



University of Pennsylvania  
ScholarlyCommons

Departmental Papers (Biology)

Department of Biology

8-15-2012

# Spatial and Functional Relationships Among Pol V-Associated Loci, Pol IV-Dependent siRNAs, and Cytosine Methylation in the *Arabidopsis* Epigenome

Andrzej T. Wierzbicki  
*University of Michigan-Ann Arbor*


Ross Cocklin  
*Indiana University - Bloomington*

Anoop Mayampurath  
*Indiana University - Bloomington*

Ryan Lister  
*Salk Institute*

M. J. Rowley  
*University of Michigan-Ann Arbor*

Follow up on additional works at: [http://repository.upenn.edu/biology\\_papers](http://repository.upenn.edu/biology_papers)

 Part of the [Biology Commons](#), [Genetics and Genomics Commons](#), and the [Nucleic Acids, Nucleotides, and Nucleosides Commons](#)

## Recommended Citation

Wierzbicki, A. T., Cocklin, R., Mayampurath, A., Lister, R., Rowley, M. J., Gregory, B. D., Ecker, J. R., Tang, H., & 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 & Development*, 26 (16), 1825-1836. <http://dx.doi.org/10.1101/gad.197772.112>

At the time of this publication, Dr. Gregory was affiliated with the Salk Institute, but he is now a faculty member of the University of Pennsylvania.

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/biology\\_papers/29](http://repository.upenn.edu/biology_papers/29)

For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# Spatial and Functional Relationships Among Pol V-Associated Loci, Pol IV-Dependent siRNAs, and Cytosine Methylation in the *Arabidopsis* Epigenome

## Abstract

Multisubunit RNA polymerases IV and V (Pols IV and V) mediate RNA-directed DNA methylation and transcriptional silencing of retrotransposons and heterochromatic repeats in plants. We identified genomic sites of Pol V occupancy in parallel with siRNA deep sequencing and methylcytosine mapping, comparing wild-type plants with mutants defective for Pol IV, Pol V, or both Pols IV and V. Approximately 60% of Pol V-associated regions encompass regions of 24-nucleotide (nt) siRNA complementarity and cytosine methylation, consistent with cytosine methylation being guided by base-pairing of Pol IV-dependent siRNAs with Pol V transcripts. However, 27% of Pol V peaks do not overlap sites of 24-nt siRNA biogenesis or cytosine methylation, indicating that Pol V alone does not specify sites of cytosine methylation. Surprisingly, the number of methylated CHH motifs, a hallmark of RNA-directed de novo methylation, is similar in wild-type plants and Pol IV or Pol V mutants. In the mutants, methylation is lost at 50%–60% of the CHH sites that are methylated in the wild type but is gained at new CHH positions, primarily in pericentromeric regions. These results indicate that Pol IV and Pol V are not required for cytosine methyltransferase activity but shape the epigenome by guiding CHH methylation to specific genomic sites.

## Keywords

DNA-dependent RNA polymerase, gene silencing, DNA methylation, epigenetics, short interfering RNA, RNA-directed DNA methylation

## Disciplines

Biology | Genetics and Genomics | Nucleic Acids, Nucleotides, and Nucleosides

## Comments

At the time of this publication, Dr. Gregory was affiliated with the Salk Institute, but he is now a faculty member of the University of Pennsylvania.

## Author(s)

Andrzej T. Wierzbicki, Ross Cocklin, Anoop Mayampurath, Ryan Lister, M. J. Rowley, Brian D. Gregory, Joseph R. Ecker, Haixu Tang, and Craig S. Pikaard

# Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the *Arabidopsis* epigenome

Andrzej T. Wierzbicki,<sup>1,8</sup> Ross Cocklin,<sup>2,3,8</sup> Anoop Mayampurath,<sup>4,8</sup> Ryan Lister,<sup>5,9</sup> M. Jordan Rowley,<sup>1</sup> Brian D. Gregory,<sup>5,10</sup> Joseph R. Ecker,<sup>5,6</sup> Haixu Tang,<sup>4</sup> and Craig S. Pikaard<sup>2,3,7,11</sup>

<sup>1</sup>Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48109, USA; <sup>2</sup>Department of Biology, <sup>3</sup>Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, Indiana 47405, USA; <sup>4</sup>School of Informatics and Computing, Indiana University, Bloomington, Indiana 47405, USA; <sup>5</sup>Salk Institute, La Jolla, California 92037, USA; <sup>6</sup>Howard Hughes Medical Institute, Salk Institute, La Jolla, California 92037, USA; <sup>7</sup>Howard Hughes Medical Institute, Indiana University, Bloomington, Indiana 47405, USA

**Multisubunit RNA polymerases IV and V (Pols IV and V) mediate RNA-directed DNA methylation and transcriptional silencing of retrotransposons and heterochromatic repeats in plants. We identified genomic sites of Pol V occupancy in parallel with siRNA deep sequencing and methylcytosine mapping, comparing wild-type plants with mutants defective for Pol IV, Pol V, or both Pols IV and V. Approximately 60% of Pol V-associated regions encompass regions of 24-nucleotide (nt) siRNA complementarity and cytosine methylation, consistent with cytosine methylation being guided by base-pairing of Pol IV-dependent siRNAs with Pol V transcripts. However, 27% of Pol V peaks do not overlap sites of 24-nt siRNA biogenesis or cytosine methylation, indicating that Pol V alone does not specify sites of cytosine methylation. Surprisingly, the number of methylated CHH motifs, a hallmark of RNA-directed de novo methylation, is similar in wild-type plants and Pol IV or Pol V mutants. In the mutants, methylation is lost at 50%–60% of the CHH sites that are methylated in the wild type but is gained at new CHH positions, primarily in pericentromeric regions. These results indicate that Pol IV and Pol V are not required for cytosine methyltransferase activity but shape the epigenome by guiding CHH methylation to specific genomic sites.**

[**Keywords:** DNA-dependent RNA polymerase; gene silencing; DNA methylation; epigenetics; short interfering RNA; RNA-directed DNA methylation]

Supplemental material is available for this article.

Received June 2, 2012; revised version accepted July 2, 2012.

RNA polymerases IV and V (Pols IV and V) are plant-specific, specialized forms of Pol II that evolved through duplication and subfunctionalization of Pol II subunit genes (Luo and Hall 2007; Huang et al. 2009; Ream et al. 2009; Tucker et al. 2010). Unlike Pols I, II, or III, Pols IV and V are not essential for viability but play important roles in transposon taming, defense against pathogens and invading nucleic acids, and RNA-mediated interalle-

lic communications involved in paramutation (Haag and Pikaard 2011).

Pol IV and Pol V functions are best understood with respect to their roles in the RNA-directed DNA methylation pathway (Herr et al. 2005; Kanno et al. 2005b; Onodera et al. 2005; Pontier et al. 2005), in which 24-nucleotide (nt) siRNAs specify sites of de novo cytosine methylation by DRM2 (DOMAINS REARRANGED METHYLTRANSFERASE 2). DRM2 methylates cytosines in all sequence contexts: CG, CHG, or CHH (where H is a base other than G) (Cao and Jacobsen 2002). Methylation at CG and CHG motifs, which are symmetrical with respect to the anti-parallel strands of duplex DNA, can then be maintained by the MET1 and CMT3 cytosine methyltransferases, which modify newly replicated daughter strands to match the methylation pattern of the parent strand (Finnegan et al. 1996; Ronemus et al.

<sup>8</sup>These authors contributed equally to this work.

Present addresses: <sup>9</sup>The University of Western Australia, Perth 6009, Australia; <sup>10</sup>University of Pennsylvania, Philadelphia, PA 19104, USA.

<sup>11</sup>Corresponding author

E-mail cpikaard@indiana.edu

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.197772.112>. Freely available online through the *Genes & Development* Open Access option.

