

ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation

Andrea D McCue, Kaushik Panda, Saivageethi Nuthikattu, Sarah G Choudury, Erica N Thomas & R Keith Slotkin*

Abstract

Transposable elements (TEs) generate mutations and chromosomal instability when active. To repress TE activity, eukaryotic cells evolved mechanisms to both degrade TE mRNAs into small interfering RNAs (siRNAs) and modify TE chromatin to epigenetically inhibit transcription. Since the populations of small RNAs that participate in TE post-transcriptional regulation differ from those that establish RNA-directed DNA methylation (RdDM), the mechanism through which transcriptionally active TEs transition from post-transcriptional RNAi regulation to chromatin level control has remained unclear. We have identified the molecular mechanism of a plant pathway that functions to direct DNA methylation to transcriptionally active TEs. We demonstrated that 21–22 nucleotide (nt) siRNA degradation products from the RNAi of TE mRNAs are directly incorporated into the ARGONAUTE 6 (AGO6) protein and direct AGO6 to TE chromatin to guide its function in RdDM. We find that this pathway functions in reproductive precursor cells to primarily target long centromeric high-copy transcriptionally active TEs for RdDM prior to gametogenesis. This study provides a direct mechanism that bridges the gap between the post-transcriptional regulation of TEs and the establishment of TE epigenetic silencing.

Keywords Argonaute; epigenetics; RNA-directed DNA methylation; small interfering RNA; transposable element

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

DOI 10.15252/emj.201489499 | Received 10 July 2014 | Revised 30 September 2014 | Accepted 6 October 2014 | Published online 11 November 2014

The EMBO Journal (2015) 34: 20–35

Introduction

The mobilization of transposable elements (TEs) results in chromosome instability and mutation. In order to maintain genome integrity, fungi, plants and animals modify TE chromatin to

epigenetically repress the production of TE mRNAs to inhibit TE mobilization (Girard & Hannon, 2008). In mammals and flowering plants, cytosine DNA methylation is critical for inhibiting TE activity (Zemach & Zilberman, 2010). Once established at TEs, DNA methylation in the symmetrical CG context is propagated by the DNMT1 family of CG methyltransferases (MET1 in plants) (Law & Jacobsen, 2010). In plants, non-CG context DNA methylation (CHG or CHH, where H = A, T or C) is also epigenetically maintained via recruitment of the CMT3 and CMT2 DNA methyltransferases through their interaction with repressive histone modifications (Stroud *et al.*, 2014). Thus, once established, robust mechanisms exist to propagate TE DNA methylation, resulting in the epigenetic transcriptional silencing of TEs (Saze *et al.*, 2003). However, the mechanism of how DNA methylation and epigenetic silencing are originally targeted to TEs is not known.

In contrast to the maintenance of methylation, the methylation of previously unmethylated cytosines (*de novo* methylation) occurs through the DNMT3 family of DNA methyltransferases (DRM2 in plants), which methylate cytosines in any sequence context (Matzke & Mosher, 2014). In plants and mammals, *de novo* TE DNA methylation is known to be targeted through an RNA-directed DNA methylation (RdDM) pathway, which utilizes TE small interfering RNAs (siRNAs) in plants and piRNAs in animals to guide ARGONAUTE (AGO) family proteins to TE chromatin (Castel & Martienssen, 2013). Similar pathways of small RNA-mediated chromatin modification exist in fission yeast, *Drosophila* and *C. elegans*, which have evolutionarily lost cytosine DNA methylation (Grewal, 2010; Zemach & Zilberman, 2010; Castel & Martienssen, 2013). In each of these models, a small RNA-targeted AGO family protein is recruited to the TE locus via a scaffolding RNA that functionally tethers the AGO protein to the target chromatin (Castel & Martienssen, 2013). The recruitment of the AGO protein to the TE locus then initiates a poorly understood cascade of chromatin modification, which includes *de novo* DNA methylation in plants and mammals.

In the reference plant *Arabidopsis thaliana*, the well-studied RdDM pathway begins with transcription of a noncoding TE RNA by RNA Polymerase IV (Pol IV). Pol IV is a plant-specific specialized derivative of RNA Polymerase II (Pol II) that utilizes multiple shared

Pol II subunits (Haag & Pikaard, 2011). Although the recruitment of Pol IV is not fully understood, transcriptionally repressive histone modifications have been shown to guide Pol IV recruitment to previously silenced TEs (Law *et al.*, 2013). Once produced, the Pol IV-derived transcript is converted into double-stranded RNA by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) and cleaved into distinctly sized 24 nucleotide (nt) siRNAs by DICER-LIKE 3 (DCL3). These 24nt siRNAs have been a long-standing hallmark of plant RdDM activity. The 24nt siRNAs are incorporated into the AGO4 and AGO6 proteins and guide these proteins to TE loci through their interaction with a TE scaffolding transcript produced from RNA Polymerase V (Pol V) (another plant-specific specialized derivative of Pol II) (Haag & Pikaard, 2011). This Pol IV-dependent RdDM pathway (Pol IV-RdDM) utilizes 24nt siRNAs to constantly re-target methylation to previously silenced TEs, particularly small euchromatic TEs located near genes in order to maintain chromatin boundaries (Zemach *et al.*, 2013), and can function in *trans* to silence homologous TE loci (Nuthikattu *et al.*, 2013).

In addition to larger small RNAs such as 24nt siRNAs in plants or 26–31nt piRNAs in animals, 21–22nt endogenous siRNAs (endo-siRNAs) have been linked to the establishment of heterochromatic states in both plants and animals (Morris *et al.*, 2004; Herr *et al.*, 2005; Kim *et al.*, 2006; Chen *et al.*, 2012; Pushpavalli *et al.*, 2012; White *et al.*, 2014). However, the mechanism responsible for chromatin modification has not been elucidated and remains controversial. Recently, several publications by independent groups have uncovered a genetic pathway of RdDM in *Arabidopsis* that utilizes 21–22nt endo-siRNAs produced from RNAi and acts independently of Pol IV, RDR2, DCL3 and 24nt siRNAs. Initially, this pathway was found to methylate *trans*-acting siRNA (TAS)-generating loci, which are regions of noncoding transcripts that are processed into 21–22nt gene-regulating *trans*-acting siRNAs (tasiRNAs) (Wu *et al.*, 2012; Kanno *et al.*, 2013). This pathway is also responsible for the methylation of transcriptionally active TEs, whose Pol II-derived mRNAs are degraded into 21–22nt siRNAs by the combined activities of RDR6, DCL2, DCL4 and AGO1 (McCue *et al.*, 2012; Pontier *et al.*, 2012; Nuthikattu *et al.*, 2013). For both the TAS and TE targets, this Pol IV- and DCL3-independent RdDM activity is dependent on RDR6; thus, this pathway is referred to as RDR6-RdDM to differentiate it from Pol IV-RdDM (Nuthikattu *et al.*, 2013). RDR6-RdDM functions specifically in the Pol II expression-dependent initiation and reestablishment of proper TE methylation levels, but not in the maintenance methylation of epigenetically silenced TEs (Nuthikattu *et al.*, 2013). Although the RDR6-RdDM pathway has been previously genetically characterized, how 21–22nt endo-siRNAs are utilized to direct DNA methylation remained unclear.

We investigated the molecular mechanism of RDR6-RdDM with particular focus on the *Athila* family of LTR retrotransposons (the largest TE family in the *Arabidopsis* genome) in a mutant background where RdDM can be visualized without the confounding obstruction of the maintenance methylation pathways. We have shown that 21–22nt endo-siRNAs are directly incorporated into the AGO6 protein, and we demonstrate that these siRNAs are sufficient to guide AGO6 to its chromatin targets to establish TE expression-dependent DNA methylation. Our findings represent a new mechanism for AGO6 in the RdDM of transcriptionally active TEs, but this function typically remains latent and has largely been overlooked due to its activity specifically in the reproductive tissue precursor

cells and the near complete epigenetic transcriptional silencing of TEs in wild-type *Arabidopsis*. These results demonstrate that the well-characterized Pol IV-RdDM pathway represents only part of the total function of RdDM.

Results

A system to detect only *de novo* TE DNA methylation

In the inflorescence (floral bud) of wild-type Columbia reference strain *Arabidopsis* (wt Col), which gives rise to germ cell differentiation, TEs are transcriptionally silenced and have dense symmetrical (CG and CHG) DNA methylation (Lister *et al.*, 2008), including at the transcriptional start site (TSS) region of the *Athila6A* subfamily of LTR retrotransposons (Fig 1A) (Nuthikattu *et al.*, 2013). This symmetrical methylation is required to epigenetically propagate the silenced TE state from generation to generation (Saze *et al.*, 2003). In contrast to the levels of symmetrical DNA methylation, the *Athila6A* TSS asymmetrical CHH methylation level is low (15.9%) (Fig 1A) (Nuthikattu *et al.*, 2013). In the wt Col epigenome when TEs are transcriptionally silenced, we find the CHH methylation of the *Athila6A* TSS to be largely independent of the siRNA-targeted *de novo* DNA methyltransferase DRM2 (Fig 1A), even though this region produces Pol IV-dependent 24nt siRNAs (Fig 1B). Instead, CHH methylation at this transcriptionally silenced region is primarily maintained at low levels by the maintenance methyltransferase CMT2 (Fig 1A). Thus, in spite of 24nt siRNA production, RdDM does not continually target methylation of this TE region.

To determine the contribution of RdDM when TEs are transcriptionally active, we utilized the global TE transcriptional reactivation found in plants that lack a functional DDM1 protein. DDM1 is a SWI/SNF family ATPase that coordinates linker histone chromatin compaction (Zemach *et al.*, 2013). In *ddm1* mutants, global maintenance of methylation fails, and TEs undergo chromatin decondensation that leads to a genome-wide transcriptional activation of TEs (Lippman *et al.*, 2004; Zemach *et al.*, 2013). When transcriptionally activated, the *Athila6A* TSS has increased CHH methylation levels (Fig 1A) (Nuthikattu *et al.*, 2013). In contrast to when silenced in wt Col, when the TE is transcriptionally active in *ddm1* plants, all the CHH methylation (and nearly all total methylation) is dependent on the RdDM-targeted DRM2 (Fig 1A). Therefore, at least at the *Athila6A* TSS, RdDM only functions when the TE is expressed. In wt Col plants, genome-wide maintenance and *de novo* DNA methylation via RdDM occur simultaneously at TEs, obscuring the molecular dissection of either pathway. However, since *ddm1* mutant plants fail to maintain TE DNA methylation, these mutants can be used to obtain an unobstructed view of the *de novo* DNA methylation pathways.

