Supplemental Figures S1-29

Supplemental Figure S1. Knockdown of RNA pol IV in rice results in pleiotropic phenotypes.

Supplemental Figure S2. Distinct sets of Agrobacterium mediated transformation events yielded knockdown lines with similar phenotypic defects.

Supplemental Figure S3. Specific lineage of Pol IV knockdown plants showed severe defects that are not due to transgenesis or knockdown dosage suggesting possible epimutation.

Supplemental Figure S4. Loss of pol IV causes specific loss of repeats and transposon associated sRNAs.

Supplemental Figure S5. RNA pol IV is responsible for maintaining DNA methylation, repeat silencing and contributes to gene regulation.

Supplemental Figure S6. Pol IV silences transposons and repeats.

Supplemental Figure S7. ChIP – Sequencing exhibits concordant signals across replicates.

Supplemental Figure S8. Representative immunofluorescence micrographs of WT nuclei.

Supplemental Figure S9. Representative immunofluorescence micrographs of *nrpd1-kd* nuclei.

Supplemental Figure S10. Loss of Pol IV impacts occupancy of histone marks genomewide.

Supplemental Figure S11. Loss of pol IV impacts occupancy of pol II and MNase accessibility.

Supplemental Figure S12. Pol IV complex suppresses peculiar sRNA production from several loci.

Supplemental Figure S13. Pol IV-dependent and Pol IV-suppressed sRNA clusters are of similar length.

Supplemental Figure S14. Pol IV-dependent and Pol IV-suppressed sRNA bins are distinct in producing sRNAs with specific sizes.

Supplemental Figure S15. Pol IV-dependent and Pol IV-suppressed sRNA bins are not due to oversampling and library normalisation.

Supplemental Figure S16. Pol IV-suppressed sRNAs from panicle are non-uniformly distributed across the genome.

Supplemental Figure S17. Pol IV-suppressed sRNAs from anther are non-uniformly distributed across the genome.

Supplemental Figure S18. Pol IV-suppressed sRNAs from endosperm are non-uniformly distributed across the genome.

Supplemental Figure S19. Pol IV-suppressed sRNAs have significant overlap with coding regions.

Supplemental Figure S20. Genes overlapping with suppressed sRNAs exhibit reduced silencing compensation by H3K27me3.

Supplemental Figure S21. Pol IV-suppressed sRNAs from *Arabidopsis* inflorescence are non-uniformly distributed across the genome similar to rice tissues.

Supplemental Figure S22. Pol IV-suppressed sRNAs from *Arabidopsis* are characteristically similar to that of rice.

Supplemental Figure S23. Canonical RdDM does not potentiate Pol IV-suppressed sRNAs in rice.

Supplemental Figure S24. Canonical RdDM does not potentiate Pol IV-suppressed sRNAs in *Arabidopsis*.

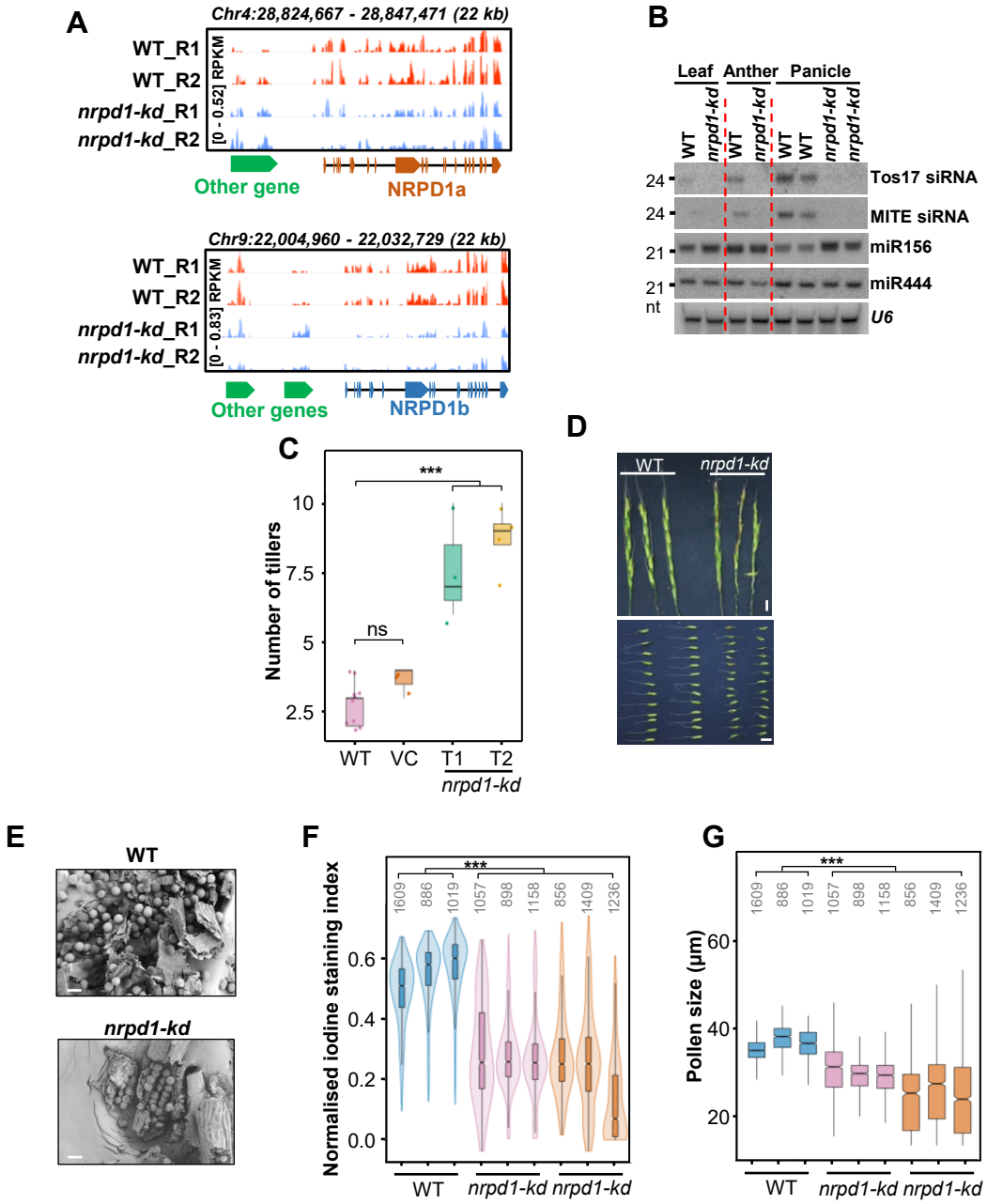
Supplemental Figure S25. Net Pol II occupancy over Pol IV-suppressed sRNA clusters does not change.

Supplemental Figure S26. Distinct H3K27me3 modifications at pol IV-dependent and suppressed sRNA clusters.

