

Distinct and concurrent pathways of Pol II- and Pol IV-dependent siRNA biogenesis at a repetitive *trans*-silencer locus in *Arabidopsis thaliana*

Taku Sasaki^{1,2,†}, Tzuu-fen Lee³, Wen-Wei Liao¹, Ulf Naumann², Jo-Ling Liao¹, Changho Eun^{2,‡}, Ya-Yi Huang¹, Jason L. Fu¹, Pao-Yang Chen¹, Blake C. Meyers³, Antonius J.M. Matzke¹ and Marjori Matzke^{1,*}

¹Institute of Plant and Microbial Biology, Academia Sinica, 128, Sec. 2, Academia Road, Nankang 115, Taipei Taiwan,

²Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Dr. Bohr-Gasse 3, Vienna Austria, and

³Department of Plant and Soil Sciences, Delaware Biotechnology Institute, University of Delaware, 15 Innovation Way, Newark DE 19711, USA

Received 6 January 2014; revised 21 April 2014; accepted 25 April 2014; published online 5 May 2014.

*For correspondence (e-mail marjorimatzke@gate.sinica.edu.tw).

[†]Present address: Plant Genomic Network Research Team, RIKEN Center for Sustainable Resource Science, 1-7-22 Suehiro-cho, Tsurumi-ku Yokohama Kanagawa 230-0045, Japan.

[‡]Present address: Subtropical Horticulture Research Institute, JEJU National University, 102 Jejudaehakro, Jeju Korea.

SUMMARY

Short interfering RNAs (siRNAs) homologous to transcriptional regulatory regions can induce RNA-directed DNA methylation (RdDM) and transcriptional gene silencing (TGS) of target genes. In our system, siRNAs are produced by transcribing an inverted DNA repeat (IR) of enhancer sequences, yielding a hairpin RNA that is processed by several Dicer activities into siRNAs of 21–24 nt. Primarily 24-nt siRNAs trigger RdDM of the target enhancer in *trans* and TGS of a downstream *GFP* reporter gene. We analyzed siRNA accumulation from two different structural forms of a *trans*-silencer locus in which tandem repeats are embedded in the enhancer IR and distinguished distinct RNA polymerase II (Pol II)- and Pol IV-dependent pathways of siRNA biogenesis. At the original silencer locus, Pol-II transcription of the IR from a *35S* promoter produces a hairpin RNA that is diced into abundant siRNAs of 21–24 nt. A silencer variant lacking the *35S* promoter revealed a normally masked Pol IV-dependent pathway that produces low levels of 24-nt siRNAs from the tandem repeats. Both pathways operate concurrently at the original silencer locus. siRNAs accrue only from specific regions of the enhancer and embedded tandem repeat. Analysis of these sequences and endogenous tandem repeats producing siRNAs revealed the preferential accumulation of siRNAs at GC-rich regions containing methylated CG dinucleotides. In addition to supporting a correlation between base composition, DNA methylation and siRNA accumulation, our results highlight the complexity of siRNA biogenesis at repetitive loci and show that Pol II and Pol IV use different promoters to transcribe the same template.

Keywords: *Arabidopsis thaliana*, inverted repeats, RNA-directed DNA methylation, Pol IV, RNA polymerase II, siRNAs, tandem repeats.

INTRODUCTION

RNA-directed DNA methylation (RdDM) is a short interfering RNA (siRNA)-mediated epigenetic pathway that contributes to the transcriptional silencing of transposons and other repetitive sequences in plants. RdDM requires specialized transcriptional machinery that includes two RNA polymerase II (Pol II)-related, plant-specific RNA polymerases, called Pol IV and Pol V, as well as a number of ancillary factors that assist Pol-IV and Pol-V transcription, and participate in the silencing effector complex. Pol IV initiates siRNA biogenesis by transcribing a single-stranded RNA that is

copied into double-stranded RNA by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2). The double-stranded RNA is processed by DICER-LIKE 3 (DCL3) to produce 24-nt siRNAs that are the major trigger for DNA methylation. The Pol-V complex acts downstream of siRNA biogenesis to generate scaffold RNAs that are believed to base-pair with complementary siRNAs incorporated into ARGONAUTE 4, resulting in the recruitment of DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) to catalyze *de novo* methylation of DNA at the target site (He *et al.*, 2011;

Pikaard *et al.*, 2012; Wierzbicki, 2012; Matzke and Mosher, 2014).

Short interfering RNAs of 24 nt can also be generated in a Pol IV–RDR2-independent pathway, wherein Pol II transcribes a DNA inverted repeat (IR), producing a single-stranded RNA that is self-complementary and folds back on itself to form an RNA hairpin. The hairpin RNA is processed redundantly by several DCL enzymes, including DCL3, into siRNAs of 21–24 nt (Dunoyer *et al.*, 2005; Kanno *et al.*, 2008; Daxinger *et al.*, 2009). Transgene IRs are frequently used to silence genes at the transcriptional (Eamens *et al.*, 2008; Kanno *et al.*, 2008; Finke *et al.*, 2012) and post-transcriptional levels (Dunoyer *et al.*, 2005; Fusaro *et al.*, 2006; Molnar *et al.*, 2010). In *Arabidopsis thaliana* (*Arabidopsis*), several endogenous IRs that give rise to hairpin-derived siRNAs have been described (Henderson *et al.*, 2006; Zhang *et al.*, 2007; Dunoyer *et al.*, 2010).

As assessed by genome-wide patterns of DNA methylation and siRNA accumulation, the primary targets of RdDM are transposons and non-protein coding repeats (Cokus *et al.*, 2008; Lister *et al.*, 2008). These sequences are most abundant in pericentromeric heterochromatin, where they are targets of both RdDM-dependent and RdDM-independent epigenetic silencing pathways (Zemach *et al.*, 2013); however, a number of targets of RdDM are also present in genic and intergenic regions in euchromatin (Huettel *et al.*, 2006; Lee *et al.*, 2012; Zemach *et al.*, 2013). Features that recruit Pol IV and Pol V are not yet fully understood, but specific nucleic acid structures, RNA and certain chromatin modifications may all play a role (Haag and Pikaard, 2011). Two recent studies suggested that Pol IV is recruited to a subset of its targets through its association with SAWADEE HOMEODOMAIN HOMOLOG 1/DNA-BINDING TRANSCRIPTION FACTOR 1 (SHH1/DTF1), which binds histone H3 methylated at lysine 9 through its SAWADEE domain (Law *et al.*, 2013; Zhang *et al.*, 2013). Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) analyses of genome-wide Pol-V occupancy identified up to two thousand peaks corresponding to transposons and genes that are distributed among all five *Arabidopsis* chromosomes (Wierzbicki *et al.*, 2012; Zhong *et al.*, 2012). A common sequence motif at these loci that might be involved in recruiting Pol V was not identified (Wierzbicki *et al.*, 2012). It therefore remains unclear whether Pol IV and Pol V recognize and use conventional promoters in a way similar to Pol II (Haag and Pikaard, 2011).

Various tandem repeats, such as those in the promoter regions of *FLOWERING WAGENINGEN* (*FWA*; Soppe *et al.*, 2000; Chan *et al.*, 2006) and *SUPPRESSOR OF DRM1 DRM2 CMT3* (*SDC*; Henderson and Jacobsen, 2008) in *Arabidopsis*, and at the *b1* locus in *Zea mays* (maize) (Sidorenko *et al.*, 2009), have been identified as targets of RdDM and sources of Pol IV–RDR2-dependent 24-nt siRNAs.

This may reflect an inherent ability of tandem repeats to sustain RDR-DCL-dependent siRNA biogenesis (Martienssen, 2003). Similarly to other types of repeat, the majority of tandem repeats with overlapping siRNAs and DNA methylation are present in pericentromeric heterochromatin. A previous survey of unique tandem repeats in euchromatic regions found that only a small percentage are associated with siRNAs and DNA methylation in *Arabidopsis* (Chan *et al.*, 2006). Thus, tandem repeats – particularly in euchromatic contexts – are not invariably targets of Pol IV- and Pol V-mediated RdDM. The reasons for the differential siRNA accumulation and methylation of euchromatic tandem repeats remain unclear.

