Title
The Genetics of de novo Methylation in Arabidopsis thaliana

Author
Greenberg, Maxim Van Cleef

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Genetics of *de novo* Methylation in *Arabidopsis thaliana*

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular, Cell, and Developmental Biology

by

Maxim Van Cleef Greenberg

2012

ABSTRACT OF THE DISSERTATION

The Genetics of *de novo*

Methylation in *Arabidopsis thaliana*

by

Maxim Van Cleef Greenberg

Doctor of Philosophy in Molecular, Cell, and Developmental Biology

University of California, Los Angeles, 2012

Professor Steven E. Jacobsen, Chair

Cytosine DNA methylation is an ancient form of transcriptional control that is conserved across all kingdoms of eukaryotes. DNA methylation plays a major role in silencing of selfish genetic elements, such as transposons. Additionally, in some instances, DNA methylation is required for genomic imprinting and regulation of endogenous genes. In the model plant *Arabidopsis thaliana*, at least three pathways, each with its own methyltransferase, maintain DNA methylation. MET1 targets CG dinucleotide sequences; due to the inherent symmetry across DNA strands, MET1 is able to recognize hemimethylated sites after DNA replication, thus maintains faithful methylation patterns. CMT3 typically has preference for CHG sites (where H is A, T, or C), and is targeted to chromatin via its chromodomain, which has specificity for histone 3 lysine 9 dimethylation—another epigenetic mark associated with heterochromatin. Finally, DRM2 maintains CHH, or asymmetric, methylation through targeting by a dual siRNA/long non-coding RNA pathway termed RNA-directed DNA methylation (RdDM). It should be noted that CMT3 and DRM2 are both capable of methylating non-CG sites. While all three

methyltransferases maintain existing DNA methylation patterns, only the RdDM pathway establishes the mark in all sequence contexts, in a process known as *de novo* methylation.

In this dissertation I will describe both forward and reverse genetic techniques I have used to uncover factors required for *de novo* methylation. For both techniques, I made use of the *FWA* transgene. In wild-type plants, the RdDM pathway is able to target, methylate, and silence the transgene at the repeats in its 5' UTR. However, in RdDM mutants, the transgene remains unmethylated and expresses, leading to a late-flowering phenotype. From a mutagenesis screen, I discovered novel mutations in 11 genes required for DNA methylation establishment. I will describe the methodologies of cloning and characterizing those mutants. Additionally, from the same study, I was able to show a *de novo* methylation phenotype from previously described RdDM mutant alleles.

In a reverse genetic screen, utilizing a collection of insertional mutations in known or putative RNA binding proteins, I helped characterize a known RNA splicing factor, the first such RNA processing protein shown to be required for RdDM. I also showed that two partially redundant paralogs of the IDN2 RNA-binding protein are required for RdDM and *de novo* methylation. Further biochemical analysis revealed that the paralogs form a complex with IDN2. In collaboration with a structural biology group, we solved the structure of the RNA binding motif of IDN2.

Finally, I will discuss the data explicating the relationship between histone 3 lysine 4 (H3K4) demethylases and the RdDM pathway. We made the surprising discovery that active demethylation is required for RdDM maintenance, but not establishment. In sum, the work in this dissertation contributes to our knowledge of the components and mechanism of RdDM, and how the RNA polymerase-dependent pathway is affected by perturbations in local chromatin.

The dissertation of Maxim Van Cleef Greenberg is approved.

Arnold J. Berk

Amander T. Clark

Chentao Lin

Steven E. Jacobsen, Committee Chair

University of California, Los Angeles

2012

I dedicate this dissertation to the first molecular biologist in my family, my beloved grandfather

Sidney ‘Zaide’ Raffel

TABLE OF CONTENTS

FIGURES AND TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

ACKNOWLEDGMENTS

First and foremost, I would like to thank my graduate advisor and committee chair, Steve Jacobsen. He has created a nurturing and collegial lab environment that has truly allowed me to grow as a scientist. Additionally, he has an optimistic nature about science that kept me excited and motivated throughout my graduate career. His zeal for science and his willingness to take risks are traits I hope to take with me as I advance to the next stage. I am grateful for the opportunity to work in his lab. I would also like to thank the other members of my committee, Arnie Berk, Amander Clark, and Chentao Lin for their support and advice as I have progressed through the stages of doctoral candidacy.

I have had the tremendous luck to work with a fabulous group of graduate students and post docs in the Jacobsen lab. I feel that I have gained knowledge, insight, and perspective from everyone I have worked with. In particular I would like to thank Julie Law and Magnolia Bostick for helping to train me at the bench, Ian Henderson, Angelique Deleris, Chris Hale, and Lianna Johnson for our stimulating discussions, and finally Israel Ausin for helping me map—literally and figuratively—the Arabidopsis genome. A brief acknowledgment in this dissertation does not do justice for all the gratitude I wish to express to them.

I have had a number of undergraduates who have assisted me with the less glamorous aspects of the job. I am tremendously thankful for all their help over the years. In order they are Carolyn Chu, Sophia Espanola, MaiThy Nguyen, Matt Browne, and Phuong Nguyen.

Finally I would like to thank my parents for their unwavering encouragement and support ever since I first decided I wanted to be a scientist at around age 5 or so.

Chapter 2:

Chapter 2 is a version of the research article published in Epigenetics 6:3, 1-11 (2011) titled “Identification of genes required for *de novo* DNA methylation in Arabidopsis.” This

study was authored by Maxim V.C. Greenberg, Israel Ausin, Simon W.L. Chan, Shawn J. Cokus, Josh T. Cuperus, Suhua Feng, Julie A. Law, Carolyn Chu, Matteo Pellegrini, James C. Carrington and Steven E. Jacobsen. We thank Thierry Lagrange for kindly sending homozygous *nrpe5a-1* mutant seeds. We thank Govind Shah for assistance with the forward screen. We thank members of the Jacobsen laboratory for supportive discussions. Jacobsen lab research was supported by US National Institutes of Health grant GM60398. M.V.C.G. was supported by USPHS National Research Service Award GM07104. I.A. was supported by a postdoctoral fellowship from the Ministerio de Educacion y Ciencia. S.F. is a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation. J.A.L. was supported the National Institutes of Health National Research Service Award 5F32GM820453. M.P. was supported by NSF Plant Genome Research Program Grant 0701745. J.C.C. was supported by grants from the National Science Foundation (MCB-0618433 and MCB-0956526). S.E.J. is an investigator of the Howard Hughes Medical Institute.

Chapter 3:

Chapter 3 is a version of the research article published in Epigenetics 7:1, 29-33 (2012) titled, “The splicing factor SR45 affects the RNA-directed DNA methylation pathway in Arabidopsis.” The study was authored by Israel Ausin, Maxim V.C. Greenberg, Carey Fei Li, and Steven E. Jacobsen. We would like to thank Craig Pikaard for kindly donating anti-AGO4 immunosera and the expression vector for the peptide used for purification. We would like to thank members of the S. Jacobsen laboratory for supportive discussions. Jacobsen lab research was supported by US National Institutes of Health grant GM60398. I.A. was supported by a postdoctoral fellowship from the Ministerio de Educacion y Ciencia. M.V.C.G. was supported by USPHS National Research Service

Award GM07104. C.F.L. was supported by the Ruth L. Kirschstein National Research Service Award GM07185.

Chapter 4:

Chapter 4 is a version of the research article published in Proceedings of the National Academy of Sciences of the United States of America doi: 10.1073/pnas.1206638109 titled “An INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in *Arabidopsis.*” The study was authored by Israel Ausin, Maxim V.C. Greenberg, Dhirendra K. Simanshu, Christopher J. Hale, Ajay A. Vashisht, Stacey A. Simon, Tzuu-fen Lee, Suhua Feng, Sophia D. Española, Blake C. Meyers, James A. Wohlschlegel, Dinshaw J. Patel, and Steven E. Jacobsen. We thank the staff of ID-24-C beamline at the Advanced Photon Source for their help with data collection. We thank the UCLA BSCRC BioSequencing Core Facility. We thank M. Browne and M. Nguyen for their help preparing plant tissue. We thank M. Akhavan for technical assistance. We thank members of the Jacobsen lab for supportive discussions. Jacobsen lab research was supported by US National Institutes of Health grant GM60398. I.A. was supported by a postdoctoral fellowship from the Ministerio de Educacion y Ciencia. M.V.C.G. was supported by USPHS National Research Service Award GM07104 and an UCLA Dissertation Year Fellowship. C.J.H. is a Howard Hughes Medical Institute fellow of the Damon Runyon Cancer Research Foundation. S.F. is a Special Fellow of the Leukemia and Lymphoma Society. Work in the Meyers lab was supported by National Science Foundation award 0701745. Wohlschlegel lab research was supported by US National Institutes of Health grant GM089778 and funds from the Jonsson Cancer Center at UCLA. Patel lab research was supported by the Abby Rockefeller Mauze Trust and the Maloris Foundation. S.E.J. is an investigator of the Howard Hughes Medical Institute.

Chapter 5:

Chapter 5 is a version of the research article published in EMBO Reports 11:12, 950-55 (2010) titled “Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation.” The study was authored by Angelique Deleris, Maxim V.C. Greenberg, Israel Ausin, Rona W.Y. Law, Guillaume Moissiard, Daniel Schubert and Steven E. Jacobsen. We thank M. Mobilia for help with fluorescent microscopy and A. Hooper for genotyping. Jacobsen lab research was funded by US National Institutes of Health grant GM60398. A.D. was funded by EMBO fellowhip ALTF 1161-2007 & HFSP fellowship LT 0056-2008L. M.V.C.G. was funded by USPHS National Research Service Award GM07104. I.A. was funded by the Ministerio de Educacion y Ciencia. R.W.Y.L. was funded by a Van Trees Scholarship under UCLA Undergraduate Research Scholars Program (URSP). S.E.J. is an investigator of the Howard Hughes Medical Institute.

Chapter 6:

Chapter 6 is a version of a manuscript in preparation titled “Interplay Between Active Chromatin Marks and RNA-directed DNA Methylation in *Arabidopsis thaliana*.” It was authored by Maxim V.C. Greenberg, Angelique Deleris, Christopher J. Hale, Ao Liu, Suhua Feng, and Steven E. Jacobsen. We thank R. Amasino for kindly sending seeds for the *ldl1-2 ldl2* double mutant. We thank P. Nguyen, M. Browne, M. Nguyen, R.Y. Law, and L. Johnson for their assistance with plant work and lab work. We thank the UCLA BSCRC BioSequencing Core Facility. We thank M. Akhavan for technical assistance. We thank members of the Jacobsen lab for supportive discussions. Jacobsen lab research was supported by US National Institutes of Health grant GM60398. M.V.C.G. was supported by USPHS National Research Service Award GM07104 and an UCLA Dissertation Year Fellowship. A.D. was funded by a European

Molecular Biology Organization fellowship ALTF 1161-2007 and Human Frontier Science Program fellowship LT 0056-2008L. C.J.H. is a Howard Hughes Medical Institute fellow of the Damon Runyon Cancer Research Foundation. A.L. is part of the Cross-disciplinary Scholars in Science and Technology Program at UCLA. S.F. is a Special Fellow of the Leukemia and Lymphoma Society. S.E.J. is an investigator of the Howard Hughes Medical Institute.

VITA

June 2, 1982	Born, Washington, District of Columbia
2001	Summer Cancer Research Training Award National Cancer Institute, National Institutes of Health Bethesda, Maryland
2004	Bachelor of Arts, Cum Laude with Distinction in Biology University of Pennsylvania Philadelphia, Pennsylvania
2004-06	Research Assistant Department of Assay & Automation Technology, Genentech, Inc. South San Francisco, California
2007-10	Genetics Training Program Fellowship University of California, Los Angeles Los Angeles, California
2008	Certificate of Distinction in Teaching, UCLA Life Sciences Division University of California, Los Angeles Los Angeles, California
2011-12	Dissertation Year Fellowship University of California, Los Angeles Los Angeles, California

PUBLICATIONS

Israel Aui Israel Ausin*, **Maxim V.C. Greenberg***, Dhirendra K. Simanshu*, Christopher J. Hale, Ajay A. Vashisht, Stacey A. Simon, Tzuu-fen Lee, Suhua Feng, Sophia D. Española, Blake C. Meyers, James A. Wohlschlegel, Dinshaw J. Patel, and Steven E. Jacobsen. (2012) An IDN2-containing complex involved in RNA-directed DNA methylation in Arabidopsis. *Proc Natl Acad Sci USA*, In Press. *These authors contributed equally to this work.

Israel Ausin*, **Maxim V.C. Greenberg***, Carey Fei Li, and Steven E. Jacobsen. (2011) The splicing factor SR45 affects the RNA-directed DNA methylation pathway in Arabidopsis. *Epigenetics*, 7:29-33 *These authors contributed equally to this work.

Maxim V.C. Greenberg, Israel Ausin, Simon W.L. Chan, Shawn J. Cokus, Josh T. Cuperus, Suhua Feng, Julie A. Law, Carolyn Chu, Matteo Pellegrini, James C. Carrington and Steven E. Jacobsen. (2011) Identification of genes required for *de novo* DNA methylation in Arabidopsis. *Epigenetics*, 6, 344-354.

Angelique Deleris*, **Maxim V.C. Greenberg***, Israel Ausin, Rona W.Y. Law, Guillaume Moissiard, Daniel Schubert and Steven E. Jacobsen. (2010) Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA Methylation. *EMBO Reports*, 11, 950-955. *These authors contributed equally to this work.

Matthew D. Ringel, **Max Greenberg**, Xuwu Chen, Nicole Hayre, Koichi Suzuki, Dennis Priebat, Motoyasu Saji, Kenneth D. Burman. (2000). Cytotoxic activity of 2',2'- difluorodeoxycytidine (gemcitabine) in poorly differentiated thyroid carcinoma cells. *Thyroid*, 10, 865-869

CHAPTER 1

Introduction

Understanding how genetic elements are regulated is at the fundamental core of molecular biology. For example, the progression of differentiating cell lineages in development of any multicellular organism requires the precise differentiation of specific gene programs. In mammals, cells that are virtually identical at the very early stages of the embryo soon form primordial cell types that are the basis for all adult tissue (Peter and Davidson, 2011). It goes without saying that if errors occurred—if genes that define the endoderm are turned on in the ectoderm, say—the result for the developing embryo would be catastrophic. What underlies the changes in gene expression? And perhaps more saliently, how are specific genes targeted for these changes?

Historically, the basis of the study of gene regulation has been at the level of transcription factors. Transcription factors have DNA binding capacity, and cue transcription by recruiting and/or activating the RNA polymerase complex (Ptashne, 1988). In more recent years, the relationship between transcription and the local chromatin state has come more into focus—that is covalent modifications to DNA and the histone proteins around which DNA is wrapped. Over a decade ago, the term “histone code” entered the scientific lexicon (Jenuwein and Allis, 2001). The basis of the code is combinatorial histone marks, which likely function as transcriptional signatures—even if the “rules” are not as strict as those for the genetic code. For example in the model plant *Arabidopsis thaliana*, histone 3 lysine 4 trimethylation (H3K4m3) and histone 3 acetylation (H3ac) correlate with actively transcribed loci. Conversely, H3K27m1, H3K9m2 and DNA cytosine methylation correlate with silent loci (Roudier et al., 2011). These chromatin signatures are often heritable across cell division, and the study thereof has garnered the moniker of “epigenetics.”

The focus of this dissertation will be the study of the epigenetic mark cytosine DNA methylation in Arabidopsis; in particular, the non-coding RNA-driven pathway that is responsible for establishing the methyl-mark, and how the pathway is affected by perturbations in histone

marks. In this chapter, the functional characteristics of cystosine DNA methylation will be explored in a broader sense.

Cytosine DNA Methylation in Prokaryotes

DNA cytosine methylation is an ancient form chromatin modification that finds its evolutionary roots in bacteria. For example, *E. coli* genomes encode for a cytosine methyltransferase from the *Dcm* gene, which methylates the internal C from CCWGG sequences, where the internal nucleotide is either A or T (May and Hattman, 1975). The role of bacterial methyltransferases is typically thought to be in defense against viral DNA (Wilson and Murray, 1991). Methylation of palindromic sequences (such as *Dcm* target sites) prevents digestion by many restriction endonucleases. Therefore, if unmethylated virus DNA gets intercalated into bacterial genomes, the potentially harmful sequences will be degraded. The cytosine DNA methyltransferases in eukaryotes—while not functioning as restriction-modifiers—find their sequence and structural origins from prokaryotic ancestors (Bujnicki and Radlinska, 1999; Cheng, 1995; Jeltsch, 2002; Kumar et al., 1994). Therefore, cytosine DNA methylation is a very ancient process. Incidentally, the sensitivity of many restriction endonucleases to cytosine-methylated sequences has become a useful tool for molecular biological analyses of DNA methylation in eukaryotic laboratory model organisms. In the age of molecular biology, the role of cytosine methyltransferases as restriction modifiers has adopted a new analytical function.

Cytosine DNA Methylation in Eukaryotes

In eukaryotic organisms, cytosine DNA methylation is largely localized to repetitive and selfish genetic elements, such as transposons (Martienssen and Colot, 2001; Yoder et al., 1997). The transcriptional repression of transposons is tremendously important for maintaining the integrity

of an organism's genome, as unchecked transposition will assuredly eventually lead to genic disruption. Therefore, it has long been thought that DNA methylation plays a role in transcriptional repression. It is interesting to note that while cystosine DNA methylation has evolved a dramatically distinct mechanism in eukaryotes, its role in protecting an organism form invasive genetic elements has remained intact from its prokaryotic ancestral function.

Interestingly, despite being present across all kingdoms of eukaryotes, cytosine DNA methylation has been lost in many commonly used laboratory model organisms, such as *C. elegans*, *S. pombe*, and *S. cerevisiae* (Chan et al., 2005). While *D. melanogaster* contains a gene encodes a protein with homology to DNA methyltransferases, it appears to have adopted a role as a tRNA methyltransferase, although there is some evidence for DNA methylation in Drosophila development (Goll et al., 2006; Phalke et al., 2009). These organisms have evolved alternative mechanisms for controlling transposable elements. Because the eukaryotic laboratory organisms with the most extensive cytosine DNA methylation are mice and Arabidopsis, this introductory chapter will delve mainly discuss the role of eukaryotic DNA methylation within the context of mammals and plants.

Both mammals and plants have made use of the transcriptional repression characteristics of the cytosine methyl-mark in order to regulate endogenous genes (Law and Jacobsen, 2010). For example, in Arabidopsis, certain pathogen resistance genes are subject to epigenetic control through DNA methylation (Stokes et al., 2002; Stokes and Richards, 2002). A number of Arabidopsis genes, such as *BONSAI* (*BNS*), *SUPERMAN* (*SUP*) and *FLOWERING WAGENINGEN* (*FWA*) all have stable "epialleles," in which the methylation state of the gene has been deviated from wild-type, which results in a change in expression and a morphological phenotype (Jacobsen and Meyerowitz, 1997; Saze and Kakutani, 2007; Soppe et al., 2000). Additionally, Arabidopsis rDNA in stochastically methylated, as a control mechanism of rRNA transcription (Chan et al., 2005). Similarly, the Kit gene in mice can form stable epialleles that are inherited even after out-crossing to wild-type animals (Rassoulzadegan et al., 2006).

Both plants and mammals also utilize cytosine methylation in a process known as genomic imprinting (Huh et al., 2008; Reik and Walter, 2001). Imprinting is defined as expression of genes in a parent-of-origin manner. In both plants and animals, the differential expression is due to methylation of a parental allele. The mechanism of genomic imprinting is quite different in plants and animals. Flowering plants, like Arabidopsis, undergo a double fertilization process in which one male gamete fuses with the female egg cell to form the zygote, and a second male gamete fertilizes with central cell nuclei of the female gametophyte to form triploid endosperm (Huh et al., 2008). In the endosperm, several female genes are demethylated, while the male copies remain methylated (Hsieh et al., 2011). Therefore, in the endosperm many female genes are expressed, while the male genes remain transcriptionally silent. One such example is the aforementioned *FWA* gene (Kinoshita et al., 2004). However, the endosperm does not contribute to somatic tissue, so adult plants do not carry a direct epigenetic signature of the imprint. In mammals, both female and male genes can be methylated in a parent-specific manner prior to the formation of mature gametes in germline cells. These marks persist through the formation of the zygote and embryonic development, thus are observed as differentially expressing alleles adults (Reik and Walter, 2001). Much has been hypothesized about the evolutionary implications of genomic imprinting, given that evolutionary theory states that there is a parental conflict over the level of maternal investment to offspring (Iwasa and Pomiankowski, 2001; Moore and Haig, 1991; Spencer et al., 1999). The continued efforts to marry evolutionary theory with molecular biological observations in the realm of genomic imprinting will undoubtedly provide fascinating insights into both fields.

As mentioned, historically cytosine DNA methylation is associated with transcriptional repression. However, as whole genome analyses of DNA methylation advanced, it became clear that in fact large swaths of transcribed loci are cytosine DNA methylated. In fact, in Arabidopsis, constitutively transcribed genes are more likely to be methylated in their bodies (Zhang et al., 2006). This phenomenon is known as gene-body methylation, and is conserved

across several species across kingdoms and phyla (Feng et al., 2010a; Zemach et al., 2010). While the function of gene-body methylation is still being elucidated, some hints are emerging as to its biological role. In Arabidopsis, the gene-body methylation is enriched on exonic DNA (Chodavarapu et al., 2010). Indeed, in mammals it is coming to light that gene-body methylation plays a role in proper gene splicing at some loci (Shukla et al., 2011). While more research must be done to cement the relationship between DNA methylation and spliceosome machinery, it nevertheless indicates that DNA methylation is not sufficient to prevent transcription, *per se*, and likely must be in combination with other chromatin marks in order to proffer transcriptional repression.

Mechanisms of Maintenance DNA Methylation

In both mammals and plants, the most prevalent sequence for cytosine DNA methylation is in the CG dinucleotide context. In fact, in mammals, DNA methylation is almost exclusively observed at CG sites, with the notable exception of embryonic stem cells, wherein there is cytosine DNA methylation observed at non-CG sites, albeit at relatively low levels (Lister et al., 2011). Plants and mammals share a highly homologous DNA methyltransferase that maintains the CG mark; in plants it is referred to as METHYLTRANSFERASE 1 (MET1), and in mammals DNA METHYLTRANSFERASE 1 (DNMT1). It has long been proposed that faithful maintenance of CG DNA methylation relies on the inherent symmetry across the double-stranded DNA (Figure 1-1 A). To wit, immediately following DNA replication, a hemimethylated CG site would form with an unmethylated daughter "CG." Because of the symmetry, either MET1/DNMT1 or an associated factor could recognize the hemimethylation, resulting in methylation of the daughter cytosine. In line with this hypothesis, DNMT1 has been shown to interact with the DNA replication machinery (Chuang et al., 1997). As mentioned, prokaryotic cytosine DNA

methyltransferases recognize palindromic sequences, thus there is a precedent for methyltransferases to have sequence recognition capability.

In fact, it is not MET1/DNMT1 that recognizes the hemimethylated CG site, but an interaction partner also found at replicating DNA (Bostick et al., 2007; Sharif et al., 2007). This protein, called UHRF1 in mammals, as a SET and RING-associated (SRA) domain, which is responsible for the recognition. Subsequent structural studies have elucidated the mechanism in which UHRF1 flips out the methylated cytosine from the DNA helix (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008). Interestingly, bacterial DNA methyltransferases are known to also have a base flipping mechanism (Cheng and Roberts, 2001; Klimasauskas et al., 1994). It appears that eukaryotes have partitioned the activities of prokaryotic DNA methyltransferases into multiple proteins. The plant MET1 pathway likely acts in virtually an identical fashion. Arabidopsis contains a group of proteins with high homology to UHRF1 termed the VARIANT IN METHYLATION (VIM) family. In higher order mutants for several of these genes, the DNA methylation phenotype recapitulates that of *met1* mutants (Woo et al., 2007; Woo et al., 2008). Moreover, the SRA domain of VIM1 has been shown to bind to hemimethylated CG sites *in vitro* (Johnson et al., 2007).

While the DNMT1 and MET1 pathways are clearly very conserved between kingdoms, plants have a evolved a second plant-specific cytosine DNA methylation pathway dependent on a protein called CHROMOMETHYLASE 3 (CMT3). CMT3 has traditionally been thought to have specificity for CHG sequences (where H = A, T, or C) (Cao and Jacobsen, 2002a; Lindroth et al., 2001). Because of the symmetry across DNA strands, like with CG sites, CMT3 could be acting in a similar method as MET1. However, genome wide methylation analyses suggest that CMT3 also plays a significant role in maintaining CHH (or asymmetric) methylation (Cokus et al., 2008). Then what accounts for the maintenance activity? The answer lies in the eponymous chromodomain, which recognizes H3K9m2 (Lindroth et al., 2004). Thus at H3K9 methylated loci, a mechanism is in place to recruit CMT3 to maintain faithful methylation patterns (Figure 1-

1 B). Moreover, three H3K9m2 methyltransferases termed KYP/SUVH4, SUVH5 and SUVH6 all contain SRA domains akin to UHRF1 and the VIM proteins. These proteins have been shown to bind methylated DNA *in vitro* (Johnson et al., 2007). Presumably, methylated DNA then recruits the histone methyltransferases. Null mutations in these histone methyltransferase affect DNA methylation at CMT3 targets (Jackson et al., 2002; Johnson et al., 2007; Malagnac et al., 2002). Therefore, the cooperation between a family of histone methyltransferases and a DNA methyltransferase form a self-reinforcing loop that contributes to transcriptional gene silencing.

De novo DNA Methylation in Mammals

Neither the plant cytosine DNA methyltransferases MET1 and CMT3 nor mammalian DNMT1 establish the methyl-mark on previously unmethylated sequences (Feng et al., 2010b). This process, known as *de novo* methylation, is dependent on DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) in plants, and DNA METHYLTRANSFERASE 3A and 3B (DNMT3A and 3B) in mammals. While the plant and mammalian *de novo* methyltransferases are orthologous, they appear to have evolved distinct pathways, unlike MET1 and DNMT1 (Law and Jacobsen, 2010). Establishing DNA methylation has important implications for targeting invasive genetic elements for silencing, as well as genomic imprinting. In mammalian development, the significance of *de novo* methylation is underscored by embryonic lethality observed in *Dnmt3A* and *Dnmt3A* knockout mice (Okano et al., 1999).

Mammals, unlike plants, undergo massive epigenetic reprogramming two times during the life cycle: in germ line cells and again after formation of the zygote (Hemberger et al., 2009; Surani et al., 2008). It is during the germ line reprogramming that the parental imprints are established, and typically primary imprints avert the reprogramming that takes place in the developing embryo (Sasaki and Matsui, 2008). These imprints are maintained throughout life of the offspring in the somatic tissue, but are erased and re-imprinted again in the primordial germ

cells (PGCs) (Sasaki and Matsui, 2008). While genomic imprinting takes place at different time points in male and female PGCs—in males it occurs in the developing embryo, and in females after birth—both sexes require DNMT3A and a non-catalytic binding partner termed DNMT3L (Bourc'his and Bestor, 2004; Kaneda et al., 2004; Kato et al., 2007). Structural analysis has suggested that the DNMT3A/3L complex preferentially methylates CG dinucleotides that are spaced 10 base pairs (bps) apart; incidentally, 10-bp CG spacing is characteristic of many imprinted loci (Jia et al., 2007). Additionally, DNMT3L binds to unmethylated H3K4 (H3K4m0), and stimulates DNA methylation activity of DNMT3A (Ooi et al., 2007). Therefore, both genetic and epigenetic information may contribute to the targeting of genomic imprinting machinery.

Given that the germ line cells contribute directly to the genome of the next generation, silencing harmful genetic elements during reprogramming is of paramount importance. In the male germ line, *de novo* methylation of transposons occurs at the same time point as genomic imprinting, and makes use of both DNMT3A and DNMT3B, in a locus-specific manner (Kato et al., 2007). DNMT3L is required for all *de novo* methylation of germ line transposons, and homozygous *Dnmt3L* knockout male mice are marked by meiotic failure and loss of germ line cells, thus are infertile (Bourc'his and Bestor, 2004). In the case of the male germ line, small RNAs from the piwi family (piRNAs) are required for *de novo* methylation of transposons (Kuramochi-Miyagawa et al., 2008). Mutants in the piRNA pathway result in a male phenotype reminiscent of the *Dnmt3L* knockout (Aravin et al., 2007; Kuramochi-Miyagawa et al., 2004). Epistasy studies indicate that the piRNA pathway likely acts upstream of DNMTs, or possibly other chromatin modifiers, suggesting it potentially guides the *de novo* targeting (Aravin et al., 2008). There is also at least one male imprinted locus, termed *Rasgrf1*, which requires both piRNAs and long non-coding RNAs in order to target the *de novo* methyltransferase (Watanabe et al., 2011). As will be discussed in the next section, long and small RNAs guiding the *de novo* methylation machinery is likely an example of convergent evolution between plants and mammals.

As mentioned, a second wave of reprogramming occurs in the embryo, immediately after formation of the zygote (Feng et al., 2010b). Active demethylation of the paternal genome occurs soon after fertilization (Reik, 2007). Although the exact mechanism of paternal demethylation is still an active area of research, recent studies implicate hydroxymethylation of methyl-cytosines in the paternal zygotic genome as a precursor to demethylation (Gu et al., 2011; Inoue and Zhang, 2011; Wossidlo et al., 2011). Specifically, the enzyme TET3 oxides the methyl group on the cytosine, generating a hydroxymethylcytosine (hmC). The enzyme has been shown to further modify the hmC into carboxylcytosine (He et al., 2011; Ito et al., 2011). This modified nucleotide is recognized and removed by thymine-DNA glycosylase (TDG), a component of the base excision repair (BER) pathway (He et al., 2011). Thus, in the final step of the BER pathway, an unmethylated cytosine is incorporated where the methyl-cytosine used to reside. The female genome seems to protect itself from active demethylation (Farthing et al., 2008; Mayer et al., 2000), which may well underscore parental conflict theory in action. Both paternal and maternal genomes undergo passive demethylation, simply by limiting DNMT1 activity in the nucleus (Feng et al., 2010b). However, a mechanism is in place to simultaneously protect imprinted loci from losing the mark, likely involving the STELLA protein (Nakamura et al., 2007). After the demethylation program has completed, the remethylation in the early embryo is almost entirely dependent on DNMT3B, unlike what is observed in the germ line (Borgel et al., 2010). DNMT3L also appears dispensable, as both male and female mice devoid of the protein are morphologically normal, with the notable exception of their germ line cell lineages, which renders both males and female *Dnmt3L* homozygous knockouts infertile (Bourc'his and Bestor, 2004; Bourc'his et al., 2001).

RNA-directed DNA Methylation in Plants

As mentioned, flowering plants do not undergo the waves of epigenetic reprogramming that is observed in mammals in the germ line and the embryo. There is however a wave of active demethylation that occurs in the endosperm, via the DNA glycosylase DEMETER (DME) (Choi et al., 2002). The endosperm provides the nutrients for the developing embryo, but does not make any genetic contribution to the adult plant. The developmental consequences of losing the active demethylation is far from trivial, as absence of the DME protein in the endosperm leads to seed abortion (Choi et al., 2002). The methylation patterns across generations of plants, however, are meiotically stable (Law and Jacobsen, 2010). Unlike mammals, wherein mutants lacking either *de novo* DNA methyltransferase die early in development, Arabidopsis *drm2* mutants are virtually morphologically wild-type, and perfectly viable (Cao and Jacobsen, 2002b). Moreover, genome wide analyses show that global methylation is minimally affected in *drm2* mutants, especially in comparison with *met1* and *cmt3* (Cokus et al., 2008). What then is the major function of the *de novo* methyltransferase in plants?

Two studies shine light on the importance of the DRM2 pathway in Arabidopsis. The first study makes use of the *FWA* imprinted locus, which is silenced when the tandem repeats in its 5' UTR are methylated (Soppe et al., 2000). When an unmethylated *FWA* transgene is introduced into wild-type plants, the tandem repeats are efficiently methylated. However, in *drm2* mutants the transgene never has methylation established, resulting in *FWA* expression and plants are marked by a late-flowering phenotype (Cao and Jacobsen, 2002b). This system indicates the DRM2's role in recognizing invasive genetic elements for silencing. Other transgene systems have recapitulated this phenomenon (Henderson and Jacobsen, 2008). A second study examined inbred lines generated in mutants for maintenance of DNA methylation. When these inbred lines also contained homozygous mutations in DRM2 pathway components, but were otherwise genetically wild-type, most of the genome was not "remethylatable" (Teixeira et al., 2009). These two studies underscore the importance of the DRM2 pathway as a

surveillance system that recognizes and targets potentially harmful genetic elements for silencing.

The basis of the recognition system is a byzantine pathway that depends both on small interfering RNAs (siRNAs) and long non-coding RNAs (lncRNAs) acting in parallel, in a system known as RNA-directed DNA methylation (RdDM) (Figure 1-2). The siRNAs are generated by the plant specific DNA-DEPENDENT RNA POLYMERASE IV (Pol IV) (Dalmay et al., 2000). The Pol IV complex includes RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), which presumably generates a double-stranded RNA species form the Pol IV transcript (Law et al., 2011). There are four dicers in Arabidopsis, and there is a degree of redundant activity RdDM (Henderson et al., 2006). However, the primary siRNAs associated with the pathway are 24nt in length, and produced by DICER-LIKE 3 (Herr et al., 2005; Xie et al., 2004). There are two ARGONAUTE (AGO) proteins associated with RdDM—AGO4 and AGO6—the stronger DNA methylation phenotype being observed in *ago4* mutants (Zheng et al., 2007; Zilberman et al., 2003; Zilberman et al., 2004). AGO4 localizes to a nuclear body, known as the Cajal (Li et al., 2006; Pontes et al., 2006). Cajal bodies are known to be involved with processing ribonucleoprotein complexes, although its exact function in the context of RdDM remains elusive.

A second plant specific RNA Polymerase, Pol V, is required for RdDM and (Kanno et al., 2005) generates lncRNAs, which are detectable at intergenic non-coding (IGN) regions (Wierzbicki et al., 2008). The transcription of Pol V is dependent on the DDR complex: the putative chromatin remodeler DRD1, an SMC-hinge containing protein DMS3, and RDM1, which has methylated DNA-binding capacity (Ausin et al., 2009; Law et al., 2010). The Pol V transcript then serves as a scaffold, which recruits AGO4 to chromatin (Wierzbicki et al., 2009). A protein with homology to the yeast elongation factor SPT5, known in Arabidopsis as KTF1/SPT5L, interacts with both Pol V and AGO4, although it apparently is not required for Pol V transcription, and its exact function remains unclear (Bies-Etheve et al., 2009; Greenberg et al., 2011; He et al., 2009;

Rowley et al., 2011). In the absence of Pol V, 24nt siRNAs are severely reduced genome wide, although it is unclear if the siRNAs are being produced off the transcript, or Pol V activity serves as some sort of feedback for Pol IV (Mosher et al., 2008). In null mutants for any of any of these proteins, DRM2-mediated methylation is largely—or completely—lost, however the exact mechanism for targeting DRM2 to the DNA is still unclear. A possible intermediary is the IDN2 complex, which binds double-stranded RNA (Ausin et al., 2009). The role of this complex will be elaborated in Chapter 4.

Importantly, all of these components have also been shown to be required for DNA methylation establishment in all sequence contexts (Ausin et al., 2009; Chan et al., 2004; Greenberg et al., 2011). Therefore, siRNAs being produced in trans in the genome can recognize specific deleterious sequences based on homology. Additionally, there must be the cis-acting Pol V transcribing to recruit AGO4. The question begs: what recruits the DDR complex and Pol V? Is it based on the genetic sequence of a given element, its chromatin state, or siRNAs? This remains a major active area of research in the field of RdDM.

Role of Histone Marks in Arabidopsis DNA Methylation

The patterns of chromatin marks that correlate with cytosine DNA methylation often times play a role in the recruitment and/or activity of methyltransferases. For example, earlier the relationship between H3K9 methyltransferases and CMT3 was explicated. In the case of mammals, DNMT3L binding to H3K4m0 stimulates DNMT3A activity. In fact, an H3K4 demethylase termed KDM1B is required for genomic imprinting in the female germ line (Ciccone et al., 2009). The data from this study suggests an epistatic relationship between a chromatin effector and a DNA methyltransferase. The regulation of RNA Pol II in all eukaryotic systems is heavily impacted by chromatin (Cairns, 2009). Given that two paralogs of Pol II are required for RdDM in

Arabidopsis—one of which acts in cis—it would not be surprising if chromatin impacts the DRM2 pathway.

There are some hints that indeed chromatin marks are involved in RdDM regulation. Two partially redundant paralogs termed SUVH2 and SUVH9 have high similarity to the H3K9 methyltransferases in the CMT3 pathway, and are required DRM2-mediated maintenance and establishment methylation (Johnson et al., 2008). Like SUVH4/5/6, these proteins contain an SRA domain that binds to methylated DNA. Although the histone methyltransferase activity of SUVH2/9 has not been demonstrated *in vitro*, there is the distinct possibility of a similar relationship between H3K9 methylation with RdDM machinery as observed with the CMT3 pathway components.

The histone 2B (H2B) deubiquitinase UBP26 has also been shown to be a necessary factor in DRM2 maintenance methylation (Sridhar et al., 2007). H2B ubiquitination (H2Bub) is a mark that is highly correlated with active transcription, and H3K4 methylation (Roudier et al., 2011). Given the relationship between between H2Bub and H3K4m2/m3, as well as the knowledge that a H3K3 demethylase is required for DNMT3A activity in some instances, we were interested in the possible role of Arabidopsis proteins performing analogous function.

In the next five chapters, I will describe my methods to further elucidate the factors required in the RdDM pathway in Arabidopsis by both a forward and reverse genetic screen. I will also describe the genetic, biochemical, and genomic analyses I employed in order to characterize a novel complex in the RdDM pathway. Finally, I will describe my efforts to resolve the relationship between H3K4 demethylases and RdDM. The findings from the latter studies revealed the surprising discovery that the requirements for DNA methylation establishment are different from those in RdDM maintenance activity.

Figure Legends

Figure 1-1. DNA Methylation Maintenance Pathways in Mammals and Arabidopsis

(A) Mammals and plants share a conserved pathway for maintaining cystosine DNA methylation in the CG context. Following DNA replication, a hemimethylated site is recognized by the SRA domain-containing protein UHRF1 in mammal or the VIM proteins in plants. UHRF1/VIM recruits DNMT1 in mammals or MET1 in plants to methylate the cytosine across the axis of symmetry of the DNA strands.

(B) The plant-specific CMT3 pathway. CMT3 contains a chromodomain that recognizes H3K9m2. Methylation of this lysine residue is maintained by the KYP family of H3K9 methyltransferases, all of which contain a methyl-cytosine binding SRA domain. Thus, a self-reinforcing loop is in place to maintain the two heterochromatic marks

Figure 1-2. RNA-directed DNA Methylation in Arabidopsis

The plant specific RNA polymerase Pol IV generates a transcript that is converted into double stranded RNA by RDR2. This double-stranded species is diced into 24nt siRNAs by DCL3, and these siRNAs are primarily loaded into AGO4. AGO4 is recruited to chromatin by the transcript generated by Pol V. Pol V transcription is dependent on the DDR complex (DRD1, DMS3, and RDM1). The IDN2 complex possibly recognizes the double stranded RNA formed between the siRNA and Pol V transcript, and serves to direct DRM2 to DNA. UBP26, JMJ14, and LDL1/2 all impact chromatin, which affects the RdDM pathway at an unknown step.

Figure 1-1

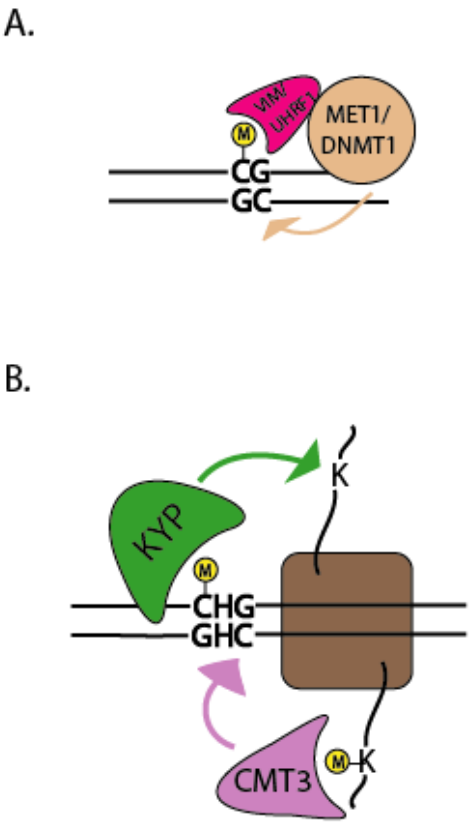
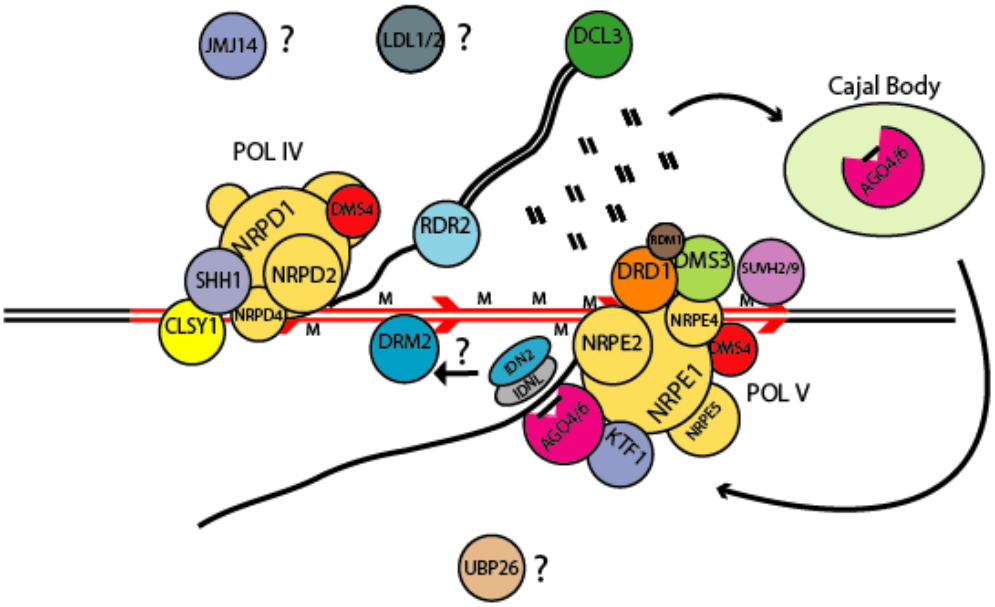


Figure 1-2


JMJ14
?
LDL1/2
?
DCL3
Cajal Body
AGO4/6
POL IV
NRPD1
DMS4
RDR2
SHH1
NRPD2
CLSY1
NRPD4
M
DRM2
?
RDM1
DRD1
DMS3
SUVH2/9
NRPE4
NRPE2
DMS4
IDN2
IDNL
NRPE1
POL V
AGO4/6
KTF1
NRPE5
UBP26
?

