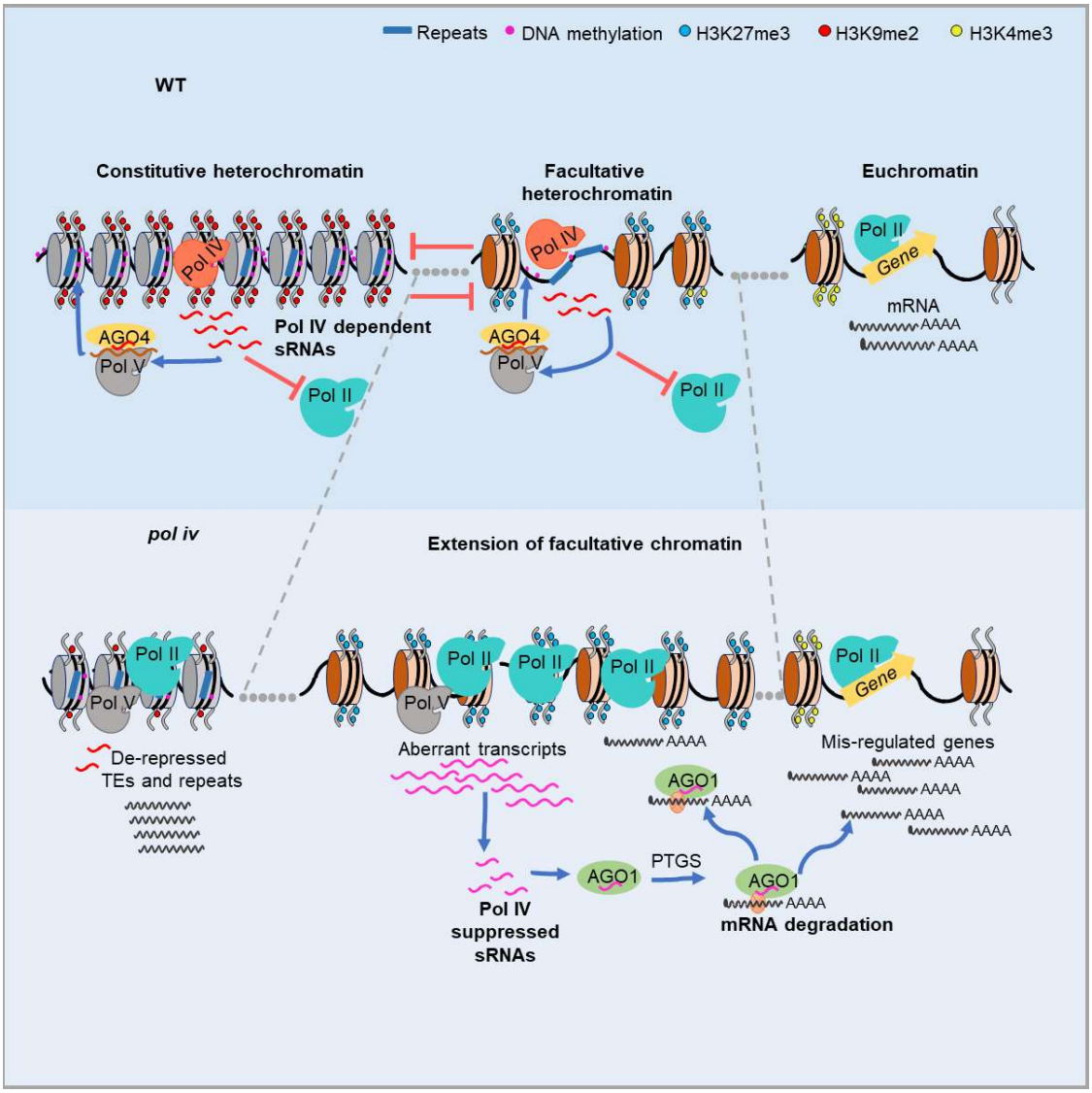
Plant Polymerase IV sensitizes chromatin through histone modifications to preclude spread of silencing into protein-coding domains

Vivek Hari Sundar G[a], Chenna Swetha[a], Debjani Basu[a], Kannan Pachamuthu[a], Tania Chakraborty[b], Rebecca A. Mosher[b], and P. V. Shivaprasad[a]*

[a]National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK Campus, Bangalore 560 065, India.

[b]School of Plant Sciences, The University of Arizona, Tucson, AZ 85721, USA.

(*Corresponding author, shivaprasad@ncbs.res.in)



One-line summary: RNA polymerase IV maintains chromatin states to influence regulation of protein coding genes.

Keywords: RNA silencing, epigenetics, histone modifications, small RNAs, RNA polymerase IV, Rice

Abstract

Heterochromatin is the predominant architectural feature of genomes that ensures genomic stability across eukaryotes. It mostly functions in restricting expression of repeats, mobile elements such as transposons and other regions. The establishment, maintenance and spreading of heterochromatin requires several factors including chromatin modifiers. However, how exactly heterochromatin formation is avoided in protein-coding domains is poorly understood. Here we show that a plant specific paralogue of RNA polymerase (pol) II, named pol IV, is involved in avoidance of facultative heterochromatic marks in protein coding genes, in addition to silencing the repeats and transposons forming constitutive heterochromatin. In its absence, H3K27 trimethylation mark intrudes the protein coding genes, more profoundly in genes embedded with repeats. In a subset of genes that lack the compensatory silencing, spurious transcriptional activity results in small(s)RNA production leading to post-transcriptional gene silencing. We show that such effects are significantly pronounced in rice, a plant with larger genome with distributed heterochromatin when compared to Arabidopsis. Our results indicate the surprising division of labour among plant-specific polymerases, not just in establishing effective silencing via small RNAs and epigenetics, but also in influencing chromatin boundaries.

Introduction

The genetic material DNA in eukaryotes is compacted as chromatin by means of association with histones and other proteins (Kornberg, 1974; Luger et al., 1997). The accessibility of the DNA for various molecular processes such as transcription, repair and replication is dynamically modulated in distinct domains of the chromatin, largely categorized into fairly accessible gene-rich euchromatin and condensed heterochromatin (Feng and Michaels, 2015; Brown, 1966). The chromatin is decorated with modifications to DNA and histones constituting epigenetic modifications (Feng and Michaels, 2015; Law and Jacobsen, 2010). Proportion of genomic regions coding for proteins, i.e., euchromatin reduces with increase in genome size across organisms and such alterations and genome expansion is mainly due to the proliferation of repeats and transposons (Pellicer et al., 2018). Plant genomes are especially enriched with latent transposons poised for transcription located proximal to the protein coding genes in the euchromatic domains (Hirsch and Springer, 2017).

The activity of RNA polymerase II (pol II) on the euchromatic domains is regulated locally by transcription factors, chaperones and chromatin remodelling enzymes that transduces the local epigenetic status (Schier and Taatjes, 2020; Hahn, 2004; Gibney and Nolan, 2010). Evolution of plants with increased instances of gene-proximal repeats in the genome necessitates articulation of silencing states to the pol II with precision. This locally silenced state is called facultative heterochromatin and is enriched with Polycomb Repressive Complex (PRC) dependent H3K27me3 marks in contrast to the constitutive heterochromatin enriched with H3K9me2 modifications (Zhang et al., 2007). Studies in early land plant genomes like that of *Marchantia polymorpha* suggest that H3K27me3 evolved as a predominant silencing mark in plants (Montgomery et al., 2020) and H3K9me2 mark took over the constitutive silencing leading to the observed dichotomy between facultative and constitutive heterochromatin (Déléris et al., 2021). Mutants of MET1, a major player in constitutive heterochromatin establishment, exhibited compensation by H3K27me3 marks at repeats and transposons (Deleris et al., 2012; Soppe et al., 2002; Mathieu et al., 2005), indicating that unknown players monitor heterochromatin states in specific domains to initiate compensatory marks in constitutive heterochromatic regions. Demarcation of facultative and constitutive heterochromatic boundaries is paramount in avoiding the intrusion of silencing states into neighbouring protein coding genes, warranting evolution of novel machineries curtailing silencing overshoot and simultaneously defending against genotoxic repeats. Although these mechanisms have been envisaged across eukaryotes, major upstream players involved in these beyond the obvious epigenetic readers, writers and erasers, are unknown.

Plants have evolved RNA silencing as an efficient mode of robust and targeted silencing of repeats and genes, both at transcriptional and post-transcriptional levels (Baulcombe, 2004). Small RNAs (sRNAs), predominant effectors of plant RNA silencing, confer both specificity and amplification modality in silencing. Production of sRNAs associated with transcriptional silencing are primarily

initiated by plant specific RNA polymerase II (pol II) paralog RNA polymerase IV (pol IV) and in peculiar cases by pol II itself (Nuthikattu et al., 2013; Cuerda-Gil and Slotkin, 2016). The pol IV transcripts, majorly originating from the repeats and transposons are acted upon by RNA DEPENDENT RNA POLYMERASE2 (RDR2) while the pol II aberrant transcripts are acted upon by RNA DEPENDENT RNA POLYMERASE6 (RDR6), thereby converting them to double stranded duplexes that become substrates for several DICER-LIKE proteins (DCLs) (Nuthikattu et al., 2013). The resultant short duplex sRNAs are picked up by specific ARGONAUTE (AGO) effectors, specified by their length and one of the strands possessing the preferred 5'-nucleotide (nt) (Mi et al., 2008). For post transcriptional gene silencing (PTGS), majorly undertaken in plants by cleavage of target mRNA, 21-22nt size class sRNAs with 5'-Uracil (U) are loaded into the AGO1 leading to slicing of the target mRNA between positions 10 and 11 in the sRNA:mRNA duplex regions (Baumberger and Baulcombe, 2005). On the contrary, for transcriptional silencing pol IV derived repeat associated 24nt sRNAs are loaded into AGO4 (Zilberman et al., 2003) with a 5'-Adenosine (A) preference and this ribonucleoprotein complex binds to the complementary region of the long scaffold transcript produced by RNA polymerase V (pol V) (Wierzbicki et al., 2008). This triggers recruitment of DNA methyl transferases like DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2) and majorly result in asymmetric DNA methylation at the target locus (Matzke and Mosher, 2014). Several other pathways culminating in production of sRNAs, for example, viral transcription, transgene induced and endogenous aberrant transcription, the pol II-RDR6 aberrant activity feedback to the similar silencing pathway with identical end results (Elmayan and Vaucheret, 1996; Herr et al., 2005; Voinnet et al., 1998; Nuthikattu et al., 2013). Numerous non-canonical modalities in the concerted activity of these RNA polymerase variants have also been reported (Cuerda-Gil and Slotkin, 2016).

Even though RNA silencing via sRNAs is an effective process to accurately delimit silencing, the loci over which different polymerases operate are dependent on the chromatin states. For instance, the pol IV is recruited by the readers of H3K9me2 marks named SAWADEE HOMEODOMAIN HOMOLOG1 and 2 (SHH1 and 2) in coordination with CLASSY type chromatin remodellers (Law et al., 2013; Zhang et al., 2013; Zhou et al., 2018). The pol V recruitment involves DNA methylation readers like SUVH2 and SUVH9 (Liu et al., 2014), chromatin remodelling complex called as DDR complex (Law et al., 2010) and several other proteins that are modulated by chromatin states. The activity of these polymerases are tightly linked to the epigenetic status of the locus of interest and hence the RNA silencing. Few reports have probed the effects of loss of DNA methylation machineries on the chromatin states in Arabidopsis and observed excessive reorganisation of chromatin domains. For example, in *met1* mutants chromatin is decondensed with loss of clear heterochromatin and euchromatin boundaries, exemplifying the impact of the cross talk between these epigenetic states (Soppe et al., 2002; Mathieu et al., 2005; Zhong et al., 2021). Similarly, *pol iv* exhibits permanent loss of silencing at selected loci that are recalcitrant to rescue by complementation of pol IV and this phenomenon is due to associated loss of H3K9me2 marks (Li et al., 2020). Unrestricted to silencing, it is also evidenced that loss of pol IV leads to transcriptional mis-regulation as observed by the accumulation of atypical nascent transcripts in maize (Erhard et al., 2015) and by increased pol II activity at 3'-ends of pol II transcriptional units (McKinlay et al., 2018). The indirect facilitation of spurious transcription by virtue of loss of epigenetic players triggers cryptic transcription units by pol II both in genes and transposons (Le et al., 2020). Prolific presence of repeat fragments in the gene coding units like introns in rice have shown to trigger transcriptional silencing like features (Espinas et al., 2020). Taken together, epigenetic landscape modulates not only the silencing of repeats directly but also encumbers the cryptic transcriptional activity.

Such counter-balancing reinforcement loops between epigenetic states must contribute to molecular and morphological phenotypes upon perturbation, especially in monocots with higher proportion of transposons. It has been conclusively documented that reproductive structures in rice, including gametes, undergo massive reprogramming in terms of sRNA production and DNA methylation in a locus-specific manner (Chenxin et al., 2020). Indeed, unlike Arabidopsis, loss of silencing players display exacerbated phenotypes in the monocot model rice in the cases of *drm2*, *met1b, pol iv, pol v, dcl3* (Moritoh et al., 2012; Yamauchi et al., 2014; Zheng et al., 2021; Wei et al., 2014). In agreement with this, grass family (*Poaceae*) members have evolved specific neo-functionalized RNA polymerase paralog called RNA polymerase VI, functions of which are still unclear (Trujillo et al., 2018). These

phenomena substantiate that complex gene arrangement interleaved by repeats mandates very robust and agile mechanism of epigenetic silencing in a very localised manner.

In this study, we resorted to examine the genome and phenome level functions of RNA pol IV in rice by generating precise knockdown (*kd*) mutants and validated them by genetic complementation. We present evidence for over-shoot of heterochromatic H3K27me3 marks into the protein coding regions in the *kd* and loss of H3K9me2 signals genome-wide. A subset of loci that do not undergo H3K27me3 mediated substitution became hotspots of aberrant sRNA production by other RNA polymerases. We show that these sRNAs from atypical loci in *kd* are present in loss of function mutants of Arabidopsis counterparts as well. We provide evidence that these sRNAs are indeed functional in PTGS of protein coding genes more specifically in larger genomes such as rice. Together, these results indicate the presence of several novel features of regulation of plant genomes, including complementation of H3K27me3 in constitutive heterochromatin, role of pol IV in preventing facultative heterochromatinization of protein-coding genes and a strong negative selection on pol IV-sRNA modules, absence of which can trigger PTGS of important genes in a species-specific manner.

Results

Knockdown of RNA polymerase IV induced pleiotropic phenotypic defects

The loss of function of the catalytically active component of the polymerase IV complex, NRPD1, causes a spectrum of effects in different plant species studied – from delayed growth transition in *Physcomitrella patens,* delayed flowering in *Arabidopsis thaliana,* reproductive defects in *Brassica rapa, Capsella rubella and Zea mays,* to increased tillering in *Japonica* rice (Coruh et al., 2015; Wang et al., 2020c; Kf et al., 2009; Grover et al., 2018). In order to understand the multifaceted roles of polymerase IV in particular and RdDM in general in rice, we generated artificial microRNA (amiR)-mediated knockdown (*kd*) transgenic lines targeting the largest subunit of RNA pol IV (NRPD1). Since rice genome encodes two isoforms of NRPD1 – NRPD1a and NRPD1b, sharing 88.8% of sequence identity, amiR was designed to target a conserved sequence in the N-terminal region of both the transcripts, incorporating optimal design parameters as put forth earlier (Fig. 1A) (Narjala et al., 2020). Efforts have already been made to target both the rice NRPD1 isoforms using RNA interference targeting the C-terminal region as well as the CRISPR-Cas9 mediated knockout of both the isoforms, the latter was reported as embryo lethal (Xu et al., 2020; Debladis et al., 2020; Zheng et al., 2021). In order to get a precise targeting and to avoid the lethality of knockout, we resorted to the amiR technology. *Agrobacterium*-mediated transformation was performed twice independently to score for consistent phenotypes and to eliminate T-DNA insertion effects. The expression of mature amiR was confirmed with northern hybridisation and independent transgenic events were identified by junction fragment Southern analyses (Supplemental Fig. 1A and 1B). The knockdown lines showed high reduction in transcript abundance of both the isoforms (Fig. 1B).

