

The NRPD1 N-terminus contains a Pol IV-specific motif that is critical for genome surveillance in Arabidopsis

Laura Ferrafiat¹, David Pflieger¹, Jasleen Singh^{2,3}, Michael Thieme⁴, Marcel Böhner¹, Christophe Humber¹, Aude Gerbaud¹, Etienne Bucher⁴, Craig S. Pikaard^{2,3} and Todd Blevins^{1,*}

¹Institut de Biologie Moléculaire des Plantes, CNRS, Université de Strasbourg, F-67084 Strasbourg, France,

²Howard Hughes Medical Institute, Indiana University, Bloomington, IN 47405, USA, ³Department of Biology, Indiana University, Bloomington, IN 47405, USA and ⁴Botanisches Institut, Universität Basel, CH-4056 Basel, Switzerland

Received March 03, 2019; Revised July 03, 2019; Editorial Decision July 06, 2019; Accepted July 11, 2019

ABSTRACT

RNA-guided surveillance systems constrain the activity of transposable elements (TEs) in host genomes. In plants, RNA polymerase IV (Pol IV) transcribes TEs into primary transcripts from which RDR2 synthesizes double-stranded RNA precursors for small interfering RNAs (siRNAs) that guide TE methylation and silencing. How the core subunits of Pol IV, homologs of RNA polymerase II subunits, diverged to support siRNA biogenesis in a TE-rich, repressive chromatin context is not well understood. Here we studied the N-terminus of Pol IV's largest subunit, NRPD1. Arabidopsis lines harboring missense mutations in this N-terminus produce wild-type (WT) levels of NRPD1, which copurifies with other Pol IV subunits and RDR2. Our *in vitro* transcription and genomic analyses reveal that the NRPD1 N-terminus is critical for robust Pol IV-dependent transcription, siRNA production and DNA methylation. However, residual RNA-directed DNA methylation observed in one mutant genotype indicates that Pol IV can operate uncoupled from the high siRNA levels typically observed in WT plants. This mutation disrupts a motif uniquely conserved in Pol IV, crippling the enzyme's ability to inhibit retrotransposon mobilization. We propose that the NRPD1 N-terminus motif evolved to regulate Pol IV function in genome surveillance.

INTRODUCTION

Genome surveillance pathways have evolved in eukaryotes to limit deleterious mutations by recognizing and si-

lencing transposable elements (TEs). Because TEs vary in sequence and replication mechanism, animals and plants use RNA silencing to guide repressive effects to TEs (1–3). A specialized non-coding RNA machinery, involving multisubunit RNA polymerases IV and V (Pol IV and Pol V), is the key nuclear pathway targeting TEs in plants (4,5). To initiate this RNA-directed DNA methylation (RdDM) process, Pol IV transcribes TEs and repeat-associated genes into primary transcripts that are templates for RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) (6). RDR2 synthesizes double-stranded RNAs (dsRNAs) that are processed into 24 nt small interfering RNAs (siRNAs) by DICER-LIKE 3 (DCL3) (6–10). These siRNAs are loaded onto ARGONAUTE 4 (AGO4)-clade proteins (11,12) and resulting AGO4–siRNA complexes appear to find TEs via base-pairing of the siRNA guide to nascent Pol V transcripts in chromatin (13,14). The *de novo* cytosine methyltransferase DRM2 is then recruited, resulting in TE methylation, repressive histone modifications and silencing (15).

In heat-stressed *Arabidopsis thaliana* (*Arabidopsis*), Pol IV-RdDM inhibits the replication and transgenerational mobilization of *ATCOPIA78/ONSEN* retrotransposons (16–18). To accomplish this, DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) methylates cytosines in CG, CHG and CHH sequence contexts, where H is A, C or T. However, most TEs are also silenced by DNA methylation and chromatin modification systems that are siRNA-independent (19–21). METHYLTRANSFERASE 1 (MET1) maintains CG methylation while influencing CHG and CHH methylation (22,23); CHROMOMETHYLASE 3 (CMT3) maintains CHG methylation while reinforcing CHH methylation at certain targets (24); and CHROMOMETHYLASE 2 (CMT2) methylates a subset of CHG and CHH sites, typically in the bodies of longer

*To whom correspondence should be addressed. Todd Blevins. Tel: +33 0 3 67 15 53 28; Fax: +33 0 3 67 15 53 00; Email: todd.blevins@ibmp-cnrs.unistra.fr
Present addresses:

Michael Thieme, Department of Plant & Microbial Biology, Universität Zürich, CH-8008 Zürich, Switzerland.

Etienne Bucher, Agroscope, Research Division Plant Breeding, CH-1260 Nyon, Switzerland.

TEs (25,26). Biochemical feedback loops link methylcytosine maintenance to repressive histone modifications (notably, histone 3 lysine 9 dimethylation; H3K9me2), reinforcing the silent chromatin states that inhibit TE activity (27,28). Moreover, the nucleosome remodeler DDM1 and histone deacetylase HDA6 facilitate a significant fraction of DNA methylation in plants (27). Because of these overlapping silencing machineries, the role of Pol IV in TE repression is most evident in MET1, DDM1 or HDA6-deficient backgrounds, where RdDM compensates for losses in H3K9me2 and CG methylation to limit TE reactivation (2,19–21,29).

Pol IV is a twelve-subunit enzyme that evolved in plants as a specialized form of eukaryotic RNA polymerase II (Pol II) (30). Pol IV localization to TEs is mediated by the SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1) protein, which recognizes H3K9me2 and unmethylated H3K4; about 50% of Pol IV-dependent siRNA clusters depend on SHH1 (31). In addition, the SNF domain-containing CLASSY proteins (CLSY1/2/3/4) facilitate locus-specific methylation via their interactions with Pol IV and SHH1 (32–35). Little is known about how the core domains of Pol IV have evolved to support its unique function in silencing. The insensitivity of Pol IV to α -amanitin, its elevated error rate, its physical coupling to RDR2, and the Pol IV-RDR2 complex's short dsRNA products all distinguish Pol IV transcription from Pol II transcription (8–10,33,36). Nevertheless, similar to Pol IV loss-of-function, inhibiting Pol II boosts the activity of certain retrotransposons because Pol II generates precursors for siRNAs that trigger 'non-canonical' RdDM (18,37–39). Many questions thus remain about how the activities of Pol II, Pol IV and Pol V are differentiated, balanced and regulated *in vivo* to prevent TE proliferation (36,39–41).

Screens for RNA interference factors using transgene-encoded silencers, like the potato virus X amplicon system or dsRNA directed against endogenous genes, have yielded numerous *pol IV* mutations in Arabidopsis (32,42–45). However, past molecular analyses of Pol IV *in vivo* function have typically focused on null alleles of its major subunits (NRPD1 and NRPD2), which either block Pol IV accumulation altogether or abolish Pol IV enzymatic activity by destroying the RNA polymerase active site (21,31,44,46–50). In addition, null alleles of the NRPD4/E4 subunit, which functions in both Pol IV and Pol V, were found to reduce siRNA production from a subset of Pol IV-dependent loci (45).

Here we studied missense mutations in the N-terminus of Arabidopsis NRPD1 that disrupt Pol IV function by changing amino acids remote from the RNA polymerase active site. The mutants do not disrupt Pol IV subunit assembly or RDR2 association but show partial derepression of TEs. Small RNA sequencing and methylome analyses of the N-terminus mutants indicate that Pol IV can mediate RdDM at subfeatures of TEs without generating the high siRNA levels typical of wild-type (WT) plants. Phylogenetic analysis of one mutation's context revealed an N-terminal motif, uniquely conserved in Pol IV, which facilitates 24 nt siRNA production and CHH methylation across the length of TEs. Disrupting this motif cripples the ability of Pol IV to inhibit *ONSEN* retrotransposon mobilization. We propose

that this NRPD1 subdomain evolved to facilitate RdDM and genome surveillance, illuminating its potential role in regulating Pol IV transcription more generally.

MATERIALS AND METHODS

Plant materials

Arabidopsis thaliana (Arabidopsis) T-DNA insertion mutants *nRPD1-3*, *nRPE1-11* and *nRPD/e2-2* were described in (46,51). The *nRPD1* point mutations were obtained in an EMS screen using the *SUC2::IR-SUL* silencing reporter transgene in a *35S::DCL4* background (SucSul D4). Each *nRPD1* point mutant was backcrossed to WT Col-0 to eliminate *35S::DCL4* and restored to a homozygous state prior to molecular analyses shown in Figures 3, 5 and 6. The *SUC2::IR-SUL* transgene was recovered in a homozygous state after backcrossing because it is linked to the *NRPD1* locus. Mutants for DNA methyltransferases (*drm2-2*, *cmt3-11*) or RNA-dependent RNA polymerases (*rdr2-2*, *rdr6-11*) were obtained from the Arabidopsis Biological Resource Center. *RDR2-FLAG rdr2-2* lines were generated by recombining a pENTR-RDR2 genomic clone into pEarleyGate302 and complementing the *rdr2-2* null mutant by Agrobacterium transformation. To enable Pol IV-RDR2 co-immunoprecipitation and Pol IV *in vitro* transcription experiments in *nRPD1* mutant backgrounds, *RDR2-FLAG rdr2-2* was crossed to each of the *nRPD1-47* through *nRPD1-51* mutants. F2 progeny were then selected in which each *nRPD1* allele was homozygous and RDR2-FLAG was robustly expressed. Double mutants of each *nRPD1* mutant with *drm2-2*, *cmt3-11*, *nRPE1-11* and *rdr6-11* were obtained by crossing the corresponding lines and selecting homozygous F2 progeny by polymerase chain reaction (PCR)-based genotyping.

Heat stress and retrotransposon detection

Surface-sterilized seed from Arabidopsis control (WT Col-0, WT SucSul) and *nRPD1* mutant lines were grown asexually in a Sanyo MLR-350 chamber on solid 0.5X MS medium (1% sucrose, 0.5% Phytigel (Sigma), pH 5.8) under long day conditions (16 h light) at 24°C (day) and 22°C (night). After one week of growth, plants were either exposed to a control stress (CS, 24 h at 6°C followed by 24 h at control conditions) or to an acute heat stress (24 h at 6°C followed by 24 h at 37°C). These treatments, tissue sampling and qPCRs to measure *ONSEN*-copy numbers were conducted as described previously (16,18) (see Supplementary Table S1 for primer sequences).

Antibodies

Native antibodies specific for the catalytic subunits of Pol IV (NRPD1 and NRPD2) were raised using a commercial service (Eurogentec). Two rabbits per target protein were inoculated, respectively, with a C-terminal peptide from NRPD1 (CLKNGTLESGGFSENPE) or with an N-terminal peptide of NRPD/E2 (MPDMDIDVKDLEEFEC). Serum aliquots from the final bleeds were affinity purified on columns using the peptide corresponding to each original inoculation (Eurogentec). Antibody specificity was tested by extracting total protein from WT, *nRPD1-3* null and

nripd/e2-2 null plants, then performing a western blot and observing the loss of the appropriately sized bands in each null mutant compared to WT samples (see Figure 1B). The monoclonal Anti-FLAG-HRP M2 antibody used to detect RDR2-FLAG was a commercial reagent (Sigma).

Western blotting

Denaturing protein extraction was performed on inflorescence tissue of Arabidopsis following (52). Resulting protein pellets were resuspended in a buffer composed of 10% glycerol, 3% sodium dodecyl sulphate (SDS), 62.3 mM Tris-HCl pH 8.0, 1× Complete ethylenediaminetetraacetic acid (EDTA)-free Protease Inhibitor Cocktail (Sigma) and 2% β-mercaptoethanol. Protein amounts were quantified by Lowry (Bio-Rad), the concentrations were adjusted and 4× Laemmli Buffer (0.25 M Tris-HCl, 8% SDS, 40% glycerol, 0.01% bromophenol blue, 10% β-mercaptoethanol) was added before storage at −20°C. Samples were thawed at 95°C for 3 min. A total of 600 μg were separated on a 6% SDS-PAGE, transferred on a Immobilon-P membrane (Millipore IPVH00010). The membrane was blocked 30', incubated overnight with the primary 'anti-NRPD1' antibody (1:5000 dilution), washed and incubated with the secondary antibody coupled to horseradish peroxidase. Chemiluminescent western signals were detected on film (Fuji Medical X-ray Medical Film) using the Lumi-Light Plus Western Blotting Substrate kit (Roche). The membrane was stripped for 12 min in Restore PLUS Western Blot Stripping Buffer (Thermo Scientific), washed, blocked and then incubated with the primary 'anti-NRPD/E2' antibody (1:2500 dilution).

