

Broad noncoding transcription suggests genome surveillance by RNA polymerase V

Masayuki Tsuzuki^{a,b,1}, Shriya Sethuraman^{c,1}, Adriana N. Coke^a, M. Hafiz Rothi^a, Alan P. Boyle^{d,e}, and Andrzej T. Wierzbicki^{a,2}

^aDepartment of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109; ^bDepartment of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 153-8902 Tokyo, Japan; ^cBioinformatics Graduate Program, University of Michigan, Ann Arbor, MI 48109; ^dDepartment of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109; and ^eDepartment of Human Genetics, University of Michigan, Ann Arbor, MI 48109

Edited by Craig S. Pikaard, Indiana University Bloomington, Bloomington, IN, and approved October 14, 2020 (received for review July 9, 2020)

Eukaryotic genomes are pervasively transcribed, yet most transcribed sequences lack conservation or known biological functions. In Arabidopsis thaliana, RNA polymerase V (Pol V) produces noncoding transcripts, which base pair with small interfering RNA (siRNA) and allow specific establishment of RNA-directed DNA methylation (RdDM) on transposable elements. Here, we show that Pol V transcribes much more broadly than previously expected, including subsets of both heterochromatic and euchromatic regions. At already established RdDM targets, Pol V and siRNA work together to maintain silencing. In contrast, some euchromatic sequences do not give rise to siRNA but are covered by low levels of Pol V transcription, which is needed to establish RdDM de novo if a transposon is reactivated. We propose a model where Pol V surveils the genome to make it competent to silence newly activated or integrated transposons. This indicates that pervasive transcription of nonconserved sequences may serve an essential role in maintenance of genome integrity.

Arabidopsis thaliana | RNA polymerase V | noncoding RNA | RdDM

Lukaryotic genomes confront a variety of threats to their in-tegrity. Transposable elements (TEs) are prevalent in most eukaryotic genomes, and their activity is repressed by a variety of gene silencing mechanisms. One of those processes is RNAmediated transcriptional gene silencing. In plants, it is known as RNA-directed DNA methylation (RdDM) and represses TEs, repeats, and other potentially harmful genetic elements by establishing repressive chromatin marks (1). This process relies on small interfering RNAs (siRNAs), which, in plants, are most commonly produced from precursors generated by a specialized RNA polymerase, Pol IV. RdDM also requires the presence of long noncoding RNA scaffolds, which are needed for the recognition of complementary target sequences by siRNA (2). While fungi and possibly also animals use RNA polymerase II (Pol II) to produce scaffold transcripts (1), plants use another specialized RNA polymerase, Pol V (3, 4). Several lines of evidence suggest that siRNA-AGO4 complexes recognize target loci by base pairing with nascent scaffold transcripts (4-6), although siRNA-DNA base pairing is also possible (7). Interaction of Pol V and its transcripts with siRNA-AGO4 leads to the recruitment of de novo DNA methyltransferase DRM2, which establishes DNA methylation (8, 9).

A key feature of RdDM is sequence specificity, which assures efficient targeting of TEs and prevents inadvertent silencing of endogenous genes. This specificity is determined by the recruitment of both Pol IV and Pol V to TEs. However, unlike most DNA-dependent RNA polymerases, Pol IV and Pol V do not rely on sequence-encoded promoters (10). Instead, they are recruited by preexisting repressive chromatin modifications (11–14). This explains how silencing of already repressed TEs is specifically maintained by a positive feedback of RdDM.

An important unresolved question about RdDM is the mechanism responsible for the initial silencing of newly inserted or reactivated TEs. This process requires the activity of Pol V

(15–20); however, it is unknown how Pol V is specifically recruited to TEs in the absence of preexisting DNA methylation. Another open question about RdDM is the functional relationship between Pol II, Pol IV, and Pol V, which are all involved in this process. This especially applies to Pol IV and Pol V, which are recruited by H3K9me2 and DNA methylation, respectively (11–14). Because these repressive chromatin modifications are closely functionally related, Pol IV and Pol V are expected to be recruited to the same loci. This negates the need for two specialized polymerases and raises the question why one polymerase cannot produce both siRNA precursors and scaffold transcripts, as is the case in fission yeast (1).

Here, we show that Pol V transcription is not limited to TEs and other known RdDM targets. Instead, Pol V transcribes more broadly and pervasively than previously expected. Specificity of silencing is not restricted by the recruitment of Pol V; instead, siRNA production is likely its primary determinant. Pol V is needed to facilitate silencing of reactivated TEs, presumably by making them competent to receive the silencing signal. This explains how RdDM may be established de novo and explains the functional relationship between Pol II, Pol IV, and Pol V. This also demonstrates that pervasive transcription of nonconserved genomic regions may serve an important role in maintaining genome integrity.

