *Proc Natl Acad Sci U S A*, vol. 74, no. 7, pp. 2725–2728, Jul. 1977, doi: 10.1073/pnas.74.7.2725.
[81] A. Razin and A. D. Riggs, “DNA methylation and gene function,” *Science*, vol. 210, no. 4470, pp. 604–610, Nov. 1980, doi: 10.1126/science.6254144.
[82] X. Cao and S. E. Jacobsen, “Role of the Arabidopsis DRM Methyltransferases in De Novo DNA Methylation and Gene Silencing,” *Current Biology*, vol. 12, no. 13, pp. 1138–1144, Jul. 2002, doi: 10.1016/S0960-9822(02)00925-9.
[83] Z. Gong, T. Morales-Ruiz, R. R. Ariza, T. Roldán-Arjona, L. David, and J. K. Zhu, “ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase,” *Cell*, vol. 111, no. 6, pp. 803–814, Dec. 2002, doi: 10.1016/s0092-8674(02)01133-9.
[84] R. M. Kohli and Y. Zhang, “TET enzymes, TDG and the dynamics of DNA demethylation,” *Nature*, vol. 502, no. 7472, pp. 472–479, Oct. 2013, doi: 10.1038/nature12750.
[85] M. Weber *et al.*, “Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome,” *Nat Genet*, vol. 39, no. 4, Art. no. 4, Apr. 2007, doi: 10.1038/ng1990.
[86] H. Zhang, Z. Lang, and J.-K. Zhu, “Dynamics and function of DNA methylation in plants,” *Nat Rev Mol Cell Biol*, vol. 19, no. 8, Art. no. 8, Aug. 2018, doi: 10.1038/s41580-018-0016-z.
[87] J. A. Law and S. E. Jacobsen, “Establishing, maintaining and modifying DNA methylation patterns in plants and animals,” *Nature Reviews Genetics*, vol. 11, no. 3, p. 204, Mar. 2010, doi: 10.1038/nrg2719.
[88] W. Reik, W. Dean, and J. Walter, “Epigenetic Reprogramming in Mammalian Development,” *Science*, vol. 293, no. 5532, pp. 1089–1093, Aug. 2001, doi: 10.1126/science.1063443.
[89] L. M. Johnson *et al.*, “The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation,” *Current Biology*, vol. 17, no. 4, pp. 379–384, Feb. 2007, doi: 10.1016/j.cub.2007.01.009.
[90] L. M. Johnson *et al.*, “SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation,” *Nature*, vol. 507, no. 7490, p. 124, Mar. 2014, doi: 10.1038/nature12931.
[91] C. I. Stoddard *et al.*, “A Nucleosome Bridging Mechanism for Activation of a Maintenance DNA Methyltransferase,” *Mol. Cell*, vol. 73, no. 1, pp. 73-83.e6, 03 2019, doi: 10.1016/j.molcel.2018.10.006.
[92] F. Malagnac, L. Bartee, and J. Bender, “An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation,” *The EMBO Journal*, vol. 21, no. 24, pp. 6842–6852, Dec. 2002, doi: 10.1093/emboj/cdf687.
[93] J. Du *et al.*, “Mechanism of DNA methylation-directed histone methylation by KRYPTONITE,” *Mol Cell*, vol. 55, no. 3, pp. 495–504, Aug. 2014, doi: 10.1016/j.molcel.2014.06.009.
[94] A. Verdel *et al.*, “RNAi-Mediated Targeting of Heterochromatin by the RITS Complex,” *Science*, vol. 303, no. 5658, pp. 672–676, Jan. 2004, doi: 10.1126/science.1093686.
[95] M. Wassenegger, S. Heimes, L. Riedel, and H. L. Sänger, “RNA-directed de novo methylation of genomic sequences in plants,” *Cell*, vol. 76, no. 3, pp. 567–576, Feb. 1994, doi: 10.1016/0092-8674(94)90119-8.

[96] M. F. Mette, J. van der Winden, M. A. Matzke, and A. J. M. Matzke, “Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters in trans,” *The EMBO Journal*, vol. 18, no. 1, pp. 241–248, Jan. 1999, doi: 10.1093/emboj/18.1.241.
[97] M. Pal-Bhadra, U. Bhadra, and J. A. Birchler, “Cosuppression of Nonhomologous Transgenes in Drosophila Involves Mutually Related Endogenous Sequences,” *Cell*, vol. 99, no. 1, pp. 35–46, Oct. 1999, doi: 10.1016/S0092-8674(00)80060-4.
[98] A. Piunti and A. Shilatifard, “The roles of Polycomb repressive complexes in mammalian development and cancer,” *Nat Rev Mol Cell Biol*, vol. 22, no. 5, Art. no. 5, May 2021, doi: 10.1038/s41580-021-00341-1.
[99] A. A. Aravin, N. M. Naumova, A. V. Tulin, V. V. Vagin, Y. M. Rozovsky, and V. A. Gvozdev, “Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline,” *Current Biology*, vol. 11, no. 13, pp. 1017–1027, Jul. 2001, doi: 10.1016/S0960-9822(01)00299-8.
[100] A. Aravin *et al.*, “A novel class of small RNAs bind to MILI protein in mouse testes,” *Nature*, vol. 442, no. 7099, Art. no. 7099, Jul. 2006, doi: 10.1038/nature04916.
[101] J. Brennecke *et al.*, “Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila,” *Cell*, vol. 128, no. 6, pp. 1089–1103, Mar. 2007, doi: 10.1016/j.cell.2007.01.043.
[102] A. A. Aravin, G. J. Hannon, and J. Brennecke, “The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race,” *Science*, vol. 318, no. 5851, pp. 761–764, Nov. 2007, doi: 10.1126/science.1146484.
[103] A. A. Aravin *et al.*, “A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice,” *Molecular Cell*, vol. 31, no. 6, pp. 785–799, Sep. 2008, doi: 10.1016/j.molcel.2008.09.003.