As observed in other studies (Xu et al., 2020), the *kd* plants showed increased tiller number (Fig. 1C and 1D). Strikingly, the percentage of viable filled grains were significantly low at 75% in T1 *kd* plants and diminished progressively over generations to reach 50% in T3 (Fig. 1E and 1F). The unfilled florets showed structural deformities and they were consistent among the distinct transgenic events proving that the defects observed were not due to transgenesis (Supplemental Fig. 1C, 1D and 1E). Specific lineages of the *kd* lines had extreme reproductive defects without viable grains (Supplemental Fig. 2A). Southern and northern analyses showed these defects are not due to the zygosity of the T-DNA or the dosage of amiR in such a lineage indicating possible effect of induced epimutations (Supplemental Fig. 2B and 2C). Since NRPD1 is reported to influence pollen development (Wang et al., 2020c), we further investigated the pollen quality by measuring size, iodine staining potential and ultrastructural characterisation using scanning electron microscopy. These revealed that compared to WT, the *kd* pollen showed structural defects, reduced size and viability (Fig. 1G and Supplemental Fig. 3A, 3B and 3C).

Taken together, amiR mediated *kd* of RNA polymerase IV complex in rice led to severe reproductive defects consistent across generations in distinct transgenic lines similar to monocots such as maize (Kf et al., 2009), unlike *Arabidopsis*.

Rice pol IV is required for biogenesis of repeat associated sRNAs

In order to explore the molecular effects of loss of polymerase IV in rice and to discern the reasons for extreme reproductive defects upon NRPD1 *kd*, we performed total sRNA profiling in three reproductive tissues that exhibited drastic defects – pre-emerged panicle, anther and endosperm. As shown earlier in other species, *kd* plants showed evident reduction of 24nt sRNAs in all the tissues (Fig. 1H and 1I, Supplemental fig. 4A) (Herr et al., 2005; Wang et al., 2020c; Coruh et al., 2015; Grover et al., 2018). As expected (Mi et al., 2008; Zhai et al., 2015), there was loss of 5'-A containing sRNAs in all the three tissues (Fig. 1H). On the contrary, bulk of the pol II transcribed miRNAs and the sRNAs mapping to their precursors remain largely unchanged (Supplemental Fig. 4B and 4C). Further, as the pol IV derived sRNAs are known to be associated with transposons and repeats, we profiled the sRNAs from major annotated repeats in the rice genome. We observed substantial loss of repeat associated sRNAs and this was also validated by northern blotting (Fig. 1I and Supplemental Fig. 4D, 4E and 4F). As shown earlier, the increased tillering was linked to over-accumulation of microRNA156 and this was observed in our investigation as well (Fig. 1I) (Xu et al., 2020).

In order to substantiate that the loss of sRNAs is directly linked to NRPD1 *kd* and to unequivocally prove that amiR is targeting only NRPD1 leading to the effects seen, we super-transformed the *kd* calli with a construct expressing amiR-targeting resistant version of NRPD1b CDS driven by its cognate promoter (Fig. 2A). It is already established that NRPD1a and NRPD1b are likely functionally redundant (Xu et al., 2020). The transgenic calli expressing both resistance markers – *HPTII* and *BlpR* retained 75% of WT *NRPD1b* transcripts depicting successful rescue (Fig. 2B and Supplemental Fig. 5A). Tos17 expression was also restored back to WT levels in the complemented lines (Fig. 2B and Supplemental Fig. 5A). In addition, the repeat associated sRNAs were also restored to WT levels in the complementation lines as proven by northern blots (Fig. 2C). In conclusion, RNA polymerase IV is essential for production of repeat associated sRNAs in different tissues of rice and genetic complementation of one of the two isoforms of NRPD1 rescues the molecular phenotypes.

RNA pol IV counters transposon proliferation and indirectly influence gene expression

Given the severity of the defects found in the rice *kd* plants, we hypothesised that the loss of sRNA mediated regulation of transposons and the concomitant mis-regulation of genes might be attributable to the defects. In order to test this we performed transcriptome profiling in panicle and anther, and this analysis showed mis-regulation of genes, both up and down regulated, in both the tissues (Fig. 2D). *kd* panicle showed upregulation of 2302 genes and downregulation of 1638 genes, while anther showed upregulation of 2655 genes and downregulation of 2762 genes with significant 4-fold difference in expression (Fig. 2D). In addition, as in Arabidopsis (Herr et al., 2005), expression of repeats increased in the *kd* plants when compared to WT (Supplemental Fig. 5A and 5B). Ribosomal RNA (rRNA) expression is regulated dynamically by sRNAs and other epigenetic players (Earley et al., 2010; Blevins et al., 2009). We performed northern hybridisation with probes for different regions of the precursor fragments (as described in Hang et al., 2018) of rRNAs and found that rDNA locus was heavily over-transcribed in *kd* plants when compared to WT (Fig. 2E). The DNA methylation levels profiled by whole genome methylome analyses over all the transcripts (genes and transposons) revealed that the DNA methylation in CHH context is reduced, in accordance with the proven roles of NRPD1 derived sRNAs in mediating asymmetric DNA methylation (Fig. 2F) (Onodera et al., 2005).

The upregulation of repeats and loss of DNA methylation prompted us to question if the transposons have gained proliferative potential upon pol IV *kd*. This possibility was previously predicted to influence phenotypes and genomic integrity in a number of reports (Debladis et al., 2020; Cui et al., 2013; Wei et al., 2014). To verify such a possibility, we performed a PCR-based assay to detect the extrachromosomal circular DNA (ECC DNA) intermediates that are produced as by-products of transposition (Lanciano et al., 2017). We observed additional bands portraying increased proliferative potential of *PopRice and Tos17* transposons (Supplemental Fig. 5C). These bands were observed even after the removal of linear DNA using a specific exonuclease (Supplemental Fig. 5D). This observation is in line with a previous report where silencing of NRPD1 caused an increase in transposon copy numbers (Debladis et al., 2020). We examined if specific transposons proliferated by performing transposon display Southern hybridisations. Among the *kd* lines tested, specific line showed discernible copy number increase of LINE1 element across generations (Fig. 2G). This

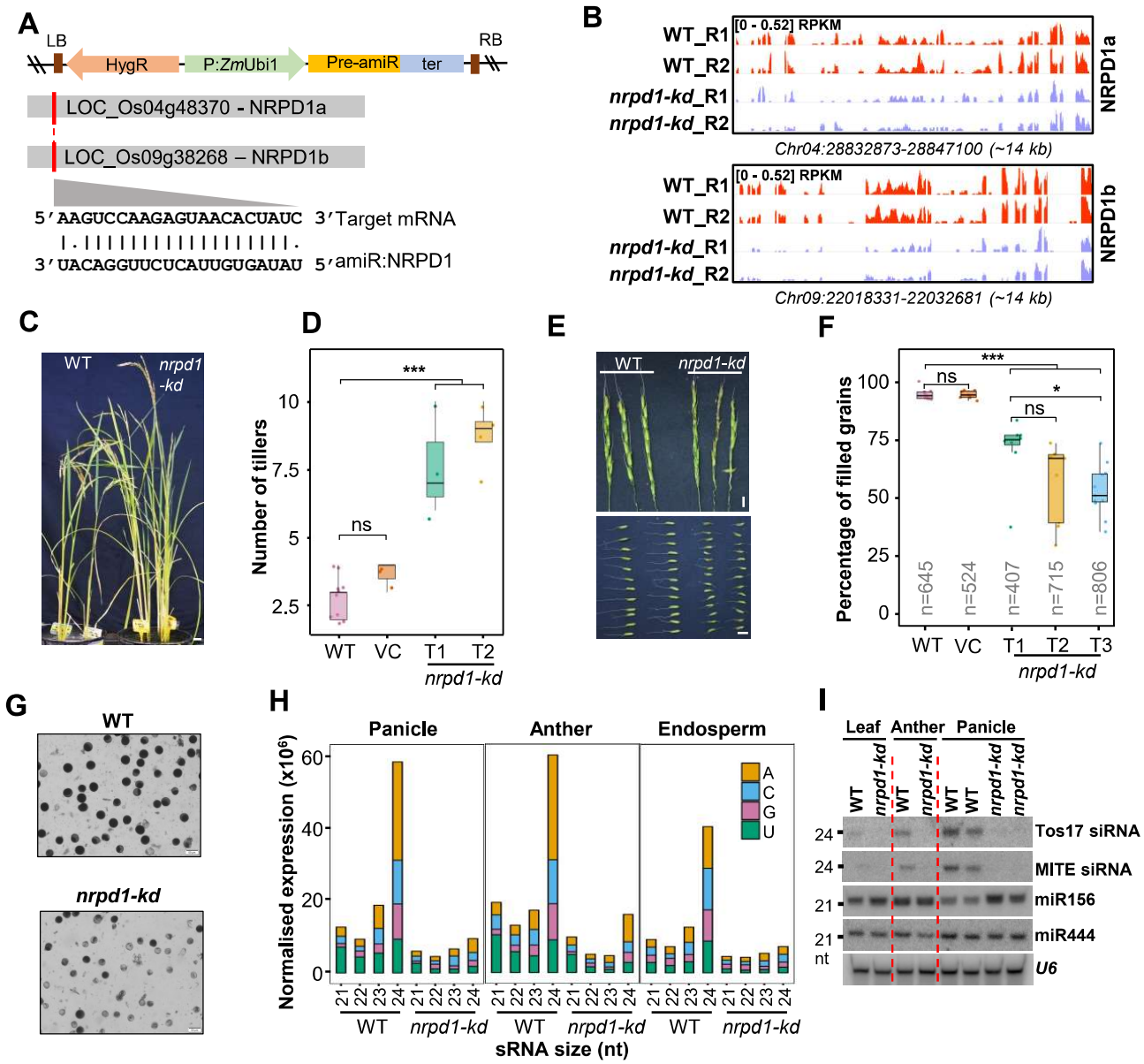


Figure 1. Knockdown of RNA pol IV in rice results in pleiotropic phenotypes.
(A) T-DNA map of amiR binary construct. The targeting regions of NRPD1a and NRPD1b is shown with the amiR sequence. The precursor-amiR (Pre:amiR) is driven by maize ubiquitin promoter (P:*Zm*Ubi1). ter – 35S-poly(A) signal; HygR – Hygromycin selection marker; RB and LB – right and left border.
(B) IGV screenshots of panicle RNA-seq coverage over the NRPD1a and NRPD1b loci in WT and *kd.*
(C) Images of the *nrpd1-kd* plants with equally grown WT controls. Scale bar:1 inch.
(D) Boxplots of number of tillers observed in WT, vector control (VC), T1 and T2 generations of *nrpd1-kd* plants. Dots represent result of each plant (Tukey's test; ***p-value of 0.001,: *p-value of 0.01, ns-non-significant).
(E) Images showing the panicle (top) and the individual florets (bottom) of WT and *nrpd1-kd* plants. Scale bar:1 cm.
(F) Boxplots showing the percentage of filled grains (n=number of grains). Dots represent average of panicles in each plant (Tukey's test ; ***p-value of 0.001, *p-value of 0.01, ns-non-significant).
(G) Representative images of pollen grains of dehisced anther stained with iodine. Scale: 50 µm.
(H) Stacked bar plots showing normalised abundance of sRNAs of different sizes with their 5' nucleotides. The replicates were merged.
(I) sRNA northern blots indicating loss of 24nt siRNA in different tissues. miR444 and miR156 were examples of miRNAs. *U6* was used as loading control.

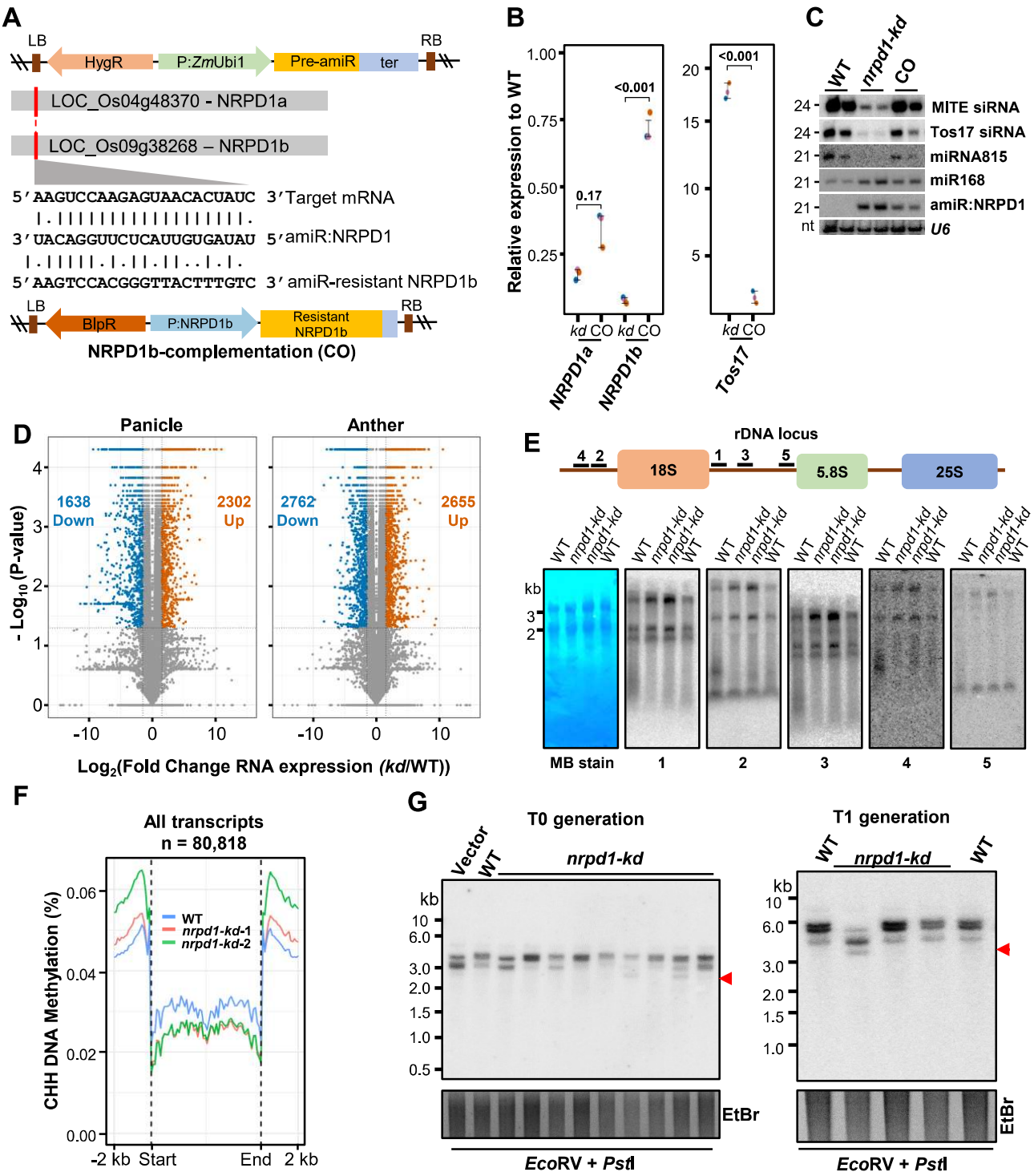


Figure 2. RNA pol IV is responsible for repeat associated sRNA production, DNA methylation and also contributes to gene regulation.

(A) T-DNA map of amiR coding binary construct as in Fig. 1A with the super-transformed NRPD1b complementation (CO) construct. The alignment depicts the modifications in miR-resistant CDS of NRPD1b that was driven by its cognate promoter (P:NRPD1b). BlpR – bialaphos resistance.

(B) Plots representing relative abundance of *NRPD1a, NRPD1b* and *Tos17* transcripts in *kd* and CO with respect to WT, measured by RT-qPCRs.

(C) sRNA northern blots showing the accumulation of 24nt siRNAs and miRNAs in WT, *kd* and CO. *U6* was used as loading control.

(D) Volcano plots showing the number of significantly upregulated ($Log_2FC > 2$) and downregulated ($Log_2FC < 2$) genes in *kd* compared to WT in panicle and anther.

(E) Northern blots showing the abundance of rRNA precursors from WT and *kd* panicle. Methylene blue (MB) stained membrane is shown as loading control.

(F) Metaplots showing CHH DNA methylation levels at all transcribed regions in WT and two *kd* panicles.