Results

Pol V Transcribes Broadly. The current model of RdDM predicts that Pol V transcribes only bona fide RdDM targets. To test this prediction, we developed a high-sensitivity method to detect Pol V transcription: immunoprecipitation followed by analysis of

Significance

Eukaryotic genomes are pervasively transcribed, yet most transcribed sequences lack conservation or known biological functions. We show that a specialized plant-specific RNA polymerase V broadly transcribes the *Arabidopsis* genome. We propose a model where Pol V transcription surveils the genome and is required to recognize and repress newly inserted or reactivated transposons. Our results indicate that pervasive transcription of nonconserved sequences may serve an essential role in maintenance of genome integrity.

Author contributions: M.T., S.S., A.N.C., M.H.R., and A.T.W. designed research; M.T., S.S., A.N.C., and M.H.R. performed research; M.T., S.S., A.N.C., A.P.B., and A.T.W. analyzed data; and M.T., S.S., and A.T.W. wrote the paper.

The authors declare no competing interest

This article is a PNAS Direct Submission.

Published under the PNAS license

¹M.T. and S.S. contributed equally to this work.

²To whom correspondence may be addressed. Email: wierzbic@umich.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2014419117/-/DCSupplemental.

RNA ends (IPARE) (Fig. 1*A*). It combines immunoprecipitation of the Pol V complex using an antibody specific toward NRPE1, the largest subunit of Pol V (6), with a modified nanoPARE RNA sequencing protocol (21). This method achieved a greatly improved sensitivity in detecting Pol V transcripts (*SI Appendix*, Fig. S1*A*) but not messenger RNA (*SI Appendix*, Fig. S1*B*) compared to previously used RNA immunoprecipitation (6) or global run-on sequencing (GRO-seq) (5).

Analysis of the IPARE results confirmed the presence of the anticipated Pol V transcription signal on known RdDM targets (*SI Appendix*, Fig. S1A) (3, 5, 6), which were previously shown to have high levels of CHH methylation and 24-nt siRNA (6). One locus is shown on a genome browser screenshot in Fig. 1B and SI Appendix, Fig. S1C. Surprisingly, IPARE sequencing reads from Col-0 wild type covered a relatively large proportion of the genome, which could be observed in three biological replicates of IPARE (SI Appendix, Fig. S1D) and in a merged dataset, where as much as 31.2% of the genome is covered by sequencing reads in Col-0 wild type (SI Appendix, Fig. S1D) with thinning and Fig. 1C without thinning). This is substantially more than expected based on the extent of siRNA accumulation, CHH methylation, or TE annotations (Fig. 1C). This indicates that Pol V transcribes more broadly than the estimates of RdDM prevalence.

Detection of Pol V transcription by IPARE relies on the availability of a negative control, the null *nrpe1-11* mutant, which does not contain the epitope for IP and has no strong developmental or physiological phenotypes (4, 6, 22, 23). The *nrpe1* mutant had 8.1% of the genome covered by IPARE sequencing reads, compared to 22.8% in the Col-0 wild-type dataset thinned to the same coverage as *nrpe1* (*SI Appendix*, Fig. S1D). To eliminate background signal originating from Pol I, Pol II, and Pol III, we filtered all sequencing reads based on known properties of different RNA polymerases (5, 6) (*SI Appendix*, Fig. S1E). This eliminated the vast majority of reads present in *nrpe1*

(1.9% genome coverage remaining) but only a small subset of reads from Col-0 wild type (23.3% genome coverage remaining without thinning) (*SI Appendix*, Fig. S1*F*). This confirms that the IPARE signal originates from bona fide Pol V transcripts, and Pol V transcribes more broadly than the estimates of RdDM prevalence.

To obtain a reliable and unbiased way to identify Pol V transcribed genomic regions, we split the genome into 200-bp-long bins and used a Hidden Markov Model (HMM) to identify bins with evidence of Pol V transcription (SI Appendix, Fig. S2A). Using raw read per million (RPM)-normalized Pol V IPARE sequencing data combined from three biological replicates, this approach split the genome into Pol V-transcribed (42.4%) and non-Pol V-transcribed (57.6%) bins (Fig. 1D). IPARE signal levels were not caused by stochastic mapping of sequencing reads (SI Appendix, Fig. S2 B and C) and were significantly correlated between three independent biological replicates (SI Appendix, Fig. S2D). Non-Pol V-transcribed bins include loci with no detectable transcription, transcription by other RNA polymerases, and a smaller subset of bona fide RdDM targets where the loss of RdDM in nrpe1 leads to increased transcription by other RNA polymerases. Together, these results further confirm that Pol V transcribes more broadly than previously expected.

