RNA-directed DNA methylation involves co-transcriptional small RNA-guided slicing of Pol V transcripts in *Arabidopsis.*

Wanlu Liu[1, 2*], Sascha H. Duttke[3,4,5,6*], Jonathan Hetzel[3,4,5], Martin Groth[2], Suhua Feng[2,7], Javier Gallego-Bartolome[2], Zhenhui Zhong[2,8], Hsuan Yu Kuo[2], Jixian Zhai[2,9], Joanne Chory[3,4,5], Steven E. Jacobsen[1,2,7, 10, †]

[1.] Molecular Biology Institute, University of California, Los Angeles, CA 90095.

[2.] Department of Molecular, Cell and Developmental Biology, University of California at Los Angeles, Los Angeles, CA 90095.

[3.] Plant Biology Laboratory, Salk Institute for Biological Studies, 10010 N Torrey Pines RD, La Jolla, CA 92037.

[4.] Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92093.

[5.] Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, CA 92037.

[6.] Department of Cellular & Molecular Medicine, School of Medicine, University of California at San Diego, La Jolla, CA, 92093.

[7.] Eli & Edythe Broad Center of Regenerative Medicine & Stem Cell Research, University of California at Los Angeles, Los Angeles, CA 90095, USA.

[8.] State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

[9.] Institute of Plant and Food Science, Department of Biology, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China.

[10.] Howard Hughes Medical Institute, University of California at Los Angeles, Los Angeles, CA 90095, USA.

*** These authors contributed equally to this work.**

† Corresponding author. Email: jacobsen@ucla.edu

Abstract

Small RNAs regulate chromatin modifications such as DNA methylation and gene silencing across eukaryotic genomes. In plants, RNA-directed DNA methylation (RdDM) requires 24-nucleotide (nt) small RNAs (siRNAs) that bind ARGONAUTE4 (AGO4) and target genomic regions for silencing. It also requires non-coding RNAs transcribed by RNA POLYMERASE V (Pol V) that likely serve as scaffolds for binding of AGO4/siRNA complexes. Here we utilized a modified global nuclear run-on (GRO) protocol followed by deep sequencing to capture Pol V nascent transcripts genome-wide. We uncovered unique characteristics of Pol V RNAs, including a uracil (U) common at position 10. This uracil was complementary to the 5' adenine found in many AGO4-bound 24-nt siRNAs and was eliminated in a siRNA-deficient mutant as well as in the *ago4/6/9* triple mutant, suggesting that the +10U signature is due to siRNA-mediated co-transcriptional slicing of Pol V transcripts. Expressing wild-type AGO4 in *ago4/6/9* was able to restore slicing of Pol V transcripts but a catalytically inactive AGO4 mutant did not correct the slicing defect. We also found that Pol V transcript slicing required the little understood elongation factor SPT5L. These results highlight the importance of Pol V transcript slicing in RNA-mediated transcriptional gene silencing, which is a conserved process in many eukaryotes.

Introduction

DNA methylation is an evolutionarily conserved epigenetic mark associated with gene silencing that plays a key role in diverse biological processes. In plants, DNA methylation is mediated by small RNAs that target specific genomic DNA sequences in a process known as RNA-directed DNA methylation (RdDM). RdDM involves RNA polymerase (Pol) IV and Pol V, both of which evolved from Pol II, and plays crucial roles in transposon silencing and maintenance of genome integrity [1]. The current model for RdDM involves several sequential steps. First, Pol IV initiates the biogenesis of siRNAs by producing 30- to 40-nt ssRNA [2-4]. These ssRNAs are then made double stranded by RNA-dependent RNA polymerase 2 (RDR2) [5,6], processed into 24-nt siRNA by DCL3 [7], and loaded into the effector protein AGO4 [8-10]. A second set of non-coding transcripts, generated by Pol V, has been proposed to serve as a targeting scaffold for the binding of AGO4-associated siRNAs through sequence complementarity [11]. Ultimately, AGO4 targeting recruits the DRM2 DNA methyltransferase to mediate *de novo* methylation of cytosines in all sequence contexts (CG, CHG, and CHH, where H represents A, C, or T) [12]. Pol V is required for DNA methylation and silencing, and has been shown to be transcriptionally active *in vitro*. A recent study of RNAs co-immunoprecipitation (RIP) with Pol V showed Pol V-associated RNAs at thousands of locations in the genome [13]. However, shearing was used in the library preparation protocol, which meant that many features of the individual Pol V transcripts were lost [13]. Thus, several characteristics of Pol V transcripts and how they mediate RdDM remain poorly characterized [11,14].

Identification of nascent Pol V transcripts genome-wide.

To enable a detailed analysis of Pol V transcripts at single nucleotide resolution, we used a modified global nuclear run-on assay [15,16] followed by deep sequencing (GRO-seq) in *Arabidopsis* (Fig. 1a). This technique captures nascent RNA from engaged RNA polymerases in a strand specific manner. Uniquely mapping paired end reads were obtained from two independent experiments (Supplementary Fig. 1a) prepared from wild-type Columbia (Col-0) plants (Table S1). GRO-seq captures transcriptionally engaged RNA polymerases [15,16], and although we selected against full length capped Pol II transcripts (Fig. 1a), we still observed a background level of signal over Pol II transcribed protein-coding genes. Thus, in order to specifically identify Pol V-dependent nascent transcripts, we also performed GRO-seq in a Pol V mutant (*nrpe1*) as well as in a Pol IV/Pol V double mutant (*nrpd1/e1*). We coupled this with a genome-wide map of the chromatin association profile of Pol V, using ChIP-seq with an endogenous antibody against NRPE1, the largest catalytic subunit of Pol V. Combining Pol V ChIP-seq and GRO-seq in Col-0, *nrpe1*, and *nrpd1/e1*, we identified GRO-seq reads that mapped to Pol V regions, including those at previously defined individual Pol V intergenic non-coding (IGN) transcripts [11](Fig. 1b). As expected, we found that GRO-seq signals generated from Pol V occupied regions were largely eliminated in the *nrpe1* mutant, while signals over mRNA regions in the *nrpe1* mutant remained unchanged (Supplementary Fig. 1b,c), confirming that we had indeed identified Pol V-dependent nascent transcripts. In addition to the tight spatial co-localization of Pol V ChIP-seq and GRO-seq signals, we also observed a positive correlation between the two in signal intensity (Supplementary Fig. 1d).

However, Pol V-dependent GRO-seq signals were much more narrowly defined compared to signals from Pol V ChIP-seq, thereby providing a higher resolution view of Pol V transcription (Fig. 1c). Unlike Pol II transcripts, which are primarily transcribed from one strand (Fig. 1b, Fig. 2a), Pol V-dependent transcripts were present roughly equally on both strands (Fig. 1b, Fig. 2b). RdDM has been shown to be enriched at short transposons as well as at the edges of long transposons [17]. Consistent with Pol V occupancy at long transposon edges [18], we found that Pol V-dependent GRO-seq transcripts were also preferentially localized over those regions (Fig. 2c, Supplementary Fig. 1e).

To investigate the relationship between Pol IV activity and Pol V transcript production, we performed Pol V ChIP-seq and GRO-seq in the *nrpd1* mutant, which specifically eliminates Pol IV activity. Although many Pol V transcripts were eliminated in the *nrpd1* mutant (Supplementary Fig. 2a), most remained (Supplementary Fig. 2b). Based on whether or not Pol V ChIP-seq signal remained in *nrpd1*, we classified Pol V regions into Pol IV/V-codependent regions (1,903 sites) or Pol IV-independent Pol V regions (2,365 sites) (Table S2). As expected, both the GRO-seq signal and the Pol V ChIP-seq signal were largely eliminated in *nrpd1* at Pol IV/V-codependent sites, while the signals at Pol IV-independent sites largely remained (Supplementary Fig. 2c,d).

The reason that some Pol V transcripts are dependent on Pol IV activity is likely because the RdDM pathway is a self-reinforcing loop [1]. For example, although Pol V

is required for DNA methylation and silencing, Pol V recruitment to chromatin requires preexisting DNA methylation via the methyl DNA binding proteins SUVH2 and SUVH9 [19]. We therefore hypothesized that the reason that Pol IV is required for Pol V activity at only some genomic sites is because it plays a larger role in DNA methylation maintenance at this subset of sites. To test this, we analyzed cytosine methylation levels as well as 24-nt siRNAs abundance at both the Pol IV/V-codependent and Pol IV-independent sites. If Pol IV actively maintains DNA methylation at specific genomic sites to enable Pol V recruitment and transcription, then loss of Pol IV should have a more dramatic effect on the methylation levels at these sites. Indeed, Pol IV/V-codependent sites showed significantly higher 24-nt siRNAs levels as well as substantial reductions of all types of cytosine methylation in *nrpd1*, while Pol IV-independent sites showed fewer 24-nt siRNAs and less reduction in DNA methylation (Supplementary Fig. 2e,f). This is likely because the other DNA methylation maintenance pathways involving MET1, CMT3, and CMT2 are active at these loci, and compensate for the loss of methylation in the Pol IV mutant. In summary, these results show that even though Pol IV and Pol V work closely together in the RdDM pathway, Pol V can transcribe independently of Pol IV at many sites in the genome. Previous studies of Pol IV transcripts have shown them to be exceedingly rare in wild type because of their efficient processing into siRNAs by DICER enzymes [2-4]. However, it remains possible that trace levels of Pol IV transcripts could be present in our GRO-seq libraries. Thus, in order to uniquely focus on the characteristics of Pol V transcripts without any complication of the

presence of small amounts of Pol IV transcripts, we focused our remaining analysis on Pol IV-independent Pol V regions.

Pol V transcripts show evidence of small RNA dependent slicing.

Because our GRO-seq method did not include the fragmentation step typical of traditional GRO-seq [15], it was possible to estimate the length of Pol V nascent transcripts and assess their 5' nucleotide composition. We observed a range of read lengths from 30- to 90-nt long with a peak at around 50-nt, and detected very few reads longer than about 120-nt (Fig. 3a). Nascent Pol V transcripts observed in *nrpd1* GRO-seq showed a similar size distribution (Supplementary Fig. 3a). GRO-seq involves an *in vitro* nuclear run-on step in which the reaction is limited by time and nucleotide concentration, meaning that the run-on is unlikely to proceed to the natural 3' end of the transcript. Thus, the average Pol V transcript length measured here is likely an underestimate of the true length of Pol V transcripts *in vivo*. Using Pol V RIP-seq, Bohmdorfer et al. recently estimated the median Pol V transcript length to be around 200 nucleotides. However, since a fragmentation step was included in their RIP protocol, this was also an estimation [13]. Nevertheless, Pol V transcripts are clearly at least 50-nt long on average, which is significantly longer than Pol IV transcripts, which have been estimated to be around 30- to 40-nt long [2,3].