(G) Transposon copy number Southern blot showing copy number variation (arrowa) of LINE1 transposon in both T0 and T1 generations in *kd*. Ethidium bromide (EtBr) staining served as control.

lineage specific effects on transposon copy number is shown by a previous study as well (Debladis et al., 2020), and this might be the reason for exacerbated reproductive defects in selected *kd* lines (Supplemental Fig. 2A) . In conclusion, the activity of RNA polymerase IV and ensuing sRNAs are essential for regulating expression of genes and repeats via DNA methylation and loss of these processes result in potentially genotoxic proliferation of transposons in the genome.

Activity of RNA pol IV potentiates silencing by H3K9me2 and H3K27me3 and modulates their relative occupancy over genes and repeats

Several reports have suggested that the effective silencing is brought about by the cross talk between DNA methylation and histone modifications in a locus specific manner in plants (Mathieu et al., 2005; Deleris et al., 2012; Gent et al., 2014; Zhou et al., 2016). These reports have found that loss of DNA methylation led to perturbation of histone Post Translational Modifications (PTMs) at the respective loci and as a whole manifests as transcriptional silencing status. Recent studies in Arabidopsis have suggested that loss of pol IV and associated sRNAs can influence the histone H3 PTMs levels at specific loci (Li et al., 2020; Parent et al., 2021). In order to understand the impact of loss of pol IV on other epigenetic marks, we profiled the histone modifications H3K4me3, H3K9me2, H3K27me3 and total H3 in WT and *kd* lines using chromatin immunoprecipitation (ChIP). We characterized the occupancy of these histone marks relative to the total H3 over the protein coding genes (PCGs) and transposons and repeats. Over the PCGs, loss of H3K9me2 marks was substantial while the H3K4me3 levels remained unchanged in *kd* lines (Fig. 3A). But strikingly, we observed a dramatic change in the profile of H3K27me3 occupancy over the PCGs in *kd* lines that is drastically different in terms of quantity as well as the spatial occupancy over the gene bodies (Fig. 3A). Selective accumulation of H3K27me3 at the start site is noteworthy and this might likely be a compensatory mechanism acting at the protein coding genes by an unknown mechanism of genome protection. The repeats of different kinds are known to be maintained in a repressed state via H3K9me2 marks and rDNA repeats are one amongst them (Lopez et al., 2021). The increased expression of rDNA locus (Supplemental Fig. 5C) appears to be manifested by loss of H3K9me2 mark. Indeed, the occupancy of H3K9me2 marks were diminished greatly over the transposons and repeats (Fig. 3B). On the contrary, we did not see any such compensation by the H3K27me3 marks over these regions (Fig. 3B). This is in line with the observation that the silencing by H3K9me2 and H3K27me3 are compartmentalized, former majorly targeting the transposons while the latter associating with the PCGs (Déléris et al., 2021). Since monocot genomes are relatively enriched with repeats and they can influence the proximal genes (Hirsch and Springer, 2017), we picked the set of PCGs that have at least 10% of their length overlapping with the annotated repeats ('Genes with repeats'), and explored the perturbation of the histone marks. As expected, the presence of the repeats in the PCGs resulted in an enhanced degree of loss of the H3K9me2 marks and a proportional compensation of H3K27me3 marks (Fig. 3C).

Further to validate the redistribution of these marks upon *kd*, we resorted to nuclear immunostaining. This substantiated that the H3K9me2 marks were indeed reduced globally except for darkly stained chromocenters and evidently H3K7me3 marks were seen to be evenly over-distributed in *kd* (Fig. 3D and 3E). In essence, as seen in other plants, we observed a significant role of pol IV in maintaining the ideal occupancy states of the silencing histone marks H3K9me2 and H3K27me3 over both protein coding genes and the repeats in rice. Surprisingly, we also found that the mechanistically demarcated histone mark H3K27me3 could potentially substitute the loss of regulation by H3K9me2.

Loss of pol IV alters chromatin permissiveness for transcription by other RNA polymerases

Since the loss of pol IV caused a drastic loss of silencing marks, we hypothesised that relaxed chromosome state must be poised extensively for upregulated transcription by other polymerases. In order to examine this, we performed ChIP-seq to profile the occupancy status of RNA polymerase II in WT and *kd* lines. We observed a marked increase in occupancy of the pol II over the PCGs and repeat regions (Fig. 4A) matching previous observations (Zheng et al., 2009). But over the PCGs, in spite of the higher occupancy of the pol II, we saw an evident drop in the pol II levels near the start sites (Fig. 4A). This encumbered occupancy of pol II at the steady state mirrored exactly the same 5'-end of the genes that over-accumulated the H3K27me3 marks upon loss of pol IV (Fig. 3A and 3C). This trend validates that the H3K27me3 compensation is indeed having a control over the

transcription at the PCGs, consequently aiding in abrogating aberrant transcription. On the other hand, over the transposons and repeats, absence of the H3K27me3 compensation resulted in unhindered over-occupancy of pol II in *kd*. This is likely the reason for upregulated transcription of transposon and repeat regions (Supplemental Fig. 5A, 5B, 5C and 5D).

Further to substantiate the observation that aberrant polymerase transcription is increased upon the loss of silencing chromatin signatures, we analysed datasets of pol II and pol V ChIP-sequencing from Arabidopsis inflorescence tissues in Col-0 and *nrpd1* backgrounds (Zhai et al., 2015). Similar to rice, the loss of pol IV indeed caused a relaxed chromatin marked by increased pol II occupancy. In addition, pol V also over-transcribed the annotated transposons and repeats (Fig. 4B). On the contrary, there was no marked difference in the occupancy of pol II and pol V over the PCGs in Arabidopsis unlike rice. Possibly this is due to the clear demarcation of euchromatin and heterochromatin and fewer instances of repeat rich regions in proximity of PCGs when compared to rice as observed previously (Espinas et al., 2020). This might also be a reason for the extensive mis-regulation of numerous PCGs in rice (Fig. 2D) when compared to Arabidopsis (Panda et al., 2020) and might explain the absence of stronger phenotypes including reproductive defects in Arabidopsis *nrpd1*. Further, to investigate the roles of pol IV derived sRNAs directly being the cause of the reduction in H3K9me2 or the over-occupancy by pol II, we analysed the pol II occupancy from *dcl234* datasets. Very surprisingly, we find that the complete loss of pol IV derived mature sRNAs in *dcl234* (Li et al., 2015), did not impact the pol II occupancy and the profile mirrored the Col-0 status (Fig. 4B). In conclusion, these observations indicate that the pol IV transcription modulates the chromatin in a manner that directs the activity of other polymerases most notably, II and V. This activity is independent of sRNAs and the effect is prominent over the repeats and transposons in both rice and Arabidopsis. However, unlike Arabidopsis, PCGs in rice indeed underwent aberrant transcription by pol II, but hindered at the transcription start sites (TSS) by the compensatory H3K27me3 marks.

Aberrant transcription in *kd* lines gives rise to atypical sRNAs that are suppressed by pol IV

Aberrant RNAs generated by polymerases are major hotspots for sRNA production to initiate PTGS (Elmayan and Vaucheret, 1996; Voinnet et al., 1998; Morel et al., 2000; Vazquez et al., 2004). We hypothesized that the unusual hyper-occupancy by the pol II or pol V over the PCGs might trigger production of sRNAs from these regions. In order to explore this avenue of mis-transcription derived sRNAs, we critically evaluated the sRNAs from WT and *kd* in three different tissues examined earlier. As anticipated, this analysis showed that the numerous ShortStack identified sRNA clusters ('clusters') were upregulated in *kd* in all the three tissues in both the size classes (Fig. 5A). The sRNAs that are upregulated in *kd* were denoted as pol IV suppressed sRNAs and those that are reduced in *kd* were named as pol IV dependent sRNAs. In order to evaluate if these clusters potentially make sRNAs of mixed sizes, we counted sRNAs of each size class in both upregulated and downregulated clusters using bedtools multicov (Quinlan and Hall, 2010). The clusters were sub-categorized as 21-22nt predominant, 23-24nt predominant and mixed clusters based on mentioned thresholds (Supplemental Fig. 6A; Methods) and all the differentially expressed clusters were found to be of similar length distribution (Supplemental Fig. 6B). We validated the abundance of both categories of sRNAs by northern hybridisation and the results conclusively indicate the accumulation of pol IV suppressed sRNAs (Fig. 5B).

In order to have an unbiased estimation of the relative distribution of these sRNAs, we performed a genome window (100bp) based analyses (Methods). We found that the pol IV dependent bins are higher in number in all the three pol IV *kd* tissues (Supplemental Fig. 7A, 7B and 7C). We verified that these atypical sRNAs are not due to oversampling of the residual sRNAs in the *kd* tissues and associated normalisation artefacts by comparing the raw abundance of sRNAs in each library to the sum total of miRNAs in them (Supplemental Fig. 8). The dependent bins were distributed over the genome while the suppressed bins were fewer and concentrated at fewer selective loci (Supplemental Fig. 9A, 10A and 11A). Further, cumulative sum plots describing the relation between the abundance of sRNAs and the number of uniformly sized bins also indicated that suppressed sRNAs were abundantly concentrated in fewer bins (Supplemental Fig. 9B, 10B and 11B).

Further to validate if this feature of aberrant sRNAs in *kd* is associated with other pol IV machineries in vegetative tissues as well, we analysed the published sRNA datasets from rice seedlings from

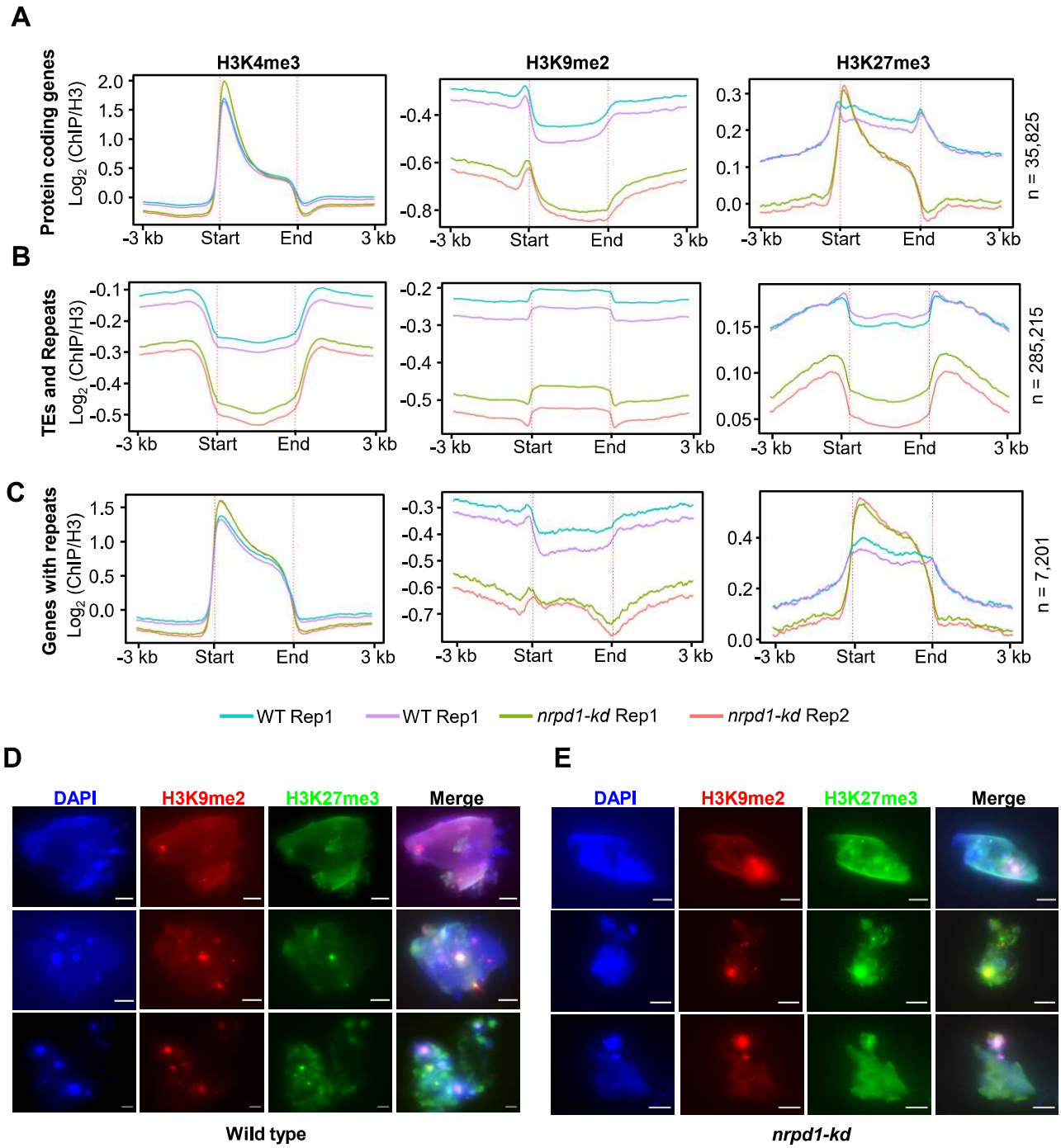


Figure 3. Pol IV maintains the genome-wide distribution of chromatin states.
(A) Metaplots describing the occupancy of histone H3 modifications H3K4me3, H3K9me2 and H3K27me3, normalised to total histone H3 occupancy plotted over the protein coding regions.
(B) Metaplots over TEs and repeats.
(C) Metaplots for genes having at least 10% of their length with annotated repeats.
(D and E) Immunostaining images of H3K9me2 and H3K27me3 marks in the nuclei extracted from WT **(D)** and *kd* **(E)** with their merged overlay. DAPI stain was used as DNA marker. Scale bar: 5 µm.

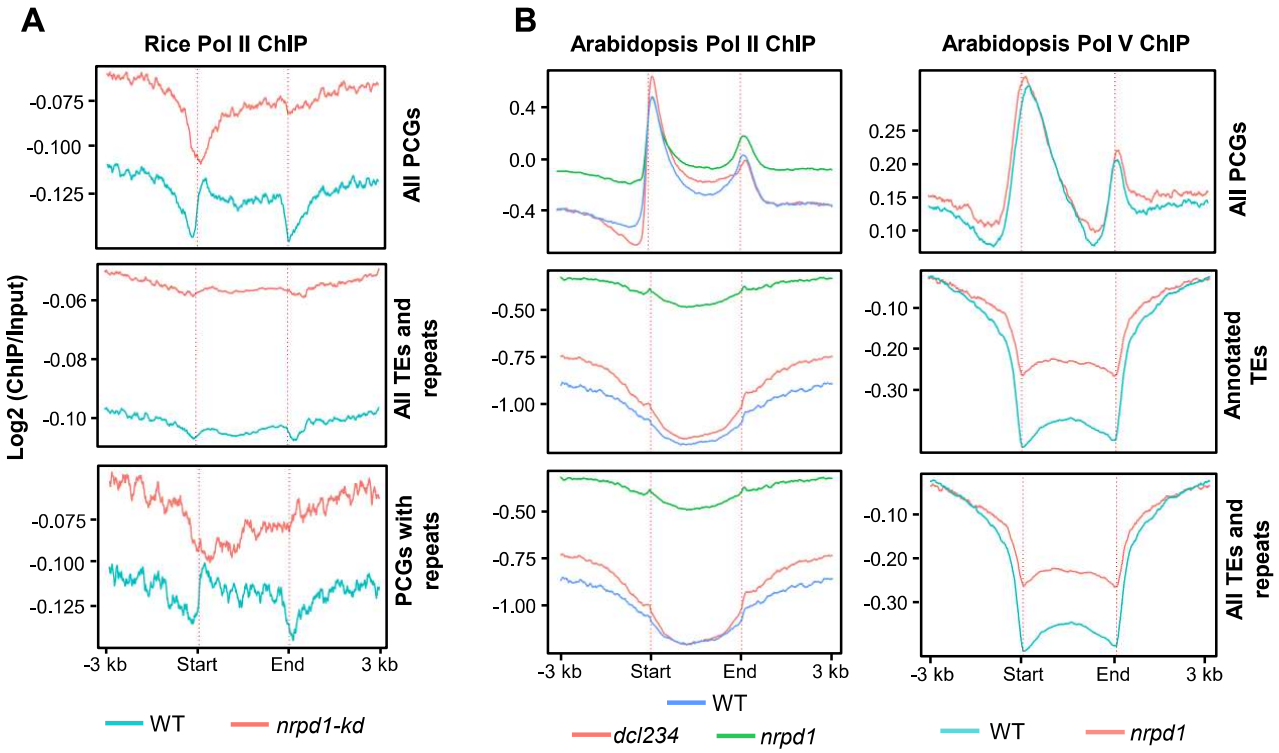


Figure 4 Loss of pol IV impacts occupancy of pol II and pol V.
(A) Metaplots depicting pol II occupancy over rice protein coding genes (PCGs) (n = 35,825), annotated TEs and repeats (n = 285,215) and genes having at least 10% of their length as repeats (n = 7,201). Signals are normalised to the input.
(B) Metaplots depicting the pol II and pol V occupancy in the mentioned genotypes over the protein coding genes (PCGs) (n = 28,775), fully annotated TEs (n = 31,189) and all the repeats and TE features (n = 68985) in Arabidopsis. Pol II datasets were taken from GSE61439 and pol V datasets were taken from GSE100010. Signals are normalised to the input.