To determine whether IPARE specifically detects Pol V transcription, we performed this assay using the *drd3-3* mutant (23), which is an allele of *NRPE1* with a point mutation in the catalytic active site (24). This mutant is expected to contain the epitope for IP but no Pol V transcripts (3). The IPARE signal in *drd3-3* was significantly lower than in Col-0 wild type (Fig. 1*E*), which confirms that IPARE specifically detects Pol V transcripts. We obtained a similar result with *dms5-1*, another allele of *NRPE1*, which contains a premature stop codon (25) (Fig. 1*E*). To further test the specificity of IPARE, we used *drd1* and *dms3* mutants, which lack subunits of the DDR complex and are expected to disrupt Pol V transcription



Fig. 1. Pol V transcribes more broadly than expected. (A) The IPARE method of detecting Pol V transcription. RT refers to reverse transcription, UMI refers to unique molecule identifiers. (*B*) Genome browser screenshot of a region transcribed by Pol V (Chromosome 1, 1878280 to 1883250). TAIR10 genome annotation, DNA methylation in CHH contexts, Pol V IPARE, and annotated Pol V transcribed regions data are shown. Data for individual biological replicates are shown in *SI Appendix*, Fig. S1C. (C) IPARE reads from Col-0 wild type cover a greater proportion of the genome than known features of RdDM including siRNA (28), annotated TEs (19), and CHH DMRs. Red bars indicate percentage of the genome covered by specific features. (*D*) HMM identifies Pol V-transcribed regions of the genome. Boxplot shows IPARE signal using combined data from three biological replicates comparing bins identified as Pol V transcribed (states 0 and 1) or non-Pol V transcribed (states 2 and 3). (*E*) Pol V IPARE signal depends on the enzymatic activity of Pol V. Boxplots show RPM-normalized IPARE signal levels at Pol V-transcribed and non-Pol V-transcribed regions in Col-0, *nrpe1* (null allele), *dms5-1* (early termination allele of *NRPE1*). Asterisk indicates Wilcoxon test *P* < 2.2e-16. (*F*) Pol V IPARE signal depends on the activity of the DDR complex. Boxplots show RPM-normalized IPARE signal levels at Pol V-transcribed and non-Pol V-transcribed and non-Pol V-transcribed regions in Col-0 and *nrpe1* as well as DDR subunit mutants *drd1* and *dms3*. Asterisk indicates Wilcoxon test *P* < 2.2e-16.

without affecting the accumulation of Pol V (3, 4, 26, 27). IPARE signal was slightly higher in *drd1* and *dms3* than in *nrpe1* (Fig. 1*F*), which is consistent with the epitope (NRPE1) being present in those mutants but could also be explained by the presence of another mechanism recruiting Pol V. It was, however, significantly lower than in Col-0 wild type (Fig. 1*F*), which further confirms that IPARE specifically captures Pol V transcripts. Although pervasive transcription is inherently difficult to demonstrate, these results indicate that alternative explanations of the broad IPARE signal are unlikely and confirm that Pol V transcribes more broadly than previously expected.

Pol V Is Not the Primary Determinant of Silencing Specificity. Broad presence of Pol V throughout the genome suggests that it may not be the primary determinant of the specificity of RdDM. To test this prediction, we designed HMM-based identification of Pol V-transcribed regions in a way that allows distinguishing RdDM targets from sequences not targeted by RdDM. This determination was possible based on whole genome bisulfite sequencing data of CHH methylation. Among four identified HMM states, two (state 0 and state 1) included Pol V-transcribed regions (Figs. 1D and 2A and SI Appendix, Fig. S3A). Regions identified as state 0 were enriched in Pol V-dependent CHH methylation (Fig. 2B), which indicates that they are bona fide RdDM targets. State 1, on the other hand, although Pol V transcribed (Fig. 2A) and abundant throughout the genome (SI Appendix, Fig. S3B), had no enrichment in Pol V-dependent CHH methylation (Fig. 2B). This indicates that there is evidence of extensive Pol V transcription outside of RdDM targets.