Eukaryotic and bacterial RNA polymerases preferentially initiate transcription at purines (A or G), commonly with a pyrimidine (C or T) present at the -1 position with respect to the transcription start site[2-4,20-22]. However, instead of this expected

enrichment at Pol V transcript 5' ends, we observed a strong U preference (on average 53.41%) at nucleotide +10 across six Col-0 biological replicates (Fig. 3b, Supplementary Fig. 3b). This characteristic was unlikely to be an artifact of the GRO-seq procedure since no such preference was observed in transcripts that mapped to mRNA regions (Supplementary Fig. 3c,d). In order to test whether the +10U signature was specific to nascent RNAs with certain lengths, we examined the nucleotide preferences within different size ranges. We found a +10U signature in all size ranges tested from 30-nt RNAs to RNAs longer than 70-nt, with the strongest signature in 40- to 50-nt long reads (Supplementary Fig. 3e-i).

In *Arabidopsis*, AGO4 shows slicer activity *in vitro* and interacts directly with Pol V [10,23]. In addition, AGO4-associated 24-nt siRNAs are highly enriched for 5' adenines [24,25]. Therefore, we hypothesized that the 5' end of Pol V transcripts is often defined by an AGO4 slicing event, and that the U at position 10 in Pol V transcripts corresponds to a 5' A in AGO4 24-nt siRNAs (Fig. 3c). We plotted the sequence composition of previously published AGO4-associated 24-nt siRNAs [26] that mapped to our identified Pol V transcript sites and observed a strong 5' enrichment for A (80.53%) (Fig. 3d). If Pol V transcripts are sliced at 10-nt from the AGO4-siRNAs 5' end, we should detect sense-antisense siRNA-Pol V transcript pairs separated by 10-nt and a corresponding 10-nt of complementary sequence (Fig. 3c). We plotted the distance between each AGO4-siRNAs 5' end and the 5' end of its Pol V transcript neighbors on the opposite strand. Consistent with our hypothesis, we found a strong peak of AGO4-associated 24-nt siRNAs 5' ends at 10 nucleotides downstream from

the Pol V 5' end (Fig. 3e). Overall, 78.07% of AGO4-associated 24-nt siRNAs had a Pol V-dependent transcripts partner detected in GRO-seq whose 5' end could be mapped 10 nucleotides away on the complementary strand.

To determine whether the slicing-associated U signature at position 10 was dependent on 24-nt siRNAs, which are transcribed by Pol IV, we examined the Pol V transcript sequence composition in the Pol IV mutant *nrpd1*. We found that in *nrpd1* the U preference at position 10 was completely abolished (Fig. 3f,g). Instead, we observed the conventional +1 A/U and a -1 U/A 5' signature (Fig. 3f) similar to other RNA polymerases [2-4,16,22,27], and also similar to mRNA GRO-seq reads in wild type or the *nrpd1* mutant (Supplementary Fig. 3c,d). These results strongly support the hypothesis that the +10U signature is due to 24-nt siRNAs dependent slicing of Pol V transcripts.

AGO4, AGO6, and AGO9 are required for the slicing of Pol V transcripts.

Given that AGO4 is the main ARGONAUTE involved in RdDM, we tested whether AGO4 is also required for slicing of Pol V transcripts by performing GRO-seq in the *ago4-5* mutant in the Col-0 background (*ago4/Col-0*) and the *ago4-4* mutant in the Ws background (*ago4/Ws*). We observed that the +10U slicing signature of Pol V transcripts was reduced 13.26% in *ago4-5* relative to wild-type Col-0 and 12.37% in *ago4-4* relative to wild-type Ws (Fig. 3b, Fig. 4a-c,i). The remaining slicing signature in *ago4* mutants is likely due to redundancy of AGO4 with two other close family members, AGO6 and AGO9 [24,28]. Therefore, we also performed GRO-seq in the *ago4-*

4/ago6-2/ago9-1 (*ago4/6/9*) triple mutant background [29]. The +10U signature in *ago4/6/9* mutants was completely abolished (Fig. 4d,i) suggesting a complete lack of slicing.

Previous work showed that the Asp-Asp-His (DDH) catalytic motif of AGO4 is required for slicing of RNA transcripts *in vitro* [10]. We therefore performed GRO-seq in plants containing either a wild-type AGO4 transgene (wtAGO4) expressed in *ago4/Ws* or the *ago4/6/9* mutant triple mutant, or a slicing defective AGO4 (D742A) mutant expressed in *ago4/Ws* or the *ago4/6/9* triple mutant [29]. We found that the wild-type AGO4 transgene largely complemented the +10U slicing signature in the *ago* mutants, while the AGO4 D742A catalytic mutant failed to restore the +10U signature (Fig. 4e-i). To rule out the possibility that the elimination of the +10U Pol V slicing signature in the *ago* mutants is caused by elimination of the +1A nucleotide preference of 24-nt siRNAs, we analyzed previously published small RNA-seq datasets corresponding to the same collection of *ago* mutant/transgene combinations [29]. We found that all mutants and mutant/transgene combinations retained a strong enrichment of A at position 1 of the 24-nt siRNAs (Supplementary Fig. 4a-h). These results further support the hypothesis that the +10U signature is due to Pol V transcript slicing, and that slicing is abolished in *ago4/6/9* triple mutants, although we cannot rule out minor levels of slicing that do not involve U-A pairing or by other AGO proteins.

SPT5L is required for the slicing of Pol V transcripts.

There are a number of proteins in the RdDM pathway whose precise function is unknown but that act at some point downstream of the biogenesis of siRNAs, including SUPPRESSOR OF TY INSERTION 5 – like/ KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (SPT5L) [30-34], DOMAINS REARRANGED METHYLTRANSFERASE3 (DRM3) [35], INVOLVED IN DE NOVO2 (IDN2) [36], IDN2-LIKE1 and 2 (IDL1 and 2) [37,38] SNF2-RING-HELICASE-LIKE1 and 2 (FRG1 and 2) [39], and SU(VAR)3-9 RELATED2 (SUVR2) [40,41]. Mutations in these genes all show a partial reduction of DNA methylation associated with the RdDM pathway, rather than a complete loss of RdDM as seen in strong mutant such as *nrpd1* or *nrpe1* [30-41]. To examine if any of these components are involved in the slicing of Pol V transcripts we performed GRO-seq in mutant backgrounds including *spt5l, drm3*, *idn2*, *idn2/idl1/idl2*, *frg1/frg2*, and *suvr2*. We observed that all mutants retained a strong +10U slicing signature (Fig. 5a-e, Fig. 6a) except for the *spt5l* mutant, which completely eliminated the slicing signature (Fig. 5f, Fig. 6a). A trivial explanation for the lack of +10U slicing signature in *spt5l* would be that this mutant eliminated 24-nt siRNAs or eliminated the enrichment of A at the 5' nucleotide of 24-nt siRNAs. However, we found only a moderate (though significant) reduction of 24-nt siRNA abundance (Fig. 6b) [30,32-34] and a strong remaining +1A nucleotide preference (Fig. 6c,d) in *spt5l*. These results reveal a novel role for SPT5L in the slicing of Pol V transcripts.

We also analyzed the effect of each of the mutants on the overall levels of Pol V GRO-seq signals (Fig. 6e), and as a control examined their effects on the background

levels of GRO-seq signals at the top 1,000 expressed Pol II genes (Supplementary Fig. 4i). While the *drm3*, *idn2*, *idn2/idl1/idl2*, *frg1/frg2*, and *suvr2* mutants showed only minor effects on overall Pol V transcript levels, *spt5l* showed a strong reduction. This reduction was even greater than that seen in the Pol IV mutant *nrpd1*, a strong RdDM mutant which shows a much greater reduction in DNA methylation than in *spt5l* [40]. This result suggests that SPT5L plays a role in Pol V transcript stability and/or production. SPT5L is a homolog of the Pol II elongation factor SPT5 [32]. It has been shown to interact with the Pol V complex, but its precise role in the RdDM pathway has been unclear [30-34]. Our finding that both slicing and Pol V transcript levels are affected in *spt5l* suggests that SPT5L plays a dual role in the processing and utilization of Pol V transcripts.

Conclusions

In this work we show that Pol V transcripts are frequently sliced in a siRNA- and SPT5L-dependent manner. Because the slicing signature is present in Pol V transcripts that are in the process of transcribing, it is clear that this slicing is occurring co-transcriptionally. AGO4 mutations that affect the catalytic residues required for slicing show a partial loss of RdDM similar to *spt5l* mutants [10,29], suggesting that the slicing step is required for efficient RNA-directed DNA methylation. However, it is also clear that slicing is not required for all RdDM, since *spt5l* mutants appear to abolish slicing, and yet show only a partial loss of CHH methylation at RdDM sites [30-33]. AGO4 can also physically interact with DRM2, which provides an alternative mechanism by which AGO4/siRNA complexes can promote

RdDM. This suggests a dual mechanism by which AGO4 can promote DRM2 activity, through both Pol V transcript slicing and through interaction with DRM2 (Model Fig. 6f).

SPT5L contains a region rich in WG repeats (called the AGO hook) that is capable of binding to AGO4 [32]. AGO4 also interacts with a similar WG repeat region within the largest subunit of Pol V [23]. It has been recently shown that deletion of the WG repeats of SPT5L, or deletion of the WG repeats of Pol V, still allow AGO4 recruitment and RdDM. However, simultaneous deletion of both WG repeat regions abolishes RdDM, indicating that the WG-rich domains of SPT5L and Pol V are redundantly required for AGO4 recruitment [42]. This genetic redundancy also indicates that SPT5L's role in AGO4 recruitment is unlikely to account for its requirement for Pol V transcript slicing. SPT5L is therefore a multifunctional protein mediating a number of steps in RdDM including AGO4 recruitment, and, as shown here, Pol V slicing and Pol V transcript abundance or stability (Model Fig. 6f)

In *Drosophila*, similar slicing patterns were observed in the AGO3-rasiRNA 'ping-pong' pathway in which AGO3 directs cleavage of its cognate mRNA target across from nucleotides 10 and 11, measured from the 5' end of the small RNA guide strand, followed by the generation of secondary small RNAs from mRNA targets [43,44]. Thus, one hypothesis is that sliced Pol V RNAs are further trimmed to generate secondary small RNAs, as was previously proposed [10]. However, we did not observe evidence suggesting secondary RNA production, suggesting that AGO4 slicing of Pol

V transcripts does not result in the production of secondary small RNAs (data not shown). This is consistent with a recent study suggesting that AGO4 dependent siRNAs result from RdDM feedback rather than from secondary siRNA production [29].

Our results also shed light on the long debate over the mechanism of action of AGO/siRNA complexes and whether the siRNAs target the nascent Pol V RNA or whether they bind directly to the DNA [11,42]. Our results demonstrating siRNA-mediated slicing of Pol V nascent transcripts clearly supports an RNA targeting model whereby the siRNAs target the nascent Pol V RNA rather than binding directly to the DNA. This is also supported by the conclusive data in fission yeast suggesting siRNA/RNA interactions [45-47]. Once the AGO4-siRNAs have bound to nascent Pol V RNAs and slicing has occurred, one possibility is that the resulting sliced RNAs or siRNA/sliced RNA duplexes play a signaling role, perhaps through specific RNA binding proteins, in the targeting of the DRM2 methyltransferase to methylate chromatin (Model Fig. 6f). This model is attractive because slicing represents the integration of the activities of the upstream Pol IV driven siRNA biogenesis pathway and the downstream Pol V driven non-coding RNA biogenesis pathway, which could provide additional accuracy and specificity for DNA methylation targeting. Another possibility is that slicing promotes the recycling of AGO/siRNA complexes, and/or Pol V transcripts to promote iterative cycles of targeting of DNA methylation through AGO4-DRM2 interactions [12]. Future studies aimed at understanding the biochemical details of the interaction of AGO4-bound siRNAs and Pol V targets are likely to shed additional light on the mechanisms of DNA methylation control.