REFERENCES

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell *31*, 785-799.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. Science *316*, 744-747.

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature *455*, 818-821.

Ausin, I., Mockler, T.C., Chory, J., and Jacobsen, S.E. (2009). IDN1 and IDN2 are required for de novo DNA methylation in Arabidopsis thaliana. Nat Struct Mol Biol *16*, 1325-1327.

Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature *455*, 822-825.

Bies-Etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., Cooke, R., and Lagrange, T. (2009). RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. EMBO Rep *10*, 649-654.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet *42*, 1093-1100.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96-99.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536-2539.

Bujnicki, J.M., and Radlinska, M. (1999). Molecular evolution of DNA-(cytosine-N4) methyltransferases: evidence for their polyphyletic origin. Nucleic Acids Res *27*, 4501-4509.

Cairns, B.R. (2009). The logic of chromatin architecture and remodelling at promoters. Nature *461*, 193-198.

Cao, X., and Jacobsen, S.E. (2002a). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A *99 Suppl 4*, 16491-16498.

Cao, X., and Jacobsen, S.E. (2002b). Role of the Arabidopsis DRM methyltransferases in *de novo* DNA methylation and gene silencing. Current Biology *12*, 1138-1144.

Chan, S.W., Henderson, I.R., and Jacobsen, S.E. (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet *6*, 351-360.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Cheng, X. (1995). Structure and function of DNA methyltransferases. Annu Rev Biophys Biomol Struct *24*, 293-318.

Cheng, X., and Roberts, R.J. (2001). AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res *29*, 3784-3795.

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

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. Cell *110*, 33-42.

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science *277*, 1996-2000.

Ciccone, D.N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E., and Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature *461*, 415-418.

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

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. (2000). An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell *101*, 543-553.

Farthing, C.R., Ficz, G., Ng, R.K., Chan, C.F., Andrews, S., Dean, W., Hemberger, M., and Reik, W. (2008). Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS Genet *4*, e1000116.

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

Feng, S., Jacobsen, S.E., and Reik, W. (2010b). Epigenetic reprogramming in plant and animal development. Science *330*, 622-627.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science *311*, 395-398.

Greenberg, M.V., Ausin, I., Chan, S.W., Cokus, S.J., Cuperus, J.T., Feng, S., Law, J.A., Chu, C., Pellegrini, M., Carrington, J.C.*, et al.* (2011). Identification of genes required for de novo DNA methylation in Arabidopsis. Epigenetics *6*.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G.*, et al.* (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606-610.

Hashimoto, H., Horton, J.R., Zhang, X., Bostick, M., Jacobsen, S.E., and Cheng, X. (2008). The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature *455*, 826-829.

He, X.J., Hsu, Y.F., Zhu, S., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H.L., Wang, C.S., Jin, H., and Zhu, J.K. (2009). An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. Cell *137*, 498-508.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L.*, et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303-1307.

Hemberger, M., Dean, W., and Reik, W. (2009). Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. Nat Rev Mol Cell Biol *10*, 526-537.

Henderson, I.R., and Jacobsen, S.E. (2008). Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. Genes Dev *22*, 1597-1606.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat Genet *38*, 721-725.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science *308*, 118-120.

Hsieh, T.F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D.*, et al.* (2011). Regulation of imprinted gene expression in Arabidopsis endosperm. Proc Natl Acad Sci U S A *108*, 1755-1762.

Huh, J.H., Bauer, M.J., Hsieh, T.F., and Fischer, R.L. (2008). Cellular programming of plant gene imprinting. Cell *132*, 735-744.

Inoue, A., and Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science *334*, 194.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300-1303.

Iwasa, Y., and Pomiankowski, A. (2001). The evolution of X-linked genomic imprinting. Genetics *158*, 1801-1809.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556-560.

Jacobsen, S.E., and Meyerowitz, E.M. (1997). Hypermethylated *SUPERMAN* epigenetic alleles in arabidopsis. Science *277*, 1100-1103.

Jeltsch, A. (2002). Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem *3*, 274-293.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature *449*, 248-251.

Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I.R., Callis, J., and Jacobsen, S. (2007). The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation. Current Biology *17*, 379.

Johnson, L.M., Law, J.A., Khattar, A., Henderson, I.R., and Jacobsen, S.E. (2008). SRA-domain proteins required for DRM2-mediated de novo DNA methylation. PLoS Genet *4*, e1000280.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature *429*, 900-903.

Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat Genet *37*, 761-765.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Hum Mol Genet *16*, 2272-2280.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science *303*, 521-523.

Klimasauskas, S., Kumar, S., Roberts, R.J., and Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. Cell *76*, 357-369.

Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J., and Wilson, G.G. (1994). The DNA (cytosine-5) methyltransferases. Nucleic Acids Res *22*, 1-10.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W.*, et al.* (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development *131*, 839-849.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W.*, et al.* (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev *22*, 908-917.

Law, J.A., Ausin, I., Johnson, L.M., Vashisht, A.A., Zhu, J.K., Wohlschlegel, J.A., and Jacobsen, S.E. (2010). A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in Arabidopsis. Curr Biol *20*, 951-956.

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

Law, J.A., Vashisht, A.A., Wohlschlegel, J.A., and Jacobsen, S.E. (2011). SHH1, a Homeodomain Protein Required for DNA Methylation, As Well As RDR2, RDM4, and Chromatin Remodeling Factors, Associate with RNA Polymerase IV. PLoS Genet *7*, e1002195.

Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. (2006). An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. Cell *126*, 93-106.

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

Lindroth, A.M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I.*, et al.* (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. Embo J *23*, 4286-4296.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S.*, et al.* (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature *471*, 68-73.

Malagnac, F., Bartee, L., and Bender, J. (2002). An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. EMBO J *21*, 6842-6852.

Martienssen, R.A., and Colot, V. (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science *293*, 1070-1074.

May, M.S., and Hattman, S. (1975). Analysis of bacteriophage deoxyribonucleic acid sequences methylated by host- and R-factor-controlled enzymes. J Bacteriol *123*, 768-770.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501-502.

Moore, T., and Haig, D. (1991). Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet *7*, 45-49.

Mosher, R.A., Schwach, F., Studholme, D., and Baulcombe, D.C. (2008). PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. Proc Natl Acad Sci U S A *105*, 3145-3150.

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M.*, et al.* (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol *9*, 64-71.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. Cell *99*, 247-257.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D.*, et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

Peter, I.S., and Davidson, E.H. (2011). Evolution of gene regulatory networks controlling body plan development. Cell *144*, 970-985.

Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M.C., and Reuter, G. (2009). Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. Nat Genet *41*, 696-702.

Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. (2006). The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. Cell *126*, 79-92.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature *335*, 683-689.

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I., and Cuzin, F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature *441*, 469-474.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425-432.

Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nat Rev Genet *2*, 21-32.

Roudier, F., Ahmed, I., Berard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L.*, et al.* (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J *30*, 1928-1938.

Rowley, M.J., Avrutsky, M.I., Sifuentes, C.J., Pereira, L., and Wierzbicki, A.T. (2011). Independent chromatin binding of ARGONAUTE4 and SPT5L/KTF1 mediates transcriptional gene silencing. PLoS Genet *7*, e1002120.

Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat Rev Genet *9*, 129-140.

Saze, H., and Kakutani, T. (2007). Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. EMBO J *26*, 3641-3652.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K.*, et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74-79.

Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The Late Flowering Phenotype of fwa Mutants Is Caused by Gain-of-Function Epigenetic Alleles of a Homeodomain Gene. Mol Cell *6*, 791-802.

Spencer, H.G., Clark, A.G., and Feldman, M.W. (1999). Genetic conflicts and the evolutionary origin of genomic imprinting. Trends Ecol Evol *14*, 197-201.

Sridhar, V.V., Kapoor, A., Zhang, K., Zhu, J., Zhou, T., Hasegawa, P.M., Bressan, R.A., and Zhu, J.K. (2007). Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. Nature *447*, 735-738.

Stokes, T.L., Kunkel, B.N., and Richards, E.J. (2002). Epigenetic variation in Arabidopsis disease resistance. Genes Dev *16*, 171-182.

Stokes, T.L., and Richards, E.J. (2002). Induced instability of two Arabidopsis constitutive pathogen-response alleles. Proc Natl Acad Sci U S A *99*, 7792-7796.

Surani, M.A., Durcova-Hills, G., Hajkova, P., Hayashi, K., and Tee, W.W. (2008). Germ line, stem cells, and epigenetic reprogramming. Cold Spring Harb Symp Quant Biol *73*, 9-15.

Teixeira, F.K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M.F.*, et al.* (2009). A role for RNAi in the selective correction of DNA methylation defects. Science *323*, 1600-1604.

Watanabe, T., Tomizawa, S., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S., Iida, N., Hoki, Y., Murphy, P.J., Toyoda, A.*, et al.* (2011). Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. Science *332*, 848-852.

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. Cell *135*, 635-648.

Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet *41*, 630-634.

Wilson, G.G., and Murray, N.E. (1991). Restriction and modification systems. Annu Rev Genet *25*, 585-627.

Woo, H., Pontes, O., Pickart, C.M., and Richards, E.J. (2007). VIM1, a methylcytosine-binding protein required for centromeric heterchromatinization. Genes & Dev *21*, 267-277.

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

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun *2*, 241.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol *2*, E104.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet *13*, 335-340.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916-919.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E.*, et al.* (2006). Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. Cell *126*, 1189-1201.

Zheng, X., Zhu, J., Kapoor, A., and Zhu, J.K. (2007). Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. Embo J *26*, 1691-1701.

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

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. (2004). Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. Curr Biol *14*, 1214-1220.

CHAPTER 2

Identification of genes required for

de novo DNA methylation in Arabidopsis.

Introduction

Cytosine DNA methylation is an ancient form of epigenetic information found in all kingdoms of eukaryotes. Functions for DNA methylation-mediated gene silencing include imprinting and repression of transposable elements. In mammals, methylation occurs almost exclusively in the CG context. Maintenance of the methyl-mark persists through function of the DNA methyltransferase, DNMT1, whereas the initial, or *de novo*, methylation of unmodified DNA occurs through DNMT3 activity. The importance of DNA methylation for proper development is underscored by embryonic lethality in knockout mice for either type of methyltransferase (Li et al., 1992; Okano et al., 1999).

The flowering plant *Arabidopsis thaliana* serves as a powerful model to analyze the function of DNA methylation since it contains orthologs to both DNMT1 and DNMT3, termed METHYLTRANSFERASE1 (MET1) and DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), respectively (Cao and Jacobsen, 2002a; Saze et al., 2003). Moreover, unlike mammals, Arabidopsis plants that contain homozygous null mutations for either gene are viable. In addition, Arabidopsis contains a third methyltransferase gene: the plant-specific *CHROMOMETHYLASE 3* (*CMT3*). Unlike mammals, plants are able to methylate cytosines in three contexts: CG, CHG (where H is A, C, or T), and CHH. Although there is a degree of redundancy, MET1 largely carries out CG methylation, CMT3 performs CHG, and DRM2 is the primary CHH methyltransferase (Cao and Jacobsen, 2002a; Lindroth et al., 2001; Saze et al., 2003).

The CHH context is of particular note because it lacks inherent symmetry across the DNA strand. Whereas methylation in symmetric contexts can be recognized as hemimethylated DNA on both daughter strands of DNA after replication and therefore be maintained passively (Bostick et al., 2007), maintenance of asymmetric methylation must operate via an active signal. In the last decade a flood of studies have implicated a 24-nucleotide species of siRNAs as

guiding DRM2 in a process known as RNA-directed DNA Methylation (RdDM) (Aufsatz et al., 2002a; Cao et al., 2003; Herr et al., 2005; Zilberman et al., 2004).

In addition to maintaining pre-existing CHH methylation, RdDM via DRM2 is needed to establish DNA methylation on unmethylated sequences, such as an incoming (*FLOWERING WAGENINGEN*) *FWA* transgene (Cao and Jacobsen, 2002b; Chan et al., 2004), *FWA* is a maternally imprinted homeodomain transcription factor. In vegetative tissue *FWA* is silenced as a result of methylation in its promoter and 5' UTR (Kinoshita et al., 2004; Soppe et al., 2000). Heritable unmethylated *fwa-1* epialleles are dominant and ectopic expression causes a late-flowering phenotype. RdDM mutant plants are incapable of *de novo* methylating *FWA* upon transformation, thus flower late.

Several aspects of the RdDM process remain a mystery. For example it is unclear what downstream effectors contribute to the silent state following *de novo* methylation. Additionally, it is uncertain how a 24-nt siRNA-loaded ARGONAUTE4 physically targets DRM2 to specific loci. However, perhaps the most compelling question is how RdDM machinery is able to recognize particular unmethylated DNA sequences. Although repetitive elements such as transposons are associated with small RNAs and DNA methylation, there is no known sequence specificity or secondary structure that is a hallmark of recruitment, let alone an endogenous factor that serves to respond to invasive DNA (Chan et al., 2006).

In order to further clarify the process of the initial establishment of methylation, we employed both forward and reverse genetic screens taking advantage of the *FWA* late-flowering phenotype. In the first forward mutagenesis screen specifically designed to discover mutations that block the initial establishment of methylation, a broad picture of the RdDM pathway emerged. We utilized both traditional and emerging whole-genome sequencing approaches to identify mutations from the screen. Of particular note, the recently described *SUPPRESSOR OF TY INSERTION 5-LIKE* (*SPT5-LIKE*) / KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (*KTF1*) (Bies-Etheve et al., 2009; He et al., 2009c; Huang et al., 2009) and

CLASSY1 (*CLSY1*) (Smith et al., 2007) genes were discovered by our approaches to be required for initial establishment of methylation.

Results and Discussion

A forward screen to discover *de novo* methylation mutants

In an effort to increase our knowledge of *de novo* methylation in Arabidopsis, we took an unbiased approach to screen for mutations that block the establishment of DNA methylation (Figure 2-1). This was accomplished by transforming plants with an *FWA* transgene and screening for mutants that block the establishment of silencing of *FWA* and therefore produce a late flowering phenotype. Columbia-0 (Col-0) seeds were treated with ethyl methanesulfonate (EMS) and roughly 900 mutagenized M1 lines were allowed to self-pollinate to produce individual M2 families. Due to concerns that unrelated flowering time mutations would mimic the *FWA* hypomethylation phenotype, approximately 110 M2 lines that exhibited late flowering were discarded. The remaining M2 lines were transformed with the *FWA* transgene and the first generation of Basta-resistant transformants (T1s) was screened for late flowering plants. Roughly 300 T1 lines displayed a late-flowering phenotype upon transformation. The large number of late flowering plants is likely due to liberal thresholds used to score the phenotype, incomplete silencing of *FWA* that occurs even in wild type plants, natural variation, and the segregation of flowering time mutants in a host of genes. RT-PCR from leaf material confirmed that *FWA* was expressed above wild-type levels in 128 lines.

As mentioned previously, all known mutants that block *de novo* methylation are components of RdDM. Therefore, to identify an assay for confirming and classifying our mutants, we searched for an endogenous locus that has an siRNA-dependent methylation state which can be examined with relative facility. The *MEDEA-INTERGENIC SUBTELOMERIC*

REPEATS (*MEA-ISR*) are a set of tandem repeats downstream of the *MEDEA* gene which are targeted by DRM2 (Cao and Jacobsen, 2002a). Using the MspI restriction endonuclease, non-CG methylation at *MEA-ISR* was examined by Southern blot analysis. Wild type plants display roughly a 1:1 ratio of methylated to unmethylated bands. In strong mutants, such as null *drm2* and *ago4* alleles, non-CG methylation is virtually eliminated (Figure 2-2 A). Weaker mutants, such as null mutations in *dicer-like 3* (*dcl3*)—the dicer protein most strongly associated with RdDM—exhibit an intermediate banding pattern (Figure 2-2 A). Thus, the 128 *de novo* methylation mutants were sub-classified into three categories based on their *MEA-ISR* phenotype: *MEA-ISR* normal (wild-type), *MEA-ISR* eliminated (phenocopy of *drm2*), and *MEA-ISR* reduced (phenocopy of *dcl3*) (Figure 2-1).

Based on the Southern blot phenotypes, seven lines fell into the *MEA-ISR* eliminated class, and ten lines fell into the *MEA-ISR* reduced class; the remaining 111 *de novo* mutant lines are termed *MEA-ISR* normal. Mutations mapped in the first two classes will be discussed.

Map based approach to discover mutations

In order to positionally clone the mutations that block *de novo* methylation, we took advantage of the maintenance methylation defect in the *MEA-ISR* eliminated and *MEA-ISR* reduced classes. Mutant lines were outcrossed to Landsberg *erecta* (L*er*), and individual plants from segregating F2 populations were assayed for a *MEA-ISR* Southern blot loss of methylation phenotype. We then utilized a series of PCR based molecular markers to map each of the mutations. We included markers that are tightly linked to the genes already known or suspected to affect *de novo* methylation in order to most efficiently identify the genes harboring mutations.

Linkage analysis followed by sequencing led to the discovery of seven mutations in six different genes that phenocopy *drm2* (*MEA-ISR* eliminated class), and two mutations in two different genes that phenocopy *dcl3* (*MEA-ISR* reduced class) (Figure 2-2 A and 2-2 B, and

described below). Allelism crosses with null T-DNA alleles were used to confirm gene identities. Included in the list are alleles of genes that heretofore have not been previously shown to be required for the initial establishment of methylation: *KTF1* and *CLSY1*.

Analysis of *de novo* methylation components upstream of small RNA production

The species of small RNAs most frequently associated with RdDM are 24 nucleotides in length (Aufsatz et al., 2002a; Zhang et al., 2007). Three components required to synthesize this class of siRNAs were isolated in the screen: *rdr2-6*, *dcl3-6*, and *clsy1-7*. RNA-DEPENDANT RNA POLYMERASE 2 (RDR2) likely produces double-stranded RNA from transcripts generated by the plant-specific RNA Polymerase IV (Pol IV) (Herr et al., 2005). Here, we identified an allele of *rdr2* that contains a premature stop codon well upstream of the RNA-dependent RNA Polymerase (RdRP) domain (Figure 2-2 B), which therefore is likely a null allele. DCL3 is proposed to process the RDR2 double stranded RNA product into 24-mer siRNAs (Henderson et al., 2006; Xie et al., 2004). We identified a glutamate-to-lysine missense mutation in a highly conserved DCL3 residue in one of the Ribonuclease III (RNAseIII) domains (Figure 2-2 B). The allele recovered in the screen appears to have a weaker effect on RdDM than a *dcl3* null allele (Figures 2-3 A and 2-3 B). This perhaps is due to the fact that DCL3 contains two RNAseIII domains, so while one is impaired, the second retains its enzymatic activity (Figure 2-2 B). Despite this, the mutation still affects *de novo* methylation of the *FWA* transgene such that it causes significantly reduced transgene methylation causing retarded flowering time compared to the wild type control (Figures 2-3 C and 2-3 D). Genetic complementation with a null *dcl3* allele confirms that the missense in *dcl3-6* causes the impaired methylation phenotype (Figures 2-3 A and 2-3 B). The values for *FWA* transgene methylation are much lower than the endogenous copies, even in wild type plants, due to incomplete methylation that is observed in the first generation of transformants.

A third mutation in a gene required for siRNA synthesis is a proline-to-leucine missense in *clsy1-7* (Figures 2-2 A and 2-2 B). CLSY1 is a member of a SNF2-like domain-containing class of proteins that also contain a helicase domain near the carboxyl-terminus, and structural modeling predicts a putative chromatin remodeling function for the CLSY1 protein (Smith et al., 2007). Genetic data suggests that CLSY1 is required for proper localization of both NRPD1—the largest subunit of Pol IV—and RDR2, suggesting it acts upstream of both of these components (Smith et al., 2007). The mutation discovered here is adjacent to both the putative ATP binding site and DNA backbone binding residues within the SNF2-like domain (Figure 2-2 B) (Smith et al., 2007). The proline is a conserved residue among Arabidopsis chromatin remodeling proteins, and its mutation in *clsy1-7* may result in an altered structural conformation that affects either or both of these functions. We found that a null *clsy1* T-DNA insertion mutant, which we termed *clsy1-8*, did not completely block *de novo* methylation, as assayed by *FWA* transformation (Figures 2-4 A and 2-4 C). Concordantly, the maintenance methylation phenotype at the *MEA-ISR* locus was also incomplete, with residual non-CG methylation present at the MspI restriction site (Figure 2-4 B). In order to confirm that CLSY1 is required for RdDM maintenance methylation, we examined the *FWA* endogene in the mutant (Figure 2-4 D). Indeed, in *clsy1* the non-CG methylation is sharply reduced. A possible explanation for the weaker phenotype of *clsy1* mutants compared to other components in the pathway may be partial redundancy with another member of the subfamily of SNF2-like proteins. A likely candidate is CLSY2, which contains 81% sequence homology with CLSY1.

Analysis of *de novo* methylation components downstream of siRNA production

Subsequent to biogenesis, 24-nt siRNAs target RdDM machinery to homologous DNA sequences for DRM2-mediated methylation. Mutations in genes downstream of siRNA production do not exhibit severe defects in small RNA accumulation at many RdDM target loci,

however non-CG methylation is sharply reduced or eliminated (Henderson and Jacobsen, 2007; Law and Jacobsen, 2009; Matzke et al., 2009). The *FWA* screening approach produced six alleles in five genes required for *de novo* DNA methylation but not strictly required for siRNA biogenesis: *ago4-4*, *nrpe1-13*, *drd1-10*, *drd1-11*, *drm2-3*, and *ktf1-5* (Figure 2-2 B).

Small RNAs associated with RdDM are primarily loaded into AGO4 (Qi et al., 2006; Zilberman et al., 2004). AGO4 contains two conserved domains termed PAZ and PIWI. The PIWI domain contains a triad of catalytic residues (D660, D742 and H874) that are conserved in argonaute proteins in plants, humans and *S. pombe*, and have been previously shown to be required for AGO4 slicer activity and RdDM (Qi et al., 2006). An allele of *ago4* identified from this screen contains a splice donor mutation that causes a frame shift and eventual premature stop downstream of amino acid 630 (Figure 2-2 B). Given that the three catalytic triad residues all lie downstream of the mutation, this allele likely renders the protein catalytically inactive, explaining the strong *de novo* methylation phenotype and the strong loss of methylation *MEA-ISR* phenotype (Figure 2-2 A). AGO4 acts partially redundantly with its close family member AGO6 in maintenance DNA methylation, and the *ago6-2* mutant shows a reduced *MEA-ISR* methylation phenotype (Figure 2-5 A) (Zheng et al., 2007). Therefore, we wanted to confirm that *ago6* similarly affects *de novo* DNA methylation. Analysis of *FWA*-transformed *ago6-2* plants confirmed that indeed there is a late-flowering phenotype (Figure 2-5 C) (Note: the variation in flowering time for Col-0 + *FWA* from different experiments is likely due to differences in growth conditions).

Previous studies have shown that AGO4 colocalizes with NRPE1, the largest subunit of the plant-specific RNA Polymerase V (Pol V) in the nucleus (Li et al., 2006). Pol V produces a transcript that is necessary to recruit AGO4 to chromatin of methylated loci(Wierzbicki et al., 2009). Additionally, NRPE1 contains a hydrophilic (S/G/A/D/E/K-rich) domain in its C-terminal that interacts with AGO4 (El-Shami et al., 2007; Li et al., 2006). The interaction is dependent on conserved tryptophan-glycine/glycine-tryptophan (WG/GW) repeats within NRPE1 (El-Shami et

al., 2007). An allele of *nrpe1* arose from the *FWA* screen containing a premature stop codon at amino acid position 1693 (Figure 2-2 B). RT-PCR analysis indicated that the *nrpe1-13* mutant transcript is also significantly reduced, possibly by nonsense-mediated mRNA decay (data not shown). This is consistent with the strong loss-of-function mutant phenotype of this allele.

Because NRPE1 is required for *de novo* methylation, we also performed *FWA* transformation on an allele of the recently described Pol V-specific subunit *NRPE5a* (Lahmy et al., 2009; Ream et al., 2009). The *nrpe5a-1* mutant indeed showed a partial late flowering phenotype after transformation, showing that NRPE5a is also required for establishment of methylation (Figure 2-5 B). Consistent with this effect, the *nrpe5a-1* mutant also showed a partial maintenance methylation phenotype at *MEA-ISR* (Figure 2-5 A). A possible explanation for why this mutant does not have as severe a methylation defect as *nrpe1* may be because of partial redundancy with its paralogs, NRPE5b and NRPE5c (Ream et al., 2009).

Pol V transcriptional activity is dependent on the SNF2-like putative chromatin remodeler DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1) (Wierzbicki et al., 2008). Null alleles of *drd1* display a severe decrease in non-CG methylation at RdDM target loci (Kanno et al., 2005; Kanno et al., 2004). Two alleles of *drd1* emerged from the *FWA* screen, both of which exhibited a strong loss of methylation at *MEA-ISR* (Figure 2-2 A). *drd1-11* contains a missense mutation in a highly conserved glycine within the SNF2-like domain (Figure 2-2 B). The second allele, *drd1-10*, also is a glycine missense, however it occurs in a non-conserved region of DRD1 (Figure 2-2 B). This glycine is therefore either critical for the function of DRD1, or it is possible that the mutation causes structural changes rendering the protein non-functional.

WG/GW repeats—such as those contained in NRPE1—have been shown to interact with argonaute proteins, forming a motif termed the ‘AGO-hook’ (Till et al., 2007). Bioinformatic, biochemical, and forward genetic approaches all have converged on an Arabidopsis AGO-hook-containing protein that is required for RdDM (Bies-Etheve et al., 2009; He et al., 2009c; Huang et al., 2009). The protein, named KTF1, also contains homology with the yeast transcription

elongation factor SPT5 (Bies-Etheve et al., 2009). The *FWA* screen approach produced an allele of *ktf1* that contains a premature stop in amino acid 378 demonstrating its requirement for *de novo* methylation, observed by late-flowering *FWA* transformants (Figure 2-6 A). Additional analysis at endogenous loci indicates a substantial loss in non-CG methylation at DRM2-dependent sites in the *ktf1-5* mutant (Figures 2-6 B, 2-6 C, and 2-6 D).

KTF1 does not appear to be required for the production of 24-nucleotide siRNAs at all RdDM-targeted loci (Figure 2-6 E) (Bies-Etheve et al., 2009; He et al., 2009c). This characteristic places it in the RdDM pathway downstream of Pol IV-RDR2 activity. Although KTF1 exhibits homology to RNA Polymerase II (Pol II) elongation factors in other eukaryotes, it does not appear to affect the accumulation of Pol V transcripts showing that it is not required for Pol V transcription (He et al., 2009c). However, KTF1 does interact with both AGO4 and Pol V, and has RNA-binding capability (Bies-Etheve et al., 2009; He et al., 2009c; Huang et al., 2009). A potential function could be recruiting AGO4 to chromatin at sites of Pol V transcription.

Finally, we isolated a strong allele of the *de novo* methyltransferase DRM2; a splice donor mutation at amino acid 85 predicted to cause a frameshift and premature stop well upstream of the DNA methyltransferase catalytic domain (Figure 2-2 B). The allele caused *FWA* late flowering and a strong loss of methylation at *MEA-ISR*. Although DRM2 is the methyltransferase targeted by RdDM machinery, no previous RdDM mutagenesis screen has isolated a mutant allele of the gene.

Whole-genome sequencing can efficiently identify mutants with weak phenotypes

Although *MEA-ISR* is a useful locus for examining defects in RdDM, we found that some late-flowering mutants from our forward screen had only a very subtle phenotype at *MEA-ISR* as assayed by Southern blot analysis. Therefore, scoring mutants in F2 mapping populations

proved extremely difficult. For example, in a population of 56 Col/Ler F2 plants scored as being homozygous for the recessive *m48* mutant, we found that the area showing linkage on chromosome 3 never fell below 25 cM, suggesting that some mutants were mis-scored and that the population was contaminated with wild-type plants. In order to map and clone *m48* we therefore employed a whole-genome sequencing approach.

First, we reduced our initial mapping population of 56 recombinants to 38 plants that exhibited homozygosity for Col-0 specific PCR based molecular markers in a wide region of chromosome three, to reduce the frequency of mis-scored plants. The DNA from the set of 38 recombinants was then pooled, and we performed paired-end shotgun sequencing using Illumina GA II technology, giving 142 million high quality reads and 21.4 times coverage of the Arabidopsis genome.

In order to locate regions that are enriched for Col-0 sequences, we utilized the Mapping and Assembly with Short Sequences (MASS) approach described by Cuperus *et a*l (Cuperus et al., 2010). The region exhibiting the highest enrichment for Col-0 SNPs fell in a 3 MB window on chromosome three (Figure 2-7 A), with the highest peaks between positions 8.25 and 8.8 MB. Within the 3 MB interval, 11 G-A or C-T mutations were identified, which are consistent with EMS mutagenesis. From these 11, 2 fell into intergenic regions, one occurred in a miRNA, and the remaining 8 were in protein-coding genes.

One mutation occurred in *NRPD/E2*, a gene previously reported to be involved in RdDM (Herr et al., 2005; Kanno et al., 2005). The mutation discovered here is an Arg-to-Gln missense mutation in the conserved Rbp2_4 domain (Figure 2-7 C). NRPD/E2 is a paralog of NRPB2, the second largest subunit of Pol II. It is shared by both Pol IV and Pol V, and null mutations cause substantial loss in 24nt siRNA accumulation and non-CG methylation (Onodera et al., 2005; Zhang et al., 2007). The allele identified from the screen does not have as severe a phenotype as a null allele at the *MEA-ISR* locus (Figure 2-7 B). In all likelihood, this is due to the missense mutation impairing the protein function without rendering it completely inactive. A genetic

complementation test confirmed that the mutation in the *nrpd/e2* gene does cause the *MEA-ISR* methylation phenotype in the *m48* line (Figure 2-7 B). Its *FWA* flowering-time phenotype provides the first evidence that NRPD/E2 is required for the initial establishment of DNA methylation.

The mutant *m61* also exhibited a reduced phenotype at the *MEA-ISR* locus (Figure 2-8 C), and we utilized whole-genome sequencing to both map and identify the gene affected in this mutant, this time with no pre-selection by traditional mapping. From a mapping population of 200 plants, we identified 42 that displayed an apparent weak *MEA-ISR* methylation phenotype by Southern blot. Pooled DNA from the 42 apparent recombinants was sequenced in the same manner as described above, giving 33.5 times coverage of the genome. Depletion of L*er* SNPs was observed only on chromosome two (Figure 2-8 A), identifying an 8 MB region as the likely location of this mutation. Mutations were analyzed from this 8 MB region, and the most likely mutation to cause the phenotype was a Glycine-to-Arginine missense occurring at amino acid position 49 in NRPE1 (*nrpe1-12*) (Figure 2-8 B). Genetic complementation confirmed that indeed this mutation causes the observed phenotype (Figure 2-8 C). The mutation was caused by a G-to-A base change consistent with EMS mutagenesis. Interestingly, out of 42 reads for this nucleotide, 38 were mutant and four were wild type, showing that roughly 10 percent of the pooled DNA was contaminated by wild-type sequences due to mis-scoring of the weak mutant phenotype. Importantly, the whole genome sequencing approach was able to compensate for the mis-scoring of this mutant phenotype and allow us to efficiently identify the affected gene. Identification of this weak mutation would have been nearly impossible by a traditional mapping strategy. Such results indicate the promise of using a whole-genome approach to identify even weak mutants with recombinant DNA from a relatively limited number of individual samples.

Conclusion

We present the first forward genetic screen that specifically searches for mutations that block the establishment of methylation of an incoming transgene. The mutations that have been characterized to date help to depict the components required for both *de novo* DNA methylation and RdDM maintenance at endogenous loci (Figure 2-9). In addition to this work, other recent studies have reported genes that are required for RdDM, including INVOLVED IN DE NOVO 1/DEFECTIVE IN MERISTEM SILENCING 3 (IDN1/DMS3) (Ausin et al., 2009; Kanno et al., 2008), INVOLVED IN DE NOVO 2 (IDN2) (Ausin et al., 2009), SU(VAR)3-9 HOMOLOG 2 (SUVH2), SU(VAR)3-9 HOMOLOG 9 (SUVH9) (Johnson et al., 2008), HISTONE DEACETYLACE 6 (HDA6) (Aufsatz et al., 2002b; He et al., 2009a), and DEFECTIVE IN MERISTEM SILENCING 4/ RNA DIRECTED DNA METHYLATION 4 (DMS4/RDM4) (Figure 2-9) (He et al., 2009b; Kanno et al.). Thus far, all RdDM mutations that have been tested, including all of those in the current study as well as mutations in *IDN2*, *DMS3*, *SUVH2*, and *SUVH9* affect *FWA de novo* DNA methylation. These results strongly support the view that components of the RdDM DNA methylation maintenance pathway are also components of the pathway that establishes *de novo* DNA methylation.

Materials and methods

Plant materials. We used the following Arabidopsis strains: the wild type WS, L*er* and Col-0; the recessive *nrpe5a-1* (FLAG_607D12) allele in the WS background; the recessive *ago4-1* allele in L*er* background (See Zilberman et al.(Zilberman et al., 2003)); recessive alleles *clsy1-8* (SALK_018319), *dcl3-1* (SALK_005512), *drd1-6* (See Kanno et al.(Kanno et al., 2004)), *drm2-2* (SALK_150863), *ktf1-1* (SALK_001294), *nrpd/e2-2* (SALK_046208), *nrpe1-11* (SALK_029919), and *rdr2-2* (SALK_059661) in the Col-0 background. The Myc-tagged complementing AGO4 line used for immunfluorescence is described in Li et al (Li et al., 2006).

EMS Mutagenesis, *FWA* transformation, and flowering time analysis. 10,000 Col-0 seeds were incubated in 0.3% EMS in a volume of 15 mL for 12 hours. Roughly 900 M2 families were then screened for flowering-time abnormalities. Any family containing late flowering mutants were discarded. The remaining lines were transformed with the *FWA* transgene. We performed *FWA* transformation using an AGL0 *Agrobacterium tumefaciens* strain carrying a pCAMBIA3300 vector with an engineered version of *FWA* in which an EcoRI site was converted into a BglII site. For selection, we sprayed the resultant T1 population with a 1:1000 dilution of Finale^{TM}. We measured flowering time of resistant plants as the total number of leaves (rosette and cauline leaves) developed by a plant.

Southern blotting. We performed Southern blotting at the *MEA-ISR* and *Ta3* loci as described in Johnson et al.(Johnson et al., 2008) and Cao et al.,(Cao and Jacobsen, 2002a) respectively.

Bisulfite analysis. We performed sodium bisulfite sequencing using EZ DNA Methylation Gold (Zymo Research) by following the manufacturer's instructions. Following amplication of bisulfite treated DNA, we cloned the resulting PCR fragments into pCR2.1-TOPO (Invitrogen) and

analyzed 15 to 22 clones per sample. All primers are listed in Table 2-1.

Whole genome mutation identification. Recombinant mutants were pooled as described in the text. *nrpd/e2-19* was identified by methods described by Cuperus et al.(Cuperus et al., 2010) *nrpe1-12* was identified as follows:

A list of 258,838 high-quality known homozygous Ler SNPs relativeto Col-0 was composed by filtering:
ftp://ftp.arabidopsis.org/Polymorphisms/Ecker_ler.homozygous_snp.txt
(downloaded 2009-06-19)

All 163,907,331 raw reads from 7 lanes from one GA II RTA SCS 2.4 55-cycle single end flow cell run were aligned with bowtie 0.10.1(using qualities, allowing <=2 mismatches in seed of first 28 cycles, and reporting all alignments of best stratum; additional parameters: maqerr=60 without rounding, maxbts=1000000, tryhard) to the TIGR5 Col-0 reference genome with mitochondrion and chloroplast; the 88,181,158 reads with exactly one alignment were retained.

Only the 4,842,337,521 read bases discretebasecalled A/C/G/T with quality score >=20 aligning to a genomic A/C/G/T were retained, giving ~42.1-fold average pooled-strand coverage of those genomic bases to which filtered alignments are possible.

All but 32 of the 258,838 known filtered L*er* SNPs were aligned to at least one read base. In the 8,705,359 base-pair linkage region of interest (all chr2 bases from 11,000,001 onward),~99.2% of genomic base-pairs were aligned to at least one read base. Candidate mutant SNPs were taken to be genomic base-pairs with at least one and >=49% non-Col-0 observations, ordered for priority by descending number of non-Col-0 observations. Of the top 40, one in NRPE1 (#13

on the list) was the most likely candidate, and was validated experimentally.

Figure Legends

Figure 2-1. Schematic representation of the *FWA* screen approach.

EMS mutagenized Col-0 plants were transformed with the *FWA* transgene. Prior to transformation, any mutagenized line that exhibited a late-flowering phenotype was discarded. The first generation of *FWA*-transformed mutant plants were screened for flowering time. All late-flowering lines were deemed potential *de novo* methylation mutants. As a secondary screen, late-flowering lines were divided into three classes based on their phenotype at the *MEA-ISR* locus. Genomic DNA was digested with the methylation sensitive enzyme MspI, and digestion was analyzed by Southern blot.

Figure 2-2. Characterization of mutant lines, and mutation identification.

(A) *MEA-ISR* Southern blot phenotype for nine *de novo* mutant lines. Seven lines were sub-categorized *MEA-ISR* eliminated, and two as *MEA-ISR* reduced. The allele resulting in each mutant phenotype is indicated

(B) Protein models for each mutation identified from the screen. The mutation is denoted above the model. *drd1* is the only gene for which we obtained multiple alleles: *drd1-10* is a Gly=>Asp missense and *drd1-11* is a Gly=>Glu missense.

Figure 2-3. Analysis of the *dcl3* allele recovered from the screen.

(A-C) Sodium bisulfite treatment and analysis of methylation state of *FWA* endogene (A), *MEA-ISR* (B), and *FWA* transgene (C). Y axis denotes percent methylation.

(D) *FWA* flowering-time assay. Average number of leaves upon flowering for both untransformed and *FWA* transformed lines.

Figure 2-4. Analysis of *clsy1 de novo* methylation phenotype.

(A) *FWA* flowering-time assay.

(B) *MEA-ISR* Southern blot from two individual Col-0 plants and two *clsy1-8* plants to show the reproducibility of the weak phenotype.

(C and D) Sodium bisulfite analysis of the *FWA* transgene (C) and endogene (D).

Figure 2-5. Reverse genetics showing that *nrpe5a* and *ago6* are required for *de novo* DNA methylation.

(A) *MEA-ISR* Southern blot.

(B) Flowering-time assay.

Figure 2-6. Molecular characterization of *ktf1*.

(A) *FWA* flowering-time assay.

(B-D) Sodium bisulfite analysis of *MEA-ISR* (B), *FWA* (C), and *IGN5* (D). *IGN5* is a third locus targeted by RdDM machinery.

(E) Small RNA blots showing the abundance of various small RNAs in different mutant backgrounds. miRNA159 serves as a loading control. *ktf1* mutant does not abolish 24nt species of siRNAs suggesting it is acting downstream of siRNA biosynthesis.

Figure 2-7. Identifcation of an *nrpd/e2* mutant allele by a whole-genome sequencing approach.

(A) Enrichment of Col-0 specific single nucleotide polymorphisms in pooled DNA from population of F2 Ler x *m48* recombinant plants. Ratios were calculated in sliding 250 KB windows at 50 KB intervals. The greatest peak of Col-0 specific polymorphisms is observed between 8 and 11 MB on chromosome 3.

(B) *MEA-ISR* Southern blot. *m48* has a weaker phenotype than the *nrpd/e2-2* null allele. However, a genetic complementation test confirms the mutation causes the observed phenotype.

(C) Protein model for NRPD/E2. Numbered domains in yellow refer to conserved Rbp2 regions across the protein. The asterisk denotes a missense mutation recovered from screen.

Figure 2-8. Identification of *nrpe1* mutant allele by a whole-genome sequencing approach.

(A) Depletion in percentage of Ler specific single nucleotide polymorphisms across chromosome 2. Note that percentage of Ler SNPs never reaches zero, suggesting that pooled DNA contains wild-type (non-mutant) contaminants. The red dot indicates the physical location of *NRPE1*.

(B) Protein model of NRPE1. Asterisk denotes missense mutation recovered from screen.

(C) *MEA-ISR* Southern blot. The mutation discovered by the whole-genome approach has a weaker phenotype than the null allele of *nrpe1* recovered from the screen.

Figure 2-9. Model for *de novo* methylation.

A repetitive invasive genetic element is marked in red. Pol IV together with CLSY1 generates a single stranded RNA, which is made double stranded by RDR2, and diced into 24nt siRNAs which are primarily loaded into AGO4. DMS3 and DRD1 act upstream of Pol V, which creates a "scaffold" transcript that recruits AGO4 to chromatin. IDN2 binds to double stranded RNA, and may stabilize the siRNA-Pol V RNA hybrid. Possibly through an intermediate protein or chromatin modification, DRM2 is recruited to *de novo* methylate the invasive DNA. Proteins with stars have been shown to be required for *FWA de novo* DNA methylation.

Table 2-1. List of primers used in this study.