We used a transgene silencing system that incorporates both an IR and tandem repeats in a forward-genetic screen to identify factors required for RdDM and transcriptional gene silencing (TGS) in *Arabidopsis*. The silencing system consists of a target (*T*) locus containing a *GFP* reporter gene under the control of an upstream target enhancer, and an unlinked silencer (*S*) locus containing an IR of target enhancer sequences downstream of the *35S* promoter. A short tandem repeat comprising three copies of a 42-bp monomer is embedded in the enhancer sequence, which is derived from an endogenous pararetrovirus (EPRV) of *Nicotiana tomentosiformis* (Gregor *et al.*, 2004; Chabannes and Iskra-Caruana, 2013). Pol-II transcription of the enhancer IR produces a hairpin RNA that is processed by several DCL activities to produce 21-, 22- and 24-nt siRNAs (Kanno *et al.*, 2008). Primarily the 24-nt size class, which is produced by DCL3, induces methylation in *trans* of the target enhancer at the *T* locus, resulting in TGS of the *GFP* gene (Daxinger *et al.*, 2009; Figure 1). To identify mutants in this RdDM pathway, we treated doubly homozygous seeds, which harbor *T* and *S* transgenes, with ethyl methanesulfonate (EMS) and screened M_2 seedlings for release of *GFP* silencing (Kanno *et al.*, 2008). We retrieved 13 *dms* (*defective in meristem silencing*) mutants harboring recessive, loss-of-function mutations in genes encoding Pol-V pathway components (Eun *et al.*, 2012).

In addition to mutations in genes encoding RdDM factors, our screen could also potentially recover so-called *cis* mutants in which structural alterations of the *S* locus abolished the silencing ability and restored *GFP* expression in otherwise wild-type plants. Here, we report two *cis* mutants identified in our screen and describe how one allowed us to dissect distinct Pol II- and Pol IV-dependent pathways of siRNA biogenesis operating at the original silencer locus. We compare these two pathways and discuss sequence features that promote siRNA accumulation at tandem repeats in euchromatin.

RESULTS

Similarly to the *dms* mutants, two *cis* mutants, named 1.6x and 15.5a, were identified by their GFP-positive phenotype

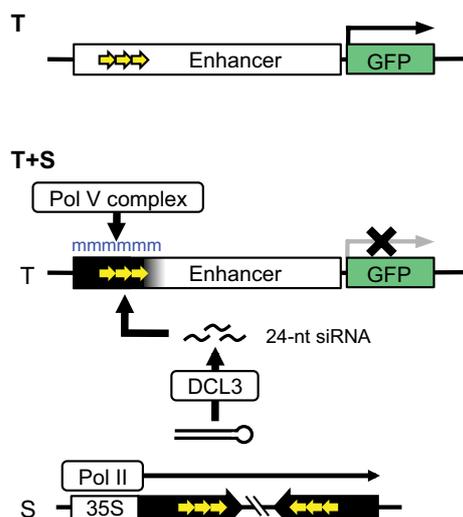


Figure 1. The *T+S* transgene silencing system. The target locus (*T*) contains a GFP reporter gene downstream of a minimal *35S* promoter (from position -46 to position $+8$ of the *35S* promoter sequence; Benfey *et al.*, 1989) and a 1276-bp enhancer, derived from an endogenous pararetrovirus (EPRV), that drives GFP expression in meristem regions (Kanno *et al.*, 2008). The upstream portion of the enhancer contains a naturally occurring short tandem repeat (three copies of a 42-bp monomer; yellow arrows). The origin of the tandem repeat in the enhancer and whether it binds transcription factors is not known. The silencer locus (*S*) comprises an inverted repeat (IR) of distal enhancer sequences containing the tandem repeats (approximately 295 bp, shaded black), separated by a 280-bp spacer (slashed, to indicate truncation; Figure S2). The enhancer IR is transcribed from the *35S* promoter by Pol II. The resulting hairpin RNA is processed by DCL3 to produce 24-nt siRNAs that trigger Pol V-mediated methylation (blue 'm') of the target enhancer at the *T* locus to elicit transcriptional gene silencing (TGS) of the GFP gene. Additionally, 21- and 22-nt siRNAs (not shown) are produced by other DICER-LIKE (DCL) activities (Daxinger *et al.*, 2009). The figure is not drawn to scale.

in a *T+S* background. As described below, these mutants did not appear to contain hairpin-derived siRNAs (21–24 nt) in an initial northern blot analysis, suggesting that structural changes at the *S* locus may be responsible for the observed deficiency in *GFP* silencing. Further investigation confirmed this hypothesis.

Structure of silencer locus in *cis* mutants

Polymerase chain reaction (PCR) analysis revealed that the 1.6x mutant contained a substantially altered *S* locus, lacking both the enhancer IR and *35S* promoter (Figure 2). It is possible that the IR is potentially unstable, rendering this region susceptible to spontaneous deletion. This potential instability should be considered when using IRs to induce gene silencing in transgenic plants. The extensive deletion in the 1.6x mutant accounts for the absence of hairpin-derived siRNAs and unmethylated target enhancer (Figure S1).

For the 15.5a mutant, PCR analysis and whole-genome sequencing revealed an unusual structural variant of the *S* locus in which only the *35S* promoter was deleted, by an

unknown mechanism, leaving the enhancer IR largely intact (Figures 2, S2). We will hereafter refer to the altered silencer locus as $S^{\Delta 35S}$ and the 15.5a *cis* mutant as $T+S^{\Delta 35S}$, and the original unaltered silencer locus as S^{WT} and plants containing this locus as $T+S^{WT}$ ('WT', standing for original unaltered or 'wild-type' silencer). The lack of readily detectable hairpin-derived siRNAs in our initial northern blot analysis of $T+S^{\Delta 35S}$ plants (Figure S1) thus reflected the loss of Pol-II transcription of the enhancer IR following the deletion of the *35S* promoter.

GFP expression and methylation of target enhancer in $T+S^{\Delta 35S}$ plants

Unexpectedly, despite the absence of hairpin-derived siRNAs in $T+S^{\Delta 35S}$ plants, *GFP* expression was still partially silenced, as assessed by the visualization of GFP fluorescence in seedlings (Figure 3a, compare $T+S^{\Delta 35S}$ with un silenced *T* and fully activated $T+S^{WT}$ *nrpe1*) and GFP protein levels on western blots (Figure 3b, compare lanes 2 and 4). Consistent with partial *GFP* silencing, the target enhancer retained residual methylation, which was restricted to the tandem repeat region (Figure 3c, $T+S^{\Delta 35S}$). This methylation pattern differed from that observed in $T+S^{WT}$ plants, where *GFP* is fully silenced and methylation extends throughout the entire target enhancer (Figure 3c, $T+S^{WT}$). Similarly to the enhancer methylation in $T+S^{WT}$ plants (Kanno *et al.*, 2008; Daxinger *et al.*, 2009), the tandem repeat methylation in $T+S^{\Delta 35S}$ plants is the result of RdDM, as indicated by its loss in a mutant defective in Pol-V function (*nrpe1*) and dependence on the $S^{\Delta 35S}$ locus (Figure S3, $T+S^{WT}$ *nrpe1*).

Accumulation of siRNAs in $T+S^{\Delta 35S}$ plants

In view of the tandem repeat methylation in the $T+S^{\Delta 35S}$ plants and previous reports that tandem repeats can be sources of Pol IV-dependent siRNAs (Chan *et al.*, 2006; Henderson and Jacobsen, 2008; Sidorenko *et al.*, 2009), we tested again for siRNAs in $T+S^{\Delta 35S}$ plants by northern blotting using a probe specific to the tandem repeat region. After a long exposure time, we were able to detect a faint 24-nt band in these plants (Figure 4, lane 2). Unlike Pol II-dependent hairpin-derived siRNAs (21–24 nt), which were detectable on northern blots in $T+S^{WT}$ plants in a Pol-IV (*nrpd1*) mutant background (Figure 4, lane 4), the 24-nt siRNAs in $T+S^{\Delta 35S}$ plants disappeared in the *nrpd1* mutant (Figure 4, lanes 5 and 6). The disappearance of the 24-nt siRNAs coincided with a complete loss of methylation in the tandem repeat at the *T* locus (Figure 3c, $T+S^{\Delta 35S}$ *nrpd1*) and with a full reactivation of *GFP* expression (Figure 3a, $T+S^{\Delta 35S}$ *nrpd1*; Figure 3b, lanes 5 and 6). By contrast, the extensive methylation induced by Pol II-dependent, hairpin-derived siRNAs originating from the intact S^{WT} locus was not eliminated in an *nrpd1* mutant (Figure 3c, $T+S^{WT}$ *nrpd1*; Daxinger *et al.*, 2009), nor was

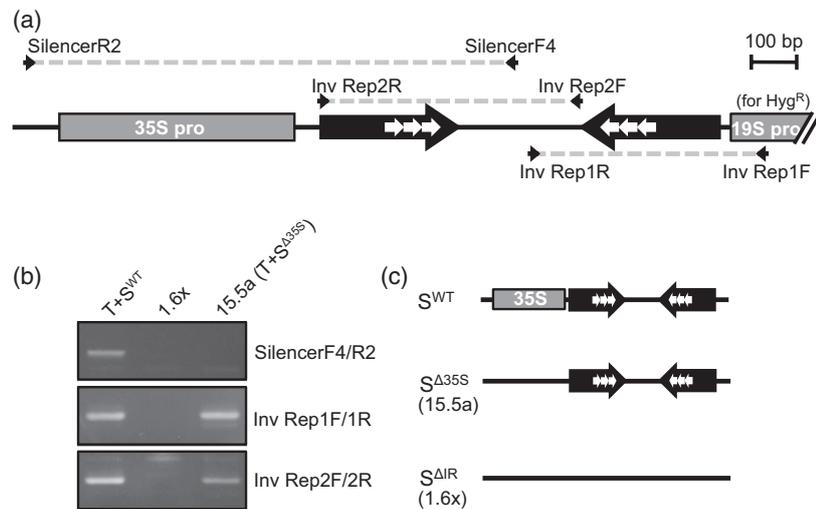


Figure 2. Analysis of silencer locus structure in *cis* mutants.

