A novel protein complex that regulates active DNA demethylation in *Arabidopsis*

Pan Liu[1,3,4], Wen-Feng Nie[1,4], Xiansong Xiong[1,3], Yuhua Wang[1], Yuwei Jiang[1,3], Pei Huang[1,3], Xueqiang Lin[1], Guochen Qin[1], Huan Huang[1], Qingfeng Niu[1], Jiamu Du[1], Zhaobo Lang[1], and Jian-Kang Zhu[1,2,*]

[1]Shanghai Center for Plant Stress Biology, Center of Excellence in Molecular Plant Sciences, Chinese Academy of Sciences (CAS), Shanghai 201602, China

[2]Department of Horticulture & Landscape Architecture, Purdue University, West Lafayette, IN 47906, USA

[3]University of Chinese Academy of Sciences，No.19(A) Yuquan Road, Shijingshan District, Beijing 100049, China

[4]These authors contributed equally

[*]Correspondence: jkzhu@sibs.ac.cn

Pan Liu, liupan@sibs.ac.cn
Wen-Feng Nie, wfnie@sibs.ac.cn
Xiansong Xiong, xiongxiansong@sibs.ac.cn
Yuhua Wang, wangyuhua@sibs.ac.cn
Yuwei Jiang, ywjiang@sibs.ac.cn
Pei Huang, peihuang@psc.ac.cn
Xueqiang Lin, xueqianglin@foxmail.com
Guochen Qin，qinguochen@sibs.ac.cn
Huan Huang, ahuanghuan@163.com
Qingfeng Niu, qfniu@psc.ac.cn
Jiamu Du，dujm@sustech.edu.cn
Zhaobo Lang, zblang@sibs.ac.cn

Jian-Kang Zhu, jkzhu@sibs.ac.cn

Abstract

Active DNA demethylation is critical for altering DNA methylation patterns and regulating gene expression. The 5-methylcytosine DNA glycosylase/lyase ROS1 initiates a base excision repair pathway for active DNA demethylation and is required for the prevention of DNA hypermethylation at thousands of genomic regions in *Arabidopsis*. How ROS1 is regulated and targeted to specific genomic regions is not well understood. Here, we report the discovery of an *Arabidopsis* protein complex that contains ROS1, that regulates *ROS1* gene expression, and that likely targets the ROS1 protein to specific genomic regions. ROS1 physically interacts with a WD40 domain protein (RWD40), which in turn interacts with a methyl-DNA binding protein (RMB1) as well as with a zinc finger- and homeobox domain protein (RHD1). RMB1 binds to DNA that is methylated in any sequence context, and this binding is necessary for RMB1 function *in vivo*. Loss-of-function mutations in *RWD40*, *RMB1*, or *RHD1* cause DNA hypermethylation at several tested genomic regions and in a manner that is independent of the known ROS1 regulator IDM1. Because the hypermethylated genomic regions include the DNA methylation monitoring sequence in the *ROS1* promoter, the mutant plants show increased *ROS1* expression. Our results demonstrate that ROS1 forms a protein complex with RWD40, RMB1, and RHD1, and that this novel complex regulates active DNA demethylation in plants.

Key words: DNA methylation, DNA demethylation, ROS1, WD40 domain, methyl-DNA binding

Introduction

DNA methylation at the fifth position of the cytosine ring is important for gene regulation, transposon silencing, and imprinting in plants and many other eukaryotic organisms (Robertson, 2005; Slotkin and Martienssen, 2007; Law and Jacobsen, 2010; He et al., 2011; Matzke and Mosher, 2014; Zhang et al., 2018). DNA methylation occurs in different sequence contexts, i.e., CG, CHG, and CHH, where H represents A, C, or T. In plants, *de novo* DNA methylation is mediated by the RNA-directed DNA methylation (RdDM) pathway, which involves small interfering RNAs (siRNAs) and scaffold RNAs in addition to an array of proteins (Haag and Pikaard, 2011; Matzke and Mosher, 2014; Zhang et al., 2018). In *Arabidopsis*, CG methylation is maintained by DNA METHYLTRANSFERASE 1 (MET1) (Finnegan and Dennis, 1993), while CHG methylation is maintained by the plant-specific DNA methyltransferase CHROMOMETHYLASE 3 (CMT3) (Cao and Jacobsen, 2002). Asymmetric CHH methylation is maintained by DOMAIN REARRANGED METHYLTRANSFERASE 2 (DRM2) through the RdDM pathway and by CHROMOMETHYLASE 2 (CMT2) (Haag and Pikaard, 2011; Zemach et al., 2013; Matzke and Mosher, 2014).

Specific DNA methylation patterns are tightly regulated by DNA methylation and demethylation pathways (Zhu, 2009; Furner and Matzke, 2011; Zhang et al., 2018). Passive DNA demethylation results from an absence of DNA

methyltransferases or a reduction in their activity, or from a shortage of methyl donors following DNA replication (Zhang et al., 2018). Active DNA demethylation is mediated by 5-methylcytosine DNA glycosylases through a DNA base-excision repair pathway in plants (Gong et al., 2002; Agius et al., 2006; Gehring et al., 2006; Zhu et al., 2007; Zhu, 2009). There are four 5-methylcytosine DNA glycosylases in *Arabidopsis,* including REPRESSOR OF SILENCING 1 (ROS1), DEMETER (DME), DEMETER-LIKE 2 (DML2), and DML3 (Gong et al., 2002; Gehring et al., 2006; Penterman et al., 2007; Hsieh et al., 2009). Because the 5-methylcytosine DNA glycosylases can recognize and directly remove the 5-mC base, they are critical enzymes in the active DNA demethylation pathway and are thus often referred to as DNA demethylases (Zhu, 2009; Zhang et al., 2018). Research on how these DNA demethylases are regulated and recruited to their target loci is required to reveal how distinct genomic DNA methylation patterns are generated and modified during organismal development and in response to environmental change.

The histone acetyltransferase IDM1 is part of the IDM protein complex, which specifically recognizes certain methylated genomic regions through the MBD domains of its subunits MBD7 and IDM1, and through the DNA-binding domain of HDP2 (Duan et al., 2017; Lang et al., 2015; Qian et al., 2014; Qian et al., 2012). Recent research showed that ROS1 physically interacts with the histone variant H2A.Z and is recruited to some genomic regions via SWR1-mediated H2A.Z deposition, which involves specific histone acetylation marks created by the IDM complex (Nie et al., 2019). Because this mechanism only applies to a subset of target genomic regions of ROS1, there must be alternative mechanisms for the targeting of ROS1.

In this study, we found that ROS1 forms a protein complex with RWD40, RMB1 and RHD1, which are proteins that contain WD40, methyl-DNA binding, and homeodomain and zinc finger domains, respectively. We show that this novel protein complex is required for the prevention of DNA hypermethylation at several genomic regions in *Arabidopsis*, and that this function is independent of the IDM-SWR1-H2A.Z pathway. One of these regions is the *ROS1* promoter, which contains the DNA methylation monitoring sequence important for DNA methylation homeostasis in *Arabidopsis* (Lei et al., 2015; Williams et al., 2015). Consistent with its role in controlling the DNA methylation of the *ROS1* promoter, we found that this protein complex negatively regulates the expression of *ROS1*. These results suggest that RWD40, RMB1, and RHD1 proteins control genomic DNA methylation by regulating *ROS1* gene expression and perhaps by targeting the ROS1 protein.

RESULTS

A WD40 protein physically interacts with ROS1 and functions in DNA demethylation

To date, two proteins, H2A.Z and MET18, have been reported to physically interact with ROS1 and to regulate active DNA demethylation in *Arabidopsis* plants (Nie et al., 2019; Duan et al., 2015). H2A.Z is required for ROS1 to target specific genomic regions defined by the IDM and SWR1 protein complexes (Qian et al., 2012; Lang et al., 2015; Nie et al., 2019). MET18 is a critical enzyme in the biosynthesis of the MoCo cofactor required for the methyl-DNA glycosylase/lyase activity of ROS1 (Duan et al., 2015). To identify additional ROS1-interacting

proteins, we used yeast two-hybrid (Y2H) assays to determine whether ROS1 may interact with several proteins that were found in the anti-FLAG pull-down products from ROS1-3xFLAG-3xHA (*ros1-13* mutant expressing 3xFLAG-3xHA-tagged ROS1 under its native promoter) *Arabidopsis* plants (Lei et al., 2015) in preliminary experiments. The Y2H assays led to the identification of RWD40, which is a WD40 domain-containing protein (named RWD40 for ROS1-associated WD40 domain-containing protein) (Figures 1A and 1B). The interaction between ROS1 and RWD40 was confirmed by a bimolecular fluorescence complementation (BiFC) assay (Figure 1C). Transient expression assays in *Nicotiana benthamiana* leaves showed that RWD40 is mainly localized in the nucleus (Figure S1A). In the anti-FLAG immunoprecipitate from ROS1-3xFLAG-3xHA plants (Lei et al., 2015) but not from Col-0 control plants, we identified multiple RWD40 peptides (Figure 1D). Furthermore, we also identified peptides corresponding to ROS1 in the anti-FLAG immunoprecipitate from transgenic plants expressing a 3xFLAG-RWD40 fusion driven by the *RWD40* native promoter in the *rwd40-1* mutant background (Figures1D and S1B-S1D). These results suggest that ROS1 physically interacts with RWD40 in *Arabidopsis*. To investigate whether RWD40 regulates ROS1-dependent active DNA methylation, we used chop-PCR to determine the DNA methylation level at the AT5G39160 locus in *rwd40-1* and *rwd40-2* mutant seedlings. Similar to *ros1* mutant plants (Qian et al., 2012), *rwd40-1* and *rwd40-2* plants showed increased DNA methylation at the 5' region of AT5G39160 (Figure S1E). Individual locus bisulfite sequencing confirmed that the DNA methylation levels at the AT5G39160 locus were higher in *rwd40-1*, *rwd40-2*, and *ros1-4* mutant plants than in the wild-type control (Figure S1F). Interestingly, the methylation level at AT5G39160 was not increased in the *idm1-2* mutant (Figures S1E-S1G). The expression of

RWD40-3xFLAG fusion from its native promoter in the *rwd40-1* mutant reduced the methylation level of the *rwd40-1* mutant (Figure S1G). These results show that RWD40 interacts with ROS1 and functions in regulating DNA demethylation.