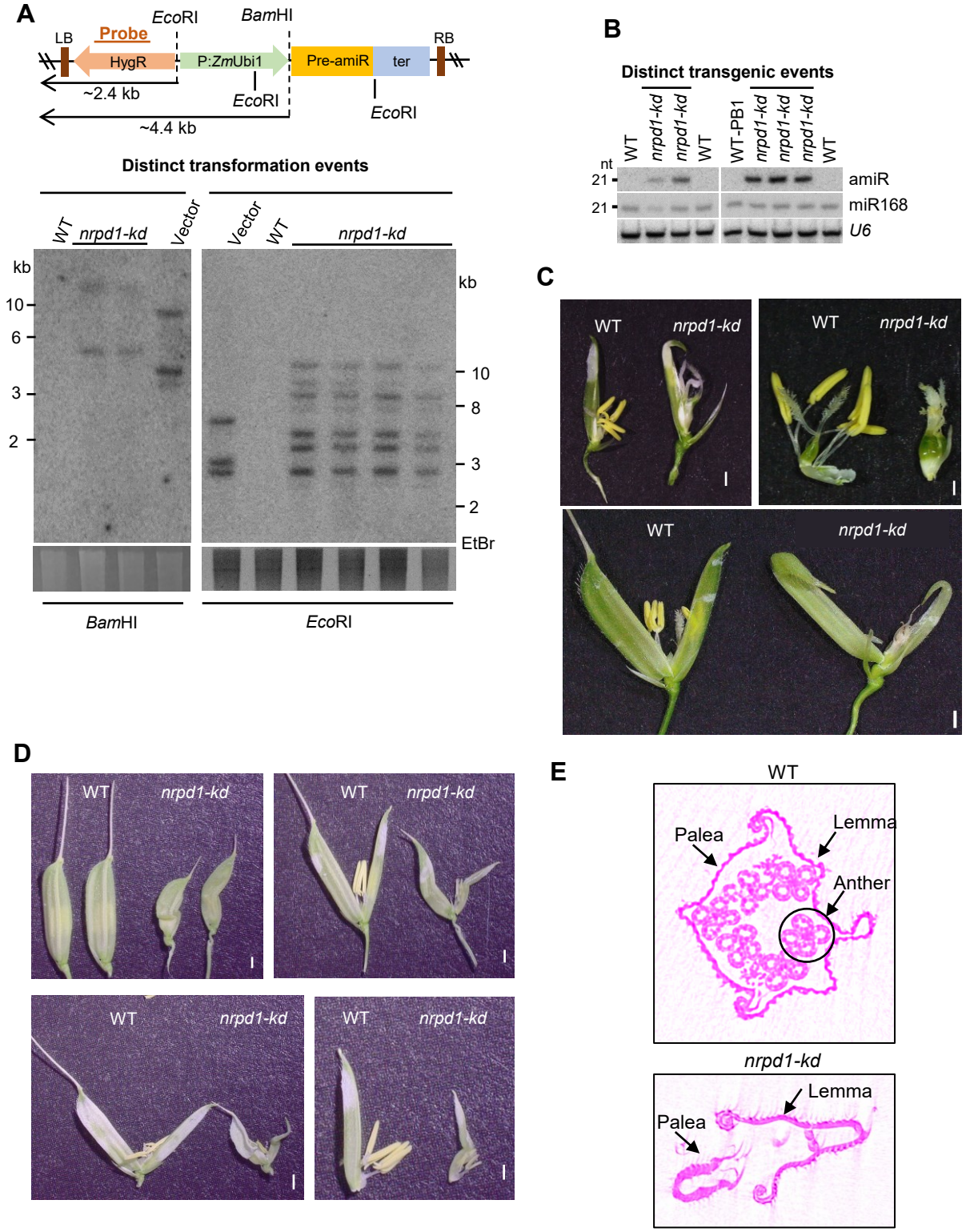
Supplemental Figure S27. Analyses of rice AGO-IP and PARE datasets exhibits expected profiles.

Supplemental Figure S28. Genes targeted by Pol IV-suppressed sRNAs show reduction in expression in *kd* lines.

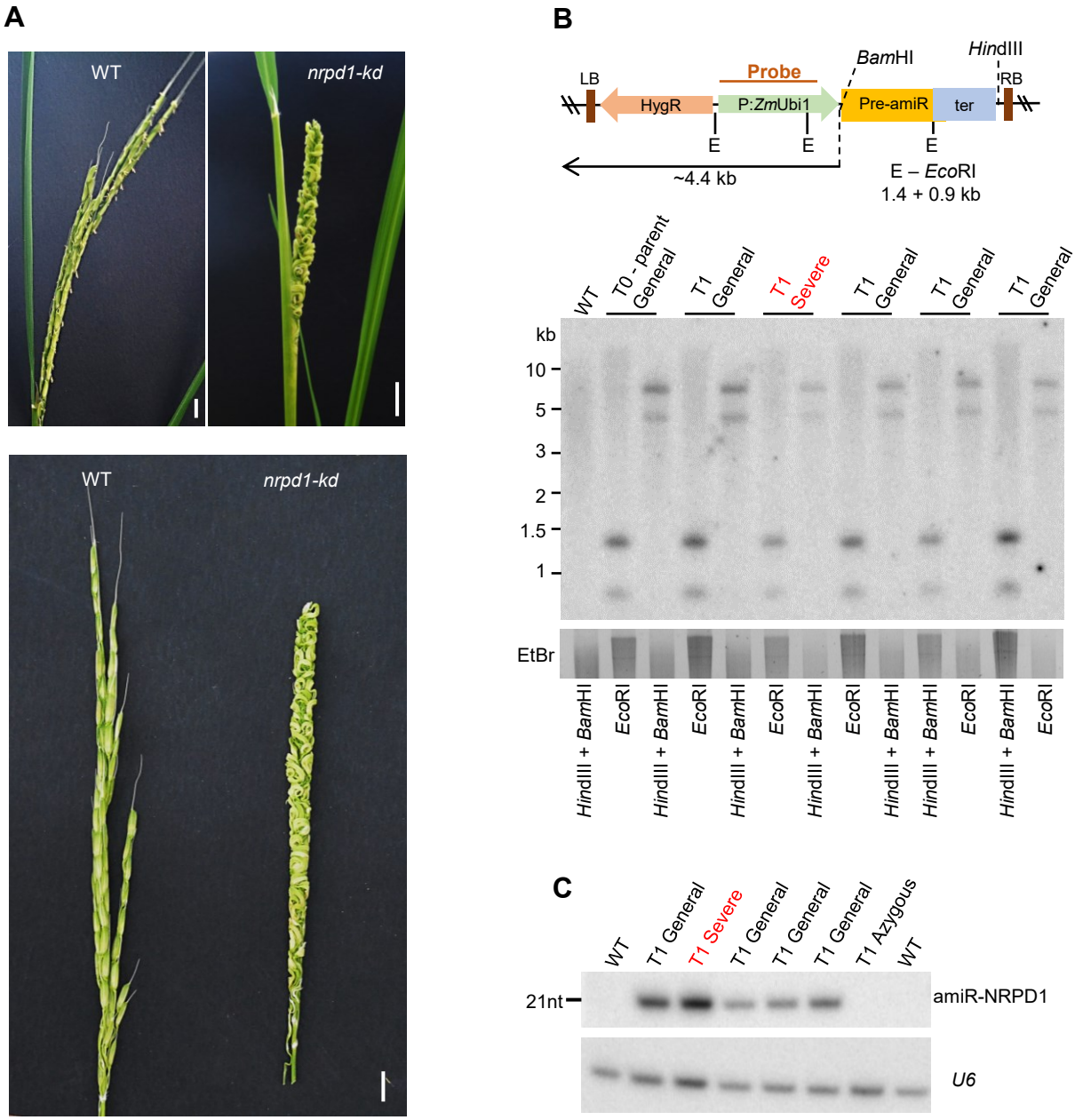
Supplemental Figure S29. *Arabidopsis* Pol IV-suppressed sRNAs are not effectively loaded into AGO1 and unlikely to target genes.