[104] M. A. Matzke and R. A. Mosher, “RNA-directed DNA methylation: an epigenetic pathway of increasing complexity,” *Nature Reviews Genetics*, vol. 15, no. 6, p. 394, Jun. 2014, doi: 10.1038/nrg3683.
[105] K. V. Morris, S. W.-L. Chan, S. E. Jacobsen, and D. J. Looney, “Small Interfering RNA-Induced Transcriptional Gene Silencing in Human Cells,” *Science*, vol. 305, no. 5688, pp. 1289–1292, Aug. 2004, doi: 10.1126/science.1101372.
[106] T. S. Ream *et al.*, “Subunit Compositions of the RNA-Silencing Enzymes Pol IV and Pol V Reveal Their Origins as Specialized Forms of RNA Polymerase II,” *Molecular Cell*, vol. 33, no. 2, pp. 192–203, Jan. 2009, doi: 10.1016/j.molcel.2008.12.015.
[107] J. R. Haag *et al.*, “In Vitro Transcription Activities of Pol IV, Pol V, and RDR2 Reveal Coupling of Pol IV and RDR2 for dsRNA Synthesis in Plant RNA Silencing,” *Molecular Cell*, vol. 48, no. 5, pp. 811–818, Dec. 2012, doi: 10.1016/j.molcel.2012.09.027.
[108] Y. Chen, J. Weeks, M. A. Mortin, and A. L. Greenleaf, “Mapping mutations in genes encoding the two large subunits of Drosophila RNA polymerase II defines domains essential for basic transcription functions and for proper expression of developmental genes,” *Molecular and Cellular Biology*, vol. 13, no. 7, pp. 4214–4222, Jul. 1993, doi: 10.1128/mcb.13.7.4214-4222.1993.
[109] W. Powell and D. Reines, “Mutations in the Second Largest Subunit of RNA Polymerase II Cause 6-Azauracil Sensitivity in Yeast and Increased Transcriptional Arrest in Vitro(∗),” *Journal of Biological Chemistry*, vol. 271, no. 12, pp. 6866–6873, Mar. 1996, doi: 10.1074/jbc.271.12.6866.

[110] O. Pontes *et al.*, "The Arabidopsis Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center," *Cell*, vol. 126, no. 1, pp. 79–92, Jul. 2006, doi: 10.1016/j.cell.2006.05.031.
[111] A. T. Wierzbicki, J. R. Haag, and C. S. Pikaard, "Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes," *Cell*, vol. 135, no. 4, pp. 635–648, Nov. 2008, doi: 10.1016/j.cell.2008.09.035.
[112] H. Stroud, M. V. C. Greenberg, S. Feng, Y. V. Bernatavichute, and S. E. Jacobsen, "Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome," *Cell*, vol. 152, no. 1, pp. 352–364, Jan. 2013, doi: 10.1016/j.cell.2012.10.054.
[113] Y. Zhu, M. J. Rowley, G. Böhmdorfer, and A. T. Wierzbicki, "A SWI/SNF Chromatin-Remodeling Complex Acts in Noncoding RNA-Mediated Transcriptional Silencing," *Molecular Cell*, vol. 49, no. 2, pp. 298–309, Jan. 2013, doi: 10.1016/j.molcel.2012.11.011.
[114] T. Kawashima and F. Berger, "Epigenetic reprogramming in plant sexual reproduction," *Nat Rev Genet*, vol. 15, no. 9, Art. no. 9, Sep. 2014, doi: 10.1038/nrg3685.
[115] H. R. Woo, O. Pontes, C. S. Pikaard, and E. J. Richards, "VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization," *Genes Dev.*, vol. 21, no. 3, pp. 267–277, Feb. 2007, doi: 10.1101/gad.1512007.
[116] H. R. Woo, T. A. Dittmer, and E. J. Richards, "Three SRA-Domain Methylcytosine-Binding Proteins Cooperate to Maintain Global CpG Methylation and Epigenetic Silencing in Arabidopsis," *PLOS Genetics*, vol. 4, no. 8, p. e1000156, Aug. 2008, doi: 10.1371/journal.pgen.1000156.
[117] H. Stroud *et al.*, "Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*," *Nature Structural & Molecular Biology*, vol. 21, no. 1, p. 64, Jan. 2014, doi: 10.1038/nsmb.2735.
[118] M. L. Ebbs and J. Bender, "Locus-Specific Control of DNA Methylation by the Arabidopsis SUVH5 Histone Methyltransferase," *The Plant Cell*, vol. 18, no. 5, pp. 1166–1176, May 2006, doi: 10.1105/tpc.106.041400.
[119] J. P. Jackson, A. M. Lindroth, X. Cao, and S. E. Jacobsen, "Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase," *Nature*, vol. 416, no. 6880, p. 556, Apr. 2002, doi: 10.1038/nature731.
[120] J. Du *et al.*, "Dual Binding of Chromomethylase Domains to H3K9me2-Containing Nucleosomes Directs DNA Methylation in Plants," *Cell*, vol. 151, no. 1, pp. 167–180, Sep. 2012, doi: 10.1016/j.cell.2012.07.034.
[121] D. Cuerda-Gil and R. K. Slotkin, "Non-canonical RNA-directed DNA methylation," *Nature Plants*, vol. 2, no. 11, pp. 1–8, Nov. 2016, doi: 10.1038/nplants.2016.163.
[122] R. K. Slotkin *et al.*, "Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen," *Cell*, vol. 136, no. 3, pp. 461–472, Feb. 2009, doi: 10.1016/j.cell.2008.12.038.
[123] K. Panda, L. Ji, D. A. Neumann, J. Daron, R. J. Schmitz, and R. K. Slotkin, "Full-length autonomous transposable elements are preferentially targeted by expression-dependent forms of RNA-directed DNA methylation," *Genome Biology*, vol. 17, no. 1, p. 170, Aug. 2016, doi: 10.1186/s13059-016-1032-y.

[124] M. A. Matzke, T. Kanno, and A. J. M. Matzke, "RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants," *Annual Review of Plant Biology*, vol. 66, no. 1, pp. 243–267, 2015, doi: 10.1146/annurev-arplant-043014-114633.
