# RNA polymerases IV and V influence the 3' boundaries of Polymerase II transcription units in *Arabidopsis*

Anastasia McKinlay[1, 6], Ram Podicheti[2,3], Ross Cocklin[1,4,5], Craig S. Pikaard[1,4] and Douglas B. Rusch[2]

[1] Department of Biology and Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, Indiana 47405, USA

[2] Center for Genomics and Bioinformatics, Indiana University, Bloomington, Indiana 47405, USA

[3] School of Informatics and Computing, Indiana University, Bloomington, IN 47405, USA

[4] Howard Hughes Medical Institute, Indiana University, Bloomington, Indiana 47405

[5] Current address: Keene State College, Keen, New Hampshire 03431

[6] Corresponding author: Department of Biology and Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, Indiana 47405, USA. E-mail: anagrid@indiana.edu

Running title: Polymerases IV and V affect Pol II transcription termination

**Key words: plant, nuclear multisubunit RNA polymerases IV and V, Pol IV, Pol V, RNA directed DNA Methylation, RdDM, transcription termination, nuclear run-on assay**

## Abstract

Nuclear multisubunit RNA polymerases IV and V (Pol IV and Pol V) evolved in plants as specialized forms of Pol II. Their functions are best understood in the context of RNA-directed DNA methylation (RdDM), a process in which Pol IV-dependent 24 nt siRNAs direct the *de novo* cytosine methylation of regions transcribed by Pol V. Pol V has additional functions, independent of Pol IV and 24 nt siRNA biogenesis, in maintaining the repression of transposons and genomic repeats whose silencing depends on maintenance cytosine methylation. Here we report that Pol IV and Pol V play unexpected roles in defining the 3' boundaries of Pol II transcription units. Nuclear run-on assays reveal that in the absence of Pol IV or Pol V, Pol II occupancy downstream of poly A sites increases for approximately 12% of protein-coding genes. This effect is most pronounced for convergently transcribed gene pairs. Although Pols IV and V are detected near transcript ends of the affected Pol II – transcribed genes, their role in limiting Pol II read-through is independent of siRNA biogenesis or cytosine methylation. We speculate that Pols IV and V (and/or their associated factors) play roles in Pol II transcription termination by influencing polymerase bypass or release at collision sites for convergent genes.

## Introduction

Nuclear multisubunit RNA polymerases IV and V (Pol IV and Pol V) play distinct roles in the transcriptional silencing of genes and transposons via RNA-directed DNA methylation (RdDM). [1–3]. Pol IV is thought to initiate the major RdDM pathway, transcribing DNA to generate relatively short RNA transcripts that are then copied into double-stranded RNAs by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) [4,5]. Resulting double-stranded RNAs are diced by DICER-LIKE 3 (DCL3) into 24 nt siRNAs that are loaded into ARGONAUTE 4 (AGO4), or a related AGO protein [6,7]. In parallel, Pol V generates transcripts to which AGO4-siRNA complexes bind [8,9], ultimately bringing about repressive cytosine methylation (in CG, CHG or CHH contexts) by the *de novo* cytosine methyltransferase, DRM2 [10–12]. There is evidence that Pol V has functions independent of RdDM, helping maintain repression of transposons and genomic repeats whose silencing depends on maintenance methylation by MET1 and the chromatin

remodeling ATPase, DDM1 [13–15]. Pol V has also been implicated in at least one alternative TE silencing pathway, the RDR6-dependent DNA methylation pathway (RDR6-RdDM) [14–16]. In this Pol IV- and 24-nt siRNA- independent pathway, Pol II-derived RNAs from TE regions are processed into 21- and 22-nt siRNAs in a RDR6, DCL2, and DCL4-dependent manner, loaded into AGO6 and targeted to Pol V scaffold transcripts [14–16]. Thus, Pol V can function as a downstream component in both 24-nt siRNA-dependent (canonical RdDM) and 21- and 22-nt siRNA-dependent (RDR6-RdDM) TE silencing pathways.

Steady-state levels of mRNAs transcribed by Pol II in *Arabidopsis thaliana* are minimally affected in *pol IV* or *pol V* mutant plants [17]. However, several observations have suggested that Pol IV and V function in proximity to Pol II transcription units. At long retrotransposons, 24-nt siRNA biogenesis and RdDM occur primarily at the ends of the elements, where promoter sequences recognized by the Pol II transcription machinery are located [14,18]. Moreover, chromatin immunoprecipitation studies have shown that approximately 30% of Pol V occupied positions are proximal to the promoters of Pol II - transcribed genes [10]. Likewise, 21-24 nt sRNAs are enriched within 1 kb of either the 5' or 3' ends of transcription units for approximately 20% of protein-coding genes, with studies suggesting that sRNA abundance correlates with mRNA abundance at these genes [19,20]. However, unlike Pol V ChIP peaks that correspond to transposons, Pol V peaks that overlap protein-coding genes typically do not correlate with 24-nt siRNAs or CHH DNA methylation, which are hallmarks of RdDM, suggesting the possibility of an alternative function of Pol V in regulation of protein-coding genes.

RNA Pol II transcription termination in eukaryotes is a complex process. Termination by Pol II occurs stochastically within a region of several hundred bases downstream of the polyA site and is functionally connected to the cleavage and polyadenylation of nascent transcript 3' ends [21–24]. Transcription pause sites, binding sites for protein termination factors, the presence of RNA-DNA hybrids, or chromatin modifications within the termination region can affect termination efficiency [25–27].

We conducted nuclear run-on assays combined with high-throughput sequencing (NRO-Seq) [28] to identify genomic regions where nascent RNAs are synthesized, comparing

wild-type plants to *pol IV* or *pol V* mutant plants, similar to an analogous study conducted in maize [29]. Interestingly, as in maize, we detect increased RNA Pol II abundance downstream from the cleavage and polyadenylation sites of protein coding genes in *pol IV* or *pol V* null mutants. By contrast, nascent transcript abundance in the region from the initiation site to the PolyA site is unaffected, as are steady-state mRNA levels. This suggests that 3' end processing of pre-mRNAs is not affected in *pol IV* or *pol V* mutants, but termination and Pol II release is impaired or delayed. Convergently-transcribed protein coding genes whose 3' regions overlap show the greatest increase in nascent RNA signals downstream of PolyA addition sites when Pol IV or Pol V are mutated. Those gene pairs also have a higher tendency to overlap with sites of Pol IV, Pol V or AGO4 occupancy, as measured by ChIP. However, these regions are not characterized by abundant 24-nt siRNAs or CHH methylation, suggesting that Pol II transcription termination does not require RdDM modification within termination zones. Collectively, our results suggest a methylation independent role for Pols IV and V in shaping the boundaries of Pol II transcription units.

## Results

## Nascent transcript levels increases downstream of mRNA polyA addition sites in *pol IV* or *pol V* mutants

Nascent RNA transcripts produced by template-engaged RNA polymerases in purified nuclei were generated by nuclear run-on (NRO) in the presence of biotin-UTP, allowing capture of the labeled RNAs using streptavidin beads. Quantitative PCR (qPCR) assays, comparing reactions conducted using biotin-UTP versus unmodified UTP, were then used to calculate the fold enrichment achieved for biotinylated RNAs versus background for a set of highly-expressed *Arabidopsis thaliana* genes (**Suppl. Figure 1**). This analysis allowed us to estimate that nascent biotinylated RNA accounts for more than 99.7% of the RNA captured on the streptavidin beads, with background binding, attributable to unlabelled RNAs, accounting for the remaining 0.3%.