21–22nt siRNAs and AGO6 specifically function in RDR6-RdDM

We previously demonstrated that in *ddm1* mutant plants, high levels of 21–22nt endo-siRNAs are produced, and we have previously genetically implicated the production of these siRNAs and RDR6 in RdDM (McCue *et al.*, 2012; Nuthikattu *et al.*, 2013). To distinguish the effects of Pol IV-RdDM from RDR6-RdDM, we utilized a *ddm1 dcl3* double mutant and have now deep sequenced its small RNAs to

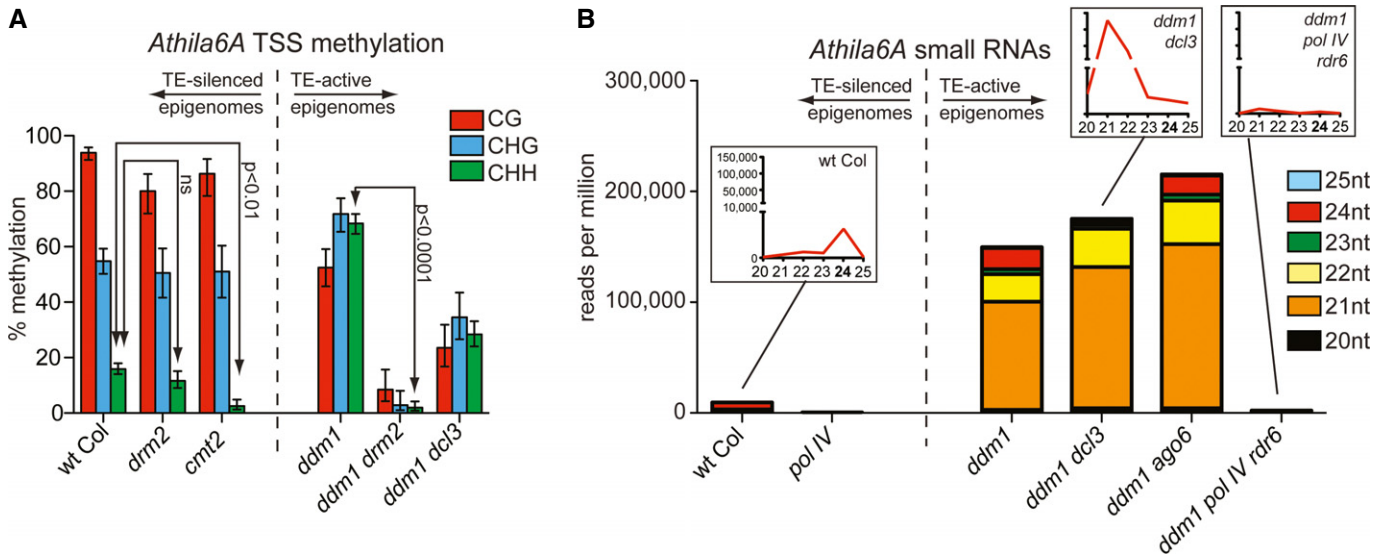


Figure 1. A system to detect *de novo* TE DNA methylation guided by 21–22nt endo-siRNAs.

A Bisulfite DNA methylation sequencing of the *Athila6A* TE transcriptional start site (TSS) in genotypes with silenced TEs (left of the dashed line) or transcriptionally active TEs (right of the dashed line). Methylation in the CG (red), CHG (blue) and CHH (green) sequence contexts are shown (H = C, T, or A). Error bars represent Wilson score interval 95% confidence limits. For multiple statistical tests, one-way ANOVA was performed using Tukey's multiple comparison test. ns, not significant.

B The reads-per-million-(RPM)-normalized accumulation of *Athila6A* TE siRNAs from TE-silenced epigenomes (left) and *ddm1* TE-active epigenomes (right). Inset line graphs reflecting the same values are shown for select genotypes. All line graphs have the same y-axis scale and are shown for increased resolution in genotypes with low siRNA levels.

determine that in multiple biological replicates, the *ddm1 dcl3* double mutant fails to produce *Athila6A* 24nt siRNAs above background levels, but still retains 28.3% CHH methylation (Fig 1A and B; Supplementary Fig S1A). We define deep sequencing background levels of 24nt siRNAs as equal to the number of 23 and 25nt siRNAs, as there is no known mechanism to generate these sized siRNAs, and they represent *in vivo* processing errors or *in vitro* sequencing artifacts. Using the high level of *de novo* CHH methylation in *ddm1* mutants, we determine that in addition to the known 24nt siRNAs that target RdDM, the 21–22nt siRNAs remaining in the *ddm1 dcl3* double mutant are also partially responsible for targeting RdDM. Thus, the 24nt and 21–22nt siRNAs function additively to produce the high levels of CHH methylation in *ddm1* mutants. We additionally deep sequenced the *ddm1 pol IV rdr6* triple mutant small RNAs and found that this mutant does not accumulate any size of *Athila6A* siRNAs (Fig 1B; Supplementary Fig S1A), while we had previously determined that this triple mutant has only background levels of *Athila6A* TSS methylation (Nuthikattu *et al*, 2013), demonstrating that when transcriptionally active all methylation of the *Athila6A* TSS is targeted by RdDM. Lastly, we know that the 21–22nt siRNAs that drive RdDM are not produced by Pol IV, as 21.0% CHH methylation at the *Athila6A* TSS accumulates in *ddm1 pol IV* double mutants (Nuthikattu *et al*, 2013). In addition, previous work demonstrated that Pol IV-derived transcripts are not processed into siRNAs in the absence of RDR2 (Kasschau *et al*, 2007), while *ddm1 rdr2* double mutants retain similar levels of *Athila6A* TSS methylation as *ddm1 dcl3* double mutants (Nuthikattu *et al*, 2013). Therefore, by starting with a series of mutant combinations and their known effects on siRNA accumulation and DNA methylation, we can use these mutants in direct assays to dissect the molecular mechanism of RDR6-RdDM.

RDR6-RdDM is mediated through AGO6

To determine the mechanism of TE RDR6-RdDM, we focused on identifying the AGO family effector protein. We previously genetically implicated AGO6 in this process, as *ddm1 ago6* double mutants lose nearly all *Athila6A* TSS CHH *de novo* methylation (Nuthikattu *et al*, 2013). Our deep sequencing of small RNAs shows that the accumulation of *Athila6A* siRNAs is not perturbed in *ddm1 ago6* double mutants (Fig 1B; Supplementary Fig S1A), demonstrating that AGO6 functions downstream of siRNA production in RDR6-RdDM (unlike AGO1 (McCue *et al*, 2012)). In Supplementary Fig S1B and C, we provide further genetic evidence that AGO6 is required for Pol IV-RdDM (which has been previously described (Zheng *et al*, 2007; Havecker *et al*, 2010; Eun *et al*, 2011)) and additionally plays a key role in RDR6-RdDM expression-dependent TE methylation and the corrective reestablishment of active TE silencing (Supplementary Fig S1B and C). In addition to AGO6, we also investigated the candidates AGO4 and AGO2 for their roles in RDR6-RdDM. We find that AGO4 only functions in Pol IV-RdDM to the same extent as Pol IV, RDR2 or DCL3 and has no function in RDR6-RdDM (Supplementary Fig S1D). In addition, AGO2 has been specifically tested and plays no role in the expression-dependent methylation of TAS loci, which are targeted by RDR6-RdDM (Wu *et al*, 2012). AGO2's known role in methylating *Arabidopsis* intergenic regions is fully dependent on Pol IV (Pontier *et al*, 2012), demonstrating that it operates in a maintenance of silencing pathway distinct from the Pol IV-independent function of RDR6-RdDM. From this combined genetic data, we determine that AGO6 is the critical effector protein of the expression-dependent RDR6-RdDM pathway.

AGO6 incorporation of 21–22nt siRNAs guides RDR6-RdDM

To determine how AGO6 functions in RDR6-RdDM, we aimed to immunoprecipitate (IP) the AGO6 protein both in the wt Col TE-silenced epigenome without active TE RDR6-RdDM and in the *ddm1* TE-active epigenome with ongoing TE RDR6-RdDM. We utilized a previously generated FLAG epitope-tagged AGO6 protein under the control of its native promoter (Havecker *et al*, 2010). We determined that this tagged protein functionally complements the loss of RDR6-RdDM in *ago6* mutants (Supplementary Fig S2A). We next determined that the successful IP of FLAG-AGO6 is dependent on the presence of the FLAG antigen, while other potentially contaminating proteins (such as AGO1, which is known to incorporate TE 21–22nt siRNAs (McCue *et al*, 2013)) are not immunoprecipitated (Supplementary Fig S2B).

We immunoprecipitated the FLAG-AGO6 protein from *ago6* and *ddm1 ago6* mutants, as these complemented mutations functionally represent the wt Col TE-silenced and *ddm1* TE-transcriptionally active epigenomes, respectively, and then deep sequenced the small RNAs isolated from these IPs. Previous publications have directly examined the small RNAs associated with AGO protein IPs; however, we determined that some background small RNAs were consistently contaminating all FLAG-IPs, mock-IPs or no-antigen control IPs in experiments from our laboratory as well as other published data (data not shown). In order to account for these contaminating small RNAs, we also sequenced small RNAs from FLAG antibody IPs performed in plants with corresponding genotypes lacking the FLAG-AGO6 transgene (no-antigen controls) and then calculated the relative enrichment of FLAG-AGO6-IP small RNAs. As a negative control for our methodology, we observe no enrichment of highly abundant microRNAs in FLAG-AGO6 (Fig 2A). Previous reports have shown low levels of microRNAs in FLAG-AGO6-IPs (Havecker *et al*, 2010); however, by calculating the level of enrichment using the no-antigen IP, we can determine that these highly abundant microRNAs are not specifically enriched in AGO6. As a positive control for FLAG-AGO6-IP enrichment, we investigated the non-autonomous TE *SimpleHat2*, as AGO6 was previously shown to be required for the accumulation and binding of *SimpleHat2* 24nt siRNAs, and for Pol IV-RdDM of *SimpleHat2* (Zheng *et al*, 2007; Havecker *et al*, 2010). In wt Col, we observe enrichment of specifically 24nt siRNAs in FLAG-AGO6 for the *SimpleHat2* TE, demonstrating that we have successfully enriched known and functional AGO6-incorporated siRNAs (Fig 2B). Next, we investigated

TAS3a-derived siRNAs, as AGO6 has been genetically implicated in methylating this locus, and Wu *et al* theorized that AGO6 is directed to methylate TAS3a via incorporation of 21–22nt siRNAs produced from the TAS3a non-protein-coding RNA (Wu *et al*, 2012). Wu *et al* detected low levels (288 reads per million (RPM)) of 21nt TAS3a-derived siRNAs present in the wt Col FLAG-AGO6-IP performed by Havecker *et al* (2010). However, they also found 21nt tasiRNAs present in four other AGO proteins, providing an unclear connection between siRNA incorporation into AGO proteins and their role in RdDM. Additionally, it was unclear from this analysis whether these siRNA RPM values represent enrichment above background level, as no background or mock-IP sequencing control was performed. Therefore, the role of AGO6 (and AGO4) direct incorporation of 21–22nt TAS siRNAs and function in RdDM was not directly demonstrated and was only speculated based on correlation. Here, we show that FLAG-AGO6 is specifically enriched for 21nt siRNAs (but not 24nt siRNAs) from the TAS3a locus (Fig 2C) and drives all TAS3a CHH methylation via RdDM in inflorescence tissue (Supplementary Fig S2A). We conclude from these data that AGO6 can incorporate siRNA size classes other than 24nt to function in RdDM.