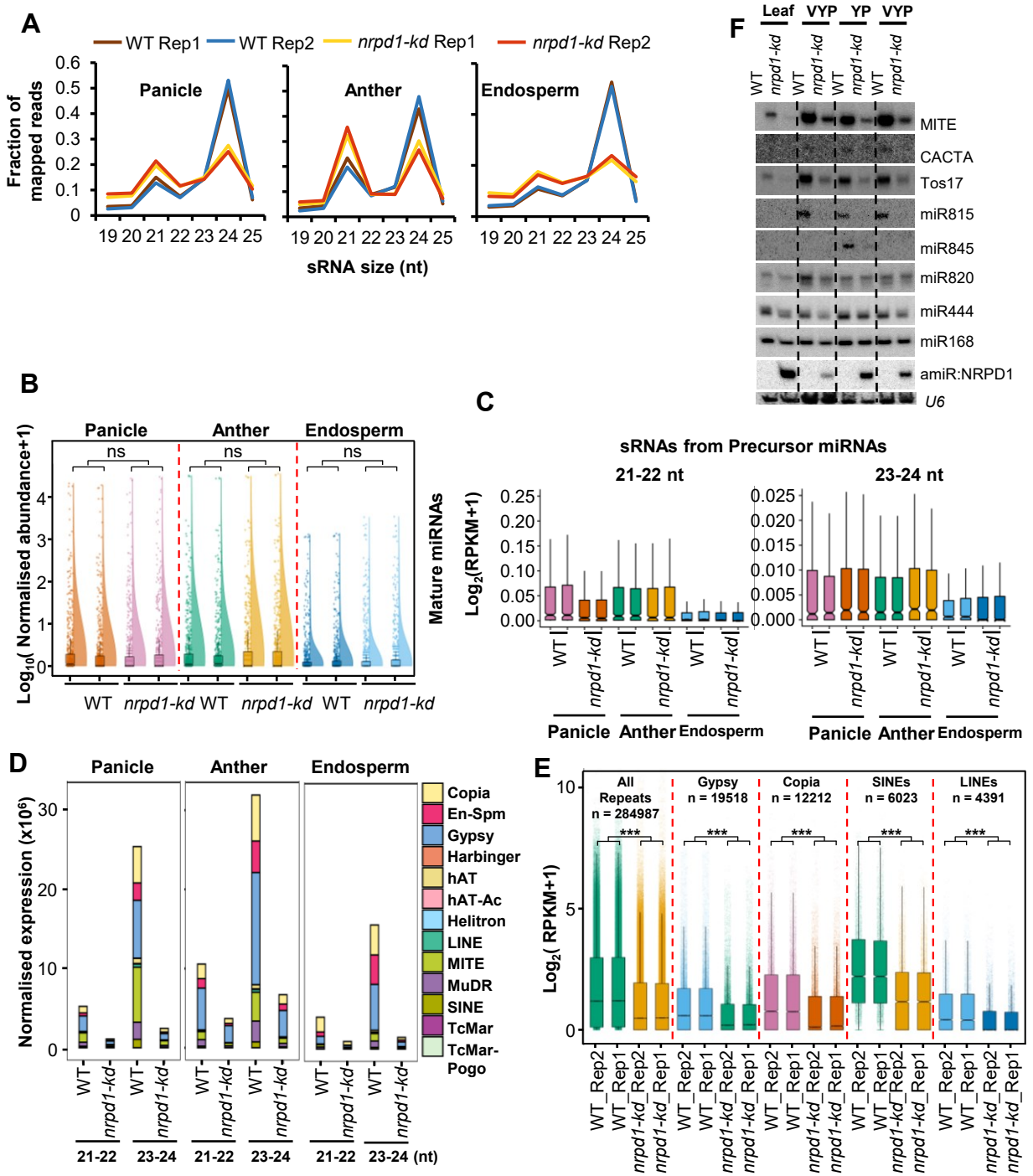
Supplemental Figure S1. Knockdown of RNA Pol IV in rice results in pleiotropic phenotypes. *(A)* IGV screenshots of panicle RNA-seq coverage over the NRPD1a and NRPD1b loci in WT and *kd. (B)* sRNA northern blots indicating loss of 24nt siRNA in different tissues. miR444 and miR156 were examples of miRNAs. *U6* was used as loading control. *(C)* 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). *(D)* Images showing the panicle (top) and the individual florets (bottom) of WT and *nrpd1-kd* plants. Scale bar:1 cm. *(E)* Representative SEM images of pollen grains from dehisced anther. Scale: 40 μm. *(F)* Violin-box plots showing the distribution of normalised iodine staining index of pollen grains. Numbers in grey represent the number of pollen grains examined. (Mann-Whitney U test; *** p-value < 2.2×10^{-16}). *(G)* Distribution of pollen sizes in WT and *nrpd1-kd* plants. The stained light microscopy images are quantified using custom ImageJ scripts to detect the sizes and degree of staining of pollen grains. (Mann-Whitney U test; *** p-value < 2.2×10^{-16}).



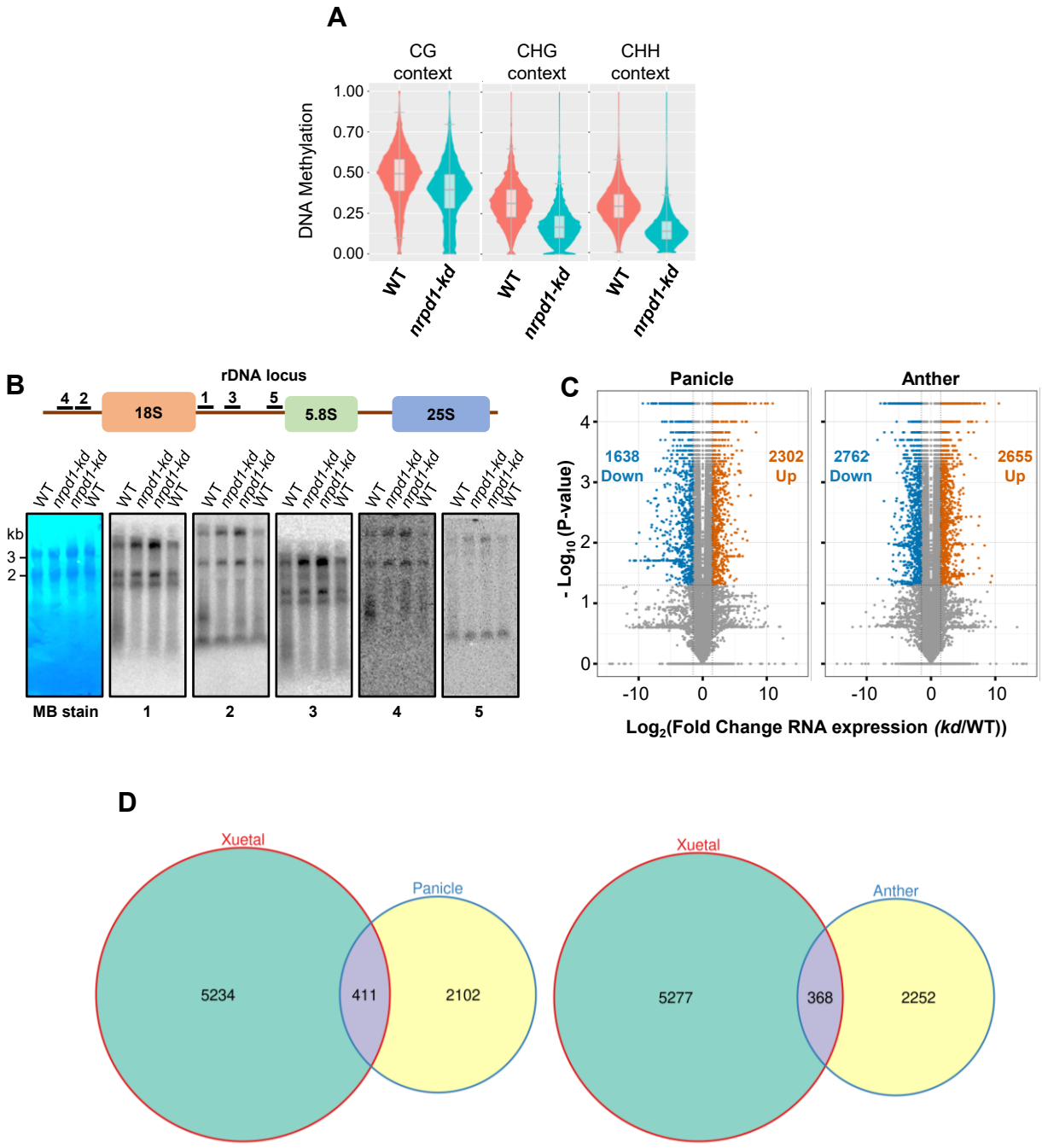
Supplemental Figure S2. Distinct sets of Agrobacterium mediated transformation events yielded knockdown lines with similar phenotypic defects. *(A)* DNA blots showing the T-DNA junction fragments corresponding to digestion of DNA with mentioned restriction sites. The distinct transformation events are marked by the line above the blots. Minimum lengths of junction fragment sizes are mentioned by arrows on the T-DNA map above. Probe (HygR) region is mentioned in the map. *(B)* sRNA northern blots indicating the accumulation of amiRNA in transgenic knockdown (*nrpd1-kd*) plants obtained from independent transgenic events. miRNA168 is used as miRNA control and *U6* as loading control. *(C – D)* Defects in the reproductive structures obtained in the plants from two independent transgenic events as seen by distinct T-DNA integration pattern in *(A)*. Scale bars correspond to 1 mm. *(E)* Micro-computed tomographs (micro-CT) of equally developed florets in WT and *kd* lines.