Methods

Plant Materials and Growth

The *A. thaliana* accession Columbia (Col-0) was used as the wild-type genetic background for this study unless specified. The mutant alleles of *nrpd1-4* (SALK_083051) [48], *nrpe1-12* (SALK_033852) *, spt5l-1* (SALK_001254) [32], *drm3-1* (SALK_136439) [35], *idn2-1* (SALK_012288) [36], *suvr2-1*(SAIL_832_E07) [39], and *ago4-5 (described in* [33]*)* used in this study have been characterized previously and were in the Col-0 background. The double mutant for NRPD1 and NRPE1 was made by crossing *nrpd1-4* (SALK_083051) and *nrpe1-11* (SALK_029919) as described [49]. *frg1/2 (SALK_027637, SALK_057016)* double mutants were described before [39]. *idn2-1*, *idnl1-1* (SALK_075378), and *idnl2-1* (SALK_012288) triple mutant were described before [37]. Ws, *ago4/Ws*, *ago4/ago6/ago9*, *ago4/wtAGO4*, *ago4/D742A*, *ago4/6/9/wtAGO4*, and *ago4/6/9/D742A* were described previously [29]. All plants were grown on soil under long day conditions (16 hours light, 8 hours dark). Inflorescence tissues with both floral buds and open flowers were collected and used for the GRO-seq procedure. T-DNAs were confirmed by PCR-based genotyping.

Nuclei Isolation

Approximately 10 grams of inflorescence and meristem tissue was collected from plants and immediately placed in ice cold grinding buffer (300 mM sucrose, 20 mM Tris, pH 8.0, 5 mM MgCl2, 5 mM KCl, 0.2% Triton X-100, 5 mM β-mercaptoethanol, and 35% glycerol). Nuclei were isolated as described previously [16]. Briefly, samples were ground with an OMNI International General Laboratory Homogenizer at 4°C

until well homogenized, filtered through a 250 µm nylon mesh, a 100 µm nylon mesh, a miracloth, and finally a 40 µm cell strainer before being split into 50 ml conical tubes. Samples were spun for 10 minutes at 5,250*g*, the supernatant was discarded, and the pellets were pooled and resuspended in 25 ml of grinding buffer using a Dounce homogenizer. The wash step was repeated at least once more and nuclei were resuspended in 1 ml of freezing buffer (50 mM Tris, pH 8.0, 5 mM $MgCl_2$, 20% glycerol, and 5 mM β-mercaptoethanol).

GRO-seq

Approximately $5x10^6$ nuclei in 200 µl of freezing buffer were run-on in 3x NRO-reaction buffer [16]. For GRO-seq in Ws, *ago4/Ws*, *ago4/ago6/ago9*, *ago4/wtAGO4*, *ago4/D742A*, *ago4/6/9/wtAGO4*, and *ago4/6/9/D742A*, approximately $3x10^5$ to $5x10^5$ nuclei were used. To minimize run-on length, the limiting CTP concentration was reduced to a final concentration of 20 nM. Reactions were stopped after 5 minutes to minimize run on length (~5-15 nt) while still incorporating BrUTP by addition of 750 µl TRIzol LS(Fisher Scientific) and RNA was purified according to the manufacturer's manual. Without fragmentation or Terminator treatment, nascent RNA was enriched twice for BrU by αBrdU (Santa Cruz Biotechnology sc-32323AC Lots #A0215 and #C1716) and immunoprecipitated as described in Hetzel et al. 2016 [16]. Subsequently, sequencing libraries were prepared from precipitated RNA using TruSeq Small RNA Library Prep kit following manufacturer instructions (Illumina). For most GRO-seq libraries, 14 cycles of PCR were used to amplify the libraries and products ranging from 100 to 500 bp were size selected by agarose gel, except for

replicate 1 and 2 of *spt5l* (replicate 3 was prepared the same way as all other GRO-seq libraries), where products were size selected by double SPRI bead purification (ratio of Ampure beads to library: 0.5:1 to 1.1:1). The libraries were sequenced on either Illumina HiSeq 2000 or 2500 platform.

ChIP-seq

Chromatin immunoprecipitation was performed from 2 grams of formaldehyde crosslinked flower tissue as previously described [18], except that half of the input was immunoprecipitated with 3 µg of affinity purified anti-NRPE1 antibody generated by Covance that recognizes the peptide N-CDKKNSETESDAAAWG- C [50], and the other half was immunoprecipitated with pre-immune serum as control. DNA libraries for Illumina sequencing were generated using the Ovation Ultralow V2 system (NuGEN), and the libraries were sequenced on a HiSeq 2000 platform for single-end 50 bp, following the manufacturers' instructions.

Small RNA-seq

Total RNA was first extracted with Zymo Direct-zol RNA mini Prep kit (ZRC200687) followed by a size selection of RNA on a 15% Urea TBE Polyacrylamide gel (Invitrogen, EC6885BOX). Gels containing 15- to 30-nt were cut for small RNA library. After gel elution, Illumina TruSeq Small RNA kit (RS-200-0012) was used for making small RNA library. Agilent D1000 ScreenTape (5067-5582) was then used for checking the size and quality of final libraries.

Bioinformatic Analysis

GRO-seq analysis

Qseq files from the sequencer were demultiplexed and converted to fastq format with a customized script for downstream analysis. For GRO-seq data, paired-end reads were first trimmed for Illumina adaptors and primers using Cutadapt (v 1.9.1). After trimming, reads less than 10 bp long were removed with a customized Perl script. Paired-end reads were then separately aligned to the reference TAIR10 genome using Bowtie (v1.1.0) [51] by allowing only unique hit (-m 1) and up to 3 mismatches (-v 3). Paired reads aligned to positions within 2,000 bp to each other were considered as correct read pairs, and reads aligned to Watson or Crick strands were separated by a customized Perl script.

ChIP-seq analysis

Qseq files from the sequencer were demultiplexed and converted to fastq format with a customized script for downstream analysis. Fastq reads were aligned to the Arabidopsis reference genome (TAIR10) with Bowtie (v1.0.0) [51], allowing only uniquely mapping reads with fewer than two mismatches, and duplicated reads were combined into one read. NRPE1 ChIP-seq peak were called using MACS2 (v 2.1.1.) [52] in Col-0 and *nrpd1*, respectively, with default parameters using ChIP-seq with pre-immune serum in each condition as control. ChIP-seq metaplots were plotted using NGSplot (v 2.41.4) [53].

Identification of Pol V-dependent transcripts from GRO-seq data

In order to remove signals from annotated gene regions, we only included GRO-seq reads aligned to defined Pol V occupied regions. Pol V ChIP-seq peak regions were split into 100 bp bins and the reads from GRO-seq in each bin were counted. To call Pol V-dependent transcripts, the R package DESeq2 [54] was used applied. Only bins with at least 4-fold enrichment in Col-0 compared to the *nrpe1* and *nrpd1/e1* mutant and FDR less than 0.05 were retained. Bins within 200 bp of each other were then merged into Pol V-dependent transcripts clusters. To characterize Pol IV dependency on those Pol V-dependent transcripts clusters, we checked NRPE1 binding in *nrpd1* mutant. If a Pol V-dependent transcripts cluster was not bound by NRPE1 in *nrpd1* mutant while also had a RPKM (Reads Per Kilobase Million) of GRO-seq in *nrpd1* greater than 2, then this site was classified as Pol IV/V codependent. On the other hand, if a Pol V-dependent transcripts cluster was also bound by NRPE1 in *nrpd1* mutant while had a RPKM of GRO-seq in *nrpd1* less than 1, then this site was classified as Pol IV-independent Pol V sites.

AGO4 RIP-seq and total small RNA analysis

Qseq files for small RNA-seq from the sequencer were demultiplexed and converted to fastq format with a customized script for downstream analysis. Raw AGO4 RIP-seq data were obtained from previously published datasets (GSM707686) [26]. Reads were then trimmed for Illumina adaptors using Cutadapt (v 1.9.1) and mapped to the TAIR10 reference genome using Bowtie(v1.1.0) [51] allowing only one unique hit (-m 1) and zero mismatch.

Whole Genome Bisulfite Sequencing (WGBS) analysis

Processed WGBS data of Col-0 and *nrpd1* were obtained from previously published datasets (GSE39901, GSE38286) [40]. CG, CHG, and CHH methylation over different regions were extracted using a customized Perl script.

Data availability

High-throughput sequencing data that support the findings in this study can be accessed through Gene Expression Omnibus (GEO) database with accession number GSE108078 and GSE100010.

Correspondence and requests for materials should be addressed to S.E.J.

Acknowledgments

The authors thank members of the Jacobsen lab for insightful discussion and Mahnaz Akhavan for technical assistance. The authors thank Life Science Editors for editing assistance. High throughput sequencing was performed at UCLA BSCRC BioSequencing Core Facility. W.L. is supported by Philip J. Whitcome Fellowship from the UCLA Molecular Biology Institute and a scholarship from the Chinese Scholarship Council. Z.Z. is supported by a scholarship from the Chinese Scholarship Council. Group of J.Z. is supported by the Thousand Talents Program for Young Scholars and by the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (2016ZT06S172). This work was supported by NIH grant

GM60398 to S.E.J. and NIH grant R01GM094428 and R01GM52413 to J.C. S.E.J. and J.C. are Investigators of the Howard Hughes Medical Institute.

Author Contributions

W.L., J.H., S.H.C.D., and S.F. performed GRO-seq experiments. M.G. performed ChIP-seq experiments. W.L., J.G.B, Z.Z., and S.F. performed small RNA-seq experiments. W.L. and M.G. performed the bioinformatics analysis. W.L. and S.E.J. wrote the manuscript. J.Z., H.Y.K., and J.C. assisted in writing and discussion.

Competing interest

The authors declare no competing financial interests.

Figure Legends

Fig. 1. Capturing Pol V-dependent transcripts with GRO-seq.

a, Procedure for constructing *Arabidopsis* GRO-seq library, which captures nascent Pol V transcripts. 7meG-capped transcripts generated by Pol II are excluded by selective ligation to the 5′ monophosphorylated (5′Pi) RNAs generated by Pol I, IV, and V. **b,** Screenshot of CG, CHG, and CHH methylation in wild-type Col-0, Pol V ChIP-seq in Col-0, and GRO-seq in Col-0, *nrpe1*, and *nrpd1/e1* over the previously identified Pol V locus IGN5 [11]. For CG, CHG, and CHH methylation, y-axis indicate the percentage of methylation. Plus (+) and Minus (-) indicate the strandness of GRO-seq signal. **c,** Metaplot of Pol V ChIP-seq signal over input and ratio of GRO-seq signal in Col-0 to *nrpe1* graphed over the centers of Pol V occupied regions defined by Pol V ChIP-seq.

