Retraction

THE EMBO JOURNAL

Retraction: 'A pair of transposon-derived proteins regulate active DNA demethylation in *Arabidopsis*'

Qi Li, Wei Yuan, Xiaokang Wang, Yannan Liu, Xinyi Ma, Han Sun, Jinjin Tong, Xueyuan Tian, Yan Li & Weiqiang Qian

Retraction of: *The EMBO Journal* DOI 10.15252/embj.201694284 | Published online 28.07.2016

The authors have requested retraction of the paper noting: 'We found a mistake for the materials used in this paper. In Fig 5A–C, Fig 6D–E and Fig 7D, WT-LUC and idm1-14 were in an rdr6-11 background, while idap1-1 and idap2-1 were in wild type background. RDR6 has a demonstrated role in DNA methylation and transcriptional gene silencing, and the rdr6 mutation could therefore have a profound impact on gene silencing in these experiments. Thus, the observed reduced LUC luminescence in idap1-1 and idap2-1 mutants could have resulted from the combined effects of rdr6 and idap1 or idap2 mutations, which might change the main conclusion of this paper. Further experiments will be performed to confirm the phenotypes. All authors concur with this statement and wish to apologize for the inconvenience caused.'

Article

TRANSPARENT PROCESS

THE EMBO JOURNAL

A pair of transposon-derived proteins regulate active DNA demethylation in *Arabidopsis*

Qi Li[†], Wei Yuan[†], Xiaokang Wang[†], Yannan Liu, Xinyi Ma, Han Sun, Jinjin Tong, Xueyuan Tian, Yan Li & Weiqiang Qian[*]

Abstract

DNA methylation and histone modifications are important epigenetic markers involved in transcriptional gene regulation in animals and plants. In *Arabidopsis*, IDM1 is a histone acetyltransferase that plays critical roles in preventing hypermethylation of DNA and transcriptional gene silencing, but the mechanism(s) by which IDM1 is targeted to specific genomic loci is still unclear. Here, we find two new components of the IDM1 complex, IDAP1 and IDAP2, which bear high similarities to the transposase and myb-like DNA-binding protein of *PIF/Harbinger* transposons, respectively. Dysfunction of *IDAP1* or *IDAP2* causes DNA hypermethylation and repression of reporter genes as well as endogenous genes. IDAP1 interacts with IDAP2, and they form a protein complex with IDM1. Furthermore, IDAP2 has DNA-binding activity, which facilitates the recruitment of IDM1 to its target loci. We propose that IDAP1 and IDAP2, co-opted from ancient *PIF/Harbinger*-like transposon proteins, act together with IDM1 to regulate active DNA demethylation in higher plants.

Keywords active DNA demethylation; anti-silencing; domestication; epigenetics; *PIF/Harbinger* transposon

Subject Categories Chromatin, Epigenetics, Genomics & Functional Genomics; Plant Biology

DOI 10.15252/embj.201694284 | Received 8 March 2016 | Revised 21 June 2016 | Accepted 5 July 2016

Introduction

Cytosine DNA methylation is an important epigenetic mechanism that regulates gene expression, silences transposable elements (TEs) and safeguards genome stability (Vaillant & Paszkowski, 2007; Law & Jacobsen, 2010; Matzke & Mosher, 2014). In plants, cytosine can be methylated in three sequence contexts: CG, CHG, and CHH (H represents A, T, or C) (Lister *et al*, 2008). The *de novo* establishment of cytosine methylation in all sequence contexts is catalyzed by domains rearranged methyltransferase 2 (DRM2, homologous to mammalian DNMT3) through an RNA-directed DNA methylation (RdDM) pathway (Cao & Jacobsen, 2002; Zhong *et al*, 2014). Once established, CG and CHG methylations are maintained by methyltransferase 1 (MET1, homologous to mammalian DNMT1) and a plant-specific DNA methyltransferase chromomethylase 3 (CMT3), respectively, during DNA replication while CHH methylation can be established *de novo* in each new cell cycle by the RdDM pathway (Law & Jacobsen, 2010; Matzke & Mosher, 2014) or maintained by chromomethylase 2 (CMT2) through a separate pathway dependent on the chromatin remodeling protein DDM1 (Zemach *et al*, 2013; Stroud *et al*, 2014). DNA methylome analyses have revealed that heavy cytosine methylation in all sequence contexts occurs in TEs and other repetitive sequences and plays an important role in transcriptional silencing of these regions (Zhang *et al*, 2006; Zhang & Zhu, 2011). Methylation at CG residues only is usually observed in the bodies of actively transcribed genes, but the role of these methylation events remains obscure (Zhang *et al*, 2006).

DNA demethylation, which can be passive or active, counteracts DNA methylation, and their balance determines the final DNA methylation pattern (Zhu, 2009; Wu & Zhang, 2010). In *Arabidopsis*, active DNA demethylation is initiated by the ROS1 family of bifunctional DNA glycosylases/lyases, which remove 5-methylcytosine and cleave the DNA backbone at the abasic site (Gong *et al*, 2002; Agius *et al*, 2006; Gehring *et al*, 2006; Morales-Ruiz *et al*, 2006; Penterman *et al*, 2007; Zhu, 2009). Then, either the DNA phosphatase ZDP or the AP endonuclease APE1L processes the 3′ end of the abasic site to generate a hydroxyl group (Martinez-Macias *et al*, 2012; Lee *et al*, 2014; Li *et al*, 2015b). Immediately following that, a yet-to-be identified DNA polymerase and DNA ligase I act sequentially to promote the insertion of an unmethylated cytosine (Li *et al*, 2015c). One of the important questions in the study of active DNA demethylation is how locus specificity is achieved. The locus-specific targeting of ROS1 has been shown to be mediated by the RNA-binding protein ROS3, which co-localizes with ROS1 (Zheng *et al*, 2008). Recently, increased DNA methylation 1 (IDM1), a histone acetyltransferase, was reported to recognize loci with low histone 3 lysine 4 (H3K4) and histone 3 arginine 2 (H3R2) methylation levels, acetylate histone 3 lysine 18 and 23 (H3K18 and H3K23) and recruit ROS1 to these loci (Li *et al*, 2012; Qian *et al*, 2012). MBD7, IDM2, and IDM3/IDL1 form a complex with IDM1 (Qian *et al*, 2014; Lang *et al*, 2015; Li *et al*, 2015a; Wang *et al*, 2015).

State Key Laboratory of Protein and Plant Gene Research, The Peking-Tsinghua Center for Life Sciences, School of Life Sciences, Peking University, Beijing, China
*Corresponding author. Tel: +86 10 62768230; E-mail: wqqian@pku.edu.cn
[†]These authors contributed equally to this work

MBD7 is required for binding of IDM1 to methylated DNA, while IDM2 and IDM3/IDL1 regulate the function of IDM1 via unknown mechanisms (Qian *et al*, 2014; Zhao *et al*, 2014; Lang *et al*, 2015).

Transposable elements are an important source of DNA sequences that can be co-opted by host genomes via an evolutionary process known as molecular domestication, which drives the emergence of new genes (Sinzelle *et al*, 2009). *PIF/Harbinger* is a superfamily of DNA transposons identified in both the plant and animal kingdoms (Kapitonov & Jurka, 1999; Zhang *et al*, 2001; Jiang *et al*, 2003). Members of the *PIF/Harbinger* superfamily carry a canonical transposase gene as well as a myb-like gene (Jiang *et al*, 2003; Zhang *et al*, 2004). The transposase harbors DNA endonuclease activity, while the myb-like protein possesses DNA-binding activity. The interaction between the transposase and the myb-like protein and their joint action are required for transposition of the *PIF/Harbinger* elements (Kapitonov & Jurka, 2004; Sinzelle *et al*, 2008). The domestication of the *PIF/Harbinger* elements correspondingly gives rise to two genes. *HARBI1* and *NAIF1* represent two known domesticated genes originated from a *PIF/Harbinger* transposon in human cells (Kapitonov & Jurka, 2004; Sinzelle *et al*, 2008). Recently, ALP1 was identified as a domesticated *PIF/Harbinger* transposase in *Arabidopsis* and was shown to play critical roles in inhibiting transcriptional silencing mediated by the polycomb group (PcG) proteins (Hartwig *et al*, 2012; Liang *et al*, 2015). However, the identity of its myb-like partner is unknown and it is unclear whether there exist other domesticated genes derived from the *PIF/Harbinger* elements in plants.

In this study, we identified IDAP1 and IDAP2 using affinity purification of the IDM1 complex. *IDAP1* and *IDAP2* encode two genes related to *PIF/Harbinger* transposases. Specifically, *IDAP1* encodes a transposase-like protein and *IDAP2* encodes a myb-like DNA-binding protein. IDAP1 and IDAP2 interact with each other and they form a complex with IDM1, IDM2, MBD7, and IDL1. IDAP1 and IDAP2 act together to mediate the recruitment of IDM1, which promotes active DNA demethylation and repression of transcriptional silencing at specific loci. Our study demonstrated for the first time that a pair of domesticated genes derived from the *PIF/Harbinger* elements in *Arabidopsis* executes important epigenetic functions.

Results

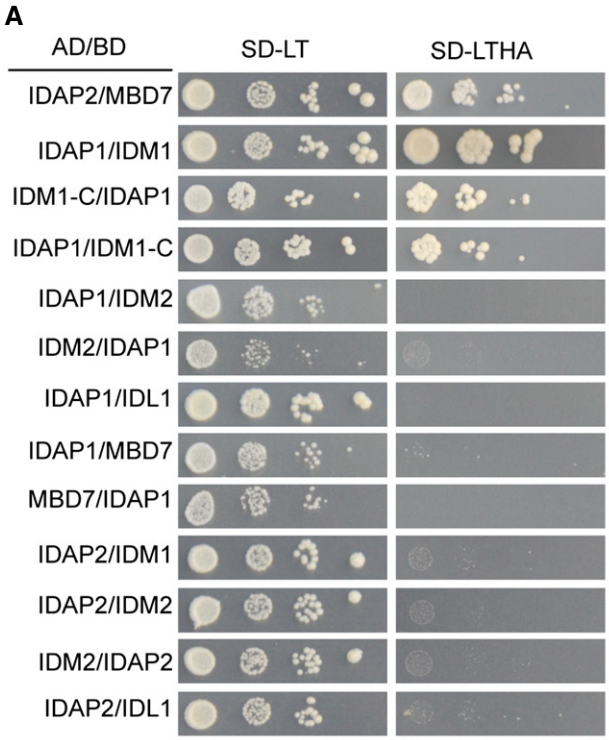
Identification of IDAP1 and IDAP2 by affinity purification of the IDM1 complex

To elucidate the mechanism(s) underlying IDM1 targeting in active DNA demethylation, we purified the IDM1 complex by affinity purification in a previous study and identified two proteins, MBD7 and IDL1, in this complex (Li *et al*, 2015a). In addition to MBD7 and IDL1, we also identified IDM1-associated protein 1 (IDAP1) and IDM1-associated protein 2 (IDAP2) from the mass spectrometric analysis of IDM1 co-purified proteins (Appendix Table S1). IDAP1 is encoded by *At1g72270* and IDAP2 is encoded by *At4g31270*. Interestingly, *At1g72270* encodes two splice variants *At1g72270.1* and *At1g72270.2* (Appendix Fig S1A). Detailed mass spectrometric analyses revealed that the peptides co-purified with IDM1 all corresponded to *At1g72270.2*, a shorter splice variant (Appendix Fig S1B and Table EV1). This indicates that only the shorter splice variant of *At1g72270* was incorporated into the IDM1 complex.

In addition to being identified from IDM1 co-purified proteins, IDAP1 and IDAP2 were also detected from IDM2, IDL1, and MBD7 co-purified proteins (Appendix Table S1). Our results suggest that IDAP1 and IDAP2 may form a complex with IDM1 and other related proteins. To confirm the protein–protein interactions identified by mass spectrometry, we performed yeast two-hybrid (Y-2-H) assays and found that IDAP2 directly interacted with MBD7, while IDAP1 directly interacted with IDM1 (Fig 1A). Interestingly, it was the C-terminus of IDM1 that mediates the interaction between IDAP1 and IDM1 (Fig 1A). To further confirm the protein–protein interactions, we performed co-immunoprecipitation (co-IP) experiments. The interaction between IDM1 and IDAP1 or IDAP2 was readily detected by co-IP in plants harboring both the *IDM1-HA-YFP* and *IDAP1-3Myc* or *IDAP2-3Flag* transgenes under the control of their native promoters (Fig 1B and Appendix Fig S1C). Additionally, we generated *IDAP1-3Myc* and *IDAP2-3Flag* transgenic plants and purified proteins associated with IDAP1-3Myc or IDAP2-3Flag by affinity purification. Mass spectrometric analyses revealed that IDAP1 and IDAP2 co-purified with each other. We also identified a large number of peptides corresponding to IDM1, IDM2, IDL1, and MBD7 (Fig 1C and Table EV2). Our data suggest that IDM1, IDM2, IDL1, MBD7, IDAP1, and IDAP2 form a complex. To further characterize the associations between IDAP1, IDAP2, and IDM1, we generated protein extracts from flower tissues of *IDAP1-3Myc*, *IDAP2-3Flag,* and *IDM1-HA-YFP* transgenic plants and analyzed these extracts by gel filtration followed by Western blotting. In gel filtration, IDAP1, IDAP2, and IDM1 were predominantly co-eluted in the fractions > 440 kDa (Fig 1D), indicating that these three proteins are present in a complex *in vivo*.