Figure 2-1

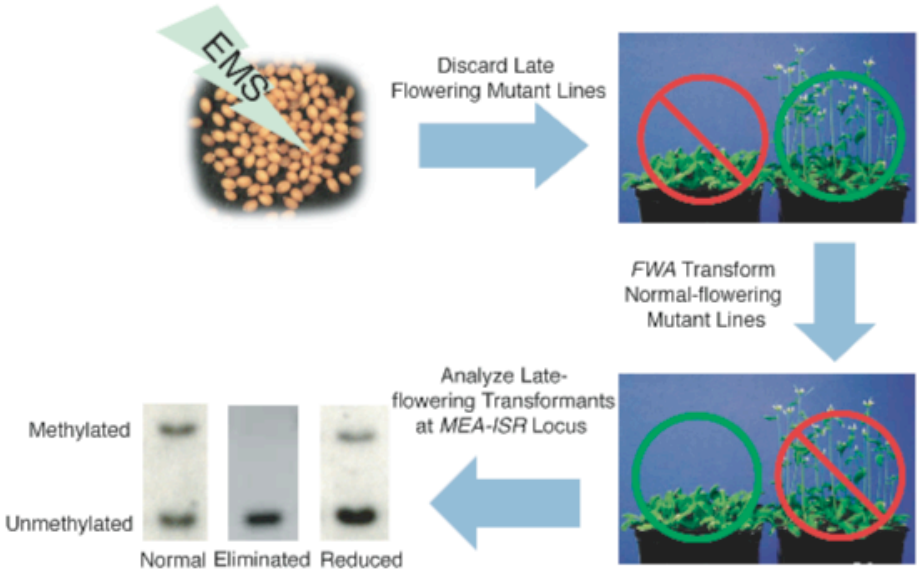
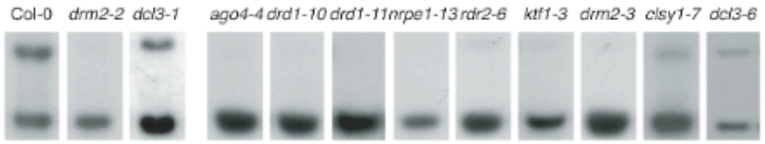

EMS
Discard Late
Flowering Mutant Lines
FWA Transform
Normal-flowering
Mutant Lines
Analyze Late-
flowering Transformants
at MEA-ISR Locus
Methylated
Unmethylated
Normal
Eliminated
Reduced

Figure 2-2

a.


Col-0
drm2-2
dcl3-1
ago4-4
drd1-10
drd1-11
nrpe1-13
rdr2-6
ktf1-3
drm2-3
clsy1-7
dcl3-6

b.

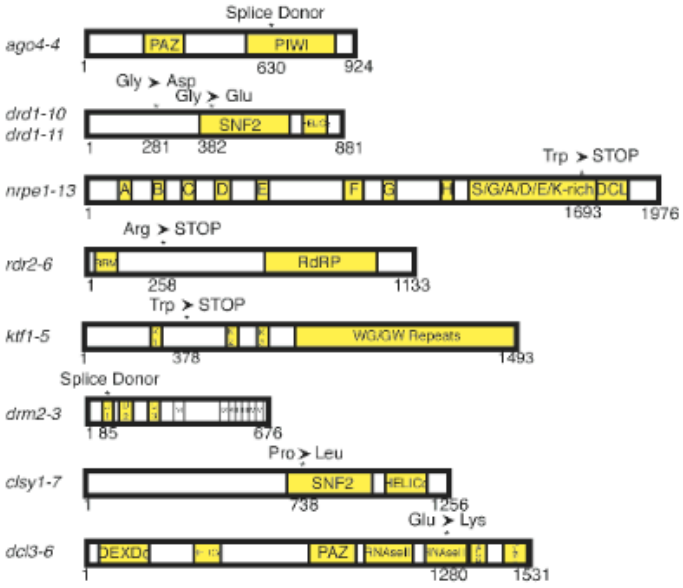

Splice Donor
ago4-4
PAZ
PIWI
1
630
924
Gly ➤ Asp
Gly ➤ Glu
drd1-10
drd1-11
SNF2
1
281
382
881
Trp ➤ STOP
nrpe1-13
A
B
C
D
E
F
G
S/G/A/D/E/K-rich
DCL
1
1693
1976
Arg ➤ STOP
rdr2-6
RdRP
1
258
1133
Trp ➤ STOP
ktf1-5
WG/GW Repeats
1
378
1493
Splice Donor
drm2-3
1 85
676
Pro ➤ Leu
clsy1-7
SNF2
1
738
1256
Glu ➤ Lys
dcl3-6
DEXDc
PAZ
1
1280
1531

Figure 2-3

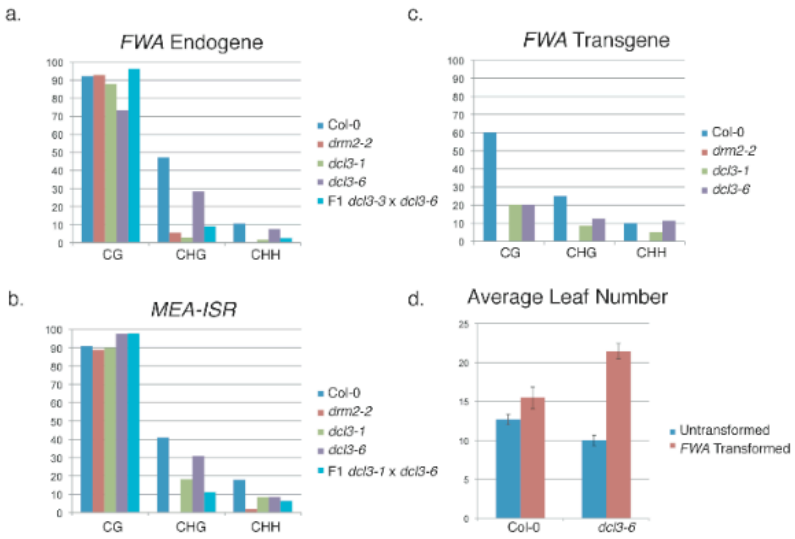

a.
FWA Endogene
100
90
80
70
60
50
40
30
20
10
0
CG
CHG
CHH
Col-0
drm2-2
dcl3-1
dcl3-6
F1 dcl3-3 x dcl3-6
c.
FWA Transgene
100
90
80
70
60
50
40
30
20
10
0
CG
CHG
CHH
Col-0
drm2-2
dcl3-1
dcl3-6
b.
MEA-ISR
100
90
80
70
60
50
40
30
20
10
0
CG
CHG
CHH
Col-0
drm2-2
dcl3-1
dcl3-6
F1 dcl3-1 x dcl3-6
d.
Average Leaf Number
25
20
15
10
5
0
Col-0
dcl3-6
Untransformed
FWA Transformed

Figure 2-4

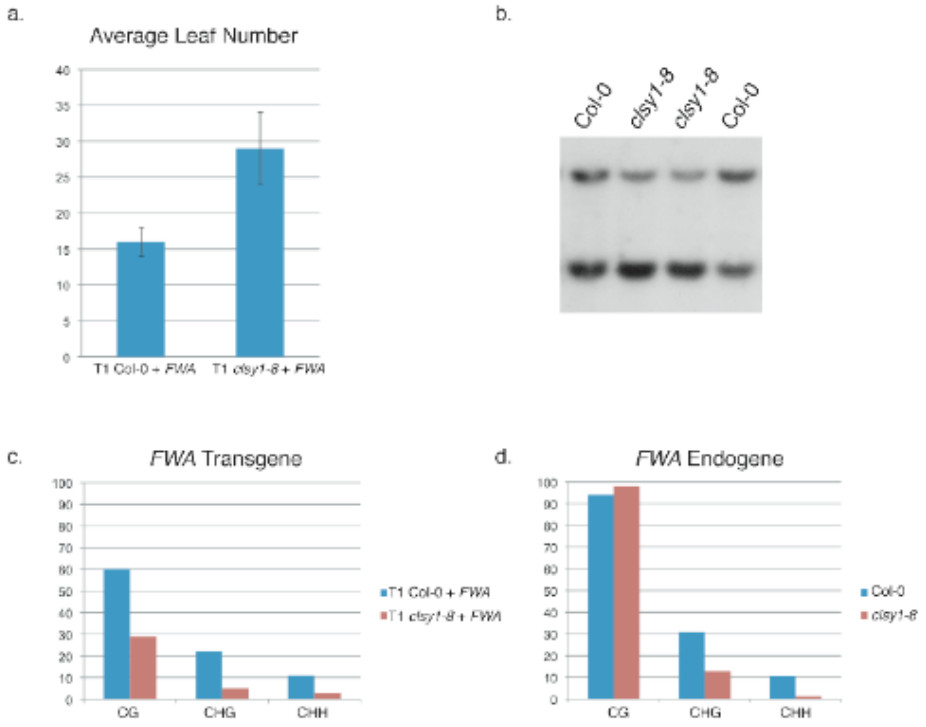

a.
Average Leaf Number
40
35
30
25
20
15
10
5
0
T1 Col-0 + FWA
T1 clsy1-8 + FWA
b.
Col-0
clsy1-8
clsy1-8
Col-0
c.
FWA Transgene
100
90
80
70
60
50
40
30
20
10
0
CG
CHG
CHH
T1 Col-0 + FWA
T1 clsy1-8 + FWA
d.
FWA Endogene
100
90
80
70
60
50
40
30
20
10
0
CG
CHG
CHH
Col-0
clsy1-8

Figure 2-5

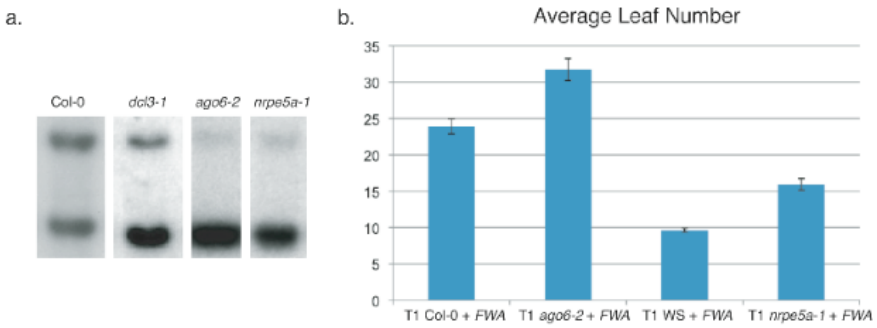

a.
Col-0
dcl3-1
ago6-2
nrpe5a-1
b.
Average Leaf Number
35
30
25
20
15
10
5
0
T1 Col-0 + FWA
T1 ago6-2 + FWA
T1 WS + FWA
T1 nrpe5a-1 + FWA

Figure 2-6

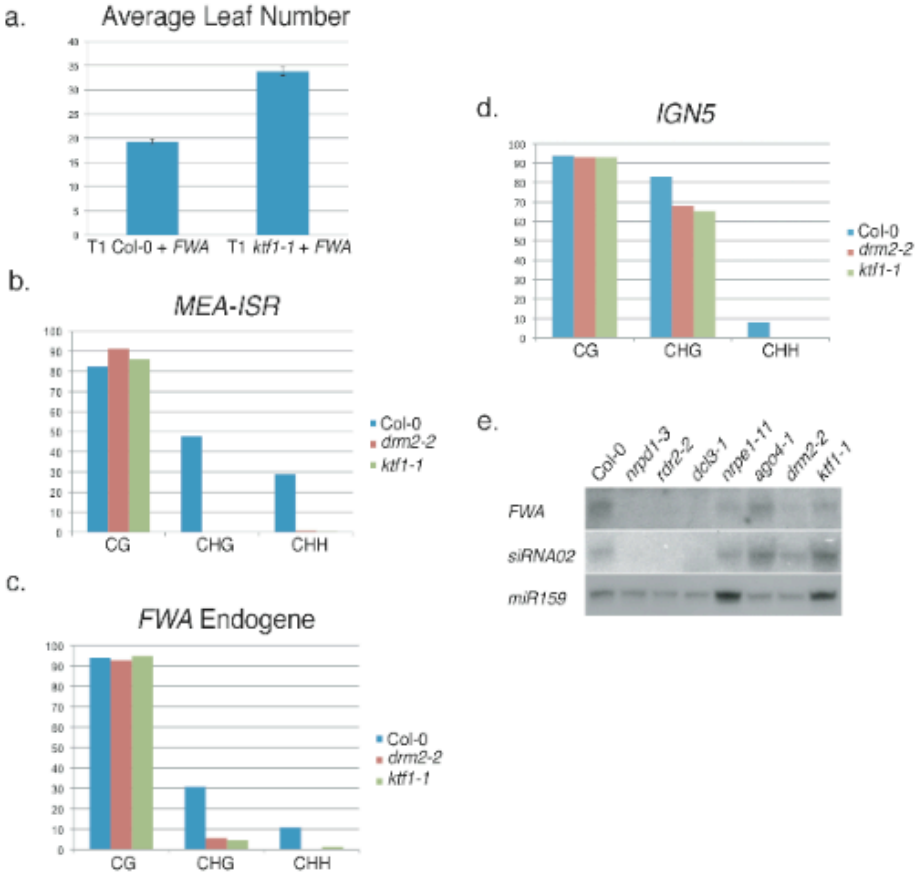

a.
Average Leaf Number
T1 Col-0 + FWA
T1 ktf1-1 + FWA
b.
MEA-ISR
CG
CHG
CHH
Col-0
drm2-2
ktf1-1
c.
FWA Endogene
CG
CHG
CHH
Col-0
drm2-2
ktf1-1
d.
IGN5
CG
CHG
CHH
Col-0
drm2-2
ktf1-1
e.
Col-0
nrpd1-3
rdr2-2
dcl3-1
nrpe1-11
ago4-1
drm2-2
ktf1-1
FWA
siRNA02
miR159

Figure 2-7

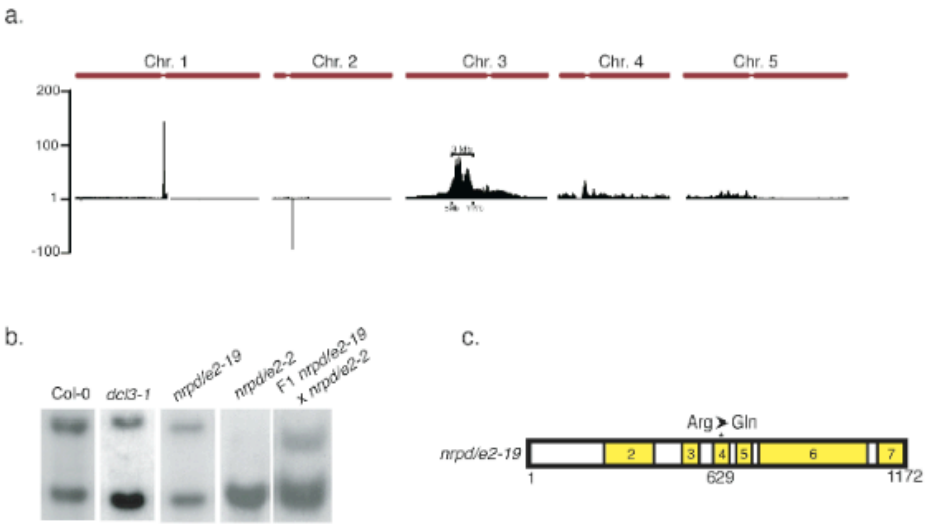

a.
Chr. 1
Chr. 2
Chr. 3
Chr. 4
Chr. 5
200
100
1
-100
b.
Col-0
dcl3-1
nrpd/e2-19
nrpd/e2-2
F1 nrpd/e2-19 x nrpd/e2-2
c.
Arg➤Gln
nrpd/e2-19
2
3
4
5
6
7
1
629
1172

Figure 2-8

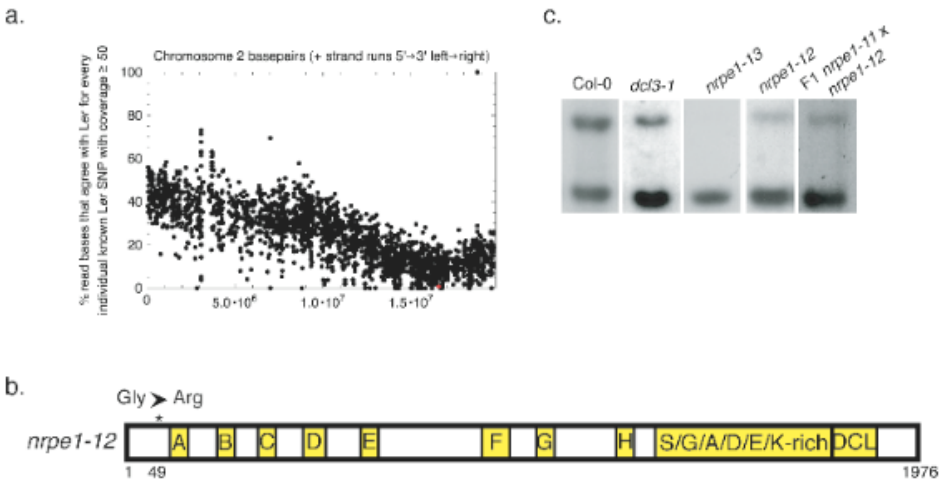

a.
Chromosome 2 basepairs (+ strand runs 5'→3' left→right)
% read bases that agree with Ler for every individual known Ler SNP with coverage ≥ 50
100
80
60
40
20
0
0
5.0·10^6
1.0·10^7
1.5·10^7
c.
Col-0
dcl3-1
nrpe1-13
nrpe1-12
F1 nrpe1-11 x nrpe1-12
b.
Gly ➤ Arg
nrpe1-12
A
B
C
D
E
F
G
H
S/G/A/D/E/K-rich
DCL
1
49
1976

Figure 2-9

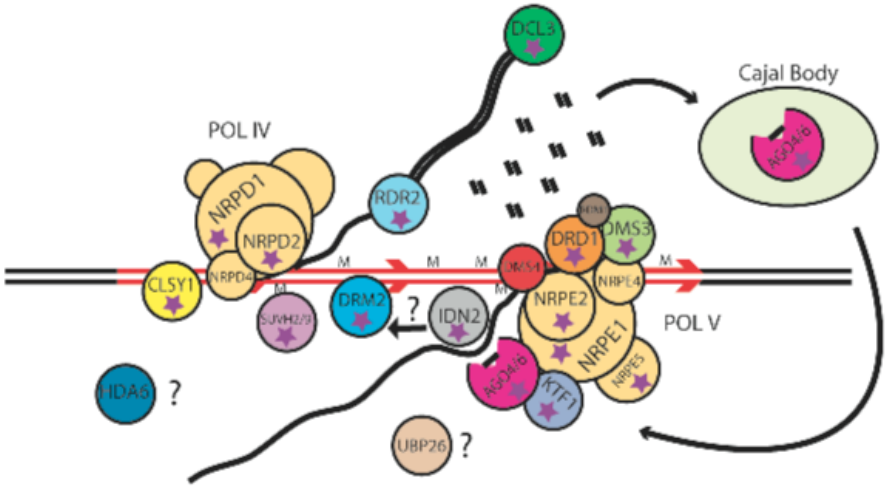

DCL3
Cajal Body
POL IV
NRPD1
RDR2
NRPD2
DRD1
DMS3
CLSY1
NRPD4
DMS4
NRPE4
DRM2
?
IDN2
NRPE2
POL V
NRPE1
HDA6
?
KTF1
NRPE5
UBP26
?

Table 2-1

Primer	Sequence
MEA-ISR Bisulfite	AAAGTGGTTGTAGTTTATGAAAGGTTTTAT
MEA-ISR Bisulfite	CTTAAAAAATTTTCAACTCATTTTTAAAAAA
FWA Bisulfite	GGTTTTATATTAATATTAAAGAGTTATGGGTYGAAGTTT
FWA Bisulfite	AACCAAAATCATTCTCTAAACAAAATATAAAAAAATC
IGN5 Bisulfite	GTTYYYGAGAAGAGTAGAAYAAATGYTAAAATGTATYATGYGGTT
IGN5 Bisulfite	RRACTAARTCTTRTCRAACAARRACCCAACCATRTCCRCTTAAAAA
MEA-ISR Southern Probe	AAACCTTTCGTAAGCTACAGCCACTTTGTT
MEA-ISR Southern Probe	TCGGATTGGTTCTTCCTACCTCTTTACCTT
FWA sRNA Northern	GCCGCTCTAGGGTTTTTGCTTTTCGCCATTGGTCCAAGTG
siR02 sRNA Northern	GTTGACCAGTCCGCCAGCCGAT
miR159 sRNA Northern	TAGAGCTCCCTTCAATCCAAA

REFERENCES

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, A.J., and Matzke, M. (2002a). RNA-directed DNA methylation in Arabidopsis. Proc Natl Acad Sci U S A *99 Suppl 4*, 16499-16506.

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, M., and Matzke, A.J. (2002b). HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. EMBO J *21*, 6832-6841.

Ausin, I., Mockler, T.C., Chory, J., and Jacobsen, S.E. (2009). IDN1 and IDN2 are required for de novo DNA methylation in Arabidopsis thaliana. Nat Struct Mol Biol *16*, 1325-1327.

Bies-Etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., Cooke, R., and Lagrange, T. (2009). RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. EMBO Rep.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M., and Jacobsen, S.E. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. Curr Biol *13*, 2212-2217.

Cao, X., and Jacobsen, S.E. (2002a). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A *99 Suppl 4*, 16491-16498.

Cao, X., and Jacobsen, S.E. (2002b). Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Curr Biol *12*, 1138-1144.

Chan, S.W., Zhang, X., Bernatavichute, Y.V., and Jacobsen, S.E. (2006). Two-step recruitment of RNA-directed DNA methylation to tandem repeats. PLoS Biol *4*, e363.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Cuperus, J.T., Montgomery, T.A., Fahlgren, N., Burke, R.T., Townsend, T., Sullivan, C.M., and Carrington, J.C. (2010). Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing. Proc Natl Acad Sci U S A *107*, 466-471.

El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R., and Lagrange, T. (2007). Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. Genes Dev *21*, 2539-2544.

He, X.J., Hsu, Y.F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.S., and Zhu, J.K. (2009a). NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. Genes Dev *23*, 318-330.

He, X.J., Hsu, Y.F., Zhu, S., Liu, H.L., Pontes, O., Zhu, J., Cui, X., Wang, C.S., and Zhu, J.K. (2009b). A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development. Genes Dev *23*, 2717-2722.

He, X.J., Hsu, Y.F., Zhu, S., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H.L., Wang, C.S., Jin, H., and Zhu, J.K. (2009c). An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. Cell *137*, 498-508.

Henderson, I.R., and Jacobsen, S.E. (2007). Epigenetic Inheritance in Plants. Nature *447*, 418-424.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat Genet *38*, 721-725.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science *308*, 118-120.

Huang, L., Jones, A.M., Searle, I., Patel, K., Vogler, H., Hubner, N.C., and Baulcombe, D.C. (2009). An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. Nat Struct Mol Biol *16*, 91-93.

Johnson, L.M., Law, J.A., Khattar, A., Henderson, I.R., and Jacobsen, S.E. (2008). SRA-domain proteins required for DRM2-mediated de novo DNA methylation. PLoS Genet *4*, e1000280.

Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Bohmdorfer, G., Gregor, W., Kreil, D.P., Matzke, M., and Matzke, A.J. (2008). A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. Nat Genet *40*, 670-675.

Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Kreil, D.P., Breinig, F., Lind, M., Schmitt, M.J., Simon, S.A., Gurazada, S.G.*, et al.* RNA-directed DNA methylation and plant development require an IWR1-type transcription factor. EMBO Rep *11*, 65-71.

Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat Genet *37*, 761-765.

Kanno, T., Mette, M.F., Kreil, D.P., Aufsatz, W., Matzke, M., and Matzke, A.J. (2004). Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. Curr Biol *14*, 801-805.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science *303*, 521-523.

Lahmy, S., Pontier, D., Cavel, E., Vega, D., El-Shami, M., Kanno, T., and Lagrange, T. (2009). PolV(PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-specific subunit. Proc Natl Acad Sci U S A *106*, 941-946.

Law, J.A., and Jacobsen, S.E. (2009). Establishing, maintaining, and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics *In Submission*.

Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. (2006). An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. Cell *126*, 93-106.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

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

Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009). RNA-mediated chromatin-based silencing in plants. Curr Opin Cell Biol *21*, 367-376.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell *120*, 613-622.

Qi, Y., He, X., Wang, X.J., Kohany, O., Jurka, J., and Hannon, G.J. (2006). Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature *443*, 1008-1012.

Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L., and Pikaard, C.S. (2009). Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. Mol Cell *33*, 192-203.

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

Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard, C.S., and Baulcombe, D.C. (2007). An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. Plant Cell *19*, 1507-1521.

Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol Cell *6*, 791-802.

Till, S., Lejeune, E., Thermann, R., Bortfeld, M., Hothorn, M., Enderle, D., Heinrich, C., Hentze, M.W., and Ladurner, A.G. (2007). A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. Nat Struct Mol Biol *14*, 897-903.

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell *135*, 635-648.

Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet *41*, 630-634.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol *2*, E104.

Zhang, X., Henderson, I.R., Lu, C., Green, P.J., and Jacobsen, S.E. (2007). Role of RNA polymerase IV in plant small RNA metabolism. Proc Natl Acad Sci U S A *104*, 4536-4541.

Zheng, X., Zhu, J., Kapoor, A., and Zhu, J.K. (2007). Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. EMBO J *26*, 1691-1701.

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

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. (2004). Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. Curr Biol *14*, 1214-1220.

CHAPTER 3

The splicing factor SR45 affects the RNA-directed DNA methylation pathway in Arabidopsis.

Results and Discussion

DRM2 carries out all known DNA methylation establishment—or *de novo* methylation—in *Arabidopsis thaliana* (Cao and Jacobsen, 2002b). Given that DRM2 is guided by both siRNAs and long non-coding RNAs (Law and Jacobsen; Wierzbicki et al., 2008), we screened a collection of homozygous lines carrying T-DNA insertion in genes containing known or predicted RNA-interacting domains (Ausin et al., 2009). For this screen we used *FWA* transgene silencing as a reporter system. *FWA* is a homeodomain transcription factor that has two tandem repeats in its promoter. In wild type plants, the endogenous *FWA* repeats are stably methylated and *FWA* is silenced. However, hypomethylation in its promoter region leads to an ectopic expression and a late-flowering phenotype (Soppe et al., 2000). After *FWA* transformation, plants with an intact *de novo* methylation machinery are able to methylate and silence the transgenic *FWA*, while mutants affecting *de novo* methylation express the transgene and flower late (Cao and Jacobsen, 2002c; Chan et al., 2004).

Following this strategy, we isolated a line containing a T-DNA insertion in ARGININE/SERINE-RICH 45 (*SR45*) that flowered slightly late before *FWA* transformation, but showed a major late flowering phenotype after transformation (Figure 3-1 A and 3-2). The *sr45-1* mutant has been reported to show a late flowering phenotype due to an increased expression of *FLOWERING LOCUS C* (*FLC*) (Ali et al., 2007). To assess whether the observed effect was mediated by *FLC* alone, we analyzed the methylation status of *FWA* in *sr45-1* mutants after transformation. Bisulfite sequencing analyses revealed that the endogenous *FWA* methylation was not affected in the CG-dinucleotide context, but did display a defect in non-CG methylation, which is consistent with other mutants in the DRM2 pathway (Greenberg et al., 2011). However, the transgenic copy of *FWA* exhibited reduced methylation levels in every sequence context, (CG, CHG, and CHH; H being A, T, or G), (Figure 3-1 B). This reduction in methylation is correlated with the late flowering phenotype and ectopic *FWA* expression in the transgenic

plants (Figures 3-3 A, B and C). Moreover, the introduction of an *SR45* copy into *sr45-1*, partially restores the ability to methylate *FWA* to wild type levels (Figures 3-3 D and E), ruling out that the sole increment in *FLC* levels accounted for the observed late flowering phenotype. It is worth noting that some individual mutant plants are able to establish methylation at or near wild-type levels, indicating a degree of stochasticity. However on average, we can safely say that *sr45* mutant plants have impaired DNA methylation establishment capacity.

We further analyzed the progression of *FWA de novo* methylation across subsequent generations in *sr45-1*. In wild type, the *FWA* transgene did not reach full levels of methylation until the T_2 generation (Figure 3-4). In *sr45-1* mutants, the levels of methylation at *FWA* transgene are reduced in the T_1 generation, but in some cases complete methylation does occur until the T_3 generation or later (Figure 3-4). Thus, the *sr45-1* mutant only partially impairs the *de novo* methylation machinery, and wild type methylation levels are regained in later generations. These observations are consistent with the slow intergenerational silencing that has been reported in other transgene systems(Cao et al., 2003).

To date, every mutant that has been shown to be defective in *de novo* methylation is also defective in the maintenance of non-CG methylation. We, therefore, examined the methylation status at known RdDM targets such as *FWA*, *MEA-ISR* and *AtSN1*. For this purpose we digested genomic DNA with methylation sensitive enzymes and performed either Southern blots or PCR, or we examined individual loci by sequencing following bisulfite treatment. Analysis revealed that the *sr45-1* mutant exhibits reduced non-CG methylation for all of the aforementioned loci (Figures 3-5 A, B and C). Due to the incompleteness of the phenotype, we generated double mutants with *dcl3-1*, which been reported to display a weak DNA methylation phenotype (Greenberg et al.). We found that *sr45-1*, *dcl3-1* double mutants exhibited an additive effect in methylation phenotype. This enhancement was found at both RdDM target loci tested (Figures 3-5 A, B and C). Concordantly, *sr45-1 dcl3-1* double mutants

also show an enhanced *de novo* methylation phenotype at *FWA* transgene (Figure 3-1). Together this data indicate that SR45 and DCL3 could cooperate within the RdDM pathway.

To assess whether the methylation defect is specific to RdDM targets, we analyzed methylation and expression levels at *Ta3* and *REPEAT12* (*REP12*) loci by Southern blot and RT-PCR–two loci that are known to be methylated in a DRM2-independent manner (Cao and Jacobsen, 2002a). We observed that *sr45-1* as well as *sr45-1, dcl3-1* double mutants showed no difference in methylation when compared to wild type, indicating that SR45 function is most likely confined to the DRM2 pathway (Figures 3-5 D and E).

In order to place SR45 within the context of the RdDM pathway, we analyzed *sr45-1* siRNA production by northern blot. The 24-nucleotide siRNAs associated with RdDM are broadly grouped into two types: type I (dependent on both plant specifc RNA polymerases: Pol IV and Pol V) and type II (only dependent on Pol IV) (Zheng et al., 2009). Regardless of type, the siRNAs abundance was reduced in *sr45-1* mutant plants (Figure 3-6 A), suggesting that SR45 acts in the pathway at steps prior to the production of small RNAs.

Previous studies have shown that ARGONAUTE 4 (AGO4) protein is destabilized in mutants upstream of siRNA biogenesis (Li et al., 2006). To test whether this holds true for *sr45-1*, we examined *AGO4* transcript levels by RT-PCR and Northern blots in three different tissues. We observed neither significant alteration in *AGO4* expression pattern, nor major splicing variants in *sr45-1* relative to wild type (Figures 3-6 B and C). However, western blot analyses revealed a slight, but reproducible decrease of AGO4 protein in *sr45-1* mutants. In addition, the effect of *sr45-1* on AGO4 levels was increased in *dcl3-1* background (Figure 3-6 C), reinforcing the hypothesis of those two genes cooperating in the regulation of the RdDM pathway. To further confirm the observed reduction of AGO4 levels, we analyzed the nuclear localization pattern of a complementing epitope-tagged version of AGO4 in the *sr45-1* background by immunoflourescence. Consistent with western blot data, we observed a decrease in AGO4

abundance in *sr45-1*, nonetheless the localization pattern of AGO4 was similar to wild type (Figure 3-6 D).

The *FLC* locus, which is silenced by a DNA methylation-independent mechanism (Jean Finnegan et al., 2005) is also partially de-repressed in the *sr45-1* mutant background (Figure 3-7) (Ali et al., 2007). It is interesting to note that DCL3 has been previously reported to be required for *FLC* silencing despite the lack of transcriptional control by DNA methylation (Schmitz and Amasino, 2007). This DCL3 regulation of *FLC* is probably through small RNAs matching its 3' region (Swiezewski et al., 2007). Furthermore, the de-repression of *FLC* is enhanced in the *dcl3-1 sr45-1* double mutant (Figure 3-7).

In sum, we have discovered a known spliceosome gene that is required for RdDM. It can not be ruled out that SR45 may be involved in the splicing of an RdDM factor, thus the methylation phenotype is a secondary effect. Alternatively, given its small RNA phenotype, it potentially has a novel function in siRNA processing. It is worth noting that the nuclear cap-binding complex, which is involved in pre-mRNA splicing, has a role in a distinct DICER-LIKE1-dependent micro RNA pathway(Laubinger et al., 2008). This suggests SR45, and perhaps other spliceosome factors, may indeed play a direct role in siRNA accumulation as well. We screened a number of known or putative splicesome factors as part of our screen, however *sr45* was the only one with an *FWA*-dependent flowering-time defect (Table 3-1) Interestingly, *sr45* shares a very similar phenotype as *dcl3*, even at the *FLC* locus which is not an RdDM target. This suggests that these two proteins likely work in concert to control RNA-mediated silencing.

Materials and Methods

Plant materials. We used the following Arabidopsis strains: The wild type Columbia; the recessive *sr45-1* (SALK_004132) and *dcl3-1* (SALK_005512); the Myc-tagged complementing AGO4 line used for immunofluorescence and western blots is described in Li.(Li et al., 2006)

***FWA* transformation and flowering-time analysis:** We performed *FWA* transformation using an AGL0 *Agrobacterium tumefaciens* strain carrying a pCAMBIA3300 vector with an engineered version of *FWA* in which an *EcoRI* site was converted into a *BglII* site. For selection, we sprayed the resultant T1 population with a 1:1000 dilution of Finale^{TM}. We measured flowering time of resistant plants as the total number of leaves (rosette and cauline leaves) developed by a plant.

Bisulfite analysis. We performed sodium bisulfite sequencing using EZ DNA Methylation Gold (Zymo Research) reagents for conversion of plant genomic DNA extracted from floral tissue using a standard CTAB protocol. Following amplication of bisulfite treated DNA by PCR, we cloned the resulting PCR fragments into pCR2.1-TOPO (Invitrogen) and analyzed 15 to 22 clones per sample. The *FWA* transgene was distinguished from the endogene by BglII digestion prior to bisulfite treatement (see FWA Transformation methods) and elimination of any clones containing Col-0 polymorphisms rom the data set after sequencing. All primers are listed in Table 3-2.

Southern blotting. DNA from young flowers was extracted using a standard CTAB protocol. 1 µg of genomic DNA was digested overnight with *MspI*. The digestion was run on a 1% agarose gel, transferred to Hybond N^+ membranes, blocked and washed according to manufacturer instructions (GE Healthcare). Membranes were probed with a PCR product radiolabeled with alpha ^{32}P-dCTP using the Megaprime DNA Labeling System. *MEA-ISR*, *Ta3* and *REP12* PCR products for probing were generated with primers listed in Table 3-2.

Small RNA northern blotting. Detection of small RNAs was performed exactly as described in Law.(Law et al., 2011) Oligonucleotide sequences used for probing can be found in Table 3-2.

Immunofluorescent microscopy. Detection of Myc-tagged AGO4 protein was performed

exactly as described in Li.(Li et al., 2006) Primary mouse monoclonal anti-Myc (Covance 9E10) was used at a 1:200 dilution. Secondary anti-mouse FITC (Abcam) was used at a 1:200 dilution. DNA was stained using Vectashield mounting medium containing DAPI (Vector Laboratories).

Northern blotting. RNA was extracted from the indicated tissue using Trizol reagent (Invitrogen). Northern blots were performed as described in Henderson et al. (Henderson et al., 2006). *AGO4* and *UBQ10* PCR products used for probing were generated with primers listed in Table 3-2.

Bisulfite cutting assay. DNA was extracted and bisulfite treated as described above. The cutting assay was performed exactly as described in Chant et al. (Chan et al., 2006).

Figure Legends

Figure 3-1. *sr45-1 de novo* DNA methylation phenotype.

(A) Flowering time of Columbia, *sr45-1*, *dcl3-1* and *sr45-1*, *dcl3-1* double before and after *FWA* transformation. Flowering time is measured as the total number of leaves at the time of flowering.

(B) Methylation levels at endogenous and transgenic *FWA* after *FWA* transformation. The 594 base-pair repeated region in the 5' UTR was analyzed. The methylation state of the *FWA* endogene should remain unaltered by the presence of the *FWA* transgene. All samples were analyzed in the T_1 generation.

Figure 3-2. Flowering time distribution of *sr45-1* versus *sr45-1+FWA.*

Figure 3-3. Confirmation of *sr45-1 de novo* phenotype.

(A) Flowering time of randomly selected T_2 *sr45-1+FWA* transformants. *FWA* construct has a Basta® resistance gene as a selection marker that allows testing for the presence of *FWA* construct. Red-dotted line depicts the flowering time of untransformed *sr45-1* mutants grown under the same conditions.

(B) RT-PCR showing *FWA* expression a selection of the above-mentioned lines. *UBQ10* expression is showed as a loading control.

(C) Bisulfite cutting assay, showing *FWA* methylation status in the above-mentioned lines. Genomic DNA is digested with *BglII* to destroy the endogenous *FWA* gene before bisulfite treatment. DNA methylation of transgenic *FWA* was assayed by PCR from bisulfite-treated DNA followed by *ClaI* digestion. CG methylation protects the *ClaI* site from bisulfite conversion. Black arrow indicates the unmethylated size.

(D) Flowering time of homozygous T_3 *sr45-1+SR45* complemented lines after *FWA* transformation.

(E) RT-PCR and bisulfite cutting assay showing SR45 expression and partial restoration of methylation at *FWA*. *UBQ10* expression is showed as a loading control.

Figure 3-4. Analysis of the *FWA* transgene methylation generationally.

Methylation levels at endogenous and transgenic *FWA* after *FWA* transformation. Endogenous *FWA* was only analyzed in the T_1 generation. Transgenic *FWA* was analyzed during three generations after transformation.

Figure 3-5. *sr45-1* maintenance DNA methylation phenotype.

(A) Sodium bisulfite analysis of an 180 base-pair region of the *MEA-ISR* locus.

(B) *AtSN1* Chop-qPCR assay. Genomic DNA was digested with the methylation sensitive enzyme *HaeIII*, which recognizes three sites in *AtSN1*. Amplification of *AtSN1* was

quantified by Real Time PCR, and signal was normalized to undigested DNA. *HaeIII*, is blocked by C methylation in GGCC context.

(C) *MEA-ISR* Southern blot. *MspI* digested genomic DNA was probed with *MEA-ISR*.

(D) *REP12* and *Ta3* Southern blot. *MspI* digested genomic DNA was probed with *REP12* or *Ta3. MspI* is blocked by methylation of the external C in CCGG context.

(E) RT-PCR showing expression levels of *REP12* and *Ta3*. *UBQ10* expression is showed as a loading control.

Figure 3-6. Placement of SR45 in RdDM pathway.

(A) RNA blots showing siRNAs abundance at both type I and II loci. Hybridization with *miR163* is shown as a loading control for *5S* siRNAs and hybridization with *miR159* is shown as a loading control for *AtSN1* and *siR02.*

(B) Northern blot showing *AGO4* expression in leaves, seedlings, and flowers. Hybridization with *UBQ10* is shown as a loading control.

(C) RT-PCR and western blot showing the expression and abundance of *AGO4*/AGO4. *UBQ10* expression is showed as a loading control for RT-PCR and amido black staining of the RUBISCO large subunit is shown as a loading control for the western blot.

(D) Immunofluorescent microscopy showing AGO4 localization in 4xmyc::AGO4 (Columbia) and 4xmyc::AGO4 (*sr45-1*) backgrounds. White arrows indicate the position of AGO4 in the *sr45-1* panel.

Figure 3-7. *FLC* de-repression enhancement.

(A) Flowering time of Columbia, *sr45-1*, *dcl3-1* and *sr45-1*, *dcl3-1* double mutant.

(B) RT-PCR showing the expression of *FLC* in the above-mentioned lines. *UBQ10* expression is showed as a loading control.

Table 3-1. A list of known or putative spliceosome factors screened in this study.

Table 3-2. Primers used in this study.