(a) S^{WT} contains an enhancer inverted repeat (IR, thick black arrows), separated by a spacer, under the transcriptional control of the 35S promoter. White arrows indicate the tandem repeat in the target enhancer sequence. A 19S promoter driving expression of a gene encoding resistance to hygromycin (Hyg^R) is present, in opposite orientation, to the right of the IR (Figure S2). Arrowheads joined by dotted lines indicate the positions of primers used to detect the two halves of the IR and the region containing the 35S promoter. (b) PCR analysis to determine the structure of the silencer locus in $T+S^{WT}$ plants and the *cis* mutants, 1.6x and 15.5a ($T+S^{\Delta 35S}$). The Inv Rep1 and Inv Rep2 primer pairs detect, respectively, the two halves of the IR. The silencer F4/R2 primers detect the 35S promoter and one-half of the IR (positions of primers indicated in part a). (c) Structures of altered silencer loci in the *cis* mutants. S^{WT} contains an intact 35S promoter and IR. The 15.5a silencer ($S^{\Delta 35S}$) lacks the 35S promoter while retaining the IR (Figure S2). The 1.6x silencer ($S^{\Delta IR}$) lacks the entire 35S promoter-IR region while maintaining the silencer transgene-plant DNA junction fragment (Figure S1D).

GFP expression reactivated (Figure 3a, $T+S^{WT} nrpd1$; Figure 3b, lane 3).

$T+S^{\Delta 35S}$ plants thus revealed a previously obscured class of Pol IV-dependent 24-nt siRNAs that only became evident following the deletion of the 35S promoter at the $S^{\Delta 35S}$ locus, which eliminated the more abundant Pol II-dependent hairpin-derived siRNAs. Despite their relatively low abundance, Pol IV-dependent 24-nt siRNAs were able to trigger methylation in the tandem repeat at the *T* locus, resulting in the partial silencing of *GFP* expression in $T+S^{\Delta 35S}$ plants. Pol IV-dependent biogenesis of tandem repeat 24-nt siRNAs relies at least to some extent on RDR2, as illustrated by the partial loss of *GFP* silencing and target enhancer methylation in $T+S^{\Delta 35S} rdr2$ plants (Figure S4).

Pol II- and Pol IV-dependent pathways appear to contribute contemporaneously to siRNA accumulation in $T+S^{WT}$ plants, as shown by a northern blot analysis of siRNA accumulation in these plants, and in mutants deficient in Pol-II and Pol-IV function. Pol II-dependent 21-nt siRNAs were the major size class in the *nrpd1* mutant, in which Pol IV-dependent 24-nt siRNAs are substantially reduced relative to $T+S^{WT}$ plants (Figure 4, compare lanes 4 and 1). By contrast, Pol IV-dependent 24-nt siRNAs were the major size class in a hypomorphic *nrpb2* mutant, which displayed lower levels of Pol II-dependent 21- and 22-nt siRNAs, compared with $T+S^{WT}$ plants (Figure 4, compare lanes 7 and 1; Figure S5). The pattern of siRNA accumulation in $T+S^{WT}$ plants thus represents the sum of Pol II- and Pol IV-dependent siRNAs,

suggesting that these pathways operate concurrently at the original unaltered silencer locus.

Biased accumulation of Pol II- and Pol IV-dependent siRNAs at target enhancer

To examine siRNA accumulation from Pol-II and Pol-IV pathways in more detail, we deeply sequenced small RNAs from $T+S^{WT}$ and $T+S^{\Delta 35S}$ plants. A marked non-uniform distribution of siRNAs from both pathways was observed along the enhancer sequence (Figure 5). A similar pattern of siRNA accumulation in $T+S^{WT}$ plants was observed in an independent study (Melnik *et al.*, 2011). Both Pol II- and Pol IV-dependent siRNAs were concentrated in the tandem repeat region and in a short upstream peak (Figure 5a, $T+S^{WT}$ and $T+S^{\Delta 35S}$). Comparable accumulation patterns were observed in the absence of the *T* locus (Figure 5a, S^{WT} and $S^{\Delta 35S}$), which produced negligible quantities of siRNAs on its own (Figure 5a, T), confirming that essentially all siRNAs originated at the enhancer sequences of the respective silencer locus. As expected from the northern blot analysis, despite the similar accumulation patterns for Pol II- and Pol IV-dependent siRNAs, the major size class differed for the two pathways, with the predominant species being 21 and 24 nt, respectively (Figure 5b). As reported previously, siRNA generation downstream of the enhancer sequence was observed only in $T+S^{WT}$ plants (Figure S6; Daxinger *et al.*, 2009), consistent with the spread of methylation into this region (Figure 3c).

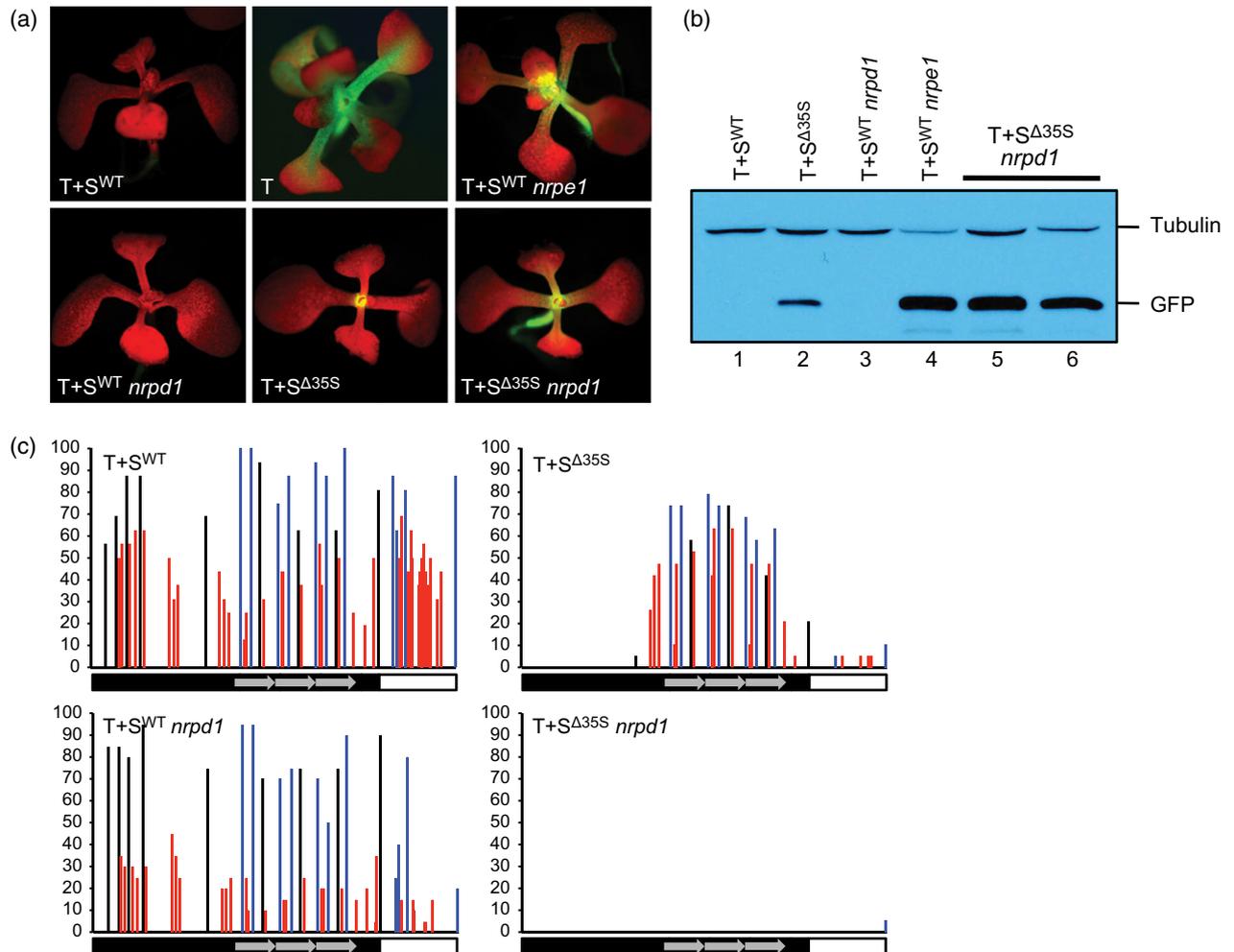


Figure 3. Analysis of GFP expression and DNA methylation status of the target enhancer.