Protein co-immunopurification

About 170 mg of Arabidopsis inflorescences were ground in liquid nitrogen and then suspended in 1.5 ml lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM MgCl₂, 0,1% NP-40, 200 μl Protease inhibitor, 1 mM PMSF) with continued grinding at 4°C for 10 min. Resulting extracts were transferred to microfuge tubes and centrifuged at 16 000 rcf and 4°C for 5 min. An aliquot of each supernatant was stored at −20°C (input), then the remainder was transferred to a tube containing 50 μl of re-suspended Miltenyi (μMACS 'DYKDDDDK' kit, equivalent to anti-FLAG) beads and this mixture was incubated at 4°C for 35 min on a wheel (8 rpm). Miltenyi columns were installed in the manufacturer's magnetic stand, prepped with 200 μl of lysis buffer and progressively loaded with 200 μl volumes of the supernatant-bead mixtures. Each column was washed six times with lysis buffer. Remaining liquid was removed and 45 μl of preheated Laemmli buffer (95°C) was added for a 5 min incubation with the columns stoppered. Three additional 30 μl aliquots of Laemmli buffer (95°C) were added, allowing protein elution from the columns. The combined eluate was mixed, heated for 5 min at 95°C and stored at −20°C. For western blotting, 30 μl of input or 30 μl of each sample were used following the previous description. Monoclonal Anti-FLAG-HRP M2 (dilution 1:15 000) was used to detect RDR2-FLAG protein.

Semi-quantitative/quantitative reverse transcription PCR (RT-PCR)

Total RNA was extracted from Arabidopsis inflorescence tissue following instructions of the RNeasy Kit (Qiagen), treated with DNase I (ThermoFisher Scientific) and then re-purified using phenol-chloroform extraction followed by ethanol precipitation. Then, 1 μg of DNase-treated RNA was aliquoted for random-primed cDNA synthesis using SuperScript IV Reverse Transcriptase (ThermoFisher Scientific) at 50°C. For each RT-PCR panel, a control omitting SuperScript IV was also prepared. For semi-quantitative RT-PCR fragments were amplified with *AtSN1* or *ACT2*-specific primers and visualized on an agarose gel after ethidium bromide staining. For quantitative RT-PCR, the synthesized cDNA was subjected to real-time PCR and detected via SYBR Green fluorescence in the LightCycler 480 II instrument (Roche Applied Science). See Supplementary Table S1 for primer sequences.

Pol IV *in vitro* transcription assays

Transcription assays were carried out as described in (6). Briefly, 3 g of 3-week-old Arabidopsis seedlings were flash frozen in liquid-nitrogen and then lysed in 14 ml of lysis buffer (20 mM Tris-Cl, pH 7.6; 150 mM sodium sulfate; 5 mM magnesium sulfate; 20 μM zinc sulfate; 1 mM PMSF; 5 mM Dithiothreitol (DTT) and 1× Plant Protease Inhibitor Cocktail (Sigma)). Crude lysates were centrifuged at 18 000 rcf for 15 min and the soluble fractions were incubated with 25 μl of anti-FLAG M2 agarose resin (Sigma) for 2.5 h at 4°C to immunoprecipitate Pol IV-RDR2-FLAG complexes. The resin was washed twice with 15 ml of lysis buffer (without Plant Protease Inhibitor Cocktail) followed by 15 ml of low salt wash buffer (20 mM HEPES-KOH, pH 7.6; 100 mM potassium acetate; 5 mM magnesium sulfate; 20 μM zinc sulfate; 10% glycerol; 1 mM PMSF and 5 mM Dithiothreitol). The resin was then resuspended in low salt wash buffer to 50 μl, followed by addition of transcription buffer to a transcription reaction volume of 100 μl.

The template DNA, non-template DNA and RNA primer oligos were synthesized by Integrated DNA Technologies and polyacrylamide gel electrophoresis (PAGE)-purified. A total of 2 μM RNA primer was end-labeled using T4 polynucleotide kinase (NEB) in the presence of 25 μCi [γ-³²P]-ATP (6000 Ci/mmol, Perkin Elmer) and in a total volume of 50 μl. Equimolar amounts of template DNA, end-labeled RNA primer and 10% excess of non-template DNA were mixed in the annealing buffer (30 mM HEPES-KOH pH 7.6, 100 mM potassium acetate), brought to 100°C and slowly cooled to room temperature to obtain the template used in the transcription assays.

Transcription reactions were carried out in a buffer with the following final composition: 20 mM HEPES-KOH pH 7.6, 100 mM potassium acetate, 60 mM ammonium sulfate, 10 mM magnesium sulfate, 10% v/v glycerol, 20 μM zinc sulfate, 0.1 mM PMSF, 1 mM DTT, 0.8U/μl Ribolock (Thermo Fisher), 1 mM each of ATP, GTP, CTP and UTP and 25 nM template. The reactions were incubated at room temperature for 1 h on a rotisserie, stopped by heat denaturation at 70°C for 5 min followed by desalting using Performa spin columns (Edgebio). Transcription reactions

were adjusted to 0.3 M sodium acetate, and 15 μ g of Glycoblue (ThermoFisher) and 3 volumes of isopropanol were added. Following centrifugation at 16 000 \times g, 15 min, nucleic acid pellets were washed 2 \times with 70% ethanol, resuspended in 2 \times RNA loading dye (NEB) and resolved on 15% denaturing polyacrylamide gels (45 \times 20 cm). The gels were dried for 2 h on a vacuum gel drier at 80°C and the signal was developed using autoradiography.

Small RNA blot hybridization

Total RNA was extracted from Arabidopsis inflorescences using TRIzol reagent (Invitrogen) and size-fractionated as described in (53). Then, 9 μ g low molecular weight RNA was resuspended in 8 μ l RNA Loading Buffer (95% formamide, 0.025% bromophenol blue, 0.025% xylene cyanol FF, 5 mM EDTA, 0.025% SDS, pH 8.5). Samples were heated to 95°C for 3 min and separated on an 16% polyacrylamide gel. RNA loading was documented using Ethidium bromide gel staining followed by UV transillumination. Size-separated RNAs were transferred to a nylon membrane (Hybond-N+, GE Healthcare) by electroblotting and UV cross-linked (140 mJ/cm²). Different ³²P 5'-end-labeled DNA oligonucleotides were used for successive hybridizations in PerfectHyb Plus Buffer (Sigma) overnight at 35–40°C, depending on the probe. The membrane was washed three times for 20 min in Wash Buffer (0.3 M NaCl, 30 mM sodium acetate, 0.5% SDS, pH 7.0), exposed to a phosphor-imager screen for 3 days, then the screen was scanned using a Typhoon Multimode-imager (GE Healthcare). Each probe was stripped with boiling 0.1% SDS (two times, 20 min) prior to the next hybridization (see Supplementary Table S1 for probe sequences).

Small RNA sequencing

Total RNA was extracted from Arabidopsis inflorescences using TRIzol reagent, treated with DNase I (ThermoFisher Scientific) and then re-purified using phenol-chloroform extraction followed by ethanol precipitation. About 5 μ g of DNase-treated RNA was sent for library preparation: ~15–94 nt RNAs were selected by polyacrylamide size-separation, Illumina TruSeq small RNA-seq libraries were prepared (2 \times replicates per genotype) and the libraries sequenced on an HiSeq 2500 platform (1 \times 125 bp, Fasteris SA, <http://www.fasteris.com>). The 3'-adapter sequences were removed by the Fasteris data pipeline. These trimmed reads were quality-filtered ($q > 30$) using Cutadapt v1.14 (<https://doi.org/10.14806/ej.17.1.200>) and mapped to the Arabidopsis TAIR10 reference genome using Bowtie v1.2.2 (<http://bowtie-bio.sourceforge.net/index.shtml>) without allowing mismatches but permitting multi-mapped reads (Supplementary Tables S7 and 8). Small RNA counts were extracted using ShortStack v3.8.5 (<http://sites.psu.edu/axtell/software/shortstack/>) and normalized by the total number of mapped reads. Boxplots of 24 nt siRNAs were generated in R, counting reads per kb per million mapped (RPKM) within *pol IV* differentially methylated regions (DMRs) (loci hypomethylated in *nppd1-51* versus WT SucSul). Wilcoxon rank sum tests were then performed using the ggpubr package (Supplementary Table S12).

DNA methylation detection

DNA was extracted from Arabidopsis inflorescences using the Nucleon Phytopure kit (GE Healthcare) following the manufacturer's recommendations and including RNase A treatment. *Chop-PCR*: 90 ng of genomic DNA was digested with HaeIII or AluI, alongside reactions aliquots from which the restriction enzyme was omitted (no digest controls), as described in (21). Target loci were then amplified by PCR or qPCR with primers flanking the restriction sites (Supplementary Table S1). *Amplicon-based bisulfite sequencing* was performed using the EpiMark Bisulfite Conversion Kit (New England Biolabs). PCR fragments amplified using bisulfite-treated DNA and the primers AtSN1-Bi-F and AtSN1-Bi-R (Supplementary Table S1) were cloned into pGEM-T-Easy and Sanger sequenced. For each genotype, at least 38 *AtSN1* bisulfite clones were aligned in Geneious, analyzed using CyMATE (<http://www.cymate.org/>) and plotted in Excel and R. Boxplots of percentage DNA methylation on three distinct *AtSN1* intervals were generated, and Wilcoxon rank sum tests were performed using the ggpubr package in R. *Whole-genome bisulfite sequencing* (WGBS) was performed by Beijing Genomics Institute (BGI, <https://www.bgi.com/>) on 18 samples (2 \times replicates, 9 genotypes) using a 2 \times 150 bp Illumina HiSeq run to obtain ~38 million reads per sample. WGBS read quality and mapping stats are provided in Supplementary Table S4.

Differentially methylated regions

Adapter and quality trimming ($q > 20$) were performed on the WGBS data using TrimGalore (v0.4.4). Clean reads were mapped to the Arabidopsis reference genome (TAIR10) using Bismark v0.18.1 (<https://www.bioinformatics.babraham.ac.uk/projects/bismark/>). Methylation information for each methylcytosine context (CG, CHG and CHH) was extracted (*bismark_methylation_extractor*) after de-duplication (*deduplicate_bismark*). DMRs were identified from the Bismark analysis files using the BSseq R package (v1.10.0). WT SucSul was used as the reference sample for calling DMRs. These DMRs were called using the BSseq default t-stats quantile cutoff and only including cytosine positions supported by at least four reads in both replicates (Supplementary Tables S5 and 6). Furthermore, the DMRs were filtered using the following stringent criteria: minimum 100 bp length, more than five total Cs and minimum differences in methylation level of 40, 20 and 10%, respectively, for the CG, CHG and CHH sequence contexts. All graphics were generated in R using ggplot2 and ggpubr packages (<https://www.r-project.org/>).

Protein sequence alignment

Amino acid (aa) sequences for the largest subunits of DNA-dependent RNA polymerases were obtained from Uniprot, NCBI and Phytozome (Supplementary Table S10), including NRPD1 (Pol IV) from 17 diverse species (54). All sequences were imported into Geneious (v11.1.5) (<https://www.geneious.com>). Sc.RPB1 was hand-annotated with evolutionarily conserved 'Domains A to H' of

Pol I/II/III/IV/V and with point mutations known to affect Pol II activity (46,48,55–57). Ath.NRPD1 was annotated with the *nripd1-47*, *nripd1-49* and *nripd1-50* mutations (Supplementary Figure S1B). Ath.NRPD1 and Ath.NRPE1 were annotated with the DeCL/DUF3223 domains and the WG repeat region (44,46,58). NRPD1 sequences were aligned using MUSCLE (v3.8.425, default parameters), then Ath.NRPB1, Sc.RPB1 and Ath.NRPE1 were introduced to this alignment using the Geneious profile-based aligner. Figure 4, and Supplementary Figures S2.A and S4 represent views of the same global alignment with species shifted top/bottom, or omitted depending on space available in each panel (e.g. a long insertion between ‘Domain A’ and ‘Domain B’ of *Ginkgo biloba* NRPD1 and *P. canariensis* NRPD1 prevented their inclusion in Figure 4). We generated a Hidden Markov Model based on the 22 aa Pol IV-specific motif region in the alignment and queried UniProt Reference Proteomes using *hmmsearch* (<https://www.ebi.ac.uk/Tools/hmmer/>). The list of high quality hits (*E*-value < 0.01) included proteins from 46 distinct plant species: all proteins >1300 aa were downloaded and re-analyzed by the same procedures as outlined above to scan for NRPB1, NRPD1 and NRPE1 domains/subdomains (Supplementary Table S11).