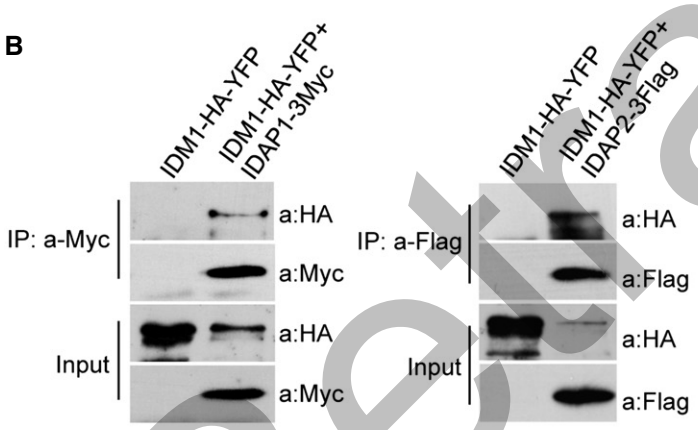
IDAP1 and IDAP2 bear high similarities to the transposase and myb-like DNA-binding protein of a *PIF/Harbinger* transposon

PIF/Harbinger transposases belong to the P instability factor superfamily of transposases and consist of two proteins. One is a transposase with DNA endonuclease activity and the other is a myb-like protein with DNA-binding activity (Walker *et al*, 1997; Jiang *et al*, 2003). *IDAP1* encodes a gene related to the *PIF/Harbinger* transposases (Hartwig *et al*, 2012; Liang *et al*, 2015). Because IDAP1 has an endonuclease domain of the DDE-Tnp superfamily in its C-terminal region (Fig 2A), *IDAP1* is believed to encode the transposase component. Alignments between plant IDAP1 proteins revealed highly conserved blocks within the regions previously identified to be conserved between the *PIF/Harbinger* transposases (Fig 2A and Appendix Fig S2A). Interestingly, IDAP2 was predicted to have a myb-like_DNA-binding domain in the N-terminal region, and presumably, *IDAP2* encodes the DNA-binding component (Fig 2B and Appendix Fig S2B). Alignments between plant IDAP2 proteins revealed that the three W residues in the myb-like DNA-binding domain (-W-(X_{28})-W-(X_{19})-W-) are conserved among plant IDAP2 proteins (Fig 2B). Taken together, our results suggest that *IDAP1* and *IDAP2* encode two proteins related to the *PIF/Harbinger* DNA transposases in plants. However, in contrast to active *PIF/Harbinger* transposases, IDAP1 and IDAP2 are most likely inactive transposases. First, *Arabidopsis* IDAP1 is unlikely to retain transposase activity as its conserved catalytic triad DDE is mutated to



C

Gene Accession	Protein	Score	Spectra	Unique Peptides
IDAP1 co-purified proteins				
AT3G14980	IDM1	804	29	15
AT1G72270	IDAP1	866	28	7
AT4G31270	IDAP2	111	7	4
AT1G20870	IDL1	70	6	2
AT1G54840	IDM2	62	4	2
IDAP2 co-purified proteins				
AT3G14980	IDM1	4594	203	42
AT4G31270	IDAP2	4347	157	19
AT1G72270	IDAP1	2127	71	19
AT1G54840	IDM2	620	38	14
AT5G59800	MBD7	308	13	9
AT1G20870	IDL1	243	15	9



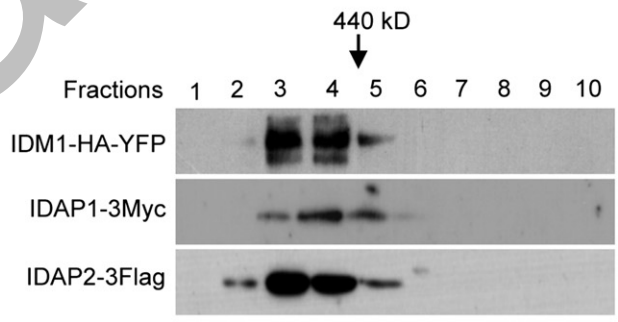


Figure 1. IDAP1 and IDAP2 interact with IDM1, IDM2, IDL1, and MBD7.

A Yeast two-hybrid assays. Yeast cells carrying different fusion protein combinations are listed on the left. Yeast cells expressing the indicated proteins from the pGAD-T7 (AD) and pCBK-T7 (BD) vectors were plated onto medium lacking Leu and Trp (SD-LT) (left) or medium lacking Leu, Trp, Ade, and His (SD-LTHA) (right).

B Co-immunoprecipitation (co-IP) of IDM1 with IDAP1 or IDAP2 in transgenic plants. Transgenic plants expressing IDM1-HA-YFP and IDAP1-3Myc or IDAP2-3Flag under their native promoters and their F1 offspring were used for co-IP analysis. Input, total protein before IP.

C Detection of proteins that associate with IDAP1 or IDAP2. Proteins were detected using LC-MS/MS following IP of Myc-tagged IDAP1 or Flag-tagged IDAP2.

D The elution profiles of IDM1-HA-YFP, IDAP1-3Myc, and IDAP2-3Flag in gel filtration assays. Anti-HA, anti-Myc, and anti-Flag antibodies were used in Western blots.

GDE (Fig 2A). Second, when we carried out phylogenetic analyses, we found that both IDAP1 and IDAP2 are encoded by single-copy genes, whereas active *PIF/Harbinger* transposases are typically present at much high copy numbers in plants (Walker *et al*, 1997; Jiang *et al*, 2003). Specifically, we identified proteins similar to IDAP1 in all major land plant groups. The IDAP1 clade includes 17 additional plant proteins of unknown function from different species. The *Arabidopsis* homolog closest in sequence to IDAP1, encoded by *At3g19120*, shows very low identity (18.5%) to IDAP1, and it is present in a sister clade of the tree (Fig 2C). In addition, IDAP1 belongs to an IDAP1 clade distinct from the clade formed by previously identified ALP1 and other proteins (Fig 2C). IDAP2 can be found in monocots as well as in dicots (Appendix Fig S2C). The plant IDAP2 proteins form a strongly supported group distinct from that of other myb-like DNA-binding proteins of the *PIF/Harbinger* transposases (Appendix Fig S2C). Third, there are no typical terminal inverted repeats identified in the 2-kb region upstream or downstream of the *IDAP1* and *IDAP2* gene loci. This is consistent with the

notion that the *IDAP1* and *IDAP2* genes are located on different chromosomes, whereas the active *PIF/Harbinger* transposases are encoded by two genes on one transposable element. Thus, we hypothesize that IDAP1 and IDAP2 are domesticated from an ancient *PIF/Harbinger* transposon but appear to have acquired new functions over time.

As the phylogenetic relationship suggested in Fig 2C, IDAP1 is highly conserved in land plants, indicating a very ancient domestication event. In non-vascular plants, the oldest homolog of IDAP1 we could find is WGTU_2005970 in *Leucostegia immerse*, which bears a high degree of similarity to IDAP1 and contains the same functional domain as IDAP1. It appears that the domestication event occurred during the speciation of *Leucostegia immerse*, about 400 million years ago (Gerrienne *et al*, 2016) or much earlier. Interestingly, in animals, the protein HARBI1 evolved from a *Harbinger* DNA transposase over 450 million years ago (Kapitonov & Jurka, 2004). So far, we did not find a species that only has IDM1 but does not have IDAP1, suggesting co-evolution of proteins in the IDM1 complex.

IDAP1 interacts with IDAP2 *in vitro* and *in vivo*

The transposase and the myb-like protein in the *Harbinger3-DR* system physically interact and cooperatively function to regulate transposition of the *PIF/Harbinger* transposon (Kapitonov & Jurka, 2004). Human HARBI1 and NAIF1, which are derived from a *PIF/Harbinger* transposon, also physically interact with each other (Sinzelle *et al*, 2008). To examine the possible physical interaction between IDAP1 and IDAP2, we first performed Y-2-H assays, and our results showed that yeast cells co-expressing either activation domain (AD)-IDAP1 and binding domain (BD)-IDAP2 or AD-IDAP2 and BD-IDAP1 could grow on minus-four media (media without leucine, tryptophan, histidine, and adenine) (Fig 3A), suggesting a strong interaction between IDAP1 and IDAP2. We further mapped the regions that mediate their interaction and found that the N-terminus of IDAP1 (amino acids 1–196) interacted with the C-terminus of IDAP2 (amino acids 125–294; Fig 3A). We also performed GST pull-down assays using co-expression systems in *E. coli*, and our results showed that glutathione S-transferase (GST)-IDAP2, but not the GST control, was able to pull down approximately equal amounts of maltose-binding protein (MBP)-IDAP1, suggesting that IDAP1 can directly interact with IDAP2 *in vitro* (Fig 3B). Finally, the interaction between IDAP1 and IDAP2 was confirmed using a split luciferase (LUC) complementation assay in *N. benthamiana* leaves (Fig 3C).

To determine the subcellular localization of the interaction between IDAP1 and IDAP2, we performed bimolecular fluorescence complementation (BiFC) assays in *N. benthamiana* leaves and found that IDAP1 and IDAP2 interacted in the nuclei (Fig 3D). To further visualize the subnuclear pattern of their localization, we immunostained nuclei isolated from the wild-type plants or the F1 heterozygous plants expressing Myc-tagged IDAP1 and Flag-tagged IDAP2 under the control of their native promoters. IDAP1 and IDAP2 co-localized in approximately 89% of the transgenic nuclei, as shown by the yellow signals resulting from an overlap of the green and red signals (Fig 3E). No other signals, except the DAPI signals, were detected in all wild-type nuclei (Fig 3E), suggesting the specificity of our staining. Taken together, our results suggest that IDAP1 and IDAP2 interact *in vitro* and *in vivo* and may function cooperatively.

Dysfunction of *IDAP1* or *IDAP2* causes DNA hypermethylation at hundreds of genomic loci

Because IDAP1 and IDAP2 were identified in the IDM1 complex that is required for ROS1 targeting in active DNA demethylation (Qian *et al*, 2012), we sought to investigate their function in the DNA demethylation pathway. To this end, we obtained two T-DNA insertion alleles *idap1-1* (GT_5_84536 in the Ler background) and *idap2-1* (Sail_1301_B12 in the Col background), in which *IDAP1* and *IDAP2* gene expression was, respectively, eliminated (Appendix Fig S3A–C). In addition, we generated the *idap1-2*, *idap1-3*, and *idap2-2* mutants (all in the Col background) using the CRISPR/Cas9 system (Feng *et al*, 2013) (Appendix Fig S3A). Each of these mutants harbors a one-nucleotide insertion mutation in the coding region of the targeted gene, resulting in an unaltered expression level of the corresponding gene but a shifted open reading frame (Appendix Fig S3A–C). The transcript levels of *ROS1* and *IDM1* were not affected by the *IDAP1* or *IDAP2* mutations that we examined, excluding the possibility that the *IDAP1* or *IDAP2* mutation causes DNA methylation level changes by altering the expression levels of *ROS1* and *IDM1* (Appendix Fig S3D).

Next, to investigate the direct effect of the *IDAP1* or *IDAP2* mutation on DNA methylation, we performed whole-genome bisulfite sequencing using DNA purified from 12-day-old *idap1-2*, *idap2-1*, *idm1-1*, and wild-type seedlings. The overall genome methylation levels were similar between the mutant plants and the wild-type control (Appendix Fig S4A). However, we identified 1,512 and 1,368 differentially methylated regions (DMRs) in *idap1-2* and *idap2-1*, respectively (Appendix Fig S4B and Table EV3). Among the 1,512 DMRs in *idap1-2*, 854 were hypermethylated (hyper-DMRs) and 658 were hypomethylated, whereas the 1,368 DMRs in *idap2-1* included 918 hypermethylated and 450 hypomethylated DMRs (Appendix Fig S4B). Similar to the situation in *ros1-4* and *idm1-1*, the hyper-DMRs

Figure 2. Characterization of the domain structures of IDAP1 and IDAP2, alignment of amino acid sequences of the predicted IDAP1 and IDAP2 homologs, and phylogenetic analysis of IDAP1 sequences from different species.

A Domain structure for IDAP1 and alignment of IDAP1 homologs with the *PIF/Harbinger* transposase. The catalytic DDE triad in the *PIF/Harbinger* nucleases is indicated by black boxes.

B Domain structure for IDAP2 and alignment of the myb-like DNA-binding domains of IDAP2 homologs. The putative triple conserved W residues in the *PIF/harbinger* myb-like DNA-binding domain are indicated by black boxes.

C Phylogenetic analysis of IDAP1. Clade I (IDAP1) is in red, Clade II (AT3G19120) is dark blue, Clade III (AT3G55350) is pale blue, Clade IV (ALP1) is yellow, Clade V is golden, and Clade VI (*Harbinger*) is light green. GenBank accession numbers are prefixed gi, and the others are accession numbers for sequences retrieved from the NCBI or OneKP databases.