RMB1 and RHD1 interact with RWD40 and function in DNA demethylation

We noticed that the anti-FLAG immunoprecipitates from 3xFLAG-RWD40 transgenic plants contained peptides corresponding to AT1G63240 and AT5G42780, in addition to peptides from RWD40 and ROS1 (Figure 1D). AT1G63240 and AT5G42780 peptides were also detected in the anti-FLAG immunoprecipitates from ROS1-3xFLAG-3xHA transgenic plants (Figure 1D). The AT1G63240 protein is predicted to have a non-canonical methyl-DNA binding domain, a CBD domain, and a domain of unknown function, and is hereafter referred to as RMB1 (for ROS1-associated methyl-DNA binding protein 1) (Figure S2A). AT5G42780 contains a zinc-finger domain and a homeodomain, and is hereafter referred to as ROS1-associated homeodomain protein 1 (RHD1) (Figure S2B). In Y2H assays, RWD40 interacted with RMB1 and RHD1, and RMB1 interacted with RHD1 (Figure 2A). These interactions as well as the interaction between RWD40 and ROS1 were confirmed by split luciferase and BiFC assays in tobacco leaves (Figures 2B and S2C). The assays also showed that neither RMB1 nor RHD1 could directly interact with ROS1 (Figures 2A，2B and S2C).

RMB1 was immunoprecipitated from transgenic plants expressing a 3xFLAG-RMBD1 fusion driven by the RMB1 native promoter in the *rmb1-1* mutant background (Figure S2A). The immunoprecipitate contained not only RMB1 but also RWD40, RHD1, and ROS1 (Figure 1D). Similarly, the RHD1 immunoprecipitate from transgenic plants expressing a 3xFLAG-RHD1 fusion

driven by the RHD1 native promoter in the *rhd1-1* background (Figure S2B) contained not only RHD1 but also RMB1, ROS1, and RWD40 (Figure 1D). These results show that RWD40, RMB1, and RHD1 proteins are associated with ROS1 *in vivo*.

Dysfunction of *RMB1* caused DNA hypermethylation at the AT5G39160 locus (Figures S2D and S2E). The DNA hypermethylation phenotype of the *rmb1-1* mutant was suppressed in transgenic plants expressing 3xFLAG-RMB1 (Figure S2F). The *rhd1-1* mutant did not show a DNA hypermethylation phenotype at the AT5G39160 locus (Figure S2G). We assessed the methylation levels at another endogenous ROS1 target, AT4G18380, by chop-PCR. Similar to *ros1-4* mutant plants, *rhd1-1* mutant plants had increased DNA methylation at AT4G18380 (Figure S2G). Expression of the 3xFLAG-RHD1 transgene suppressed this hypermethylation phenotype of the *rhd1-1* mutant (Figure S2G). These results suggest that RMB1 and RHD1 function in regulating locus-specific demethylation in *Arabidopsis.*

RWD40, RMB1, RHD1, and ROS1 form a protein complex

The interactions of RWD40 with ROS1, RMB1, and RHD1, and between RMB1 and RHD1 in the Y2H, BiFC, and split luciferase assays, together with their *in vivo* association as indicated by immunoprecipitation-mass spectrometry analyses, suggested that the four proteins may form a protein complex in plant cells. To assess this possibility, we carried out gel filtration assays with protein extracts from 3xFLAG-RMB1, 3xFLAG-RWD40, 3xFLAG-RHD1, and ROS1-3xFLAG-3xHA transgenic plants. The results showed that RMB1, RWD40, RHD1, and ROS1 were eluted in the same fractions, and indicated that they exist

in a complex with an estimated size of approximately 350 kDa (Figure 3A). This is close to the predicted total size of 328 kDa if the four proteins are in a 1:1:1:1 ratio. We co-expressed full-length RWD40, RMB1, and RHD1, and a truncated ROS1 protein (amino acid residues 1 to 100) in Sf9 insect cells using baculovirus vectors. After conducting histidine affinity purification and gel filtration, we found that the four proteins were eluted in a single peak fraction (Figure 3B). All the gel bands (Figure 3B) were extracted and their identities were confirmed by mass spectrometry (Table S1). Taken together, these results suggest that RWD40, RMB1, RHD1, and ROS1 form a protein complex *in vivo*.

Protein domains that mediate the interactions within the protein complex

ROS1 contains an HhH-GPD domain, an End-III domain, a CXXC domain, an RRM_DME domain, and an N-terminal domain of unknown function (Figure S3A). The ROS1 N-terminal region (amino acid residues 41 to 100) has sequence similarity to the EAR (Ethylene response factor–associated Amphiphilic Repression) motif (Yang et al., 2018) (Figure S3A), and is thus referred to as the ELD (EAR like Domain) domain. To determine which domain in ROS1 may mediate the interaction between ROS1 and RWD40, we performed Y2H assays with various deletion mutants of ROS1. We found that the N-terminal half (amino acid residues 1 to 696) of ROS1 interacted with RWD40, whereas the C-terminal half (amino acid residues 697 to 1393) did not (Figure S3B). Further truncation of the N-terminal half of ROS1 showed that the ELD domain (amino acid residues 41 to 100) was sufficient to interact with RWD40 (Figure S3B).

We also performed Y2H assays with various deletion mutants of RWD40, and found that the N-terminal half including two LISH-CTLH domains but not the

WD40 domain of RWD40 interacted with ROS1 (Figures 1B and S3B). On the other hand, the WD40 domain of RWD40 interacted with RMB1 in Y2H assays (Figure 2D). The C-terminal region of RMB1 interacted with the WD40 domain of RWD40 (Figures 2C and 2D), and is thus referred to as the WID (WD40 interaction domain) domain (Figure 2C). The WD40 domain of RWD40 was also capable of mediating the interaction between RWD40 and RHD1 (Figures S3C and S3D). The zinc-finger domain of RHD1 was responsible for the interaction with RWD40, while both the zinc finger domain and homeodomain of RHD1 were necessary for the interaction with RMB1 (Figures S3C and S3E). In the Y2H assay, RHD1 also interacted with itself, and this self-interaction was mediated through the zinc finger domain (Figures S3C and S3D).

RMB1 binds to methylated DNA through the MBD domain

Arabidopsis has 13 canonical MBD proteins (Zemach and Grafi, 2007). Our analysis indicates that Arabidopsis also has three non-canonical MBD proteins, including RMB1, IDM1, and the protein encoded by AT4G14920 (Figure S4A). The IDM complex contains two MBD domain proteins, i.e., the histone acetyltransferase IDM1 and MBD7 (Qian et al., 2012; Lang et al., 2015). The two MBD domain proteins recognize methylated cytosine and thus ensure that the IDM histone acetyltransferase complex is directed only to methylated sequences (Lang et al., 2015; Qian et al., 2012). Many of the amino acid residues in the MBD domain of RMB1 are not conserved compared to the other MBD proteins (Figure S4B). We carried out electrophoretic mobility shift assays (EMSA) to determine whether RMB1 can bind to methylated DNA. The EMSA assays showed that RMB1 was capable of binding to DNA probes containing CG, CHG, or CHH methylation, and that the binding was competitively blocked by unlabeled

methylated DNA but not by unmethylated DNA of the same sequence (Figures 4A and 4B). To confirm RMB1 binding to methylated DNA, we carried out microscale thermophoresis (MST) assays and found that the MBD domain of RMB1 was capable of binding to DNA methylated in any sequence context (Figure 4C). We constructed four mutant versions of the MBD domain, including W22G, Y38F, T49A, and K50T (Figure S4B), and found that the W22G, Y38F, and T49A mutations decreased the methyl-DNA binding activity, while the K50T mutation abolished the methyl-DNA binding activity (Figure 4C). The inability of the K50T mutant version of RMB1 to bind to DNA containing CG, CHG, or CHH methylation was confirmed by EMSA assays (Figure S5A). Expression of the wild type but not of the K50T mutant version of RMB1 under its native promoter complemented the At5g39160 DNA hypermethylation phenotype of the *rmb1-1* mutant (Figure S5B). These results show that RMB1 is a novel methyl-DNA-binding protein and that its methyl-DNA binding is critical for RMB1 function in active DNA demethylation.

RWD40, RMB1, and RHD1 regulate *ROS1* expression

To begin to explore the function of the protein complex of RWD40, RMB1, RHD1 and ROS1 in plants, we assessed the DNA methylation levels at the *ROS1* promoter region, which contains a DNA methylation monitoring sequence (MEMS) that senses methylation and demethylation activities and that regulates *ROS1* expression (Lei et al., 2015; Williams et al., 2015). Individual locus bisulfite sequencing showed that the DNA methylation level at the MEMS was increased in *rwd40-1*, *rmb1-1*, *rhd1-1*, and *ros1-4* compared to the wild-type control (Figure 5A). In contrast, the MEMS DNA methylation level was not increased in the *idm1-2* mutant (Figure 5A). Consistent with the changes in DNA methylation levels at the MEMS, *ROS1* expression was increased in *rwd40*, *rmb1*, and *rhd1*

mutant plants but not in *idm1-2* mutant plants (Figure 5B). In the *ros1-4* mutant, *ROS1* expression measured using the 5' primer pair (Figure 5B) was also increased, as expected (Lei et al., 2015). The expression of the 3xFLAG-RWD40 transgene driven by the *RWD40* native promoter in the *rwd40-1* mutant restored the expression of *ROS1* to the wild-type level (Figure 5B). These results show that RWD40, RMB1, and RHD1 regulate the expression of *ROS1* by modulating the methylation level of MEMS at the *ROS1* promoter.