1996; Lindroth et al. 2001). However, nonsymmetrical CHH methylation is thought to require de novo methylation by DRM2 in each cell cycle. In the current model for the RNA-directed DNA methylation pathway (Fig. 1), Pol IV generates transcripts that are made double-stranded through the action of an RNA-dependent RNA polymerase, RDR2 (RNA-DEPENDENT RNA POLYMERASE 2) (Xie et al. 2004). The resulting dsRNA precursors are diced by DICER-LIKE 3 (DCL3) and loaded into ARGONAUTE 4 (AGO4) or closely related argonaute proteins (Qi et al. 2005, 2006; Zheng et al. 2007; Olmedo-Monfil et al. 2010). Independent of Pol IV transcription and siRNA biogenesis, Pol V generates transcripts at target loci with assistance from DRD1 (Kanno et al. 2005a; Wierzbicki et al. 2008), a putative chromatin remodeling ATPase; DMS3 (Kanno et al. 2008; Wierzbicki et al. 2009), a protein related to the hinge domain of cohesins and condensins; and RDM1, an ssDNA-binding protein with a preference for methylated DNA (Gao et al. 2010; Law et al. 2010). These latter three activities associate within the so-called DDR complex, named for the first initials of the component proteins (Law et al. 2010). AGO4 can be cross-linked to Pol V transcripts, suggesting that Pol V noncoding RNAs serve as “scaffolds” to which AGO4–siRNA complexes bind through RNA–RNA base-pairing (Wierzbicki et al. 2009). AGO4 can also interact with the C-terminal domain of the Pol V largest subunit, presumably stabilizing the complex through protein–protein interactions (Li et al. 2006; El-Shami et al. 2007). Subsequent recruitment of DRM2 and other chromatin-modifying machinery may be mediated by RDM1, which apparently has functions apart from the DDR complex based on its ability to physically interact with AGO4 and DRM2, which have not been detected in the DDR complex (Gao et al. 2010). RDM1 may thus be a key bridging protein that recruits DRM2 to AGO4–siRNA–Pol V transcript complexes, thereby facilitating the de novo cytosine methylation of Pol V transcribed loci.

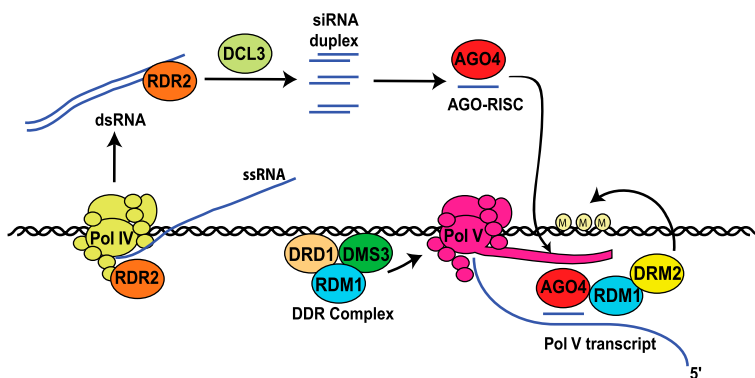
A prediction of the model above, which is based on a relatively small number of loci, is that de novo cytosine methylation occurs at sites where Pol IV-dependent 24-nt siRNAs coincide with Pol V transcripts (Wierzbicki et al. 2008). To test this model genome-wide, we conducted Pol V ChIP-seq (chromatin immunoprecipitation [ChIP] followed by deep sequencing) to identify sites of Pol V occupancy and also carried out genome-wide bisulfite

mapping of methylcytosine positions in parallel with small RNA deep sequencing (RNA-seq) in wild-type, Pol IV mutant, Pol V mutant, or Pol IV/V double-mutant plants. Collectively, these data sets allow the degree of congruence between sites of Pol V transcription, small RNA biogenesis, and cytosine methylation to be evaluated. We show that the majority of Pol V-associated loci overlap sites of 24-nt siRNA biogenesis and cytosine methylation, in agreement with the current model. However, approximately one-quarter of Pol V peaks occur at loci (mostly genes) lacking significant cytosine methylation or siRNA signatures. Among the loci associated with Pol V are genes encoding proline- and glycine-rich proteins enriched for a consensus sequence, CCNCCNCC NCC (or its complement, GGNGGNGGNGG), which is detected at least once within 50% of all Pol V peaks. However, this sequence alone is not necessary or sufficient to explain genomic sites of Pol V occupancy. Interestingly, we found that the occurrence of methylated CHH sites in the genome is not abolished in Pol IV or Pol V mutants or their double mutant but is shifted to new positions concentrated in the pericentromeric regions. These results suggest that the occurrence of CHH methylation does not require Pol IV or Pol V activity but is not guided properly, thereby disrupting silencing of transposons and other elements subjected to RNA-directed DNA methylation.

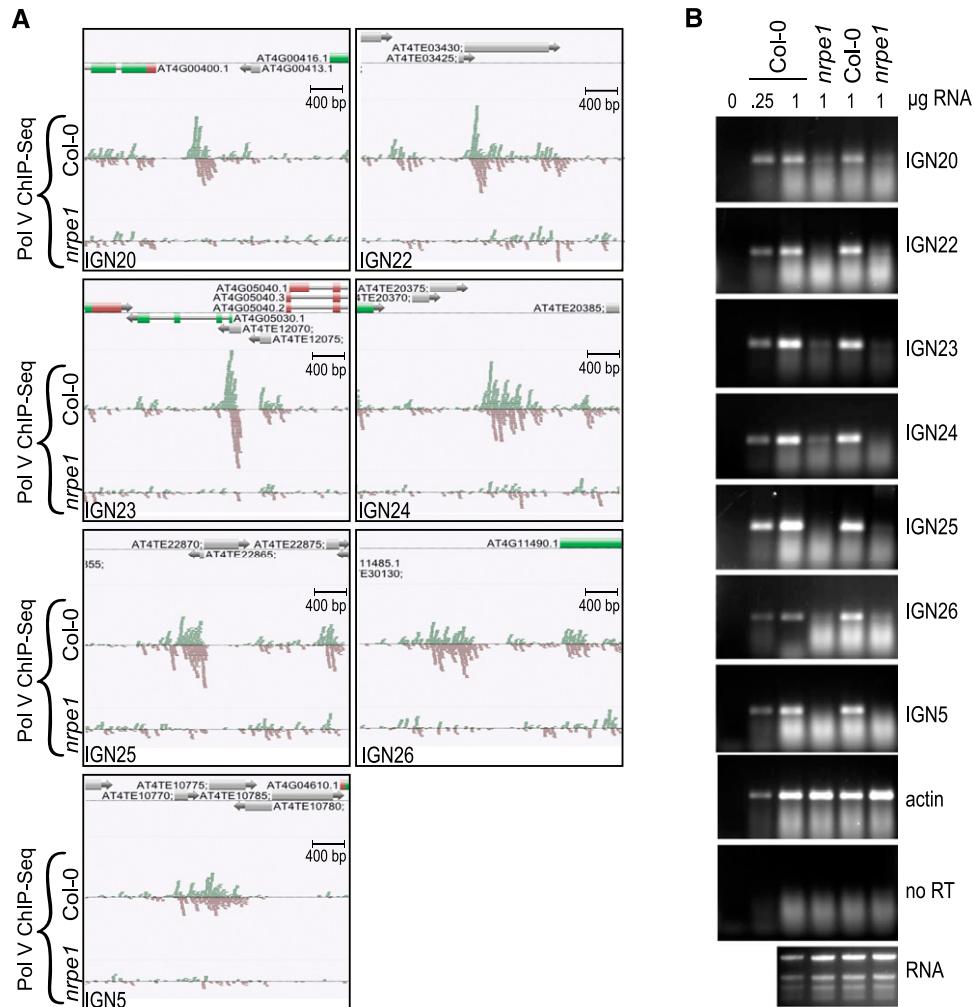
## Results

### Pol V ChIP-seq

To identify genomic loci associated with Pol V, we conducted ChIP-seq using an antibody that recognizes a C-terminal domain of the Pol V largest subunit NRPE1; similar sequences are not present in RNA Pol I, II, III, or IV. ChIP-seq was conducted using both wild-type (ecotype Col-0) plants and an *nrpe1*-null mutant (*nrpe1-11*) that lacks Pol V. By comparing these data sets, we identified high-confidence regions of Pol V occupancy (Pol V peaks), defined as positions where signal to noise ratios (i.e., signals in wild type vs. *nrpe1* mutant) yielded an estimated false discovery rate no higher than 4%. According to this criterion, 1157 high-confidence Pol V peaks were identified (Supplemental Table S1), including previously identified Pol V transcribed loci, such as IG5 (Fig. 2A).



**Figure 1.** A model for RNA-directed DNA methylation in *Arabidopsis thaliana*. RNA Pol IV transcripts are used as templates by the Pol IV-interacting protein RNA-DEPENDENT RNA POLYMERASE 2 (RDR2). DICER-LIKE 3 (DCL3) cleaves resulting dsRNAs into 24-nt siRNA products, one strand of which is loaded into an ARGONAUTE 4 (AGO4) RISC complex. Independent of siRNA biogenesis, the DDR complex enables transcription by RNA Pol V, whose nascent transcripts serve as scaffolds for the binding of AGO–RISC complexes. AGO4 also interacts with the C-terminal domain of the Pol V largest subunit and RDM1. In turn, RDM1 interacts with the de novo DNA methyltransferase DRM2.