nrpd1, rdr2 and *nrpe1* backgrounds (Wang et al., 2020b; Zheng et al., 2021). We indeed observed the same trend as in reproductive tissues in the *nrpd1* and *rdr2* but not in *nrpe1* seedlings (Fig. 5C). On the other hand, analyses of *rdr2* panicle sRNAs mimicked our *kd* profiles, substantiating our results (Fig. 5C). Further, we show that the *nrp(d/e)2* leaf sRNA datasets indeed retained the pol IV suppressed loci suggesting a minor role of pol V in accumulating these atypical sRNAs (Fig. 5C).

sRNAs in plants majorly contribute to silencing transcriptionally or post-transcriptionally (Baulcombe, 2004) and these sRNAs exhibit specific characteristics that define their functions. We examined these features for the pol IV suppressed sRNAs. Unlike the pol IV dependent sRNAs that are majorly loaded into AGO4 and possess 5'-A, the pol IV suppressed sRNAs did not show any bias at the first nucleotide position (Fig. 6A). Further, genomic repeat features that are hallmarks of dependent sRNAs were completely devoid of suppressed sRNAs (Fig. 6B). On the other hand, they accumulated many-folds over the protein coding features in *kd* lines, suggesting that loss of pol IV facilitated the enhanced production of these sRNAs, likely due to possible relaxed and permissive chromatin state (Fig. 6B). In addition we examined the status of these sRNAs in the NRPD1 complemented lines in rice, and found that they were retained (Fig. 6C). This adds to the evidence that the these pol IV suppressed sRNA arose due to loss of epigenetic silencing at the chromatin level and they might require multiple generations of rescue by the NRPD1 complementation to be silenced again as observed in other epigenetic mutants (Vongs et al., 1993; Martienssen and Colot, 2001; Li et al., 2020; Wang et al., 2021).

We speculated how conserved these suppressed sRNAs are and if they are seen in loss of function mutants of NRPD1 in other plant species. By analysing datasets derived from Arabidopsis WT and *nrpd1* inflorescence tissues (Zhou et al., 2018), we observed that loss of pol IV resulted in suppressed sRNAs, similar to rice, but from fewer bins (Supplemental Fig. 12A). Unlike pol IV dependent bins, suppressed sRNAs showed non-uniform distribution similar to rice tissues (Supplemental Fig. 12B). Availability of various genetic mutants in Arabidopsis prompted us to check the mechanistic origins of these sRNAs. In this direction, we proved that the suppressed sRNAs are exclusively dependent on pol IV complex using sRNA northern blots from various mutants in sRNA machinery (Fig.6D). The suppressed sRNA was unique to *nrpd1* and was not observed in sRNA processing mutants such as *dcl3, dcl234, ago4* or a mutant of chromatin remodeller *ddm1* (Fig. 6D). One of the pol IV suppressed loci accumulated aberrant transcripts of ~200nt detected with specific probes designed to bind to suppressed sRNAs leading to the notion that these were indeed products of aberrant transcription (Fig. 6E). Extending these observations, we obtained sRNA datasets from the similar stage inflorescence tissues from a set of mutants involved in RNA silencing (Zhai et al., 2015). This independent set of *nrpd1* datasets also confirmed the increase of suppressed sRNAs by at least 10 fold compared to WT mainly from non-repeat regions (Fig. 6F, Supplemental Fig. 13A and 13B). Interestingly, the pattern is seen in the other *nrpd1* associated mutants like *rdr2, nrp(d/e)2, nrpd1dcl3* and *rdr2dcl3* (Fig. 6F). This indicates that the suppressed sRNAs are directly coupled to the absence of pol IV complex (NRPD1 or RDR2). However, in Arabidopsis, most of the suppressed sRNAs were brought back to the WT levels upon NRPD1 complementation (Supplementary Fig. 13B). We also counted the abundance of sRNAs from up and downregulated bins in Col-0, *nrpd1, nrpe1* and *nrpd1nrpe1* seedlings, revealing a dependence of suppressed sRNAs in the *nrpd1* mutant partially with pol V like rice (Supplemental Fig. 13C, 13D and 13E). These analyses revealed that the pol IV suppressed sRNAs display antagonistic occupancy over the repeats and transposons but accumulate over the PCGs. In addition, this feature of pol IV suppressed sRNAs is conserved in Arabidopsis *nrpd1* and associated mutants and they display similar characteristics as in rice, but with a milder degree.

Pol IV suppressed sRNAs correlate with loss of silencing histone marks but not with DNA methylation changes

We presumed that the suppressed sRNAs must be a result of an epigenetic phenomenon due to loss of pol IV, supposedly due to revamping of histone or DNA methylation marks. In order to test this conjecture, we profiled the occupancy of the histone marks and the status of DNA methylation at differentially expressed sRNA clusters. We observed that over the pol IV dependent clusters H3K9me2 levels drastically dropped over all size classes (Fig. 7A). On the other hand, at the pol IV

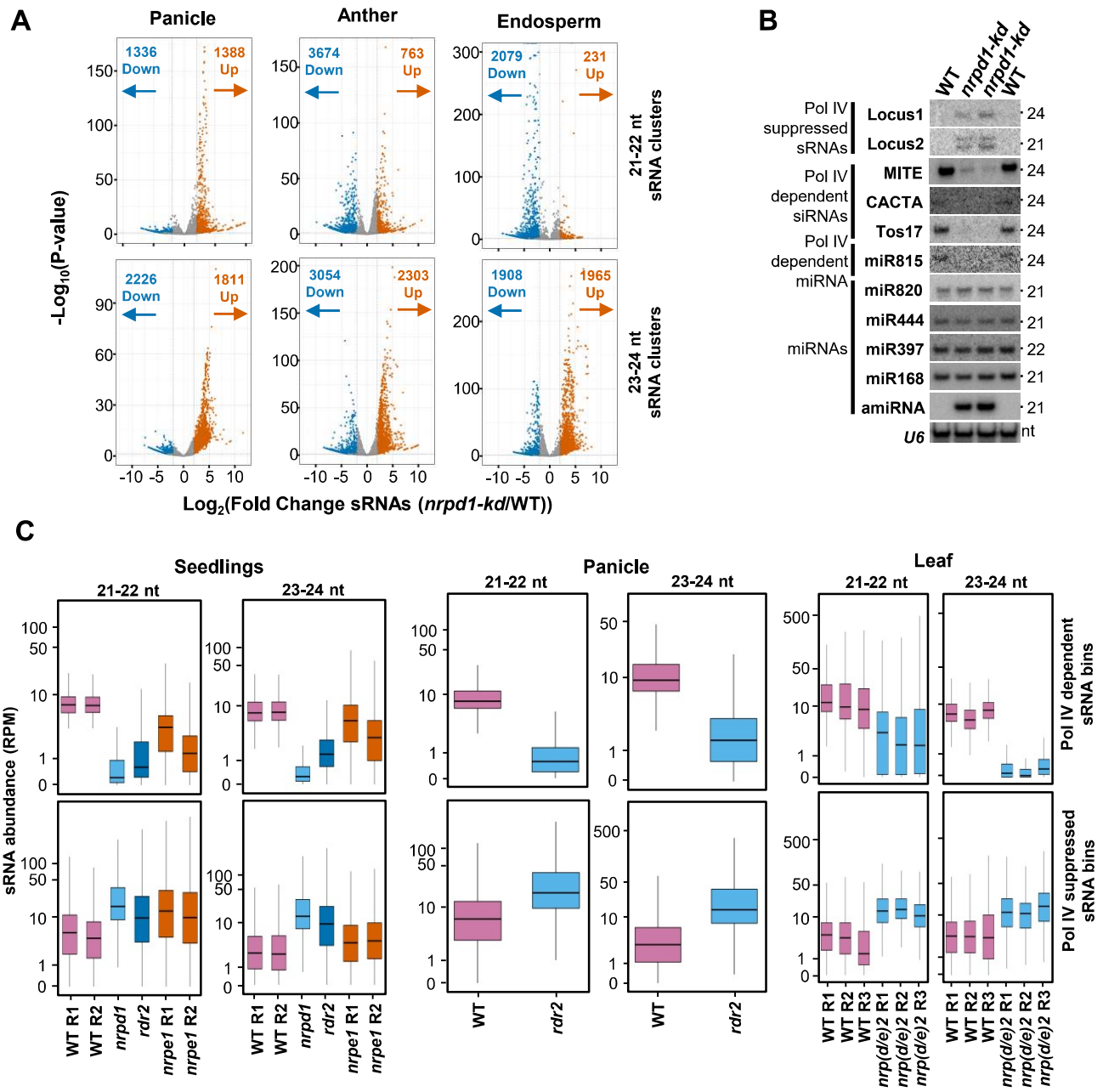


Figure 5. Pol IV complex suppresses small RNA production from several loci.
(A) Volcano plots showing levels of deregulation of sRNAs over clusters identified by shortstack from panicle, anther and endosperm sRNA datasets. The clusters were identified after subcategorizing into 21-22 nt and 23-24 nt size classes. Clusters showing statistically significant difference of 2-folds relative to WT are highlighted. The number of up- and downregulated clusters are mentioned with arrows.
(B) Small RNA northern blots validating the presence of pol IV suppressed sRNAs, dependent sRNAs and independent miRNAs. *U6* was used as loading control.
(C) Boxplots showing the abundance of sRNAs from different mutants in rice. sRNAs were size-categorized into 21-22nt and 23-24nt and counted in 100bp non-overlapping windows. Plots depict the abundance of sRNAs in each size class over *nrp(d/e)2* (leaf), *nrpd1* (seedlings) and *rdr2* (panicle) dependent and suppressed bins. The datasets are from GSE158709 (seedlings - *nrpd1, nrpe1*) and GSE130166 (seedlings and panicle – *rdr2*). The Y-axis is scaled to inverse sine hyperbolic function of RPM values.

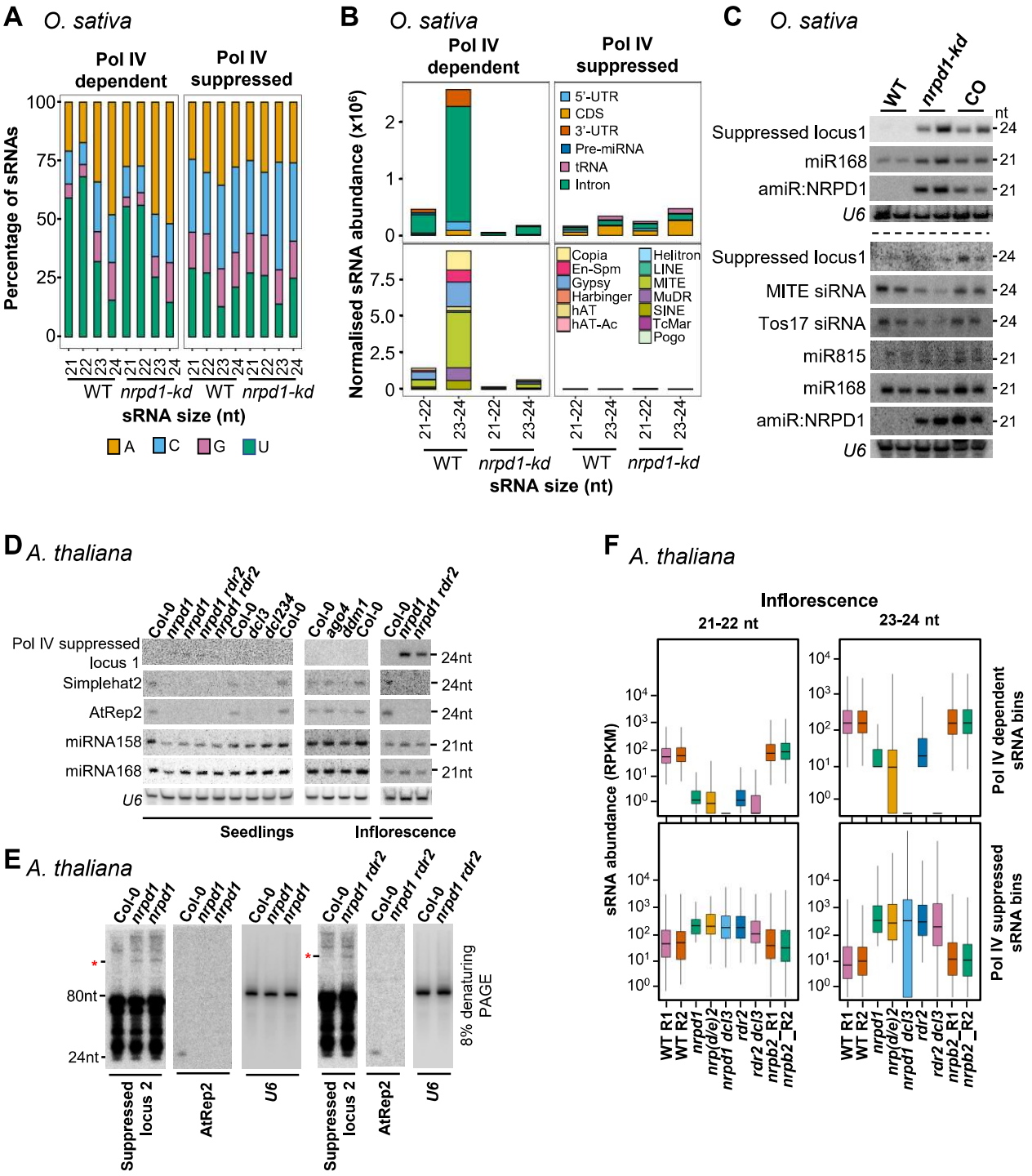


Figure 6. Polymerase IV suppressed sRNAs are conserved in Arabidopsis.
(A) Stacked bar plots showing normalised abundance of rice sRNAs of different sizes from pol IV suppressed and dependent bins displaying the 5' nucleotide bias.
(B) Stacked bar plots showing abundance of rice sRNAs from genomic features (coding features in the top and transposon categories in the bottom).
(C) Small RNA northern blots showing the levels of pol IV suppressed sRNAs in *kd* and NRPD1b complemented tissues (CO). Top set of images are the same as Fig. 2C and the membrane was stripped and reprobed for suppressed locus 1. Bottom blots (below the dotted line) are from another biological replicate.
(D) Small RNA northern blots showing the levels of pol IV suppressed sRNAs and other sRNAs from Arabidopsis seedlings and inflorescence tissues.
(E) Northern blot made with 8% denaturing PAGE gel depicting the resolved higher sized RNAs (*).
(F) Boxplots showing the sRNA abundance from different genotypes of Arabidopsis with counts from pol IV dependent and suppressed bins categorized into 21-22nt and 23-24nt size classes. The sRNA datasets were obtained from GSE61439 and analyses depicts sRNAs from inflorescence tissues. The Y-axis is scaled to inverse sine hyperbolic function of RPKM values.

dependent clusters, we observed over occupancy of H3K27me3 marks, comparatively more evident in 23-24nt and mixed size clusters (Fig. 7A). This unusual phenomenon is a proof that H3K27me3 marks can compensate for the loss of other silencing marks over transposons and repeats, that are the classical targets of pol IV activity. This also substantiates the latent redundancy of silencing brought about by H3K9me2 and H3K27me3 as a compensatory mechanism as seen in *met1* mutants but not reported in *nrpd1* mutants in Arabidopsis or other plants (Mathieu et al., 2005; Deleris et al., 2012). This also potentially hindered the aberrant transcription by pol II or pol V.