RWD40, RMB1, and RHD1 regulate DNA demethylation in an IDM1-independent manner

RWD40, *RMB1*, *RHD1*, and *ROS1* but not *IDM1* were necessary for the prevention of DNA hypermethylation at the MEMS in the *ROS1* promoter (Figure 5A). The results suggested that RWD40, RMB1, and RHD1 might regulate ROS1-mediated DNA demethylation in a manner independent of IDM1. To further investigate the role of RWD40, RMB1, and RHD1 in the regulation of active DNA demethylation, we used chop-PCR to assess the DNA methylation levels in *rwd40-1*, *rmb1-1*, and *rhd1-1* single mutants and in the *rwd40 rmb1 rhd1* triple mutant, as well as in *idm1-2* at several genomic targets of ROS1. Similar to DNA methylation levels in the *ros1-4* mutant, DNA methylation levels at the AT1G35140, AT4G00660, AT4G18380, and AT5G55875 loci were increased in *rwd40-1*, *rmb1-1*, and *rhd1-1* but not *idm1-2* mutant plants relative to the Col-0 control (Figure 5C). Individual locus bisulfite sequencing data confirmed that the DNA methylation levels at these ROS1 targets were increased in *rwd40-1*, *rmb1-1*, *rhd1-1,* and *ros1-4* mutant plants but not in *idm1-2* mutant plants (Figures 5D-5G). Analysis of the *rwd40 rmb1 rhd1* triple mutant indicated that the *rwd40-1*, *rmb1-1*, and *rhd1-1* mutations are not additive in causing the DNA

hypermethylation phenotype (Figures 5D-5G). We also determined the DNA methylation levels at several genomic regions (including DT-75, DT-76, DT-77, and DT-78) that are known to be hypermethylated in *idm1* and *ros1* mutants (Qian et al., 2012) and that are amenable to chop-PCR assays (Duan et al., 2017). The data showed that the DNA methylation levels were not increased at these regions in *rwd40-1*, *rmb1-1*, or *rhd1-1* mutant plants (Figure 5H). Together, these results support the inference that RWD40, RMB1, and RHD1 function in regulating active DNA demethylation in a manner independent of IDM1.

ROS1 and the IDM1 complex are required for the prevention of silencing of the *35S:SUC2* reporter gene in *Arabidopsis* (Duan et al., 2017; Lang et al., 2015; Nie et al., 2019; Qian et al., 2014). When we introgressed the *35S:SUC2* reporter gene into *rwd40-1*, *rmb1-1*, or *rhd1-1* mutant plants through genetic crosses, we found that the reporter gene was not silenced in any of these mutants (Figure S6). Therefore, unlike IDM1 or ROS1, RWD40, RMB1 and RHD1 are not required for the prevention of silencing of the *35S:SUC2* reporter gene in Arabidopsis, further indicating that RWD40, RMB1, and RHD1 function in an IDM1-independent manner.

DISCUSSION

In this study, we identified RWD40, RMB1, and RHD1 as cellular factors critical for the regulation of *ROS1* expression and for the prevention of DNA hypermethylation at several endogenous genomic regions (Figure 5). DNA methylation patterns are important for organismal development, carcinogenesis and many other diseases, and even for human aging since changes in DNA methylation levels in a number of genomic loci have been found to correlate

nearly perfectly with age (Jung and Pfeifer, 2015; Klutstein et al., 2016). It is therefore important to understand how DNA methylation patterns are controlled by active DNA demethylation (Zhang et al., 2019). Although the biochemistry of ROS1-mediated enzymatic removal of DNA methylation has been extensively studied, the mechanisms by which the enzymatic machinery is regulated and recruited to specific target sites are poorly understood (Lang et al., 2015; Nie et al., 2019; Qian et al., 2012). Our results indicate that RWD40, RMB1, and RHD1 contribute to locus-specific DNA demethylation and in regulating the expression of *ROS1*. These findings provide insights into the regulation and targeting of active DNA demethylation in plants.

In mammals, active DNA demethylation is initiated by the Tet dioxygenases (Wu and Zhang, 2017). Several transcription factors have been shown to interact with and to target Tet enzymes to genes critical for cell differentiation and reprogramming (Costa et al., 2013; de la Rica et al., 2013; Sardina et al., 2018; Wang et al., 2015; Xiong et al., 2016). In *Arabidopsis*, the IDM histone acetyltransferase complex is required for directing ROS1 to a subset of active DNA demethylation target regions (Duan et al., 2017; Lang et al., 2015; Nie et al., 2019; Qian et al., 2014; Qian et al., 2012). MBD7 in the IDM complex recognizes dense mCpG sites and thus ensures that ROS1 is eventually targeted to genomic regions with densely methylated CpG sequences (Lang et al., 2015). On the other hand, the SANT/Myb/trihelix DNA-binding motif-containing protein HDP2 in the IDM complex probably helps direct the complex to possible regulatory sequences, thus avoiding heavily methylated transposon body regions and some genic regions that should not be demethylated (Duan et al., 2017). The IDM complex creates histone acetylation marks that attract the SWR complex, which then deposits the histone variant H2A.Z. Once deposited at the targeted site, H2A.Z

helps to recruit ROS1 through a direct physical interaction (Nie et al., 2019). The protein complex identified in this study is similar to the IDM complex in that it also contains a protein (i.e., RMB1) that recognizes methylated DNA and a transcription factor-like protein (i.e., RHD1). RMB1 ensures that the complex is directed to methylated genomic regions, while RHD1 probably helps to target the complex to regulatory sequences. Like the IDM complex, the RWD40 protein complex is likely also important for targeting active DNA demethylation to specific genomic regions (Figures 5A and 5C-5G). Our results suggest that the RWD40 complex targets active DNA demethylation to genomic regions that are different from those targeted by the IDM complex. This difference can be explained by the fact that RMB1 recognizes mC marks in all sequence contexts (Figure 4), whereas MBD7 and IDM1 bind to mCpG sequences only (Lang et al., 2015; Qian et al., 2012) (Figures S5C and S5D). In addition, HDP2 and RHD1 likely have different sequence specificities in their interactions with DNA, which probably also contributes to the different targeting preferences of the two protein complexes. The lack of DNA hypermethylation at the AT5G39160 locus in *rhd1* mutant plants (Figure S2G) indicated that RHD1 may be dispensable for some targets of the RWD40 protein complex, either because of genetic redundancy with another transcription factor-like protein, or because it is simply not needed at some genomic regions.

Another important difference between the RWD40 complex and the IDM complex is that the RWD40 complex contains ROS1 and thus has a direct role in targeting active DNA demethylation. Furthermore, because the RWD40 complex targets the DNA methylation-monitoring sequence in the *ROS1* promoter (Figure 5A), it can regulate *ROS1* gene expression (Figure 5B). Through the regulation of *ROS1*

gene expression, the RWD40 complex is expected to have a broad effect on genomic DNA methylation patterns. The genome-wide effects of DNA demethylation factors can be difficult to analyze because DNA methylation changes caused by mutations in these factors are often quite subtle (Nie et al., 2019). Future studies are needed to determine the genome-wide effects of *rwd40*, *rmb1,* and *rhd1* mutations.

The current research, together with our previous work (Lang et al., 2015; Nie et al., 2019), suggests that active DNA demethylation activities are targeted only to methylated genomic regions by proteins that bind to methyl-DNA. Such targeting makes sense, because only methylated genomic regions may require demethylation. Our findings also suggest that the presence of a transcription factor-like protein is important for the targeting of active DNA demethylation. Perhaps the transcription factor-like proteins preferentially bind to regulatory sequences and thus help target active DNA demethylation to regulatory sequences in order to prevent harmful silencing.

SUPPLEMENTAL INFORMATION

Supplemental Information includes 6 figures and 2 tables, and can be found with this article online.

AUTHOR CONTRIBUTIONS

J.-K.Z. and P.L. designed the study. P.L., W.-F.N., X.X., Y.W., Y.J., P.H., X.L., G.Q., H.H., J.D., and Z.L. performed the experiments. W.-F.N., P.L., and J.-K.Z analyzed the data and wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by the Chinese Academy of Sciences (to J.-K.Z.).

STAR METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Flag antibody	Sigma	Cat # F1804
Anti-His antibody	Novoprotein	Cat # AB002-01A
Anti-Myc antibody	Millipore	Cat # 05-724
Chemicals, Peptides, and Recombinant Proteins		
5-Aza-2'-deoxycytidine	Sigma	Cat # A3656
DO Supplement-Leu/-Trp	Clontech	Cat # 630417
DO Supplement-His/-Leu/-Trp	Clontech	Cat # 630419
DO Supplement--Ade/-His/-Leu/-Trp	Clontech	Cat # 630322
pENTR/D-TOPO	Invitrogen	Cat # K240020
Protease Inhibitor Cocktail	Sigma	Cat # P9599
iQ SYBR green supermix	Bio-RAD	Cat # 1708880
Critical Commercial Assays		
RNeasy Plant Mini Kit	Qiagen	Cat # 74904
DNeasy Plant Mini Kit	Qiagen	Cat # 69104
BisulFlash DNA Modification Kit	Epigentek	Cat # P-1026-050
pMD18-T Cloning Kit	Takara.	Cat # 6011
TransScript® First-Strand cDNA Synthesis SuperMix	TRNSGENE Bio.	Cat # AT301-02
Oligonucleotides		
See Table S2		
Experimental Models: Organisms/Strains		
Arabidopsis thaliana accession Col-0	N/A	N/A
ros1-13/pROS1:ROS1-3xFLAG-3xHA	Lei et al., 2015	N/A
rwd40-1/RWD40-3xFLAG	This paper	N/A
rmb1-1/RMB1-3xMYC	This paper	N/A

rhd1-1/RHD-3XMYC	This paper	N/A
Escherichia coli BL21	Takara	D90120-9125
Escherichia coli DH5α	N/A	N/A
Saccharomyces cerevisiae: Y2H Gold	Clontech	Cat # 630489
A. tumefaciens GV3101	N/A	N/A
Recombinant DNA		
Plasmid: AD-ROS1	This paper	N/A
Plasmid: BD-ROS1	This paper	N/A
Plasmid: BD-RWD40	This paper	N/A
Plasmid: BD-RMB1	This paper	N/A
Plasmid: BD-RHD1	This paper	N/A
Plasmid: AD-RWD40	This paper	N/A
Plasmid: AD-RMB1	This paper	N/A
Plasmid: AD-RHD1	This paper	N/A
Plasmid: YFP-N-ROS1	This paper	N/A
Plasmid: YFP-N-RMB1	This paper	N/A
Plasmid: YFP-N-RHD1	This paper	N/A
Plasmid: YFP-C-RWD40	This paper	N/A
Plasmid: YFP-C-RMB1	This paper	N/A
Plasmid: YFP-C-RHD1	This paper	N/A
Plasmid: pDEX-RWD40-YFP	This paper	N/A
Plasmid: 35S:GFP-RWD40	This paper	N/A
Plasmid: pCAMBIA-cLuc-ROS1	This paper	N/A
Plasmid: pCAMBIA-cLuc-RWD40	This paper	N/A
Plasmid: pCAMBIA-cLuc-RMB1	This paper	N/A
Plasmid: pCAMBIA-cLuc-RHD1	This paper	N/A
Plasmid: pCAMBIA-nLuc-RWD40	This paper	N/A
Plasmid: pCAMBIA-nLuc-RMB1	This paper	N/A
Plasmid: pCAMBIA-nLuc-RHD1	This paper	N/A