(a, b) GFP expression in the shoot meristem region was visualized in seedlings under a fluorescence microscope (a) and by performing Western blotting using a GFP antibody (b). In *T* plants, GFP is expressed. In *T+S^{WT}* plants, GFP is silenced. Silencing persists in a Pol-IV mutant (*nrpd1*) but is fully released in a Pol-V mutant (*nrpe1*). In *T+S^{Δ35S}* plants, GFP silencing is partially released. Introducing an *nrpd1* mutation into *T+S^{Δ35S}* plants fully releases GFP silencing (*T+S^{Δ35S} nrpd1*, two examples shown, lanes 5 and 6 in panel b). The accumulation of constitutively expressed tubulin protein is shown as a loading control in (b). (c) Bisulfite sequencing was used to analyze DNA methylation at the target enhancer (black bar, gray arrows represent the tandem repeat) in the indicated genotypes. The primers used to amplify the enhancer fragment included one that was specific for the *T* locus. The y-axis shows the percentage methylation at individual cytosines. In *T+S^{WT}* plants, the entire enhancer was methylated (CG, black; CHG, blue; CHH, red), with some spreading into the downstream region (white bar). A similar pattern was observed in an *nrpd1* mutant (*T+S^{WT} nrpd1*) with reduced CHH methylation, particularly in the downstream region, owing to the loss of Pol IV-dependent secondary siRNAs (Daxinger *et al.*, 2009). In *T+S^{Δ35S}* plants, methylation was restricted to the tandem repeats and was dependent on Pol IV (compare *T+S^{Δ35S}* with *T+S^{Δ35S} nrpd1*).

In principle, the concentration of siRNAs in the tandem repeat region can be explained by the proposed affinity of Pol IV for such repeats (discussed further below); however, the uneven distribution of Pol II-dependent siRNAs was unanticipated because they are produced from a perfectly base-paired hairpin RNA that is expected to be processed homogeneously by DCL activities (Dalakouras and Wassenegger, 2013). One possibility is that Pol II-dependent siRNAs may accumulate preferentially from internal regions of the IR, which contain the tandem repeats (Figure 1), as suggested in an earlier study (Stam *et al.*, 1998). Another factor to consider is sequence composition. This possibility

was suggested by the siRNA-depleted spacer between the tandem repeats and upstream peak (Figure 5a, all genotypes). This spacer region is relatively GC-poor compared with the flanking upstream peak and tandem repeats (Figure S7).

siRNA accumulation biases at another transgene IR and endogenous tandem repeats

To extend our analysis to additional repeats, we examined siRNA accumulation at another transgenic IR for which deep sequencing data are available and at endogenous tandem repeats in euchromatin.

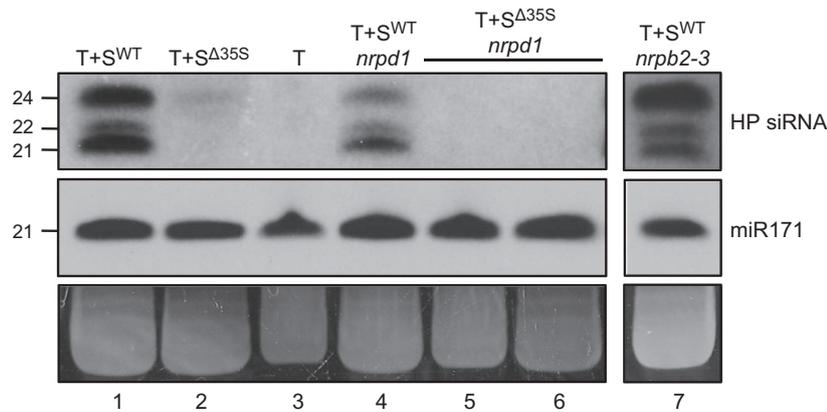


Figure 4. Northern blot analysis of siRNAs. siRNAs were analyzed using a probe for the enhancer tandem repeat in plants containing the *T* locus and either S^{WT} or $S^{\Delta35S}$ in different mutant backgrounds. $T+S^{WT}$ plants contain the full complement of siRNAs (21–24 nt; lane 1). In $T+S^{WT}$ *nrpd1* plants, only Pol II-dependent siRNAs are present (lane 4), whereas in a $T+S^{WT}$ hypomorphic *nrpb2* mutant, specifically Pol II-dependent siRNAs are reduced (lane 7). In $T+S^{\Delta35S}$ plants (lane 2), only a faint band of 24-nt siRNAs was detected in $T+S^{\Delta35S}$ plants (lane 2), and these disappeared in an *nrpd1* background (lanes 5 and 6), demonstrating their dependence on Pol IV. In $T+S^{WT}$ plants the levels of 24-nt siRNAs do not exceed the levels of 21-nt siRNAs (lane 1), suggesting the presence of both the Pol II-dependent siRNAs (21–24 nt, with 21 nt predominating; lane 4) and Pol IV-dependent siRNAs (24 nt, lane 2). By contrast, in the hypomorphic *nrpb2-3* mutant, 24-nt siRNAs, which include those produced by the Pol-IV pathway, predominate (lane 7; see also Figure S5). siRNAs were not detectable by this approach in the *T* line (lane 3). Note that hairpin-derived siRNAs (21–24 nt) are not reduced in an *nrpe1* background (Kanno *et al.*, 2008), indicating that Pol V contributes negligibly to their biogenesis. The same blot was hybridized with a miR171 probe as a hybridization control. The major band on the gel staining with ethidium bromide is shown as a loading control.

In a previous study, deep sequencing of siRNAs in a transgenic line expressing an IR of '*GF*' sequences to silence a *GFP* reporter gene revealed distinct 'hot spots' of siRNA accumulation (Molnar *et al.*, 2010). We analyzed the sequencing data and found that the siRNA peaks represent the more internal parts of the IR and generally correspond to regions that contain the most CG dinucleotides (Figure S8).

Using `ETANDEM` (<http://emboss.sourceforge.net/apps/release/6.0/emboss/apps/etandem.html>) to screen the Arabidopsis genome (TAIR10 version), we identified 786 euchromatic tandem repeats, defined as being at least 4 Mb away from the centromere, with a minimum monomer size of 40 bp and a score >80 (calculated directly by `ETANDEM`; Table S1). Of these, 282 are in intergenic (IGN) regions, 486 are in genes, four are in pseudogenes and 14 are in transposons (Figure S9; Table S1, column I). The majority (68.1%) of IGN tandem repeats are sources of Pol IV-dependent 24-nt siRNAs, compared with only around 9.5% of the tandem repeats in genic regions (Figure S9; Table S1, column L). Consistent with a sequence bias of siRNA accumulation, regions of tandem repeats (all categories) that accumulate more siRNAs are significantly GC-rich (Figure 6a, $P < 10^{-10}$). Moreover, of the IGN and genic tandem repeats that accumulate Pol IV-dependent 24-nt siRNAs, most (91.1 and 78.3%, respectively) contain methylated CG dinucleotides, whereas the majority of IGN and genic tandem repeats that do not accumulate 24-nt siRNAs (82.2 and 72.3%, respectively) lack CG methylation (Figures 6b, S9A) or CG dinucleotides altogether (Table S1, column C, red zeroes). Tandem repeats in pseudogenes

and transposons showed a similar trend (Figure S9A; Table S1, column C). Examples of the correlation between siRNA accumulation and regions of tandem repeats that are relatively GC-rich and contain methylated cytosines are shown in Figure S10. Note that in these examples CG methylation is not lost in an *nrpd1* mutant.

DISCUSSION

We exploited a structural variant identified in a forward genetic screen to dissect distinct Pol II- and Pol IV-dependent pathways of siRNA biogenesis at a *trans*-silencer locus containing inverted and tandem repeats. In addition to highlighting the complexity of siRNA accumulation at compound repetitive loci, our studies allow comparisons between the two pathways operating at a single locus and help to identify general features influencing siRNA accumulation at repeats in euchromatin.

