



OPEN

# An siRNA-guided ARGONAUTE protein directs RNA polymerase V to initiate DNA methylation

Meredith J. Sigman<sup>1,2</sup>, Kaushik Panda<sup>2</sup>, Rachel Kirchner<sup>1,5</sup>, Lauren L. McLain<sup>2</sup>, Hayden Payne<sup>2,6</sup>, John Reddy Peasari<sup>2,3</sup>, Aman Y. Husbands<sup>1</sup>, R. Keith Slotkin<sup>2,4</sup>✉ and Andrea D. McCue<sup>1,2</sup>

**In mammals and plants, cytosine DNA methylation is essential for the epigenetic repression of transposable elements and foreign DNA. In plants, DNA methylation is guided by small interfering RNAs (siRNAs) in a self-reinforcing cycle termed RNA-directed DNA methylation (RdDM). RdDM requires the specialized RNA polymerase V (Pol V), and the key unanswered question is how Pol V is first recruited to new target sites without pre-existing DNA methylation. We find that Pol V follows and is dependent on the recruitment of an AGO4-clade ARGONAUTE protein, and any siRNA can guide the ARGONAUTE protein to the new target locus independent of pre-existing DNA methylation. These findings reject long-standing models of RdDM initiation and instead demonstrate that siRNA-guided ARGONAUTE targeting is necessary, sufficient and first to target Pol V recruitment and trigger the cycle of RdDM at a transcribed target locus, thereby establishing epigenetic silencing.**

Chromatin marks segregate genomes into expressed domains and regions that remain transcriptionally silenced. In mammals and plants, DNA methylation provides information such as which regions are transposable elements (TEs) and should not be expressed (reviewed in ref. <sup>1</sup>), while integrated transgenes are often subject to this regulation as well<sup>2</sup>. Most studies have focused on how DNA methylation is epigenetically maintained, resulting in heritable transcriptional repression. However, how DNA methylation is initially established at individual loci is less understood.

In both plants and mammals, DNA methylation is targeted via the action of small RNAs (piRNAs in mammals)<sup>3,4</sup>. Specifically in plants, siRNAs are produced from TEs, viruses and transgenes, targeting them for RdDM (reviewed in ref. <sup>5</sup>). RdDM is a feed-forward cycle that reinforces DNA methylation and results in epigenetic transcriptional repression. The mechanism of RdDM is split between an upstream siRNA-generating phase and a downstream chromatin-linked phase. In the upstream phase, siRNAs are generated from either RNA polymerase IV (Pol IV)- or Pol II-derived transcripts. In the downstream phase, these siRNAs are incorporated into one of the closely related ARGONAUTE proteins AGO4, AGO6 or AGO9 (ref. <sup>6</sup>). Base complementarity between the siRNA and a chromatin-linked nascent transcript results in recruitment of the de novo DNA methyltransferases DRM1 and DRM2 (ref. <sup>7</sup>). The nascent transcript, produced by RNA polymerase V (Pol V), provides the scaffold for AGO-siRNA complex interaction<sup>8,9</sup>. Pol IV and Pol V are derived from Pol II, as subunits of these holoenzymes duplicated early in plant evolution and subfunctionalized into their respective roles in siRNA biogenesis (Pol IV) and scaffolding RNA production (Pol V)<sup>10</sup>.

Pol V is continually recruited to RdDM loci through its interaction with the proteins SUVH2 and SUVH9, which bind existing cytosine DNA methylation<sup>11,12</sup>. However, how Pol V is recruited to new unmethylated DNA to first trigger RdDM and establish chromatin marks is not understood. One popular model in the literature proposes that, in the absence of Pol V, a Pol II-derived transcript

acts as the scaffold and can set the initial round of DNA methylation<sup>13,14</sup>. A second model suggests that Pol V ubiquitously surveys the genome at low levels, and the first round of RdDM occurs upon the addition of siRNAs. This ‘Pol V surveillance’ model is supported by a recent publication demonstrating that Pol V is present at a second set of loci that do not undergo RdDM due to lack of siRNAs<sup>15</sup>. In both models, after the first round of methylation, Pol V would then be recruited through the activity of SUVH2/SUVH9, leading to the positive feed-forward loop of RdDM. Neither of the two models have been examined in the context of true ‘de novo’ silencing as in ref. <sup>16</sup>, and this leaves a formidable gap in our understanding of how the first round of DNA methylation is established.

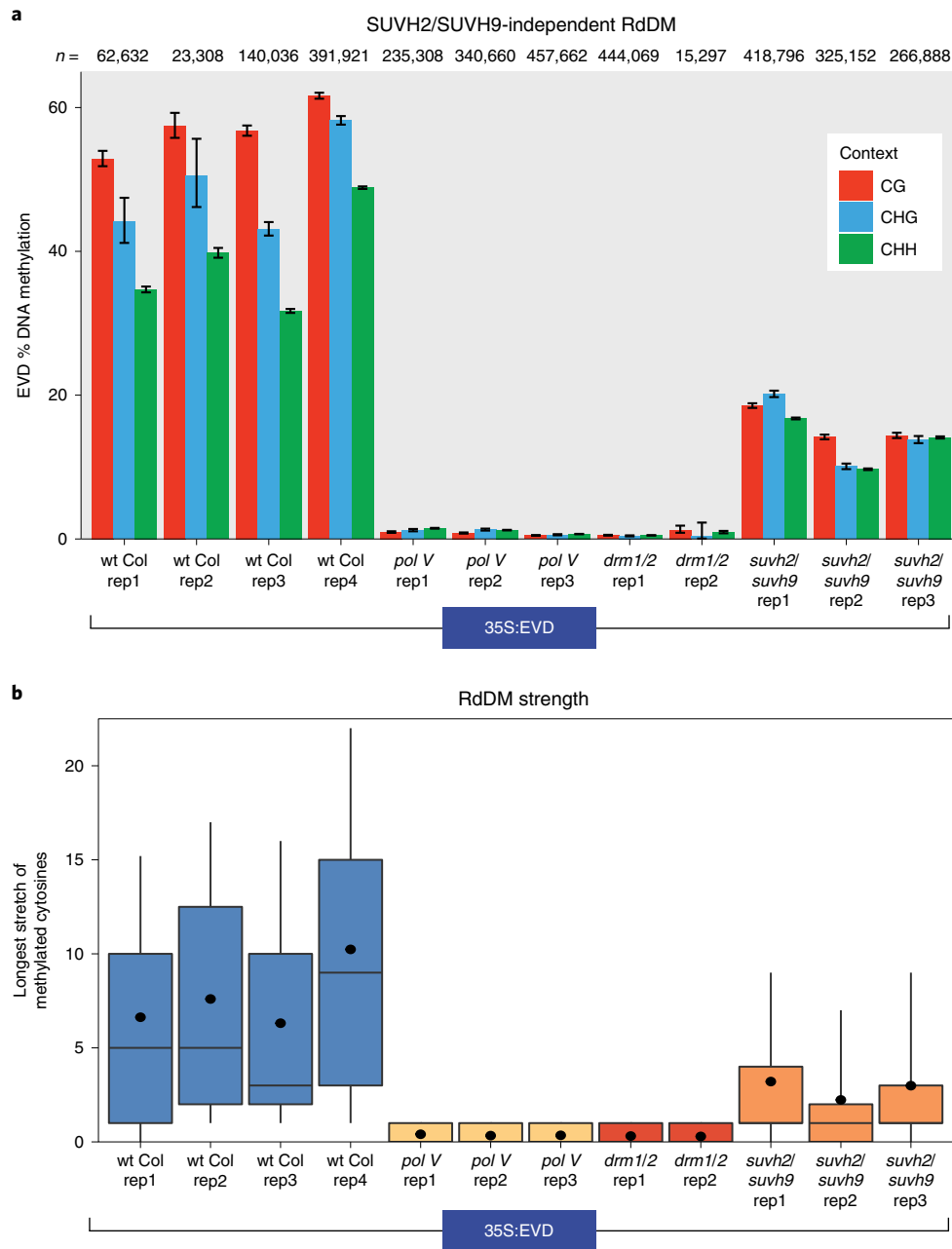
Since RdDM is a self-reinforcing cycle, it is impossible to study the first round for endogenous regions of the genome that are already engaged. To address this, we developed an approach that interrogates de novo DNA methylation of newly transformed transgene DNA, so that no pre-existing chromatin marks are guiding RdDM. We used this system to address a critical question in plant epigenetics: how is Pol V first recruited to new unmethylated sites for the first round of RdDM? Our findings demonstrate that, contrary to previous models, AGO4-clade proteins precede Pol V recruitment to new targets of RdDM. This mechanism provides the missing link between unmethylated DNA and the initiation of chromatin modification towards epigenetic silencing.

## Results

**Transgenic system to investigate the first round of DNA methylation.** The expression of TE-derived sequences reproducibly triggers RdDM<sup>17,18</sup>. To study the initiation of DNA methylation, we recreated the ‘35S:EVD’ transgene consisting of a broadly expressed promoter driving the full-length coding sequence of the *Arabidopsis Evade* TE<sup>19</sup>. This transgene was stably integrated into wild-type Columbia (wt Col) *Arabidopsis thaliana* genomes, and transgene silencing in the first-generation (T1) transformants was assayed for all experiments. To determine whether our TE transgene triggers RdDM,

<sup>1</sup>Department of Molecular Genetics, The Ohio State University, Columbus, OH, USA. <sup>2</sup>Donald Danforth Plant Science Center, St. Louis, MO, USA.

<sup>3</sup>Bioinformatics and Computational Biology Program, Saint Louis University, St. Louis, MO, USA. <sup>4</sup>Division of Biological Sciences, University of Missouri, Columbia, MO, USA. <sup>5</sup>Present address: Medical Scientist Training Program, University of Wisconsin, Madison, WI, USA. <sup>6</sup>Present address: Graduate Program in the School of Plant Sciences, University of Arizona, Tucson, AZ, USA. ✉e-mail: [kslotkin@danforthcenter.org](mailto:kslotkin@danforthcenter.org)



**Fig. 1 | SUVH2/SUVH9-independent function of Pol V. a**, BSAS of the 35S:EVD transgene in independent biological replicates of T1 plants. The bar displays the methylation percentage and the error bars represent the 95% confidence interval calculated using the Wilson score interval method.  $n$  is the number of total cytosines assayed for each amplicon. H is A, C or T. **b**, RdDM strength plot of DNA methylation. RdDM strength was measured by calculating the length of stretches of consecutively methylated cytosines along individual sequence reads. All box plots in this paper represent 25th and 75th percentile values with whiskers at the 10th and 90th percentiles, the median is represented by a line and the mean is denoted by a filled circle.

we used bisulfite amplicon sequencing (BSAS)<sup>20</sup>, a technique with high sequencing depth. In a side-by-side comparison, the average methylation remains the same for BSAS, Sanger sequencing and whole-genome sequencing techniques at target loci after conversion with sodium bisulfite, but the resolution of the data with BSAS is superior (roughly 3,800 average coverage) (Supplementary Fig. 1). When BSAS and small RNA sequencing were applied to T1 35S:EVD plants, we found that siRNAs are generated from the transgene (Supplementary Fig. 2), which results in high levels of de novo DNA methylation (Fig. 1a and Supplementary Fig. 2). Furthermore, although the coding sequence in 35S:EVD is an exact copy of the *Evadé* element EVD5 from the *Arabidopsis* genome (At5TE20395),

the transgene is not competent to transpose (Supplementary Fig. 2), and transgene methylation is not influenced in *trans* by siRNAs from the endogenous EVD elements (Supplementary Fig. 2).

**SUVH2/SUVH9-independent Pol V function.** Pol II has been proposed to substitute for Pol V during the first round of RdDM. To test this, we interrogated the methylation of T1 35S:EVD in *pol V* mutants (*nprE1* subunit mutation). We found that Pol V is necessary to trigger RdDM, as plants lacking Pol V have only background levels of methylation similar to plants lacking de novo DNA methyltransferase activity (in the *drm1/2* double mutant), and similar to the non-conversion rate during sodium bisulfite treatment

(Fig. 1a) (non-conversion rates calculated in Supplementary Fig. 1). BSAS affords increased data resolution<sup>20</sup>, which reveals that the low level of DNA methylation in *pol V* mutants exists as sporadic unconverted cytosines (stochastic), rather than as consecutive runs of methylated cytosines (indicative of DRM1/DRM2 activity in RdDM) (Supplementary Fig. 1). We assayed RdDM strength by quantifying these strings of consecutive methylated bases (Methods and controls in Supplementary Fig. 1), which takes advantage of the large number of reads generated by BSAS to individually score each read, rather than averaging the data as in other methylation analysis methods. Using this improved methodology, we found no evidence of RdDM activity in *pol V* mutants (Fig. 1b). Our data demonstrate that, contrary to the Pol II substitution model, Pol V is essential to initiate the first round of DNA methylation.