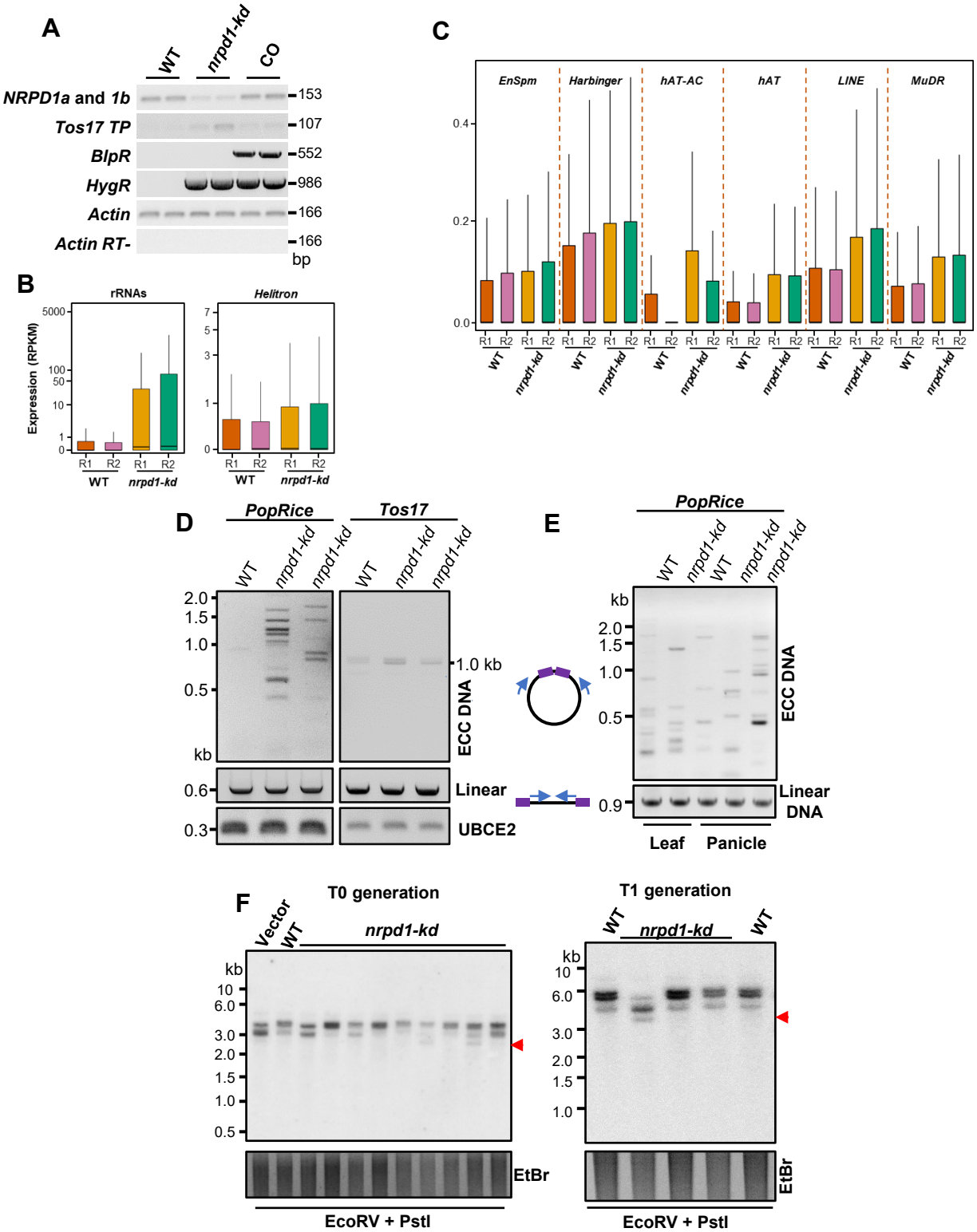
Supplemental Figure S3. Specific lineage of Pol IV knockdown plants showed severe defects that are not due to transgenesis or knockdown dosage suggesting possible epimutation. *(A)* Phenotype of panicle displaying the severe defect in T1 generation (T1 severe) that is more drastic than the other offsprings of the same parent. Scale bar: 1 cm. *(B)* DNA blots showing the T-DNA profiles of the T1 severe line compared to that of the parent and other offsprings of the same parent that showed characteristic (general) *kd* like phenotypes. T-DNA map shows the location of the restriction sites used and the expected band sizes upon enzymatic digestion. *(C)* sRNA northern blots showing the levels of artificial miRNA (amiR-NRPD1) levels in the severe as well as the general phenotype displaying lines.



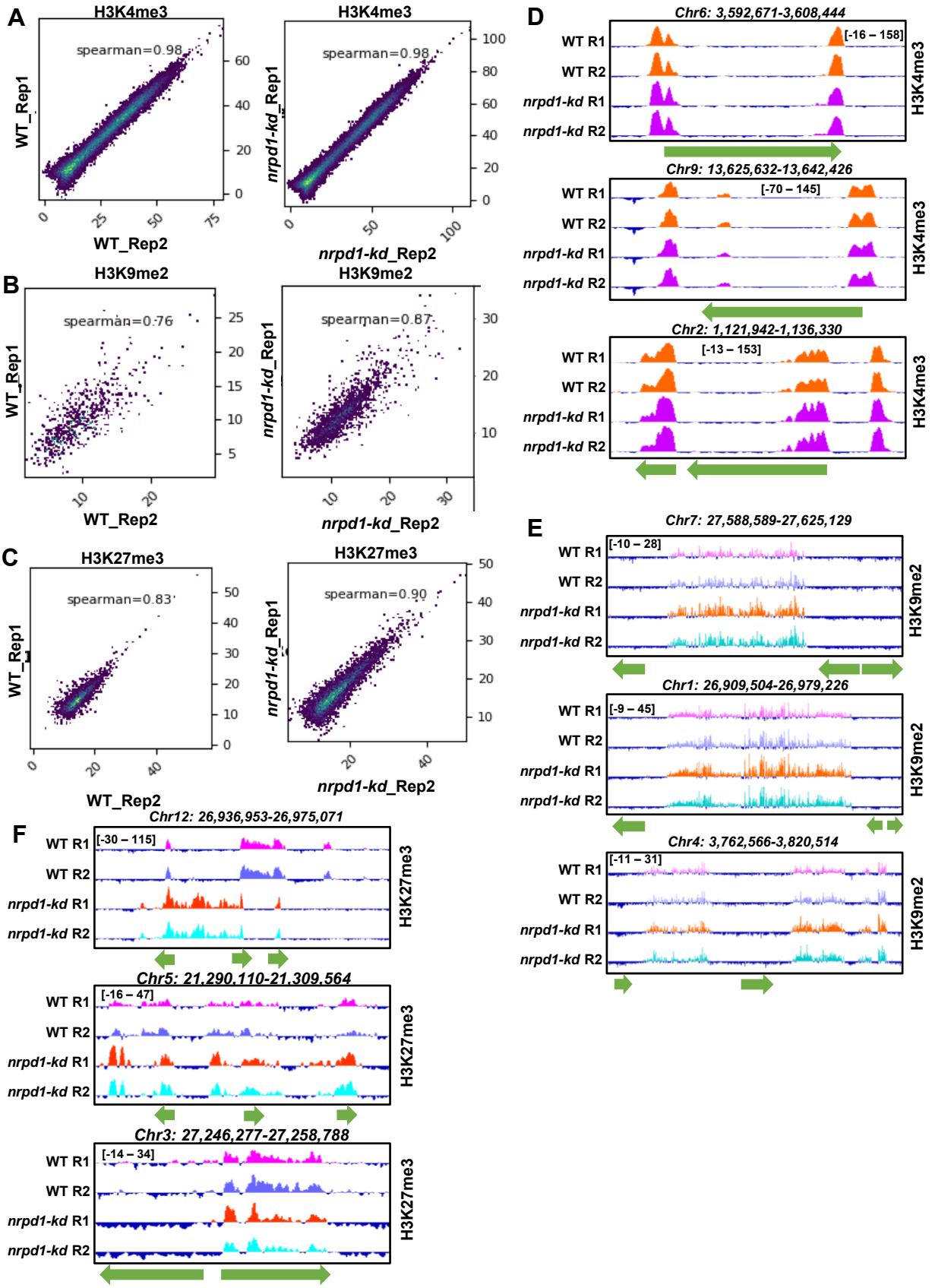
Supplemental Figure S4. Loss of Pol IV causes specific loss of repeats and transposon associated sRNAs. *(A)* Comparison of relative accumulation of sRNAs of different sizes profiled by small RNA sequencing performed in replicates from different tissues. *(B)* Normalised levels of mature miRNAs in the different tissues, in replicates, of WT and *kd* plants as estimated by miRProf. Rain cloud plots represent the median and the distribution of the datapoints (n = 633). (Mann-Whitney U test; ns:p-value > 0.05). *(C)* Boxplots showing the 21-22 nt and 23-24 nt sRNA abundance counted over the miRbase annotated miRNA precursor loci from rice (n = 586). *(D)* Abundance of small RNAs from different transposon categories. 21-22 nt and 23-24 nt sizes of sRNAs from corresponding replicates were merged. *(E)* Abundance of sRNAs from different TE categories in replicates. (Mann-Whitney U test; *** p-value < 2.2×10^{-16}). *(F)* sRNA northern blots showing the abundance of sRNAs of different categories (miRNAs, siRNAs and TE derived miRNA845) in WT and *kd* tissues – leaves, very young panicle (VYP) and pre-emerged young panicle (YP).