**Figure 2.** Pol V ChIP-seq and evidence for Pol V transcription at associated loci. (A) Browser views for a subset of genomic regions associated with Pol V, identified by ChIP-seq. Individual reads corresponding to the two DNA strands are represented by green or red horizontal bars. Annotated genes or transposable elements (TEs) are shown. (B) RT-PCR detection of Pol V-dependent transcripts at the Pol V peak loci shown in A.

Pol V peaks are similarly distributed among the five nuclear chromosomes (281 peaks on chromosome I, 215 on chromosome II, 236 on chromosome III, 171 on chromosome IV, and 254 on chromosome V) and range in size from 414 base pairs (bp) to 3175 bp, with an average peak spanning 1165 bp (Supplemental Table S1). No Pol V peaks were detected in the chloroplast or mitochondrial genomes, consistent with Pol V's nuclear localization (Onodera et al. 2005; Pontier et al. 2005; Li et al. 2006; Pontes et al. 2006).

The validity of Pol V peaks identified by ChIP-seq was tested for a subset of representative loci using RT-PCR and ChIP-qPCR. Figure 2A shows genome browser displays of ChIP-seq results for seven Pol V peaks, including the known Pol V transcribed locus IGN5 (Wierzbicki et al. 2008). At each of these Pol V peak regions, transcripts detected by RT-PCR in wild-type (Col-0) plants are lost or substantially reduced in *nrpe1* mutants (Fig. 2B), consistent with Pol V transcription of these loci. Quantitative ChIP-PCR assays also confirm that Pol V associates with

these loci as well as with the well-characterized Pol V-regulated transposon loci *AtSN1* and *soloLTR* (Supplemental Fig. S1).

#### *Spatial relationships among Pol V peaks, cytosine methylation, and small RNAs*

In parallel with Pol V ChIP-seq, we mapped genome-wide positions of cytosine methylation and conducted deep sequencing of small RNAs in wild type (Col-0); Pol IV mutants (*nrpd1*); Pol V (*nrpe1*) mutants; or double mutants resulting from disruption of the gene encoding the second subunit of both Pols IV and V, *nrp(d/e)2*.

Cytosine methylation mapping was conducted using inflorescence DNA and sodium bisulfite treatment to convert unmethylated cytosines to uracil. Comparison of Illumina sequencing data for bisulfite-treated and untreated DNA allows the positions and frequency of methylcytosines to be determined genome-wide. The bisulfite

conversion efficiency, assessed by the number of converted versus unconverted cytosines for the unmethylated chloroplast genome, was between 99.10% and 99.43% for all samples (Supplemental Table S2). Sequence reads were filtered so as to include only those reads that uniquely mapped to the *Arabidopsis thaliana* Col-0 ecotype (strain) reference genome and to eliminate artifacts arising from clonal duplication during library preparation and PCR amplification, as previously described (Lister et al. 2008). For all libraries, the nuclear genome was represented by ~22 million unique, nonclonal reads with an average read length of 48 bases and an average coverage of five reads per base (Supplemental Table S3). This analysis identified 2,410,770 unique methylated cytosines in the nuclear genome of Col-0, representing 5.6% of all nuclear cytosines, consistent with prior studies (Lister et al. 2008). Of these methylated cytosines, 55% are in the CG context, 27% are in the CHG context, and 18% are in the CHH context, where H represents A, C, or T (Supplemental Table S4). The number of methylated cytosines is only slightly reduced in the mutants relative to Col-0, with 2,313,395 methylcytosines mapped in *nripd1-3*, 2,206,736 mapped in *nripe1-11*, and 2,166,470 mapped in *nrip(d/e)2*.

Small RNAs were identified by Illumina sequencing using libraries of size-fractionated 15- to 30-nt small RNAs purified from inflorescence tissues. Consistent with previous reports (Zhang et al. 2007; Mosher et al. 2008), Pol IV is required for the production of >90% of all 24-nt siRNAs, and Pol V reinforces or amplifies siRNA levels at ~50% of siRNA producing loci (Supplemental Tables S5–S8).

Coordinated display of the Pol V ChIP-seq, methylcytosine mapping, and siRNA-seq data sets using the AnnoJ genome browser platform allows relationships between sites of Pol V occupancy, cytosine methylation, and small RNAs to be visualized ([http://neomorph.salk.edu/pol\\_epigenomes/browser.html](http://neomorph.salk.edu/pol_epigenomes/browser.html)). A representative browser display is shown in Figure 3A for the RathE2 SINE retrotransposon locus At2TE82000. Transcription of this element is suppressed in wild-type plants (Col-0), but silencing is lost in the *nripe1* Pol V mutant (Fig. 3B, cf. first and second lanes). A transgene expressing NRPE1 fused to a C-terminal Flag tag restores silencing of the transposon in the *nripe1* mutant background (Fig. 3B, third lane). As can be seen in the browser display, a region of high methylcytosine density coincides with a hot spot of 24-nt siRNA production, and reductions in these siRNAs in the *nripd1*, *nripe1*, and *nrip(d/e)2* mutants correlates with decreased cytosine methylation at the locus, primarily in CHH and CHG contexts.

For locus At2TE82000 (Fig. 3A), the cytosine methylation and small RNA hot spots are centrally located within the Pol V peak. This is generally the case genome-wide (Fig. 3C). To generate the Figure 3C histogram, all Pol V peaks that associate with small RNAs and cytosine methylation were subdivided into five bins, each comprising 20% of the Pol V peak region. The number of small RNAs and methylcytosines within each bin was then tallied. This analysis reveals that small RNAs and methylcytosines occur most frequently near the middle

of Pol V peaks (bin 3) and are least abundant at the edges of the Pol V peaks (bins 1 and 5).

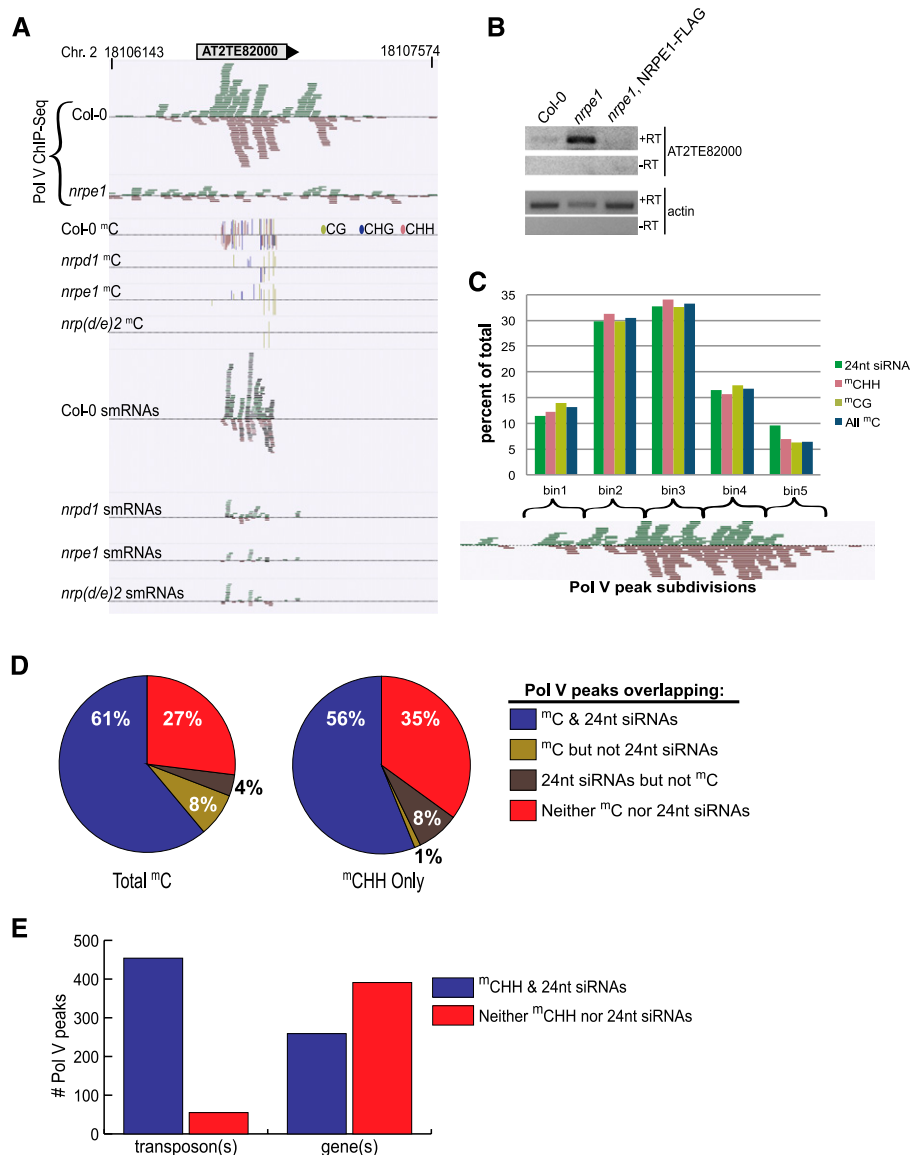
Most, but not all, Pol V peaks overlap sites of siRNA biogenesis and cytosine methylation. A focus on CHH methylation, the hallmark of de novo cytosine methylation, shows that 56% of all Pol V peaks overlap sites of both CHH methylation and 24-nt siRNA complementarity (Fig. 3D). Only 1% of Pol V peaks (12 peaks) overlap sites of CHH methylation but not 24-nt siRNAs. Collectively, these results fit the model's prediction that CHH methylation occurs where Pol V transcripts and 24-nt siRNAs coincide (see Fig. 1).