The only known mechanism for directly recruiting Pol V is through the DNA methyl-binding proteins SUVH2/SUVH9 (refs. <sup>11,12</sup>). We found that the DNA methylation of 35S:EVD in plants lacking both SUVH2 and SUVH9 does not phenocopy the total loss of RdDM in *pol V* or *drm1/2* mutants (Fig. 1a). Rather, *suvh2/suvh9* double mutants have an intermediate level of DNA methylation and RdDM strength (Fig. 1a,b), suggesting that these proteins function to amplify DNA methylation levels rather than trigger RdDM at new locations. This demonstrates that there is a mechanism other than SUVH2/SUVH9 for the first round function of Pol V at an unmethylated target locus.

#### AGO4-clade proteins are required for the first round of RdDM.

To determine which other pathway components are essential for the first round of RdDM, we transformed 35S:EVD into a series of known DNA methylation, RdDM and RNA interference (RNAi) mutants. As expected, 35S:EVD methylation and siRNA production are not dependent on DNA methyltransferases that propagate DNA methylation during DNA replication (maintenance methyltransferases) (Supplementary Fig. 3). Rather, its methylation is dependent on at least one of the closely related AGO4-clade proteins (AGO4, AGO6 or AGO9 in the *ago4/6/9* triple mutant) (Fig. 2a, Supplementary Fig. 4). Because AGO2 has a published role in RdDM<sup>21</sup>, we additionally tested *ago2* mutants and found only partial reductions in 35S:EVD 21–22-nucleotide (nt) siRNAs, DNA methylation and RdDM strength, suggesting it plays only a secondary role during the initiation of DNA methylation (Supplementary Fig. 5). Furthermore, we used IP followed by small RNA sequencing and found that 35S:EVD-derived siRNAs are enriched in AGO4 protein complexes (Fig. 2c and controls in Supplementary Fig. 6), demonstrating the direct role of AGO4 in the RdDM of 35S:EVD.

siRNA biogenesis is not dependent on Pol V, DRM1/DRM2 or AGO4/AGO6/AGO9 (Fig. 2b and Supplementary Fig. 4), confirming that these proteins function downstream of siRNA production during the chromatin-linked phase of RdDM. Instead, mutations in the downstream factors Pol V, DRM1/DRM2 and

AGO4 have increased siRNA accumulation (Fig. 2b). We found that this increase in siRNAs positively correlates with the level of Pol II transcription of 35S:EVD in *drm1/2* and *ago4/6/9* mutants (Fig. 2d), and is inversely correlated with the level of the transcriptionally repressive histone mark H3K9me2 (Fig. 2e). This demonstrates that in mutants of RdDM downstream machinery (*pol V*, *drm1/2*, *ago4/6/9*), without DNA methylation and H3K9me2, 35S:EVD expression is uninhibited, generating more transcription that leads to increased siRNA production. Additional regions of the endogenous genome also generate more siRNAs in *pol V*, *drm1/2* and *ago4/6/9* mutants (Supplementary Fig. 7), suggesting that a lack of downstream RdDM function broadly leads to enhanced Pol II transcription and increased siRNA production.

#### Pol II-derived transcripts cleaved by DICER proteins trigger RdDM.

Our data indicate that siRNA production from 35S:EVD is from Pol II, since siRNA accumulation is not lost in *pol IV* or *pol V* mutants (Fig. 2b), and the abundance of siRNAs (Fig. 2b) positively correlates with the level of Pol II transcription of 35S:EVD (Fig. 2d). In plants, siRNA sizes and categories are determined by the specific DCL protein that produces them<sup>22</sup>. Our data refute an existing ‘saturation’ model based on the identical 35S:EVD transgene, in which RdDM begins only when DCL2 and DCL4 are overwhelmed with double-stranded RNA substrate, thus activating DCL3 to generate 24-nt siRNAs<sup>19</sup>. Instead, we find that *dcl2/4* double mutants result in a roughly 50% reduction in DNA methylation (Fig. 2f) rather than being hyper-methylated as previously posited<sup>19</sup>. Further, DNA methylation is not entirely dependent on the presence of 24-nt siRNAs, as *dcl3* mutants that lack 24-nt siRNAs (Fig. 2g) do not lose all methylation (Fig. 2f) (as in ref. <sup>16</sup>). Additionally, DNA methylation still persists in *dcl2/3/4* triple mutants, where only DCL1-dependent 21-nt siRNAs remain (Fig. 2f,g). These data indicate that 21, 22 and 24-nt siRNAs are all sufficient to trigger the initiation of RdDM. DNA methylation is nearly absent only when siRNAs are severely reduced in *dcl1/2/3/4* quadruple mutants (Fig. 2f,g). Therefore, the initiation of DNA methylation is fully dependent on the presence of Pol II-derived DCL-processed siRNAs in a size-independent manner.

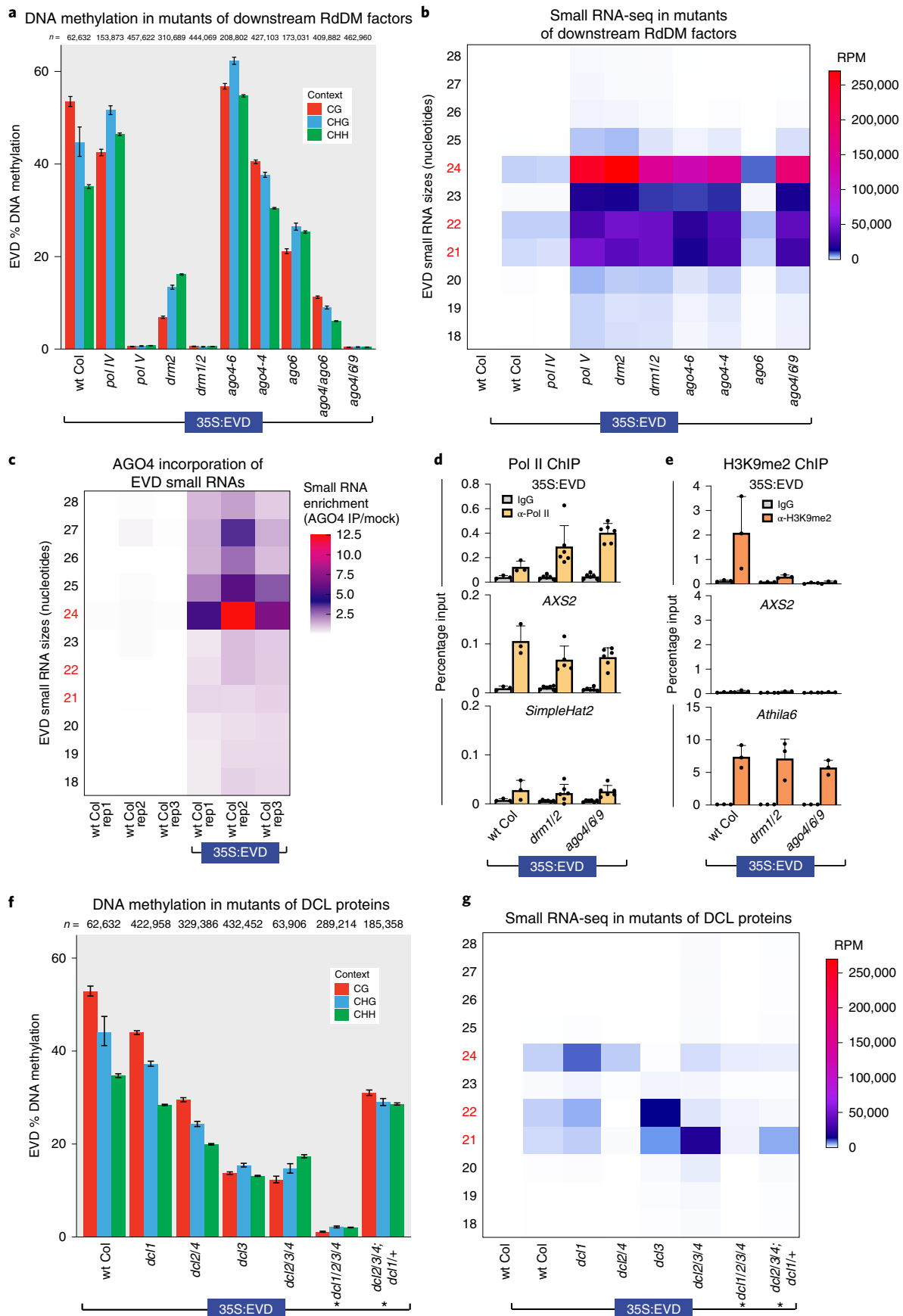
#### Initial Pol V localization requires an AGO4-clade protein.

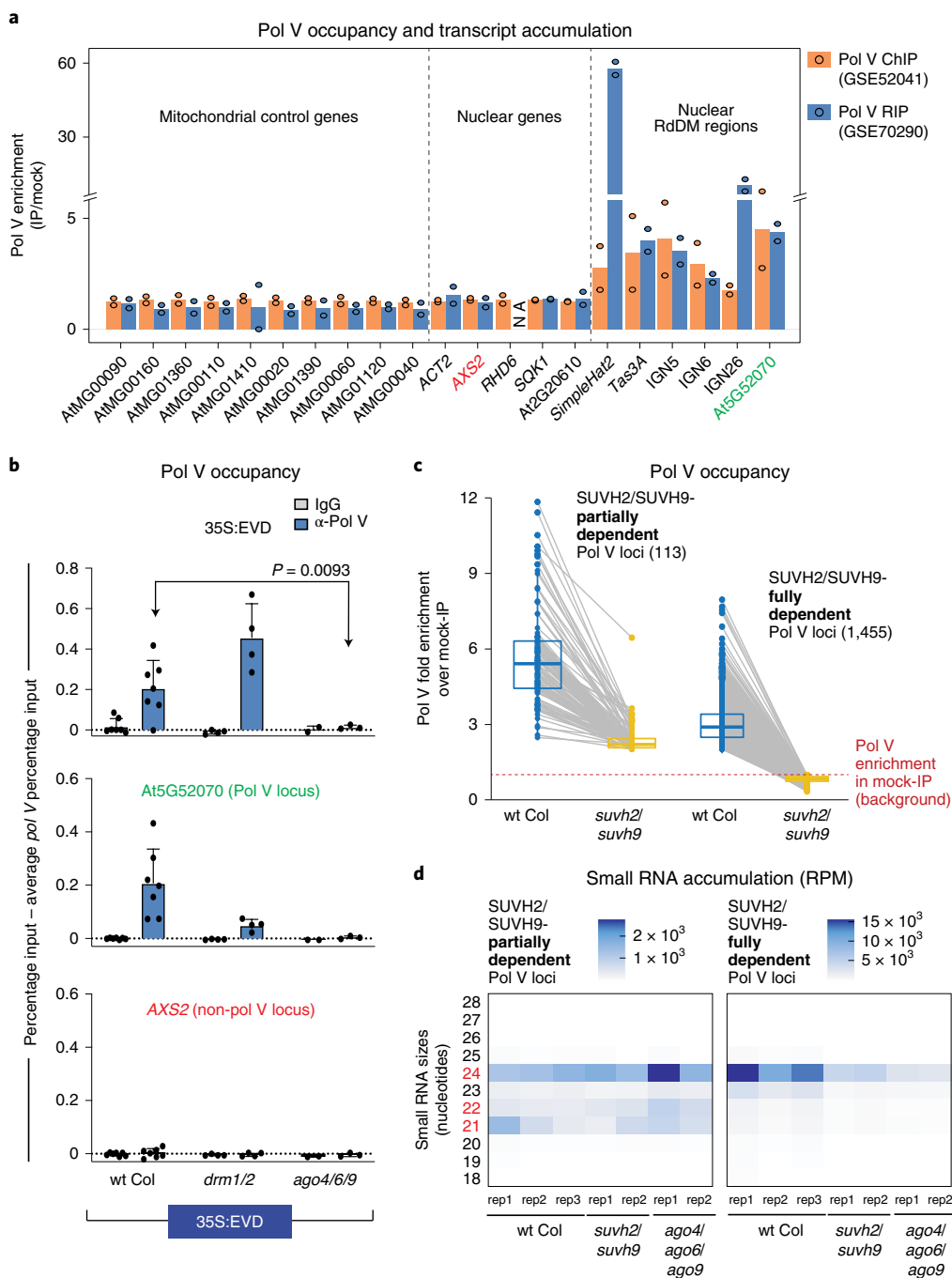
As Pol V is essential to establish RdDM, (Fig. 1), we therefore aimed to understand how Pol V is recruited to a locus in the absence of pre-existing DNA methylation. We reanalysed published Pol V chromatin immunoprecipitation (ChIP) and RNA IP (RIP) data<sup>9,12</sup> to determine whether Pol V is present (ChIP) and transcribing (RIP) at low levels throughout the genome. Our reanalysis centred on using mitochondrial genes as additional negative baseline controls, as Pol V subunits are located exclusively in the nucleus<sup>23,24</sup>. We find that Pol V signal is detected at non-RdDM nuclear loci only at the same rate as the mitochondrial negative control genes