[125] J. A. Law, A. A. Vashisht, J. A. Wohlschlegel, and S. E. Jacobsen, "SHH1, a Homeodomain Protein Required for DNA Methylation, As Well As RDR2, RDM4, and Chromatin Remodeling Factors, Associate with RNA Polymerase IV," *PLOS Genetics*, vol. 7, no. 7, p. e1002195, Jul. 2011, doi: 10.1371/journal.pgen.1002195.
[126] J. A. Law *et al.*, "Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1," *Nature*, vol. 498, no. 7454, p. 385, Jun. 2013, doi: 10.1038/nature12178.
[127] T. Blevins *et al.*, "Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis," *eLife Sciences*, vol. 4, p. e09591, Oct. 2015, doi: 10.7554/eLife.09591.
[128] J. Zhai *et al.*, "A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis," *Cell*, vol. 163, no. 2, pp. 445–455, Oct. 2015, doi: 10.1016/j.cell.2015.09.032.
[129] Z. Yang, Y. W. Ebright, B. Yu, and X. Chen, "HEN1 recognizes 21–24 nt small RNA duplexes and deposits a methyl group onto the 2′ OH of the 3′ terminal nucleotide," *Nucleic Acids Res*, vol. 34, no. 2, pp. 667–675, Jan. 2006, doi: 10.1093/nar/gkj474.
[130] L. Ji and X. Chen, "Regulation of small RNA stability: methylation and beyond," *Cell Research*, vol. 22, no. 4, p. 624, Apr. 2012, doi: 10.1038/cr.2012.36.
[131] R. Ye *et al.*, "Cytoplasmic Assembly and Selective Nuclear Import of Arabidopsis ARGONAUTE4/siRNA Complexes," *Molecular Cell*, vol. 46, no. 6, pp. 859–870, Jun. 2012, doi: 10.1016/j.molcel.2012.04.013.
[132] X. Zhong *et al.*, "DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons," *Nature Structural & Molecular Biology*, vol. 19, no. 9, p. 870, Sep. 2012, doi: 10.1038/nsmb.2354.
[133] Q. Zheng, M. J. Rowley, G. Böhmdorfer, D. Sandhu, B. D. Gregory, and A. T. Wierzbicki, "RNA polymerase V targets transcriptional silencing components to promoters of protein-coding genes," *Plant J*, vol. 73, no. 2, pp. 179–189, Jan. 2013, doi: 10.1111/tpj.12034.
[134] X.-J. He *et al.*, "An Effector of RNA-Directed DNA Methylation in Arabidopsis Is an ARGONAUTE 4- and RNA-Binding Protein," *Cell*, vol. 137, no. 3, pp. 498–508, May 2009, doi: 10.1016/j.cell.2009.04.028.
[135] M. J. Rowley, M. I. Avrutsky, C. J. Sifuentes, L. Pereira, and A. T. Wierzbicki, "Independent Chromatin Binding of ARGONAUTE4 and SPT5L/KTF1 Mediates Transcriptional Gene Silencing," *PLoS Genet*, vol. 7, no. 6, Jun. 2011, doi: 10.1371/journal.pgen.1002120.
[136] A. T. Wierzbicki, T. S. Ream, J. R. Haag, and C. S. Pikaard, "RNA polymerase V transcription guides ARGONAUTE4 to chromatin," *Nature Genetics*, vol. 41, no. 5, p. 630, May 2009, doi: 10.1038/ng.365.
[137] I. Ausin, T. C. Mockler, J. Chory, and S. E. Jacobsen, "IDN1 and IDN2 are required for *de novo* DNA methylation in *Arabidopsis thaliana*," *Nature Structural & Molecular Biology*, vol. 16, no. 12, p. 1325, Dec. 2009, doi: 10.1038/nsmb.1690.
[138] X. Cao and S. E. Jacobsen, "Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes," *PNAS*, vol. 99, no. suppl 4, pp. 16491–16498, Dec. 2002, doi: 10.1073/pnas.162371599.

[139] A. Janssen, S. U. Colmenares, and G. H. Karpen, “Heterochromatin: Guardian of the Genome,” *Annual Review of Cell and Developmental Biology*, vol. 34, no. 1, pp. 265–288, 2018, doi: 10.1146/annurev-cellbio-100617-062653.
[140] L. Ichino *et al.*, “MBD5 and MBD6 couple DNA methylation to gene silencing through the J-domain protein SILENZIO,” *Science*, vol. 372, no. 6549, pp. 1434–1439, Jun. 2021, doi: 10.1126/science.abg6130.
[141] Z.-W. Liu *et al.*, “Two Components of the RNA-Directed DNA Methylation Pathway Associate with MORC6 and Silence Loci Targeted by MORC6 in Arabidopsis,” *PLOS Genetics*, vol. 12, no. 5, p. e1006026, May 2016, doi: 10.1371/journal.pgen.1006026.
[142] T. Zhang, W. Zhang, and J. Jiang, “Genome-Wide Nucleosome Occupancy and Positioning and Their Impact on Gene Expression and Evolution in Plants,” *Plant Physiology*, vol. 168, no. 4, pp. 1406–1416, Aug. 2015, doi: 10.1104/pp.15.00125.
[143] R. K. Chodavarapu *et al.*, “Relationship between nucleosome positioning and DNA methylation,” *Nature*, vol. 466, no. 7304, p. 388, Jul. 2010, doi: 10.1038/nature09147.
[144] C. K. Collings and J. N. Anderson, “Links between DNA methylation and nucleosome occupancy in the human genome,” *Epigenetics & Chromatin*, vol. 10, no. 1, p. 18, Apr. 2017, doi: 10.1186/s13072-017-0125-5.
[145] T. T. M. Ngo *et al.*, “Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability,” *Nature Communications*, vol. 7, no. 1, Art. no. 1, Feb. 2016, doi: 10.1038/ncomms10813.
[146] D. B. Lyons and D. Zilberman, “DDM1 and Lsh remodelers allow methylation of DNA wrapped in nucleosomes,” *eLife Sciences*, vol. 6, p. e30674, Nov. 2017, doi: 10.7554/eLife.30674.
