12-2016

Dynamic regulation of DNA demethylation and RNA-directed DNA methylation in Arabidopsis

Kai Tang
Purdue University

DYNAMIC REGULATION OF DNA DEMETHYLATION AND RNA-DIRECTED DNA METHYLATION IN *ARABIDOPSIS*

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Kai Tang

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2016

Purdue University

West Lafayette, Indiana

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Kai Tang

Entitled

DYNAMIC REGULATION OF DNA DEMETHYLATION AND RNA-DIRECTED DNA METHYLATION IN ARABIDOPSIS

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Jian-Kang Zhu
Chair

Zhixiang Chen

Brian P. Dilkes

Kashchandra G. Raghothama

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Jian-Kang Zhu

Approved by: Hazel Y. Wetzstein 11/09/2016

Head of the Departmental Graduate Program Date

Dedicated to my parents: Tang Chunlin and Xu Jialing.

ACKNOWLEDGMENTS

I would like to thank my major advisor, Dr. Jian-Kang Zhu, whose enormous support and insightful comments are invaluable during the course of my Ph.D. career. I have learned how to think critically and independently like a scientist from Dr. Zhu. I also want to thank Dr. Renyi Liu, my former co-advisor in University of California, Riverside (UCR), who gave me beginning training as a bioinformatician.

I am grateful to my advisory committee: Dr. Zhixiang Chen, Dr. Brian P Dilkes, and Dr. Kashchandra G Raghothama, for their support and help.

I would like to thank Dr. Paul Michael Hasegawa for serving as the chair of my preliminary examination committee. I also would like to give thanks to professors in UCR who accepted me as rotation student: Dr. Leonard Nunney, Dr. Brian A Federici, and Dr. Jason E Stajich.

Let me express my sincere gratitude for my colleagues: Dr. Cheng-Guo Duan, Dr. Huiming Zhang, and Dr. Zhaobo Lang. Without their collaboration and help, this dissertation cannot be finished.

Many thanks go to Zhu lab members: Rebecca Ann Stevenson, Dr. Chao-Feng Huang, Dr. Heng Zhang, Dr. Dong-Lei Yang, Xingang Wang, Dr. Yueh-Ju Hou, Dr. Mingguang Lei, Dr. Pengcheng Wang, Dr. Yihai Wang, Dr. Shaojun Xie, Dr. Jun Yan, Dr. Cuijun Zhang, Dr. Hui Zhang, Dr. Chunzhao Zhao, Dr. Yang Zhao, Dr. Xiaohong Zhu, Dr. Yingfang Zhu, Russell S Julian, Omar Mohamed Zayed, Sam Shenyu Zhang, *etc.*

I would also like to express my gratitude to my family for their moral support and warm encouragement.

Finally, thank you Lu Xing for loving me, supporting me, and helping me.

Chapter 2

Chapter 2 is a version of the research article published in *Nature Plants* titled "The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications". This study was authored by Kai Tang, Zhaobo Lang, Heng Zhang, and Jian-Kang Zhu. This work was supported by National Institutes of Health Grant R01GM070795 and by the Chinese Academy of Sciences.

Chapter 3

Chapter 3 is a version of the research article published in *The EMBO Journal* titled "Specific but interdependent functions for *Arabidopsis* AGO4 and AGO6 in RNA-directed DNA methylation". The study was authored by Cheng-Guo Duan, Huiming Zhang, Kai Tang, Xiaohong Zhu, Weiqiang Qian, Yueh-Ju Hou, Bangshing Wang, Zhaobo Lang, Yang Zhao, Xingang Wang, Pengcheng Wang, Jianping Zhou, Gaimei Liang, Na Liu, Chunguo Wang, and Jian-Kang Zhu. We thank Dr. David Baulcombe for FLAG-tagged AGO4 and AGO6 transgenic plants, Dr. Craig Pikaard for anti-NRPE1 antibody. The work in Zhu lab is supported by grants from National Institutes of Health, USA, and by the Chinese Academy of Sciences, China.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

5mC	5-methylcytosine
AGO	ARGONAUTE
APE1L	APURINIC/APYRIMIDINIC ENDONUCLEASE 1-LIKE
AtLIG1	*Arabidopsis thaliana* DNA LIGASE 1
bp	base pair
ChIP	chromatin immunoprecipitation
CLSY1	CLASSY1
CMT3	CHROMOMETHYLASE 3
DCL3	DICER-LIKE 3
ddcmt2	*drm1/drm2/cmt2*
ddcmt3	*drm1/drm2/cmt3*
ddcc	*drm1/drm2/cmt2/cmt3*
DMCs	Differentially Methylated Cytosines
DME	DEMETER
DML2	DEMETER-LIKE 2
DMRs	Differentially Methylated Regions
DMS3	DEFECTIVE IN MERISTEM SILENCING 3
DRD1	DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 2
DTF1(SHH1)	DNA-BINDING TRANSCRIPTION FACTOR 1 (SAWADEE HOMEODOMAIN HOMOLOG 1)
EPF2	EPIDERMAL PATTERNING FACTOR 2
H3	Histone H3
H3K18Ac	Histone H3 lysine 18 acetylation

H3K27me1	Histone H3 lysine 27 monomethylation
H3K27me3	Histone H3 lysine 27 trimethylation
H3K36me2	Histone H3 lysine 36 dimethylation
H3K36me3	Histone H3 lysine 36 trimethylation
H3K4me2	Histone H3 lysine 4 dimethylation
H3K4me3	Histone H3 lysine 4 trimethylation
H3K9Ac	Histone H3 lysine 9 acetylation
H3K9me2	Histone H3 lysine 9 dimethylation
IDM1	INCREASED DNA METHYLATION 1
IG	intergenic
MBD7	METHYL-CPG-BINDING DOMAIN 7
MEMS	DNA MEthylation Monitoring Sequence
MET1	METHYLTRANSFERASE 1
NRPD1	NUCLEAR RNA POLYMERASE D 1
nt	nucleotide
Pol IV	RNA POLYMERASE IV
PTGS	Post-Transcriptional Gene Silencing
RDM1	RNA-DIRECTED DNA METHYLATION 1
RDR2	RNA-DEPENDENT RNA POLYMERASE 2
ROS1	REPRESSOR OF SILENCING 1
RdDM	RNA-Directed DNA Methylation
scRNA	scaffold RNA
siRNA	small interfering RNA
SUVH2	SUPPRESSOR OF VARIEGATION 3-9 HOMOLOG 2
TE	Transposable Element
TGS	Transcriptional Gene Silencing
WT	Wild Type
ZDP	ZINC FINGER DNA 3' PHOSPHOESTERASE

ABSTRACT

Tang, Kai PhD, Purdue University, December 2016. Dynamic Regulation of DNA Demethylation and RNA-Directed DNA Methylation in *Arabidopsis*. Major Professor: Jian-Kang Zhu.

DNA methylation is an important epigenetic mark present in many eukaryotes, and is involved in many crucial biological processes, such as gene imprinting, regulation of gene expression, and genome stability. Proper genomic DNA methylation patterns are achieved through the concerted action of DNA methylation and demethylation pathways. In the model plant species *Arabidopsis thaliana*, ROS1 (REPRESSOR OF SILENCING 1) is one of the DNA demethylases and the key component in the demethylation pathway. Dysfunction of ROS1 leads to increase in DNA methylation level at thousands of genomic loci. However, the features of ROS1 targets are not well understood.

In the first part of this dissertation, I will describe a study in which we identified and characterized genome-wide ROS1 target loci in *Arabidopsis* Col-0 and C24 ecotypes. In this study, we showed that ROS1 targets are associated with an enrichment of H3K18ac and H3K27me3, and with depletion of H3K27me1 and H3K9me2. Also we found that ROS1 can prevent the spreading of DNA methylation from highly methylated transposons to their nearby genes. Unexpectedly, we uncovered thousands of previously unidentified RdDM (RNA-directed DNA methylation) targets by analyzing the DNA methylome of *ros1/nrpd1* double mutant plants. In addition, we showed that ROS1 also antagonizes RdDM-independent DNA methylation at more than one thousand genomic loci. Our results provide significant insights into the genome-wide effects of both ROS1-mediated active DNA demethylation and RNA-directed DNA methylation, as well as their interactions in plants.

In the second part of this dissertation, I will describe another study in which we focused on AGO6 and AGO4, two Argonaute proteins involved in RdDM. AGO6 is generally considered to be redundant with AGO4 in RdDM. However, our genome-wide DNA methylation profiles and immunofluorescence localization analyses showed that redundancy between AGO4 and AGO6 is unexpectedly negligible in the genetic interactions and AGO4 and AGO6 mainly act sequentially in mediating RdDM.

1. INTRODUCTION

1.1 Transcriptional Gene Silencing

Gene silencing is important for development, stress responses, and suppression of viruses, transposons or transgenes [1–3] . Several epigenetic phenomena such as genome imprinting [4] and X chromosome inactivation [5] are caused by transcriptional gene silencing (TGS). The expression of some transgenes as well as developmentally or environmentally regulated endogenous genes can also be affected by TGS [6]. DNA methylation is an important contributor to TGS events. Mutations in DNA methylation factors have been shown to release TGS of a number of genes [7] . Analysis of gene silencing often uses *Arabidopsis* as a model system because mutations in the relevant components in this plant are not lethal, thus allowing their effects on development and physiology to be analyzed throughout the entire life cycle of a multicellular organism [8, 9].

1.2 DNA Methylation

In prokaryotes, DNA methylation is important in DNA repair and replication, and in recognition and protection of self DNA [10]. In eukaryotes, DNA methylation is essential for gene repression, genome organization and stability, genomic imprinting, X chromosome inactivation, and other developmental aspects [6]. Aberrant methylation patterns of tumor suppressor genes and oncogenes are common features of many cancers [11] . In mammals, methylation occurs almost exclusively at the sites of CG dinucleotides, although some non-CG methylation was also observed in mammalian embryonic stem cells, fibroblast cells, and brain tissues [12]. In most vertebrates, 60-90% of the cytosines at CG dinucleotides are methylated [13] . In contrast, methy-

lation in plants can occur at all three contexts; *i.e.*, the symmetric CG and CHG contexts (H = A, T or C) and the asymmetric CHH context.

. In *Arabidopsis*, DNA methylation is established and maintained by different pathways. *de novo* DNA methylation is carried out by DRM2 (DOMAINS REARRANGED METHYLTRANSFERASE 2) through the RdDM pathway. Four different DNA methyltransferases maintain DNA methylation after DNA replication, depending on the sequence context: mCG is maintained by MET1 (METHYLTRANSFERASE 1) , mCHG is maintained by CMT3 (CHROMOMETHYLASE 3), and mCHH is maintained by CMT2 or DRM2.

Small interfering RNAs (siRNAs) and scaffold RNAs (scRNAs) play important roles in guiding DRM2 to its target loci during the establishment of DNA methylation by RdDM. The production of siRNA and scRNA depends on plant specific RNA polymerase IV and V (Pol IV and Pol V), respectively. Pol IV is presumed to produce transcripts that serve as templates for RDR2 (RNA-DEPENDENT RNA POLYMERASE 2) to synthesize double-stranded RNAs, with the assistance of DTF1/SHH1 (DNA-BINDING TRANSCRIPTION FACTOR 1/SAWADEE HOMEODOMAIN HOMOLOG 1) and ClSY1 (CLASSY1) . Subsequently, DCL3 (DICER-LIKE 3) cleaves the double-stranded RNAs to form 24 nt (nucleotide) siRNAs, which are loaded onto AGO4 (ARGONAUTE 4) and AGO6 [14, 15]. The RdDM scRNAsare produced by Pol V, recruitment of which is facilitated by SUVH2 (SUPPRESSOR OF VARIEGATION 3-9 HOMOLOG 2) and SUVH9, and a protein complex containing DRD1 (DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1), DMS3 (DEFECTIVE IN MERISTEM SILENCING 3) and RDM1 (RNA-DIRECTED DNA METHYLATION 1) [14, 16]. The scRNAs recruit their complementary siRNAs in the siRNA-AGO4/6 complex to RdDM target loci and guides DRM2 to catalyze DNA methylation. Recently, it was found that Pol IV-dependent 25-35 nt precursors of the 24 nt siRNAs can trigger DNA methylation independently of the 24 nt siRNAs [17–20].

1.3 Active DNA Demethylation

DNA methylation influences diverse biological processes, and its levels are dynamically regulated. DNA methylation can be passively lost if it is not maintained after DNA replication, or can be actively removed by DNA demethylases [21]. In plants, active DNA demethylation is initiated by the ROS1 family proteins including ROS1, DEMETER (DME), DML2, and DML3. In contrast to DME, which is required for genomic imprinting during reproductive development, *ROS1*, *DML2*, and *DML3* are mainly expressed in vegetative tissues and can prevent hypermethylation of specific genomic regions, thus protecting these regions from transcriptional silencing [21].

ROS1 is the first genetically characterized DNA demethylase in eukaryotes [22]. The bifunctional glycosylase/lyase activities of ROS1 have been well studied in *Arabidopsis*. Its DNA glycosylase activity removes 5mC base from the DNA backbone, and then its lyase activity cleaves the backbone by either β- or β, δ-elimination, leaving a single nucleotide gap with 3'-phosphor-α, β-unsaturated aldehyde (3'-PUA) or 3' phosphate termini, respectively [23, 24]. The AP endonuclease-like protein APE1L (APURINIC/APYRIMIDINIC ENDONUCLEASE 1-LIKE) and the 3' DNA phosphatase ZDP (ZINC FINGER DNA 3' PHOSPHOESTERASE) convert the 3'-PUA and 3' phosphate groups, respectively, to generate a 3'-OH end, which allows a DNA polymerase and the ligase, AtLIG1 (*Arabidopsis thaliana* DNA LIGASE 1),to fill the gap with an unmethylated cytosine [25–27]. An anti-silencing protein complex containing MBD7 (METHYL-CPG-BINDING DOMAIN 7), IDM1 (INCREASED DNA METHYLATION 1),IDM2, and IDM3, was recently discovered to regulate ROS1 targeting, and in turn DNA demethylation. MBD7 binds highly methylated DNA and recruits the α-crystallin domain containing proteins IDM2 and IDM3 [28, 29], which then recruit the histone acetyltransferase IDM1. IDM1-mediated acetylation creates a permissible chromatin environment for ROS1 to initiate DNA demethylation [28, 30].

In plants, transposable elements (TEs) and other repetitive sequences are usually heavily methylated, and their methylation is known to be regulated by DNA methyla-

tion pathways and can also be regulated by ROS1-mediated DNA demethylation. In *Arabidopsis*, some TEs show lower expression levels in ros1 mutants due to increased DNA methylation [30,31]. Some methylated TEs are located near actively transcribed genes. For example, a transposable element, AT1TE40605, is located upstream of the promoter of *EPF2* (*EPIDERMAL PATTERNING FACTOR 2*) that encodes a negative regulator of stomata formation [32]. A recent study showed that *ros1* mutation results in spreading of methylation from AT1TE40605, and consequently silencing of *EPF2* and abnormal epidermal cell patterning [33]. Several other genes are also silenced in *ros1* mutants due to DNA hypermethylation [30, 31]. In addition, ROS1 family demethylases were found to positively regulate many fungal pathogen responsive genes via demethylating TEs located in their promoters. As a result, the DNA demethylase mutants showed enhanced susceptibility to fungal infection [34]. These studies suggest that ROS1-mediated demethylation of TEs is important for regulation of gene expression by preventing nearby genes from being silenced.

Proper genomic DNA methylation patterns are achieved through the concerted action of DNA methyltransferases and demethylases. Previous studies have suggested interactions between ROS1-mediated DNA demethylation and RdDM. ROS1 can antagonize RdDM at a transgene locus and several endogenous loci [35, 36]. Interestingly, *ROS1* is silenced in mutant plants defective in MET1 [37] or components of RdDM [38]. Recent studies revealed that the expression of *ROS1* is fine-tuned by RdDM and DNA demethylation pathways through dynamic regulation of methylation status of a methylation monitoring sequence (MEMS) in the *ROS1* promoter [39, 40].

1.4 Argonaute Protein Family

Argonaute (AGO) proteins are core components of small RNA-mediated silencing pathways, which negatively regulate gene expression in a sequence-specific manner. In the post-transcriptional gene silencing (PTGS) pathway, AGO proteins bind to small RNAs such as miRNAs, tasiRNAs, or piRNAs and mediate gene silencing through

degradation and/or translational suppression of the target mRNAs [41, 42]. In the TGS pathway, AGO proteins bind to siRNAs or piRNAs and subsequently facilitate the formation of repressive chromatin at loci that show complementarities to the small RNAs [41, 42] .

AGO proteins are highly conserved and present in various life forms ranging from Archaea to humans [43–46]. Each AGO protein is characterized by three conserved domains: PAZ, MID, and PIWI. The PAZ domain recognizes and binds the 3 end of small RNAs [47–49]. The MID domain contains a 5'-phosphate-binding pocket that binds to a guide small RNA at the 5' phosphate [50–53]. The PIWI domain is structurally similar to RNase H and, in some AGO proteins, exhibits endonuclease activity [44,54,55]. While the capacity to bind small RNAs is a common feature for all AGO proteins, different eukaryotic AGO proteins often have unique functions, which can arise from multiple factors such as distinct biochemical activities or differential association with specific types of small RNAs [56].

1.5 Argonaute function in *Arabidopsis*

Arabidopsis encodes 10 AGO proteins that are only partially understood. Among the characterized *Arabidopsis* AGO proteins, AGO1 is the major AGO protein that mediates miRNA- and tasiRNA-induced PTGS [41, 57]; AGO2 is required for the DNA methylation induced by 21 nt siRNAs at a subset of non-canonical RdDM target loci [58]; AGO2 also binds viral siRNAs and is involved in antiviral defense response [59,60]. In addition, AGO2 has been implicated in binding small RNAs that direct DNA double-stranded break repair through a process that also involves Pol IV and V [61]. AGO7 binds miR390 and controls the production of TAS3 tasiRNAs [52]; AGO10 specifically sequesters miR166/miR165 to regulate shoot apical meristem development [62]; and AGO4, AGO6, and AGO9 are involved in the RdDM pathway that requires 24 nt siRNAs.

RdDM confers transcriptional repression via the formation of heterochromatin, characterized by DNA methylation and repressive histone modifications [35, 63, 64]. In the canonical RdDM pathway, the complementary pairing of 24 nt siRNAs with nascent scaffold RNAs guides the DNA methylation complex to its target loci [15, 65–67].

Production of nearly all the 24 nt siRNAs requires Pol IV, a plant-specific DNA-dependent RNA polymerase [68]. According to the current paradigm, Pol IV transcription produces single-stranded non-coding RNAs, which serve as substrates for RDR2 production of double-stranded RNAs that are subsequently cleaved by DCL3 and loaded onto AGO4, AGO6, or AGO9 [65–67,69]. In parallel to siRNA production, Pol V generates long, non-coding RNAs that, before being released from the chromatin, function as scaffold RNAs for the recruitment of AGO-siRNA complexes [70]. In addition to Pol IV and Pol V, Pol II can also generate 24 nt siRNAs and scaffold RNAs at some intergenic, low-copy-number repeat loci [71]. AGO4 can be cross-linked to scaffold RNAs, supporting the model of siRNA-scaffold RNA pairing [72]. Both Pol V and Pol II possess an Argonaute-binding motif and have been shown to interact with AGO4 [71, 73]. In addition, AGO4 co-localizes with Pol V in perinucleolar foci [74] and with Pol II in the nucleoplasm [75]. Cytological analyses revealed co-localization of AGO4 and DRM2, together with other RdDM components in distinct subnuclear foci [74–77]. A recent study revealed that AGO4 can be co-purified with DRM2 in co-immunoprecipitation assays [78]. Additionally, AGO4 and DRM2 both co-immunoprecipitate with RDM1 [75]. Therefore, AGO4 has been assigned a critical role in targeting de novo DNA methylation in the RdDM pathway.

Arabidopsis AGO4, AGO6, and AGO9 are closely related family members [79]. AGO9 specifically silences TEs in the female gametophyte in a non-cell-autonomous manner [64]. Although the DNA methylation phenotype is unclear in the *ago9* mutant, AGO9 has been suggested to function through Pol IV-dependent 24 nt siRNAs because it preferentially binds 24 nt siRNAs and because a double mutant with dysfunctional Pol IV and Pol V resembles the *ago9* mutant in that they both are defective

in specifying the cell fate in the ovule [64]. Like AGO4, AGO6 has been clearly shown to regulate DNA methylation through the RdDM pathway [35, 63, 80]. Mutants defective in either AGO4 or AGO6 were both isolated in the same genetic screen for mutants with defective RdDM [35, 81]. This indicates that AGO4 and AGO6 can have non-redundant roles in regulating the same RdDM target, although AGO6 is generally considered to be redundant with AGO4 in regulating RdDM [80, 82]. It remains unclear, on a whole-genome scale, to what extent AGO4 and AGO6 function non-redundantly in the RdDM pathway. In fact, although AGO4 and AGO6 are known to differ in preference for siRNAs from different heterochromatin-associated loci [83], the genome-wide functional integration and/or diversification of these two closely related paralogs in regulating DNA methylation has yet to be explored.

2. THE DNA DEMETHYLASE ROS1 TARGETS GENOMIC REGIONS WITH DISTINCT CHROMATIN MODIFICATIONS

2.1 Abstract

Arabidopsis ROS1 is the first genetically characterized DNA demethylase in eukaryotes. Dysfunction of ROS1 leads to increase in DNA methylation level at thousands of genomic loci. However, the features of ROS1 targets are not well understood. In this study, we identified and characterized ROS1 target loci in *Arabidopsis* Col-0 and C24 ecotypes. Most ROS1 targets are TEs and intergenic regions. Compared to other TEs, ROS1-targeted TEs are closer to protein coding genes, suggesting a role for ROS1 in preventing the spreading of DNA methylation from highly methylated TEs to nearby genes. Interestingly, we found that unlike general TEs, ROS1 targets are associated with an enrichment of H3K18ac and H3K27me3, and depletion of H3K27me1 and H3K9me2. We investigated the antagonism between ROS1 and RdDM by identifying and characterizing thousands of genomic regions regulated by both ROS1 and RdDM. Unexpectedly, we uncovered thousands of previously unidentified RdDM targets by analyzing the DNA methylome of *ros1/nrpd1* double mutant plants. In addition, we show that ROS1 also antagonizes RdDM-independent DNA methylation at more than one thousand genomic loci. Our results provide significant insights into the genome-wide effects of both ROS1-mediated active DNA demethylation and RNA-directed DNA methylation as well as their interaction in plants.

2.2 Results

2.2.1 Characterization of *ros1* mutant methylomes in Col-0 and C24 ecotypes

ros1-4 is an *Arabidopsis* mutant of Col-0 ecotype with T-DNA insertion in the *ROS1* gene, causing complete loss of function of ROS1 [36]. *ros1-1* is a loss of function mutant of ROS1 in C24 ecotype and has a single nucleotide substitution in *ROS1* leading a premature stop codon, and is likely a null allele [22]. In this study, we generated single-base resolution maps of DNA methylomes of two-week-old seedlings of *ros1-4* and *ros1-1* mutants. Methylomes of Col-0 and C24 wild types at the same developmental stage were sequenced and served as controls.

To identify potential genomic targets of ROS1 and compare the targets in different ecotypes, we identified differentially methylated regions (DMRs) in *ros1-4* and *ros1-1* mutants relative to their respective wild type plants. *ros1-4* has 6902 hypermethylated DMRs (hyper DMRs) with an average length of 495 bp (base pair), and 1469 hypomethylated DMRs (hypo DMRs) with an average length of 193 bp, while 5011 hyper DMRs and 332 hypo DMRs were identified in *ros1-1*. The overwhelmingly higher numbers of hyper DMRs compared to hypo DMRs of ros1 mutants in both Col-0 and C24 are consistent with the ROS1 function in DNA demethylation. In *ros1-4*, 1887 (27%) hyper DMRs are in genic regions, 2878 (42%) in TE regions, 2010 (29%) in intergenic (IG) regions, and 127 (2%) in the category of others (Fig. 2.1). Compared to the composition of randomly selected control regions that are composed of 27% TEs, 54% genes, and 18% IG regions (Fig. 2.1), *ros1-4* hyper DMRs have a decreased percentage in genes and increased percentages in TEs and IG regions, which is also observed in *ros1-1* hyper DMRs (Fig. 2.1). This indicates that ROS1 preferentially targets TEs and IG regions. *ros1-1* and *ros1-4* hyper DMRs are distributed throughout the five chromosomes of *Arabidopsis* (Fig. 2.2). For both *ros1-1* and *ros1-4* hyper DMRs, DNA hypermethylation was detected in all three contexts (CG, CHG and CHH) (Fig. 2.3). The length distribution of ROS1 targeted TEs

is similar to that of all TEs (Table 2.1), suggesting that ROS1 has no preference for short or long TEs. Interestingly, by analyzing the distance between TEs and its nearest genes, we found that ROS1-targeted TEs in both *ros1-1* and *ros1-4* tend to be located closer to genes relative to TEs that are not targeted by ROS1 (Fig. 2.4).

In both *ros1-1* and *ros1-4*, DNA methylation is substantially increased around the boarders of TEs targeted by ROS1 (Fig. 2.5). As expected, these TEs display decreased methylation in *nrpd1* mutants, which are dysfunctional for RdDM due to disruption of NRPD1, the largest subunit of RNA polymerase IV (Fig. 2.5). Interestingly, we noticed that, the hypermethylation in *ros1* mutants extends from the TE borders to neighboring sequences before tempering off (Fig. 2.5). These patterns support our previous hypothesis that ROS1 may counteract RdDM to prevent the spreading of methylation from highly methylated regions, such as TEs, to nearby genes [21].

To investigate the influence of different genetic backgrounds on ROS1 targeting, we compared DMRs of *ros1-4* and *ros1-1*, which are mutants in Col-0 and C24 ecotypes respectively. We found that only 27% hyper DMRs in *ros1-4* are also hyper DMRs in *ros1-1*, suggesting that ROS1 targeting is greatly influenced by genetic backgrounds. Interestingly, 35% of TE-type and 28% of IG-type hyper DMRs are shared between *ros1-4* and *ros1-1*, but only 15% of genic-type hyper DMRs are shared between the two mutants. Thus, ROS1 targeting seems relatively conserved in TE regions in Col-0 and C24 ecotypes. Since TEs and genes typically display similar levels of genetic variation, these findings suggest that chromatin features important for active DNA demethylation might be more conserved at TEs than genes between the two ecotypes. Several examples of shared hyper DMRs and non-shared hyper DMRs are displayed in Fig. 2.6.

In summary, we identified, characterized, and compared targets of ROS1 in Col-0 and C24 genetic backgrounds. ROS1 targets in Col-0 and C24 display a preference for TEs and intergenic regions and the targeted TEs being located near genes. However,

the specific genomic regions targeted by ROS1 are largely different in Col-0 and C24 backgrounds.

2.2.2 Chromatin features associated with ROS1 targets

Histone modifications, such as histone methylation and acetylation, are known to interact with DNA methylation, therefore we determined which histone marks associated with ROS1 targets. Compared to control regions, which are randomly selected genomic regions with the same length distribution as the DMRs, both total TEs and ROS1 hyper DMRs show a slightly decrease in the level of H3 (Fig. 2.7 and 2.8), indicating a lower nucleosome density in TEs and ROS1 targets. ROS1 targets are negatively associated with most active histone marks compared to control regions, such as H3K36 di-/tri-methylation (H3K36me2/3), H3K4me2/3 and H3K9 acetylation (H3K9Ac) (Fig. 2.7, 2.8, and 2.9) which was expected since a large proportion of ROS1 targets are within TEs (Fig. 2.1). However, in contrast with most TEs, ROS1 targets are positively associated with the active histone mark H3K18Ac compared to control regions (Fig. 2.10, 2.11, and 2.12). Because 42% of *ros1-4* hyper DMRs are TE regions (Fig. 2.1), it is possible that the remaining 58% of *ros1* targets that are not within TEs account for the enrichment of H3K18Ac. To investigate this possibility, we compared TE, intergenic and genic types of *ros1* DMRs with simulated TE, intergenic, and genic regions, respectively. Consistently, we found that H3K18Ac is enriched in all types of *ros1* DMRs (Fig. 2.13), suggesting that *ros1* targets are indeed generally characterized by enrichment of H3K18Ac. The association with H3K18Ac is fully consistent with our previous finding that IDM1, an H3K14/18/23 acetyltransferase, is required for the demethylation of a subset of ROS1 targets [30].

We also identified additional histone marks that distinguish ROS1 target regions. As shown in Fig. 2.10, 2.11, and 2.12, TEs in general are negatively associated with H3K27me3, and are positively associated with H3K27me1 and H3K9me2. In contrast, *ros1* DMRs have the opposite features, in that they are associated with enrichment

of H3K27me3 and depletion of H3K27me1 and H3K9me2 (Fig. 2.10, 2.11, and 2.12). Similarly, we compared these chromatin features of *ros1* targets and corresponding simulation for each type of region (TE, intergenic and genic regions). All types of *ros1* targets are enriched of H3K27me3 compared to their respective simulated regions (Fig. 2.13). We didn't observed decreased H3K27me1 and H3K9me2 signal for genic and intergenic *ros1* targets, since the levels of these histone marks are already have very low in simulated genic and intergenic regions (Fig. 2.13). These results support that *ros1* targets are associated with enrichment of H3K27me3 and depletion of H3K27me1 and H3K9me2.

2.2.3 A new class of RdDM targets

De novo DNA methylation, especially in the CHH context, is established through the RdDM pathway, which requires DNA dependent RNA polymerase IV for small RNA production [15]. In previous studies, RdDM targets have been identified through the identification of hypo DMRs in RdDM mutants compared to wild type plants. In this study, we identified 4580 hypo DMRs and 2348 hyper DMRs in *nrpd1* mutant (RdDM Pol IV mutant) compared to wild type plants; and the large number of hypo DMRs in the Pol IV mutant is consistent with the role of Pol IV in DNA methylation. The more than 2000 hyper DMRs in Pol IV may be related to reduced expression level of *ROS1* in Pol IV mutant [39, 40] (see below).

Homeostasis of DNA methylation is regulated by DNA methylation and active DNA demethylation processes [21,84]. As diagrammed in Fig. 2.14, regions identified as hypo DMRs in Pol IV mutant must be methylated in the wild type. The presence of methylation in the wild type implies that RdDM dominates over active DNA demethylation at these loci or that active DNA demethylation does not occur at these loci. We refer to the 4580 hypo DMRs in Pol IV mutant as "type I" RdDM targets.

We hypothesized that DNA demethylation may be dominant over RdDM at some loci. These RdDM targets (we denote as "type II" RdDM targets) would not be methylated in wild type plants due to abundant active DNA demethylation (Fig. 2.14 and 2.15). To uncover type II RdDM targets, we introduced *nrpd1* mutation to *ros1-4* mutant, and compared the methylome of *ros1/nrpd1* double mutant with that of *ros1-4* mutant (Fig. 2.14 and 2.15). Type II regions would be predicted to gain cytosine methylation in *ros1-4* mutant due to loss of *ROS1* function. However, the gained cytosine methylation would be lost in in *ros1/nrpd1* mutant due to dysfunction in RdDM. In total, we identified 6069 hypo DMRs in in *ros1/nrpd1* compared to *ros1*. Out of the 6069 hypo DMRs, 3750 display DNA methylation in wild type (mC% $\geqslant$ 2%). Interestingly, consistent with our hypothesis, there are 2319 hypo DMRs that do not display DNA methylation in the wild type (mC% $<$ 2%). These represent the type II RdDM targets, which have not been identified previously. Similarly, by using published methylome data of *ros1-1/nrpd1* in C24 background, we found 4966 hypo DMRs in *ros1-1/nrpd1* double mutant compared to *ros1-1*, and 1656 of them are type II RdDM targets, suggesting that type II RdDM targets exist in both Col-0 and C24 ecotypes. As shown in Fig. 2.15, 2.16, 2.17, and 2.18, methylation of type I loci is decreased in *nrpd1*. In contrast, type II loci do not display a change in methylation level in *nrpd1* (Fig. 2.15, 2.16, 2.19, and 2.20). However, introducing the *nrpd1* mutation into the *ros1-4* mutant revealed the role of RdDM in DNA methylation at the type II loci (Fig. 2.14, 2.15, 2.16, 2.19, and 2.20).

We evaluated 24 nt siRNA enrichment for each type II locus and found that 39% of type II loci had 24 nt siRNAs (N $>$ 0 in either WT replicates), while 61% of type II loci did not have 24 nt siRNA reads (N = 0 in both WT replicates). We cannot exclude the possibility that siRNA levels at these loci were too low to be detected by siRNA-seq. Similar with type I targets, the type II targets also have decreased siRNA level in *nrpd1* mutant relative to wild type (Fig. 2.21). This result further supports that type II loci are targets of RdDM. However, the siRNA level in type II

loci is much lower than that in type I loci (Fig. 2.21), indicating weak RdDM at type II loci. This is consistent with a lack of methylation at type II loci in the wild type.

We examined Pol IV occupancy at type I and type II loci using previously published Pol IV ChIP-seq data [85]. Consistent with previous study [85], Pol IV was enriched in type I loci (Fig. 2.22). However, we did not observe a significant enrichment of Pol IV in type II loci. Both low siRNA level and low Pol IV enrichment were consistent with weak RdDM effect at these type II loci. However, type II loci have Pol IVdependent increased DNA methylation in *ros1* mutant (Fig. 2.16), suggesting stronger RdDM effect at these loci in *ros1* mutant.

We performed small RNA-Seq in *ros1-4* and *ros1-4/nrpd1* double mutant plants. We found that type II RdDM targets have significantly elevated 24 nt siRNA level in *ros1* relative to WT plants, and this increase in siRNAs can be suppressed by *nrpd1* mutation (Fig. 2.23). In contrast, type I RdDM targets did not display increased 24 nt siRNA level in *ros1* mutant compared to WT (Fig. 2.24). The results suggest that RdDM becomes stronger at type II loci when ROS1 is removed.

DRD1 is a component of the RdDM pathway, and a previous study showed that DRD1- mediated CHH methylation was positively correlated with euchromatic histone marks including H3K27me3, H3K4me2, H3K4me3, and H3K36me3, and was negatively correlated with H3, H3K9me2 and H3K27me1 [86]. Consistent with the previous study, type I and type II RdDM targets display a slight decrease in H3 enrichment (Fig. 2.25, 2.26, and 2.27), suggesting a reduced nucleosome density in RdDM target loci than in control regions. Our results showed that both type I and type II targets are associated with a depletion of euchromatic histone marks, including H3K4me2, H3K36me2 and H3K36me3 (Fig. 2.25, 2.26, and 2.27).

Type I and type II targets also display distinct chromatin features as shown in Fig. 2.28, 2.29, and 2.30. Chromatin features of type II RdDM targets are similar to ROS1 targets, including enrichment of H3K18Ac and H3K27me3 (Fig. 2.10 and 2.28), as expected. In contrast, type I targets display decreased H3K18Ac and slightly decreased H3K27me3 (Fig. 2.28). These distinct chromatin features are supported

by examination of type I and type II targets for different categories of regions (TE, IG and genic regions) (Fig. 2.31). Consistently, we found enrichment of H3K18Ac and H3K27me3 in all categories of type II targets compared to the corresponding categories of type I targets. Also unlike type I targets, type II targets are depleted of H3K9me2 and H3K27me1. The depletion of H3K9me2 and H3K27me1 was found only in TE and intergenic regions of type II compared to type I targets, but not in genic regions of type II targets (Fig. 2.31).