**Fig. 2 | Genetic requirements for the first round of RdDM.** **a**, BSAS of 35S:EVD in T1 plants lacking key RdDM factors demonstrates that the initiation of DNA methylation is dependent on Pol V, DRM1/2 and at least one member of the AGO4-clade of proteins. The data display is the same as Fig. 1a. The wt Col and *pol V* data are replicates 1 and 3, respectively, from Fig. 1a. **b**, Heatmap of small RNA sequencing for transgenic lines from **a**. The accumulation of differently sized siRNAs is shown on the y axis, with the key sizes 21, 22 and 24 nt in red. The siRNAs shown were mapped to the entire length of the EVD coding region, not just the region analysed in **a** with BSAS. **c**, Heatmap of EVD small RNA enrichment in AGO4. Enrichment is calculated as the ratio of small RNA accumulation (RPM) in AGO4-IP over mock-IP samples for each size class of small RNA. **d**, ChIP analysis of Pol II (Ser5P) in T1 transgenic plants from **a** shown as percentage input. Error bars represent the standard deviation of the mean. Three biological replicates were used for wt Col, and six biological replicates were used for the other two genotypes. *SimpleHat2* is a transcriptionally silenced TE negative control, and *AXS2* is a transcribed positive control gene. **e**, ChIP analysis of H3K9me2 levels in T1 transgenic plants shown as percentage input. Three biological replicates were used for each genotype. *Athila6* is a silenced TE with known accumulation of H3K9me2. Error bars represent the standard deviation of the mean of three biological replicates. **f**, BSAS of 35S:EVD in T1 plants lacking various combinations of DCL family proteins. The *dcl1/2/3/4* quadruple mutant plants are siblings of the *dcl2/3/4*; *dcl1/+* plants (siblings are denoted with an asterisk). The data display is the same as Fig. 1a. The wt Col data are replicate 1 from Fig. 1a. **g**, Heatmap of small RNA sequencing for the same transgenic lines as in **f**. The wt Col data are the same from **b**.

(Fig. 3a). Therefore, rather than patrolling all loci at low levels, Pol V is actively recruited to its new target sites for the initiation of RdDM.

We identified three mutant combinations of downstream RdDM factors that completely lose the ability to target T1 35S:EVD for RdDM (*pol V*, *drm1/2* and *ago4/6/9*) (Fig. 2). To test the require-

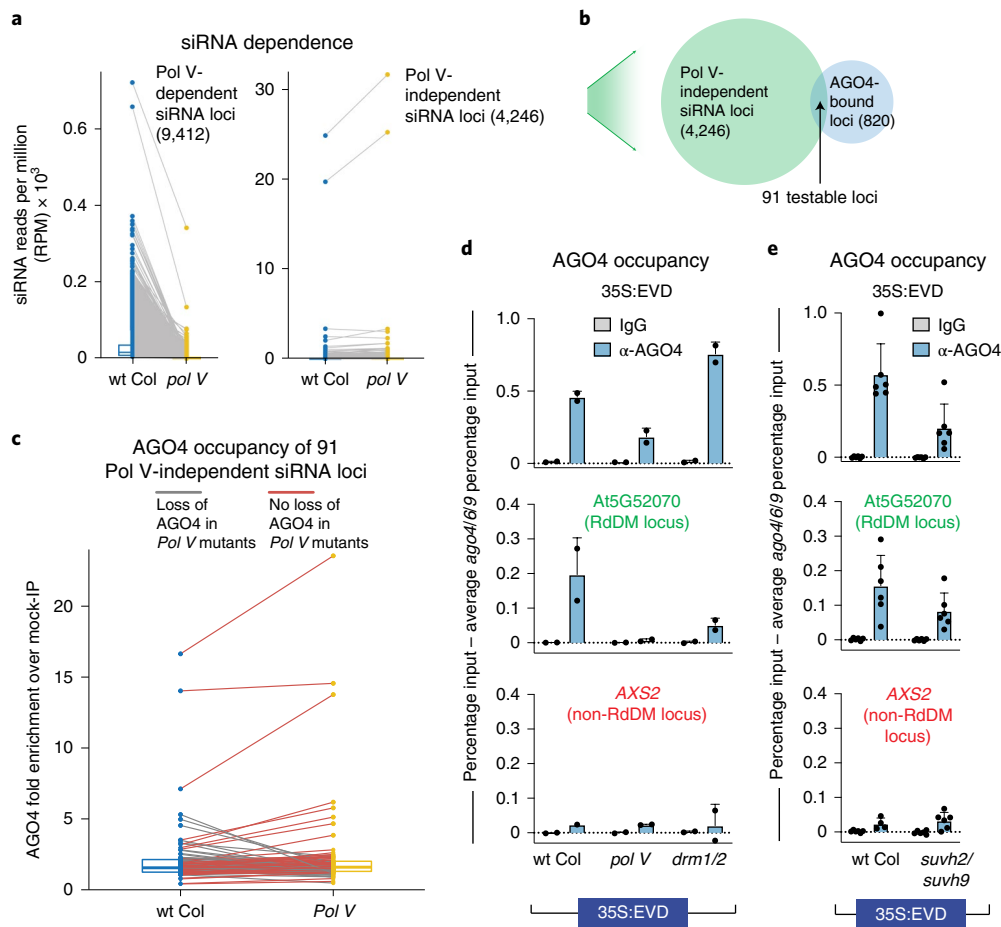




**Fig. 3 | The first round of Pol V recruitment requires an AGO4-clade protein.** **a**, Reanalysis of ChIP and RIP data of Pol V occupancy and transcription. Mitochondrial negative control genes, nuclear genes and nuclear positive control RdDM regions are shown. Two replicates are shown as dots, and their average is the height of the bar. NA, not applicable due to absence of any mapping reads in the IP or the mock samples. Genes with coloured labels are used as controls in ChIP experiments. **b**, Pol V ChIP of 35S:EVD T1 plants as percentage input minus the background of percentage input in T1 *pol V* mutants. At least three biological replicates were used for each genotype: seven for wt Col, four for *drm1/2* and three for *ago4/6/9*. *At5G52070* is a locus that undergoes RdDM and serves as a positive control for Pol V occupancy<sup>32</sup>, while *AXS2* is a gene that does not undergo RdDM and serves as a negative control. Error bars represent the standard deviation of the mean. Statistical significance was determined by a two-tailed unpaired t-test with Welch’s correction. **c**, Box plot with connected data points of Pol V occupancy comparing wt Col (blue) and *suvh2/suvh9* double mutants (yellow). Pol V occupied regions are categorized into SUVH2/SUVH9-partially and -fully dependent loci (Methods). Pol V occupancy is displayed as fold enrichment over mock-IP. Grey lines represent the change in Pol V enrichment between wt Col and *suvh2/suvh9* for each locus. Box plots represent 25th and 75th percentile with whiskers at 10th and 90th percentile values, and the median is represented by a line. **d**, Heatmap of small RNA sequencing for Pol V occupied loci that are categorized as either SUVH2/SUVH9-partially or -fully dependent from **c**.

ments of Pol V recruitment to new loci, we performed Pol V (subunit NRPE1) ChIP in these mutant backgrounds in the T1 generation of 35S:EVD plants. We found that the Pol V protein still accumulates

in each of these mutant backgrounds (Supplementary Fig. 7) and is recruited in the first generation to the TE transgene (Fig. 3b). This recruitment is not dependent on existing DNA methylation, as Pol



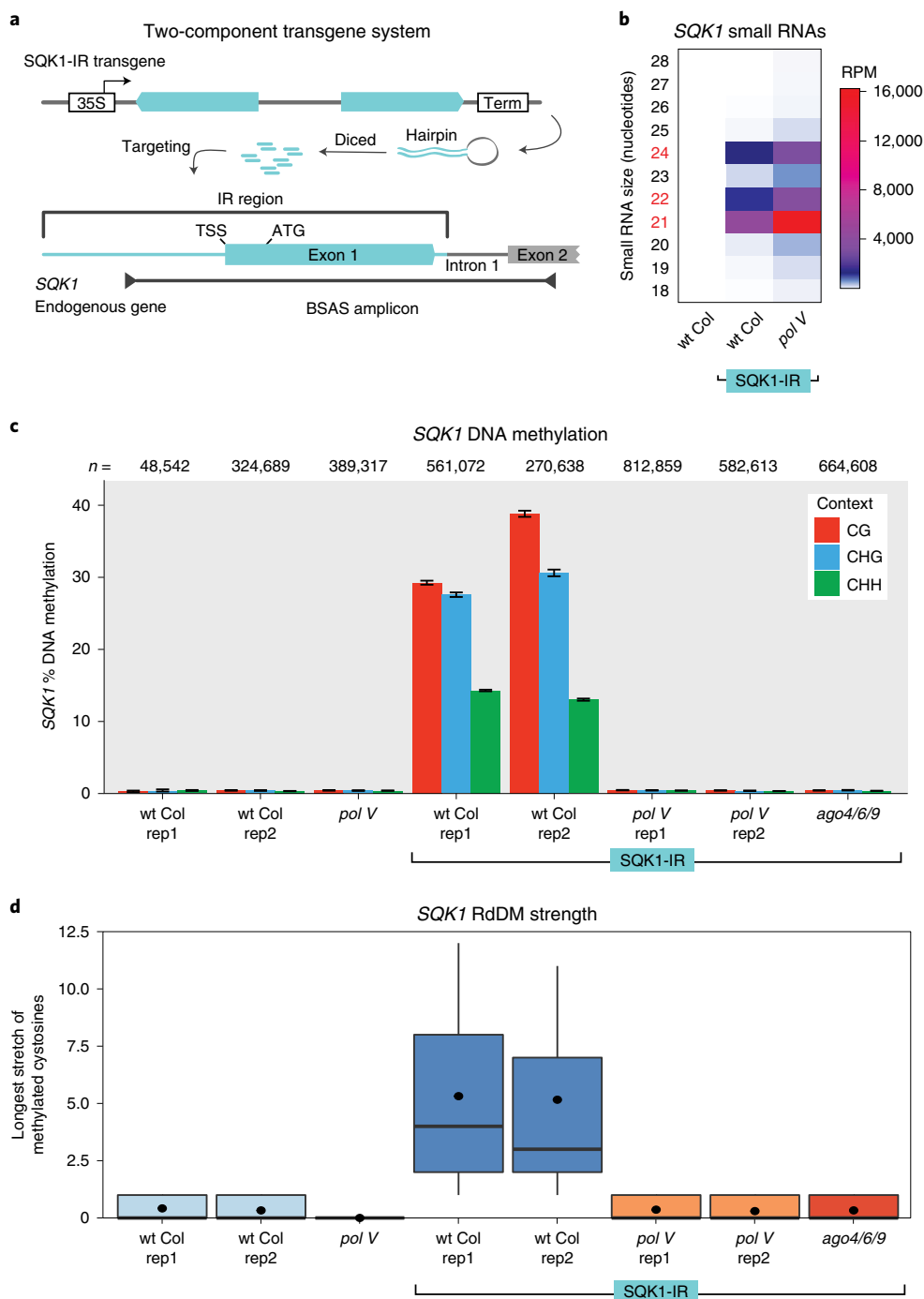
**Fig. 4 | AGO4 can localize to chromatin loci independent of Pol V.** **a**, Box plot with connected data points of 23–24-nt siRNA accumulation in wt Col (blue) and *pol V* mutants (yellow). The siRNA loci are defined as Pol V-dependent or -independent on the basis of the change in siRNA accumulation in *pol V* compared to wt Col. Box plot percentiles are the same as in Fig. 3c. **b**, Venn diagram showing the overlap of Pol V-independent siRNA loci from **a** and AGO4-enriched loci as previously defined<sup>27</sup>. The overlap provides 91 testable loci for **c**. **c**, Box plot with connected data points of AGO4 protein enrichment. AGO4 occupancy is displayed as fold enrichment over mock-IP. Grey lines display loci that lose AGO4 occupancy in *pol V* mutants (31%). Red lines display loci that retain AGO4 occupancy in the *pol V* mutant background (69%). Box plot display is the same as in Fig. 3c. **d**, AGO4 ChIP of 35S:EVD T1 plants as percentage input minus the background of percentage input in T1 *ago4/6/9* mutants. Bar graph display is the same as in Fig. 3b. At5G52070 is a positive control locus that undergoes RdDM, and AXS2 is a negative control gene that does not go through RdDM. Error bars represent the standard deviation of the mean of two biological replicates. This experiment was performed twice with similar results each time. **e**, AGO4 ChIP of 35S:EVD T1 plants as percentage input minus the background of percentage input in T1 *ago4/6/9* mutant plants. Bar graph display is the same as in Fig. 4d. Error bars represent the standard deviation of the mean of six biological replicates.