[147] M. Tsuzuki, S. Sethuraman, A. N. Coke, M. H. Rothi, A. P. Boyle, and A. T. Wierzbicki, “Broad noncoding transcription suggests genome surveillance by RNA polymerase V,” *PNAS*, Nov. 2020, doi: 10.1073/pnas.2014419117.
[148] G. Böhmdorfer, M. J. Rowley, J. Kuciński, Y. Zhu, I. Amies, and A. T. Wierzbicki, “RNA-directed DNA methylation requires stepwise binding of silencing factors to long non-coding RNA,” *Plant J*, vol. 79, no. 2, pp. 181–191, Jul. 2014, doi: 10.1111/tpj.12563.
[149] X.-J. He *et al.*, “A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development,” *Genes Dev.*, vol. 23, no. 23, pp. 2717–2722, Dec. 2009, doi: 10.1101/gad.1851809.
[150] M. F. Mette, W. Aufsatz, J. van der Winden, M. A. Matzke, and A. J. M. Matzke, “Transcriptional silencing and promoter methylation triggered by double-stranded RNA,” *The EMBO Journal*, vol. 19, no. 19, pp. 5194–5201, Oct. 2000, doi: 10.1093/emboj/19.19.5194.
[151] S. Lahmy *et al.*, “Evidence for ARGONAUTE4–DNA interactions in RNA-directed DNA methylation in plants,” *Genes Dev.*, vol. 30, no. 23, pp. 2565–2570, Dec. 2016, doi: 10.1101/gad.289553.116.
[152] D. Zilberman, X. Cao, and S. E. Jacobsen, “ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation,” *Science*, vol. 299, no. 5607, pp. 716–719, Jan. 2003, doi: 10.1126/science.1079695.
[153] J. Yang *et al.*, “SWI3B and HDA6 interact and are required for transposon silencing in Arabidopsis,” *The Plant Journal*, vol. 102, no. 4, pp. 809–822, 2020, doi: 10.1111/tpj.14666.

[154] J. Brzeski, W. Podstolski, K. Olczak, and A. Jerzmanowski, “Identification and analysis of the Arabidopsis thaliana BSH gene, a member of the SNF5 gene family,” *Nucleic Acids Res*, vol. 27, no. 11, pp. 2393–2399, Jun. 1999, doi: 10.1093/nar/27.11.2393.
[155] D. Wagner and E. M. Meyerowitz, “SPLAYED, a Novel SWI/SNF ATPase Homolog, Controls Reproductive Development in Arabidopsis,” *Current Biology*, vol. 12, no. 2, pp. 85–94, Jan. 2002, doi: 10.1016/S0960-9822(01)00651-0.
[156] T. J. Sarnowski *et al.*, “SWI3 Subunits of Putative SWI/SNF Chromatin-Remodeling Complexes Play Distinct Roles during Arabidopsis Development,” *The Plant Cell*, vol. 17, no. 9, pp. 2454–2472, Sep. 2005, doi: 10.1105/tpc.105.031203.
[157] S.-K. Han *et al.*, “The SWI2/SNF2 Chromatin Remodeling ATPase BRAHMA Represses Abscisic Acid Responses in the Absence of the Stress Stimulus in Arabidopsis,” *The Plant Cell*, vol. 24, no. 12, pp. 4892–4906, Dec. 2012, doi: 10.1105/tpc.112.105114.
[158] S. P. Sacharowski *et al.*, “SWP73 Subunits of Arabidopsis SWI/SNF Chromatin Remodeling Complexes Play Distinct Roles in Leaf and Flower Development,” *The Plant Cell*, vol. 27, no. 7, pp. 1889–1906, Jul. 2015, doi: 10.1105/tpc.15.00233.
[159] C. Li *et al.*, “Concerted genomic targeting of H3K27 demethylase REF6 and chromatin-remodeling ATPase BRM in *Arabidopsis*,” *Nature Genetics*, vol. 48, no. 6, p. 687, Jun. 2016, doi: 10.1038/ng.3555.
[160] R. Archacki *et al.*, “Arabidopsis SWI/SNF chromatin remodeling complex binds both promoters and terminators to regulate gene expression,” *Nucleic Acids Res*, vol. 45, no. 6, pp. 3116–3129, Apr. 2017, doi: 10.1093/nar/gkw1273.
[161] J. T. Huff and D. Zilberman, “Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes,” *Cell*, vol. 156, no. 6, pp. 1286–1297, Mar. 2014, doi: 10.1016/j.cell.2014.01.029.
[162] A. Zemach *et al.*, “The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin,” *Cell*, vol. 153, no. 1, pp. 193–205, Mar. 2013, doi: 10.1016/j.cell.2013.02.033.
[163] M. Felle, H. Hoffmeister, J. Rothammer, A. Fuchs, J. H. Exler, and G. Längst, “Nucleosomes protect DNA from DNA methylation in vivo and in vitro,” *Nucleic Acids Res*, vol. 39, no. 16, pp. 6956–6969, Sep. 2011, doi: 10.1093/nar/gkr263.
[164] C. K. Collings, P. J. Waddell, and J. N. Anderson, “Effects of DNA methylation on nucleosome stability,” *Nucleic Acids Res*, vol. 41, no. 5, pp. 2918–2931, Mar. 2013, doi: 10.1093/nar/gks893.
[165] G. Böhmdorfer *et al.*, “Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin,” *eLife Sciences*, vol. 5, p. e19092, Oct. 2016, doi: 10.7554/eLife.19092.
[166] A. T. Wierzbicki *et al.*, “Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome,” *Genes Dev.*, vol. 26, no. 16, pp. 1825–1836, Aug. 2012, doi: 10.1101/gad.197772.112.
[167] M. Tsuzuki, S. Sethuraman, A. N. Coke, M. H. Rothi, A. P. Boyle, and A. T. Wierzbicki, “Broad noncoding transcription suggests genome surveillance by RNA polymerase V,” *Proc Natl Acad Sci U S A*, vol. 117, no. 48, pp. 30799–30804, Dec. 2020, doi: 10.1073/pnas.2014419117.