In summary, type I RdDM targets show DNA methylation in the wild type, and they may or may not be regulated by ROS1. In contrast, the newly discovered type II RdDM targets are all regulated by ROS1 and are essentially depleted of DNA methylation in the wild type due to ROS1 activity. The two types of RdDM targets are also characterized by distinct small RNA profiles and histone modification marks.

2.2.4 Relationship between ROS1-mediated DNA demethylation and RdDM pathway

ROS1-mediated active DNA demethylation counteracts RdDM pathway to prevent DNA hypermethylation at some specific loci [22, 75, 87]. However, the crosstalk between these two pathways genome wide has not been studied. To identify genomic regions targeted by both ROS1 and RdDM, we compared two groups of DMRs: hyper DMRs in *ros1* mutant and hypo DMRs in *nrpd1* mutants. We found that there are 1136 shared DMRs between *ros1* hyper DMRs and *nrpd1* hypo DMRs, suggesting that 16.5% (1136/6902) of ROS1 targets is antagonized by RdDM in wild type plants. However, this ratio increased to 60.1% (4146/6902) by using hypo DMRs identified in *ros1/nrpd1* mutant relative to *ros1*. This result suggests that the antagonistic effects between ROS1-mediated active DNA demethylation and RdDM have been underestimated, since type II RdDM targets were previously unappreciated.

It has been reported recently that there is a regulatory link between RdDM and ROS1-mediated active DNA demethylation. It was found that *ROS1* expression is

dramatically reduced in RdDM mutants, including *nrpd1*, due to the change of DNA methylation at the promoter region of *ROS1* gene [39, 40]. Thus, we speculated that the hyper DMRs in *nrpd1* might be caused by reduced *ROS1* expression. We found that 1026 of the 2348 hyper DMRs in *nrpd1* overlapped with hyper DMRs in *ros1*, suggesting that nearly half of the hyper methylated loci in *nrpd1* might be caused by reduction of *ROS1* expression in *nrpd1*. *ROS1* expression is reduced in not only *nrpd1*, but also other RdDM mutants, such as *nrpe1* [39]. As shown in Fig. 2.32, hyper DMRs of different RdDM mutants, including *nrpd1* and *nrpe1*, also have increased DNA methylation in *ros1* mutant, suggesting that the decreased *ROS1* expression level contributes to the hyper methylation in the two examined RdDM mutants. We then determined whether the methylome of the *nrpd1* single mutant, which has a dramatically reduced ROS1 expression level, is similar with the methylome of *ros1/nrpd1* double mutant (where *ROS1* was knocked out). After comparing *ros1/nrpd1* with *nrpd1*, we identified 3411 hyper DMRs in *ros1/nrpd1* relative to *nrpd1*. This finding suggests that the remaining *ROS1* expression in *nrpd1* still functions at thousands of loci, although other DMLs may also contribute to these hyper DMRs.

2.2.5 ROS1 antagonizes RdDM-independent DNA methylation

We identified 1026 shared hypermethylated loci in *nrpd1* and *ros1* (Fig. 2.33), which are distributed across five chromosomes (Fig. 2.34). Although these loci have similar chromatin features to ROS1 targets, such as H3K18Ac, H3K4me2, and H3K4me3 (Fig. 2.35), they display slightly increased levels of H3 compared to general ROS1 targets. At these loci, ROS1 prevents hypermethylation, and the methylation must be independent of RdDM since the methylation can occur in *nrpd1* mutant. This indicates that there are RdDM-independent pathways responsible for the methylation and are antagonistic to ROS1 at these loci. Using previously published methylome data (Table 2.2), we examined methylation levels of these 1026 loci in wild type,

nrpd1, *ros1*, *drm2*, *drm1drm2*, *cmt2*, *cmt3*, *cmt2cmt3*, *drm1/drm2/cmt2* (*ddcmt2*), *drm1/drm2/cmt3* (*ddcmt3*), *drm1/drm2/cmt2/cmt3* (*ddcc*), and *met1* mutants. The *nrpd1*, *ros1*, and *drm2* mutants display increased mCG, mCHG and mCHH levels at these loci (Fig. 2.36), suggesting that MET1, DRM1, CMT2 and CMT3 may all contribute to the methylation at these loci. Indeed, we found that the mCG level of these loci was significantly reduced in *met1* (Fig. 2.36), whereas the mCHG levels were significantly reduced in *cmt3*. Although CMT2 and CMT3 have been shown to function redundantly in mCHG methylation [88], it seems that the mCHG methylation at these loci mainly depends on CMT3 (Fig. 2.36). For mCHH methylation level, there were no significant changes in *cmt2* and *cmt2cmt3* double mutants. However, mCHH was significantly reduced in *ddcmt2* and *ddcc* mutants but was increased in *drm2*, *drm1drm2* and *ddcmt3* (Fig. 2.36). This indicates that DRM1 and CMT2 may function redundantly at these regions.

At these loci, mCHG and mCHH levels were increased in *met1* mutant plants (Fig. 2.36). This may be caused by the reduction in *ROS1* expression in *met1* mutants [37], such that the mCHG and mCHH methylation by CMT3, CMT2 and DRM1 could not be removed by ROS1. These results suggested that ROS1 antagonizes CMT3-, CMT2-, DRM1-, and MET1-mediated DNA methylation, which are independent of DRM2, the major DNA methyltransferase in the RdDM pathway (Fig. 2.37).

We examined siRNA levels at these 1026 loci in wild type and *nrpd1*, and found that 24 nt siRNAs accumulated at the loci in the wild type, but were lost in *nrpd1* mutant plants (Fig. 2.38). Since the siRNAs but not DNA methylation at these loci were dependent on Pol IV, the siRNAs at these loci would not be required for the methylation. This is consistent with a recent study showing that a reduction of siRNA levels in RdDM mutants does not substantially reduce CMT2-dependent CHH methylation [88].

In summary, our study revealed that, besides RdDM, ROS1 can antagonize DNA methylation mediated by MET1, DRM1 and CMTs in an siRNA-independent manner.

2.3 Discussion

Among the four proteins in the ROS1 family in *Arabidopsis*, ROS1 is the major DNA demethylase in vegetative tissues. In this study, we showed that genome-wide, ROS1 preferentially targets TEs that are close to protein coding genes (Fig. 2.5). We also showed that the sequences just outside the boarders of ROS1-targeted TEs have increased DNA methylation in *ros1* mutants (Fig. 2.5), suggesting that ROS1 prevents the spreading of DNA methylation from highly methylated TEs. Consistently, Yamamuro *et al.* reported that ROS1 is required for the expression of the *EPF2* gene by preventing the spreading of methylation from a TE near the promoter of *EPF2* [33]. In addition, ROS1 family demethylases can positively regulate fungal pathogen responsive genes via demethylating TEs located in or near their promoters [34]. Together with these previous studies, our data support that ROS1 is involved in the regulation of gene expression by preventing DNA methylation spreading from nearby TEs.

H3K18Ac is an active histone mark correlated with transcriptional activation [89]. We found that ROS1 targets are positively associated with H3K18Ac (Fig. 2.10 and 2.11), supporting our previous work showing that IDM1, an H3K18/23 acetyltransferase, can create a permissive chromatin environment important for ROS1 to access target loci 20. ROS1 targets were also found to be enriched with H3K27me3 (Fig. 2.10 and 2.11), but depleted of H3K27me1 and H3K9me2, which is opposite to general TEs. This is consistent with a previous finding that there was a strong correlation between H3K18Ac and H3K27me3 in *Arabidopsis* [90], and is also consistent with findings in mammals that DNA demethylation process is coupled with decreased H3K27me1 and H3K9me2 [91] .

Previous studies showed that several genomic loci were regulated by both RdDM and ROS1-mediated DNA demethylation, suggesting an antagonism between RdDM and ROS1. Consistently, as shown in Fig. 2.5, ROS1-targeted TEs had decreases in DNA methylation in *nrpd1* mutants. At some of these genomic regions, it is possible

that ROS1-mediated DNA demethylation is so strong that there is no methylation in the wild type at the regions. Therefore, these potential RdDM targets cannot be identified through comparison between RdDM mutants and wild type plants. In this study, we discovered over two thousand of this type of RdDM targets, named type II RdDM targets, by comparing *ros1* and *ros1/nrpd1* mutants. These RdDM targets have eluded previous attempts of RdDM target identification. Our identification of these type II loci suggested that the number of RdDM targets has been greatly underestimated.

To compare this newly uncovered group of RdDM targets with previously known RdDM targets, we examined their siRNA accumulation and Pol IV occupancy. Overall siRNA enrichment and Pol IV occupancy were lower at type II targets, consistent with weaker RdDM at these loci. More than half of the type II loci did not have any siRNA reads, and we did not observe any significant Pol IV enrichment at type II loci. However, we cannot exclude the possibility that siRNA level and Pol IV occupancy at these type loci were too low to be detected through siRNA sequencing and ChIP-seq, Interestingly, we observed increased siRNA level at type II loci in *ros1-4* mutant, indicating stronger RdDM effect in *ros1* mutant. It is possible that the demethylation effect by ROS1 at these loci in wild type can limit RdDM accessibility to these loci thus leading to weak RdDM effect in wild type, and this suppression of RdDM in wild type can be removed by ROS1 mutation. ROS1 has been shown to have similar binding affinity to both methylated and non-methylated DNA through a Lysine-rich Domain at the N terminal [92]. Thus, it is possible that ROS1 antagonizes RdDM not only by removal of DNA methylation, but also by preventing the access of RdDM machinery to the target loci. In the future, it will be interesting to compare Pol IV occupancy at these loci in wild type and *ros1* mutant plants to further investigate this possibility.

It is well known that ROS1 expression is dramatically reduced in RdDM mutants [39]. Our results suggested that the reduction in ROS1 expression in *nrpd1* mutant plants causes DNA hypermethylation at more than one thousand genomic re-

gions in the mutant. The DNA hypermethylation in RdDM mutants must be caused by some RdDM-independent DNA methylation pathways. Our analysis suggested that four DNA methylases including DRM1, CMT2, CMT3, and MET1 contribute to the hypermethylation in RdDM mutants. This finding implies that ROS1 can also antagonize RdDM-independent DNA methylation. Interestingly, we noticed that the *nrpd1* mutant had a slight increase in DNA methylation at type II RdDM loci compared to the wild type (Fig. 2.16). Actually, we found that 198 out of 2319 type II loci were overlapped with 1026 hyper DMR in *nrpd1* mutant, indicating that at some of these type II loci, RdDM-independent DNA methylation may compensate to methylate DNA when RdDM is lacking. Our findings suggested that the fine tuning of the plant methylome is complex and involves interactions among RdDM, RdDM-independent DNA methylation, and ROS1 family demethylase-mediated DNA demethylation.

2.4 Material and methods

2.4.1 Plant materials

Mutants including *ros1-4*, *nrpd1-3* (SALK_128428), *ros1/nrpd1* double mutant and *nrpe1-11* (SALK_029919) are in the Columbia-0 (Col-0) background. *ros1-4* and *nrpd1-3* were crossed to generate *ros1/nrpd1* double mutant. *ros1-1*, *nrpd1* (C24) and *ros1-1/nrpd1* (C24) are *Arabidopsis* mutants of C24 ecotype. The wild-type C24 and *ros1-1* mutant plants carried a homozygous RD29A promoter-driven luciferase transgene and a 35S promoter-driven NPTII transgene.

Seeds were stratified for 2-3 d at 4 °C before being sown on 1/2 MS plates containing 2% (wt/vol) sucrose and 0.7% (wt/vol) agar. All of the plants were grown under long day conditions at 22 °C. DNA was extracted from 14-d-old seedlings.

2.4.2 Whole genome bisulfite sequencing and analysis

The genomic DNA was extracted from 1 gram of 14-day-old seedlings using the Plant DNeasy Maxi Kit from Qiagen. And 5 g of gDNA was used for library construction using Illumina's standard DNA methylation analyses protocol and the TruSeq DNA sample preparation kit. The samples in Col-0 background were sequenced in the Genomics Core Facilities of the Shanghai Center for Plant Stress Biology, SIBS, CAS (Shanghai, China) with Illumina HiSeq2500. The samples in C24 background were sequenced in the Biosciences Core Laboratory of King Abdullah University of Science & Technology (KAUST) with Illumina HiSeq2000.

For Col-0 background data analysis, low quality sequences ($q < 20$) were trimmed using trim in BRAT-BW [93], and clean reads were mapped to the TAIR10 genome using BRAT-BW and allowing two mismatches. To remove potential PCR duplicates, the remove-dupl command of BRAT-BW was used. DNA hypomethylated regions were identified according to Ausin *et al.* [94] with minor modification. In brief, only cytosines with 4× coverage in all libraries in the same background were considered. A sliding-window approach with a 200-bp window sliding at 50-bp intervals was used to identify DMRs. Fisher's exact test was performed for methylated versus unmethylated cytosines for each context, within each window, with FDRs estimated using a Benjamini-Hochberg adjustment of Fisher's p-values calculated in the R environment. Windows with an FDR ≤ 0.05 were considered for further analysis, and windows within 100 bp of each other were merged to larger regions. Regions were then adjusted to shrink to the first and last differentially methylated cytosines (DMCs). A cytosine was considered differentially methylated if it showed at least a two-fold change in methylation percentage in the mutant. The regions were then filtered to include only those with at least 10 DMCs and with at least a two-fold change in arithmetic mean of methylation percentage of all cytosine.

For *ros1-1* and C24 data, clean reads were mapped to a pseudo-C24 genome using BRAT-BW allowing two mismatches. We used public data set of *ros1-1/nrpd1* (C24)

double mutant in C24 background [95] to analyze type II RdDM targets. The pseudo-C24 genome was generated through the replacement of SNPs in the Col-0 genome (TAIR10) with C24 variants (1001genomes.org/data/MPI/MPISchneeberger2011/releases/current/C24/Marker/C24.SNPs.TAIR9.txt).

2.4.3 TE border analysis

The analysis was according to previously described method [96]: *ros1-4* hyper DMR associated TE were aligned at the 5' end or the 3' end. We discarded from the analysis 250 bp from the end opposite to the one used for alignment to avoid averaging the edges of shorter TEs with the middles of longer sequences.

2.4.4 Histone feature analysis

Histone features were analyzed according to a previously described method [17] with a minor modification: Briefly, the public data used for the analysis were downloaded from Gene Expression Omnibus (Accession No: GSE28398) [90]. The color-space reads were aligned to TAIR10 genome using Bowtie [97] allowing no more than 3 mismatches. Only reads that are uniquely mapped to the genome were retained for the downstream analysis. To generate the relative histone signal distribution in the flanking 5-kb region of the mid-point of DMRs, the whole region (10050 bp long) was divided into 201 bins with a size of 50 bp and the 101st bin aligning at the middle point of each DMR. The number of depth in each of the 201 bins was summed. The relative histone modification signal (y axis) in each of the 201 bins was defined as: n(Histone_modification) $\times$ N(Input) / [N(Histone_modification) $\times$ n(Input)], where n is the sum of depth of the corresponding library in each bin and N is the number of mapped reads of the corresponding library.

For box plots, DMRs were considered as the 1050 bp region from the DMR mid-point (+/-10 bins plus the mid-bin). In each region, the relative histone modification signal was calculated as above. The box plots were generated in R using function

"boxplot" with parameter "range = 1.5, outline = F, notch = T". The p-values were calculated in R using function "wilcox.test".

2.4.5 Small RNA analysis

Small RNA samples were prepared from 14-day-old seedlings. The analysis pipeline was similar as Zhang *et al.* [98]: Briefly, after adapter sequences were trimmed, clean reads with sizes ranging from 18 to 31 nt were mapped to the *Arabidopsis* genome (TAIR10) using Bowtie [97] with parameters "-v 0 k 10". Read counts were normalized to Reads Per Ten Million (RPTM) based on the total abundance of genome-matched small RNA reads, excluding structural small RNAs originating from annotated tRNAs, rRNAs, snRNAs, and snoRNAs. The hits-normalized-abundance (HNA) values were calculated by dividing the normalized abundance (in RPTM) for each small RNA hit, where a hit is defined as simply the number of loci at which a given sequence perfectly matches the genome [99]. The HNA values of 24 nt small RNAs within DMRs throughout the whole genome were calculated.

2.4.6 Pol IV ChIP-seq analysis

The data sets we used are from published paper [85]. According to this paper, WT is pure wild-type plants without any transgenes. *nrpd1*/NRPD1-3×FLAG is *nrpd1* mutant with NRPD1-3×FLAG transgene.

2.5 Figures

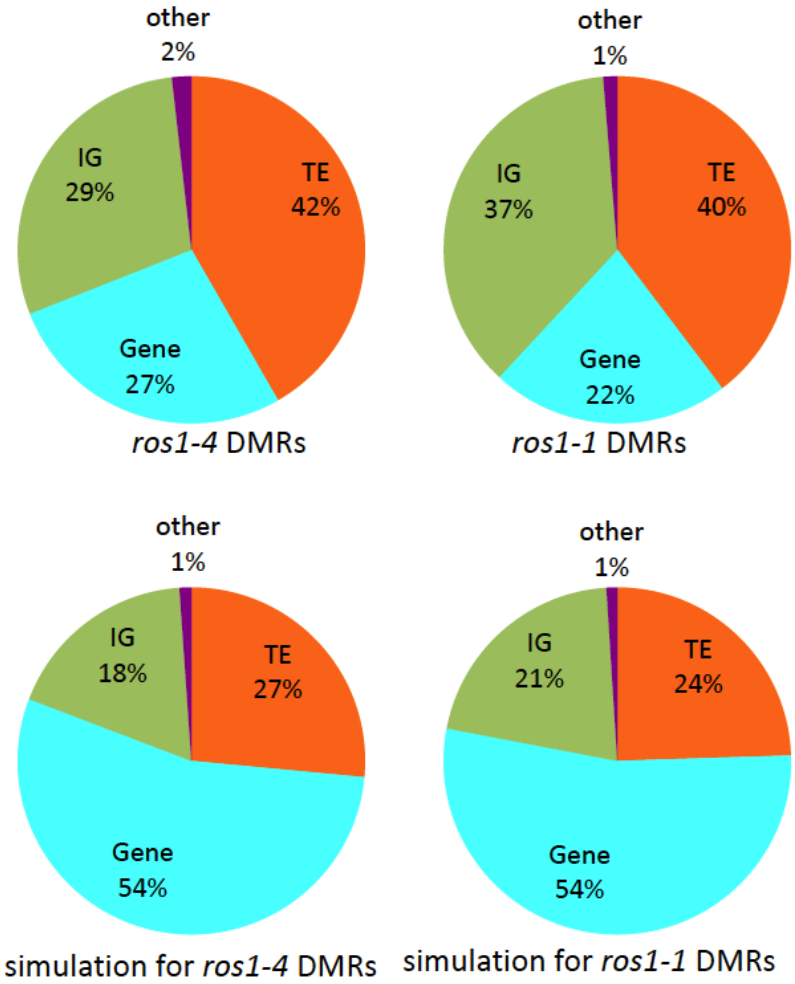

Fig. 2.1. Composition of *ros1* hyper DMRs. Composition of the hyper DMRs in *ros1-4*, *ros1-1* and of the simulated genomic regions. Simulations are randomly selected regions with the same length distribution as the DMRs.

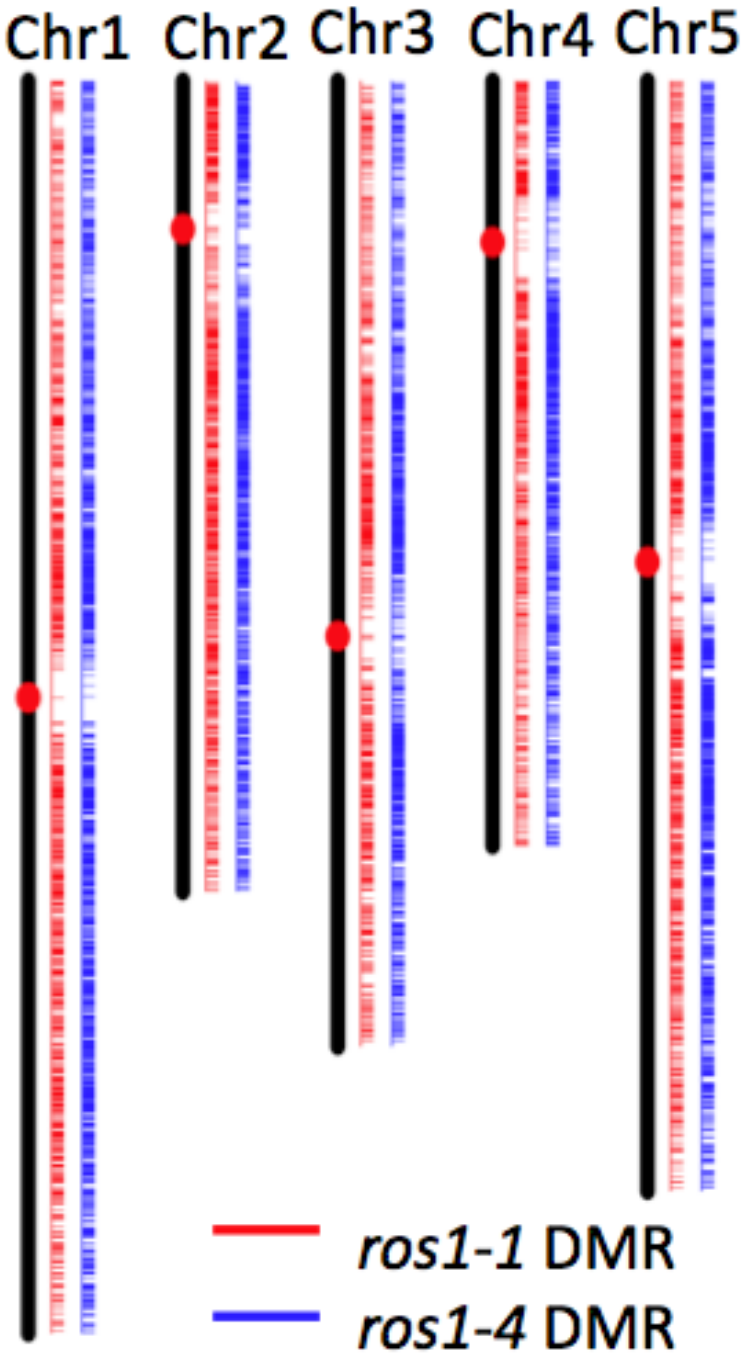

Fig. 2.2. Chromosomal distribution of *ros1* hyper DMRs. Each red horizontal bar represents a DMR in *ros1-1* and each blue horizontal bar represents a DMR in *ros1-4*. The red dots indicate the centromeres.

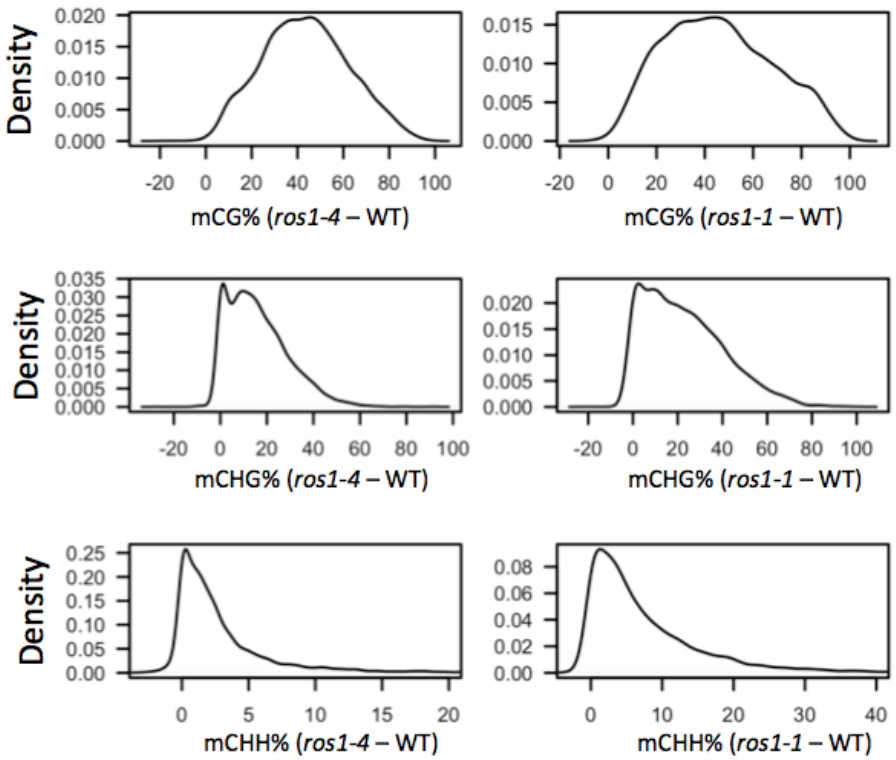

Fig. 2.3. Kernel density plots of methylation changes. Methylation changes in *ros1* mutants relative to wild type at *ros1* hyper DMRs are shown in the kernel density plots. Changes in mCG, mCHG and mCHH are shown separately.

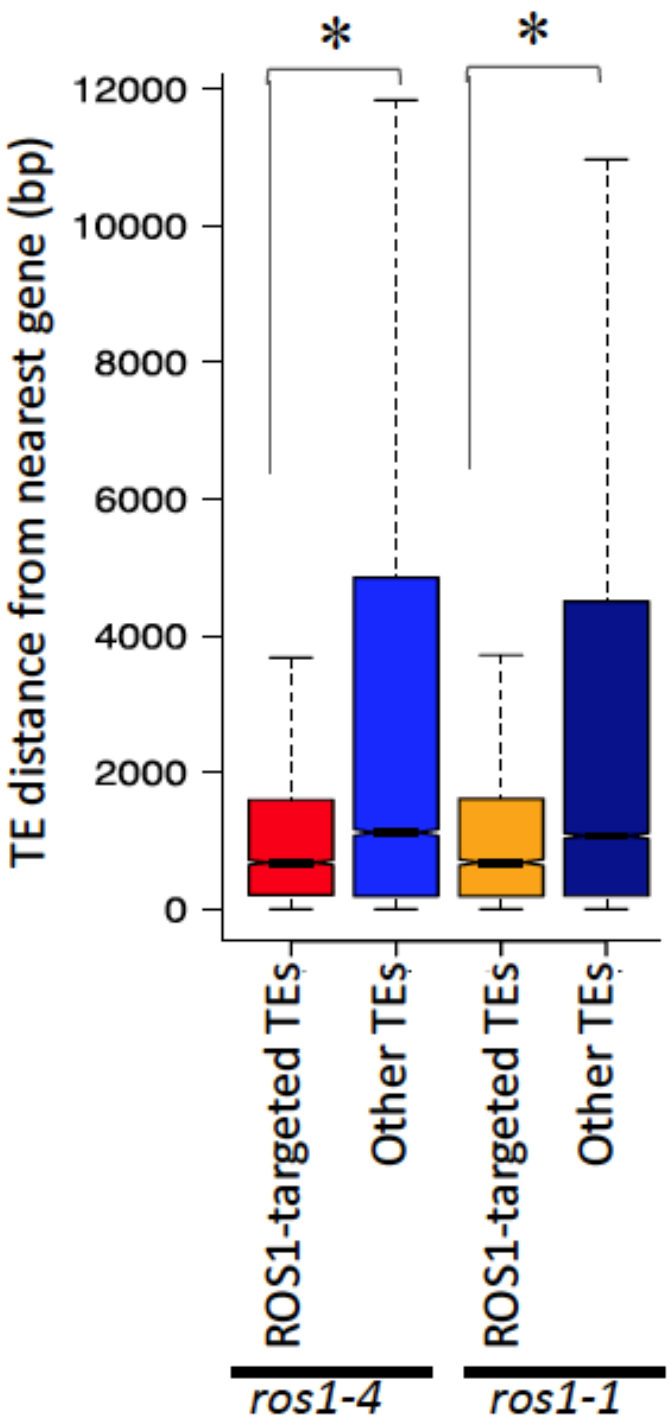

Fig. 2.4. Distances between ROS1-targeted TE and their nearest genes. Box plot showing the distances between ROS1-targeted or non-targeted TEs and their nearest protein coding genes. For both Col-0 and C24 ecotypes, ROS1-targeted TEs are significantly closer to genes relative to TEs that are not targeted by ROS1 (* $p\text{-}value < 2.2e-16$, one-tailed Wilcoxon rank sum test).

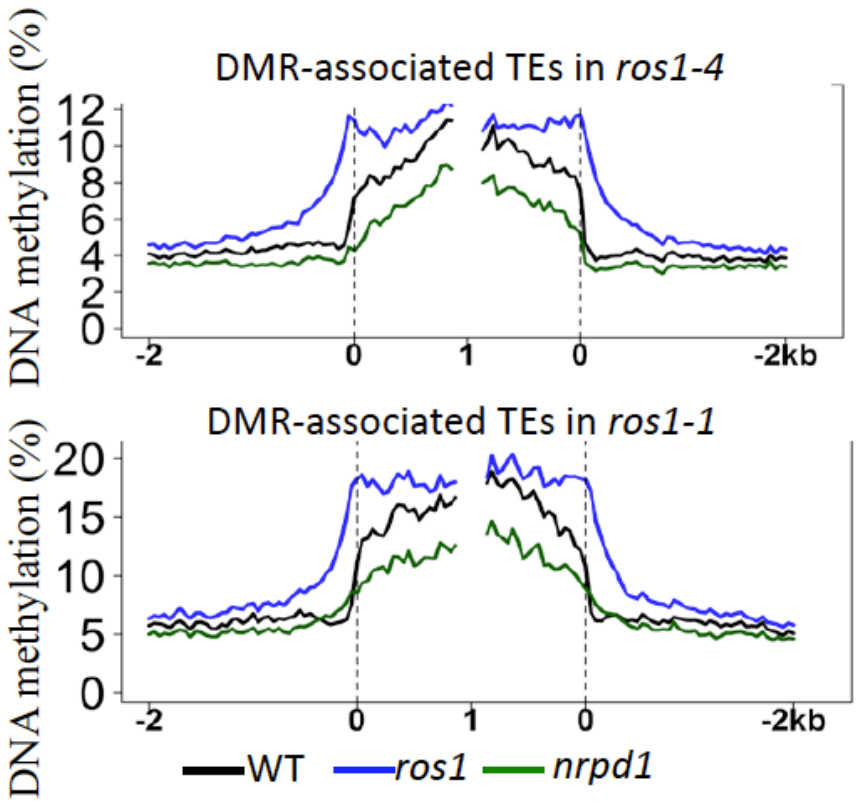

Fig. 2.5. Increased DNA methylation around the boarders of TEs targeted by ROS1. DNA methylation levels of *ros1* hyper DMR-associated TEs in wild type, *ros1* and *nrpd1* mutants. The upper panel shows methylation levels in Col-0, and the lower panel shows the levels in C24. TEs were aligned at the 5' end or the 3' end, and average methylation for all cytosines within each 50 bp interval was plotted.

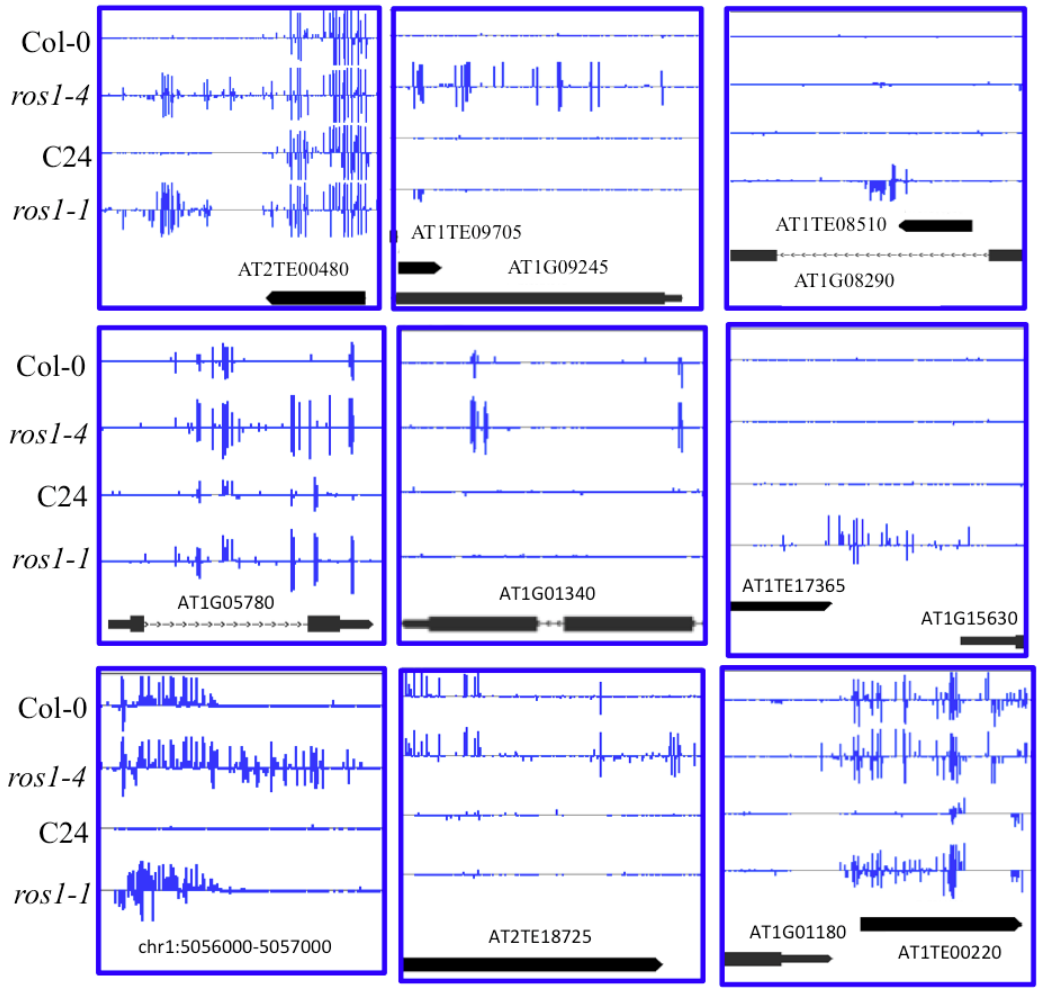

Fig. 2.6. Screenshots of *ros1* hyper DMR. Levels of DNA methylation at shared or non-shared hyper DMRs between *ros1-1* and *ros1-4*. Integrated Genome Browser (IGB) display of whole-genome bisulfite sequencing data is shown in the screenshots. DNA methylation levels of cytosines were indicated with the heights of vertical bars on each track. Left three panels show hyper DMRs in both *ros1-4* and *ros1-1*. Middle three panels show hyper DMRs in only *ros1-4* but not *ros1-1*. Right three panels show hyper DMRs in only *ros1-1* but not *ros1-4*.