V recruitment still occurs in the *drm1/2* mutant (Fig. 3b) that lacks DNA methylation (Fig. 1). Therefore, this system provides the ability to dissect the methylation-independent recruitment of Pol V. We found that Pol V is not recruited for the first round of RdDM in the *ago4/6/9* triple mutant (Fig. 3b). This was surprising, as the prevailing dogma suggests that Pol V presence and transcription occurs first and is required to position AGO4-clade proteins at the chromatin target<sup>25</sup>. Conversely, our data demonstrate that AGO4-clade proteins are required to localize Pol V during the initiation of RdDM.

AGO4-clade proteins are directed to their targets by the complementarity of their incorporated siRNAs<sup>26</sup>. To determine whether there are regions of the endogenous genome where Pol V is positioned independently of SUVH2/SUVH9 and instead on the basis of AGO4/siRNAs, we first identified regions of the genome where Pol V occupancy is only partially dependent on the methylation-dependent SUVH2/SUVH9 mechanism (left, Fig. 3c). These 113 regions retain siRNA accumulation in *suvh2/suvh9* and *ago4/6/9* mutants (Fig. 3d). Conversely, the 1,455 regions of the genome where Pol V recruitment is fully dependent on SUVH2/

SUVH9 tend to have reduced siRNA accumulation in *suvh2/suvh9* mutants (Fig. 3d). These data indicate that when the SUVH2/SUVH9 recruitment method is absent, the continual production of siRNAs is necessary for Pol V occupancy at a small number of loci in the *Arabidopsis* genome.

**AGO4 localizes independently of Pol V and DNA methylation.** We aimed to order the events of recruitment at the chromatin that result in the first round of RdDM. Since Pol V recruitment to 35S:EVD is dependent on a protein of the AGO4-clade (Fig. 3), we aimed to determine whether the converse was true: is AGO4's interaction with chromatin dependent on Pol V? Our data support this idea that there are a small number of regions of the endogenous *Arabidopsis* genome where AGO4 interacts with target chromatin loci independent of a Pol V-derived scaffolding transcript. Since AGO4 targeting is dependent on siRNA complementarity, we began by identifying loci that continue to produce 23–24-nt siRNAs in a *pol V* mutant (Fig. 4a). We overlapped these 4,246 Pol V-independent siRNA loci with 820 previously identified AGO4-bound loci<sup>27</sup>, resulting in 91



**Fig. 5 | siRNAs are sufficient to target the first round of RdDM. a**, Two-component system with siRNA production from a T1 inverted repeat (IR) transgene targeting the endogenous gene *SQK1* for RdDM. **b**, Heatmap of small RNA sequencing. **c**, BSAS for T1 transgenic *SQK1*-IR plants. The data display is the same as Fig. 1a. **d**, Box plots of RdDM strength for plants from **c**. Box plot percentiles are the same as in Fig. 1b.

testable AGO4-bound regions of the genome that do not lose siRNAs (Fig. 4b). Of these testable AGO4-bound regions, 63 (69%) retain AGO4 occupancy in the *pol V* mutant (Fig. 4c), demonstrating AGO4 recruitment without Pol V.

To determine whether AGO4 is localized to the 35S:EVD transgene during the initiation of RdDM, we performed AGO4 ChIP on T1 35S:EVD plants. As a control, we confirmed that the AGO4 protein is present in the various mutants we tested (Supplementary Fig. 7). In the ChIP experiment, we detected AGO4 at 35S:EVD during the initiation of RdDM in wild-type plants, as well as in *pol V*, *svh2/svh9* and *drm1/2* mutants (Fig.

4d,e). The level of AGO4 at 35S:EVD in *pol V* and *svh2/svh9* mutants is not as high as in wt Col plants, but is nonetheless substantially higher than the negative control *AXS2* gene and the IgG negative control (Fig. 4d,e). In addition, the AGO4 ChIP signal in the *drm1/2* mutant may be higher due to the increased siRNAs produced in this mutant (Fig. 2b). We conclude that AGO4 is directed to the target transgene without the requirement of a Pol V scaffolding transcript or DNA methylation, and is independent of the known SUVH2/SUVH9 recruitment mechanism of Pol V (Fig. 4e), placing AGO4 interaction with the target locus before and not reliant on Pol V activity.

**siRNAs are sufficient to direct Pol V function.** Since AGO4 can be recruited to a new target locus independent of Pol V, we aimed to determine whether siRNAs are sufficient to direct Pol V activity and the first round of RdDM. Our 35S:EVD transgene system cannot address this question, since the source locus of the siRNAs and the target of Pol V action are the same (*cis*-acting). Instead, we generated a *trans*-acting two-component system to uncouple siRNA production from Pol V recruitment. The siRNAs are generated from an inverted repeat transgene that targets an unmethylated endogenous gene in *trans* (Fig. 5a) (as in refs. 28–30). The endogenous gene we targeted (At3G12210) encodes a broadly expressed DNA-binding protein that does not produce siRNAs (Fig. 5b), is not methylated (Fig. 5c) and is not bound by Pol V (Fig. 3a) in wt Col plants. We named this uncharacterized gene *SQUEAKY1* (*SQK1*). On addition of the *SQK1* inverted repeat transgene (*SQK1*-IR), 21, 22 and 24-nt siRNAs accumulate from the hairpin region (Fig. 5b) and these siRNAs function in *trans* to direct RdDM to the *SQK1* endogenous gene (Fig. 5c,d). In a *pol V* mutant, siRNAs are still generated from the inverted repeat transgene (Fig. 5b), but RdDM does not occur (Fig. 5c,d), again demonstrating that Pol V acts downstream of siRNA production and is essential for RdDM. As with 35S:EVD, the initiation of RdDM is fully dependent on an AGO4-clade protein (Fig. 5c,d). Taken together, these data demonstrate that the production of new siRNAs and the presence of an AGO4-clade protein are sufficient to target Pol V-dependent RdDM to a new non-transgenic locus independent of existing methylation.

**Pol II expression is necessary for the first round of RdDM.** While Pol II cannot substitute for Pol V (Fig. 1), we have identified that Pol II does play a role in making the target locus receptive to the first round of RdDM. While investigating the targeted methylation of the *SQK1* gene (Fig. 5), we observed that the CHH methylation (indicative of RdDM) was not distributed evenly across the region targeted by siRNAs. Instead, methylation peaks upstream of the transcriptional start site near the proximal *SQK1* promoter (Fig. 6a). This pattern of methylation does not correlate with the abundance of siRNAs across this region (Fig. 6a), suggesting that the observed methylation pattern is based on the *SQK1* promoter activity. To test the role of Pol II expression during the first round of RdDM, we used Cas9 to generate a 1,298 bp deletion of the *SQK1* promoter (*sqk1-1*) (shown in Fig. 6a). This severely reduces, but does not completely eliminate *SQK1* expression (Fig. 6b). Where on the *sqk1-1* allele Pol II initiates and the direction of transcription is not known. When the homozygous *sqk1-1* mutation is combined with the *SQK1*-IR, although the siRNAs still target *sqk1-1*, methylation is substantially reduced (Fig. 6a,c). In a separate test, we generated a second inverted repeat targeting system (as in Fig. 5a) for a differ-

ent gene that has a specific and limited developmental expression pattern (Supplementary Fig. 8). Our data for both two-component inverted repeat systems (Fig. 6a–c and Supplementary Fig. 8) demonstrate that the strength of RdDM positively correlates with the level of Pol II expression at the target locus.

To decisively test the necessity of Pol II expression in the first round of RdDM, we created two new EVD transgenes that are definitively either expressed or not. These include the ‘EVD-only’ (EVD coding sequence with no promoter) and ‘T3A-EVD’ transgenes (T3A terminator directly 5′ of the EVD coding sequence, to ensure no read-through Pol II transcription) (Fig. 6d). Reverse transcription with quantitative PCR (RT–qPCR) of EVD transgenes confirms that only 35S:EVD has appreciable messenger RNA accumulation (Fig. 6e), although the expression levels generated by the 35S promoter are highly variable due to the nature of T1 transgenesis<sup>31</sup>. We confirmed that RdDM of 35S:EVD is dependent on Pol II expression, as only the expressed transgene version is targeted for methylation (Fig. 6f and biological replicates in Supplementary Fig. 8). The key test was when we combined expressed and unexpressed EVD transgenes in the same genome. We found that whenever the expressed 35S:EVD transgene is present, it produces abundant siRNAs (Fig. 2b and Fig. 6g, lanes 1, 4 and 5), which are incorporated into AGO4 (Supplementary Fig. 6) and drive RdDM to 35S:EVD itself (Fig. 1 and Fig. 6h, lanes 1, 4 and 5). When a second promoterless EVD transgene is introduced into the same plant genome that has 35S:EVD, these unexpressed transgenes do not become methylated (Fig. 6h lanes 4–5, biological replicates in Supplementary Fig. 8). These double-transgenic plants have EVD siRNAs that are incorporated into AGO4 and are competent to perform RdDM (evidenced by the fact that 35S:EVD is targeted by RdDM in the same plant (Fig. 6g, lanes 4–5)). The only difference as to why one transgene is methylated and the other is not, is the Pol II activity at the target transgene. Therefore, Pol II transcription is not sufficient to substitute for the absence of Pol V, but Pol II activity is necessary to make a locus receptive to the first round of RdDM.