[168] I. Ausin *et al.*, “INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in Arabidopsis,” *PNAS*, vol. 109, no. 22, pp. 8374–8381, May 2012, doi: 10.1073/pnas.1206638109.

[169] M. Xie, G. Ren, C. Zhang, and B. Yu, “The DNA- and RNA-binding protein FACTOR of DNA METHYLATION 1 requires XH domain-mediated complex formation for its function in RNA-directed DNA methylation,” *Plant J.*, vol. 72, no. 3, pp. 491–500, Nov. 2012, doi: 10.1111/j.1365-313X.2012.05092.x.
[170] T. J. Sarnowski, S. Świeżewski, K. Pawlikowska, S. Kaczanowski, and A. Jerzmanowski, “AtSWI3B, an Arabidopsis homolog of SWI3, a core subunit of yeast Swi/Snf chromatin remodeling complex, interacts with FCA, a regulator of flowering time,” *Nucleic Acids Res*, vol. 30, no. 15, pp. 3412–3421, Aug. 2002.
[171] R. Lister *et al.*, “Highly integrated single-base resolution maps of the epigenome in Arabidopsis,” *Cell*, vol. 133, no. 3, pp. 523–536, May 2008, doi: 10.1016/j.cell.2008.03.029.
[172] Y. Zhang *et al.*, “Large-scale comparative epigenomics reveals hierarchical regulation of non-CG methylation in Arabidopsis,” *PNAS*, vol. 115, no. 5, pp. E1069–E1074, Jan. 2018, doi: 10.1073/pnas.1716300115.
[173] C. Zhang *et al.*, “Arabidopsis AGDP1 links H3K9me2 to DNA methylation in heterochromatin,” *Nature Communications*, vol. 9, no. 1, Art. no. 1, Oct. 2018, doi: 10.1038/s41467-018-06965-w.
[174] M. El-Shami *et al.*, “Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components,” *Genes Dev.*, vol. 21, no. 20, pp. 2539–2544, Oct. 2007, doi: 10.1101/gad.451207.
[175] Z. Gao *et al.*, “An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation,” *Nature*, vol. 465, no. 7294, pp. 106–109, May 2010, doi: 10.1038/nature09025.
[176] X. Zhong *et al.*, “Molecular mechanism of action of plant DRM de novo DNA methyltransferases,” *Cell*, vol. 157, no. 5, pp. 1050–1060, May 2014, doi: 10.1016/j.cell.2014.03.056.
[177] T. Kanno, M. F. Mette, D. P. Kreil, W. Aufsatz, M. Matzke, and A. J. M. Matzke, “Involvement of Putative SNF2 Chromatin Remodeling Protein DRD1 in RNA-Directed DNA Methylation,” *Current Biology*, vol. 14, no. 9, pp. 801–805, May 2004, doi: 10.1016/j.cub.2004.04.037.
[178] L. M. Smith *et al.*, “An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis,” *Plant Cell*, vol. 19, no. 5, pp. 1507–1521, May 2007, doi: 10.1105/tpc.107.051540.
[179] M. Zhou, A. M. S. Palanca, and J. A. Law, “Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family,” *Nat. Genet.*, vol. 50, no. 6, pp. 865–873, 2018, doi: 10.1038/s41588-018-0115-y.
[180] H. Saze, O. M. Scheid, and J. Paszkowski, “Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis,” *Nature Genetics*, vol. 34, no. 1, pp. 65–69, May 2003, doi: 10.1038/ng1138.
[181] S. K. Bowman, M. D. Simon, A. M. Deaton, M. Tolstorukov, M. L. Borowsky, and R. E. Kingston, “Multiplexed Illumina sequencing libraries from picogram quantities of DNA,” *BMC Genomics*, vol. 14, no. 1, p. 466, Jul. 2013, doi: 10.1186/1471-2164-14-466.
[182] B. Langmead and S. L. Salzberg, “Fast gapped-read alignment with Bowtie 2,” *Nat Methods*, vol. 9, no. 4, Art. no. 4, Apr. 2012, doi: 10.1038/nmeth.1923.

[183] K. Chen *et al.*, "DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing," *Genome Res.*, vol. 23, no. 2, pp. 341–351, Feb. 2013, doi: 10.1101/gr.142067.112.
[184] Y. Di, D. W. Schafer, J. S. Cumbie, and J. H. Chang, "The NBP Negative Binomial Model for Assessing Differential Gene Expression from RNA-Seq," *Statistical Applications in Genetics and Molecular Biology*, vol. 10, no. 1, 2011, doi: 10.2202/1544-6115.1637.
[185] W. Chen, Y. Liu, S. Zhu, C. D. Green, G. Wei, and J.-D. J. Han, "Improved nucleosome-positioning algorithm iNPS for accurate nucleosome positioning from sequencing data," *Nat Commun*, vol. 5, no. 1, pp. 1–14, Sep. 2014, doi: 10.1038/ncomms5909.
[186] A. R. Quinlan and I. M. Hall, "BEDTools: a flexible suite of utilities for comparing genomic features," *Bioinformatics*, vol. 26, no. 6, pp. 841–842, Mar. 2010, doi: 10.1093/bioinformatics/btq033.
[187] F. Krueger and S. R. Andrews, "Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications," *Bioinformatics*, vol. 27, no. 11, pp. 1571–1572, Jun. 2011, doi: 10.1093/bioinformatics/btr167.
[188] A. Akalin *et al.*, "methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles," *Genome Biology*, vol. 13, no. 10, p. R87, Oct. 2012, doi: 10.1186/gb-2012-13-10-r87.
[189] S. V. Kumar and P. A. Wigge, "H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in Arabidopsis," *Cell*, vol. 140, no. 1, pp. 136–147, Jan. 2010, doi: 10.1016/j.cell.2009.11.006.
[190] A. T. Wierzbicki *et al.*, "Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome," *Genes Dev.*, vol. 26, no. 16, pp. 1825–1836, Aug. 2012, doi: 10.1101/gad.197772.112.