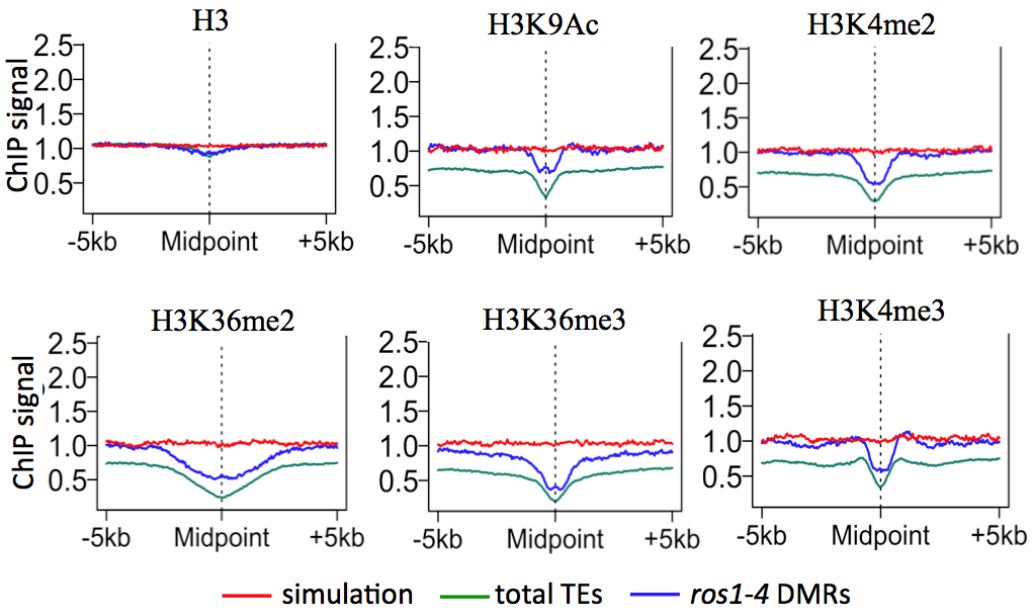

Fig. 2.7. Similar chromatin features shared by ROS1 targets and TEs. Association of different histone modifications surrounding the mid-point of *ros1-4* hyper DMRs, all TEs and simulated regions. Similar with general TEs, *ros1-4* hyper DMRs are negatively associated with H3K36me2, H3K36me3, H3K4me2, H3K4me3 and H3K9Ac.

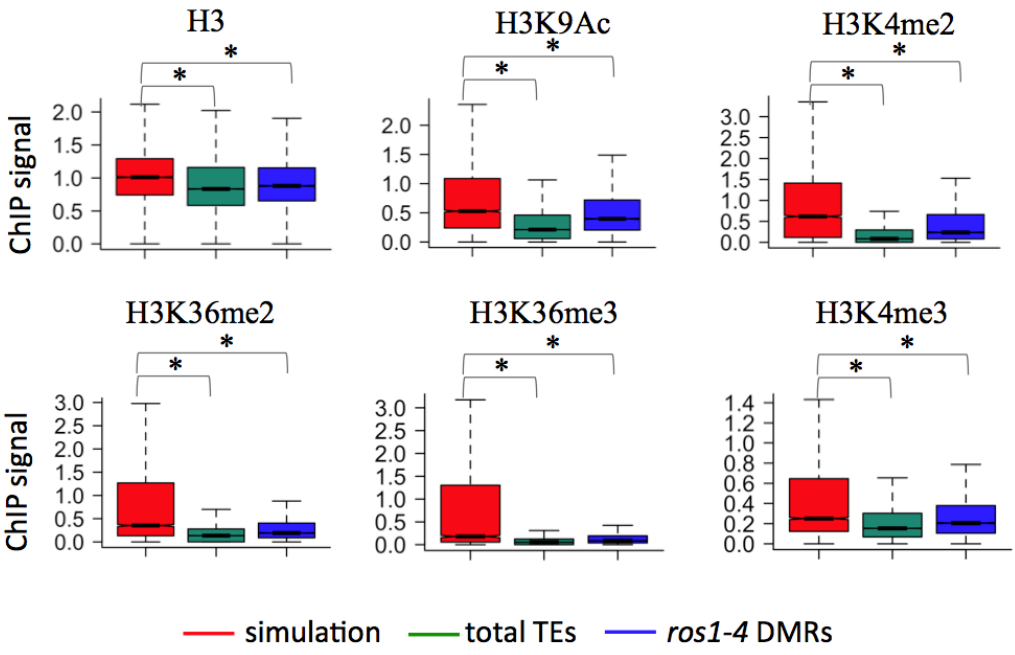

Fig. 2.8. Box plots of ChIP signal of similar chromatin features shared by ROS1 targets and TEs. Box plots display of the results in Fig. 2.7 (* $p\text{-}value < 1e-15$, one-tailed Wilcoxon rank sum test).

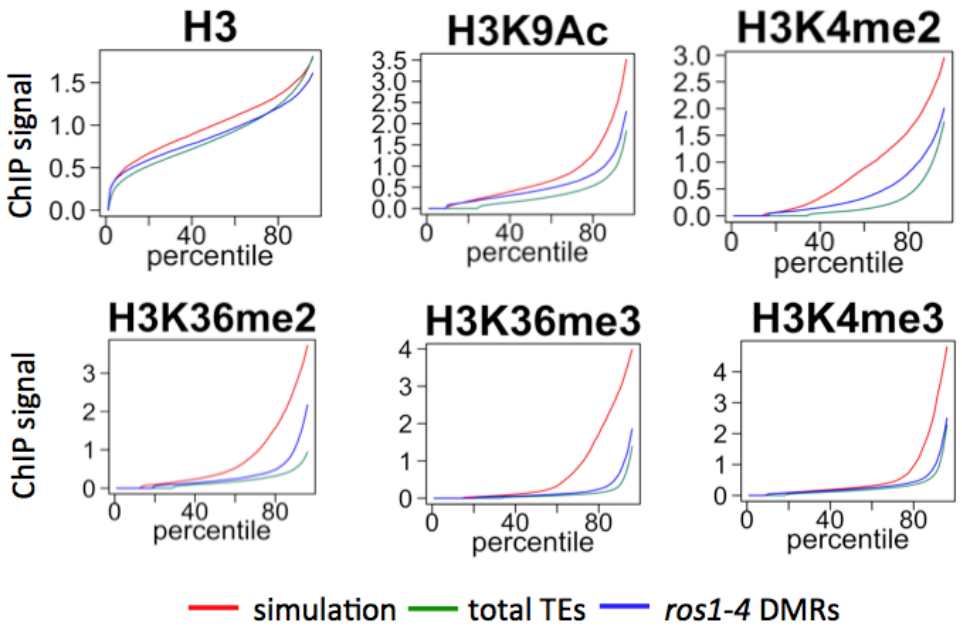

Fig. 2.9. Percentile plots of ChIP signal of similar chromatin features shared by ROS1 targets and TEs. Percentile plots of the same data as in Fig. 2.8. For each histone mark, simulated regions, TEs and *ros1-4* DMRs were ranked based on their histone ChIP signals from low (left) to high (right) along X-axis. X-axis is ranking percentile, and Y-axis is ChIP signal.

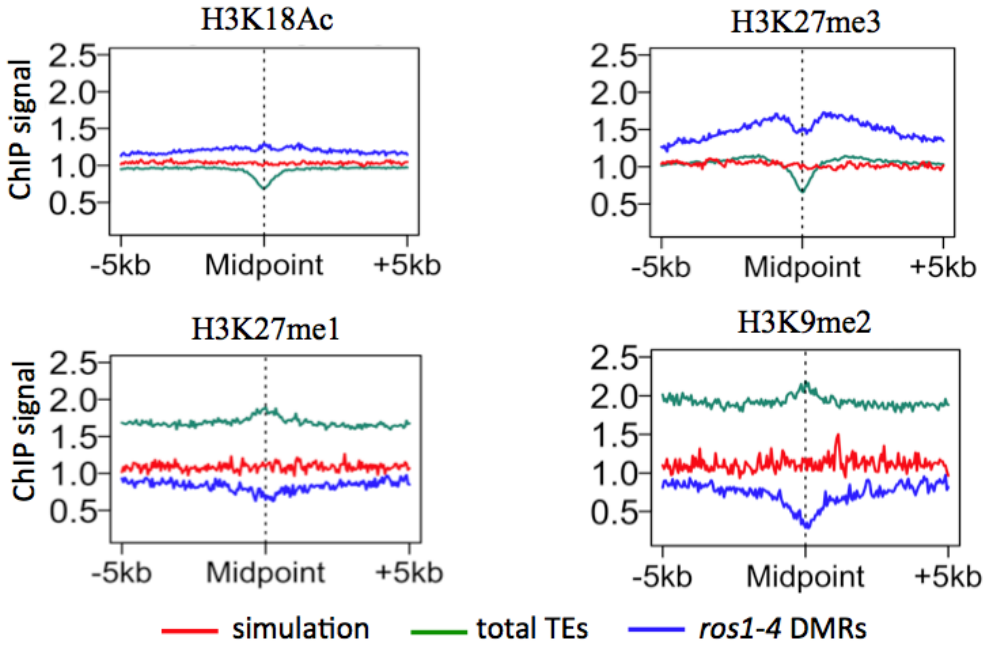

Fig. 2.10. Distinct chromatin features associated with ROS1 targets and TEs. Association of different histone modifications surrounding *ros1-4* hyper DMRs. Association of histone modifications at total TEs and simulated regions served as controls. In contrast to total TEs, *ros1-4* hyper DMRs are positively associated with H3K18Ac, H3K27me3, and negatively associated with H3K27me1 and H3K9me2.

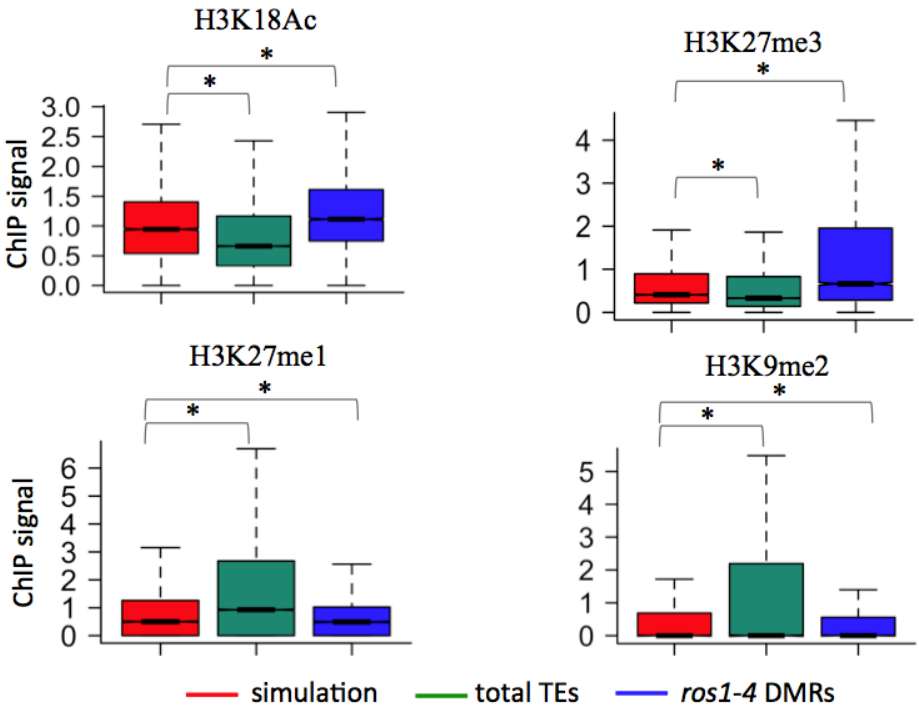

Fig. 2.11. Box plots of ChIP signal of distinct chromatin features associated with ROS1 targets and TEs. Box plots display of the results in Fig. 2.10 (* *p-value* $< 1e-15$, one-tailed Wilcoxon rank sum test).

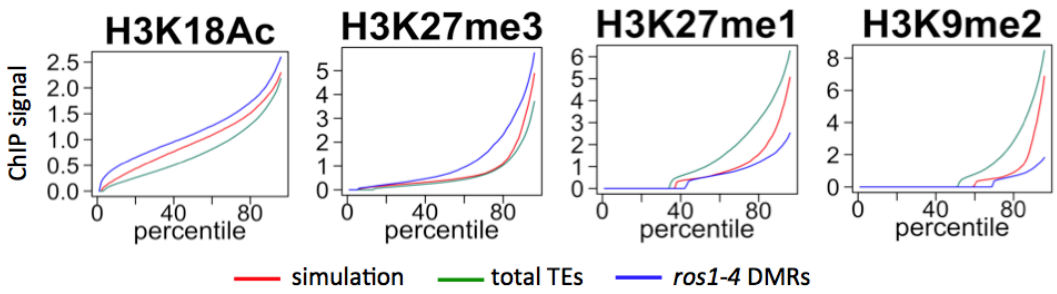

Fig. 2.12. Percentile plots of ChIP signal of distinct chromatin features associated with ROS1 targets and TEs. Percentile plots of the same data as in Fig. 2.11. For each histone mark, simulated regions, TEs and *ros1-4* DMRs were ranked based on their histone ChIP signals from low (left) to high (right) along X-axis. X-axis is ranking percentile, and Y-axis is ChIP signal.

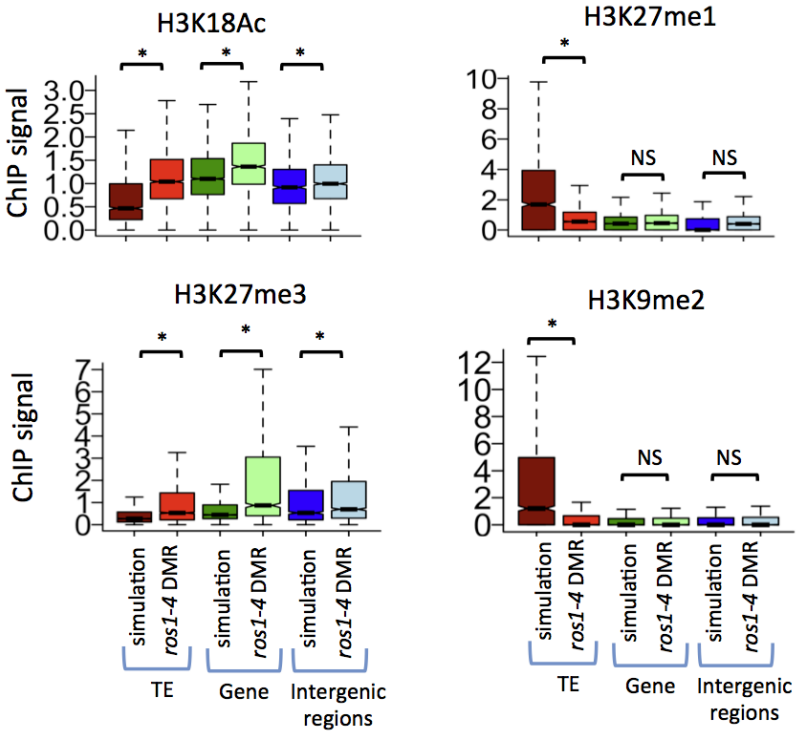

Fig. 2.13. Box plots of histone ChIP signal in three types of genomic regions of *ros1-4* hyper DMRs. Box plot display of histone modifications of different types of *ros1-4* hyper DMRs and corresponding simulation regions (* $p\text{-}value < 1e-6$, one-tailed Wilcoxon rank sum test).

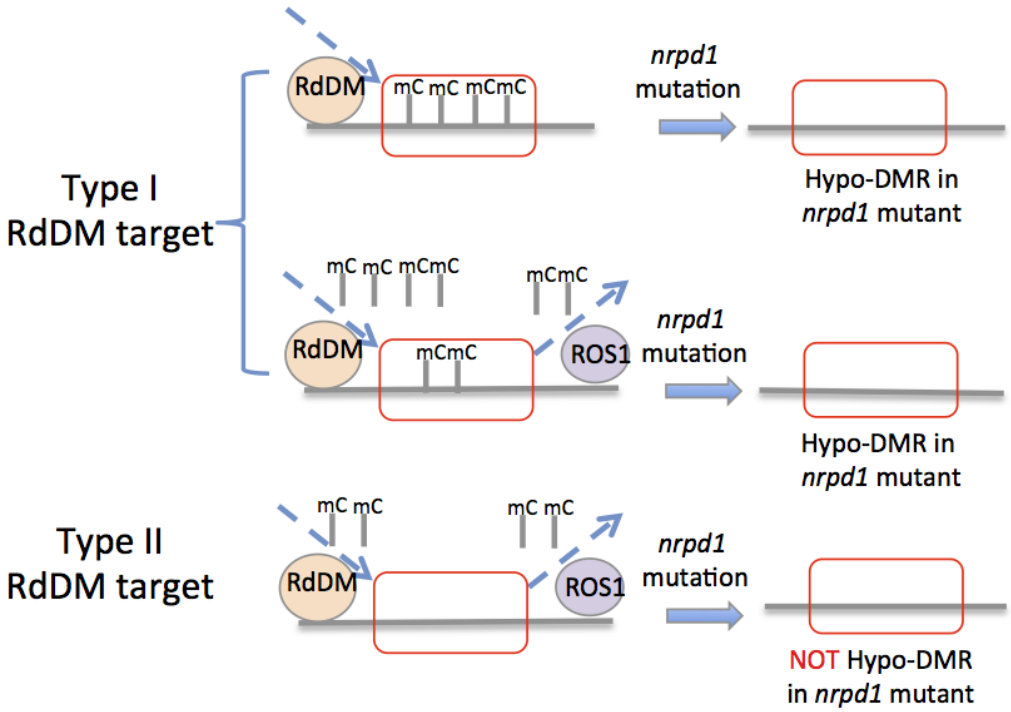

Fig. 2.14. Schematic diagram of different RdDM targets. Schematic hypothesis that different RdDM targets may be regulated differently by ROS1. Some type I RdDM targets are not regulated by ROS1 (Upper panel) whereas other type I RdDM targets are regulated by ROS1, although RdDM is more dominant at these loci (Middle panel). Type II RdDM targets are always regulated by ROS1, and ROS1 is more dominant at these loci (Lower panel).

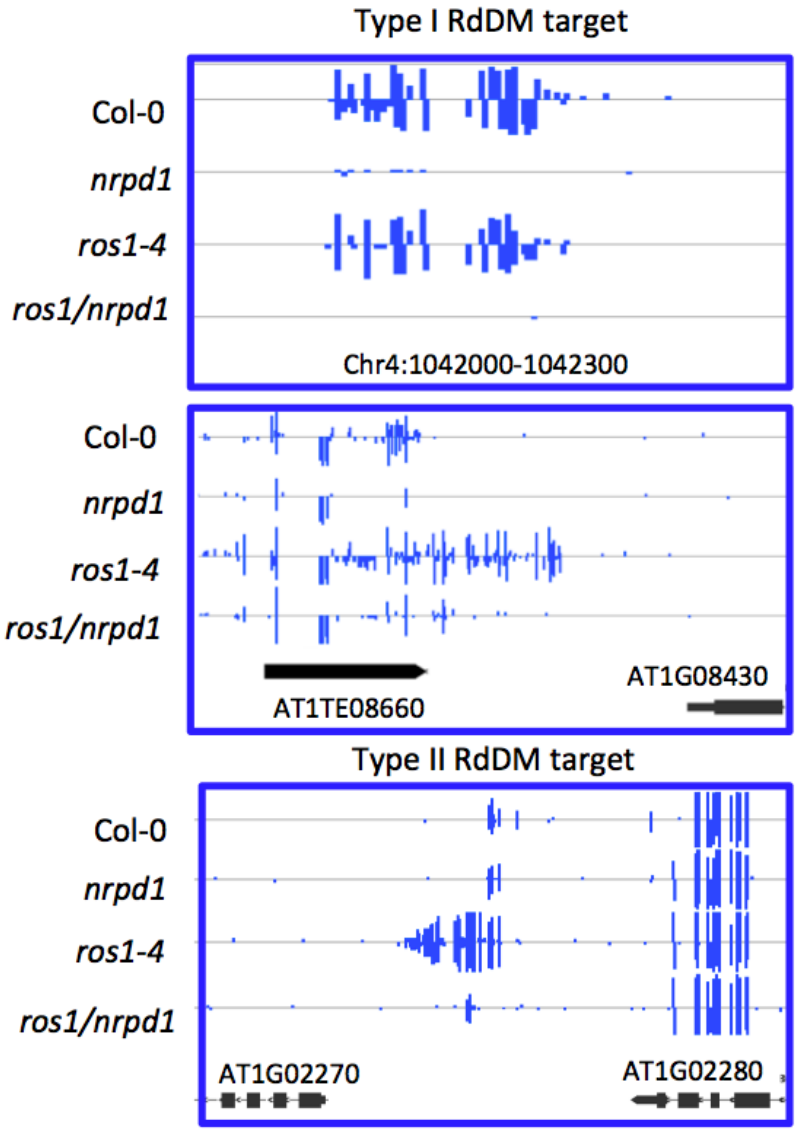

Fig. 2.15. Screenshots of type I and type II RdDM targets. Methylation levels of type I and type II RdDM targets in Col-0, *nrpd1*, *ros1-4*, and *ros1/nrpd1*. The three regions, from top to the bottom, are representative regions as diagramed in the upper, middle, and bottom panels of Fig. 2.14 respectively.

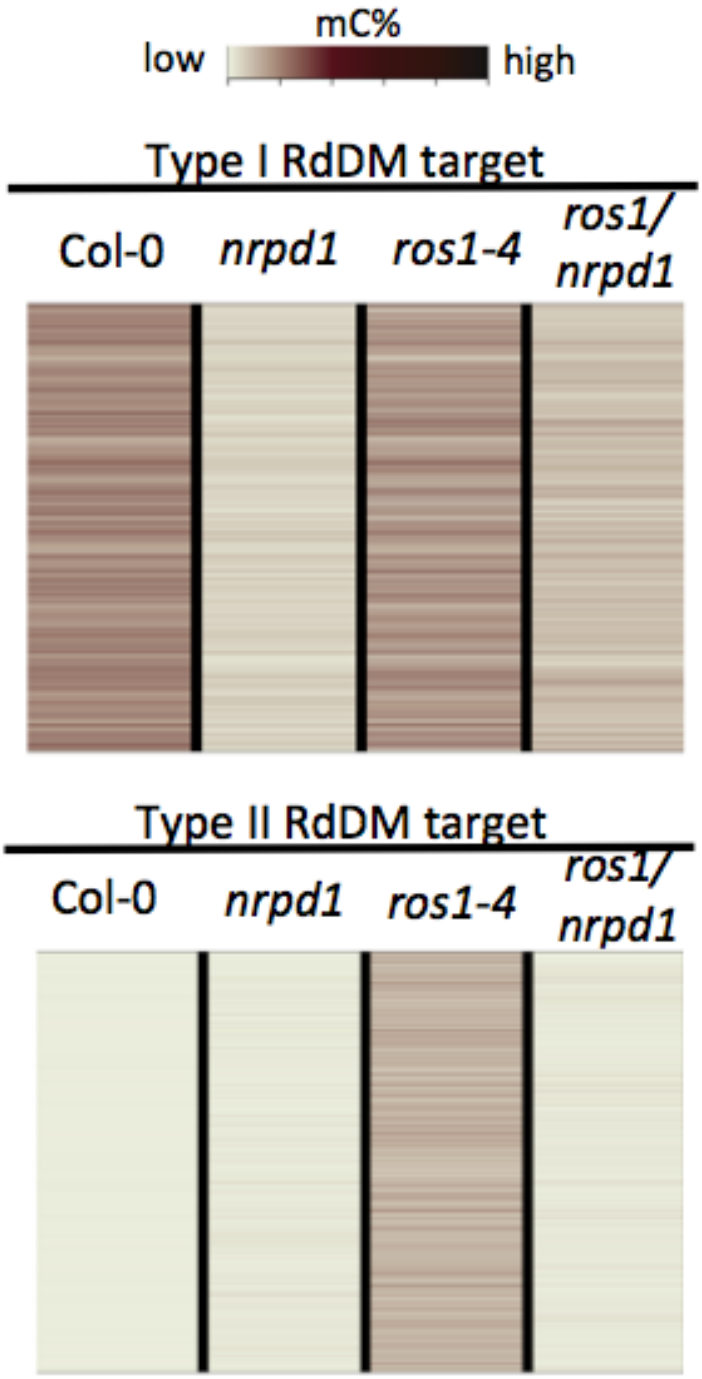

Fig. 2.16. Heat maps of mC methylation in type I and type II targets. Heat maps showing DNA methylation levels of all type I and type II RdDM target loci in Col-0, *nrpd1*, *ros1-4*, and *ros1/nrpd1*.

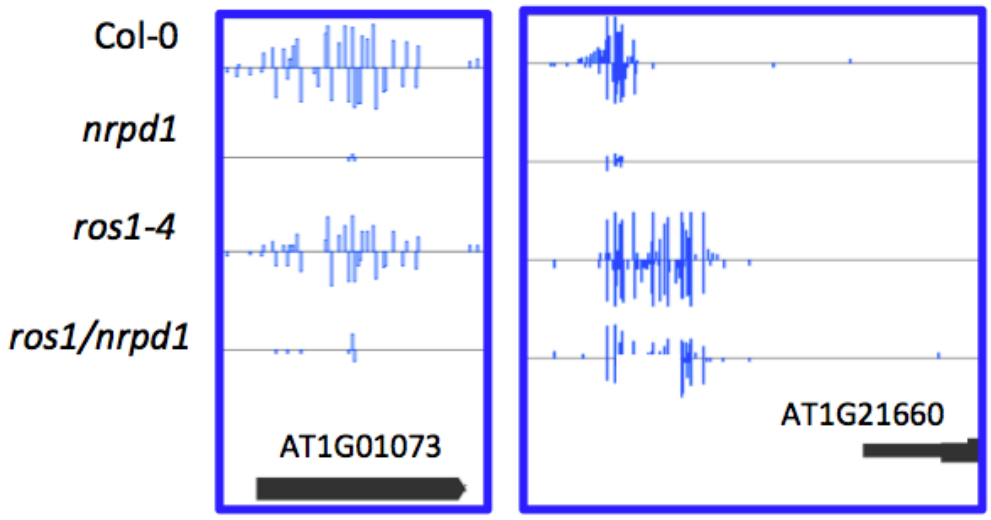

Fig. 2.17. Representative screenshots of type I target. Methylation levels of two type I RdDM targets in Col-0, *nrpd1*, *ros1-4*, and *ros1/nrpd1*. Whole genome bisulfite sequencing data are shown for two type I RdDM targets. The region on the left is the type of RdDM target shown in the upper panel of Fig. 2.14, while the region on the right is the type of RdDM targets shown in the middle panel of Fig. 2.14.

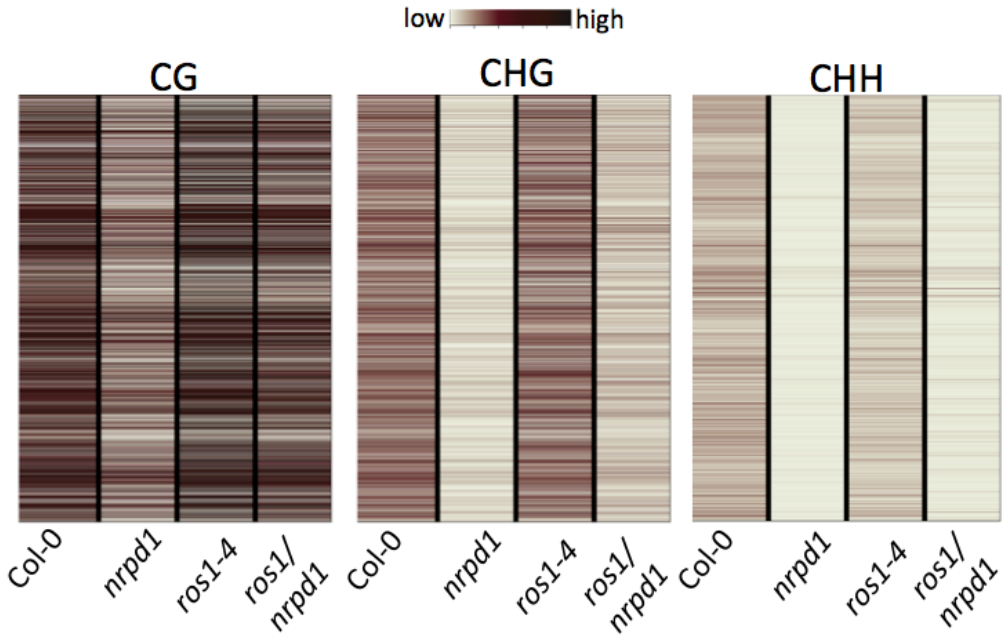

Fig. 2.18. Heat maps of mCG, mCHG, mCHH methylation in type I target. Heat maps showing CG, CHG and CHH methylation levels of all type I RdDM targets in Col-0, *nrpd1*, *ros1-4*, and *ros1/nrpd1*.

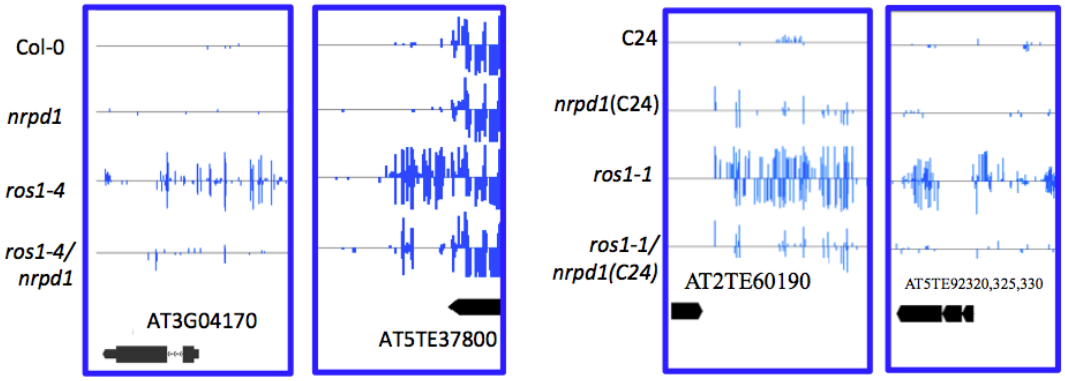

Fig. 2.19. Representative screenshots of type II target. Methylation levels of type II RdDM targets. Methylation levels of two type II RdDM targets in Col-0, *nrpd1*, *ros1-4*, and *ros1/nrpd1* were shown in left panels. Methylation levels of two type II RdDM targets in C24, *nrpd1* (C24), *ros1-1*, and *ros1-1/nrpd1* (C24) were shown in right panels.

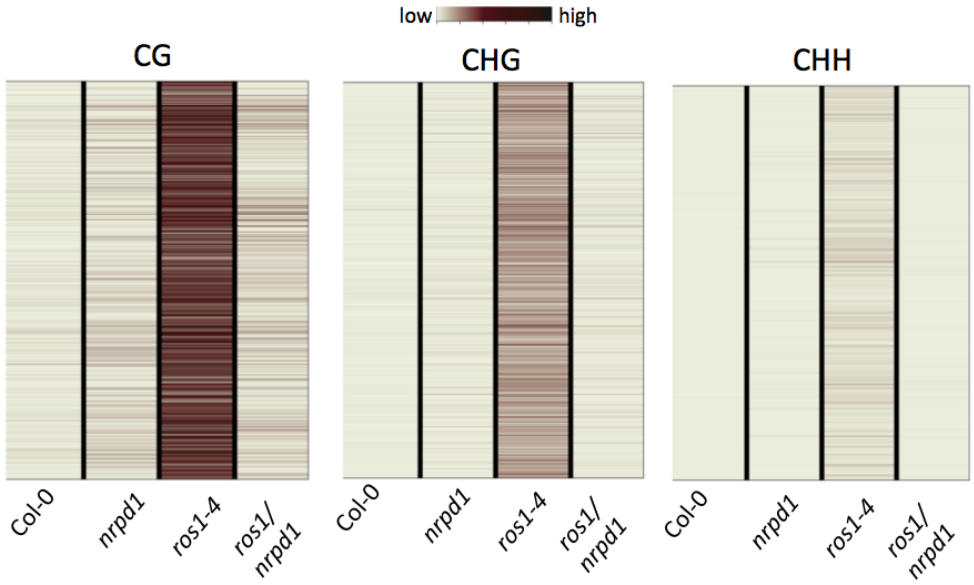

Fig. 2.20. Heat maps of mCG, mCHG, mCHH methylation in type II target. Heat maps showing CG, CHG and CHH methylation levels of all type II RdDM target in Col-0, *nrpd1*, *ros1-4*, and *ros1/nrpd1*, C24, *nrpd1* (C24), *ros1-1*, and *ros1-1/nrpd1* (C24).