Next, we investigated the siRNAs enriched in FLAG-AGO6 produced from *Athila6A*, as *Athila6A* is a known target of RDR6-RdDM. In the wt Col TE-silenced epigenome, *Athila6A* produces primarily 24nt siRNAs (Fig 1B), and we detect AGO6 enrichment of these siRNAs for the *Athila6A* LTR, *gag/pol* protein-coding region, and in the degenerate *env*-coding region (Fig 2D). In the *ddm1* TE transcriptionally active epigenome, *Athila6A* 21–22nt siRNAs accumulate (Fig 1B), and we find significant FLAG-AGO6 enrichment of 22nt siRNAs from the *Athila6A* LTR, in addition to enrichment of 21–22nt siRNAs from the 3' half of the *gag/pol* protein-coding region and from the intergenic region (IR) between *gag/pol* and *env* (Fig 2E). Interestingly, we find that in *ddm1*, FLAG-AGO6-enriched small RNAs do not simply correlate with total 21–22nt siRNA abundance, which is highest in the *env* and 3' non-protein-coding region (Fig 2E–G). To determine whether the AGO6 enrichment of siRNAs from particular regions of *Athila6A* drives specificity in RdDM targeting, we performed bisulfite DNA sequencing of the *Athila6A* IR promoter and the 3' portion of the *env*-coding region. We found that regions that display enrichment of 21–22nt siRNAs in AGO6 (the IR promoter and LTR TSS) undergo expression-dependent RDR6-RdDM, while a region with low 21–22nt siRNA enrichment (the 3' region of *env*) does not (Figs 1A and 2H).

Figure 2. AGO6 incorporates 21–22nt siRNAs.

AGO6 relative enrichment of small RNAs (solid lines) and total small RNAs (dashed lines) are shown for each small RNA size class (21, 22 and 24nt) at four genomic loci.

- A A representative microRNA, miR158B. The location of miR158B from this locus is shown in orange.
- B The *SimpleHat2* non-autonomous TE targeted by Pol IV-RdDM.
- C The RDR6-RdDM target locus TAS3a.
- D The *Athila6A* consensus element in the wt Col epigenome.
- E The same *Athila6A* consensus element as in (D) upon transcriptional activation in the *ddm1* epigenome.
- F The total small RNAs produced in the wt Col TE-silenced epigenome.
- G The total small RNAs in *ddm1* mutants.
- H Bisulfite sequencing of different regions of the *Athila6A* element. Error bars represent Wilson score interval 95% confidence limits. Differences in the methylation status were analyzed with a two-tailed unpaired Student's *t*-test.
- I A second biological replicate of the FLAG-AGO6-IP and no-antigen control followed by deep sequencing shown for the *Athila6A* consensus element in *ddm1* mutants. The transcriptional start sites of the LTR TSS- and IR-driven transcripts are shown.
- J Analysis of the first base of AGO6-enriched siRNAs from the wt Col TE-silenced epigenome (left) and the *ddm1* TE-transcriptionally active epigenome (right).

Data information: Maps of the *Athila6A* element shown below panels (F) and (G) align with AGO6 enrichment shown in (D) and (E).

The AGO6 enrichment of *Athila6A* 21–22nt siRNAs in the *ddm1* TE transcriptionally active epigenome was verified in a biological replicate of the FLAG-AGO6-IP (Fig 2I). Additionally, we demonstrated that our detection of the 21–22nt *Athila6A* siRNAs is not a result of FLAG-AGO6 binding these siRNAs after cell lysis

(Supplementary Fig S2C). Lastly, we have verified these AGO6 small RNA enrichment patterns using a different antibody that recognizes the native AGO6 protein. We confirmed previous findings that this antibody detects the AGO6 protein, but also detects other unknown proteins (Supplementary Fig S3A), and for this reason, this antibody

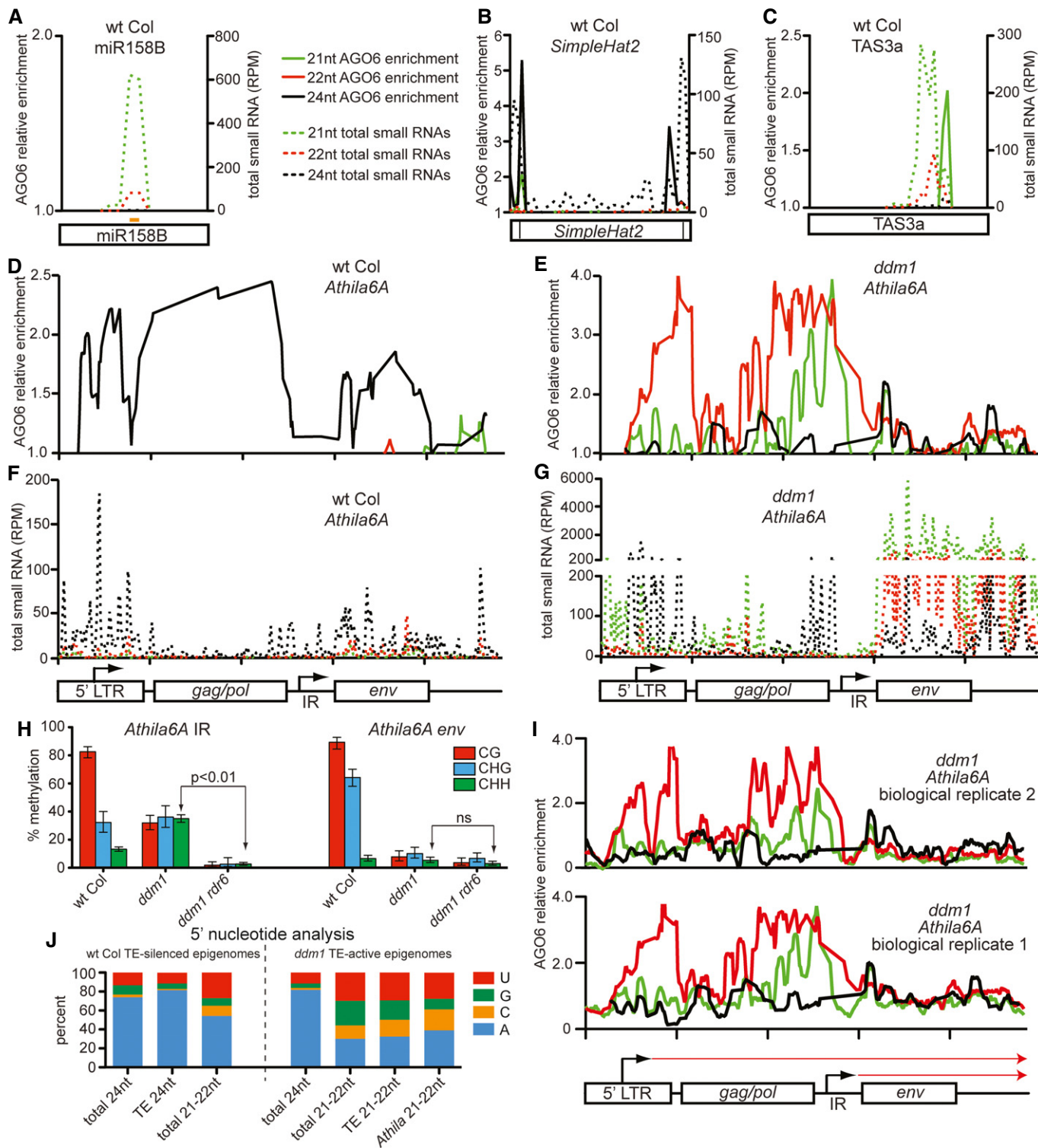


Figure 2.

was not previously used for IPs of small RNAs (Havecker *et al*, 2010). In Supplementary Fig S3B–F, we demonstrate that the AGO6 native antibody-specific enrichment of small RNAs is similar to the FLAG-AGO6-IP: 21–22nt siRNAs for TAS3a in wt Col and for *Athila6A* in *ddm1* are enriched in AGO6. Therefore, by using multiple controls, biological replicates and two independent antibodies to IP the AGO6 protein, we have shown that the expression-dependent production of 21–22nt *Athila6A* and TAS3a endo-siRNAs results in their incorporation into AGO6 protein complexes and their function in RDR6-RdDM. These data demonstrate that AGO6 incorporation of 21–22nt siRNAs is a requirement for RDR6-RdDM function.

We investigated why particular small RNAs are incorporated into AGO6 while others are not. First, we determined the 5' nucleotide of the siRNAs enriched in AGO6, as AGO proteins show preferences for particular 5' nucleotides, and this is thought to play an important role in sorting of small RNAs into particular AGO complexes (Mi *et al*, 2008). In wt Col, we observe the same 5' adenosine bias in AGO6 for both total and TE 24nt siRNAs as Havecker *et al* (2010) observed; however, this bias is not complete as siRNAs with all four different 5' nucleotides are highly enriched in AGO6 (Fig 2J). In addition, the 21–22nt siRNAs present in AGO6 in wt Col (mostly tasiRNAs) also have a 5' adenosine bias, although it is not as strong as the 24nt siRNA bias. In *ddm1* mutant plants, the 24nt siRNAs show the same 5' adenosine bias, while the 21–22nt total, TE and *Athila* siRNA pools display a less biased accumulation (Fig 2J). These data suggest that while AGO6 demonstrates an overall bias, it can accumulate 21, 22 or 24nt siRNAs with any 5' nucleotide, particularly when it is functioning in expression-dependent RDR6-RdDM. Second, we investigated why some regions of *Athila6A* display AGO6 enrichment of 21–22nt siRNAs in *ddm1* while others do not. We find that the *env* and 3' regions do not enrich 21–22nt siRNAs in AGO6, and this coincides with the location of the IR promoter, which drives *env* expression (shown on Fig 2I). Our data suggest that the transcript generated by the 5' LTR promoter is degraded, and the resulting 21–22nt siRNAs from at least the LTR, *gag/pol* and IR regions are incorporated into AGO6. On the other hand, the transcript generated by the IR promoter generates high levels of 21–22nt siRNAs (Fig 2G) that are not enriched in AGO6 and thus reduce the overall enrichment values of this region (seen in Fig 2I). Using previously published AGO1-IPs from the same tissue of *ddm1* plants (McCue *et al*, 2013), we observe that AGO6 and AGO1 associate with different populations of *Athila6A* siRNAs that potentially come from the different transcripts (Supplementary Fig S2D). AGO6 incorporates 21–22nt siRNAs from the 5' LTR-driven transcript, while AGO1 incorporates 21–22nt siRNAs derived from the IR-driven transcript. This suggests that AGO6 and AGO1 may not compete for the exact same siRNAs. Rather, particular transcripts may feed into these specific pathways destined for either post-transcriptional silencing (via AGO1) or RdDM (via AGO6). How particular transcripts are differentially fed into these different pathways leading to AGO6 loading and enrichment of particular siRNAs is currently unknown.

21–22nt siRNAs direct AGO6 to RdDM target chromatin

To determine whether the incorporation of 21–22nt siRNAs into AGO6 is sufficient to direct AGO6 to its chromatin targets, we performed chromatin immunoprecipitation (ChIP) of the AGO6

protein in plants that produce TE 21–22nt siRNAs. ChIP requires formaldehyde crosslinking to capture *in vivo* protein/DNA interactions. However, AGO6 is not a DNA-binding protein, and it is only tethered to chromatin by the base pairing interaction between its incorporated siRNA and a Pol V-derived scaffolding transcript (see Fig 3A). Therefore, the fold enrichment of AGO6 in ChIP experiments is much closer to background and must be carefully controlled (Wierzbicki *et al*, 2009). We performed ChIP for the AGO6 protein and acetylated histone H3 (H3Ac, a transcriptionally active chromatin mark) as a control for wt Col and *ddm1* biological replicates (Fig 3B). As additional controls, we also determined the level of AGO6 association with the constitutively expressed gene At1g08200 and the *SimpleHat2* TE in wt Col and a series of mutants including *ago6* (no-antigen control), *pol V* (no scaffolding transcript) and *dcl3* (no 24nt siRNAs). We determined that AGO6 is not associated with the genic At1g08200 locus in wt Col, while it is associated with the *SimpleHat2* TE (Fig 3C). The RdDM and accumulation of *SimpleHat2* 24nt siRNAs is dependent on AGO6 (Zheng *et al*, 2007), and we have verified that 24nt siRNAs from *SimpleHat2* are enriched in AGO6 (Fig 2B) (Havecker *et al*, 2010). We find that the association of AGO6 with *SimpleHat2* is dependent on the presence of the AGO6 antigen, Pol V-derived scaffolding transcript and DCL3 to produce 24nt siRNAs (Fig 3C), supporting the Pol IV-RdDM model of *SimpleHat2* silencing.