At 8% of Pol V peaks, 24-nt siRNAs are generated, but CHH methylation is not found (Fig. 3D). An example is the IGN23 locus overlapping the AT4TE12070 transposon and AT4G05030 gene. Small RNAs uniquely mapping at low abundance to this Pol V peak region in wild-type plants (six 24-nt, four 23-nt, and one 21-nt small RNAs) are not detected in Pol IV or Pol V mutants. All methylation at IGN23 is in the CG and CHG contexts and persists in Pol IV or Pol V mutants (Supplemental Fig. S2).

At 35% of the Pol V peaks, neither CHH methylation nor 24-nt siRNAs are detected. This fraction decreases to 27% if total methylcytosine (CG, CHG, and CHH) is considered, and not just CHH methylation (Fig. 3D, cf. the two pie charts). This decrease is due primarily to Pol V's association with CG or CHG methylated sites, but not CHH sites, at 8% of the Pol V peaks, including IGN23 discussed above. Other examples include transposons and pericentromeric repeats that are silenced by Pol V in collaboration with MET1 (the CG maintenance methyltransferase) and the SWI2/SNF2-like chromatin remodeling ATPases DDM1 and DRD1, but independent of Pol IV, RDR2, DCL3, AGO4, or DRM2 (Pontes et al. 2009).

#### Sequence motifs at Pol V-associated loci

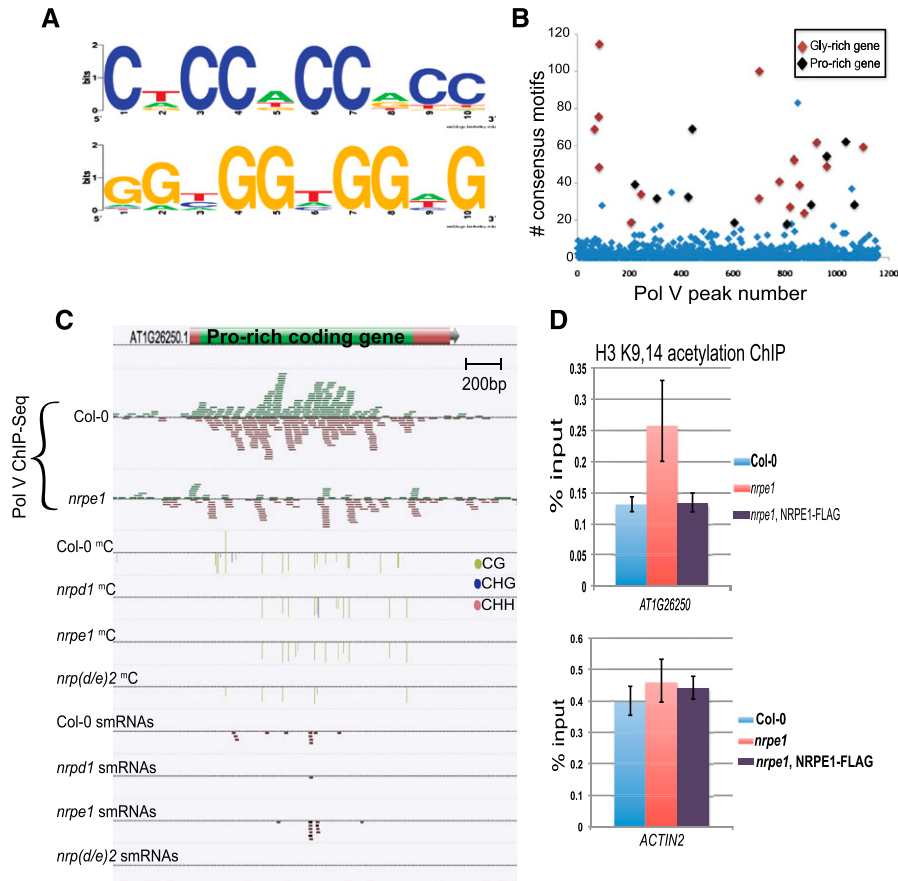
We examined annotated elements that lie within or overlap Pol V peaks to estimate Pol V's prevalence for associating with transposons or genes. Among the 1157 high-confidence Pol V peaks, 565 include or overlap at least one transposon (annotated in Supplemental Table S1). Of these, 455 (81%) are associated with CHH methylation and 24-nt siRNAs (Fig. 3E), both of which are hallmarks of RNA-directed DNA methylation in facultative heterochromatin. Pol V peaks also overlap genes in 724 cases (annotated in Supplemental Table S1), the majority of which (54%; 392 cases) lack associated CHH methylation or complementarity to 24-nt siRNAs (Fig. 3E). However, 260 (36%) of the gene-associated Pol V peaks are associated with CHH methylation and 24-nt siRNAs, primarily due to the fact that in 245 cases, the Pol V peaks also overlap at least one transposon, as in the case of the AT4TE12070 transposon and AT4G05030 gene discussed above (Supplemental Fig. S2). Another example is the AT3G51820 gene that has the AT3TE77985 transposable element inserted in its promoter region such that Pol IV- and Pol V-dependent siRNA biogenesis and CHH methylation occurs in this region.



**Figure 3.** Correlations among Pol V peaks, cytosine methylation, and 24-nt siRNAs. (A) Example of Pol V ChIP-seq, DNA methylation, and small RNA data tracks displayed in the AnnoJ browser for the region including the SINE element *AT2TE82000*. (B) RT-PCR analysis of *AT2TE82000* expression in wild type or an *nrpe1* mutant or *nrpe1* mutant rescued by an *NRPE1* transgene. (C) Genomic Pol V peaks were subdivided into five bins, each corresponding to 20% of the peak. The numbers of methylcytosines (in different contexts) or 24-nt siRNAs within each bin are plotted. (D) Pie chart depicting the relationships among 24-nt siRNAs, methylcytosines in CHH motifs, or total methylcytosines within Pol V peaks. (E) The number of Pol V peaks overlapping transposons or genes among the blue or red subclasses of Pol V peaks shown in D (for CHH methylation) is shown.

We searched for sequence motifs that might be enriched within Pol V peaks using the motif-finding tools BioProspector (Liu et al. 2001) and BioOptimizer (Jensen and Liu 2004). Two 10-nt motifs that are complementary to one another were identified as top-scoring hits using BioProspector: C(T/A)CC(A/T)CC(A/T)CC and GG(T/C)GG(T/A)GG(A/T)G (Fig 4A). Using this initial motif, BioOptimizer identified a 1-nt-longer motif, CC(T/A)CC(A/T)CC(A/T)CC and GG(T/C)GG(T/A)GG(A/T)GG, as the best-scoring motif among all optimizations. The 11-bp position weight matrix of this motif was used for motif pattern searching using PATSER. Fifty percent

(578) of the Pol V peaks have at least one consensus motif, and 27 Pol V peaks include 20 or more motif copies. Twenty-three of these 27 loci correspond to genes encoding glycine-rich or proline-rich proteins (Fig. 4B), which fits with the CCN/GGN repeats of the consensus motif being codons for proline or glycine, respectively. CHH methylation is not detected at most of these proline- or glycine-rich genes; however, low-abundance Pol IV-dependent 24-nt siRNAs are detected at some of these genes, as at the proline-rich protein gene *AT1G26250* (Fig. 4C). Twenty-two small RNAs perfectly match the *At1g26250* sequence; 11 of these small RNAs uniquely match *At1g26250*,



**Figure 4.** Motifs enriched at Pol V peak regions. (A) Complementary sequences that were identified using BioProspector as being overrepresented at Pol V-associated loci. (B) Number of consensus repeats (vertical axis) within the 1157 Pol V peaks (horizontal axis). Peaks overlapping genes encoding proline- or glycine-rich proteins are colored black or rust, respectively. (C) AnnoJ browser tracks of the region surrounding *AT1G26250*, a gene encoding a proline-rich protein with 115 copies of the CNCCNCCNCC motif. (D) Histone H3 Lys 9 and Lys 14 acetylation at the *AT1G26250* and *ACTIN2* genes in Col-0, *nrpe1*, or *nrpe1* complemented with an NRPE1-Flag transgene.

whereas the remainder might also map to other locations in the genome. At the *At1g26250* locus, *nrpe1* mutants have elevated levels of Histone H3 Lys 9 and Lys 14 acetylation, which are marks of an active transcriptional state. Complementation of the *nrpe1* mutant with an NRPE1 transgene under the control of its own promoter returns H3K9 and H3K14 acetylation to wild-type levels (Fig. 4D). In contrast, at a control locus encoding *ACTIN2*, H3K9 and H3K14 acetylation are unaffected in *nrpe1* mutants. Thus, Pol V may play some role in affecting histone modifications independent of DNA methylation, which is unchanged in *nrpd1*, *nrpe1*, or *nrp(d/e)2* mutants relative to wild type at *AT1G26250* (Fig. 4C).