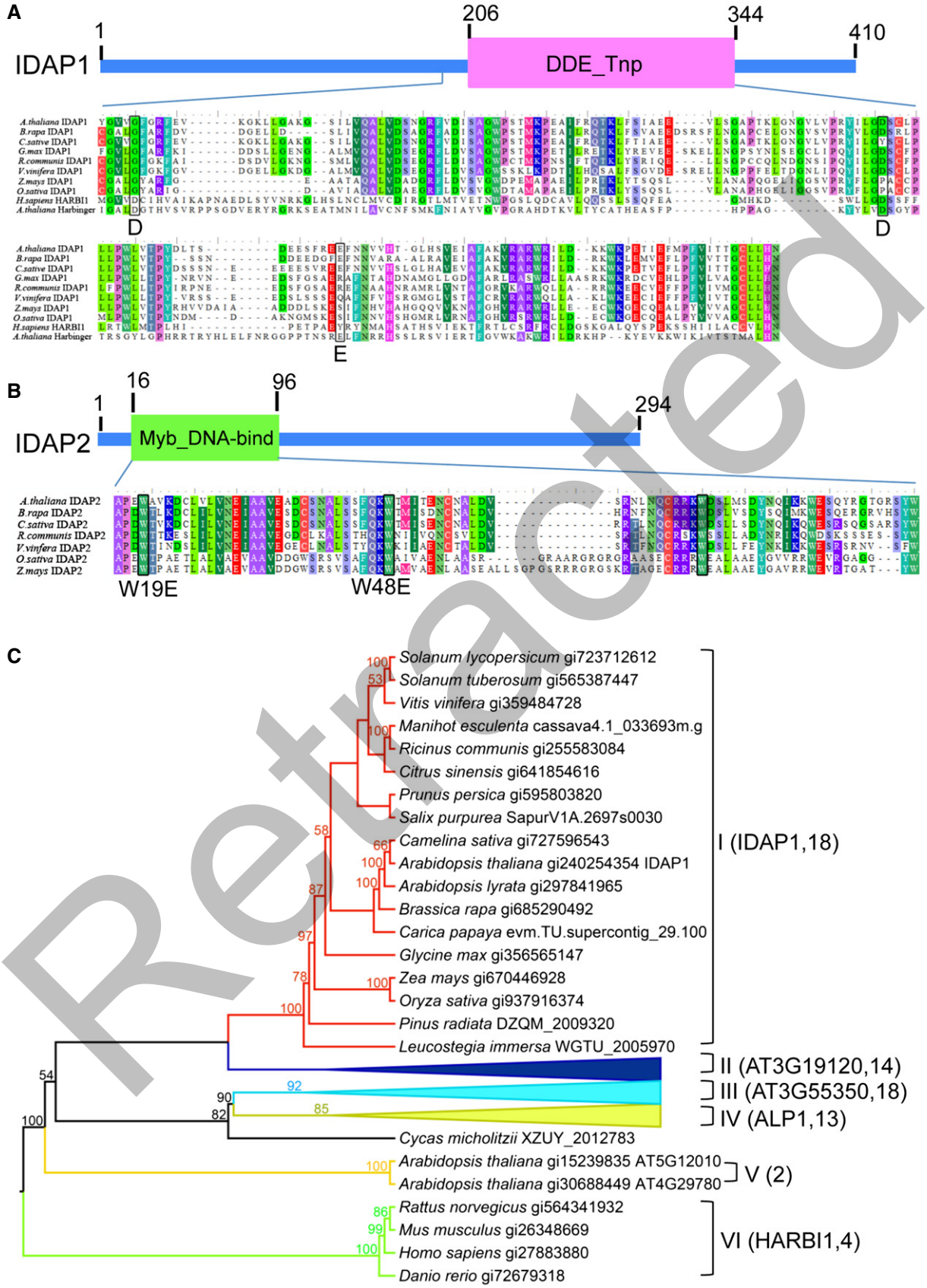


Figure 2.

in *idap1-2* and *idap2-1* were almost evenly distributed across the five chromosomes, with no apparent enrichment at the chromocenters (Appendix Fig S4C). Since IDAP1 and IDAP2 form a complex in the nuclei and may function together, we compared the hyper-DMRs in *idap1-2* and *idap2-1*. There were 327 overlapping hyper-DMRs, accounting for about 40 and 36% of the hyper-DMRs in *idap1-2* and *idap2-1*, respectively (Fig EV1A). This overlap between the hyper-DMRs in *idap1-2* and *idap2-1* arose significantly more often than expected by chance ($P < 0.01$) (Appendix Table S2). Box plot analyses indicated that the DNA methylation levels in all sequence contexts showed increases at the overlapping hyper-DMRs (Fig EV1A). The same analysis also revealed that the actual percentages of overlap in our comparisons could be larger. For example, although *idap1-2*-specific hyper-DMRs were not defined as hyper-DMRs in *idap2-1* due to our stringent criteria, the DNA methylation levels at these loci in *idap2-1* appeared to be higher than those in the wild-type controls (Fig EV1A). Taken together, our results indicate that IDAP1 and IDAP2 function together in the active DNA demethylation pathway.

IDAP1 and IDAP2 affect a subset of loci targeted by active DNA demethylation

To ascertain whether IDAP1 and IDAP2 function in ROS1-mediated active DNA demethylation, we compared the hyper-DMRs in *idap1-2* and *idap2-1* to those in *rdd* (a triple mutant for *ros1dml2dml3*). Approximately 40 and 50% of the hyper-DMRs in *idap1-2* and *idap2-1*, respectively, overlapped with the hyper-DMRs in *rdd* (Fig 4A and B), which were significantly enriched compared to the percentages expected by random chance from a dataset of this size ($P < 0.01$) (Appendix Table S2). The DNA methylation levels in all sequence contexts showed increases at these overlapping hyper-DMRs (Fig 4A and B). Notably, the DNA methylation levels at the *idap1-2*-affected (or *idap2-1*-affected) loci were also significantly increased in *rdd* (Fig 4A and B), although the increases were not large enough to be defined as hyper-DMRs according to the criteria used in this study. These results indicate that IDAP1, IDAP2, and ROS1 may function in the same active DNA demethylation pathway. We further categorized the hyper-DMRs from *idap1-2*, *idap2-1*, and *rdd* into TE_in_gene, TE_out_gene, gene and intergenic regions and found that the TE_out_gene was overrepresented in the overlapping hyper-DMRs between *idap1* or *idap2* and *rdd* (Appendix Fig S4D and E). Moreover, we found that the hyper-DMRs in *idap1-2* or *idap2-1* were enriched with H3K9me2 histone mark (Appendix Fig S4F and G). To examine whether IDAP1 and IDAP2 mediate active DNA demethylation at all loci targeted by ROS1, we made further analysis but found no real methylation level changes at some *rdd*-affected loci in *idap1-2* or *idap2-1* (Fig 4A and B), suggesting that IDAP1 and IDAP2 may only control DNA demethylation at a subset of ROS1-targeted loci.

To further test the hypothesis that IDAP1 and IDAP2 function in the same pathway as IDM1, we examined the methylation levels of *idap1-2*, *idap2-1*, and *ros1-4* in regions that are hypermethylated in *idm1-1* in CG, CHG, and CHH contexts and found that the overall CG, CHG, and CHH DNA methylation levels in these regions were significantly higher in the indicated mutants than those in the wild-type controls (Fig 4C). Moreover, many of the hypermethylated regions in *idm1-1* were also hypermethylated to different degrees in *idap1-2*, *idap2-1*, and *ros1-4* (Appendix Fig S4H–L). Specifically, approximately 48 and 63% of the hyper-DMRs in *idm1-1* were also hypermethylated in the *idap1-2* and *idap2-1* mutants, respectively (Fig EV1B and C), and these overlaps were statistically significant ($P < 0.01$) (Appendix Table S2). Our data suggest that IDAP1, IDAP2, and IDM1 function in the same active DNA demethylation pathway. Since MBD7 is also involved in IDM1-mediated active DNA demethylation, we compared the hyper-DMRs in *mbd7-1* with those in *idap2-1*. There are 407 overlapping hyper-DMRs, accounting for about 54 and 44% of the hyper-DMRs in *mbd7-1* and *idap2-1*, respectively (Fig EV2). This overlap was also significantly higher than what was expected by chance ($P < 0.01$) (Appendix Table S2). Box plot analyses indicated that the actual percentages of overlap in our comparisons could be larger (Fig EV2). Thus, IDAP1, IDAP2, and MBD7 may collaborate to regulate IDM1-mediated active DNA demethylation.

IDAP1 and IDAP2 antagonize transcriptional gene silencing

Histone modifications correlated with transcriptional activation and active DNA demethylation are required for the expression of many reporter and endogenous genes. Previous reports showed that IDM1 and other IDM1 complex components (IDM2, IDL1/IDM3, and MBD7) are required to prevent transcriptional silencing of specific transgenes and endogenous genes (Qian *et al*, 2012; Lang *et al*, 2015; Li *et al*, 2015a). To determine whether IDAP1 and IDAP2

Figure 3. IDAP1 interacts with IDAP2, and they co-localize in the nucleus.

A Yeast two-hybrid analysis of IDAP1 and IDAP2 interactions. Yeast cells expressing the indicated proteins from the pGAD-T7 (AD) and pGBK-T7 (BD) vectors were plated onto medium lacking Leu and Trp (SD-LT) (left) or medium lacking Leu, Trp, Ade, and His (SD-LTHA) (right).

B Pull-down assays showing that IDAP1 and IDAP2 interact with each other. MBP-IDAP1 and GST or GST-IDAP2 were co-expressed in *E. coli* BL21 (DE3). GST or GST-IDAP2 was purified with GST affinity beads. The eluted proteins and total proteins with or without IPTG treatment were separated by SDS–PAGE and stained with Coomassie Brilliant Blue. A, GST + MBP-IDAP1; B, GST-IDAP2 + MBP-IDAP1.

C Split luciferase complementation assays showing that IDAP1 interacts with IDAP2 in *N. benthamiana* leaves. 1, IDAP1-nLUC + cLUC; 2, IDAP1-nLUC + IDAP2-cLUC; 3, nLUC + IDAP2-cLUC; 4, nLUC + cLUC; 5, IDAP2-nLUC + cLUC; 6, IDAP2-nLUC + IDAP1-cLUC; 7, nLUC + IDAP1-cLUC. Three biological replicates were performed, and similar results were obtained.

D BiFC analysis of the interaction between IDAP1 and IDAP2 in *Agrobacterium*-infiltrated *N. benthamiana* leaves. YFP signals (green) and the overlay of the YFP signals and the DAPI nuclear stain (blue) are shown. YFP^C, C-terminal region of YFP; YFP^N, N-terminal region of YFP. Three biological replicates were performed, and similar results were obtained. Scale bar: 50 μm.

E Co-localization of IDAP1 and IDAP2 in the nucleus. The nuclear distribution of IDAP1 and IDAP2 was analyzed by immunostaining using anti-Myc (green) and anti-Flag (red) in wild-type plants or the F1 heterozygous plants expressing Myc-tagged IDAP1 and Flag-tagged IDAP2 under their native promoters. DNA was stained with DAPI (blue). The frequency of nuclei displaying each interphase pattern is shown on the right. Upper panel: WT; lower panel: F1 offspring of a cross between the *IDAP1-3Myc* and *IDAP2-3Flag* transgenic plants.

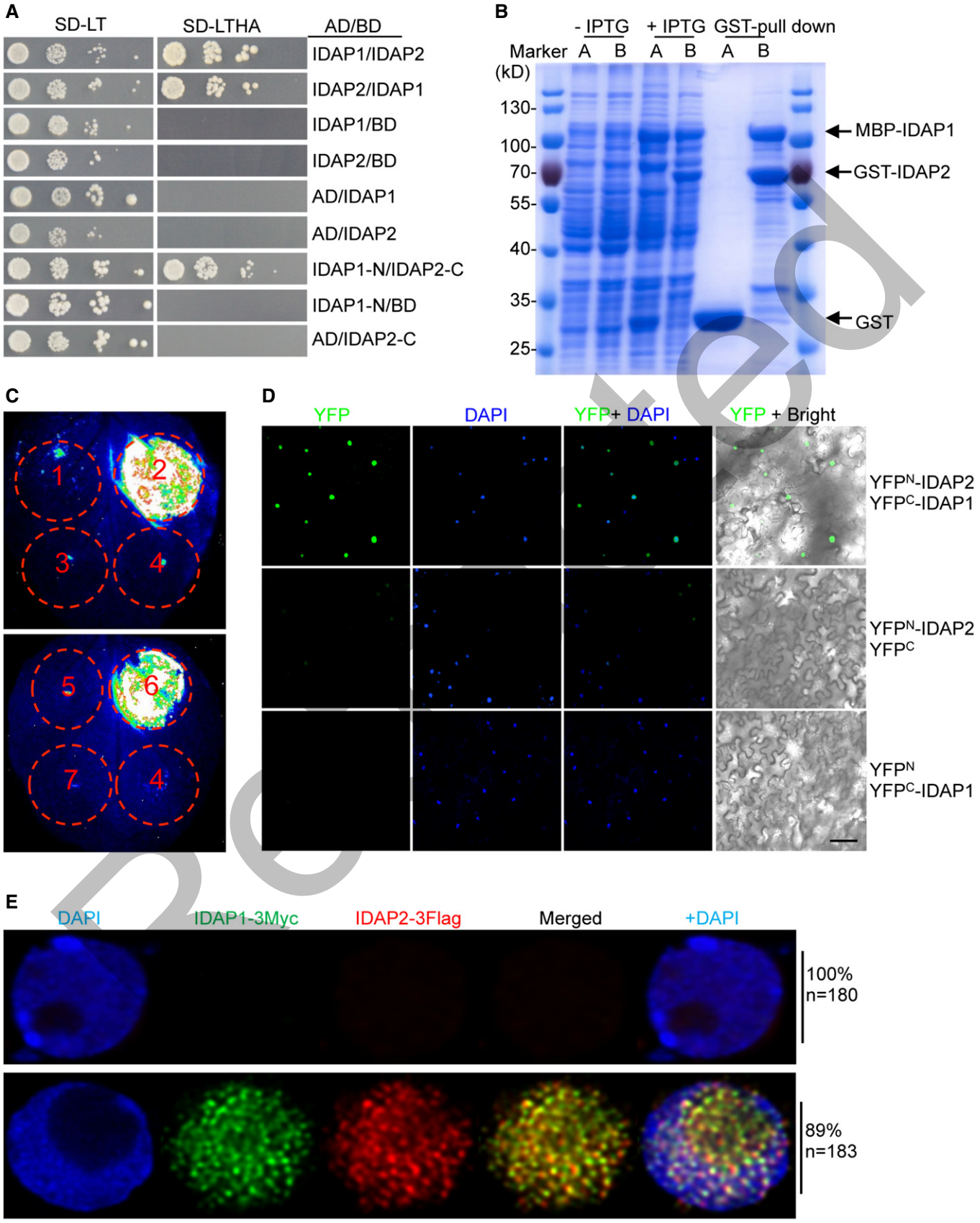


Figure 3.

participate in the anti-silencing of genes, we first introduced the *idap1-1* and *idap2-1* mutations into a transgenic line harboring the *LUC* and neomycin phosphotransferase II (*NPTII*) transgenes under the control of a dual mosaic virus *35S* promoter (Li *et al*, 2016). As expected, both mutations resulted in reduced LUC luminescence, although the phenotype in *idap2-1* was weaker compared to that in *idm1-14* and *idap1-1* (Fig 5A). Real-time PCR analysis showed that the transcript levels of *LUC* and *NPTII* were greatly reduced in *idap1-1* and *idap2-1* (Fig 5B), suggesting that IDAP1 and IDAP2 are required for expression of the transgenes. Bisulfite sequencing analysis showed that compared to the wild-type control, the DNA methylation level of the *35S* promoter region was slightly increased in *idap1-1* and *idap2-1* (Fig 5C).