Software and Algorithms		
Kismeth	http://katahdin.mssm.edu/kismeth/revpage.pl	N/A
NanoTemper Technologies	(Harris et al., 2018)	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further requests for reagents and resources should be directed to and will be fulfilled by the Lead Contact, Jian-Kang Zhu (jkzhu@sibs.ac.cn).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plant materials and growth conditions

Plants were grown under long-day conditions (16-h light/8-h dark) at 22°C. The T-DNA insertion mutants *rwd40-1* (SALK_068825), *rwd40-2* (SALK_012947), *rmb1-1* (SALK_110885), *rmb1-2* (SALK_100783), *rhd1-1* (SALK_092897), *ros1-4* (SALK_045303), and *idm1-2* (SALK_062999) were obtained from the *Arabidopsis* Biological Resource Center (http://www.arabidopsis.org) and were genotyped by PCR with the primers listed in Table S2. For 5-Aza treatments, 5 μM 5-Aza was added to 1/2 MS medium containing 1% glucose. An equal volume of DMSO was added to the medium as the control. Two-week-old plants grown on MS plates were imaged for luminescence after cold treatment at 4°C for 2 days.

METHOD DETAILS

Mutant plant complementation

For the complementation of mutants, *RWD40*, *RMB1*, or *RHD1* genomic DNA with the 2-kb upstream region (as the native promoter region) was amplified from genomic DNA of Col-0 with the primers listed in Table S2. The amplification products were cloned into the pCambia1305 vector (with a 3xFLAG tag at the C-terminus) by T4 DNA ligase (NEB). The constructs were transformed into mutants using *Agrobacterium tumefaciens* GV3101 by the standard floral dip

452 method (Clough and Bent, 1998). Unsegregated T3 homozygous
453 complementation lines were then used for further experiments.

454 **Real-time qPCR and chop-PCR**

455 For real-time qPCR analysis, total RNA was extracted from 0.1 g of 14-day-old
456 *Arabidopsis* seedlings with the RNeasy plant kit (Qiagen). A 2-μg quantity of the
457 mRNA was converted into cDNA with the First-Strand cDNA Synthesis Kit
458 (TRNSGENE Bio.) following the manufacturer's instructions, and the cDNAs were
459 used as templates for real-time PCR with iQ SYBR green supermix (Bio-RAD).
460 For chop-PCR, genomic DNA was extracted from 14-day-old seedlings, and 500
461 ng of DNA was digested with the indicated methylation-sensitive enzyme in a
462 20-μL reaction mixture. After digestion, PCR was performed using 2 μl of the
463 digested DNA as template in a 20-μl reaction mixture with the primers listed in
464 Table S2. Undigested DNA was amplified as the loading control.

465 **IP and LC-MS/MS analysis**

466 For IP, about 5 g of floral tissue for each epitope-tagged transgenic line was used.
467 Dynabeads (10003D, Invitrogen) conjugated with FLAG antibody (F1804, Sigma)
468 were applied for IP. Affinity purification was performed as previously described
469 (Law et al., 2010), and the protein samples were subjected to LC-MS/MS analysis
470 as previously described (Lang et al., 2015; Nie et al., 2019; Qin et al., 2017).

471 **Protein extraction and western blot detection**

472 About 0.1 g of each transgenic plant was harvested and ground to a fine power in
473 liquid N_2. Total protein was extracted by protein lysate buffer (LB: 0.5 mM DTT, 5
474 mM $MgCl_2$, 50 mM Tris [pH 7.6], 10% glycerol, 150 mM NaCl, 0.1% NP-40, 1 mM
475 PMSF, and protease inhibitor cocktail (Roche)). The proteins were then detected
476 by western blot.

477 **Split luciferase complementation assays**

The coding sequences of RWD40, RMB1, RHD1, and ROS1 proteins were cloned into pCAMBIA-cLUC and/or pCAMBIA-nLUC vectors. *Agrobacterium tumefaciens* GV3101 carrying different constructs were infiltrated into 4-weeks-old *N. benthamiana* leaves. Luciferase activity was detected 48 h post infiltration.

Individual locus bisulfite sequencing

A 200-ng quantity of genomic DNA extracted from indicated mutants was treated with the BisulFlash DNA Modification Kit (Epigentek) according to the manufacturer's protocol. A 2-μL volume of bisulfite-treated DNA was used for PCR amplification with the primers listed in Table S2. The PCR products were cloned into the pMD18-T vector (Takara) according to the supplier's instructions. At least 15 independent clones in each sample were sequenced and analyzed using the online tool Kismeth (http://katahdin.mssm.edu/kismeth/revpage.pl).

Bimolecular fluorescence complementation (BiFC) assay

The coding sequences of RWD40, RMB1, RHD1, and ROS1 were cloned into the p2YN or p2YC vector to generate fused split YFP constructs. For protein–protein interaction analysis, *Agrobacterium tumefaciens* GV3101 carrying the indicated constructs were cultured overnight. When the cultures obtained an OD_{600} of 1.0, they were suspended in buffer containing 10 mM $MgCl_2$, 100 μM acetosyringone, and 10 mM MES (pH 5.6). The suspensions were then infiltrated into *N. benthamiana* leaves. Fluorescence was examined at 2 days post infiltration.

Gel filtration assays and western blot analysis

Western blotting of gel filtration assays was performed as previously described (Duan et al., 2017). In brief, 2 g of flower tissue from Col-0 or from the indicated transgenic plants was harvested and ground into a fine power in liquid N_2. The fine powder was suspended in 3 ml of lysis buffer (0.5 mM DTT, 5 mM MgCl2, 50

mM Tris [pH 7.6], 10% glycerol, 150 mM NaCl, 0.1% NP-40, 1 mM PMSF), and the suspension was kept at 4°C for 5 min without shaking. The supernatant was loaded into a HiPrep 16/60 Sephacryl S-300 HR column (GE Healthcare), and 10 fractions were collected. Each fraction was run on 10% SDS–PAGE for western blot detection.

Y2H assay

In brief, the cDNA sequences were cloned into pGADT7-AD or pGBKT7-BD vectors (Clontech), and the pair of genes to be tested for interaction were co-transformed into the yeast strain Gold (Clontech). GAL4 activation domain-linked *Arabidopsis* cDNA libraries were ordered from a commercial company (Clontech). Y2H assays were performed as previously described (Bai et al., 2013).

Protein expression and purification

Arabidopsis RWD40, RMB1, and RHD1 were cloned into pFastBac1 with Flag tag on the N-terminus. Arabidopsis ROS1 (amino acids 1-100) was cloned into pFastBac1 with a double His tag (6xHis-13aa-10xHis) fused with GFP on the N-terminus. Recombinant baculoviruses were generated by the Bac-to-Bac system in Sf9 insect cells. For protein co-expression, the insect cells were grown to a density of 2.0E+06 cells per mL and were then infected with four separate viruses. The infected cells were cultured for 60–72 h at 27°C before collection. Collected cell pellets were stored at −80°C before use. Each cell pellet was resuspended in buffer A (20 mM HEPES pH 7.5, 150 mM NaCl, 10% v/v glycerol, 50 mM imidazole, and 0.1 mM Tris (2-carboxyethyl) phosphine hydrochloride (TCEP), supplemented with protease inhibitor cocktail and then lysed by sonication. The suspension was centrifuged at 18,000 × *g* for 30 min at 4°C. The supernatant was incubated with Ni-NTA resin at 4°C for 1 h. The resin was

washed with 20 column volumes of buffer A in a gravity column. The complex was eluted by buffer B (buffer A with 250 mM imidazole) and was then loaded onto a Superdex200 10/300 GL column equilibrated with buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, and 0.1 mM TCEP. Fractions containing the ROS1 complex were collected and analyzed with SDS-PAGE and Coomassie brilliant blue staining.

Microscale thermophoresis (MST)

An MBD domain construct of *Arabidopsis thaliana RMB1* (amino acids 135-669) was cloned into a self-modified pFast-Bac-MBP vector to generate the His-MBP-tagged target protein. Protein expression was assessed using the standard Bac-to-Bac baculovirus expression system in Sf9 insect cells. The recombinant expressed proteins were purified with a HisTrap column (GE Healthcare) and were further purified with Heparin column and Superdex G200column (GE Healthcare). All of the mutant proteins were expressed and purified using the same protocol as their wild-type counterparts. The detailed procedures were previously described (Harris et al., 2018).

References

Agius, F., Kapoor, A., and Zhu, J.K. 2006. Role of the Arabidopsis DNAglycosylase/lyase ROS1 in active DNA demethylation. Proc. Natl. Acad. Sci. U S A 103:11796–801

Bai, G., Yang, D.H., Zhao, Y., Ha, S., Yang, F., Ma, J., Gao, X.S., Wang, Z.M., and Zhu, J.K. (2013). Interactions between soybean ABA receptors and type 2C protein phosphatases. Plant Mol Biol *83*, 651-664.