RESULTS

Point mutations in the N-terminus of Pol IV’s largest subunit

To obtain an allelic series of *nripd1* mutations in Arabidopsis, EMS-mutagenized seed pools were screened using a *SUC2::IR-SUL* transgenic reporter. In this system (43), the Arabidopsis *SUL* mRNA is silenced by *SUL* dsRNA arising from an inverted-repeat (*IR-SUL*) under control of the Arabidopsis *SUC2* promoter (Supplementary Figure S1A). Plants carrying the silencer (WT SucSul) show vascular bleaching due to knock-down of the *SUL* magnesium chelatase, in contrast to uniformly green leaves of untransformed plants (WT Col-0) (Figure 1A). To preclude recovery of *dcl4* mutations, which are frequently retrieved in *SUC2::IR-SUL* screens (59), the mutagenized parental line also harbored a *DCL4* transgene driven by the CaMV 35S promoter (WT SucSul D4). The subsequent M2 plant generation was screened for individuals that lack vascular bleaching despite the presence of *SUC2::IR-SUL*; these candidates were analyzed for *NRPD1* gene mutations by PCR amplification and Sanger sequencing (Supplementary Figure S1B).

The five point mutations isolated in *NRPD1* were designated *nripd1-47* to *nripd1-51*, numbering up from the last reported *nripd1* alleles (42). The *nripd1-48* mutation generates stop codons in Exon 3 due to a splicing defect, whereas the *nripd1-51* mutation substitutes an early stop for tryptophan (W664*), suggesting that both are *pol IV* null alleles. By contrast, *nripd1-47*, *nripd1-49* and *nripd1-50* are missense mutations in the NRPD1 N-terminus within and adjacent to the evolutionarily conserved ‘Domain A’ of multisubunit RNA polymerases (57). The corresponding WT NRPD1 amino acids are not thought to contribute to the Pol IV active site (Figure 1A) (9,48). As expected, given the premature stop codons, NRPD1 protein was not detectable in *nripd1-48* or *nripd1-51* point mutants, pheno-

copying the *nripd1-3* null mutant (T-DNA insertion). However, *nripd1-47*, *nripd1-49* and *nripd1-50* missense mutants expressed WT or slightly higher levels of NRPD1 protein (Figure 1B, top panel). The second largest subunit of Pol IV, NRPD2, accumulated equally in WT controls and in all the *nripd1* mutants, but was not detectable in the *nripd1/e2-2* null mutant control (Figure 1B, middle panel).

Genetic lesions in conserved ‘Domain A’ of NRPD1

To explore the impact of NRPD1 N-terminus mutations on Pol IV, we aligned protein sequences of RPB1 from yeast (*Saccharomyces cerevisiae*, Pol II), NRPB1 from Arabidopsis (Pol II), NRPD1 from 17 phylogenetically diverse plant species and NRPE1 from Arabidopsis (Pol V). Focusing on conserved ‘Domain A’ in this alignment (Supplementary Figure S2A), we noticed that *nripd1-47* and *nripd1-49* affect amino acids corresponding to a zinc-binding domain of yeast Pol II (56,57). The *nripd1-50* mutation does not change a residue in ‘Domain A’ so we address its context in a later section. Comparison of the Pol IV amino acids mutated in *nripd1-47* and *nripd1-49* to homologous positions in yeast Pol II situates these highly conserved residues near the RNA exit channel of the Pol II quaternary structure. This inference led us to hypothesize that the *nripd1-47* and *nripd1-49* missense mutations could disrupt a putatively similar zinc-binding domain in Pol IV, perhaps interfering with Pol IV-RDR2 assembly (Figure 1C).

Pol IV-RDR2 assembly in NRPD1 N-terminus missense mutants

To test whether NRPD1 carrying N-terminal mutations can assemble with other Pol IV subunits and RDR2, we generated plants expressing FLAG-epitope tagged RDR2 that rescues the *rdr2-2* null mutant (Supplementary Figure S2B). Crossing this *RDR2-FLAG* line to *nripd1-47*, *nripd1-49*, *nripd1-50* and *nripd1-51* mutants, respectively, we selected F2 progeny in which each *nripd1* mutation was homozygous and the RDR2-FLAG protein was expressed. Anti-FLAG beads were used to immunopurify RDR2-FLAG from protein extracts obtained from the *RDR2-FLAG* WT *NRPD1* or *RDR2-FLAG nripd1* mutant plants. After SDS-PAGE and western blotting, RDR2-FLAG was detected in the WT *NRPD1* background and in all the *nripd1* mutants (Figure 1D, α -FLAG IP lanes, top panel). An NRPD1-specific antibody detected the largest subunit of Pol IV in all samples except the *nripd1-51* null mutant. Likewise, an NRPD2-specific antibody detected the second largest subunit of Pol IV in all samples except *nripd1-51* (Figure 1D, α -FLAG IP lanes, middle panels). Based on these data, we conclude that the NRPD1 N-terminus residues mutated in *nripd1-47*, *nripd1-49* and *nripd1-50* are not individually essential for Pol IV-RDR2 assembly.

N-terminal *nripd1* missense mutations disrupt *AtSNI* retroelement silencing

To assess how distinct *nripd1* mutations affect Pol IV function in TE silencing, we measured the derepression of a known Pol IV target, *AtSNI*, by qRT-PCR (Figure 1E). In

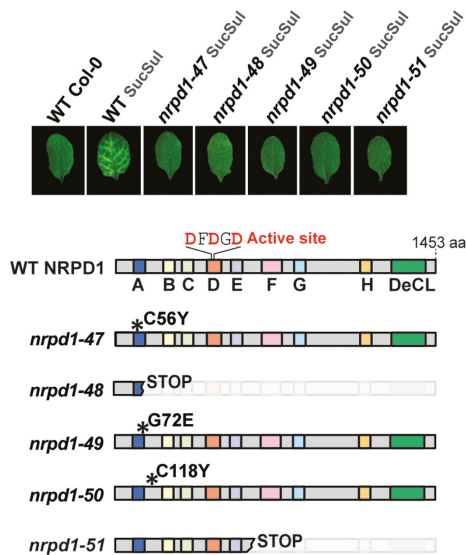
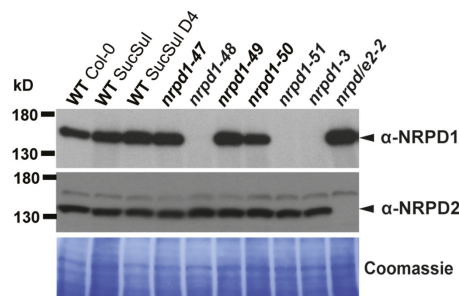
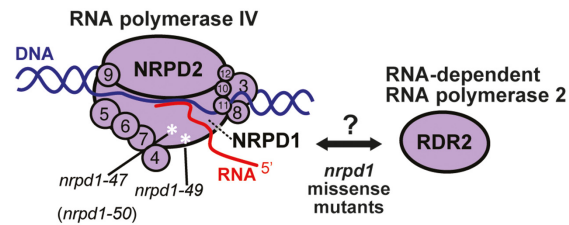
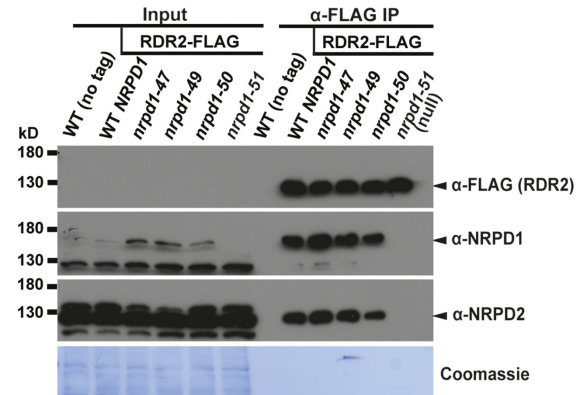
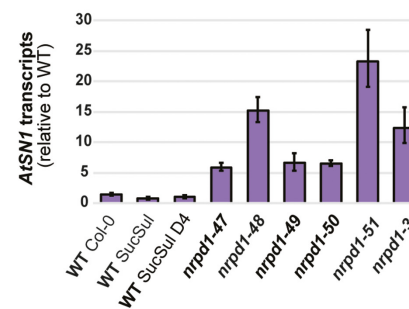
A Mutations in Pol IV's largest subunit**B Western blot****C Speculative view of the Pol IV-RDR2 complex****D Co-purification of the Pol IV complex with RDR2****E qRT-PCR detection of *AtSN1* expression**

Figure 1. N-terminal missense alleles in NRPD1 disrupt the *in vivo* function of Pol IV-RDR2. (A) Above: Rosette leaves are shown from WT Col-0, WT SucSul and from *nrpd1-47*, *nrpd1-48*, *nrpd1-49*, *nrpd1-50* and *nrpd1-51* mutants (all in the SucSul background). Below: Diagram showing the position and predicted effect of the five mutations on NRPD1, Pol IV's largest subunit. Denoted along the length of NRPD1 are the evolutionarily conserved Domains A to H of multisubunit RNA polymerase largest subunits (46,57). The 'Defective in Chloroplasts and Leaves' (DeCL) domain is specific to Pol IV and Pol V and not found outside plants. (B) Western blot detection of Pol IV subunit accumulation in crude protein isolated from inflorescences of WT controls (Col-0, SucSul and SucSul D4), *nrpd1* mutants and the *nrpd1e2* mutant. Antibodies used here were raised against peptides from the NRPD1 C-terminus or the NRPD2 N-terminus, respectively. (C) Speculative view of the Pol IV complex (modified from (70)) showing estimated positions (*) of the amino acid substitutions *nrpd1-47* (C56Y) and *nrpd1-49* (G72E) inferred from alignment of Arabidopsis NRPD1 to yeast RPB1 and the related Pol II structure (57). (D) WT and mutant forms of Pol IV co-purified under native conditions via epitope-tagged RDR2 (RDR2-FLAG). Total protein input and Anti-FLAG immunopurified protein were separated on an SDS 6% polyacrylamide gel and subjected to western blotting. NRPD2 and NRPD1-specific antibodies (validated in panel B), were used to successively detect corresponding Pol IV subunits in protein fractions co-purified with RDR2-FLAG (middle panels, right-hand lanes), then an anti-FLAG antibody was used to detect RDR2-FLAG (top panel, right-hand lanes). Staining with Coomassie solution was used to evaluate protein loading in the input fractions. (E) qRT-PCR performed on total RNA from inflorescences using random-primed cDNA synthesis and qPCR primers specific for the *AtSN1* retroelement. Following *ACT2* normalization, *AtSN1* transcript expression is plotted as the ratio of WT SucSul D4/sample. Error bars indicate the standard error of the mean for three technical replicates.

WT plants *AtSN1* is silenced and transcripts do not accumulate, but transcripts are detected in *nrpd1* null mutants (44). The NRPD1 N-terminus missense mutants (*nrpd1-47*, *nrpd1-49*, *nrpd1-50*) showed less *AtSN1* transcript accumulation than null mutants (*nrpd1-48*, *nrpd1-51*, *nrpd1-3*) (Figure 1E). This suggests that *AtSN1* is still partially silenced by Pol IV in plants expressing N-terminally-mutated

NRPD1, in contrast to full *AtSN1* derepression in null mutants (Figure 1E, Supplementary Figure S2C). Accordingly, cytosine methylation assayed at *Hae* III sites in *AtSN1* was only partially lost in *nrpd1-47*, *nrpd1-49* and *nrpd1-50*, but more severely reduced in *nrpd1-48*, *nrpd1-51* and *nrpd1-3* (Supplementary Figure S2D).

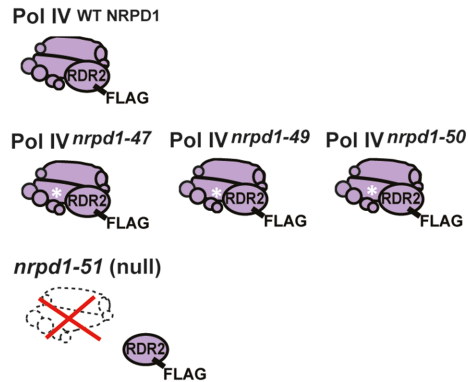
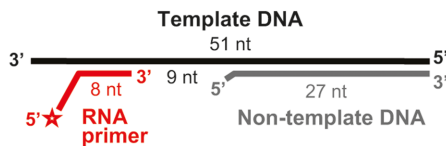
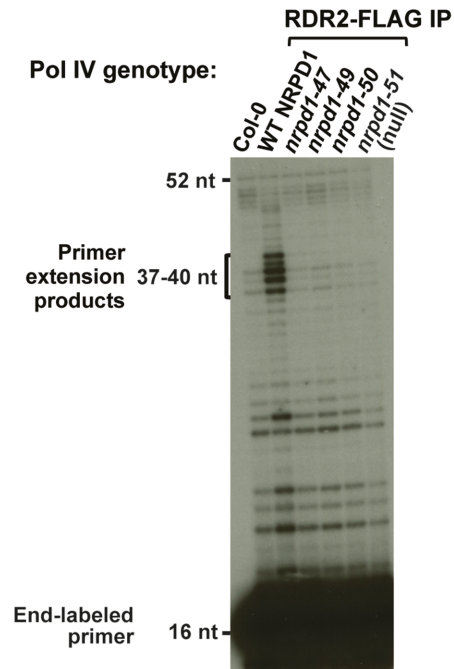
A Pol IV variants used in assays**B Tripartite substrate****C Detection of RNA extension**

Figure 2. Pol IV *in vitro* transcription assays. (A) The Pol IV complexes assembled with WT NRPD1 or with *nrpd1-47*, *nrpd1-49* or *nrpd1-50* mutant variants (Pol IV^{WT NRPD1}, Pol IV^{*nrpd1-47*}, Pol IV^{*nrpd1-49*} or Pol IV^{*nrpd1-50*}) were co-purified with RDR2-FLAG and then used for *in vitro* transcription assays. *nrpd1-51* is a null mutation in Pol IV, abolishing its co-purification with RDR2-FLAG and thus serves as a negative control for Pol IV activity. (B) The tripartite substrate used for the assays is composed of a template DNA, a non-template DNA and a radiolabeled RNA primer. (C) Transcription assays for each Pol IV variant (Pol IV^{WT NRPD1}, Pol IV^{*nrpd1-47*}, Pol IV^{*nrpd1-49*}, Pol IV^{*nrpd1-50*} or the null mutant *nrpd1-51*). An assay using protein from non-transgenic plants (Col-0) provides an additional negative control. The 5'-end radiolabeled, 16 nt RNA primer is seen at the gel bottom and *in vitro* RNA extension products are visible in the 37–40 nt range near the top third of the gel.