To further confirm the anti-silencing role of IDAP1 and IDAP2, we also generated *idap1-3* or *idap2-2* mutation in a transgenic line (C24 background) with the *RD29A-LUC* and *35S-NPTII* reporter genes (Li *et al*, 2012) (Appendix Fig S3A). Similar to *idm1-3*, the *idap1-3* and *idap2-2* mutants exhibited a kanamycin-sensitive growth phenotype but produced normal LUC luminescence (Fig 5D). Real-time PCR analysis showed that the expression of *LUC* in *idap1-3* and *idap2-2* remained the same as that in the wild-type control, but the expression of *NPTII* in *idap1-3* and *idap2-2* background was highly repressed (Fig 5E). Our results are consistent with previous findings indicating that IDM1, IDM2, and MBD7 are required for the expression of *35S-NPTII* but not *Rd29A-LUC* and suggest that IDAP1 and IDAP2 are specifically involved in the anti-silencing of transgenes driven by the *35S* promoters (Li *et al*, 2012; Zhao *et al*, 2014; Wang *et al*, 2015).

To analyze the effects of the *IDAP1* and *IDAP2* mutations on endogenous gene expression, we randomly selected genes with DNA hypermethylation in *ros1-4*, *idap1-1*, and *idap2-1* (Appendix Fig S5) and measured their expression levels in multiple mutants using real-time RT–PCR. Similar to *ros1-4*, all *idap1* and *idap2* mutants we examined were defective in fully activating the expression of *AT1G28480, AT1G53265, AT1G66780, AT1G70840, AT1G77790*, *AT2G40420, AT2G42460, AT3G16930,* and *AT3G62470* (Fig 5F), indicating that IDAP1 and IDAP2 are important for the anti-silencing of these endogenous genes. All *idap1* and *idap2* mutants also had repressed expression of *AT1G62760* (Fig 5F). However, the extents of the repression in these mutants were significantly lower than that in *ros1-4*, suggesting that in some cases, ROS1-dependent but IDAP1- and IDAP2-independent mechanisms are responsible for the activation of genes. Thus, our results suggest that IDAP1 and IDAP2, similar to IDM1, are critical for preventing transcriptional silencing of some endogenous genes.

The DNA-binding activity is essential for the function of IDAP2 *in vivo*

The *PIF/Harbinger* transposases and HARBI1 were predicted to have DNA-binding capabilities because they contain a single HTH motif (Casola *et al*, 2007). However, results of electrophoretic mobility shift assays (EMSAs) demonstrated that they are incompetent to bind DNA (Sinzelle *et al*, 2008). Instead, the myb-like proteins and NAIF1 were found to bind the 9-bp palindromic sequence motif in transposon DNA, and the myb-like proteins contribute to transposition of *PIF/Harbinger* by recruiting the *PIF/Harbinger* transposases. The trihelix motif is important for the DNA-binding activity of the myb-like proteins. To determine whether IDAP1 and IDAP2 resemble the *PIF/Harbinger* transposases and the myb-like proteins, respectively, in binding DNA, we purified IDAP1 and IDAP2 from *E. coli* (Fig EV3A) and performed EMSA using a probe corresponding to the 9-bp palindromic sequence motif in transposon DNA (Sinzelle *et al*, 2008). Indeed, the addition of IDAP2 retarded the bands, and the binding was competitively blocked by unlabeled DNA of the same sequence (Fig 6A). The W19E and W48E mutations abolished the retarded bands (Fig 6B). We also tested DNA corresponding to the *35S* promoter locus and six of the hypermethylated regions identified from *idap2-1* and found that IDAP2 bound all of these DNA sequences (Figs 6C and EV3B and Appendix Fig S6). Unexpectedly, IDAP2 also bound to the control DNA sequences whose DNA methylation levels were not altered by the *idap2-1* mutation (Fig EV3B). Our results suggest that IDAP2 binds DNA without sequence specificity *in vitro*. In contrast to IDAP2, IDAP1 could not bind DNA *in vitro* (Fig 5B).

To explore whether the DNA-binding capability of IDAP2 is important for its role in active DNA demethylation and transcriptional gene activation, we performed complementation experiments in the *idap2-1* mutant with wild-type Flag-tagged IDAP2 or mutated IDAP2 (W48E) under the control of their native promoters. Although this mutation did not affect the interaction between IDAP2 and IDAP1 (Fig EV4A), only wild-type IDAP2, but not mutated IDAP2, could rescue the low LUC, repression of endogenous gene expression, and DNA hypermethylation phenotypes of the *idap2-1* mutant (Figs 6D and E and EV4B), suggesting that DNA binding is important for IDAP2 to execute its function in active DNA demethylation and transcriptional gene activation.

IDAP1 and IDAP2 associate with targeted loci and are important for the recruitment of IDM1 to specific targets

Since IDAP2 can bind to DNA *in vitro* and IDAP1 and IDAP2 function in the same pathway as IDM1, we hypothesized that IDAP1 and IDAP2 act together to recognize target loci and recruit the IDM1 complex to the target region. To test this, we set out to determine whether IDAP1 and IDAP2 associate with the IDM1 target loci. We performed chromatin immunoprecipitation (ChIP) using transgenic plants expressing Myc-tagged IDAP1 or Flag-tagged IDAP2 and found that both IDAP1 and IDAP2 were enriched at a subset of loci targeted by IDM1 (Fig 7A and B). Since IDAP1 could not bind DNA *in vitro*, IDAP1 may be recruited to the IDM1 target loci by IDAP2. We then assessed whether the accumulation of IDM1 at its target loci was affected in *idap2-1*. The *idap2-1* mutation markedly reduced the accumulation of IDM1 at IDM1 target loci (Fig 7C and Appendix Fig S7A), although approximately equal amounts of IDM1-HA-YFP protein were expressed in *idm1-3*, *idap1-1*, and *idap2-1* (Appendix Fig S7B), suggesting that IDAP2 is indeed important for IDM1 targeting to specific loci.

IDAP1 and IDAP2 are important for the *in vivo* function of IDM1

Previous studies showed that IDM1 is a histone H3 acetyltransferase that catalyzes H3K18 and H3K23 acetylation (Qian *et al*, 2012), and the levels of acetylated H3K18 and H3K23 are reduced at silenced *35S* promoter and endogenous targets (Qian *et al*, 2014). To

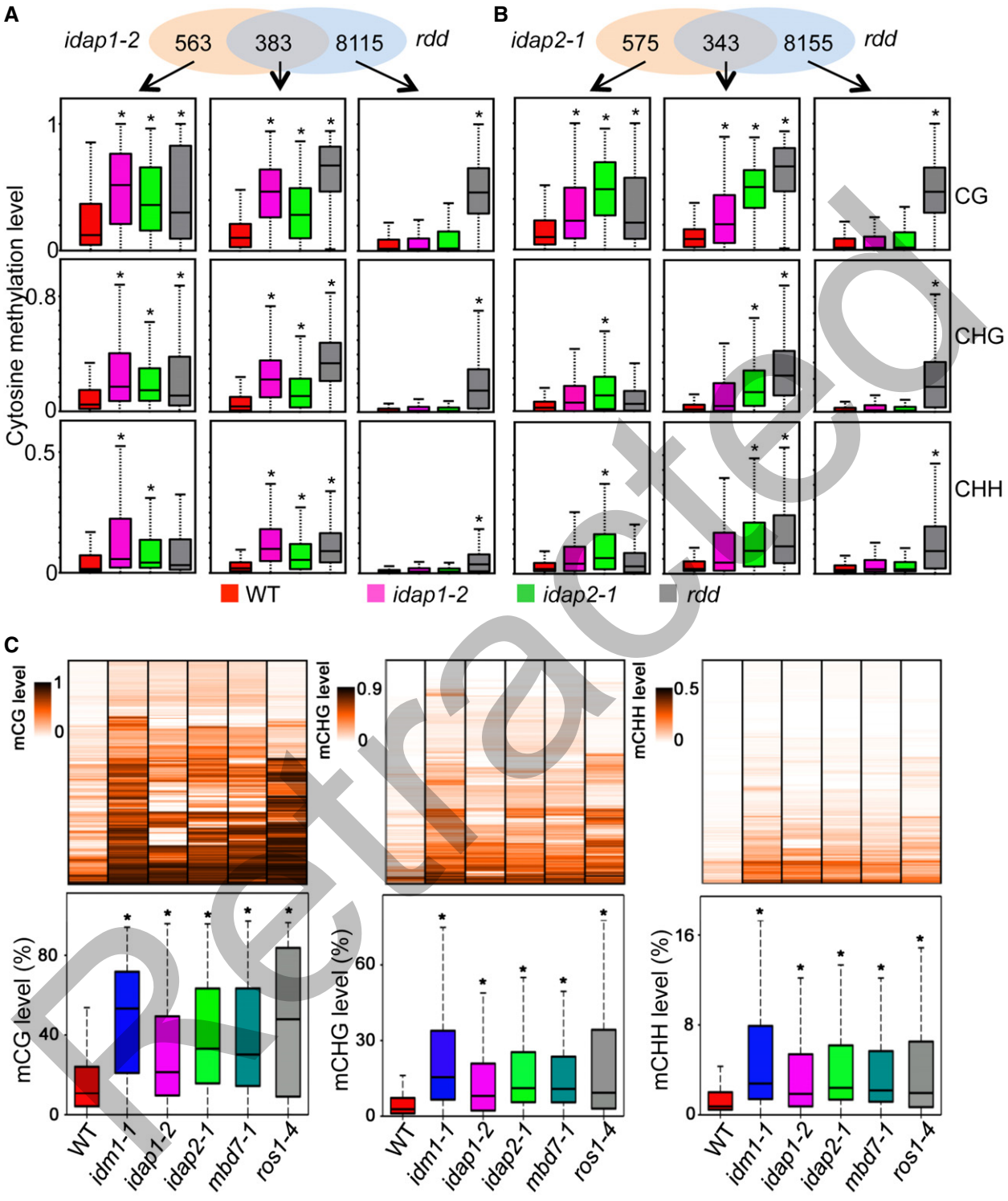


Figure 4. IDAP1 and IDAP2 function in active DNA demethylation.

A, B Venn diagram showing the number of hyper-DMRs that overlap between *idap1-2* and *rdd* (A) and between *idap2-1* and *rdd* (B). Box plots show the methylation level of each class of DMRs. Dark horizontal line, median; edges of boxes, 25th (bottom) and 75th (top) percentiles; whiskers, minimum and maximum percentage of DNA methylation. Asterisks indicate significant differences between the wild-type and the indicated mutants (*$P < 10^{-15}$, Mann–Whitney *U*-test).

C DNA methylation analysis in different mutant plants. Upper panel: heat map showing the methylation levels of IDM1 targeted hyper-DMRs in *idap1-2*, *idap2-1*, *mbd7-1*, *idm1-1*, and *ros1-4* in three contexts. Light yellow indicates low methylation, and black indicates high methylation. Lower panel: box plots showing CG, CHG, and CHH methylation levels at IDM1 targeted hyper-DMRs in the indicated mutants. Dark horizontal line, median; edges of boxes, 25th (bottom) and 75th (top) percentiles; whiskers, minimum and maximum percentage of DNA methylation. Asterisks indicate significant differences between the wild-type and the indicated mutants (*$P < 10^{-15}$, Mann–Whitney *U*-test).

determine whether IDAP1 and IDAP2 are required for IDM1 to acetylate H3K18 and H3K23, we examined the accumulation of acetylated H3K18 at several representative loci in the *idap1* and *idap2* mutants. Our results revealed that the level of acetylated H3K18 was significantly reduced in *idap1-2* and *idap2-2* at the *35S* promoter, as in *idm1-14* (Fig 7D). In addition, the level of acetylated H3K18 was substantially reduced not only in *idm1-1* but also in *idap2-1* at several endogenous IDM1 target loci (Fig 7E and Appendix Fig S7C). Our results suggest that IDAP1 and IDAP2 are important for the *in vivo* function of IDM1.