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

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

560 Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for
561 Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J *16*, 735-743.
562 Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P.V., Fidalgo, M.,
563 Saunders, A., Lawrence, M., Dietmann, S., *et al.* (2013). NANOG-dependent function of
564 TET1 and TET2 in establishment of pluripotency. Nature *495*, 370-374.
565 de la Rica, L., Rodriguez-Ubreva, J., Garcia, M., Islam, A.B., Urquiza, J.M., Hernando, H.,
566 Christensen, J., Helin, K., Gomez-Vaquero, C., and Ballestar, E. (2013). PU.1 target
567 genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in
568 monocyte-to-osteoclast differentiation. Genome Biol *14*, R99.
569 Duan, C.G., Wang, X., Tang, K., Zhang, H., Mangrauthia, S.K., Lei, M., Hsu, C.C., Hou, Y.J.,
570 Wang, C., Li, Y., *et al.* (2015). MET18 Connects the Cytosolic Iron-Sulfur Cluster
571 Assembly Pathway to Active DNA Demethylation in Arabidopsis. PLoS Genet *11*,
572 e1005559.
573 Duan, C.G., Wang, X., Xie, S., Pan, L., Miki, D., Tang, K., Hsu, C.C., Lei, M., Zhong, Y., Hou,
574 Y.J., *et al.* (2017). A pair of transposon-derived proteins function in a histone
575 acetyltransferase complex for active DNA demethylation. Cell Res *27*, 226-240.
576 Finnegan, E.J., and Dennis, E.S. (1993). Isolation and identification by sequence
577 homology of a putative cytosine methyltransferase from Arabidopsis thaliana. Nucleic
578 Acids Res *21*, 2383-2388.
579 Furner, I.J., and Matzke, M. (2011). Methylation and demethylation of the Arabidopsis
580 genome. Curr Opin Plant Biol *14*, 137-141.
581 Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B.,
582 and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb
583 gene self-imprinting by allele-specific demethylation. Cell *124*, 495-506.
584 Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., David, L., and Zhu, J.K. (2002).
585 ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA
586 glycosylase/lyase. Cell *111*, 803-814.
587 Haag, J.R., and Pikaard, C.S. (2011). Multisubunit RNA polymerases IV and V:
588 purveyors of non-coding RNA for plant gene silencing. Nat Rev Mol Cell Biol *12*,
589 483-492.
590 Harris, C.J., Scheibe, M., Wongpalee, S.P., Liu, W., Cornett, E.M., Vaughan, R.M., Li, X.,
591 Chen, W., Xue, Y., Zhong, Z., *et al.* (2018). A DNA methylation reader complex that
592 enhances gene transcription. Science *362*, 1182-1186.
593 He, X.J., Chen, T., and Zhu, J.K. (2011). Regulation and function of DNA methylation in
594 plants and animals. Cell Res *21*, 442-465.
595 Hsieh, T.F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and
596 Zilberman, D. (2009). Genome-wide demethylation of Arabidopsis endosperm.
597 Science *324*, 1451-1454.
598 Jung, M., and Pfeifer, G.P. (2015). Aging and DNA methylation. BMC Biol *13*, 7.

599 Klutstein, M., Nejman, D., Greenfield, R., and Cedar, H. (2016). DNA Methylation in
600 Cancer and Aging. Cancer Res *76*, 3446-3450.
601 Lang, Z., Lei, M., Wang, X., Tang, K., Miki, D., Zhang, H., Mangrauthia, S.K., Liu, W., Nie,
602 W., Ma, G., *et al.* (2015). The methyl-CpG-binding protein MBD7 facilitates active DNA
603 demethylation to limit DNA hyper-methylation and transcriptional gene silencing. Mol
604 Cell *57*, 971-983.
605 Law, J.A., Ausin, I., Johnson, L.M., Vashisht, A.A., Zhu, J.K., Wohlschlegel, J.A., and
606 Jacobsen, S.E. (2010). A protein complex required for polymerase V transcripts and
607 RNA- directed DNA methylation in Arabidopsis. Curr Biol *20*, 951-956.
608 Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA
609 methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.
610 Lei, M., Zhang, H., Julian, R., Tang, K., Xie, S., and Zhu, J.K. (2015). Regulatory link
611 between DNA methylation and active demethylation in Arabidopsis. Proc Natl Acad
612 Sci U S A *112*, 3553-3557.
613 Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic
614 pathway of increasing complexity. Nat Rev Genet *15*, 394-408.
615 Nie, W.F., Lei, M., Zhang, M., Tang, K., Huang, H., Zhang, C., Miki, D., Liu, P., Yang, Y.,
616 Wang, X., *et al.* (2019). Histone acetylation recruits the SWR1 complex to regulate
617 active DNA demethylation in Arabidopsis. Proc Natl Acad Sci U S A *116*, 16641-16650.
618 Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L.
619 (2007). DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci U S A *104*,
620 6752-6757.
621 Qian, W., Miki, D., Lei, M., Zhu, X., Zhang, H., Liu, Y., Li, Y., Lang, Z., Wang, J., Tang, K., *et*
622 *al.* (2014). Regulation of active DNA demethylation by an alpha-crystallin domain
623 protein in Arabidopsis. Mol Cell *55*, 361-371.
624 Qian, W., Miki, D., Zhang, H., Liu, Y., Zhang, X., Tang, K., Kan, Y., La, H., Li, X., Li, S., *et al.*
625 (2012). A histone acetyltransferase regulates active DNA demethylation in
626 Arabidopsis. Science *336*, 1445-1448.
627 Qin, G., Ma, J., Chen, X., Chu, Z., and She, Y.M. (2017). Methylated-antibody affinity
628 purification to improve proteomic identification of plant RNA polymerase Pol V
629 complex and the interacting proteins. Sci Rep *7*, 42943.
630 Robertson, K.D. (2005). DNA methylation and human disease. Nat Rev Genet *6*,
631 597-610.
632 Sardina, J.L., Collombet, S., Tian, T.V., Gomez, A., Di Stefano, B., Berenguer, C.,
633 Brumbaugh, J., Stadhouders, R., Segura-Morales, C., Gut, M., *et al.* (2018).
634 Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram
635 Cell Fate. Cell Stem Cell *23*, 905-906.
636 Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic
637 regulation of the genome. Nat Rev Genet *8*, 272-285.

638 Wang, D., Xia, X., Weiss, R.E., Refetoff, S., and Yen, P.M. (2010). Distinct and
639 histone-specific modifications mediate positive versus negative transcriptional
640 regulation of TSHalpha promoter. PLoS One *5*, e9853.
641 Wang, Y., Xiao, M., Chen, X., Chen, L., Xu, Y., Lv, L., Wang, P., Yang, H., Ma, S., Lin, H., *et al.*
642 (2015). WT1 recruits TET2 to regulate its target gene expression and suppress
643 leukemia cell proliferation. Mol Cell *57*, 662-673.
644 Williams, B.P., Pignatta, D., Henikoff, S., and Gehring, M. (2015). Methylation-sensitive
645 expression of a DNA demethylase gene serves as an epigenetic rheostat. PLoS Genet
646 *11*, e1005142.
647 Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism,
648 function and beyond. Nat Rev Genet *18*, 517-534.
649 Xiong, J., Zhang, Z., Chen, J., Huang, H., Xu, Y., Ding, X., Zheng, Y., Nishinakamura, R., Xu,
650 G.L., Wang, H., *et al.* (2016). Cooperative Action between SALL4A and TET Proteins in
651 Stepwise Oxidation of 5-Methylcytosine. Mol Cell *64*, 913-925.
652 Yang, J., Liu, Y., Yan, H., Tian, T., You, Q., Zhang, L., Xu, W., and Su, Z. (2018). PlantEAR:
653 Functional Analysis Platform for Plant EAR Motif-Containing Proteins. Front Genet *9*,
654 590.
655 Zemach, A., and Grafi, G. (2007). Methyl-CpG-binding domain proteins in plants:
656 interpreters of DNA methylation. Trends Plant Sci *12*, 80-85.
657 Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K.,
658 Harmer, S.L., and Zilberman, D. (2013). The Arabidopsis nucleosome remodeler
659 DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell
660 *153*, 193-205.
661 Zhang, H., Lang, Z., and Zhu, J.K. (2018). Dynamics and function of DNA methylation
662 in plants. Nat Rev Mol Cell Biol *19*, 489-506.
663 Zhang, H., Zhang, K., and Zhu, J.K. (2019). A model for the aberrant DNA methylomes
664 in aging cells and cancer cells. Biochem Soc Trans *47*, 997-1003.
665 Zhu, J., Kapoor, A., Sridhar, V.V., Agius, F., and Zhu, J.K. (2007). The DNA
666 glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in
667 Arabidopsis. Curr Biol *17*, 54-59.
668 Zhu, J.K. (2009). Active DNA demethylation mediated by DNA glycosylases. Annu Rev
669 Genet *43*, 143-166.

670

671

672

673

674

FIGURE LEGENDS

Figure. 1 RWD40 interacts with ROS1.

(A) Schematic representation of the RWD40 protein (top panel) and of the truncated forms of RWD40 used in Y2H assays (lower panels). The two LISH-CTLH-containing regions of RWD40 encompass amino acid residues 7 to 95 and 191 to 279, respectively. The WD40 domain encompasses amino acid residues 388 to 741. LI-CT indicates the LISH-CTLH domain. aa, amino acid residue.

(B) RWD40 interacts with ROS1 in Y2H assays. The N-terminal region, including the two LISH-CTLH domains of RWD40, interacts with ROS1.

(C) RWD40 interacts with ROS1 in BiFC assays.

(D) Proteins detected by LC-MS/MS following immunoprecipitation (IP) using an anti-Flag antibody, which immunopurifies ROS1, RWD40, RMB1, and RHD1 from transgenic Arabidopsis lines. Results obtained in one to four independent biological replicates are shown (Rep1–Rep4).

Figure 2. RMB1 and RHD1 interact with RWD40 but not with ROS1.

(A) Interactions between RWD40, RMB1, RHD1, and ROS1 in Y2H assays. BD, GAL4-binding domain; AD, GAL4-activation domain.

(B) Interactions between RWD40, RMB1, RHD1, and ROS1 in split luciferase complementation assays. The indicated proteins were fused to the N-terminal or the C-terminal part of the luciferase protein (nLuc or cLuc, respectively) and

transiently expressed in *Nicotiana benthamiana* leaves. White circles indicate regions infiltrated with *Agrobacterium*. The luminescent signal indicates a protein–protein interaction.

(C) Schematic representation of the truncated forms of RWD40 used in the Y2H assays. WID, WD40-interaction domain. aa, amino acid residue.

(D) The WID domain of RMB1 interacts with the WD40 domain of RWD40 in Y2H assays.

Figure 3. ROS1, RWD40, RMB1, and RHD1 form a protein complex.