We next determined the association of AGO6 with the *Athila6A* TSS. We found that in wt Col, AGO6 is not significantly enriched at the *Athila6A* TSS compared to the *ago6* no-antigen control (Fig 3D), which is supported by the observation that the *Athila6A* TSS is not targeted by RdDM when it is transcriptionally silent (Fig 1A). In contrast, when *Athila6A* is transcriptionally active, we detect an AGO6 association with the *Athila6A* TSS ($P < 0.05$) (Fig 3D). We find that the recruitment of AGO6 to this locus is TE expression-dependent and requires the presence of a Pol V scaffolding transcript (see *ddm1 pol V*), as well as the production of siRNAs. In *ddm1* mutants, both Pol IV-RdDM (via 24nt siRNAs) and RDR6-RdDM (via 21–22nt siRNAs) target the *Athila6A* TSS for *de novo* methylation (Nuthikattu *et al*, 2013), and correspondingly, we observe 24nt-driven AGO6 enrichment at the *Athila6A* TSS in a *ddm1 rdr6* mutant that lacks 21–22nt siRNAs (Fig 3D). In the *ddm1 pol IV rdr6* triple mutant, siRNA production of all sizes is abolished (Fig 1B), and we find that AGO6 is not recruited to *Athila6A* chromatin (Fig 3D). Importantly, when the *Athila6A* TE is transcriptionally active but 24nt siRNAs are absent (see *ddm1 dcl3*, Fig 1B; Supplementary Fig S1A), AGO6 is still directed to the *Athila6A* TSS (Fig 3D). This demonstrates that 21–22nt siRNAs are sufficient to direct AGO6 to its chromatin targets, where we have shown that AGO6 is required for RdDM (Supplementary Fig S1D) (Nuthikattu *et al*, 2013). The 21–22nt siRNA direction of AGO6 to chromatin targets can also be observed at the TAS3a locus, where AGO6 association is not dependent on 24nt siRNAs (see *dcl3*, Fig 3E). Therefore, we conclude that 21–22nt siRNAs direct AGO6 to its chromatin targets to establish DNA methylation.

Pol V presence at transcriptionally active TEs

For both Pol IV- and RDR6-RdDM to take place, Pol V function must be present (Wierzbicki *et al*, 2009; Nuthikattu *et al*, 2013), and presumably, a scaffold transcript generated by Pol V must be present

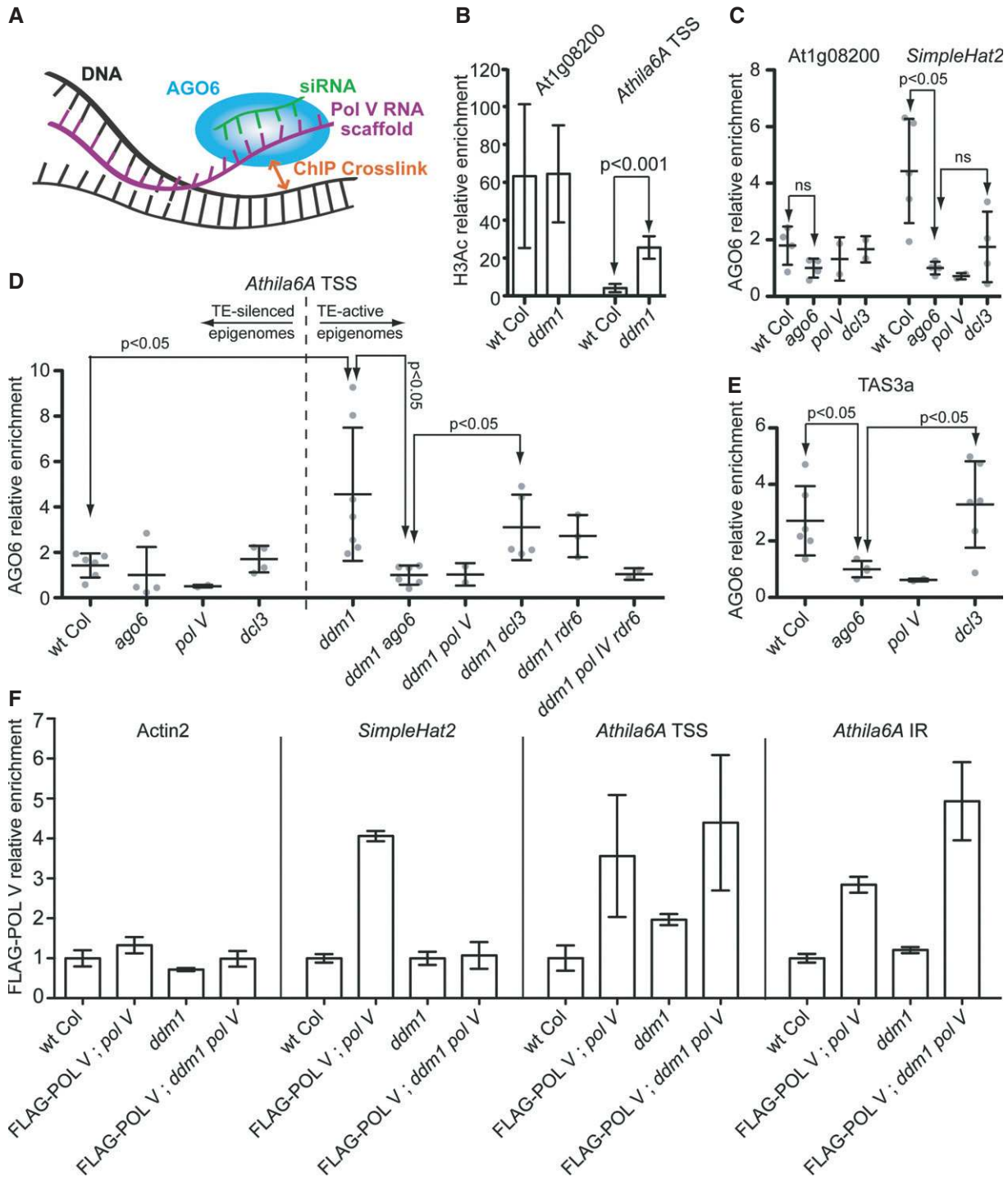


Figure 3. 21–22nt siRNAs guide AGO6 to target chromatin.

A Schematic of the ChIP crosslinking between chromatin and the AGO6 protein which is tethered to its targets by a Pol V-derived RNA scaffold.

B H3Ac control ChIP from the same chromatin used in the AGO6-IP for the At1g08200 constitutively expressed gene and the *Athila6A* TSS. Enrichment of H3Ac at the *Athila6A* TSS is increased upon transcriptional activation in *ddm1* mutants, while enrichment of H3Ac at the gene At1g08200 is consistently high in both wt Col and *ddm1*. Data are represented as mean \pm SD of six biological replicates for each genotype.

C–E AGO6-ChIP qPCR analysis of (C) the gene At1g08200 and the Pol IV-RdDM target *SimpleHat2*, (D) the *Athila6A* TSS in wt Col TE-silenced epigenomes (left) and *ddm1* TE-active epigenomes (right) and (E) TAS3a. ChIP signals in (D) are normalized to the *ago6* or *ddm1 ago6* no-antigen controls. ChIP signals in (C) and (E) are normalized to the *ago6* no-antigen control. In (C–E), each biological replicate data point is shown as a gray dot. The data are represented as mean \pm SD. For multiple statistical tests, one-way ANOVA was performed using Tukey's multiple comparison test.

F ChIP qPCR analysis of the FLAG-tagged POL V protein in wt Col and *ddm1*. ChIP signals are normalized to the wt Col no-antigen control. Data are represented as mean \pm SD of three biological replicates for each genotype.

at the target locus. Pol V is known to be recruited to heterochromatic regions of the genome (Johnson *et al*, 2014); however, it is unknown whether Pol V is still recruited to transcriptionally active TEs. Therefore, we set out to determine whether Pol V is present at the Pol II transcriptionally active *Athila6A* TSS. We crossed a FLAG epitope-tagged POL V protein into a *ddm1 pol V* TE transcriptionally active background (where FLAG-POL V complements the *pol V* mutation (Wierzbicki *et al*, 2012)). We performed ChIP for FLAG-POL V both in plants with the FLAG-POL V transgene and, as a no-antigen negative control, in wt Col plants without the transgene. We found that FLAG-POL V is not enriched at the constitutively expressed Actin-2 gene, while it is present in wt Col at a positive control undergoing Pol IV-RdDM, the *SimpleHat2* TE (Fig 3F). Next, we find that Pol V is present at the silenced *Athila6A* TSS in wt Col, as suggested by previous work that demonstrates that POL V is present at most silenced TEs (Johnson *et al*, 2014), even though this region is not undergoing significant RdDM (Figs 1A and 3F). Importantly, we detect FLAG-POL V enrichment at the *Athila6A* TSS and IR in *ddm1* mutants when *Athila6A* is transcriptionally active and targeted by RDR6-RdDM (Fig 3F). This demonstrates that POL V recruitment and function, which is a second prerequisite for RDR6-RdDM activity (Nuthikattu *et al*, 2013), still occurs when TE maintenance methylation is lost and TEs are transcriptionally activated.

Genome-wide targets of RDR6-RdDM

To understand the total contribution of RDR6-RdDM, we aimed to characterize the genome-wide association between RDR6-dependent CHH methylation and the incorporation of 21–22nt siRNAs into AGO6. Using our FLAG-AGO6-IP small RNA sequencing data, we began by assaying AGO6 enrichment of 21–22nt siRNAs for the entire genome as 100-bp tiles. Using a cutoff of ≥ 2 -fold enrichment of 21–22nt siRNAs, we identified enriched tiles and annotated them by their genomic origin: genic, intergenic or TE regions. As a control, we also identified genomic tiles that were either depleted for AGO6 or displayed intermediate levels of AGO6 incorporation of 21–22nt siRNAs. We found few AGO6-enriched 21–22nt siRNA tiles in the wt Col TE-silenced epigenome, most of which were intergenic (Fig 4A). In the *ddm1* TE transcriptionally active epigenome, we identified a sevenfold increase in the total number of tiles enriched, which primarily constitute TE and intergenic regions (Fig 4A). Additionally, we identified higher numbers of TE tiles that have intermediate or depleted levels of AGO6 enrichment of 21–22nt siRNAs, demonstrating that not all TEs or TE regions undergo RDR6-RdDM.