Figure 3-1

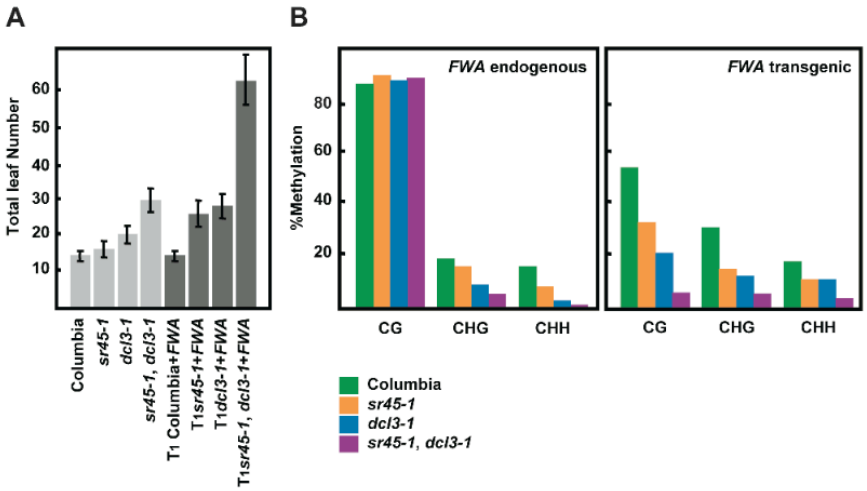

A
Total leaf Number
60
50
40
30
20
10
Columbia
sr45-1
dcl3-1
sr45-1, dcl3-1
T1 Columbia+FWA
T1sr45-1+FWA
T1dcl3-1+FWA
T1sr45-1, dcl3-1+FWA
B
%Methylation
80
60
40
20
FWA endogenous
FWA transgenic
CG
CHG
CHH
CG
CHG
CHH
Columbia
sr45-1
dcl3-1
sr45-1, dcl3-1

Figure 3-2

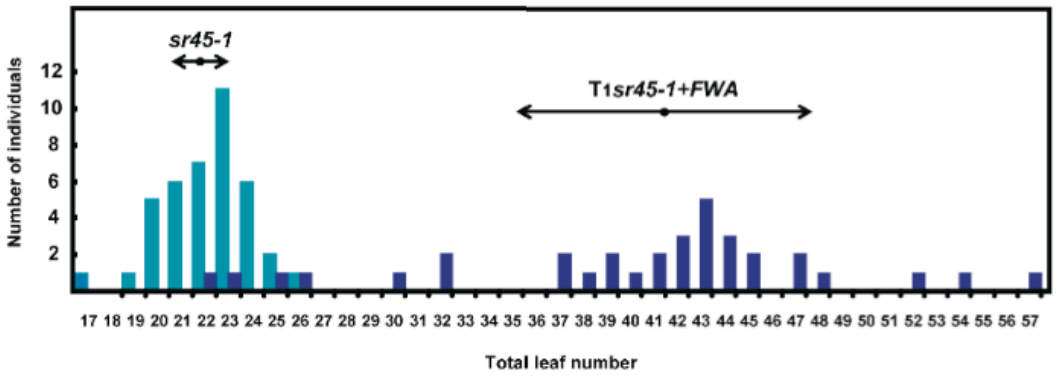

sr45-1
T1sr45-1+FWA
Number of individuals
12
10
8
6
4
2
17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57
Total leaf number

Figure 3-3

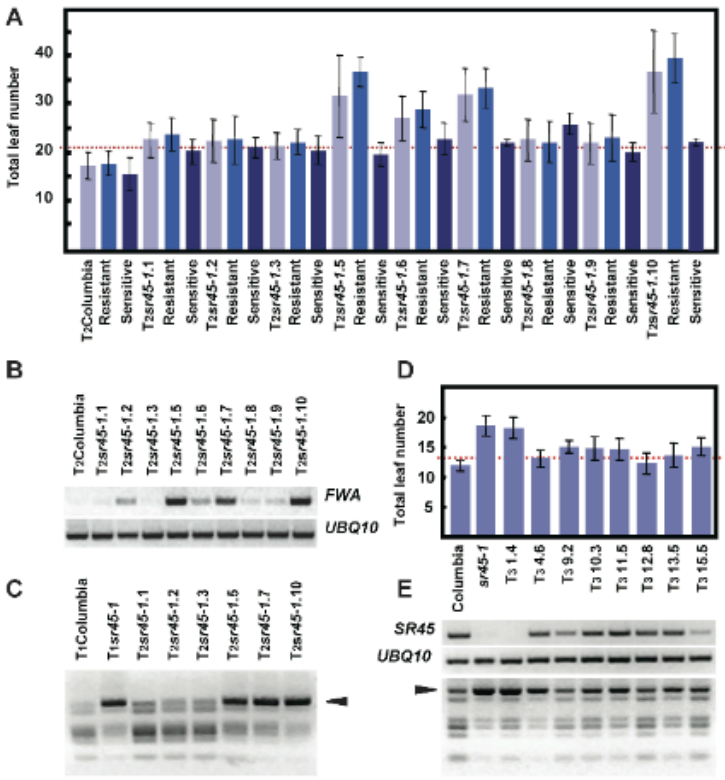

A
Total leaf number
40
30
20
10
T2Columbia
T2sr45-1 Resistant
T2sr45-1 Sensitive
T2sr45-1.1 Resistant
T2sr45-1.1 Sensitive
T2sr45-1.2 Resistant
T2sr45-1.2 Sensitive
T2sr45-1.3 Resistant
T2sr45-1.3 Sensitive
T2sr45-1.5 Resistant
T2sr45-1.5 Sensitive
T2sr45-1.6 Resistant
T2sr45-1.6 Sensitive
T2sr45-1.7 Resistant
T2sr45-1.7 Sensitive
T2sr45-1.8 Resistant
T2sr45-1.8 Sensitive
T2sr45-1.9 Resistant
T2sr45-1.9 Sensitive
T2sr45-1.10 Resistant
T2sr45-1.10 Sensitive
B
T2Columbia
T2sr45-1.1
T2sr45-1.2
T2sr45-1.3
T2sr45-1.5
T2sr45-1.6
T2sr45-1.7
T2sr45-1.8
T2sr45-1.9
T2sr45-1.10
FWA
UBQ10
C
T1Columbia
T1sr45-1
T2sr45-1.1
T2sr45-1.2
T2sr45-1.3
T2sr45-1.5
T2sr45-1.7
T2sr45-1.10
D
Total leaf number
20
15
10
5
Columbia
sr45-1
T3 1.4
T3 4.6
T3 9.2
T3 10.3
T3 11.5
T3 12.8
T3 13.5
T3 15.5
E
SR45
UBQ10

Figure 3-4

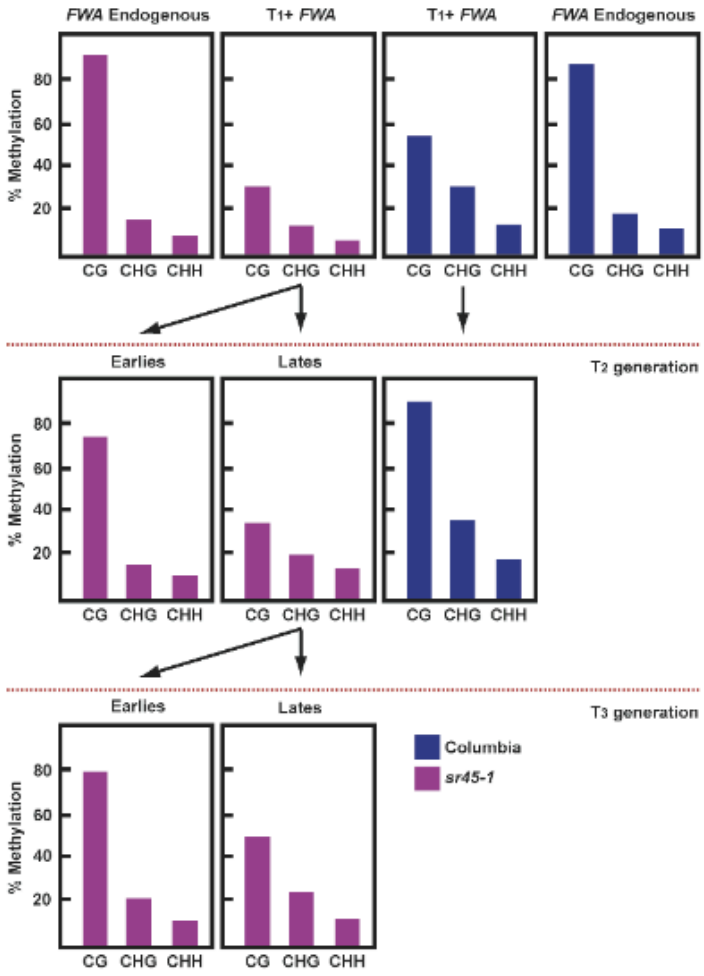

FWA Endogenous
T1+ FWA
T1+ FWA
FWA Endogenous
% Methylation
80
60
40
20
CG CHG CHH
T2 generation
Earlies
Lates
T3 generation
Columbia
sr45-1

Figure 3-5

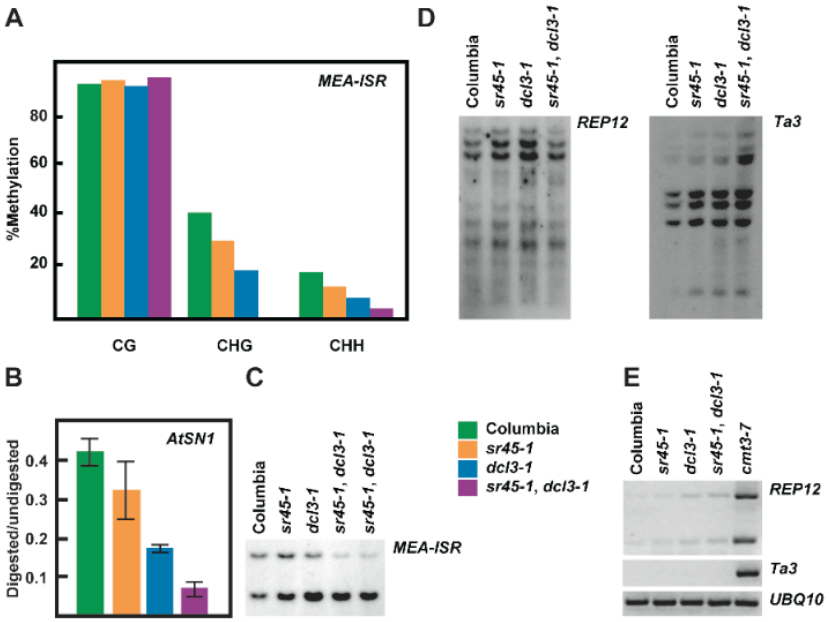

A
MEA-ISR
%Methylation
80
60
40
20
CG
CHG
CHH
D
Columbia
sr45-1
dcl3-1
sr45-1, dcl3-1
REP12
Ta3
B
AtSN1
Digested/undigested
0.4
0.3
0.2
0.1
C
Columbia
sr45-1
dcl3-1
sr45-1, dcl3-1
MEA-ISR
E
cmt3-7
REP12
Ta3
UBQ10

Figure 3-6

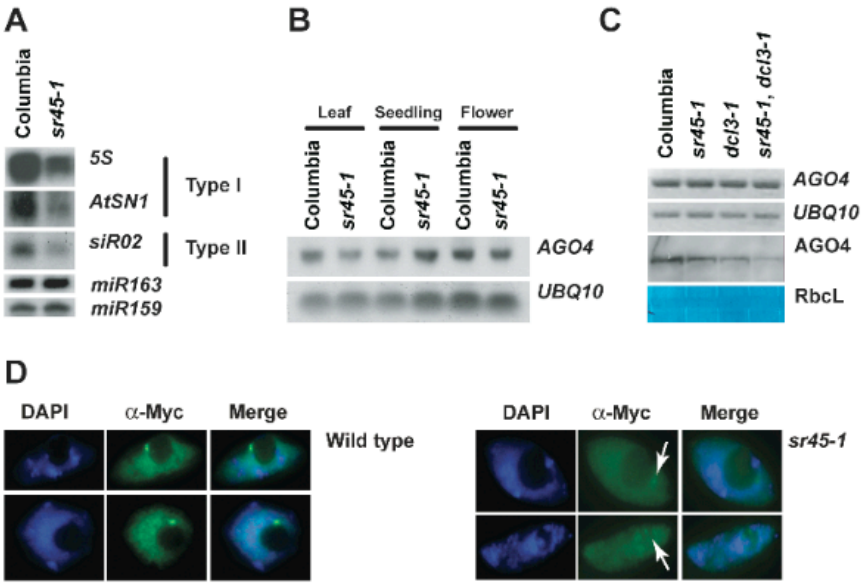

A
Columbia
sr45-1
5S
AtSN1
siR02
miR163
miR159
Type I
Type II
B
Leaf
Seedling
Flower
Columbia
sr45-1
AGO4
UBQ10
C
Columbia
sr45-1
dcl3-1
sr45-1, dcl3-1
AGO4
UBQ10
AGO4
RbcL
D
DAPI
α-Myc
Merge
Wild type
sr45-1

Figure 3-7

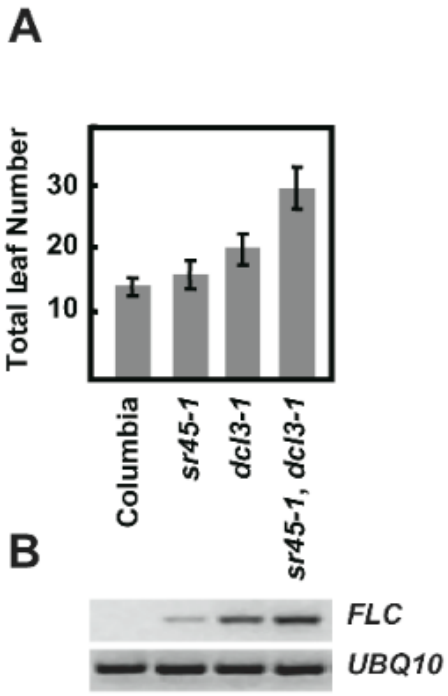

A
Total leaf Number
30
20
10
Columbia
sr45-1
dcl3-1
sr45-1, dcl3-1
B
FLC
UBQ10

Table 3-1

Line	Insertion	AGI	Annotation
99	SALK_034824	At5g37370	pre-mRNA splicing factor PRP38 family protein (SRL1)
114	SALK_040864	At1g55310	SC35-like splicing factor, 33 kD (SCL33)
118	SALK_041849	At3g13570	SC35-like splicing factor, 30a kD (SCL30a)
159	SALK_059236	At2g33440	splicing factor family protein
167	SALK_062177	At5g51300	splicing factor-related
288	SALK_125057	At1g03140	splicing factor Prp18 family protein
299	SALK_130919	At3g45950	splicing factor-related
336	SALK_147534	At3g55930	RNA splicing factor-related
357	SALK_004132	At1g16610	arginine/serine-rich protein, putative (SR45)
367	SALK_007933	At4g03430	pre-mRNA splicing factor-related
453	SALK_035759	At2g29210	splicing factor PWI domain-containing protein
518	SALK_055030	At1g60200	splicing factor PWI domain-containing protein / RNA recognition motif (RRM)-containing protein
549	SALK_063749	At4g38780	splicing factor, putative
562	SALK_069027	At5g37370	pre-mRNA splicing factor PRP38 family protein (SRL1)
589	SALK_085295	At4g38780	splicing factor, putative
591	SALK_086134	At3g06455	splicing factor-related
620	SALK_106715	At1g60200	splicing factor PWI domain-containing protein / RNA recognition motif (RRM)-containing protein
640	SALK_121879	At3g25440	group II intron splicing factor CRS1-related
662	SALK_132986	At1g09140	SF2/ASF-like splicing modulator (SRP30)
666	SALK_135024	At2g18510	pre-mRNA splicing factor, putative
667	SALK_135314	At2g29210	splicing factor PWI domain-containing protein
668	SALK_135626	At5g64270	splicing factor, putative
704	SALK_149863	At2g33440	splicing factor family protein
864	WiscDsLox289_292D17	At1g80070	splicing factor, putative
880	WiscDsLox384B10	At5g18810	SC35-like splicing factor, 28 kD (SCL28)
898	WiscDsLox246B03	At5g64200	arginine/serine-rich splicing factor SC35
899	WiscDsLox247E04	At4g02430	pre-mRNA splicing factor, putative / SR1 protein, putative
914	WiscDsLox382G12	At4g25500	arginine/serine-rich splicing factor RSP40 (RSP40)
916	WiscDsLox384G9	At4g13070	group II intron splicing factor CRS1-related

Table 3-2

Primer	Sequence 5' to 3'
5S siRNA Probe	ATGCCAAGTTTGGCCTCACGGTCT
AGO4 RT-PCR and Northern	CAGTGCCATTTCTGTTGTTGC
AGO4 RT-PCR and Northern	TGGCGACGTTGTCTTTGAGT
AtSN1 Chop-qPCR	TTTAAACATAAGAAGAAGTTCCTTTTTCATCTAC
AtSN1 Chop-qPCR	ACTTAATTAGCACTCAAATTAAACAAAATAAGT
AtSN1 siRNA Probe	ACCAACGTGTTGTTGGCCCAGTGGTAAATCTCTCAGATAGAGG
FLC RT-PCR	TGTGGATAGCAAGCTTGTGG
FLC RT-PCR	TAGTCACGGAGAGGGCAGTC
FWA Bisulfite	GGTTTTATATTAATATTAAAGAGTTATGGGTYGAAGTTT
FWA Bisulfite	AACCAAAATCATTCTCTAAACAAAATATAAAAAAATC
FWA Bisulfite Cutting	GGTTTTATTTTAATGTAAATATGTATTTGATGTATT
FWA Bisulfite Cutting	CTAAATTAATATAACRTAATATAACATTATA
FWA RT-PCR	TAGTCCAGGATTGTCTGCAAAAGG
FWA RT-PCR	CCATTATTAACGATTTTCAGAAGAGAGA
MEA-ISR Bisulfite	AAAGTGGTTGTAGTTTATGAAAGGTTTTAT
MEA-ISR Bisulfite	CTTAAAAAATTTTCAACTCATTTTTAAAAAA
MEA-ISR Southern Probe	AAACCTTTCGTAAGCTACAGCCACTTTGTT
MEA-ISR Southern Probe	TCGGATTGGTTCTTCCTACCTCTTTACCTT
miR159 miRNA Probe	TAGAGCTCCCTTCAATCCAAA
miR163 miRNA Probe	ATCGAAGTTGGAAGTCCTCTTCAA
REP12 Southern probe	TCCTCTTTCTCCCTCCTCTCCC
REP12 Southern probe	TCCCACCAAGAAGCACACGC
siR02 siRNA Probe	GTTGACCAGTCCGCCAGCCGAT
Ta3 Southern Probe	GATCTATCTGGCCCCAGACGTAGATCTAA
Ta3 Southern Probe	CCGGCAATCTACTATATGAGATCTTTACAA
UBQ10 RT-PCR and Northern	GATCTTTGCCGGAAAACAATTGGAGG
UBQ10 RT-PCR and Northern	CGACTTGTCATTAGAAAGAAAGAGAT

References

Ali, G.S., Palusa, S.G., Golovkin, M., Prasad, J., Manley, J.L., and Reddy, A.S. (2007). Regulation of plant developmental processes by a novel splicing factor. PLoS One *2*, e471.

Ausin, I., Mockler, T.C., Chory, J., and Jacobsen, S.E. (2009). IDN1 and IDN2 are required for de novo DNA methylation in Arabidopsis thaliana. Nat Struct Mol Biol *16*, 1325-1327.

Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M., and Jacobsen, S.E. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. Curr Biol *13*, 2212-2217.

Cao, X., and Jacobsen, S.E. (2002a). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A *99 Suppl 4*, 16491-16498.

Cao, X., and Jacobsen, S.E. (2002b). Role of the Arabidopsis DRM methyltransferases in *de novo* DNA methylation and gene silencing. Current Biology *12*, 1138-1144.

Cao, X., and Jacobsen, S.E. (2002c). Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Curr Biol *12*, 1138-1144.

Chan, S.W., Zhang, X., Bernatavichute, Y.V., and Jacobsen, S.E. (2006). Two-step recruitment of RNA-directed DNA methylation to tandem repeats. PLoS Biol *4*, 1923-1933.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Greenberg, M.V., Ausin, I., Chan, S.W., Cokus, S.J., Cuperus, J.T., Feng, S., Law, J.A., Chu, C., Pellegrini, M., Carrington, J.C*., et al.* Identification of genes required for de novo DNA methylation in Arabidopsis. Epigenetics *6*, 344-354.

Greenberg, M.V., Ausin, I., Chan, S.W., Cokus, S.J., Cuperus, J.T., Feng, S., Law, J.A., Chu, C., Pellegrini, M., Carrington, J.C*., et al.* (2011). Identification of genes required for de novo DNA methylation in Arabidopsis. Epigenetics *6*.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat Genet *38*, 721-725.

Jean Finnegan, E., Kovac, K.A., Jaligot, E., Sheldon, C.C., James Peacock, W., and Dennis, E.S. (2005). The downregulation of FLOWERING LOCUS C (FLC) expression in plants with low levels of DNA methylation and by vernalization occurs by distinct mechanisms. Plant J *44*, 420-432.

Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J.U., Ratsch, G., and Weigel, D. (2008). Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. Proc Natl Acad Sci U S A *105*, 8795-8800.

Law, J.A., and Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.

Law, J.A., Vashisht, A.A., Wohlschlegel, J.A., and Jacobsen, S.E. (2011). SHH1, a Homeodomain Protein Required for DNA Methylation, As Well As RDR2, RDM4, and Chromatin Remodeling Factors, Associate with RNA Polymerase IV. PLoS Genet *7*, e1002195.

Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. (2006). An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. Cell *126*, 93-106.

Schmitz, R.J., and Amasino, R.M. (2007). Vernalization: a model for investigating epigenetics and eukaryotic gene regulation in plants. Biochim Biophys Acta *1769*, 269-275.

Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol Cell *6*, 791-802.

Swiezewski, S., Crevillen, P., Liu, F., Ecker, J.R., Jerzmanowski, A., and Dean, C. (2007). Small RNA-mediated chromatin silencing directed to the 3' region of the Arabidopsis gene encoding the developmental regulator, FLC. Proc Natl Acad Sci U S A *104*, 3633-3638.

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. Cell *135*, 635-648.

Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J.Y., and Chen, X. (2009). Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes Dev *23*, 2850-2860.

CHAPTER 4

An IDN2-containing complex involved in RNA-directed DNA methylation in Arabidopsis

Introduction

DNA methylation is a stable epigenetic mark that is associated with the repression of genes and transposable elements. In *Arabidopsis thaliana*, maintenance of DNA methylation at silent loci is carried out by at least three methyltransferases: METHYLTRANSFERASE 1 (MET1), CHROMOMETHYLTRANSFERASE 3 (CMT3) and DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2). However, DRM2 is solely responsible for establishment of DNA methylation—or *de novo* methylation—of silent elements (Cao and Jacobsen, 2002). DRM2 is guided to chromatin by small interfering RNAs (siRNAs) in a process known as RNA-directed DNA methylation (RdDM) (Law and Jacobsen, 2010). It has been proposed that RdDM also requires intergenic non-coding (IGN) transcripts that are synthesized by RNA Polymerase V (Pol V). These transcripts likely serve as platforms for the recruitment of siRNA-loaded ARGONAUTE 4 (AGO4) to methylated loci (Wierzbicki et al., 2008; Wierzbicki et al., 2009).

Recently, we discovered the requirement of INVOLVED IN DE NOVO 2 (IDN2) for RdDM from a forward genetic screen (Ausin et al., 2009). Alleles of this same gene were later reported from another screen for DNA methylation mutants (Zheng et al., 2010). We previously demonstrated that IDN2 binds to double-stranded RNA through its XS domain, which is also observed in the XS-domain containing protein SUPPRESSOR OF GENE SILENCING 3 (SGS3) (Fukunaga and Doudna, 2009). We also found that IDN2 was likely to act in a step downstream of initial siRNA biogenesis. The XS domain is conserved throughout the plant kingdom, and XS-domain containing proteins are involved in a wide range of processes such as viral defense (Glick et al., 2008) and stress response (Qin et al., 2009).

To gain a better understanding of the *in vivo* role of IDN2, we performed affinity purifications from complementing transgenic lines expressing epitope-tagged full-length IDN2 protein. We found that IDN2 forms a complex with two novel proteins from the same family as IDN2, which we termed IDN2-LIKE1 (IDNL1) and IDN2-LIKE2 (IDNL2). Through single-locus

approaches as well as whole-genome bisulfite sequencing we find that IDNL1 and IDNL2 are also essential components of the RdDM pathway. In higher order mutants, the methylation phenotype is not enhanced further than what is observed in the single *idn2-1* mutant, suggesting a non-redundant role between IDN2 and its paralogs in the IDN2 complex. Comparisons with known RdDM mutants in genome-wide methylation and expression analyses solidify the role of the IDN2 complex as a core component of RdDM machinery.

Results

Structural Analysis of IDN2 XS domain.

We previously showed that IDN2 binds to double-stranded RNA with 5' overhangs *in vitro* through its XS domain (Ausin et al., 2009). Bioinformatic analysis has suggested that the XS domain is likely to adopt a unique RNA-recognition motif (RRM) fold, which would be consistent with its *in vitro* activity (Zhang and Trudeau, 2008). In order to gain further insights into the structure and mechanism of the XS domain we determined the structure of the IDN2-XS domain along with a small segment of adjacent coiled-coil region (120-292) by X-ray crystallography (Figure 4-1). We found that the core structure of the XS domain superimposes closely over a known RRM domain. However, insertions in the XS domain form a few additional secondary structural elements: a β-strand (β_N) at the N-terminus, a longer loop having anti-parallel β-sheet (formed by β1a and β1b) between α1 and β2, a longer loop having a small α-helix (α3) between α2 and β4, and two additional α-helices (α3 and α4) at the C-terminal end of the XS domain (Figures 4-1 B and 4-2 A). Since RNA binding specificity by RRM fold proteins depends on the loops present between α-helices and β-strands, it is likely that insertions in these regions in the XS domain results in its unique specificity towards 5' overhang containing dsRNA. This structural study provides the first empirical evidence that the XS domain has a unique RRM motif. Additionally, by examining the electrostatic surface of the XS domain, we observed an

exposed positively charged basic patch (Figure 4-1 D). This likely interacts directly with negatively charged RNA molecules.

As seen in the crystal structure, IDN2 (120-292) dimerizes mainly via coiled-coil segment (Figures 4-1 *C* and *D*). The fact that the coiled-coil region readily dimerizes suggests that IDN2 potentially exists *in vivo* in higher order complexes with itself. However, as discussed below, the interaction is more likely between IDN2 and closely related homologs, which also contain a coiled-coil domain. Interestingly, the interacting hydrophobic residues on the respective coiled-coils are conserved between IDN2 and its homologs (Figure 4-2 B).

IDN2 forms a complex with paralogs IDNL1 and IDNL2.

In order to determine the interacting partners of IDN2, we produced transgenic lines expressing *IDN2* fused to different epitope tags under the control of the *IDN2* promoter region. These IDN2 epitope-tagged transgenic lines were able to complement the methylation defect of *idn2-1* mutant at the *MEDEA-INTERGENIC SUBTELOMERIC REPEATS* (*MEA-ISR*) indicating normal protein functionality (Figure 4-3). Upon establishing homozygous lines, we prepared protein extracts from the apical tissue of *BLRP::9xMyc::IDN2* complementing lines and performed affinity purification with streptavidin. Purified extracts were analyzed by multidimensional protein identification technology (MudPIT) (Law et al., 2010). MudPIT analysis from two independent purifications revealed the presence of abundant peptides of the proteins At1g15910 and At4g00380, indicating that those two proteins and IDN2 could form a complex *in vivo* (Table 4-1). Much less abundant peptides from a few other proteins were also found in both replicas, though it is not known if these are of any significance. At1g15910 and At4g00380 are in the same gene family as IDN2, and share 92% amino acid identity with each other (Figure 4-4 A). The members of this gene family all contain the same domain architecture and organization as IDN2 (Figure 4-1 A). We termed At1g15910 and At4g00380 IDNL1 and IDNL2, respectively.

In order to confirm the interaction between IDN2 and IDNL1, we performed several assays including gel filtration, IDNL1 co-immunoprecipitation (Co-IP), and IDNL1 affinity purifications (Figures 4-5 A and B and Table 1). For gel filtration assays we first generated a complementing transgenic line expressing IDN2*::3xFlag::BLRP* in two different genetic backgrounds: the *idn2-1* mutant and the *idn2-1 idnl-1 idnl2-1* triple mutant. After gel filtration and western blotting, the elution profile revealed a significant delay in elution of the complex in the triple *idn2-1 idnl-1 idnl2-1* mutant background when compared to *idn2-1* (Figure 4-5 A). Given that both *idnl1-1* and *idnl2-1* are insertion mutants that do not generate a transcript (Figures 4-4 B and C), this delay can be explained by the absence of these two proteins in the complex.

To examine the IDN2-complex in further detail, we generated complementing transgenic lines carrying the *IDNL1*::9xMyc fusion under the control of *IDNL1* promoter region (Figure 4-3). This line was crossed to *IDN2::3xFLAG::BLRP*, and plants from subsequent F_1 generation were analyzed by co-IP confirming the *in vivo* interaction (Figure 4-5 B). Additionally, we affinity purified the transgenic IDNL1 using an anti-Myc antibody and performed MudPIT analysis. Consistent with both the co-IP and IDN2 MudPIT, we observed a substantial interaction with IDN2 (Table 4-1). In fact, over two independent replicates, IDN2 was the only protein to be recovered from both. Taken together, these data confirm the IDNL1-IDN2 *in vivo* interaction.

IDNL1 and IDNL2 are required for RdDM.

Given the fact that IDN2 affects *de novo* methylation (Ausin et al., 2009) and both IDNL1 and 2 co-purify with IDN2, it seemed likely that those two genes might be also required for *de novo* methylation. To assess this hypothesis, we used the well-studied gene *FWA*. *FWA* is heritably silenced by methylation, however, unmethylated *fwa* epialleles exhibit ectopic expression that results in a dominant late flowering phenotype (Soppe et al., 2000). After *FWA* transformation, wild type plants are able to methylate and silence *FWA* transgenes, while RdDM mutants fail to methylate *FWA* and thus flower late (Cao and Jacobsen, 2002; Soppe et al., 2000). Using *FWA*,

we transformed individual *idnl* mutants as well as the *idnl* double mutant and *idn2-1 idnl1-1 idnl2-1* triple mutant. After *FWA* transformation, *idnl2-1* did not show any flowering defect, while *idnl1-1* displayed a slightly late flowering phenotype (Figure 4-5 C). This delay in flowering was correlated with a small decrease in methylation in all contexts, however pre-established CG methylation remained unaffected at the *FWA* endogenous gene (Figure 4-6 D). Interestingly *idnl1-1 idnl2-1* double mutant plants showed a late flowering phenotype as strong as *idn2-1*. However, the *idn2-1 idnl1-1 idnl2-1* triple mutant did not show a defect any stronger than *idn2-1* (Figures 4-5 C and D). These data suggest partial redundancy between IDNL1 and IDNL2, which is consistent with their high similarity at the amino acid level. The function of IDNL2 can be fully compensated by IDNL1, but IDNL2 seemingly cannot entirely replace the function of IDNL1.

All known genes affecting *de novo* methylation are involved in the maintenance of DRM2-mediated methylation at several loci (Chan et al., 2004; Greenberg et al., 2011). Combining methylation sensitive enzymes and PCR or Southern blot analysis, as well as bisulfite sequencing techniques, we examined the methylation of status of known RdDM targets: *MEA-ISR*, *FWA* repeats, and the *AtSN1* transposon (Figures 4-6 A, B, C, and D). As expected based in its minor *de novo* methylation defect, *idnl1-1* caused a slight reduction in non-CG methylation at all tested loci. Similarly with the observed data for *de novo* methylation, *idnl2-1* did not display any defect in methylation whereas *idnl1-1 idnl2-1* double mutants and *idn2-1 idnl1-1 idnl2-1* triple mutants showed a drastic reduction in non-CG methylation levels. Again, this severe reduction was comparable to that observed in *idn2-1* mutant, reinforcing the hypothesis that IDNL1 and IDNL2 act redundantly, together with the required factor IDN2.

The IDN2 complex acts at a downstream step of RdDM

To determine where in the pathway IDNL proteins are acting, we analyzed the abundance of siRNAs at several loci. IDN2 complex members contain the double-stranded RNA-binding XS

domain, shared with another Arabidopsis protein, SGS3 (Bateman, 2002). SGS3 acts upstream of RNA-DEPENDENT RNA POLYMERASE 6 (RDR6) in a small RNA pathway that is distinct from RdDM (Mourrain et al., 2000). However, in previous work we have shown that IDN2 is not required to generate 24 nucleotide siRNAs associated with RdDM, thus does not act upstream of RDR2 (Ausin et al., 2009). The *idnl1-1 idnl2-1* double mutant has a similar effect as *idn2-1,* in that siRNAs are reduced at some loci, however not eliminated as is observed in mutants in *NUCLEAR RNA POLYMERASE D 1* (*NRPD1*)—the largest subunit of Pol IV (Figure 4-7 A). The siRNA levels were not further decreased in the triple mutant background, providing strong confirmation that the IDN2 complex acts at a downstream step from siRNA biogenesis. Instead, the siRNA pattern is reminiscent of the *NUCLEAR RNA POLYMERASE E 1* (*NRPE1*) mutants—the largest subunit of Pol V—which affects siRNA levels at a subset of RdDM targets (Zhang et al., 2007).

We next wanted to determine if the IDN2 complex is required for the accumulation of Pol V transcripts. Pol V has been shown to be an active polymerase that transcribes intergenic non-coding (IGN) regions that are necessary for the recruitment of downstream RdDM components such as AGO4 (Wierzbicki et al., 2008; Wierzbicki et al., 2009). Using quantitative reverse-transcriptase PCR (Q-RTPCR), we tested the accumulation of IGN transcripts at two loci—*IGN5* and *MEA-ISR*—in *idnl* single mutants, the *idnl1-1 idnl2-1* double mutant and the *idn2-1 idnl1-1 idnl2-1* triple mutant. Comparison with wild type levels showed no significant differences, placing the IDN2 complex downstream of the production of *IGN* transcripts (Figure 4-7 B).

Genome-wide methylation and expression analysis.

In order to gain a broader understanding of how the IDN2 complex affects the Arabidopsis epigenome, we performed shotgun bisulfite sequencing in wild type Columbia, the various *idn* mutants, and *drm2-2* as a positive control (Cokus et al., 2008). We defined differentially methylated regions (DMRs) that showed reduced CHH context methylation in the *drm2-2* and

idn libraries relative to Columbia, and plotted the respective densities across the five chromosomes in the Arabidopsis genome (Figure 4-8 A). We found that the patterns of DMRs were markedly similar for all mutants tested, indicating that there are few, if any, loci affected by the IDN2 complex outside of those affected by DRM2. This last point is further supported by comparing the *drm2-2* and *idn2-1* DMRs (Figure 4-8 B). We observed a striking overlap between the regions identified in these datasets, which is consistent with the *idn2-1* methylation effect observed at the known RdDM target loci tested (Figure 4-6). Moreover, when analyzing the methylation state of the "non-overlapping regions", we do in fact see a somewhat reduced DNA methylation state in *idn* mutants at *drm2-2* DMRs, and vice versa (Figure 4-8 C). This indicates that virtually all DRM2 targets are affected by loss of IDN2, even though some were excluded from the threshold used to call overlap of regions (Figure 4-8 B) due to the stringency cutoffs used to define DMRs.

An important aspect of the genome-wide bisulfite sequencing data comes from comparing the methylation of various *idn* mutants in DMRs defined in both *drm2-2* and *idn2-1* (Figure 4-8 C). In every instance, the patterns for *idn2-1,* the *idnl1-1 idnl2-1* double mutant and the *idn2-1 idnl1-1 idnl2-1* triple mutant phenocopy one another. The genomic methylation data serves as strong additional evidence that the components of the IDN2 complex are likely to function together, and are consistent with our methylation analysis at individual loci (Figures 4-5 B and C, and Figure 4-6).

We also performed whole-genome mRNA sequencing (mRNA-Seq) in the *idn* class of mutants and various other RdDM mutants to better understand the regulatory role of the IDN2 complex (Figures 4-9 and 4-10). Initially we performed the analysis in floral tissue, and observed a direct correlation ($P<1 \times 10^{-15}$) between *idn2-1* and *nrpe1-11* affected genes (Figures 4-9 A and B). Comparing the fold-change of genes affected in RdDM mutants as well as *idn* family mutants, we clearly can see a similar pattern of transcriptional control (Figure 4-9 C). We performed a second replicate in three-week-old leaf tissue—the same tissue-type used for

shotgun bisulfite sequencing—and observed the same trend as in the floral tissue (Figure 4-10). Together, the data indicates that the IDN2 complex helps mediate DNA methylation and transcriptional control of RdDM targets genome-wide.

Discussion

In this study we have described a complex containing IDN2 and two partially redundant paralogs to IDN2 that we named IDNL1 and IDNL2. While we were preparing this manuscript, an independent group corroborated our results indicating a role in RdDM for IDN2 paralogs (Xie et al., 2012). Structural analysis of the IDN2 XS + coiled-coil show that *in vitro* the protein tends to homodimerize. We propose that *in vivo*, the coiled-coil domain mediates the interaction between IDN2 and the coiled-coils of either IDNL1 or IDNL2 (Figure 4-11). It appears that IDNL1 and IDNL2 are interchangeable components of the complex, but IDN2 does not have a redundant partner. Genomic RNA-seq and BS-seq data strongly support the conclusion that null mutations in all three genes do not further enhance the RdDM phenotype of the IDN2 single mutant. Moreover, even if IDN2 has the ability to complex with itself *in vivo*—as it appears to be able to under crystallization conditions—this complex does not compensate for the loss of IDNL1/IDNL2. Therefore, the IDN2 together with either IDNL1 or IDNL2 are required for complete DRM2-mediated genome methylation.

In the study describing the discovery of IDN2, we noted that the DNA methylation phenotype of the *idn2-1* mutation is not as strong as the *drm2-2* mutation (Ausin et al., 2009). We hypothesized that this may be due to redundant activity of one or more XS+XH proteins in the IDN2 family. Given that we do not see an enhanced phenotype even in the triple mutant, apparently DRM2 still maintains some minimal activity in the absence of the IDN2 complex. However, we cannot discount the possibility that some other XS+XH protein(s) that do not interact with the components of the IDN2 complex fulfill a partial functional redundancy.

The fact that the IDN2 and IDNL1/IDNL2 proteins with such similar domain architecture are not only tightly complexed with each other, but are non-redundant with each other, raises the interesting question of what distinguishes the different activities of IDN2 and IDNL1/IDNL2. We have previously shown that the XS domain of IDN2 binds to double-stranded RNA with 5' overhangs (Ausin et al., 2009). Perhaps the XS domain of IDNL1/IDNL2 contains a slightly different RNA binding preference, which allows the IDN2 complex to bind to its target. Alternatively, all three proteins contain the conserved—but uncharacterized—XH domain (Bateman, 2002). Until the domain's function is better understood, we can only speculate on its role. However, we cannot rule out that different, or dual, XH activity is required for IDN2 complex function.

As previously mentioned, the XS-domain-containing protein SGS3 acts upstream of RDR6 in a distinct small RNA pathway (Mourrain et al., 2000). We have shown that the IDN2 complex is not required for RDR2 activity for primary siRNA generation (Figure 4-7 A). It is also known that SGS3 acts downstream of ARGONAUTE 1 (AGO1) cleavage of *TAS* transcripts (Yoshikawa et al., 2005). To extend the analogy to the RdDM pathway, it is possible that the IDN2 complex may act downstream of AGO4 (Figure 4-11). Given that IDN2 complex proteins contain zinc fingers, which may serve as DNA binding motifs, one possibility is that the IDN2 complex binds to RNA and DNA simultaneously (Figure 4-11). This could serve to anchor DNA methylation effectors to chromatin that is producing long non-coding Pol V transcripts. However, since some zinc fingers have been shown to bind RNA, we cannot rule out that the zinc fingers of IDN proteins might serve as further RNA binding motifs (Burdach et al., 2012; Font and Mackay, 2010; Hall, 2005). Although it has not been conclusively demonstrated that AGO4 slices Pol V transcripts, these transcripts are required to recruit AGO4 to chromatin (Wierzbicki et al., 2009). This model is consistent with the IDN2 complex acting downstream of Pol V transcription (Figure 4-7 B). If the IDN2 complex does interact directly with AGO4, the interaction is too transient to be reliably detected by our methods. Perhaps, the siRNA

disengages from AGO4, and binds to the Pol V transcript leaving double stranded RNA with a 5' overhang (Figure 4-11).

It is also unclear what purpose binding the Pol V transcripts by IDN2 complex would serve with regard to DNA methylation. The IDN2 complex does not seem to be required for Pol V transcript degradation, since we did not observed an increased abundance of *IGN* transcripts in the mutant backgrounds (Figure 4-7 B). If the IDN2 complex serves to bind the junction formed by the hybridization of an AGO4 bound siRNA and a Pol V transcript, this could serve to integrate the information from both the upstream siRNA generation part of the pathway (driven by the location of Pol IV transcription) and the downstream long non-coding RNA portion of the pathway (driven by the location of Pol V transcription). IDN2 complex could then serve as a signaling molecule for the recruitment or activation of chromatin modification enzymes, ultimately culminating in the recruitment or activation of DRM2 to methylate DNA. Continued studies on the role of XS+XH containing proteins will be an exciting area of research in understanding RNA-mediated transcriptional control.

MATERIALS AND METHODS

Plant materials

All plants utilized in this study are in the Columbia ecotype, and grown under long day conditions. The following mutant lines were used: *ago4-4* (described in (Greenberg et al., 2011)), *drm2-2* (SALK_150863), *nrpd1-4* (SALK_08305), *nrpe1-11* (SALK_02991), *idn2-1* (described in (Ausin et al., 2009)), *idnl1-1* (SALK_075378), and *idnl2-1* (SALK_012288). Information about the *idnl1-1* and *idnl2-1* T-DNA insertions can be found in Figure 4-4 C.

Protein expression and purification

The PCR-amplified cDNA fragment of *Arabidopsis* IDN2 (120-292) encoding the XS domain and a small segment of coiled coil region was cloned in pET-Sumo vector (Invitrogen) and overexpressed in *E. coli* BL21 (DE3) cells. The protein was purified from the soluble fraction by a Ni-affinity column which was followed by overnight treatment with SUMO protease Ulp1 at 4° C to cleave the His_6-SUMO tag. Cleaved His_6-SUMO tag was removed from the protein by second round of Ni-affinity column chromatography. Protein was further purified by gel-filtration chromatography using HiLoad 16/60 Superdex-75 prep grade column (GE healthcare). Purified proteins were concentrated to 15–20 mg/ml in 25 mM Tris-HCl, pH 8.0, 100 mM NaCl and 1 mM DTT, immediately frozen in liquid N_2, and stored at -80° C. L-selenomethionine (Se-Met)-labeled protein for *ab initio* phasing was produced by feedback inhibition of the methionine synthesis pathway.

Crystallization and structure determination

Crystals of IDN2 (120-292) was grown using the vapor-diffusion method by mixing the protein with an equal volume of reservoir solution containing 0.1 M BIS-TRIS pH 6.5, 2.0 M ammonium sulfate. Small crystals appeared overnight at 20 °C and grew to full size within 1 week. For data collection, crystals were flash frozen (100 K) in reservoir solution supplemented with 20% (v/v) ethylene glycol. Diffraction data sets were collected on 24-ID-C beamline at the Advanced Photon Source (APS). Data sets were integrated and scaled using the HKL2000 suite (Otwinowski and Minor, 1997).

Our attempt to solve the structure of IDN2 (120-292) using datasets collected on SeMet-labeled protein-containing crystal was unsuccessful due to presence of only one methionine (lack of sufficient phasing power). Phasing was finally carried out using an IDN2 construct

having residues 120-270 with residues Leu198 and Thr247 mutated to methionine to improve the phasing power (Table 4-2). The structure of Se-Met labeled IDN2 (120-270; L198M and T247M) was determined by the single-wavelength anomalous dispersion (SAD) method using Phenix.Autosol (Adams et al., 2010). Phase improvement was carried out using density modification producing a clearly interpretable electron density map from which an initial model was built manually using Coot (Emsley and Cowtan, 2004). The structures of IDN2 (120-292) was solved by molecular replacement program MOLREP (Vagin and Teplyakov, 2010) using the structure of the Se-Met labeled IDN2 (120-270; L198M and T247M). The model was completed using several rounds of manual model building in Coot (Emsley and Cowtan, 2004) and refinement using Phenix.Refine (Adams et al., 2010). The majority of the model has a clear and well-interpretable electron density map with the exception of a few solvent-exposed side chains, which were omitted in the final model. The geometry of the final model was checked using Procheck (Laskowski et al., 1993). The data collection and refinement statistics for IDN2 (120-292) is given in Table 4-2.