Pol IV *in vitro* transcription assays

A previously reported C67S mutation in yeast RPB1 (*rpo21-27*) causes growth defects (56) and corresponds to the NRPD1 position mutated in the Arabidopsis *nrpd1-47* (C56Y) mutant. Another such yeast mutation, *rpo21-30* (H80Y), affects a position adjacent to the NRPD1 amino acid mutated in *nrpd1-49* (Supplementary Figure S2A). These *rpo21-27* and *rpo21-30* mutations both reduce the transcriptional activity of Pol II in yeast (56). Therefore, an alternative hypothesis to Pol IV-RDR2 assembly being disrupted in *nrpd1-47* or *nrpd1-49* would be that Pol IV transcriptional activity is reduced by these genetic lesions.

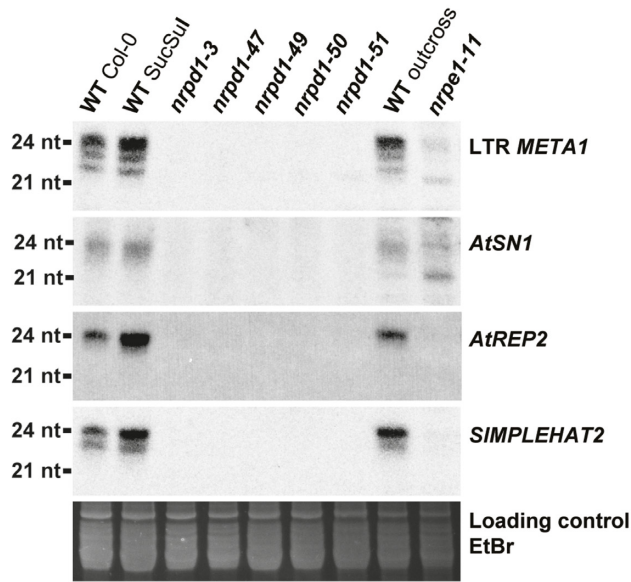
To test whether Pol IV enzymatic activity is affected by NRPD1 N-terminus mutations, we performed *in vitro* transcription assays following an established protocol (6,9). Precipitation of RDR2-FLAG using Anti-FLAG beads allowed co-purification of Pol IV^{WT NRPD1}, Pol IV^{*nrpd1-47*}, Pol IV^{*nrpd1-49*} or Pol IV^{*nrpd1-50*} using the respective WT or *nrpd1* mutant backgrounds (Figure 2A). *nrpd1-51* is the *pol IV* null mutant in which Pol IV-RDR2 assembly is not possible (Figure 1D). The assays utilize a 51 nt DNA template oligonucleotide annealed for 27 bp with a non-template DNA strand and hybridized for 8 bp with an end-labeled RNA oligonucleotide primer (Figure 2B), thus mimicking a transcription elongation complex. Pol IV can elongate the RNA primer in a DNA templated fashion, but

terminates 12–16 nt after encountering the base-paired non-template DNA strand (6), generating 37–40 nt transcripts, as observed for Pol IV assembled using WT NRPD1 (Figure 2C). By contrast, Pol IV co-purified with RDR2-FLAG in the *nrpd1-47*, *nrpd1-49* or *nrpd1-50* backgrounds, respectively, showed little activity, resembling the *nrpd1-51* null mutant or Col-0 negative control (Figure 2C). We conclude that the Pol IV *in vitro* activity is crippled or abolished by the mutations in the NRPD1 N-terminus.

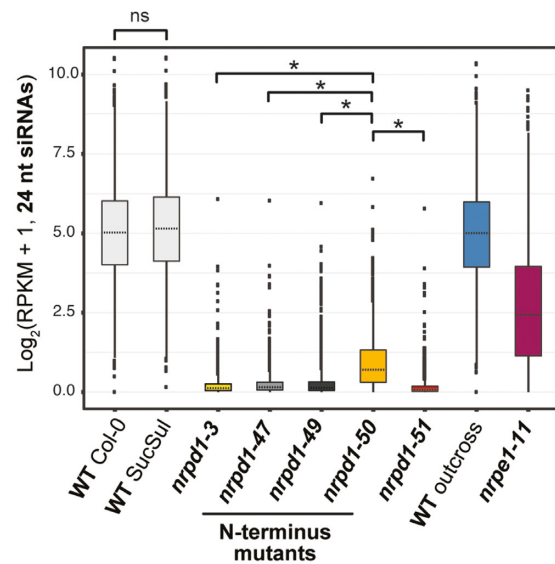
siRNA biogenesis and DNA methylation in NRPD1 N-terminus mutants

Defects in Pol IV transcription would limit production of 24 nt siRNAs *in vivo* (8,46,48), so we used RNA blot hybridization to test whether siRNA levels changed in the *nrpd1* N-terminus mutants. A probe for the LTRs of *MET1 Copia* retrotransposons detected 24 nt siRNAs in WT Col-0 and WT SucSul (Figure 3A). These siRNAs were not detected in *nrpd1-3* or *nrpd1-51* null mutants, but they were also not observed in any N-terminus missense line (Figure 3A). Similar results were obtained for siRNAs from the *AtSN1* retroelement, as well as from the *AtREP2* and *SIMPLEHAT2* DNA transposons. Outcrossing *nrpd1-51* to WT Col-0 followed by selection of homozygous WT NRPD1 in the F2 generation restored siRNA levels (Figure 3A).

A Small RNA northern blot



C Small RNA-seq analysis of *pol IV* DMRs



B Overlap of Differentially-Methylated Regions (DMRs) with genomic annotations

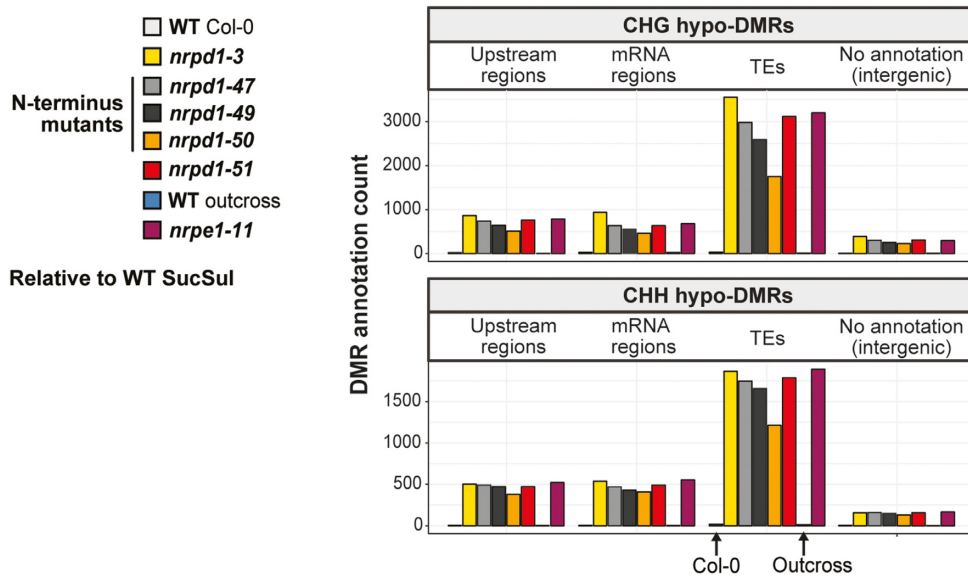


Figure 3. N-terminal mutations in NRPD1 nearly abolish 24 nt siRNA accumulation and RdDM. (A) Small RNA northern blot using several probes at Pol IV-RdDM targets: LTR *META1*, *AtSN1*, *AtREP2*, *SIMPLEHAT2*. Equal loading was tested by imaging the gel stained with ethidium bromide (EtBr) prior to blotting. (B) Overlap of Differentially-Methylated Regions (DMRs) with genomic annotations. DMRs were identified in each sample relative to WT SucSul using whole-genome bisulfite sequencing (2× replicates per sample genotype). Regions with less methylation than WT SucSul (hypo-DMRs) were counted upstream of genes, in mRNA regions, in TEs and in regions lacking annotation (intergenic regions). CHG sequence and CHH sequence context DMRs are tabulated separately (above and below, respectively). (C) Small RNA-seq analysis of *pol IV* DMRs. Small RNA-seq was performed on the same genotypes as analyzed in panel B. After mapping, 24 nt siRNAs were counted in putative Pol IV-RdDM regions (i.e. *nrpd1-51* hypo-DMRs; see Supplementary Tables S5 and S9). Pairwise statistical comparisons were performed using the Wilcoxon rank sum test: *P*-values < 0.01 were treated as not significant (ns), whereas *P*-values ≥ 0.01 were treated as significant (*). Only comparisons between WT Col-0 and WT SucSul controls, as well as comparisons between *nrpd1-50* and the other *nrpd1* mutant alleles are shown explicitly with brackets in Figure 3.

The apparent loss of siRNAs (and undetectable Pol IV activity *in vitro*) in *nRPD1-47*, *nRPD1-49* and *nRPD1-50* mis-sense lines was surprising because these mutants only showed a partial loss of *AtSN1* silencing *in vivo*. We thus pursued genome-scale analyses to chart the global impact of NRPD1 N-terminus mutations on RdDM. Whole-genome bisulfite sequencing (WGBS) was performed on DNA from WT Col-0 and SucSul controls, *nRPD1* mutants, *nRPD1-51* outcrossed to Col-0 and the *nrpe1-11* (*pol V*) null mutant. WGBS reads were mapped to the Arabidopsis genome and Differentially-Methylated Regions (DMRs) were called relative to WT SucSul (2x replicates per sample, Supplementary Table S5). Comparison of WT Col-0 to WT SucSul revealed only 100 regions with reduced methylation (hypo-DMRs), indicating that both controls display similar patterns of DNA methylation. In the *nRPD1-3* null mutant, 3553 TE regions, 863 regions upstream of genes and 935 mRNA regions were detected overlapping CHG hypo-DMRs (relative to WT SucSul); in addition, 1866 TEs, 502 upstream regions and 538 mRNA regions were found overlapping CHH hypo-DMRs in the *nRPD1-3* mutant (Figure 3B, yellow bars). *nRPD1-51* showed frequencies of hypo-DMRs comparable to *nRPD1-3* (Figure 3B, red bars). Slightly fewer hypo-DMRs were detected in *nRPD1-47* and *nRPD1-49* N-terminus mutants (Figure 3B, dark grey bars), although these mutations nearly phenocopied *pol IV* null alleles in the CHH methylation context. By contrast, the *nRPD1-50* N-terminus mutant displayed far fewer hypo-DMRs than either null allele, an effect most apparent in TEs (Figure 3B, orange bars). Pol IV-dependent DNA methylation was globally resettable, with only 57 total hypo-DMRs recovered after *nRPD1-51* outcross (Figure 3B, blue bars). Finally, the number of hypo-DMRs in the *pol V* null mutant (Figure 3B, purple bars) was similar to *pol IV* null alleles (*nRPD1-3* and *nRPD1-51*) and larger than any of the *nRPD1* N-terminus mis-sense alleles.