On the contrary, the suppressed sRNA clusters exhibited no such compensation by H3K27me3 marks, even though the H3K9me2 marks are lost over these loci (Fig. 7B). We also tested the status of the DNA methylation over these clusters revealing that it does not change significantly over CG, CHG or CHH contexts in both rice and Arabidopsis (Supplemental Figure 14 and 15). Taken together, these results indicate that loss of Pol IV relaxes the chromatin globally and the latent H3K27me3 compensation mechanism prevents aberrant transcription at pol IV dependent loci. Lack of this compensation in suppressed loci, in the absence of commensurate DNA methylation changes, likely results in aberrantly produced pol IV suppressed sRNAs.

Pol IV suppressed sRNAs are loaded into AGO1 to induce PTGS in protein coding loci

We undertook a set of analyses to identify if the pol IV suppressed sRNAs are functional. Annotation of the differentially expressed sRNA bins from different tissues identified that nearly half of the pol IV suppressed sRNA bins mapped to coding regions while dependent counterparts map to the distal intergenic regions as implicated in Fig. 6B (Supplemental Fig. 16A). In addition, the sRNAs from the suppressed bins overlapping with the PCGs mapped uniformly to both the sense and antisense strand of the gene, suggesting activity of one of the RNA dependent RNA polymerases (RDRs) (Supplemental Fig. 16B). Distribution of dependent and suppressed sRNAs over the protein coding genes matched the above observations (Supplemental Fig. 16C). We tested the status of various histone marks at these genes and found that unlike the bulk of all genes, these genes did not undergo H3K27me3 effectively, possibly triggering an aberrant transcription that is abrogated in other genes (Supplemental Fig. 17A).

Given that these atypical sRNAs are from gene coding regions and they do not change the DNA methylation signature of the loci, we hypothesised if they can target genes post transcriptionally by associating with argonautes (AGOs) such as AGO1. We tested this by checking if the sRNAs from these sRNA bins get loaded into AGOs, employing the publicly available IP datasets (Wu et al., 2009, 2010). As anticipated, the sRNAs from pol IV suppressed sRNAs get loaded into AGO1a and AGO1c preferentially as opposed to the pol IV dependent sRNAs (Fig. 8a). In line with the previous observations, pol IV dependent sRNAs were specifically enriched in the AGO4a, AGO4b and AGO16 IP datasets (Supplemental Fig. 18a). To test if these AGO-loaded sRNAs indeed target genes, we performed whole genome degradome sequencing in the WT and *kd* panicle tissues. Among the genes that overlap with the suppressed sRNAs (389 genes), target prediction tools identified 154 genes as potential targets of suppressed loci derived sRNAs (Supplemental Table 3). Of these genes, 58 genes showed reduction in mRNA accumulation in RNA-Seq (Supplemental Table 3 and Fig. 8D). Moreover, the degradome tag density at the predicted target loci was substantially increased in the *kd* lines, with enhanced variation among the genes showing reduced expression, thereby suggesting precise slicing at these sites (Fig. 8B, Supplemental Fig. 18B). In further agreement with the targeting process, the AGO1 IP was found to enrich sRNAs at the degradome predicted targeting position (Fig. 8C). These mRNA-targeting pol IV suppressed sRNAs were reduced in the rice *dcl1* mutants when compared to *dcl3* dataset (Wu et al., 2010), emphasizing the role of PTGS in regulating the genes (Supplemental Fig. 18C). We also observed that many of the genes that underwent PTGS by suppressed sRNA-mediated targeting in rice were previously implicated in reproductive growth and development (Fig. 8D). For instance, OsMADS18 (APETALA1 homolog in rice), a member involved in floral architecture establishment (Wang et al., 2020a), a close homolog of fertility restorer (RF) (Os10t0497366), glycine rich interaction partner of RF5 (Os12t0632000) (Hu et al., 2012) and pollen specific desiccation associated protein (Os11t0167800) were targeted for degradation in *kd* lines (Supplemental Table 3). In addition, these genes had reduced mRNA expression in *kd* lines. For example, RF5 interaction protein (Os12t0632000), mutant of which shows pollen non-viability due to

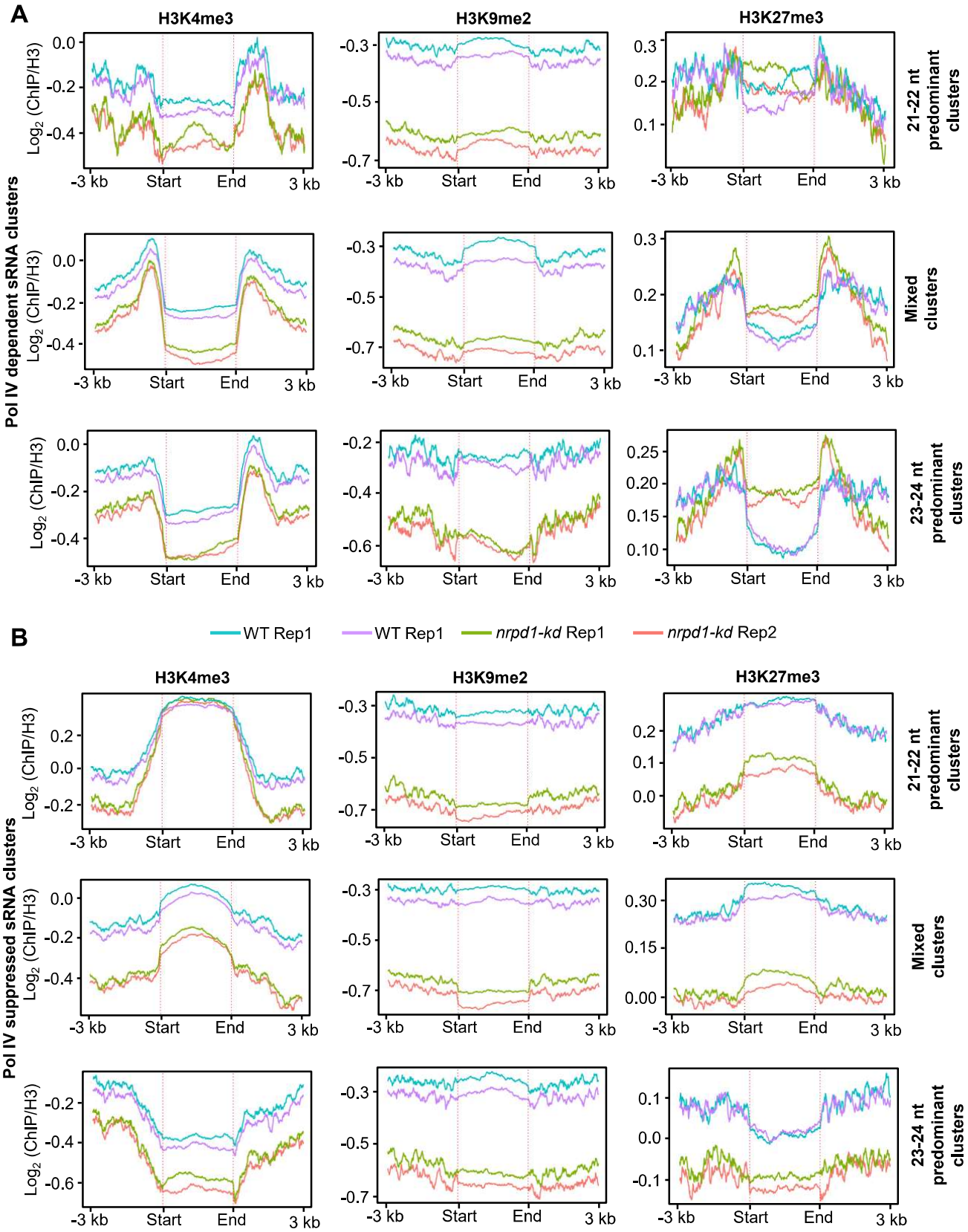


Figure 7. Distinct chromatin modifications at pol IV dependent and suppressed sRNA clusters in rice.

(A) Metaplots describing histone H3 PTMs (H3K4me3, H3K9me2 and H3K27me3) occupancy normalised to total H3 signal over the pol IV dependent sRNA clusters, categorized as 21-22 nt predominant clusters, 23-24 nt predominant clusters and the mixed sized clusters.

(B) Metaplots for suppressed clusters.

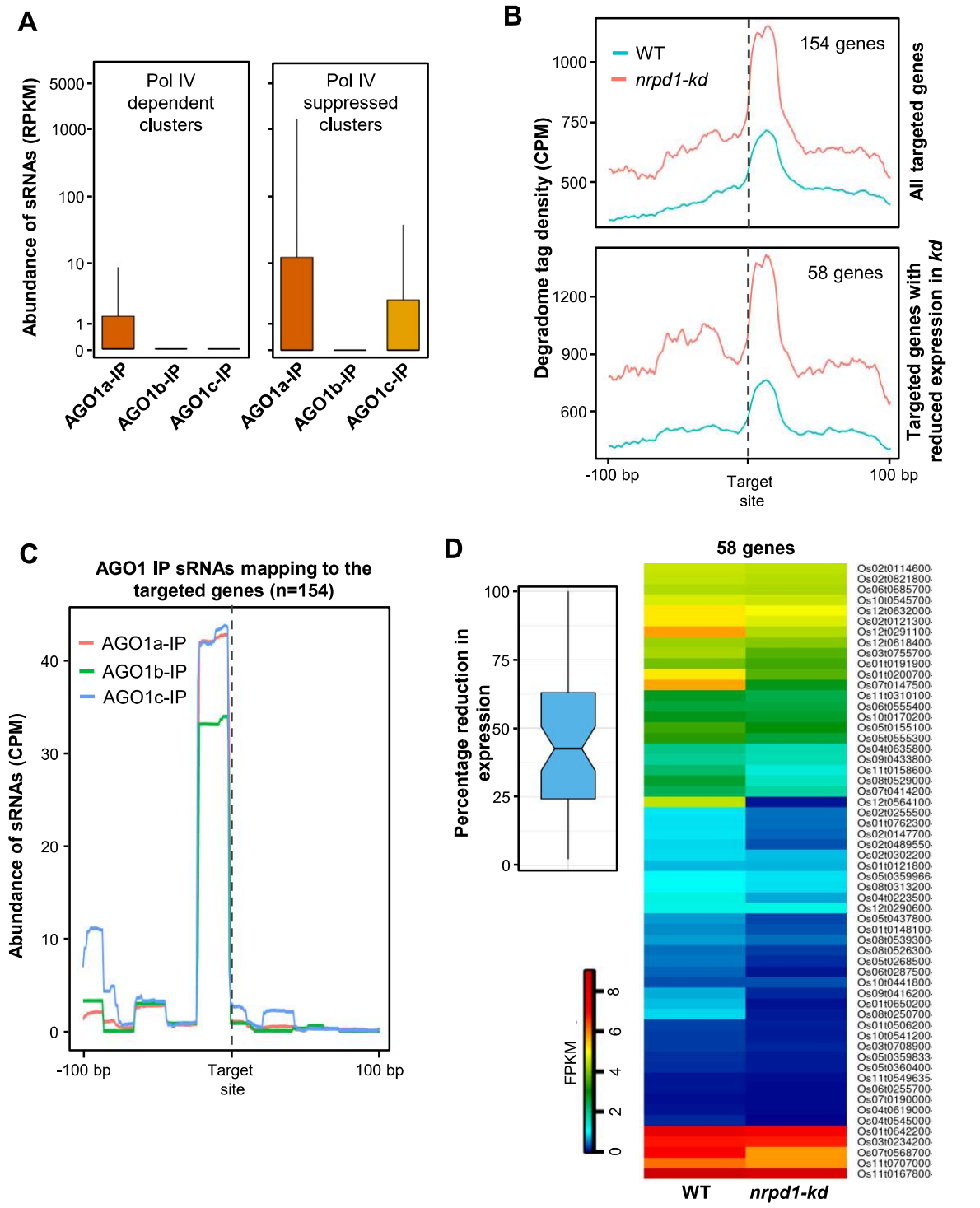


Figure 8. Pol IV suppressed sRNAs get loaded into AGO1 to mediate PTGS.
(A) Box plots describing the abundance of sRNAs enriched in AGO1 IPs of AGO1a, AGO1b and AGO1c from pol IV dependent and pol IV suppressed sRNA clusters. The datasets are taken from GSE18250. The Y-axis is scaled to inverse sine hyperbolic function of RPKM values.
(B) Metaplots showing the degradome tag density over the genes identified as targets in *nrpd1-kd* panicle degradome (top panel) and the same for the subset of targeted genes that show reduced expression in *kd* compared to WT. Targeted locations are centred and 100bp on either sides is displayed.
(C) Metaplots showing the abundance of AGO1 IP enriched sRNAs over the identified targeted genes centred at the targeting site.
(D) Heatmap showing the expression of 58 genes that are identified as targets in *nrpd1-kd* panicle degradome showing reduced expression in *kd* compared to WT. The inset boxplot shows the distribution of reduction in expression observed in the 58 genes in *kd* lines.

loss of cytoplasmic male sterility (Hu et al., 2012) was not only effectively targeted in *kd* lines (category 2), but also had reduced mRNA expression in these lines (Supplemental Table 3).

On the other hand, the suppressed sRNAs in Arabidopsis did not accumulate in the AGO1 IP datasets in both Col-0 and *nrpd1* plants while the dependent sRNAs were abundantly found in AGO4 IP datasets (Panda et al., 2020; Zhai et al., 2015) (Supplemental Fig. 19). While such targeting is not possible in Arabidopsis where suppressed sRNAs are neither abundant nor get loaded into AGO1 (Supplemental Fig.19), it is tempting to speculate that abundant sRNAs target same or similar key genes in other plants where perturbation of NRPD1 showed strong phenotypes (Wang et al., 2020c; Grover et al., 2018). Such a targeting mediated by suppressed sRNAs might act as strong deterrent for the loss of pol IV activity, and might contribute towards evolution of additional copies of this polymerase. In fact, monocots that have large genomes, also code for pol VI that is expressed and appears to have neo-functionalized carrying signatures of pol IV activity at least in part (Trujillo et al., 2018).

Discussion

RNA polymerases play a crucial role in the central dogma of life by transcribing the genetic material into functional RNA entities of different types. Activity of these polymerases are under tight control of several layers – epigenetic aspects of the substrate chromatin as well as the availability of necessary cofactors like transcription factors, enhancers and other chaperones. Two major conserved contributors for epigenetic silencing of this nature are DNA methylation and histone modifications that relay the regulatory outcome. These marks mainly prevent the spurious transcription by RNA polymerases at unintended locations. The precise and dynamic regulation of these modifications in a coordinated manner hallmarks the centre-stage of regulation. A mediator of such a synergistic response modulating both of these aspects is paramount. Such a mediator should not only obstruct illegit transcription but also prevent unwarranted occlusion of transcription of necessary genes and loci. Our investigation uncovers a plant specific RNA polymerase, pol IV, to be endowed with such a potential.

Evolutionary analyses suggest that plant-specific RNA polymerases IV, V, and in specific cases VI are novel machineries evolved in conjunction with the complexity of plant genomes and the degrees of importance of these polymerases varies across plants. For instance, in early land plants like *Physcomitrella patens* the pathway is completely redundant with other sRNA-independent DNA methyltransferases introducing *de novo* DNA methylation (Yaari et al., 2019) as opposed to the widespread defects observed in other plants with severity in rice (Fig. 1 and Supplemental Fig. 1, 2 and 3) (Zheng et al., 2021). The functions of these polymerases are pronounced in reproductive tissues as they are involved in faithful transmission of epigenetic information, antagonizing genome dosage aberrations and hybridisation (Satyaki and Gehring, 2019; Erdmann et al., 2017; Martinez et al., 2018; Zhang et al., 2016). It is apparent that the green-lineage specific RNA polymerases have neo-functionalized to perform additional roles with increasing genome complexity. The comparisons we present with respect to compensation of epigenetic marks between rice and Arabidopsis serves as a strong evidences to these predictions.