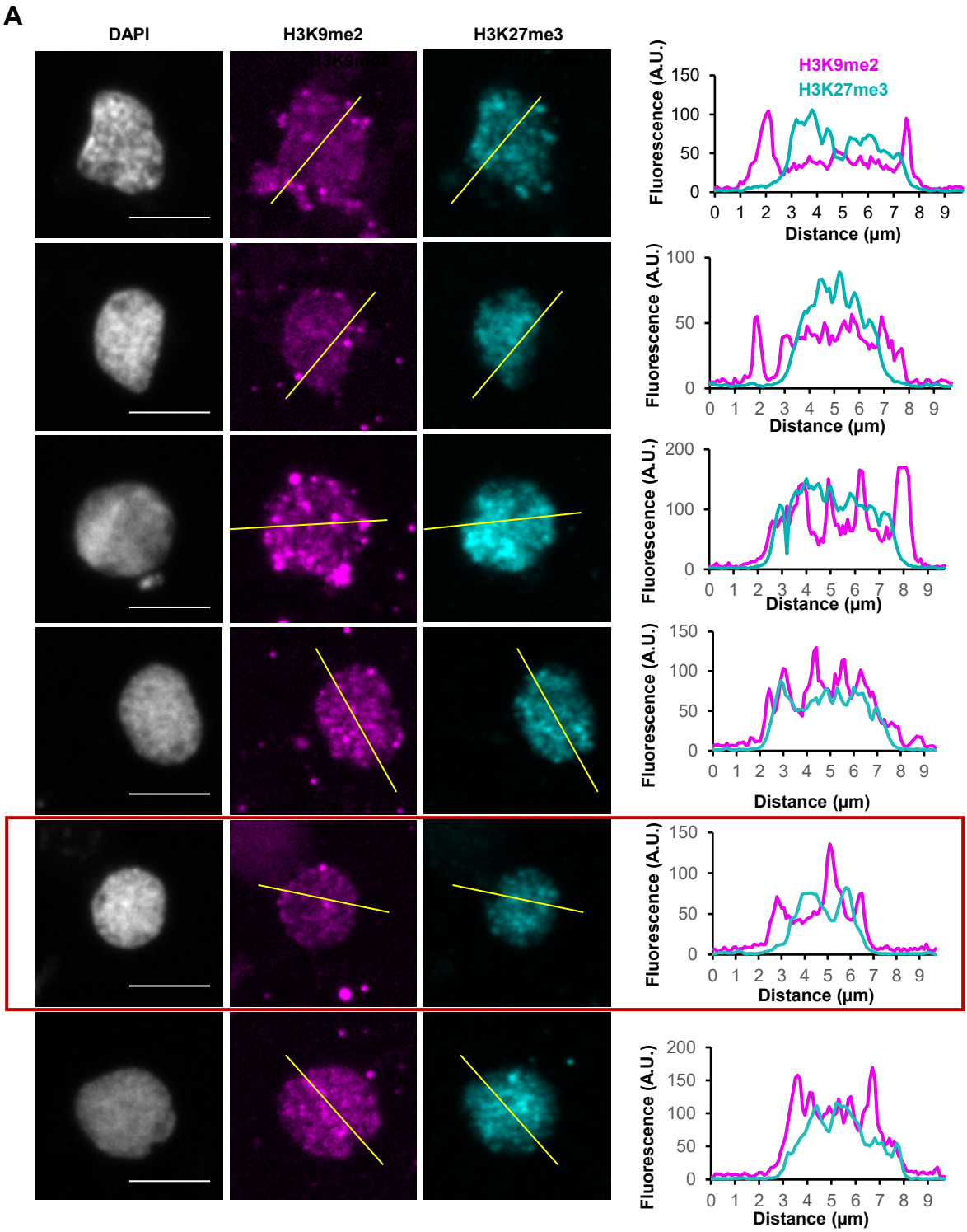
Supplemental Figure S5. RNA Pol IV is responsible for maintaining DNA methylation, repeat silencing and contributes to gene regulation. *(A)* Box-violin plots showing DNA methylation levels at transposons and repeat features (285215 loci) in WT and kd lines. *(B)* Northern blots showing the abundance of rRNA precursors from WT and *kd* panicle. Methylene blue (MB) stained membrane is shown as loading control. *(C)* Volcano plots showing the number of significantly up-regullated ($Log_2FC > 2$) and downregulated ($Log_2FC < 2$) genes in *kd* compared to WT in panicle and anther. *(D)* Venn plots showing overlap of DEGs in shoot bases of *nrpd1-RNAi* (GSEGSE131319) and that of panicle/anther DEGs.