Although 24 nt siRNAs could not be detected by northern blot in *nRPD1-50* plants (Figure 3A), the fewer DMRs in *nRPD1-50* compared to *pol IV* null plants suggested that Pol IV-dependent DNA methylation continued at hundreds of chromosomal targets in *nRPD1-50*. To better understand the role of the NRPD1 N-terminus in Pol IV function, we performed small RNA-seq on the same samples as were analyzed by WGBS. The abundance of 24 nt siRNAs was quantified at all regions of Pol IV-dependent DNA methylation (Figure 3C) (2x replicates per sample, Supplementary Table S9). Boxplots for WT Col-0 and SucSul controls were indistinguishable, with median values near 35 reads per kilobase per million reads mapped (RPKM), whereas medians for *nRPD1-3* and *nRPD1-51* null mutants were drastically reduced (Figure 3C, note log₂ scale). Similar to the DMR results above, Pol IV-dependent siRNA production was restored after the *nRPD1-51* null mutation was outcrossed (Figure 3C, blue boxplot). However, 24 nt siRNA levels in the *nRPD1-50* N-terminus mutant (Figure 3C, Supplementary Figure S3) were significantly higher than those of *pol IV* null mutants. This *nRPD1-50* median represented trace amounts of 24 nt siRNAs, ~55 times less than WT SucSul, explaining why the less sensitive RNA blot technique detected no 24 nt siRNAs (Figure 3A).

In summary, NRPD1 N-terminus mutations impair 24 nt siRNA accumulation and DNA methylation with varying degrees of potency. The strongest alleles, *nRPD1-47* and *nRPD1-49*, disrupted a putative zinc-binding domain in NRPD1 'Domain A' that is conserved in the largest subunits of all multisubunit RNA polymerases (46,56). Thus, both *nRPD1-47* and *nRPD1-49* nearly phenocopied the siRNA and DNA methylation defects of *pol IV* null mutants. By contrast, the *nRPD1-50* mutation adjacent to 'Domain A' (Figure 1A, Supplementary Figure S4) caused less severe deficiencies, preserving trace 24 nt siRNA levels and residual DNA methylation at certain Pol IV-RdDM targets.

A signature Pol IV motif uniquely conserved in NRPD1

Inspecting the *nRPD1-50* mutation (C118Y) in our NRPD1 alignment, we discovered an evolutionarily conserved protein motif composed of a C[KR]YC box followed by a 5–10 amino acid (aa) spacer and then by a YPx[MV][KR]F[KR] box (Figure 4). We initially found this motif in NRPD1 of 17 species ranging from the basal angiosperm *Amborella trichopoda* to the conifer *Pinus canariensis* (Figure 4, Supplementary Figure S4; Supplementary Table S10). However, it was absent in the corresponding regions of Arabidopsis NRPB1 (Pol II), Arabidopsis NRPE1 (Pol V) and yeast RPB1 (*S. cerevisiae*, Pol II). To identify more proteins with this motif, we generated a hidden Markov model based only on the motif region in our NRPD1 alignment and queried UniProt Reference Proteomes using hmmsearch (<https://www.ebi.ac.uk/Tools/hmmer/>). High-confidence matches to RNA polymerase subunits (*E*-value < 0.01, length > 1300 aa) included proteins from 46 plant species. These proteins contained close matches to Domains 'A through H' and the 'Defective in Chloroplasts and Leaves' domain (DeCL), as expected for NRPD1 or NRPE1 but not for NRPB1. Moreover, all identified subunits lacked the C-terminal extension containing WG/GW motifs, which is important for NRPE1 function in Pol V (54,58,60) (Supplementary Figure S4 and Table S11). We conclude that the amino acid position mutated in *nRPD1-50* is part of a signature Pol IV motif not found in Pol II, Pol V or other eukaryotic RNA polymerases.

The Pol IV-specific motif safeguards robust TE methylation patterning

Comparison of the hypo-DMRs common to *nRPD1-50* mis-sense and *nRPD1-51* null mutants showed that the Pol IV-specific motif was critical for CHG methylation at 1942 loci and for CHH methylation at 1286 loci. An additional 1408 CHG and 620 CHH hypo-DMRs were detected only in the *nRPD1-51* null plants, in which Pol IV does not assemble (Figure 5A, pie charts). We used amplicon-based bisulfite sequencing to precisely quantify DNA methylation changes across the *AtSN1* retrotransposon locus. In the WT control 86% of CG sites, 73% of CHG sites and 24% of CHH sites were methylated (Figure 5A, bar chart). All three cytosine contexts in *AtSN1* showed less methylation in *nRPD1-50*, but these levels were reduced much further in

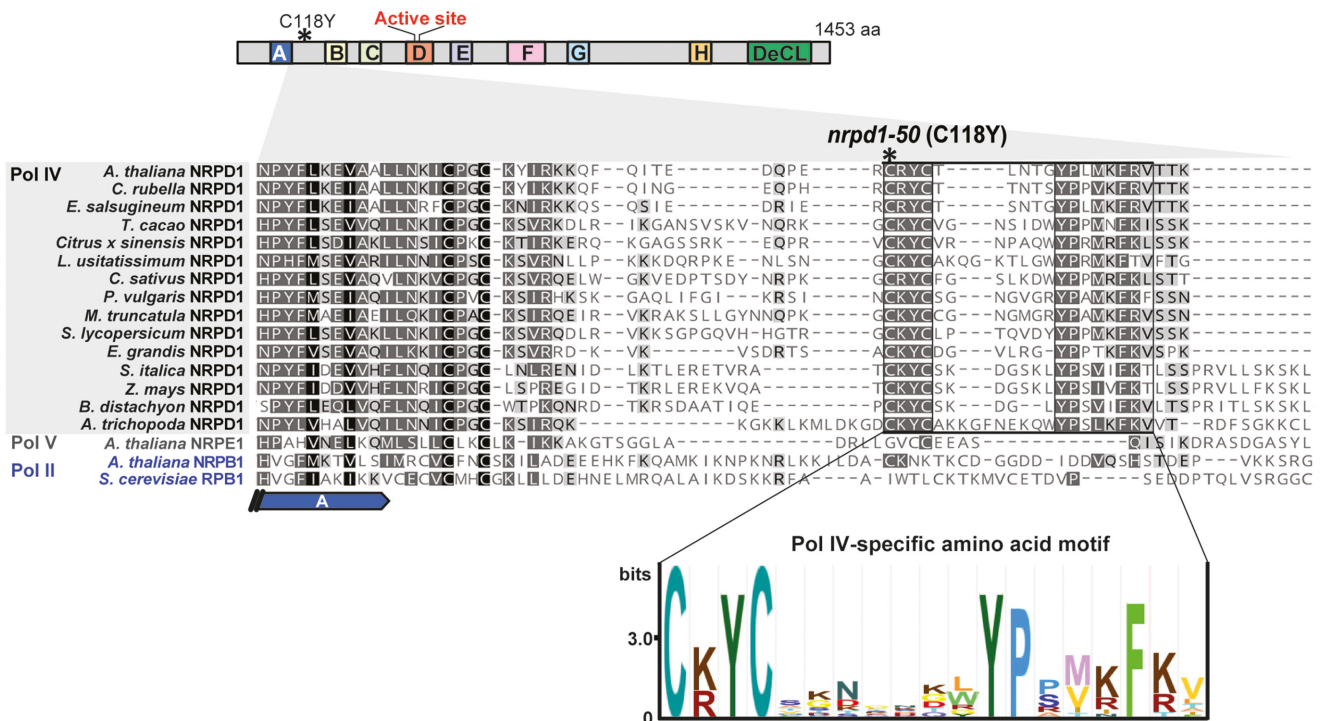
NRPD1 from *Arabidopsis thaliana*

Figure 4. Mutation in NRPD1 adjacent to conserved 'Domain A' reveals a Pol IV-specific motif. Above: Diagram showing the position of the *nrpd1-50* mutation (C118Y) in NRPD1, the largest subunit of Pol IV, relative to this subunit's Domains A to H (conserved in the largest subunits of all nuclear RNA polymerases), and its C-terminal 'Defective in Chloroplasts and Leaves' (DeCL) domain. Middle: Alignment of amino acids adjacent to 'Domain A' in NRPD1 from the species *Arabidopsis thaliana*, *Capsella rubella*, *Eutrema salsugineum*, *Theobroma cacao*, *Citrus x sinensis*, *Linum usitatissimum*, *Crocus sativus*, *Phaseolus vulgaris*, *Medicago truncatula*, *Eucalyptus grandis*, *Solanum lycopersicum*, *Setaria italica*, *Zea mays*, *Brachypodium distachyon* and *Amborella trichopoda*. Included for comparison are NRPE1, the largest subunit of Pol V from *Arabidopsis*; NRPB1, the largest subunit of Pol II from *Arabidopsis*; and RPB1, the largest subunit of Pol II from *Saccharomyces cerevisiae*. Below: The alignment reveals a Pol IV-specific protein motif starting at *Arabidopsis* NRPD1 amino acid 118, which is composed of a C[KR]YC box followed by a 5–10 amino acid (aa) spacer and then by a YP[MV][KR]F[KR] box.

nrpd1-51 null plants. Pol IV-RdDM, scored as CHH methylation, was distributed evenly across *AtSN1* in WT SucSul and was uniformly lost in *nrpd1-51* (Figure 5B). By contrast, the *nrpd1-50* methylation pattern displayed a striking discontinuity: CHH methylation was erased at the *AtSN1* 5'-end, over A and B-box promoter elements, but it remained intact near the 3' polyA tract. Wilcoxon rank sum analysis of WT versus *nrpd1-50* amplicons supports the assessment that CHH methylation remained within the 3' *AtSN1* interval (i1) in the mutant, unlike at upstream *AtSN1* subfeatures (i2, i3; Figure 5B and Supplementary Figure S5A). This trend was also noted in our WGBS data (Supplementary Figure S5B). Thus, residual *AtSN1* methylation and partial *AtSN1* silencing (see Figure 1E) support the hypothesis that the *nrpd1-50* mutant expresses a partially functional Pol IV^{*nrpd1-50*} enzyme.

To further explore patterns of residual methylation in *nrpd1-50*, we surveyed other TE annotations in our WGBS data. Metaplots of CHH methylation across LTR/*Copia*, *MuDR* and *Helitron* elements revealed symmetrical peaks at the TE extremities in WT plants (Figure 5C and Supplementary Figure S5). These peaks remained sharp in *nrpd1-50*, whereas the overall methylation profile flattened in *nrpd1-51* null plants. Notably, CHH methylation at *Copia*

long-terminal repeats (LTRs) was less sensitive to *nrpd1-50* than to *nrpd1-51*, whereas TE body methylation was reduced in *nrpd1-50* and *nrpd1-51* (Figure 5C and Supplementary Figure S5). Similar effects were evident at individual *Copia* elements: AT1TE14315 body methylation was reduced in both *nrpd1-50* and *nrpd1-51* mutants, whereas LTR methylation peaks remained intact in *nrpd1-50* (Figure 5D, green tracks). This residual LTR methylation correlated with trace 24 nt siRNA accumulation in *nrpd1-50*; by contrast, the *pol V* mutant (*nrpe1-11*) erased CHH methylation with only a slight loss in siRNAs (Figure 5D, purple tracks). Numerous TEs had profiles similar to Figure 5D, with swaths of DNA methylation lost in *nrpd1-51* but remaining partly intact in the *nrpd1-50* N-terminus mutant (Supplementary Figure S6).