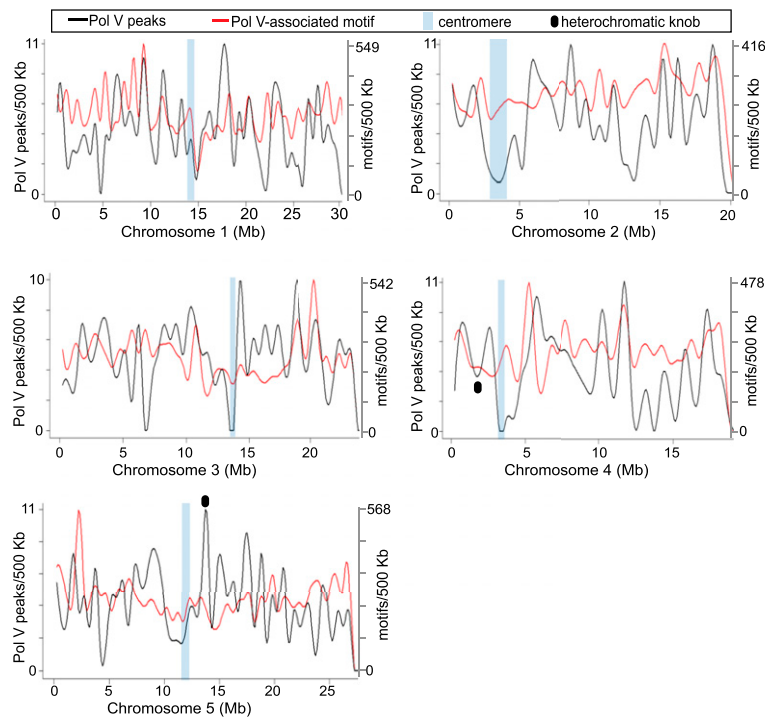
In addition to motif finding, we performed gene ontology (GO) analysis for genes associated with Pol V peaks. Only two GO terms were significantly enriched; namely, genes encoding “structural constituent of cell wall” and “protein homo-oligomerization” proteins. In total, 25 Pol V-associated genes fell into these two groups. Fifteen encode proline-rich proteins that are a subset of the “structural constituent of cell wall” GO category (89 genes), and 10 encode glycine-rich proteins that are a subset of the “protein homo-oligomerization” GO category with 61

members, (Supplemental Tables S9–S11). The coding regions of Pol V-associated genes in these two GO categories are internally repetitive (see dot plots in Supplemental Fig. S3) and enriched for the proline- or glycine-encoding C(T/A)CC(A/T)CC(A/T)CC or GG(T/C)GG(T/A)GG(A/T)G motifs. Importantly, “structural constituent of cell wall” or “protein homo-oligomerization” proteins are diverse, and many are not proline- or glycine-rich or Pol V-associated.

To visualize the spatial distribution of Pol V peaks and consensus motifs on the five chromosomes, we plotted the frequency of their occurrence within a 500-kb sliding window along each chromosome (Fig. 5). Pol V peaks are depleted at the presumptive centromere regions (depicted by the blue vertical bars in Fig. 5) but are otherwise distributed throughout the chromosomes, averaging approximately four high-confidence Pol V peaks per 500 kb but ranging from zero to 11 peaks per 500 kb. Pol V is enriched at the heterochromatic knob of chromosome 5 (denoted by a black oval in Fig. 5) but is not similarly enriched at the knob on chromosome 4 (also denoted by a black oval in Fig. 5).

The C(T/A)CC(A/T)CC(A/T)CC motifs are present throughout the chromosomes, with only occasional concordance with sites of Pol V enrichment. Importantly, the





**Figure 5.** Genome-wide distribution of Pol V and a Pol V-associated motif. The number of Pol V (black line) and Pol V-associated motifs (red line) within a 500-kb fixed window is plotted for each chromosome. Data were normalized relative to the highest value position for each chromosome and smoothed using spline interpolation. Light-blue shading indicates the approximate position of each centromere. Black circles mark the approximate position of heterochromatic knobs on chromosomes 4 and 5.

consensus motif is present 65,071 times in the genome, outnumbering high-confidence Pol V peaks  $\sim 60$ -fold. Moreover, 50% of all high-confidence Pol V peaks lack the motif. Therefore, despite being the only motif significantly enriched among the Pol V peaks, we conclude that the motif is not necessary or sufficient to explain Pol V recruitment.

#### *Effects of Pol IV and Pol V mutations on siRNAs and methylcytosine positions*

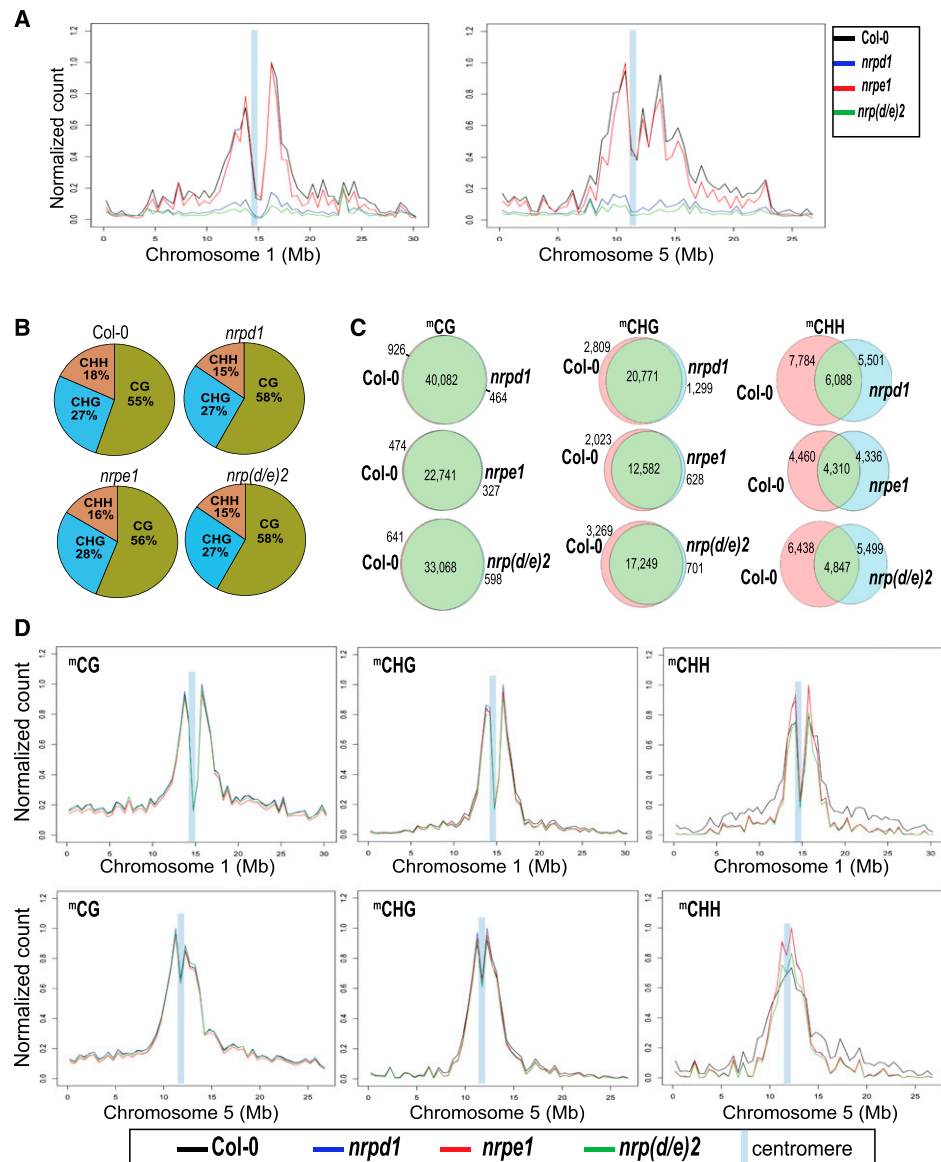
Small RNAs and DNA methylation in all sequence contexts are most abundant in the pericentromeric regions of *Arabidopsis* chromosomes (Lu et al. 2005; Zhang et al. 2006). Our sequencing analyses of siRNAs in wild-type, Pol IV mutant [*nripd1*], Pol V mutant [*nripe1*], and Pol IV/V double-mutant [*nrip(d/e)2*] plants are consistent with these prior studies, showing that siRNAs corresponding to pericentromeric regions are the most abundant and are dramatically lost in these regions in *nripd1* and *nrip(d/e)2* mutants (Fig. 6A). In contrast, siRNA abundance in Pol V mutants (*nripe1*) resembles wild type, although siRNA abundance is somewhat reduced, especially in the chromosome arms beyond the pericentromeric regions. In *nrip(d/e)2* mutants, which are defective for both Pols IV and V, chromosomal profiles of 24-nt siRNA abundance resemble those of *nripd1* mutants, attributable to the absence of Pol IV activity.