Fig. 2. Characteristics of Pol V-dependent transcripts.

a, Distribution of ratios of plus strand GRO-seq signals over minus strand GRO-seq signals in Col-0 over the top 500 expressed mRNAs. **b,** Distribution of ratios of plus strand GRO-seq signals over minus strand GRO-seq signals in Col-0 over the top 500 Pol V enriched regions defined by Pol V ChIP-seq. **c,** Pol V ChIP-seq signals over inputs and the ratio of GRO-seq signal in Col-0 to *nrpe1* plotted over Pol V-associated transposons with different lengths.

Fig. 3. Pol V transcripts is sliced in a small RNA dependent manner.

a, Size distribution of nascent transcripts in Col-0 over Pol V-dependent regions. All replicates for Col-0 GRO-seq were merged for this plot. **b,** The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent transcripts captured in Col-0. All replicates for Col-0 GRO-seq were merged for this plot. **c,** A predicted model indicating the first 10-nt of AGO4/6/9 associated small RNAs show complementarities to the first 10-nt of sliced nascent transcripts over Pol V-dependent regions captured in GRO-seq library. **d,** The relative nucleotide bias of each position for all AGO4-associated 24-nt siRNAs over regions that generated Pol V-dependent transcripts. **e,** Frequency map of the separation of 5' of Pol V-dependent RNAs mapping to AGO4-associated 24-nt siRNAs on the opposite strand. **f,** The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent transcripts captured in *nrpd1*. **g,** The percentage of U presented over genomic average at position 10 from the 5' ends of nascent transcripts captured with GRO-seq in Col-0, *nrpd1*, *nrpe1*, and *nrpd1/e1*.

Fig. 4. Slicing of Pol V transcripts requires AGO4/6/9.

a-h, The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent transcripts captured in Ws (**a**), *ago4/Col-0* (**b**), *ago4/Ws* (**c**), *ago4/6/9* (**d**), *ago4/wtAGO4* (**e**), *ago4/D742A* (**f**), *ago4/6/9/wtAGO4* (**g**) and *ago4/6/9/D742A* (**h**). Replicates were merged for plot (**a-h**). **i,** The percentage of U presented over genomic average at position 10 from the 5' end of nascent transcripts captured with GRO-seq in Col-0, *ago4/Col-0*, Ws, *ago4/Ws*, *ago4/6/9*, *ago4/wtAGO4*, *ago4/D742A*, *ago4/6/9/wtAGO4*, and *ago4/6/9/D742A*.

Fig. 5. Slicing signature of Pol V transcripts is eliminated in *spt5l* mutants.

a-f, The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent transcripts captured in *idn2* (**a**), *idn2/idl1/idl2* (**b**), *drm3* (**c**), *suvr2* (**d**), *frg1/2* (**e**), *spt5l* (**f**). Replicates were merged for plot (**a-f**).

Fig. 6. SPT5L is required for slicing of Pol V transcripts.

a, The percentage of U presented over genomic average at position 10 from the 5' end of nascent transcripts captured with GRO-seq in Col-0, *spt5l*, *drm3*, *frg1/2*, *idn2/idl1/2*, *idn2*, and *suvr2*. **b,** Normalized 24-nt siRNAs abundance in Col-0, *spt5l*, and *nrpd1*. *p-value < 0.05* (Welch Two Sample t-test). **c,d,** The relative nucleotide bias of each position for all 24-nt siRNAs in Col-0 (**c**) and *spt5l* (**d**) generated over Pol V-dependent regions. **e,** Nascent transcripts abundance over Pol V-dependent regions in Col-0, *nrpd1*, *nrpe1*, *nrpd1/e1*, *spt5l*, *drm3*, *frg1/2*, *idn2/idl1/2*, *idn2*, and *suvr2*. *p-value < 0.05* (Welch Two Sample t-test). **f,** Proposed model for slicing of Pol V transcripts.

References

1. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11,** 204–220 (2010).
2. Blevins, T. *et al.* Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. *Elife* **4,** e09591 (2015).
3. Zhai, J. *et al.* A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis. *Cell* **163,** 445–455 (2015).
4. 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).
5. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2,** E104 (2004).
6. Haag, J. R. *et al.* In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. *Molecular Cell* **48,** 811–818 (2012).
7. Qi, Y., Denli, A. M. & Hannon, G. J. Biochemical specialization within Arabidopsis RNA silencing pathways. *Molecular Cell* **19,** 421–428 (2005).
8. Zilberman, D., Cao, X. & Jacobsen, S. E. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299,** 716–719 (2003).
9. Li, C. F. *et al.* An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. *Cell* **126,** 93–106 (2006).
10. Qi, Y. *et al.* Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443,** 1008–1012 (2006).
11. 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).
12. Zhong, X. *et al.* Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell* **157,** 1050–1060 (2014).
13. Böhmdorfer, G. *et al.* Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. *Elife* **5,** 1325 (2016).
14. Wierzbicki, A. T., Ream, T. S., Haag, J. R. & Pikaard, C. S. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nature Genetics* **41,** 630–634 (2009).
15. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322,** 1845–1848 (2008).
16. Hetzel, J., Duttke, S. H., Benner, C. & Chory, J. Nascent RNA sequencing reveals distinct features in plant transcription. *Proc. Natl. Acad. Sci. U.S.A.* **113,** 12316–12321 (2016).
17. Zemach, A. *et al.* The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153,** 193–

205 (2013).
18. Zhong, X. *et al.* DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat. Struct. Mol. Biol.* **19,** 870–875 (2012).
19. Johnson, L. M. *et al.* SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507,** 124–128 (2014).
20. Smale, S. T. & Kadonaga, J. T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **72,** 449–479 (2003).
21. Sollner-Webb, B. & Reeder, R. H. The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in X. laevis. *Cell* **18,** 485–499 (1979).
22. Zecherle, G. N., Whelen, S. & Hall, B. D. Purines are required at the 5' ends of newly initiated RNAs for optimal RNA polymerase III gene expression. *Mol. Cell. Biol.* **16,** 5801–5810 (1996).
23. El-Shami, M. *et al.* Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* **21,** 2539–2544 (2007).
24. Mi, S. *et al.* Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell* **133,** 116–127 (2008).
25. Havecker, E. R. *et al.* The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *The Plant Cell* **22,** 321–334 (2010).
26. Wang, H. *et al.* Deep sequencing of small RNAs specifically associated with Arabidopsis AGO1 and AGO4 uncovers new AGO functions. *The Plant Journal* **67,** 292–304 (2011).
27. Vo Ngoc, L., Cassidy, C. J., Huang, C. Y., Duttke, S. H. C. & Kadonaga, J. T. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev.* **31,** 6–11 (2017).
28. Eun, C. *et al.* AGO6 functions in RNA-mediated transcriptional gene silencing in shoot and root meristems in Arabidopsis thaliana. *PLoS ONE* **6,** e25730 (2011).
29. Wang, F. & Axtell, M. J. AGO4 is specifically required for heterochromatic siRNA accumulation at Pol V-dependent loci in Arabidopsis thaliana. *The Plant Journal* (2016). doi:10.1111/tpj.13463
30. He, X.-J. *et al.* An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. *Cell* **137,** 498–508 (2009).
31. 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).
32. Bies-Etheve, N. *et al.* RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep.* **10,** 649–654 (2009).
33. Greenberg, M. V. C. *et al.* Identification of genes required for de novo DNA methylation in Arabidopsis. *Epigenetics* **6,** 344–354 (2011).
34. Huang, L. *et al.* An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat. Struct. Mol. Biol.* **16,** 91–93

(2009).
35. Zhong, X. *et al.* Domains rearranged methyltransferase3 controls DNA methylation and regulates RNA polymerase V transcript abundance in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 911–916 (2015).
36. Ausin, I., Mockler, T. C., Chory, J. & Jacobsen, S. E. IDN1 and IDN2 are required for de novo DNA methylation in Arabidopsis thaliana. *Nat. Struct. Mol. Biol.* **16,** 1325–1327 (2009).
37. Ausin, I. *et al.* INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 8374–8381 (2012).
38. Zhang, C.-J. *et al.* IDN2 and its paralogs form a complex required for RNA-directed DNA methylation. *PLoS Genet.* **8,** e1002693 (2012).
39. Groth, M. *et al.* SNF2 chromatin remodeler-family proteins FRG1 and -2 are required for RNA-directed DNA methylation. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 17666–17671 (2014).
40. Stroud, H., Greenberg, M. V. C., Feng, S., Bernatavichute, Y. V. & Jacobsen, S. E. Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the *Arabidopsis* Methylome. *Cell* **152,** 352–364 (2013).
41. Han, Y.-F. *et al.* SUVR2 is involved in transcriptional gene silencing by associating with SNF2-related chromatin-remodeling proteins in Arabidopsis. *Cell Res.* **24,** 1445–1465 (2014).
42. Lahmy, S. *et al.* Evidence for ARGONAUTE4-DNA interactions in RNA-directed DNA methylation in plants. *Genes Dev.* **30,** 2565–2570 (2016).
43. Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science* **315,** 1587–1590 (2007).
44. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128,** 1089–1103 (2007).
45. Shimada, Y., Mohn, F. & Bühler, M. The RNA-induced transcriptional silencing complex targets chromatin exclusively via interacting with nascent transcripts. *Genes Dev.* **30,** 2571–2580 (2016).
46. Noma, K.-I. *et al.* RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nature Genetics* **36,** 1174–1180 (2004).
47. Zofall, M. *et al.* RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. *Science* **335,** 96–100 (2012).
48. Herr, A. J., Jensen, M. B., Dalmay, T. & Baulcombe, D. C. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308,** 118–120 (2005).
49. Pontier, D. *et al.* Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev.* **19,** 2030–2040 (2005).
50. 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. *Molecular Cell* **33,** 192–203 (2009).
51. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome*

Biol. **10,** R25 (2009).
52. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
53. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15,** 284 (2014).
54. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).

Supplementary Figure 1. Modified GRO-seq is able to capture nascent Pol V-dependent transcripts. **a,** Scatterplot of signals from two independent GRO-seq experiments in Col-0. The Pearson's correlation coefficient is calculated and shown on the plot. **b,** Metaplot showing GRO-seq signals over Pol V-occupied regions in Col-0 and *nrpe1*. **c,** Metaplot showing GRO-seq signals over annotated genes in Col-0 and *nrpe1*. **d,** Scatterplot of normalized signals from Pol V ChIP-seq versus GRO-seq in Col-0. The Pearson's correlation coefficient is calculated and shown on the plot. **e,** Genome browser screenshot for CG, CHG, and CHH methylation in Col-0, Pol V ChIP-seq signals in Col-0, and GRO-seq signals in Col-0, *nrpe1*, and *nrpd1/e1* of a representative long TE and a representative short TE. Plus (+) and Minus (-) indicate the strandness of GRO-seq signal.