We performed nuclear run-on using nuclei of wild-type plants as well as *pol IV* and *pol V* mutant plants and constructed sequencing libraries for resulting RNAs (NRO-Seq library), as well as for purified total nuclear RNAs (RNA-Seq library), thus allowing comparisons between nascent and steady-state RNA levels (**Suppl. Figure 2**). The NRO-Seq and RNA-Seq data were aligned to *Arabidopsis thaliana* TAIR10 RefSeq transcription start sites (TSSs) and transcription end sites (TESs), corresponding to polyA addition sites, for 27859 protein-coding genes (**Figure 1**). Both nascent RNA (**Figure 1A, in black**) and steady-state RNA (**Figure 1B, in black**) profiles show a sharp increase in RNA sequence reads starting at the TSS, followed by a plateau that begins ~200 bps downstream. This pattern contrasts with nascent RNA profiles reported for yeast and other eukaryotic organisms, in which peaks are often observed downstream of the TSS, reflecting promoter-proximal pausing of Pol II [28,30–34]. The absence of a promoter-proximal peak in the nascent RNA profile in *Arabidopsis* suggests that promoter-proximal pausing is not a general feature of *Arabidopsis* genes transcribed by Pol II. Similarly, no evidence of promoter-proximal pausing was observed in maize NRO data [29].

Loss of Pol IV or V activity does not affect RNA Pol II read density profiles near TSSs for either nascent RNA (**Figure 1A,** compare black with red and green) or steady-state RNA (**Figure 1B**), suggesting that RNA polymerases IV and V exert no detectable influence on Pol II transcription initiation at protein-coding genes on a genome-wide scale.

Analysis of the nascent RNA (NRO) read distribution in the 1 kb region surrounding the transcript end sites (TESs) revealed a gradual decrease in read abundance prior to the TES, followed by a small increase in reads ~100 bp post-TES (**Figure 1C, in black**). This read density peak just downstream of TES in the nascent RNA population is not observed for steady-state RNA (**Figure 1D, in black**), suggesting that post-TES RNA is removed by cleavage at the polyA addition site and thus absent from mature mRNAs. Interestingly, a further increase in nascent transcript abundance occurs downstream of TESs in *pol IV* or *pol V* mutant plants, suggesting increased in Pol II occupancy, possibly due to delayed termination or Pol II release (**Figure 1C**).

To determine if specific genes have dramatically altered Pol II occupancy downstream of TESs in *pol IV* or *pol V* mutants, we calculated a transcript end site - associated termination index (TES-TI) for each gene as the number of total reads in the first 500 bp downstream of the TES (interval 2) divided by the number of total reads in the 500 bp preceding the TES (interval 1) (**Figure 2A**). Approximately 12% of the 18,419 protein coding genes analyzed exhibit a two fold or greater increase in their termination index in *pol IV* (*nrpd1-3*) or *pol V* (*nrpe1-11*) mutants, compared to wild-type (**Figure 2B**). Profiles for these genes, plotted as in Figure 1, are shown in **Figure 2C**. A strong concordance (~80% overlap) is observed for genes affected in *pol IV* or *pol V* mutant plants (**Figure 2B, Venn diagram**). A set of 1312 genes, affected at least two fold by both *pol IV* and *pol V*, was chosen for all subsequent analyses. Loss of Pol IV or Pol V activities did not cause detectable lengthening of mature mRNAs for these genes, suggesting that RNA processing still occurs (**Figure 2D**), as illustrated in the gene browser profiles shown in **Figure 2E**.

To further examine whether changes in *pol IV* or *pol V* mutants affect qualitative or temporal aspects of 3' end processing, we performed circular RT-PCR analyses at two loci, *At5g24670* and *At5g36700*, using primers that allow discrimination between full-

length processed transcripts and transcripts that extend post-TES (**Figure 3A**). Longer circularized transcripts were detected in *pol IV* and *pol V* mutant plants, compared to wild-type Col-0 plants (**Figure 3B**). Cloning and sequencing of these PCR products revealed that the majority of clones support the annotated gene models, but for the *At5g24670* transcript 14% of the cloned cDNAs in *nrpd1* and 24% of the cDNA clones in *nrpe1* include sequences downstream of the annotated TES (**Figure 3C**). Combined with our observation that steady-state mRNAs do not display changes in their 3' ends, this suggests that 3' end processing may be less efficient in *pol IV* or *pol V* mutants.

### Pol II occupancy correlates with nuclear run-on signal intensity

Using chromatin immunoprecipitation (ChIP) of exonuclease-trimmed chromatin-protein complexes (ChIP-exo), we examined Pol II association with genes displaying increased (>2x) post-TES nascent RNA-seq read density in *pol IV* or *pol V* mutants (**Suppl. Figure 3**). The ChIP-exo method allows high resolution mapping of DNA binding factors due to the exonuclease trimming of sequences not protected by the bound protein [35]. Consistent with NRO-seq results, the ChIP-exo sequence data revealed increased Pol II association with gene 3' ends in *pol IV* mutants (**Suppl. Figure 3A**), compared to Col-0 wild-type, with a strong peak downstream of the TESs. In contrast, no differences in RNA Pol II occupancy was observed near the transcription start sites for the same set of genes (**Suppl. Figure 3B**), suggesting an asymmetric affect at gene 3' ends.

### Pol II accumulation in the post-TES region is increased for convergent gene pairs

We next examined whether changes in RNA expression levels correlate with increased post-TES read density in *pol IV* or *pol V* mutants. These analyses revealed that in mutant plants, genes with increased post-TES signals tend to be highly expressed, compared to total protein-coding genes (**Suppl. Figure 4**).

Interestingly, convergent gene pairs were significantly enriched among genes with increased post-TES read density in *pol IV* or *pol V* mutant plants (**Figure 4A**). Furthermore, these genes tend to have neighboring genes that initiate or end within 100 bps downstream or 100 bps upstream of their TESs (**Figure 4B**).

## Genes with Pol IV and V – affected transcription termination associate with RdDM pathway components

We examined published Pol IV [36], Pol V [10], and AGO4 [37] ChIP-seq datasets to see if these proteins are enriched near the 3' ends of genes that show increased Pol II occupancy post TES in *pol IV* or *V* mutant plants. Interestingly, for 60% of the 1312 genes with increased post-TES read density, Pol IV, Pol V, or AGO4 binding sites overlap the 1 kb region surrounding their TESs, with Pol V being present at 50% of these genes (**Figure 5A**). Analysis of Pol V ChIP-seq [10] read distribution in wild-type (Col-0) plants revealed Pol V association primarily upstream of TESs, with an additional peak, compared to background (*nrpe1-11*), 150-250 bp downstream of the TES (**Figure 5B**). Analysis of Pol IV ChIP-seq [36] data did not reveal evidence for a similar pattern of enrichment (data not shown).

Fewer than 1% (12 out of 1312) of the analyzed genes displayed overlapping Pol IV, Pol V, and AGO4 occupancy post-TES, as would be expected of loci subjected to RNA-directed DNA methylation. This led us to further examine whether 23-24 nt siRNAs map near TESs of genes affected by the loss of Pol IV or Pol V. 18% of affected genes overlap with 23-24 nt and 25% with 21-22 nt siRNA reads within a 2 kb region surrounding the TESs (**Suppl. Figure 5A**). But only about 10% (130 out of 1312) of genes showed a two-fold or more reduction of 23-24 nt sRNA in *pol IV* or *pol V* mutant plants (**Suppl. Figure 5B**). Similarly, no correlation between post-TES read density and CHH methylation was detected in mutant plants (**Suppl. Figure 6A**). This suggests that effect of Pol IV and Pol V on Pol II transcription termination is unlikely to be due to RNA-directed DNA methylation. Interestingly, CG methylation profile mimics Pol V occupancy at genes that show increased Pol II signal post TES in *pol IV* or *V* mutant plants (**Suppl.**

**Figure 6B**). Consistent with previous report [38], CG methylation might guide Pol V to Pol II transcript end sites. Whereas lack of CHH methylation and siRNA accumulation at these sites suggests that Pol V is not transcriptionally engaged.