[191] M. A. Matzke and R. A. Mosher, "RNA-directed DNA methylation: an epigenetic pathway of increasing complexity," *Nat. Rev. Genet.*, vol. 15, no. 6, pp. 394–408, Jun. 2014, doi: 10.1038/nrg3683.
[192] G. Böhmdorfer and A. T. Wierzbicki, "Control of Chromatin Structure by Long Noncoding RNA," *Trends Cell Biol.*, vol. 25, no. 10, pp. 623–632, Oct. 2015, doi: 10.1016/j.tcb.2015.07.002.
[193] R. Martienssen and D. Moazed, "RNAi and heterochromatin assembly," *Cold Spring Harb Perspect Biol*, vol. 7, no. 8, p. a019323, Aug. 2015, doi: 10.1101/cshperspect.a019323.
[194] J. M. Wendte and C. S. Pikaard, "The RNAs of RNA-directed DNA methylation," *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, vol. 1860, no. 1, pp. 140–148, Jan. 2017, doi: 10.1016/j.bbagrm.2016.08.004.
[195] M. Zhou and J. A. Law, "RNA Pol IV and V in gene silencing: Rebel polymerases evolving away from Pol II's rules," *Current Opinion in Plant Biology*, vol. 27, pp. 154–164, Oct. 2015, doi: 10.1016/j.pbi.2015.07.005.
[196] D. Cuerda-Gil and R. K. Slotkin, "Non-canonical RNA-directed DNA methylation," *Nature Plants*, vol. 2, no. 11, pp. 1–8, Nov. 2016, doi: 10.1038/nplants.2016.163.
[197] S. Lahmy *et al.*, "Evidence for ARGONAUTE4-DNA interactions in RNA-directed DNA methylation in plants," *Genes Dev.*, vol. 30, no. 23, pp. 2565–2570, 01 2016, doi: 10.1101/gad.289553.116.

[198] A. T. Wierzbicki, T. S. Ream, J. R. Haag, and C. S. Pikaard, “RNA polymerase V transcription guides ARGONAUTE4 to chromatin,” *Nat. Genet.*, vol. 41, no. 5, pp. 630–634, May 2009, doi: 10.1038/ng.365.
[199] G. Böhmdorfer *et al.*, “Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin,” *Elife*, vol. 5, 25 2016, doi: 10.7554/eLife.19092.
[200] W. Liu *et al.*, “RNA-directed DNA methylation involves co-transcriptional small-RNA-guided slicing of polymerase V transcripts in Arabidopsis,” *Nat Plants*, vol. 4, no. 3, pp. 181–188, 2018, doi: 10.1038/s41477-017-0100-y.
[201] A. T. Wierzbicki, J. R. Haag, and C. S. Pikaard, “Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes,” *Cell*, vol. 135, no. 4, pp. 635–648, Nov. 2008, doi: 10.1016/j.cell.2008.09.035.
[202] Y. Qi, X. He, X.-J. Wang, O. Kohany, J. Jurka, and G. J. Hannon, “Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation,” *Nature*, vol. 443, no. 7114, pp. 1008–1012, Oct. 2006, doi: 10.1038/nature05198.
[203] Q. Zheng, M. J. Rowley, G. Böhmdorfer, D. Sandhu, B. D. Gregory, and A. T. Wierzbicki, “RNA polymerase V targets transcriptional silencing components to promoters of protein-coding genes,” *Plant J.*, vol. 73, no. 2, pp. 179–189, Jan. 2013, doi: 10.1111/tpj.12034.
[204] G. Böhmdorfer, M. J. Rowley, J. Kuciński, Y. Zhu, I. Amies, and A. T. Wierzbicki, “RNA-directed DNA methylation requires stepwise binding of silencing factors to long non-coding RNA,” *Plant J.*, vol. 79, no. 2, pp. 181–191, Jul. 2014, doi: 10.1111/tpj.12563.
[205] M. J. Rowley, M. H. Rothi, G. Böhmdorfer, J. Kuciński, and A. T. Wierzbicki, “Long-range control of gene expression via RNA-directed DNA methylation,” *PLoS Genet.*, vol. 13, no. 5, p. e1006749, May 2017, doi: 10.1371/journal.pgen.1006749.
[206] Y. Zhu, M. J. Rowley, G. Böhmdorfer, and A. T. Wierzbicki, “A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing,” *Mol. Cell*, vol. 49, no. 2, pp. 298–309, Jan. 2013, doi: 10.1016/j.molcel.2012.11.011.
[207] Y.-H. Hung and R. K. Slotkin, “The initiation of RNA interference (RNAi) in plants,” *Curr Opin Plant Biol*, vol. 61, p. 102014, Feb. 2021, doi: 10.1016/j.pbi.2021.102014.
[208] T. Blevins *et al.*, “Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis,” *Elife*, vol. 4, p. e09591, Oct. 2015, doi: 10.7554/eLife.09591.
[209] J. R. Haag *et al.*, “In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing,” *Mol Cell*, vol. 48, no. 5, pp. 811–818, Dec. 2012, doi: 10.1016/j.molcel.2012.09.027.
[210] J. A. Law *et al.*, “Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1.,” *Nature*, vol. 498, no. 7454, pp. 385–389, Jun. 2013, doi: 10.1038/nature12178.
[211] H. Zhang *et al.*, “DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, no. 20, pp. 8290–8295, May 2013, doi: 10.1073/pnas.1300585110.
[212] L. M. Johnson *et al.*, “SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation,” *Nature*, vol. 507, no. 7490, pp. 124–128, 06 2014, doi: 10.1038/nature12931.

[213] Z.-W. Liu *et al.*, “The SET Domain Proteins SUVH2 and SUVH9 Are Required for Pol V Occupancy at RNA-Directed DNA Methylation Loci,” *PLOS Genetics*, vol. 10, no. 1, p. e1003948, Jan. 2014, doi: 10.1371/journal.pgen.1003948.