(A) Western blot analysis of gel filtration eluates. The indicated fractions eluted from the gel filtration column were probed with anti-MYC or anti-FLAG to detect the epitope-tagged RWD40, RMB1, RHD1, and ROS1. The arrow indicates the estimated molecular weight of the protein complex.

(B) Co-expression of full-length RWD40, RMB1, RHD, and a truncated ROS1 protein (amino acid residues 1-100) in an insect cell expression system. RWD40, RMB1, RHD, and the truncated ROS1 were obtained in a single peak fraction. Arrows and numbers indicate elution markers and their molecular masses (in kDa), respectively (a.u.: arbitrary unit).

Figure 4. RMB1 binds to methylated DNA.

(A,B) Electrophoretic mobility shift assay (EMSA) showing that the MBD domain of RMB1 (residues 1 to 81) binds to methylated oligonucleotides corresponding to the AT5G39160 locus (A) and the MEMS in the *ROS1* promoter (B). M, methylated; U, unmethylated. (C) RMB1 binding to methyl-DNA assessed by microscale thermophoresis (MST). MST binding assays were used to quantify the interaction of RMB1 with unmethylated (Um) probe or probes methylated (m) in

CG, CHG, or CHH context. The m1 (W22G), m2 (Y38F), m3 (T49A), or m4 (K50T) mutation in conserved residues in the MBD domain abolished or diminished the binding to methylated DNA. Values are means ± SD (n=3). Curves indicate calculated fits, and binding affinities are listed. NBD indicates 'no binding detected'.

Figure 5. RWD40, RMB1, and RHD1 function in active DNA methylation.

(A) Bisulfite sequencing data showing DNA methylation in different sequence contexts at the MEMS in the ROS1 promoter in Col-0 and indicated mutants.

(B) Relative expression of ROS1 in the indicated samples. Values are means ± SD of three biological replicates. *P < 0.05, **P < 0.01, compared with Col-0 plants; NS, not significantly different compared with Col-0 plants (2-tailed t test).

(C) Chop-PCR showing that *rwd40-1*, *rmb1-1*, *rhd1-1*, and *ros1-4* mutants displayed an increased methylation phenotype at the indicated loci. Amplification of non-digested DNA served as a control.

(D-G) Analysis of DNA methylation levels at several DNA demethylation target loci in Col-0 control plants and indicated mutants as indicated by individual locus bisulfite sequencing analysis.

(H) Chop-PCR showing that *rwd40-1*, *rmb1-1*, and *rhd1-1* mutants did not display an increased methylation phenotype in IDM1-dependent DNA demethylation target loci (Qian et al., 2012). Amplification of non-digested DNA served as a control.

Figure S1. DNA demethylation phenotypes of *rwd40-1* and *rwd40-2* mutants and mutant complementation.

(A) RWD40 localization in *Nicotiana benthamiana.*

(B) T-DNA insertion positions in *rwd40-1* and *rwd40-2* mutants. Boxes and lines denote exons and introns, respectively.

(C) Expression of *RWD40* in mutants relative to that in Col-0 control plants, as measured by qPCR. Values are means ± SD of three biological replicates.

(D) Western blot analysis of the RWD40 in the T2 transgenic lines.

(E) DNA methylation level in Col-0 and indicated mutants as determined by chop-PCR.

(F) DNA methylation levels at the AT5G39160 locus in *rwd40-1* and *rwd40-2* mutants as determined by individual bisulfite sequencing.

(G) DNA methylation levels in *rwd40* mutant alleles and complementation lines at the AT5G39160 locus.

Figure S2. Characterization of RMB1 and RHD1.

(A) RMB1 domains and mutants. Upper panel, schematic representation of the RMB1 protein. CBD, C(AT)-rich DNA-binding domain. Middle panel, T-DNA insertion positions in *rmb1-1* and *rmb1-2* mutants. Boxes and lines denote exons and introns, respectively. Lower panel, RMB1 transcript levels in the mutants.

(B) Schematic representation of the RMB1 protein (upper panel) and T-DNA insertion position in the *rhd1-1* mutant (lower panel).

(C) Interactions between RWD40, RMB1, and RHD1 in BiFC assays.

(D) DNA methylation levels in Col-0 and indicated mutants as determined by chop-PCR.

(E) DNA methylation levels at the AT5G39160 locus in *rmb1-1* and *rmb1-2* mutants as determined by individual locus bisulfite sequencing.

(F) DNA methylation levels in *rmb1* mutant alleles and in complementation lines at the AT5G39160 locus.

(G) DNA methylation levels in Col-0 and indicated mutants at the AT5G39160 locus (upper panel), and in *rhd1* mutant alleles and complementation line at the AT4G18380 locus as determined by chop-PCR (lower panel).

Figure S3. Interactions between RWD40, RMB1, RHD1, and ROS1 in Y2H assays.

(A) Schematic representation of the truncated forms of ROS1 used in the Y2H assays and of the ELM domain identified in N-terminal regions of ROS1. HhH-GPD, hallmark Helix-hairpin-helix and Gly/Pro-rich loop domain; EndIIII, endonuclease III; Perm-CXXC, permuted version of a single unit of the zinc finger-CXXC; RRM_DME, RRM-fold domain present at the C-terminus of Demeter-like glycoslyases; EAR, ethylene response factor-associated amphiphilic repression domain; ELD, EAR-like motif.

(B) The ELD domain of ROS1 interacts with RWD40 in Y2H assays.

(C) Schematic representation of the truncated forms of RHD1 used in Y2H assays.

(D) RHD1 interacts with RHD1 through the zinc-finger domain, and the WD40 domain of RWD40 interacts with the zinc-finger domain of RHD1 in Y2H assays.

(E) RMB1 interacts with full-length RHD1 but does not interact with its zinc-finger domain or homeodomain.

Figure S4. Analysis of MBD domain-containing proteins in *Arabidopsis.*

(A) Phylogenetic tree of MBD domain-containing proteins in *Arabidopsis thaliana.*

(B) Alignment of MBD domain-containing proteins in *Arabidopsis thaliana*. The amino acid residues targeted for mutations in the MBD domain are indicated by triangles.

Figure S5. Mutation of the MBD domain reduces the binding of RMB1 to methylated DNA and abolishes the demethylation function of RMB1.

(A) K50T mutation in the MBD domain abolishes binding to methylated oligonucleotides corresponding to the AT5G39160 locus. Um, unmethylated.

(B) DNA methylation level at AT5G39160 in the *rmb1-1* mutant and in the *rmbd1-1* mutant expressing WT (wild type) or K50T-mutated RMB1 as determined by chop PCR.

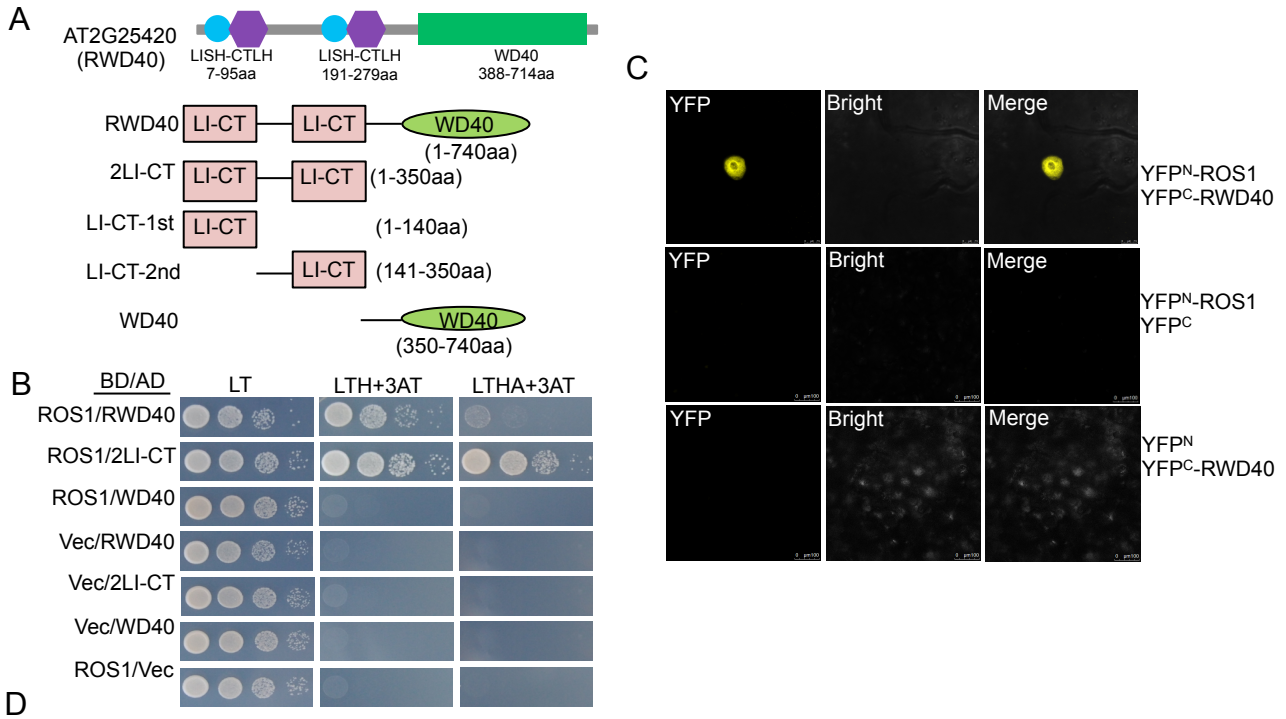
(C,D) EMSA of the MBD domain of MBD7 (C) and IDM1 (D) binding to methylated oligonucleotides corresponding to the AT5G39160 locus. M, methylated. Um, unmethylated.

Figure S6. Dysfunction of RWD40, RMB1, or RHD1 does not cause silencing of the *35S:SUC2* transgene.

(A) Root phenotype in *rwd40-1*, *rmb1-1*, and *rhd1-1* in the *35S:SUC2* transgene background.

(B) Relative expression of the *35S:SUC2* transgene in the indicated mutants. Values are relative to transcript levels in *35S:SUC2* control plants and are means ± SD of three biological replicates. *P < 0.01, compared with *35S:SUC2* plants; NS, not significantly different compared with *35S:SUC2* plants (2-tailed t test).