## Discussion

Nuclear run-on combined with high-throughput sequencing (NRO-Seq)[28] and RNA-Seq [39] revealed that Pol IV and Pol V affect Pol II occupancy downstream of polyA addition sites (TESs), for approximately 12% of protein-coding genes. Consistently, Pol IV or Pol V interaction sites include the TES regions for genes that display increased NRO signals downstream of TESs in *pol IV* or *pol V* mutant plants. However, for the majority of these genes there is no enrichment for 24-nt siRNAs and/or changes in CHH methylation, a hallmark of Pol IV and Pol V - dependent RdDM. Thus, any influence of Pol IV and Pol V on Pol II transcription termination is likely independent of DNA methylation.

Convergent gene pairs tend to show the highest post-TES NRO signals in *pol IV* or *pol V* mutant plants. Convergent transcription can result in head-to-head collisions of transcribing Pol II enzymes [40]. Though such events are expected to be rare, Pol II collision doesn't result in polymerase dissociation from the template but leads to transcriptional arrest and gene blockage [40]. To date, not much is known about how cells deal with collision-induced transcription blocks. However, Pol II ubiquitination-mediated degradation has been shown to be involved in clearing Pol II from the region between convergent genes, such that lack of the yeast RNA Pol II ubiquitination/degradation protein ELC1 results in a peak of RNA Pol II density between convergent genes [40]. A possibility is that Pol IV or Pol V occupancy at intergenic DNA may reduce RNA polymerase II collisions by slowing or dissociating Pol II, thus allowing more efficient cleavage at polyA sites and subsequent termination (**Figure 6**). Consistently, we observed lower Pol II occupancy just upstream of TES (**Figure 1C and Figure 2C**) as well as higher levels of 3' unprocessed transcripts (**Figure 3**) in *pol IV* or *pol V* mutant plants.

Pol V transcription can recruit the SWI/SNF chromatin remodeling complex, such that a loss of Pol V transcripts correlates with changes in nucleosome positioning [41]. Whether or not Pol IV can similarly affect chromatin states has not yet been determined. However, CLASSY1, a putative ATP-dependent DNA translocase, has been implicated in Pol IV function and speculated to affect nucleosome repositioning, similar to SWI/SNF [42,43]. Thus, Pol IV or Pol V – mediated chromatin modifications near Pol II TESs may also influence Pol II termination or release.

# References

1.    Mosher, R. A. & Melnyk, C. W. siRNAs and DNA methylation: seedy epigenetics. *Trends Plant Sci.* **15,** 204–210 (2010).

2.    Almeida, R. & Allshire, R. C. RNA silencing and genome regulation. *Trends Cell Biol.* **15,** 251–258 (2005).

3.    Pontier, D. *et al.* Reinforcement of silencing at transposons and highly repeated sequences requires  the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev.* **19,** 2030–2040 (2005).

4.    Blevins, T. *et al.* Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. *Elife* **4,** e09591 (2015).

5.    Haag, J. R. *et al.* In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. *Mol. Cell* **48,** 811–818 (2012).

6.    Zilberman, D., Cao, X. & Jacobsen, S. E. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299,** 716–719 (2003).

7.    Zheng, X., Zhu, J., Kapoor, A. & Zhu, J.-K. Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J.* **26,** 1691–1701 (2007).

8.    Wierzbicki, A. T., Haag, J. R. & Pikaard, C. S. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional  silencing of overlapping and adjacent genes. *Cell* **135,** 635–648 (2008).

9.    Wierzbicki, A. T., Ream, T. S., Haag, J. R. & Pikaard, C. S. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat. Genet.* **41,** 630–634 (2009).

10.   Wierzbicki, A. T. *et al.* Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. *Genes Dev.* **26,** 1825–1836 (2012).

11.   Onodera, Y. *et al.* Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120,** 613–622 (2005).

12.   Pikaard, C. S., Haag, J. R., Pontes, O. M. F., Blevins, T. & Cocklin, R. A transcription fork model for Pol IV and Pol V-dependent RNA-directed DNA methylation. *Cold Spring Harb. Symp. Quant. Biol.* **77,** 205–212 (2012).

13. Pontes, O., Costa-Nunes, P., Vithayathil, P. & Pikaard, C. S. RNA polymerase V functions in Arabidopsis interphase heterochromatin organization independently of the 24-nt siRNA-directed DNA methylation pathway. *Mol. Plant* **2,** 700–710 (2009).

14. Fultz, D., Choudury, S. G. & Slotkin, R. K. Silencing of active transposable elements in plants. *Curr. Opin. Plant Biol.* **27,** 67–76 (2015).

15. Nuthikattu, S. *et al.* The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol.* **162,** 116–131 (2013).

16. Wu, L., Mao, L. & Qi, Y. Roles of dicer-like and argonaute proteins in TAS-derived small interfering RNA-triggered DNA methylation. *Plant Physiol.* **160,** 990–999 (2012).

17. Kasschau, K. D. *et al.* Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biol.* **5,** e57 (2007).

18. Stroud, H. *et al.* Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat. Struct. Mol. Biol.* **21,** 64–72 (2014).

19. Hollister, J. D. *et al.* Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 2322–2327 (2011).

20. Greaves, I. K. *et al.* Trans chromosomal methylation in Arabidopsis hybrids. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 3570–3575 (2012).

21. Proudfoot, N. J., Furger, A. & Dye, M. J. Integrating mRNA processing with transcription. *Cell* **108,** 501–512 (2002).

22. Connelly, S. & Manley, J. L. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.* **2,** 440–452 (1988).

23. Buratowski, S. Connections between mRNA 3' end processing and transcription termination. *Curr. Opin. Cell Biol.* **17,** 257–261 (2005).

24. Logan, J., Falck-Pedersen, E., Darnell, J. E. J. & Shenk, T. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc. Natl. Acad. Sci. U. S. A.* **84,** 8306–8310 (1987).

25. Proudfoot, N. J. Ending the message: poly(A) signals then and now. *Genes Dev.* **25,** 1770–1782 (2011).

26. Richard, P. & Manley, J. L. Transcription termination by nuclear RNA polymerases. *Genes Dev.* **23,** 1247–1269 (2009).

27. Rosonina, E., Kaneko, S. & Manley, J. L. Terminating the transcript: breaking up is hard to do. *Genes Dev.* **20,** 1050–1056 (2006).

28. McKinlay, A., Araya, C. L. & Fields, S. Genome-Wide Analysis of Nascent Transcription in Saccharomyces cerevisiae. *G3 (Bethesda).* **1,** 549–558 (2011).

29. Erhard, K. F. J., Talbot, J.-E. R. B., Deans, N. C., McClish, A. E. & Hollick, J. B. Nascent transcription affected by RNA polymerase IV in Zea mays. *Genetics* **199,** 1107–1125 (2015).

30. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322,** 1845–1848 (2008).

31. Churchman, L. S. & Weissman, J. S. Native elongating transcript sequencing (NET-seq). *Curr. Protoc. Mol. Biol.* **Chapter 4,** Unit 4.14.1–17 (2012).

32. Nechaev, S. *et al.* Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science* **327,** 335–338 (2010).

33. Lis, J. T. Imaging Drosophila gene activation and polymerase pausing in vivo. *Nature* **450,** 198–202 (2007).

34. Core, L. J. & Lis, J. T. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science* **319,** 1791–1792 (2008).

35. Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147,** 1408–1419 (2011).

36. Law, J. A. *et al.* Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498,** 385–389 (2013).

37. Zheng, Q. *et al.* RNA polymerase V targets transcriptional silencing components to promoters of protein-coding genes. *Plant J.* **73,** 179–189 (2013).