Supplementary Figure 2. Characterization of Pol IV/V-codependent sites and Pol IV-independent Pol V sites. **a,b,** Genome browser screenshot for Pol V ChIP-seq signals in Col-0 and GRO-seq signals in Col-0, *nrpe1, nrpd1,* and *nrpd1/e1* of a representative Pol IV/V-codependent site (**a**) and Pol IV-independent Pol V site (**b**) . Plus (+) and Minus (-) indicate the strandness of GRO-seq signal. **c,d,** Heatmap of log2 ratio of GRO-seq in Col-0 vs. *nrpe1*, GRO-seq in *nrpd1* vs. *nrpd1*, Pol V ChIP signals in Col-0, and Pol V ChIP-seq signals in *nrpd1* plotted over Pol IV/V-codependent sites (**c**) and Pol IV-independent Pol V sites (**d**). **e,** Boxplot of CG, CHG, and CHH methylation difference in *nrpd1* vs. Col-0. **p-value < 0.05* (Welch Two Sample t-test). **f,** Normalized 24-nt siRNAs abundance in Col-0 over Pol IV/V-

codependent sites and Pol IV-independent Pol V sites. **p-value < 0.05* (Welch Two Sample t-test).

Supplementary Figure 3. Pol V transcripts with different lengths are sliced. **a,** Size distribution of nascent transcripts in *nrpd1* over Pol V-dependent regions. Replicates were merged for this plot. **b,** The percentage of U presented over genomic average at position 10 from the 5' ends of nascent transcripts captured with GRO-seq in six biological replicates for Col-0. **c,d,** The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent RNAs generated from the top 1,000 expressed annotated gene regions in Col-0 (**c**) and *nrpd1* (**d**). Replicates were merged for plot (**c-d**). **e-i,** The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent transcripts of 30- to 40-nt long (**e**), 40- to 50-nt long (**f**), 50- to 60-nt long (**g**), 60- to 70-nt long (**h**) and 70-nt and longer (**i**) captured in Col-0. Replicates were merged for plot (**e-i**).

Supplementary Figure 4. 24nt-siRNAs retain strong enrichment of A at position 1 for *ago4*, *ago4/6/9* mutant and *ago4* or *ago4/6/9* mutant expressing wtAGO4 or D742A. **a-h,** The relative nucleotide bias of each position for 24-nt siRNAs over Pol V dependent regions in Col-0 (**a**), Ws (**b**), *ago4/Ws* (**c**), *ago4/wtAGO4* (**d**), *ago4/D742A* (**e**), *ago4/6/9* (**f**), *ago4/6/9/wtAGO4* (**g**) and *ago4/6/9/D742A* (**h**). **i,** Boxplot of normalized GRO-seq signals from top 1,000 expressed annotated gene in Col-0, *nrpd1*, *nrpe1*, *nrpd1/e1*, *spt5l*, *drm3*, *frg1/2*, *idn2/idl1/idl2*, *idn2*, and *suvr2*. N.S., not significant.

Fig.1

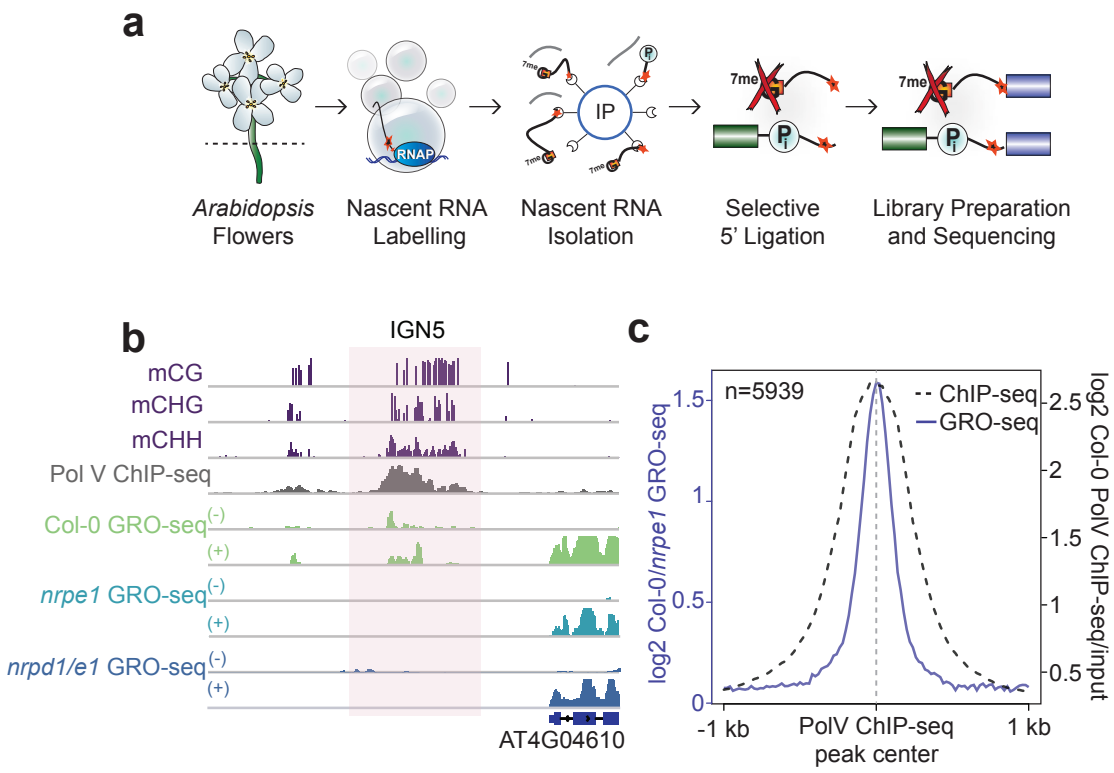


Fig. 1. Capturing Pol V-dependent transcripts with GRO-seq.
a, Procedure for constructing Arabidopsis GRO-seq library, which captures nascent Pol V transcripts. 7meG-capped transcripts generated by Pol II are excluded by selective ligation to the 5′ monophosphorylated (5′Pi) RNAs generated by Pol I, IV, and V. **b,** Screenshot of CG, CHG, and CHH methylation in wild-type Col-0, Pol V ChIP-seq in Col-0, and GRO-seq in Col-0, *nrpe1*, and *nrpd1/e1* over the previously identified Pol V locus IGN5 [11]. For CG, CHG, and CHH methylation, y-axis indicate the percentage of methylation. Plus (+) and Minus (-) indicate the strandness of GRO-seq signal. **c,** Metaplot of Pol V ChIP-seq signal over input and ratio of GRO-seq signal in Col-0 to *nrpe1* graphed over the centers of Pol V occupied regions defined by Pol V ChIP-seq.

Fig. 2

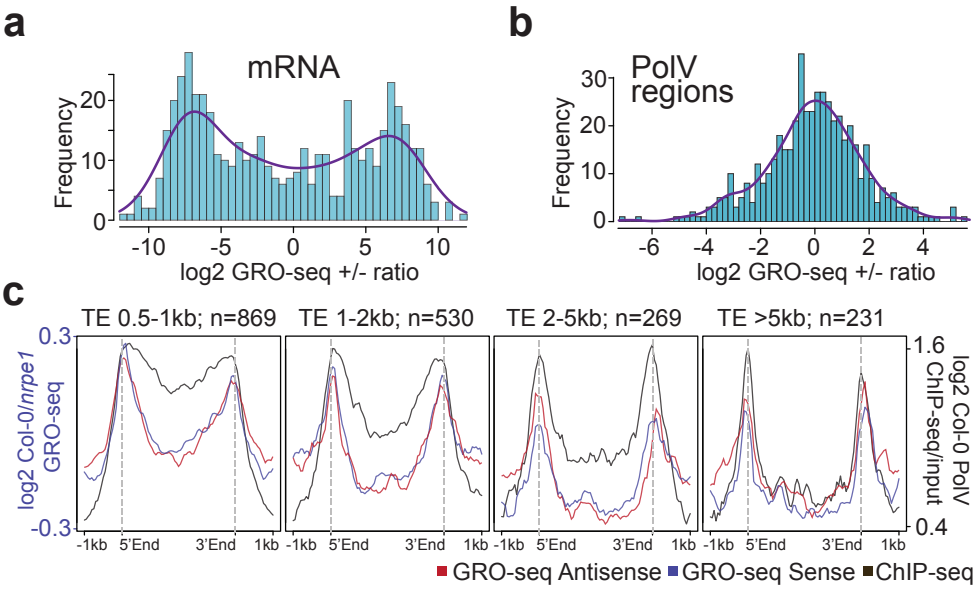


Fig. 2. Characteristics of Pol V-dependent transcripts.
a, Distribution of ratios of plus strand GRO-seq signals over minus strand GRO-seq signals in Col-0 over the top 500 expressed mRNAs. **b,** Distribution of ratios of plus strand GRO-seq signals over minus strand GRO-seq signals in Col-0 over the top 500 Pol V enriched regions defined by Pol V ChIP-seq. **c,** Pol V ChIP-seq signals over inputs and the ratio of GRO-seq signal in Col-0 to *nrpe1* plotted over Pol V-associated transposons with different lengths.