Figure 1



D

		ROS1 (AT2G36490)		RWD40(AT2G25420)		RMB1 (AT1G63240)		RHD1 (AT5G42780)	
		Score	Unique peptides	Score	Unique peptides	Score	Unique peptides	Score	Unique peptides
IP ROS1	Rep.1	5574.0	83	152.0	10	81.0	8	55.0	4
	Rep.2	3643.0	86	139.0	11	148.0	4	78.0	4
IP RWD40	Rep.1	816.0	22	605.0	16	202.0	7	205.0	9
IP RMB1	Rep.1	76.0	4	71.0	3	902.0	20	69.0	3
	Rep.2	1371.0	48	176.0	8	591.0	13	147.0	7
IP RHD1	Rep.1	94.0	7	242.0	9	106.0	4	1085.0	14
IP Col-0 control	Rep.1	0	0	0	0	0	0	0	0
	Rep.2	0	0	0	0	0	0	0	0
	Rep.3	0	0	0	0	0	0	0	0
	Rep.4	0	0	0	0	0	0	0	0

Figure 2

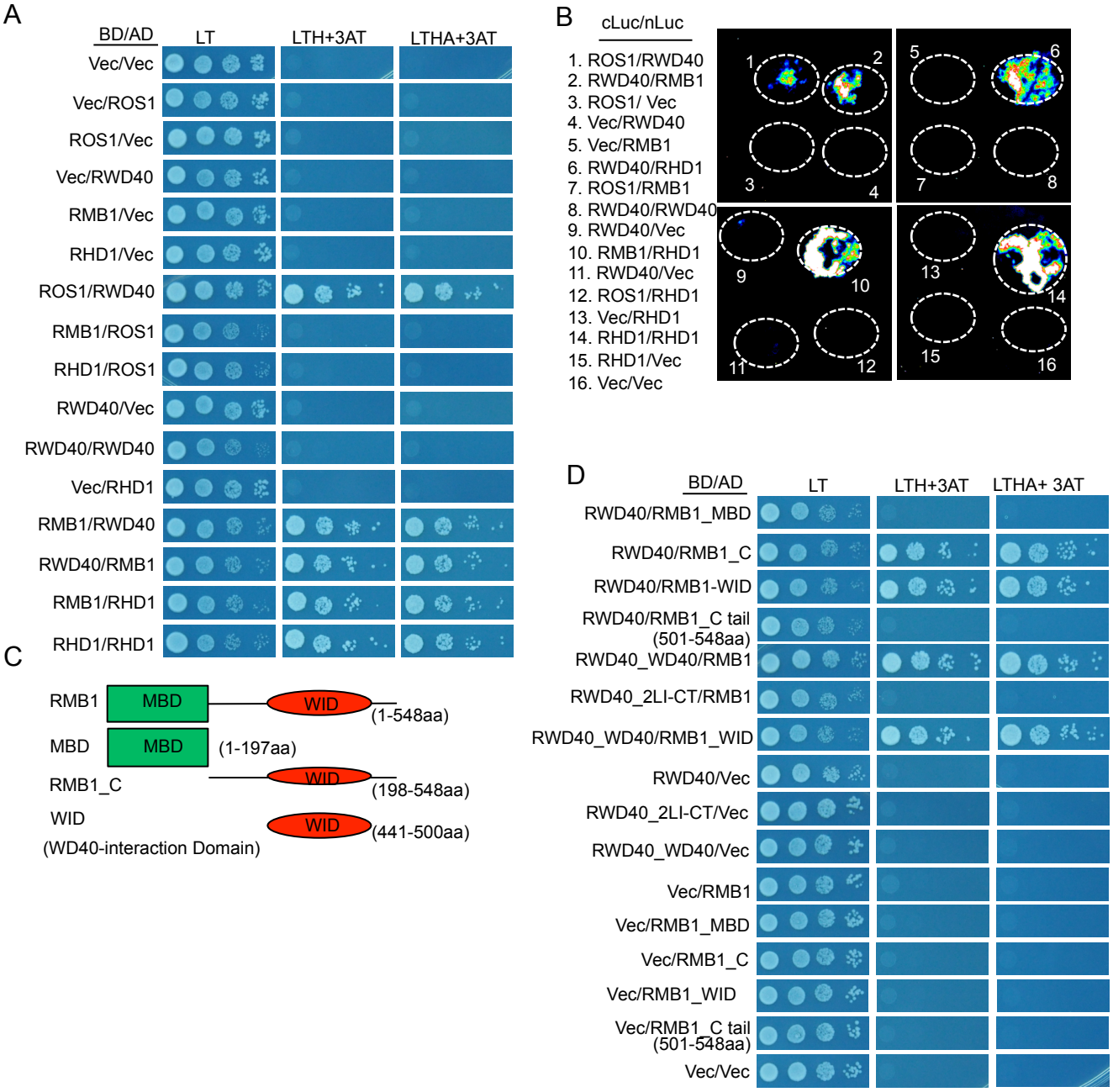

A
BD/AD
LT
LTH+3AT
LTHA+3AT
Vec/Vec
Vec/ROS1
ROS1/Vec
Vec/RWD40
RMB1/Vec
RHD1/Vec
ROS1/RWD40
RMB1/ROS1
RHD1/ROS1
RWD40/Vec
RWD40/RWD40
Vec/RHD1
RMB1/RWD40
RWD40/RMB1
RMB1/RHD1
RHD1/RHD1
B
cLuc/nLuc
1. ROS1/RWD40
2. RWD40/RMB1
3. ROS1/ Vec
4. Vec/RWD40
5. Vec/RMB1
6. RWD40/RHD1
7. ROS1/RMB1
8. RWD40/RWD40
9. RWD40/Vec
10. RMB1/RHD1
11. RWD40/Vec
12. ROS1/RHD1
13. Vec/RHD1
14. RHD1/RHD1
15. RHD1/Vec
16. Vec/Vec
C
RMB1
MBD
WID
(1-548aa)
MBD
MBD
(1-197aa)
RMB1_C
WID
(198-548aa)
WID
WID
(441-500aa)
(WD40-interaction Domain)
D
BD/AD
LT
LTH+3AT
LTHA+ 3AT
RWD40/RMB1_MBD
RWD40/RMB1_C
RWD40/RMB1-WID
RWD40/RMB1_C tail
(501-548aa)
RWD40_WD40/RMB1
RWD40_2LI-CT/RMB1
RWD40_WD40/RMB1_WID
RWD40/Vec
RWD40_2LI-CT/Vec
RWD40_WD40/Vec
Vec/RMB1
Vec/RMB1_MBD
Vec/RMB1_C
Vec/RMB1_WID
Vec/RMB1_C tail
(501-548aa)
Vec/Vec

Figure 3

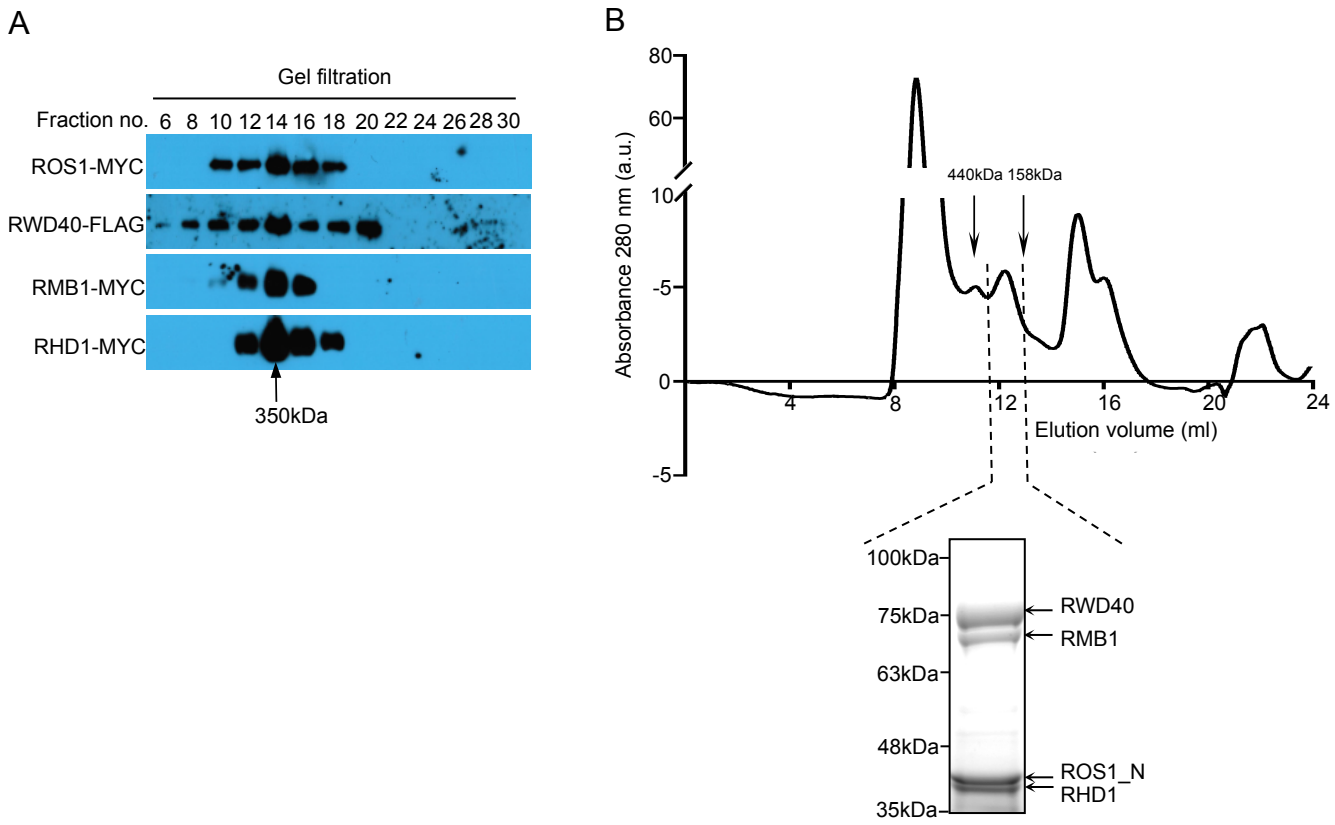
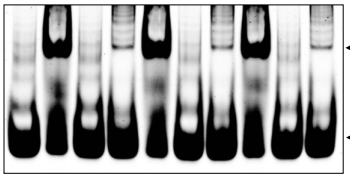