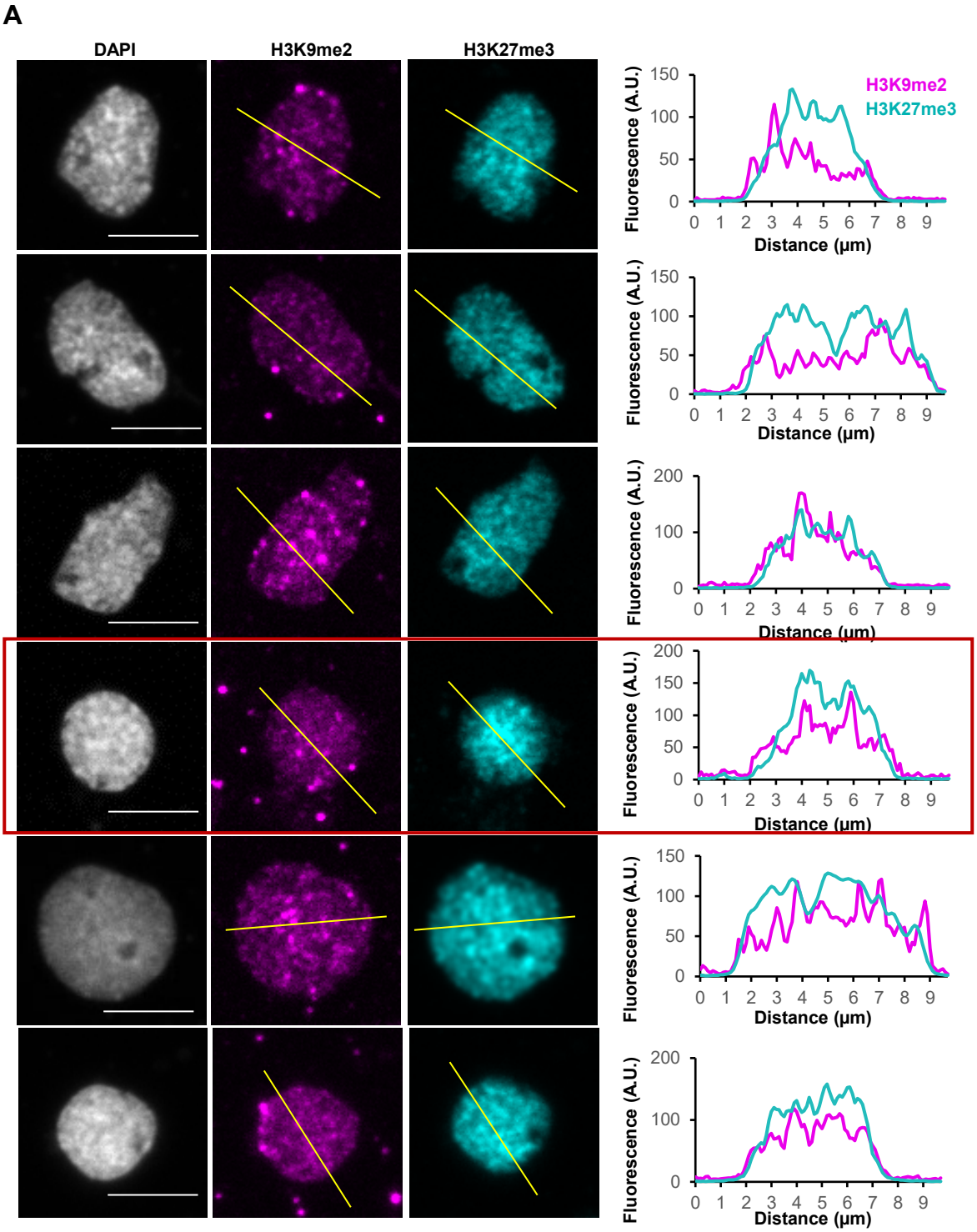
Supplemental Figure S6. Pol IV silences transposons and repeats. *(A)* Semi-quantitative RT-PCRs showing levels of the selection markers (HPTII-*HygR* and Bialaphos resistance-*BlpR*), *NRPD1a* and *NRPD1b* (primers binding a conserved region) and *Tos17 transposase (Tos7 TP)* in WT, *kd* and NRPD1b complementation (CO). *(B-C)* Boxplots showing the normalised abundance of transcripts from annotated repeat loci in the replicates (R1 and R2) of WT and *kd* plants. The Y-axis is scaled to inverse sine hyperbolic function of RPKM values. *(D)* PCRs showing accumulation of extrachromosomal circular (ECC) DNA intermediates of TEs *PopRice* and *Tos17* with intact DNA from panicle. UBCE2 fragment is used as linear DNA loading control. *(E)* PCRs showing accumulation of extrachromosomal circular (ECC) DNA intermediates of *PopRice* DNA from panicle pre-digested with Plasmid-safe™ DNase. *(F)* Transposon copy number Southern blot showing copy number variation (arrows) of LINE-1 transposon in both T0 and T1 generations in *kd*. Ethidium bromide (EtBr) staining served as control.



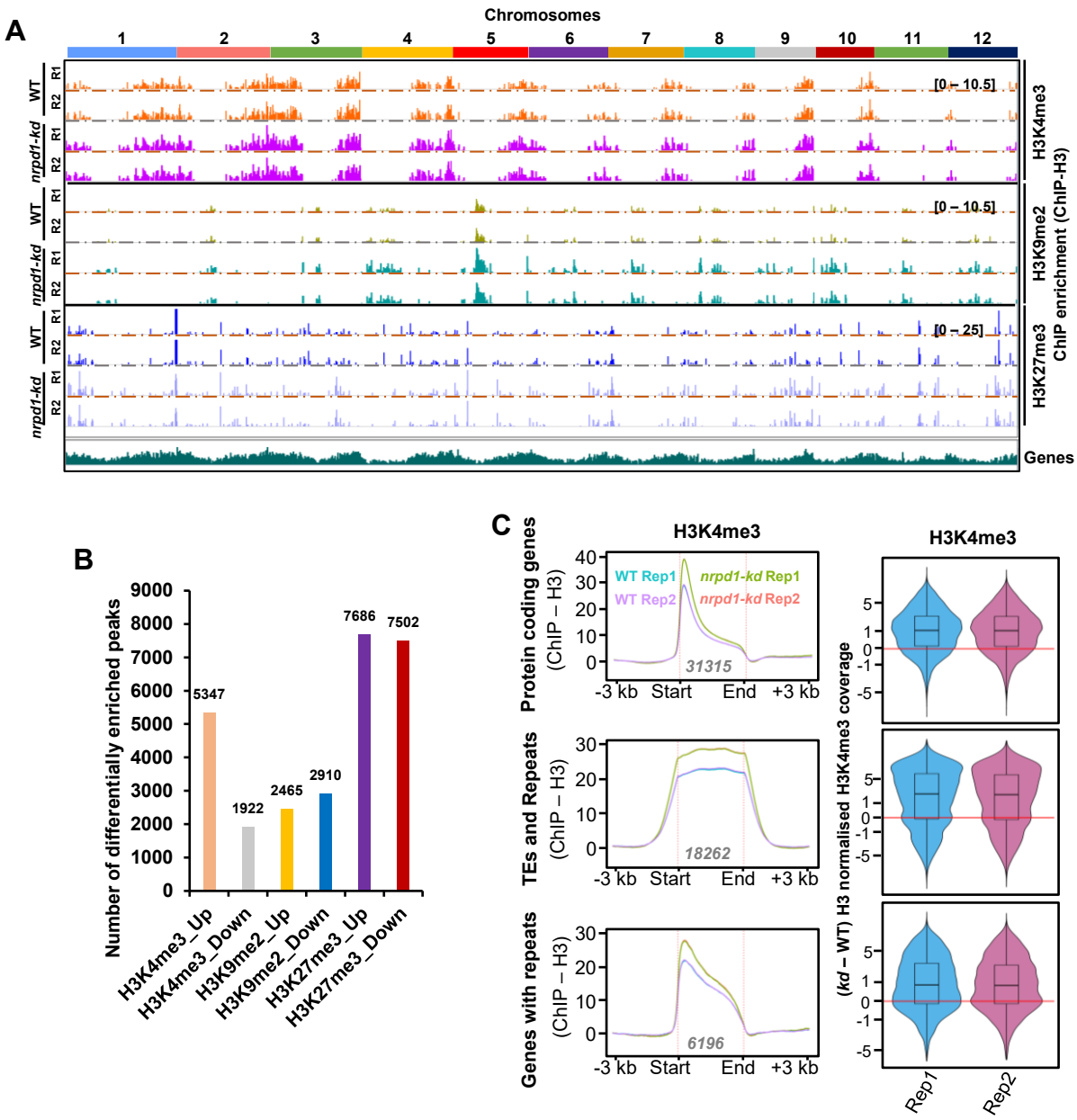
Supplemental Figure S7. ChIP – Sequencing exhibits concordant signals across replicates. *(A-C)* Scatter plots showing the ChIP signals normalised to H3 (ChIP-H3) over the combined sets of peaks identified by MACS2 in WT and *nrpd1-kd* for H3K4me3 *(A)*, H3K9me2 *(B)* and H3K27me3 *(C)*. Spearman correlation coefficients for the ChIP signals between the replicates are inside the plots for pairwise comparisons. *(D-F)* Genome browser screen shots of H3K4me3 *(D)*, H3K9me2 *(E)* and H3K27me3 *(F)* occupancy with normalised signals (ChIP-H3) depicted in square brackets. Genes are shown as green arrows.