Southern blots

Approximately 4 µg of genomic DNA was separated on a 1% agarose gel, and then we transferred to Hybond N+ membranes. We blocked and washed the blot according to manufacturer's instructions (GE Healthcare). We probed the membranes using PCR products radiolabeled with alpha ^{32}P-dCTP using the Megaprime DNA Labeling System (GE cat. No. RPN1606). Primers used for probe amplification are listed in Table 4-3.

Bisulfite sequencing and analysis

For sodium bisulfite sequencing, DNA was treated using the EZ DNA Methylation Gold kit (Zymo Research) by following the manufacturer's instructions. Amplified PCR fragments from each analyzed locus were cloned into pCR2.1-TOPO (Invitrogen) and sequenced. We analyzed

15 to 22 clone sequences per sample using Lasergene SeqMan software. In order to distinguish the *FWA* transgene from the endogene, we destroyed a *BglII* restriction site in the transgenic copy in the region of PCR amplification. We then bisulfite treated genomic DNA of transgenic plants following a *BglII* digestion (37 degrees, overnight), which prevented amplification of the endogenous gene. Additionally, the transgenic copy of *FWA* was derived from the Landsberg ecotype, thus we could distinguish between the transgene and endogene based on the existence of three single nucleotide polymorphisms within the amplicon in case *BglII* digestion was not complete. Primers used for amplification are listed in Table 4-3.

HaeIII cutting assay

Analysis of asymmetric methylation at the *AtSN1* locus was performed exactly as described in (Deleris et al., 2010). Primers used for amplification are listed in Table 4-3.

Flowering time

We measured flowering time of plants as the total number of leaves (rosette and cauline leaves) developed by a plant at the time of flowering.

Generation of transgenic plants

Transgenic plants were generated as described in (Clough and Bent, 1998).

Gel filtration

We collected about 300 mg of young inflorescence tissue and homogenized in IP buffer, then spun in microfuge tubes for 5 min at 4C at 10,000rpm. Then we transferred the supernatant to a fresh tube and spun again. The supernatant was then filtered through a 0.2 micron filter and 500 ml were loaded onto a Superdex 200 (GE healthcare, 17-5175-01). 250 ml fractions were

collected and loaded in on a 4%–12% SDS-PAGE and then probed with anti-FLAG antibody, following standard western blot procedures.

Affinity purification and mass spectrometric analysis

~10 g of flower tissue from transgenic 9xMyc-BLRP-IDN2, IDNL1-9xMyc, or Columbia (negative control) was ground in liquid nitrogen, and resuspended in 50 ml of lysis buffer (LB: 50 mM Tris pH7.6, 150 mM NaCL, 5 mM MgCl2, 10% glycerol, 0.1% NP-40, 0.5 mM DTT, 1 mg/mL pepstatin, 1 mM PMSF and 1 protease inhibitor cocktail tablet (Roche, 14696200)). Each supernatant was incubated at 4° C for 2.5 hours with 200 ml of Dynabeads MyOne Streptavidin C1 (Invitrogen) for 9xMyc-BLRP-IDN2 or Monoclonal 9E10 agarose beads (Covance AFC-150P) for IDNL1-9xMyc; both incubations were also used for the negative control. The respective bead-bound complexes were then washed twice with 40 mL of LB and five times with 1 mL of LB. For each wash, the beads were rotated at 4° C for 5 minutes. Proteins were then released from the Streptavidin beads by 3C cleavage or from the 9E10 agarose beads with two 10 minutes incubation with 400 uL 8M urea. Mass spectrometric analyses were conducted as described in (Law et al., 2010).

Co-immunoprecipitation analysis

Co-immunoprecipitation and western blotting was performed as described in (Law et al., 2011). IP was performed with M2 Flag agarose (50% slurry, Sigma A2220). Western blotting was performed with ANTI-FLAG M2 Monoclonal Antibody-Peroxidase Conjugate (Sigma A 8592) and c-Myc 9E10 mouse monoclonal antibody (Santa Cruz Biotechnology, sc-40).

RNA analysis

Pol V transcript assays and siRNA northern blots were performed exactly as described in (Law et al., 2011). Sequences used for amplification and probing are listed in Table 4-3.

Genome-wide RNA sequencing and computational analysis

We prepared mRNA for Illumina sequencing from both floral and three-week-old leaf tissue. Summary of read counts for each library can be found in.

Floral tissue: Total RNA was isolated from mixed stage inflorescence tissues of Columbia and the various *idn* mutants using TRIzol reagent (Invitrogen). Total RNA (10 ug) for each sample were used for purifying the poly-A containing mRNA molecules, RNA amplification and synthesis of double stranded cDNAs that were ligated to adapters, thus preparing the libraries for sequencing on the Illumina GAII. The mRNA-Seq library preparation protocol was followed as described in Illumina's mRNA Sequencing Sample Preparation Guide (Illumina). Libraries were sequenced on an Illumina GAII at the Delaware Biotechnology Institute (Newark, DE).

Three-week-old leaf tissue: Total RNA was prepared using a TRIzol (Invitrogen) extraction from 0.5 grams of 3-week old plant aerial tissue. 4 µg of total RNA was then used to prepare libraries for Illumina sequencing, following the Illumina TruSeq RNA Sample Prep guidelines. Multiplexed samples were sequenced at 50bp length on an Illumina HiSeq 2000 instrument.

For data analysis, 50 bp sequences called by the Illumina pipeline, were mapped to the Arabidopsis genome (TAIR8) using Bowtie (Langmead et al., 2009). Only reads mapping uniquely to the genome with a maximum of 2 mismatches were used for further analysis. To quantify changes in gene expression, read counts over each Arabidopsis gene model were used to perform Fisher Exact Tests between genotypes. False discovery rates (FDR) were estimated by applying a Benjamini-Hochberg adjustment to resulting Fisher p-values. All statistical analysis was conducted with the R environment.

The mRNA sequence data are available from NCBI's Gene Expression Omnibus (GEO) and are accessible via GEO Series accession numbers GSE37206 (floral) and GSE36129 (leaf).

Shotgun bisulfite sequencing

Genomic DNA was extracted from one gram of 3-week old plant aerial tissue using a DNeasy Plant Maxi Kit (Qiagen). Libraries for bisulfite sequencing were generated and sequenced as described in (Feng et al., 2011), with the change that sequencing was carried out on an Illumina HiSeq 2000 instrument. Summary of library read counts can be found in Table S6. Reads were subsequently mapped to the Arabidopsis genome (TAIR8) using the BSseeker wrapper (Chen et al., 2010) of the Bowtie aligner (Langmead et al., 2009).

For data analysis, only cytosines with 5X coverage in all libraries compared were considered. DMRs were discovered using a sliding window approach with 200 bp window sliding at 50 bp intervals. The Fisher Exact Test was performed for methylated versus unmethylated cytosines for each context using the resultant windows, with FDRs estimated using a Benjamini-Hochberg adjustment of Fisher p-values calculated in the R environment. Windows with a FDR<0.05 were considered for further analysis and windows within 100 bp of each other were condensed to larger regions. Regions were then adjusted to extend to differentially methylated cytosines at each border. A cytosine was considered differentially methylated if it showed at least a 2-fold reduction in methylation percentage in the mutant. Finally regions were also filtered to have at least 10 differentially methylated cytosines and have an average 2-fold reduction in methylation percentage per cytosine.

The bisulfite sequence data are available from NCBI's Gene Expression Omnibus (GEO) and are accessible via GEO Series accession number GSE36143.

Protein databank accession number

Atomic coordinates and structure factor for the crystal structure of *Arabidopsis* IDN2 (120-292) has been deposited in the Protein Data Bank with the following accession code 4E8U.

FIGURE LEGENDS

Figure 4-1. *Arabidopsis* IDN2 domain architecture and crystal structure of IDN2-XS domain (120-292) of IDN2.

(A) Domain architecture of Arabidopsis IDN2 (top) and XS domain and a small segment of coiled-coil region (below) used for structural studies.

(B) Stereo view of structural superposition of IDN2-XS domain with RRM domain of FBP-interacting repressor (PDB ID: 2QFJ). The IDN2-XS domain is colored in cyan whereas the RRM domain is colored in brown.

(C) Quaternary (dimeric) structure of IDN2-XS domain. The two molecules are colored in cyan and light orange. The two subunits interact mainly via hydrophobic interactions formed by residues present in the α-helix of coiled coil segment. A few hydrophobic residues present on the surface of XS domain also interact with the residues present at the C-terminal end of the terminal helix.

(D) Electrostatic surface representation of dimeric IDN2-XS domain highlighting basic (blue) and acidic (red) regions on the IDN2-XS domain.

Figure 4-2. IDN2 amino acid sequence alignments.

(A) Structure-based sequence alignment of IDN2-XS domain with the RRM domain of FBP-interacting repressor. Secondary structural elements of IDN2-XS domain and RRM domain are shown above and below the aligned sequence, respectively. Additional secondary structural elements observed in IDN2-XS domain are labelled as βN, β1a, β1b, α3 and α4.

(B) Alignment of IDN2 region highlighted in middle square from Figure 4-1 B with the same regions of IDNL1 and IDNL2, respectively. Interacting hydrophobic residues (shaded in green) are conserved among the paralogs.

Figure 4-3. ***idn2-1 and idnl1-1*** **genetic complementation.**

The figure shows the *MEA-ISR* Southern blot phenotype of wild type, *idn2-1* and *idnl1-1 idnl2-1* mutants and several individual transgenic lines expressing either the IDN2::tag or IDNL1::tag.

Figure 4-4. IDNL1/IDNL2 proteins and mutant lines.

(A) Tree showing relative distances of XH/XS domain containing proteins in Arabidopsis. The protein sequences were aligned using ClustalW2, and a tree was generated using the Phylodendron Phylogenetic tree printer.

(B) RT-PCR of *idnl1-1* and *idnl2-1* alleles. *UBQ10* was used as a loading control. Primers for amplification can be found in Table 4-3.

(C) Gene models showing the location of the TDNA insertions for *idnl1-1* and *idnl2-1*.

Figure 4-5. IDN2-IDNLs *in vivo* interaction, and *de novo* methylation phenotype of *idnl* mutants.

(A) Gel filtration showing the elution profile of IDN2::3xFLAG fusion in two different mutant backgrounds. Arrows indicate the fraction in which size standards peaks.

(B) FLAG pull-down and co-immunoprecipitation assays confirming IDN2-IDNL1 interaction. Input lanes confirm expression of the epitope fusion proteins in the parental lines indicated above each lane. F1 represents a cross between the two parental lines.

(C) Flowering time measured as total number of leaves produced by wild type (Columbia), *idnI* mutants and *FWA*-transformed T_1 plants under long-day conditions. Error bars depict standard error.

(D) Methylation status of wild type and *idnI* mutants at the transgenic copy of *FWA*.

Figure 4-6. Maintenance methylation phenotype of the *idnI* mutants.

(A) Methylation-sensitive enzyme Southern hybridization assay at the *MEA-ISR* locus. *MspI* is blocked by methylation of the external C in CCGG context.

(B) *HaeIII* cutting assay at *AtSN1*. Genomic DNA is digested with *Hae III* and then Quantitative PCR is performed. X-values are relative to uncut and then to Columbia. Errors bar represent standard error from two independent experiments, each with two technical replicates. *HaeIII* is blocked by cytosine methylation in GGCC context.

(C) Methylation status of wild type (Columbia) and *idnI* mutants at endogenous *MEA-ISR.*

(D) Methylation status of Columbia and *idnI* mutants at endogenous *FWA.*

Figure 4-7. **Analysis of siRNAs and *IGN* transcripts.**

(A) Northern blot analysis showing siRNAs abundance in wild type, *idnI* mutants, and other RdDM mutants at several loci. Hybridization with *miR159* probe is shown as loading control.

(B) Quantitative RT-PCR analysis showing the relative abundance of *IGN* transcripts at *MEA-ISR* and *IGN5*. Y-values are first normalized to *ACTIN* and then normalized to wild type (Columbia). Error bars represent standard error from three independent replicas.

Figure 4-8. DMRs identified in *drm2-2* overlap with differentially methylated regions in *idn2* family mutants.

(A) Density of DMRs across the genome.

(B) Venn diagram showing overlap of *drm2-2* and *idn2-2* DMRs.

(C) Boxplots representing methylation levels of each DMR class represented by the Venn diagram for Columbia and mutant genomes.

Figure 4-9. The *idn* transcriptomes are similar to other RdDM mutants in floral tissue.

(A) Plot of log2 ratios of normalized read counts in *idn2-1* and *nrpe1-11* mutants for genes affected (FDR<1e-5) in the *idn2* mutant.

(B) Plot of log2 ratios for genes affected in the *nrpe1-11* mutant.

(C) Heat map of log2 ratios (mutant / Columbia) for various RdDM mutants and *idn* mutants for genes affected in the *idn2* mutant.

Figure 4-10. The *idn* transcriptomes are similar to other RdDM mutants in three-week-old leaf tissue.

(A) Plot of log2 ratios of normalized read counts in *idn2-1* and *nrpe1-11* mutants for genes affected (FDR<1e-5) in the *idn2* mutant.

(B) Plot of log2 ratios for genes affected in the *nrpe1-11* mutant.

(C) Heat map of log2 ratios (mutant / Columbia) for *nrpe1-11* mutants and *idn* mutants for genes affected in the *idn2* mutant.

Figure 4-11. Model for the role of the IDN2 complex in RdDM.

Pol V produces a transcript that recruits siRNA-loaded AGO4. The siRNA loaded in AGO4 hybridizes with the nascent Pol V transcript, and the AGO4 protein is released. The XS domain of IDN2 (magenta) is able to recognize this species of RNA. In this interpretation, IDNL1/IDNL2 (cyan) forms a complex with IDN2 in an anti-parallel orientation along their respective coiled-coils. It is possible that the XS domain of IDNL1/IDNL2 binds to a separate siRNA-Pol V transcript hybrid as well. The riboprotein

complex is stabilized to chromatin by the IDN2 complex proteins binding to DNA via their zinc fingers. Through an unknown mechanism, the IDN2 complex and bound RNA promote DNA methylation by DRM2.

Table 4-1. **Mass Spectrometric Analyses of IDN2 and IDNL1 Affinity Purifications.**

NSAF = Normalized Spectral Abundance Factor. Only proteins appearing in two experiments are shown.

Table 4-2. Crystallographic statistics for Se-Met labelled IDN2-XS (120-270; L198M and T247M) and IDN2-XS (120-292) constructs.

Table 4-3. List of primers and probes used in this study.

Figure 4-1

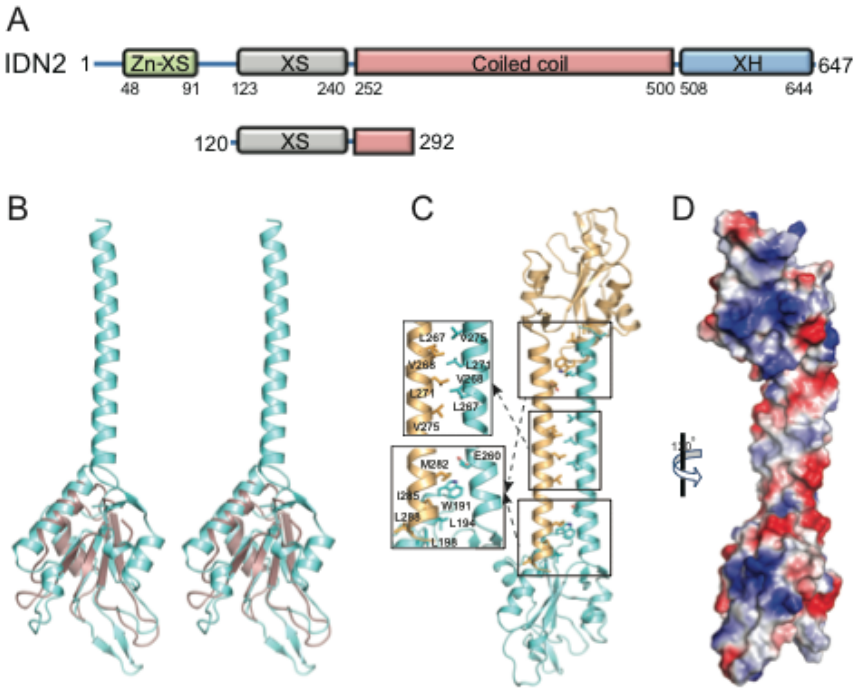

A
IDN2
1
Zn-XS
XS
Coiled coil
XH
647
48
91
123
240
252
500
508
644
120
XS
292
B
C
D
L267
V275
V268
L271
V268
L271
L267
V275
M282
E280
I285
W191
L284
L194
L198

Figure 4-2

A


XS_IDN2
RRM_2QFJ
XS_IDN2
RRM_2QFJ
XS_IDN2
RRM_2QFJ

B

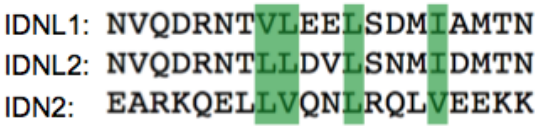

IDNL1: NVQDRNTVLEELSDMIAMTN
IDNL2: NVQDRNTLLDVLSNMIDMTN
IDN2: EARKQELLVQNLRQLVEEKK

Figure 4-3

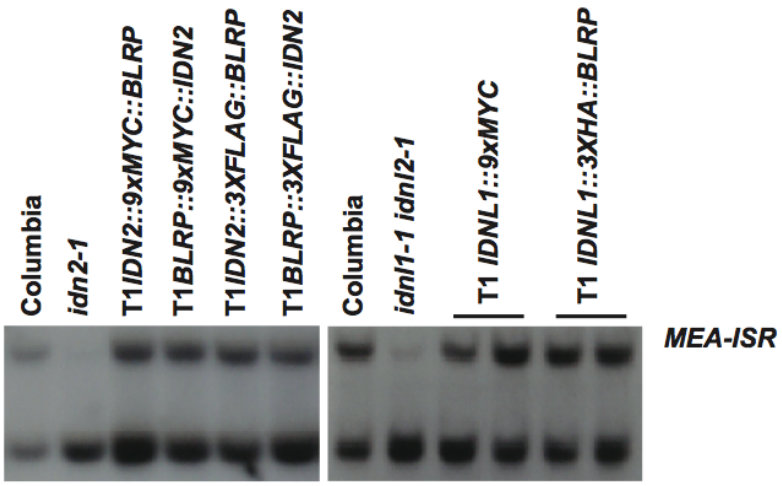

Columbia
idn2-1
T1IDN2::9xMYC::BLRP
T1BLRP::9xMYC::IDN2
T1IDN2::3XFLAG::BLRP
T1BLRP::3XFLAG::IDN2
Columbia
idnl1-1 idnl2-1
T1 IDNL1::9xMYC
T1 IDNL1::3XHA::BLRP
MEA-ISR

Figure 4-4

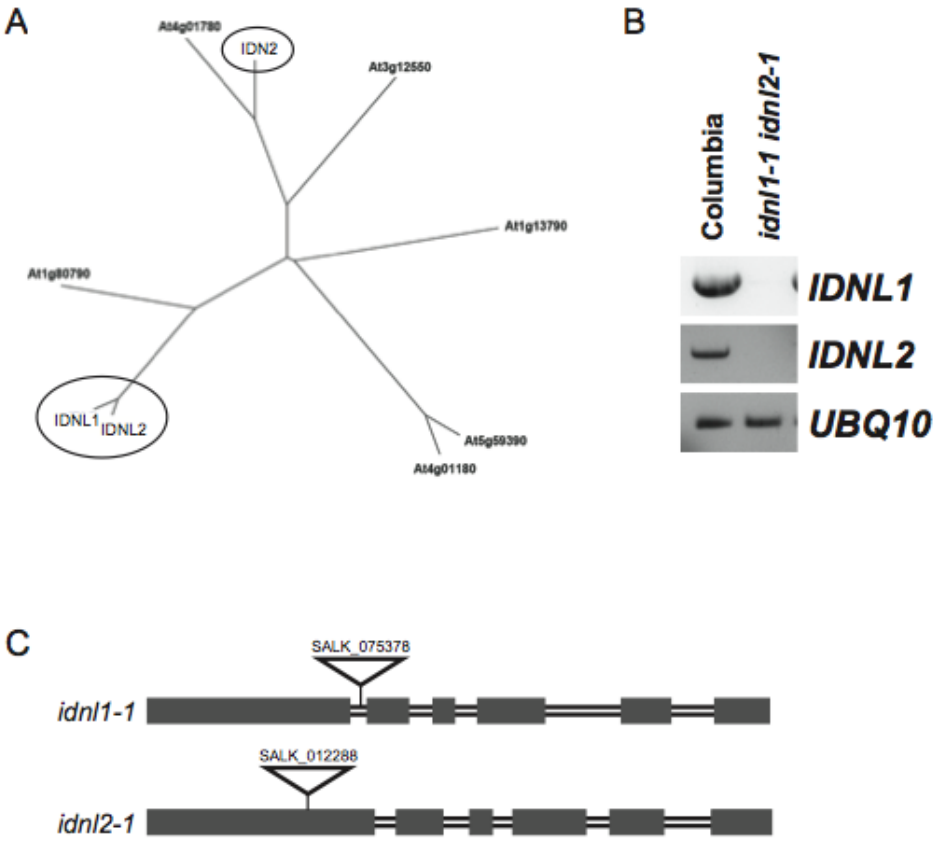

A
At4g01780
IDN2
At3g12550
At1g13790
At1g80790
IDNL1
IDNL2
At5g59390
At4g01180
B
Columbia
idnl1-1 idnl2-1
IDNL1
IDNL2
UBQ10
C
SALK_075378
idnl1-1
SALK_012288
idnl2-1

Figure 4-5

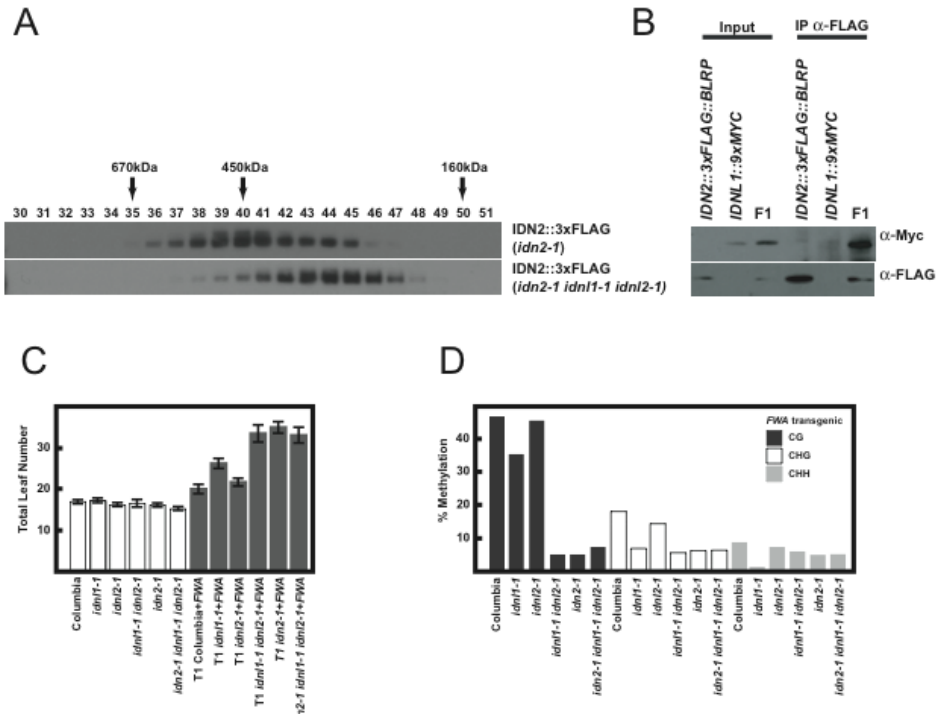

A
670kDa
450kDa
160kDa
30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51
IDN2::3xFLAG (idn2-1)
IDN2::3xFLAG (idn2-1 idnl1-1 idnl2-1)
B
Input
IP α-FLAG
IDN2::3xFLAG::BLRP
IDNL1::9xMYC
F1
α-Myc
α-FLAG
C
Total Leaf Number
Columbia
idnl1-1
idnl2-1
idnl1-1 idnl2-1
idn2-1
idn2-1 idnl1-1 idnl2-1
T1 Columbia+FWA
T1 idnl1-1+FWA
T1 idnl2-1+FWA
T1 idnl1-1 idnl2-1+FWA
T1 idn2-1+FWA
T1 idn2-1 idnl1-1 idnl2-1+FWA
D
% Methylation
FWA transgenic
CG
CHG
CHH

Figure 4-6

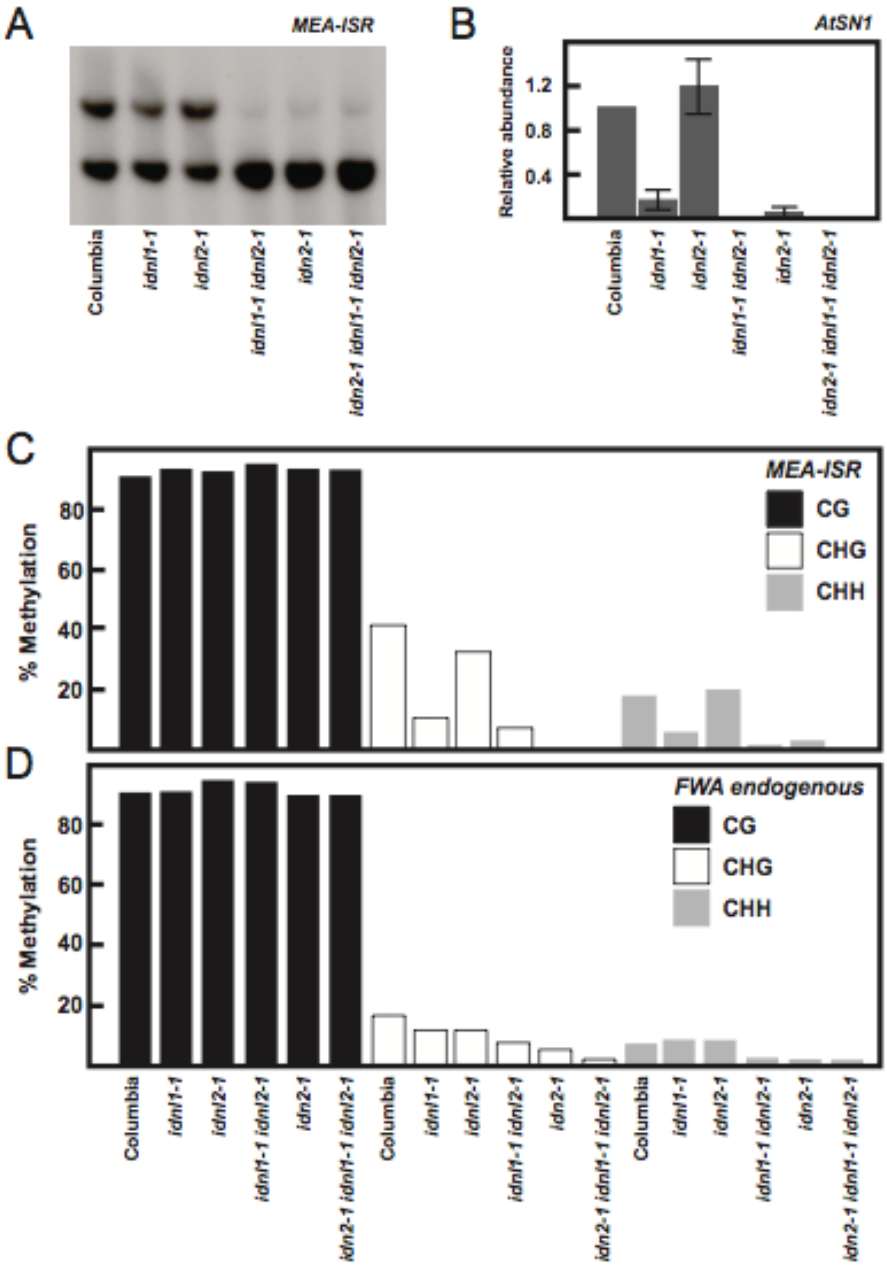

A
MEA-ISR
Columbia
idnl1-1
idnl2-1
idnl1-1 idnl2-1
idn2-1
idn2-1 idnl1-1 idnl2-1
B
AtSN1
Relative abundance
1.2
0.8
0.4
C
MEA-ISR
CG
CHG
CHH
% Methylation
80
60
40
20
D
FWA endogenous
CG
CHG
CHH
% Methylation

Figure 4-7

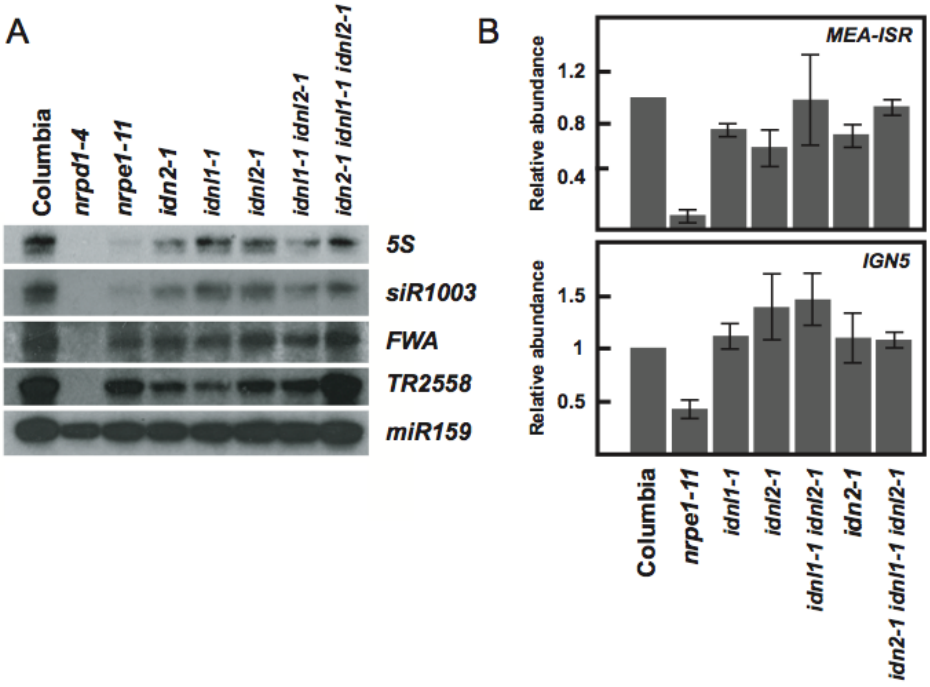

A
Columbia
nrpd1-4
nrpe1-11
idn2-1
idnl1-1
idnl2-1
idnl1-1 idnl2-1
idn2-1 idnl1-1 idnl2-1
5S
siR1003
FWA
TR2558
miR159
B
MEA-ISR
Relative abundance
1.2
0.8
0.4
IGN5
Relative abundance
1.5
1
0.5
Columbia
nrpe1-11
idnl1-1
idnl2-1
idnl1-1 idnl2-1
idn2-1
idn2-1 idnl1-1 idnl2-1

Figure 4-8

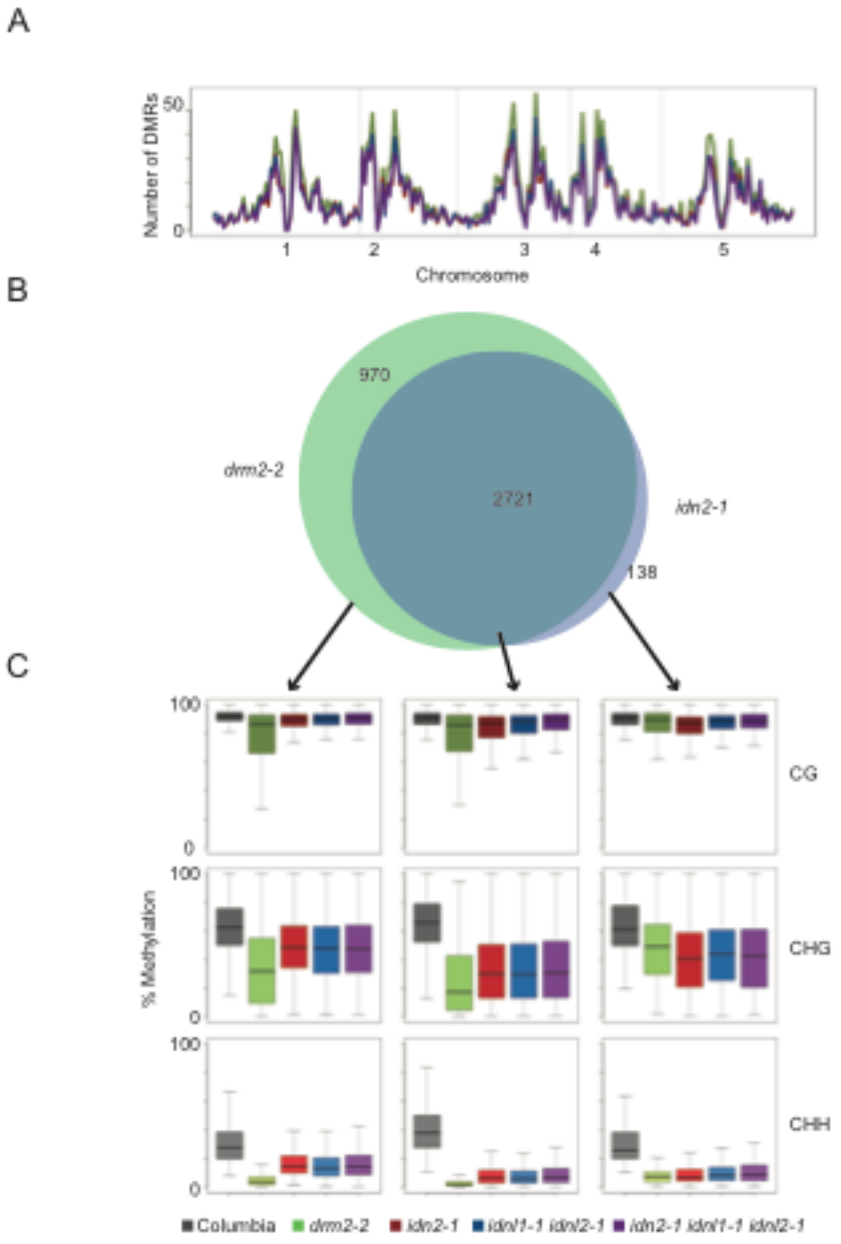

A
Number of DMRs
50
0
1
2
3
4
5
Chromosome
B
970
drm2-2
2721
idn2-1
138
C
100
0
CG
CHG
CHH
% Methylation
Columbia
drm2-2
idn2-1
idn1-1 idn2-1
idn2-1 idn1-1 idn2-1

Figure 4-9

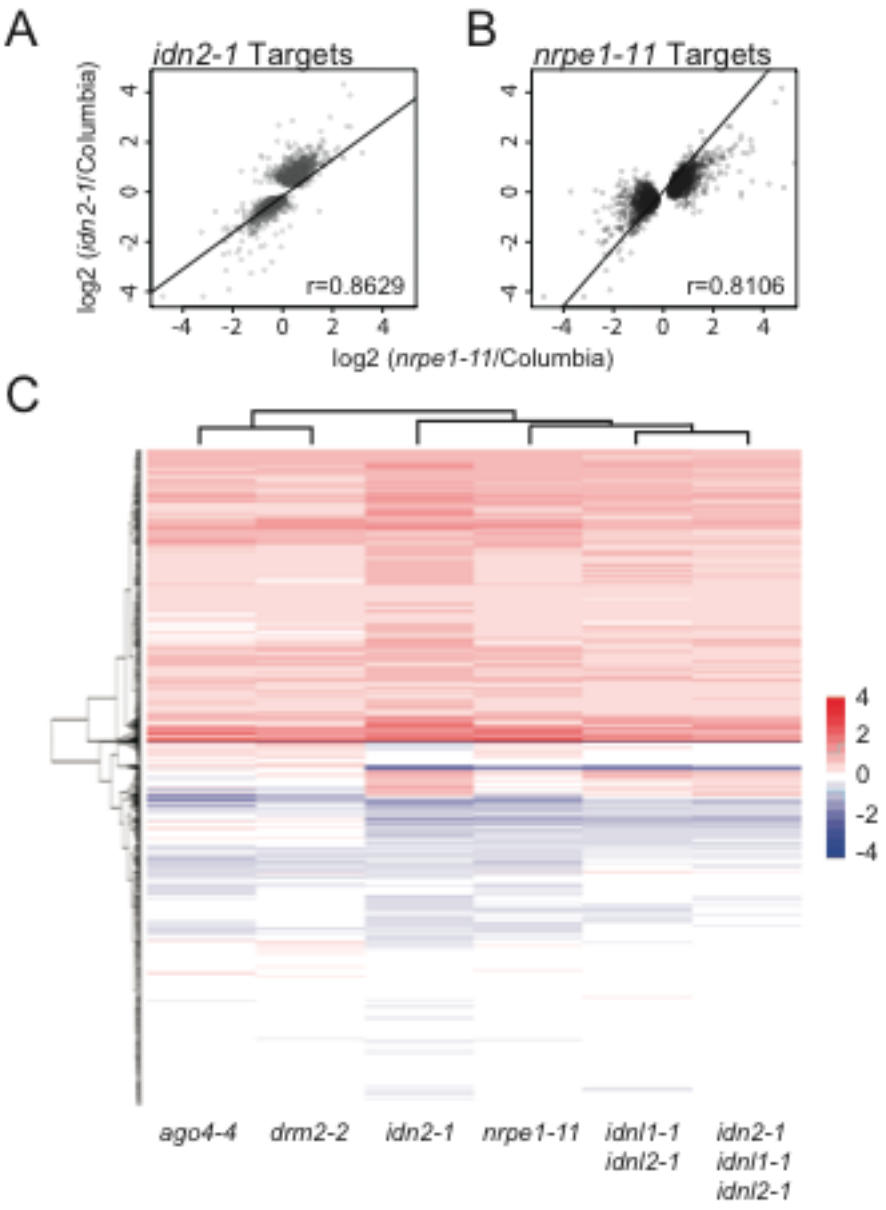

A
idn2-1 Targets
log2 (idn2-1/Columbia)
r=0.8629
B
nrpe1-11 Targets
r=0.8106
log2 (nrpe1-11/Columbia)
C
4
2
0
-2
-4
ago4-4
drm2-2
idn2-1
nrpe1-11
idnl1-1 idnl2-1
idn2-1 idnl1-1 idnl2-1

Figure 4-10

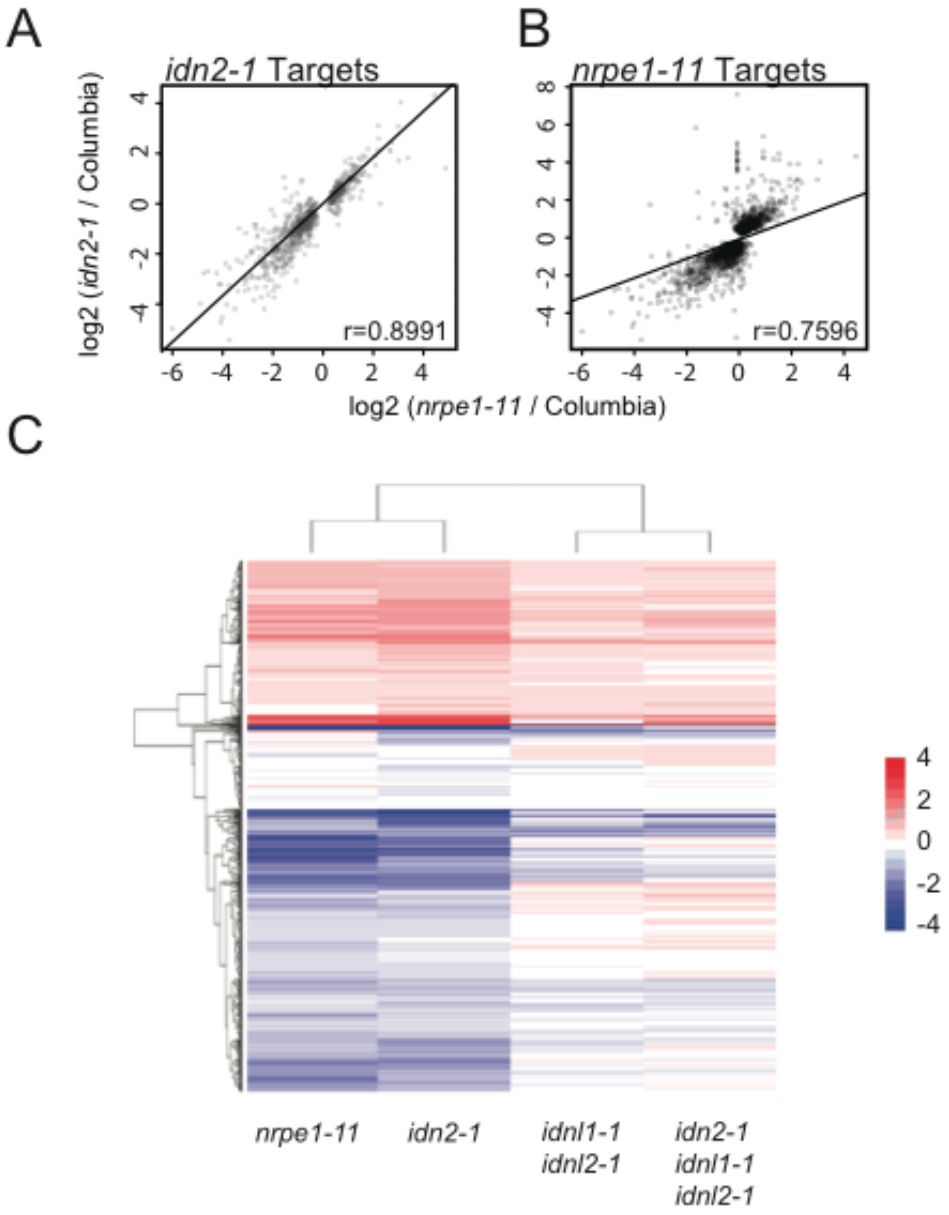

A
idn2-1 Targets
log2 (idn2-1 / Columbia)
r=0.8991
B
nrpe1-11 Targets
r=0.7596
log2 (nrpe1-11 / Columbia)
C
nrpe1-11
idn2-1
idnl1-1
idnl2-1
idn2-1
idnl1-1
idnl2-1
4
2
0
-2
-4