Fig. 3

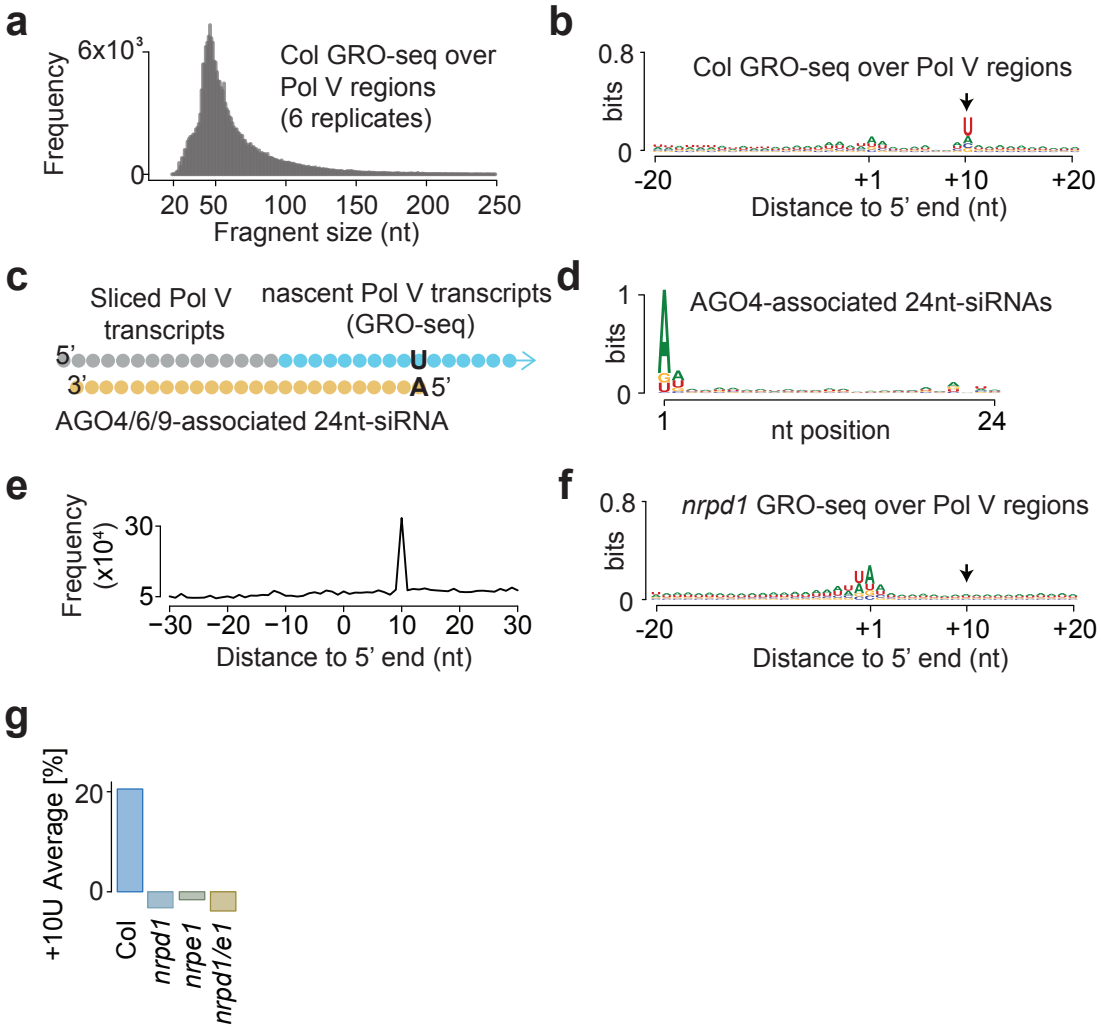


Fig. 3. Pol V transcripts is sliced in a small RNA dependent manner.
a, Size distribution of nascent transcripts in Col-0 over Pol V-dependent regions. All replicates for Col-0 GRO-seq were merged for this plot. **b,** The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent transcripts captured in Col-0. All replicates for Col-0 GRO-seq were merged for this plot. **c,** A predicted model indicating the first 10-nt of AGO4/6/9 associated small RNAs show complementarities to the first 10-nt of sliced nascent transcripts over Pol V-dependent regions captured in GRO-seq library. **d,** The relative nucleotide bias of each position for all AGO4-associated 24-nt siRNAs over regions that generated Pol V-dependent transcripts. **e,** Frequency map of the separation of 5' of Pol V-dependent RNAs mapping to AGO4-associated 24-nt siRNAs on the opposite strand. **f,** The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent transcripts captured in *nrpd1*. **g,** The percentage of U presented over genomic average at position 10 from the 5' ends of nascent transcripts captured with GRO-seq in Col-0, *nrpd1*, *nrpe1*, and *nrpd1/e1*.

Fig. 4

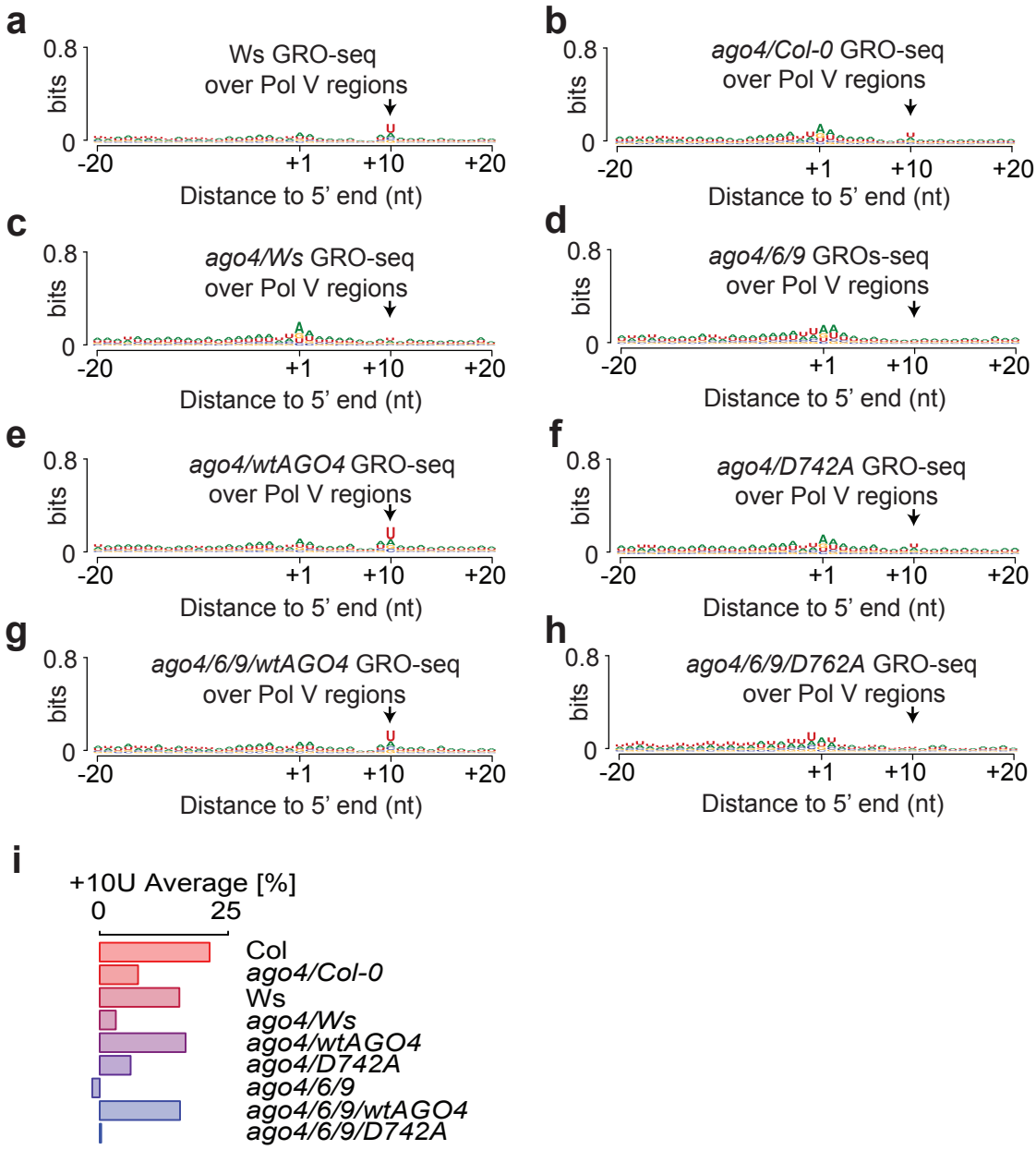


Fig. 4. Slicing of Pol V transcripts requires AGO4/6/9.
a-h, The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent transcripts captured in Ws (**a**), *ago4/Col-0* (**b**), *ago4/Ws* (**c**), *ago4/6/9* (**d**), *ago4/wtAGO4* (**e**), *ago4/D742A* (**f**), *ago4/6/9/wtAGO4* (**g**) and *ago4/6/9/D742A* (**h**). Replicates were merged for plot (**a-h**). **i,** The percentage of U presented over genomic average at position 10 from the 5' end of nascent transcripts captured with GRO-seq in Col-0, *ago4/Col-0*, Ws, *ago4/Ws*, *ago4/6/9*, *ago4/wtAGO4*, *ago4/D742A*, *ago4/6/9/wtAGO4*, and *ago4/6/9/D742A*.

Fig. 5

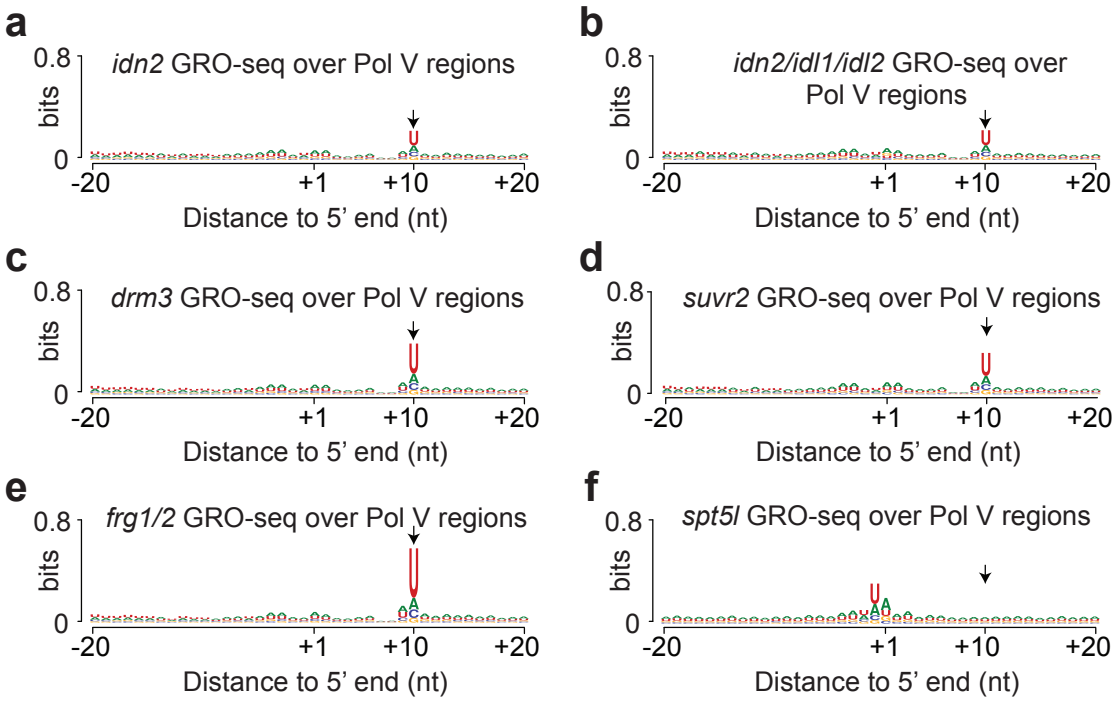


Fig. 5. Slicing signature of Pol V transcripts is eliminated in spt5l mutants.
a-f, The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent transcripts captured in *idn2* (**a**), *idn2/idl1/idl2* (**b**), *drm3* (**c**), *suvr2* (**d**), *frg1/2* (**e**), *spt5l* (**f**). Replicates were merged for plot (**a-f**).