38. Bohmdorfer, G. *et al.* Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. *Elife* **5,** (2016).

39. Nagalakshmi, U., Waern, K. & Snyder, M. RNA-Seq: a method for comprehensive transcriptome analysis. *Curr. Protoc. Mol. Biol.* **Chapter 4,** Unit 4.11.1–13 (2010).

40. Hobson, D. J., Wei, W., Steinmetz, L. M. & Svejstrup, J. Q. RNA polymerase II collision interrupts convergent transcription. *Mol. Cell* **48,** 365–374 (2012).

41.  Zhu, Y., Rowley, M. J., Bohmdorfer, G. & Wierzbicki, A. T. A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing. *Mol. Cell* **49,** 298–309 (2013).

42.  Smith, L. M. *et al.* An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. *Plant Cell* **19,** 1507–1521 (2007).

43.  Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78,** 273–304 (2009).

44.  Gaudino, R. J. & Pikaard, C. S. Cytokinin induction of RNA polymerase I transcription in Arabidopsis thaliana. *J. Biol. Chem.* **272,** 6799–6804 (1997).

45.  Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483,** 295–301 (2012).

## Figure Legends

**Figure 1: Loss of Pol IV or Pol V results in increased NRO-Seq read density downstream of protein-coding gene transcript end sites. (A,B)** Average read densities are plotted relative to the transcription start sites (TSS) of 27,859 genes for nascent RNA **(A)** or steady-state RNA **(B)** isolated from Col-0, *pol IV* or *pol V* mutant plants. **(C)** Nascent RNA read-density plotted relative to transcript end sites (TES). **(D)** As in **(C)** but for steady-state RNA.

**Figure 2: Genes with increased read density downstream of their transcript end sites (TES) in *pol IV* or *pol V* mutant plants. (A)** Schematic for calculating the TES-termination index (TES-TI). **(B)** Histogram showing the number of genes with increased or decreased TES-TI values in *pol IV* or *pol V* mutant plants, relative to wild-type (Col-0) plants. **(Insert in B)** Venn diagram showing the overlap between genes with 2 fold or more increase in post-TES read density in *pol IV* (*nrpd1-3,* in red) or *pol V* (*nrpe1-11,* in green) plants relative to Col-0 plants. **(C)** Nascent RNA read-density plotted relative to TESs for the 1312 genes affected by both Pol IV and Pol V (see Venn diagram of panel **B**). **(D)** As in **(C)** but for steady-state RNA. **(E)** Examples of protein-coding genes with increased read density downstream of the TES (marked with blue dashed vertical line) in *pol IV* (*nrpd1-3*) or *pol V* (*nrpe1-11*) mutants, displayed using JBrowse.

**Figure 3: Transcripts with increased read density post-TES in *pol IV* or *pol V* mutant plants exhibit a longer termination zone. (A)** Schematics of the circular RT-PCR approach used for independent confirmation of transcript ends. **(B)** Circular RT-PCR results for *At5g24670* and *At5g36700*. Top panels show PCR products obtained using primers placed near transcription start sites (TSS). Bottom panels show both mature and 3'-extended PCR products. Primers for each PCR reactions are color-coded as in **(A)** and are shown on the right. **(C)** Sequencing results for cloned *At5g24670* PCR products (from **B**, bottom panel). Number of clones supporting transcript ends at the marked (in red) genomic positions are shown for Col-0, *nrpd1,* and *nrpe1* plants. Numbers on the right represent observed increase (in nt) in the length of the transcripts relatively to mature transcripts.

**Figure 4: post-TES read accumulation is more prevalent among convergent gene pairs. (A)** Histogram showing the percent of convergent genes among all protein-coding genes in *Arabidopsis thaliana* versus genes displaying increased post-TES read density in *pol IV* or *pol V* mutant plants. *P*-value was calculated using Fisher's exact test. **(B)** Histogram showing the percent of genes with increased post-TES read density in *pol IV* or *pol V* mutant plants that have the nearest neighbor gene initiating or ending at various distances (x-axes) from their transcript end sites (TES).

**Figure 5: Correspondence between genes with increased read density post-TES and RdDM pathway protein chromatin interaction. (A)** Venn diagram showing the number of genes with increased read density post-TES that exhibit enrichment (≥ 2 fold) for Pol IV, Pol V, or AGO4 within 2 kb regions centered at their TESs. **(B)** Pol V ChIP-seq read density plotted for genes with increased nascent RNA reads post-TES in *pol IV* or *pol V* mutant plants.

**Figure 6: Model for RNA Pol IV or Pol V involvement in RNA Pol II transcription termination.** Pol IV or Pol V may impede Pol II to promote transcript cleavage and release at Pol II convergent genes. In the absence of Pol IV or Pol V, Pol II undergoes head-to-head collision and pausing.

**Suppl Figure 1: Specificity of nascent biotinylated RNA enrichment on streptavidin beads.** Nuclear run-on (NRO) reactions were performed in triplicates with either biotin-UTP or with non-biotinylated UTP (as a negative control). Nascent RNA was captured on streptavidin beads, eluted and used in qPCR assays. The table shows averages of crossing point (Cp) values for several highly expressed *Arabidopsis thaliana* genes. Fold enrichment was calculated as the difference in Cp values between UTP samples and biotin-UTP samples: as Fold enrichment = $2^{\wedge}(Cp_{UTP}-Cp_{biotin-UTP})$.

**Suppl Figure 2: Sequencing data mapping statistics.** Acquired sequence reads were mapped to TAIR10 RefSeq. All non-unique reads were filtered to obtain high-quality sequence reads for metagene analyses.

**Suppl Figure 3: Pol II accumulates post-TES in *pol IV* mutant plants. (A)** RNA Pol II ChIP-exo reads are plotted relative to transcript end sites (TES) for genes with increased post-TES read density in *pol IV* or *pol V* mutants. **(B)** As in **(A)** but reads are plotted relative to TSSs.

**Suppl Figure 4: Genes with increased post-TES read density in *pol IV* or *pol V* mutants are highly expressed.** The graph compares levels of transcription for genes with increased post-TES read density in *pol IV* or *pol V* mutants and a random set of 1312 genes in Col-0 plants. Genes were sorted by the level of expression (x-axes).

**Suppl Figure 5: Increase in post-TES read density in *pol IV* or *pol V* mutants is independent of siRNA biogenesis. (A)** The Venn diagram shows the overlap between genes with increased post-TES read density and sRNA-associated genes. **(B)** Number of genes that show changes in 23-24nt sRNA read in *pol IV* mutant plants relatively to wild-type Col-0.

**Suppl Figure 6:** CG and CHH DNA methylation profile for overlapped set of 1312 genes with increased post-TES read density (by 2 fold or more) in mutant plants relative to wild-type Col-0 plants.

## Materials and Methods

*Plant material*

*A. thaliana* Col-0 and *nrpd1-3* and *nrpe1-11* have been described previously [9,11]. Plants were cultivated at 22°C under long-day conditions (16 h day/8 h night).