Several investigations in the model plant Arabidopsis have established that pol IV initiates biogenesis of sRNAs from repeats and transposons mediating establishing *de novo* DNA methylation aiding against their genotoxic proliferation. Our studies in rice corroborate this function of pol IV in multiple reproductive tissues (Fig. 1, 2 and Supplemental Fig. 1, 2, 3 and 4). Several reports have mentioned the synergistic effects and cross-dependence of histone modifications and global CG DNA methylation in establishing epigenetic modalities in Arabidopsis (Soppe et al., 2002; Mathieu et al., 2005; Deleris et al., 2012; Li et al., 2018; Zhong et al., 2021). These studies had employed *met1* and *ddm1* mediated DNA methylation loss occurring at the CG contexts globally resulting in loss of compaction at the pericentromeres and constitutive heterochromatin domains. On the other hand the RdDM is not restricted to heterochromatin and its purview extends into the gene rich regions as well, whether modulation of the sRNAs can directly perturb the chromatin states attributable to the gene expression was unclear. Especially, in monocots like rice, the interspersion of repeat fragments within genes mandates gene regulation not at extended length scales but in localised compartments even within the euchromatin (Espinas et al., 2020). Probing the effect of loss of RdDM on the chromatin states over the genes in our study uncovered a distinct redistribution of H3K9me2 and H3K27me3 marks (Fig. 3). Predominantly, H3K9me2 marks, attributed with well-suppressed constitutively silenced loci, are lost over these loci, whereas H3K27me3 associated with silenced gene coding units

by action of Polycomb Represive Complex2 showed over-occupancy in *kd* over the genes. Strikingly, the authentic pol IV transcribing loci that result in sRNAs showed significant H3K27me3 occupancy, likely an ectopic compensatory mode of silencing (Fig. 7). This feature is evolutionarily conserved between plants and animals and loss of transposon methylation can contribute to the redistribution of these marks into transposon territories (Deleris et al., 2012; Déléris et al., 2021). Close proximity of the protein coding genes and transposons might be the trigger causing the intrusion of facultative silencing marks on the genes in species with larger genomes such as rice. This is well-supported by the fact that upon *kd* of the pol IV, genes that are dispersed with the repeat fragments showed increased degree of H3K27me3 intrusion (Fig. 3C). PRC2 mediated H3K27me3 on genes impacts reproductive success in plants by controlling several imprinted genes that scale the genome dosage (Köhler et al., 2003; Roszak and Köhler, 2011; D. et al., 2013; Jiang et al., 2017). Such indirect effects on the protein coding genes by unwarranted H3K27me3 marks, to a certain extent, might explain the defects and stochastic nature of these variations in specific pol IV *kd* lines (Supplemental Fig. 2).

Loss of H3K9me2 at transposons, precise gain of H3K27me3 over the pol IV transcribing sRNA loci and redistribution of H3K27me3 marks over the protein coding genes upon loss of RdDM in rice should result in differential accessibility of these loci to RNA polymerases. Interestingly, probing the pol II status over genes and transposons revealed the distinct occupancy, not only in rice but also in Arabidopsis (Fig. 4). Such conservation in Arabidopsis might imply the importance of polymerase IV activity directly in influencing transcription of genes and transposons as the *dcl234* did not mirror these defects. Competitive activity of pol II and pol V, in the absence of pol IV, might trigger aberrant transcripts from rampant loci. Ungauged transcription in the absence of pol IV might feed to the small RNA pools via activity of DICER-LIKE proteins that are devoid of pol IV precursor load. Exploration of sRNA pools in the *nrpd1* mutants of rice and Arabidopsis indeed showed resultant aberrant sRNAs, likely triggered by spurious transcripts. Interestingly, such spuriously transcribed loci were not compensated by H3K27me3 or DNA methylation. This commonality of spurious sRNA transcripts in RdDM mutants is observed in *nrpe1* very recently (Zheng et al., 2021) where a similar notion of chromatin relaxation triggering pol IV and pol II transcription is promulgated. Similarly, studies in maize and Arabidopsis have suggested atypical transcripts in the *nrpd1* mutation (Erhard et al., 2015; McKinlay et al., 2018). Our studies mechanistically delineates the causative chromatin features in the *nrpd1* mutant. Not limiting to that we find that the resultant pol IV suppressed sRNAs are capable of targeting genes post transcriptionally by loading into specific AGOs.

These analyses reveal that the atypical pol IV suppressed sRNAs from diverse plants such as rice and Arabidopsis, were a result of open chromatin due to the perturbations to pol IV, and at least in rice, these are capable of targeting protein-coding genes post transcriptionally. Such atypical sRNAs are limited to selective loci as the poised loci for aberrant transcription from repeats are abrogated by compensatory H3K27me3 silencing. We catalogue reciprocity of the silencing between transcriptional and post-transcriptional modes, repurposing an existing machinery. Arabidopsis *nrpd1* related mutants also exhibited production of these sRNAs, nevertheless, they did not get loaded into AGO1 that potentiates them for active targeting. Absence of efficient PTGS in Arabidopsis *nrpd1* mutants might have alleviated the extent of defects unlike in rice. These results suggest a strong and multifaceted impact of pol IV in the expression of protein coding genes, especially in rice while encouraging evolution of additional genomic complexities, architecture and heterogeneity.

Materials and methods

Plant material

Indica variety rice (*Oryza sativa indica sp.*) Pusa Basmati 1 (PB1) plants were grown in growth chamber at 24°C/70% RH/16h-8h light-dark cycle for hardening before transferring to greenhouse maintained at 28°C with natural day-night cycle. Arabidopsis plants were grown in controlled growth chamber maintained at 22°C/70% RH/16h-8h light-dark cycle. Different mutants used were reported in Supplemental Table 5.

Binary vector construction and *Agrobacterium* mobilisation

For amiR precursor binary construct, the artificial miRNA was designed using WMD3 tool (Ossowski et al., 2008). The amiR was chosen so that it targets both the NRPD1 isoforms (NCBI Ids: NRPD1a - XM_015781553.2 and NRPD1b - XM_015756207.1) and optimal miRNA parameters were satisfied (Narjala et al., 2020). The mature amiR sequence (5'-UAUAGUGUUACUCUUGGACAU-3') was embedded in the *Osa*miR528 precursor in the pNW55 plasmid using WMD3 recommended primers

(Warthmann et al., 2008; Ossowski et al., 2008) (Supplemental Table 4). The precursor region was sub-cloned into a binary vector (pCAMBIA1300 backbone) between *Zm*Ubiquitin1 promoter and 35S poly(A) signal and mobilised into *Agrobacterium* cells using electroporation method.

For the complementation construct , NRPD1b promoter and 5'-UTR were amplified from the genomic DNA (chr09:22015503-22019281) and fused to the amiR-resistant CDS of OsNRPD1b (obtained by site directed mutagenesis) and this was cloned into pCAMBIA3300 vector (with Bialaphos resistance marker, BlpR). The construct was super-transformed to the HygR *kd* calli using the same method described.

Transgenic plant generation and segregation analysis

Agrobacterium tumefaciens strain LBA4404 with a helper plasmid pSB1 (containing extra copies of virulence genes) harbouring the binary vector of interest was used for infecting embryogenic calli derived from scutellum. Transformation of PB1 calli was performed as per previous methods (Hiei et al., 1994; Sridevi et al., 2003). About 21 days old scutellar calli maintained on the callus induction media (CIM) (Murashige-Skoog (MS) medium, 2.5 mg/L 2,4-D, 300 mg/L casein hydrolysate, 500 mg/L L-proline, pH 5.8) was incubated for 4 days(d) after cutting and infected with *Agrobacterium* culture of OD_{600nm} around 1.0. The calli were co-cultivated with *Agrobacterium* in dark and washed with liquid CIM and then incubated in selection medium (CIM with 50 mg/L hygromycin or 5 mg/L phosphinothricin and 250 mg/L cefotaxime) for two rounds of 21 days each. The selected calli were moved to regeneration medium (MS medium, 3 mg/L kinetin, 2 mg/L NAA) in light. The regenerated shoots were moved to half-strength MS medium for rooting before hardening in growth chamber and then transferred to a transgene-compliant green house.

For segregation to subsequent generations, de-husked seeds were surface sterilised using ethanol, bleach and 0.1% mercuric chloride and placed on half-strength MS medium with 50mg/L hygromycin.

Plant phenotyping

WT and *kd* plants of different generations were grown uniformly after genotyping for the transgene. Tillers were counted on a per plant basis and the grain filling rate was calculated as number of filled grains per tiller.

Pollen staining and imaging was done as described in (Pedersen et al., 2004). Equally developed, pre dehisced anthers were collected and stained as described. Three biological replicates were chosen per transgenic line (2 transgenic lines - 6 plants) and multiple images were taken post staining under light microscope. The obtained images were analysed using custom ImageJ scripts for detecting the pollen staining and size of the pollen. Staining index was calculated as percentage of negative stain normalised to the background.

RNA-seq and analyses

RNA seq was performed in pre-dehisced anthers and pre-emerged panicle tissues. The total RNA was extracted using TRIzol method and was poly(A) enriched before library preparation. Library preparation was done with NEBNext® Ultra™ II Directional RNA Library Prep kit (E7765L) as per manufacturer's instructions. The obtained libraries were sequenced in paired end mode (100bp) on a Illumina Hiseq2500 platform.

The obtained reads were adapter trimmed using Trimmomatic (Bolger et al., 2014) and rRNA depletion was done using SortmeRNA (Kopylova et al., 2012). The reads are mapped to IRGSP1.0 genome using HISAT2 (Kim et al., 2015) with default parameters. Cufflinks (Trapnell et al., 2012) was used to perform differential gene expression analyses and statistical testing. The volcano plots were generated for DEGs using custom R scripts with the p-value cut-off of less than 0.05 and absolute log_2 (fold change) expression cut-off of more than 2. For quantifying the expression of genes and transposons, bedtools multicov (Quinlan and Hall, 2010) was used to obtain raw abundance and then normalised to RPKM values. These values are plotted as box-plots using custom R scripts.

Reverse Transcriptase – quantitative Polymerase Chain Reaction (RT-qPCR)

Total RNA isolated from the respective plant tissues using TRIzol method was converted to cDNA using reverse transcriptase kit (Thermo RevertAid RT kit) following manufacturer's protocol after treating with DNaseA. The cDNA template was used for the qPCRs using SYBR green qPCR master mix (Solis Biodyne - 5x HOT Firepol Evagreen qPCR Master Mix). Rice *Actin* (LOC_Os03g50885.1)

was used as an internal control. The primers used for quantification are listed in Supplemental Table 4.

Southern hybridisation

Southern hybridisation was performed as described previously (Ramanathan and Veluthambi, 1995). Total DNA (around 5 µg) was isolated from plant tissues using CTAB method and digested using 30 units of appropriate restriction enzyme overnight. The digested DNA was electrophoresed on a 0.8% agarose gel in 1x TBE buffer and capillary transferred to Zeta probe nylon membrane (Biorad). The transferred membrane was UV crosslinked and taken for hybridisation. The probes (PCR amplified from gDNA or binary plasmid using the oligos in Supplemental Table 4) were internally labelled using [α-P^{32}] dCTP (BRIT India) using Rediprime labelling kit (GE healthcare) and hybridisation and washes were performed as described. The blots were exposed to phosphor imaging screen and scanned using Typhoon scanner (GE healthcare).

sRNA northern hybridisation

sRNA northern blots were performed as described earlier (Tirumalai et al., 2020; Shivaprasad et al., 2012). Around 15 µg of TRIzol extracted total RNA from different tissues was electrophoresed on a denaturing 15% acrylamide gel. The gel was electroblotted onto Hybond N+ membrane (GE healthcare) and UV crosslinked. The membrane was hybridised with the T4 PNK end labelled oligonucleotides (with [γ-P32]-ATP) in hybridisation buffer (Ultrahyb buffer - Invitrogen). The blots were washed, exposed to phosphor imaging screen and scanned using Typhoon scanner (GE healthcare). Post-scanning, blots were stripped at 80°C in stripping buffer and proceeded with repeat hybridisations with subsequent probes.

mRNA northern hybridisation

Total mRNA northern hybridisation was performed as described previously (Shivaprasad et al., 2006). Around 20ug of total RNA was ethanol precipitated and resuspended in denaturing loading buffer. RNA was electrophoresed in 1.5% denaturing agarose gel with 1% formaldehyde in MOPS buffer. The gel was blotted onto Hybond N+ membrane (GE healthcare) by capillary elution and UV crosslinked. The crosslinked membrane was hybridised using rRNA oligo probes described previously (Hang et al., 2018) end labelled with [γ-P32]-ATP in ultrahyb-buffer (Invitrogen). The blots are hybridised at 42°C, washed, exposed to phosphor screen (GE healthcare) and scanned using Typhoon scanner. After scanning stripping and re-probing was performed as done for sRNA northern blots. The membrane was stained with methylene blue solution and shown as loading control.

Extrachromosomal Circular DNA (ECC DNA) PCR

Profiling of extrachromosomal fragments were performed as described previously (Lanciano et al., 2017). The primers were designed to amplify either only the circular DNA products or linear products. UBCE2 gene was used as loading control. To further validate the results, the total DNA (100ng) was digested with PlasmidSafe DNase (Lucigen) to eliminate the linear DNA and the same PCRs were repeated again. The primers used are listed in Supplemental Table 4.

DNA methylome analyses

Total DNA was isolated using CTAB method. The library was constructed using the NEBNext® Enzymatic Methyl-seq Kit (Catalog no-E7120) using manufacturer's instructions and protocols described previously (Feng et al., 2020). The libraries were sequenced in a paired end mode (100bp) on a Hiseq2500 platform.

The obtained reads were quality checked and trimmed using cutadapt (Martin, 2011) followed by alignment to IRGSP1.0 genome using Bismark aligner tool with default parameters (Krueger and Andrews, 2011). DNA methylation status was extracted and coverage reports were generated using the Bismark tools. The obtained results are analysed using methylation package ViewBS (Huang et al., 2018). The chloroplast and mitochondrial genomes' cytosine conversion rate was used as a quality control. ViewBS tools were used for estimating DNA methylation in different contexts over the defined regions.

For Arabidopsis datasets (Supplemental Table 2), same analyses pipeline was used except that the reads are aligned to the TAIR10 genome.

Scanning electron microscopy

SEM imaging was performed as described previously (Das et al., 2020). Equally grown WT and *kd* spikelets were collected before anthesis and fixed in 0.1 M sodium phosphate buffer (pH 7.0) with 3% paraformaldehyde and 0.25% glutaraldehyde. After incubation at 4°C for 24 h,spikelets were rinsed with 0.1 M sodium phosphate buffer and serially dehydrated with ethanol (30% to 100%). Samples were further dehydrated using critical point drying (LEICA EM CPD300), gold coated and the images were acquired using a Carl ZEISS scanning electron microscope at an accelerating voltage of 2 kV.

Micro computed tomography (CT) scan

Micro CT scanning of developing florets was performed as described earlier (Kannan et al., 2021). Florets from panicles that were equally developed before anthesis was scanned on Bruker Skyscan-1272 (Kontich, Belgium) at 40 kV, 250 µA, 4 µm image pixel size, without filters. The obtained image series was merged into a montage after colour adjustments using ImageJ.

Chromatin Immunoprecipitation - sequencing (ChIP-seq) and analyses

ChIP was performed as described by earlier methods (Song et al., 2016; Saleh et al., 2008). Around 1.2 grams of pre-emerged panicle tissues were taken and crosslinked with 1% formaldehyde. The crosslinked tissues were pulverized in liquid nitrogen and nuclei were isolated. Equal amount of nuclei were lysed and sheared using ultrasonication (Covaris) until fragments substantially accumulate in the 150-350bp size range. The sheared chromatin was incubated overnight with 50ul of protein G dynabeads (Thermo Fisher) at 4 degrees that were pre-bound to the appropriate antibodies. The beads were pre-bound with 4-5 micrograms of antibodies (H3K4me3 - Merck 07-473; H3K9me2 Abcam ab1220; H3K27me3 - Acive Motif 39155; H3 – Merck 07-10254; PolII – Abcam ab817) before incubation with the sheared chromatin. Washes, elution, decrosslinking and purification were performed as described. The purified IP products were taken for library preparation using NEBNext® Ultra™ II DNA Library Prep with Sample Purification Beads (NEB, E7103L) as per manufacturer's protocol. The libraries (with replicates) were sequenced on Illumina HiSeq 2500 platform in single end mode (50bp).

The obtained datasets were adapter trimmed using cutadapt (Martin, 2011) and aligned to the IRGSP 1.0 genome using Bowtie2 tool (Langmead and Salzberg, 2012) with the following parameters: -v 1 -m 100 -y -a --best –strata. PCR duplicates were removed before further analyses. The alignment files were converted to coverage files and compared (Log2 fold ratio) to total H3 signal (for histone H3 PTMs) or to sheared chromatin input (for Pol II) using bamcompare utility of Deeptools (Ramírez et al., 2014). The average signals over the desired regions/ annotations were estimated using computematrix utility of deeptools and the coverage signal metaplots were plotted using plotprofile utility that were modified using custom R script.