[214] X.-J. He *et al.*, “An Effector of RNA-Directed DNA Methylation in Arabidopsis Is an ARGONAUTE 4- and RNA-Binding Protein,” *Cell*, vol. 137, no. 3, pp. 498–508, May 2009, doi: 10.1016/j.cell.2009.04.028.
[215] T. Blevins *et al.*, “A Two-Step Process for Epigenetic Inheritance in Arabidopsis,” *Molecular Cell*, vol. 54, no. 1, pp. 30–42, Apr. 2014, doi: 10.1016/j.molcel.2014.02.019.
[216] K. Domb *et al.*, “DNA methylation mutants in Physcomitrella patens elucidate individual roles of CG and non-CG methylation in genome regulation,” *Proc Natl Acad Sci U S A*, vol. 117, no. 52, pp. 33700–33710, Dec. 2020, doi: 10.1073/pnas.2011361117.
[217] H. Stroud *et al.*, “Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis,” *Nature Structural & Molecular Biology*, vol. 21, no. 1, Art. no. 1, Jan. 2014, doi: 10.1038/nsmb.2735.
[218] H. Stroud, M. V. C. Greenberg, S. Feng, Y. V. Bernatavichute, and S. E. Jacobsen, “Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome,” *Cell*, vol. 152, no. 1–2, pp. 352–364, Jan. 2013, doi: 10.1016/j.cell.2012.10.054.
[219] J. Feng, C. A. Meyer, Q. Wang, J. S. Liu, X. Shirley Liu, and Y. Zhang, “GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data,” *Bioinformatics*, vol. 28, no. 21, pp. 2782–2788, Nov. 2012, doi: 10.1093/bioinformatics/bts515.
[220] L. M. Johnson, J. A. Law, A. Khattar, I. R. Henderson, and S. E. Jacobsen, “SRA-Domain Proteins Required for DRM2-Mediated De Novo DNA Methylation,” *PLOS Genetics*, vol. 4, no. 11, p. e1000280, Nov. 2008, doi: 10.1371/journal.pgen.1000280.
[221] A. Zemach *et al.*, “The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin,” *Cell*, vol. 153, no. 1, pp. 193–205, Mar. 2013, doi: 10.1016/j.cell.2013.02.033.
[222] K. Panda and R. K. Slotkin, “Long-Read cDNA Sequencing Enables a ‘Gene-Like’ Transcript Annotation of Transposable Elements,” *Plant Cell*, vol. 32, no. 9, pp. 2687–2698, Sep. 2020, doi: 10.1105/tpc.20.00115.
[223] J. P. Jackson, A. M. Lindroth, X. Cao, and S. E. Jacobsen, “Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase,” *Nature*, vol. 416, no. 6880, pp. 556–560, Apr. 2002, doi: 10.1038/nature731.
[224] R. A. Mosher, F. Schwach, D. Studholme, and D. C. Baulcombe, “PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis,” *Proc Natl Acad Sci U S A*, vol. 105, no. 8, pp. 3145–3150, Feb. 2008, doi: 10.1073/pnas.0709632105.
[225] F. Wang and M. J. Axtell, “AGO4 is specifically required for heterochromatic siRNA accumulation at Pol V-dependent loci in Arabidopsis thaliana,” *The Plant Journal*, vol. 90, no. 1, pp. 37–47, 2017, doi: https://doi.org/10.1111/tpj.13463.
[226] Q. Li *et al.*, “RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 112, no. 47, pp. 14728–14733, Nov. 2015, doi: 10.1073/pnas.1514680112.

[227] J. Gallego-Bartolomé *et al.*, "Co-targeting RNA Polymerases IV and V Promotes Efficient De Novo DNA Methylation in Arabidopsis," *Cell*, vol. 176, no. 5, pp. 1068-1082.e19, 21 2019, doi: 10.1016/j.cell.2019.01.029.
[228] H. Zhang, Z. Lang, and J.-K. Zhu, "Dynamics and function of DNA methylation in plants," *Nat Rev Mol Cell Biol*, vol. 19, no. 8, pp. 489–506, Aug. 2018, doi: 10.1038/s41580-018-0016-z.
[229] D. Pontier *et al.*, "Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis," *Genes Dev.*, vol. 19, no. 17, pp. 2030–2040, Sep. 2005, doi: 10.1101/gad.348405.
[230] M. V. C. Greenberg and D. Bourc'his, "The diverse roles of DNA methylation in mammalian development and disease," *Nat Rev Mol Cell Biol*, vol. 20, no. 10, pp. 590–607, Oct. 2019, doi: 10.1038/s41580-019-0159-6.
[231] C. Luo, P. Hajkova, and J. R. Ecker, "Dynamic DNA methylation: In the right place at the right time," *Science*, vol. 361, no. 6409, pp. 1336–1340, Sep. 2018, doi: 10.1126/science.aat6806.
[232] E. J. Finnegan, R. K. Genger, W. J. Peacock, and E. S. Dennis, "Dna methylation in plants," *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.*, vol. 49, no. 1, pp. 223–247, Jun. 1998, doi: 10.1146/annurev.arplant.49.1.223.
[233] T. L. Dunwell and G. P. Pfeifer, "Drosophila genomic methylation: new evidence and new questions," *Epigenomics*, vol. 6, no. 5, pp. 459–461, 2014, doi: 10.2217/epi.14.46.
[234] S. V. Yi, "Birds do it, bees do it, worms and ciliates do it too: DNA methylation from unexpected corners of the tree of life," *Genome Biology*, vol. 13, no. 10, p. 174, Oct. 2012, doi: 10.1186/gb-2012-13-10-174.
[235] E. Heard and R. A. Martienssen, "Transgenerational Epigenetic Inheritance: Myths and Mechanisms," *Cell*, vol. 157, no. 1, pp. 95–109, Mar. 2014, doi: 10.1016/j.cell.2014.02.045.
[236] M. G. Izban and D. S. Luse, "Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates.," *Journal of Biological Chemistry*, vol. 267, no. 19, pp. 13647–13655, Jul. 1992, doi: 10.1016/S0021-9258(18)42262-4.