As expected, abundant hairpin-derived siRNAs (21–24 nt) were detected in $T+S^{WT}$ plants, where Pol II transcribes the enhancer IR from a *35S* promoter. Because the siRNAs (21–24 nt) accrued from both the 5' and 3' ends of the target enhancer (upstream peak and tandem repeat, respectively), we presume that the hairpin RNA formed from a full-length, or nearly full-length, single-stranded transcript of the enhancer IR. Pol IV–RDR2-dependent 24-nt siRNAs detected in the $T+S^{\Delta35S}$ mutant derived primarily from the tandem repeat, with lesser quantities accumulating from the upstream peak. The ability of Pol IV to carry out transcription after deletion of the *35S* promoter indicates that it can function in the absence of a constitutive Pol-II promoter. The identity of the promoter used by

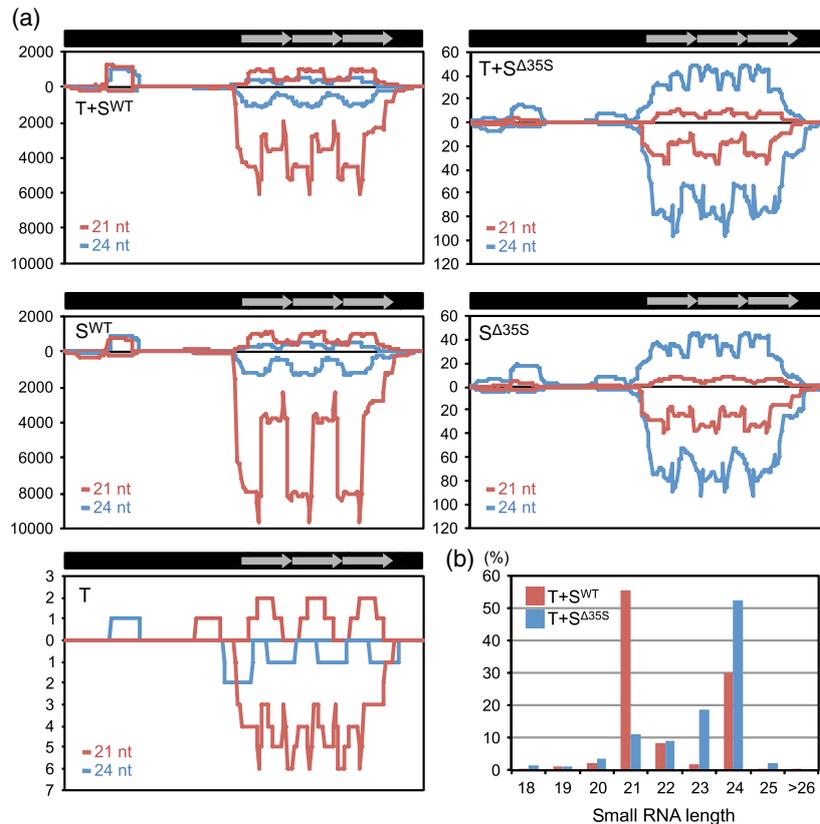


Figure 5. Deep-sequencing analysis of siRNAs accumulating from enhancer sequences at silencer loci.

(a) Accumulation of siRNAs (21 nt in red and 24 nt in blue) at the enhancer sequence (approximately 295 bp, black bar) in both polarities (top and bottom) in the indicated genotypes. The enhancer sequence includes only those sequences present in the two halves of the silencer inverted repeat (IR; Figure S2). The levels are shown in the number of siRNA reads in reads per five million (RP5M), where the x-axis ('per nucleotide') is the enhancer sequence, and the abundance of the sequenced siRNAs is attributed to the nucleotide at the 5' end of the small RNA. Note that the y-axis scales differ for each panel. Accumulation occurred preferentially at the tandem repeat (gray arrows) and in a short segment upstream. $T+S^{WT}$ plants contain abundant 21- and 24-nt siRNAs, with the 21-nt size class accumulating to a higher level (55 versus 30%, respectively; in b, note that the northern blot analysis in Figure 4, lane 1, showing roughly similar levels of 21- and 24-nt siRNAs from the top strand is not strictly quantitative). $T+S^{\Delta 35S}$ plants contained substantially lower levels of siRNAs than $T+S^{WT}$ plants, and were enriched in the 24-nt size class (52 versus 11% for 21-nt siRNAs, part b). The presence of the target locus (*T*) did not influence siRNA accumulation (compare $T+S^{WT}$ and $T+S^{\Delta 35S}$ with S^{WT} and $S^{\Delta 35S}$, respectively), and negligible siRNAs were detected in plants containing only the *T* locus (*T*). (b) The y-axis indicates the percentage of siRNAs accumulating from the top strand in a specific size category. In $T+S^{WT}$ plants, Pol II-generated hairpin-derived 21-nt siRNAs predominated, whereas in the $T+S^{\Delta 35S}$ *cis* mutant, Pol IV-dependent 24-nt siRNAs were the major species.

Pol IV is not clear, but it is likely to reside in or near the tandem repeat itself. Whether this tandem repeat binds transcription factors is not known. Tandem repeats that mediate paramutation at the *b1* locus in maize contain transcriptional regulatory sequences and give rise to Pol IV-dependent siRNAs (Belele *et al.*, 2013).

Pol II-dependent siRNAs were considerably more abundant than Pol IV-dependent siRNAs, which became noticeable only when the Pol-II pathway was eliminated in $T+S^{\Delta 35S}$ plants. Although they produce different quantities of siRNAs, the Pol II- and Pol IV-dependent pathways are not mutually exclusive, and can apparently coexist at the unaltered silencer locus in $T+S^{WT}$ plants (Figure 7). Both pathways generate siRNAs from the tandem repeat, indicating that each polymerase can use this region as a template for transcription. Thus, chromatin features

facilitating transcription by one polymerase do not necessarily preclude transcription by the other; however, the lower levels of siRNAs produced in the Pol-IV pathway suggest that it is much less efficient than the Pol-II pathway.

Although our study does not identify the promoter used by Pol IV, it does illuminate features promoting Pol IV-dependent accumulation of siRNAs at tandem repeats. Tandem repeats are often sites of amplification of Pol IV-RDR2-dependent siRNAs (Chan *et al.*, 2006; Henderson and Jacobsen, 2008; Sidorenko *et al.*, 2009), which is attributed to the ability of RDR and DCL to more efficiently generate siRNAs from a tandem array than from a single copy sequence (Martienssen, 2003). In our study, the major peaks of accumulation of Pol IV-dependent 24-nt siRNAs were observed at the tandem repeats, suggesting that

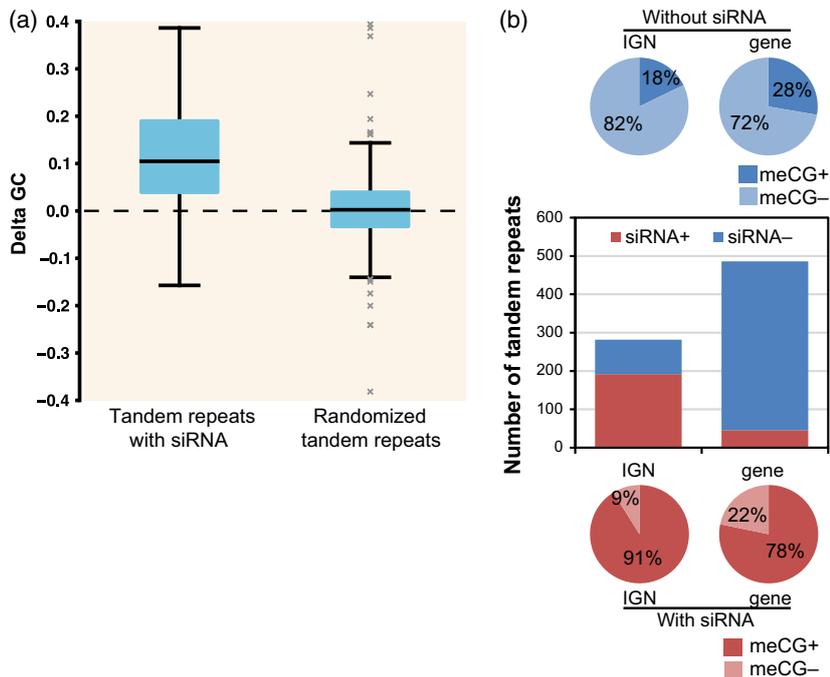


Figure 6. Preferential accumulation of Pol IV-dependent 24-nt siRNA at GC-rich portions of endogenous tandem repeats with CG methylation. (a) siRNA accumulation and GC content. The box plot shows the difference in GC ratio between siRNA-accumulating and non-accumulating regions of the tandem repeats (all categories; Figure S9A, 'siRNA+', 248') and randomized sequences of these tandem repeats. The regions of tandem repeats that accumulate siRNAs are significantly GC-enriched (Student's *t*-test, $P < 10^{-10}$). (b) siRNA accumulation and CG methylation. The bar graph shows the numbers of endogenous tandem repeats with (red) and without (blue) siRNA accumulation in intergenic (IGN) regions and genes. The pie charts show the percentages of tandem repeats with (dark blue/red) and without (light blue/red) CG methylation. Detailed information on the categorization of tandem repeats is shown in Figure S9.