Using available whole-genome bisulfite DNA methylation data produced from the same tissue of wt Col, *rdm6*, *ddm1* and

ddm1 rdm6 genotypes (Creasey *et al*, 2014), we determined the CHH methylation level for each of the AGO6 21–22nt siRNA enriched, depleted and intermediate tiles. We next identified the tiles that had specifically RDR6-dependent CHH methylation by comparing the methylation of tiles in wt Col to *rdm6* and in *ddm1* to *ddm1 rdm6*. We found that only transcriptionally active TEs from the *ddm1* epigenome significantly accumulate tiles that have both AGO6 enrichment of 21–22nt siRNAs and RDR6-dependent CHH methylation ($P < 0.0001$, chi-square test) (Fig 4B), while the depleted and intermediate tiles failed to show RDR6-dependent methylation. From this analysis, we identify transcriptionally active TEs as the genome-wide target of RDR6-RdDM.

Although AGO6 incorporation of 21–22nt siRNAs in *ddm1* mutants shows a significant positive correlation with RDR6-dependent methylation, we wondered why only 22.7% of the 5,262 total AGO6-enriched *ddm1* TE tiles from Fig 4A showed evidence of RDR6-RdDM in Fig 4B. First, not all small RNAs incorporated into AGO6 will successfully target RdDM, similar to how the *Athila6A* TSS is not targeted by Pol IV-RdDM in wt Col, even though 24nt siRNAs are produced from this region and enriched in AGO6 (Figs 1, 2D and 3D). Even in the case of successful AGO6 incorporation of 21–22nt siRNAs, a required Pol V-derived scaffolding transcript may not be present, preventing RdDM. Second, whole-genome bisulfite sequencing less efficiently determines the methylation levels of highly repetitive regions such as TEs (Krueger & Andrews, 2011). Thus, we calculated the average fold coverage for genes and TEs in this bisulfite deep sequencing dataset and found that while more than 72% of genic tiles have a higher than fourfold sequencing coverage, only 50% of TE tiles have this same sequencing depth. Therefore, the total number of repetitive TE tiles that we identified in Fig 4B is an underestimate of the global TE targets of RDR6-RdDM.

We next performed a genome-wide correlation between small RNAs (and AGO6-enriched small RNAs) with MET1-maintained CG methylation, CMT3-maintained CHG, CMT2-maintained CHH and DRM2-dependent CHH-methylated regions (see Supplementary Fig S4 for further information and controls). We find that overall CHH methylation maintained by CMT2 does not correlate with small RNA accumulation or their AGO6 incorporation, suggesting that regions undergoing CMT2 maintenance of methylation are not specifically targeted by Pol IV-RdDM in wt Col or by RDR6-RdDM in *ddm1* mutants (Supplementary Fig S4A and B). Importantly, a higher number of genomic regions of the wt Col epigenome that have MET1-dependent CG methylation, CMT3-dependent CHG methylation and/or DRM2-dependent CHH methylation positively correlate with regions that produce AGO6-incorporated 21–22nt small RNAs in *ddm1* (Supplementary Fig S4A and B). Thus, the

Figure 4. Genome-wide TE targets of RDR6-RdDM.

- A AGO6 enrichment and depletion of 21–22nt siRNAs for the entire genome in 100-bp tiles for both wt Col and *ddm1*. Tiles were categorized by annotation feature.
- B Of the tiles in (A) in both wt Col and *ddm1*, we determined the number of tiles with ≥ 2 -fold CHH methylation dependent on RDR6.
- C Distribution of AGO6 enrichment across TEs (solid lines) and genes (dashed lines) for different sizes of small RNAs in either the wt Col TE-silenced epigenome or the *ddm1* TE-transcriptionally active epigenome. Dashed lines have the same color codes as the solid lines. 'Start' and 'Stop' represent the edges of the TE and the transcriptional start and stop sites of the genes.
- D–G Heatmaps of AGO6 enrichment of TE siRNA size classes in wt Col and *ddm1* epigenomes. Analyzed were (D) the position of the TE on a chromosome and TE size, (E) the copy number of the TE, (F) the TE type and (G) the TE family name. The color of the TE family name in (G) corresponds to the TE type in (F).
- H Bisulfite sequencing of TE targets. Error bars represent Wilson score interval 95% confidence limits.

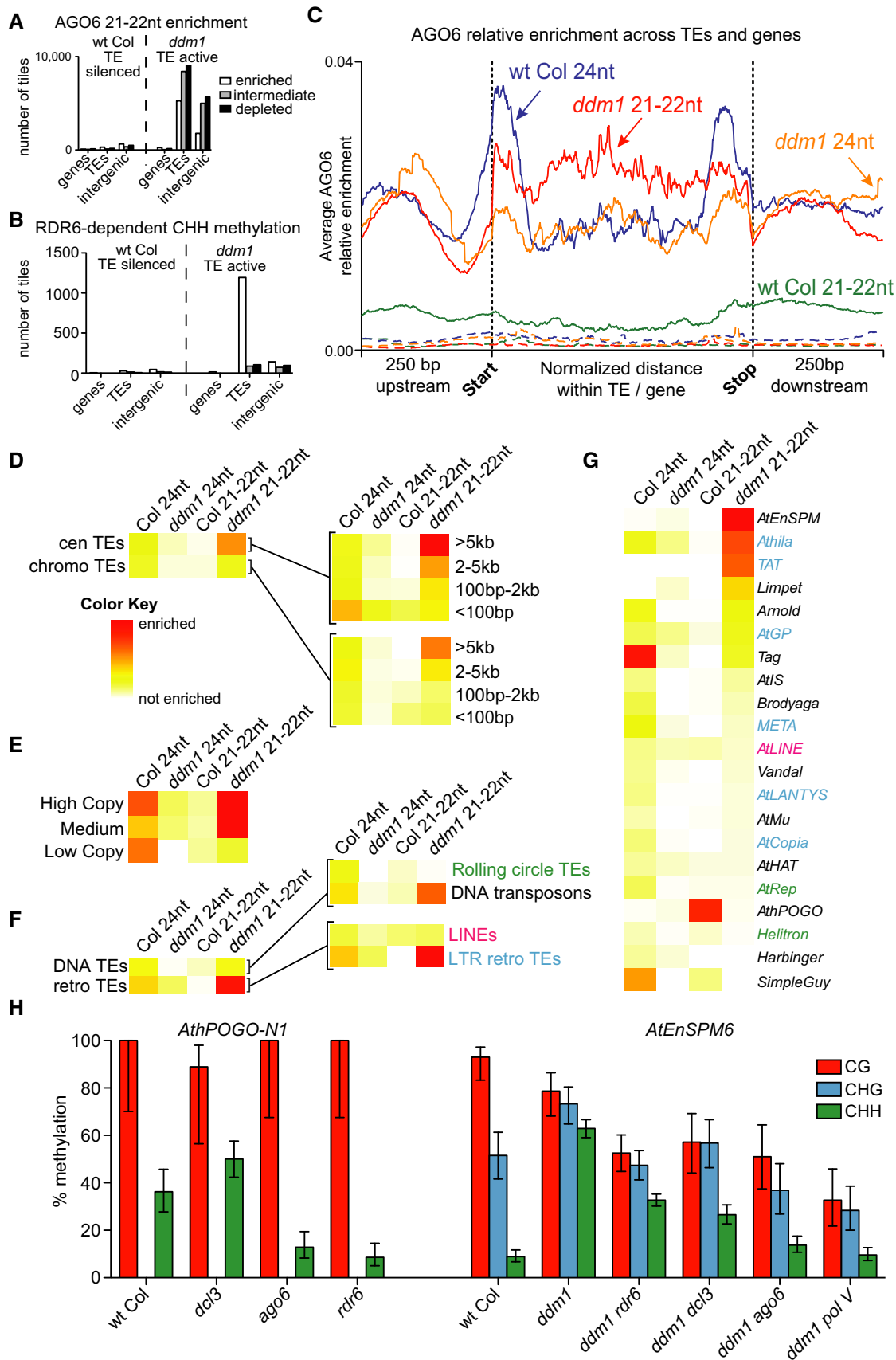


Figure 4.

correlation of methylation and small RNA enrichment supports the RDR6-RdDM model that regions of the wt Col genome that are associated with maintenance CG and CHG methylation and/or CHH methylation targeted by Pol IV-RdDM (mostly TEs) are subject to RDR6-RdDM driven by AGO6-incorporated 21–22nt siRNAs when the TEs are transcriptionally activated.

We next aimed to visualize the different functions of AGO6 in RDR6-RdDM and Pol IV-RdDM genome-wide for all TEs and genes. Figure 4C shows the relative enrichment of 21–22nt siRNAs or 24nt siRNAs in AGO6 for both the wt Col and *ddm1* epigenomes. As in Fig 4A and B, we find no accumulation of genic small RNAs (not including TAS loci) in AGO6 (dotted lines, Fig 4C). As previously reported, Pol IV-RdDM guided by 24nt siRNAs in the wt Col epigenome targets the edges of TEs (Zemach *et al*, 2013), and we find the same trend of high enrichment of 24nt siRNAs in AGO6 at the TE edges in wt Col (blue line, Fig 4C). When TEs are silent in wt Col, TE 21–22nt siRNAs do not accumulate in AGO6 (green line, Fig 4C); however, when TEs are transcriptionally activated in *ddm1* mutants, enrichment of 21–22nt siRNAs in AGO6 is found not only at the TE edges, but throughout the TE internal region as well (red line, Fig 4C). This constitutes a major functional difference between Pol IV-RdDM and RDR6-RdDM: Pol IV-RdDM targets the edges of TEs to maintain their transcriptional silencing, likely guided to these regions by the patterns of Pol IV transcription. In contrast, RDR6-RdDM uses Pol II-derived mRNAs and therefore potentially targets a large portion of the TE protein-coding body, with AGO6 selective incorporation of particular siRNAs (see Fig 2) and Pol V occupancy (see Fig 3F) likely determining the specificity of RDR6-RdDM targets.

To identify which types of TEs are targets of RDR6-RdDM, we measured the level of AGO6 enrichment of 21–22nt siRNAs for TEs split into several categories, including chromosomal position, copy number and type of TE (Fig 4D–G). We found that in wt Col, AGO6 incorporates 24nt siRNAs for most TE types, with a slight preference for 24nt siRNAs that come from short centromeric TEs (Fig 4D). Most of this AGO6 incorporation of 24nt siRNAs is lost when the TEs are transcriptionally activated in *ddm1* mutants (Fig 4D–G). In addition, the incorporation of TE 21–22nt siRNAs is very low in wt Col when TEs are silenced (Fig 4D–G). However, we did identify one TE family, the *AthPOGO* DNA transposons, that have AGO6 enrichment of 21–22nt siRNAs in wt Col (Fig 4G). We investigated this TE family and found that a non-autonomous version of *AthPOGO*, *AthPOGON1*, has AGO6 enrichment of 21–22nt siRNAs in wt Col. We show that the *AthPOGON1* TE undergoes RdDM in wt Col that is dependent on AGO6 and RDR6 but not DCL3 (see CHH methylation, Fig 4H). Thus, both the *AthPOGON1* and TAS loci undergo RDR6-RdDM in wt Col, demonstrating that RDR6-RdDM is not specific to *ddm1* mutants, and although the function of RDR6-RdDM is reduced in wt Col due to the transcriptional silencing of nearly all TEs, it is still functional for at least one TE.