Discussion

Our study demonstrates that IDAP1 and IDAP2 are two key components of the IDM1 complex required for IDM1 targeting. This is the first time that the domesticated transposase and myb-like genes of *PIF/Harbinger* were found to regulate active DNA demethylation and transcriptional gene activation. Our mass spectrometry results revealed that in addition to IDM2, MBD7, and IDL1, IDAP1 and IDAP2 are also associated with IDM1 (Appendix Table S1). Thus, these six proteins form a protein complex, with IDM1 as the core histone acetyltransferase that catalyzes H3K18 and H3K23 acetylation and assists in ROS1 targeting (Qian *et al*, 2012). To date, MBD7, IDAP1, and IDAP2 were found to regulate IDM1 targeting. However, we could not exclude the possibility that there exist other factors that interact with IDM1 and affect IDM1 targeting. Changing the purification method or improving the sensitivity of protein identification may reveal more factors that weakly or transiently associate with IDM1 but play important roles in regulating IDM1 targeting.

Since IDAP1 and IDAP2 physically interact, and when either *IDAP1* or *IDAP2* mutates, IDM1 targeting, active DNA demethylation, and transcriptional gene silencing at specific loci were affected (Figs 4 and 5), we concluded that IDAP1 and IDAP2 cooperate to promote IDM1 targeting and H3K18 and H3K23 acetylation at many endogenous loci. MBD7 was previously found to assist in IDM1 targeting by binding to methylated DNA via its MBD motifs (Lang *et al*, 2015; Li *et al*, 2015a; Wang *et al*, 2015). To summarize our findings on IDM1 targeting regulated by these three proteins, we proposed a model in which domesticated IDAP1 and IDAP2 coordinate with MBD7 to regulate IDM1 targeting (Fig 8). The acetylated H3K18 and H3K23 generated by IDM1 are then recognized by ROS1 itself or ROS1-interacting proteins. Eventually ROS1 and other related DNA demethylases are recruited to specific loci to prevent DNA hypermethylation and transcriptional gene silencing. It is interesting that both the IDAP1 and IDAP2 complex and MBD7 are required for IDM1 targeting in most cases (Fig EV2), suggesting that MBD7 alone is not sufficient to recruit IDM1 to its target loci. IDAP2 contains a myb-like DNA-binding domain at its N-terminus, and our EMSA assays confirmed that IDAP2 could bind DNA sequences *in vitro* (Figs 6A–C and EV3B). However, we could not find conserved DNA sequences for IDAP2 to bind based on our EMSA results and bioinformatics analysis of the hyper-DMRs in *idap2-1*. Our results suggest that, unlike typical myb transcription factors that bind to specific DNA sequence *in vitro* (Dubos *et al*, 2010), IDAP2 lost the sequence specificity for DNA binding during evolution. Since transposases naturally target their own TEs, losing sequence specificity may increase the diversity of their substrates. In support, our whole-genome bisulfite sequencing data revealed that *PIF/Harbinger* TEs are not IDAP1's and IDAP2's targets anymore (Appendix Fig S4M). Instead, IDAP1 and IDAP2 can target many gene loci. This could be beneficial for plant hosts that recruit these transposon-derived genes as important epigenetic regulators. Because active DNA demethylation occurs in the context of the chromatin structure, we propose that local chromatin environment, rather than sequence itself, is important for determining the specificity. Since the mutations of *IDAP1* and *IDAP2* preferentially affect DNA methylation levels in TEs and repeats, IDAP1 and IDAP2 may be responsible for the recognition of genes and TEs that contain highly repetitive sequences, whereas MBD7 preferentially binds and targets IDM1 to DNA regions with high CG methylation. As a result of the joint action of MBD7, IDAP1, and IDAP2, IDM1 can then be targeted to regions that have both highly CG-methylated and highly repetitive features. A previous study showed that IDM1 predominantly targets highly homologous genes and repetitive sequences, which is consistent with this model (Qian *et al*, 2012). Although IDAP1, IDAP2, and MBD7 jointly act to mediate IDM1 targeting at most IDM1-targeted loci, it is still possible that not all IDM1-targeted loci require IDAP1, IDAP2, and MBD7 for IDM1 targeting and either IDAP1, IDAP2, or MBD7 alone may be sufficient to regulate IDM1 targeting. Alternatively, IDAP1, IDAP2, MBD7, and other factors may form different protein complexes using different combinations, allowing diversification and specification of IDM1 targeting regulation. In addition, like MBD7, the IDAP1 and IDAP2 complex may contribute to active DNA demethylation by recruiting histone modification enzymes other than IDM1 because IDM1 only regulates DNA demethylation at a subset of ROS1 targets.

Figure 5. IDAP1 and IDAP2 antagonize transcriptional gene silencing.

A The *IDAP1* or *IDAP2* mutation causes silencing of the *35S-LUC* reporter gene. The reporter gene was introduced into *idap1-1* or *idap2-1* by crossing. Seedlings grown in MS plates were imaged after being sprayed with luciferase substrate.

B Real-time PCR analysis of the expression levels of the *LUC* and *NPTII* reporter genes in the different genotypes in (A).

C Bisulfite sequencing results showing the methylation levels of the *35S* promoter in wild-type and different mutant plants.

D Dysfunction of *IDAP1* or *IDAP2* causes silencing of the *35S-NPTII* but not the *RD29A-LUC* transgene. Seedlings grown in MS plates were imaged after cold treatment at 4°C for 48 h. For kanamycin resistance tests, the seeds were planted on MS medium supplemented with 50 mg/l kanamycin and incubated for 2 weeks before being photographed. The *idap1-3* and *idap2-2* mutants in C24 background (harboring the *RD29A-LUC* and *35S-NPTII* transgenes) were generated using the CRISPR/Cas9 system.

E Real-time PCR analysis of the expression levels of *LUC* and *NPTII* reporter genes in the different genotypes in (D).

F The effects of the *ros1-4*, *idap1*, and *idap2* mutations on expression of the hypermethylated genes or nearby genes. Gene expression levels in the indicated mutants were determined by real-time PCR. *TUB8* was used as an internal control.

Data information: Standard deviations (error bars) were calculated from three biological replicates, $*P < 0.05$ (Student's *t*-test).

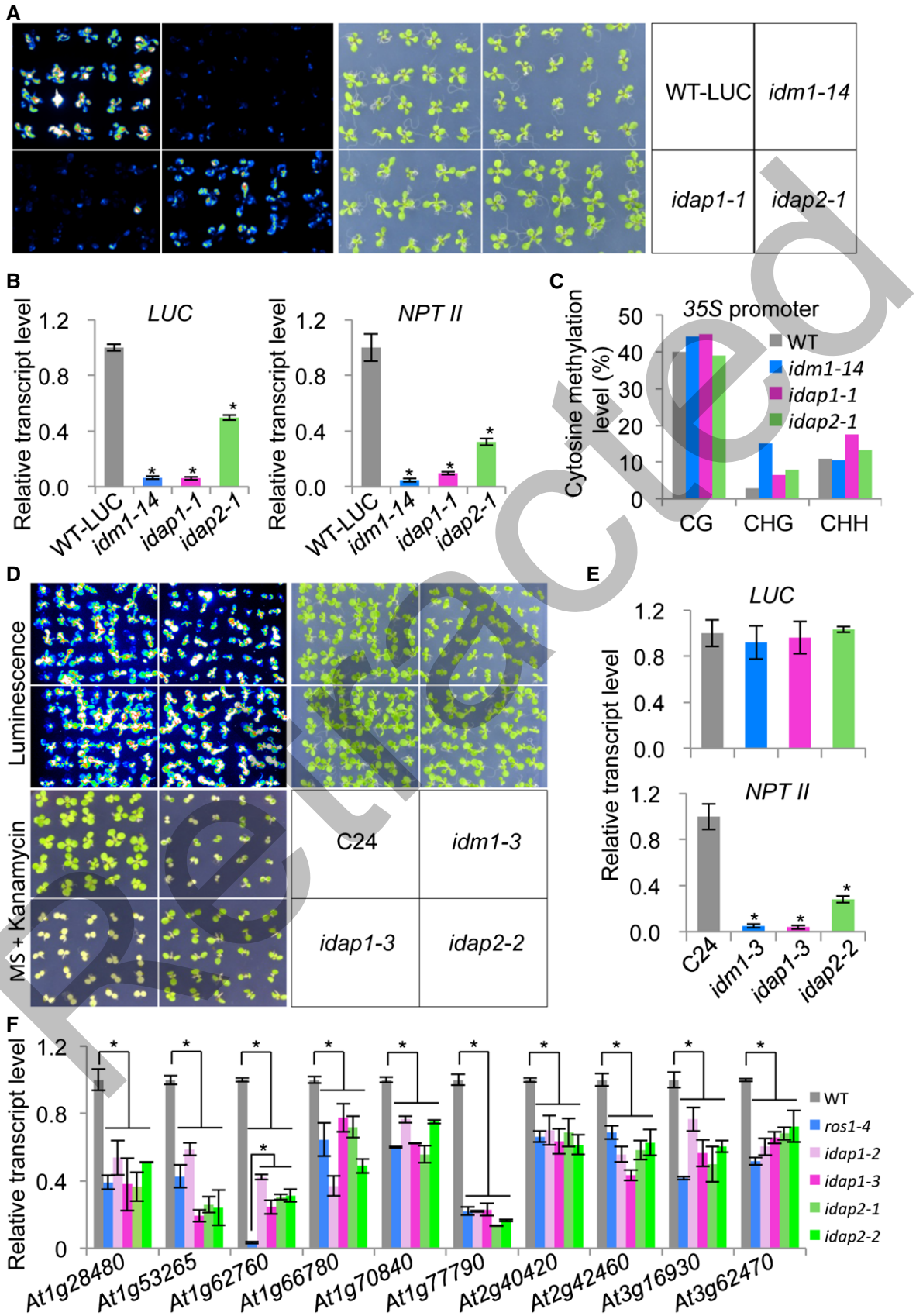


Figure 5.

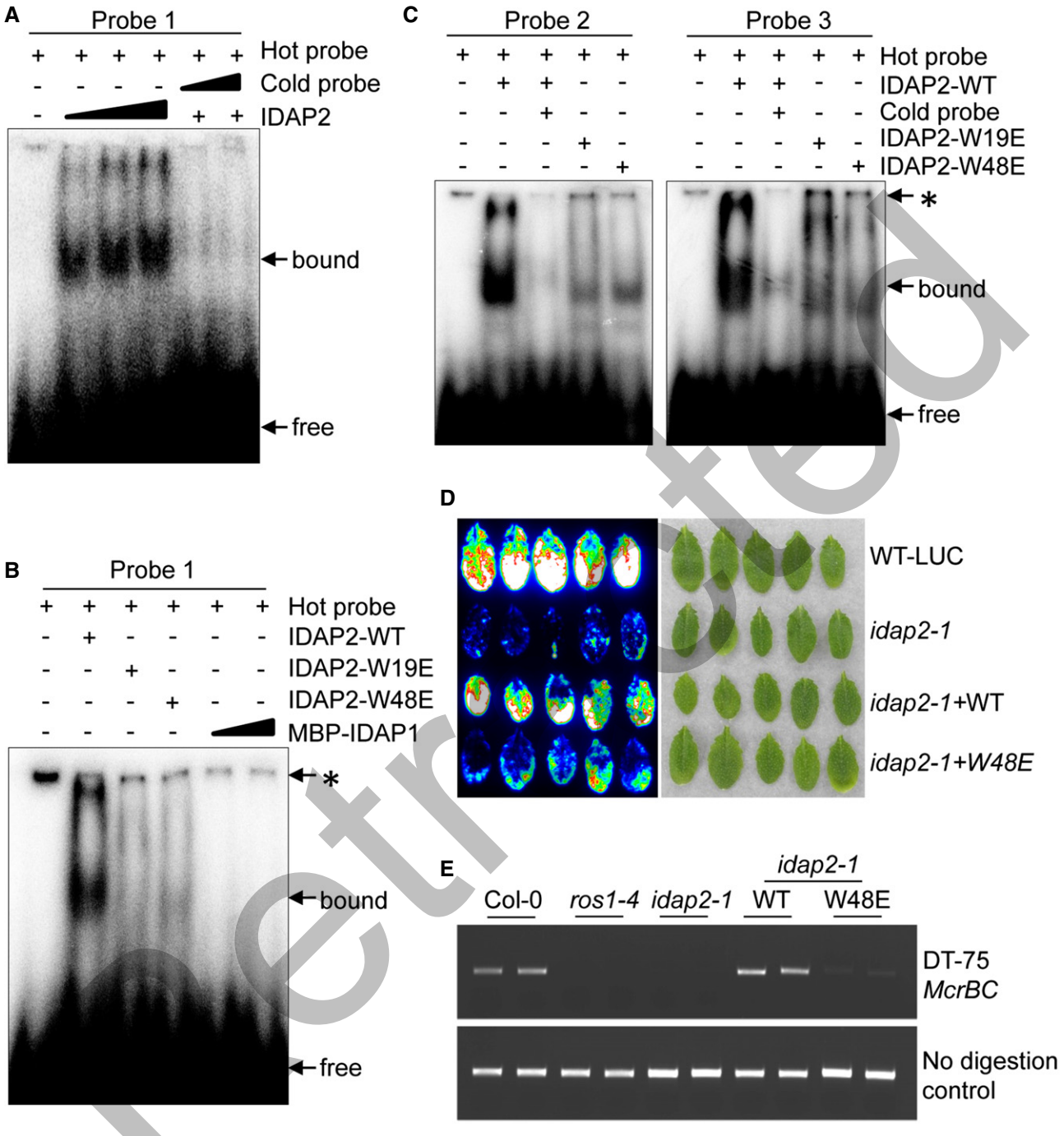


Figure 6. The DNA-binding activity is essential for the proper function of IDAP2.

A EMSA showing IDAP2 binding to DNA probes and competition by unlabeled DNA probes.

B, C The effects of the W19E and W48E mutations on IDAP2 binding to DNA probes. Asterisks indicate non-specific bands.

D Complementation assay showing that wild-type, but not the mutant, form (W48E) of IDAP2-3Flag can restore *LUC* expression in *idap2-1*.