Arabidopsis ChIP datasets (Supplemental Table 2) were handled in the same way as the rice datasets except that the reads were mapped to TAIR10 genome. The annotations were taken from TAIR10 GTF file.

Nuclei immunostaining

Immunofluorescence labelling and microscopy was performed as described earlier (Yelagandula et al., 2014) with the following modifications. The nuclei were isolated exactly as described for ChIP with fixation using 4% formaldehyde. The isolated nuclei are fixed again using 4% formaldehyde in PBS and taken for immunostaining performed in a tube. The post-fixed, washed nuclei are blocked with blocking solution (1xPBS with 0.1% Triton-X 100, 1%BSA, 10% normal goat serum and 100mM glycine) for 1 hour at room temperature. Primary antibodies (H3K9me2 Abcam ab1220 [1:100]; H3K27me3 - Acive Motif 39155 [1:300]) incubations were done at 4 degrees for 16 hours in antibody binding solution (1xPBS with 0.1%BSA, 1% normal goat serum, 0.1% Triton-X 100 and 100mM glycine). The nuclei are washed with 1xPBS with 0.1% Triton X 100 and incubated with appropriate secondary antibodies (Invitrogen; goat raised Anti-mouse 488 and Anti-Rabbit 555) in antibody binding solution at 1:1000 dilution in dark at RT. Post washing, the nuclei are DAPI stained, mounted with fluoromount anti-fade reagent and imaged on a fluorescence microscope (Nikon).

sRNA sequencing and analyses

Small RNA sequencing was performed from pre-emerged panicle, anther and endosperm of equally developed WT and *kd* plants. The size fractionation and library preparation was done as described in

Tirumalai et al., 2019. Reads obtained were quality checked, adapter removed and size selected using UEA small RNA workbench (Stocks et al., 2018). The reads obtained are aligned after categorisation into 21-22nt or 23-24nt sizes to IRGSP1.0 genome using Bowtie (Langmead and Salzberg, 2012) with the following parameters: -v 1 -m 100 -y -a --best –strata. Step-wise analyses for genomic annotations and 5'-nucleotide abundance was performed as described earlier (Swetha et al., 2018). Only the mapped reads were used for further analyses. MicroRNA (miRNA) abundance was calculated using miRProf tool (Stocks et al., 2018).Arabidopsis datasets were handled the same way except that the reads were aligned to TAIR10 genome.

sRNA differential expression analyses

The filtered sRNAs, size classified into 21-22nt and 23-24nt, were aligned and sRNA loci identified using Shortstack (Axtell, 2013) with following parameters: --nohp --mmap f --mismatches 1 --mincov 0.2 rpmm. The raw counts obtained in each cluster was used for differential expression analyses using DESeq (Love et al., 2014). Only the clusters with p-value less than 0.05 and absolute $\log_2$ abundance fold change more than 1 were called as differential sRNA clusters (Supplemental Dataset 1). Volcano plots generated using custom R scripts. The clusters identified were further subcategorized into predominant 21-22nt clusters (if the 21-22nt population in a particular cluster exceeds 65% of total small RNAs), predominant 23-24nt (if 23-24nt sRNAs exceed 65%) and mixed clusters (all the other DE clusters).

Genome window based sRNA analyses and plotting

The genome was split into 100bp non-overlapping windows using bedtools makewindows (Quinlan and Hall, 2010) and the sRNAs (after dividing into 21-22nt and 23-24nt) were counted in each of the bins in the respective sized alignment files using bedtools multicov . The raw counts were normalised to RPM by considering the total reads mapped in each size class. All the bins that accumulate non-zero RPM values in both genotypes and at least minimum of 2 RPM when summed were taken for differential analysis. 2-fold change in RPM and absolute difference in RPM as minimum 5 were considered as the cut-off for differential calling for a bin (upregulated or downregulated in *kd* w.r.t. WT) (Supplemental Dataset 2). The bins identified from different tissues were taken for further analyses of mapping to genomic feature (overlap with the annotation), 5'-nucleotide bias of the sRNAs mapping to them, circos plots and the cumulative sum plots.

ChIPseekeR (Yu et al., 2015) was used for plotting the Vennpie diagrams of annotations of the suppressed and dependent sRNA bins after merging the different size category bins. Circos plots displaying the heatmaps of the sRNA abundance of size-categorized bins were plotted using shinyCircos (Yu et al., 2018). The Upset plots were created in R using Intervene (Khan and Mathelier, 2017).

Arabidopsis sRNA analysis was also performed the same way with the same thresholds for differential bin calling and associated analyses as handled for rice bins. The differential expressed sRNA bins were calculated with the same cut-off as used for rice when comparing WT and *nrpd1* sRNAs in different tissues (Supplemental Dataset 3). The chromosome-wide heatmaps of the sRNA bins were plotted using shinyChromosome (Yu et al., 2019). All the Arabidopsis datasets were processed the same way as rice and the bedtools multicov was used to estimate the sRNA abundance in each of the upregulated and downregulated sRNA bins. The data was plotted as a boxplot using ggplot in R. Statistical analyses are performed using R.

Degradome analysis

The obtained reads were processed for adapter removal and size filtering from 18 to 21 nt using UEA small RNA workbench (Stocks et al., 2018). CleaveLand version 4 (Addo-Quaye and Axtell 2008) with default parameters was used to identify the degradome validated target genes. Three fasta files were given as input to CleaveLand, sRNA from the pol IV suppressed loci, genes overlapping with pol IV suppressed loci and degradome reads. The cut-offs of Allen score 8 and mfe ratio of 0.65 were used to obtain the valid gene targeting list. The list of targeted genes are in Supplemental Table 3. The target sites are mapped onto the target transcripts and deeptools was used to generate the metaplots as described earlier centering at the slicing site.

AGO-IP data analyses

AGO IP datasets of various AGOs from rice and Arabidopsis were processed in the same way as the sRNA datasets, mapping to their corresponding genomes. The IP sRNAs mapping to the upregulated and downregulated bins/clusters were counted using bedtools multicov and box-plots were plotted using custom R scripts.

Supplementary Information

Supplementary figures

Supplemental Figure 1. Distinct sets of Agrobacterium mediated transformation events yielded knockdown lines with similar phenotypic defects.

Supplemental Figure 2. Specific lineage of Pol IV knockdown plants showed severe defects that are not due to transgenesis or knockdown dosage suggesting possible epimutation.

Supplemental Figure 3. Knockdown of pol IV in rice produces defects in reproductive development and loss of sRNAs in reproductive tissues.

Supplemental Figure 4. Loss of pol IV causes specific loss of repeats and transposon associated sRNAs.

Supplemental Figure 5. Pol IV regulates genes and silences transposons and repeats.

Supplemental Figure 6. Pol IV dependent and pol IV suppressed sRNA clusters are of similar length.

Supplemental Figure 7. Pol IV dependent and pol IV suppressed sRNA bins are distinct in producing sRNAs with specific sizes.

Supplemental Figure 8. Pol IV dependent and pol IV suppressed sRNA bins are not due to oversampling and library normalisation.

Supplemental Figure 9. Pol IV suppressed sRNAs from panicle are non-uniformly distributed across the genome.

Supplemental Figure 10. Pol IV suppressed sRNAs from anther are non-uniformly distributed across the genome.

Supplemental Figure 11. Pol IV suppressed sRNAs from endosperm are non-uniformly distributed across the genome.

Supplemental Figure 12. Pol IV suppressed sRNAs from Arabidopsis inflorescence are non-uniformly distributed across the genome similar to rice tissues.

Supplemental Figure 13. Pol IV suppressed sRNAs from Arabidopsis are characteristically similar to that of rice.

Supplemental Figure 14. Canonical RdDM does not potentiate pol IV suppressed sRNAs in rice.

Supplemental Figure 15 Canonical RdDM does not potentiate pol IV suppressed sRNAs in Arabidopsis.

Supplemental Figure 16. Pol IV suppressed sRNAs have significant overlap with coding regions.

Supplemental Figure 17. Genes overlapping with suppressed sRNAs exhibit reduced silencing compensation by H3K27me3.

Supplemental Figure 18. Rice pol IV suppressed sRNAs can target genes effectively.

Supplemental Figure 19. Arabidopsis pol IV suppressed sRNAs are not effectively loaded into AGO1 and unlikely to target genes.

Supplementary Tables and Datasets

Supplemental Table 1: Details of high-throughput genomics data generated in this study.

Supplemental Table 2: Details of high-throughput genomics data obtained from publicly available datasets.

Supplemental Table 3: List of genes targeted by the sRNAs from overlapping suppressed sRNA bins.

Supplemental Table 4: List of oligos and probes used in this study.

Supplemental Table 5: List of Arabidopsis mutants used in this study.

Supplemental Dataset 1: Bed co-ordinates of pol IV dependent and suppressed sRNA clusters in different rice tissues.

Supplemental Dataset 2: Bed co-ordinates of pol IV dependent and suppressed sRNA 100bp bins in different rice tissues.

Supplemental Dataset 3: Bed co-ordinates of pol IV dependent and suppressed sRNA 100bp bins identified in Arabidopsis tissues.

Declarations

Ethics approval and consent to participate

Not applicable.

Availability of data and materials

Most data generated or analysed during this study are included in this article and its supplemental files. Deep sequencing datasets have been deposited in GEO (https://www.ncbi.nlm.nih.gov/geo/) under the super-series GSEXXXXXX.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported Ramanujan Fellowship (SR/S2/RJN-109/2012; Department of Science and Technology, Government of India) and NCBS-TIFR core funding and a grant

(BT/PR12394/AGIII/103/891/2014; BT/IN/Swiss/47/JGK/2018-19; BT/PR25767/GET/ 119/151/2017) from Department of Biotechnology (DBT), Government of India. This study was also supported by Department of Atomic Energy, Government of India, under Project Identification No. RTI 4006 (1303/3/2019/R&D-II/DAE/4749 dated 16.7.2020). Swetha Chenna and Kannan Pachamuthu acknowledge fellowship from DBT, India. These funding agencies did not participate in the designing of experiments, analysis or interpretation of data, or in writing of the manuscript.

Author contributions

PVS and VHS designed all experiments and discussed results and wrote the manuscript. VHS performed most of the experiments and bioinformatics analyses. SC performed bioinformatics analysis. DB generated transgenic lines. KP performed electron microscopy and micro CT scanning. TC generated and shared the *nrp(d/e)2* sRNA datasets. RAM helped with the analysis and discussion. All authors have read and approved the manuscript.

Acknowledgements

We thank Professor K. Veluthambi for *Agrobacterium* strains, PB1 seeds and binary plasmids. We are grateful to Prof. David Baulcombe for the Arabidopsis mutants. We thank genomics, electron microscopy, IT, radiation, greenhouse and lab-kitchen facilities at the NCBS campus. We acknowledge the guidance of Anushree Narjala in bioinformatics analysis. We thank Dr. Dimple Notani for sharing reagents. We thank Rahul Raj Singh, M. Rajagopalan and Sumvit Goyal for amiR-NRPD1 binary plasmid cloning. We acknowledge the R based meta-plotting script from the Github page of Jeffrey Grover (https://github.com/groverj3). We are grateful to Nitish Dua and Mohammad Shariq for the help with microscopy. We thank all the lab members for discussions and comments.

References

Axtell, M.J. (2013). ShortStack: comprehensive annotation and quantification of small RNA genes. RNA **19**: 740–751.

Baulcombe, D. (2004). RNA silencing in plants. Nature **431**: 356–363.

Baumberger, N. and Baulcombe, D.C. (2005). Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc. Natl. Acad. Sci. U. S. A. **102**: 11928–11933.

Blevins, T., Pontes, O., Pikaard, C.S., and Meins, F., Jr (2009). Heterochromatic siRNAs and DDM1 independently silence aberrant 5S rDNA transcripts in Arabidopsis. PLoS One **4**: e5932.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114–2120.

Brown, S.W. (1966). Heterochromatin. Science **151**: 417–425.

Chenxin, L., Hengping, X., Fang-Fang, F., Scott D, R., Venkatesan, S., and Jonathan I, G. (2020). Genome-wide redistribution of 24-nt siRNAs in rice gametes. Genome Res. **30**: 173–184.

Coruh, C., Cho, S.H., Shahid, S., Liu, Q., Wierzbicki, A., and Axtell, M.J. (2015). Comprehensive annotation of Physcomitrella patens small RNA loci reveals that the heterochromatic short interfering RNA pathway is largely conserved in land plants. Plant Cell **27**: 2148–2162.

Cuerda-Gil, D. and Slotkin, R.K. (2016). Non-canonical RNA-directed DNA methylation. Nat. Plants **2**.

Cui, X., Jin, P., Cui, X., Gu, L., Lu, Z., Xue, Y., Wei, L., Qi, J., Song, X., Luo, M., An, G., and Cao, X. (2013). Control of transposon activity by a histone H3K4 demethylase in rice. Proc. Natl. Acad. Sci. U. S. A. **110**: 1953–1958.

D., K., Hennig, L., and Köhler, C. (2013). Increased Maternal Genome Dosage Bypasses the Requirement of the FIS Polycomb Repressive Complex 2 in Arabidopsis Seed Development. PLoS Genetics **9:e1003163**.

Das, S., Swetha, C., Pachamuthu, K., Nair, A., and Shivaprasad, P.V. (2020). Loss of function of Oryza sativa Argonaute 18 induces male sterility and reduction in phased small RNAs. Plant Reprod. **33**: 59–73.

Debladis, E. et al. (2020). Construction and characterization of a knock-down RNA interference line of OsNRPD1 in rice (Oryza sativa ssp japonica cv Nipponbare). Philos. Trans. R. Soc. Lond. B Biol. Sci. **375**: 20190338.

Déléris, A., Berger, F., and Duharcourt, S. (2021). Role of Polycomb in the control of transposable elements. Trends Genet.

Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D., and Jacobsen, S.E. (2012). Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis thaliana. PLoS Genet. **8**: e1003062.

Earley, K.W., Pontvianne, F., Wierzbicki, A.T., Blevins, T., Tucker, S., Costa-Nunes, P., Pontes, O., and Pikaard, C.S. (2010). Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. Genes Dev. **24**: 1119–1132.

Elmayan, T. and Vaucheret, H. (1996). Expression of single copies of a strongly expressed 35S transgene can be silenced post-transcriptionally. Plant J. **9**: 787–797.

Erdmann, R.M., Satyaki, P.R.V., Klosinska, M., and Gehring, M. (2017). A small RNA pathway mediates Allelic dosage in endosperm. Cell Rep. **21**: 3364–3372.

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

Espinas, N.A., Tu, L.N., Furci, L., Shimajiri, Y., Harukawa, Y., Miura, S., Takuno, S., and Saze, H. (2020). Transcriptional regulation of genes bearing intronic heterochromatin in the rice genome. PLoS Genet. **16**: e1008637.

Feng, S., Zhong, Z., Wang, M., and Jacobsen, S.E. (2020). Efficient and accurate determination of genome-wide DNA methylation patterns in Arabidopsis thaliana with enzymatic methyl sequencing. Epigenetics Chromatin **13**: 42.

Feng, W. and Michaels, S.D. (2015). Accessing the inaccessible: The organization, transcription, replication, and repair of heterochromatin in plants. Annu. Rev. Genet. **49**: 439–459.

Gent, J.I., Madzima, T.F., Bader, R., Kent, M.R., Zhang, X., Stam, M., McGinnis, K.M., and Dawe, R.K. (2014). Accessible DNA and relative depletion of H3K9me2 at maize loci undergoing RNA-directed DNA methylation. Plant Cell **26**: 4903–4917.

Gibney, E.R. and Nolan, C.M. (2010). Epigenetics and gene expression. Heredity (Edinb.) **105**: 4–13.

Grover, J.W., Kendall, T., Baten, A., Burgess, D., Freeling, M., King, G.J., and Mosher, R.A. (2018). Maternal components of RNA -directed DNA methylation are required for seed development in Brassica rapa. Plant J. **94**: 575–582.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nat. Struct. Mol. Biol. **11**: 394–403.

Hang, R., Wang, Z., Deng, X., Liu, C., Yan, B., Yang, C., Song, X., Mo, B., and Cao, X. (2018). Ribosomal RNA biogenesis and its response to chilling stress in Oryza sativa. Plant Physiol. **177**: 381–397.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science **308**: 118–120.