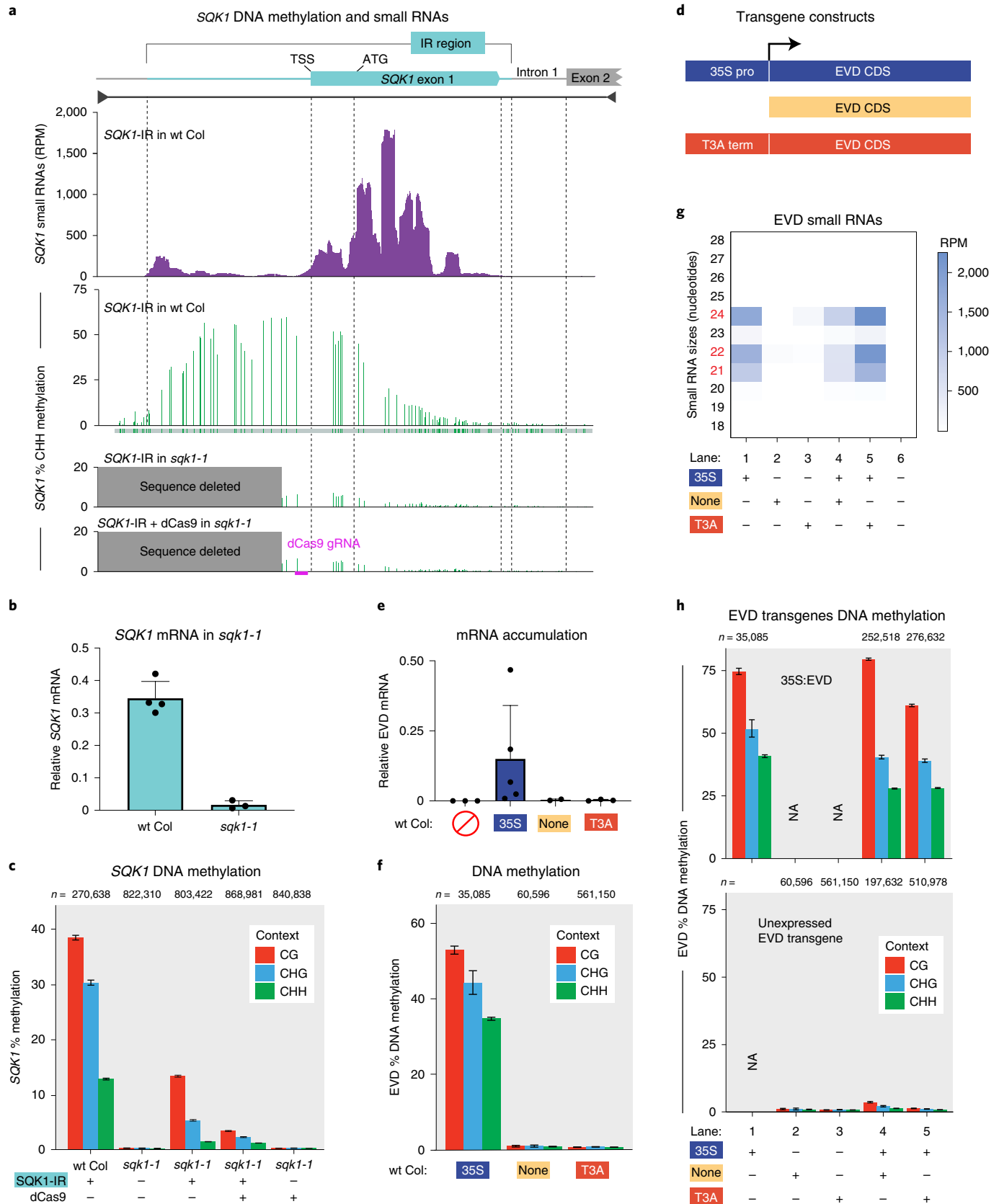
**The role of Pol II in the initiation of RdDM.** We aimed to identify the exact function of Pol II during the first round of RdDM. Pol II could be involved in producing a scaffolding transcript for AGO–siRNA interaction. Alternatively, AGO–siRNA complexes are known to interact with single-stranded DNA<sup>32</sup>. Pol II’s function at the target locus could be to open the double-stranded structure of the DNA, allowing for the AGO–siRNA complex to base pair with single-stranded DNA, which fits the theory that RdDM is focused at DNA replication forks<sup>33</sup>. To separate these models, we attempted to increase the RdDM of the low-expressed *sqk1-1* allele by targeting dCas9 to this locus. We used a gRNA to target dCas9 and R-loop

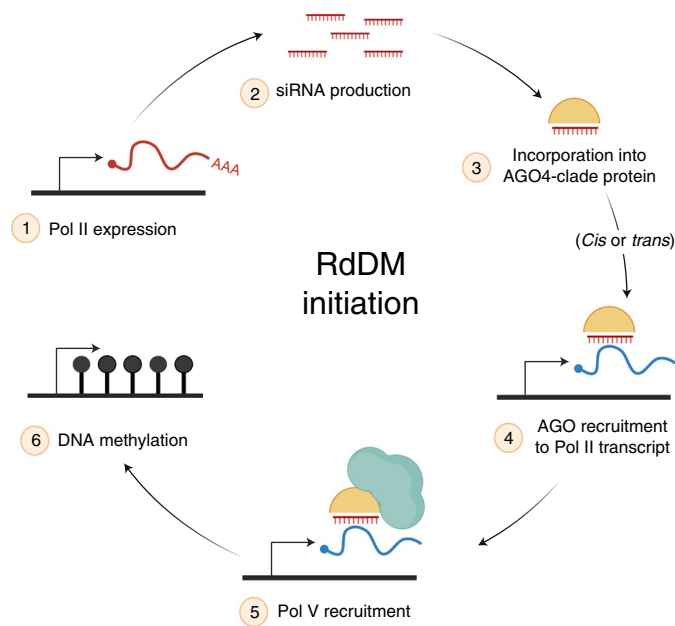
**Fig. 6 | The role of Pol II in the first round of RdDM.** **a**, CHH context DNA methylation and siRNAs (combined 21, 22 and 24 nt) aligned to the endogenous *SQK1* gene. The region of *SQK1* that matches the *SQK1*-IR transgene is annotated. The track below the *SQK1*-IR of wt Col CHH methylation identifies where each CHH context cytosine is located (green), even if unmethylated. The DNA methylation track for *SQK1*-IR in wt Col is produced from BSAS of two overlapping amplicons. Dashed lines align key annotation features of the *SQK1* gene, and the location of the CRISPR gRNA used in the dCas9 experiment is annotated. **b**, RT–qPCR of plants homozygous for the *SQK1* promoter deletion allele *sqk1-1*. The error bar represents the standard deviation of the mean in three biological replicates. **c**, BSAS of plants with and without the *SQK1*-IR, the homozygous *sqk1-1* deletion and dCas9. The data display is the same as Fig. 1a. The wt Col sample is the same as replicate 2 in Fig. 5c. **d**, Depiction of three EVD TE transgenes. These transgenes share the exact coding sequence of EVD5 (AT5TE20395) lacking any vector-encoded terminator sequence while differing in their upstream elements. These upstream elements include the 35S constitutive promoter (from Fig. 1), no promoter and the T3A terminator. **e**, RT–qPCR of EVD mRNA in T1 transgenic plants indicates that only 35S:EVD has Pol II expression. The error bar represents the standard deviation of the mean of three biological replicates. **f**, BSAS of T1 transformants for each respective EVD transgene in wt Col. The data display is the same as Fig. 1a. The 35S:EVD sample is the same as replicate 1 from Fig. 1a. **g**, Heatmap of small RNA sequencing from single and double EVD transgenic plants. The siRNAs shown were mapped to the entire length of the EVD coding region, not just the region analysed by BSAS in **f**. The non-transgenic wt Col sample is the same as from Fig. 2a, whereas the 35S:EVD wt Col sample (lane 1) is from the T3 generation, to match the same 35S:EVD generation as in the double-transgenic plants (lanes 4 and 5). **h**, BSAS for the 35S:EVD transgene (top) and the second unexpressed EVD transgene (bottom) in the same transgenic individuals. NA, not applicable, as that transgene is not in this plant line. Biological replicates are shown in Supplementary Fig. 8. The data display is the same as Fig. 1a. The wt Col 35S:EVD sample (lane 1) is from the T3 generation, to match the same 35S:EVD generation as in the double-transgenic plants (lanes 4 and 5).



formation<sup>34</sup> to the homozygous *sqk1-1* locus in the presence of the SQK1-IR (see gRNA location in Fig. 6a). This did not result in higher methylation (Fig. 6a,c) as would be expected if DNA opening

was the RdDM function of Pol II at the target locus. Instead, methylation decreased (Fig. 6c), as though dCas9 interferes with the low level of Pol II transcription at this locus. Since this dCas9 experiment





**Fig. 7 | Model of RNA-driven recruitment of Pol V to new RdDM targets.**

The initiation of RdDM begins with Pol II expression of the unmethylated locus (step 1), followed by siRNA production from that transcript by any member of the DCL family (step 2). The siRNAs are incorporated into a member of the AGO4-clade of proteins (step 3) and this can only target RdDM to complementary regions of the genome that are expressed (step 4). *Cis* action refers to a region that generates siRNAs and targets its own methylation, while in *trans* action the siRNAs are acting on a separate locus. The AGO4 targeting of a Pol II expressed region of the genome results in Pol V recruitment (step 5) and the first round of RdDM (step 6). This final methylated product can then enter the methylation-dependent pathway of continual Pol V recruitment guided by SUVH2/SUVH9. See Discussion for details. Created with BioRender.com.

provided negative results, several alternative interpretations exist. However, this experiment indicates that the function of Pol II at a new RdDM target locus is likely to generate the RNA transcript that acts in AGO–siRNA complementary base pairing.

## Discussion

The roles of RNA polymerase IV and V at silenced TE fragments undergoing the self-reinforcing cycle of RdDM are well described<sup>35</sup>. However, the specific roles of these polymerases and Pol II during the initiation phase of TE and transgene silencing have remained enigmatic. Here we dissect polymerase function using newly integrated transgenes. This strategy permits the discrimination of the first round of RdDM from those already engaged in the RdDM cycle.

We find that Pol V is required for all de novo RdDM, and Pol II cannot substitute for this function. Our data refute the model whereby Pol II transcripts are able to initiate the first round of DNA methylation. The essential function of Pol V during the initiation of RdDM correlates with the recent finding that Pol V is a key factor in the evolutionary repression of TEs in the *Arabidopsis* lineage<sup>36</sup>. Our data also refute the ‘Pol V surveillance’ model, whereby Pol V produces scaffolding transcripts that, when a siRNA is present, will trigger the first round of methylation. It has been recently demonstrated that the artificial tethering of a Pol V-recruitment factor to a genomic locus triggers RdDM<sup>37</sup>. Therefore, if Pol V surveilled everywhere, we would expect to observe spurious low-level methylation across the genome, which is not detected, even in DNA glycosylase mutants that fail to remove DNA methylation from the

genome<sup>38</sup>. Although our data refute the Pol V surveillance model, our findings agree with the point that siRNAs are the key determinant to target new RdDM<sup>15</sup>.

Our data demonstrate a methylation-independent mechanism for recruitment of Pol V to a new target locus. Previous reports suggested that Pol V is recruited independently of AGO4 and siRNAs, chromatin-associated Pol V recruits AGO4 in a strictly protein-dependent manner<sup>32</sup>, and RdDM would occur only when both Pol V and AGO4 were recruited to a locus<sup>39</sup>. In this previous model, the likelihood that the recruited AGO4 would contain the complementary siRNA that matches the locus of recruitment is low. Our results indicate a more linear ordered pathway of recruitment, with base complementarity of the incorporated siRNA first and independently guiding AGO4 to the chromatin, followed by the recruitment of Pol V. We propose an siRNA-directed pathway of Pol V recruitment in Fig. 7.

Pol II’s function during the first round of RdDM is twofold. First, it is necessary to produce the raw transcripts for siRNA production (step 1, Fig. 7). During the self-reinforcing cycle of RdDM at endogenous silenced TEs, this function is taken over by Pol IV. However, without existing heterochromatic marks to recruit Pol IV, the new region must be transcribed by Pol II to generate siRNAs in *cis* or siRNAs provided from a separate locus in *trans*. We find that in cases where Pol II is generating siRNAs, mutations in the downstream RdDM factors (such as *pol V*, *drm1/2* and *ago4/6/9*) result in the lack of DNA and H3K9 methylation, leading to more Pol II expression and more substrate for the higher accumulation of siRNAs (Fig. 2).

Once the raw transcripts are produced by Pol II, we find that they can be processed by any DCL protein into 21–24-nt siRNAs (step 2, Fig. 7). All these siRNAs sizes, generated by any of the four *Arabidopsis* DCL proteins, are capable and sufficient to target the first round of RdDM. Our data agree with the ‘all hands on deck’ model whereby during the initiation of RdDM all available DCL proteins function to process the large volume of Pol II transcripts, producing 21, 22 and 24-nt siRNAs that are all capable of guiding RdDM<sup>16</sup>. These siRNAs are loaded into an AGO4-clade protein (step 3, Fig. 7), as RdDM function was abolished in the *ago4/6/9* triple mutant. While all DCL and AGO4-clade proteins are sufficient to initiate RdDM, our data indicate that DCL3, 24-nt siRNAs and AGO4 primarily perform this function in wild-type plants.