*Nuclei isolation*

Nuclei were isolated from 3 week old Col-0, *nrpd1-3* and *nrpe1-11* plant seedlings as described previously [44] with a few modifications. Briefly, 3 week-old plants were harvested and washed in ice-cold water. Then plants were submerged in cold diethyl ether for 5 min and washed twice in ice-cold water. Homogenization of plant material was performed in 100 mL of nuclei isolation buffer (1M sucrose, 50 mM Tris-HCl [pH 7.2], 5 mM MgCl, 5 mM KCl, 10 mM 2-mercaptoethnol, 0.2 mM PMSF) with a motorized homogenizer (Fisher Scientific Powergen 700). Homogenate was filtered through 1 layer of Miracloth and 1 layer of 50 um of nylon mesh. After centrifugation at 14,000 g for 15 min at 4°C the pellet was resuspended in 10 min of nuclei isolation buffer and prepared for loading into discontinuous Percoll gradient (15%, 30%, 45%, and 60%) by adding one volume of "100% Percoll solution" (34.23 g sucrose, 5 mL 1 M Tris-HCl [pH 7.2], 0.5 mL 1 M MgCl$_2$, 0.5 mL 1 M KCl, 34 uL 2-mercaptoethnol, 100 uL 2 mM PMSF and Percoll to 100 mL) to 19 volumes of nuclei. Gradients were centrifuged at 2,000 rpm for 10 min, then 8,000 rpm for 20 min. Nuclei were harvested from 30%/45% and 45%/60% boundaries and pooled. Nuclei were diluted with 5 volumes of nuclei isolation buffer, mixed by inversion and collected by centrifugation at 1,500 g, 10 min, 4°C. Second wash of nuclei was performed with 25 volumes of nuclei isolation buffer. Final resuspension of nuclei was in 1 mL of nuclei storage buffer (50 mM HEPES [pH 7.2], 5 mM MgCl$_2$, 5 mM KCl, 2 mM DTT, 0.2 mM PMSF, 50% glycerol). Aliquots were frozen in liquid nitrogen and stored at -80°C.

*Nuclear run-on (NRO)*

NRO reactions were performed as described previously [28] with a few modifications. Briefly, 10 *in vitro* transcription reactions (each containing190 uL of nuclei, 60 uL 5x Txn

buffer (100 mM Tris-HCl, [pH 7.7], 6 uL 0.1 M DTT, 5 uL RiboLock, 12 uL Biotin-16-UTP (10 mM), 16 uL AGC (10 mM each)) were set up in parallel for each plant line examined and incubated at 30°C for 60 min. Total RNA was extracted using Trizol method, treated with Turbo DNase I. An aliquot of total RNA was used for total RNA library contruction (see Materials and Methods). Biotinylated RNA was precipitated from total RNA with Dynabeads MyOne Streptavidin C1 (Invitrogen, Calsbad, CA). Binding was performed in 5mM Tris-HCl [pH 7.7], 0.5 mM EDTA and 1 M NaCl for 20 min at 42°C and 2 hours at room temperature with rotation. After washes, RNA was eluted from beads by adding 50 uL of 10 uM EDTA/95% formamide and incubation at 65°C for 5 min.

## RNA library construction

Briefly, total and eluted biotinylated RNA was fragmented and used for strand-specific libraries preparations and HiSeq PE50 sequencing.

## exoChIP library construction

Chromatin cross-linking, isolation and shearing were carried out as previously described with slight modifications [9]. Briefly, three grams of three-week old leaf tissue was suspended in 35 ml SH Buffer (100 mM sucrose, 20mM HEPES-NaOH) containing 1% formaldehyde and placed under vacuum for 55 minutes. Cross-linking was stopped by the addition of 1.25 ml 2M Glycine. Leaf tissue was washed three times with water. Tissue was ground in mortar with pestle and suspended in 15 ml Honda buffer (0.44M Sucrose, 1.25% Ficoll, 2.5% Dextran T40, 20mM HEPES-NaOH pH7.4, 10mM $MgCl_2$, 0.5% Triton X-100, 5mM DTT, 1mM PMSF, 1% Plant Protease Inhibitor Cocktail). Slurry was passed through Miracloth and centrifuged at 2,000 x g for 15 minutes at 4C. The pellet was resuspended in 1 ml Honda buffer and centrifuged as above. Again, the pellet was resuspended in Honda buffer, centrifuged as above and the pellet was resuspended in 575 microliters of Nuclei Lysis Buffer (50mM Tris-HCl pH8, 10mM EDTA, 1% SDS, 1mM PMSF, 1% Plant Protease Inhibitor Cocktail). Chromatin was sheared using a Bioruptor sonicator (Diagenode) to generate fragment sizes between 200 and 600 bp. Sheared chromatin was centrifuged for 10 minutes at 4C at 12,000 x g. The

supernatant was removed and diluted ten fold with ChIP Dilution Buffer (1.1% Triton X-100, 1.2mM EDTA, 16.7mM NaCl, 16.7mM Tris-HCl pH8). Aliquots containing approximately 6 mg of chromatin were shipped to Peconic, LLC for ChIP-exo Sequencing. ChIP-exo was carried out as described in [45]. Five microliters of the 4H8 monoclonal antibody (Cell Signaling Technology), which recognizes the Pol II CTD, was used for immunoprecipitation.

## Sequence processing

Reads were adapter trimmed and quality filtered (q20) using trimmomatic and then mapped to TAIR10 genome reference using tophat. Only read alignments mapped concordantly to non-ribosomal regions of nuclear chromosomes were selected for subsequent analysis. Number of reads uniquely mapped to each individual gene was obtained using htseq-count. The read counts between samples were normalized based on the total number of reads mapped for each sample.

## Read-through analysis

TES-termination index (TES-TI) was determined in the NRO and total RNA libraries by calculating ratios of read density 500 bp downstream of the TES to read density in the body of the gene 500 bp upstream of TES. For each gene, TES-TI was calculated in Col-0, *nrpd1-3* and *nrpe1-11* backgrounds.

## ChIP data processing

Among the annotated TAIR10 nuclear protein coding genes, pairs of consecutive genes with their 3' ends facing each other and separated by at least 1kb were identified as convergent genes. Among these, any gene pairs overlapping with miRNA or tRNA or snRNA or snoRNA or rRNA or other non-coding RNA were excluded, resulting in 1,453 gene pairs (2,906 genes) that were used in the analysis. For each of these 2,906 genes, coverage information at each base position starting from 1000 nt upstream of the 3' end into the gene body to 1000 nt downstream of the 3' end was obtained and total coverage contributed by all 2,906 genes at base position was computed and plotted.

## Circular RT-PCR

Total RNA was extracted using Trizol reagent. 10 ug of RNA was used for circularization with RNA ligase I (NEB, 10,000 U/uL) according to manufacture instructions. After incubation at 37°C for 1 hour, enzymes were heat inactivated at 65°C for 15 min and reactions were cleand up and concentrated with Zymogen kit. The resulted RNA was treated with Turbo DNase I, reverse-transcribed and used in two rounds of PCR amplification followed by PCR clean up (Quigen PCR clean up kit) and DNA cloning into pGEMT easy vector according to manufacturer's instructions. Individual colonies were grown and used for DNA extraction and sequencing.