Although non-RdDM Pol V transcription (state 1) is clearly detectable, it accumulates at substantially lower levels than RdDM Pol V transcription (state 0; Fig. 24). To confirm that non-RdDM Pol V transcription is not an artifact of IPARE, we performed real-time RT-PCR using total RNA on arbitrarily selected non-RdDM Pol V-transcribed regions. We found 10 primer pairs that showed a substantial signal reduction in *nrpe1 (SI Appendix*, Fig. S3C), which is consistent with the presence of Pol V-dependent transcription. We further used IPARE to analyze the *drd3-3* and *dms5-1* alleles of *NRPE1*, which had significant reductions of signal on both RdDM (state 0) and non-RdDM (state 1) Pol V-transcribes both RdDM and non-RdDM genomic regions, and, therefore, presence or absence of Pol V may not be essential for the determination of the specificity of RdDM.

We further analyzed regions of non-RdDM Pol V transcription to determine where in the genome Pol V transcribes independently of RdDM. We first performed HMM-based identification of Pol V transcription independently in each individual biological replicate of IPARE and determined the percentage of genomic bins identified in at least two independent replicates. While 77% of RdDM Pol V transcripts were identified more than once, 50% of non-RdDM Pol V transcripts were identified more than once. This is consistent with detection of non-RdDM Pol V transcription being limited by sequencing depth and this category of transcripts being possibly more widespread than detected at the sequencing coverage we used. Further analysis of non-RdDM Pol V transcription indicated that it is not associated with proximity to RdDM loci (SI Appendix, Fig. S4A), and it is mostly euchromatic (SI Appendix, Fig. S4B) and C) and enriched on intergenic regions (SI Appendix, Fig. S4D). While RdDM regions had the expected high levels of CG methylation, a minor subset of non-RdDM Pol V-transcribed regions also had elevated levels of CG methylation (SI Appendix, Fig. S4E), which may be explained by RdDM-independent silencing or gene body methylation. Together, these results support the possibility that non-RdDM Pol V transcription is produced stochastically and mostly nonspecifically over a substantial fraction of the genome.

If Pol V has a limited role in determining the specificity of RdDM, Pol IV-dependent production of siRNA remains the expected alternative determinant of specificity (2). To test this possibility, we quantified previously published Pol IV-dependent 24-nt siRNA (28) on regions corresponding to four states identified by HMM. State 0, which corresponds to RdDM Pol V transcription, was enriched in Pol IV-dependent 24-nt siRNA (Fig. 2C). However, state 1, which corresponds to non-RdDM Pol V transcription, was not enriched in siRNA (Fig. 2C). We further tested the enrichment of small RNA clusters detected in Col-0 wild type on four states and found that small RNAs were enriched on RdDM Pol V transcripts but not non-RdDM Pol V transcripts or non-Pol V-transcribed regions (SI Appendix, Fig. S4F). These results indicate that the presence of siRNA is associated with RdDM, which is consistent with siRNA being the primary determinant of RdDM specificity. We conclude that Pol V is unlikely to be the primary determinant of sequence specificity of RdDM.

Pol V Is Needed for TE Resilencing. Our observations that Pol V transcribes broadly and is not the primary determinant of RdDM explains a key inconsistency in the mechanistic understanding of this process. Although de novo silencing of newly integrated or reactivated TEs seems to always require Pol V (19), no mechanisms recruiting Pol V to unsilenced TEs are known. Our data indicate that non-RdDM Pol V transcription may occur broadly enough to facilitate silencing of TEs even if preexisting repressive marks have been fully or partially lost. To test this possibility, we took advantage of the *ddm1* mutant, which disrupts the maintenance of CG methylation (29). Because, in *Arabidopsis*, a subset of TEs is silenced by CG methylation leads to reactivation of those TEs, which are subsequently resilenced by the establishment of de novo RdDM, manifested as CHH methylation (19). We



Fig. 2. Pol V transcription is not the primary determinant of silencing specificity. (*A*) HMM analysis of Pol V IPARE identified two distinct states of Pol V transcription. Boxplot shows Pol V IPARE signal from all three biological replicates combined at each of the four emission states of the HMM output. States 0 and 1 are transcribed by Pol V. States 2 and 3 show no evidence of Pol V transcription. (*B*) DNA methylation in the CHH context is present only on a subset of Pol V-transcribed regions (state 0 but not on state 1). Boxplot shows Pol V-dependent DNA methylation in the CHH context at each of the four emission states of HMM output. (*C*) RdDM is associated with the presence of siRNA. The siRNA is enriched only on a subset of Pol V-transcribed regions (state 0 but not on state 1). Boxplot shows Pol IV-dependent siRNA signal (28) at each of the four emission states of HMM output.