Hiei, Y., Ohta, S., Komari, T., and Kumashiro, T. (1994). Efficient transformation of rice (Oryza sativa L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. Plant J. **6**: 271–282.

Hirsch, C.D. and Springer, N.M. (2017). Transposable element influences on gene expression in plants. Biochim. Biophys. Acta Gene Regul. Mech. **1860**: 157–165.

Hu, J. et al. (2012). The rice pentatricopeptide repeat protein RF5 restores fertility in Hong-Lian cytoplasmic male-sterile lines via a complex with the glycine-rich protein GRP162. Plant Cell **24**: 109–122.

Huang, X., Zhang, S., Li, K., Thimmapuram, J., and Xie, S. (2018). ViewBS: a powerful toolkit for visualization of high-throughput bisulfite sequencing data. Bioinformatics **34**: 708–709.

Jiang, H., Moreno-Romero, J., Santos-González, J., De Jaeger, G., Gevaert, K., Van De Slijke, E., and Köhler, C. (2017). Ectopic application of the repressive histone modification H3K9me2 establishes post-zygotic reproductive isolation in Arabidopsis thaliana. Genes Dev. **31**: 1272–1287.

Kannan, P., Chongloi, G.L., Majhi, B.B., Basu, D., Veluthambi, K., and Vijayraghavan, U. (2021). Characterization of a new rice OsMADS1 null mutant generated by homologous recombination-mediated gene targeting. Planta **253**: 39.

Kf, E., Jl, S., Se, P., Jp, L., and Cj, H. (2009). RNA polymerase IV functions in paramutation in Zea mays. Science **323**: 1201–1205.

Khan, A. and Mathelier, A. (2017). Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics **18**.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods **12**: 357–360.

Köhler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W., and Grossniklaus, U. (2003). The Polycomb group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. Genes Dev **17**: 1540–1553.

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics **28**: 3211–3217.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science **184**: 868–871.

Krueger, F. and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics **27**: 1571–1572.

Lanciano, S., Carpentier, M.-C., Llauro, C., Jobet, E., Robakowska-Hyzorek, D., Lasserre, E., Ghesquière, A., Panaud, O., and Mirouze, M. (2017). Sequencing the extrachromosomal circular mobilome reveals retrotransposon activity in plants. PLoS Genet. **13**: e1006630.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods **9**: 357–359.

Law, J.A., Ausin, I., Johnson, L.M., Vashisht, A.A., Zhu, J.-K., Wohlschlegel, J.A., and Jacobsen, S.E. (2010). A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in Arabidopsis. Curr. Biol. **20**: 951–956.

Law, J.A., Du, J., Hale, C.J., Feng, S., Krajewski, K., Palanca, A.M.S., Strahl, B.D., Patel, D.J., and Jacobsen, S.E. (2013). Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature **498**: 385–389.

Law, J.A. and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. **11**: 204–220.

Le, N.T., Harukawa, Y., Miura, S., Boer, D., Kawabe, A., and Saze, H. (2020). Epigenetic regulation of spurious transcription initiation in Arabidopsis. Nat. Commun. **11**: 3224.

Li, J., Yang, D.-L., Huang, H., Zhang, G., He, L., Pang, J., Lozano-Durán, R., Lang, Z., and Zhu, J.-K. (2020). Epigenetic memory marks determine epiallele stability at loci targeted by de novo DNA methylation. Nat. Plants **6**: 661–674.

Li, S., Vandivier, L.E., Tu, B., Gao, L., Won, S.Y., Li, S., Zheng, B., Gregory, B.D., and Chen, X. (2015). Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features and regulation of siRNA biogenesis. Genome Res. **25**: 235–245.

Li, X., Harris, C.J., Zhong, Z., Chen, W., Liu, R., Jia, B., Wang, Z., Li, S., Jacobsen, S.E., and Du, J. (2018). Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. Proc. Natl. Acad. Sci. U. S. A. **115**: E8793–E8802.

Liu, Z.-W., Shao, C.-R., Zhang, C.-J., Zhou, J.-X., Zhang, S.-W., Li, L., Chen, S., Huang, H.-W., Cai, T., and He, X.-J. (2014). The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genet. **10**: e1003948.

Lopez, F.B. et al. (2021). Gene dosage compensation of rRNA transcript levels in Arabidopsis thaliana lines with reduced ribosomal gene copy number. Plant Cell **33**: 1135–1150.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**: 550.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature **389**: 251–260.

Martienssen, R.A. and Colot, V. (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science **293**: 1070–1074.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. **17**: 10.

Martinez, G., Wolff, P., Wang, Z., Moreno-Romero, J., Santos-González, J., Conze, L.L., DeFraia, C., Slotkin, R.K., and Köhler, C. (2018). Paternal easiRNAs regulate parental genome dosage in Arabidopsis. Nat. Genet. **50**: 193–198.

Mathieu, O., Probst, A.V., and Paszkowski, J. (2005). Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. EMBO J. **24**: 2783–2791.

Matzke, M.A. and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. **15**: 394–408.

McKinlay, A., Podicheti, R., Wendte, J.M., Cocklin, R., and Rusch, D.B. (2018). RNA polymerases IV and V influence the 3' boundaries of Polymerase II transcription units in Arabidopsis. RNA Biol. **15**: 269–279.

Mi, S. et al. (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell **133**: 116–127.

Montgomery, S.A. et al. (2020). Chromatin organization in early land plants reveals an ancestral association between H3K27me3, transposons, and constitutive heterochromatin. Curr. Biol. **30**: 573-588.e7.

Morel, J.B., Mourrain, P., Béclin, C., and Vaucheret, H. (2000). DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. Curr. Biol. **10**: 1591–1594.

Moritoh, S., Eun, C.-H., Ono, A., Asao, H., Okano, Y., Yamaguchi, K., Shimatani, Z., Koizumi, A., and Terada, R. (2012). Targeted disruption of an orthologue of DOMAINS REARRANGED METHYLASE 2, OsDRM2, impairs the growth of rice plants by abnormal DNA methylation. Plant J. **71**: 85–98.

Narjala, A., Nair, A., Tirumalai, V., Hari Sundar, G.V., and Shivaprasad, P.V. (2020). A conserved sequence signature is essential for robust plant miRNA biogenesis. Nucleic Acids Res. **48**: 3103–3118.

Nuthikattu, S., McCue, A.D., Panda, K., Fultz, D., DeFraia, C., Thomas, E.N., and Slotkin, R.K. (2013). The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. Plant Physiol. **162**: 116–131.

Onodera, Y., Haag, J.R., Ream, T., Costa Nunes, P., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell **120**: 613–622.

Ossowski, S., Schwab, R., and Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. Plant J. **53**: 674–690.

Panda, K., McCue, A.D., and Slotkin, R.K. (2020). Arabidopsis RNA Polymerase IV generates 21-22 nucleotide small RNAs that can participate in RNA-directed DNA methylation and may regulate genes. Philos. Trans. R. Soc. Lond. B Biol. Sci. **375**: 20190417.

Parent, J.-S., Cahn, J., Herridge, R.P., Grimanelli, D., and Martienssen, R.A. (2021). Small RNAs guide histone methylation in Arabidopsis embryos. Genes Dev. **35**: 841–846.

Pedersen, J.F., Bean, S.R., Funnell, D.L., and Graybosch, R.A. (2004). Rapid iodine staining techniques for identifying the waxy phenotype in sorghum grain and waxy genotype in sorghum pollen. Crop Sci. **44**: 764–767.

Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I.J. (2018). Genome size diversity and its impact on the evolution of land plants. Genes (Basel) **9**.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**: 841–842.

Ramanathan, V. and Veluthambi, K. (1995). Transfer of non-T-DNA portions of the Agrobacterium tumefaciens Ti plasmid pTiA6 from the left terminus of TL-DNA. Plant Mol. Biol. **28**: 1149–1154.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. **42**: W187-91.

Roszak, P. and Köhler, C. (2011). Polycomb group proteins are required to couple seed coat initiation to fertilization. Proc. Natl. Acad. Sci. U. S. A. **108**: 20826–20831.

Saleh, A., Alvarez-Venegas, R., and Avramova, Z. (2008). An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in Arabidopsis plants. Nat. Protoc. **3**: 1018–1025.

Satyaki, P.R.V. and Gehring, M. (2019). Paternally-acting canoncial RNA-directed DNA methylation pathway genes sensitize Arabidopsis endosperm to paternal genome dosage. Plant Cell **31**: 1563–1578.

Schier, A.C. and Taatjes, D.J. (2020). Structure and mechanism of the RNA polymerase II transcription machinery. Genes Dev. **34**: 465–488.

Shivaprasad, P.V., Chen, H.-M., Patel, K., Bond, D.M., Santos, B.A.C.M., and Baulcombe, D.C. (2012). A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. Plant Cell **24**: 859–874.

Shivaprasad, P.V., Thillaichidambaram, P., Balaji, V., and Veluthambi, K. (2006). Expression of full-length and truncated Rep genes from Mungbean yellow mosaic virus-Vigna inhibits viral replication in transgenic tobacco. Virus Genes **33**: 365–374.

Song, L., Koga, Y., and Ecker, J.R. (2016). Profiling of transcription factor binding events by chromatin immunoprecipitation sequencing (ChIP-seq). Curr. Protoc. Plant Biol. **1**: 293–306.

Soppe, W.J.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. EMBO J. **21**: 6549–6559.

Sridevi, G., Sabapathi, N., Meena, P., Nandakumar, R., Samiyappan, R., Muthukrishnan, S., and Veluthambi, K. (2003). Transgenic indica rice variety Pusa basmati 1 constitutively expressing a rice chitinase gene exhibits enhanced resistance to Rhizoctonia solani. J. Plant Biochem. Biotechnol. **12**: 93–101.

Stocks, M.B., Mohorianu, I., Beckers, M., Paicu, C., Moxon, S., Thody, J., Dalmay, T., and Moulton, V. (2018). The UEA sRNA Workbench (version 4. 4): a comprehensive suite of tools for analyzing miRNAs and sRNAs.

Swetha, C., Basu, D., Pachamuthu, K., Tirumalai, V., Nair, A., Prasad, M., and Shivaprasad, P.V. (2018). Major domestication-related phenotypes in Indica rice are due to loss of miRNA-mediated laccase silencing. Plant Cell **30**: 2649–2662.

Tirumalai, V., Prasad, M., and Shivaprasad, P.V. (2020). RNA blot analysis for the detection and quantification of plant MicroRNAs. J. Vis. Exp.

Tirumalai, V., Swetha, C., Nair, A., Pandit, A., and Shivaprasad, P.V. (2019). miR828 and miR858 regulate VvMYB114 to promote anthocyanin and flavonol accumulation in grapes. J. Exp. Bot. **70**: 4775–4792.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. **7**: 562–578.

Trujillo, J.T., Seetharam, A.S., Hufford, M.B., Beilstein, M.A., and Mosher, R.A. (2018). Evidence for a unique DNA-dependent RNA polymerase in cereal crops. Mol. Biol. Evol. **35**: 2454–2462.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.-L., Bartel, D.P., and Crété, P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol. Cell **16**: 69–79.

Voinnet, O., Vain, P., Angell, S., and Baulcombe, D.C. (1998). Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. Cell **95**: 177–187.

Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J. (1993). Arabidopsis thaliana DNA methylation mutants. Science **260**: 1926–1928.

Wang, F., Yano, K., Nagamatsu, S., Inari-Ikeda, M., Koketsu, E., Hirano, K., Aya, K., and Matsuoka, M. (2020a). Genome-wide expression quantitative trait locus studies facilitate isolation of causal genes controlling panicle structure. Plant J. **103**: 266–278.

Wang, L., Zeng, L., Zheng, K., Zhu, T., Yin, Y., Xu, D., Zhan, H., Wu, Y., and Yang, D.-L. (2020b). Global reinforcement of DNA methylation through enhancement of RNA-directed DNA methylation ensures sexual reproduction in rice. bioRxiv.

Wang, Z., Butel, N., Santos-González, J., Borges, F., Yi, J., Martienssen, R.A., Martinez, G., and Köhler, C. (2020c). Polymerase IV plays a crucial role in pollen development in Capsella. Plant Cell **32**: 950–966.

Wang, Z., Butel, N., Santos-González, J., Simon, L., Wärdig, C., and Köhler, C. (2021). Transgenerational effect of mutants in the RNA-directed DNA methylation pathway on the triploid block in Arabidopsis. Genome Biol. **22**: 141.

Warthmann, N., Chen, H., Ossowski, S., Weigel, D., and Hervé, P. (2008). Highly specific gene silencing by artificial miRNAs in rice. PLoS One **3**: e1829.

Wei, L., Gu, L., Song, X., Cui, X., Lu, Z., Zhou, M., Wang, L., Hu, F., Zhai, J., Meyers, B.C., and Cao, X. (2014). Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. Proc. Natl. Acad. Sci. U. S. A. **111**: 3877–3882.

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

Wu, L., Zhang, Q., Zhou, H., Ni, F., Wu, X., and Qi, Y. (2009). Rice MicroRNA effector complexes and targets. Plant Cell **21**: 3421–3435.

Wu, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C., and Qi, Y. (2010). DNA methylation mediated by a microRNA pathway. Mol. Cell **38**: 465–475.

Xu, L., Yuan, K., Yuan, M., Meng, X., Chen, M., Wu, J., Li, J., and Qi, Y. (2020). Regulation of rice tillering by RNA-directed DNA methylation at miniature inverted-repeat transposable elements. Mol. Plant **13**: 851–863.

Yaari, R., Katz, A., Domb, K., Harris, K.D., Zemach, A., and Ohad, N. (2019). RdDM-independent de novo and heterochromatin DNA methylation by plant CMT and DNMT3 orthologs. Nat. Commun. **10**: 1613.

Yamauchi, T., Johzuka-Hisatomi, Y., Terada, R., Nakamura, I., and Iida, S. (2014). The MET1b gene encoding a maintenance DNA methyltransferase is indispensable for normal development in rice. Plant Mol. Biol. **85**: 219–232.

Yelagandula, R. et al. (2014). The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. Cell **158**: 98–109.

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics **31**: 2382–2383.

Yu, Y., Ouyang, Y., and Yao, W. (2018). shinyCircos: an R/Shiny application for interactive creation of Circos plot. Bioinformatics **34**: 1229–1231.

Yu, Y., Yao, W., Wang, Y., and Huang, F. (2019). ShinyChromosome: An R/Shiny application for interactive creation of non-circular plots of whole genomes. Genomics Proteomics Bioinformatics **17**: 535–539.

Zhai, J. et al. (2015). A one precursor one siRNA model for Pol IV-dependent siRNA biogenesis. Cell **163**: 445–455.

Zhang, H. et al. (2013). DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proc. Natl. Acad. Sci. U. S. A. **110**: 8290–8295.

Zhang, Q., Wang, D., Lang, Z., He, L., Yang, L., Zeng, L., Li, Y., Zhao, C., Huang, H., Zhang, H., Zhang, H., and Zhu, J.-K. (2016). Methylation interactions in Arabidopsis hybrids require RNA-directed DNA methylation and are influenced by genetic variation. Proc. Natl. Acad. Sci. U. S. A. **113**: E4248–E4256.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biol. **5**: e129.

Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J.-Y., and Chen, X. (2009). Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes Dev. **23**: 2850–2860.

Zheng, K., Wang, L., Zeng, L., Xu, D., Guo, Z., Gao, X., and Yang, D.-L. (2021). The effect of RNA polymerase V on 24-nt siRNA accumulation depends on DNA methylation contexts and histone modifications in rice. Proc. Natl. Acad. Sci. U. S. A. **118**: e2100709118.

Zhong, Z., Feng, S., Duttke, S.H., Potok, M.E., Zhang, Y., Gallego-Bartolomé, J., Liu, W., and Jacobsen, S.E. (2021). DNA methylation-linked chromatin accessibility affects genomic architecture in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. **118**: e2023347118.

Zhou, M., Palanca, A.M.S., and Law, J.A. (2018). Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. Nat. Genet. **50**: 865–873.

Zhou, S., Liu, X., Zhou, C., Zhou, Q., Zhao, Y., Li, G., and Zhou, D.-X. (2016). Cooperation between the H3K27me3 chromatin Mark and non-CG methylation in epigenetic regulation. Plant Physiol. **172**: 1131–1141.

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