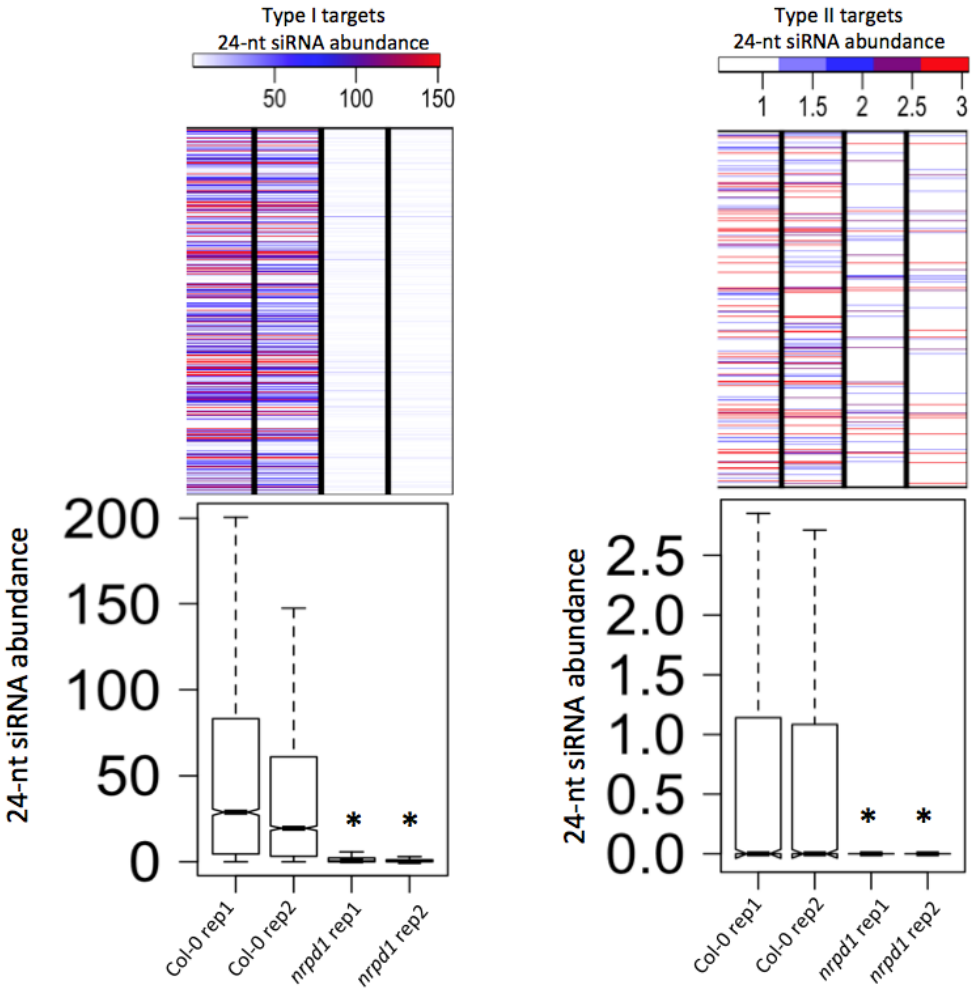

Fig. 2.21. Heat maps and box plots of 24 nt siRNA abundance in Col-0 and *nrpd1*. Heat maps and box plots showing 24 nt siRNA abundance of type I (left panel) and type II (right panel) RdDM targets in Col-0 and *nrpd1* (* $p\text{-}value < 1e-7$, paired two sample t-test)

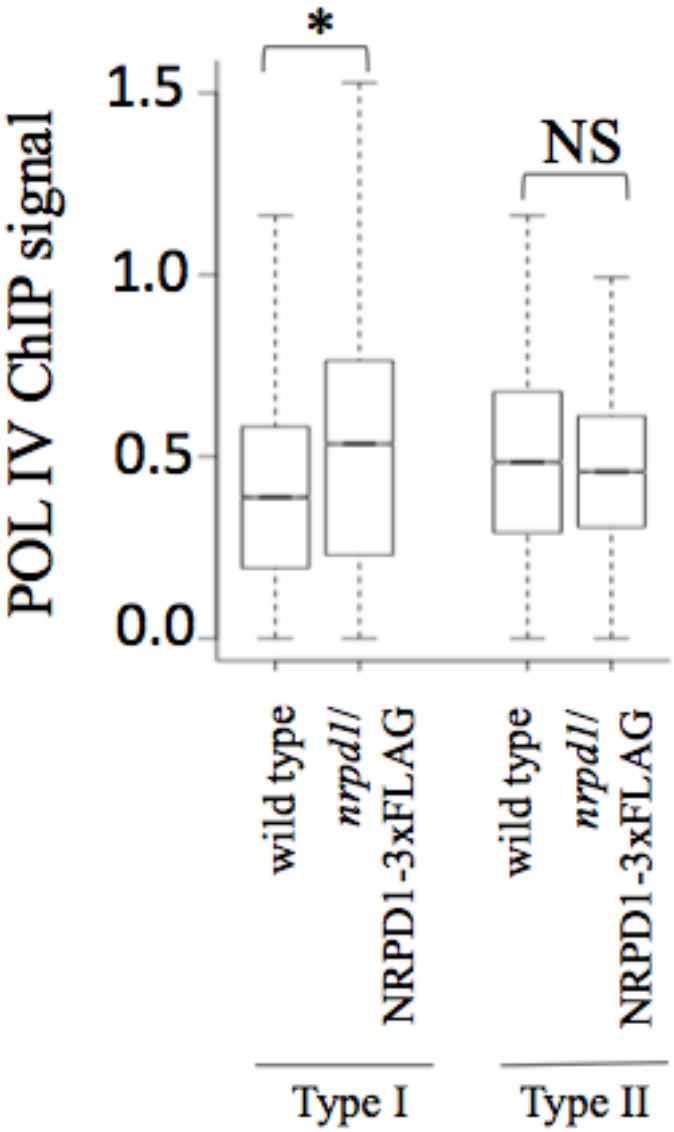

Fig. 2.22. Pol IV enriched at type I loci but not type II. Box plot showing Pol IV enrichment at type I and type II RdDM targets. Pol IV signal in wild type plants served as control. Pol IV is significantly enriched at type I but not type II RdDM targets. * $p\text{-}value < 2.2e-16$; NS, not significant (one-tailed Wilcoxon rank sum test).

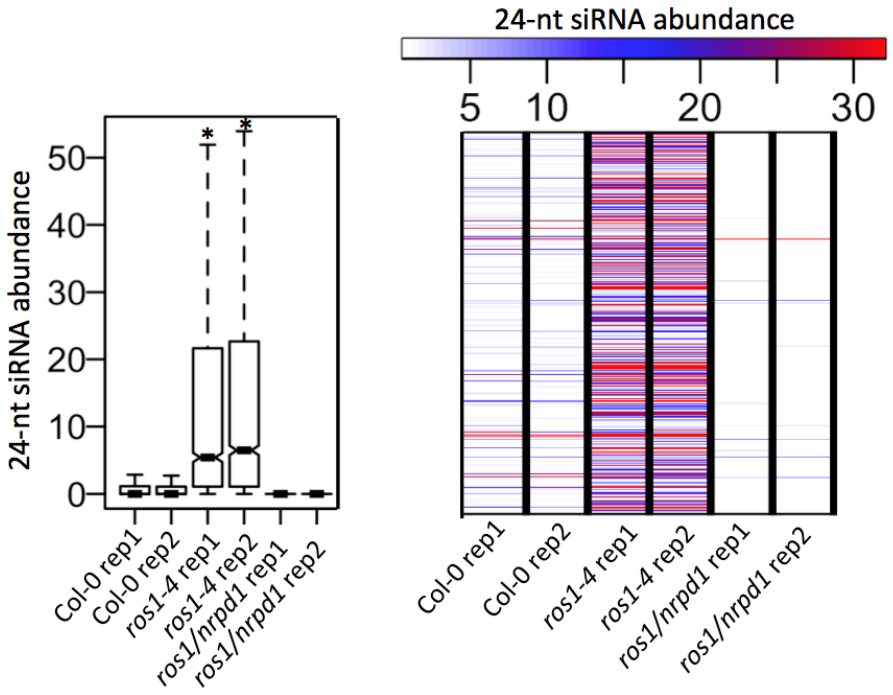

Fig. 2.23. Heat maps and box plots of 24 nt siRNA abundance at type II targets. Box plot and heat map showing 24 nt siRNA abundance of type II RdDM target loci in Col-0, *ros1-4*, and *ros1/nrpd1* (* *p-value* $< 2.2e-16$, paired two-sample t-test).

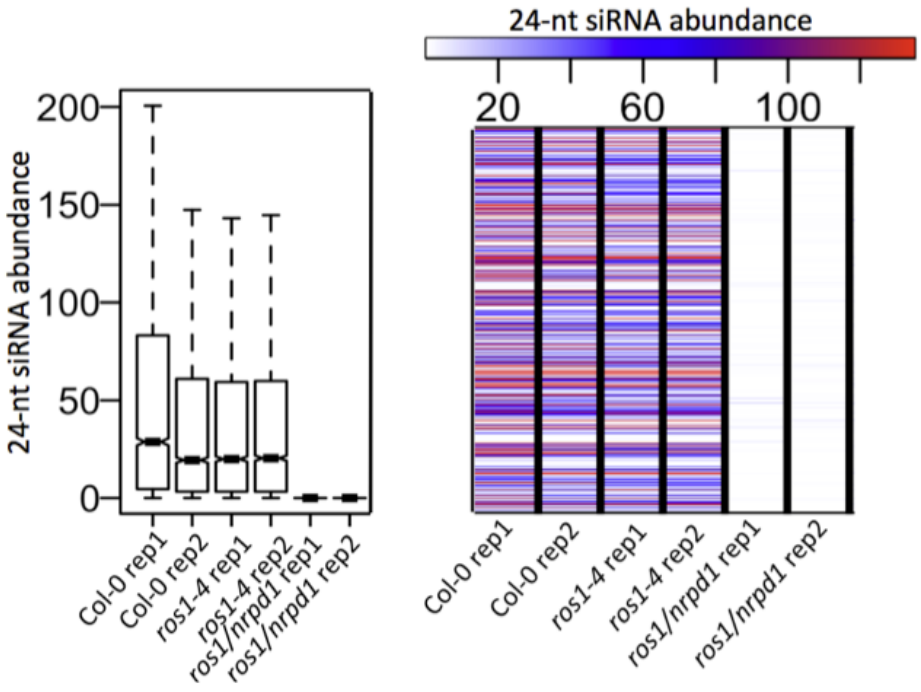

Fig. 2.24. Heat maps and box plots of 24 nt siRNA abundance at type I targets. Box plot and heat map showing 24 nt siRNA abundance of type I RdDM target loci in Col-0, *ros1-4*, and *ros1/nrpd1*.

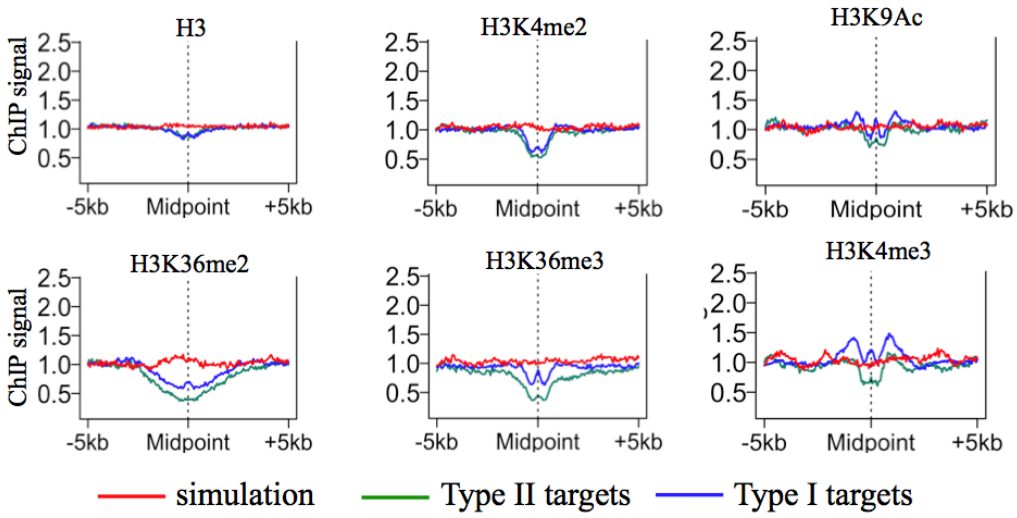

Fig. 2.25. Similar chromatin features shared by type II and type I targets. Association of different histone modifications at regions surrounding the mid-points of type I and type II targets and simulation regions.

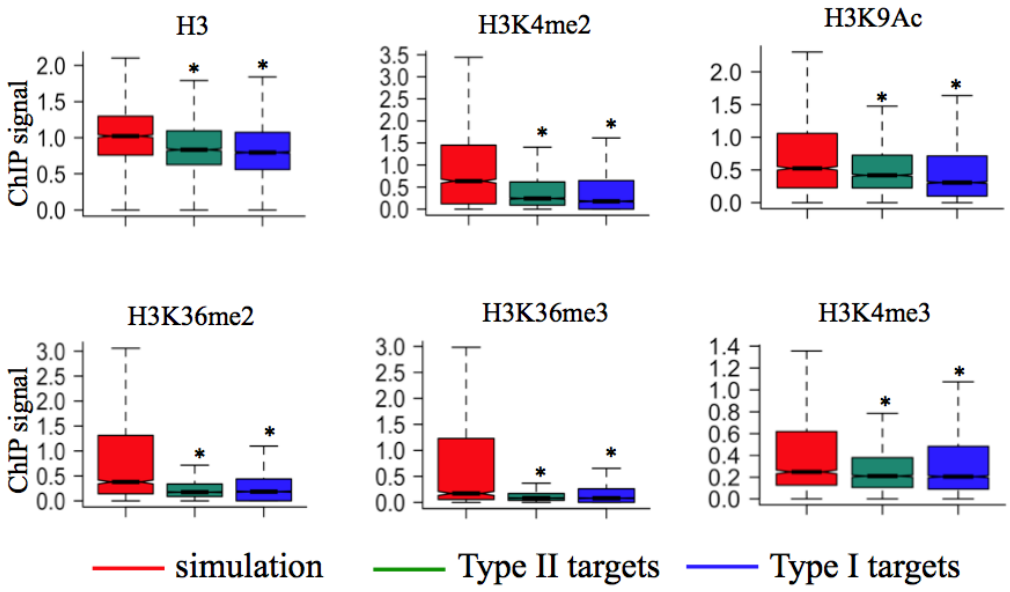

Fig. 2.26. Box plots of similar chromatin features shared by type II and type I targets. Box plots showing the same results as in Fig. 2.25. * *p-value* < 0.005, significantly lower than that of simulation (one-tailed Wilcoxon rank sum test).

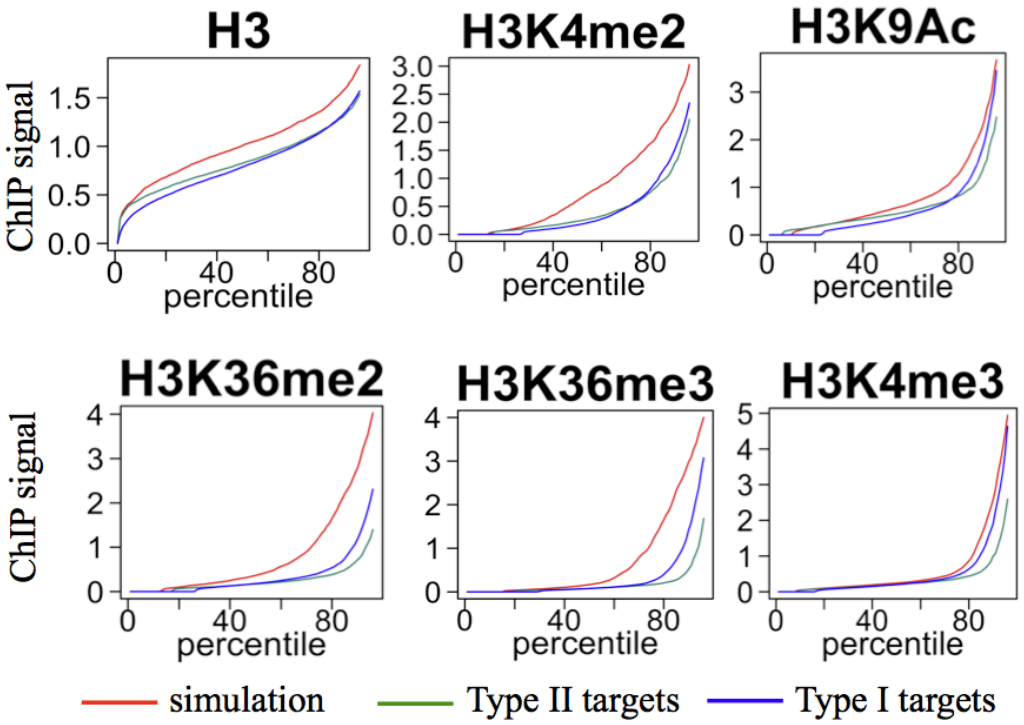

Fig. 2.27. Percentile plots of similar chromatin features shared by type II and type I targets. Percentile plots of the same data as in Fig. 2.26. For each histone mark, simulated regions, type I and type II RdDM targets were ranked based on their histone ChIP signals from low (left) to high (right) along X-axis. X-axis is ranking percentile, and Y-axis is ChIP signal.

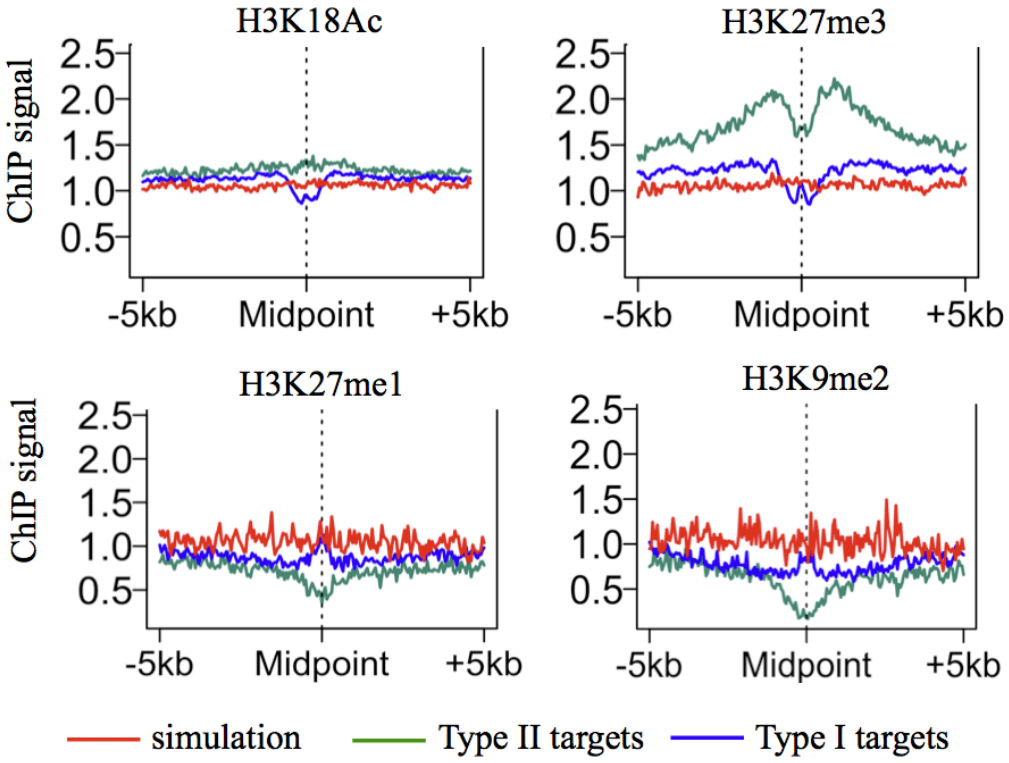

Fig. 2.28. Distinct chromatin features associated with type II and type I targets. Association of different histone modifications at regions surrounding the mid-points of type I and type II targets. Simulation regions served as control regions.

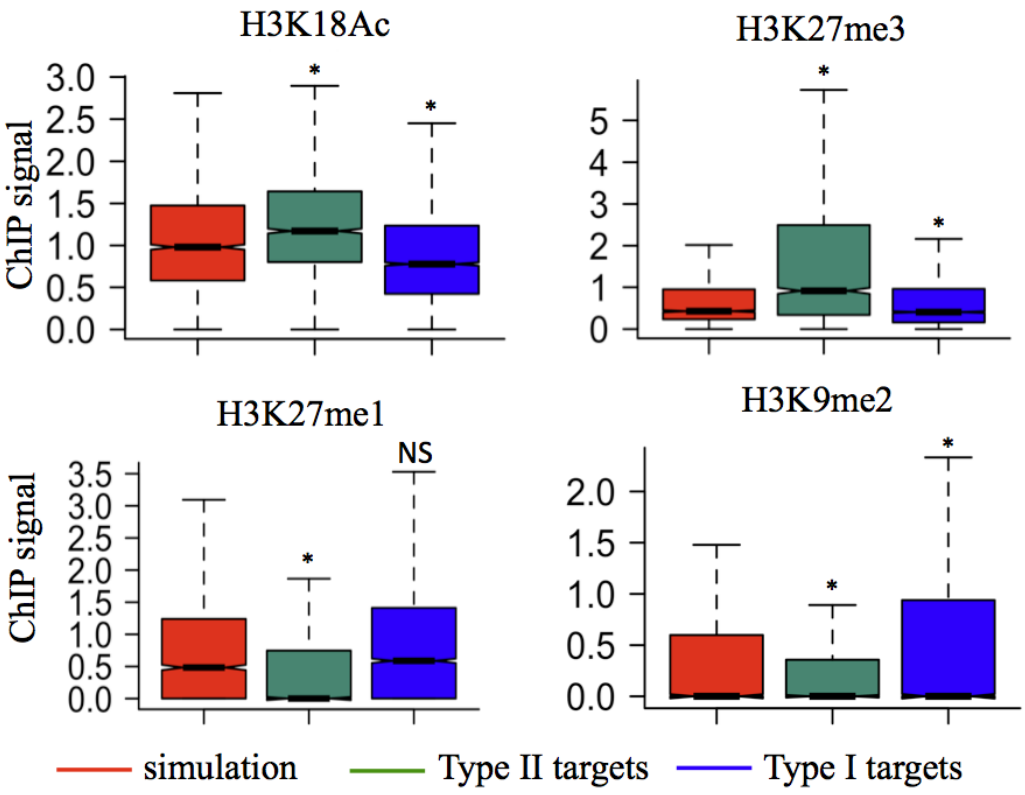

Fig. 2.29. Box plots of distinct chromatin features associated with type II and type I targets. Box plots showing the same results as in Fig. 2.28. * *p-value* < 0.005, NS, not significant compared to simulation (one-tailed Wilcoxon rank sum test).

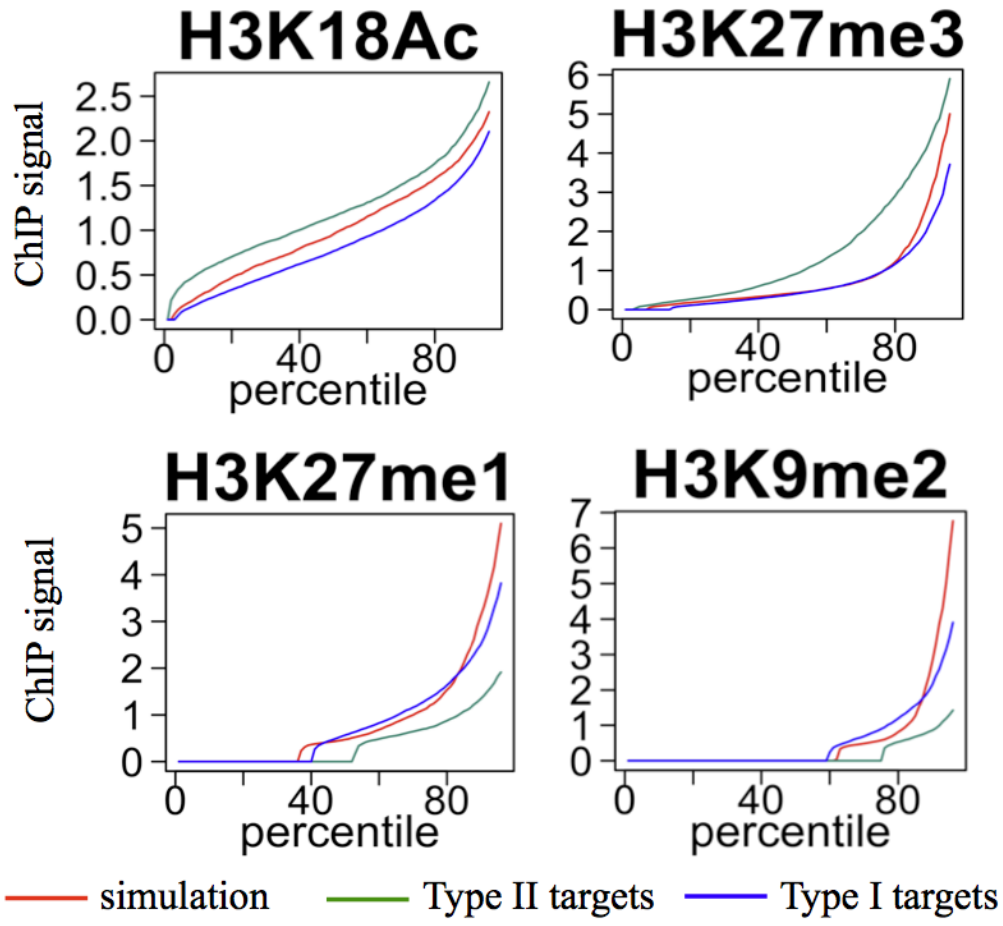

Fig. 2.30. Percentile plots of distinct chromatin features associated with type II and type I targets. Percentile plots of the same data as in Fig. 2.29. For each histone mark, simulated regions, type I and type II RdDM targets were ranked based on their histone ChIP signals from low (left) to high (right) along X-axis. X-axis is ranking percentile, and Y-axis is ChIP signal.

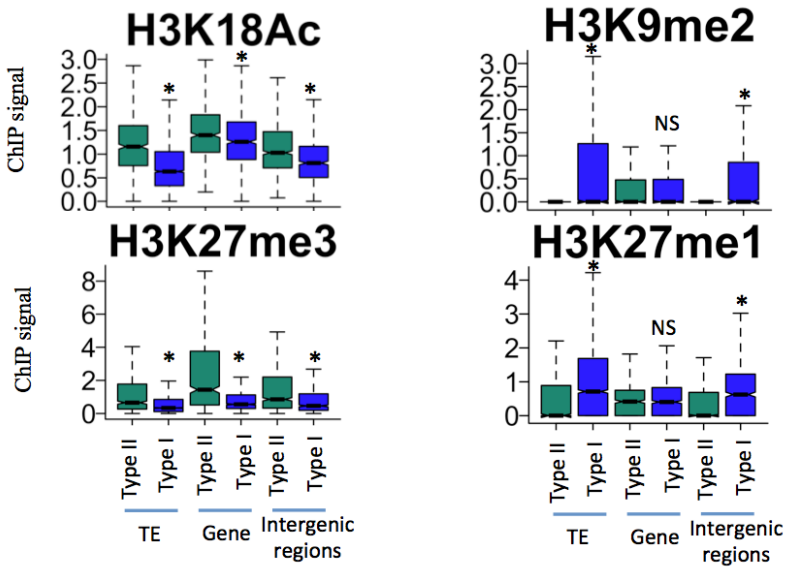

Fig. 2.31. Box plots of histone ChIP signal in three types of genomic regions of type II and type I targets. Box plots showing histone features of different categories of type I and type II RdDM targets (* *p-value* $< 1e-6$, one-tailed Wilcoxon rank sum test; NS, not significant).

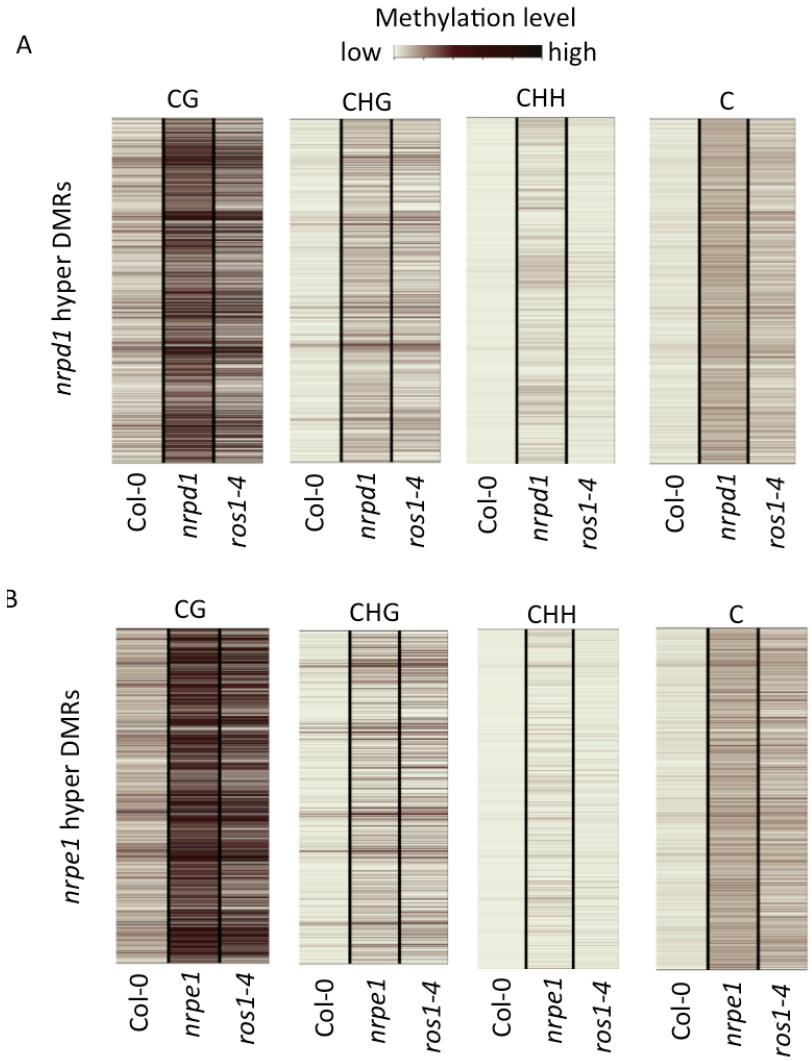

Fig. 2.32. Reduced *ROS1* expression contributes to DNA hypermethylation in RdDM mutants. **A:** Heat maps showing CG,CHG, CHH, and total C methylation levels of *nrpd1* hyper DMRs in Col-0, *nrpd1* and *ros1-4*. **B:** Heat maps showing CG,CHG, CHH, and total C methylation levels of *nrpe1* hyper DMRs in Col-0, *nrpe1* and *ros1-4*.

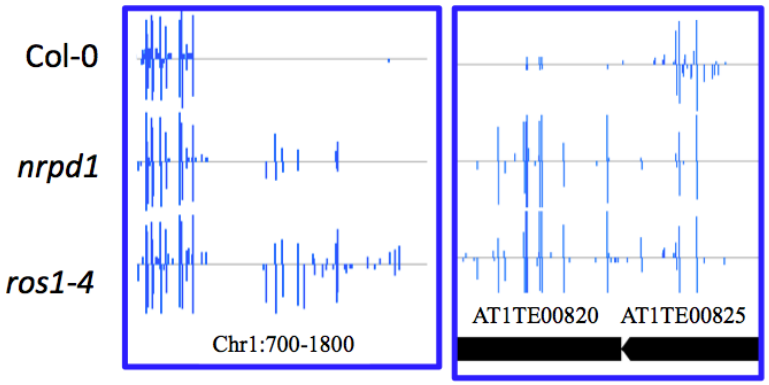

Fig. 2.33. Screenshots of shared hyper DMRs between *nrpd1* and *ros1-4*. Levels of DNA methylation at shared hyper DMRs between *nrpd1* and *ros1-4*. Integrated Genome Browser (IGB) display of whole-genome bisulfite sequencing data is shown in the screenshots. DNA methylation levels of cytosines were indicated with the heights of vertical bars on each track.

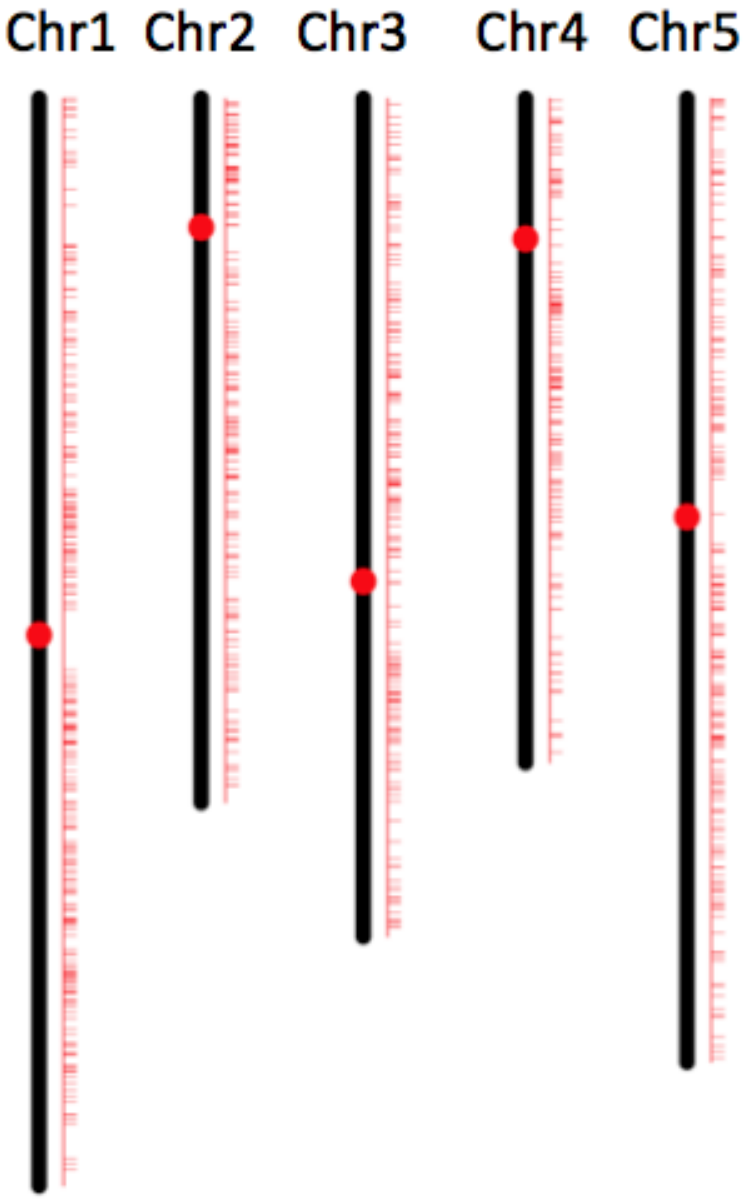

Fig. 2.34. Chromosomal distribution of shared hyper DMRs between *nrpd1* and *ros1-4*. Each red horizontal bar represents a hyper DMR shared between *nrpd1* and *ros1-4*. The red dots indicate the centromeres.

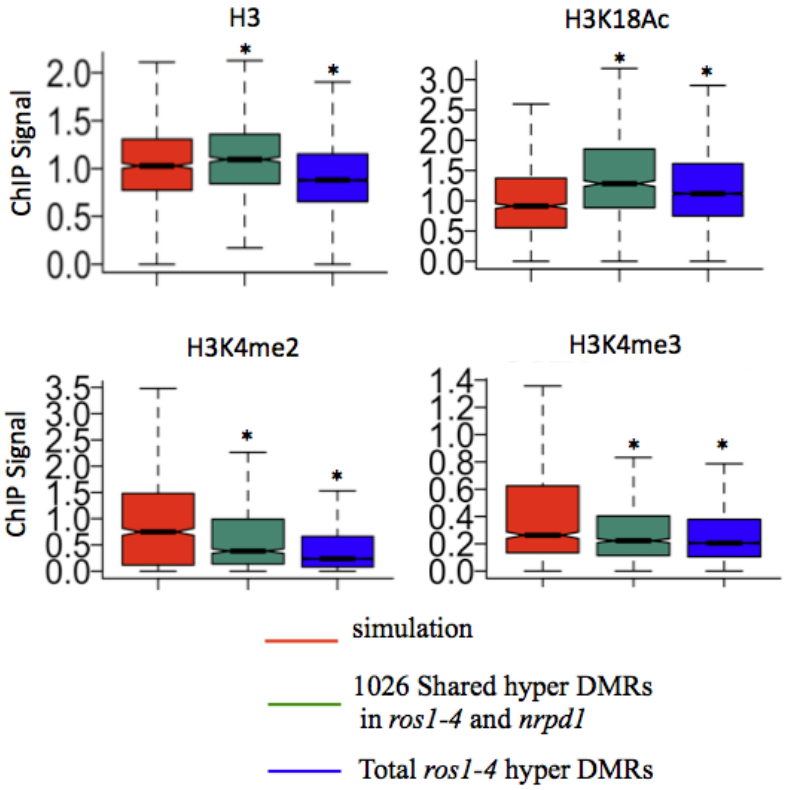

Fig. 2.35. Comparison of different histone modifications between 1026 shared DMRs and *ros1* targets. Box plots display of different histone modifications surrounding 1026 shared hyper DMRs, all *ros1-4* DMRs and simulated regions (* *p-value* < 0.0005, one-tailed Wilcoxon rank sum test).