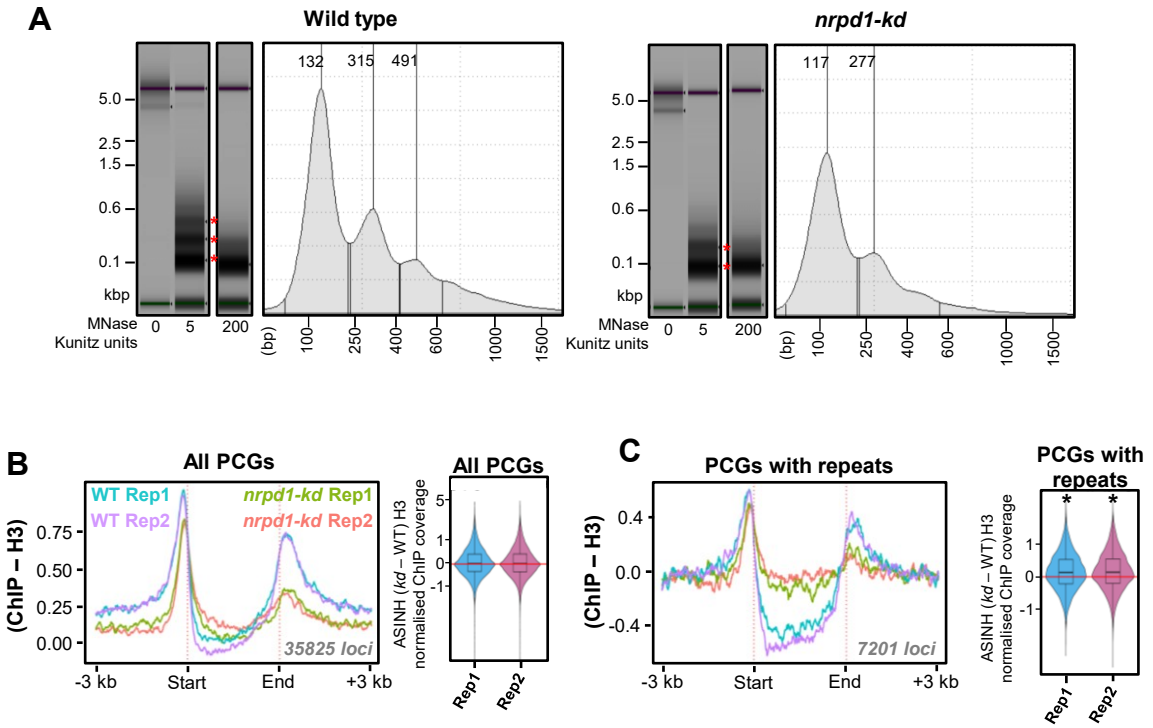
Supplemental Figure S8. Representative immunofluorescence micrographs of WT nuclei. Nuclei isolated from WT seedlings (14 days old) were stained with H3K9me2 and H3K27me3 specific antibodies. DAPI stained the DNA. Scale bar: 5 μm. Fluorescence signals across the region of interest (yellow line) was plotted. Red box shows the image presented in Fig. 2A, with its region of interest marked.



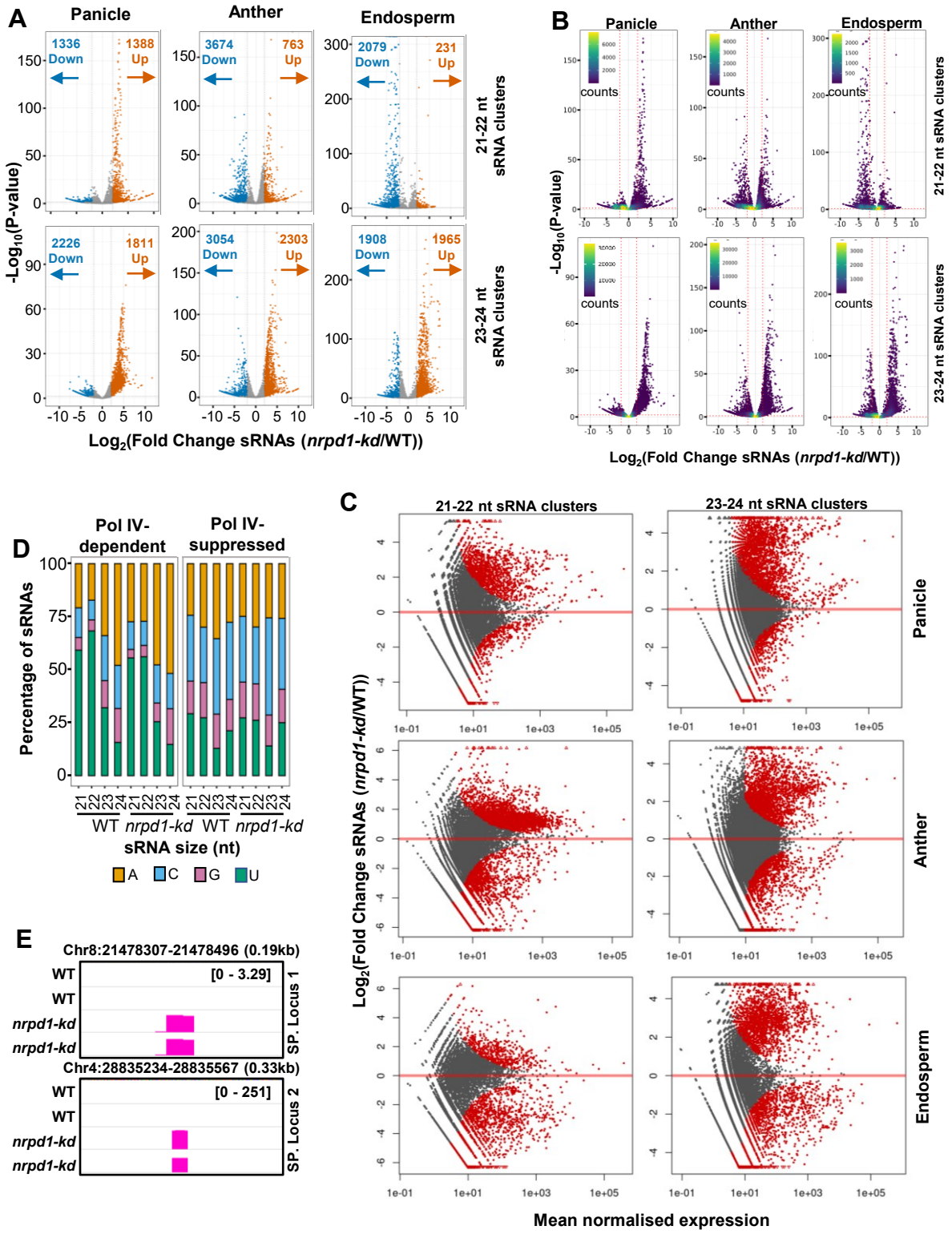
Supplemental Figure S9. Representative immunofluorescence micrographs of *nrpd1-kd* nuclei. Nuclei isolated from *nrpd1-kd* seedlings (14 days old) were stained with H3K9me2 and H3K27me3 specific antibodies. DAPI stained the DNA. Scale bar: 5 μm. Fluorescence signals across the region of interest (yellow line) was plotted. Red box shows the image presented in Fig. 2A, with its region of interest marked.