## Data availability

Raw Illumina reads will be deposited to the NCBI Gene Expression Omnibus (GEO) upon acceptance for publication and available through http://www.ncbi.nlm.nih.gov/geo/.
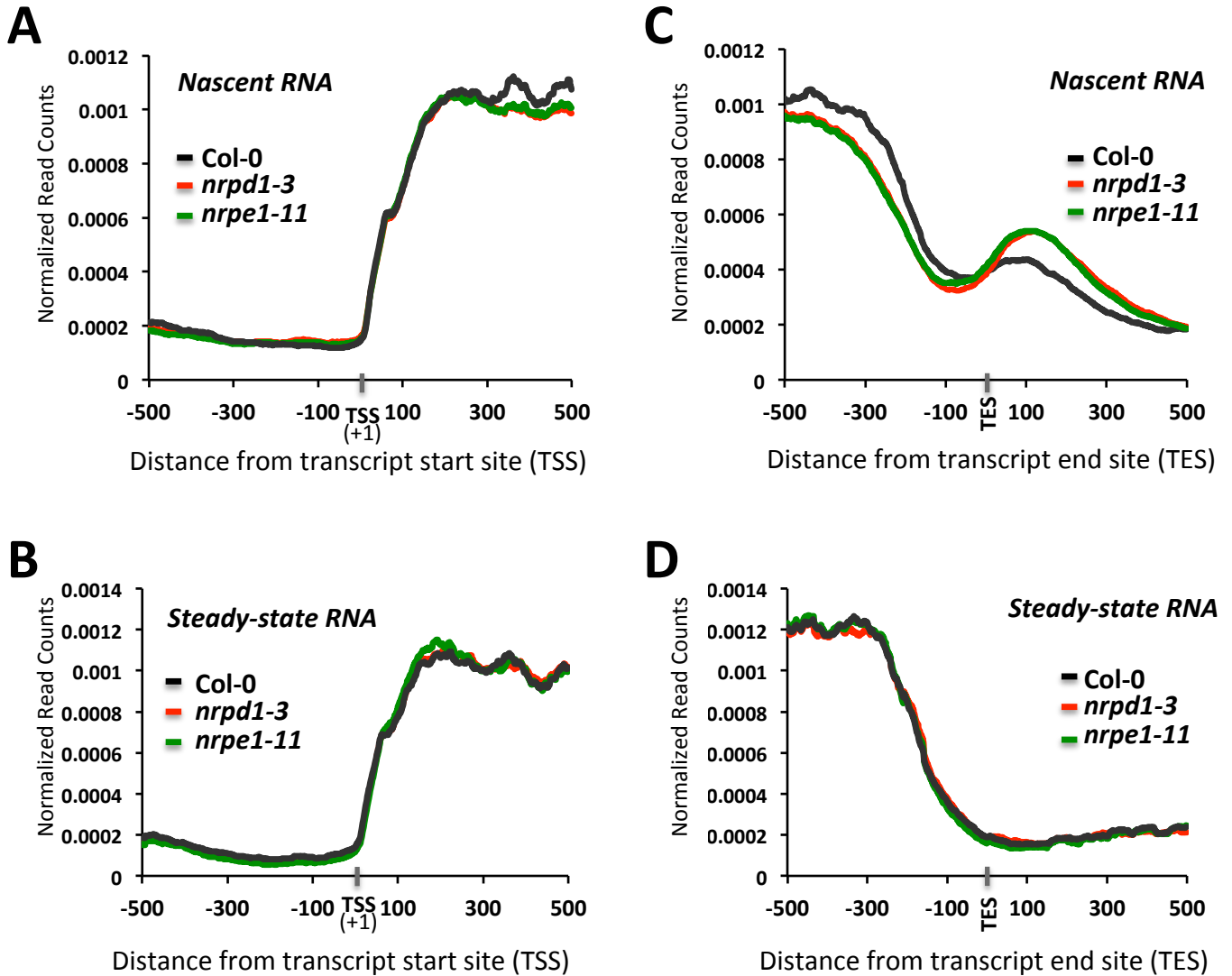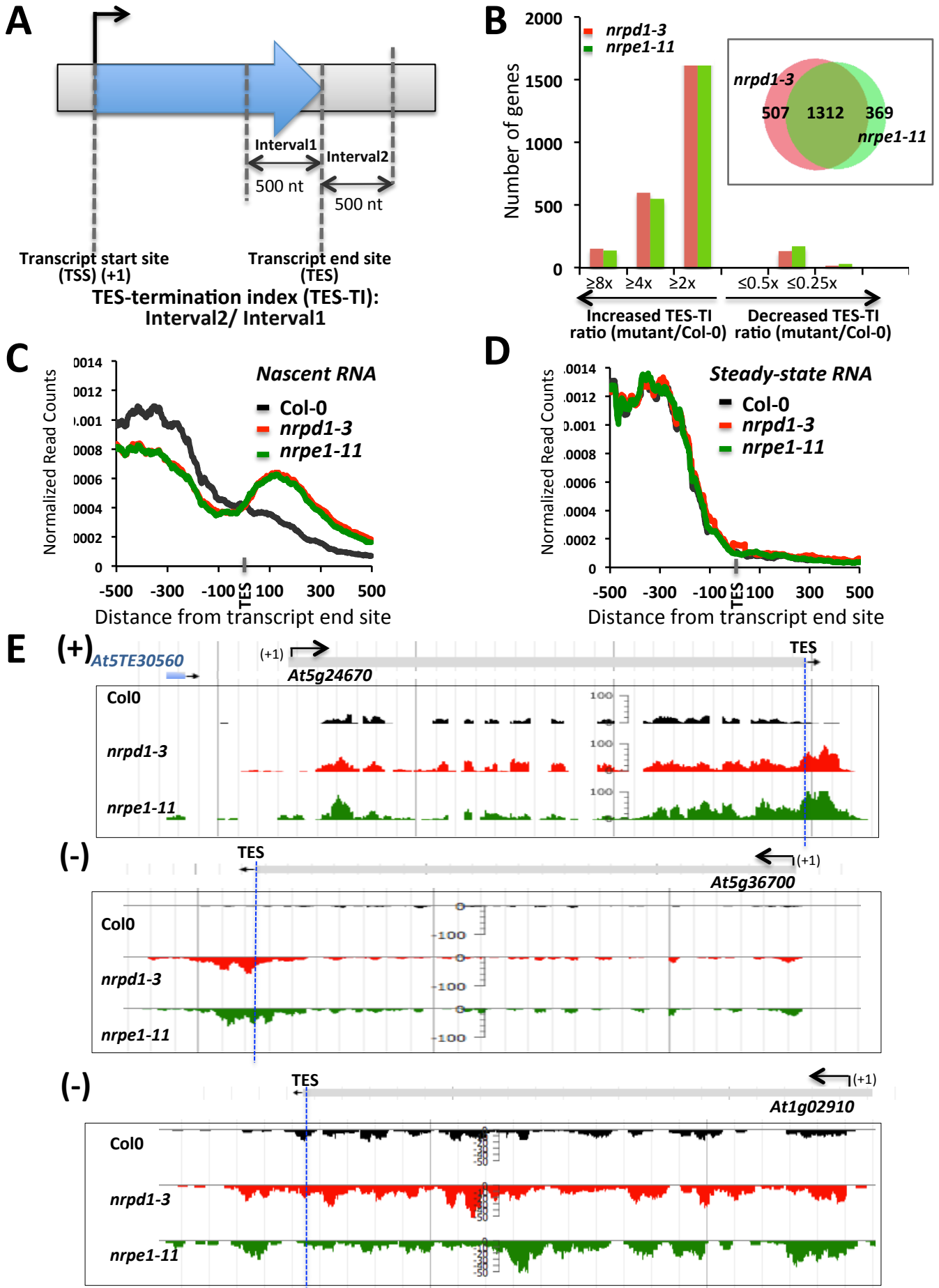
McKinlay et al. Figure 2
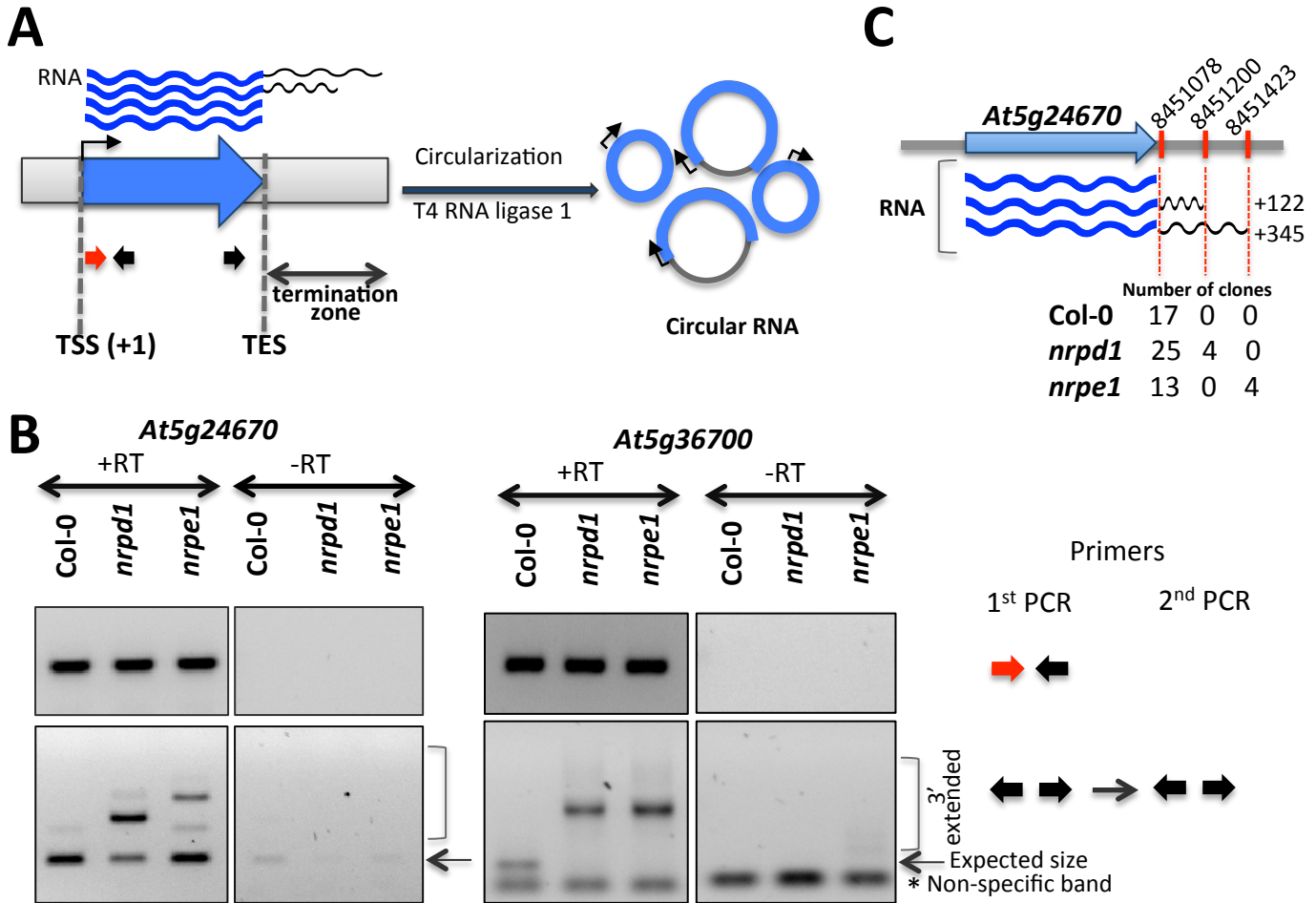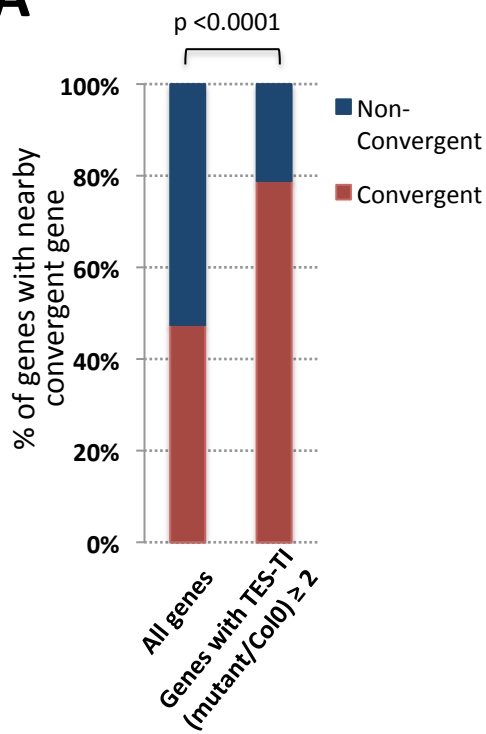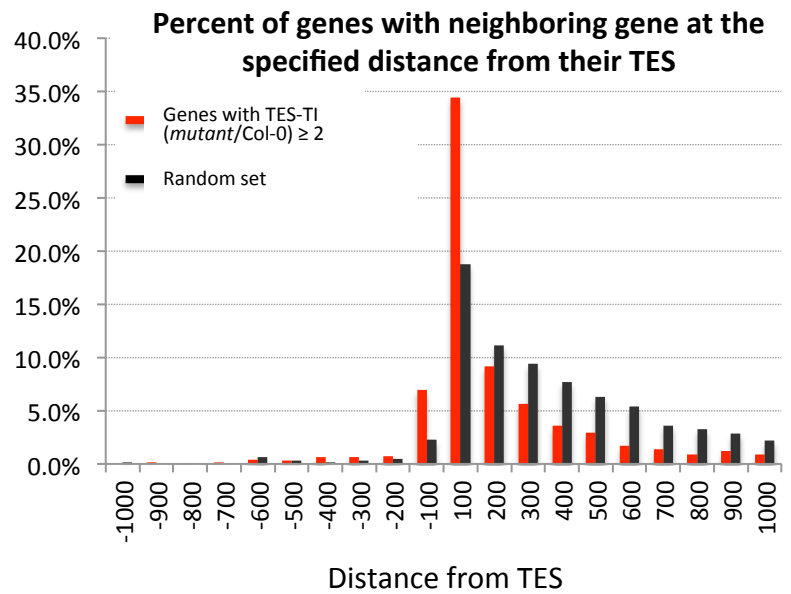
**A**