Second, Pol II action is required downstream of siRNA production. Even when complementary siRNAs and AGO4 are present, the first round of RdDM is dependent on Pol II activity at both transgenes and endogenous genes (Fig. 6). Our data indicate that the role of Pol II at the target locus is to generate the first scaffolding transcript for AGO–siRNA interaction (step 4, Fig. 7). This is supported by the protein–protein interaction of the Pol V-recruitment factor RDM1 with Pol II, AGO4 and DRM2 (ref. 24). This function of Pol II suggests that it retains some of its ancestral ability to generate scaffolding transcripts, which has otherwise been subfunctionalized and relegated to Pol V. Even though the Pol II transcript interaction with the AGO–siRNA complex does not result in methylation, it is necessary for the recruitment of Pol V (step 5, Fig. 7) and subsequent RdDM (step 6, Fig. 7). A recent study demonstrated that Pol II’s C-terminal domain can act to recruit more Pol II protein<sup>40</sup>, and during the initiation of RdDM Pol II may act in a similar manner to recruit Pol V, target RdDM and result in the establishment of epigenetic inheritance.

## Methods

**Plant growth and lines.** *A. thaliana* plants were grown at 22 °C on Pro-Mix FPX soil in Conviron MTPS-120 growth chambers with 16-h per 200 μmol m<sup>-2</sup> s<sup>-1</sup> light. The specific alleles of all mutants are shown in Supplementary Table 1. Inflorescence (flower buds, stages 1–12) tissue was used for all experiments unless otherwise noted. All transgenic material is the first generation (T1) after integration of the T-DNA into the genome unless otherwise stated. All transgenic

lines were stably integrated and produced by the *Agrobacterium*-mediated floral dip method and subsequent selection for Basta- or Hygromycin-resistant plants. Biological replicates are non-overlapping pools of tissue collected from T1 transgenic plants. Double EVD transgenic lines (Fig. 6g,h) were generated by floral dipping T2 Basta-resistant 35S:EVD plants with Hygromycin-resistant EVD-only and T3A-EVD transgenes. Screening and genotyping in the next generation resulted in T3 35S:EVD + T1 EVD-only or T1 T3A-EVD plants. To create the *dcl1/2/3/4* 35S:EVD transgenic plants, we transformed a line that was homozygous *dcl2/3/4* and heterozygous for the *dcl1-9* allele. Resulting transformants were genotyped for DCL1, and divided into those that were homozygous mutants and those that were DCL1 heterozygotes.

**Transgene generation.** Transgenes were generated from primers listed in Supplementary Table 1. The EVD coding sequence (At5TE20395) was directly amplified from wt Col DNA using primers containing appropriate plasmid homology for In-Fusion Cloning (Clontech) into their respective digested binary vector backbones. Both pB2GW7 (Basta selection) and pH2GW7 (Hygromycin selection) digested with SpeI/HindIII were used for 35S:EVD, as some mutants that were transformed with this transgene already contain a Basta resistance cassette. pH2GW7 digested with SacI/HindIII was used for EVD only. For T3A-EVD, the T3A terminator and EVD coding sequence were separately amplified and then joined together using overlapping PCR before In-Fusion Cloning into pH2GW7 digested with SacI/HindIII.

The inverted repeat regions of transgenes were synthesized by Thermo Fisher and cloned into the binary vector pEG100 via SpeI/XhoI restriction digest followed by ligation using T4 DNA Ligase (NEB). The SQK1 inverted repeat contains bases AtChr3 3895083:3894528 (order reversed to prevent open reading frame) followed by the PDK intron and then AtChr3 33894528:3895083 (resulting plasmid named pMJS064). The RHD6 inverted repeat was fashioned similarly containing bases AtChr1 24795653:24795105. See Supplementary Table 1 for inverted repeat sequences. A Hygromycin-resistant version of each of these transgenes was created by amplifying the 2× 35S promoter: Hygromycin b phosphotransferase: 35S terminator cassette from the pDIRECT\_21B vector using primers from Supplementary Table 1 to add MluI and SacI restriction sites before being digested and ligated into the inverted repeat-pEG100 vectors in place of the Basta cassette.

The *sqk1-1* allele was created using the egg-cell promoter/enhancer CRISPR-Cas9 system described by ref. 41. The two-gRNA cassette was created using a four primer overlapping PCR (Supplementary Table 1) to amplify the 20 nt gRNA: U6-26 terminator: U6-29 promoter: 20 nt gRNA from the pCBC-DT1T2 plasmid<sup>42</sup>. This PCR product was cloned via BsaI Golden Gate reaction into a version of the pHEE401E plasmid containing an additional NapA promoter: dsRED: Nos terminator at the MfeI restriction site. The dsRED cassette was amplified out of the Traffic Lines<sup>43</sup> (primers in Supplementary Table 1) and provides a method for negative selection of the Cas9 transgene. T1 transformants were selected on Hygromycin plates and screened for SQK1 promoter deletion via PCR (primers in Supplementary Table 1).

To create the dCas9+SQK1-IR system, the pHEE401E plasmid was digested with NcoI and EcoRI to remove the active Cas9 cassette. The dCas9 sequence was amplified from the pDIRECT\_21B plasmid<sup>44</sup>. The Rps5a promoter for dCas9 was synthesized by Thermo Fisher and cloned into the pHEE401E backbone via In-Fusion reaction (Clontech). A gRNA cassette was created and inserted via Golden Gate reaction as described above using primers in Supplementary Table 1. dCas9 targets SQK1 at the sequence ATACTCAAAAATTAATAGA. The gRNA-dCas9 complex was amplified (primers in Supplementary Table 1) and In-Fusion cloned (Clontech) into a SacI digested pMJS064 (described above) that contains the SQK1-IR. The Rbs terminator for dCas9 was amplified from pHEE401E and In-Fusion cloned into the pMJS064 + dCas9 plasmid following digestion with AvrII (resulting plasmid named pMJS082). The dCas9 construct without the SQK1-IR was created by removing the SQK1-IR from pMJS082 via restriction digest with SpeI and XhoI.

**Expression analysis.** RNA was extracted from three biological replicates using TRIzol reagent (Thermo Fisher) and then DNase I treated and cleaned-up with the TURBO DNase DNA-free kit (Thermo Fisher) according to the manufacturer's protocol. Complementary DNA synthesis was performed using an oligo d(T) primer and Tetro Reverse Transcriptase (Bioline). We performed RT-qPCR using the primers in Supplementary Table 1 and SYBR green supermix. Relative expression was determined using the 2<sup>-ΔΔCt</sup> method, comparing the gene of interest to a housekeeping control gene (*AXS2* or *ACTIN 2*, Supplementary Table 1). Mean and standard deviations of three biological replicates are calculated by Graphpad Prism and shown as bar graphs with error bars.

**BSAS and analysis.** Genomic DNA was treated with RNase A, then cleaned and recovered by phenol chloroform isolation. Then 500 pg–2 μg DNA was treated with sodium bisulfite using the EZ DNA Methylation kit (Zymo Research). Amplicons for sequencing were generated by PCR using degenerate primers (Supplementary Table 1) for each locus using My Taq HS mix (Bioline) and gel purification. For each bisulfite conversion reaction performed, we also performed PCR on a control unmethylated locus (At2G20610) to calculate the conversion rate (see Supplementary Fig. 1 for an example).

To generate amplicon libraries, purified amplicons were pooled in equimolar ratios. Here, 100 ng of purified amplicon DNA in 30 μl volume were used for library preparation with the Illumina Nextera DNA Flex kit. BSAS libraries from December 2019 to present were created using a modified Nextera Flex protocol (first described by ref. 45). Per index, 1 μl of Tagmentation (BLT) beads from the Nextera Flex kit was diluted with 19 μl of ultrapure water and combined with 30 μl of sample DNA and 50 μl of laboratory-made 2× tagmentation buffer (20 mM Tris, 20 mM MgCl<sub>2</sub>, 50% DMF) before being tagmented at 55 °C for 15 min. Tagmentation was stopped by adding 20 μl of 0.2% SDS and incubating at 37 °C for 15 min. Beads were washed three times with 100 μl of laboratory-made tagmentation wash buffer (10% PEG8000 and 0.25 M NaCl in TE buffer). Tagmented DNA was amplified directly from the beads via six cycles of PCR amplification using the PrimeSTAR GXL Polymerase kit (Takara) and dual-indexed adapters (Illumina) in 45 μl reactions. Finally, libraries were purified using 81 μl of SPRIselect beads (Beckman Coulter) or Nextera Purification Beads (Illumina), washed twice in 200 μl of 80% EtOH and eluted off beads in 32 μl of ultrapure water.

We sequenced the resulting libraries with 300-nt single-end reads on the Illumina MiSeq platform at the University of Delaware DNA Sequencing and Genotyping Center. Raw reads were trimmed for adapters and mapped to all the amplicons pooled together for sequencing using *methylpy* (<https://github.com/yupenghe/methylpy>)<sup>46</sup>. Reference DNA sequences have primer sequences removed. The allC files (Data and Code availability) containing methylation data for each cytosine were used as input for Bedtools<sup>47</sup> to calculate methylation percentage for each locus and plotted using ggplot2 in R. The 95% Wilson confidence intervals were calculated as in ref. 48 for error bars. BSAS data of DNA methylation levels were quantified per amplicon as the average methylation at each cytosine sequence context (CG, CHG, CHH). The total number of reads reporting the methylation status of a locus is noted as BSAS coverage in Supplementary Fig. 1. Data were analysed in R and plotted with ggplot2. In the data display, the bar displays the methylation percentage and the error bars represent the 95% confidence interval calculated using the Wilson score interval method. 'n' denotes the total number of cytosines, which is calculated as the sum of read coverage for all assayed cytosines in an amplicon.

**Bisulfite Sanger sequencing and analysis.** Amplicons for Sanger sequencing from Supplementary Fig. 1 were generated as above for BSAS. Purified amplicons were subjected to single colony purification by TOPO TA cloning into pCR4 (Thermo Fisher) and transformation into *Escherichia coli*. Individual colonies were sequenced by Sanger sequencing (Eton Biosciences) and analysed in Kismeth<sup>49</sup> using default parameters.

**Analysis of whole-genome bisulfite sequencing data.** Genome-wide MethylC-seq data are publicly available for wt Col *Arabidopsis* inflorescence<sup>50</sup>. Processed data were downloaded from GEO (GSM2101949). Similar to BSAS analysis, the allC file was used as input for Bedtools<sup>47</sup> to calculate methylation percentage for each locus shown in Supplementary Fig. 1.

**Determination of RdDM strength.** To measure RdDM strength, we calculated the number of consecutive cytosines that are methylated for a given locus per sequencing read. Using methylpy<sup>46</sup> we isolated the reads that mapped to a specific locus and used these as input for Kismeth<sup>49</sup>. The Kismeth output displays an image of methylation status of each individual cytosine along each individual read, and this was used for image analysis. We used a custom python script (<https://github.com/jpeasari/Dot-Plot-Analysis-OpenCV>) to analyse each read in the image, represented by one row of methylation data and determined the longest stretch of consecutive methylated cytosines in each row. We summarized the longest stretch counts for each locus in box plots using ggplot2 in R. Box plots of RdDM strength represent 25th and 75th percentile values with whiskers at the 10th and 90th percentiles, the median is represented by a line and the mean is denoted by a filled circle.

**Small RNA sequencing and analysis.** Small RNA was sequenced as in ref. 17. Briefly, Trizol reagent (Thermo Fisher) was used to isolate total RNA. The mirVana microRNA isolation kit (Thermo Fisher) was used to enrich small RNAs. The TruSeq Small RNA Library Preparation kit (Illumina) was used to generate libraries and multiplexed for sequencing on a HiSeq 4000, HiSeq X or NextSeq 550 system at the University of Delaware DNA Sequencing and Genotyping Center or Novogene Inc.

Postsequencing, the Illumina universal adapter was removed from the demultiplexed libraries using the fastx toolkit. Total genome matching reads were calculated using bowtie (parameters: -v 0), and this value is used for library depth normalization. The sRNA Workbench<sup>51</sup> was used to filter out transfer/ribosomal RNA reads, low complexity reads and retain only 18–28-nt small RNA reads that map the *Arabidopsis* Araport11 genome. To map the small RNAs to the genome, ShortStack<sup>52</sup> was used with parameters: -nohp -mmap f -bowtie\_m all -align\_only. For assaying siRNAs from a specific transgene, bowtie (parameter: -v 0 -best -strata -M 1) was used to map the small RNAs to the full transgene sequence. Bedtools was used to count the number of reads mapping a specific locus. In Fig. 3d, replicate 2 of *ago4/6/9* contains the 35S:EVD transgene (from Fig. 2b). The small RNAs

mapping the transgene were removed before analysis for Fig. 3d. In Fig. 4a, de novo clusters of 23–24-nt siRNAs were called using ShortStack with a threshold of at least ten raw reads in wt Col and Pol V-dependence was determined by at least a  $\geq 2$ -fold loss of siRNAs in *pol V* mutants. Conversely, loci that did not lose siRNAs in *pol V* compared to wt Col were categorized as Pol V-independent siRNA loci. *ggplot2* in R was used to generate siRNA heatmaps.