Sites of residual methylation included *AtSN1*, *Copia*, *MuDR* and *Helitron* elements at loci across all five *Arabidopsis* chromosomes. Nine such hypo-DMRs were validated by Chop-PCR (Figure 5E). For these assays, genomic DNA from WT or mutant plants was digested with a methylation-sensitive restriction enzyme, then PCR was performed spanning the enzyme's recognition sites. Successful PCR indicated that the template DNA was methylated

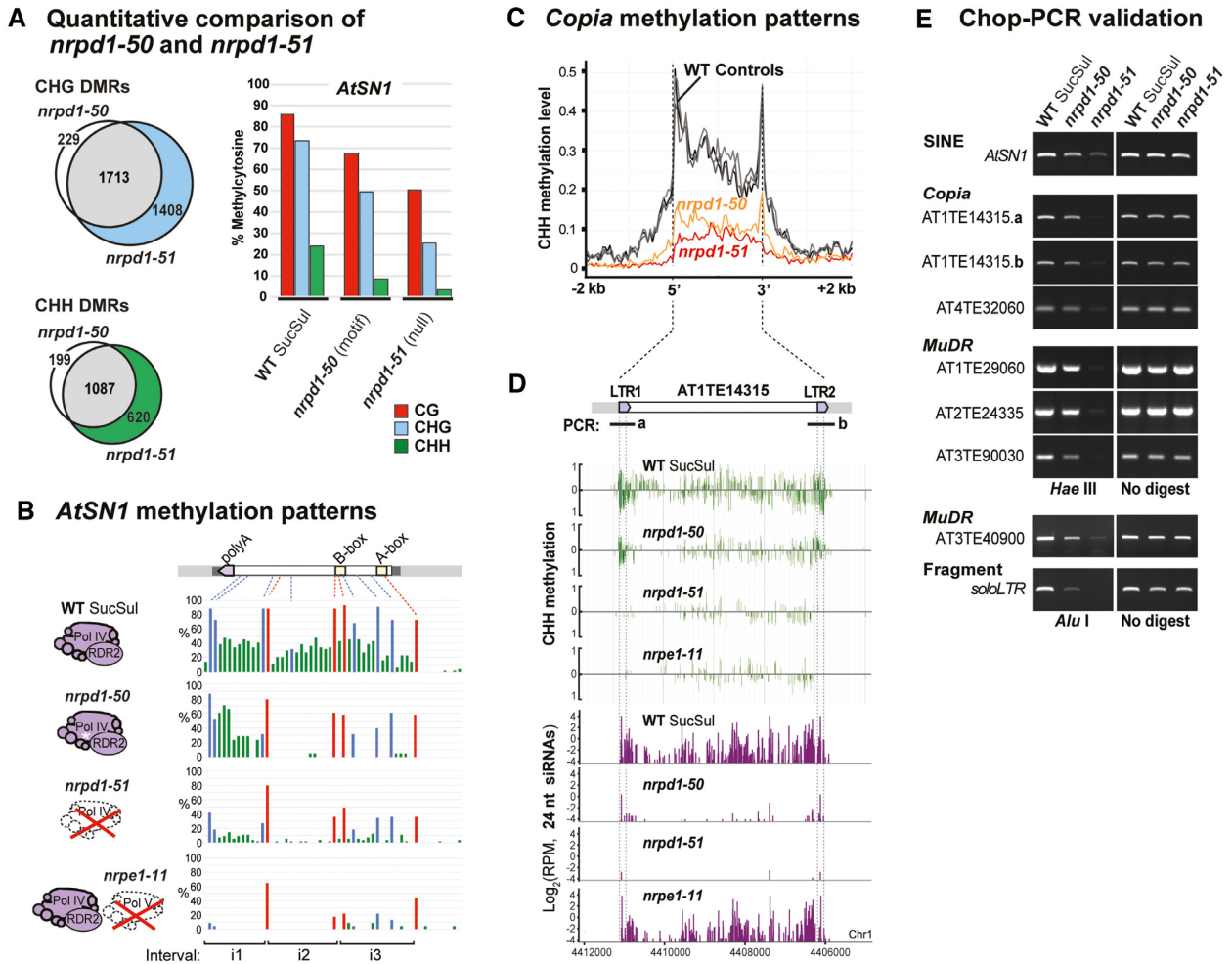


Figure 5. The Pol IV-specific motif in NRPD1 is critical for robust TE methylation patterning. (A) Left: Venn diagram of DMRs for CHG and CHH methylation contexts in *nrpd1-50* and *nrpd1-51* mutants. Right: total percentage of methylcytosine at the *AtSN1* locus in the three cytosine sequence contexts (CG, CHG and CHH) for WT SucSul, *nrpd1-50* and *nrpd1-51* mutants ($n \geq 38$ clones/sample). (B) *AtSN1* methylation patterns: per site CG, CHG and CHH methylation, respectively, are shown as red, blue and green vertical lines. Wilcoxon rank sum tests were performed to compare WT SucSul, *nrpd1-50* and *nrpd1-51* methylation within three different intervals of the *AtSN1* region (i1, i2 and i3; see Supplementary Figure S5A). (C) *Copia* methylation patterns: metaplot of CHH methylation levels on *Copia* elements between the annotated 5' and 3'-ends, with 2 kb of genomic context included upstream and downstream. (D) Detail of CHH methylation (green vertical lines) and 24 nt siRNAs (purple vertical lines) for the *Copia* retrotransposon AT1TE14315 that differ at five genomic positions and each contain one *Hae* III site. (E) Chop-PCR: genomic DNA from WT SucSul, *nrpd1-50* and *nrpd1-51* was digested with methylation-sensitive restriction enzymes (rows 1–7 with *Hae* III; rows 8 and 9 with *Alu* I) followed by PCR. Successful PCR indicated that the template DNA was methylated (protected), whereas less amplification indicated less DNA methylation. Loci tested were: the SINE element *AtSN1*; the *Copia* elements AT1TE14315 (a and b correspond to LTR1 and LTR2, panel D) and AT4TE32060; the *MuDR* elements AT1TE29060, AT2TE24335, AT3TE90030 and AT3TE40900; and the *soloLTR* fragment. ‘No digest’ controls are PCR products from assays omitting the restriction enzyme.

(protected), whereas weaker amplification indicated that little or no DNA methylation was present. The *nrpd1-50* mutant displayed somewhat less CHH methylation than WT SucSul, but methylation was nearly undetectable at these sites in the *nrpd1-51* null mutant (Figure 5E). We hypothesize that the Pol IV^{*nrpd1-50*} enzyme continues to target many TEs in the Arabidopsis genome. However, because siRNA biogenesis is impaired, Pol IV^{*nrpd1-50*}-RdDM only partially methylates these loci.

Residual RdDM and the loss of genome surveillance in NRPD1 N-terminus mutants

We considered three alternatives to the working hypothe-

sis that residual siRNAs and methylation in *nrpd1-50* stem from partially operative Pol IV-RdDM. First, these differentially methylated loci could be epigenetic variants segregating independently of Pol IV function. This alternative predicts that outcrossing a *pol IV* mutation to Col-0 would fail to restore siRNA production and methylation. To test this prediction, we plotted 24 nt siRNA abundance (x-axis) versus CHH methylation (y-axis) at all putative Pol IV targets and compared the WT Col-0, SucSul and outcross controls to *nrpd1* mutants (Figure 6A). All the WT samples showed Pol IV targets ranging from high siRNA and methylation levels (Figure 6A, top right-hand) to moderate siRNA and methylation levels (Figure

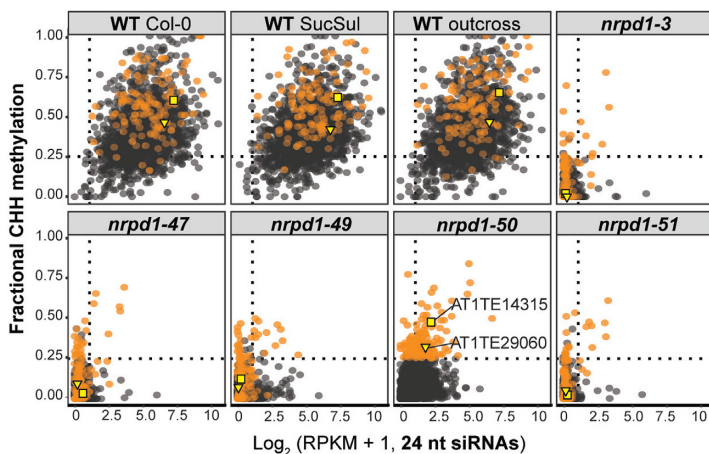
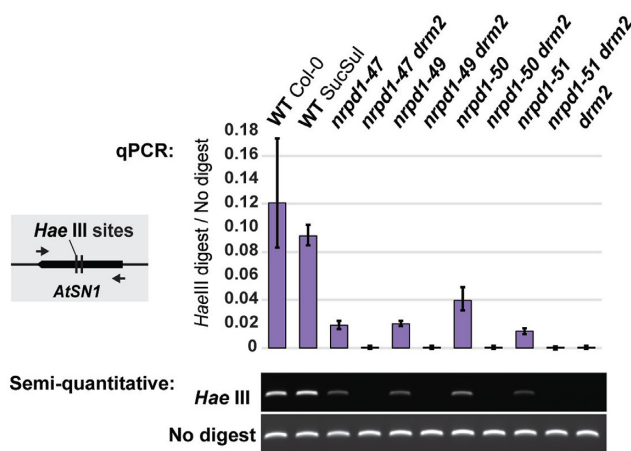
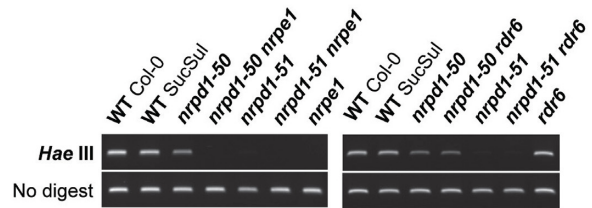
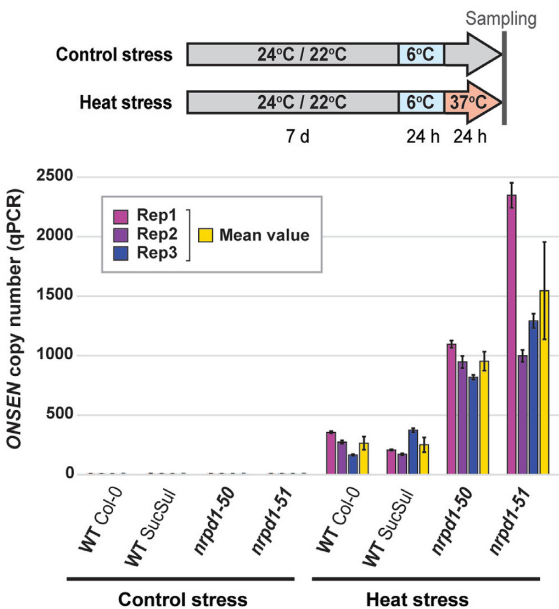
A Methylated loci versus siRNA abundance**B DRM2 mediates TE methylation in *nrpd1* hypomorphs****C Pol V, but not RDR6, mediates residual TE methylation detected in *nrpd1-50*****D *ONSEN* retrotransposon accumulation**

Figure 6. Residual Pol IV, DRM2 and Pol V-dependent genome surveillance in *nrpd1* hypomorphs. (A) Methylated loci versus siRNA abundance. Fractional CHH methylation (y -axis) is shown plotted versus 24 nt siRNA abundance (x -axis) at each Pol IV-RdDM locus. The dotted lines enclose 98% of all loci in the *nrpd1-51* subplot, demarcating a complete loss of Pol IV-RdDM. The orange dots represent loci that display residual methylation in the *nrpd1-50* mutant. The yellow square and triangle represent AT1TE14315 and AT1TE29060, respectively, two TEs showing residual methylation validated by Chop-PCR (Figure 5E). (B) DRM2 mediates TE methylation in *nrpd1* hypomorphs. Chop-PCR assays at the *AtSN1* locus using the *Hae* III methylation-sensitive restriction enzyme, followed by either qPCR (above) or semi-quantitative PCR (below). Mutant alleles of *nrpd1* were tested alongside double mutant combinations with the *drm2* null mutant. ‘No digest’ controls are PCR products from assays omitting the restriction enzyme. (C) Pol V, but not RDR6, mediates residual TE methylation detected in *nrpd1-50*. Chop-PCR assays at the *AtSN1* locus, like in panel B, except using *nrpd1* double mutant combinations with *nrpe1* (*pol V*) or *nrpd1* double mutant combinations with *rdr6*. (D) *ONSEN* retrotransposon accumulation. Above: Design of the assay: ‘control stress’ plants were grown at 24°C/22°C (day/night) with a single 24 h exposure to 6°C, whereas ‘heat stress’ plants were grown at 24°C/22°C, exposed for 24 h to 6°C, then heat-stressed for 24 h at 37°C. Below: Extrachromosomal *ONSEN* accumulation measured by qPCR for WT (Col-0 and SucSul), *nrpd1-50* and *nrpd1-51* plants grown under control and heat stress conditions. Error bars indicate the standard error of the mean for the three qPCR technical replicates used to analyze each biological replicate.

6A, dotted lines). Over 135 targets showed high (>25%) fractional CHH methylation in *nrpd1-50* plants. These loci (Figure 6A, orange points) were consistently methylated in WT Col-0, SucSul and outcross samples, and the vast majority dropped below 25% fractional methylation in *nrpd1-3* and *nrpd1-51* null mutants, indicating that prominent loci showing residual methylation in *nrpd1-50* are *bone fide* Pol IV targets (Figures 6A). For example, *META1* retrotransposons (e.g. AT1TE14315) occupy coordinates at the top right-hand of the WT plots, signifying high siRNA and CHH methylation levels. These TEs are depleted of

siRNAs/CHH methylation in *nrpd1-3* and *nrpd1-51* null mutants but display residual siRNAs/CHH methylation in *nrpd1-50* (Figure 6A, yellow markers). These data are all consistent with Pol IV^{*nrpd1-50*} mediating limited RdDM at certain TEs.