Analysis of bisulfite-mediated genomic sequencing data reveals that the sequence context in which methylated cytosines occur is similar in wild type and *nripd1*, *nripe1*, and *nrip(d/e)2* mutants (Fig. 6B; Supplemental Table S4). Because bisulfite conversion of unmethylated cytosines to uracils alters the GC content of DNA fragments,

thereby affecting the read depth achieved by Illumina sequencing, we considered for these analyses only those cytosines covered by at least six reads but no more than 10 reads on both DNA strands. This filtering eliminates bias arising from unusually high or low sequencing depth of any particular cytosine.

Interestingly, the proportion of cytosine methylation in CG, CHG, or CHH contexts genome-wide is nearly identical in wild-type (Col-0), *nripd1* (Pol IV), *nripe1* (Pol V), or *nrip(d/e)2* (double mutant) plants, with 55%–58% of all methylcytosines occurring in the CG context, 27%–28% occurring in the CHG context, and 15%–18% occurring in the CHH context (Fig. 6B). In fact, 98% of the methylated CG dinucleotide positions in wild type were unchanged in *nripd1*, *nripe1*, and *nrip(d/e)2* mutants, as shown in the Venn diagrams of Figure 6C. For instance, comparison of the wild-type and *nripd1* data sets showed that 926 methylcytosines in CG motifs were lost in the *nripd1* mutant, and 464 were gained, but 40,082 methylcytosine positions were unchanged. Similarly, 84%–88% of methylcytosines in CHG contexts were unaffected by the loss of Pol IV, Pol V, or both Pol IV and Pol V (Fig. 6C). CHH methylation was most affected by the mutations eliminating Pol IV or Pol V, with  $\sim 50\%$ – $55\%$  of CHH methylcytosine positions in wild type losing their methylation in the *nripd1*, *nripe1*, or *nrip(d/e)2* mutants. However, the total number of methylcytosines in the CHH context is not substantially altered in the *nripd1*, *nripe1*, or *nrip(d/e)2* mutants, due to the ectopic occurrence of CHH methylation at thousands of new cytosine positions (Fig. 6C). In total, 47%–53% of the CHH methylcytosines in *nripd1*, *nripe1*, or *nrip(d/e)2* mutants correspond to cytosines that were not methylated in wild-type Col-0.

Wierzbicki et al.



**Figure 6.** Pol IV- and Pol V-dependent small RNAs and DNA methylation. (A) Chromosomal distribution of uniquely mapping small RNAs within a 500-kb sliding window, shown for chromosomes 1 and 5. Wild-type Col-0 and *nrpd1*, *nrpe1*, and *nrp(d/e)2* mutants are compared. Peak heights are normalized relative to the highest level detected for each chromosome. (B) Relative amounts of methylcytosines in CG, CHG, or CHH contexts (where H represents A, C, or T) in wild-type Col-0 and *nrpd1*, *nrpe1*, or *nrp(d/e)2* mutants. (C) Venn diagrams comparing the number and genomic positions of methylated cytosines in CG, CHG, or CHH contexts in wild-type Col-0 and *nrpd1*, *nrpe1*, or *nrp(d/e)2* mutants. (D) Chromosomal distributions of methylated CH, CHG, or CHH sites, shown for chromosomes 1 and 5. Wild-type Col-0 and *nrpd1*, *nrpe1*, and *nrp(d/e)2* mutants are compared using a 500-kb fixed window size. Peak heights are normalized relative to the highest level detected for each chromosome.

To visualize where losses and ectopic gains in cytosine methylation occur, we compared the chromosomal distributions of CG, CHG, and CHH methylation in wild type versus *nrpd1*, *nrpe1*, or *nrp(d/e)2* mutants (Fig. 6D). CG and CHG methylation profiles are essentially unaffected in the Pol IV or Pol V mutants compared with wild type. However, Pol IV and Pol V mutants show increased CHH methylation in the pericentromeric regions and decreased CHH methylation in the chromosome arms (see also Supplemental Fig. S4). Collectively,

the results of Figure 6 indicate that Pol IV and Pol V are not required for the catalytic activities of cytosine methyltransferases acting at CG, CHG, or CHH motifs but are important for guiding CHH methylation to selected sites.

## Discussion

By combining Pol V ChIP-seq data with small RNA-seq and methylcytosine mapping in Pol IV, Pol V, and Pol IV/ V double mutants, several new insights relevant to RNA-

directed DNA methylation and the spatial control of cytosine methylation have emerged. With respect to our current model for siRNA-directed DNA methylation (see Fig. 1), the majority of Pol V-occupied sites in the genome fit with the basic tenets of the model, with Pol V peaks typically spanning regions where CHH methylation and complementarity to Pol IV-dependent 24-nt siRNAs occurs. siRNAs and cytosine methylation are most abundant near the center of the Pol V peaks, consistent with the hypothesis that long Pol V transcripts serve as scaffolds for the recruitment of chromatin-modifying machinery via siRNA–Pol V transcript base-pairing (Wierzbicki et al. 2008, 2009). Previous work has shown that production of Pol V transcripts is not dependent on 24-nt siRNA biogenesis or Pol IV function, suggesting that Pol IV and Pol V generate their transcripts independently (Wierzbicki et al. 2008). The present study, showing that Pol V-associated regions are broader than the zones of siRNA biogenesis, is also consistent with this conclusion.

Although most of the overlap between Pol V and cytosine methylation occurs at sites of CHH methylation and 24-nt siRNA biogenesis, ~7% of Pol V peaks correspond to sites where CG or CHG motifs are methylated, but CHH methylation and 24-nt siRNAs are not detected. One possibility is that de novo cytosine methylation by DRM2, which can methylate cytosines in any sequence context, initially methylated these loci in an RNA-directed fashion, but the symmetrical CG and CHG methylation has since been maintained by MET1 and CMT3. However, we previously identified loci that are silenced by Pol V in partnership with DRD1 (which is required for Pol V transcription), MET1, and DDM1 but independent of proteins required for 24-nt siRNA-directed DNA methylation, including Pol IV, RDR2, DCL3, AGO4, or DRM2 (Pontes et al. 2009). These results indicate that Pol V plays a role at loci silenced by maintenance methylation in addition to its roles at loci silenced by de novo methylation.

At some sites where Pol V peaks do not coincide with sites of cytosine methylation, active cytosine demethylation may be occurring, thus counteracting RNA-directed DNA methylation. For instance, Pol V peak #111, on chromosome 1 and overlapping AT1G31690 (see Supplemental Table S1), is not associated with DNA methylation or small RNAs in Col-0. However, CHH and CG methylation is detected in this region in a triple mutant defective for the DNA glycosylases *ros1*, *dml2*, and *dml3* (<http://neomorph.salk.edu/epigenome/epigenome.html>), involved in the active removal of methylated cytosines (Lister et al. 2008). Cytosine methylation can also feed back on siRNA production such that cytosine demethylation might also mask sites capable of siRNA generation. A potential locus where this might occur is Pol V peak #576 on chromosome 3 (see Supplemental Table S1), which overlaps 11 methylated CHH motifs yet lacks detectable 24-nt siRNAs in wild-type Col-0. In the rdd demethylase triple mutant, eight distinct 24-nt siRNAs produced from the region overlapping the CHH methylation are detected (<http://neomorph.salk.edu/epigenome/epigenome.html>). However, at many of the Pol V-associated regions that do not correspond to sites of CHH methylation or siRNA

biogenesis, there is no evidence for cryptic cytosine methylation or increased siRNA production as revealed by the rdd demethylase triple mutant.