**AGO4-incorporated small RNA library preparation and analysis.** Frozen inflorescence tissue was ground with liquid nitrogen and resuspended in lysis buffer (50 mM Tris pH 8, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 10% glycerol, 1% IGEPAL, 0.5 mM DTT, 1 mM PMSF, 1× Roche protease inhibitor cocktail) and homogenized with mixing for 15 min at 4°C. Lysates were then clarified with a spin. Clarified lysates were combined with 2 µl of AGO4 antibody (Agriser) or 2 µl of rabbit IgG as a mock-IP (Cell Signaling Technology) and rotated at 4°C for 1 h. Immune complexes were harvested with 40 µl of Protein G Dynabeads (Thermo), prewashed in 1× TBS, rotating 30 min at 4°C. Beads and immune complexes were washed three times with 1 ml of cold wash buffer (50 mM Tris pH 8, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.5 mM DTT). Immunoprecipitated small RNA (bound by AGO4) was released from beads and isolated by TRIsure (Bioline) extraction. All RNA recovered from AGO4 IPs went directly into small RNA library preparation.

The small RNA library was prepared using TruSeq Small RNA Library Preparation kit (Illumina) as described above for total small RNAs with the exception of using 14 cycles of PCR amplification. Small RNAs were processed exactly as described earlier and accumulation was calculated in reads per million (RPM) genome-mapped sequenced reads for each size class of small RNAs in AGO4-IP and mock-IP samples. The small RNA enrichment was calculated as the ratio of RPM values for AGO4-IP over mock samples for each size class. This enrichment value is displayed as a heatmap in Fig. 2c and Supplementary Fig. 6.

**Plasmid-safe PCR assay.** Frozen inflorescence tissue was ground with liquid nitrogen, and total DNA was purified using the QIAGEN DNeasy Plant Mini Kit. Then 1 µg of the resulting RNase-free DNA was digested using 1 µl of Plasmid-Safe ATP-dependent DNase (Lucigen) in a 50-µl reaction for 16 h at 37°C. Digestion was completed by twice adding 1 µl of additional DNase and 1 µl of additional ATP followed by 2 h of incubation at 37°C for a total of 20 h of digestion. DNase was inactivated at 70°C for 30 min. Digested and undigested DNA from each line was amplified using PCR primers for the EVD coding region (Supplementary Table 1).

**Western blotting.** Frozen inflorescence tissue was ground with liquid nitrogen and resuspended in lysis buffer (50 mM Tris pH 8, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 10% glycerol, 1% IGEPAL, 0.5 mM DTT, 1 mM PMSF, 1× Roche protease inhibitor cocktail) and homogenized with mixing for 15 min at 4°C. Lysates were clarified with a spin, combined with 2× loading buffer, denatured and then loaded onto a 4–20% gradient Tris-Glycine gel (Thermo). Protein was transferred from the gel to a polyvinylidene difluoride membrane using the BioRad semidry transblot. Membranes were blocked for 1 h at room temperature in 3% milk powder 1× PBS-T. Primary antibodies, which include Pol V (Wierzbicki laboratory), AGO4 (Agriser) and ACT11 (Agriser), were all diluted 1:1,000 in 3% milk 1× PBS-T solution and incubated on blots overnight. Washes were performed at room temperature with 1× PBS-T. Antirabbit secondary antibody (Sigma) was used for visualization of Pol V and AGO4, while antimouse secondary (Sigma) was used for ACT11. Blots were visualized using HRP chemiluminescence (Thermo), with exposures ranging from 5 s to 5 min.

**ChIP and qPCR.** Nuclei were crosslinked as follows: frozen inflorescences (300 mg per biorep) were ground with liquid nitrogen and resuspend in nuclear isolation buffer (10 mM HEPES, 1 M sucrose, 5 mM KCl, 5 mM MgCl<sub>2</sub>, 0.6% Triton X-100, 0.4 mM PMSF, 1× Roche protease inhibitor cocktail) and homogenized with mixing for 15 min at 4°C. Methanol-free formaldehyde (Pierce) was added to a final concentration of 1% with end-over-end mixing for 15 min at room temperature. Formaldehyde crosslinking was quenched with glycine (125 mM), rotating 5 min at room temperature. Crosslinked nuclei were then filtered through two layers of Miracloth to remove large particles before centrifugation at 3,000g for 15 min at 4°C. The resulting nuclear pellet was resuspended in wash buffer (10 mM Tris pH 8, 0.25 M sucrose, 10 mM MgCl<sub>2</sub>, 1 mM EDTA, 1% Triton X-100, 1× Roche protease inhibitor cocktail) and nuclei were cleaned and pelleted at 12,000g for 10 min at 4°C. The final clean nuclear pellet was resuspended in 1 ml of nuclear lysis buffer (20 mM Tris pH 8, 2 mM EDTA, 0.1% SDS, 1 mM PMSF, 1× Roche protease inhibitor cocktail) and sonicated on the Covaris E220 (150 W peak power, 20% duty factor, 200 cycles per burst, 6 min). Insoluble debris was removed from the sonicated soluble chromatin by centrifugation at 12,000g for 10 min at 4°C. Then 30 µl of 5 M NaCl and 20 µl of 30% Triton X-100 were both added to 920 µl of sonicated chromatin to sequester SDS before IP. Then 3% input by volume was set aside for each sample, and the remaining volume of sonicated chromatin was divided evenly among IPs and IgG negative controls for overnight IP with respective antibodies or IgG.

Pol V (Lagrange laboratory antibody) and Ser5P Pol II (Abcam) ChIPs were performed with 2 µl of antibody, whereas AGO4 (Agriser) and H3K9me2 (Abcam) ChIPs used 5 µl of antibody per IP. Each experiment used the same

volume of rabbit IgG as a negative control (Cell Signaling Technology). Immune complexes were collected using 40 µl of washed protein A/G magnetic beads (Pierce), rotating at 4°C for 2 h. After collecting immune complexes, beads were washed at 4°C using one rinse and two 5-min washes of each of the following buffers: low salt (150 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris pH 8), high salt (500 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris pH 8), LiCl buffer (250 mM LiCl, 1% Igepal, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris pH 8) and TE + 0.1% Igepal. Chromatin was eluted from the beads using 250 µl of elution buffer (1% SDS, 0.1 M NaHCO<sub>3</sub>) at 65°C for 15 min with agitation. Overnight reverse crosslinking was accomplished with the addition of 20 µl of 5 M NaCl at 65°C for all samples including inputs. Proteinase K digestion was performed by the addition of 10 µl of 0.5 M EDTA, 20 µl of 1 M Tris pH 7, 20 µg Proteinase K (Thermo Scientific) and 50 ng RNase A, incubating at 42°C for 1 h. DNA was then purified using DNA Clean and Concentrator-25 columns (Zymo Research) and eluted in 100 µl of elution buffer.

qPCR was performed using primers in Supplementary Table 1 and Sso Universal SYBR (BioRad). Percentage input was calculated by first normalizing the Ct values of the diluted input samples to 100% input by the following calculation:  $Ct_{100\%} = Ct_{diluted} - (\log(\text{dilution factor}, 2))$ . Then the IgG and IP Ct values were normalized to this new 100% input Ct value using the  $2^{-\Delta\Delta Ct}$  method, which represents percentage input. For Pol V and AGO4 ChIP, the *pol V* mutant and *ago4/6/9* mutants (*ago4-4* allele, a complete null) were, respectively, used to calculate background levels for each of these antibodies. These background levels were averaged for each PCR target. They were then subtracted from the percentage input values of each of the other genotypes. Mean and standard deviations of the biological replicates are shown as bar graphs with error bars.

**ChIP-seq and RIP-seq data analysis.** Raw reads were downloaded from National Center for Biotechnology Information (NCBI) GEO (GSE52041 and GSE70290), trimmed for adapters and mapped to the Araport11 *Arabidopsis* genome. The ChIP-seq reads were mapped using Shortstack (parameters: `-nohp -mmap f -bowtie_m all`). The RIP-seq reads were mapped using SoapSplice v.1.10 (ref. 33) using parameters: `-t 10300` (maximum distance between two segments) as mentioned in the original study<sup>9</sup>. For Fig. 3a, reads were counted using Bedtools for each individual locus. To determine enrichment, the normalized ratio of IP over mock sample reads was calculated (for both ChIP-seq and RIP-seq). In eukaryotes, organelle-to-nucleus DNA transfer is known to occur, which has resulted in some mitochondrial genes being duplicated and part of the nuclear genome. To avoid such genes causing an artefact of Pol V enrichment in the negative control dataset, we analysed only those mitochondrial genes that do not have any similarity to nuclear genes.

In Fig. 3c, genome-wide Pol V enriched loci were determined using the *Macs2* ChIP peak caller<sup>34</sup>. Pol V fold enrichment was calculated using the ratio of IP over mock samples. The two replicates were averaged for wt Col and *svuh2/svuh9* mutant. The loci that had  $\geq 2$ -fold Pol V enrichment in wt Col and only background levels of Pol V enrichment in *svuh2/svuh9* were categorized as SUVH2/SUVH9 fully dependent Pol V loci. The loci that retained at least  $\geq 2$ -fold Pol V enrichment in *svuh2/svuh9* mutants were categorized as SUVH2/SUVH9 partially dependent Pol V loci. In Fig. 4c, AGO4 enrichment was calculated as a ratio of IP over mock normalized read accumulation. For genome-wide analyses, the two biological replicates were averaged, whereas the replicates are also displayed individually in Fig. 3a.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Raw Illumina sequencing data produced for this study is available without restriction from NCBI as GSE165575. Additional small RNA datasets were downloaded from GSE118705. Genome-wide MethylC-seq data is publicly available for wt Col *Arabidopsis* inflorescence<sup>30</sup>. Processed data were downloaded from GEO (GSM2101949). ChIP-seq and RIP-seq reads were downloaded from NCBI GEO (GSE52041 and GSE70290). Processed BSAS data of DNA methylation levels are available as Supplementary Dataset 1. Sanger sequencing results are available as Supplementary Dataset 2. Biological materials can be obtained from the corresponding author without restriction.

## Code availability

The custom script used to generate RdDM strength plots is available on Github at <https://github.com/jpeasari/Dot-Plot-Anaysis-OpenCV>.

Received: 12 March 2021; Accepted: 9 September 2021;  
Published online: 8 November 2021

## References

- Deniz, Ö., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* **20**, 417–431 (2019).