A second alternative to our working hypothesis is that Pol IV deficiency in *nrpd1* N-terminus mutants is compensated by another DNA methyltransferase: i.e. instead of the Pol IV^{*nrpd1-50*}-Pol V-DRM2 machinery, we might be detecting an ectopic, *de novo* DNA methylation activity of CMT3 (61). To test which methyltransferase is re-

quired for CHH methylation in *nRPD1* N-terminus mutants, we crossed *nRPD1-47*, *nRPD1-49*, *nRPD1-50* and *nRPD1-51*, respectively, to *drm2* and *cmt3* null alleles. The CHH methylation detected in *nRPD1* N-terminus hypomorphs was lost in *nRPD1 drm2* double mutants (Figure 6B and Supplementary Figure S7A) but persisted in *nRPD1 cmt3* double mutants (Supplementary Figure S7B), showing that this residual methylation is primarily DRM2-dependent. Analysis of *nRPD1-50 nrpe1* and *nRPD1-51 nrpe1* double mutants confirmed that Pol V is also required for the residual methylation in *nRPD1-50* (Figure 6C and Supplementary Figure S7C; left-hand panels).

Having established that DRM2 and Pol V are required for residual DNA methylation in *nRPD1-50*, there remained a third alternative. Non-canonical RdDM can occur when Pol II transcription of an active TE generates substrates for the enzyme RDR6. dsRNA precursors produced in this alternative pathway are diced into 21 nt siRNAs, rather than 24 nt siRNAs, which may guide TE transcript cleavage and/or DNA methylation (37,39). Arguing against this alternative here, siRNAs in the *nRPD1-50* mutant were not predominantly 21 nt in length (Supplementary Figure S3) and the methylation persisted in *nRPD1-50 rdr6* double mutants (Figure 6C and Supplementary Figure S7C; right-hand panels). Instead, we propose that Pol IV^{*nRPD1-50*} mediates canonical RdDM restricted to TE subfeatures, either because the Pol IV motif disrupted in *nRPD1-50* is required for methylation to spread across TEs, or because the threshold quantity of siRNAs needed for localized RdDM can still be produced at certain subfeatures despite the *nRPD1-50* mutation (see Figure 7, models).

DNA methylation facilitated by the NRPD1 N-terminus could be key to preventing TE proliferation. To test whether the Pol IV-specific motif is required to prevent retrotransposon activity, we measured the accumulation of extrachromosomal *ONSEN* DNA in plants exposed to heat stress (24 h at 37°C) and compared these levels to plants treated with control stress. Under control conditions WT and *nRPD1* plants showed the same low *ONSEN* copy number (Figure 6D, left-hand panel). By contrast, the *nRPD1-50* N-terminus mutant displayed a 4-fold higher *ONSEN* copy number than WT Col-0 or SucSul plants (Figure 6D, right-hand panel). *ONSEN* accumulation in *nRPD1-51* null plants was even higher than *nRPD1-50* when averaged over three biological replicates (Figure 6D, yellow bars). In conclusion, despite residual traces of Pol IV-RdDM in *nRPD1-50* plants, the Pol IV signature motif in the NRPD1 N-terminus is absolutely critical for genome surveillance.

DISCUSSION

Pol IV transcribes chromosomal DNA into primary precursors for siRNAs that guide TE methylation in plants (6,8,9). Until now, little was known about novel domains in the Pol IV core that could govern this specialized function. The Pol IV active center undoubtedly includes NRPD1 'Domain D' with its aspartate triad orthologous to the Pol II residues that coordinate Mg²⁺ for catalysis of phosphodiester bonds in RNA (48,49,57); however, like Pol V, Pol IV has deletions impacting the 'trigger loop' and 'bridge-helix' subdomains that are found in Pol II and most other multi-

subunit RNA polymerases (9,48) (see Supplementary Figure S4). The latter NRPD1 sequence polymorphisms explain why Pol IV is α -amanitin insensitive and likely contribute to its high error rate (9,36,62), but they are not alterations exclusive to the Pol IV enzyme. Our present study reveals that the NRPD1 N-terminus harbors a motif that is uniquely conserved in Pol IV (i.e. absent in Pools I/II/III/V) and required for robust 24 nt siRNA biogenesis, RdDM and genome surveillance.

Remarkably, plants expressing the Pol IV^{*nRPD1-50*}-RDR2 complex produce 55 times fewer 24 nt siRNAs at Pol IV-RdDM targets than observed in WT plants (see Figure 3C). Despite being undetectable by northern blot, these trace siRNAs appear to be sufficient to direct CHH methylation to sites throughout the Arabidopsis genome. The presence of 24 nt siRNA peaks at TE extremities and a limited number of other hotspots in *nRPD1-50* (see Figure 5D; Supplementary Figures S5 and S6) suggests that these sites are regions of RdDM initiation from which an amplified genome surveillance response could expand. One possible model is that the Pol IV-specific motif that we discovered in the NRPD1 N-terminus governs such a process in WT plants. RNA-induced transcriptional silencing (RITS) in *Schizosaccharomyces pombe* resembles plant RdDM in many respects. During RITS, positive feedback couples siRNA biogenesis to H3K9 methylation and drives cis-spreading of heterochromatin across repeats (63). An analogous mechanism could potentially facilitate Pol IV function in plant genome surveillance (Figure 7A).

In WT Arabidopsis, one of several RNA-triggered initiation mechanisms (37,38,64,65) could seed H3K9me2 for SHH1-based Pol IV recruitment. The coupled activities of Pol IV and RDR2 are known to generate ~30 bp dsRNAs that are diced into 24 nt siRNAs corresponding to the DNA template (6,8–10). After siRNAs guide AGO4 to Pol V scaffold transcripts (14), positive feedback in the SHH1-Pol IV-Pol V-DRM2 system could promote RdDM spreading across TEs (Figure 7A). Supporting this 'spreading model', we found that the *nRPD1-50* mutation attenuates RdDM without disrupting Pol IV-RDR2 assembly, leaving TE subfeatures still targeted by the Pol IV^{*nRPD1-50*}-Pol V-DRM2 machinery *in vivo*. An alternative 'threshold model', however, could explain sites of methylation loss in *nRPD1-50* as sequences that are less subject to RdDM in WT plants. Loss of RdDM would occur where siRNA steady-state levels drop below a putative threshold needed to target AGO4 to Pol V transcripts for DRM2 recruitment (Figure 7B). The lower WT level of 24 nt siRNAs arising from long TE bodies compared to TE edges (26,66) could make RdDM targeting these TE bodies more sensitive to the *nRPD1-50* allele (e.g., *Copia* body versus LTRs, see Figure 5).

To balance genome surveillance with growth and development, DNA methylation must be deposited and maintained on TEs without spreading to the transcription start sites of essential genes (3,67). The selective, reliable and mutually exclusive recruitment of Pol II and Pol V to defined sequences and chromatin states helps to define otherwise fluid boundaries between TEs and genes (13,25,68,69). Modulating Pol IV activity within the confines of RdDM targets could provide an additional safeguard by delimiting regions of 24 nt siRNA biogenesis (Figure 7A). Nevertheless,

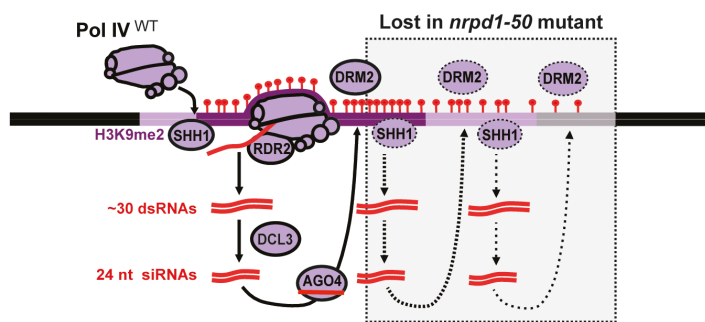
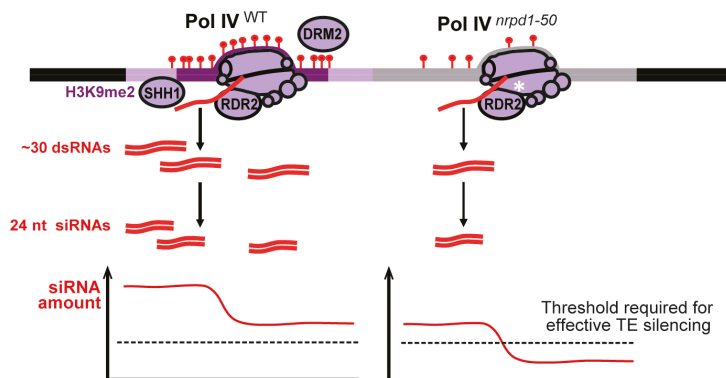
A NRPD1 N-terminus function permits DNA methylation to spread across TEs.**B** NRPD1 N-terminus allows Pol IV activity to reach a threshold for TE silencing.

Figure 7. Distinct models for NRPD1 N-terminus function in genome surveillance. Pol IV can be recruited to chromosomal targets via an accessory protein, SHH1, that recognizes dimethylated histone 3 lysine 9 (H3K9me2) (31). Pol IV and its partner enzyme RDR2 synthesize ~30 bp dsRNAs, which are cleaved by DCL3 into 24 nt siRNAs (6,8–10). AGO4 loaded with a 24 nt siRNA then guides *de novo* DNA methylation (red markers) through Pol V and DRM2 (14,15). (A) In one model, RdDM feeds forward, amplifies and spreads across TEs, leading to robust DNA methylation patterning in WT plants (Pol IV^{WT}). In plants expressing Pol IV^{*nrpd1-50*}, by contrast, RdDM initiates at particular regions but is unable to spread across TEs (gray box), providing for only sporadic siRNA production. With the NRPD1 N-terminus disrupted, the residual DNA methylation in *nrpd1-50* plants is thus insufficient to maintain genome surveillance. (B) In an alternative model, Pol IV-dependent siRNA biogenesis requires a fully functional NRPD1 N-terminus (Pol IV^{WT}) in order to reach the threshold for robust RdDM and TE silencing. In plants expressing Pol IV^{*nrpd1-50*}, the overall siRNA accumulation (thin red curve) is frequently below this threshold (dotted black line), leading to dramatic but variable DNA methylation losses and reduced genome surveillance in *nrpd1-50* plants.

the mechanisms that negatively regulate Pol IV to prevent deleterious spreading of RdDM remain unclear. A potential Pol IV regulatory function for the NRPD1 N-terminus motif (see Figure 4) should be explored. Moreover, SHH1 and CLSY proteins enhance Pol IV activity *in vivo* (31,34), and could thus permit siRNA levels to reach the putative threshold for TE silencing of the alternative model (Figure 7B). Further studies will be needed to determine whether the NRPD1 N-terminus motif is required for Pol IV partnerships with SHH1 or CLSY proteins, for Pol IV recruitment, or for other steps such as transcription start-site scanning, initiation or elongation.

DATA AVAILABILITY

Next-generation sequencing data generated for this study have been deposited in the NCBI Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) under accession number PRJNA510791.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

L.F. and T.B. designed the study. L.F., C.H., M.B. and T.B. performed all benchtop experimental procedures, except the assays in Figure 2 and Supplementary Figure S2F, which were performed by J.S., and the retrotransposon heat-activation assays in Figure 6D, which were performed by M.T. D.P. conducted the bioinformatic analyses in Figures 3B-C, 5A, 5C, 6A, Supplementary Figures S3 and S5. A.G. and T.B. performed the protein alignments in Supplementary Figures S2A and S4, as well as contributing to NGS analysis and data display. E.B. and C.S.P. provided critical advice for the study and helped edit the manuscript. L.F. and T.B. mounted the final figures and wrote the manuscript. Bart Rymen provided helpful comments on the

work. The authors thank Patrice Dunoyer and Olivier Voinet for providing access to Arabidopsis seed stocks from their *SUC2::IR-SUL* genetic screen, allowing us to characterize NRPD1-deficient lines. This study relied upon the dedicated support of the IBMP gardeners, bioinformatics platform and sequencing facility staff. At last, the authors thank our peer reviewers for improvements to this study afforded by their critical reading of the manuscript.

FUNDING

LabEx NetRNA Consortium, French Investissements d'Avenir funds [ANR-10-LABX-0036_NETRNA to T.B.]; French Agence Nationale de la Recherche Grant [ANR-17-CE20-0004-01 to T.B.]; Howard Hughes Medical Institute funds [to C.S.P.]; European Commission Grants [PITN-GA-2013-608422-IDP BRIDGES to M.T., 725701 BUNGEE to E.B.]. Funding for open access charge: Agence Nationale de la Recherche Grant [ANR-17-CE20-0004-01].

Conflict of interest statement. None declared.

This paper is linked to: <https://doi.org/10.1093/nar/gkz615>.