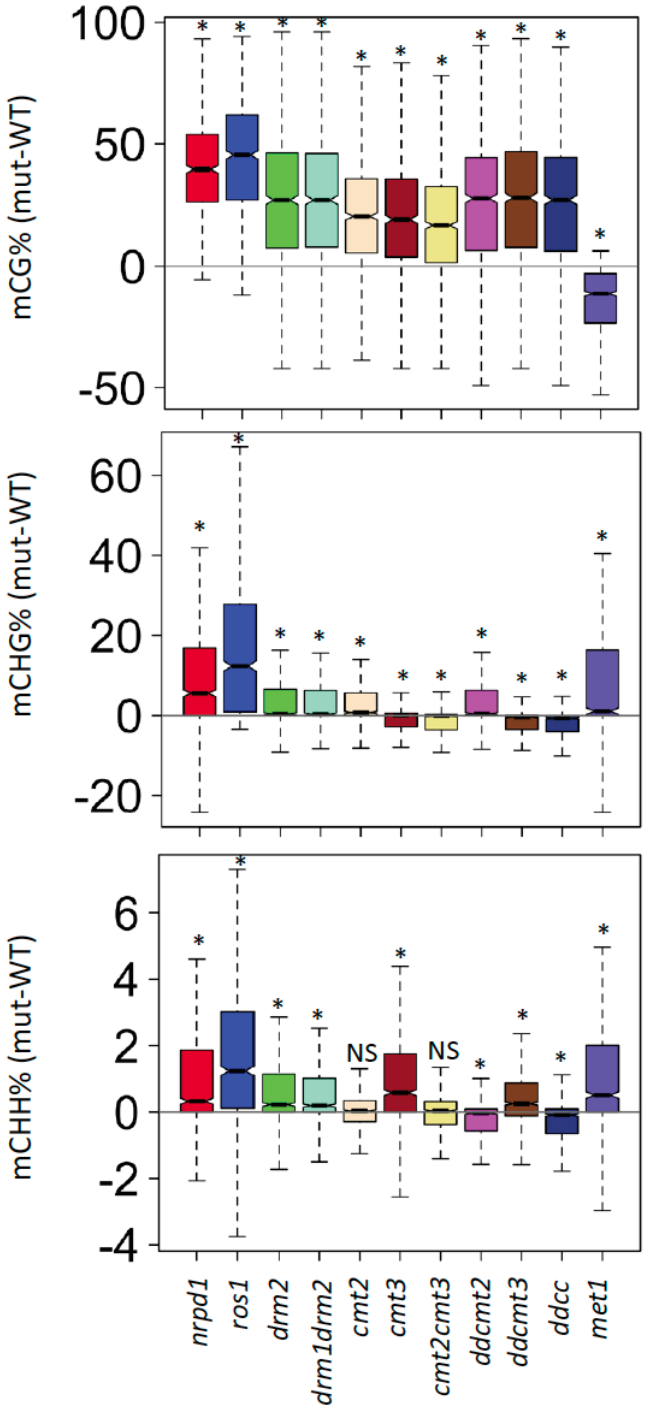

Fig. 2.36. Box plots of methylation changes in 1026 shared DMRs. The 1026 genomic regions with increased DNA methylation levels in both *ros1* and *nrpd1* mutants were used in this analysis. Box plots of CG, CHG, and CHH methylation level changes (mutant - WT) of these regions were shown in different mutants. *ddcmt2* is *drm1drm2cmt2* triple mutant, *ddcmt3* is *drm1drm2cmt3* triple mutant and *ddcc* is *drm1drm2cmt2cmt3* quadruple mutant. (* $p\text{-}value < 1e-10$, one sample one-tailed Student's t-test).

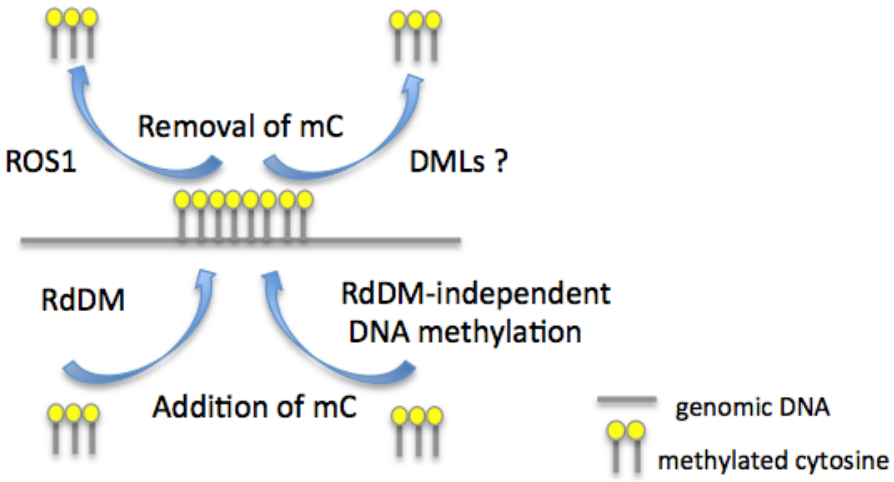

Fig. 2.37. Schematic diagram of DNA regulation in 1026 shared DMRs. Diagram showing that the 1026 shared hyper DMRs are regulated by ROS1, RdDM, and RdDM-independent DNA methylation pathways in wild type, and might be also regulated by DMLs-mediated demethylation pathways.

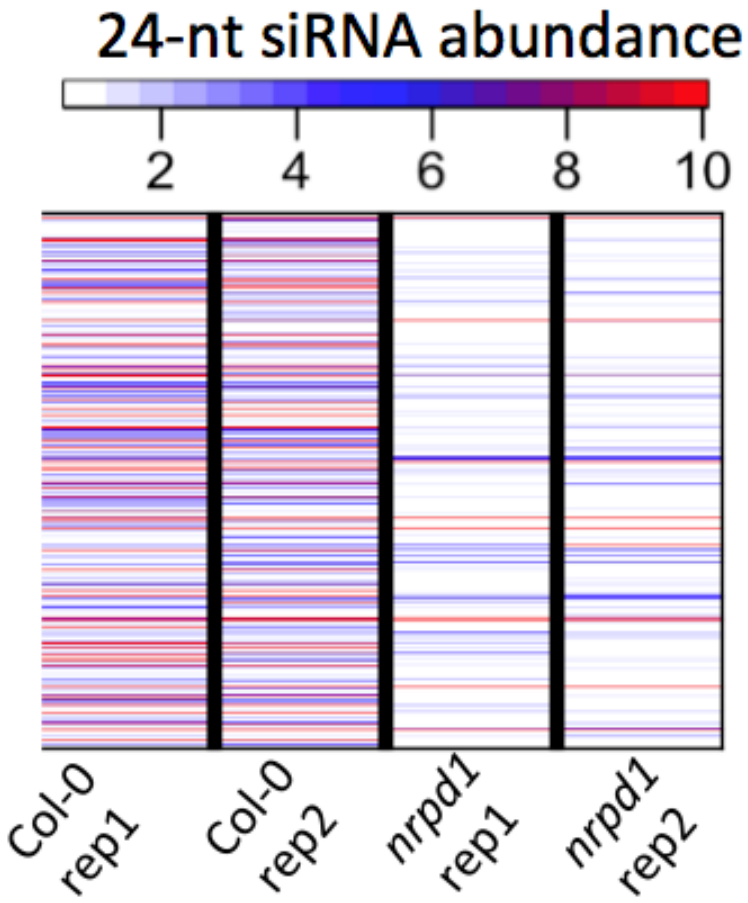

Fig. 2.38. Heat map of 24 nt siRNA abundance in 1026 shared DMRs. Heat map showing 24 nt siRNA abundance of the 1026 shared hyper DMRs in Col-0 and *nrpd1*.

Table 2.1.
The percentiles of the length or ROS1-targeted TEs and all TEs

percentile	**length of *ros1*-TEs (bp)**	**length all-TEs (bp)**
0%	20	10
5%	37	31
10%	58	46
15%	87	63
20%	118	80
25%	154	103
30%	192	133
35%	239	165
40%	292	204
45%	351	251
50%	426	309
55%	514	375
60%	598	453
65%	698	549
70%	797	646
75%	918	780
80%	1091	955
85%	1294	1215
90%	1707	1709
95%	3504	3306
100%	22996	31019

Table 2.2.

Published data used in Chapter 2 study

Data set	**GEO/GSM number**	**Reference**
ros1-1/nrpd1 BS-Seq	GSE44417	Huang *et al.* 2013. PLoS Genetics
histone modifications ChIP data set	GSE28398	Luo *et al.* 2013. Plant J
Pol IV ChIP data set	GSE45368	Law *et al.* 2013. Nature
drm2 BS-Seq	GSM1499354	Zhong *et al.* 2015. PNAS
cmt2 BS-Seq	GSM1242405	Stroud *et al.* 2014. Nat Struct Mol Biol
cmt3 BS-Seq	GSM981003	Stroud *et al.* 2013. Cell
cmt2cmt3 BS-Seq	GSM1242402	Stroud *et al.* 2014. Nat Struct Mol Biol
ddcmt2 BS-Seq	GSM1242403	Stroud *et al.* 2014. Nat Struct Mol Biol
ddcc BS-Seq	GSM1242404	Stroud *et al.* 2014. Nat Struct Mol Biol
met1 BS-Seq	GSM981031	Stroud *et al.* 2013. Cell
drm1drm2 BS-Seq	GSM981015	Stroud *et al.* 2013. Cell
ddcmt3 BS-Seq	GSM981016	Stroud *et al.* 2013. Cell

3. SPECIFIC BUT INTERDEPENDENT FUNCTIONS FOR *ARABIDOPSIS* AGO4 AND AGO6 IN RNA-DIRECTED DNA METHYLATION

3.1 Abstract

Argonaute family proteins are conserved key components of small RNA-induced silencing pathways. In the RdDM pathway in *Arabidopsis*, AGO6 is generally considered to be redundant with AGO4. In this report, our comprehensive, genome-wide analyses of AGO4- and AGO6-dependent DNA methylation revealed that redundancy is unexpectedly negligible in the genetic interactions between AGO4 and AGO6. Immunofluorescence revealed that AGO4 and AGO6 differ in their subnuclear co-localization with RNA polymerases required for RdDM. Pol II and AGO6 are absent from perinucleolar foci, where Pol V and AGO4 are co-localized. In the nucleoplasm, AGO4 displays a strong co-localization with Pol II, whereas AGO6 co-localizes with Pol V. These patterns suggest that RdDM is mediated by distinct, spatially regulated combinations of AGO proteins and RNA polymerases. Consistently, Pol II physically interacts with AGO4 but not AGO6, and the levels of Pol V-dependent scaffold RNAs and Pol V chromatin occupancy are strongly correlated with AGO6 but not AGO4. Our results suggest that AGO4 and AGO6 mainly act sequentially in mediating small RNA-directed DNA methylation.

3.2 Results

3.2.1 Genome-wide profiling of AGO4- and AGO6-dependent DNA methylation

To profile the genome-wide methylation regulated by AGO4 and/or AGO6, we performed whole-genome bisulfite sequencing using *ago4-6* and *ago6-2* mutant plants in the Col-0 background. Similar to the AGO4 mutation, the AGO6 mutation resulted in DNA hypomethylation in mature leaves (Fig. 3.1), even though it was primarily expressed in root and shoot meristems [80,83]. Dysfunction of AGO4 or AGO6 causes DNA hypomethylation at 3678 or 3731 loci, respectively (Fig. 3.2). Both methylome results were validated by individual bisulfite sequencing of a group of randomly chosen hypomethylation loci (Fig. 3.3). In both *ago4-6* and *ago6-2* mutants, DNA hypomethylation was most obvious in the CHH context (Fig. 3.4 and 3.5), which is consistent with CHH methylation being the hallmark of RdDM activities and with these genes functioning in RdDM.

Approximately 80 and 82% of loci with hypomethylated DNA identified in *ago4-6* and *ago6-2* mutant plants, respectively, overlap with loci where DNA methylation is Pol IV dependent (Fig. 3.6). Similar patterns were observed when AGO4 and AGO6 target loci were compared with Pol V target loci (Fig. 3.6). These results further support the functions of AGO4 and AGO6 in the RdDM pathway. *ROS1* gene expression is reduced in many *Arabidopsis* mutants defective in RdDM components [26, 38, 100]. We found that dysfunction of either AGO4 or AGO6 decreases *ROS1* expression (Fig. 3.7), further suggesting that both AGO proteins are required for RdDM. AGO6 dysfunction causes global DNA hypomethylation at loci where DNA methylation is AGO4 dependent (Fig. 3.2 and 3.4). Similarly, AGO4 dysfunction causes DNA hypomethylation at loci where DNA methylation is AGO6 dependent (Fig. 3.2 and 3.5). Even with a stringent filtering criterion (loci that are called "overlap loci" must show $\geqslant$ twofold reduction in methylation levels in both mutants compared to Col-0), both AGO4 and AGO6 are clearly required for DNA methylation

at 2537 loci (Fig. 3.2), which accounts for 68 and 69% of DNA hypomethylation loci identified in *ago4-6* and *ago6-2*, respectively. Therefore, AGO4 and AGO6 mostly act non-redundantly in regulating DNA methylation in the RdDM pathway.

3.2.2 AGO4 and AGO6 predominantly function non-redundantly in the RdDM pathway

To dissect the genetic interactions between AGO4 and AGO6, we generated the *ago4-6 ago6-2* double mutant plants and compared its methylome with those of the single mutants. A total of 4097 loci were identified as hypomethylated in the *ago4-6 ago6-2* double mutant compared to wild-type Col-0 (Fig. 3.8). These hypomethylated loci were categorized into four groups based on DNA methylation patterns: Group I are loci where DNA hypomethylation is observed in the double mutant but not in the single mutants, indicating that DNA methylation at these loci is redundantly regulated by AGO4 and AGO6; Group II are loci where DNA hypomethylation occurs in one of the two single mutants but not in the other, and DNA methylation is not further decreased in the double mutant, indicating that DNA methylation at these loci requires AGO4 or AGO6 specifically; Group III are loci where DNA methylation is similarly reduced in the two single mutants and the double mutant shows no additive effects on DNA hypomethylation, indicating that AGO4 and AGO6 are each required to mediate RdDM at these loci; and Group IV are loci where DNA hypomethylation is observed in the single mutants while the double mutant shows more severe DNA hypomethylation, indicating more complex genetic interactions between the two AGO proteins.

Among the 4097 hypomethylated loci in the *ago4-6 ago6-2* double mutant, only 15 (0.4%) were categorized into Group I (Fig. 3.8, 3.9 and 3.10), indicating that AGO4 and AGO6 function mostly non-redundantly in the RdDM pathway. Group II consists of 191 loci. The numbers of AGO4- and AGO6-specific loci are 66 (1.6%) and 125 (3.1%), respectively (Fig. 3.8, 3.9 and 3.10). Group III consists of 2174 regions

that account for the majority (53.1%) of the 4097 hypomethylated regions (Fig. 3.8, 3.9 and 3.10). AGO4 and AGO6 are each required to confer DNA methylation at Group III loci; thus, AGO4 and AGO6 mainly play distinct but cooperative roles, for instance, with a sequential relationship, in the RdDM pathway. Group IV consists of the remaining 1717 (41.9%) loci, where DNA hypomethylation was observed in *ago4-6*, *ago6-2*, and *ago4-6 ago6-2*, while the double mutant exhibited a further reduction in DNA methylation levels compared to the single mutants (Fig. 3.8, 3.9 and 3.10). Together, these results indicate that AGO4 and AGO6 are mostly mutually dependent, instead of being redundant, in their regulation of DNA methylation.

To examine whether the mutual dependence between AGO4 and AGO6 might be preferentially associated with certain types of genomic loci, we categorized the DNA hypomethylation loci into transposons (TEs), genic regions (Genes), or intergenic regions (IGRs). Among the 4097 hypomethylation loci identified in the *ago4-6 ago6-2* double mutant, 33% (1347) are gene loci, 44% are TE loci, and 23% are IGR loci (Fig. 3.11). Notably, 76% (1002) of the hypomethylated gene loci exhibited the pattern that is characteristic of Group III loci (Fig. 3.11). Among these 1002 loci, 94% loci are DNA methylation targets of Pol IV and/or Pol V (Fig. 3.12). Thus, DNA hypomethylation at these loci indicates defects in AGO4/6-mediated RdDM rather than natural epialleles.

The canonical RdDM pathway involves 24 nt siRNAs that are almost exclusively Pol IV dependent. The Pol IV mutant *nrpd1-3* and the Pol V mutant *nrpe1-11* share 83% of their hypomethylated loci (Fig. 3.13), whereas 82% of hypomethylated loci in the *ago4-6 ago6-2* double mutant overlap with those identified in *nrpd1-3* or *nrpe1-11* (Fig. 3.13). These patterns suggest that AGO4- and AGO6-mediated DNA methylation is predominantly dependent on Pol IV and Pol V function in the canonical RdDM pathway. On the other hand, 62% of hypomethylated loci in *nrpd1-3* or *nrpe1-11* overlap with hypomethylation regions in the *ago4-6 ago6-2* double mutant (Fig. 3.13). Although this pattern supports the importance of AGO4 and AGO6 in RdDM, it also indicates that DNA methylation at some RdDM target loci does not require

AGO4 and AGO6. At loci where DNA methylation is regulated by AGO4 and/or AGO6, simultaneous dysfunction of AGO4 and AGO6 drastically reduces but does not abolish DNA methylation in the CHH context (Fig. 3.14 and 3.15). At many of these loci, *nrpd1-3* and *nrpe1-11* mutants display a further reduction in DNA methylation levels compared to the *ago4-6 ago6-2* double mutant (Fig. 3.14 and 3.15). These results suggest that, besides AGO4 and AGO6, other AGO proteins such as AGO9 may contribute to DNA methylation at these loci.

In addition to regulating 24 nt siRNA-dependent canonical RdDM, AGO4 and AGO6 also mediate a non-canonical RdDM pathway that depends on 21-22 nt tasiRNAs [101], which do not require Pol IV for their production. DNA methylation at the TAS3a locus is tasiRNA dependent and is substantially decreased by mutation of either AGO4 or AGO6 [101]. We consistently observed that AGO4 and AGO6 mutations both reduce DNA methylation at the TAS3a locus (Fig. 3.16). In addition to TAS3a, TAS1a and TAS1c are two other loci that are subjected to tasiRNA-dependent DNA methylation [101]. We observed that AGO6 mutation strongly depletes DNA methylation at TAS1a and TAS1c, whereas AGO4 mutation causes only a slight decrease in DNA methylation at these two loci (Fig. 3.16). Thus, AGO6 has a more prominent role in mediating tasiRNA-dependent non-canonical RdDM than AGO4.

The non-redundant relationship between AGO4 and AGO6 was also observed at the level of transcriptional silencing. Dysfunction of either AGO4 or AGO6 relieves transcriptional silencing at several transposon loci tested, including AtSN1, AtGP1, IGN5, and LINE1-4 (Fig. 3.17). Derepression occurs to different degrees at some loci in *ago4-6* and *ago6-2* single mutants (Fig. 3.17), but is not additive in the *ago4-6 ago6-2* double mutant (Fig. 3.17). Therefore, AGO4 and AGO6 function non-redundantly to promote transcriptional silencing of a subset of loci.

3.2.3 Subnuclear spatial segregation of AGO4 and AGO6 in mediating RdDM

Co-localization of distinct RdDM components can be observed within nucleoplasmic and perinucleolar foci [74–77] . We therefore asked whether AGO4 and AGO6 display similar subnuclear localization patterns in *Arabidopsis* seedlings. Immunostaining of AGO4 and AGO6 in the same nuclei revealed spatial segregation, as indicated by the absence of co-localized signals (n = 105) (Fig. 3.18A). Production of non-coding RNAs for RdDM involves the RNA polymerases Pol IV, Pol V, and Pol II. Biogenesis of siRNAs depends predominantly on Pol IV, which does not affect scaffold RNA levels [70] . In all cells examined (n $\geq$ 112), neither AGO4 nor AGO6 exhibited co-localization with NRPD1, the largest subunit of Pol IV (Fig. 3.18B). These results are consistent with the inference that AGO proteins function downstream of Pol IV-dependent siRNA production. Pol V not only transcribes scaffold RNAs that recruit AGO-siRNA complexes, but also reinforces production of some siRNAs [102] . AGO6 displays partial co-localization with NRPE1, the largest subunit of Pol V, in the nucleoplasm but not in the perinucleolar foci (Fig. 3.19A). Consistent with previous reports [74, 76] , we observed that AGO4 signals overlapped with NRPE1 signals almost exclusively in the perinucleolar foci (Fig. 3.19B), where other key RdDM components such as RDR2 and DCL3 have also been observed [74–77].

In addition to Pol IV and Pol V, Pol II can produce scaffold RNAs at some intergenic low-copy-number repeat loci [71] and can also initiate siRNA production at inverted repeats [103,104]. Pol II localizes in the nucleoplasm but not at perinucleolar foci(Fig. 3.20A) [75]. Pol II was previously shown to co-localize with AGO4 in the nucleoplasm [75]. We also consistently observed a strong co-localization between AGO4 and Pol II in the nucleoplasm (Fig. 3.20A) . In contrast, Pol II does not co-localize with AGO6 (Fig. 3.19B), which displays extensive co-localization with Pol V in the nucleoplasm but not at perinucleolar foci (Fig. 3.19B). To test whether AGO4-Pol II co-localization is associated with Pol II's function in mRNA transcription, we

examined subnuclear localization patterns of TBP, a transcription factor that binds to the TATA box in gene promoters and is thereby indicative of mRNA transcription in TBP-dependent Pol II promoters [105] . AGO4 signals did not overlap with signals of TBP (Fig. 3.20B), but TBP-Pol II overlapped signals were observed (Fig. 3.20C), suggesting that, at least at those TBP-dependent promoters, the co-localization between AGO4 and Pol II is not associated with mRNA transcription. Further, the largest subunit of Pol II, NRPB1, co-immunoprecipitates with AGO4 but not AGO6, which supports a physical interaction between Pol II and AGO4 and is consistent with the immunostaining results (Fig. 3.20C). Together, these data reveal a subnuclear spatial segregation of AGO4 and AGO6 in mediating RdDM. Because AGO4 and AGO6 are simultaneously required for DNA methylation at most of the RdDM target loci, these results further suggest the existence of spatially separated yet cooperative RdDM steps in the nucleus.

3.2.4 Functional divergence between AGO4 and AGO6 in regulating siRNAs and scaffold RNAs

Pol V physically interacts with AGO proteins and synthesizes scaffold RNAs that recruit complementary siRNAs. Thus, it is critical for the recruitment of AGO-siRNA complexes. To examine the potential effects of AGO proteins on Pol V-mediated RdDM steps, we quantified scaffold RNA levels in mutants defective in AGO4 or AGO6. We evaluated RdDM loci known to produce Pol V-dependent non-coding RNAs and found that the *ago6-2* mutant displayed reduced levels of Pol V transcripts relative to wild-type plants (Fig. 3.21 and 3.22). In addition, we performed chromatin immunoprecipitation for Pol V and observed reduced occupancy at these loci in *ago6-2* mutant plants compared to wild-type (Fig. 3.23). In contrast, AGO4 dysfunction does not reduce Pol V transcript levels at the examined loci except the AtSN1 locus, where Pol V transcripts were moderately decreased in both *ago4-6* and *ago6-2* (Fig. 3.21). Meanwhile, AGO4 dysfunction does not affect Pol V occupancy in chromatin

as examined at the same loci including AtSN1 (Fig. 3.23). These results further support the functional divergence between AGO4 and AGO6 and the hypothesis that the two AGO proteins function non-redundantly in the RdDM pathway.

Although most siRNAs are Pol IV dependent, the production of some siRNAs is reinforced by Pol V [68, 102]. We wanted to determine whether AGO4 or AGO6 may preferentially bind to siRNAs that are Pol V dependent. AGO4- and AGO6-bound siRNAs have been identified previously [83]. We retrieved these siRNAs and examined their levels in our *nrpd1-3* and *nrpe1-11* whole-genome, small RNA sequencing datasets. As expected, Pol IV dysfunction decreases the levels of almost all AGO4- and AGO6-bound siRNAs (Fig. 3.24), consistent with the inference that both AGO4 and AGO6 function downstream of Pol IV. In contrast, only about 44% of AGO4- or AGO6-bound siRNAs showed reduced levels in *nrpe1-11* (Fig. 3.25), indicating that neither AGO4 nor AGO6 preferentially associates with Pol V-dependent siRNAs.

In addition to guiding DNA methylation by pairing siRNAs with scaffold RNAs, AGO4 can also contribute to siRNA production through its endonuclease activity within the PIWI domain [45]. We consistently observed decreased siRNA levels in *ago4-6* at several RdDM loci (Fig. 3.26), where individual 24 nt siRNAs can be quantitatively detected by RT-PCR [98]. The *ago6-2* mutant also showed reduced 24 nt siRNA levels at the examined loci (Fig. 3.26), indicating that AGO6 contributes to siRNA production as well. The *ago4-6 ago6-2* double mutant displayed an additive effect on siRNA accumulation at the At1TE40810 locus (Fig. 3.26); such an additive effect was not observed at other examined RdDM loci (Fig. 3.26). Together, these observations are consistent with non-redundant roles of AGO4 and AGO6 in regulating siRNA production at the majority of RdDM target loci.

3.3 Discussion

AGO4 and AGO6 have been considered redundant in regulating RNA-directed DNA methylation. In this study, we performed genome-wide quantitative analyses of

DNA methylation in *Arabidopsis* mutants defective in AGO4, AGO6, or both. Our results revealed that redundancy is unexpectedly negligible (0.4%) between AGO4 and AGO6. At most of their target loci, AGO4 and AGO6 are each required to confer DNA methylation, indicating that the two AGO proteins have distinct yet cooperative functions. Consistent with these proteins functioning non-redundantly, AGO6 and AGO4 display differential effects on Pol V-dependent scaffold RNA levels and Pol V occupancy. In addition, subnuclear localization patterns of AGO4 and AGO6 provide support for spatially segregated activities in the RdDM pathway. AGO4 and AGO6 differ in their subnuclear co-localization with Pol II and Pol V. In perinucleolar foci where Pol V and AGO4 show co-localization, Pol II and AGO6 are absent. In the nucleoplasm where AGO4 co-localizes with Pol II, AGO6 preferentially co-localizes with Pol V. These patterns suggest spatially segregated RdDM activities, which appear to be mediated through different AGO-Pol combinations (Fig. 3.27).

AGO4 and AGO6 have different expression patterns. A GUS reporter gene under the control of the AGO4 promoter (P_{AGO4}:GUS) showed ubiquitous expression in the embryo and in mature leaves [83]. In the embryo, PAGO6:GUS expression was concentrated in the shoot and root apical meristems and the vascular tissues, whereas PAGO6:GUS expression was not observed in mature leaves [80, 83]. Nevertheless, AGO6 dysfunction resulted in DNA hypomethylation in leaves at transgenic reporter genes [80] and at endogenous RdDM target loci (Fig. 3.1 and 3.3). AGO4 protein levels are reduced in mutants defective in Pol IV, RDR2, or DCL3, which are RdDM regulators upstream of AGO proteins [72, 76]. AGO6 dysfunction does not decrease AGO4 protein level (Fig. 3.28) [83]. Thus, DNA hypomethylation caused by AGO6 dysfunction is not likely mediated through the down-regulation of AGO4 in leaves. Because meristems are the hubs of AGO6 gene expression, it is possible that DNA hypomethylation in the mature leaves of *ago6* mutant is a consequence of hypomethylation in the shoot apical meristem, from which aerial portions develop. Analysis of several commonly studied RdDM loci revealed similar DNA methylation levels between *ago4-5* and *ago4-6* (Fig. 3.29), indicating that, at least at the exam-

ined loci, *ago4-6* is not a weaker allele compared to *ago4-5*. When comparing *ago4-5* and *ago6-2* effects on RdDM, Stroud [88] examined DRM1/2-dependent loci where DNA methylation showed dependence on both AGO4 and AGO6; DNA methylation patterns were examined in the CHH context within 100-bp genomic segments. We re-analyzed *ago4-5* and *ago6-2* methylomes [88] by using the same analysis pipeline described in this study. The re-analysis showed that DNA methylation at 1636 genomic loci, of varying sizes up to 2065 bp, is dependent on both AGO4 and AGO6. The majority (62%) of these loci displayed similar cytosine (including CG, CHG, and CHH) methylation levels between *ago4-5* and *ago6-2*. Loci where AGO4 and AGO6 are mutually dependent in regulating DNA methylation are characterized by the following: (1) *ago4* and *ago6* mutants show similar reduction in DNA methylation levels, and (2) *ago4ago6* double mutant does not show additive effects compared to the single mutants. Therefore, the results of re-analysis of Stroud *et al.* [88] are consistent with our observation of the interdependence between AGO4 and AGO6, although an *ago4-5 ago6-2* double mutant is unavailable for further analysis.

The function of AGO proteins in RdDM may be affected by Pol V in different ways, because Pol V controls RdDM in two ways, that is, by contributing to the transcription of scaffold RNAs and by reinforcing the production of some siRNAs. Our bioinformatics analysis revealed that AGO6-bound siRNAs are not preferentially Pol V dependent. AGO4 interacts with Pol V [72, 83] and co-localizes with Pol V at perinucleolar foci [76] where the siRNA biogenesis proteins RDR2 and DCL3 were also observed [74]. However, like AGO6, AGO4 does not preferentially bind Pol V-dependent siRNAs, as only 43% of AGO4-bound siRNAs were down-regulated by Pol V dysfunction. Therefore, the dependence of siRNAs on Pol V does not distinguish the roles of AGO4 and AGO6 in the RdDM pathway. Except at the AtSN1 locus, we also observed that dysfunction of AGO6, but not AGO4, partially decreases Pol V-dependent transcripts that can serve as scaffold RNAs to recruit AGO-siRNA complexes. It thus appears that AGO6 is more tightly connected with Pol V function than AGO4. Consistent with this notion, dysfunction of Pol V decreases the protein

level of AGO6 but not of AGO4 (Fig. 3.28) [83], while AGO6 displayed a stronger physical association with Pol V than AGO4 [83].

AGO4 and AGO6 may act sequentially to mediate siRNA-guided DNA methylation at the majority of RdDM target loci. Perhaps, when loaded with siRNAs, one of the two AGOs guides the formation of a heterochromatic histone mark, and then, the other AGO can guide DNA methylation. It is possible that AGO4-siRNA guides histone modification through association with Pol II-generated scaffold RNAs. The resulting histone mark may allow Pol V to be recruited to generate new scaffold RNAs to pair with AGO6-bound siRNAs. Through the combined actions of AGO4-siRNA and AGO6-siRNA, DRM2 may then be recruited to trigger DNA methylation. These actions would all take place in the nucleoplasm. At the perinucleolar foci where AGO4 is co-localized with Pol V, AGO4 may also directly guide DNA methylation. Future experiments to test this and other models should help us understand how siRNAs promote DNA methylation.

3.4 Material and methods

3.4.1 Plant materials and growth conditions

Arabidopsis was grown at 23°C and with 16-h light/8-h dark. T-DNA insertion lines salk_071772 (*ago4-6*) and salk_031553 (*ago6-2*) were ordered from the SALK Institute Genomic Analysis Laboratory. The native promoter-driven FLAG-tagged AGO4 and AGO6 transgenic plants were obtained from the laboratory of Dr. David Baulcombe.

3.4.2 Whole-genome bisulfite sequencing and analysis

DNA was extracted from 12-day-old seedlings and sent to BGI (Shenzhen, China) for bisulfite treatment, library preparation, and sequencing. Bisulfite sequencing libraries were prepared using standard Illumina protocols. The brief pipeline of BGI

is as follows: (1) fragment genome DNA to 100-300 bp by sonication; (2) DNA-end repair, 3-dA overhang and ligation of methylated sequencing adapters; (3) bisulfite treatment by ZYMO EZ DNA Methylation-Gold kit; (4) desalting, size selection, PCR amplification and size selection again; and (5) qualified library for sequencing. The bisulfite conversion rate of Col-0 and the five mutants used in this study are Col-0 99.37%, *ago4-6* 99.63%, *ago-2* 99.61%, *ago4-6 ago6-2* 99.59%, *nrpd1-3* 99.51% and *nrpe1-11* 99.32%.

For data analysis, adapter and low-quality sequences ($q < 20$) were trimmed, and clean reads were mapped to the *Arabidopsis* genome (TAIR10) using BRAT-BW [93] and allowing two mismatches. DNA hypomethylated regions were identified according to Ausin [94] with minor modifications. In brief, only cytosines with 4 coverage in all libraries were considered. A sliding-window approach with a 200-bp window sliding at 50-bp intervals was used to identify DMRs. Fisher's exact test was performed for methylated versus unmethylated cytosines for each context, within each window, with FDRs estimated using a Benjamini-Hochberg adjustment of Fisher's P-values calculated in the R environment. Windows with an FDR $\leqslant 0.05$ 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 (DMC) at each border. A cytosine was considered differentially methylated if it showed at least a twofold reduction in methylation percentage in the mutant. The regions were then filtered to include only those with at least 10 DMCs and with at least an average of a twofold reduction in methylation percentage per cytosine.

To dissect genetic interactions between AGO4 and AGO6, the 4097 hypomethylated loci in *ago4-6 ago6-2* were categorized into four groups based on the DNA methylation (meC) levels in the wild-type, *ago4-6*, *ago6-2*, and *ago4-6 ago6-2* following these steps: (1) loci in which reduction of meC in both *ago4-6* and *ago6-2* is $< 25\%$ were categorized as Group I loci; (2) the remaining loci were then filtered for AGO4-specific loci (meC reduction in *ago4-6* $\geqslant 25\%$ while meC reduction in *ago6-2* $< 25\%$) and AGO6-specific loci (meC reduction in *ago6-2* $\geqslant 25\%$ while meC reduction

in *ago4-6* $< 25\%$), which are collectively categorized into Group II; (3) then, Group III loci were defined by "meC reduction in *ago4-6 ago6-2* $\geqslant$ 125% meC reduction in either *ago4-6* or *ago6-2*" and "the ratio of meC reduction in *ago4-6* to meC reduction in *ago6-2* is between 75 and 125%"; and (4) the loci remaining from step 3 were then assigned to Group IV. To compare *ago* mutants with *nrpd1-3* and *nrpe1-11*, we re-analyzed methylome data of *nrpd1-3* and *nrpe1-11* [106]. A DNA hypomethylated region was categorized as "TE" if it overlapped ($\geqslant$ 1 bp) a TE. A DNA hypomethylation region was categorized as "gene" if it overlapped a gene and did not overlap any TE. An intergenic region was categorized as "IGR" if it did not fall into the groups of "TE" or "gene". For TE annotation, we used both ATxTE and ATxG numbers. If a region overlapped with ATxTE, it was classified as TE region. If a region did not overlap with ATxTE but overlapped with ATxG which was annotated as transposable element gene, it also was classified as TE region. Regions having no overlap with ATxTE or transposable element gene (ATxG) and overlapping with protein coding genes were classified as gene region.

3.4.3 Individual bisulfite sequencing

Individual bisulfite sequencing was performed as described previously [35].In brief, 2-week-old seedlings were collected for genomic DNA extraction using DNeasy Plant Mini Kit (Qiagen). Purified DNA was subjected to bisulfite conversion reaction using BisulFlash DNA Modification Kit (EPIGENTEK) according to manufacturer's protocol. Bisulfite-converted DNA was used as template to amplify target loci, and PCR products were cloned into pGEM-T easy vector (Promega) for DNA sequencing. For each locus, 15-20 clones were selected for sequencing.