When TEs are transcriptionally activated in *ddm1* mutants, TE 21–22nt siRNAs accumulate in AGO6 (Figs 2E and 4A), and we find that the TEs with the greatest levels of AGO6 enrichment of 21–22nt siRNAs are predominantly long (over 5 kb) centromeric TEs (Fig 4D). The TEs with AGO6 enrichment of 21–22nt siRNAs are also high in copy number (Fig 4E) and represent both DNA transposons and LTR retrotransposons (Fig 4F). We determined that when they are transcriptionally activated, the *AtEnSPM* DNA transposon and *Athila* and *TAT* LTR retrotransposon families represent the

major contributors of 21–22nt siRNAs enriched in AGO6 (Fig 4G). We have previously demonstrated that an *AtEnSPM* subfamily element, *AtEnSPM6*, undergoes expression-dependent RDR6-RdDM (Nuthikattu *et al*, 2013), and here, we show that the full RdDM of *AtEnSPM6* (as shown by the *ddm1 pol V* mutant) is dependent on AGO6 (see CHH methylation, Fig 4H). Similar to the *Athila6A* TSS, both RDR6-RdDM (via RDR6) and Pol IV-RdDM (via DCL3) contribute to the methylation of this TE by functioning through the AGO6 protein (Fig 4H). From the combined data in Fig 4, we conclude that the RDR6-RdDM pathway primarily targets long and centromeric transcriptionally active TEs (as well as TAS loci) over the length of the TE to establish DNA methylation.

Tissue-specific function of RDR6-RdDM in the gamete precursor cells

Not all TEs that undergo RDR6-RdDM do so in all tissues. We found that the *Athila6A* TSS is a strong target of RDR6-RdDM in *ddm1* inflorescence (floral bud) tissue (Fig 1A), but it does not undergo RDR6-RdDM in juvenile leaves (Fig 5A) despite the production of the *Athila6A* 21–22nt siRNAs in this tissue (McCue *et al*, 2012). In leaf tissue, the level of RdDM-based methylation present in *ddm1* plants is lower than in inflorescence, and this leaf RdDM is dependent on POL IV but not RDR6 (Fig 5A). This demonstrates that Pol IV-RdDM is functioning in the leaf while RDR6-RdDM is not, and since Pol IV-RdDM and RDR6-RdDM both additively function in the inflorescence, methylation of the *Athila6A* TSS is higher in inflorescence tissue. This additive increase in RdDM activity and DNA methylation level in inflorescence correlates with reduced *Athila6A gag/pol* steady-state mRNA levels compared to leaves (Fig 5B).

The explanation for this tissue specificity of RDR6-RdDM is the expression pattern of AGO6 itself. The AGO6 promoter has been previously demonstrated to express in the root and shoot apical meristems (Zheng *et al*, 2007; Eun *et al*, 2011). We show that AGO6 mRNA accumulates in inflorescences but 4.1-fold less in juvenile leaves (Fig 5C). We performed a Western blot and found AGO6 protein in inflorescences but not in leaves (Fig 5D). The lack of AGO6 expression and protein in some tissues suggests that part of the reason why RDR6-RdDM has eluded detection is that it is not occurring in the leaf tissue that has been the focus of previous studies. Alternatively, the related protein AGO4, which we find only functions in Pol IV-RdDM (Supplementary Fig S1D), accumulates in both the leaf and inflorescence (Fig 5C–D). These data demonstrate that Pol IV-RdDM functions in both the leaf and inflorescence tissue, while RDR6-RdDM is specific to the tissues where AGO6 protein accumulates.

Inflorescence is a complex mixture of reproductive and non-reproductive tissues at various stages of flower development. To determine when and where in the inflorescence AGO6 protein accumulates, we created an AGO6 translational fusion to GFP and placed this under the control of the AGO6 endogenous promoter. We found that this AGO6-GFP protein accumulates in young flower buds (up to stage 8), while the fluorescence from the AGO6-GFP protein dissipates before the floral buds open and pollination takes place (stages 11–12) (Fig 5E–H). To determine where in the young flower bud AGO6-GFP protein accumulates, we focused on a stage 4–5 floral bud, where the sepals have emerged but not yet enclosed the bud and the carpel and stamen have not yet begun to develop. We

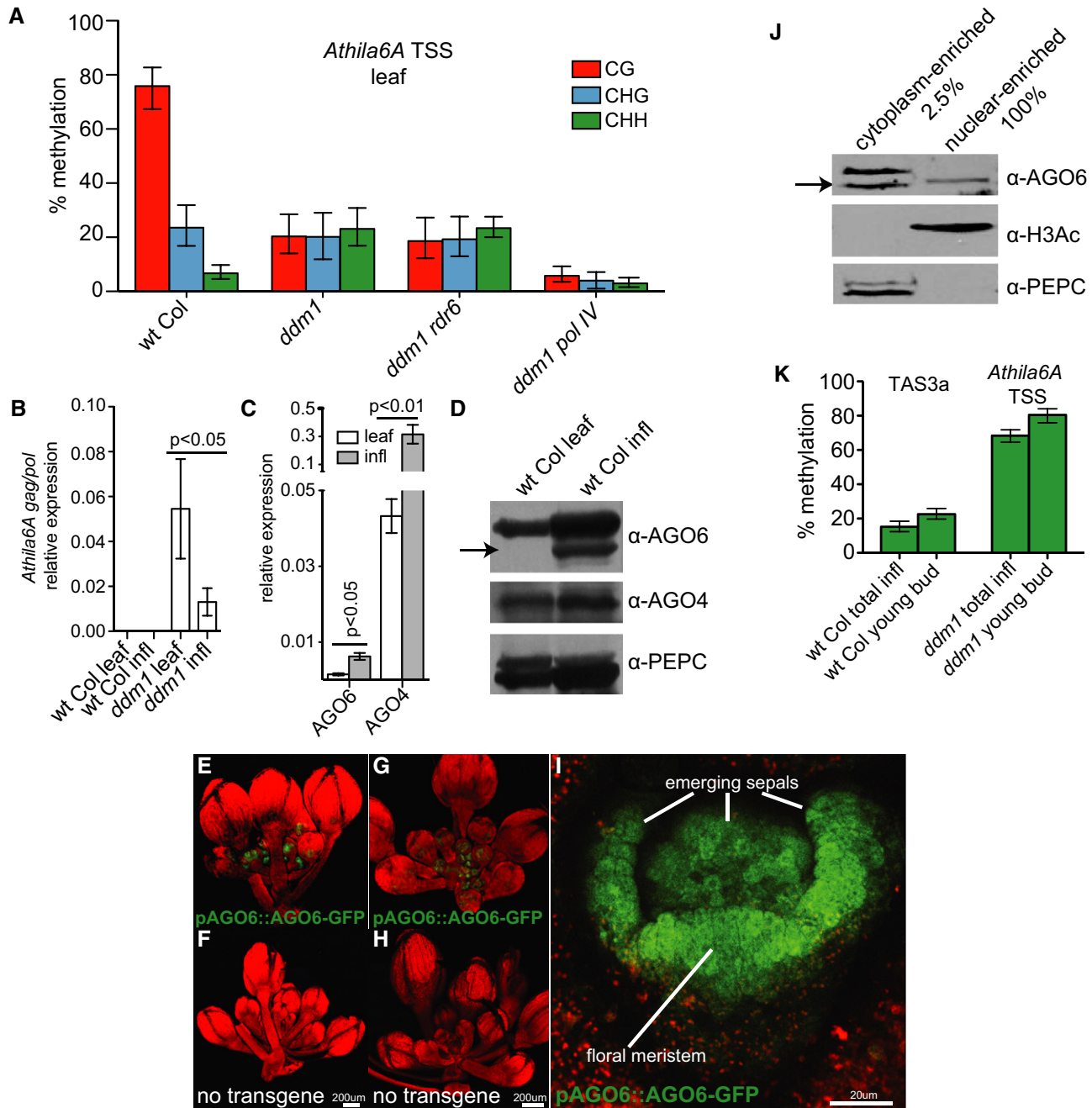


Figure 5. RDR6-RdDM is tissue specific.

- A Bisulfite sequencing of the *Athila6A* TSS in juvenile leaf tissue.
- B qRT-PCR analysis of *Athila6A gag/pol* steady-state mRNA transcript levels in inflorescence and leaf tissue.
- C qRT-PCR analysis of AGO6 and AGO4 mRNA transcript levels in inflorescence and juvenile leaf tissue.
- D Western blot analysis of wt Col leaf and inflorescence. The arrow denotes the AGO6 protein; the higher band is a previously observed cross-reactive band (see also Supplementary Fig S3A).
- E, F Side view of an inflorescence with (E) or without (F) the AGO6-GFP transgene.
- G, H Top-down view of an inflorescence with (G) or without (H) the AGO6-GFP transgene.
- I Magnified view of a stage 4–5 flower bud carrying the AGO6-GFP transgene.
- J Nuclear and cytoplasmic fractionation of inflorescence protein. The cytoplasm-enriched protein represents 2.5% of the total nuclear-enriched protein. H3Ac is a nuclear protein marker, while PEPC is a cytoplasmic protein.
- K CHH methylation levels of the RDR6-RdDM targets TAS3a and *Athila6A* TSS in whole-inflorescence-tissue samples compared to dissected young buds corresponding to floral stages 6–8.

Data information: In (B) and (C), data are represented as mean \pm SD of three biological replicates for each genotype, and differences were analyzed with a two-tailed unpaired Student's *t*-test. In (E–I), auto-fluorescence of chlorophyll is shown in red. Infl, inflorescence.

observe AGO6-GFP protein in the emerging sepals as well as in the top layers of the floral meristem (Fig 5I), which are the cells that give rise to the reproductive organs and gametes. In addition, we noted that unlike a previous report of primarily nuclear AGO6 cellular localization (Zheng *et al*, 2007), in some tissues, we observe the AGO6-GFP protein accumulation in the cytoplasm (Fig 5I). Since this AGO6-GFP transgene functionally complements the RDR6-RdDM function of AGO6 at the TAS3a locus in inflorescence tissue (Supplementary Fig S2A), our data suggest that the AGO6 protein may partially reside in the cytoplasm. To determine the cellular localization of the native AGO6 protein, we fractionated inflorescence cells into cytoplasmic-enriched and nuclear-enriched portions and confirmed significantly higher levels of endogenous AGO6 protein in the cytoplasmic fraction (Fig 5J). This is reminiscent of the finding that the related AGO protein, AGO4, is cytoplasmically localized until it is loaded with an siRNA, at which point it transits into the nucleus to participate in Pol IV-RdDM function (Ye *et al*, 2012).

Lastly, since the AGO6-GFP protein is found in young flower buds, we dissected young floral buds (stages 6–8) to determine whether we could detect increased RdDM in this purified tissue. We detect increased CHH methylation of both TAS3a in wt Col and the *Athila6A* TSS in *ddm1* (Fig 5K), demonstrating that RDR6-RdDM has its greatest effect in the tissues where AGO6 protein accumulates and where DNA methylation patterns are likely established and carried to the next generation. Recent data have shown that meristematic tissues are critical for safeguarding against TE activity (Baubec *et al*, 2014), and our data suggest that RDR6-RdDM is an AGO6-dependent mechanism that functions in floral meristem cells to establish TE methylation prior to the development of reproductive organs and gametes.