A
Gel filtration
Fraction no. 6 8 10 12 14 16 18 20 22 24 26 28 30
ROS1-MYC
RWD40-FLAG
RMB1-MYC
RHD1-MYC
350kDa
B
Absorbance 280 nm (a.u.)
80
60
10
-5
0
-5
440kDa 158kDa
4
8
12
16
20
24
Elution volume (ml)
100kDa
75kDa
63kDa
48kDa
35kDa
RWD40
RMB1
ROS1_N
RHD1

Figure 4

A

AT5G39160 locus

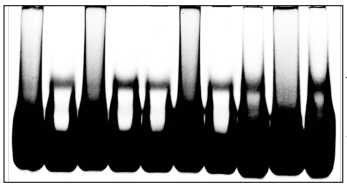
-	+	+	+	+	+	+	+	+	+	RMB1-MBD
+	-	-	-	-	-	-	-	-	-	Um probe
-	+	+	+	-	-	-	-	-	-	mCG probe
-	-	-	-	+	+	+	-	-	-	mCHG probe
-	-	-	-	-	-	-	+	+	+	mCHH probe
-	-	M	U	-	M	U	-	M	U	Competitor



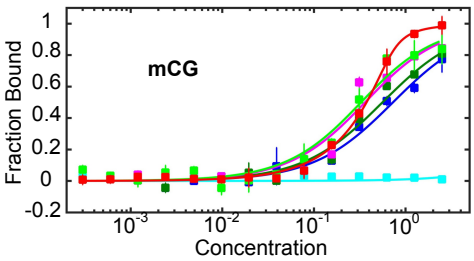
B

MEMS in the *ROS1* promoter

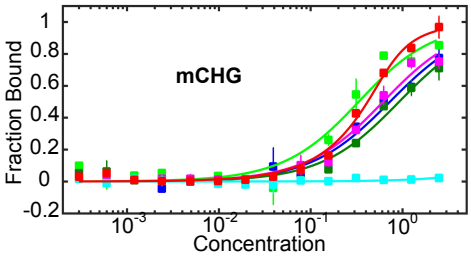
-	+	+	+	+	+	+	+	+	+	RMB1-MBD
+	-	-	-	-	-	-	-	-	-	Um probe
-	+	+	+	-	-	-	-	-	-	mCG probe
-	-	-	-	+	+	+	-	-	-	mCHG probe
-	-	-	-	-	-	-	+	+	+	mCHH probe
-	-	M	U	-	M	U	-	M	U	Competitor



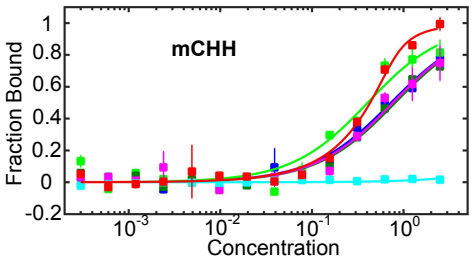
C



MBD	dsDNA	Kd(μM)
WT	mCG	0.04±0.02
WT	Um	>5.27
W22G	mCG	0.29±0.09
Y38F	mCG	0.56±0.17
T49A	mCG	0.33±0.12
K50T	mCG	No binding



MBD	dsDNA	Kd(μM)
WT	mCHG	0.11±0.07
WT	Um	>5.27
W22G	mCHG	0.30±0.11
Y38F	mCHG	0.89±0.27
T49A	mCHG	0.58±0.14
K50T	mCHG	No binding



MBD	dsDNA	Kd(μM)
WT	mCHH	0.06±0.07
WT	Um	>5.27
W22G	mCHH	0.37±0.15
Y38F	mCHH	0.78±0.20
T49A	mCHH	0.74±0.25
K50T	mCHH	No binding

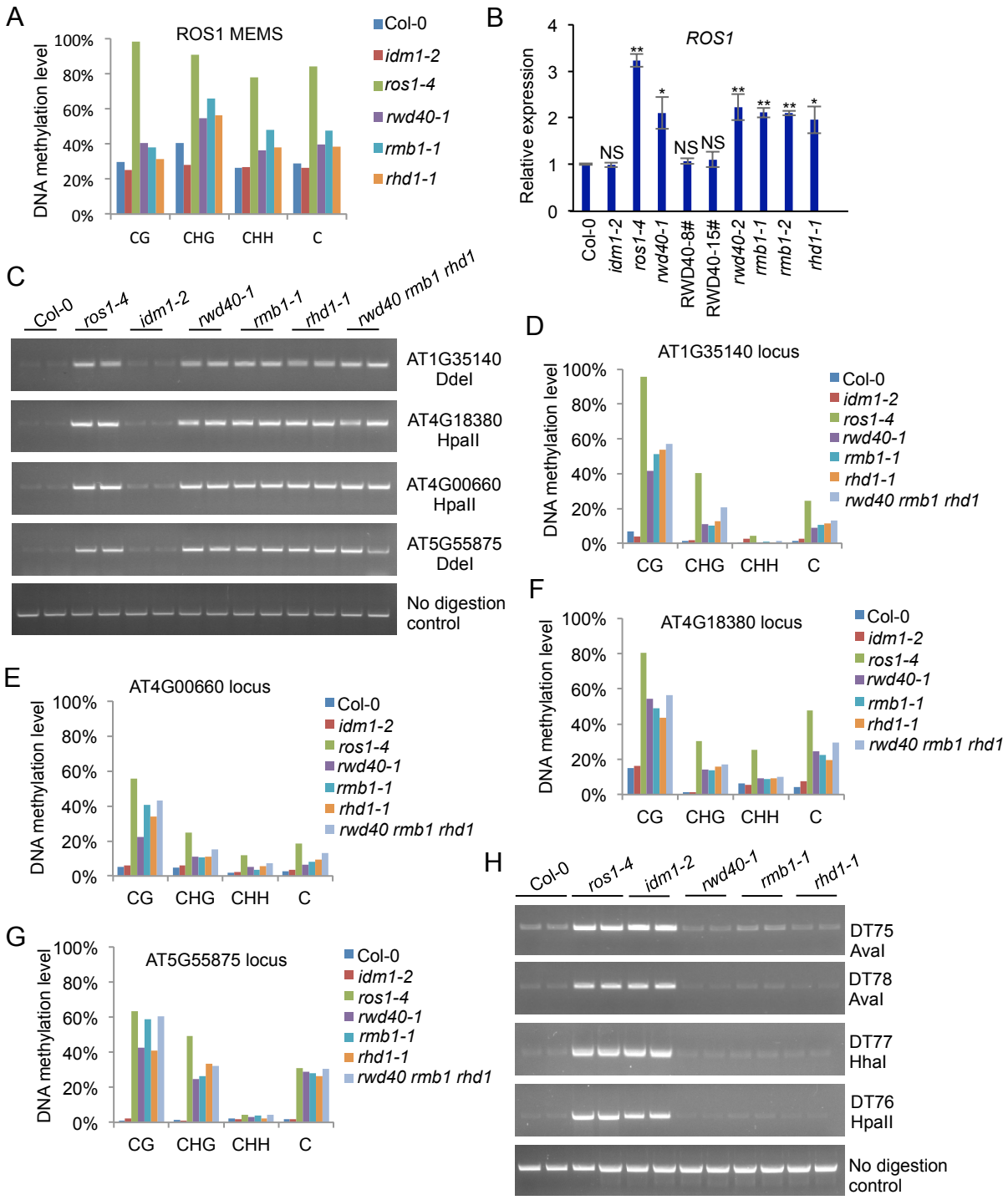
mCG probe F: GGTACT^{m}CGACAGTAT
mCG probe R:ATACTGT^{m}CGAGTACC

mCHG probe F: GGTACT^{m}CAGCAGTAT
mCHG probe R: ATACTG^{m}CTGAGTACC

mCHH probe F: GGTACT^{m}CATCAGTAT
mCHH probe R: ATACTGATGAGTACC

Um probe F: GGTACTCGACAGTAT
Um probe R: ATACTGATGAGTACC

Figure 5


A
ROS1 MEMS
DNA methylation level
100%
80%
60%
40%
20%
0%
CG
CHG
CHH
C
Col-0
idm1-2
ros1-4
rwd40-1
rmb1-1
rhd1-1
B
ROS1
Relative expression
4
3
2
1
0
NS
**
*
NS
NS
**
**
**
*
Col-0
idm1-2
ros1-4
rwd40-1
RWD40-8#
RWD40-15#
rwd40-2
rmb1-1
rmb1-2
rhd1-1
C
Col-0
ros1-4
idm1-2
rwd40-1
rmb1-1
rhd1-1
rwd40 rmb1 rhd1
AT1G35140
DdeI
AT4G18380
HpaII
AT4G00660
HpaII
AT5G55875
DdeI
No digestion
control
D
AT1G35140 locus
DNA methylation level
100%
80%
60%
40%
20%
0%
CG
CHG
CHH
C
Col-0
idm1-2
ros1-4
rwd40-1
rmb1-1
rhd1-1
rwd40 rmb1 rhd1
E
AT4G00660 locus
DNA methylation level
100%
80%
60%
40%
20%
0%
CG
CHG
CHH
C
Col-0
idm1-2
ros1-4
rwd40-1
rmb1-1
rhd1-1
rwd40 rmb1 rhd1
F
AT4G18380 locus
DNA methylation level
100%
80%
60%
40%
20%
0%
CG
CHG
CHH
C
Col-0
idm1-2
ros1-4
rwd40-1
rmb1-1
rhd1-1
rwd40 rmb1 rhd1
G
AT5G55875 locus
DNA methylation level
100%
80%
60%
40%
20%
0%
CG
CHG
CHH
C
Col-0
idm1-2
ros1-4
rwd40-1
rmb1-1
rhd1-1
rwd40 rmb1 rhd1
H
Col-0
ros1-4
idm1-2
rwd40-1
rmb1-1
rhd1-1
DT75
AvaI
DT78
AvaI
DT77
HhaI
DT76
HpaII
No digestion
control