E Analysis of the DNA methylation level at the DT-75 locus by chop-PCR. *McrBC* is a methylation-sensitive restriction enzyme, and DNA hypermethylation results in a reduced level of the PCR product. Undigested DNA was amplified as a control.

To date, a few cases of domesticated transposons in *Arabidopsis* have been reported. *DAYSLEEPER* was found to be derived from *hAT* transposons and was essential for normal plant development (Bundock & Hooykaas, 2005). It was proposed that DAYSLEEPER may recruit DNA repair factors, such as KU70 and chromatin remodeling factors, following DNA binding. However, the detailed mechanisms underlying DAYSLEEPER promotion of plant development are unclear. FHY3 and FAR1 are two domesticated

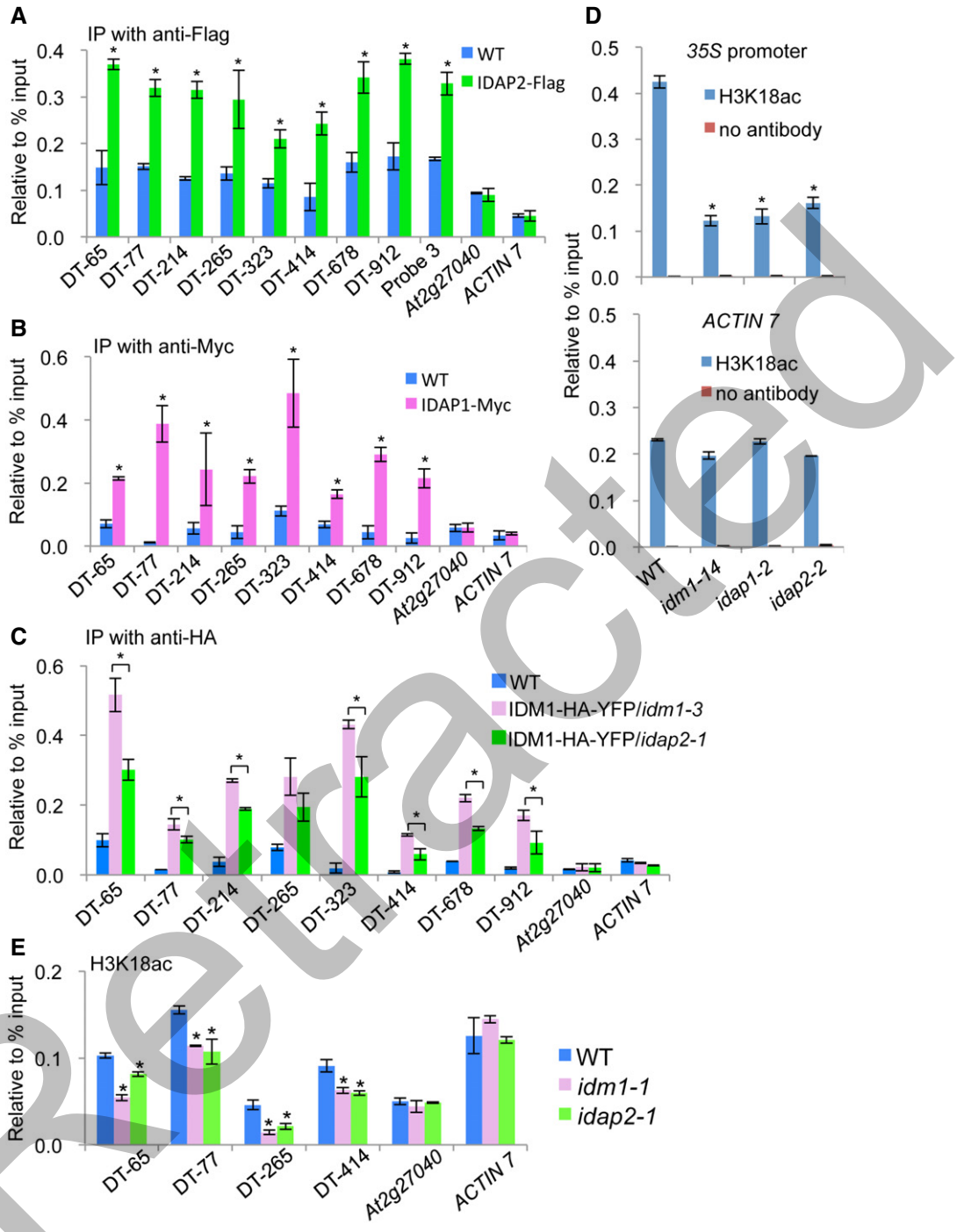


Figure 7. IDAP1 and IDAP2 associate with their target loci and are required for the recruitment of IDM1.

A Association of IDAP2 with hyper-DMR loci. The DT loci are shared hyper-DMRs between *idap2-1* and *idm1-1*. ChIP was performed in the wild-type and *IDAP2-3Flag* transgenic plants with an anti-Flag antibody.

B Association of IDAP1 with IDM1 targeted hyper-DMRs. ChIP was performed in the wild-type and *IDAP1-3Myc* transgenic plants with an anti-Myc antibody.

C The effect of the *IDAP2* mutation on the recruitment of IDM1 to its target loci. ChIP was performed in the wild-type, *IDM1-HA-YFP/idm1-3*, and *IDM1-HA-YFP/idap2-1* transgenic plants with an anti-HA antibody. The transgene was introduced into *idap2-1* by crossing.

D Detection of the levels of acetylated H3K18 at the *35S* promoter and *ACTIN7* in the wild-type, *idm1-14*, *idap1-2*, and *idap2-2* plants. ChIP was performed using an antibody against acetylated H3K18.

E The effect of the *IDAP2* mutation on the levels of acetylated H3K18. Acetylated H3K18 levels at the DMRs and control regions in the wild-type, *idm1-1*, and *idap2-1* plants were determined by ChIP with an antibody against acetylated H3K18.

Data information: In (A–E), the ChIP signal was quantified relative to input DNA. Standard deviations (error bars) were calculated from three biological replicates, $^*P < 0.05$ (Student's *t*-test).

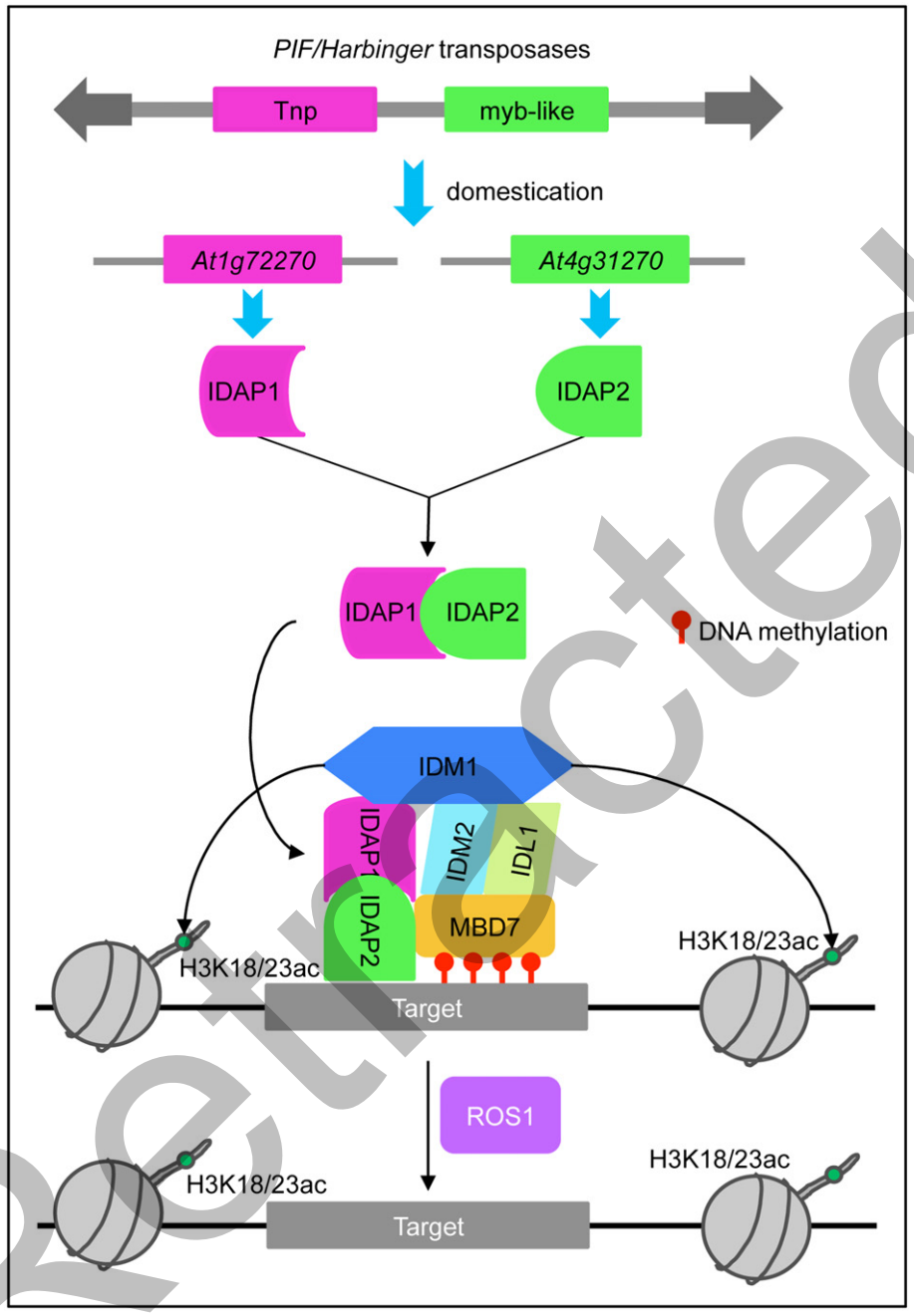


Figure 8. A working model for the functions of IDAP1 and IDAP2 in active DNA demethylation in *Arabidopsis*.
The domestication of the *PIF/Harbinger* DNA transposon gave rise to two host genes, *IDAP1* and *IDAP2*, which encode a putative DNA transposase and a myb-like DNA-binding protein, respectively. IDAP1 and IDAP2 form a complex in the nucleus where they cooperatively recognize target loci with TE-like structures and recruit the histone acetyltransferase IDM1 to these target loci. Since a previously identified protein MBD7, which recognizes highly methylated regions, also regulates IDM1 targeting, the joint action of MBD7, IDAP1, and IDAP2 may facilitate the targeting of the IDM1 complex to specific loci with multiple epigenetic features. The IDM1 complex then acetylates H3K18 and H3K23 at those loci and facilitates active DNA demethylation mediated by ROS1.

Mutator-like transposases that regulate light signaling. They have been co-opted to become transcription factors that directly activate the expression of *FHY1* and *FHL*, two important genes in light responses (Lin *et al*, 2007). Recently, ALP1 was shown to be a domesticated *PIF/Harbinger* transposase that antagonizes PcG protein-mediated transcriptional gene silencing (Liang *et al*, 2015). Similar to ALP1, IDAP1 evolved from the *PIF/Harbinger* class of transposases, lost its DNA endonuclease activity, and plays an important role in transcriptional anti-silencing. However, ALP1 exerts its anti-silencing role by associating with the PcG and trithorax group (trxG) proteins and inhibiting trimethylation of H3K27 by the PRC2 complex, the major catalytic protein complex of the PcG

proteins (Liang *et al*, 2015); in contrast, IDAP1 associates with the IDM1 complex and facilitates IDM1 targeting, thereby contributing to active DNA demethylation and transcriptional anti-silencing. Despite using different mechanisms, ALP1 and IDAP1 represent two important cases of domesticated transposases that became regulatory components of the epigenetic machinery.

Transposable elements are typically silenced by siRNA-mediated DNA methylation in plant hosts (Law & Jacobsen, 2010; Matzke & Mosher, 2014). The involvement of ALP1 in blocking PRC2-mediated H3K27 trimethylation and the involvement of IDAP1 and IDAP2 in IDM1-mediated active DNA demethylation may represent important mechanisms for the transposons to evade DNA methylation-mediated transcriptional silencing by the host. Recently, a transposon-encoded protein VANC was also found to play an important role in transcriptional anti-silencing by inhibiting the host DNA methylation machinery, although it is not domesticated. *VANC* genes are widespread in the *Mutator*-like elements (*MULEs*) without degenerated terminal inverted repeats, and VANC may have facilitated the success of *MULE* transposition in *Arabidopsis* (Fu *et al*, 2013). Alternatively, plant hosts may recruit these transposon-derived genes as important epigenetic regulators. The advantages of using transposon-derived proteins as epigenetic regulators may be that these proteins are usually capable of recruiting the machinery for gene silencing or activation. Moreover, in plants, many transposons are stress inducible, and their activation is strongly correlated with activation of host genes in response to various types of stresses (Ito *et al*, 2011; Cavrak *et al*, 2014; Sanchez & Paszkowski, 2014; Makarevitch *et al*, 2015). It appears that recruiting transposon-derived proteins is beneficial for the hosts in responding to different stresses. Although dysfunction of IDAP1 and IDAP2 does not result in obvious developmental phenotype under normal growth conditions, their anti-silencing role may contribute to the plant host responses to different stresses. It will be interesting to investigate the roles of this complex in stress responses in the future.

Materials and Methods

Plant materials and growth conditions

The T-DNA insertion mutant line Sail_1301_B12 (*idap2-1*, Col background) for *IDAP2* and the Ds transposon insertion mutant line GT_5_84536 (*idap1-1*, Ler background) for *IDAP1* were obtained from the European *Arabidopsis* Stock Centre (http://arabidopsis.info). After treatment at 4°C for 2 days, plants were grown on 1/2 Murashige–Skoog (MS) solid medium and later in soil under long-day conditions (16 h of light and 8 h of darkness) at 23°C.