Fig. 6

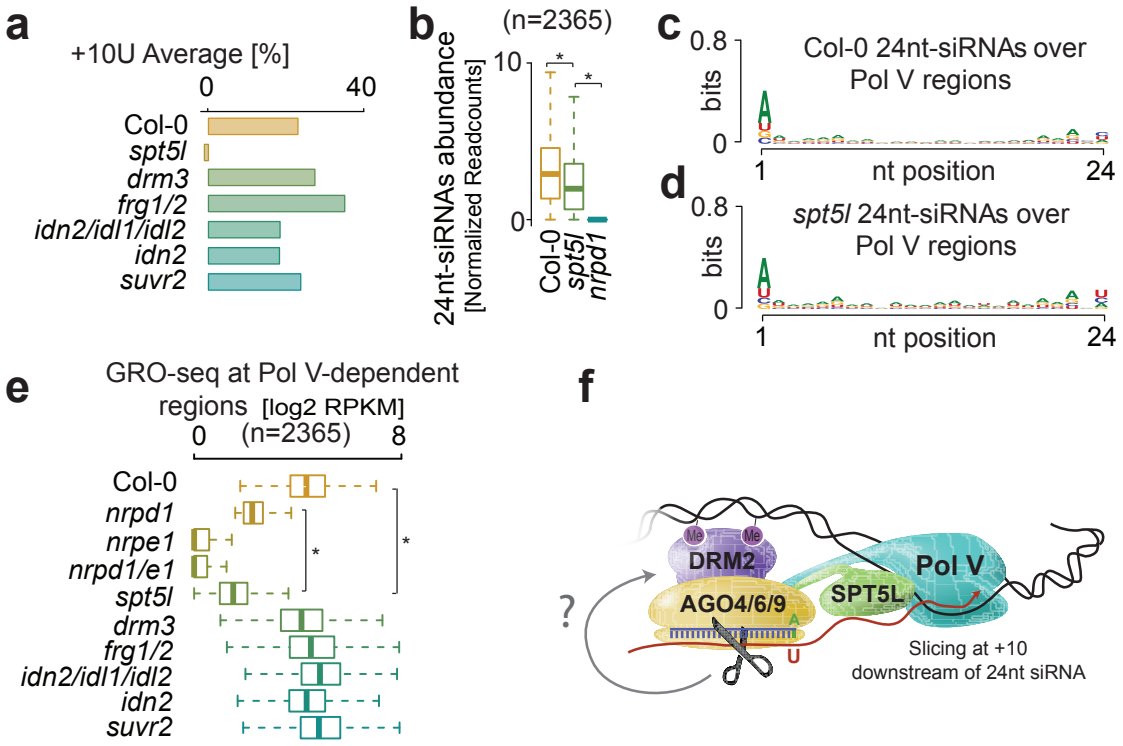
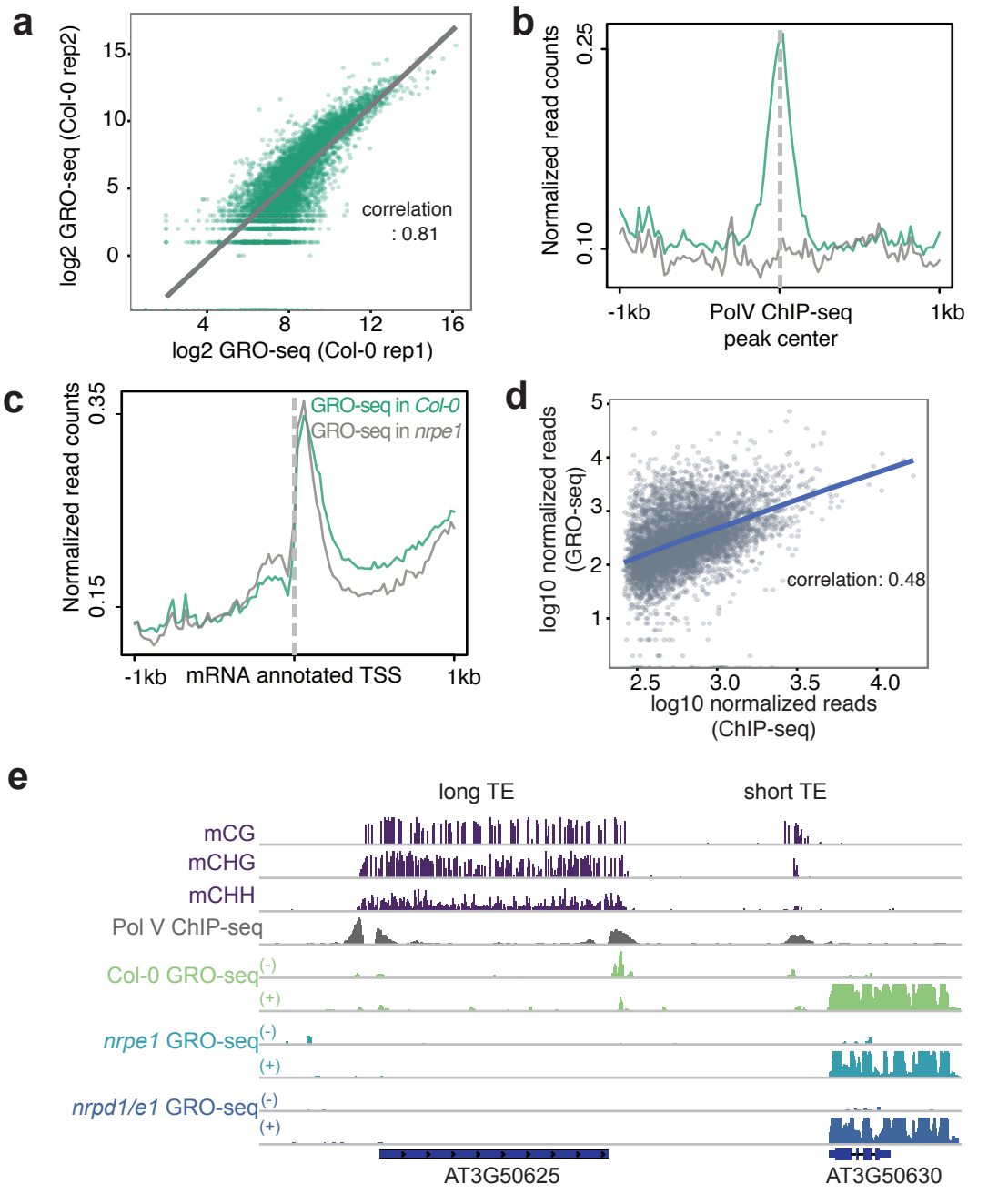


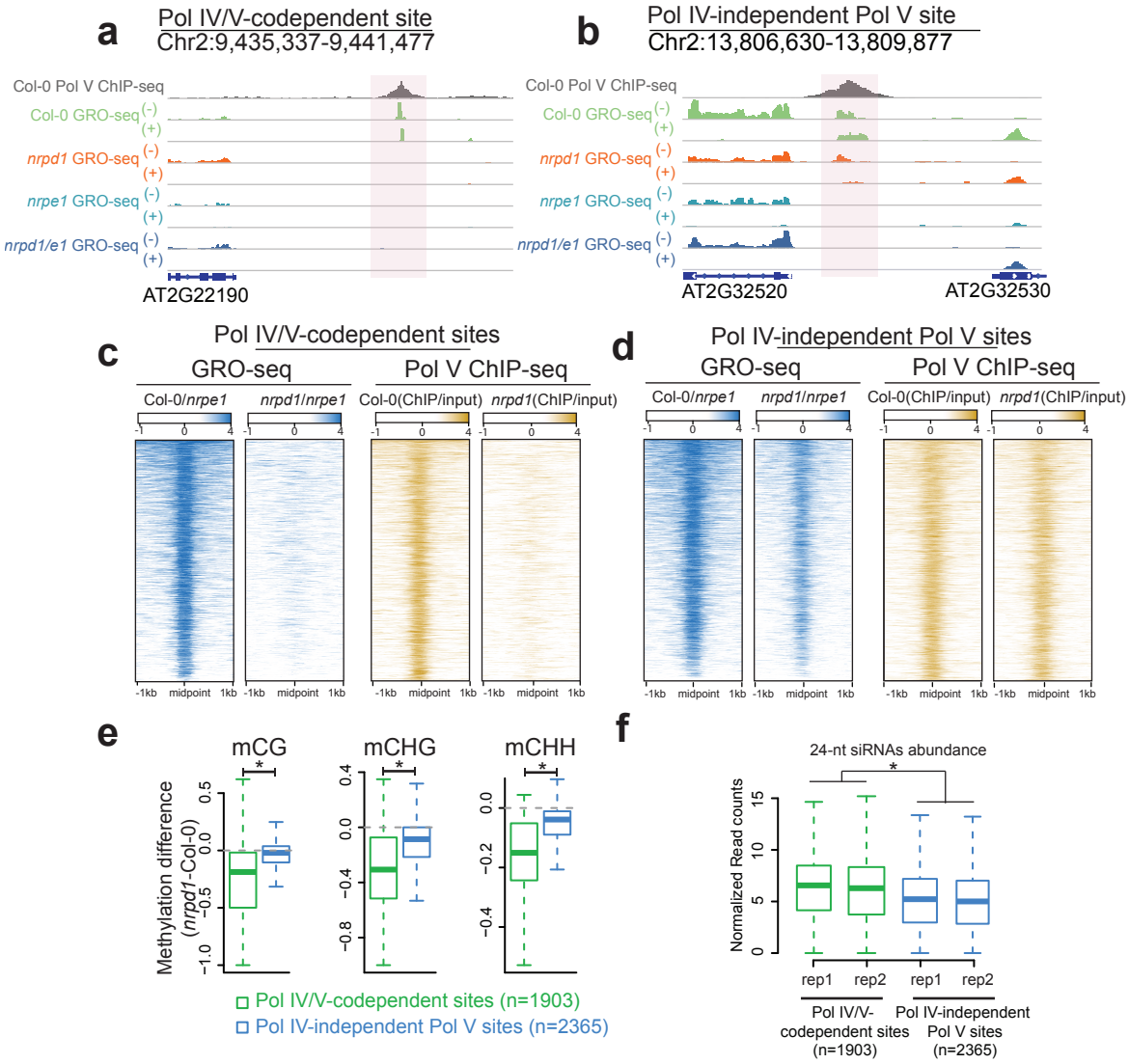
Fig. 6. SPT5L is required for slicing of Pol V transcripts.
a, The percentage of U presented over genomic average at position 10 from the 5' end of nascent transcripts captured with GRO-seq in Col-0, *spt5l*, *drm3*, *frg1/2*, *idn2/idl1/2*, *idn2*, and *suvr2*. **b**, Normalized 24-nt siRNAs abundance in Col-0, *spt5l*, and *nrpd1*. **p-value < 0.05* (Welch Two Sample t-test). **c,d,** The relative nucleotide bias of each position for all 24-nt siRNAs in Col-0 (**c**) and *spt5l* (**d**) generated over Pol V-dependent regions. **e,** Nascent transcripts abundance over Pol V-dependent regions in Col-0, *nrpd1*, *nrpe1*, *nrpd1/e1*, *spt5l*, *drm3*, *frg1/2*, *idn2/idl1/2*, *idn2*, and *suvr2*. **p-value < 0.05* (Welch Two Sample t-test). **f**, Proposed model for slicing of Pol V transcripts.