3.4.4 Immunostaining analysis

To prepare nuclei, 4 g of 2-week-old seedings grown on 1/2 MS plates was chopped, and nuclei were isolated as previously described [74]with minor modification. The

slides were first fixed in 4% formaldehyde/PBST buffer for 30 min at room temperature and washed three times with PBST. Slides were then blocked in 1% BSA/PBST at 37C for 30 min. Slides were then exposed to primary antibody in blocking solution overnight at 4C. After they were washed three times with PBST, slides were incubated with Alexa Fluor fluorescent-labeled secondary antibody. The nuclei were counterstained by ProLong Gold Antifade Reagent with DAPI (P36931, Invitrogen). For double immunostaining, slides were repeatedly blotted with primary and secondary antibodies. Images of nuclear protein localization were captured with a Nikon A1R MP confocal microscope. Images were analyzed using NIS-Elements Ar Microscope Imaging Software and processed by Adobe Photoshop (Adobe Systems). The primary antibodies used in this study included monoclonal antibodies of anti-AGO4 (1:200, Agrisera), anti-AGO6 (1:200, Agrisera), anti-H3K9me2 (1:400, Ab1220, Abcam), anti-H3K27me3 (1:400, Abcam), anti-NRPD1 (1:200, rabbit antibody, a gift from Craig S. Pikaard), and NRPE1 (1:200, rabbit antibody, a gift from Craig S. Pikaard). The following secondary antibodies were diluted at a ratio of 1:400: Alexa fluor 488 goat anti-rabbit IgG (A11008, Invitrogen), Alexa fluor 488 goat anti-mouse IgG (A11001, Invitrogen), Alexa fluor 568 goat anti-mouse IgG (H+L) (A11004, Invitrogen), and Alexa fluor 568 donkey anti-rabbit (A10042, Invitrogen).

3.4.5 Co-immunoprecipitation

Total proteins were extracted from 1 g of inflorescence tissues using IP buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, 2 mM EDTA, 1 mM PMSF, and 1 Complete Protease Inhibitor Cocktail Tablets [Roche]). Total proteins were first precleared with Protein G Dynabeads (Invitrogen) at 4C for 1 h. Anti-FLAG monoclonal antibody (F1804, Sigma) was equilibrated and prebound to Dynabeads at room temperature according to the manufacturer's manual. The anti-FLAG-Dynabead complexes were then added to the precleared total protein supernatant, and the preparation was incubated with rotation at 4°C for 2 h. After the precipitant was washed

with low/high salt (150/500 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.2% SDS, 0.5% Triton X-100, and 2 mM EDTA) and TE buffers, SDS sample buffer was added to the precipitant and boiled for 8% SDS-PAGE electrophoresis. NRPB1 protein was detected using anti-NRPB1 antibody (ab24758, Abcam).

3.4.6 Chromatin immunoprecipitation (ChIP) assay

Chromatin immunoprecipitation was performed as described previously [107] In brief, 4 g of 2-week-old seedling was harvested and cross-linked using 1% formaldehyde. Nuclei were isolated, and lysis nuclei were sonicated using Bioruptor Plus Sonicator (Diagenode). The sonicated chromatin was precleared using Protein IgG Dynabeads (Invitrogen), and antibody was added to the precleared supernatant. After overnight incubation with rotation at 4°C, the chromatin-antibody-Dynabeads were sequentially washed with low salt, high salt, LiCl, and TE buffer. The chromatin-antibody was eluted from Dynabeads using elution buffer (1% SDS and 0.1 M NaHCO3), and DNA was recovered using standard procedures. Real-time PCR was performed using recovered DNA in the CFX96 Touch Real-Time PCR Detection System (185-5196, Bio-Rad).

3.4.7 Quantification of Pol V-dependent scaffold RNAs

Total RNA was extracted from 12-day-old seedlings with TRIzol (Invitrogen). DNase-treated RNAs were reverse-transcribed with random hexameter using SuperScript III (Invitrogen). Bio-Rad SYBGreen was used for quantitative RT-PCR.

3.4.8 Quantification of individual siRNAs

Small RNAs were extracted by using RNAzol RT (Molecular Research Center). The abundance of 24 nt siRNAs was quantitatively detected by using TaqMan Small RNA Assays (Applied Bio-Science) as described previously [98].

3.4.9 Comparison of DNA methylation levels by Chop-PCR

Genomic DNA was extracted from 12-day-old seedlings with the DNeasy Kit (Qiagen). A 500-ng quantity of genomic DNA was incubated overnight with the methylation-sensitive restriction enzyme HaeIII. The digested DNA was used to amplify the RdDM targets by semi-quantitative RT-PCR. Non-digested genomic DNA was simultaneously amplified as controls.

3.5 Figures

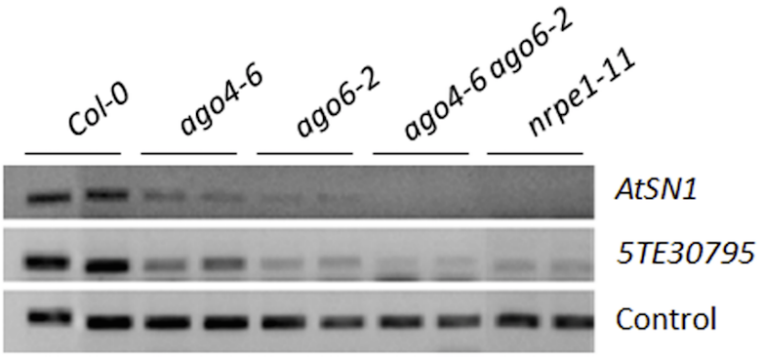

Fig. 3.1. Chop PCR results showing DNA hypomethylation in *ago4-6* and *ago6-2*. Dysfunction of AGO6 caused DNA hypomethylation in mature leaves, as in other RdDM mutants. Chop PCR results are shown. In a Chop PCR assay, DNA is first digested by methylation-sensitive restriction enzyme(MSRE) and then amplified by PCR. As MSRE cannot cleave methylated cytosines, PCR products can be successfully amplified for methylated DNA. If the methylation is lost, the DNA is digested and no PCR product will be amplified.

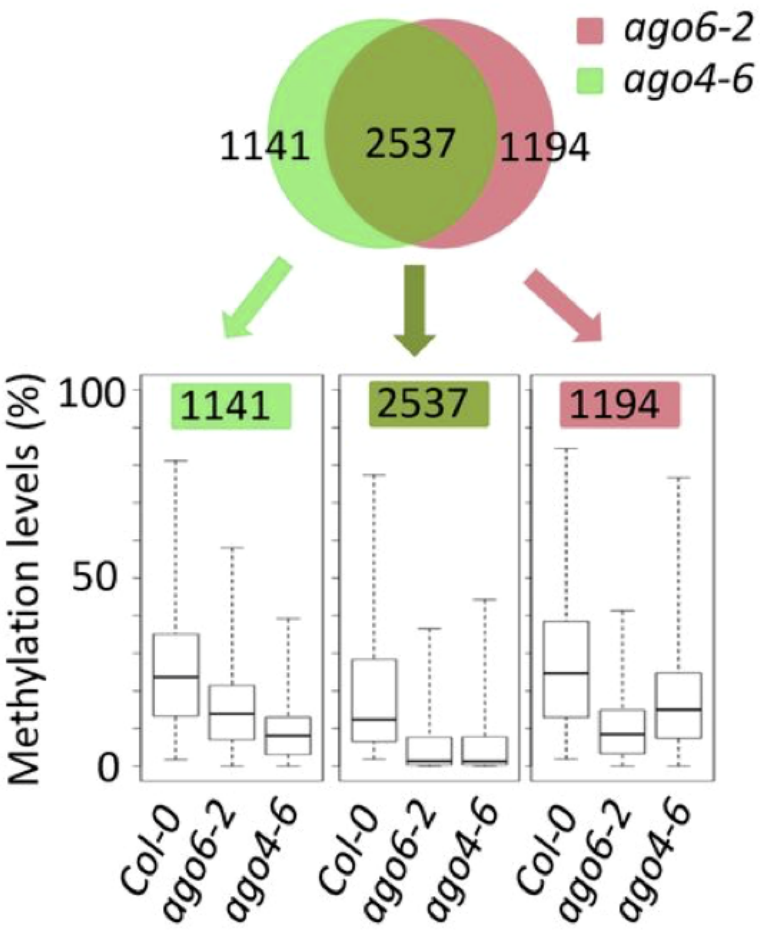

Fig. 3.2. Overlapping patterns of hypomethylated loci in *ago4-6* and *ago6-2*. Upper panel, Venn diagram showing the numbers of the following three types of hypomethylation loci: (1) loci where *ago4-6* shows lower methylation levels than *ago6-2* (green alone); (2) loci where *ago4-6* and *ago6-2* show equal methylation levels (green and red overlap); and (3) loci where *ago6-2* shows lower methylation levels than *ago4-6* (red alone). Lower panel, box plots showing methylation levels at the three groups of loci.

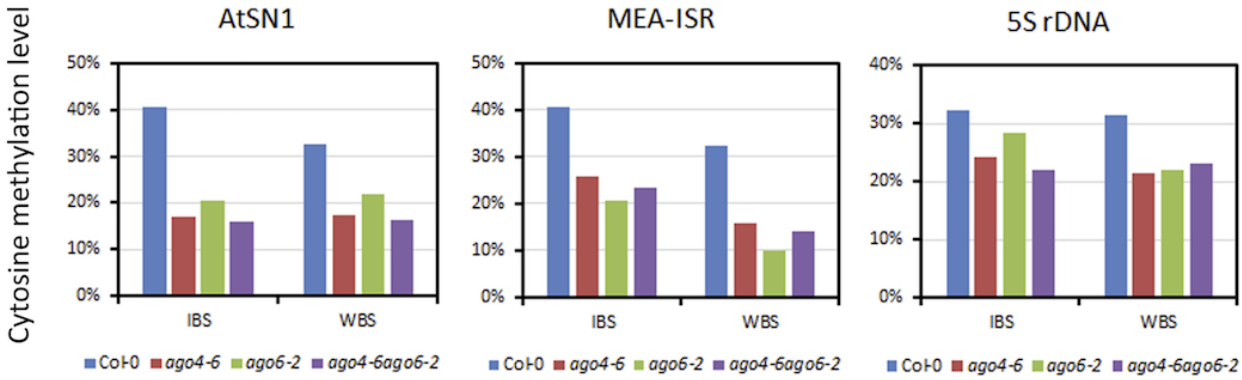

Fig. 3.3. DNA methylation levels at selected loci. The DNA methylation levels in selected ATSN1, MEA-ISR and 5S rDNA loci calculated from individual bisulfite sequencing (IBS) and whole genome bisulfite sequencing (WBS).

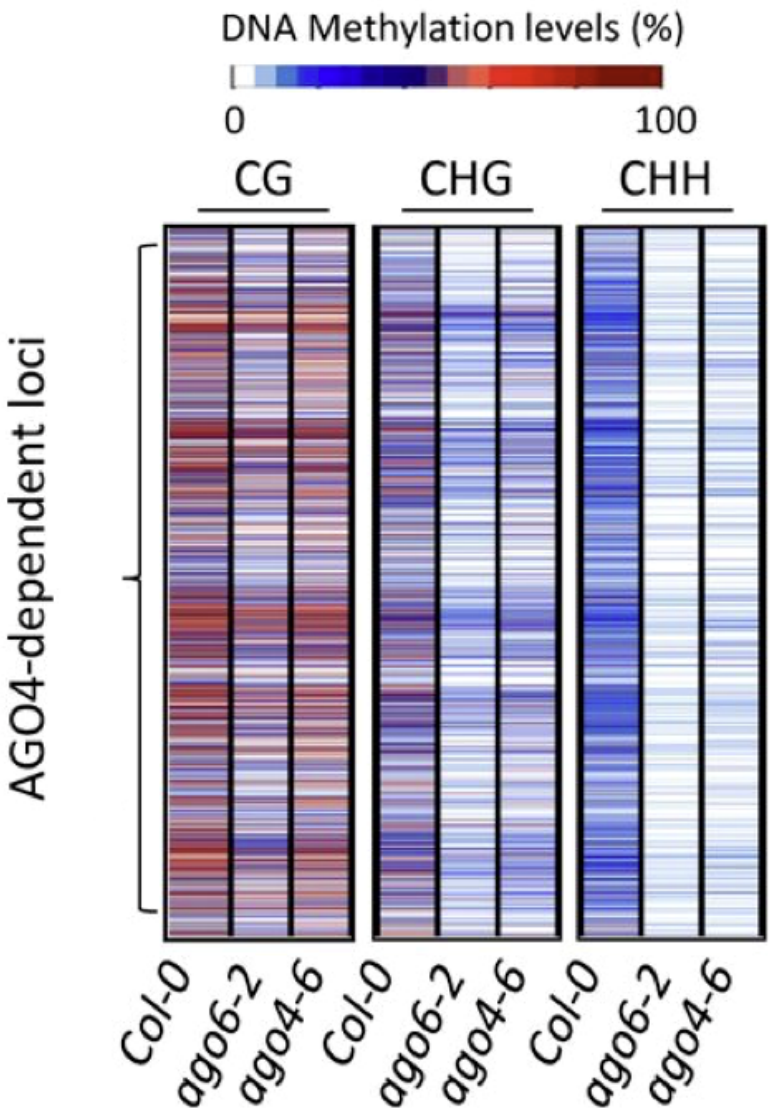

Fig. 3.4. Heat map depiction of the 3731 DNA hypomethylation loci in *ago4-6*. Each hypomethylated region corresponds to a colored horizontal bar, and the bars are clustered numerically into a column (y-axis). Cytosines were examined as CG, CHG, and CHH. The color-scaled methylation levels indicate the ratios of each type of methylated cytosines over total cytosines of the same type within the examined hypomethylated regions.

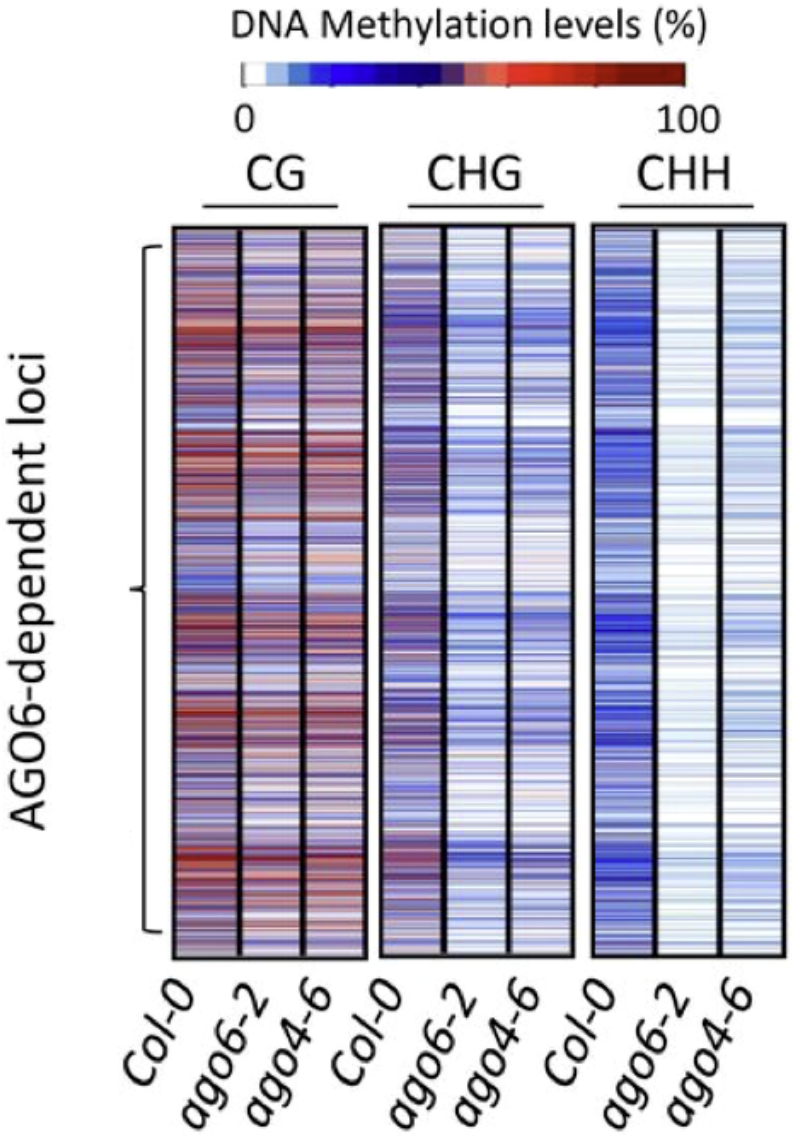

Fig. 3.5. Heat map depiction of the 3678 DNA hypomethylation loci in *ago6-2*.

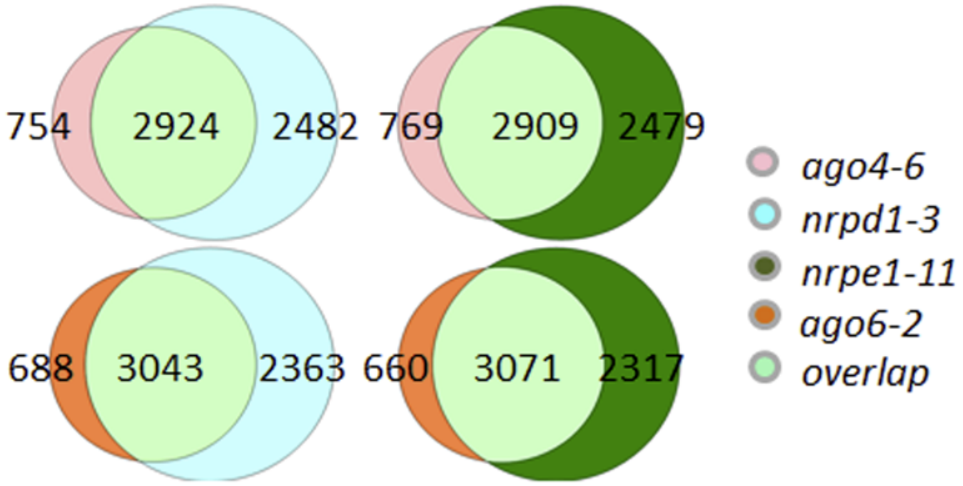

Fig. 3.6. Venn diagram showing overlapping patterns between mutants. Venn diagram showing overlapping patterns between *ago4-6* and *nrpd1-3*; *ago4-6* and *nrpe1-11*; *ago6-2* and *nrpd1-3*; *ago6-2* and *nrpe1-11*.

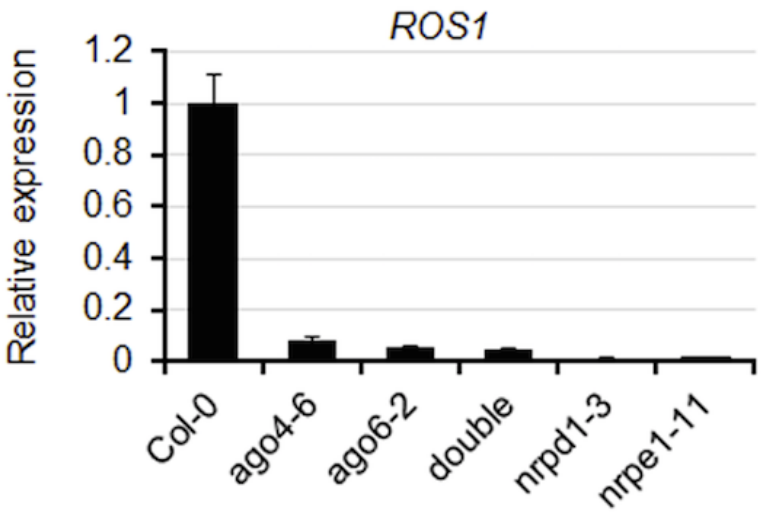

Fig. 3.7. AGO4 or AGO6 dysfunction reduces *ROS1* gene expression. *ROS1* gene expression is down-regulated in *ago4-6*, *ago6-2*, *ago4-6ago6-2* double, *nrpd1-3*, and *nrpe1-11* mutants compared with Col-0. ACTIN2 was used as the control in RT-qPCR. Means ± SD, n=3.

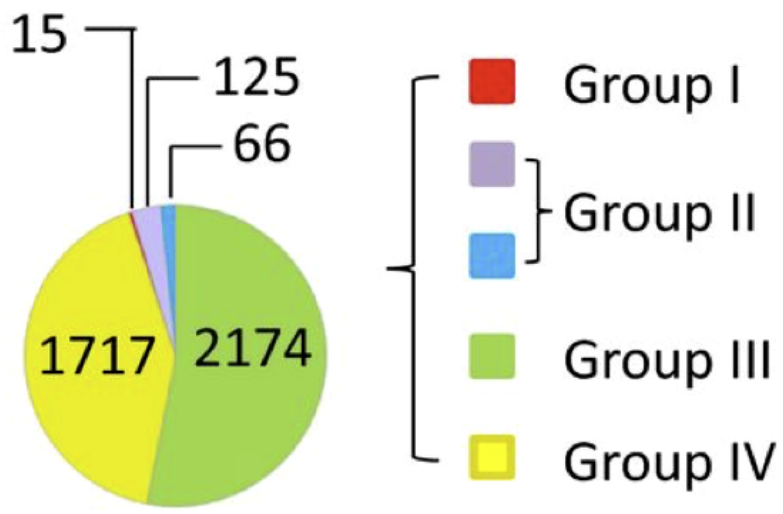

Fig. 3.8. Different groups of hypomethylation loci in the *ago4-6ago6-2* double mutant. Pie chart showing proportions of different groups of hypomethylation loci in the *ago4-6ago6-2* double mutant. Group I are loci where DNA hypomethylation is observed in the double mutant but not in the single mutants, indicating that DNA methylation at these loci is redundantly regulated by AGO4 and AGO6; Group II are loci where DNA hypomethylation occurs in one of the two single mutants but not in the other, and DNA methylation is not further decreased in the double mutant, indicating that DNA methylation at these loci requires AGO4 or AGO6 specifically; Group III are loci where DNA methylation is similarly reduced in the two single mutants, and the double mutant shows no additive effects on DNA hypomethylation, indicating that AGO4 and AGO6 are each required to mediate RdDM at these loci; and Group IV are loci where DNA hypomethylation is observed in the single mutants while the double mutant shows more severe DNA hypomethylation, indicating more complex genetic interactions between the two AGO proteins.

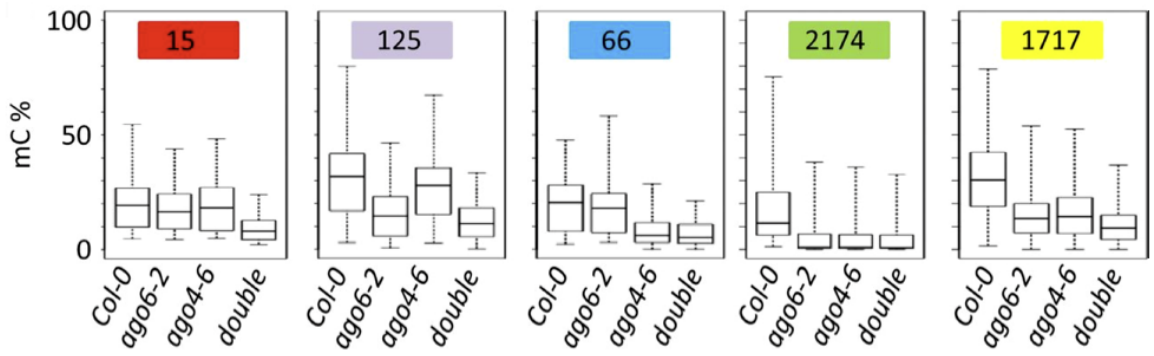

Fig. 3.9. Box plots of different groups of hypomethylation loci in double mutant. Box plots showing methylation patterns of different groups of loci as categorized in Fig. 3.8.

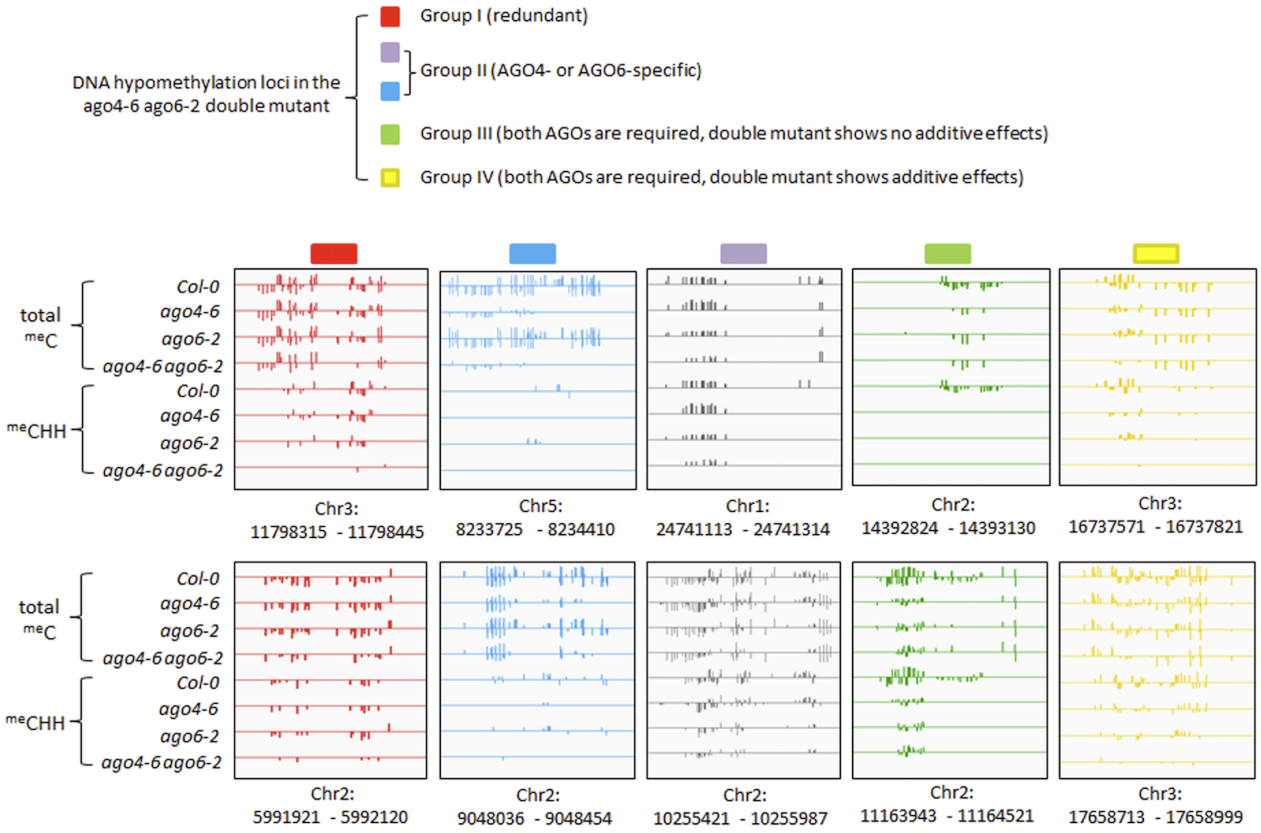

Fig. 3.10. Examples of different types of loci. Examples of different types of loci (related to Fig. 3.8). Screenshots from whole-genome bisulfite sequencing results are shown.

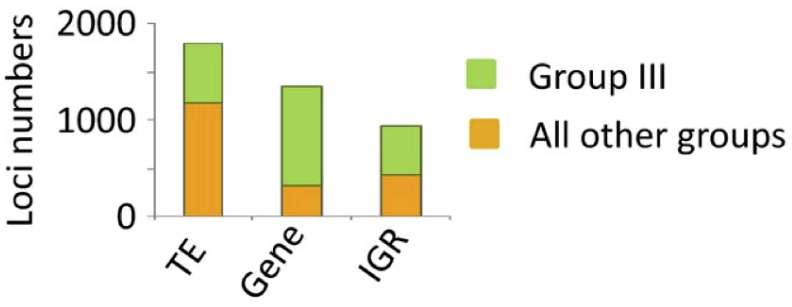

Fig. 3.11. Number of composition in different groups. Gene regions targeted by AGO4 and AGO6 are enriched in Group III loci. Y-axis shows the numbers of DNA hypomethylation loci in *ago4-6 ago6-2*. TEs: transposons; Gene: genic regions; IGR: intergenic regions that do not overlap with TEs

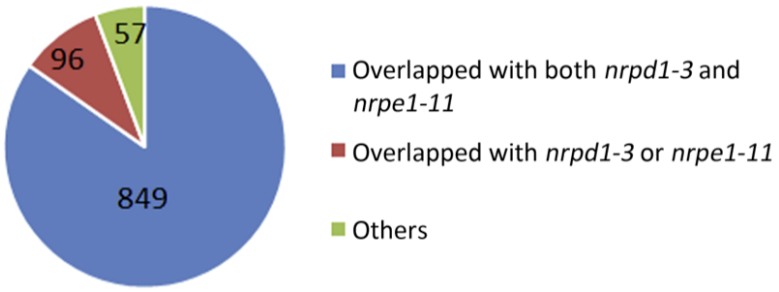

Fig. 3.12. Characterization of hypo DMRs in Group III gene loci (1002). The hypo-DMRs were overlapped with that identified in *nrpd1-3* and/or *nrpe1-11* mutants. The number of overlapped DMRs were shown here.

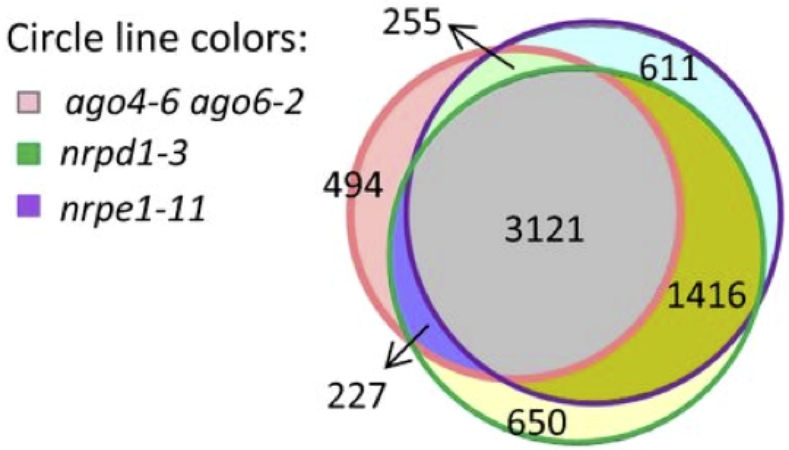

Fig. 3.13. Venn diagram showing the overlapping patterns. Venn diagram showing the overlapping patterns among DNA hypomethylation loci identified in *ago4-6ago6-2*, *nrpd1-3*, and *nrpe1-11*.

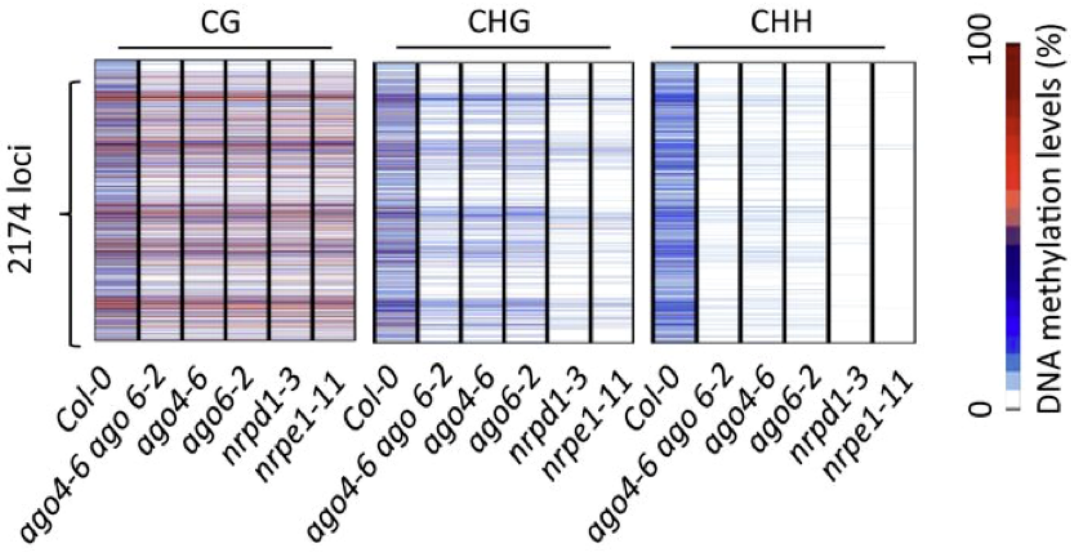

Fig. 3.14. DNA methylation levels in different mutants. The 2174 loci where AGO4 and AGO6 are mutually dependent are numerically clustered (y-axis). Cytosines were examined as CG, CHG, and CHH.

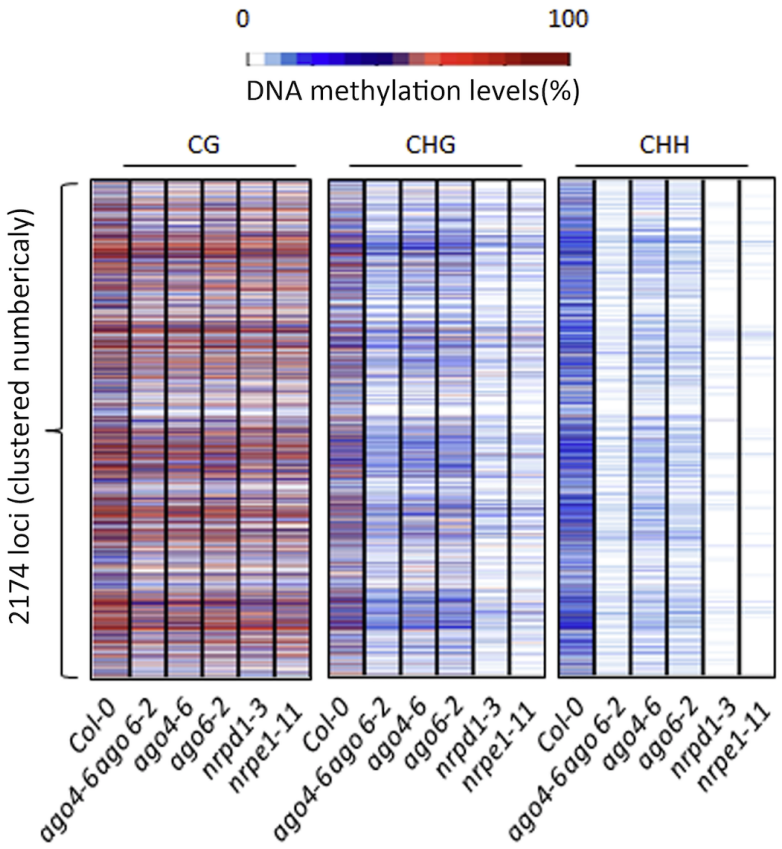

Fig. 3.15. DNA methylation levels in different mutants. The heat maps depict the 4097 DNA hypomethylation loci identified in *ago4-6ago6-2*. Each hypomethylated region corresponds to a colored horizontal bar, and the bars are clustered numerically into a column (Y-axis). Cytosines were examined as CG, CHG, and CHH. The color-scaled methylation levels indicate the ratios of each type of methylated cytosines over total cytosines of the same type within the examined hypomethylated regions.