Interestingly, transposons and genes within Pol V peaks differ with respect to the hallmarks of RNA-directed DNA methylation; namely, CHH methylation and the occurrence of 24-nt siRNAs. Transposons within Pol V peaks typically display these hallmarks, whereas Pol V-associated regions that lack CHH methylation and small RNA signatures (refer to Fig. 3E) are heavily biased toward genes. Therefore, Pol V association does not predispose a locus to RNA-directed DNA methylation. Instead, recruitment of Pol IV in order to generate precursors for 24-nt siRNAs is more likely to be critical.

Our search for consensus sequences that might be common to Pol V-associated loci revealed only one motif that was enriched to a statistically significant extent; namely, C(T/A)CC(A/T)CC(A/T)CC or its complement, GG(T/C)GG(T/A)GG(A/T)G. Consistent with CCN and GGN being codons for proline or glycine, respectively, a number of genes encoding proline-rich or glycine-rich proteins are enriched for the motifs, and Pol V associates with these loci. However, the putative consensus sequences are absent at >50% of Pol V peaks, either within the peak region or within 200 bp upstream or downstream, and are present >65,000 times in the genome, suggesting that the motif is not necessary or sufficient to explain genomic sites of Pol V occupancy or recruitment. Thus, it remains unclear how or why Pol V is recruited to specific genomic locations. The possibility that chromatin features, rather than DNA sequences, are key to Pol V recruitment cannot be excluded.

A noteworthy aspect of Pol V's association with proline-rich "structural constituent of cell wall" proteins is that two genes encoding extensins (proline-rich cell wall proteins) were previously identified in a subtractive hybridization screen for genes regulated by DRD1 (Huettel et al. 2006), a putative chromatin remodeler that facilitates Pol V transcription (Wierzbicki et al. 2008). One of these extensin genes, At1g21310, is associated with a Pol V peak, consistent with DRD1 and Pol V playing some role at the locus.

Numerous studies have shown that loss of Pol IV or Pol V activity causes a loss of CHH methylation at specific sites, including methylation-sensitive restriction endonuclease sites, and results in the failure to silence a variety of transposable elements. However, our study shows that ~50% of methylated CHH positions do not require Pol IV or Pol V, particularly in the pericentromeric heterochromatin (Supplemental Fig. S4). These data fit with a previous study in which a *cmt3 drm1 drm2* triple mutant was shown to retain a substantial level of CHH methylation, especially in the pericentromeric heterochromatin, which suggests that MET1 or another uncharacterized cytosine methyltransferase enzyme can methylate CHH motifs in the absence of CMT3 or DRM activity (Cokus et al. 2008). Unexpectedly, we also found that the total number of methylated CHH motifs is little changed in *nrd1*, *nrd1*, or *nrd1(d/e)2* mutants relative to wild-type plants as a result of the occurrence of new, ectopic sites of

CHH methylation in the mutants that are similar in number to the Pol IV/V-dependent sites of CHH methylation. This suggests that Pol IV and Pol V play a role in the targeting of CHH methylation to specific sites but not in the overall activity of cytosine methyltransferases responsible for CHH methylation. The majority of CHH methylation occurs in pericentromeric regions, and, in the absence of Pol IV and Pol V guidance, CHH methylation is even more concentrated in these regions. How and why CHH methylation persists and even accumulates at pericentromeric regions in the absence of Pol IV and Pol V guidance is unclear. However, these regions are enriched in heterochromatin, suggesting that interactions between cytosine methyltransferases and other chromatin-modifying enzymes or chromatin modifications may be responsible.

## Materials and methods

### Access to epigenomic data sets

The nine data sets for the Pol V ChIP-seq, small RNA-seq, or bisulfite-sequencing experiments conducted in this study are accessible via the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) sequence read archive, accession number SRA054962.

### Pol V ChIP and ChIP-seq library generation

Pol V ChIP was performed as described previously (Wierzbicki et al. 2009), except that protein A Dynabeads were used to capture antibody–Pol V complexes. ChIP quantitative PCR analyses were performed as described previously (Wierzbicki et al. 2008). ChIP-seq libraries for sequencing were prepared following standard ChIP-seq library generation protocols from Illumina.

### MethylC-seq library generation

Genomic DNA was extracted from immature (unopened) flower buds using the DNeasy Plant minikit (Qiagen), and 5  $\mu$ g was fragmented by sonication to 50–500 bp with a Bioruptor (Diagenode). Sheared ends were repaired (End-It DNA End Repair kit, Epicentre) using a nucleotide triphosphate mix free of dCTP, followed by 3' A-tailing with 0.2 mM dATP and 15 U of Klenow fragment of DNA polymerase and ligation of 10  $\mu$ L of single-read methylated adapters (Illumina) using 15 U of T4 DNA ligase (New England Biolabs). Adapter-ligated genomic DNA of 150–200 bp was isolated from a 2% (w/w) agarose gel, and sodium bisulfite conversion was then performed using the MethylEasy Xceed kit (Human Genetic Signatures) according to the manufacturer's instructions. Library amplification was performed by emulsion PCR as follows: 20% of the bisulfite-converted, adapter-ligated DNA molecules was mixed with 1 $\times$  PfuTurbo reaction buffer, 4% v/v DMSO, 100 mg/mL BSA, 0.2 mM dNTPs, 0.5 mM primer 1.1, and 0.5 mM primer 2.1 in a final volume of 196  $\mu$ L. After incubation for 3 min at 72°C, 10 U of PfuTurboC<sub>x</sub> Hotstart DNA polymerase (Stratagene) was added, mixed, and incubated for 5 min at 72°C, then cooled to room temperature. The PCR reaction was then added to 400  $\mu$ L of oil surfactant solution (4.5% [v/v] Span80, 0.4% Tween80, 0.05% Triton X-100, 95.05% mineral oil) while mixing with a magnetic stirrer to form an emulsion. PCR was performed on 50- $\mu$ L aliquots of the emulsion overlaid with mineral oil as follows: 5 min at 94°C; then 40 cycles of 1 min at 94°C, 30 sec at 65°C, and 3 min at

72°C; followed by a final incubation for 10 min at 72°C. Reactions were frozen for  $\geq 1$  h at  $-20^\circ\text{C}$ , then thawed at room temperature. After centrifugation at 20,000g for 5 min, the upper (oil) phase was discarded, and the emulsion PCR products were extracted by adding 0.5 mL of water-saturated diethyl ether, vortexing for 10 sec, centrifugation at 20,000g for 5 min, and discarding of the upper (diethyl ether) phase. Diethyl ether extraction was repeated once, after which 1 mL of water-saturated chloroform and 100  $\mu$ L of H<sub>2</sub>O were added to the aqueous solution, vortexed for 10 sec, and centrifuged at 20,000g for 5 min. The top, aqueous phase was removed, retained, and purified using the MinElute PCR purification kit (Qiagen).

### Small RNA library construction

Total RNA was isolated from immature (unopened) flower buds using the RNeasy Plant minikit (Qiagen). Small RNAs were enriched from the total RNA and small RNA sequencing libraries generated as described previously (Lister et al. 2008).

### MethylC-seq data analysis

MethylC-seq sequencing data was processed, filtered, aligned to the TAIR9 reference genome, and analyzed to identify sites of DNA methylation as described previously (Lister et al. 2008), except reads that contained more than three cytosines in a non-CG context were not discarded. The sodium bisulfite non-conversion rate was calculated as the percentage of cytosines sequenced at cytosine reference positions in the chloroplast genome.

### Sequencing read alignment and processing

All small RNA-seq reads that contained at least five bases of the 3' adapter sequence were selected, and the adapter sequence was removed, retaining the trimmed reads that were 16–37 nt in length. These processed reads were aligned to the *A. thaliana* reference genome (TAIR9) with the Bowtie alignment algorithm using the following parameters: `–solexa-quals –e 1 –l 20 –n 0 –a –m 1000–best–nomaqround`. Consequently, any read that aligned with no mismatches and to no more than 1000 locations in the reference genome sequence was retained for subsequent analysis. Read alignment of ChIP-seq data to the *A. thaliana* reference genome (TAIR9) was performed with the ELAND algorithm.

### Pol V peak calling and bioinformatic analyses

The software tool MACS version 1.4 was used to find enriched regions in the ChIP-seq data (Zhang et al. 2008). Col-0 was used as the treatment, and *nripe1-11* was used as the control, thereby ensuring detection of enriched regions in Col-0 that were absent in the Pol V mutant. MACS was developed for the prediction of binding sites using ChIP-seq data. The tool models tag distribution in the enriched sample (Col-0) as a Poisson distribution and measures the enrichment with respect to the background (*nripe1-11*). MACS also determines false positives by inverting the treatment and control files and calculates false discovery rate for a particular *P*-value cutoff. The default values for most MACS parameters were used, except for genome size (set to 150 Mb) and the “shifting model” (turned off). This resulted in detection of 1157 enriched regions or “peaks” with a reported false discovery rate ranging from 0% to 4% at a *P*-value cutoff on  $1 \times 10^{-5}$ .