Supplementary Fig.1



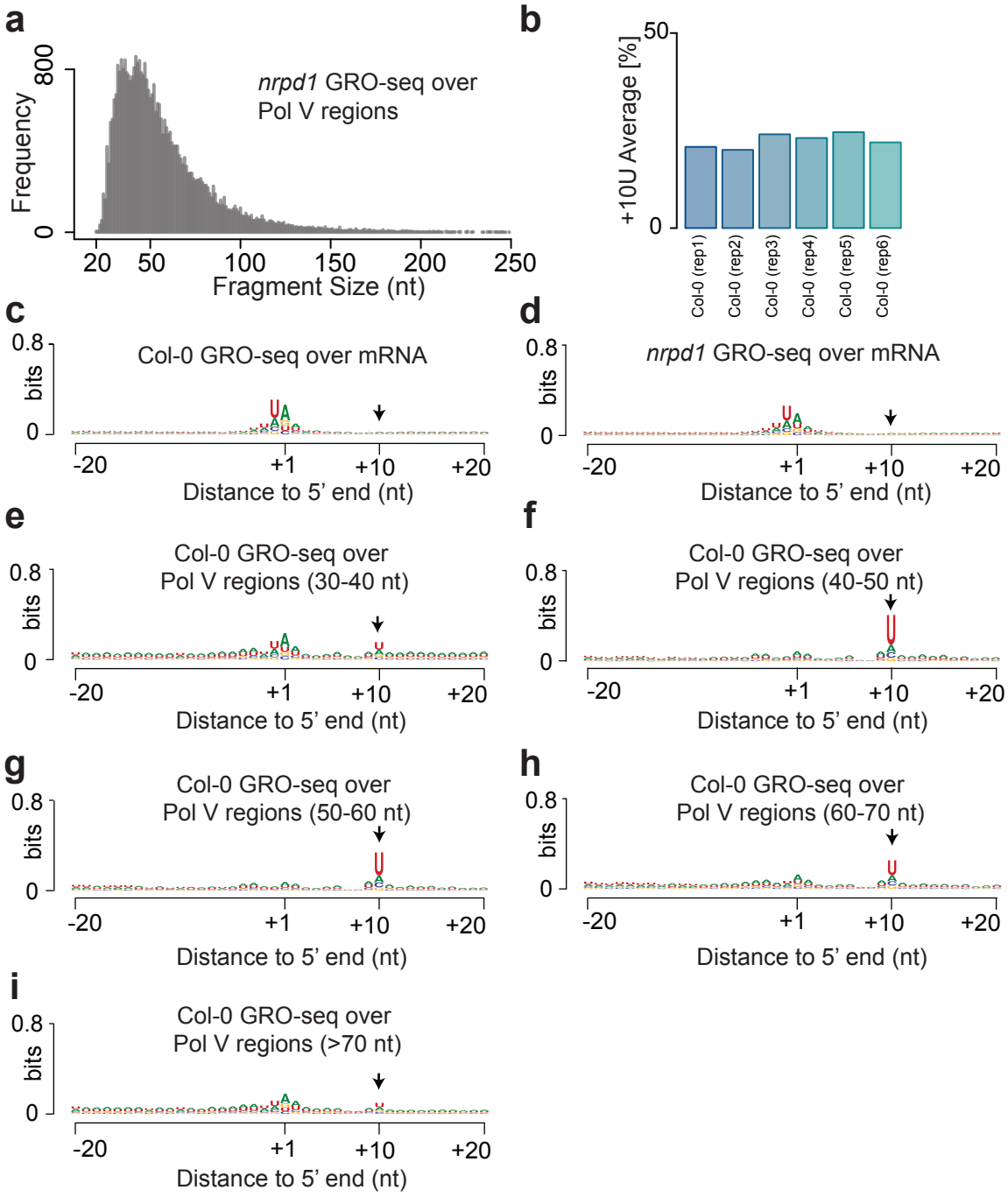
Supplementary Figure 1. Modified GRO-seq is able to capture nascent Pol V-dependent transcripts. **a,** Scatterplot of signals from two independent GRO-seq experiments in Col-0. The Pearson's correlation coefficient is calculated and shown on the plot. **b,** Metaplot showing GRO-seq signals over Pol V-occupied regions in Col-0 and *nrpe1*. **c,** Metaplot showing GRO-seq signals over annotated genes in Col-0 and *nrpe1*. **d,** Scatterplot of normalized signals from Pol V ChIP-seq versus GRO-seq in Col-0. The Pearson's correlation coefficient are calculated and shown on the plot. **e,** Genome browser screenshot for CG, CHG, and CHH methylation in Col-0, Pol V ChIP-seq signals in Col-0, and GRO-seq signals in Col-0, *nrpe1*, and *nrpd1/e1* of a representative long TE and a representative short TE. Plus (+) and Minus (-) indicate the strandness of GRO-seq signal.

Supplementary Fig.2



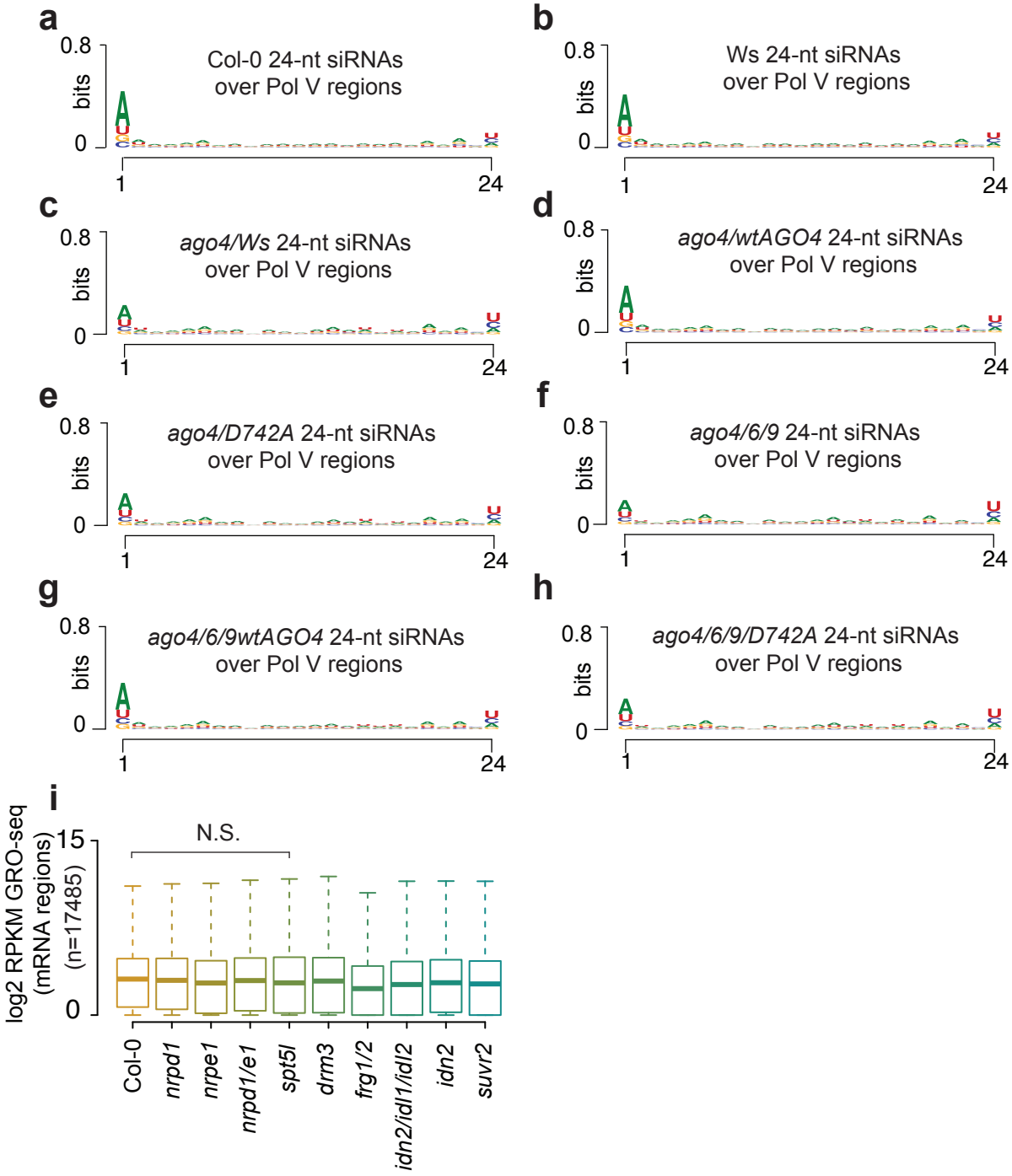
Supplementary Figure 2. Characterization of Pol IV/V-codependent sites and Pol IV-independent Pol V sites. **a,b,** Genome browser screenshot for Pol V ChIP-seq signals in Col-0 and GRO-seq signals in Col-0, *nrpe1*, *nrpd1*, and *nrpd1/e1* of a representative Pol IV/V-codependent site (**a**) and Pol IV-independent Pol V site (**b**) . Plus (+) and Minus (-) indicate the strandness of GRO-seq signal. **c,d,** Heatmap of log2 ratio of GRO-seq in Col-0 vs. *nrpe1*, GRO-seq in *nrpd1* vs. *nrpd1*, Pol V ChIP signals in Col-0, and Pol V ChIP-seq signals in *nrpd1* plotted over Pol IV/V-codependent sites (**c**) and Pol IV-independent Pol V sites (**d**). **e,** Boxplot of CG, CHG, and CHH methylation difference in *nrpd1* vs. Col-0. **p-value < 0.05* (Welch Two Sample t-test). **f,** Normalized 24-nt siRNAs abundance in Col-0 over Pol IV/V-codependent sites and Pol IV-independent Pol V sites. **p-value < 0.05* (Welch Two Sample t-test).

Supplementary Fig.3



Supplementary Figure 3. Pol V transcripts with different lengths are sliced. **a,** Size distribution of nascent transcripts in *nrpd1* over Pol V-dependent regions. Replicates were merged for this plot. **b,** The percentage of U presented over genomic average at position 10 from the 5' ends of nascent transcripts captured with GRO-seq in six biological replicates for Col-0. **c,d,** The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent RNAs generated from the top 1,000 expressed annotated gene regions in Col-0 (**c**) and nrpd1 (**d**). Replicates were merged for plot (**c-d**). **e-i**, The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent transcripts of 30- to 40-nt long (**e**), 40- to 50-nt long (**f**), 50- to 60-nt long (**g**), 60- to 70-nt long (**h**) and 70-nt and longer (**i**) captured in Col-0. Replicates were merged for plot (**e-i**).

Supplementary Fig.4



Supplementary Figure 4. 24nt-siRNAs retain strong enrichment of A at position 1 for *ago4*, *ago4/6/9* mutant and *ago4* or *ago4/6/9* mutant expressing wtAGO4 or D742A. **a-h,** The relative nucleotide bias of each position for 24-nt siRNAs over Pol V dependent regions in Col-0 (**a**), Ws (**b**), *ago4/Ws* (**c**), *ago4/wtAGO4* (**d**), *ago4/D742A* (**e**), *ago4/6/9* (**f**), *ago4/6/9/wtAGO4* (**g**) and *ago4/6/9/D742A* (**h**). **i,** Boxplot of normalized GRO-seq signals from top 1,000 expressed annotated gene in Col-0, nrpd1, *nrpe1*, *nrpd1/e1*, *spt5l*, *drm3*, *frg1/2*, *idn2/idl1/idl2*, *idn2*, and *suvr2*. N.S., not significant.