analyzed previously published methylome datasets from Col-0 wild type, nrpe1, ddm1, and ddm1 nrpe1 double mutant (19). Using these datasets, we identified differentially methylated regions (DMRs) between Col-0 wild type and *ddm1*, where CHH methvlation is increased in *ddm1*. We then selected DMRs that overlap non-RdDM Pol V transcription identified in Col-0 wild type by IPARE and HMM (state 1). As expected, these DMRs have reduced levels of CG methylation in *ddm1* (Fig. 3A). We then tested whether the increased CHH methylation on the tested loci requires Pol V, by analyzing the *ddm1 nrpe1* double mutant. Levels of CHH methylation were significantly lower in ddm1 nrpe1 compared to ddm1 (Fig. 3B). This suggests that even loci that are subject to non-RdDM Pol V transcription in Col-0 wild type require Pol V for TE resilencing by the establishment of RdDM in the *ddm1* mutant. This is consistent with Pol V transcribing the genome to make it competent for silencing if a transposon becomes reactivated. Although reactivated TEs are clearly distinct from newly integrated TEs, we speculate that a similar mechanism may occur when new transposons are integrated into the genome.

Non-RdDM Pol V Transcription Requires the DDR Complex. Non-RdDM Pol V transcription results in lower IPARE signals than RdDM transcription and therefore may be controlled in a unique manner. RdDM Pol V transcription requires the DDR complex, which is involved in transcription initiation and/or elongation (3, 4, 26, 27). To test whether non-RdDM Pol V transcription also depends on the DDR complex, we analyzed the RdDM and non-RdDM Pol V transcription IPARE signal in DDR mutants, *drd1* and *dms3*. Both mutants showed strong reductions of Pol V transcription on both RdDM and non-RdDM sites (Fig. 4A). This indicates that Pol V requires the DDR complex even on non-RdDM sites.

One known mechanism of Pol V recruitment involves SUVH2 and SUVH9 proteins, which recognize preexisting DNA methylation (11, 13). To test whether this mechanism is involved in both RdDM and non-RdDM Pol V transcription, we performed IPARE in *suvh2*, *suvh9*, and *suvh2/suvh9* mutants. RdDM Pol V transcription was reduced in *suvh2* and *suvh2/suvh9* mutants (Fig. 4B), which is consistent with previously published data (11, 13). Non-RdDM Pol V transcription was slightly reduced in both



Fig. 3. Pol V transcription is required for TE resilencing. (A) Loss of DNA methylation in the CG context in *ddm1*. Heatmap shows average CG methylation levels on non-RdDM Pol V-transcribed loci, which gain CHH methylation in *ddm1*. Boxplot shows the distribution of data points shown in the heatmap. Asterisk indicates Wilcoxon test P < 2.2e-16. (B) De novo CHH methylation of TEs reactivated in *ddm1* requires Pol V. Heatmap shows average CHH methylation levels on non-RdDM Pol V-transcribed loci which gain CHH methylation in *ddm1*. Boxplot shows the distribution of data points shown in the heatmap. Asterisk indicates Wilcoxon test P < 2.2e-16.

suvh2 and *suvh9* mutants and more substantially reduced in the *suvh2/suvh9* double mutant (Fig. 4B). This suggests that SUVH2 and SUVH9 might play a role in non-RdDM Pol V transcription. Signal observed in the *suvh2/suvh9* double mutant was still substantially stronger than in *nrpe1* (Fig. 4B), which indicates that other factors may also contribute to the initiation of both RdDM and non-RdDM Pol V transcription.

Proteins that work downstream of Pol V have been shown to affect processing of Pol V transcripts through slicing by AGO4, which requires SPT5L (5). To determine whether non-RdDM Pol V transcription is affected by these downstream factors, we performed IPARE in *spt5l* and *ago4* mutants. Both mutants had no major effects on the median levels of accumulation of both RdDM and non-RdDM Pol V transcripts detected by IPARE (Fig. 4*C*). This indicates that downstream factors do not strongly affect the accumulation of nascent Pol V transcripts detected in our assay.