2. Morel, J.-B., Mourrain, P., Béclin, C. & Vaucheret, H. DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in *Arabidopsis*. *Curr. Biol.* **10**, 1591–1594 (2000).
3. Hamilton, A., Voinnet, O., Chappell, L. & Baulcombe, D. Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21**, 4671–4679 (2002).
4. Thomas, A. L. et al. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Gene Dev.* **27**, 390–399 (2013).
5. Zhang, H., Lang, Z. & Zhu, J.-K. Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Bio.* **19**, 489–506 (2018).
6. Havecker, E. R. et al. The *Arabidopsis* RNA-directed DNA methylation Argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* **22**, 321–334 (2010).
7. Cao, X. et al. Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* **13**, 2212–2217 (2003).
8. Wierzbicki, A. T., Haag, J. R. & Pikaard, C. S. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**, 635–648 (2008).
9. Böhmendorfer, G. et al. Long non-coding RNA produced by RNA Polymerase V determines boundaries of heterochromatin. *eLife* **5**, e19092 (2016).
10. Ream, T. S. et al. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol. Cell* **33**, 192–203 (2009).
11. Liu, Z.-W. et al. The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. *PLoS Genet.* **10**, e1003948 (2014).
12. Johnson, L. M. et al. SRA- and SET-domain-containing proteins link RNA Polymerase V occupancy to DNA methylation. *Nature* **507**, 124–128 (2014).
13. Zheng, B. et al. Intergenic transcription by RNA Polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in *Arabidopsis*. *Gene Dev.* **23**, 2850–2860 (2009).
14. Li, S. et al. Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in *Arabidopsis* reveals features and regulation of siRNA biogenesis. *Genome Res* **25**, 235–245 (2015).
15. Tsuzuki, M. et al. Broad noncoding transcription suggests genome surveillance by RNA Polymerase V. *Proc. Natl Acad. Sci. USA* **117**, 30799–30804 (2020).
16. Bond, D. M. & Baulcombe, D. C. Epigenetic transitions leading to heritable, RNA-mediated de novo silencing in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **112**, 917–922 (2015).
17. Nuthikattu, S. et al. 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 (2013).
18. Fultz, D. & Slotkin, R. K. Exogenous transposable elements circumvent identity-based silencing, permitting the dissection of expression-dependent silencing. *Plant Cell* **29**, 360–376 (2017).
19. Mari-Ordóñez, A. et al. Reconstructing de novo silencing of an active plant retrotransposon. *Nat. Genet.* **45**, 1029–1039 (2013).
20. Masser, D. R., Berg, A. S. & Freeman, W. M. Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenet Chromatin* **6**, 33 (2013).
21. Pontier, D. et al. NERD, a plant-specific GW protein, defines an additional RNAi-dependent chromatin-based pathway in *Arabidopsis*. *Mol. Cell* **48**, 121–132 (2012).
22. Xie, Z. et al. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, e104 (2004).
23. Onodera, Y. et al. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**, 613–622 (2005).
24. Gao, Z. et al. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* **465**, 106–109 (2010).
25. Wierzbicki, A. T., Ream, T. S., Haag, J. R. & Pikaard, C. S. RNA Polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat. Genet.* **41**, 630–634 (2009).
26. Qi, Y. et al. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**, 1008–1012 (2006).
27. Zheng, Q. et al. RNA Polymerase V targets transcriptional silencing components to promoters of protein-coding genes. *Plant J.* **73**, 179–189 (2013).
28. Béclin, C., Boutet, S., Waterhouse, P. & Vaucheret, H. A branched pathway for transgene-induced RNA silencing in plants. *Curr. Biol.* **12**, 684–688 (2002).
29. Zilberman, D. et al. Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr. Biol.* **14**, 1214–1220 (2004).
30. Sasaki, T. et al. Distinct and concurrent pathways of Pol II- and Pol IV-dependent siRNA biogenesis at a repetitive trans-silencer locus in *Arabidopsis thaliana*. *Plant J.* **79**, 127–138 (2014).
31. Butaye, K. M. J. et al. Stable high-level transgene expression in *Arabidopsis thaliana* using gene silencing mutants and matrix attachment regions. *Plant J.* **39**, 440–449 (2004).
32. Lahmy, S. et al. Evidence for ARGONAUTE4–DNA interactions in RNA-directed DNA methylation in plants. *Gene Dev.* **30**, 2565–2570 (2016).
33. Pikaard, C. S., Haag, J. R., Pontes, O. M. F., Blevins, T. & Cocklin, R. A transcription fork model for Pol IV and Pol V-dependent RNA-directed DNA methylation. *Cold Spring Harb. Symp.* **77**, 205–212 (2012).
34. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
35. Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408 (2014).
36. Baduel, P. et al. Genetic and environmental modulation of transposition shapes the evolutionary potential of *Arabidopsis thaliana*. *Genome Biol.* **22**, 138 (2021).
37. Papikian, A., Liu, W., Gallego-Bartolomé, J. & Jacobsen, S. E. Site-specific manipulation of *Arabidopsis* loci using CRISPR-Cas9 SunTag systems. *Nat. Commun.* **10**, 729 (2019).
38. Lister, R. et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
39. Rowley, M. J., Avrutsky, M. I., Sifuentes, C. J., Pereira, L. & Wierzbicki, A. T. Independent chromatin binding of ARGONAUTE4 and SPT5L/KTF1 mediates transcriptional gene silencing. *PLoS Genet.* **7**, e1002120 (2011).
40. Quintero-Cadena, P., Lenstra, T. L. & Sternberg, P. W. RNA Pol II length and disorder enable cooperative scaling of transcriptional bursting. *Mol. Cell* **79**, 207–220.e8 (2020).
41. Wang, Z.-P. et al. Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in *Arabidopsis* in a single generation. *Genome Biol.* **16**, 144 (2015).
42. Xing, H.-L. et al. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* **14**, 327 (2014).
43. Wu, G., Rossidivito, G., Hu, T., Berlyand, Y. & Poethig, R. S. Traffic lines: new tools for genetic analysis in *Arabidopsis thaliana*. *Genetics* **200**, 35–45 (2015).
44. Čermák, T. et al. A multipurpose toolkit to enable advanced genome engineering in plants. *Plant Cell* **29**, 1196–1217 (2017).
45. Gaio, D. et al. Hackflex: low cost Illumina sequencing library construction for high sample counts. Preprint at *bioRxiv* <https://doi.org/10.1101/779215> (2019).
46. Schultz, M. D. et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015).
47. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
48. Henderson, I. R. et al. The de novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in *Arabidopsis thaliana*. *PLoS Genet.* **6**, e1001182 (2010).
49. Gruntman, E. et al. Kismeth: analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinf.* **9**, 371 (2008).
50. Panda, K. et al. Full-length autonomous transposable elements are preferentially targeted by expression-dependent forms of RNA-directed DNA methylation. *Genome Biol.* **17**, 170 (2016).
51. Stocks, M. B. et al. The UEA sRNA Workbench (version 4.4): a comprehensive suite of tools for analyzing miRNAs and sRNAs. *Bioinformatics* **34**, 3382–3384 (2018).
52. Johnson, N. R., Yeoh, J. M., Coruh, C. & Axtell, M. J. Improved placement of multi-mapping small RNAs. *G3 Genes Genomes Genet.* **6**, 2103–2111 (2016).
53. Huang, S. et al. SOAPsplice: genome-wide ab initio detection of splice junctions from RNA-seq data. *Front. Genet.* **2**, 46 (2011).
54. Gaspar, J. M. Improved peak-calling with MACS2. Preprint at *bioRxiv* <https://doi.org/10.1101/496521> (2018).

## Acknowledgements

We thank D. Cuerda-Gil, S. Edwards and P. Liu for generating small RNA sequencing libraries. We thank T. Lagrange, C. Pikaard and A. Wierzbicki for sharing NRPE1 seed and antibody resources, M. Axtell for *ago4-4/6/9* seed, Y. Qi for *dcl1/2/3/4* seed and R. Mosher for helpful comments. We thank J. Mele at The Ohio State Genomics Shared Resource, the University of Delaware DNA Sequencing and Genotyping Center and the Donald Danforth Plant Science Center Data Science Facility for computing support. We also thank the Donald Danforth Plant Science Center Plant Growth Facility, G. Posey at the Center for Applied Plant Sciences Greenhouse, and E. Yoders-Horn and D. Snodgrass at the Biological Sciences Greenhouse. M.J.S. was supported as a fellow of The Ohio State Cellular, Molecular and Biochemical Sciences Training Program (grant no. NIH T32 GM086252). This work is supported by National Science Foundation Federation grant no. MCB-1908521 to R.K.S.

## Author contributions

The concept was developed by M.J.S., K.P., R.K.S. and A.D.M. Methodology was developed by M.J.S., K.P., R.K.S. and A.D.M. Software was developed by J.R.P. and K.P. The investigation was carried out by M.J.S., R.K., L.L.M., H.P., K.P. and A.D.M. Resources were obtained by A.Y.H. Data curation was the responsibility of K.P., M.J.S., K.P., R.K.S. and A.D.M. wrote the paper. Funding was acquired by R.K.S.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41477-021-01008-7>.

**Correspondence and requests for materials** should be addressed to R. Keith Slotkin.

**Peer review information** *Nature Plants* thanks David Baulcombe, Zhaobo Lang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.  
© The Author(s) 2021

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |     |           |
|-----|-----------|
| n/a | Confirmed |
|-----|-----------|
- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
  - A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
  - The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
  - A description of all covariates tested
  - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
  - A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
  - For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
  - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
  - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
  - Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection Confirmed - Described in the manuscript.

Data analysis Confirmed - Described in the manuscript.  
MethylPy v1.4.6 (<https://github.com/yupenghe/methylpy>), BedTools v2.27.1, R v4.0.2, Kismeth ((v1.1.0) <http://katahdin.mssm.edu//kismeth/revpage.pl>), Bowtie v1.2.3, SoapSplice 1.10, ShortStack v3.8.5, Macs2 v2.2.6, GraphPad Prism 9.0.0 were used for processing of data. The custom script used to generate strength of RdDM plots is available on Github - <https://github.com/jpeasari/Dot-Plot-Anaysis-OpenCV>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

These have been provided in the manuscript. Raw Illumina sequencing data produced for this study is available without restriction from NCBI as GSE165575. Additional small RNA datasets were downloaded from GSE118705. Genome-wide MethylC-seq data is publicly available for wt Col Arabidopsis inflorescence50. Processed data was downloaded from GEO (GSM2101949). CHIP-seq and RIP-seq reads were downloaded from NCBI GEO (GSE52041 and GSE70290). Processed BSAS data of DNA methylation levels is available as Supplemental Dataset 1. Sanger sequencing results are available as Supplemental Dataset 2.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size depends on which experiment. For qRT-PCR we use 3 biological replicates per genotype in order to be able to test significance. For ChIP, sample size is determined by how many samples were able to be grown and collected and how many samples the researcher can process at the same time. These sample sizes are sufficient for ChIP experiments because they either allow for significance testing or the experiments have been repeated independently. Statistical analysis was not used to determine sample size for any experiment.
Data exclusions	No data was excluded.
Replication	Biological replicates and replicate experiments where used. Entire experimental replication information has been added to the manuscript.
Randomization	Randomization was not applicable to the study.
Blinding	Blinding was not applicable to the study.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	Information is provided in the manuscript. Antibodies used include Pol V (custom-made antibody provided as a gift from the lab of Thierry Lagrange, used at 1:1000 dilution), AGO4 (Agrisera cat. no. AS09-617, lot # 1602, used at 1:1000 dilution), ACT11 (Agrisera cat. no. AS10-702, lot # 1105, used at 1:1000 dilution), Pol II Ser5P (Abcam cat. no. ab5131, lot # GR3264297-1), H3K9me2 (Abcam cat. no. ab1220), Rabbit IgG (Cell Signaling Technology cat. no. 2729, lot #10), and Protein A/G magnetic beads (Thermo cat. no. 88802, lot #VK308442)
Validation	The commercially available antibodies have been validated on their respective manufacturers websites. The Pol V antibody was validated in a previous manuscript (Lahmy et al., Genes & Development 2006), as well as in Supplemental Figure 7E using a pol V mutant plant.

## ChIP-seq

### Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](https://www.ncbi.nlm.nih.gov/geo/).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

### Data access links

May remain private before publication.

All ChIP-seq data in this manuscript are reanalyses of publicly available data. The GEO links and data availability are provided in the manuscript. We will add the BED files for called peaks to the potential revision.



Files in database submission	All the relevant files for the ChIP-seq analysis have already been uploaded to the correct NCBI repository by the authors that originally created that data.
Genome browser session (e.g. <a href="#">UCSC</a> )	N/A

## Methodology

Replicates	All ChIP-seq data in this manuscript is reanalysis of publicly available data.
Sequencing depth	All ChIP-seq data in this manuscript is reanalysis of publicly available data.
Antibodies	All ChIP-seq data in this manuscript is reanalysis of publicly available data.
Peak calling parameters	ChIP-seq reads were mapped using ShortStack parameters: --nohp --mmap f --bowtie_m all. The RIP-seq reads were mapped using SoapSplice 1.10 using parameters: -t 10300. Peaks were called with Macs2 using default parameters.
Data quality	We displayed all biological replicates to ensure reproducibility. We also used new negative controls of mitochondrial genes.
Software	MethylPy v1.4.6, BedTools v2.27.1, R v4.0.2, Kismeth, Bowtie v1.2.3, Samtools 1.9, SoapSplice 1.10, ShortStack v3.8.5, FastX toolkit v0.0.13, UEA small RNA workbench v4.5, Macs2 v2.2.6