**B**



**C**



| | Number of clones | | |
|---|---|---|---|
| Col-0 | 17 | 0 | 0 |
| *nrpd1* | 25 | 4 | 0 |
| *nrpe1* | 13 | 0 | 4 |

**A**



**B**



Percent of genes with neighboring gene at the specified distance from their TES

**A**

1312 genes with TES-TI (*mutant*/Col-0) ≥ 2

Ago4 – associated
48

42

12    11

Pol V – associated
488

Pol IV – associated
80

106

**B**

*RNAP V ChIP*

Col-0
*nrpe1-11*

3000

2000

1000

0

Normalized Read Counts

-1000    -500    0    500    1000

Distance from TES

**Convergent**
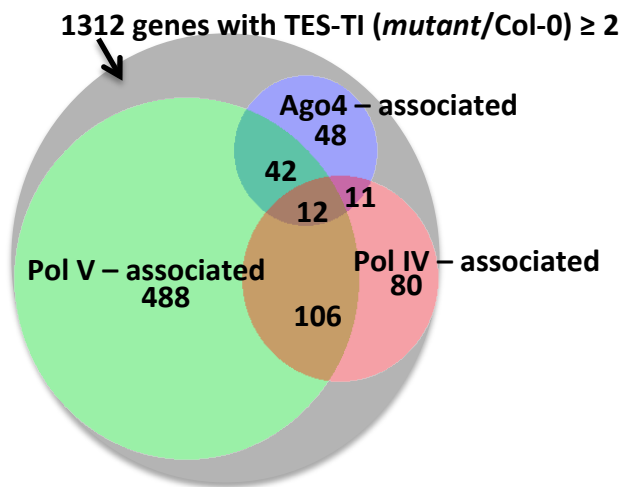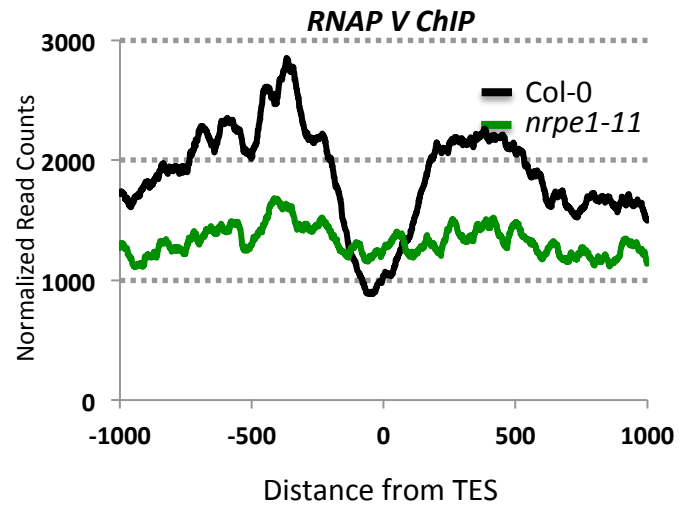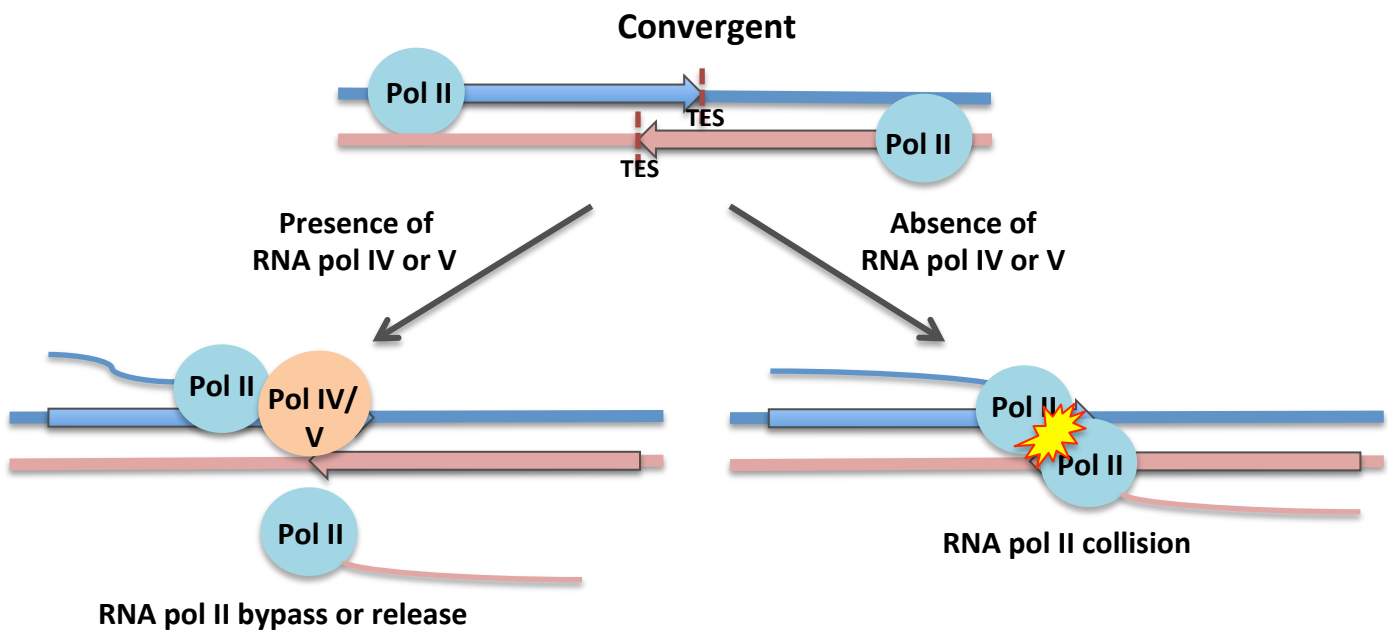
| GENE | $Cp_{biotin-UTP}$ | $Cp_{UTP}$ | Fold enrichment = $2^{(Cp_{UTP}-Cp_{biotin-UTP})}$ |
|------|------------------|-----------|-------------------------------------------------|
| 5'ETS (45S rRNA) | 20.2 | 26.9 | 100.43 |
| LHCB1.3 | 26.3 | 33.5 | 144.01 |
| PSAD1 | 28.8 | 39.6 | 1820.35 |
| LHCB6 | 30.1 | 38.3 | 292.71 |
| ACT2 | 33.1 | 39.6 | 478.82 |
| RBSC2B | 33.3 | 37.2 | 15.93 |
| AT2G45180.1 | 33.3 | 40.0 | 106.4 |
| DRT112 | 34.1 | 38.8 | 24.76 |

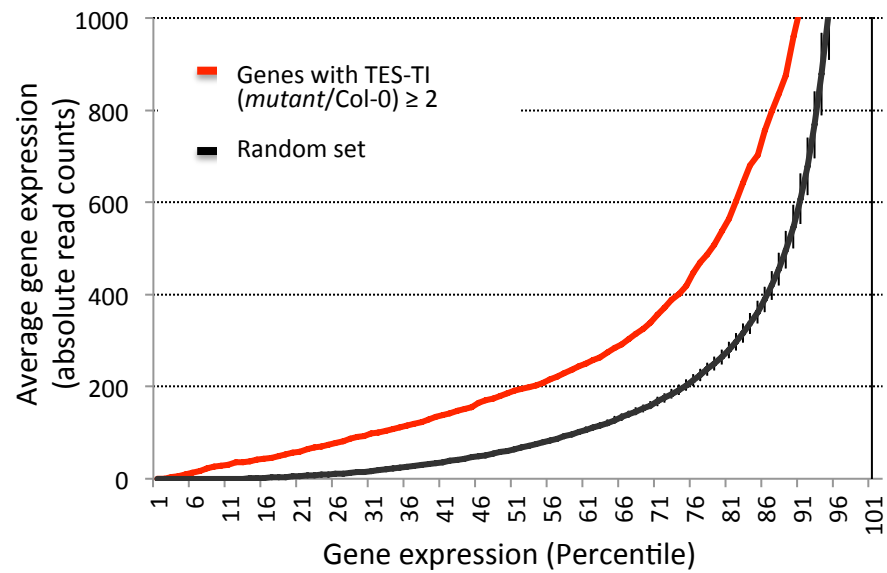Average Cp values correspond to the average cross-point values from triplicate NRO reactions

| | Library | Reads acquired | Reads mapped | Percent mapped | Unique, non-rRNA | Percent unique, non-rRNA |
|---|---|---|---|---|---|---|
| **Nascent RNA** | **Col-O** | 122,430,770 | 78,553,786 | 64% | 46,360,884 | 38% |
| | *nrpd1-3* | 120,989,174 | 84,478,400 | 70% | 59,447,048 | 49% |
| | *nrpe1-11* | 123,608,974 | 86,523,216 | 70% | 64,830,960 | 52% |
| **Nuclear RNA** | **Col-O** | 105,944,110 | 74,134,460 | 70% | 57,626,421 | 54% |
| | *nrpd1-3* | 75,280,875 | 53,325,738 | 71% | 44,609,286 | 59% |
| | *nrpe1-11* | 98,004,168 | 69,178,201 | 71% | 59,032,505 | 60% |

**A**

RNAP II ChIP-exo

- Col-0
- *nrpd1-3*

Normalized Read Counts

Distance from TES

**B**

RNAP II ChIP-exo

- Col-0
- *nrpd1-3*

Normalized Read Counts

Distance from TSS

**A**



TES-TI (mutant/Col-0) ≥ 2

922

147   187   56

1228   3844   1319

21-22 nt sRNA
-associated

23-24 nt sRNA
-associated

**B**



23-24 nt sRNA read counts ratios (Col0/*nrpd1-3*)