Discussion

We propose a speculative model where Pol V stochastically transcribes a significant fraction of the genome to make it competent for silencing (Fig. 5). This includes surveillance of euchromatic sequences, which may harbor inactive transposons or may become landing sites for random integration of new transposons (Fig. 5.4). If there is no complementary siRNA, chromatin modifiers are not recruited, and Pol V transcripts are expected to quickly degrade with no consequences. However, any newly integrated or reactivated transposon triggers one of several pathways to produce siRNA (30). Newly synthesized siRNA base pairs with already available Pol V transcripts to establish initial DNA methyl marks (Fig. 5*B*). This leads to the recruitment of Pol IV and further siRNA production. At the same time, Pol V transitions from a lowlevel surveillance status to a higher rate of transcription associated with maintenance of RdDM (Fig. 5*C*).

Surveillance Pol V transcription occurs much more broadly than siRNA production and RdDM, including euchromatic loci and possibly also heterochromatic loci repressed by pathways other than RdDM. This indicates that previous studies of Pol V localization by chromatin immunoprecipitation sequencing (31) were not sensitive enough to detect the actual breadth of Pol V transcription. Although we detected Pol V transcription on 42.4% of the genome, the absence of Pol V on any of the remaining 57.6% cannot be conclusively proven, especially on loci transcribed by Pol I, Pol II, or Pol III. It is therefore possible that Pol V transcription. This would be consistent with the fact that Pol V appears to be universally required for de novo RdDM (15, 17–20, 32).

Initiation of surveillance Pol V transcription is likely to be stochastic, which is consistent with lack of sequence specificity detected for Pol V (5, 6, 27, 31). The DDR complex and SUVH2/9 are likely responsible for sequence-independent initiation and/or elongation of Pol V transcription (3, 4, 27). In contrast to loci where RdDM is already established (11), surveillance Pol V transcription is expected to be independent of preexisting chromatin modifications, which indicates that SUVH2 and SUVH9 proteins may have a broader role than binding methylated DNA (11).

In our model, surveillance Pol V transcription is expected to have no independent impact on RdDM. However, Pol V has been proposed to have other roles independent of 24-nt siRNA (33) or gene silencing (34), both of which are consistent with our model. The role of genome surveillance by Pol V is tied to the inability of siRNA-AGO4 complexes to recognize complementary target loci in the absence of Pol V (4, 7). Widespread Pol V transcription lets siRNA-AGO4 recognize target loci even if they were not previously silenced. This indicates that the

2020

loaded at UNIVERSITY OF MICHIGAN on November 16,



Fig. 4. Non-RdDM Pol V transcription requires the DDR complex. (*A*) Both RdDM and non-RdDM Pol V transcription require the DDR complex. Boxplots show Pol V IPARE signal levels in Col-0, *nrpe1*, *drd1*, and *drns3* at RdDM and non-RdDM Pol V-transcribed loci. Asterisks indicate Wilcoxon test *P* < 2.2e-16. (*B*) SUVH2 and SUVH9 contribute to both RdDM and non-RdDM Pol V transcription. Boxplots show Pol V IPARE signal levels in Col-0, *nrpe1*, *suvh2*, *suvh9*, and *suvh2/suvh9* double mutant at RdDM and non-RdDM Pol V-transcribed loci. Asterisks indicate Wilcoxon test *P* < 2.2e-16. (*C*) Proteins that work downstream of Pol V have no strong genome-wide effects on Pol V transcription. Boxplots show Pol V IPARE signal levels in Col-0, *nrpe1*, *spt51*, and *ago4* at RdDM and non-RdDM Pol V transcribed loci.

specificity of base pairing between siRNA and a highly complex pool of Pol V transcripts is essential for precise establishment of RdDM, which is consistent with high accuracy of ribonucleotide incorporation by Pol V (35). Frequency of surveillance Pol V transcription remains unknown, but low levels of those transcripts suggests that it is not a frequent event, which is consistent with a relatively low rate of ribonucleotide incorporation by Pol V (35).



Fig. 5. Speculative model explaining the role of Pol V in de novo and maintenance RdDM. (*A*) A large fraction of the genome is subject to infrequent surveillance transcription by Pol V. This includes euchromatic loci with no active TEs. The role of this transcription is to make the genome competent to initiate silencing if siRNAs become available. (*B*) Insertion and/or activation of a TE leads to siRNA production. This siRNA may initiate silencing by basepairing with already available surveillance Pol V transcripts. This leads to the establishment of first repressive chromatin marks (me). (C) Presence of repressive chromatin marks (me) leads to recruitment of Pol IV and enhanced production of siRNAs. At the same time, Pol V transitions to a higher rate of transcription, which facilitates efficient maintenance of RdDM.