Discussion

In this manuscript, we have provided the direct mechanism of an overlooked branch of RdDM activity responsible for the expression-dependent methylation of TE and TAS endo-siRNA-generating loci. RDR6-RdDM can function in the complete absence of Pol IV-RdDM and therefore is an independent mechanism. However, RDR6-RdDM and Pol IV-RdDM often function on the same targets and may act interdependently to fully silence active TEs. For a particular subset of TEs, RdDM is much more active when the TE is transcriptionally active compared to when it is transcriptionally silenced. We have demonstrated that the RDR6-RdDM of transcriptionally active TEs operates specifically through 21–22nt siRNAs produced from TE mRNA transcripts that have been degraded by RNAi through the activity of RDR6, DCL2, DCL4 and AGO1 (McCue *et al*, 2012; Nuthikattu *et al*, 2013). The key feature allowing 21–22nt siRNAs to participate in RdDM is their incorporation into the AGO6 protein. In a background devoid of 24nt siRNAs, 21–22nt siRNAs act to guide AGO6 to its chromatin targets. Although the current dogma is that AGO6 only incorporates 24nt siRNAs, there is wide support in the literature for the incorporation of a diverse size range of siRNAs into individual AGO proteins (Farazi *et al*, 2008). In *Arabidopsis*, even after a multiple step purification of the AGO1 protein, AGO1 was shown *in vivo* to bind small RNAs of both 21–22nt and 24nt size classes (Wang *et al*, 2011). Selection of small RNAs to be

incorporated into AGO proteins is likely more dependent on the proteins that facilitate the transfer of the small RNAs to the AGO proteins than the binding specificity or interaction between the AGO protein and the small RNA itself (Meister, 2013).

Known unknowns

First, why particular TE 21–22nt siRNAs are enriched in AGO6 is currently enigmatic. We speculate that there must be a specific processing event or feature of the parent RNA transcripts that destine them for degradation and incorporation into AGO6 but not into AGO1 or other AGO proteins. Several potential mechanisms could be responsible for the biased loading of AGO6 with distinct TE siRNAs. One potential mechanism is that the cellular location of the parent mRNAs or siRNAs dictates AGO loading. For example, cytoplasmic siRNAs may be loaded into AGO1, while nuclear siRNAs may be loaded into AGO6. Although this is an attractive theory, even the nuclear-acting AGO4 is loaded with 24nt siRNAs in the cytoplasm before it enters the nucleus to function in Pol IV-RdDM (Ye *et al*, 2012), and we find AGO6 accumulation in the cytoplasm (Fig 5I and J). Another model suggests that because AGO1 expression and protein levels are not increased in *ddm1* mutants to accommodate the very high level of new TE mRNAs or 21–22nt TE siRNAs (McCue *et al*, 2013), AGO1 is overwhelmed, and the overflow mRNAs or siRNAs are directed toward other available AGO proteins such as AGO6. However, from the data presented in Fig 2 and Supplementary Fig S2, we favor a model by which individual transcripts are pre-sorted toward either AGO6 or AGO1 before subsequent degradation into siRNAs.

Second, it is unclear why the first nucleotide bias of siRNAs incorporated into AGO6 is weaker for RDR6-RdDM (Fig 2J). Constant retargeting of TE regions by Pol IV-RdDM or tasiRNA regions by RDR6-RdDM over long evolutionary time spans may result in a selection of efficient siRNAs by AGO6 and the production of a distinct 5' nucleotide bias for the incorporated siRNAs. On the other hand, evolutionarily transient TE activity would produce highly enriched 21–22nt siRNAs in AGO6, similar to those produced in *ddm1* mutants, which do not display a 5' nucleotide bias. Thus, without sufficient evolutionary time to select for a 5' bias, AGO6 is able to utilize any siRNA independent of the 5' nucleotide when it functions in RdDM to methylate transcriptionally active TEs. Regardless of how this AGO sorting and loading is accomplished, it is clear that incorporation of 21–22nt siRNAs into different AGO proteins results in vastly different biological outcomes: AGO1 loading results in post-transcriptional degradation, while AGO6 loading can lead to RdDM.

Third, as in Pol IV-RdDM, we suggest that RDR6-RdDM requires both an siRNA-loaded AGO protein and a Pol V scaffolding transcript at the target locus. However, enigmatic examples exist where both of these criteria are fulfilled, but AGO6 recruitment and RdDM do not take place (such as at the *Athila6A* TSS in wt Col (Figs 1A and B, 2D and 3F)), suggesting that these two criteria are necessary but not sufficient for AGO6 recruitment and RdDM (Fig 3D). Therefore, additional recruitment factors and steps must be missing from our current understanding of AGO targeting to chromatin.

A fourth major unknown question is how a fully unmethylated TE is originally triggered for *de novo* methylation. In our current

understanding of RdDM, Pol V must be recruited to this locus to act as a scaffolding transcript. However, Pol V itself is recruited to sites of previously existing DNA methylation (Johnson *et al*, 2014), which establishes a chicken-and-egg dilemma of how Pol V originally recognizes TEs when they are fully un-methylated. The answer to this question remains unknown. Either there must be a methylation-independent mechanism of Pol V recruitment, or another polymerase (such as Pol II) can substitute for Pol V in the initial round of methylation in models similar to the scaffolding transcript-specific function of Pol II in fission yeast and animals (Zheng *et al*, 2009; Castel & Martienssen, 2013). If Pol II can substitute for Pol V during the initial round of methylation, it is unclear why Pol II does not substitute for Pol V in later rounds of RdDM when Pol V is mutated. Recent data suggest two non-mutually exclusive models of how TE epigenetic silencing is initiated. One model suggests a 24nt siRNA-dependent pathway, while the other is 24nt siRNA-independent (Marí-Ordóñez *et al*, 2013; Panda & Slotkin, 2013). Points of agreement in these alternative models are that the TE mRNA must first be degraded into 21–22nt endo-siRNAs prior to triggering RdDM, and once RdDM occurs on the TE promoter, the maintenance of methylation will epigenetically repress TE expression.

An Achilles heel of TE activity?

How TEs are initially identified and methylated is a critical question. We believe that the tendency of TEs to transcribe their regulatory regions within their protein-coding mRNA transcripts, along with the RDR6-RdDM mechanism, constitutes an Achilles heel of TE expression. Although RDR6-RdDM functions on non-TE TAS targets, we hypothesize that the RDR6-RdDM pathway specifically evolved to affect the expression and silencing of LTR retrotransposons and DNA transposons with terminal repeats. It has been shown that the targeting of TAS loci by the RDR6-RdDM pathway results in TAS methylation confined to the transcribed region, which does not affect the expression of these loci (Wu *et al*, 2012; Kanno *et al*, 2013). If other non-TE regions, such as protein-coding genes, were targeted by this mechanism, the impact of this off-target effect would be minimal, because the methylation of coding regions (in contrast to promoters) exerts no regulatory effect on transcription (Wang *et al*, 2004). Many highly expressed genes have CG symmetrical methylation of their gene bodies (Cokus *et al*, 2008; Lister *et al*, 2008). One potential mechanism of how this off-target methylation was established follows that in the past, RNAi targeted these genes' mRNAs and these genes were subject to expression-dependent RDR6-RdDM in all cytosine sequence contexts. Subsequently, this methylation was propagated only in the CG context by the activity of the MET1 maintenance methyltransferase, while without the corresponding transcriptionally repressive histone modification, CHG and CHH methylation was not maintained.

Unlike the methylation of coding regions, the methylation of promoter and regulatory regions alters downstream transcription levels (Mette *et al*, 2000). Both LTR retrotransposons and at least some DNA transposons have repeats at their ends and transcribe their downstream LTR or TIR repeats in their protein-coding mRNAs, which may be the Achilles heel of these TEs. When these TE mRNAs are processed into 21–22nt siRNAs via RNAi, some siRNAs within this pool will match the upstream 5' LTR or TIR

repeat, which contains the TE promoter regulatory sequences. Unlike the potential methylation of gene bodies by RDR6-RdDM, the methylation of TE mRNA-coding regions will also result in methylation of their terminal repeats or subterminal regulatory elements, establishing transcriptional regulation and epigenetic repression. Therefore, because many TEs transcribe sequences that are identical to their promoters (and many TEs are found in nested configurations), the RDR6-RdDM expression-dependent mechanism of DNA methylation may have evolved to affect the transcriptional regulation of specifically TEs. A similar theory has recently been described by Inagaki and Kakutani (2013), as they suggest that the methylation and repression of *cis*-acting regulatory elements contained within TEs (in contrast to *cis*-acting regulatory elements exterior to genes) could reinforce TE silencing. We suggest the mechanism of mRNA transcript targeting that establishes this silencing is RDR6-RdDM, providing a mechanistic link for why TEs are preferentially silenced by RdDM activity.

Evolution of RdDM

We hypothesize that the Pol IV-RdDM pathway derived from RDR6-RdDM. Upon the duplication and sub-functionalization of Pol IV in plants, this second pathway arose to transcribe the RNA destined specifically for RdDM. At the same time, RDR6-RdDM still exists in *Arabidopsis* in a mostly latent form. The contribution of RDR6-RdDM in wt Col *Arabidopsis* is likely very low. The RDR6-RdDM methylation of TAS loci may represent an 'off-target' effect that triggers a mechanism meant for TE regulation (see above), and we have only identified one TE family in wt Col *Arabidopsis* that is targeted by RDR6-RdDM (*AthPOGO*). Additionally, *ago6* mutants (and *ago4* mutants) do not have morphological phenotypes and only display limited TE transcriptional activation (compared to *met1* or *ddm1* mutants, for example), suggesting that RdDM mechanisms as a whole are not as consequential when TEs are deeply epigenetically silenced via symmetrical DNA methylation. However, at the same time in wt Col plants, we believe that RDR6-RdDM surveys the transcriptome for Pol II-derived siRNAs if and when TEs are reactivated, and functions to methylate transcriptionally active TEs. AGO6 is the key effector protein of this pathway. In all well-characterized examples of RDR6-RdDM in inflorescence tissue (TAS loci, *AtEnSpm6*, *AthPOGON1* and *Athila6A*), the AGO6 protein functions to the same extent or greater than RDR6, suggesting that the vast majority of RDR6 function requires AGO6. With the mechanistic understanding of how the endo-siRNA products from TE mRNAs establish DNA methylation, the RDR6-RdDM pathway will serve as a guide for understanding endo-siRNA-mediated epigenetic silencing in other organisms.

Materials and Methods

Plant material

Plants were grown in long-day (18 h light) conditions at 23°C. FLAG-AGO6 lines were constructed from plants described in Havecker *et al* (2010). AGO6-GFP lines were constructed by amplifying the AGO6 promoter and coding region from wt Col DNA in a translational fusion to GFP in the binary vector pMDC107 (primers

listed in Supplementary Table S1). Mutant alleles used are listed in Supplementary Table S1. Unless otherwise noted, inflorescence tissue was used for each experiment.

DNA analysis

DNA was extracted via fractional precipitation, treated with RNase A and purified using phenol–chloroform extraction and precipitation. DNA was modified using the EZ DNA Methylation-Gold Kit (Zymo Research). For each modification reaction, a control unmethylated region was sequenced as in Nuthikattu *et al* (2013) to ensure that the conversion rate was above 97% (data not shown). Bisulfite-treated DNA was amplified using primers shown in Supplementary Table S1, and PCR products were cloned and individually sequenced as in Nuthikattu *et al* (2013). Data analysis was performed using Kismeth (Gruntman *et al*, 2008). For each bisulfite sequencing target/genotype combination, two or more biological replicates were performed and combined eight or more technical replicate sequences were used to determine the average methylation values.

RNA analysis

Total RNA was extracted using TRIzol reagent (Life Technologies). Quantitative RT–PCR was performed as in Nuthikattu *et al* (2013) with primers from Supplementary Table S1. Small RNAs were enriched from total RNA using the mirVana miRNA Isolation Kit (Life Technologies). Either size-enriched or IP-enriched small RNAs were used in the TruSeq Small RNA Library Preparation Kit (Illumina), and the multiplexed libraries were sequenced on a HiSeq2500 (Illumina).