Figure 4-11

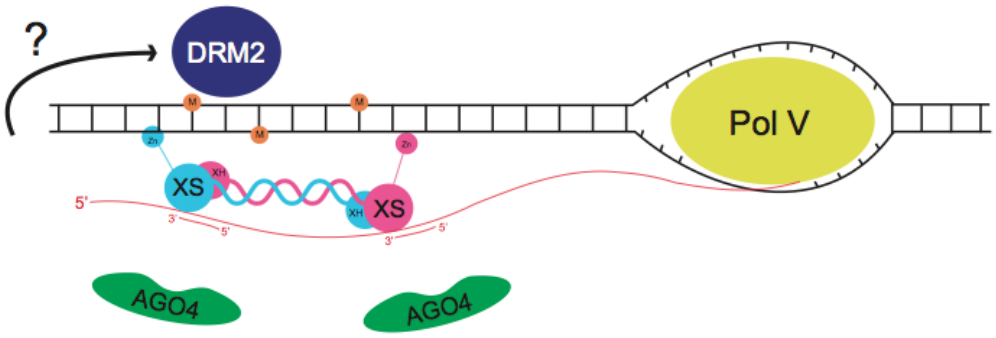


Table 4-1

IDN2 Purification	**Experiment I**				**Experiment II**			
	Spectra	UniPepts	Coverage	NSAF	Spectra	UniPepts	Coverage	NSAF
IDN2	651	91	68.0%	16261	398	52	53.0%	26405
AT1G15910 (IDNL1)	408	59	60.9%	10400	127	30	40.4%	8598
AT4G00380 (IDNL2)	163	39	49.4%	4148	82	24	32.3%	5543
AT5G24780	3	3	12.2%	179	5	3	10.4%	794
AT1G78300	7	6	27.0%	436	3	3	17.8%	497
AT4G16143	2	2	7.0%	72	3	2	7.0%	290
IDNL1 Purification								
	Spectra	UniPepts	Coverage	NSAF	Spectra	UniPepts	Coverage	NSAF
IDNL1	32	13	30.0%	755	103	20	33.0%	2238
IDN2	6	3	6.2%	138	87	22	31.4%	1852

Table 4-2

	Se-Met IDN2-XS (120-270; L198M and T247M)	IDN2-XS (120-290)
Data collection		
Wavelength (Å)	0.9790 (Se-Peak)	0.9792
Space group	$P3_121$	$C222_1$
Unit cell a, b, c (Å) α, β, γ (°)	 107.58, 107.58, 40.48 90, 90, 120	 52.37, 207.63, 119.63 90, 90, 90
Resolution[a] (Å)	50.0-2.25 (2.33-2.25)	50.0-2.7 (2.8-2.7)
Total reflections	975349	758144
Unique reflections	12971	18230
No of molecules/AU	1	2
Completeness (%)	99.8 (100.0)	99.0 (96.3)
$I/\sigma I$	42.6 (3.6)	38.5 (4.1)
Redundancy	8.2 (8.0)	10.3 (10.4)
R_{merge}	6.4 (60.2)	8.7 (54.7)
Refinement		
R_{work} / R_{free}		25.1/30.6
RMSD Bond lengths (Å) Bond angles (°)		 0.008 1.175
No. of atoms Protein/Water		 2615/25
Average B-factors (Å^2) Protein/Water		82.1/67.8
Ramachandran statistics (%) mostly allowed regions allowed regions generously allowed regions disallowed regions		 91.5 7.4 1.1 0.0

Table 4-3

Description	Sequence
5S siRNA Probe	ATGCCAAGTTTGGCCTCACGGTCT
ACTIN qRT-PCR	TCGTGGTGGTGAGTTTGTTAC
ACTIN qRT-PCR	CAGCATCATCACAAGCATCC
ACTIN qRT-PCR HEX Probe	TTTTCCCTAGTTGAGATGGGAATT-HEX
AtSN1 Chop-qPCR	TTTAAACATAAGAAGAAGTTCCTTTTTCATCTAC
AtSN1 Chop-qPCR	ACTTAATTAGCACTCAAATTAAACAAAATAAGT
FWA Bisulfite	GGTTTTATATTAATATTAAAGAGTTATGGGTYGAAGTTT
FWA Bisulfite	AACCAAAATCATTCTCTAAACAAAATATAAAAAAATC
FWA siRNA Probe	AGCAACCTTAAACAACCAAATAGCACTTGGACCAATGGCGAA
IDNL1 RT-PCR	ATGAGCATTTCTGATGAAGAGGC
IDNL1 RT-PCR	TCAGGTTCTTTTGCGTTTCAGA
IDNL2 RT-PCR	ATGGACATTTCTGATGAAGAGTCTG
IDNL2 RT-PCR	TCAGGTTCTTTTGCGTTTCAGA
IGN5 Pol V HEX Probe	TTGGGCCGAATAACAGCAAGTCC-HEX
IGN5 Pol V Transcript	TCCCGAGAAGAGTAGAACAAATGCTAAAA
IGN5 Pol V Transcript	CTGAGGTATTCCATAGCCCCTGATCC
MEA-ISR Bisulfite	AAAGTGGTTGTAGTTTATGAAAGGTTTTAT
MEA-ISR Bisulfite	CTTAAAAAATTTTCAACTCATTTTTAAAAAA
MEA-ISR Pol V HEX Probe	TGACCACGGTTAAATGGCGGG-HEX
MEA-ISR Pol V Transcript	CGCGAACGACTATTGCTAAA
MEA-ISR Pol V Transcript	TGAAATCTAACCGGATTTTGG
MEA-ISR Southern Probe	AAACCTTTCGTAAGCTACAGCCACTTTGTT
MEA-ISR Southern Probe	TCGGATTGGTTCTTCCTACCTCTTTACCTT
miR159 miRNA Probe	TAGAGCTCCCTTCAATCCAAA
siR1003 Probe	ATGCCAAGTTTGGCCTCACGGTCT
TR2558 siRNA Probe	AAGCTATCGGTCATGCTGATGAATATGAGGAGGAA
UBQ10 RT-PCR	GATCTTTGCCGGAAAACAATTGGAGG
UBQ10 RT-PCR	CGACTTGTCATTAGAAAGAAAGAGAT

REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Ausin, I., Mockler, T.C., Chory, J., and Jacobsen, S.E. (2009). IDN1 and IDN2 are required for de novo DNA methylation in Arabidopsis thaliana. Nat Struct Mol Biol *16*, 1325-1327.

Bateman, A. (2002). The SGS3 protein involved in PTGS finds a family. BMC Bioinformatics *3*, 21.

Burdach, J., O'Connell, M.R., Mackay, J.P., and Crossley, M. (2012). Two-timing zinc finger transcription factors liaising with RNA. Trends Biochem Sci.

Cao, X., and Jacobsen, S.E. (2002). Role of the Arabidopsis DRM methyltransferases in *de novo* DNA methylation and gene silencing. Current Biology *12*, 1138-1144.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Chen, P.Y., Cokus, S.J., and Pellegrini, M. (2010). BS Seeker: precise mapping for bisulfite sequencing. BMC Bioinformatics *11*, 203.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J *16*, 735-743.

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

Deleris, A., Greenberg, M.V., Ausin, I., Law, R.W., Moissiard, G., Schubert, D., and Jacobsen, S.E. (2010). Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation. EMBO Rep *11*, 950-955.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Feng, S., Rubbi, L., Jacobsen, S.E., and Pellegrini, M. (2011). Determining DNA methylation profiles using sequencing. Methods Mol Biol *733*, 223-238.

Font, J., and Mackay, J.P. (2010). Beyond DNA: zinc finger domains as RNA-binding modules. Methods Mol Biol *649*, 479-491.

Fukunaga, R., and Doudna, J.A. (2009). dsRNA with 5' overhangs contributes to endogenous and antiviral RNA silencing pathways in plants. EMBO J *28*, 545-555.

Glick, E., Zrachya, A., Levy, Y., Mett, A., Gidoni, D., Belausov, E., Citovsky, V., and Gafni, Y. (2008). Interaction with host SGS3 is required for suppression of RNA silencing by tomato yellow leaf curl virus V2 protein. Proc Natl Acad Sci U S A *105*, 157-161.

Greenberg, M.V., Ausin, I., Chan, S.W., Cokus, S.J., Cuperus, J.T., Feng, S., Law, J.A., Chu, C., Pellegrini, M., Carrington, J.C.*, et al.* (2011). Identification of genes required for de novo DNA methylation in Arabidopsis. Epigenetics *6*.

Hall, T.M. (2005). Multiple modes of RNA recognition by zinc finger proteins. Curr Opin Struct Biol *15*, 367-373.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). Procheck - a Program to Check the Stereochemical Quality of Protein Structures. J Appl Crystallogr *26*, 283-291.

Law, J.A., Ausin, I., Johnson, L.M., Vashisht, A.A., Zhu, J.K., Wohlschlegel, J.A., and Jacobsen, S.E. (2010). A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in Arabidopsis. Curr Biol *20*, 951-956.

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

Law, J.A., Vashisht, A.A., Wohlschlegel, J.A., and Jacobsen, S.E. (2011). SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS Genet *7*, e1002195.

Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Jouette, D., Lacombe, A.M., Nikic, S., Picault, N.*, et al.* (2000). Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell *101*, 533-542.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Method Enzymol *276*, 307-326.

Qin, Y., Ye, H., Tang, N., and Xiong, L. (2009). Systematic identification of X1-homologous genes reveals a family involved in stress responses in rice. Plant Mol Biol *71*, 483-496.

Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The Late Flowering Phenotype of fwa Mutants Is Caused by Gain-of-Function Epigenetic Alleles of a Homeodomain Gene. Mol Cell *6*, 791-802.

Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr *66*, 22-25.

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell *135*, 635-648.

Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet *41*, 630-634.

Xie, M., Ren, G., Costa-Nunes, P., Pontes, O., and Yu, B. (2012). A subgroup of SGS3-like proteins act redundantly in RNA-directed DNA methylation. Nucleic Acids Res.

Yoshikawa, M., Peragine, A., Park, M.Y., and Poethig, R.S. (2005). A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes Dev *19*, 2164-2175.

Zhang, D., and Trudeau, V.L. (2008). The XS domain of a plant specific SGS3 protein adopts a unique RNA recognition motif (RRM) fold. Cell Cycle *7*, 2268-2270.

Zhang, X., Henderson, I.R., Lu, C., Green, P.J., and Jacobsen, S.E. (2007). Role of RNA polymerase IV in plant small RNA metabolism. Proc Natl Acad Sci U S A *104*, 4536-4541.

Zheng, Z., Xing, Y., He, X.J., Li, W., Hu, Y., Yadav, S.K., Oh, J., and Zhu, J.K. (2010). An SGS3-like protein functions in RNA-directed DNA methylation and transcriptional gene silencing in Arabidopsis. Plant J *62*, 92-99.

CHAPTER 5

Involvement of a JmjC domain-containing

histone demethylase in DRM2-mediated maintenance DNA Methylation.

INTRODUCTION

Cytosine DNA methylation is an epigenetic modification that is conserved in all kingdoms of eukaryotes and is largely associated with heterochromatic regions undergoing transcriptional gene silencing. In the model plant *Arabidopsis thaliana*, at least three methylation pathways exist and each is associated with a specific methyltransferase. METHYLTRANSFERASE 1 (MET1) is a homolog of mammalian DNMT1, and maintains methylation in the CG dinucleotide context. CHROMOMETHYLASE 3 (CMT3) is a plant specific methyltransferase that preferentially deposits the methyl-mark in CHG contexts (where H is A, T, or C). Finally the mammalian DNMT3 homolog DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) performs *de novo* methylation, and maintains CHH—or asymmetric—methylation through an siRNA-driven signal in a process known as RNA-directed DNA methylation (RdDM) (Law and Jacobsen, 2010). At some loci, CMT3 and DRM2 act redundantly to control the maintenance of both the CHG and CHH methylation, however DRM2 alone is responsible for *de novo* DNA methylation (Cao and Jacobsen, 2002a; Chan et al., 2004).

Methylation patterns are tightly correlated with specific histone modification signatures. For example, genome-wide studies in Arabidopsis have shown that Histone 3 Lysine 9 dimethylation (H3K9m2) is a histone mark that is highly co-incidental with CHG methylation and endogenous clusters of siRNAs (Bernatavichute et al., 2008). H3K9m2 directed by the KRYPTONITE (KYP), SUVH5, and SUVH6 histone methyltransferases is required for the maintenance of CHG DNA methylation (Ebbs and Bender, 2006; Jackson et al., 2002; Malagnac et al., 2002), likely through direct targeting of CMT3 (Lindroth et al., 2004). Conversely, Histone 3 Lysine 4 mono/di/trimethylation (H3K4m1/2/3) is strongly anti-correlated with DNA methylation at non-genic silent loci (Zhang et al., 2009).

The discovery in mammals of two classes of enzymes capable of demethylating histones, Lysine-Specific Demethylase1 (LSD1) (Shi et al., 2004) and Jumonji-C (JmjC) domain

containing proteins (Klose et al., 2006), revealed that active removal of methyl-marks from histones is necessary for proper epigenetic regulation. Recently two plant homologs of the mammalian histone demethylase LSD1 termed LSD1-LIKE1 (LDL1) and LSD1-LIKE 2 (LDL2) have been shown to be required for H3K4 demethylation at the *FLC* and *FWA* loci (Jiang et al., 2007). While *FLC* is not a DNA methylated gene, *FWA* transcription is controlled by DNA methylation at the tandem repeats in its 5' UTR, and *FWA* hypomethylation results in ectopic expression and a late-flowering phenotype (Soppe et al., 2000). Interestingly, *ldl1 ldl2* double mutants flower late, and molecular analysis revealed hypomethylation at *FWA*. These data suggest that persistent H3K4 demethylation is required to maintain DNA methylation at some loci in the genome. To gain further insight into the relationship between active histone demethylation and DNA methylation at silent loci, we compiled a collection of homozygous T-DNA insertion mutants in genes containing JmjC domains in *Arabidopsis*. We show that JMJ14 is required to maintain full levels of non-CG methylation at sites controlled by DRM2. We were further able to show that the loss of non-CG methylation in *jmj14* mutants corresponded with increases in H3K4m3 marks suggesting that JMJ14 targets DNA methylated loci. Interestingly, *jmj14* mutants showed no effect on DRM2-mediated establishment of methylation of an incoming *FWA* transgene, which is in contrast with all other mutants tested in the DRM2 pathway (Ausin et al., 2009; Chan et al., 2004; Johnson et al., 2008; Law and Jacobsen, 2010). These results suggest that establishment and maintenance of methylation mediated by DRM2 can be differentially regulated and that JMJ14 has a specific role in the maintenance of RdDM.

RESULTS

jmj14 mutations affect non-CG maintenance methylation

Arabidopsis contains 21 genes with domains homologous to JmjC histone demethylases (Hong et al., 2009; Lu et al., 2008). In order to examine potential effects on DNA methylation, we

analyzed 17 JmjC mutants, for which null alleles were available, at the *MEDEA-INTERGENIC SUBTELOMERIC REPEATS* (*MEA-ISR*) locus by Southern blotting (Table 5-1). *MEA-ISR* is a set of seven tandem repeats downstream of the *MEDEA* (*MEA*) gene. Both MET1 (CG methylation) and DRM2 (CHG and CHH methylation) maintain DNA methylation at *MEA-ISR*, and hypomethylation phenotypes can be observed after digestion with the methylation-sensitive enzyme MspI (Cao and Jacobsen, 2002a). By Southern blot, we were able to observe a consistent reduction of *MEA-ISR* methylation in two null alleles of *jmj14* (Figure 5-1 A). JMJ14 was the first name given to the protein encoded by At4g20400 (Lu et al., 2008). Subsequent publications have referred to the protein as JMJ4 and PKDM7B (Jeong et al., 2009; Yang et al., 2010), but here we use the original nomenclature. In order to confirm the *jmj14* methylation defect, we performed bisulfite sequencing at the *MEA-ISR* locus (Figure 5-1 B). The bisulfite data showed a reduction in non-CG methylation, but CG methylation was unchanged compared to the wild type control. This indicates that the *jmj14* mutation interacts with the DRM2 pathway, but not the MET1 pathway.

In order to confirm the genetic interaction of JMJ14 with the DRM2 pathway, we examined the effect of the mutation on other RdDM targets. Analysis of the methylation state of the 5' UTR of *FWA* was performed by bisulfite sequencing. *FWA*, like *MEA-ISR*, is mainly targeted by MET1 and DRM2 (Cao and Jacobsen, 2002a). Similar to the bisulfite data at *MEA-ISR*, we observed a reduction in non-CG methylation but no effect at CG sites at *FWA* (Figure 5-1 C). Finally, to examine DRM2-dependent methylation at the transposable element *AtSN1*, DNA from both wild type and *jmj14* mutants was digested with the restriction endonuclease HaeIII, and analyzed by real-time quantitative PCR using primers that amplify a region spanning three restriction sites (Figure 5-1 D). HaeIII recognizes GGCC sequences for cleavage, but it is sensitive to methylation of the internal cytosine. In the region amplified, such methylation would give rise to three mCHH sites: GGmCCC, GGmCCT, and GGmCCA. Relative quantification of uncut DNA in the digested samples showed a significant decrease in asymmetric CHH

methylation in *jmj14* mutants compared to wild type, although not to the same extent as in *drm2*. To examine if the *jmj14* mutant defects were specific to the DRM2 pathway, we also analyzed the methylation state of *Ta3*—a single-copy transposable element that is methylated by CMT3 but not DRM2 (Cao and Jacobsen, 2002a). We observed no effect on methylation in any context for *jmj14* compared to the wild type control (Figure 5-1 E). This indicates that JMJ14 acts primarily in the DRM2 pathway.

jmj14 affects chromatin at RdDM target loci

To examine the localization of JMJ14, we created a C-terminal epitope tagged (9X Myc) *JMJ14* transgene driven by the endogenous *JMJ14* promoter and showed that this transgene fully complements the early-flowering phenotype (Jeong et al., 2009) of the *jmj14* mutant (Figures 5-2 A and B). Immunostaining for the Myc epitope revealed strong nuclear staining, consistent with the function of JMJ14 as a histone demethylase. Interestingly, we observed a specific pattern in which staining was rather uniformly present throughout the nucleoplasm but not in the nucleolus and the chromocenters (areas of dense heterochromatin that are highly enriched for H3K9m2) (Figure 5-2 C). This pattern is very similar to that previously found for DRM2 (Li et al., 2006), consistent with the hypothesis that JMJ14 acts in the DRM2 pathway.

Phylogenetic analyses have shown that JMJ14 shows the highest sequence similarity to human KDM5/JARID1 family histone demethylases (Lu et al., 2008) which are capable of specifically demethylating H3K4m1, H3K4m2, H3K4m3 (Christensen et al., 2007; Iwase et al., 2007; Lee et al., 2007; Seward et al., 2007). A recombinant JMJ14 was shown to demethylate efficiently H3K4m3 *in vitro* and to a lesser extent H3K4m2 and H3K4m1 (Jeong et al., 2009; Lu et al., 2010; Yang et al., 2010). This H3K4 demethylase activity was confirmed by an *in vivo* assay in *N. benthamiana* where overexpression of *JMJ14* correlated with a strong reduction of H3K4m3 and H3K4m2 marks (Lu et al., 2010). Finally, in Arabidopsis, JMJ14 was shown to demethylate H3K4m3 and H3K4m2 at two loci involved in floral transition (and not controlled by

DNA methylation) (Jeong et al., 2009; Yang et al., 2010).

This suggests that the defect in DNA methylation at non-CG sites was caused by an increase in H3K4 methylation in *jmj14* mutants. To confirm this hypothesis, we used chromatin immunoprecipitation (ChIP) to assess the levels of H3K4m2 and H3K4m3 at silent loci analyzed for DNA methylation, in wild type and *jmj14*. We observed a consistent increase in H3K4m3 marks at *AtSN1*, *FWA*, and *MEA-ISR* (Figures 5-3 A, B, and C). The extent of this increase was in the range of what was previously published in *jmj14* mutants at the floral transition loci *FLOWERING LOCUS T* (*FT*) and *TWIN SISTER OF FT* (*TSF*) (Jeong et al., 2009; Yang et al., 2010). We also saw a small but significant increase in H3K4m2 at the *FWA* locus, but not at *AtSN1* or *MEA-ISR* (Figures 5-3 A, B, and C). The minor effects on H3K4m2 might be due to the redundant activity of other demethylases such as LDL1 and LDL2. In *ldl1 ldl2*, H3K4m2 marks at *FWA* tandem-repeats were increased by 1.6 times compared to wild type and this was correlated with a defect in *FWA* methylation (Jiang et al., 2007). In general, these results show that JMJ14 may directly target silent chromatin, and suggest that the active removal of H3K4 methyl-marks at silent loci may be necessary for DRM2 to maintain proper DNA methylation patterns.

jmj14 does not affect *de novo* DNA methylation

All components of the RdDM machinery that have been tested thus far have been shown to be required both for DRM2-dependent non-CG maintenance DNA methylation at *MEA-ISR* and other loci, and for establishment of methylation in all sequence contexts on previously unmethylated sequences—or *de novo* methylation—of an incoming transgene (Ausin et al., 2009; Chan et al., 2004; Johnson et al., 2008; Law and Jacobsen, 2010). When *FWA* is introduced into wild type plants, siRNAs are able to target the repeats in the 5'UTR and the incoming transgene becomes methylated, and thus silenced. However, in RdDM mutants, the transgene remains unmethylated in all sequence contexts and is expressed (Cao and

Jacobsen, 2002b; Chan et al., 2004). Given that we observed non-CG maintenance methylation phenotypes at known RdDM targets in *jmj14*, we utilized the *FWA* transgene system to test for a role of JMJ14 in *de novo* methylation. Ectopic *FWA* expression leads to a late-flowering phenotype that gives a quantitative readout of the methylation establishment phenotype.

The *jmj14* mutant flowers earlier than wild type plants, which has previously been shown to be due to de-repression of *FT* (Figure 5-4 A) (Jeong et al., 2009; Lu et al., 2010; Yang et al., 2010). Surprisingly, we found that after *FWA* transformation, the *jmj14* mutant did not flower late, and in fact continued to flower earlier than wild type control plants that had been transformed with *FWA* (Figure 5-4 A). We note that other mutants with weak RdDM phenotypes such as *dicer-like 3 (dcl3)* that show only partial losses of *MEA-ISR* methylation (equivalent to those of *jmj14*), do show substantial effects on *FWA de novo* DNA methylation establishment, and thus flower late (Henderson et al., 2006). These results suggest that the *jmj14* mutation does not affect *FWA de novo* DNA methylation.

To confirm these findings, we analyzed the methylation state of the newly introduced *FWA* transgene by bisulfite sequencing (Figure 5-4 B). We found that in the *FWA* transgene, CG methylation levels of the *jmj14* mutant were comparable to wild type, however there was a significant decrease in non-CG methylation. By contrast, the *dcl3* mutant shows substantially less *de novo* methylation than wild type in all three sequence contexts even though it exhibited a similar non-CG maintenance phenotype (Henderson et al., 2006). These results show that the CG DNA methylation that is primarily responsible for silencing *FWA* is fully established in *jmj14*. Once CG methylation is established, it is maintained by the MET1 pathway independently of DRM2, while DRM2 maintains non-CG marks. Consistent with a role in DRM2-mediated maintenance of non-CG methylation, and like at the *FWA* endogene (Figure 5-1 C), we observed that maintenance of CHG and CHH methylation at the *FWA* transgene was reduced in the *jmj14* mutant (Figure 5-4 B).

DISCUSSION

JMJ14 is required for the maintenance of DRM2-mediated non-CG DNA methylation. Consistent with our findings, a recent report described the identification of JMJ14 through a forward-genetic screen for mutants impaired in hairpin-induced transcriptional silencing of the *PHYTOENE DESATURASE* (*PDS*) endogene (Searle et al., 2010).

We observed a moderate but consistent increase in H3K4m3 levels at RdDM targets analyzed in *jmj14,* suggesting that active demethylation of H3K4 is required for proper DRM2-pathway function, perhaps due to competition of the active H3K4 methylation mark with repressive marks such as DNA methylation (Figure 5-5). The fact that two different enzyme families—JmjC domain and LSD-like (Jiang et al., 2007)—have roles in the demethylation of H3K4 methyl-marks at silent loci/RdDM targets underlies the importance of the persistent removal of those marks for the maintenance of proper DNA methylation patterns.

Interestingly, *jmj14* mutants showed no effect on DRM2-mediated *de novo* methylation of an incoming *FWA* transgene. This is in contrast to all other mutants tested in the DRM2 pathway: *nrpd1*, *nrpe1, dcl3*, *rdr2*, *ago4*, *drd1*, *suvh2*, *dms3*, and *idn2* (Ausin et al., 2009; Chan et al., 2004; Johnson et al., 2008; Law and Jacobsen, 2010). Thus, JMJ14 is required to maintain non-CG methylation patterns but does not seem to be involved in the initial targeting of DNA methylation. This is an interesting finding as it implies that the maintenance activity of DRM2 can be mechanistically distinguished from its *de novo* methylation establishment activity, suggesting that during the maintenance phase there is another level of regulation of DRM2 activity by histones. The relationship between DRM2 activity and H3K4 methylation status is also interesting in light of mechanisms of action of the mammalian DRM2 homolog, DNMT3A. DNMT3A is in part recruited to silent loci through the interaction with a related protein (DNMT3L) that can bind H3 specifically when K4 is unmethylated (Jia et al., 2007; Ooi et al.,

2007). Future analyses should shed light on how H3K4 methyl-marks antagonize the DRM2 pathway in Arabidopsis.

METHODS

Plant Materials. We used the following Arabidopsis strains: wild type Col-0 and the recessive alleles *dcl3-1* and *drm2-2* in the Col-0 background. The list of alleles of JmjC mutants tested is presented in Table 5-1.

Southern blotting. DNA from young flowers was extracted using a standard CTAB protocol. 1 mg of genomic DNA was digested overnight with MspI. The digestion was run on a 1% agarose gel, transferred to Hybond N+ membranes, blocked and washed according to manufacturer instructions (GE Healthcare). Membranes were probed with a PCR product radiolabeled with alpha ^{32}P-dCTP using the Megaprime DNA Labeling System. *MEA-ISR* PCR product was generated with primers listed in Table 5-2.

Bisulfite analysis. DNA from young flowers was extracted using a standard CTAB protocol. We performed sodium bisulfite sequencing using EZ DNA Methylation Gold (Zymo Research) by following the manufacturer's instructions. Following amplication of bisulfite treated DNA, we cloned the resulting PCR fragments into pCR2.1-TOPO (Invitrogen) and analyzed 15 to 30 clones per sample. The *FWA* transgene was distinguished from the endogene by BglII digestion prior to bisulfite treatement (see *FWA* Transformation methods) and elimination of any clones containing Col-0 polymorphisms from the data set after sequencing. To compare the converted clones to the original unconverted sequence, we used the sequence alignment tool of CLC Workbench^{TM} software. We counted the converted/unconverted cytosines at each site manually and subsequently calculated the percent of methylation. All primers are listed in Table 5-2.

HaeIII Chop-qPCR. DNA from young flowers was extracted using a standard CTAB protocol. 200ng of genomic DNA was digested overnight at 37 degrees with HaeIII side-by-side with samples containing buffer and no enzyme (undigested). Quantitative real-time PCR validation of uncut DNA after HaeIII digestion was performed using the Biorad SYBR Green SuperMix on a MX3000 Stratagene cycler. The PCR parameters are as follows: 1 cycle of 10 minutes at 95° C, 40 cycles of 30 seconds at 95° C, 1 min at 55° C, 1 min at 72° C. PCR primers sequences are listed in Table 5-2.

***FWA* Transformation.** We performed *FWA* transformation using either an AGL0 *Agrobacterium tumefaciens* strain carrying an empty pCAMBIA3300 vector or a pCAMBIA3300 vector with an engineered version of the L*er* copy of *FWA* in which a BglII site was converted into a EcoRI site. For selection, we plated T1 seeds on MS media containing a 1:12,000 dilution of Finale^{TM}.

Flowering-time Analysis. We measured flowering time as the total number of leaves (rosette and cauline leaves) developed by a plant.

Generation of Epitope-tagged Complementing Lines. Epitope-tagged protein constructs were made using a modified Gateway cloning system for expression in plants. 1.6 kb of genomic DNA upstream the *JMJ14* ORF and the entire ORF was cloned into pENTR. An AscI restriction site was introduced at the TAA and a 9X Myc epitope tag (pLJ217) was introduced. The tagged construct was then recombined into a modified pDEST vector and introduced into Agrobacterium strain AGL1.

Protein immunofluorescence analysis. We prepared nuclei for immunofluorescent imaging as described in Li et al., 2006. Primary mouse monoclonal anti-Myc (Covance 9E10) and rabbit

polyclonal anti-H3K9m2 (Upstate) were used at a 1:200 dilution. Secondary anti-mouse FITC (Abcam) and anti-rabbit Rodamine (Jackson Immuno Research) were used at a 1:200 dilution. DNA was stained using Vectashield mounting medium containing DAPI (Vector Laboratories). Images were captured with the Zeiss AxioImager Z1 microscope with the Hamamatsu Orca-er camera at 100X magnification and analyzed using the Zeiss Axiovision software. Zeiss FL filter sets used in this study: Zeiss 49 (DAPI), Zeiss 38 (EGFP), and Zeiss 43 (Cy 3).

Chromatin immunoprecipitation (ChIP). The ChIP experiments were performed as previously described (Johnson et al., 2002) with the following modifications. Five grams of cross-linked rosette leaves (just before bolting) were ground under liquid nitrogen and resuspended in 30 mL of lysis buffer plus protease inhibitors. Cells were disrupted for 12 minutes with a Dianode Bioruptor (30 seconds on, 30 seconds off; high setting). The chromatin isolated was used for 2 IPs: H3K4m3-containing chromatin was immunoprecipitated with 5 mg of anti-H3K4m3 from Diagenode (pAb-003-050) and H3K4m2-containing chromatin was immunoprecipitated with 5 mg of anti-H3K4m2 from Abcam (AB32356). After reversal of crosslinking overnight, the immunoprecipitated DNA was purified by a regular DNA extraction protocol and analyzed by real-time PCR with TaqMan Probes (*AtSN1*, *FWA*, *ICDH*) or using the Biorad SYBR-Green Supermix (*MEA-ISR*, *ICDH*) on a MX3000. Sequences of primers and probes can be found in Table 5-2.

FIGURE LEGENDS

Figure 5-1. DNA methylation analysis of *jmj14* mutants.

(A) *MEA-ISR* Southern blot. Genomic DNA was digested with the non-CG methylation-sensitive restriction endonuclease MspI, and probed for *MEA-ISR*. The high molecular weight band represents methylated DNA and the low molecular weight band represents

unmethylated DNA. Two alleles of *jmj14* exhibit a methylation phenotype intermediate between wild type and the *drm2* mutant.

(B) *MEA-ISR* bisulfite sequencing. Genomic DNA was treated with sodium bisulfite and amplified with primers specific for *MEA-ISR*. Sequencing reveals an effect at non-CG sites compared to wild type, but not in the CG context.

(C) *FWA* endogene bisulfite sequencing. The *FWA* locus exhibits a similar pattern to *MEA-ISR* in the *jmj14-1* mutant.

(D) *AtSN1* HaeIII Chop-qPCR. Genomic DNA was digested with non-CG methylation-sensitive restriction endonuclease HaeIII. Digested DNA was quantified by real-time PCR with primers specific for a region of *AtSN1* spanning three restriction sites, and the signal was normalized to an undigested control. Two *jmj14* alleles exhibit significantly more digestion compared to the wild type control, thus less methylation. (E) *Ta3* bisulfite sequencing. The methylation state of *Ta3* shows no discernable defect in the *jmj14* mutant compared to wild type.

Figure 5-2. Analysis of complementing Myc-tagged lines.

(A) Myc-tagged JMJ14 constructs complement the early-flowering phenotype observed in the *jmj14-1* mutant background.

(B) Flowering-time assay. Quantification of complementation for tagged JMJ14 lines. Note: Line #10 was used for immunofluorescence assay.

(C) Immunolocalization of epitope-tagged JMJ14. A transgenic line expressing Myc-tagged complementing JMJ14 under its endogenous promoter was analyzed by fluorescent microscopy. JMJ14 is localized in the nucleus, but is depleted from the chromocenters (marked by H3K9m2 enrichment and dense DAPI staining).

Figure 5-3. Analysis of H3K4m2 and H3K4m3 state at RdDM targets by chromatin immunoprecipitation (ChIP).

The immunoprecipitated DNA corresponding to *AtSN1* (A), *FWA* (B), and *MEA-ISR* (C) was quantified by real-time PCR and normalized to an internal control—we used an intergenic region upstream of the *ISOCITRATE DEHYDROGENASE* gene (*ICDH-IGR*) and unlikely to be targeted by JMJ14. The fold enrichment in *jmj14-1* over wild type is shown at each locus (the wild type values being set to one). The values are the average ratio obtained from three independent ChIP experiments +/- standard error.

Figure 5-4. *de novo* DNA methylation analysis.

(A) *FWA* flowering-time assay. Total leaf number upon flowering was assessed for wild type Col-0 and *jmj14-1* for both *FWA* and empty-vector transformants.

(B) *FWA* transgene bisulfite sequencing. *jmj14-1* transformants exhibit a minimal effect on CG methylation compared to *dcl3-1*. The effect on non-CG may be due to a maintenance defect after the initial methylation has been established.

Figure 5-5. Model for the role of JMJ14 in DRM2-mediated maintenance methylation.

It is proposed that H3K4 methylation inhibits DRM2 pathway components. Active demethylation of the residue is needed for complete DRM2 maintenance activity.

Table 5-1. List of homozygous T-DNA insertion lines for JmjC-domain containing genes used in this study.

Table 5-2. List of primers and probes used in this study.

Figure 5-1

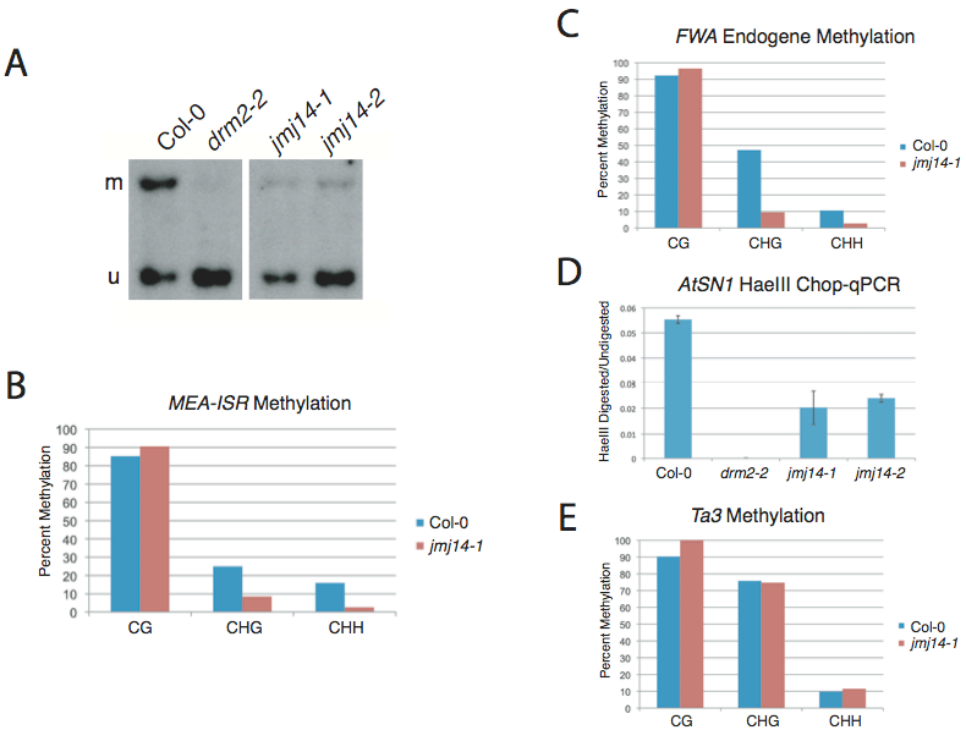

A
Col-0
drm2-2
jmj14-1
jmj14-2
m
u
B
MEA-ISR Methylation
Percent Methylation
CG
CHG
CHH
Col-0
jmj14-1
C
FWA Endogene Methylation
Percent Methylation
CG
CHG
CHH
Col-0
jmj14-1
D
AtSN1 HaeIII Chop-qPCR
HaeIII Digested/Undigested
Col-0
drm2-2
jmj14-1
jmj14-2
E
Ta3 Methylation
Percent Methylation
CG
CHG
CHH
Col-0
jmj14-1

Figure 5-2

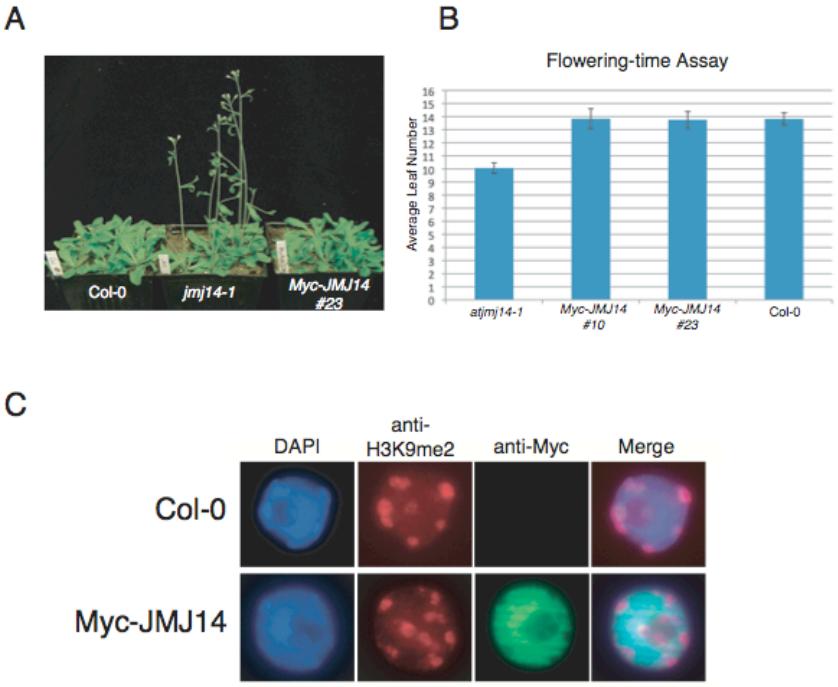

A
Col-0
jmj14-1
Myc-JMJ14 #23
B
Flowering-time Assay
Average Leaf Number
atjmj14-1
Myc-JMJ14 #10
Myc-JMJ14 #23
Col-0
C
DAPI
anti-H3K9me2
anti-Myc
Merge
Col-0
Myc-JMJ14

Figure 5-3

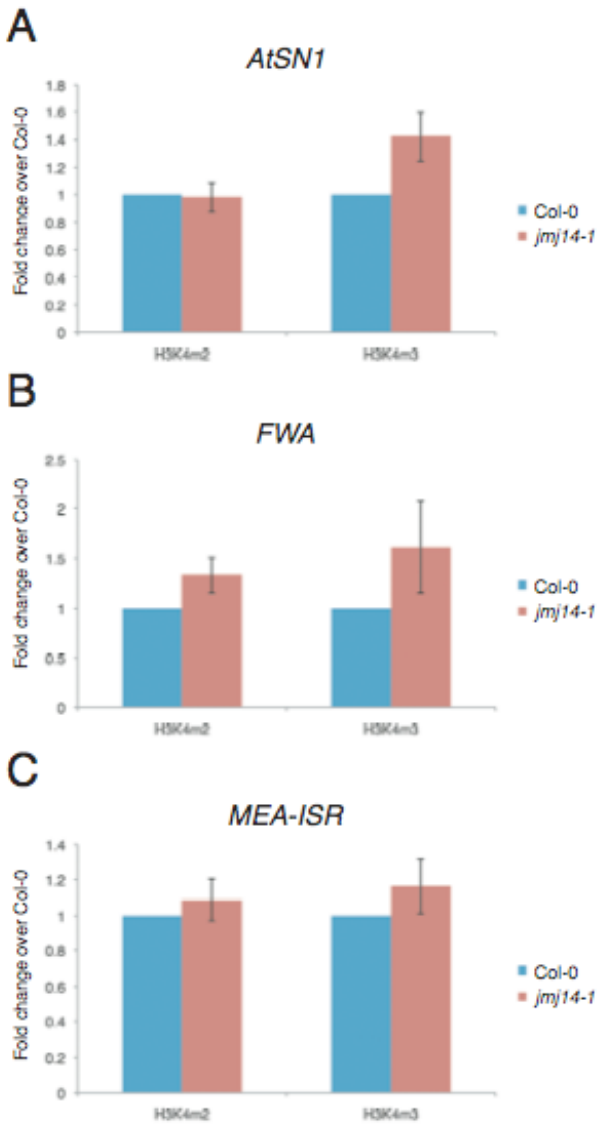

A
AtSN1
Fold change over Col-0
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
H3K4m2
H3K4m3
Col-0
jmj14-1
B
FWA
Fold change over Col-0
2.5
2
1.5
1
0.5
0
H3K4m2
H3K4m3
Col-0
jmj14-1
C
MEA-ISR
Fold change over Col-0
1.4
1.2
1
0.8
0.6
0.4
0.2
0
H3K4m2
H3K4m3
Col-0
jmj14-1