Binding of siRNA-AGO4 to euchromatic surveillance Pol V transcripts leads to the recruitment of chromatin-modifying machinery (8, 9) and the establishment of initial repressive chromatin marks. This leads to a series of events that result in robust and stable RdDM. First of those events is the repression of Pol II transcription and activation of Pol IV. This stops the production of initiating siRNAs (30), which are replaced by a strong accumulation of 24-nt siRNA produced by Pol IV, RDR2, and DCL3. This is consistent with Pol IV being recruited by H3K9me2 recognized by SHH1 (12, 14). The second event is a strong increase in the level of Pol V transcription, which allows robust reestablishment of repressive chromatin marks and efficient maintenance of RdDM. Because assays used in previous studies were not sensitive enough to detect surveillance Pol V transcription and only reported more abundant Pol V transcription on already silenced loci, this is consistent with the reported importance of preexisting DNA methylation for Pol V recruitment (11, 13). The mechanism responsible for transition of Pol V from low-level surveillance transcription to a higher rate of transcription may involve DNA methylation; however, the presence of CG methylation (6) and partial involvement of SUVH2 and SUVH9 proteins (Fig. 4B) do not fully explain this transition. This indicates that other properties of chromatin are likely to be important for transition of Pol V into high transcription rate.

The surveillance model of Pol V transcription predicts that siRNA incorporated into a proper AGO protein should be sufficient to initiate RdDM. This suggests the presence of a threshold mechanism, which prevents silencing by stochastically produced siRNAs. Existence of a threshold could explain why artificial tethering of Pol V to a reactivated *FWA* locus results in reestablishment of silencing (11, 16). Active *FWA* in *fwa-4* epiallele may accumulate a low level of siRNAs of atypical lengths, which are insufficient to initiate RdDM (16). We propose that the enhancement of Pol V transcription by tethering a DDR subunit to *FWA* lowers the threshold and allows reinitiation of RdDM.

Transition of polymerase activities during the initial establishment of RdDM explains the functional relationship between three RNA polymerases involved in this process. Aberrant Pol II transcription is the initial signal that targets newly inserted or activated TEs for silencing and is replaced by Pol IV, as previously demonstrated (30). A role of Pol II at this step of silencing is further supported by Zheng et al. (36). Pol V is functionally distinct in that it provides less or possibly even no sequence specificity for de novo RdDM. However, after silencing has been

Tsuzuki et al.

established, a higher level of Pol V transcription facilitates efficient maintenance of repressive chromatin states.

Broad transcription by Pol V changes our understanding of pervasive transcription, which, in the absence of sequence conservation, has been interpreted as nonfunctional (37). It provides evidence that noncoding transcription of nonconserved sequences may serve an important role in maintaining genome integrity. Given the conservation of transcriptional silencing mechanisms (38), a similar process may exist outside of the plant kingdom.

Methods

SI Appendix, Materials and Methods contains descriptions of performed experiments, including plant material used in this study and the IPARE method. It also includes descriptions of data analysis, including processing of

- 1. R. Martienssen, D. Moazed, RNAi and heterochromatin assembly. *Cold Spring Harb. Perspect. Biol.* 7, a019323 (2015).
- M. A. Matzke, R. A. Mosher, RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* 15, 394–408 (2014).
- A. T. Wierzbicki, J. R. Haag, C. S. Pikaard, Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135, 635–648 (2008).
- A. T. Wierzbicki, T. S. Ream, J. R. Haag, C. S. Pikaard, RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat. Genet.* 41, 630–634 (2009).
- W. Liu et al., RNA-directed DNA methylation involves co-transcriptional small-RNA-guided slicing of polymerase V transcripts in Arabidopsis. Nat. Plants 4, 181–188 (2018).
- G. Böhmdorfer et al., Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. eLife 5, e19092 (2016).
- S. Lahmy et al., Evidence for ARGONAUTE4-DNA interactions in RNA-directed DNA methylation in plants. Genes Dev. 30, 2565–2570 (2016).
- G. Böhmdorfer et al., RNA-directed DNA methylation requires stepwise binding of silencing factors to long non-coding RNA. Plant J. 79, 181–191 (2014).
- X. Zhong et al., Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell 157, 1050–1060 (2014).
- J. M. Wendte, C. S. Pikaard, The RNAs of RNA-directed DNA methylation. *Biochim. Biophys. Acta. Gene Regul. Mech.* 1860, 140–148 (2017).
- L. M. Johnson et al., SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature 507, 124–128 (2014).
- J. A. Law et al., Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature 498, 385–389 (2013).
- Z.-W. Liu 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).
- H. Zhang et al., DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proc. Natl. Acad. Sci. U.S.A. 110, 8290–8295 (2013).
- D. Fultz, R. K. Slotkin, Exogenous transposable elements circumvent identity-based silencing, permitting the dissection of expression-dependent silencing. *Plant Cell* 29, 360–376 (2017).
- J. Gallego-Bartolomé et al., Co-targeting RNA polymerases IV and V promotes efficient de novo DNA methylation in Arabidopsis. Cell 176, 1068–1082.e19 (2019).
- A. D. McCue et al., ARGONAUTE 6 bridges transposable element mRNA-derived siR-NAs to the establishment of DNA methylation. EMBO J. 34, 20–35 (2015).
- S. Nuthikattu 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).
- K. Panda *et al.*, Full-length autonomous transposable elements are preferentially targeted by expression-dependent forms of RNA-directed DNA methylation. *Genome Biol.* **17**, 170 (2016).
- L. Wu, L. Mao, Y. Qi, Roles of dicer-like and argonaute proteins in TAS-derived small interfering RNA-triggered DNA methylation. *Plant Physiol.* 160, 990–999 (2012).