FLAG-AGO6 immunoprecipitation

Lysate was prepared as in Havecker *et al* (2010) and pre-cleared with α -mouse agarose beads (Sigma) for 1 h at 4°C. The FLAG antigen was immunoprecipitated using α -FLAG magnetic bead slurry (Sigma), and immune complexes were washed as described in Havecker *et al* (2010). A no-antigen control was used for each genotype as a mock-IP. Immune complexes were eluted from the magnetic beads using the 3 \times FLAG peptide (Sigma). For small RNA deep sequencing, RNA was directly extracted from the eluates using TRIzol LS (Life Technologies).

Cellular fractionation and Western analysis

When used for Western analysis only, proteins were extracted from tissue ground in liquid nitrogen using 20 mM Tris–HCl pH 7.5, 5 mM MgCl₂, 300 mM NaCl, 0.1% NP-40 and 1% plant protease inhibitor (GoldBio). Cellular fractionation was performed as described in the Supplementary Materials and Methods. Western blots were performed as in McCue *et al* (2013). Normalization of protein input was performed with DC Protein Assay (Bio-Rad). For immunoblotting, the following antibody concentrations in 1% milk and 1 \times PBS were used: 1:10,000 α -FLAG (Sigma), 1:10,000 α -AGO1 (Agrisera), 1:1,000 α -AGO6 (Agrisera), 1:4,000 α -AGO4 (Agrisera), 1:2,000 α -H3Ac (Millipore) and 1:10,000 α -PEPC (Rockland).

Chromatin immunoprecipitation

ChIP experiments were performed as described in Huettel *et al* (2006), except for the crosslinking (see Supplementary Materials and Methods). Chromatin was immunoprecipitated with antibodies to AGO6 (Agrisera), FLAG (Sigma) and H3Ac (Millipore) at 5 μ g per IP, and immune complexes were collected with salmon sperm DNA-blocked protein A agarose beads (Millipore). The qPCR was carried out with primers shown in Supplementary Table S1. The results shown represent independent biological replicates for each genotype.

Microscopy

Microscopy was performed using a Nikon C1 confocal microscope and the NIS-Elements and EZ-C1 software packages. GFP was visualized using a 488 nm laser and 515/30 detector/filter. Chlorophyll autofluorescence was visualized using a 638 nm laser and 650 LP detector/filter. Images in Fig 5E–H were collected using a 4 \times objective (0.13 NA) and represent a maximum intensity projection of multiple z-stacks. The image in Fig 5I was collected using a 60 \times objective (1.4 NA) and is a single confocal plane through a developing flower bud. In Fig 5E–I, representative images from over 30 observed are shown.

Bioinformatic analysis

All small RNA sequencing libraries were filtered for size (18–28nt), exact matches to the *Arabidopsis* genome (TAIR10) and the removal of tRNA and rRNA reads. The TE annotation of small RNAs was performed as in Nuthikattu *et al* (2013).

For FLAG-AGO6 small RNA enrichment, filtered small RNAs from the total (non-IP), FLAG-IP and mock (no antigen)-IP were aligned to a consensus TE element, to all TEs or to the *Arabidopsis* genome and then split into 100-bp tiles with a 25-bp overlapping step. For each tile, the RPM value of IP and mock-IP small RNAs was normalized to the total (non-IP) small RNAs for that genotype. The relative enrichment of the FLAG-IP over mock-IP was calculated for each tile. Lines showing AGO6 enrichment and total small RNAs in Fig 2; Supplementary Figs S2 and S3 represent a 7-point moving average. For the first base analysis in Fig 2J, we identified specific siRNAs (21–22nt or 24nt in size) that showed at least twofold relative enrichment in AGO6 and plotted the nucleotide distribution of the first base. See the Supplementary Materials and Methods for additional details on the informatic analysis performed in each figure.

Accession Numbers

Deep sequencing files are available from NCBI GEO (GSE41755 and GSE57191).

Supplementary information for this article is available online: <http://emboj.embopress.org>

Acknowledgements

The authors thank Dalen Fultz, Eric Roose and Brian Giacomelli for their data contributions, Joy-El Talbot, Jay Hollick, Dave Bisaro, Bob Schmitz, Nathan

Springer and Kotaro Nakanishi for stimulating discussion, Xiao Zhou and Norman Groves for technical assistance and David Baulcombe and Craig Pikaard for sharing materials. This work was funded by The Ohio State Presidential Fellowship and Center for RNA Biology Fellowship to A.D.M., Ohio State, and American Society of Plant Biology undergraduate research fellowships to E.N.T., and U.S. National Science Foundation grants MCB-1020499 and MCB-1252370 to R.K.S.

Author contributions

ADM and RKS designed the experimental approach. ADM, SN, SGC and ENT performed the experiments. KP performed the bioinformatic analysis. ADM, KP and RKS wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Baubec T, Finke A, Mittelsten Scheid O, Pecinka A (2014) Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. *EMBO Rep* 15: 446–452
- Castel SE, Martienssen RA (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet* 14: 100–112
- Chen L, Dahlstrom JE, Lee S-H, Rangasamy D (2012) Naturally occurring endo-siRNA silences LINE-1 retrotransposons in human cells through DNA methylation. *Epigenetics* 7: 758–771
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452: 215–219
- Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA (2014) miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature* 508: 411–415
- Eun C, Lorković ZJ, Naumann U, Long Q, Havecker ER, Simon SA, Meyers BC, Matzke AJM, Matzke M (2011) AGO6 functions in RNA-mediated transcriptional gene silencing in shoot and root meristems in Arabidopsis thaliana. *PLoS One* 6: e25730
- Farazi TA, Juranek SA, Tuschl T (2008) The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* 135: 1201–1214
- Girard A, Hannon GJ (2008) Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol* 18: 136–148
- Grewal SI (2010) RNAi-dependent formation of heterochromatin and its diverse functions. *Curr Opin Genet Dev* 20: 134–141
- Gruntman E, Qi Y, Slotkin RK, Roeder T, Martienssen RA, Sachidanandam R (2008) Kismeth: analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinformatics* 9: 371
- Haag JR, Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol* 12: 483–492
- Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH, Baulcombe DC (2010) The Arabidopsis RNA-directed DNA methylation Argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* 22: 321–334
- Herr AJ, Jensen MB, Dalmay T, Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science* 308: 118–120
- Huetzel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJM, Matzke M (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *EMBO J* 25: 2828–2836
- Inagaki S, Kakutani T (2013) What triggers differential DNA methylation of genes and TEs: contribution of body methylation? *Cold Spring Harb Symp Quant Biol* 77: 155–160
- Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, Zhong X, Marson G, Pellegrini M, Segal DJ, Patel DJ, Jacobsen SE (2014) SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* 507: 124–128
- Kanno T, Yoshikawa M, Habu Y (2013) Locus-specific requirements of DDR complexes for gene-body methylation of TAS genes in Arabidopsis thaliana. *Plant Mol Biol Rep* 31: 1048–1052
- Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Carrington JC (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biol* 5: e57
- Kim DH, Villeneuve LM, Morris KV, Rossi JJ (2006) Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol* 13: 793–797
- Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27: 1571–1572
- Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, Strahl BD, Patel DJ, Jacobsen SE (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* 498: 385–389
- Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11: 204–220
- Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430: 471–476
- 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
- Marí-Ordóñez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O (2013) Reconstructing de novo silencing of an active plant retrotransposon. *Nat Genet* 45: 1029–1039
- Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet* 15: 394–408
- McCue AD, Nuthikattu S, Reeder SH, Slotkin RK (2012) Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genet* 8: e1002474
- McCue AD, Nuthikattu S, Slotkin RK (2013) Genome-wide identification of genes regulated in trans by transposable element small interfering RNAs. *RNA Biol* 10: 1379–1395
- Meister G (2013) Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* 14: 447–459
- Mette MF, Aufsatz W, van der Winden J, Matzke MA, Matzke AJ (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J* 19: 5194–5201
- Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C, Chen S, Hannon GJ, Qi Y (2008) Sorting of small RNAs into Arabidopsis Argonaute complexes is directed by the 5' terminal nucleotide. *Cell* 133: 116–127
- Morris KV, Chan SW-L, Jacobsen SE, Looney DJ (2004) Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* 305: 1289–1292
- Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, Slotkin RK (2013) The initiation of epigenetic silencing of active transposable

- elements is triggered by RDR6 and 21–22 nucleotide small interfering RNAs. *Plant Physiol* 162: 116–131
- Panda K, Slotkin RK (2013) Proposed mechanism for the initiation of transposable element silencing by the RDR6-directed DNA methylation pathway. *Plant Signal Behav* 8: e25206
- Pontier D, Picart C, Roudier F, Garcia D, Lahmy S, Azevedo J, Alart E, Laudie M, Karlowski WM, Cooke R, Colot V, Voinnet O, Lagrange T (2012) NERD, a plant-specific GW protein, defines an additional RNAi-dependent chromatin-based pathway in Arabidopsis. *Mol Cell* 48: 121–132
- Pushpavalli SNCVL, Bag I, Pal-Bhadra M, Bhadra U (2012) Drosophila Argonaute-1 is critical for transcriptional cosuppression and heterochromatin formation. *Chromosome Res* 20: 333–351
- Saze H, Mittelsten Scheid O, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat Genet* 34: 65–69
- 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
- Wang H, Zhang X, Liu J, Kiba T, Woo J, Ojo T, Hafner M, Tuschl T, Chua N-H, Wang X-J (2011) Deep sequencing of small RNAs specifically associated with Arabidopsis AGO1 and AGO4 uncovers new AGO functions. *Plant J* 67: 292–304
- Wang M-B, Bian X-Y, Wu L-M, Liu L-X, Smith NA, Isenegger D, Wu R-M, Masuta C, Vance VB, Watson JM, Rezaian A, Dennis ES, Waterhouse PM (2004) On the role of RNA silencing in the pathogenicity and evolution of viroids and viral satellites. *Proc Natl Acad Sci USA* 101: 3275–3280
- White E, Schlackow M, Kamieniarz-Gdula K, Proudfoot NJ, Gullerova M (2014) Human nuclear Dicer restricts the deleterious accumulation of endogenous double-stranded RNA. *Nat Struct Mol Biol* 21: 552–559
- Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, Gregory BD, Ecker JR, Tang H, Pikaard CS (2012) Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. *Genes Dev* 26: 1825–1836
- Wierzbicki AT, Ream TS, Haag JR, Pikaard CS (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* 41: 630–634
- Wu L, Mao L, Qi Y (2012) Roles of dicer-Like and ARGONAUTE proteins in TAS-derived small interfering RNA-triggered DNA methylation. *Plant Physiol* 160: 990–999
- Ye R, Wang W, Iki T, Liu C, Wu Y, Ishikawa M, Zhou X, Qi Y (2012) Cytoplasmic assembly and selective nuclear import of Arabidopsis ARGONAUTE4/siRNA complexes. *Mol Cell* 46: 859–870
- Zemach A, Kim MY, Hsieh P-H, 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
- Zemach A, Zilberman D (2010) Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr Biol* 20: R780–R785
- Zheng B, Wang Z, Li S, Yu B, Liu J-Y, Chen X (2009) Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. *Genes Dev* 23: 2850–2860
- Zheng X, Zhu J, Kapoor A, Zhu J-K (2007) Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J* 26: 1691–1701