Figure 5-4

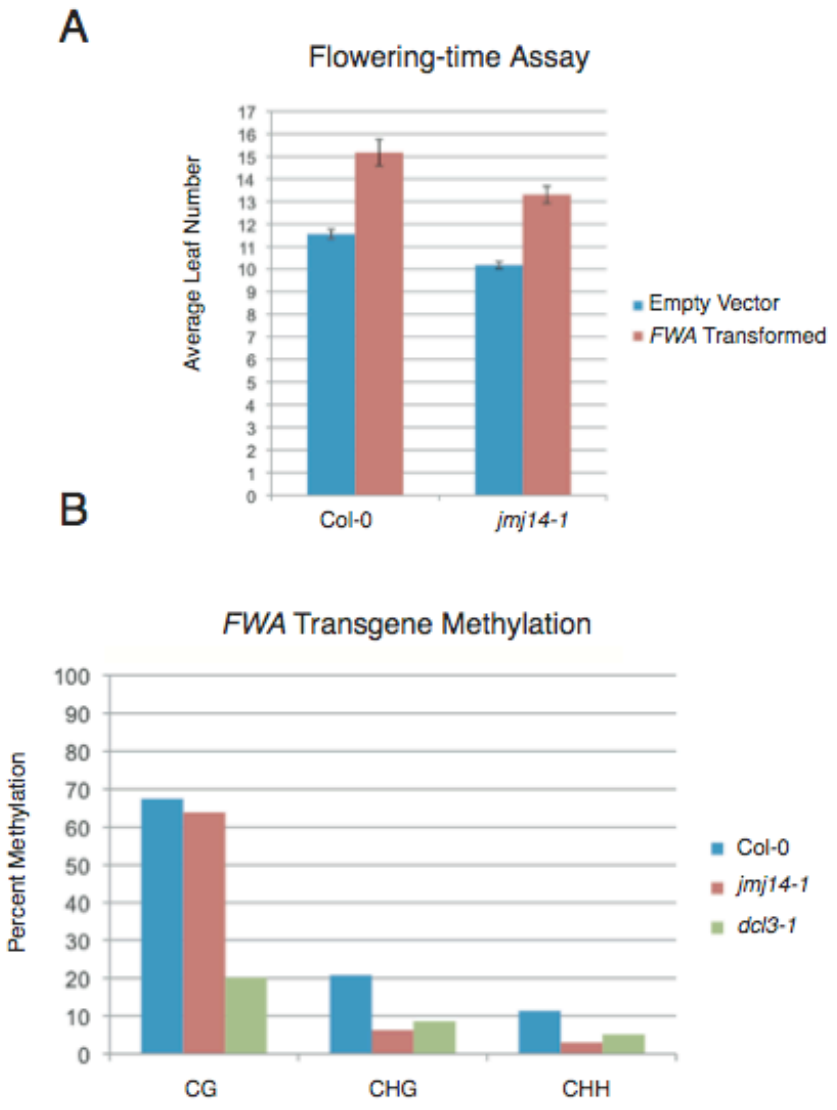

A
Flowering-time Assay
Average Leaf Number
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
Col-0
jmj14-1
Empty Vector
FWA Transformed
B
FWA Transgene Methylation
Percent Methylation
100
90
80
70
60
50
40
30
20
10
0
CG
CHG
CHH
Col-0
jmj14-1
dcl3-1

Figure 5-5

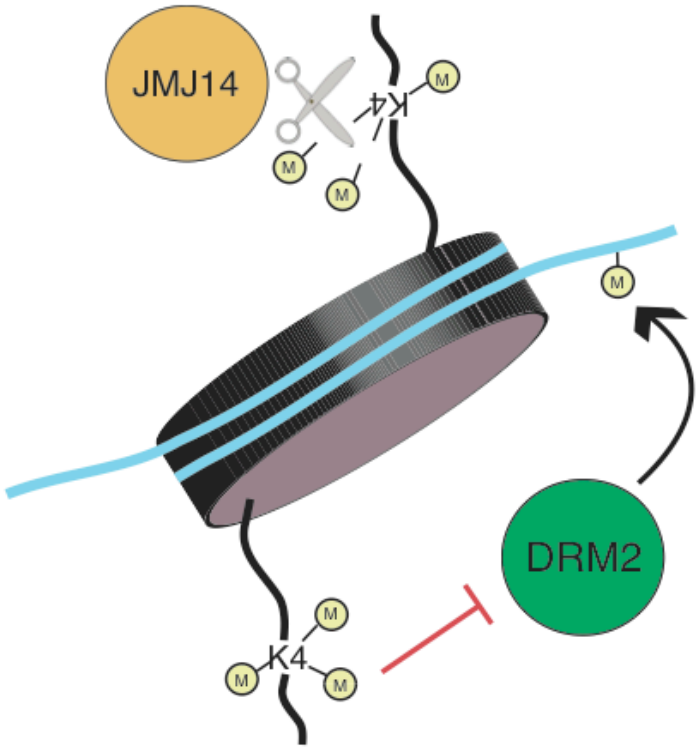


Table 5-1

Gene	AGI	Allele	Ecotype
JMJ16	At1g08620	SALK_029530	Col-0
JMJ19	At2g38950	SALK_025269	Col-0
JMJ18	At1g30810	null allele not available	
JMJ15	At2g34880	null allele not available	
JMJ14	At4g20400	SALK_135712 (*atjmj14-1*)	Col-0
		SALK_136058 (*atjmj14-2*)	Col-0
JMJ11	At5g04240	SALK_074694	Col-0
JMJ12	At3g48430	SALK_122006	Col-0
JMJ13	At5g46910	null allele not available	
JMJ17	At1g63490	FLAG_076A07	Col-0
JMJ20	At5g63080	FLAG_316E04	Col-0
JMJ21	At1g78280	SALK_000651	Col-0
JMJ22	At5g06550	SAIL_680_G02	Col-0
JMJ26	At1g11950	FLAG_484A05	Col-0
JMJ29	At1g62310	FLAG_390G09	Col-0
JMJ25	At3g07610	SALK_004652	Col-0
JMJ27	At4g00990	SALK_103092	Col-0
JMJ24	At1g09060	GABI_085H03	Col-0
JMJ28	At4g21430	WiscDsLox263E02	Col-0
unnamed	At3g45880	SALK_003313	Col-0
JMJ31	At5g19840	null allele not available	
JMJ30	At3g20810	GABI_454C10	Col-0

Table 5-2

MEA-ISR Southern Probe	AAACCTTTCGTAAGCTACAGCCACTTTGTT
MEA-ISR Southern Probe	TCGGATTGGTTCTTCCTACCTCTTTACCTT
MEA-ISR Bisulfite	AAAGTGGTTGTAGTTTATGAAAGGTTTTAT
MEA-ISR Bisulfite	CTTAAAAAATTTTCAACTCATTTTTAAAAAA
FWA Bisulfite	GGTTTTATATTAATATTAAAGAGTTATGGGTYGAAGTTT
FWA Bisulfite	AACCAAAATCATTCTCTAAACAAAATATAAAAAAATC
Ta3 Bisulfite	GAGAATYAGGTTAATAAGAAAGTGAAGTGTT
Ta3 Bisulfite	CCACTRATTCCTRAAACACAACATTTCTRCTRATA
AtSN1 Chop-qPCR	ACTTAATTAGCACTCAAATTAAACAAAATAAGT
AtSN1 Chop-qPCR	TTTAAACATAAGAAGAAGTTCCTTTTTCATCTAC
MEA-ISR ChIP	TTTAGGTATTAGCTCGTTTGGTTTTA
MEA-ISR ChIP	TCCCGCCATTTAACCGTG
FWA ChIP	ATAAAGAGCGGCGCAAGAT
FWA ChIP	CGCTCTAGGGTTTTTGCTTT
FWA ChIP Probe	CAAATAGCACTTGGACCAATGGCG
AtSN1 ChIP	GTTGGCCCAGTGGTAAATCT
AtSN1 ChIP	TGGTGGTTGTACAAGCCTAGTT
AtSN1 ChIP Probe	ATCTCCCAGAGGCGGGACCC
ICDH-IGR ChIP	AGGCCCCATCTCACAAATAC
ICDH-IGR ChIP	GTCGCCAGGTAGATTTGGTT
ICDH-IGR ChIP Probe	TCCGGTTAGACCTTAACGTGGGTCA

REFERENCES

Ausin, I., Mockler, T.C., Chory, J., and Jacobsen, S.E. (2009). IDN1 and IDN2 are required for de novo DNA methylation in Arabidopsis thaliana. Nat Struct Mol Biol *16*, 1325-1327.

Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One *3*, e3156.

Cao, X., and Jacobsen, S.E. (2002a). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A *99 Suppl 4*, 16491-16498.

Cao, X., and Jacobsen, S.E. (2002b). Role of the Arabidopsis DRM methyltransferases in *de novo* DNA methylation and gene silencing. Current Biology *12*, 1138-1144.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Christensen, J., Agger, K., Cloos, P.A., Pasini, D., Rose, S., Sennels, L., Rappsilber, J., Hansen, K.H., Salcini, A.E., and Helin, K. (2007). RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. Cell *128*, 1063-1076.

Ebbs, M.L., and Bender, J. (2006). Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. Plant Cell *18*, 1166-1176.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat Genet *38*, 721-725.

Hong, E.H., Jeong, Y.M., Ryu, J.Y., Amasino, R.M., Noh, B., and Noh, Y.S. (2009). Temporal and spatial expression patterns of nine Arabidopsis genes encoding Jumonji C-domain proteins. Mol Cells *27*, 481-490.

Iwase, S., Lan, F., Bayliss, P., de la Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A., Roberts, T.M., and Shi, Y. (2007). The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. Cell *128*, 1077-1088.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556-560.

Jeong, J.H., Song, H.R., Ko, J.H., Jeong, Y.M., Kwon, Y.E., Seol, J.H., Amasino, R.M., Noh, B., and Noh, Y.S. (2009). Repression of FLOWERING LOCUS T chromatin by functionally redundant histone H3 lysine 4 demethylases in Arabidopsis. PLoS One *4*, e8033.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature *449*, 248-251.

Jiang, D., Yang, W., He, Y., and Amasino, R.M. (2007). Arabidopsis relatives of the human lysine-specific Demethylase1 repress the expression of FWA and FLOWERING LOCUS C and thus promote the floral transition. Plant Cell *19*, 2975-2987.

Johnson, L., Cao, X., and Jacobsen, S. (2002). Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. Curr Biol *12*, 1360-1367.

Johnson, L.M., Law, J.A., Khattar, A., Henderson, I.R., and Jacobsen, S.E. (2008). SRA-domain proteins required for DRM2-mediated de novo DNA methylation. PLoS Genet *4*, e1000280.

Klose, R.J., Kallin, E.M., and Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. Nat Rev Genet *7*, 715-727.

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

Lee, M.G., Norman, J., Shilatifard, A., and Shiekhattar, R. (2007). Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBLR, a polycomb-like protein. Cell *128*, 877-887.

Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. (2006). An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. Cell *126*, 93-106.

Lindroth, A.M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I.*, et al.* (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. Embo J *23*, 4286-4296.

Lu, F., Cui, X., Zhang, S., Liu, C., and Cao, X. (2010). JMJ14 is an H3K4 demethylase regulating flowering time in Arabidopsis. Cell Res *20*, 387-390.

Lu, F., Li, G., Cui, X., Liu, C., Wang, X.J., and Cao, X. (2008). Comparative analysis of JmjC domain-containing proteins reveals the potential histone demethylases in Arabidopsis and rice. J Integr Plant Biol *50*, 886-896.

Malagnac, F., Bartee, L., and Bender, J. (2002). An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. EMBO J *21*, 6842-6852.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D.*, et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

Searle, I.R., Pontes, O., Melnyk, C.W., Smith, L.M., and Baulcombe, D.C. (2010). JMJ14, a JmjC domain protein, is required for RNA silencing and cell-to-cell movement of an RNA silencing signal in Arabidopsis. Genes Dev *24*, 986-991.

Seward, D.J., Cubberley, G., Kim, S., Schonewald, M., Zhang, L., Tripet, B., and Bentley, D.L. (2007). Demethylation of trimethylated histone H3 Lys4 in vivo by JARID1 JmjC proteins. Nat Struct Mol Biol *14*, 240-242.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol Cell *6*, 791-802.

Yang, W., Jiang, D., Jiang, J., and He, Y. (2010). A plant-specific histone H3 lysine 4 demethylase represses the floral transition in Arabidopsis. Plant J *62*, 663-673.

Zhang, X., Bernatavichute, Y.V., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol *10*, R62.

CHAPTER 6

Interplay Between Active Chromatin Marks and RNA-directed DNA Methylation in *Arabidopsis thaliana*.

INTRODUCTION

Cytosine DNA methylation is an epigenetic mark that is conserved across all kingdoms of eukaryotes. Depending on its location in the genome, DNA methylation can be partitioned into two types: gene-body and non-genic. Gene-body methylation has been observed in several species of plants and animals, and generally correlates with transcriptionally active loci (Feng et al., 2010; Zemach et al., 2010; Zhang et al., 2006). Conversely, non-genic methylation is associated with transcriptional repression at repetitive elements such as transposons (Law and Jacobsen, 2010). Both plants and animals also have instances of non-genic DNA methylation repressing protein-coding gene transcription when the mark is present in the gene's regulatory regions (Bird, 2002; Chan et al., 2005).

In the model plant *Arabidopsis thaliana*, gene-body methylation is maintained almost exclusively by METHYLTRANSFERASE 1 (MET1), the plant ortholog of mammalian DNA Methyltransferase 1 (DNMT1) (Cokus et al., 2008). Non-genic methylation is maintained by at least three methyltransferases, each with its own sequence preference: MET1, CHROMOMETHYLASE 3 (CMT3), and DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2). Much like its mammalian counterpart, MET1 primarily methylates cytosines in the CG dinucleotide context. CMT3 is a plant-specific DNA methyltransferase that has a preference for CHG sequences, where H is any base that is not a G (Cao and Jacobsen, 2002a). Finally, DRM2—the ortholog of mammalian DNA Methyltransferases 3A and 3B (DNMT3A/B)—maintains CHH, or asymmetric, methylation (Cao and Jacobsen, 2002a). It should be noted that there is a degree of redundancy for maintenance of non-CG methylation between CMT3 and DRM2. While all three methyltransferases maintain the methylcytosine mark across DNA replication events, only DRM2 establishes the mark on previously unmethylated loci (Cao and Jacobsen, 2002b).

In Arabidopsis, DNA methylation is correlated with specific histone marks depending on class. Gene-body DNA methylation, for example, is co-incidental with histone 3 lysine 4 monomethylation (H3K4m1) (Zhang et al., 2009). Conversely, non-genic DNA methylation is strongly enriched in regions of histone 3 lysine 9 dimethylation (H3K9m2) (Bernatavichute et al., 2008). Non-genic methylation is also inversely correlated with H3K4m2/3 and H2Bub—two marks that are associated with RNA Polymerase II (Pol II) transcription (Roudier et al., 2011; Zhang et al., 2009).

The nature of the respective DNA methyltransferase pathways gives some indication on the relationship between DNA methylation and histone modifications. DNMT1 is known to be recruited to hemimethylated CG sites through the SET and RING Finger Associated (SRA) domain-containing protein, Ubiquitin-like with PHD and RING Finger Domains 1 (UHRF1) (Bostick et al., 2007; Sharif et al., 2007). The UHRF1-DNMT1 complex is found at replication forks, thus therein lies a tidy mechanism to maintain CG methylation across the axis of symmetry following synthesis of the daughter strand DNA. In all likelihood, the MET1 protein in Arabidopsis operates in an identical manner as DNMT1 through the activity of the VARIANT IN METHYLATION (VIM) family proteins, which are orthologs of UHRF1 (Woo et al., 2008). Given that the DNA polymerase complex synthesizes DNA regardless of chromatin state, MET1 may be more resistant to regulation at the level of histone modifications. Therefore, it is not surprising that MET1 is active in both gene-body and non-genic loci. Conversely, CMT3 binds to H3K9m2 through its eponymous chromodomain (Lindroth et al., 2004). Null mutant lines for the H3K9m2 histone methyltransferases recapitulate the *cmt3* DNA methylation phenotype, which elucidates the tight correlation between the two marks (Cokus et al., 2008).

Chromatin regulation of the DRM2 pathway is not as clear. DRM2-dependent methylation depends on two plant specific RNA polymerases: RNA Polymerase IV and V (Pol IV and V). The former generates a transcript that is processed into 24 nucleotide small interfering RNAs (siRNAs), and the latter produces a transcript that serves a scaffold for ARGONAUTE 4

loaded siRNAs that are generated by Pol IV (Herr et al., 2005; Wierzbicki et al., 2009). This dual-RNA polymerase system targets DRM2 to DNA, although the specific mechanism for the targeting is not clear. Given the expanding understanding of the tremendous impact of histone marks on Pol II function, it is within reason that the RdDM pathway may be regulated at the chromatin level as well.

We previously showed along with an independent group that the H3K4m2/3 demethylase JMJ14 is required for complete DRM2 activity, but does not affect the MET1 or CMT3 pathways (Deleris et al., 2010; Searle et al., 2010). This correlates with a minor gain in H3K4 methylation. We concluded that H3K4m2/3 might negatively impact DRM2 maintenance activity. We were curious if the modest DNA methylation phenotype was due to redundant activity with other histone demethylases. Arabidopsis contains a family of H3K4 demethylases distinct from JUMONJI proteins known as LYSINE-SPECIFIC DEMETHYLASE 1-LIKE (LDL). In this paper we show by both single locus and whole genome analyses that two partially redundant members of the LDL family, *ldl1* and *ldl2*, exhibit a DNA methylation phenotype very similar to *jmj14*, which is further enhanced in the triple mutant. At the genome wide level, the decrease in DNA methylation correlates with an increase in both di- and tri-H3K4 methylation.

We also were interested in the relationship between H3K4m2/3 and H2Bub—histone marks that are co-incidental in the Arabidopsis genome. Moreover, an H2B deubiquitinase, UBIQUITIN-SPECIFIC PROTEASE 26 (UBP26), is already known to have a DRM2 methylation phenotype (Sridhar et al., 2007). Interestingly, global H3K4m2/3 levels appear to rise in *ubp26* mutants, suggesting a crosstalk between the marks (Sridhar et al., 2007). In a quadruple mutant generated among *jmj14 ldl1 ldl2 ubp26*, we observed a dramatic DNA methylation phenotype that is nearly as strong as *drm2* itself. The DNA methylation loss was in concordance with gains in H2Bub and H3K4m2/m3. This suggests that indeed combinatorial histone marks directly regulate RdDM.

We previously reported that *jmj14* mutants do not affect DRM2 establishment methylation (Deleris et al., 2010). Strikingly, even in the histone demethylase triple mutant, there is a minimal effect on DNA methylation establishment confirming our previous unexpected observations with jmj14. This suggests that either there are different chromatin requirements for establishment methylation, or there are fundamental differences of the nature of chromatin itself at the time of establishment. Altogether, data from this report help clarify the importance of chromatin in an RNA polymerase-dependent pathway.

RESULTS AND DISCUSSION

LDL1 and *LDL2* Impact DRM2-mediated DNA Methylation

We previously screened TDNA insertional mutant lines in genes containing JmjC putative histone demethylase domains in search of a DNA…with the hope of discovering a DNA methylation phenotype (Deleris et al., 2010). Indeed, we discovered *jmj14* mutants affect the DRM2 pathway, but not MET1 or CMT3. However, the DNA methylation phenotype is not as strong as the one observed in *drm2* mutants, and none of the other lines in the screen had an apparent phenotype. While it is possible that the effect of the *jmj14* mutation, and presumably H3K4m2/3, on DRM2 cannot further be enhanced, we were interested to know if JMJ14 had redundant activity with other histone demethylases.

We were intrigued by published reports of a separate family of H3K4 demethylases also impact DNA methylation (Jiang et al., 2007). Lysine Specific Demethylase 1 (LSD1) is a well-characterized H3K4 demethylase in mammals, which incidentally also regulates DNMT1 at the protein level through its demethylase activity (Shi et al., 2004; Wang et al., 2009). LSD1 contains four Arabidopsis homologs termed LDL1, LDL2, LDL3 and FLOWERING LOCUS D (FLD) (Jiang et al., 2007). The report from Jiang et al. described LDL1 and LDL2 as partially

redundant H3K4 demethylases that regulate the homeodomain transcription factor *FLOWERING WAGENINGEN* (*FWA*). Most interestingly to us, Jiang et al. reported a MET1-dependent CG-methylation phenotype at *FWA* in *ldl1-2 ldl2* mutants. We observed that *jmj14-1* mutants have reduced DNA methylation at *FWA* and an increase in H3K4 methylation at that locus within the same range reported in *ldl1-2 ldl2* double mutants (Deleris et al., 2010; Jiang et al., 2007). We were interested in a possible genetic interaction between the two families of demethylases, and generated a triple *jmj14-1 ldl1-2 ldl2* line.

In conducting our analyses, we observed some clear discrepancies between our data and the data reported by Jiang et al. Consistent with their results, we did observe an increase in *FWA* expression and H3K4m2/3, however we did not observe any effect on CG methylation by bisulfite analysis (Figures 6-1 A, 6-2 A, and 6-3). In fact, we see a reduction in non-CG methylation that is much more similar to what is observed in *jmj14-1* (Figure 6-1 A). In analysis of other RdDM targets, we observed the same phenomenon. At the *MEDEA-INTERGENIC SUBTELOMERIC REPEATS* (*MEA-ISR*), there is no reduction in MET1-dependent CG methylation, a decrease in non-CG contexts, once again, similar to the effect observed in *jmj14-1* (Figure 6-1 B). We also analyzed the *AtSN1* transposon using a quantitative PCR (qPCR) based assay in which we digest genomic DNA with the *HaeIII* endonuclease that is sensitive to CHH methylation at three restriction sites within the amplicon (Figure 6-1 C). Consistently, we see increased digestion in the *ldl1-2 ldl2* double mutant. Finally, we analyzed the *Ta3* locus by bisulfite sequencing (Figure 6-1 D). *Ta3* is methylated by MET1 and CMT3, but not DRM2 (Cao and Jacobsen, 2002a); as with *jmj14-1*, the *ldl1-2 ldl2* double mutant has no impact on the methylation state. Altogether, the data implies that in fact LDL1 and LDL2 impact the DRM2 pathway exclusively.

In the triple *jmj14-1 ldl1-2 ldl2* mutant plants, we do not observe an enhancement of the DNA methylation phenotype (Figures 6-1 A, B, and C). This raises the possibility that the two families of demethylases act in an epistatic manner. Alternatively, both demethylases could

target these loci independently, and the H3K4m2/3 state cannot further negatively impact DRM2-mediated DNA methylation, despite the absence of all three proteins.

ChIP Analysis at RdDM Targets

In order to further understand the relationship between H3K4 methylation and DNA methylation at RdDM targets, we performed chromatin immunoprecipitation (ChIP) against histone marks in variety of histone methylation mutants (Figure 6-2). At the loci analyzed, we observe only moderate gains in H3K4m2/3. Interestingly, even in the triple mutant, this effect does not appear to be greatly enhanced, or enhanced at all (Figure 6-2). Given that in all three cases, the H3K4 methylation increase is very moderate, we posit that only slight increases in the H3K4m2 or H3K4m3 state are enough to perturb the activity of DRM2. It also appears that there is partial redundancy at all three loci between JMJ14 and LDL1/LDL2, but in the triple mutant this redundancy does not significantly impact the DNA methylation phenotype.

Particularly interesting is the case of *FWA*, where expression is de-repressed in the *ldl1-2 ldl2* mutant background, whereas it does not appear to be affected at all in *jmj14-1* (Figure 6-3). However, their DNA methylation phenotypes are virtually identical (Figure 6-1 A). Neither the expression nor the DNA methylation phenotypes are aggravated in the triple mutant background. From this data, we can glean that LDL1/LDL2 are in a regulatory complex that is probably distinct from RdDM, at least at this locus, as canonical RdDM mutants do not de-repress *FWA* (Cao and Jacobsen, 2002b). JMJ14 may be targeted along with RdDM machinery, which would explain the modest gain in H3K4m2/m3 in *jmj14-1* mutants. One hypothesis for the DNA methylation phenotypes observed in H3K4 demethylase mutants was that Pol II transcription—as a result of increased H3K4m2/3—impacted RdDM, not the histone mark *per se*. However, at *FWA* JMJ14 appears to affect H3K4 methylation, but not expression. Therefore, it is more likely the chromatin itself that affects DRM2-mediated DNA methylation.

Genome-wide Expression and H3K4m2/3 ChIP Analysis

In order to gain a better understanding of the genetic interaction between the histone demethylases, we performed genome-wide mRNA sequencing using Illumina GAIIx technology (mRNA-seq). In the case of both families of histone demethylases, at the RNA level they regulate a broad range of targets (Figure 6-4 A). There is a sizeable overlap between *jmj14-1* and *ldl1-2 ldl2* mutant targets, indicating that both families likely regulate many of the same targets at the transcriptional level. Interestingly, there is a decrease in the total number of upregulated genes in the *jmj14-1 ldl1-2 ldl2* triple mutant. Other studies have established that *jmj14* and *ldl1 ldl2* have the opposite phenotype in regards to flowering time: JMJ14 promotes late flowering through repression of *FLOWERING LOCUS T* (*FT*) (Jeong et al., 2009; Lu et al., 2010; Yang et al., 2010), whereas LDL1 and LDL2 promote early flowering through repression of *FWA* and *FLOWERING LOCUS C* (*FLC*) (Jiang et al., 2007; Krichevsky et al., 2007). Therefore, in at least one developmental regulatory process, JMJ14 and LDL1/LDL2 are functionally antagonistic. The reduced number of upregulated genes in the triple mutant may reflect antagonism in flowering-time genes, and potentially as well in other distinct pathways.

In order to better understand the role of the H3K4 methylation state in the histone demethylase mutants, we performed H3K4m2 and H3K4m3 ChIP followed by Illumina sequencing (ChIP-seq). In focusing our analysis on the pool of genes that are upregulated in *jmj14-1 ldl1-2 ldl2* triple mutants, which are likely regulated by both families of demethylases, a distinct patter emerges (Figures 6-4 B and C). While it appears that JMJ14 affects both H3K4m2 and H3K4m3, LDL1/LDL2 seem to only have an impact on H3K4m2 (Figure 6-4 B). Mammalian LSD1 is known to have a preference for mono-and dimethylated H3K4 demethylation (Shi et al., 2004), therefore perhaps the genome-wide ChIP data suggests a similar preference for the Arabidopsis homologs. Conversely, in this same set of upregulated genes, there appears to be

a decrease in H3K4m3 in the *ldl1-2 ldl2* mutants compared to all genes in the same background (Figure 6-4 C). It is somewhat surprising that there is an observable decrease in H3K4m3 in a set of derepressed genes, but this may speak to the above notion that LDL1/LDL2's primary function in the genome is removing H3K4m2.

It should be noted that the majority of upregulated genes in all three mutant backgrounds are not targets of DNA methylation. In fact, the transcriptional profile observed for histone demethylase mutants is much more dramatic than even strong RdDM mutants, such as *drm2* (Chapter 4). Both JMJ14 and LDL1/LDL2 have wide-ranging and pleiotropic effects on a diverse set of targets, while the contribution on maintaining non-CG methylation is comparatively minor (discussed below). However, we do believe that genome-wide there is a degree of redundancy between the two families o histone demethylases. At genes upregulated in the *jmj14-1 ldl2-1 ldl2* triple mutant, we see increases in both H3K4m2 and H3K4m3 (Figures 6-4 B and C).

Genome-wide Bisulfite Analysis

In order to ascertain a global perspective on the impact on DNA methylation in the various histone demethylase mutants we performed shotgun bisulfite sequencing (BS-seq). As a control, we sequenced the *drm2-2* mutant line in parallel in order to draw a direct comparison at DRM2 targets (Figure 6-5 A). In differentially methylated regions (DMRs) defined in *drm2-2*, there is a decrease in non-CG DNA methylation in all classes of mutants, while no observable effect in CG methylation. The phenotype is not as strong as in *drm2-2*, which is consistent with analysis at canonical RdDM loci (Figures 6-1 A, B, and C). Moreover, as with the individual loci analyzed, we did not see an enhancement of the DNA methylation phenotype in the *jmj14-1 ldl1-2 ldl2* triple mutant.

We next wanted to assess whether the there was an effect at H3K4m2/m3 at the *drm2-2* DMRs in the various histone demethylase mutants, as this would give strong affirmation that

indeed the increased H3K4m2/m3 most likely results in the DNA methylation effect (Figures 6-5 B and C). We see the most dramatic effect in the triple mutant, with a significant gain in both H3K4m2 and H3K4m3 at the DMRs based on our ChIP-Seq data. And consistent with the hypothesis that LDL1/LDL2 has a greater impact on H3K4m2, we see a significant increase in that mark at DMRs in the *ldl1-2 ldl2* background (Figure 6-5 B). Interestingly, only a very minor increase in both H3K4m3 is observed in the *jmj14-1* mutant background, and no increase H3K4m2. This may be due to a minimal effect that becomes detectable only when amplified in the triple mutant. Furthermore, once the H3K4m2/m3 enhancement occurs genome-wide in the triple mutant, it is not followed by an aggravated loss of non-CG methylation (Figure 6-5 A). Once again, this is consistent with our analyses at individual loci (Figures 6-1 A, B, C, and 6-2).

Browsing the genome, there are certain loci that appear to be especially susceptible to mutations in histone demethylase mutants. For example, a Ty1/Copia element at locus *AT5G35935* shows a dramatic loss of methylation in all three sequence contexts in *jmj14-1*, *ldl1-2 ldl2*, and the triple mutant (Figure 6-6 A). We performed a PCR based assay with the methylation sensitive enzyme *MspI* to confirm the DNA methylation defect (Figure 6-6 B). The loss in DNA methylation corresponds to a gain in H3K4 methylation (Figure 6-6 C)—much greater than the gains observed the well studied RdDM loci assayed in Figure 6-2. Interestingly, there is apparently an equal gain in both *jmj14-1* and *ldl1-2 ldl2*, suggesting that in this case both are required for maintaining the low H3K4m2/m3 state of this locus. The effect is not further enhanced in the triple mutant. This transposon might be particularly susceptible to epigenetic perturbations, as previously it has been reported to acquire a significant gain of DNA methylation in DNA glycosylase mutants (Penterman et al., 2007). It also demonstrates that above a certain threshold, chromatin can strongly affect non-CG DNA methylation.

ubp26 Mutation exhibits similar phenotype as histone demethylases

In our analysis of the relationship between H3K4m2/3 and DNA methylation, we were intrigued about the potential relationship with other histone marks. In particular, it is known that the H2B deubiquitinase UBP26 is also required for RdDM (Sridhar et al., 2007). In *S. pombe*, the relationship between H2B ubiquitination and H3K4 methylation has been well studied. Specifically, the histone 2B lysine 123 deubiquitinase Ubp10 is required to maintain telomere silencing, and *ubp10*Δ deletion cells have increased steady-state H3K4m3 levels (Emre et al., 2005; Gardner et al., 2005). Moreover, increased global H3K4m2/m3 levels were reported in the Arabidopsis *ubp26* mutant (Sridhar et al., 2007). Therefore, in order to examine the possible genetic relationship between UBP26 and the histone demethylases, we generated a quadruple mutant.

While *ubp26-2* mutants have a similar effect on RdDM targets as *jmj14-1* and *ldl1-2 ldl2*, in the *jmj14-1 ldl1-2 ldl2 ubp26-2* background we observed a dramatic loss in non-CG methylation (Figures 6-7 A, B, and C). Once again, the effect of chromatin does not seem to effect methylation outside of the context of RdDM, as CG methylation remains at wild-type levels (Figures 6-7 A and B). Interested in how the DNA methylation phenotype related with chromatin, we performed H3K4m2, H3K4m3, and H2Bub ChIP in the various mutant backgrounds. We see an apparent increase in H2ub levels at most loci in the various histone demethylase backgrounds (Figure 6-7 D). Interestingly, in the *ubp26-2* mutant, there does not appear to be any impact on H3K4m2, and only a slight impact on H3K4m3 (Figures 6-7 E and F). UBP26 affects a number of processes, as exhibited by the pleiotropic phenotype in the mutant plants (Luo et al., 2008; Schmitz et al., 2009; Sridhar et al., 2007), therefore the global increase in H3K4m2/m3 may be unrelated to the effect of UBP26 at sites of RdDM.

The H3K4 methylation levels in the quadruple mutant are elevated in the quadruple mutant (Figures 6-7 E and F). Surprisingly, there does not seem to be a major increase in H2Bub in the quadruple, however (Figure 6-7 D). It is possible the increased H3K4 methylation

the cause of the severe DNA methylation phenotype. Otherwise, perhaps there are other histone marks affected that impact RdDM machinery.

Chromatin effectors do not impact DNA methylation establishment

Prior to our initial study describing JMJ14, all RdDM effectors tested were shown to be required for establishment of DNA methylation (Greenberg et al., 2011). In order to examine the requirements of DNA methylation establishment we take advantage of a transgenic version of the *FWA* gene. *FWA* is a homeodomain transcription factor with tandem repeats in its 5' UTR. In unmethylated epialleles, the *FWA* gene expresses ectopically, and the resulting phenotype is delay in flowering time (Soppe et al., 2000). Unmethylated transgenes introduced into wild-type plants are recognized by the RdDM machinery, and methylated (Cao and Jacobsen, 2002b; Chan et al., 2004). However, in mutants that are unable to establish DNA methylation, transgenic *FWA* expression leads to late flowering.

When we performed the *FWA* transformation assay on the *jmj14-1* mutant, we were surprised to discover that flowering time and DNA methylation establishment were not affected (Deleris et al., 2010). We were curious if this was due to activity of another H3K4 demethylase at the time of DNA methylation establishment, which probably is early in zygotic development (Chan et al., 2006). Otherwise, it is possible that the nature of chromatin itself at DNA methylation establishment is such that histone demethylases are not required. Therefore we transformed *ldl1-2 ldl2* and *jmj14-1 ldl1-2 ldl2* with *FWA* and scored for flowering time (Figure 6-8 A). Each untransformed mutant line has a flowering time that phenotype (Jeong et al., 2009; Jiang et al., 2007; Lu et al., 2010; Yang et al., 2010), and the flowering time subsequent to transformation is comparable to the change sin Col-0—most likely due to some relatively minor *FWA* transgene expression. A clearer picture emerges upon examination of the methylation state of the transgene at the molecular level (Figure 6-8 B). The CG methylation looks

unaffected in *jmj14* and *ldl1-2 ldl2* mutants, and only slightly reduced in the *jmj14-1 ldl1-2 ldl2* triple mutant. This suggests that perhaps that the H3K4 methylation state at some level is important for DNA methylation establishment. However, the effect is much less dramatic than for *dicer-like 3-1* (*dcl3-1*), which has a similar RdDM maintenance phenotype (Henderson et al., 2006). The non-CG methylation defect observed across the mutant lines is in all probability due to a maintenance methylation phenotype that occurs after establishment has occurred.

Given that the histone methylation proteins are chromatin effectors, we were interested to know if *ubp26* impacts DNA methylation establishment. Based on flowering time and bisulfite analysis, *ubp26-2* virtually mimics what was observed in *jmj14-1* and *ldl1-2 ldl2* (Figures 6-8 A and B). This solidifies the hypothesis that the modifications on histones do not seem to impede RdDM targeting sequences for methylation. One possible explanation is that the chromatin state is less dense, or more permissible at the time of DNA methylation establishment. Or perhaps, the histones at the time of establishment are not modified themselves, therefore do not affect DRM2 access to the DNA sequence. The data presented here does bolster the idea that regulation of the RdDM pathway by chromatin can be mechanistically distinguished between the maintenance and establishment phases.

CONCLUSION

In this study we have described four different histone modifiers, and the relationship between chromatin state and DRM2-mediated DNA methylation. Genome-wide ChIP and bisulfite analysis show that H3K4m2/m3 and H2Bub synergistically antagonize the RdDM pathway. Interestingly, the establishment of DNA methylation is much less sensitive than the maintenance methylation carried out by the DRM2 pathway. Given that all of the proteins described in this study have wide-ranging, and pleiotropic effects, perhaps instead of core components of the RdDM pathway, their effect—while direct—is more ancillary. In any event it raises interesting

questions about which RdDM protein(s) are being negatively impacted. In mammals, Dnmt3 specifically binds to unmodified H3K4 (Otani et al., 2009). Further tests must be carried out if indeed the analogous system takes place in Arabidopsis with DRM2 being having a similar relationship with histone tails.

Material and Methods

Plant material

All plants utilized in this study are in the Col-0 ecotype, and grown under long day conditions. The following mutant lines were used: *jmj14-1* (SALK_135712), *ldl1-2* (SALK_034869), *ldl2* (SALK_135831), *ubp26-2* (SALK_024392), and *drm2-2* (SALK_150863).

Bisulfite sequencing and analysis

For sodium bisulfite sequencing, DNA was treated using the EZ DNA Methylation Gold kit (Zymo Research) by following the manufacturer's instructions. Amplified PCR fragments from each analyzed locus were cloned into pCR2.1-TOPO (Invitrogen) and sequenced. We analyzed 15 to 22 clone sequences per sample using Lasergene SeqMan software. In order to distinguish the *FWA* transgene from the endogene, we destroyed a *BglII* restriction site in the transgenic copy in the region of PCR amplification. We then bisulfite treated genomic DNA of transgenic plants following a *BglII* digestion (37 degrees, overnight), which prevented amplification of the endogenous gene. Additionally, the transgenic copy of *FWA* was derived from the Landsberg ecotype, thus we could distinguish between the transgene and endogene based on the existence of three single nucleotide polymorphisms within the amplicon in case *BglII* digestion was not complete. Primers used for amplification are listed in Table 6-1.

qPCR-Chop assays

Analysis of asymmetric methylation at the *AtSN1* locus was performed exactly as described in (Deleris et al., 2010). Primers used for amplification are listed in Table 6-1. Analysis of non-CG methylation at *AT5G35935* was performed by extracting DNA from young flowers using a standard Cetyl trimethyl ammonium bromide protocol. A total of 200ng of genomic DNA was digested overnight at 371C with *MspI* side-by-side with samples containing buffer and no enzyme (undigested). Quantitative real-time PCR validation of uncut DNA after *MspI* digestion was performed using the Bio-Rad Synergy Brands Green SuperMix on a MX3000 Stratagene cycler. The PCR parameters are as follows: one cycle of 10min at 951C, 40 cycles of 30s at 951C, 1min at 551C and 1 min at 72 1C. PCR primers sequences are listed in Table 6-1.

Chromatin immunoprecipitation

ChIP assays were performed as described in (Johnson et al., 2007) with modifications. For immunoprecipitation, the following antibodies were used: H3K4m2, Abcam AB32356; H3K4m3, Diagenode pAb-003-050; H2Bub, Medimabs MM-0029. Primers used for amplification of ChIP targets are listed in Table 6-1.

Genome-wide mRNA Sequencing

Total RNA was prepared using a TRIzol (Invitrogen) extraction from 0.5 grams of 3-week old plant aerial tissue. 4 µg of total RNA was then used to prepare libraries for Illumina sequencing, following the Illumina TruSeq RNA Sample Prep guidelines. Multiplexed samples were sequenced at 50bp length on an Illumina HiSeq 2000 instrument.

Genome-wide ChIP and library generation

ChIP was performed as described above. Libraries were generated as described in Stroud et al. (Stroud et al., 2012).

Shotgun Bisulfite Sequencing

Genomic DNA was extracted from one gram of 3-week old plant aerial tissue using a DNeasy Plant Maxi Kit (Qiagen). Libraries for bisulfite sequencing were generated and sequenced as described in (Feng et al., 2011), with the change that sequencing was carried out on an Illumina HiSeq 2000 instrument.

Data Analyses

Sequenced reads were base-called using the standard Illumina pipeline. For ChIP-seq and BS-seq libraries, only full 50nt reads were retained. For ChIP-seq reads were mapped to the Arabidopsis genome (TAIR8 – www.arabidopsis.org) with Bowtie (Langmead et al., 2009) allowing up to 2 mismatches and retaining only reads mapping uniquely to the genome. For the biological replicate of the ChIP-seq experiment and the ChIP-seq of DDR mutants, 50 million reads were subset from the initial approximate 200 million reads for further analysis in the interest of computational time. For BS-seq libraries, reads were mapped using the BSseeker wrapper for Bowtie (Chen et al., 2010). For ChIP-seq and BS-seq, identical reads were collapsed into one read, whereas for smRNA-seq identical reads were retained.

For methylation analysis, percent methylation was calculated as previously reported (Cokus et al., 2008), with only cytosines having at least 5X coverage in libraries included in any analysis. For all libraries the list of mRNA and transposons along with genomic coordinates were obtained from TAIR (TAIR8). For all analyses, only transposons greater than 100bp in length were used.

For mRNA-seq analysis, 50 bp sequences called by the Illumina pipeline, were mapped to the Arabidopsis genome (TAIR8) using Bowtie (Langmead et al., 2009). Only reads mapping uniquely to the genome with a maximum of 2 mismatches were used for further analysis. To quantify changes in gene expression, read counts over each Arabidopsis gene model were used to perform Fisher Exact Tests between genotypes. False discovery rates (FDR) were

estimated by applying a Benjamini-Hochberg adjustment to resulting Fisher p-values. All statistical analysis was conducted with the R environment.

Generation of transgenic plants

Transgenic plants were generated as described in (Clough and Bent, 1998).

Flowering time

We measured flowering time of plants as the total number of leaves (rosette and cauline leaves) developed by a plant at the time of flowering. Plants transformed with the *FWA* transgene were selected for by spraying with a 1:1000 dilution of Basta^{TM} soon after germination.

Figure Legends

Figure 6-1. DNA methylation of RdDM targets in histone demethylase mutants.

(A) Bisulfite analysis of the *FWA* endogene.

(B) Bisulfite analysis of the *MEA-ISR* tandem repeats.

(C) *AtSN1* Chop-qPCR. Genomic DNA was digested with *HaeIII*, which recognizes GGCC sites, but it sensitive to cytosine DNA methylated. In the region amplified, there are three *HaeIII* sites, all corresponding to asymmetric cytosine contexts. The signal is relative to undigested DNA for each genotype.

(D) Bisulfite analysis of the *Ta3* transposon.

Figure 6-2. H3K4m2/m3 ChIP analysis of RdDM targets in histone demethylase mutants.

(A) H3K4m2 and H3K4m3 ChIP at *FWA*.

(B) H3K4m2 and H3K4m3 ChIP at *MEA-ISR*.

(C) H3K4m2 and H3K4m3 ChIP at *AtSN1*.