sequencing reads, read scoring, identification of Pol V transcripts by HMM, analysis of Pol V transcripts, and data visualization.

Data Availability. High-throughput sequencing data have been deposited in Gene Expression Omnibus (GEO) database (accession no. GSE146913).

ACKNOWLEDGMENTS. We thank Györgyi Csankovszki for critical reading of the manuscript, University of Michigan Advanced Genomics Core for their support with high-throughput sequencing, and Jakub Dolata for providing material for whole genome bisulfite sequencing. Research reported in this publication was supported by the National Institute of General Medical Sciences of the NIH under Award R01GM108722 to A.T.W. M.T. was supported by Japan Society for the Promotion of Science (JSPS) Overseas Research Fellowship number 201860706 and JSPS KAKENHI Grant 19K23715.

- M. A. Schon, M. J. Kellner, A. Plotnikova, F. Hofmann, M. D. Nodine, NanoPARE: Parallel analysis of RNA 5' ends from low-input RNA. *Genome Res.* 28, 1931–1942 (2018).
- D. Pontier 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).
- T. Kanno et al., Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat. Genet. 37, 761–765 (2005).
- S. Lahmy et al., PolV(PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-specific subunit. Proc. Natl. Acad. Sci. U.S.A. 106, 941–946 (2009).
- T. Kanno et al., A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat. Genet.* 40, 670–675 (2008).
- J. A. Law et al., A protein complex required for polymerase V transcripts and RNAdirected DNA methylation in Arabidopsis. Curr. Biol. 20, 951–956 (2010).
- X. Zhong 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).
- M. Zhou, A. M. S. Palanca, J. A. Law, Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. *Nat. Genet.* 50, 865–873 (2018).
- A. Vongs, T. Kakutani, R. A. Martienssen, E. J. Richards, Arabidopsis thaliana DNA methylation mutants. Science 260, 1926–1928 (1993).
- D. Cuerda-Gil, R. K. Slotkin, Non-canonical RNA-directed DNA methylation. *Nat. Plants* 2, 16163 (2016).
- A. T. Wierzbicki et al., Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. Genes Dev. 26, 1825–1836 (2012).
- D. M. Bond, D. C. Baulcombe, Epigenetic transitions leading to heritable, RNAmediated de novo silencing in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U.S.A.* 112, 917–922 (2015).
- O. Pontes, P. Costa-Nunes, P. Vithayathil, C. S. Pikaard, RNA polymerase V functions in Arabidopsis interphase heterochromatin organization independently of the 24-nt siRNA-directed DNA methylation pathway. *Mol. Plant* 2, 700–710 (2009).
- 34. W. Wei et al., A role for small RNAs in DNA double-strand break repair. Cell 149, 101–112 (2012).
- M. Marasco, W. Li, M. Lynch, C. S. Pikaard, Catalytic properties of RNA polymerases IV and V: Accuracy, nucleotide incorporation and rNTP/dNTP discrimination. *Nucleic Acids Res.* 45, 11315–11326 (2017).
- B. Zheng et al., Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes Dev. 23, 2850–2860 (2009).
- T. H. Jensen, A. Jacquier, D. Libri, Dealing with pervasive transcription. *Mol. Cell* 52, 473–484 (2013).
- A. Zoch et al., SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation. Nature 584, 635–639 (2020).