[237] K. Adelman and J. T. Lis, "How Does Pol II Overcome the Nucleosome Barrier?," *Molecular Cell*, vol. 9, no. 3, pp. 451–452, Mar. 2002, doi: 10.1016/S1097-2765(02)00487-2.
[238] Z.-W. Liu *et al.*, "The SET Domain Proteins SUVH2 and SUVH9 Are Required for Pol V Occupancy at RNA-Directed DNA Methylation Loci," *PLOS Genetics*, vol. 10, no. 1, p. e1003948, Jan. 2014, doi: 10.1371/journal.pgen.1003948.
[239] W. Liu *et al.*, "RNA-directed DNA methylation involves co-transcriptional small-RNA-guided slicing of polymerase V transcripts in Arabidopsis," *Nature Plants*, p. 1, Jan. 2018, doi: 10.1038/s41477-017-0100-y.
[240] R. Martienssen and D. Moazed, "RNAi and Heterochromatin Assembly," *Cold Spring Harb Perspect Biol*, vol. 7, no. 8, p. a019323, Aug. 2015, doi: 10.1101/cshperspect.a019323.
[241] F. Capuano, M. Mülleder, R. Kok, H. J. Blom, and M. Ralser, "Cytosine DNA Methylation Is Found in Drosophila melanogaster but Absent in Saccharomyces

cerevisiae, Schizosaccharomyces pombe, and Other Yeast Species," *Anal Chem*, vol. 86, no. 8, pp. 3697–3702, Apr. 2014, doi: 10.1021/ac500447w.
[242] J. F. Partridge, K. S. C. Scott, A. J. Bannister, T. Kouzarides, and R. C. Allshire, "cis-Acting DNA from Fission Yeast Centromeres Mediates Histone H3 Methylation and Recruitment of Silencing Factors and Cohesin to an Ectopic Site," *Current Biology*, vol. 12, no. 19, pp. 1652–1660, Oct. 2002, doi: 10.1016/S0960-9822(02)01177-6.
[243] K. Noma *et al.*, "RITS acts in cis to promote RNA interference–mediated transcriptional and post-transcriptional silencing," *Nat Genet*, vol. 36, no. 11, pp. 1174–1180, Nov. 2004, doi: 10.1038/ng1452.
[244] R. Yu, X. Wang, and D. Moazed, "Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation," *Nature*, vol. 558, no. 7711, pp. 615–619, Jun. 2018, doi: 10.1038/s41586-018-0239-3.
[245] C. Kadoch and G. R. Crabtree, "Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics," *Sci Adv*, vol. 1, no. 5, Jun. 2015, doi: 10.1126/sciadv.1500447.
[246] L. Tang, E. Nogales, and C. Ciferri, "Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription," *Progress in Biophysics and Molecular Biology*, vol. 102, no. 2, pp. 122–128, Jun. 2010, doi: 10.1016/j.pbiomolbio.2010.05.001.
[247] S. Bezhani *et al.*, "Unique, Shared, and Redundant Roles for the Arabidopsis SWI/SNF Chromatin Remodeling ATPases BRAHMA and SPLAYED," *The Plant Cell*, vol. 19, no. 2, pp. 403–416, Feb. 2007, doi: 10.1105/tpc.106.048272.
[248] S.-K. Han, M.-F. Wu, S. Cui, and D. Wagner, "Roles and activities of chromatin remodeling ATPases in plants," *Plant J*, vol. 83, no. 1, pp. 62–77, Jul. 2015, doi: 10.1111/tpj.12877.
[249] A. Portela, J. Liz, V. Nogales, F. Setién, A. Villanueva, and M. Esteller, "DNA methylation determines nucleosome occupancy in the 5′-CpG islands of tumor suppressor genes," *Oncogene*, vol. 32, no. 47, pp. 5421–5428, Nov. 2013, doi: 10.1038/onc.2013.162.
[250] B. Ghoshal, C. L. Picard, B. Vong, S. Feng, and S. E. Jacobsen, "CRISPR-based targeting of DNA methylation in Arabidopsis thaliana by a bacterial CG-specific DNA methyltransferase," *PNAS*, vol. 118, no. 23, Jun. 2021, doi: 10.1073/pnas.2125016118.
[251] W. Liu *et al.*, "Ectopic targeting of CG DNA methylation in Arabidopsis with the bacterial SssI methyltransferase," *Nat Commun*, vol. 12, no. 1, p. 3130, May 2021, doi: 10.1038/s41467-021-23346-y.
[252] J. M. Franco-Zorrilla, I. López-Vidriero, J. L. Carrasco, M. Godoy, P. Vera, and R. Solano, "DNA-binding specificities of plant transcription factors and their potential to define target genes," *PNAS*, vol. 111, no. 6, pp. 2367–2372, Feb. 2014.
[253] M. Galli, F. Feng, and A. Gallavotti, "Mapping Regulatory Determinants in Plants," *Front. Genet.*, vol. 11, 2020, doi: 10.3389/fgene.2020.591194.
[254] M. J. Rowley, M. H. Rothi, G. Böhmdorfer, J. Kuciński, and A. T. Wierzbicki, "Long-range control of gene expression via RNA-directed DNA methylation," *PLoS Genet*, vol. 13, no. 5, May 2017, doi: 10.1371/journal.pgen.1006749.
[255] V. B. Teif *et al.*, "Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development," *Genome Res.*, vol. 24, no. 8, pp. 1285–1295, Aug. 2014, doi: 10.1101/gr.164418.113.

[256] V. Haberle and A. Stark, “Eukaryotic core promoters and the functional basis of transcription initiation,” *Nat Rev Mol Cell Biol*, vol. 19, no. 10, pp. 621–637, Oct. 2018, doi: 10.1038/s41580-018-0028-8.
[257] M. Marasco, W. Li, M. Lynch, and C. S. Pikaard, “Catalytic properties of RNA polymerases IV and V: accuracy, nucleotide incorporation and rNTP/dNTP discrimination,” *Nucleic Acids Res*, vol. 45, no. 19, pp. 11315–11326, Nov. 2017, doi: 10.1093/nar/gkx794.