The DNA sequence for each of the 1157 regions of Pol V enrichment were extracted from the TAIR9 genome annotation and submitted to BioProspector and BioOptimizer for motif finding

(Liu et al. 2001; Jensen and Liu 2004). The entire genome was used to construct a background model that was also given to BioProspector, which uses a Gibbs sampling technique for motif detection. The motif width was set to 10 bp. A second tool, BioOptimizer, was used to further validate the motifs detected (Jensen and Liu 2004). BioOptimizer takes the output of BioProspector, varies the length of the detected motif, and reports a consensus sequence with a final score. Five similar consensus sequences were reported, out of which the consensus sequence with the greatest difference between final and null score was used for further analyses. The position weight matrix (PWM) of the motif cCtCCaCCacc was given as input to Patser, and motif searching across the entire genome was accomplished using default settings (Hertz and Stormo 1999). FatiGO (<http://babelomics.bioinfo.cipf.es>) was used for GO analysis of genes overlapped by Pol V peaks.

#### RT-PCR

RT-PCR detection of Pol V-dependent transcripts was performed using primers listed in Supplemental Table S12 and as previously described (Wierzbicki et al. 2008). For RT-PCR analysis of transcription at the AT2TE82000 locus, 1  $\mu$ g of RNA was digested with DNase RQ1 (Promega) following the manufacturer's instructions. RNA was reverse-transcribed using random hexamer primers and SuperScript III RT (Invitrogen) following the manufacturer's instructions. One microliter of the resulting cDNA was amplified with GoTaq Green master mix (Promega) and the primers listed in Supplemental Table S12.

#### Acknowledgments

We thank Murali (Ram) Podichetti for identifying genes and/or transposons overlapping Pol V peaks. This work was supported by National Institutes of Health grant GM07759 to C.S.P., an Indiana University Competitive Research Grant to C.S.P. and H.T., and National Science Foundation Grant MCB 1120271 to A.T.W. R.C. and M.J.R. were supported by Ruth L. Kirschstein National Research Service Awards. C.S.P. and J.R.E. are Investigators of the Howard Hughes Medical Institute and the Gordon and Betty Moore Foundation. R.C. and C.S.P. wrote the manuscript.

#### References

- Cao X, Jacobsen SE. 2002. Role of the *Arabidopsis* DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* **12**: 1138–1144.
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**: 215–219.
- El-Shami M, Pontier D, Lahmy S, Braun L, Picart C, Vega D, Hakimi MA, Jacobsen SE, Cooke R, Lagrange T. 2007. Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev* **21**: 2539–2544.
- Finnegan EJ, Peacock WJ, Dennis ES. 1996. Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. *Proc Natl Acad Sci* **93**: 8449–8454.
- Gao Z, Liu HL, Daxinger L, Pontes O, He X, Qian W, Lin H, Xie M, Lorkovic ZJ, Zhang S, et al. 2010. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* **465**: 106–109.
- Haag JR, Pikaard CS. 2011. Multisubunit RNA polymerases IV and V: Purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol* **12**: 483–492.
- Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**: 118–120.
- Hertz GZ, Stormo GD. 1999. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15**: 563–577.
- Huang L, Jones AM, Searle I, Patel K, Vogler H, Hubner NC, Baulcombe DC. 2009. An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat Struct Mol Biol* **16**: 91–93.
- Huettel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJ, Matzke M. 2006. Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *EMBO J* **25**: 2828–2836.
- Jensen ST, Liu JS. 2004. BioOptimizer: A Bayesian scoring function approach to motif discovery. *Bioinformatics* **20**: 1557–1564.
- Kanno T, Aufsatz W, Jaligot E, Mette MF, Matzke M, Matzke AJ. 2005a. A SNF2-like protein facilitates dynamic control of DNA methylation. *EMBO Rep* **6**: 649–655.
- Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, Daxinger L, Kreil DP, Matzke M, Matzke AJ. 2005b. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* **37**: 761–765.
- Kanno T, Bucher E, Daxinger L, Huettel B, Bohmdorfer G, Gregor W, Kreil DP, Matzke M, Matzke AJ. 2008. A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat Genet* **40**: 670–675.
- Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu JK, Wohlschlegel JA, Jacobsen SE. 2010. A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in *Arabidopsis*. *Curr Biol* **20**: 951–956.
- Li CF, Pontes O, El-Shami M, Henderson IR, Bernatavichute YV, Chan SW, Lagrange T, Pikaard CS, Jacobsen SE. 2006. An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* **126**: 93–106.
- Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE. 2001. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* **292**: 2077–2080.
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**: 523–536.
- Liu X, Brutlag DL, Liu JS. 2001. BioProspector: Discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput* **6**: 127–138.
- Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ. 2005. Elucidation of the small RNA component of the transcriptome. *Science* **309**: 1567–1569.
- Luo J, Hall BD. 2007. A multistep process gave rise to RNA polymerase IV of land plants. *J Mol Evol* **64**: 101–112.
- Mosher RA, Schwach F, Studholme D, Baulcombe DC. 2008. PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc Natl Acad Sci* **105**: 3145–3150.
- Olmedo-Monfil V, Duran-Figueroa N, Arteaga-Vazquez M, Demesa-Arevalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada J-P. 2010. Control of female gamete formation by a small RNA pathway in *Arabidopsis*. *Nature* **464**: 628–632.
- Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, Pikaard CS. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**: 613–622.

Wierzbicki et al.

- Pontes O, Li CF, Nunes PC, Haag J, Ream T, Vitins A, Jacobsen SE, Pikaard CS. 2006. The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126**: 79–92.
- Pontes O, Costa-Nunes P, Vithayathil P, Pikaard CS. 2009. RNA polymerase V functions in *Arabidopsis* interphase heterochromatin organization independently of the 24-nt siRNA-directed DNA methylation pathway. *Mol Plant* **2**: 700–710.
- Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, Hakimi MA, Lerbs-Mache S, Colot V, 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.
- Qi Y, Denli AM, Hannon GJ. 2005. Biochemical specialization within *Arabidopsis* RNA silencing pathways. *Mol Cell* **19**: 421–428.
- Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**: 1008–1012.
- Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, Zhu JK, Hagen G, Guilfoyle TJ, Pasa-Tolic L, Pikaard CS. 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.
- Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL. 1996. Demethylation-induced developmental pleiotropy in *Arabidopsis*. *Science* **273**: 654–657.
- Tucker SL, Reece J, Ream TS, Pikaard CS. 2010. Evolutionary history of plant multisubunit RNA polymerases IV and V: Subunit origins via genome-wide and segmental gene duplications, retrotransposition, and lineage-specific subfunctionalization. *Cold Spring Harb Symp Quant Biol* **75**: 285–297.
- Wierzbicki AT, Haag JR, Pikaard CS. 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**: 635–648.
- Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* **41**: 630–634.
- Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC. 2004. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* **2**: 642–652.
- Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**: 1189–1201.
- Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE. 2007. Role of RNA polymerase IV in plant small RNA metabolism. *Proc Natl Acad Sci* **104**: 4536–4541.
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-seq (MACS). *Genome Biol* **9**: R137. doi: 10.1186/gb-2008-9-9-r137.
- Zheng X, Zhu J, Kapoor A, Zhu J-K. 2007. Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J* **26**: 1691–1701.



## Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the *Arabidopsis* epigenome

Andrzej T. Wierzbicki, Ross Cocklin, Anoop Mayampurath, et al.

*Genes Dev.* 2012 26: 1825-1836 originally published online August 1, 2012  
Access the most recent version at doi:[10.1101/gad.197772.112](https://doi.org/10.1101/gad.197772.112)

---

**Supplemental Material** <http://genesdev.cshlp.org/content/suppl/2012/07/25/gad.197772.112.DC1>

**References** This article cites 39 articles, 13 of which can be accessed free at:  
<http://genesdev.cshlp.org/content/26/16/1825.full.html#ref-list-1>


Articles cited in:  
<http://genesdev.cshlp.org/content/26/16/1825.full.html#related-urls>

**Related Content** **Seeing the forest for the trees: a wide perspective on RNA-directed DNA methylation**  
Huiming Zhang and Jian-Kang Zhu  
*Genes Dev.* August 15, 2012 26: 1769-1773

**Open Access** Freely available online through the *Genes & Development* Open Access option.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---



Biofluids too dilute to detect microRNAs? See what to do. **EXIQON**

---

To subscribe to *Genes & Development* go to:  
<http://genesdev.cshlp.org/subscriptions>