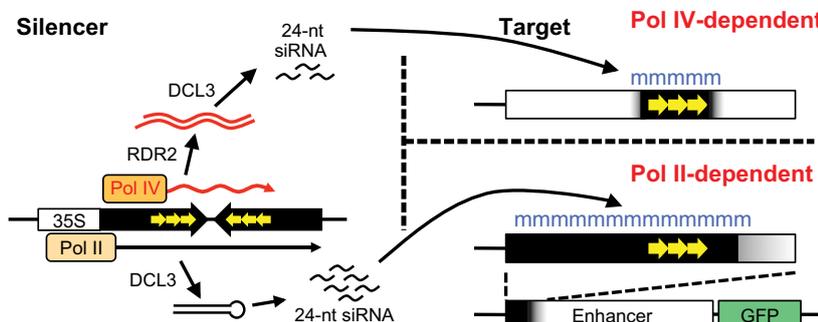


Figure 7. Distinct and concurrent pathways of siRNA biogenesis at a complex repetitive locus. At the unaltered silencer locus (S^{WT}), Pol II initiates transcription at the 35S promoter to produce a hairpin RNA from the enhancer inverted repeat (IR; black), which contains a tandem repeat (yellow arrows). The hairpin RNA is processed by DCL3 to produce 24-nt siRNAs that trigger the methylation (blue 'm') of the entire the target enhancer and *GFP* silencing. As revealed by the S^{L35S} locus, a minor secondary pathway exists in which Pol IV transcribes the tandem repeats to generate a single-stranded RNA that is presumably copied by RDR2 into double-stranded RNA, which is diced by DCL3 to generate 24-nt siRNAs that trigger methylation only at the tandem repeats and partial *GFP* silencing.

Pol IV indeed recognizes these repeats and amplifies siRNAs preferentially in this region. However, we also identified a potential contribution of sequence composition to preferential siRNA accumulation. The tandem repeats as well as the upstream 'peak', which also preferentially accumulates siRNAs, are generally GC-rich compared with the intervening siRNA-depleted spacer region. A similar sequence bias in siRNA accumulation within a tandem

repeat monomer was reported for the *b1* tandem array in maize, where siRNAs accumulated preferentially from the 5' half of the monomer unit, and not from the AT-rich 3' end (Arteaga-Vazquez *et al.*, 2010).

Prompted by these findings, we expanded our analysis by examining siRNA accumulation at 786 endogenous tandem repeats located in euchromatin. This survey revealed that nearly 70% do not accumulate Pol IV-dependent 24-nt

siRNAs, substantiating a previous study showing that euchromatic tandem repeats are not invariably targeted by Pol IV–RDR2 activities (Chan *et al.*, 2006). Around three-quarters of the tandem repeats lacking siRNAs are present in genes, where they might be protected from Pol–IV activity by unidentified chromatin features associated with Pol–II transcription of protein-coding genes. By contrast, the majority of tandem repeats in IGN regions accumulate Pol IV-dependent 24-nt siRNAs and, similarly to the enhancer tandem repeat in the *T+S* system, most of these accrue from more GC-rich segments containing methylated CG dinucleotides. The CG methylation persists in an *nrip1* mutant, indicating that it is maintained independently of RdDM. A scenario consistent with these findings is that Pol IV is recruited to tandem repeats in IGN regions by pre-existing CG methylation (Chan *et al.*, 2006), perhaps established by 21-nt siRNAs and Pol II-dependent mechanisms (Zheng *et al.*, 2009; Nuthikattu *et al.*, 2013; Stroud *et al.*, 2013), and subsequent siRNA amplification occurs through RDR2 and DCL3 activities (Martienssen, 2003). The resulting 24-nt siRNAs induce non-CG methylation to create the specific pattern of methylation of cytosines in all sequence contexts observed at repeats and transposons (Saze and Kakutani, 2011).

As discussed above, the presence of tandem repeats and variations in GC content and CG methylation appear to bias the accumulation of Pol IV-dependent 24-nt siRNAs in the *T+S* system; however, a similar non-uniform siRNA profile was observed for Pol II-generated, hairpin-derived siRNAs (21–24 nt) that are largely independent of Pol–IV activity. This result contrasts to the expectation that Pol II-dependent hairpin-derived siRNAs would be distributed more or less evenly along the entire inverted repeat that constitutes the Pol–II transcription unit (Dalakouras and Wassenegger, 2013). Given the comparable distribution patterns for both Pol IV- and Pol II-dependent siRNAs, we presume that similar factors, including GC content and CG methylation, influence siRNA accumulation of all sizes in both pathways. The enhanced siRNA accumulation at sequences with a higher GC content and methylated CG dinucleotides is consistent with previous studies showing that CG and non-CG methylation can feedback positively on 24-nt siRNA accumulation, perhaps reflecting an affinity of Pol IV for methylated DNA (Law and Jacobsen, 2010; Haag and Pikaard, 2011; Stroud *et al.*, 2014). Our data suggest that, in addition, the accumulation of 21-nt siRNAs is also influenced by sequence composition. In accordance with this result, a recent study revealed a significant correlation between high GC content and increased expression of plant siRNAs of all size classes (Zhang *et al.*, 2014).

A further factor that may influence the preferential accumulation of Pol II-dependent siRNAs from the tandem repeats in the *T+S* system is their position close to the

center of IR, which might be more stably base-paired than distal regions in the hairpin RNA (Stam *et al.*, 1998). This idea is supported by another transgenic silencer locus we examined, in which the hot spots of 'GF siRNA accumulation are also derived from internal regions of the IR (Molnar *et al.*, 2010). This internal region does not contain tandem repeats but is relatively rich in CG dinucleotides. An siRNA peak was also detected close to the center of endogenous IR-71 and, similarly to the enhancer IR in the *T+S* system, this central region contains a short tandem repeat (two copies of a 60-bp monomer; Henderson *et al.*, 2006). Conceivably, this tandem repeat could promote preferential siRNA accumulation by further stabilizing base pairing in this region of the RNA hairpin or attracting Pol IV–RDR2. Local sequence elements or structural features have also been proposed to promote preferential generation of siRNAs involved in post-transcriptional gene silencing (De Paoli *et al.*, 2009). Further work is required to understand fully the siRNA sequence biases that are observed in deep-sequencing efforts. Knowledge gained from such investigations will guide efforts to optimize siRNA production for gene-silencing applications in basic and applied research.

EXPERIMENTAL PROCEDURES

Plant materials

Previous publications describe the isolation of *nrip2-3* (Zheng *et al.*, 2009), *nrip1-7* (Smith *et al.*, 2007) and *nrip1-10* (Kanno *et al.*, 2008) mutants, which are defective in Pol–II (hypomorphic allele), Pol–IV and Pol–V function, respectively. The T-DNA insertion line *rdr2-2* (SALK_059661) was obtained from the Arabidopsis Biological Resource Center (ABRC, <http://abrc.osu.edu>). The 1.6x and 15.5a (*T+S*^{Δ35S}) *cis*-mutants (identified in seed batches 1 and 15, in the sixth and fifth screens, respectively) were screened out by the GFP-positive phenotype of M₂ seedlings derived from EMS-mutagenized M₁ seeds of our *T+S* transgenic line (Kanno *et al.*, 2008; Eun *et al.*, 2012). The *T* and *S* transgene loci are present in intergenic regions on the top and bottom arms of chromosome 1 at nucleotide positions 10 343 986 and approximately 17 480 000, respectively. Silencer (*S*) strains (*S*^{WT} and *S*^{Δ35S}) were generated by crossing out the target (*T*) transgene from *T+S*^{WT} and *T+S*^{Δ35S} lines. Silencer and target transgenes were genotyped using primers shown in Table S2. The sequences of the *T* and *S* transgene constructs are available under the accession numbers HE582394 and HE584556, respectively.