To generate the *idap1* and *idap2* mutant alleles using the CRISPR/Cas9 system, specific 20-bp sgRNAs targeting *IDAP1* or *IDAP2* were cloned into the BbsI site of the AtU6-26SK vector (Feng *et al*, 2013) and then subcloned into the pCAMBIA1300 binary vector together with *Cas9* driven by the *35S* promoter. The constructs were transformed into the wild-type Col-0 plants or plants in the C24 background (harboring the *RD29A-LUC* and *35S-NPTII* transgenes) using *Agrobacterium tumefaciens* GV3101 by the standard floral dipping method (Clough & Bent, 1998). Homozygous mutant plants with the *Cas9* transgene outcrossed were used for experiments. For complementation of the mutants, *IDAP1* and *IDAP2* genomic DNA fragments with approximately 2-kb promoter regions were amplified from wild-type Col-0 genomic DNA by PCR and cloned into the pCAMBIA1305 vector for plant transformation. All mutations were introduced into the *proIDAP2::IDAP2-3Flag* construct through site-directed mutagenesis with the QuikChange II XL Site-Directed Mutagenesis kit according to the manufacturer's instructions (Agilent Technologies). *Agrobacterium tumefaciens* strain GV3101 carrying various *IDAP1* or *IDAP2* constructs was used to transform the mutant plants via the standard floral dipping method (Clough & Bent, 1998). Primary transformants were selected on 1/2 MS plates containing 25 mg/l hygromycin. Homozygous complemented lines were used for phenotypic analysis and ChIP assays. The *IDM1-3HA-YFP/idm1-3* transgenic plants used for ChIP assays were previously described (Li *et al*, 2015a). The *IDM1-3HA-YFP/idap2-1* transgenic plants were generated by crossing the *IDM1-3HA-YFP/idm1-3* with the *idap2-1* mutant plants.

Affinity purification and mass spectrometry

Flower tissues (about 10 g) collected from the *IDAP1-3Myc*, *IDAP2-3Flag*, or *MBD7-3Myc* transgenic plants were ground in liquid N_2 to a fine powder. The flower powder was suspended in 50 ml lysis buffer (50 mM Tris–HCl [pH 8.0], 230 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 0.2% NP-40, 0.5 mM DTT, 1 mM PMSF, and 1 protease inhibitor cocktail tablet [14696200; Roche]) and then incubated at 4°C for 5 min. After centrifugation for 15 min at 18,300 *g* at 4°C, the whole protein crude sample was incubated with anti-Myc agarose (A7470; Sigma) or anti-Flag magnetic beads (M8823; Sigma) for at least 3 h. Proteins purified with anti-Myc agarose were eluted with 0.1 M ammonium hydroxide (pH 11.5). Proteins purified with anti-Flag beads were eluted with 3× Flag peptide (F4799; Sigma).

Mass spectrometric identification of the affinity-purified proteins was performed as described (Li *et al*, 2015a). Briefly, after SDS–PAGE, the proteins were digested in-gel with trypsin (0.5 ng/μl), and the digested peptides were extracted with 50% formic acid/50% acetonitrile. The extracted peptides were separated by HPLC and were then sprayed into a Velos Pro Orbitrap Elite mass spectrometer (Thermo Scientific, USA). Database searches were performed on the MASCOT server (Matrix Science Ltd., London, UK) against the IPI (International Protein Index) *Arabidopsis* protein database.

Co-immunoprecipitation

Flower tissues (about 1.5 g) from F1 hybrids of the *IDAP1-3Myc* and *IDM1-HA-YFP* transgenic plants, F1 hybrids of the *IDAP2-3Flag* and *IDM1-HA-YFP* transgenic plants, or the *IDM1-HA-YFP* transgenic plants were ground in liquid N_2. The fine powder was suspended in 7 ml lysis buffer (50 mM Tris–HCl [pH 8.0], 230 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 0.2% NP-40, 0.5 mM DTT, 1 mM PMSF, 100 μM MG132 [Gene Operation], and 1 protease inhibitor cocktail tablet [14696200; Roche]). The crude protein extract was incubated with anti-Myc agarose (A7470; Sigma) or anti-Flag magnetic beads (M8823; Sigma) for about 3 h. Anti-Flag (F1804; Sigma), anti-HA (H9658; Sigma), and anti-Myc (M4439; Sigma) antibodies were used as primary antibodies for Western blot analysis.

Gel filtration

For gel filtration, the crude protein extract was prepared from flower tissues (about 1 g) as described in the affinity purification section above and then passed through a 0.22-μm filter. About 1 ml of protein extract was loaded onto a Superdex 200 (10/300) column (17-5175-01; GE Healthcare), and 10 fractions were collected. A 30-μl aliquot of each fraction was run on 10% SDS–PAGE for Western blot analysis.

Immunolocalization

Immunofluorescence localization was performed as described by Martinez-Macias *et al* (2012). Briefly, flower tissues were used for nuclei preparation. Nuclei preparations were incubated overnight at room temperature with rabbit anti-Flag (05-724; Millipore) and mouse anti-Myc (02022-1; Abbkine) and then incubated with mouse Alexa 488 (A11001m; Invitrogen)- and rabbit Alexa 594 (A11012m; Invitrogen)-conjugated secondary antibodies for 2 h at 37°C. After PBS washes, DNA was counterstained using DAPI in Prolong Gold (Invitrogen). Nuclei were observed with a Leica TCS SP8 STED 3X (confocal) at 100× magnification.

Phylogenetic analysis of IDAP1 and IDAP2 with *PIF*/*Pong*/*Harbinger* proteins

The analysis involved 70 (IDAP1)/19 (IDAP2) amino acid sequences that were identified by BLAST against non-redundant proteins in the NCBI and OneKP database and included IDAP1/2 proteins, *Pong* proteins, *PIF* proteins, and *Harbinger* proteins to represent the different lineages within each group. These sequences were aligned with PROMALS3D according to the 3D structures of IDAP1 and IDAP2. The evolutionary history was inferred using the maximum-likelihood method with RAxML7.0.4 for IDAP1 and the neighbor-joining method with MEGA6.06 for IDAP2. There are a total of 191 (IDAP1)/678 (IDAP2) positions in the final matrix. Bootstrap values were calculated from 1,000 replicates and the bootstraps < 50% were hidden.

Real-time RT–PCR/RT–PCR

Total RNA was extracted from 14-day-old seedlings using the RNeasy Plant Mini Kit (Qiagen), and contaminating DNA was removed with the RNase-free DNase Kit (Qiagen). About 1 μg mRNA was used for first-strand cDNA synthesis with the SuperScript™ III First-Strand Synthesis System (Invitrogen) for RT–PCR following the manufacturer's instructions. The cDNA reaction mixture was then diluted 5 times, and 1 μl was used as a template in a 25 μl PCR with iQ SYBR Green Supermix. The RNA transcript levels were determined by semiquantitative RT–PCR or quantitative RT–PCR. *ACTIN4* was used as an internal control. The primers used for PCR are listed in Appendix Table S3.

Yeast two-hybrid assay

The coding sequences of *IDAP1*, *IDAP2*, *MBD7*, *IDM2*, *IDL1*, and *IDM1* were amplified by PCR and then subcloned into pGBK-T7 or pGAD-T7 (Clontech) to generate DNA-binding or activation domain fusion constructs, respectively. For protein interaction analysis, two combinatory constructs were transformed simultaneously into the yeast strain AH109 (Clontech) and tested for Leu, Trp, Ade, and His auxotrophy according to the manufacturer's protocols.

Pull-down assay

MBP-IDAP1 and GST or GST-IDAP2 were cloned into self-modified pQLink vectors and transformed into *E. coli* BL21 (DE3). Protein co-expression was induced with 0.5 mM IPTG after the cultures reached an OD_{600} of 0.8 for 16 h at 18°C. The harvested bacteria were resuspended in lysis buffer (30 mM HEPES (pH 7.4), 500 mM NaCl, 5 mM beta-mercaptoethanol), sonicated on ice (Bioruptor® Plus) and centrifuged for 15 min at 10,000 *g*. The supernatant was incubated with pre-equilibrated GST magnetic beads (BeaverBeads™, Beaver Nano-Technologies Co., Ltd.) for 1 h. After a wash step, the proteins were eluted with the lysis buffer containing 10 mM GSH (pH 8.0). The eluted proteins and protein lysates with or without IPTG treatment were separated by SDS–PAGE and stained with Coomassie Brilliant Blue.

Split luciferase complementation assay

The coding sequences of *IDAP1*, *IDAP2*, *MBD7*, *IDM2*, and *IDL1* were amplified by PCR and subcloned into the pCAMBIA1300-NLUC or pCAMBIA1300-CLUC vector (Chen *et al*, 2008) to generate N-terminal or C-terminal luciferase fusion constructs, respectively. For protein interaction analysis, *Agrobacterium tumefaciens* GV3101 carrying different constructs was cultured overnight. After resuspension in buffer containing 10 mM $MgCl_2$, 150 μM acetosyringone, and 10 mM MES (pH 5.7) at an OD_{600} of 1.0 and incubation for 3 h at room temperature in the dark, equal amounts of culture were mixed in different combinations, and the mixture was infiltrated into *N. benthamiana* leaves. To prevent silencing of these genes, a construct encoding the viral p19 protein was infiltrated at the same time (Shamloul *et al*, 2014). Two days after infiltration, luciferase activity was detected with a luminescence imaging system (Princeton Instrument).

Bimolecular fluorescence complementation (BiFC) assay

BiFC assays were performed as described by Walter *et al* (2004).

Whole-genome bisulfite sequencing and data analysis

Genomic DNA was extracted from 14-day-old seedlings using the DNeasy Plant Mini Kit (Qiagen) and sent to the Center for High Throughput Sequencing (Biodynamic Optical Imaging Center, Peking University) for bisulfite treatment, library preparation, and sequencing (Wu *et al*, 2015). Briefly, approximately 4 μg of genomic DNA combined with 1% spiked-in λDNA was sonicated into small fragments (approximately 250 bp). The DNA fragments were end-repaired and ligated with adenine at their 3′ end. After ligation with Illumina DNA adaptors, the DNA fragments were treated with a MethylCode™ Bisulfite Conversion Kit (Invitrogen) to induce chemical conversion of unmethylated cytosines into uracils. Subsequently, the bisulfite-converted DNAs were

amplified by high-fidelity PCR for 4–7 cycles. Finally, the quality-ensured libraries were used for paired-end deep sequencing on an Illumina HiSeq 2000 system. Adapter and low-quality sequences were discarded using the following criteria: (i) reads containing more than 10% N bases; (ii) reads containing more than 50% bases with a sequencing quality score below 5; (iii) reads with length shorter than 37 bases after trimming of the adaptor sequences.

For data analysis, clean reads were mapped to the *Arabidopsis thaliana* TAIR 10 genome using BSMAP allowing two mismatches. Identification of DMRs was performed as described (Zhang *et al*, 2013) with some modifications. The DNA methylation level in every 200-bp window with a step size of 50 bp was compared between the wild-type and mutant plants using Fisher's exact test with a *P*-value cutoff of 0.05. The *P*-values were then adjusted using the Benjamini–Hochberg method to control for FDRs. Windows with ≥ 7 DMCs (differentially methylated cytosines, defined as C with $P < 0.01$ in Fisher's exact test) and ≥ twofold change in DNA methylation levels were retained and combined if the gap size was < 100 bp to generate DMRs. The length of the DMR was adjusted to start from the first 5mC and end at the last 5mC.

Chop-PCR

For the chop-PCR assay, genomic DNA was digested with the methylation-sensitive restriction enzyme McrBC prior to PCR. DNA hypermethylation results in reduced levels of the PCR product. Undigested DNA was amplified as a control.

EMSA

GST-IDAP2 (WT), GST-IDAP2 (W19E), and GST-IDAP2 (W48E) recombinant fusion proteins were expressed in the *E. coli* BL21 (DE3) strain and purified using a GSTrap HP column (17-5131-01; GE Healthcare). The purified proteins were then digested by TEV protease. After digestion, GST-tagged wild-type IDAP2 and mutant IDAP2 were separated by Superdex 75 10/300GL (17-5174-01; GE Healthcare). The complementary oligonucleotides were annealed and labeled with γ-^{32}P-dATP using T4 polynucleotide kinase (M0201V; NEB). Labeled DNA oligonucleotides (100 fmol) were incubated with IDAP2 (WT), IDAP2 (W19E), or IDAP2 (W48E) in the presence or absence of cold competitors (3 or 5 pmol) in a 20 μl binding reaction buffer (20 mM HEPES (pH 7.6), 1 mM EDTA, 10 mM $(NH_4)_2SO_4$, 1 mM DTT, 0.2% Tween-20, 30 mM KCl, 50 ng/μl poly (dI-dC)). After incubation for 45 min at room temperature, 5 μl 5× loading buffer (0.25× TBE, 34% glycerol) was added to each reaction mixture which was then subjected to 6% PAGE with 0.5× TBE running buffer. The DNA–protein binding was detected with Typhoon FLA 7000. The binding of MBP-IDAP1 to DNA was detected in parallel as described above. The DNA probes used for EMSA are listed in Appendix Table S3.