REFERENCES

- Ozata, D.M., Gainetdinov, I., Zoch, A., O'Carroll, D. and Zamore, P.D. (2019) PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.*, **20**, 89–108.
- Fultz, D., Choudhury, S.G. and Slotkin, R.K. (2015) Silencing of active transposable elements in plants. *Curr. Opin. Plant Biol.*, **27**, 67–76.
- Bucher, E., Reinders, J. and Mirouze, M. (2012) Epigenetic control of transposon transcription and mobility in Arabidopsis. *Curr. Opin. Plant Biol.*, **15**, 503–510.
- Matzke, M.A., Kanno, T. and Matzke, A.J. (2015) RNA-Directed DNA Methylation: The evolution of a complex epigenetic pathway in flowering plants. *Annu. Rev. Plant Biol.*, **66**, 243–267.
- Pikaard, C.S., Haag, J.R., Pontes, O.M., Blevins, T. and Cocklin, R. (2012) A transcription fork model for Pol IV and Pol V-dependent RNA-directed DNA methylation. *Cold Spring Harb. Symp. Quant. Biol.*, **77**, 205–212.
- Singh, J., Mishra, V., Wang, F., Huang, H.-Y. and Pikaard, C. (2019) Reaction mechanisms of Pol IV, RDR2 and DCL3 drive RNA channeling in the siRNA-directed DNA methylation pathway. *Mol. Cell*, **75**, doi:10.1016/j.molcel.2019.07.008.
- Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E. and Carrington, J.C. (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.*, **2**, E104.
- Blevins, T., Podicheti, R., Mishra, V., Marasco, M., Wang, J., Rusch, D., Tang, H. and Pikaard, C.S. (2015) Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. *Elife*, **4**, e09591.
- Haag, J.R., Ream, T.S., Marasco, M., Nicora, C.D., Norbeck, A.D., Pasa-Tolic, L. and Pikaard, C.S. (2012) 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. *Mol. Cell*, **48**, 811–818.
- Zhai, J., Bischof, S., Wang, H., Feng, S., Lee, T.F., Teng, C., Chen, X., Park, S.Y., Liu, L., Gallego-Bartolome, J. et al. (2015) A one precursor one siRNA model for Pol IV-Dependent siRNA biogenesis. *Cell*, **163**, 445–455.
- Zilberman, D., Cao, X. and Jacobsen, S.E. (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*, **299**, 716–719.
- Qi, Y., He, X., Wang, X.J., Kohony, O., Jurka, J. and Hannon, G.J. (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature*, **443**, 1008–1012.
- Wierzbicki, A.T., Haag, J.R. and Pikaard, C.S. (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell*, **135**, 635–648.
- Wierzbicki, A.T., Ream, T.S., Haag, J.R. and Pikaard, C.S. (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat. Genet.*, **41**, 630–634.
- Zhong, X., Du, J., Hale, C.J., Gallego-Bartolome, J., Feng, S., Vashisht, A.A., Chory, J., Wohlschlegel, J.A., Patel, D.J. and Jacobsen, S.E. (2014) Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell*, **157**, 1050–1060.
- Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I. and Paszkowski, J. (2011) An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, **472**, 115–119.
- Cavrak, V.V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L.M. and Mittelsten Scheid, O. (2014) How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genet.*, **10**, e1004115.
- Thieme, M., Lanciano, S., Balzergue, S., Daccord, N., Mirouze, M. and Bucher, E. (2017) Inhibition of RNA polymerase II allows controlled mobilisation of retrotransposons for plant breeding. *Genome Biol.*, **18**, 134.
- Teixeira, F.K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M.F. et al. (2009) A role for RNAi in the selective correction of DNA methylation defects. *Science*, **323**, 1600–1604.
- Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel, D., Paszkowski, J. and Mathieu, O. (2009) Selective epigenetic control of retrotransposition in Arabidopsis. *Nature*, **461**, 427–430.
- Blevins, T., Pontvianne, F., Cocklin, R., Podicheti, R., Chandrasekhara, C., Yerneni, S., Braun, C., Lee, B., Rusch, D., Mockaitis, K. et al. (2014) A Two-Step process for epigenetic inheritance in Arabidopsis. *Mol. Cell*, **54**, 30–42.
- Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C. and Paszkowski, J. (2007) Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell*, **130**, 851–862.
- Saze, H., Mittelsten Scheid, O. and Paszkowski, J. (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat. Genet.*, **34**, 65–69.
- Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M. and Jacobsen, S.E. (2003) Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.*, **13**, 2212–2217.
- Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J. and Jacobsen, S.E. (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat. Struct. Mol. Biol.*, **21**, 64–72.
- Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L. and Zilberman, D. (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*, **153**, 193–205.
- Pikaard, C.S. and Mittelsten Scheid, O. (2014) Epigenetic regulation in plants. *Cold Spring Harb. Perspect. Biol.*, **6**, a019315.
- Du, J., Johnson, L.M., Jacobsen, S.E. and Patel, D.J. (2015) DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.*, **16**, 519–532.
- Tsukahara, S., Kobayashi, A., Kawabe, A., Mathieu, O., Miura, A. and Kakutani, T. (2009) Bursts of retrotransposition reproduced in Arabidopsis. *Nature*, **461**, 423–426.
- Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L. and Pikaard, C.S. (2009) Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol. Cell*, **33**, 192–203.
- Law, J.A., Du, J., Hale, C.J., Feng, S., Krajewski, K., Palanca, A.M., Strahl, B.D., Patel, D.J. and Jacobsen, S.E. (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature*, **498**, 385–389.
- Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard, C.S. and Baulcombe, D.C. (2007) An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. *Plant Cell*, **19**, 1507–1521.
- Law, J.A., Vashisht, A.A., Wohlschlegel, J.A. and Jacobsen, S.E. (2011) SHH1, a homeodomain protein required for DNA methylation, as

- well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. *PLoS Genet.*, **7**, e1002195.
34. Zhou, M., Palanca, A.M.S. and Law, J.A. (2018) Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. *Nat. Genet.*, **50**, 865–873.
 35. Yang, D.L., Zhang, G., Wang, L., Li, J., Xu, D., Di, C., Tang, K., Yang, L., Zeng, L., Miki, D. *et al.* (2018) Four putative SWI2/SNF2 chromatin remodelers have dual roles in regulating DNA methylation in Arabidopsis. *Cell Discov.*, **4**, 55.
 36. Marasco, M., Li, W., Lynch, M. and Pikaard, C.S. (2017) Catalytic properties of RNA polymerases IV and V: accuracy, nucleotide incorporation and rNTP/dNTP discrimination. *Nucleic Acids Res.*, **45**, 11315–11326.
 37. Mari-Ordonez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V. and Voinnet, O. (2013) Reconstructing de novo silencing of an active plant retrotransposon. *Nat. Genet.*, **45**, 1029–1039.
 38. Fultz, D. and Slotkin, R.K. (2017) Exogenous transposable elements circumvent Identity-Based silencing, permitting the dissection of Expression-Dependent silencing. *Plant Cell*, **29**, 360–376.
 39. Cuerda-Gil, D. and Slotkin, R.K. (2016) Non-canonical RNA-directed DNA methylation. *Nat. Plants*, **2**, 16163.
 40. Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J.Y. and Chen, X. (2009) Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. *Genes Dev.*, **23**, 2850–2860.
 41. Sasaki, T., Lee, T.F., Liao, W.W., Naumann, U., Liao, J.L., Eun, C., Huang, Y.Y., Fu, J.L., Chen, P.Y., Meyers, B.C. *et al.* (2014) Distinct and concurrent pathways of Pol II- and Pol IV-dependent siRNA biogenesis at a repetitive trans-silencer locus in Arabidopsis thaliana. *Plant J.*, **79**, 127–138.
 42. Li, Y., Yuan, Y., Fang, X., Lu, X., Lian, B., Zhao, G. and Qi, Y. (2018) A role for MINIYO and QUATRE-QUART2 in the assembly of RNA polymerases II, IV, and V in Arabidopsis. *Plant Cell*, **30**, 466–480.
 43. Dunoyer, P., Himber, C., Ruiz-Ferrer, V., Alioua, A. and Voinnet, O. (2007) Intra- and intercellular RNA interference in Arabidopsis thaliana requires components of the microRNA and heterochromatic silencing pathways. *Nat. Genet.*, **39**, 848–856.
 44. Herr, A.J., Jensen, M.B., Dalmay, T. and Baulcombe, D.C. (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science*, **308**, 118–120.
 45. He, X.J., Hsu, Y.F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.S. and Zhu, J.K. (2009) NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. *Genes Dev.*, **23**, 318–330.
 46. Onodera, Y., Haag, J.R., Ream, T., Costa Nunes, P., Pontes, O. and Pikaard, C.S. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell*, **120**, 613–622.
 47. Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V. and Lagrange, T. (2005) 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.
 48. Haag, J.R., Pontes, O. and Pikaard, C.S. (2009) Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. *PLoS One*, **4**, e4110.
 49. Lahmy, S., Pontier, D., Cavel, E., Vega, D., El-Shami, M., Kanno, T. and Lagrange, T. (2009) 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.
 50. Mosher, R.A., Schwach, F., Studholme, D. and Baulcombe, D.C. (2008) PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 3145–3150.
 51. Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E. and Pikaard, C.S. (2006) The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell*, **126**, 79–92.
 52. Faurobert, M., Pelpoir, E. and Chaib, J. (2007) Phenol extraction of proteins for proteomic studies of recalcitrant plant tissues. *Methods Mol. Biol.*, **355**, 9–14.
 53. Blevins, T. (2017) Northern blotting techniques for small RNAs. *Methods Mol. Biol.*, **1456**, 141–162.
 54. Huang, Y., Kendall, T., Forsythe, E.S., Dorantes-Acosta, A., Li, S., Caballero-Perez, J., Chen, X., Arteaga-Vazquez, M., Beilstein, M.A. and Mosher, R.A. (2015) Ancient origin and recent innovations of RNA Polymerase IV and V. *Mol. Biol. Evol.*, **32**, 1788–1799.
 55. Archambault, J. and Friesen, J.D. (1993) Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiol. Rev.*, **57**, 703–724.
 56. Donaldson, I.M. and Friesen, J.D. (2000) Zinc stoichiometry of yeast RNA polymerase II and characterization of mutations in the zinc-binding domain of the largest subunit. *J. Biol. Chem.*, **275**, 13780–13788.
 57. Cramer, P., Bushnell, D.A. and Kornberg, R.D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, **292**, 1863–1876.
 58. El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R. and Lagrange, T. (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.*, **21**, 2539–2544.
 59. Montavon, T., Kwon, Y., Zimmermann, A., Hammann, P., Vincent, T., Cognat, V., Bergdoll, M., Michel, F. and Dunoyer, P. (2018) Characterization of DCL4 missense alleles provides insights into its ability to process distinct classes of dsRNA substrates. *Plant J.*, **95**, 204–218.
 60. Lahmy, S., Pontier, D., Bies-Etheve, N., Laudie, M., Feng, S., Jobet, E., Hale, C.J., Cooke, R., Hakimi, M.A., Angelov, D. *et al.* (2016) Evidence for ARGONAUTE4-DNA interactions in RNA-directed DNA methylation in plants. *Genes Dev.*, **30**, 2565–2570.
 61. Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A., Terragni, J., Chin, H.G., Tu, A. *et al.* (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*, **151**, 167–180.
 62. Kaplan, C.D., Larsson, K.M. and Kornberg, R.D. (2008) The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol. Cell*, **30**, 547–556.
 63. Yu, R., Wang, X. and Moazed, D. (2018) Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation. *Nature*, **558**, 615–619.
 64. Creasey, K.M., Zhai, J., Borges, F., Van Ex, F., Regulski, M., Meyers, B.C. and Martienssen, R.A. (2014) miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature*, **508**, 411–415.
 65. Borges, F., Parent, J.S., van Ex, F., Wolff, P., Martinez, G., Kohler, C. and Martienssen, R.A. (2018) Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in Arabidopsis. *Nat. Genet.*, **50**, 186–192.
 66. Panda, K., Ji, L., Neumann, D.A., Daron, J., Schmitz, R.J. and Slotkin, R.K. (2016) Full-length autonomous transposable elements are preferentially targeted by expression-dependent forms of RNA-directed DNA methylation. *Genome Biol.*, **17**, 170.
 67. Saze, H., Shiraishi, A., Miura, A. and Kakutani, T. (2008) Control of genetic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana. *Science*, **319**, 462–465.
 68. Bohmdorfer, G., Sethuraman, S., Rowley, M.J., Krzyszton, M., Roth, M.H., Bouzit, L. and Wierzbicki, A.T. (2016) Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. *Elife*, **5**, e19092.
 69. Gallego-Bartolome, J., Liu, W., Kuo, P.H., Feng, S., Ghoshal, B., Gardiner, J., Zhao, J.M., Park, S.Y., Chory, J. and Jacobsen, S.E. (2019) Co-targeting RNA polymerases IV and V promotes efficient de novo DNA methylation in Arabidopsis. *Cell*, **176**, 1068–1082.
 70. Zhou, M. and Law, J.A. (2015) RNA Pol IV and V in gene silencing: rebel polymerases evolving away from Pol II's rules. *Curr. Opin. Plant Biol.*, **27**, 154–164.
 71. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M. and Cramer, P. (2016) Structure of transcribing mammalian RNA polymerase II. *Nature*, **529**, 551–554.