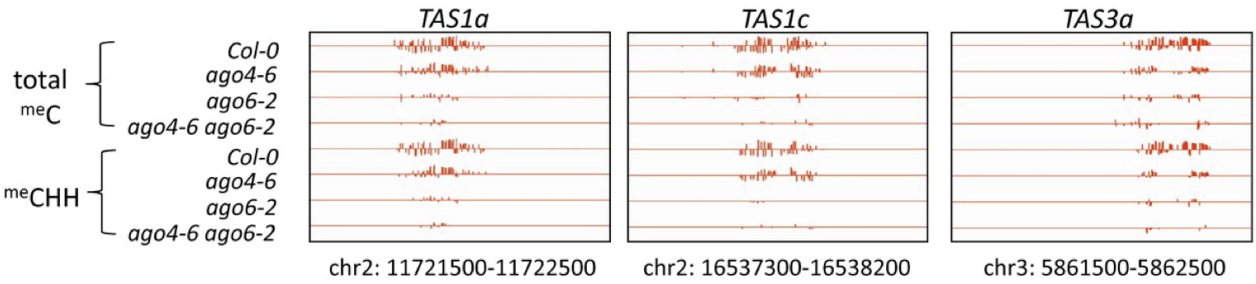

Fig. 3.16. DNA methylation levels at selected loci. DNA methylation levels at TAS1a, TAS1c, and TAS3a loci. Snapshots from whole-genome bisulfite sequencing results are shown.

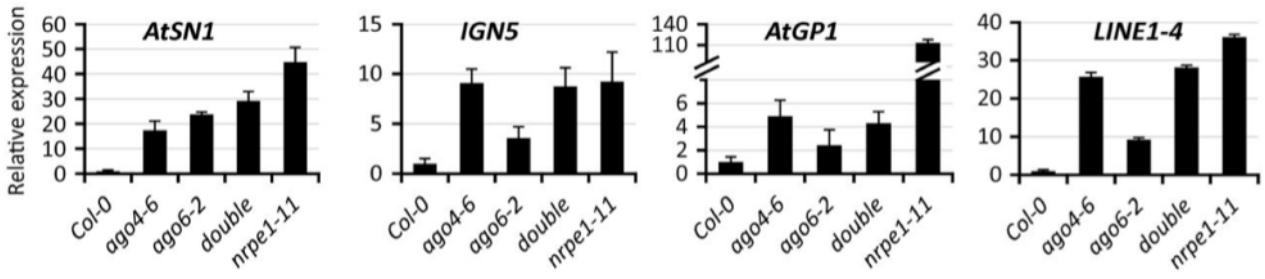

Fig. 3.17. RT-qPCR measurements of transposon RNA levels. Double: *ago4-6ago6-2* double mutant. ACTIN2 was used as an internal control. RNA levels in the mutants are relative to those in the wild-type (Col-0). Means ± SD are shown, n = 3.

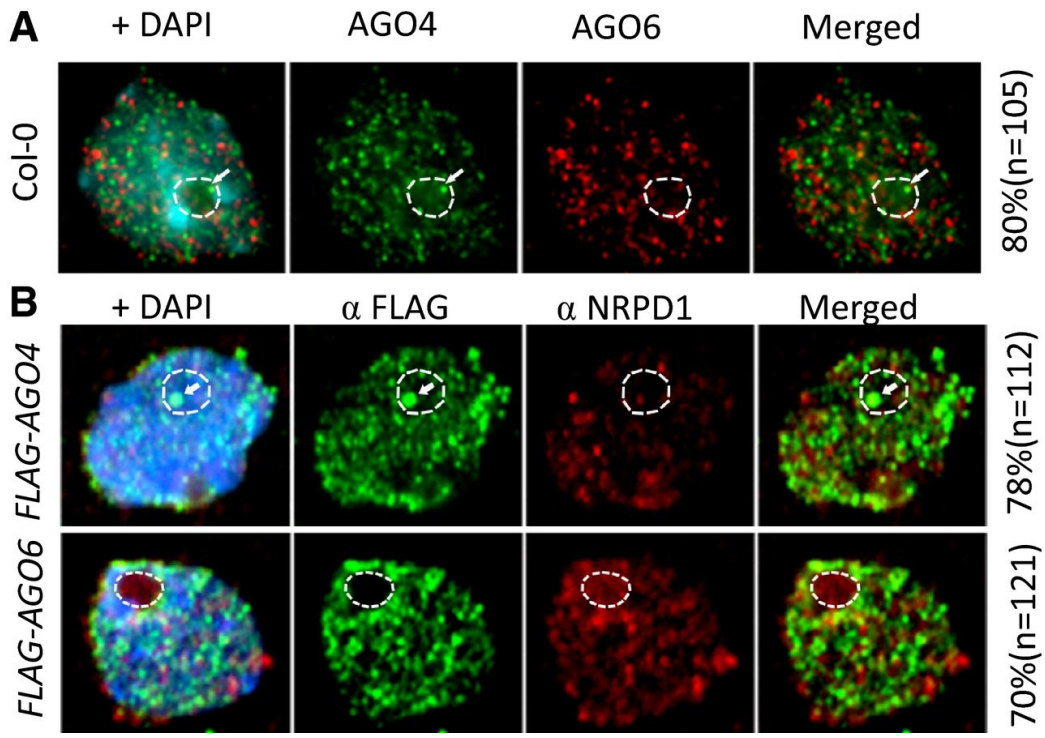

Fig. 3.18. **A:** AGO4 and AGO6 do not co-localize with each other or with Pol IV. AGO4 and AGO6 were visualized by immunofluorescence using their specific antibodies. Yellow signals would be expected in the merged images if the two proteins co-localize, as a result of the overlap of red and green channels. DNA (blue) was stained with DAPI. **B:** The largest subunit of Pol IV, NRPD1 (red), was visualized by its specific antibody in cells expressing FLAG-tagged AGO4 or AGO6 (green). Data information: The perinucleolar part was marked with dashed white lines, and nucleolar dot was marked with white arrow. The number next to the images showing the total number of nuclei with the presence of both red and green fluorescence signals. The percentage shows the proportion of the nuclei with similar co-localization patterns. Qualitative images are shown here.

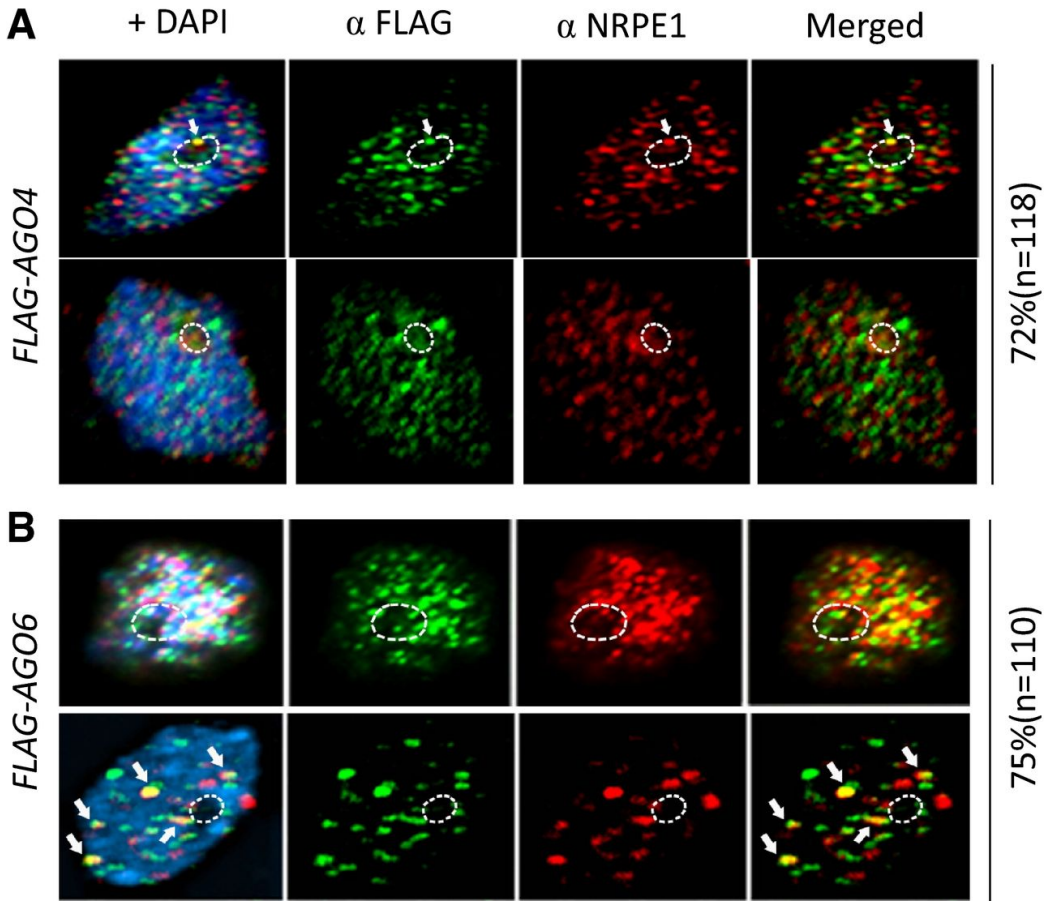

Fig. 3.19. **A:** The largest subunit of Pol V, NRPE1 (red), was visualized by immunofluorescence using its specific antibody in cells expressing FLAG-tagged AGO4 (green). AGO4-NRPE1 co-localized to nucleolar dot (yellow dot, marked with white arrow, upper panel) but not in nucleoplasmic foci (lower panel). **B:** NRPE1 (red) was visualized by its specific antibody in cells expressing FLAG-tagged AGO6 (green). The yellow signals due to the overlap of red and green channels in merged images indicate protein co-localization. DNA (blue) was stained with DAPI. AGO6-NRPE1 co-localized to nucleoplasmic foci (yellow dots, marked with white arrow). Data information: The perinucleolar part was marked with dashed white lines. The number next to the images showing the total number of nuclei with the presence of both red and green fluorescence signals. The percentage shows the proportion of the nuclei with similar co-localization patterns. Representative images are shown here.

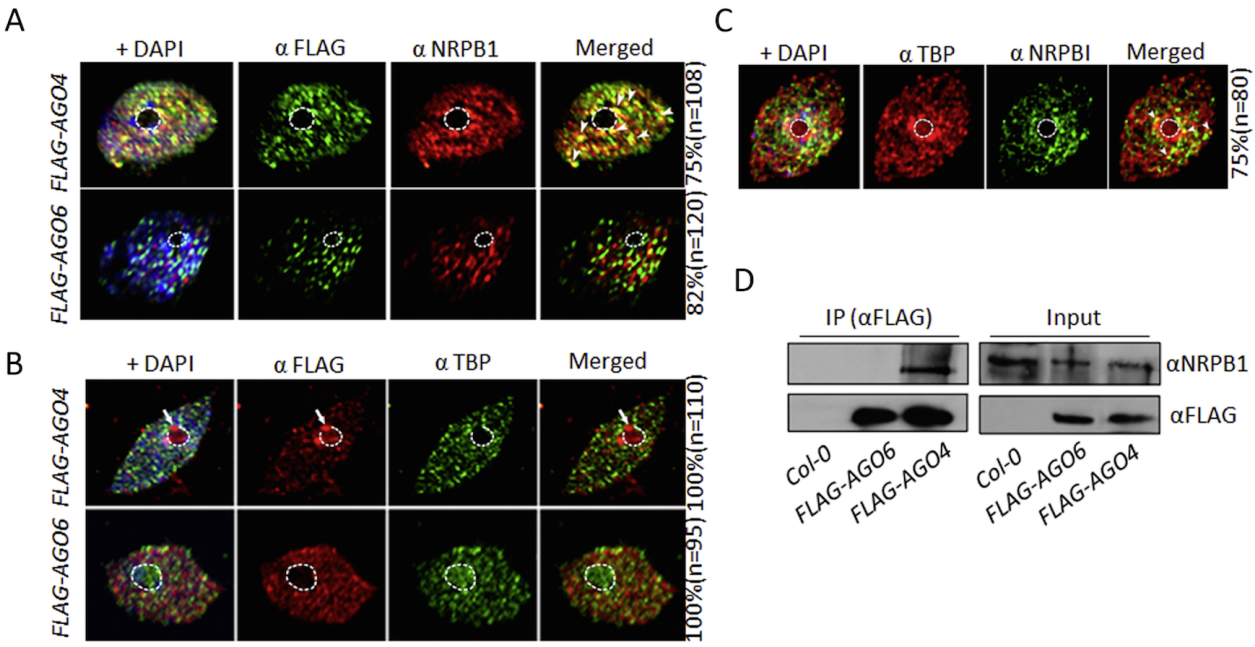

Fig. 3.20. Pol II co-localizes and physically interacts with AGO4 but not with AGO6. **A:** The largest subunit of Pol II, NRPB1 (red), was visualized by its specific antibody in cells expressing FLAG-tagged AGO4 or AGO6 (green). AGO4-NRPB1 co-localized to nucleoplasmic foci (yellow dots, marked with white arrow). **B:** AGO4 and AGO6 do not co-localize with TBP, a transcription factor that is indicative of Pol II transcription of mRNAs. **C:** TBP-NRPB1 co-localized to some nucleoplasmic foci (yellow dots, marked with white arrow). The perinucleolar part was marked with dashed white lines. The number next to the images showing the total number of nuclei with the presence of both red and green fluorescence signals. The percentage shows the proportion of the nuclei with similar co-localization patterns. Representative images were showed here. **D:** AGO4, but not AGO6, co-immunoprecipitates with NRPB1.

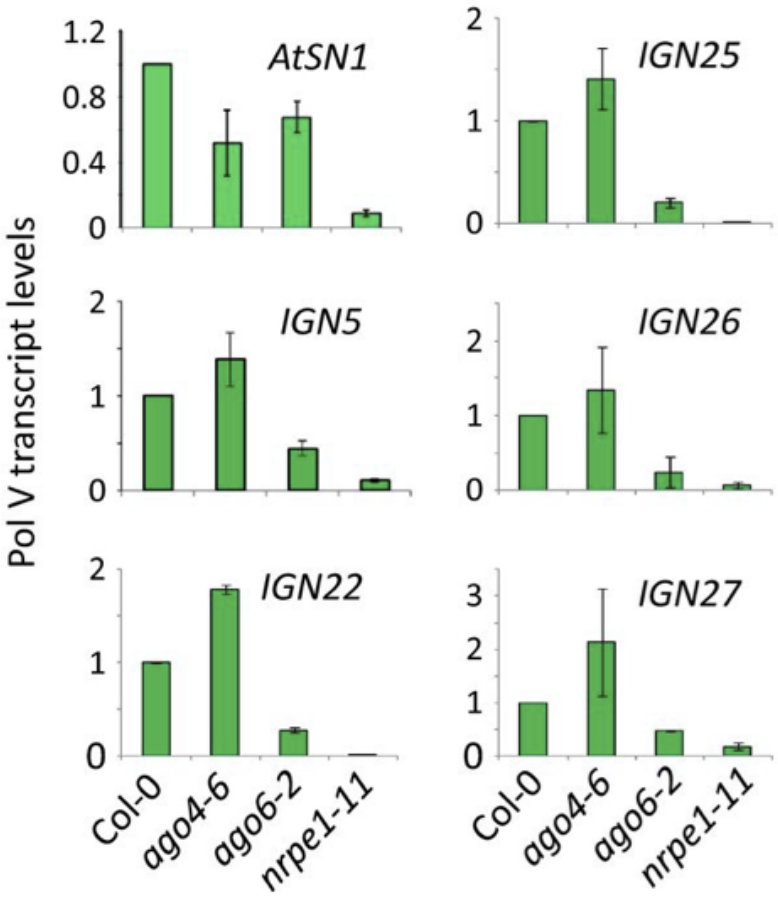

Fig. 3.21. AGO6 dysfunction reduces levels of Pol V-dependent transcripts. RNA levels were measured by RT-qPCR. Means ± SD are shown, n = 3.

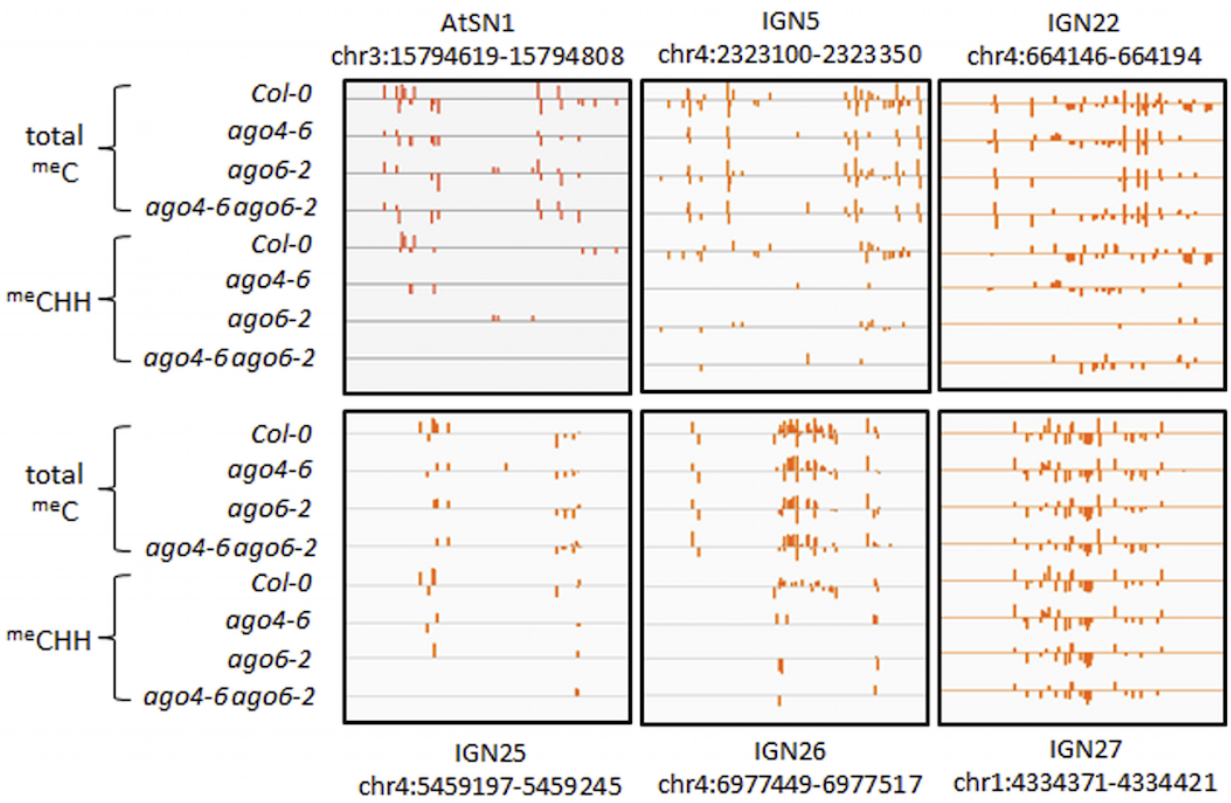

Fig. 3.22. Screenshots of methylation level at selected loci. DNA methylation patterns in *ago* mutants at loci where Pol V-dependent scaffold RNAs are affected by AGO4 or AGO6. Screenshots from whole-genome bisulfite sequencing results are shown.

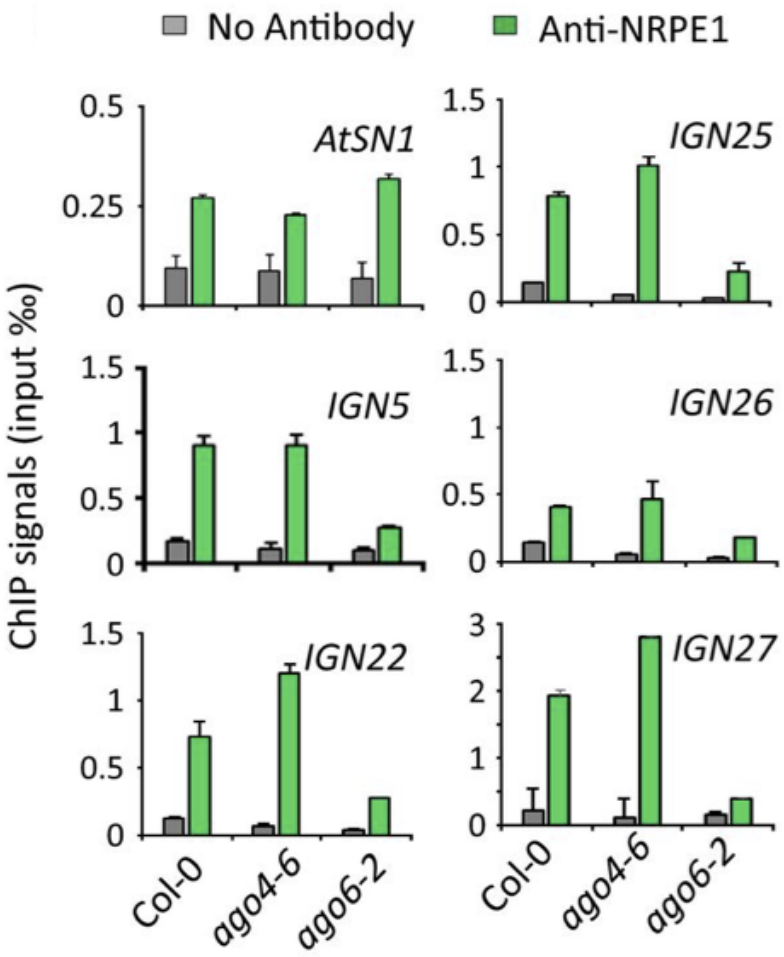

Fig. 3.23. AGO6 dysfunction decreases Pol V occupancy. AGO6 dysfunction decreases Pol V occupancy at chromatin of loci where Pol V transcript levels are affected by AGO6. Anti-NRPE1 antibody was used for ChIP assays. Means ± SD are shown, n = 3.

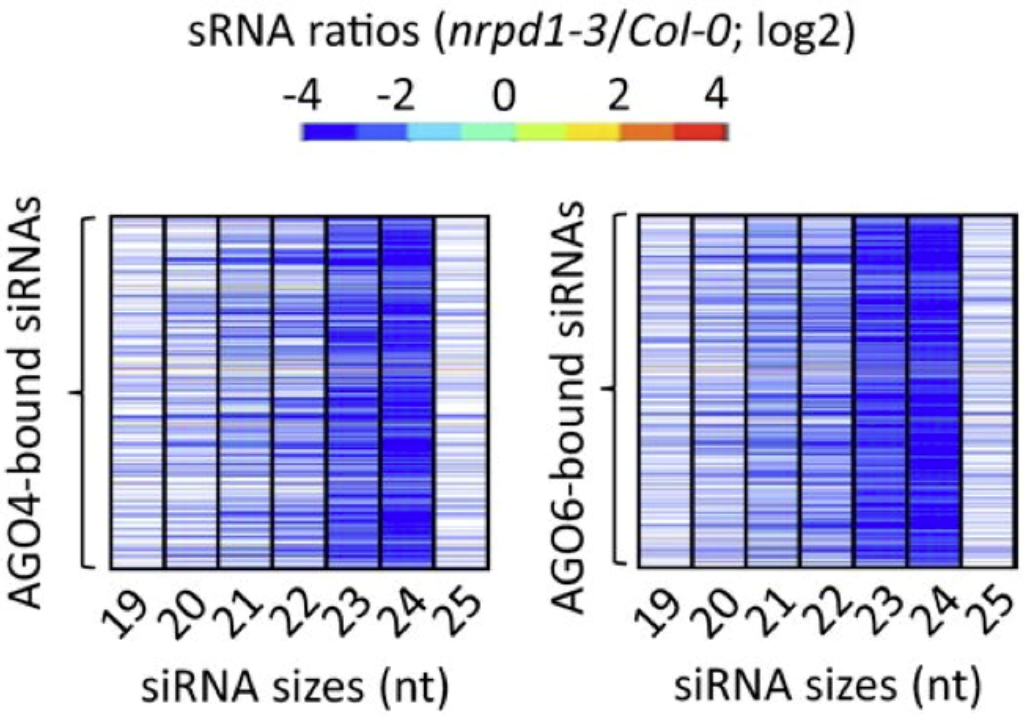

Fig. 3.24. Heat map depiction of the dependence of AGO4- and AGO6-bound siRNAs on Pol IV. Identities of siRNAs were retrieved from Havecker *et al.* [83], and siRNAs were grouped into 1210 AGO4-bound clusters and 1486 AGO6-bound clusters (see Materials and Methods for details), each of which corresponds to a colored horizontal bar; the bars are stacked numerically into a column (y-axis).

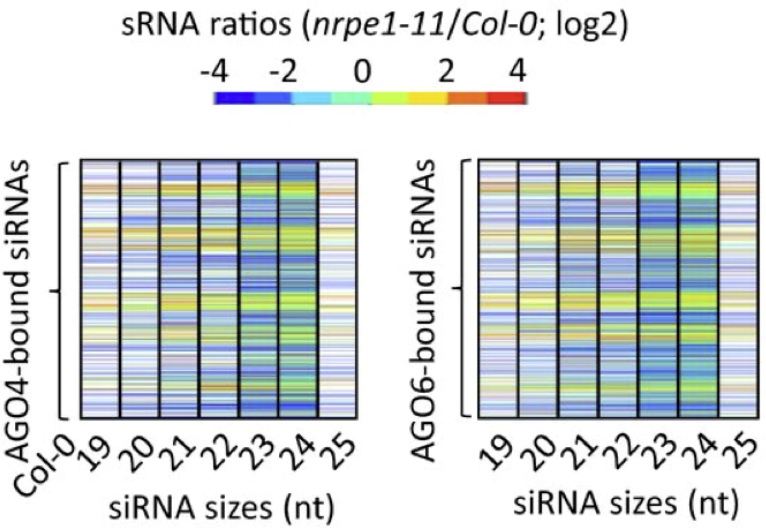

Fig. 3.25. Heat map depiction of the dependence of AGO4- and AGO6-bound siRNAs on Pol V.

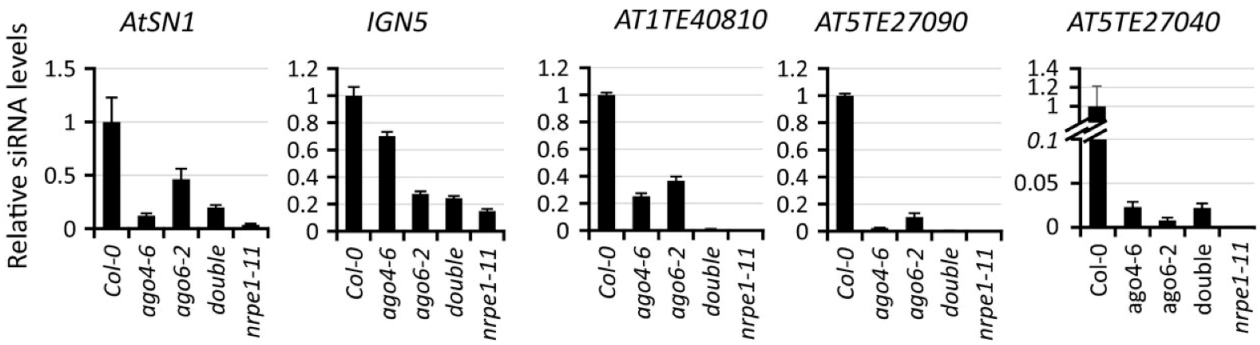

Fig. 3.26. Quantification of individual 24 nt siRNAs. Quantification of individual 24 nt siRNAs by TaqMan small RNA assays. Double: *ago4-6ago6-2* double mutant. snoR101 was used as an internal control. RNA levels in the mutants were relative to those of the wild-type (Col-0). Error bars indicate SD, n $\geq$ 3.

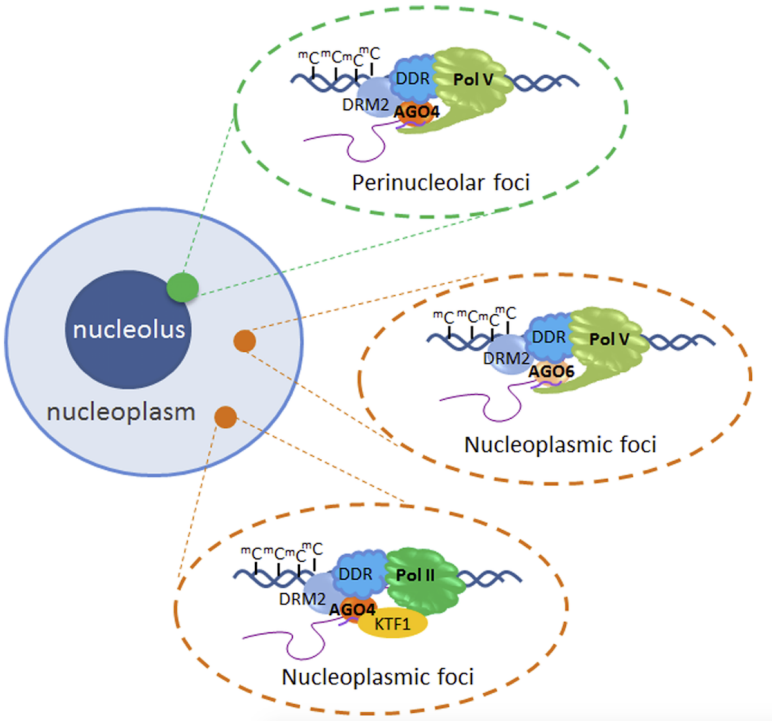

Fig. 3.27. A model of spatial segregation of nuclear RdDM activities. RdDM activities at perinucleolar foci involve AGO4 and Pol V, while RdDM activities in the nucleoplasm are mediated by AGO6 and Pol V, and by AGO4 and Pol II. DNA methylation is annotated as mC. DDR, the DRD1-DMS3-RDM1 complex. For simplicity, not all known RdDM components are shown.

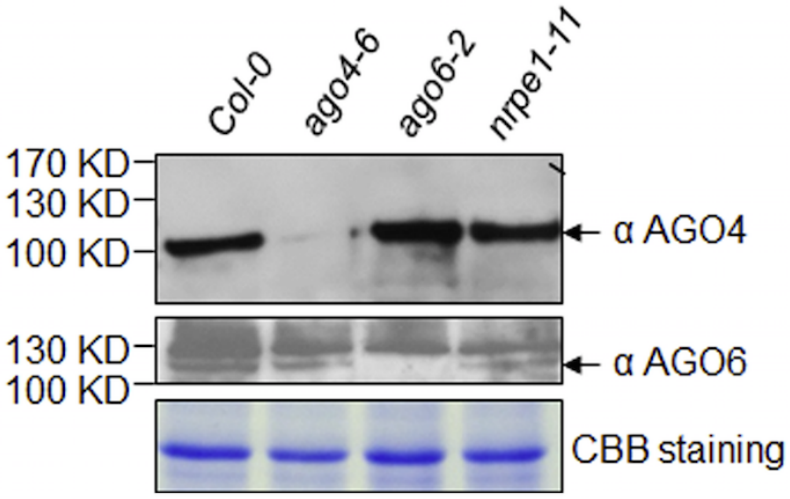

Fig. 3.28. Protein level of AGO6 is reduced in *nrpe1-11* mutant. Pol V dysfunction reduces the protein level of AGO6 but not AGO4. Antibodies specific to AGO4 and AGO6 were used in Western blot analyses.

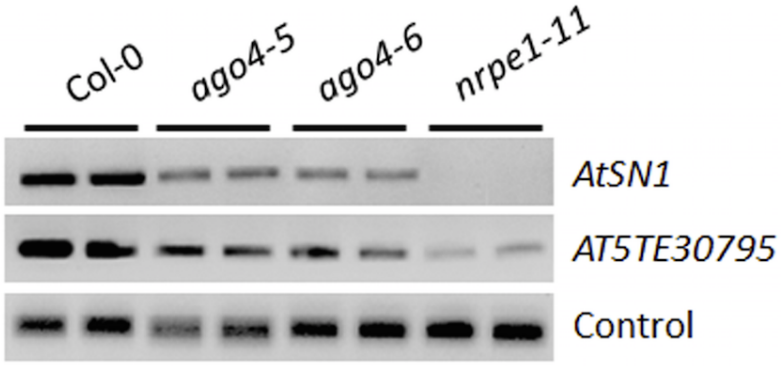

Fig. 3.29. Hypo-methylation phenotype in *ago4-6* and *ago4-5* mutant alleles. DNA methylation-sensitive Chop-PCR assay showing hypomethylation phenotype in two selected loci in *ago4-6* mutant as well as *ago4-5* mutant. HaeIII was the methylation-sensitive restriction enzyme. Non-digested DNA was amplified in parallel as the loading control.

LIST OF REFERENCES

LIST OF REFERENCES

[1] D. Baulcombe, "RNA silencing in plants." *Nature*, vol. 431, no. 7006, pp. 356–63, 2004.

[2] S. Feng, S. E. Jacobsen, and W. Reik, "Epigenetic reprogramming in plant and animal development." *Science*, vol. 330, no. 6004, pp. 622–7, 2010.

[3] B. Khraiwesh, J.-K. Zhu, and J. Zhu, "Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants." *Biochimica et Biophysica Acta*, vol. 1819, no. 2, pp. 137–48, 2012.

[4] W. Reik and J. Walter, "Genomic imprinting: parental influence on the genome." *Nature Reviews Genetics*, vol. 2, no. 1, pp. 21–32, 2001.

[5] R. Maxfield Boumil and J. T. Lee, "Forty years of decoding the silence in X-chromosome inactivation." *Human Molecular Genetics*, vol. 10, no. 20, pp. 2225–2232, 2001.

[6] J. Bender, "DNA methylation and epigenetics." *Annual Review of Plant Biology*, vol. 55, pp. 41–68, 2004.

[7] S. W.-L. Chan, I. R. Henderson, and S. E. Jacobsen, "Gardening the genome: DNA methylation in *Arabidopsis thaliana*." *Nature Reviews Genetics*, vol. 6, no. 5, pp. 351–60, 2005.

[8] E. M. Meyerowitz, "Prehistory and History of *Arabidopsis* Research." *Plant Physiology*, vol. 125, no. 1, pp. 15–19, 2001.

[9] M. Koornneef and D. Meinke, "The development of *Arabidopsis* as a model plant." *The Plant Journal*, vol. 61, no. 6, pp. 909–21, 2010.

[10] M. Noyer-Weidner and T. A. Trautner, "Methylation of DNA in prokaryotes." *EXS*, vol. 64, pp. 39–108, 1993.

[11] S. B. Baylin and P. A. Jones, "A decade of exploring the cancer epigenome - biological and translational implications." *Nature Reviews Cancer*, vol. 11, no. 10, pp. 726–34, 2011.

[12] V. Patil, R. L. Ward, and L. B. Hesson, "The evidence for functional non-CpG methylation in mammalian cells." *Epigenetics*, vol. 9, no. 6, pp. 823–8, 2014.

[13] S. Tweedie, J. Charlton, V. Clark, and A. Bird, "Methylation of genomes and genes at the invertebrate-vertebrate boundary." *Molecular and Cellular Biology*, vol. 17, no. 3, pp. 1469–1475, 1997.