Figure 6-3. Gel-based RT-PCR of *FWA* in various histone demethylase mutants.

Figure 6-4. Genome-wide expression and H3K4m2/m3 ChIP Analysis.

(A) Venn diagram depicting overlap of genes upregulated >2 fold compared to Col-0 control in various histone demethylase mutants.

(B) H3K4m2-seq analysis at genes that are upregulated >2 fold in the *jmj14-1 ldl1-2 ldl2* triple mutant. Analysis was performed in various histone demethylase mutants.

(C) H3K4m3-seq analysis at genes that are upregulated >2 fold in the *jmj14-1 ldl1-2 ldl2* triple mutant. Analysis was performed in various histone demethylase mutants.

Figure 6-5. Shotgun bisulfite and H3K4m2/3 effect at DMRs.

(A) Methylation analysis of various histone demethylase mutants centered on DMRs defined in *drm2-2* mutant. Top panel: CG methylation. Middle panel: CHG methylation. Bottom panel: CHH methylation.

(B) H3K4m2-seq analysis at *drm2-2* DMRs in various histone demethylase mutants.

(C) H3K4m3-seq analysis at *drm2-2* DMRs in various histone demethylase mutants.

Figure 6-6. DNA methylation and H3K4m2/m3 ChIP at *AT5G35935* locus.

(A) Integrated genome browser (IGB) images of non-CG DNA methylation (left panel) and H3K4m2/m3 (right panel) in Col-0 and various histone demethylase mutants.

(B) *AT5G35935* Chop-qPCR. Genomic DNA was digested with *MspI*, which recognizes CCGG sites, but it sensitive to methylation at the external cytosine. The signal is relative to undigested DNA for each genotype.

(C) H3K4m2 and H3K4m3 ChIP at *AT5G35935*.

Figure 6-7. *ubp26-2* DNA methylation and histone ChIP analysis.

(A) Bisulfite analysis of the *FWA* endogene.

(B) Bisulfite analysis of the *MEA-ISR* tandem repeats.

(C) *AtSN1* Chop-qPCR.

(D) H2Bub ChIP at RdDM targets in *ubp26-2* and various histone demethylase mutants.

(E) H3K4m2 ChIP at RdDM targets in *ubp26-2 and jmj14-1 ldl1-2 ldl2 ubp26-2*.

(F) H3K4m3 ChIP at RdDM targets in *ubp26-2 and jmj14-1 ldl1-2 ldl2 ubp26-2*.

Figure 6-8. *FWA* methylation establishment assays.

(A) *FWA* flowering-time analysis in chromatin effector mutants. Flowering-time is determined by the total number of rosette and cauline leaves when the first florescence appears. *FWA* transformed lines are compared to untransformed lines of the same genotype. The graph depicts averages from populations of >20 individual plants.

(B) Bisulfite analysis of the *FWA* transgene in various mutants.

Table 6-1. Primers and probes used in this study.

Figure 6-1

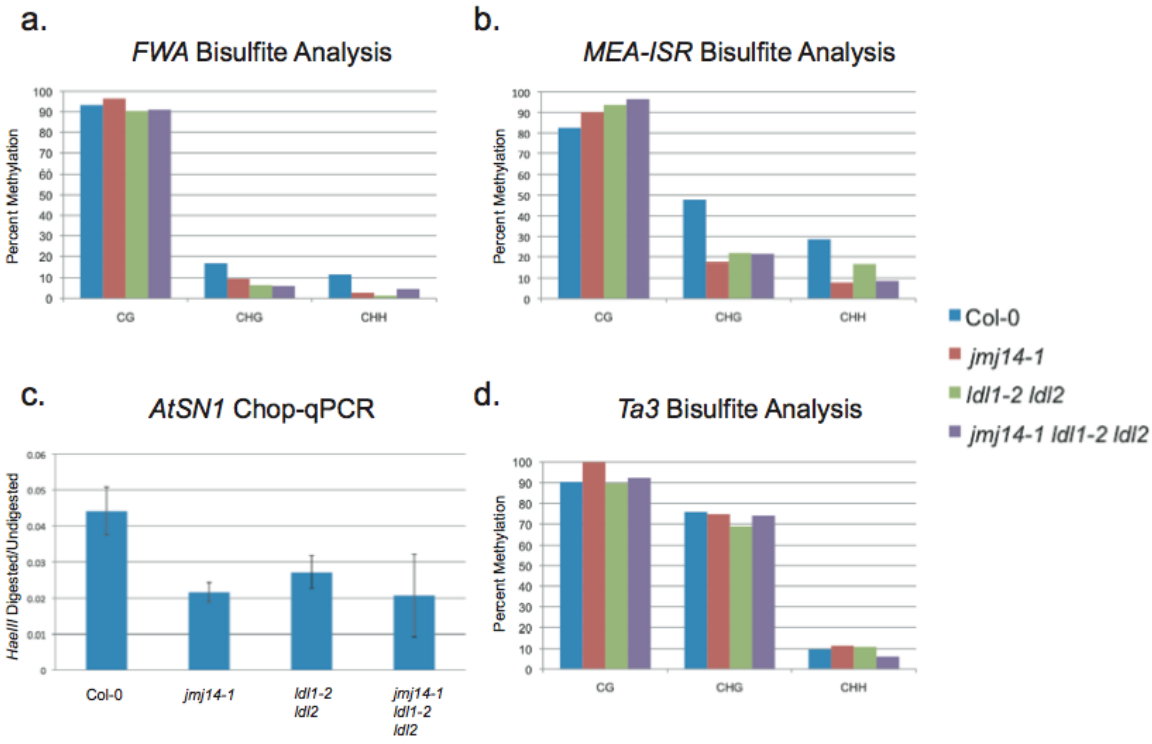

a.
FWA Bisulfite Analysis
Percent Methylation
CG
CHG
CHH
b.
MEA-ISR Bisulfite Analysis
Percent Methylation
CG
CHG
CHH
Col-0
jmj14-1
ldl1-2 ldl2
jmj14-1 ldl1-2 ldl2
c.
AtSN1 Chop-qPCR
HaeIII Digested/Undigested
Col-0
jmj14-1
ldl1-2 ldl2
jmj14-1 ldl1-2 ldl2
d.
Ta3 Bisulfite Analysis
Percent Methylation
CG
CHG
CHH

Figure 6-2

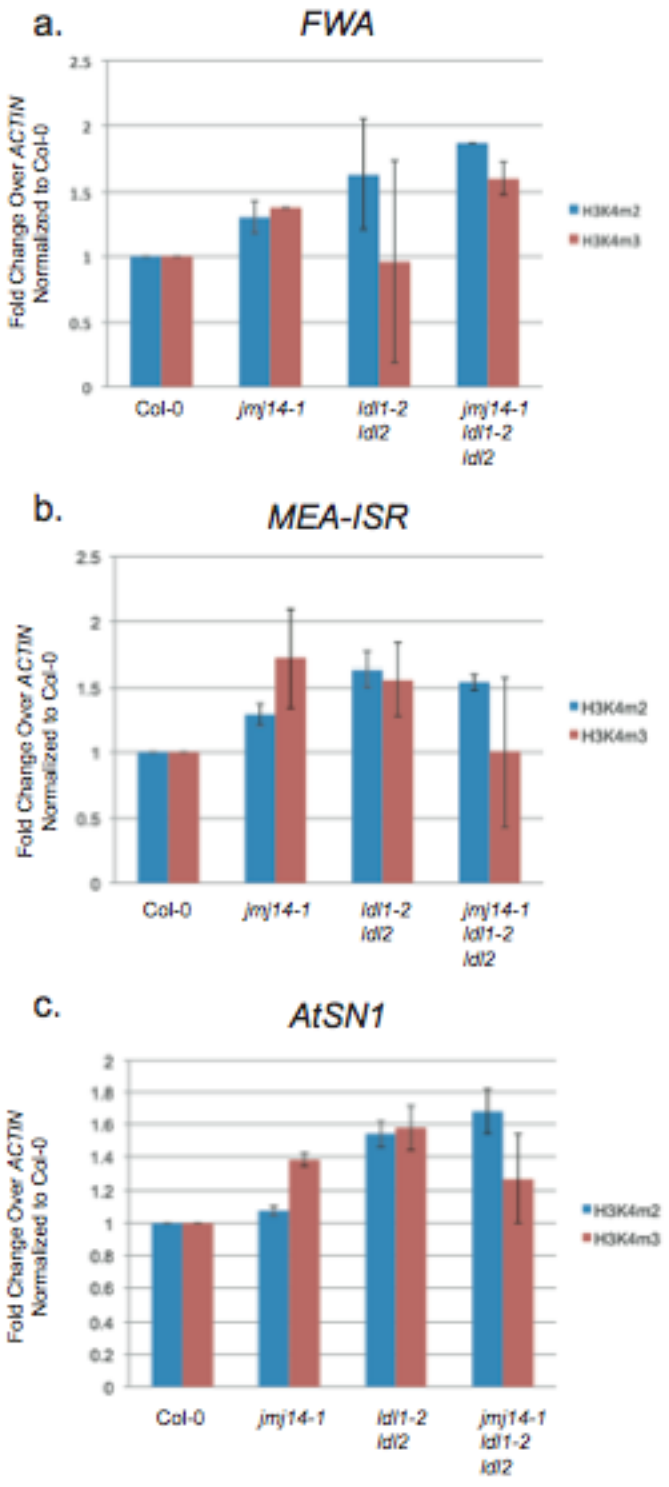

a.
FWA
Fold Change Over ACTIN Normalized to Col-0
2.5
2
1.5
1
0.5
0
Col-0
jmj14-1
idl1-2 idl2
jmj14-1 idl1-2 idl2
H3K4m2
H3K4m3
b.
MEA-ISR
Fold Change Over ACTIN Normalized to Col-0
2.5
2
1.5
1
0.5
0
Col-0
jmj14-1
idl1-2 idl2
jmj14-1 idl1-2 idl2
H3K4m2
H3K4m3
c.
AtSN1
Fold Change Over ACTIN Normalized to Col-0
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Col-0
jmj14-1
idl1-2 idl2
jmj14-1 idl1-2 idl2
H3K4m2
H3K4m3

Figure 6-3

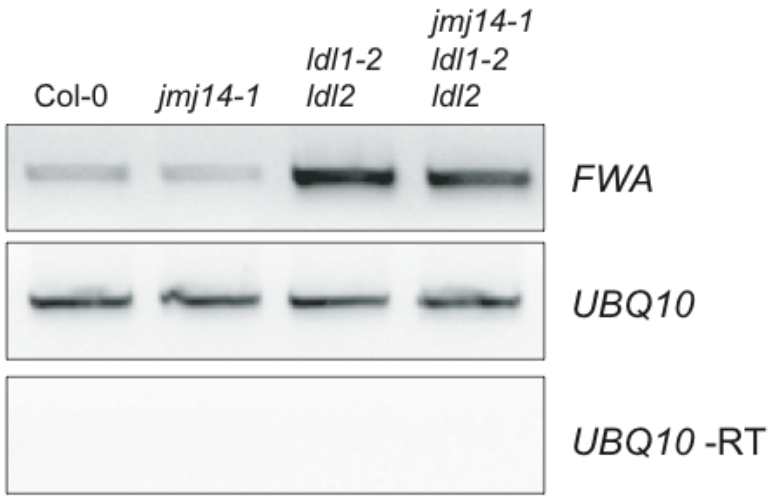


Figure 6-4

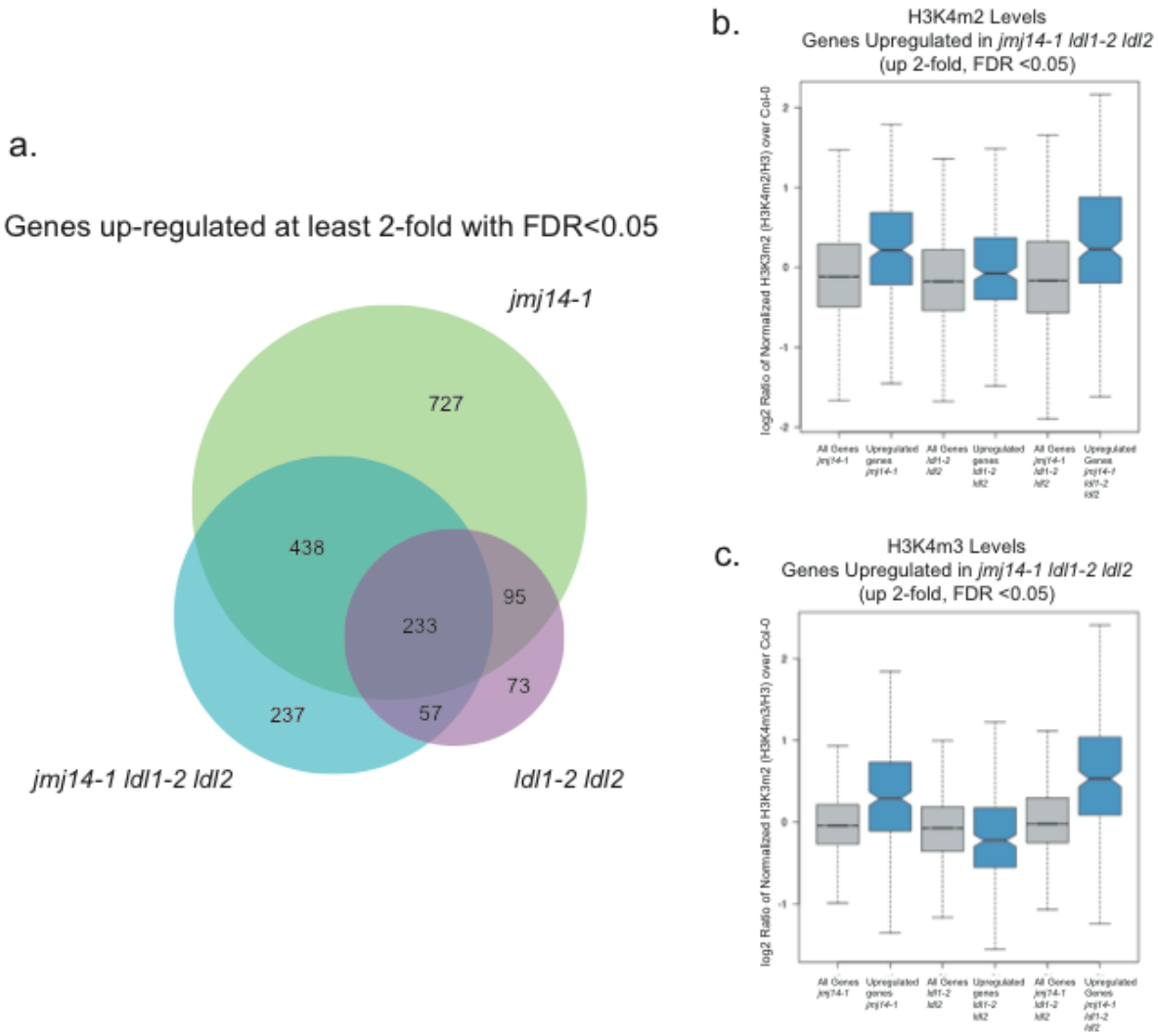

a.
Genes up-regulated at least 2-fold with FDR<0.05
jmj14-1
727
438
95
233
73
237
57
jmj14-1 ldl1-2 ldl2
ldl1-2 ldl2
b.
H3K4m2 Levels
Genes Upregulated in jmj14-1 ldl1-2 ldl2
(up 2-fold, FDR <0.05)
log2 Ratio of Normalized H3K3m2 (H3K4m2/H3) over Col-0
c.
H3K4m3 Levels
Genes Upregulated in jmj14-1 ldl1-2 ldl2
(up 2-fold, FDR <0.05)
log2 Ratio of Normalized H3K3m2 (H3K4m3/H3) over Col-0

Figure 6-5

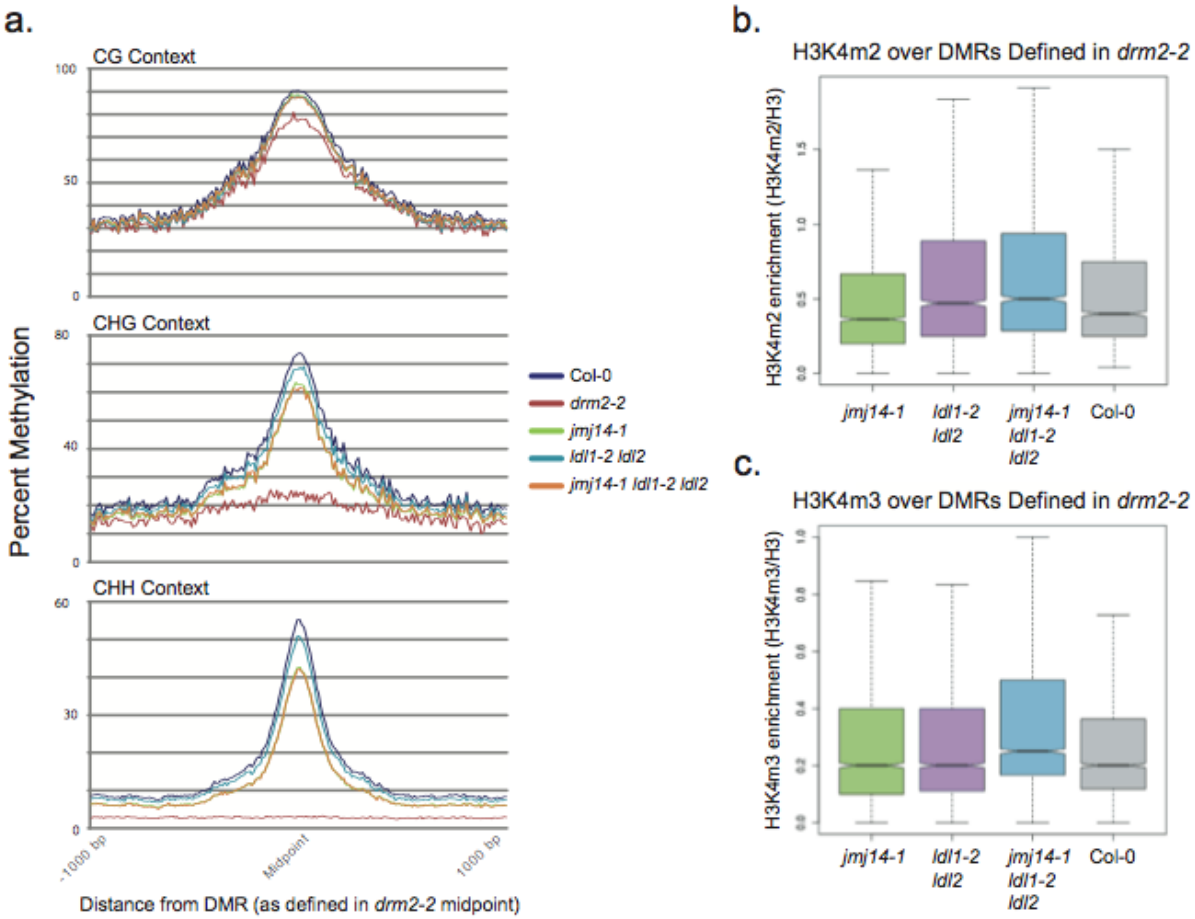

a.
CG Context
CHG Context
CHH Context
Percent Methylation
Col-0
drm2-2
jmj14-1
ldl1-2 ldl2
jmj14-1 ldl1-2 ldl2
-1000 bp
Midpoint
1000 bp
Distance from DMR (as defined in drm2-2 midpoint)
b.
H3K4m2 over DMRs Defined in drm2-2
H3K4m2 enrichment (H3K4m2/H3)
jmj14-1
ldl1-2 ldl2
jmj14-1 ldl1-2 ldl2
Col-0
c.
H3K4m3 over DMRs Defined in drm2-2
H3K4m3 enrichment (H3K4m3/H3)
jmj14-1
ldl1-2 ldl2
jmj14-1 ldl1-2 ldl2
Col-0

Figure 6-6

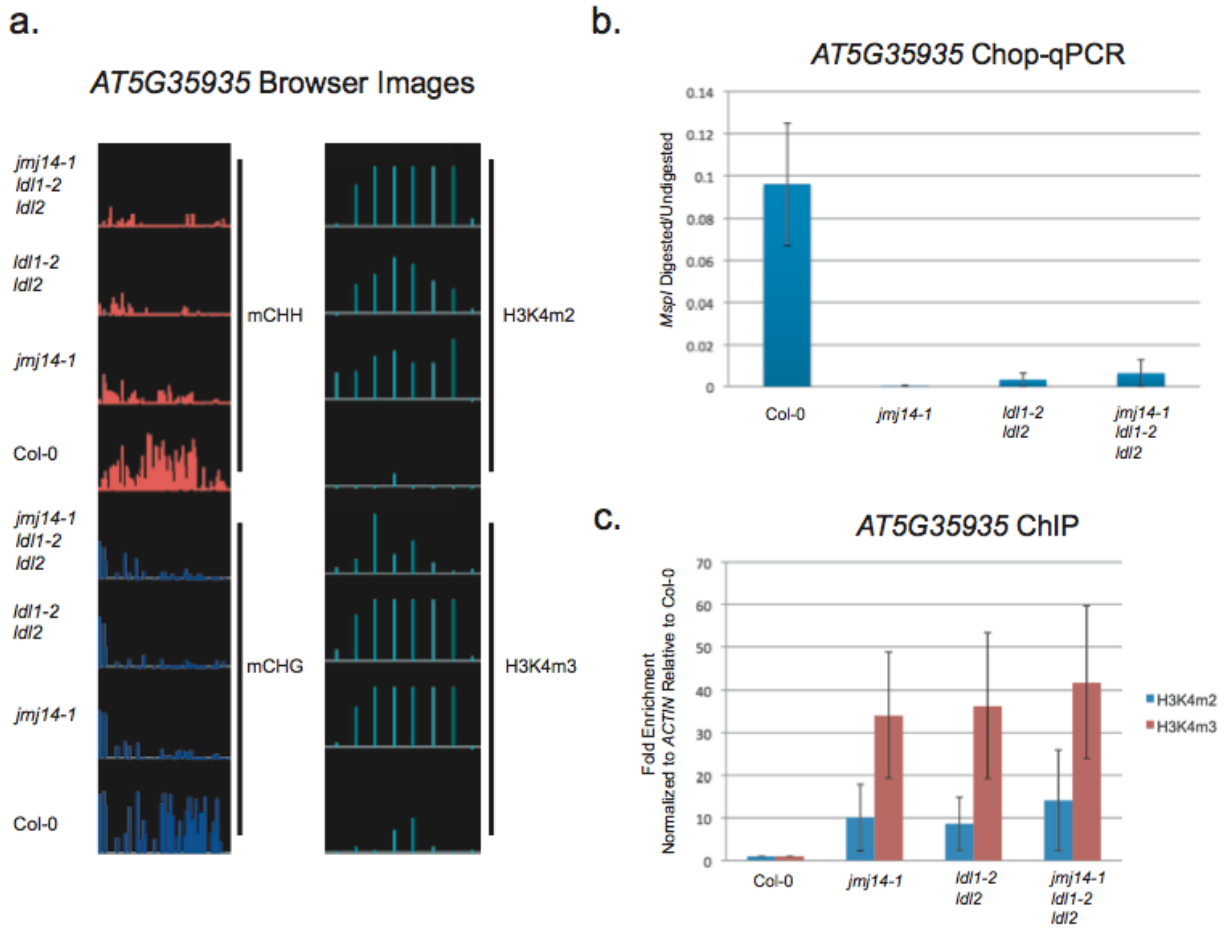

a.
AT5G35935 Browser Images
jmj14-1 ldl1-2 ldl2
ldl1-2 ldl2
jmj14-1
Col-0
mCHH
H3K4m2
jmj14-1 ldl1-2 ldl2
ldl1-2 ldl2
jmj14-1
Col-0
mCHG
H3K4m3
b.
AT5G35935 Chop-qPCR
MspI Digested/Undigested
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Col-0
jmj14-1
ldl1-2 ldl2
jmj14-1 ldl1-2 ldl2
c.
AT5G35935 ChIP
Fold Enrichment Normalized to ACTIN Relative to Col-0
70
60
50
40
30
20
10
0
Col-0
jmj14-1
ldl1-2 ldl2
jmj14-1 ldl1-2 ldl2
H3K4m2
H3K4m3

Figure 6-7

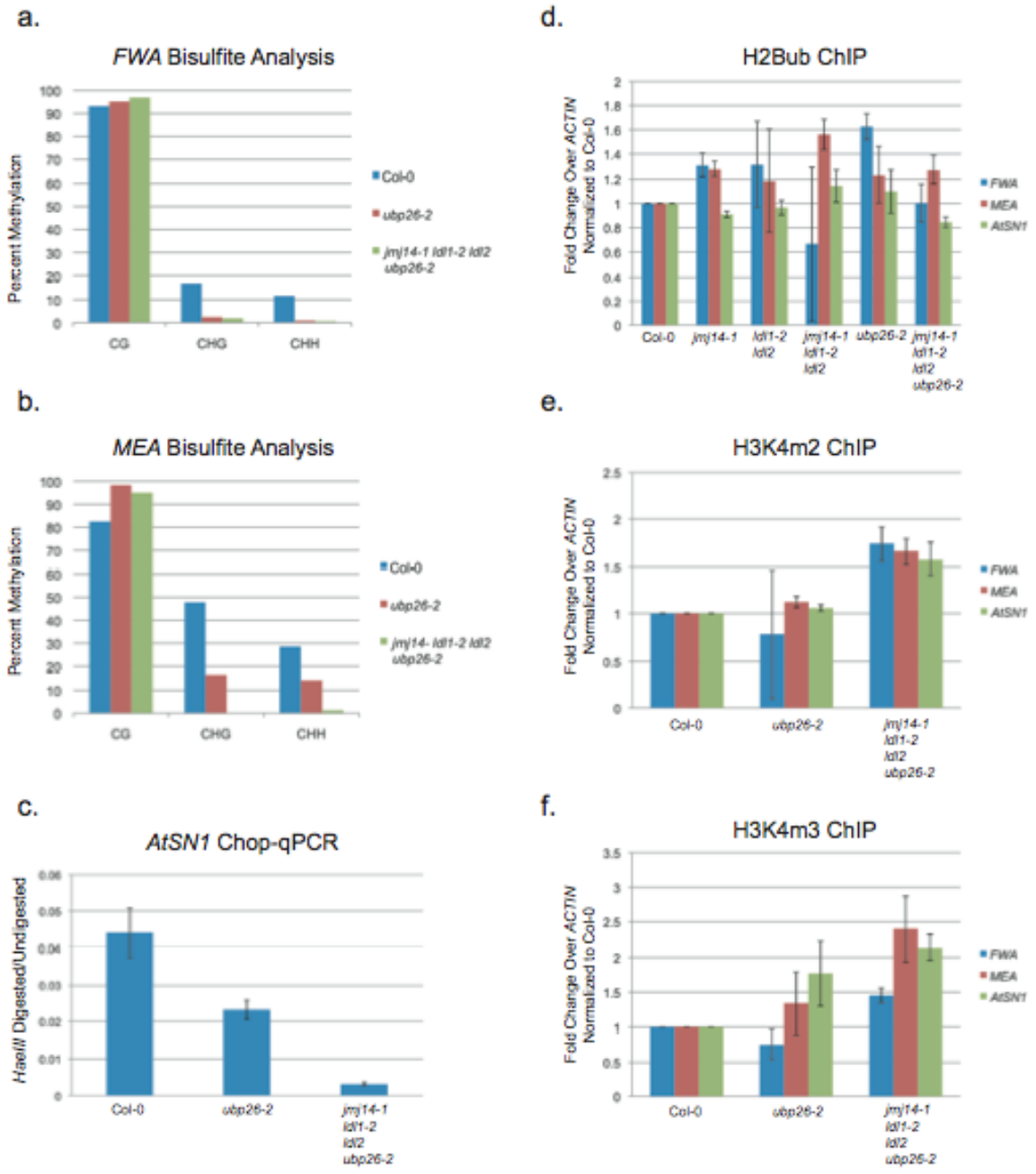

a.
FWA Bisulfite Analysis
Percent Methylation
CG
CHG
CHH
Col-0
ubp26-2
jmj14-1 ldl1-2 ldl2 ubp26-2
b.
MEA Bisulfite Analysis
Percent Methylation
CG
CHG
CHH
Col-0
ubp26-2
jmj14- ldl1-2 ldl2 ubp26-2
c.
AtSN1 Chop-qPCR
HaeIII Digested/Undigested
Col-0
ubp26-2
jmj14-1 ldl1-2 ldl2 ubp26-2
d.
H2Bub ChIP
Fold Change Over ACTIN Normalized to Col-0
Col-0
jmj14-1
ldl1-2 ldl2
jmj14-1 ldl1-2 ldl2
ubp26-2
jmj14-1 ldl1-2 ldl2 ubp26-2
FWA
MEA
AtSN1
e.
H3K4m2 ChIP
Fold Change Over ACTIN Normalized to Col-0
Col-0
ubp26-2
jmj14-1 ldl1-2 ldl2 ubp26-2
FWA
MEA
AtSN1
f.
H3K4m3 ChIP
Fold Change Over ACTIN Normalized to Col-0
Col-0
ubp26-2
jmj14-1 ldl1-2 ldl2 ubp26-2
FWA
MEA
AtSN1

Figure 6-8

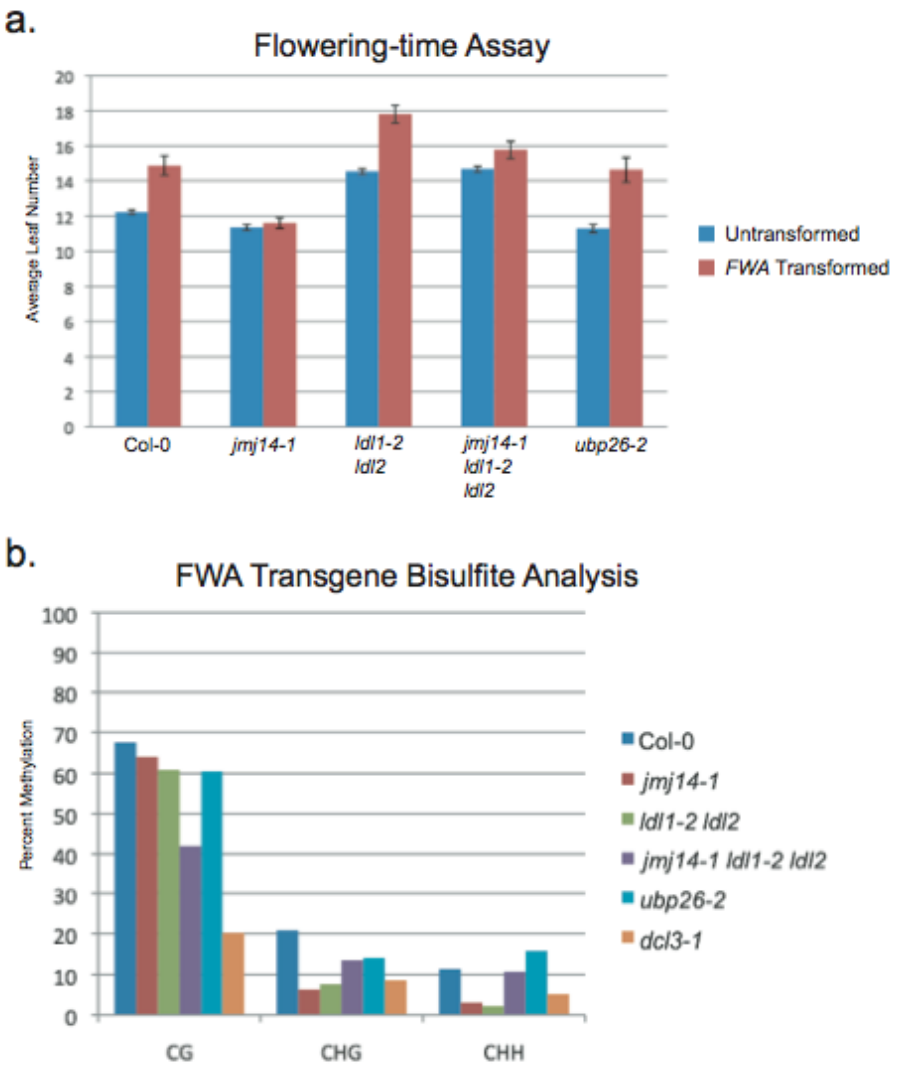

a.
Flowering-time Assay
Average Leaf Number
0
2
4
6
8
10
12
14
16
18
20
Col-0
jmj14-1
ldl1-2 ldl2
jmj14-1 ldl1-2 ldl2
ubp26-2
Untransformed
FWA Transformed
b.
FWA Transgene Bisulfite Analysis
Percent Methylation
0
10
20
30
40
50
60
70
80
90
100
CG
CHG
CHH
Col-0
jmj14-1
ldl1-2 ldl2
jmj14-1 ldl1-2 ldl2
ubp26-2
dcl3-1

Table 6-1

MEA-ISR Bisulfite	AAAGTGGTTGTAGTTTATGAAAGGTTTTAT
MEA-ISR Bisulfite	CTTAAAAAATTTTCAACTCATTTTTAAAAAA
FWA Bisulfite	GGTTTTATATTAATATTAAAGAGTTATGGGTYGAAGTTT
FWA Bisulfite	AACCAAAATCATTCTCTAAACAAAATATAAAAAAATC
Ta3 Bisulfite	GAGAATYAGGTTAATAAGAAAGTGAAGTGTT
Ta3 Bisulfite	CCACTRATTCCTRAAACACAACATTTCTRCTRATA
AtSN1 Chop-qPCR	ACTTAATTAGCACTCAAATTAAACAAAATAAGT
AtSN1 Chop-qPCR	TTTAAACATAAGAAGAAGTTCCTTTTTCATCTAC
AT5G35935 Chop-qPCR	CTAGGGTTGAAGTTCTGAAA
AT5G35935 Chop-qPCR	CGGCTCTTTACCTCAAACAT
MEA-ISR ChIP	TTTAGGTATTAGCTCGTTTGGTTTTA
MEA-ISR ChIP	TCCCGCCATTTAACCGTG
FWA ChIP	ATAAAGAGCGGCGCAAGAT
FWA ChIP	CGCTCTAGGGTTTTTGCTTT
FWA ChIP Probe	CAAATAGCACTTGGACCAATGGCG
AtSN1 ChIP	GTTGGCCCAGTGGTAAATCT
AtSN1 ChIP	TGGTGGTTGTACAAGCCTAGTT
AtSN1 ChIP Probe	ATCTCCCAGAGGCGGGACCC
ACTIN ChIP	CGCTGCTTCTCGAATCTTCT
ACTIN ChIP	AAGCACGGATCGAATCACAT
AT5G35935 ChIP	TGTCCTGCCCAACTGATGTCC
AT5G35935 ChIP	TGGATCTTGTGAACCTTGCTGC
FWA RT	TTAGATCCAAAGGAGTATCAAAG
FWA RT	CTTTGGTACCAGCGGAGA
UBQ10 RT	GATCTTTGCCGGAAAACAATTGGAGG
UBQ10 RT	CGACTTGTCATTAGAAAGAAAGAGAT

REFERENCES

Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One *3*, e3156.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev *16*, 6-21.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

Cao, X., and Jacobsen, S.E. (2002a). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A *99 Suppl 4*, 16491-16498.

Cao, X., and Jacobsen, S.E. (2002b). Role of the Arabidopsis DRM methyltransferases in *de novo* DNA methylation and gene silencing. Current Biology *12*, 1138-1144.

Chan, S.W., Henderson, I.R., and Jacobsen, S.E. (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet *6*, 351-360.

Chan, S.W., Zhang, X., Bernatavichute, Y.V., and Jacobsen, S.E. (2006). Two-step recruitment of RNA-directed DNA methylation to tandem repeats. PLoS Biol *4*, 1923-1933.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Chen, P.Y., Cokus, S.J., and Pellegrini, M. (2010). BS Seeker: precise mapping for bisulfite sequencing. BMC Bioinformatics *11*, 203.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J *16*, 735-743.

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

Deleris, A., Greenberg, M.V., Ausin, I., Law, R.W., Moissiard, G., Schubert, D., and Jacobsen, S.E. (2010). Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation. EMBO Rep *11*, 950-955.

Emre, N.C., Ingvarsdottir, K., Wyce, A., Wood, A., Krogan, N.J., Henry, K.W., Li, K., Marmorstein, R., Greenblatt, J.F., Shilatifard, A.*, et al.* (2005). Maintenance of low histone ubiquitylation by Ubp10 correlates with telomere-proximal Sir2 association and gene silencing. Mol Cell *17*, 585-594.

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

Feng, S., Rubbi, L., Jacobsen, S.E., and Pellegrini, M. (2011). Determining DNA methylation profiles using sequencing. Methods Mol Biol *733*, 223-238.

Gardner, R.G., Nelson, Z.W., and Gottschling, D.E. (2005). Ubp10/Dot4p regulates the persistence of ubiquitinated histone H2B: distinct roles in telomeric silencing and general chromatin. Mol Cell Biol *25*, 6123-6139.

Greenberg, M.V., Ausin, I., Chan, S.W., Cokus, S.J., Cuperus, J.T., Feng, S., Law, J.A., Chu, C., Pellegrini, M., Carrington, J.C.*, et al.* (2011). Identification of genes required for de novo DNA methylation in Arabidopsis. Epigenetics *6*.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat Genet *38*, 721-725.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science *308*, 118-120.

Jeong, J.H., Song, H.R., Ko, J.H., Jeong, Y.M., Kwon, Y.E., Seol, J.H., Amasino, R.M., Noh, B., and Noh, Y.S. (2009). Repression of FLOWERING LOCUS T chromatin by functionally redundant histone H3 lysine 4 demethylases in Arabidopsis. PLoS One *4*, e8033.

Jiang, D., Yang, W., He, Y., and Amasino, R.M. (2007). Arabidopsis relatives of the human lysine-specific Demethylase1 repress the expression of FWA and FLOWERING LOCUS C and thus promote the floral transition. Plant Cell *19*, 2975-2987.

Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I.R., Callis, J., and Jacobsen, S. (2007). The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation. Current Biology *17*, 379.

Krichevsky, A., Gutgarts, H., Kozlovsky, S.V., Tzfira, T., Sutton, A., Sternglanz, R., Mandel, G., and Citovsky, V. (2007). C2H2 zinc finger-SET histone methyltransferase is a plant-specific chromatin modifier. Dev Biol *303*, 259-269.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

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

Lindroth, A.M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I.*, et al.* (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. Embo J *23*, 4286-4296.

Lu, F., Cui, X., Zhang, S., Liu, C., and Cao, X. (2010). JMJ14 is an H3K4 demethylase regulating flowering time in Arabidopsis. Cell Res *20*, 387-390.

Luo, M., Luo, M.Z., Buzas, D., Finnegan, J., Helliwell, C., Dennis, E.S., Peacock, W.J., and Chaudhury, A. (2008). UBIQUITIN-SPECIFIC PROTEASE 26 is required for seed development and the repression of PHERES1 in Arabidopsis. Genetics *180*, 229-236.

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep *10*, 1235-1241.

Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007). DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci U S A *104*, 6752-6757.

Roudier, F., Ahmed, I., Berard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L.*, et al.* (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J *30*, 1928-1938.

Schmitz, R.J., Tamada, Y., Doyle, M.R., Zhang, X., and Amasino, R.M. (2009). Histone H2B deubiquitination is required for transcriptional activation of FLOWERING LOCUS C and for proper control of flowering in Arabidopsis. Plant Physiol *149*, 1196-1204.

Searle, I.R., Pontes, O., Melnyk, C.W., Smith, L.M., and Baulcombe, D.C. (2010). JMJ14, a JmjC domain protein, is required for RNA silencing and cell-to-cell movement of an RNA silencing signal in Arabidopsis. Genes Dev *24*, 986-991.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K.*, et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The Late Flowering Phenotype of fwa Mutants Is Caused by Gain-of-Function Epigenetic Alleles of a Homeodomain Gene. Mol Cell *6*, 791-802.

Sridhar, V.V., Kapoor, A., Zhang, K., Zhu, J., Zhou, T., Hasegawa, P.M., Bressan, R.A., and Zhu, J.K. (2007). Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. Nature *447*, 735-738.

Stroud, H., Otero, S., Desvoyes, B., Ramirez-Parra, E., Jacobsen, S.E., and Gutierrez, C. (2012). Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. Proc Natl Acad Sci U S A *109*, 5370-5375.

Wang, J., Hevi, S., Kurash, J.K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G.*, et al.* (2009). The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. Nat Genet *41*, 125-129.

Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet *41*, 630-634.

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

Yang, W., Jiang, D., Jiang, J., and He, Y. (2010). A plant-specific histone H3 lysine 4 demethylase represses the floral transition in Arabidopsis. Plant J *62*, 663-673.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916-919.

Zhang, X., Bernatavichute, Y.V., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol *10*, R62.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E.*, et al.* (2006). Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. Cell *126*, 1189-1201.

Chapter 7

Concluding Remarks

In this dissertation, I have described the methodologies I have employed to shed light on RNA-directed DNA methylation (RdDM) in *Arabidopsis thaliana.* In chapters 2 and 3 I describe how the *FWA* transgene system is a tool to discover mutants that are required for *de novo* DNA methylation. Both forward and reverse genetic screens were used to uncover novel mutations. In chapter 4, I describe utilize genetic, biochemical, and genomic tools in order to characterize a complex required for RdDM. Finally, in chapters 5 and 6, I try and find a link between chromatin state and an epigenetic process that is RNA polymerase dependent.

Over the course of my graduate career, a there have a flood of studies from numerous labs characterizing the RdDM pathway. It was an exciting time to be working in the field, and I am proud to have made contributions to understanding this complex RNA-driven system. I am happy to have partially assisted in the development of analytic techniques such as IP-Mass Spec and mapping-by-genomics, which are now becoming a powerful methods for discovery in Arabipopsis. I am also fortunate to have performed my graduate work in an era with relatively facile genomics, which provides a wealth of information of how these pathway components contribute genome-wide at the level of DNA methylation, histone modification, and expression.

A number of questions remain, such as how the RNA polymerases are targeted to sequences, and if post-translational regulation of RdDM pathway components contributes to pathway function. I believe that Arabidopsis is a powerful model for DNA methylation studies, and I look forward to the breakthroughs in the coming years.