Whole-genome sequencing

The sequence of the silencer locus in the *T+S*^{Δ35S} line (Figure S2) was deduced from whole-genome sequence data. Paired-end sequencing of a genomic DNA library was performed with 72-bp read length using an Illumina Genome Analyzer II, as described previously (Eun *et al.*, 2011; Sasaki *et al.*, 2012). For assembling the sequence of the silencer (*S*^{Δ35S}) locus in the *T+S*^{Δ35S} line, the sequence of the *S*^{WT} locus (accession number HE584556) was used as a reference sequence (Eun *et al.*, 2011). The whole-genome sequencing data from the *T+S*^{Δ35S} line are available from the National Center for Biotechnology Information (NCBI, <http://>

www.ncbi.nlm.nih.gov) Sequence Read Archive (SRA), under accession number SRX312270.

DNA methylation analysis of target enhancer

Bisulfite sequencing was used to determine the DNA methylation status of the target enhancer at the *T* locus. Genomic DNAs were extracted using a DNeasy Plant Mini Kit (Qiagen, <http://www.qiagen.com>). For bisulfite sequencing, 1 µg of genomic DNA was digested by *HindIII* and purified by QIAquick PCR purification kit (Qiagen). DNA (500 ng) was used for bisulfite conversion of unmethylated cytosines to uracils using the EpiTect Bisulfite Kit (Qiagen). Converted DNA was used as a template for PCR using the primers listed in Table S2. One of the primers used was specific for the target locus so that no fragment from the copies of the enhancer IR at the silencer locus was amplified. Amplified fragments were cloned using the T-Easy Vector System (Promega, <http://worldwide.promega.com>), and 14–20 clones were sequenced.

Small RNA northern blotting analysis

Small RNA was isolated from 250 mg of mixed floral inflorescences using a mirVana miRNA Isolation Kit (Ambion, now Life Technologies, <http://www.lifetechnologies.com>). RNA was separated by electrophoresis on a 15% denaturing polyacrylamide/urea gel and transferred to Hybond-N+ (GE Healthcare, <http://www.gehealthcare.com>) by electroblotting using a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad, <http://www.bio-rad.com>). The membrane was hybridized overnight with [γ - 32 P]ATP-labeled oligonucleotides (Table S2) using Ultra Hyb oligo Buffer (Ambion) at 40°C. The two probes used are for the top strand, and cover the entire 42-bp tandem repeat monomer. The membrane was washed twice at 40°C in 2X SSC, 0.5% SDS for 15 min, and exposed to X-ray film for 4 days for hairpin-derived siRNAs and 5 h for miR171 at –80°C.

Western blotting analysis

The protein extraction, SDS-PAGE and western blotting to detect GFP protein were carried out as previously described (Sasaki *et al.*, 2012).

Small RNA deep sequencing

For small RNA libraries, total RNA from the materials described above was isolated using Tri Reagent™ (Molecular Research Center, Inc., <http://www.mrcgene.com>). Small RNA libraries were constructed using the Illumina TruSeq Small RNA Sample Preparation Kit (RS-200-0012) and sequenced on an Illumina HiSeq2000 instrument at the Delaware Biotechnology Institute of the University of Delaware. Raw sequencing data was first trimmed of adapter sequences, then the read counts were normalized based on the total abundance of genome-matched small RNA reads, excluding structural sRNAs originating from annotated tRNA, rRNA, small nuclear and small nucleolar RNAs. For the transgene loci (wild-type or mutated Silencer), BOWTIE (<http://bowtie-bio.sourceforge.net/index.shtml>) was used to map small RNA reads to the transgene sequences, with the small RNA abundance as reads per five million total genome-matched reads (RP5M). For whole-genome small RNA analyses, the ‘hits normalized abundance’ (HNA) values were calculated by dividing the normalized abundance (in RP5M) for each small RNA read hit, where a hit is defined as the number of locations at which a given sequence perfectly matches the genome. The small RNA sequence data are available from the NCBI Gene Expression Omnibus (GEO) under GEO series accession number GSE47852.

Statistical analysis of GC contents of regions preferentially accumulating siRNAs

We identified 248 tandem repeats that accumulated 24-nt siRNA reads (Figure S9A; Table S1, column I). Within these 248 tandem repeats, the sequences are further classified into two regions, with or without siRNA accumulation, for the GC-content analysis. The GC-content percentage (GC%) based on the genomic sequence is defined as:

$$\text{GC}\% = (\text{G} + \text{C}) / (\text{A} + \text{T} + \text{C} + \text{G}) \times 100\%.$$

For each repeat, we calculated the difference of GC% (Δ GC) between the siRNA-accumulating regions (GC_{siRNA}%) and the non-siRNA-accumulating regions (GC_{non}%), as:

$$\Delta\text{GC} = \text{GC}_{\text{siRNA}}\% - \text{GC}_{\text{non}}\%.$$

We compared the Δ GCs in repeats with the randomized genomic sequences, in which the genomic sequences of each of the 248 tandem repeats are shuffled (http://github.com/wwliao/del_tagc_repeat). The result shows that the Δ GCs in repeats are significantly greater than those in randomized sequences, suggesting that the siRNA-accumulating regions are significantly GC-rich.

Whole-genome bisulfite sequencing

The methylation of endogenous tandem repeats (Figure S10) was extracted from whole-genome bisulfite sequencing data following the standard BS-seq protocol (Schmitz *et al.*, 2013). Approximately 3 µg of *T+S*^{WT} genomic DNA was sonicated to approximately 250 bp before it was ligated to Illumina adaptors, then size-selected, denatured and treated with sodium bisulfite to reveal their methylation status. The BS-seq libraries were sequenced using the Illumina HiSeq 2000 platform, according to the manufacturer's instructions. Sequencing was carried out for up to 100 cycles in paired ends. The reads were aligned to the reference genome (TAIR10) using the modified bisulfite aligner, BS SEEKER (Chen *et al.*, 2010). To produce genome-wide DNA methylation profiles, the methylation level for each covered cytosine in the genome is calculated. Because bisulfite treatment converts unmethylated cytosines (Cs) to thymines (Ts), the methylation level at individual cytosines is estimated as $\#C / (\#C + \#T)$, where $\#C$ represents the number of methylated reads and $\#T$ corresponds to the number of unmethylated reads. The methylation level per cytosine provides an estimate of the percentage of cells containing methylation at this cytosine. The raw reads and the processed data set for the *T+S*^{WT} methylome can be downloaded from NCBI GEO under accession GSE47453.

ACKNOWLEDGEMENTS

Financial support was provided by Academia Sinica and the Austrian Academy of Sciences. T.S. was supported by the Japan Society for the Promotion of Science Postdoctoral Fellowships for Research Abroad. Work in the Meyers Laboratory was supported by NSF award #1051576. We thank the sequencing services provided by the National Center for Genome Medicine of the National Core Facility Program for Biotechnology, National Science Council, Taiwan. We are grateful to Lucia Daxinger for performing the initial northern blot analysis to detect hairpin-derived siRNAs in the 15.5a mutant, Xuemei Chen for *nprp2-3* seeds, Christine Ying for editorial assistance, and Ming-Tsung Louis Wu and Attila Molnar for helpful discussions. The authors have no conflicts of interest to declare.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Analysis of the 1.6x *cis* mutant.

Figure S2. Sequence of the S^{135S} locus, as derived from whole-genome sequencing of the 15.5a ($T+S^{\Delta 355}$) *cis* mutant.

Figure S3. Bisulfite sequence analysis of DNA methylation at the target enhancer in the indicated genotypes.

Figure S4. Effect of an *rd2* mutation on *GFP* silencing and target enhancer methylation.

Figure S5. Small RNA accumulation detected by northern blotting.

Figure S6. siRNA accumulation in the downstream region of the target enhancer.

Figure S7. Sequence of the target enhancer region and peaks of siRNA accumulation.

Figure S8. Accumulation of siRNAs at the 'GF' inverted repeat.

Figure S9. Correlation between Pol IV-dependent 24-nt siRNA accumulation and CG methylation at endogenous tandem repeats.

Figure S10. Correlation between siRNA accumulation, GC content and cytosine methylation at selected endogenous intergenic (IGN) tandem repeats.

Table S1. List of endogenous euchromatic tandem repeats.

Table S2. Primers used in this study.