[14] X.-J. He, T. Chen, and J.-K. Zhu, "Regulation and function of DNA methylation in plants and animals." *Cell Research*, vol. 21, no. 3, pp. 442–65, 2011.

[15] M. A. Matzke and R. A. Mosher, "RNA-directed DNA methylation: an epigenetic pathway of increasing complexity." *Nature Reviews Genetics*, vol. 15, no. 6, pp. 394–408, 2014.

[16] L. M. Johnson, J. Du, C. J. Hale, S. Bischof, S. Feng, R. K. Chodavarapu, X. Zhong, G. Marson, M. Pellegrini, D. J. Segal, D. J. Patel, and S. E. Jacobsen, "SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation." *Nature*, vol. 507, 2014.

[17] D.-L. Yang, G. Zhang, K. Tang, J. Li, L. Yang, H. Huang, H. Zhang, and J.-K. Zhu, "Dicer-independent RNA-directed DNA methylation in *Arabidopsis*." *Cell Research*, 2015.

[18] J. Zhai, S. Bischof, H. Wang, S. Feng, T.-f. Lee, C. Teng, X. Chen, S. Park, L. Liu, J. Gallego-Bartolome, W. Liu, I. Henderson, B. Meyers, I. Ausin, and S. Jacobsen, "A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis." *Cell*, vol. 163, no. 2, pp. 445–455, 2015.

[19] R. Ye, Z. Chen, B. Lian, M. Rowley, N. Xia, J. Chai, Y. Li, X.-J. He, A. Wierzbicki, and Y. Qi, "A Dicer-Independent Route for Biogenesis of siRNAs that Direct DNA Methylation in *Arabidopsis*." *Molecular Cell*, vol. 61, no. 2, pp. 222–235, 2016.

[20] T. Blevins, R. Podicheti, V. Mishra, M. Marasco, H. Tang, and C. S. Pikaard, "Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in *Arabidopsis*." *eLife*, vol. 4, p. e09591, 2015.

[21] J.-K. Zhu, "Active DNA demethylation mediated by DNA glycosylases." *Annual Review of Genetics*, vol. 43, pp. 143–66, 2009.

[22] Z. Gong, T. Morales-Ruiz, R. R. Ariza, T. Roldán-Arjona, L. David, and J. K. Zhu, "ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase." *Cell*, vol. 111, no. 6, pp. 803–14, 2002.

[23] F. Agius, A. Kapoor, and J.-K. Zhu, "Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation." *Proceedings of the National Academy of Sciences*, vol. 103, no. 31, pp. 11 796–801, 2006.

[24] T. Morales-Ruiz, A. P. Ortega-Galisteo, M. I. Ponferrada-Marín, M. I. Martínez-Macías, R. R. Ariza, and T. Roldán-Arjona, "DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases." *Proceedings of the National Academy of Sciences*, vol. 103, no. 18, pp. 6853–8, 2006.

[25] Y. Li, D. Córdoba-Cañero, W. Qian, X. Zhu, K. Tang, H. Zhang, R. R. Ariza, T. Roldán-Arjona, and J.-K. Zhu, "An AP endonuclease functions in active DNA demethylation and gene imprinting in *Arabidopsis*." *PLoS Genetics*, vol. 11, no. 1, p. e1004905, 2015.

[26] M. I. Martínez-Macías, W. Qian, D. Miki, O. Pontes, Y. Liu, K. Tang, R. Liu, T. Morales-Ruiz, R. R. Ariza, T. Roldán-Arjona, and J.-K. Zhu, "A DNA 3' phosphatase functions in active DNA demethylation in *Arabidopsis*." *Molecular Cell*, vol. 45, no. 3, pp. 357–70, 2012.

[27] Y. Li, C.-G. Duan, X. Zhu, W. Qian, and J.-K. Zhu, "A DNA ligase required for active DNA demethylation and genomic imprinting in *Arabidopsis*." *Cell Research*, vol. 25, no. 6, pp. 757–60, 2015.

[28] Z. Lang, M. Lei, X. Wang, K. Tang, D. Miki, H. Zhang, S. K. Mangrauthia, W. Liu, W. Nie, G. Ma, J. Yan, C.-G. Duan, C.-C. Hsu, C. Wang, W. A. Tao, Z. Gong, and J.-K. Zhu, "The methyl-CpG-binding protein MBD7 facilitates active DNA demethylation to limit DNA hyper-methylation and transcriptional gene silencing." *Molecular Cell*, vol. 57, no. 6, pp. 971–83, 2015.

[29] C. Wang, X. Dong, D. Jin, Y. Zhao, S. Xie, X. Li, X. He, Z. Lang, J. Lai, J.-K. Zhu, and Z. Gong, "Methyl-CpG-binding domain protein MBD7 is required for active DNA demethylation in *Arabidopsis*." *Plant Physiology*, vol. 167, no. 3, pp. 905–14, 2015.

[30] W. Qian, D. Miki, H. Zhang, Y. Liu, X. Zhang, K. Tang, Y. Kan, H. La, X. Li, S. Li, X. Zhu, X. Shi, K. Zhang, O. Pontes, X. Chen, R. Liu, Z. Gong, and J.-K. Zhu, "A histone acetyltransferase regulates active DNA demethylation in *Arabidopsis*." *Science*, vol. 336, no. 6087, pp. 1445–8, 2012.

[31] J. Zhu, A. Kapoor, V. V. Sridhar, F. Agius, and J.-K. Zhu, "The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in *Arabidopsis*." *Current Biology*, vol. 17, no. 1, pp. 54–9, 2007.

[32] L. Hunt and J. E. Gray, "The Signaling Peptide EPF2 Controls Asymmetric Cell Divisions during Stomatal Development." *Current Biology*, vol. 19, no. 10, pp. 864–869, 2009.

[33] C. Yamamuro, D. Miki, Z. Zheng, J. Ma, J. Wang, Z. Yang, J. Dong, and J.-K. Zhu, "Overproduction of stomatal lineage cells in *Arabidopsis* mutants defective in active DNA demethylation." *Nature Communications*, vol. 5, p. 4062, 2014.

[34] T.-N. Le, U. Schumann, N. A. Smith, S. Tiwari, P. C. K. Au, Q.-H. Zhu, J. M. Taylor, K. Kazan, D. J. Llewellyn, R. Zhang, E. S. Dennis, and M.-B. Wang, "DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in *Arabidopsis*." *Genome Biology*, vol. 15, no. 9, p. 458, 2014.

[35] X. Zheng, J. Zhu, A. Kapoor, and J.-K. Zhu, "Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing." *The EMBO Journal*, vol. 26, no. 6, pp. 1691–701, 2007.

[36] J. Penterman, D. Zilberman, J. H. Huh, T. Ballinger, S. Henikoff, and R. L. Fischer, "DNA demethylation in the *Arabidopsis* genome." *Proceedings of the National Academy of Sciences*, vol. 104, no. 16, pp. 6752–7, 2007.

[37] O. Mathieu, J. Reinders, M. Čaikovski, C. Smathajitt, and J. Paszkowski, "Transgenerational Stability of the *Arabidopsis* Epigenome Is Coordinated by CG Methylation." *Cell*, vol. 130, no. 5, pp. 851–862, 2007.

[38] B. Huettel, T. Kanno, L. Daxinger, W. Aufsatz, A. J. M. Matzke, and M. Matzke, "Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*." *The EMBO Journal*, vol. 25, no. 12, pp. 2828–36, 2006.

[39] M. Lei, H. Zhang, R. Julian, K. Tang, S. Xie, and J.-K. Zhu, "Regulatory link between DNA methylation and active demethylation in *Arabidopsis*." *Proceedings of the National Academy of Sciences*, vol. 112, no. 11, pp. 3553–3557, 2015.

[40] B. P. Williams, D. Pignatta, S. Henikoff, and M. Gehring, "Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat." *PLoS Genetics*, vol. 11, no. 3, p. e1005142, 2015.

[41] A. Mallory and H. Vaucheret, "Form, function, and regulation of ARGONAUTE proteins." *The Plant Cell*, vol. 22, no. 12, pp. 3879–89, 2010.

[42] S. E. Castel and R. A. Martienssen, "RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond." *Nature Reviews Genetics*, vol. 14, no. 2, pp. 100–12, 2013.

[43] M. Fagard, S. Boutet, J. B. Morel, C. Bellini, and H. Vaucheret, "AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals." *Proceedings of the National Academy of Sciences*, vol. 97, no. 21, pp. 11 650–4, 2000.

[44] J.-J. Song, S. K. Smith, G. J. Hannon, and L. Joshua-Tor, "Crystal structure of Argonaute and its implications for RISC slicer activity." *Science*, vol. 305, no. 5689, pp. 1434–7, 2004.

[45] Y. Qi, X. He, X.-J. Wang, O. Kohany, J. Jurka, and G. J. Hannon, "Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation." *Nature*, vol. 443, no. 7114, pp. 1008–12, 2006.

[46] N. H. Tolia and L. Joshua-Tor, "Slicer and the argonautes." *Nature Chemical Biology*, vol. 3, no. 1, pp. 36–43, 2007.

[47] K. S. Yan, S. Yan, A. Farooq, A. Han, L. Zeng, and M.-M. Zhou, "Structure and conserved RNA binding of the PAZ domain." *Nature*, vol. 426, no. 6965, pp. 468–74, 2003.

[48] A. Lingel, B. Simon, E. Izaurralde, and M. Sattler, "Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain." *Nature Structural & Molecular Biology*, vol. 11, no. 6, pp. 576–7, 2004.

[49] J.-B. Ma, K. Ye, and D. J. Patel, "Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain." *Nature*, vol. 429, no. 6989, pp. 318–22, 2004.

[50] J.-B. Ma, Y.-R. Yuan, G. Meister, Y. Pei, T. Tuschl, and D. J. Patel, "Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein." *Nature*, vol. 434, no. 7033, pp. 666–70, 2005.

[51] S. Mi, T. Cai, Y. Hu, Y. Chen, E. Hodges, F. Ni, L. Wu, S. Li, H. Zhou, C. Long, S. Chen, G. J. Hannon, and Y. Qi, "Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5' terminal nucleotide." *Cell*, vol. 133, no. 1, pp. 116–27, 2008.

[52] T. A. Montgomery, M. D. Howell, J. T. Cuperus, D. Li, J. E. Hansen, A. L. Alexander, E. J. Chapman, N. Fahlgren, E. Allen, and J. C. Carrington, "Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation." *Cell*, vol. 133, no. 1, pp. 128–41, 2008.

[53] F. Frank, N. Sonenberg, and B. Nagar, "Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2." *Nature*, vol. 465, no. 7299, pp. 818–22, 2010.

[54] F. V. Rivas, N. H. Tolia, J.-J. Song, J. P. Aragon, J. Liu, G. J. Hannon, and L. Joshua-Tor, "Purified Argonaute2 and an siRNA form recombinant human RISC." *Nature Structural & Molecular Biology*, vol. 12, no. 4, pp. 340–9, 2005.

[55] Y. Wang, S. Juranek, H. Li, G. Sheng, T. Tuschl, and D. J. Patel, "Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex." *Nature*, vol. 456, no. 7224, pp. 921–6, 2008.

[56] T. A. Farazi, S. A. Juranek, and T. Tuschl, "The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members." *Development*, vol. 135, no. 7, pp. 1201–14, 2008.

[57] K. Bohmert, I. Camus, C. Bellini, D. Bouchez, M. Caboche, and C. Benning, "AGO1 defines a novel locus of *Arabidopsis* controlling leaf development." *The EMBO Journal*, vol. 17, no. 1, pp. 170–80, 1998.

[58] D. Pontier, C. Picart, F. Roudier, D. Garcia, S. Lahmy, J. Azevedo, E. Alart, M. Laudié., W. M. Karlowski, R. Cooke, V. Colot, O. Voinnet, and T. Lagrange, "NERD, a plant-specific GW protein, defines an additional RNAi-dependent chromatin-based pathway in *Arabidopsis*." *Molecular Cell*, vol. 48, no. 1, pp. 121–32, 2012.

[59] J. J. W. Harvey, M. G. Lewsey, K. Patel, J. Westwood, S. Heimstädt, J. P. Carr, and D. C. Baulcombe, "An antiviral defense role of AGO2 in plants." *PloS One*, vol. 6, no. 1, p. e14639, 2011.

[60] M. Jaubert, S. Bhattacharjee, A. F. S. Mello, K. L. Perry, and P. Moffett, "ARGONAUTE2 mediates RNA-silencing antiviral defenses against Potato virus X in *Arabidopsis*." *Plant Physiology*, vol. 156, no. 3, pp. 1556–64, 2011.

[61] W. Wei, Z. Ba, M. Gao, Y. Wu, Y. Ma, S. Amiard, C. White, J. Rendtlew-Danielsen, Y.-G. Yang, and Y. Qi, "A Role for Small RNAs in DNA Double-Strand Break Repair." *Cell*, vol. 149, no. 1, pp. 101–112, 2012.

[62] H. Zhu, F. Hu, R. Wang, X. Zhou, S.-H. Sze, L. W. Liou, A. Barefoot, M. Dickman, and X. Zhang, "*Arabidopsis* Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development." *Cell*, vol. 145, no. 2, pp. 242–56, 2011.

[63] D. Zilberman, X. Cao, and S. E. Jacobsen, "ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation." *Science*, vol. 299, no. 5607, pp. 716–9, 2003.

[64] V. Olmedo-Monfil, N. Durán-Figueroa, M. Arteaga-Vázquez, E. Demesa-Arévalo, D. Autran, D. Grimanelli, R. K. Slotkin, R. A. Martienssen, and J.-P. Vielle-Calzada, "Control of female gamete formation by a small RNA pathway in *Arabidopsis*." *Nature*, vol. 464, no. 7288, pp. 628–632, 2010.

[65] J. a. Law and S. E. Jacobsen, "Establishing, maintaining and modifying DNA methylation patterns in plants and animals." *Nature Reviews Genetics*, vol. 11, no. 3, pp. 204–20, 2010.

[66] H. Zhang and J.-K. Zhu, "RNA-directed DNA methylation." *Current Opinion in Plant Biology*, vol. 14, no. 2, pp. 142–7, 2011.

[67] C. S. Pikaard, J. R. Haag, O. M. F. Pontes, T. Blevins, and R. Cocklin, "A Transcription Fork Model for Pol IV and Pol V-dependent RNA-Directed DNA Methylation." *Cold Spring Harbor Symposia on Quantitative Biology*, 2013.

[68] X. Zhang, I. R. Henderson, C. Lu, P. J. Green, and S. E. Jacobsen, "Role of RNA polymerase IV in plant small RNA metabolism." *Proceedings of the National Academy of Sciences*, vol. 104, no. 11, pp. 4536–41, 2007.

[69] M. Matzke, T. Kanno, L. Daxinger, B. Huettel, and A. J. M. Matzke, "RNA-mediated chromatin-based silencing in plants." *Current Opinion in Cell Biology*, vol. 21, no. 3, pp. 367–76, 2009.

[70] A. T. Wierzbicki, J. R. Haag, and C. S. Pikaard, "Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes." *Cell*, vol. 135, no. 4, pp. 635–648, 2008.

[71] B. Zheng, Z. Wang, S. Li, B. Yu, J.-Y. Liu, and X. Chen, "Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in *Arabidopsis*." *Genes & Development*, vol. 23, no. 24, pp. 2850–60, 2009.

[72] A. T. Wierzbicki, T. S. Ream, J. R. Haag, and C. S. Pikaard, "RNA polymerase V transcription guides ARGONAUTE4 to chromatin." *Nature Genetics*, vol. 41, no. 5, pp. 630–4, 2009.

[73] M. El-Shami, D. Pontier, S. Lahmy, L. Braun, C. Picart, D. Vega, M.-A. Hakimi, S. E. Jacobsen, R. Cooke, and T. Lagrange, "Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components." *Genes & Development*, vol. 21, no. 20, pp. 2539–44, 2007.

[74] O. Pontes, C. F. Li, P. Costa Nunes, J. Haag, T. Ream, A. Vitins, S. E. Jacobsen, and C. S. Pikaard, "The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center." *Cell*, vol. 126, no. 1, pp. 79–92, 2006.

[75] Z. Gao, H.-L. Liu, L. Daxinger, O. Pontes, X. He, W. Qian, H. Lin, M. Xie, Z. J. Lorkovic, S. Zhang, D. Miki, X. Zhan, D. Pontier, T. Lagrange, H. Jin, A. J. M. Matzke, M. Matzke, C. S. Pikaard, and J.-K. Zhu, "An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation." *Nature*, vol. 465, no. 7294, pp. 106–9, 2010.

[76] C. F. Li, O. Pontes, M. El-Shami, I. R. Henderson, Y. V. Bernatavichute, S. W.-L. Chan, T. Lagrange, C. S. Pikaard, and S. E. Jacobsen, "An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*." *Cell*, vol. 126, no. 1, pp. 93–106, 2006.

[77] C. F. Li, I. R. Henderson, L. Song, N. Fedoroff, T. Lagrange, and S. E. Jacobsen, "Dynamic regulation of ARGONAUTE4 within multiple nuclear bodies in *Arabidopsis thaliana*." *PLoS Genetics*, vol. 4, no. 2, p. e27, 2008.

[78] X. Zhong, J. Du, C. Hale, J. Gallego-Bartolome, S. Feng, A. Vashisht, J. Chory, J. Wohlschlegel, D. Patel, and S. Jacobsen, "Molecular Mechanism of Action of Plant DRM De Novo DNA Methyltransferases." *Cell*, vol. 157, no. 5, pp. 1050–1060, 2014.

[79] J.-B. Morel, "Fertile Hypomorphic *ARGONAUTE (ago1)* Mutants Impaired in Post-Transcriptional Gene Silencing and Virus Resistance." *The Plant Cell*, vol. 14, no. 3, pp. 629–639, 2002.

[80] C. Eun, Z. J. Lorkovic, U. Naumann, Q. Long, E. R. Havecker, S. A. Simon, B. C. Meyers, A. J. M. Matzke, and M. Matzke, "AGO6 functions in RNA-mediated transcriptional gene silencing in shoot and root meristems in *Arabidopsis thaliana*." *PloS One*, vol. 6, no. 10, p. e25730, 2011.

[81] X.-J. He, Y.-F. Hsu, S. Zhu, A. T. Wierzbicki, O. Pontes, C. S. Pikaard, H.-L. Liu, C.-S. Wang, H. Jin, and J.-K. Zhu, "An effector of RNA-directed DNA methylation in *arabidopsis* is an ARGONAUTE 4- and RNA-binding protein." *Cell*, vol. 137, no. 3, pp. 498–508, 2009.

[82] M. Ghildiyal and P. D. Zamore, "Small silencing RNAs: an expanding universe." *Nature Reviews Genetics*, vol. 10, no. 2, pp. 94–108, 2009.

[83] E. R. Havecker, L. M. Wallbridge, T. J. Hardcastle, M. S. Bush, K. A. Kelly, R. M. Dunn, F. Schwach, J. H. Doonan, and D. C. Baulcombe, "The *Arabidopsis* RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci." *The Plant Cell*, vol. 22, no. 2, pp. 321–34, 2010.

[84] H. Zhang and J.-K. Zhu, "Active DNA demethylation in plants and animals." *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 77, no. 0, pp. 161–73, 2012.

[85] J. A. Law, J. Du, C. J. Hale, S. Feng, K. Krajewski, A. M. S. Palanca, B. D. Strahl, D. J. Patel, and S. E. Jacobsen, "Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1." *Nature*, 2013.

[86] A. Zemach, M. Kim, P.-H. Hsieh, D. Coleman-Derr, L. Eshed-Williams, K. Thao, S. Harmer, and D. Zilberman, "The *Arabidopsis* Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin." *Cell*, vol. 153, no. 1, pp. 193–205, 2013.

[87] X.-J. He, Y.-F. Hsu, S. Zhu, H.-L. Liu, O. Pontes, J. Zhu, X. Cui, C.-S. Wang, and J.-K. Zhu, "A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development." *Genes & Development*, vol. 23, no. 23, pp. 2717–22, 2009.

[88] H. Stroud, M. Greenberg, S. Feng, Y. Bernatavichute, and S. Jacobsen, "Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the *Arabidopsis* Methylome." *Cell*, vol. null, no. null, 2013.

[89] H. A. Eskandarian, F. Impens, M.-A. Nahori, G. Soubigou, J.-Y. Coppée, P. Cossart, and M. A. Hamon, "A role for SIRT2-dependent histone H3K18 deacetylation in bacterial infection." *Science*, vol. 341, no. 6145, p. 1238858, 2013.

[90] C. Luo, D. J. Sidote, Y. Zhang, R. A. Kerstetter, T. P. Michael, and E. Lam, "Integrative analysis of chromatin states in *Arabidopsis* identified potential regulatory mechanisms for natural antisense transcript production." *The Plant Journal*, vol. 73, no. 1, pp. 77–90, 2013.

[91] C. M. Rose, S. van den Driesche, R. R. Meehan, and A. J. Drake, "Epigenetic reprogramming: preparing the epigenome for the next generation." *Biochemical Society Transactions*, vol. 41, no. 3, pp. 809–14, 2013.

[92] M. I. Ponferrada-Marín, M. I. Martínez-Macías, T. Morales-Ruiz, T. Roldán-Arjona, and R. R. Ariza, "Methylation-independent DNA binding modulates specificity of Repressor of Silencing 1 (ROS1) and facilitates demethylation in long substrates." *The Journal of Biological Chemistry*, vol. 285, no. 30, pp. 23 032–9, 2010.

[93] E. Y. Harris, N. Ponts, K. G. Le Roch, and S. Lonardi, "BRAT-BW: efficient and accurate mapping of bisulfite-treated reads." *Bioinformatics*, vol. 28, no. 13, pp. 1795–6, 2012.

[94] I. Ausin, M. V. C. Greenberg, D. K. Simanshu, C. J. Hale, A. A. Vashisht, S. A. Simon, T.-f. Lee, S. Feng, S. D. Española, B. C. Meyers, J. A. Wohlschlegel, D. J. Patel, and S. E. Jacobsen, "INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in *Arabidopsis*." *Proceedings of the National Academy of Sciences*, vol. 109, no. 22, pp. 8374–81, 2012.

[95] C.-F. Huang, D. Miki, K. Tang, H.-R. Zhou, Z. Zheng, W. Chen, Z.-Y. Ma, L. Yang, H. Zhang, R. Liu, X.-J. He, and J.-K. Zhu, "A Pre-mRNA-Splicing Factor Is Required for RNA-Directed DNA Methylation in *Arabidopsis*." *PLoS Genetics*, vol. 9, no. 9, p. e1003779, 2013.

[96] C. A. Ibarra, X. Feng, V. K. Schoft, T.-F. Hsieh, R. Uzawa, J. A. Rodrigues, A. Zemach, N. Chumak, A. Machlicova, T. Nishimura, D. Rojas, R. L. Fischer, H. Tamaru, and D. Zilberman, "Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes." *Science*, vol. 337, no. 6100, pp. 1360–4, 2012.

[97] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg, "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." *Genome Biology*, vol. 10, no. 3, p. R25, 2009.

[98] H. H. Zhang, K. Tang, W. Qian, C.-G. Duan, B.-S. B. Wang, P. Wang, X. Zhu, Z. Lang, Y. Yang, and J.-K. Zhu, "An Rrp6-like protein positively regulates non-coding RNA levels and DNA methylation in *Arabidopsis*." *Molecular Cell*, vol. 54, no. 3, pp. 418–30, 2014.

[99] T.-f. Lee, S. G. R. Gurazada, J. Zhai, S. Li, S. A. Simon, M. A. Matzke, X. Chen, and B. C. Meyers, "RNA polymerase V-dependent small RNAs in *Arabidopsis* originate from small, intergenic loci including most SINE repeats." *Epigenetics*, vol. 7, no. 7, pp. 781–95, 2012.

[100] X. Li, W. Qian, Y. Zhao, C. Wang, J. Shen, J.-K. Zhu, and Z. Gong, "Antisilencing role of the RNA-directed DNA methylation pathway and a histone acetyltransferase in *Arabidopsis*." *Proceedings of the National Academy of Sciences*, vol. 109, no. 28, pp. 11 425–30, 2012.

[101] L. Wu, L. Mao, and Y. Qi, "Roles of dicer-like and argonaute proteins in TAS-derived small interfering RNA-triggered DNA methylation." *Plant Physiology*, vol. 160, no. 2, pp. 990–9, 2012.

[102] R. A. Mosher, F. Schwach, D. Studholme, and D. C. Baulcombe, "PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis." *Proceedings of the National Academy of Sciences*, vol. 105, no. 8, pp. 3145–50, 2008.

[103] C. S. Pikaard, J. R. Haag, T. Ream, and A. T. Wierzbicki, "Roles of RNA polymerase IV in gene silencing." *Trends in Plant Science*, vol. 13, no. 7, pp. 390–397, 2008.

[104] P. Dunoyer, G. Schott, C. Himber, D. Meyer, A. Takeda, J. C. Carrington, and O. Voinnet, "Small RNA duplexes function as mobile silencing signals between plant cells." *Science*, vol. 328, no. 5980, pp. 912–6, 2010.

[105] A. Vannini and P. Cramer, "Conservation between the RNA polymerase I, II, and III transcription initiation machineries." *Molecular Cell*, vol. 45, no. 4, pp. 439–46, 2012.

[106] H. Zhang, Z.-Y. Ma, L. Zeng, K. Tanaka, C.-J. Zhang, J. Ma, G. Bai, P. Wang, S.-W. Zhang, Z.-W. Liu, T. Cai, K. Tang, R. Liu, X. Shi, X.-J. He, and J.-K. Zhu, "DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV." *Proceedings of the National Academy of Sciences*, vol. 110, no. 20, pp. 8290–5, 2013.

[107] A. Saleh, R. Alvarez-Venegas, and Z. Avramova, "An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in *Arabidopsis* plants." *Nature Protocols*, vol. 3, no. 6, pp. 1018–25, 2008.

VITA

VITA

Kai Tang

Purdue University

Department of Horticulture and
Landscape Architecture (HLA)

Phone: (765) 496-7603
Email: tang58@purdue.edu
Address: 625 Agriculture Mall Dr.
West Lafayette, IN 47907

Education

Jan. 2011-Dec. 2016	Ph.D. Horticulture and Landscape Architecture, Purdue University Advisor: Dr. Jian-Kang Zhu
May 2016	M.S. Statistics and Computer Science joint program Department of Statistics, Purdue University
Sep. 2009-Dec. 2010	Ph.D. student Genetics, Genomics and Bioinformatics(GGB) Program University of California, Riverside (UCR) Major: Bioinformatics Advisor: Dr. Jian-Kang Zhu Co-advisor: Dr. Renyi Liu
Sep. 2005-Jul. 2009	B.S. School of Life Sciences University of Science and Technology of China Advisor: Dr. Fei Sun

Publications

1. J. Yan, P. Wang, B. Wang, C.-C. Hsu, H. Zhang, Y.-J. Hou, Y. Zhao, **K. Tang**, Q. Wang, C. Zhao, X. Zhu, W. A. Tao, J. Li, and J.-K. Zhu (2016) The SnRK2 kinases modulate miRNA accumulation in *Arabidopsis*. *PLoS Genetics*. (Submitted)
2. **K. Tang**, Z. Lang, H. Zhang, and J.-K. Zhu (2016) The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. *Nature Plants* 2(11):16169.
3. C.-G. Duan, X. Wang, S. Xie, L. Pan, D. Miki, **K. Tang**, C.-C. Hsu, M. Lei, Y. Zhong, Y.-J. Hou, Z. Wang, Z. Zhang, S. Mangrauthia, H. Xu, H. Zhang, B. Dilkes, and W. A. Tao, and J.-K. Zhu (2016) A Pair of Transposon-derived Proteins Function in A Histone Acetyltransferase Complex for Active DNA Demethylation. *Cell Research*. (Accepted)
4. Y. Yang, H. La, **K. Tang**, D. Miki, L. Yang, B. Wang, C.-G. Duan, W. Nie, X. Wang, S. Wang, Y. Pan, E. J. Tran, L. An, H. Zhang, J.-K. Zhu (2016) SAC3B, a central component of the mRNA export complex TREX-2, is required for prevention of epigenetic gene silencing in *Arabidopsis*. *Nucleic Acids Research* gkw850.
5. D.-L. Yang, G. Zhang, **K. Tang**, J. Li, L. Yang, H. Huang, H. Zhang, and J.-K. Zhu (2016) Dicer-independent RNA-directed DNA methylation in *Arabidopsis*. *Cell Research* 26:6682.
6. N. Liu, J. Yang, X. Fu, L. Zhang, **K. Tang**, K. M. Guy, Z. Hu, S. Guo, Y. Xu, and M. Zhang (2016) Genome-wide identification and comparative analysis of grafting-responsive mRNA in watermelon grafted onto bottle gourd and squash rootstocks by high-throughput sequencing. *Molecular Genetics and Genomics* 291(2):621-633.
7. C.-G. Duan*, H. Zhang*, **K. Tang***, X. Zhu, W. Qian, Y.-J. Hou, B. Wang, Z. Lang, Y. Zhao, X. Wang, P. Wang, J. Zhou, G. Liang, N. Liu, C. Wang, and J.-

K. Zhu (2015) Specific but interdependent functions for *Arabidopsis* AGO4 and AGO6 in RNAdirected DNA methylation. *The EMBO Journal* 34(5):581-592.
* Contributed equally

8. Y. Wan*, **K. Tang***, D. Zhang*, S. Xie, X. Zhu, Z. Wang, and Z. Lang (2015) Transcriptome-wide high-throughput deep m^6A-seq reveals unique differential m^6A methylation patterns between three organs in *Arabidopsis thaliana*. *Genome Biology* 16:272.
* Contributed equally

9. C.-G. Duan, X. Wang, **K. Tang**, H. Zhang, S. K. Mangrauthia, M. Lei, C.-C. Hsu, Y.-J. Hou, C. Wang, Y. Li, W. A. Tao, and J.-K. Zhu (2015) MET18 Connects the Cytosolic Iron-Sulfur Cluster Assembly Pathway to Active DNA Demethylation in *Arabidopsis*. *PLoS Genetics* 11(10):e1005559.

10. Z. Lang, M. Lei, X. Wang, **K. Tang**, D. Miki, H. Zhang, S. K. Mangrauthia, W. Liu, W. Nie, G. Ma, J. Yan, C.-G. Duan, C.-C. Hsu, C. Wang, W. A. Tao, Z. Gong, and J.-K. Zhu (2015) The Methyl-CpG-Binding Protein MBD7 Facilitates Active DNA Demethylation to Limit DNA Hyper-Methylation and Transcriptional Gene Silencing. *Molecular Cell* 57(6):971-983.

11. M. Lei, H. Zhang, R. Julian, **K. Tang**, S. Xie, and J.-K. Zhu (2015) Regulatory link between DNA methylation and active demethylation in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 112(11):3553-3557.

12. Y. Li, D. Córdoba-Cañero, W. Qian, X. Zhu, **K. Tang**, H. Zhang, R. R. Ariza, T. Roldán-Arjona, and J.-K. Zhu (2015) An AP Endonuclease Functions in Active DNA Demethylation and Gene Imprinting in *Arabidopsis*. *PLoS Genetics* 11(1):e1004905.

13. W. Qian, D. Miki, M. Lei, X. Zhu, H. Zhang, Y. Liu, Y. Li, Z. Lang, J. Wang, **K. Tang**, R. Liu, and J.-K. Zhu (2014) Regulation of Active DNA Demethylation by an α-Crystallin Domain Protein in *Arabidopsis*. *Molecular Cell* 55(3):361-371.

14. H. Zhang, **K. Tang**, B. Wang, C.-G. Duan, Z. Lang, and J.-K. Zhu (2014) Protocol: a beginners guide to the analysis of RNA-directed DNA methylation in plants. *Plant Methods* 10:18.

15. H. Zhang, **K. Tang**, W. Qian, C.-G. Duan, B. Wang, P. Wang, X. Zhu, Z. Lang, Y. Yang, and J.-K. Zhu (2014) An Rrp6-like Protein Positively Regulates Noncoding RNA Levels and DNA Methylation in *Arabidopsis*. *Molecular Cell* 54(3):418-430.

16. M. Lei, H. La, K. Lu, P. Wang, D. Miki, Z. Ren, C.-G. Duan, X. Wang, **K. Tang**, L. Zeng, L. Yang, H. Zhang, W. Nie, P. Liu, J. Zhou, R. Liu, Y. Zhong, D. Liu, and J.-K. Zhu (2014) *Arabidopsis* EDM2 promotes *IBM1* distal polyadenylation and regulates genome DNA methylation patterns. *Proceedings of the National Academy of Sciences of the United States of America* 111(1):527-532.

17. K. Dou, C.-F. Huang, Z.-Y. Ma, C.-J. Zhang, J.-X. Zhou, H.-W. Huang, T. Cai, **K. Tang**, J.-K. Zhu, and X.-J. He (2013) The PRP6-like splicing factor STA1 is involved in RNA-directed DNA methylation by facilitating the production of Pol V-dependent scaffold RNAs. *Nucleic Acids Research* 41(18):8489-8502.

18. X. Wang, C.-G. Duan, **K. Tang**, B. Wang, H. Zhang, M. Lei, K. Lu, S. K. Mangrauthia, P. Wang, G. Zhu, Y. Zhao, and J.-K. Zhu (2013) RNA-binding protein regulates plant DNA methylation by controlling mRNA processing at the intronic heterochromatin-containing gene *IBM1*. *Proceedings of the National Academy of Sciences of the United States of America* 110(38):15467-15472.

19. C.-F. Huang, D. Miki, **K. Tang**, H.-R. Zhou, Z. Zheng, W. Chen, Z.-Y. Ma, L. Yang, H. Zhang, R. Liu, X.-J. He, and J.-K. Zhu (2013) A Pre-mRNA-Splicing Factor Is Required for RNA-Directed DNA Methylation in *Arabidopsis*. *PLoS Genetics* 9(9):e1003779.

20. H. Zhang, Z.-Y. Ma, L. Zeng, K. Tanaka, C.-J. Zhang, J. Ma, G. Bai, P. Wang, S.-W. Zhang, Z.-W. Liu, T. Cai, **K. Tang**, R. Liu, X. Shi, X.-J. He, and J.-K. Zhu (2013) DTF1 is a core component of RNA-directed DNA methylation and

may assist in the recruitment of Pol IV. *Proceedings of the National Academy of Sciences of the United States of America* 110(20):8290-8295.

21. W. Qian, D. Miki, H. Zhang, Y. Liu, X. Zhang, **K. Tang**, Y. Kan, H. La, X. Li, S. Li, X. Zhu, X. Shi, K. Zhang, O. Pontes, X. Chen, R. Liu, Z. Gong, and J.-K. Zhu (2012) A Histone Acetyltransferase Regulates Active DNA Demethylation in *Arabidopsis*. *Science* 336(6087):1445-1448.

22. M. I. Martínez-Macías, W. Qian, D. Miki, O. Pontes, Y. Liu, **K. Tang**, R. Liu, T. Morales-Ruiz, R. R. Ariza, T. Roldán-Arjona, and J.-K. Zhu (2012) A DNA 3' phosphatase functions in active DNA demethylation in *Arabidopsis*. *Molecular Cell* 45(3):357-370.