ChIP assay

Chromatin immunoprecipitation assays were performed according to a published protocol (Saleh *et al*, 2008) with minor modifications. Briefly, flower tissues (about 1.5 g) were ground in liquid N_2 and suspended in 6 ml nuclear extraction buffer (20 mM Tris–HCl [pH 7.5], 20 mM KCl, 2 mM EDTA, 2.5 mM $MgCl_2$, 25% glycerol, 250 mM sucrose). The nuclei were fixed with 1% formaldehyde for 20 min at 4°C and then neutralized with 0.125 M glycine for 5 min. The nuclei were washed several times with NRBT buffer until it turned white, suspended in nuclear lysis buffer (50 mM Tris–HCl [pH 8.0], 10 mM EDTA, 1% SDS), and sonicated 5 times. After centrifugation, the chromatin supernatant was diluted 1:10 with dilution buffer (16.7 mM Tris–HCl [pH 8.0], 668 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100). The following beads or antibodies were used for ChIP assays: anti-Flag beads (M8823; Sigma), anti-acetylated H3K18 (ab1191; Abcam), anti-HA (H9658; Sigma), and anti-Myc (M4439; Sigma). The purified DNA was suspended in 50 μl ddH_2O and diluted 1:5, and a 1-μl aliquot was used for quantitative PCRs. The primers used for quantitative PCR are listed in Appendix Table S3.

Data availability

Primary data

We used whole-genome bisulfite sequencing data to analyze the genome-wide methylation status of the *idm1-1*, *idap1-1*, and *idap2-1* mutant plants. The dataset was deposited at NCBI (PRJNA307202). The WT and *mbd7-1* whole-genome bisulfite sequencing data were deposited at NCBI (Col-0:SRX747290, *mbd7-1*:SRX833671) (Li *et al*, 2015a).

Referenced data

ChIP-seq data of H3K4me2, H3K4me3, and H3K9me2 in *Arabidopsis*, published by Luo *et al* (2013), were retrieved from the GEO database, under the accession ID GSE28398. The *ros1-4* and *rdd* whole-genome bisulfite sequencing data were from GEO accession GSE33071 (Qian *et al*, 2012).

Expanded View for this article is available online.

Acknowledgements

We thank Dr. Jian-kang Zhu for the CRISPR/Cas9 system, the Mass Spectrometry Facility of the National Center for Protein Sciences at Peking University for the mass spectrometric analysis and Dr. Meng Rong for assistance with the data analyses. This work was supported by the National Natural Science Foundation of China (31571326 and 31522005) and the Recruitment Program of Global Youth Experts of China to WQ.

Author contributions

QL and WQ designed the study; QL, WY, XW, YL, XM, HS, JT, XT, and WQ performed the experiments; QL, WY, XW, and WQ analyzed the data; and YL and WQ wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Agius F, Kapoor A, Zhu JK (2006) Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc Natl Acad Sci USA* 103: 11796 – 11801

Bundock P, Hooykaas P (2005) An *Arabidopsis* hAT-like transposase is essential for plant development. *Nature* 436: 282 – 284

Cao X, Jacobsen SE (2002) Role of the *Arabidopsis DRM* methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* 12: 1138 – 1144

Casola C, Lawing AM, Betran E, Feschotte C (2007) PIF-like transposons are common in drosophila and have been repeatedly domesticated to generate new host genes. *Mol Bio Evol* 24: 1872 – 1888

Cavrak VV, Lettner N, Jamge S, Kosarewicz A, Bayer LM, Mittelsten Scheid O (2014) How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genet* 10: e1004115

Chen H, Zou Y, Shang Y, Lin H, Wang Y, Cai R, Tang X, Zhou JM (2008) Firefly luciferase complementation imaging assay for protein-protein interactions in plants. *Plant Physiol* 146: 368 – 376

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16: 735 – 743

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB transcription factors in *Arabidopsis*. *Trends Plant Sci* 15: 573 – 581

Feng Z, Zhang B, Ding W, Liu X, Yang DL, Wei P, Cao F, Zhu S, Zhang F, Mao Y, Zhu JK (2013) Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res* 23: 1229 – 1232

Fu Y, Kawabe A, Etcheverry M, Ito T, Toyoda A, Fujiyama A, Colot V, Tarutani Y, Kakutani T (2013) Mobilization of a plant transposon by expression of the transposon-encoded anti-silencing factor. *EMBO J* 32: 2407 – 2417

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL (2006) DEMETER DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation. *Cell* 124: 495 – 506

Gerrienne P, Servais T, Vecoli M (2016) Plant evolution and terrestrialization during Palaeozoic times-The phylogenetic context. *Rev Palaeobot Palynol* 227: 4 – 18

Gong Z, Morales-Ruiz T, Ariza RR, Roldan-Arjona T, David L, Zhu JK (2002) ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell* 111: 803 – 814

Hartwig B, James GV, Konrad K, Schneeberger K, Turck F (2012) Fast isogenic mapping-by-sequencing of ethyl methanesulfonate-induced mutant bulks. *Plant Physiol* 160: 591 – 600

Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J (2011) An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* 472: 115 – 119

Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) An active DNA transposon family in rice. *Nature* 421: 163 – 167

Kapitonov VV, Jurka J (1999) Molecular paleontology of transposable elements from *Arabidopsis thaliana*. *Genetica* 107: 27 – 37

Kapitonov VV, Jurka J (2004) Harbinger transposons and an ancient HARBI1 gene derived from a transposase. *DNA Cell Biol* 23: 311 – 324

Lang Z, Lei M, Wang X, Tang K, Miki D, Zhang H, Mangrauthia SK, Liu W, Nie W, Ma G, Yan J, Duan CG, Hsu CC, Wang C, Tao WA, Gong Z, Zhu JK (2015) The methyl-CpG-binding protein MBD7 facilitates active DNA demethylation to limit DNA hyper-methylation and transcriptional gene silencing. *Mol Cell* 57: 971 – 983

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11: 204 – 220

Lee J, Jang H, Shin H, Choi WL, Mok YG, Huh JH (2014) AP endonucleases process 5-methylcytosine excision intermediates during active DNA demethylation in *Arabidopsis*. *Nucleic Acids Res* 42: 11408 – 11418

Li X, Qian W, Zhao Y, Wang C, Shen J, Zhu JK, Gong Z (2012) Antisilencing role of the RNA-directed DNA methylation pathway and a histone acetyltransferase in *Arabidopsis*. *Proc Natl Acad Sci USA* 109: 11425 – 11430

Li Q, Wang X, Sun H, Zeng J, Cao Z, Li Y, Qian W (2015a) Regulation of Active DNA Demethylation by a Methyl-CpG-Binding Domain Protein in *Arabidopsis thaliana*. *PLoS Genet* 11: e1005210

Li Y, Cordoba-Canero D, Qian W, Zhu X, Tang K, Zhang H, Ariza RR, Roldan-Arjona T, Zhu JK (2015b) An AP endonuclease functions in active DNA demethylation and gene imprinting in *Arabidopsis*. *PLoS Genet* 11: e1004905

Li Y, Duan CG, Zhu X, Qian W, Zhu JK (2015c) A DNA ligase required for active DNA demethylation and genomic imprinting in *Arabidopsis*. *Cell Res* 25: 757 – 760

Li S, Liu L, Gao L, Zhao Y, Kim YJ, Chen X (2016) SUVH1, a Su(var)3-9 family member, promotes the expression of genes targeted by DNA methylation. *Nucleic Acids Res* 44: 608 – 620

Liang SC, Hartwig B, Perera P, Mora-Garcia S, de Leau E, Thornton H, de Alves FL, Rapsilber J, Yang S, James GV, Schneeberger K, Finnegan EJ, Turck F, Goodrich J (2015) Kicking against the PRCs – A Domesticated Transposase Antagonises Silencing Mediated by Polycomb Group Proteins and Is an Accessory Component of Polycomb Repressive Complex 2. *PLoS Genet* 11: e1005660

Lin R, Ding L, Casola C, Ripoll DR, Feschotte C, Wang H (2007) Transposase-derived transcription factors regulate light signaling in *Arabidopsis*. *Science* 318: 1302 – 1305

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133: 523 – 536

Luo C, Sidote DJ, Zhang Y, Kerstetter RA, Michael TP, Lam E (2013) Integrative analysis of chromatin states in *Arabidopsis* identified potential regulatory mechanisms for natural antisense transcript production. *Plant J* 73: 77 – 90

Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM (2015) Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet* 11: e1004915

Martinez-Macias MI, Qian W, Miki D, Pontes O, Liu Y, Tang K, Liu R, Morales-Ruiz T, Ariza RR, Roldan-Arjona T, Zhu JK (2012) A DNA 3′ phosphatase functions in active DNA demethylation in *Arabidopsis*. *Mol Cell* 45: 357 – 370

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet* 15: 394 – 408

Morales-Ruiz T, Ortega-Galisteo AP, Ponferrada-Marin MI, Martinez-Macias MI, Ariza RR, Roldan-Arjona T (2006) DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proc Natl Acad Sci USA* 103: 6853 – 6858

Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL (2007) DNA demethylation in the *Arabidopsis* genome. *Proc Natl Acad Sci USA* 104: 6752 – 6757

Qian W, Miki D, Zhang H, Liu Y, Zhang X, Tang K, Kan Y, La H, Li X, Li S, Zhu X, Shi X, Zhang K, Pontes O, Chen X, Liu R, Gong Z, Zhu JK (2012) A histone acetyltransferase regulates active DNA demethylation in *Arabidopsis*. *Science* 336: 1445 – 1448

Qian W, Miki D, Lei M, Zhu X, Zhang H, Liu Y, Li Y, Lang Z, Wang J, Tang K, Liu R, Zhu JK (2014) Regulation of active DNA demethylation by an α-crystallin domain protein in *Arabidopsis*. *Mol Cell* 55: 361 – 371

Saleh A, Alvarez-Venegas R, Avramova Z (2008) An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in *Arabidopsis* plants. *Nat Protoc* 3: 1018 – 1025

Sanchez DH, Paszkowski J (2014) Heat-induced release of epigenetic silencing reveals the concealed role of an imprinted plant gene. *PLoS Genet* 10: e1004806

Shamloul M, Trusa J, Mett V, Yusibov V (2014) Optimization and utilization of *Agrobacterium*-mediated transient protein production in *Nicotiana*. *J Vis Exp* 86: e51204

Sinzelle L, Kapitonov VV, Grzela DP, Jursch T, Jurka J, Izsvak Z, Ivics Z (2008) Transposition of a reconstructed Harbinger element in human cells and functional homology with two transposon-derived cellular genes. *Proc Natl Acad Sci USA* 105: 4715 – 4720

Sinzelle L, Izsvak Z, Ivics Z (2009) Molecular domestication of transposable elements: from detrimental parasites to useful host genes. *Cell Mol Life Sci* 66: 1073 – 1093

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE (2014) Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat Struct Mol Biol* 21: 64 – 72

Vaillant I, Paszkowski J (2007) Role of histone and DNA methylation in gene regulation. *Curr Opin Plant Biol* 10: 528 – 533

Walker EL, Eggleston WB, Demopulos D, Kermicle J, Dellaporta SL (1997) Insertions of a novel class of transposable elements with a strong target site preference at the r locus of maize. *Genetics* 146: 681 – 693

Walter M, Chaban C, Schutze K, Batistic O, Weckermann K, Nake C, Blazevic D, Grefen C, Schumacher K, Oecking C, Harter K, Kudla J (2004) Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J* 40: 428 – 438

Wang C, Dong X, Jin D, Zhao Y, Xie S, Li X, He X, Lang Z, Lai J, Zhu JK, Gong Z (2015) Methyl-CpG-binding domain protein MBD7 is required for active DNA demethylation in *Arabidopsis*. *Plant Physiol* 167: 905 – 914

Wu SC, Zhang Y (2010) Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* 11: 607 – 620

Wu Y, Zhou H, Fan X, Zhang Y, Zhang M, Wang Y, Xie Z, Bai M, Yin Q, Liang D, Tang W, Liao J, Zhou C, Liu W, Zhu P, Guo H, Pan H, Wu C, Shi H, Wu L *et al* (2015) Correction of a genetic disease by CRISPR-Cas9-mediated gene editing in mouse spermatogonial stem cells. *Cell Res* 25: 67 – 79

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D (2013) The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153: 193 – 205

Zhang X, Feschotte C, Zhang Q, Jiang N, Eggleston WB, Wessler SR (2001) P instability factor: an active maize transposon system associated with the amplification of Tourist-like MITEs and a new superfamily of transposases. *Proc Natl Acad Sci USA* 98: 12572 – 12577

Zhang X, Jiang N, Feschotte C, Wessler SR (2004) PIF- and Pong-like transposable elements: distribution, evolution and relationship with Tourist-like miniature inverted-repeat transposable elements. *Genetics* 166: 971 – 986

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* 126: 1189 – 1201

Zhang H, Zhu JK (2011) RNA-directed DNA methylation. *Curr Opin Plant Biol* 14: 142 – 147

Zhang H, Ma ZY, Zeng L, Tanaka K, Zhang CJ, Ma J, Bai G, Wang P, Zhang SW, Liu ZW, Cai T, Tang K, Liu R, Shi X, He XJ, Zhu JK (2013) DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. *Proc Natl Acad Sci USA* 110: 8290 – 8295

Zhao Y, Xie S, Li X, Wang C, Chen Z, Lai J, Gong Z (2014) REPRESSOR OF SILENCING5 Encodes a Member of the Small Heat Shock Protein Family and Is Required for DNA Demethylation in *Arabidopsis*. *Plant Cell* 26: 2660 – 2675

Zheng X, Pontes O, Zhu J, Miki D, Zhang F, Li WX, Iida K, Kapoor A, Pikaard CS, Zhu JK (2008) ROS3 is an RNA-binding protein required for DNA demethylation in *Arabidopsis*. *Nature* 455: 1259 – 1262

Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, Chory J, Wohlschlegel JA, Patel DJ, Jacobsen SE (2014) Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell* 157: 1050 – 1060

Zhu JK (2009) Active DNA demethylation mediated by DNA glycosylases. *Annu Rev Genet* 43: 143 – 166