REFERENCES

- Arteaga-Vazquez, M., Sidorenko, L., Rabanal, F.A., Shrivastava, R., Nobuta, K., Green, P.J., Meyers, B.C. and Chandler, V.L. (2010) RNA-mediated trans-communication can establish paramutation at the *b1* locus in maize. *Proc. Natl Acad. Sci. USA*, **107**, 12986–12991.
- Belele, C.L., Sidorenko, L., Stam, M., Bader, R., Arteaga-Vazquez, M.A. and Chandler, V.L. (2013) Specific tandem repeats are sufficient for paramutation-induced trans-generational silencing. *PLoS Genet.* **9**, e1003773.
- Benfey, P.N., Ren, L. and Chua, N.H. (1989) The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. *EMBO J.* **8**, 2195–2202.
- Chabannes, M. and Iskra-Caruana, M.L. (2013) Endogenous pararetroviruses—a reservoir of virus infection in plants. *Curr. Opin. Virol.* **6**, 615–620.
- Chan, S.W., Zhang, X., Bernatavichute, Y.V. and Jacobsen, S.E. (2006) Two-step recruitment of RNA-directed DNA methylation to tandem repeats. *PLoS Biol.* **4**, e363.
- Chen, P.Y., Cokus, S.J. and Pellegrini, M. (2010) BS Seeker: precise mapping for bisulfite sequencing. *BMC Bioinformatics*, **11**, 203.
- Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M. and Jacobsen, S.E. (2008) Shotgun bisulfite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, **452**, 215–219.
- Dalakouras, A. and Wassenegger, M. (2013) Revisiting RNA-directed DNA methylation. *RNA Biol.* **10**, 453–455.
- Daxinger, L., Kanno, T., Bucher, E., van der Winden, J., Naumann, U., Matzke, A.J. and Matzke, M. (2009) A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *EMBO J.* **28**, 48–57.
- De Paoli, E., Dormantes-Acosta, A., Zhai, J., Accerbi, M., Jeong, D.H., Park, S., Meyers, B.C., Jorgensen, R.A. and Green, P.J. (2009) Distinct extremely abundant siRNAs associated with cosuppression in petunia. *RNA*, **15**, 1965–1970.
- Dunoyer, P., Himber, C. and Voinnet, O. (2005) DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nat. Genet.* **37**, 1356–1360.
- Dunoyer, P., Brosnan, C.A., Schott, G., Wang, Y., Jay, F., Alioua, A., Himber, C. and Voinnet, O. (2010) An endogenous, systemic RNAi pathway in plants. *EMBO J.* **29**, 1699–1712.
- Eamens, A., Vaistij, F.E. and Jones, L. (2008) NRPD1a and NRPD1b are required to maintain post-transcriptional RNA silencing and RNA-directed DNA methylation in Arabidopsis. *Plant J.* **55**, 596–606.
- Eun, C., Lorkovic, Z.J., Naumann, U., Long, Q., Havecker, E.R., Simon, S.A., Meyers, B.C., Matzke, A.J. and Matzke, M. (2011) AGO6 functions in RNA-mediated transcriptional gene silencing in shoot and root meristems in Arabidopsis thaliana. *PLoS One*, **6**, e25730.
- Eun, C., Lorkovic, Z.J., Sasaki, T., Naumann, U., Matzke, A.J.M. and Matzke, M. (2012) Use of forward genetic screens to identify genes required for RNA-Directed DNA methylation in Arabidopsis thaliana. *Cold Spring Harb. Symp. Quant. Biol.* **77**, 195–204.
- Finke, A., Kuhlmann, M. and Mette, M.F. (2012) IDN2 has a role downstream of siRNA formation in RNA-directed DNA methylation. *Epigenetics*, **7**, 950–960.
- Fusaro, A.F., Matthew, L., Smith, N.A. et al. (2006) RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. *EMBO Rep.* **7**, 1168–1175.
- Gregor, W., Mette, M.F., Staginnus, C., Matzke, M. and Matzke, A.J.M. (2004) A distinct endogenous pararetrovirus family in Nicotiana glauca, a diploid progenitor of diploid tobacco. *Plant Physiol.* **134**, 1191–1199.
- Haag, J.R. and Pikaard, C.S. (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* **12**, 483–492.
- He, X.J., Chen, T. and Zhu, J.K. (2011) Regulation and function of DNA methylation in plants and animals. *Cell Res.* **21**, 442–465.
- Henderson, I.R. and Jacobsen, S.E. (2008) Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev.* **22**, 1597–1606.
- Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J. and Jacobsen, S.E. (2006) Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.* **38**, 721–726.
- Huetzel, B., Kanno, T., Daxinger, L., Aufsatz, W., Matzke, A.J. and Matzke, M. (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *EMBO J.* **25**, 2828–2836.
- Kanno, T., Bucher, E., Daxinger, L., Huetzel, B., Böhmendorfer, G., Gregor, W., Kreil, D.P., Matzke, M. and Matzke, A.J. (2008) A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat. Genet.* **40**, 670–675.
- Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220.
- 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.
- Lee, T.F., Gurazada, S.G., Zhai, J., Li, S., Simon, S.A., Matzke, M.A., Chen, X. and Meyers, B.C. (2012) RNA polymerase V-dependent small RNAs in Arabidopsis originate from small, intergenic loci including most SIN repeats. *Epigenetics*, **7**, 781–795.
- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. and Ecker, J.R. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, **133**, 523–536.
- Martienssen, R.A. (2003) Maintenance of heterochromatin by RNA interference of tandem repeats. *Nat. Genet.* **35**, 213–214.
- Matzke, M.A. and Mosher, R.A. (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408.
- Melnik, C.W., Molnar, A., Bassett, A. and Baulcombe, D.C. (2011) Mobile 24 nt small RNAs direct transcriptional gene silencing in the root meristems of Arabidopsis thaliana. *Curr. Biol.* **21**, 1678–1683.
- Molnar, A., Melnyk, C.W., Bassett, A., Hardcastle, T.J., Dunn, R. and Baulcombe, D.C. (2010) Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. *Science*, **328**, 872–875.
- Nuthikattu, S., McCue, A.D., Panda, K., Fultz, D., DeFraia, C., Thomas, E.N. and Slotkin, R.K. (2013) The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21–22 nucleotide small interfering RNAs. *Plant Physiol.* **162**, 116–131.
- 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.

- Sasaki, T., Naumann, U., Forai, P., Matzke, A.J.M. and Matzke, M. (2012) Unusual case of apparent hypermutation in *Arabidopsis thaliana*. *Genetics*, **192**, 1271–1280.
- Saze, H. and Kakutani, T. (2011) Differentiation of epigenetic modifications between transposons and genes. *Curr. Opin. Plant Biol.* **14**, 81–87.
- Schmitz, R.J., Schultz, M.D., Urich, M.A. et al. (2013) Patterns of population epigenomic diversity. *Nature*, **495**, 193–198.
- Sidorenko, L., Dorweiler, J.E., Cigan, A.M., Arteaga-Vazquez, M., Vyas, M., Kermicle, J., Jurcin, D., Brzeski, J., Cai, Y. and Chandler, V.L. (2009) A dominant mutation in mediator of paramutation2, one of three-second-largest subunits of a plant-specific RNA polymerase, disrupts multiple siRNA silencing processes. *PLoS Genet.* **5**, e10000725.
- 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 silencing signal between cells in *Arabidopsis*. *Plant Cell*, **19**, 1507–1521.
- Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M. and Peeters, A.J. (2000) The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell*, **6**, 791–802.
- Stam, M., Viterbo, A., Mol, J.N.M. and Kooter, J.M. (1998) Position-dependent methylation and transcriptional silencing of transgenes in inverted T-DNA repeats: implications for posttranscriptional silencing of homologous host genes in plants. *Mol. Cell. Biol.* **18**, 6165–6177.
- Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V. and Jacobsen, S.E. (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell*, **152**, 352–364.
- 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.
- Wierzbicki, A.T. (2012) The role of long non-coding RNA in transcriptional gene silencing. *Curr. Opin. Plant Biol.* **15**, 517–522.
- Wierzbicki, A.T., Cocklin, R., Mayampurath, A., Lister, R., Rowley, M.J., Gregory, B.D., Ecker, J.R., Tang, H. and Pikaard, C.S. (2012) Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. *Genes Dev.* **26**, 1825–1836.
- 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.
- Zhang, X., Henderson, I.R., Lu, C., Green, P.J. and Jacobsen, S.E. (2007) Role of RNA polymerase IV in plant small RNA metabolism. *Proc. Natl Acad. Sci. USA*, **104**, 4536–4541.
- Zhang, H., Ma, Z.Y., Zeng, L. et al. (2013) DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. *Proc. Natl Acad. Sci. USA*, **110**, 8290–8295.
- Zhang, Y., Yan, C. and Kuang, H. (2014) GC content fluctuation around plant small RNA-generating sites. *FEBS Lett.* **588**, 764–769.
- 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.
- Zhong, X., Hale, C.J., Law, J.A., Johnson, L.M., Feng, S., Tu, A. and Jacobsen, S.E. (2012) DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat. Struct. Mol. Biol.* **19**, 870–876.