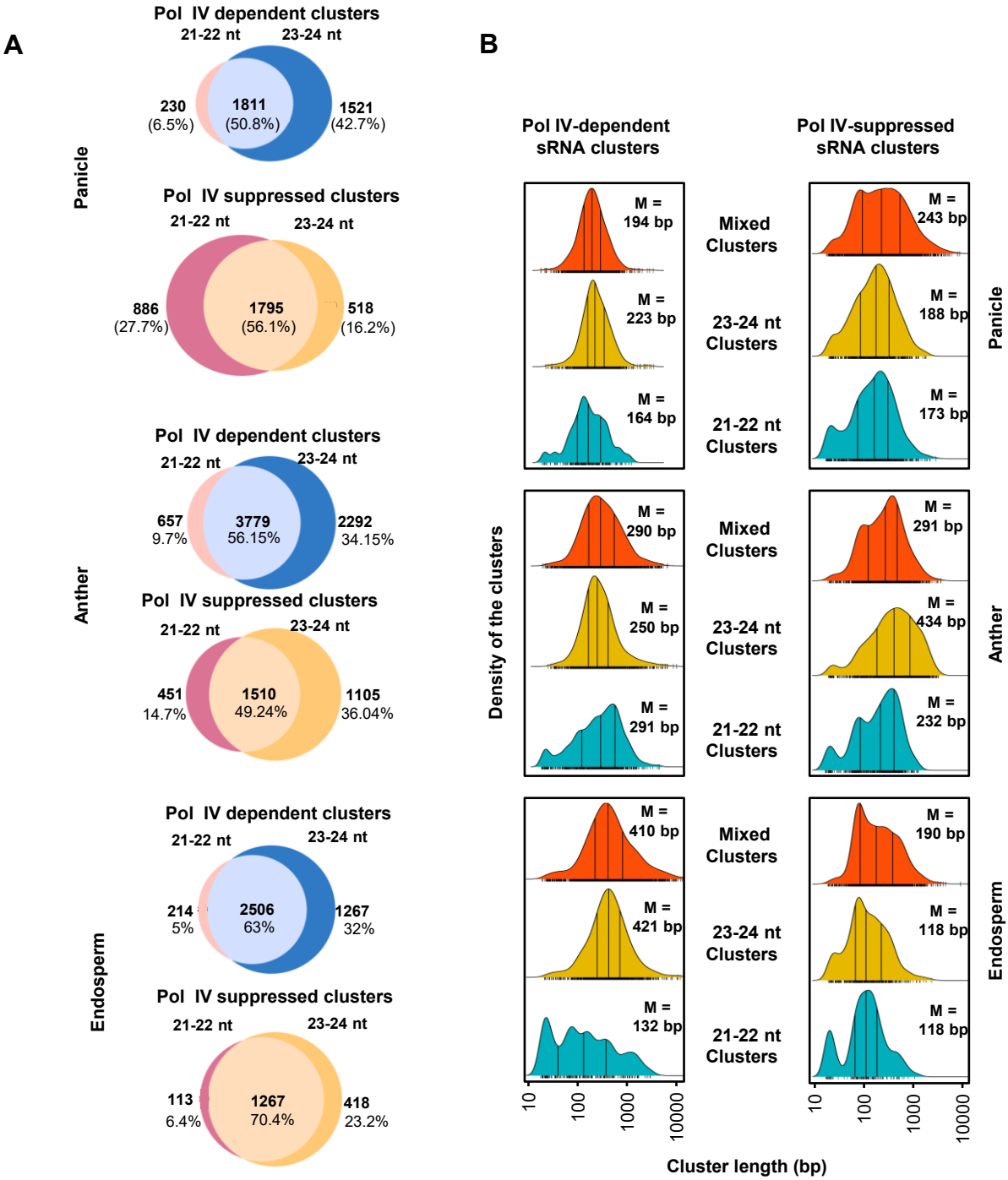
Supplemental Figure S10. Loss of Pol IV impacts occupancy of histone marks genomewide. *(A)* Genomewide profiles (across 12 chromosomes) of H3K4me3, H3K9me2 and H3K27me3 marks in WT and *kd* plants. ChIP enrichment (ChIP-H3) signals (positive) are depicted in square brackets. *(B)* Bar plots showing number of differentially enriched histone modification peaks in WT and *kd* lines for H3K4me3, H3K9me2 and H3K27me3. *(C)* Metaplots showing the levels of H3K4me3 enrichment (normalised to H3) over the annotated features overlapping with the identified H3K4me3 peaks. Numbers in grey depict the number of loci. Box-violin plots show the difference in enrichment in *kd* compared to WT. The Y-axis is scaled to inverse sine hyperbolic function of H3 ChIP normalised values.



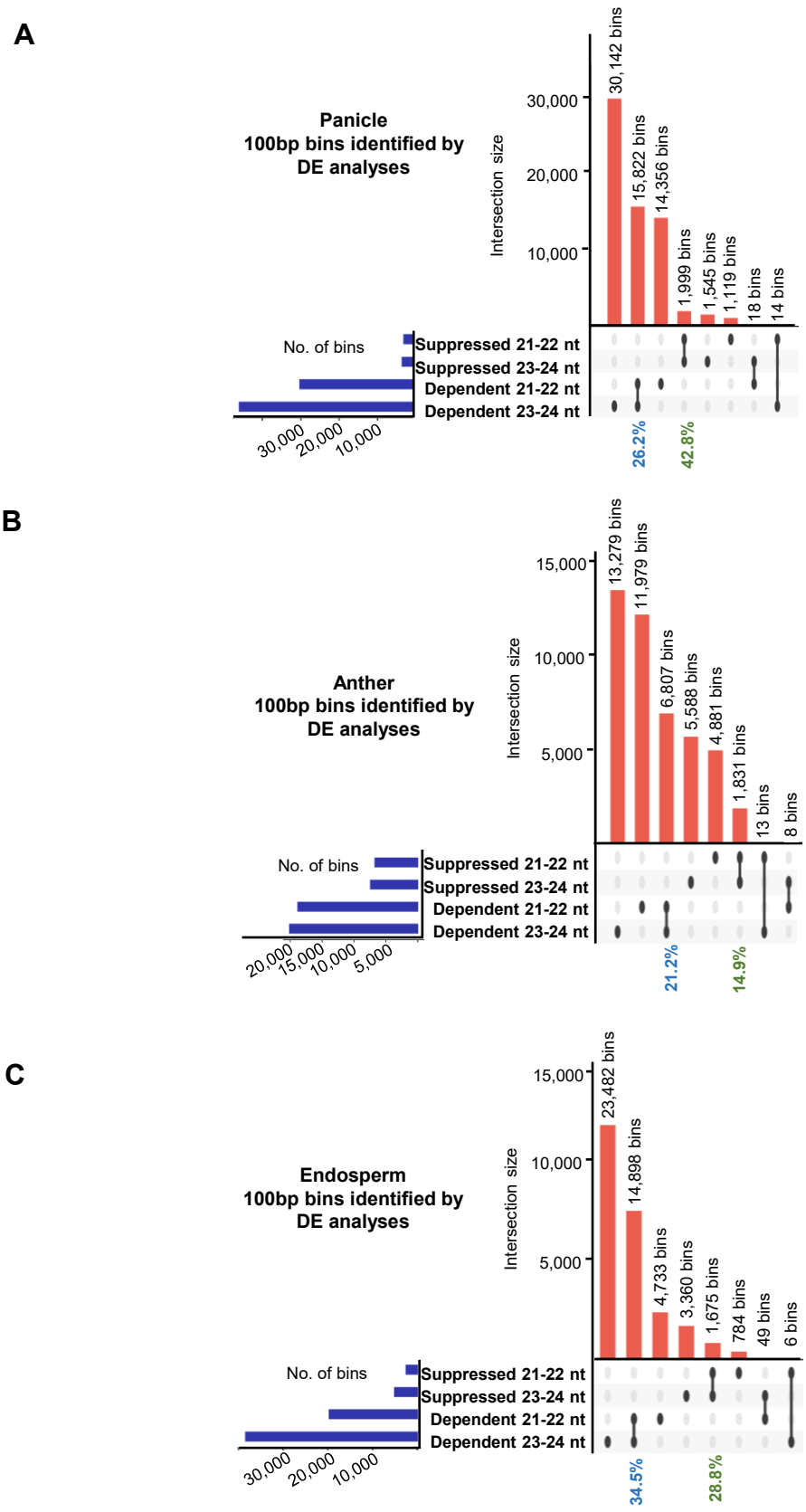
Supplemental Figure S11. Loss of Pol IV impacts occupancy of Pol II and MNase accessibility. *(A)* DNA fragmentation profiles after treatment with denoted amounts (Kunitz units) of micrococcal nuclease in WT and *kd* derived chromatin samples. Sizes of the major population of fragments denoted over the peaks. Red stars indicate the location of the peaks on the bioanalyser profiles (Refer Supplemental methods). *(B)* Metaplot depicting Pol II occupancy over rice protein coding genes normalised to the H3 ChIP. Numbers in grey describe the number of loci taken for analyses. Box-violin plots shows the difference in enrichment in *kd* compared to WT over the described sets of loci. The Y-axis is scaled to inverse sine hyperbolic function of enrichment values. *(C)* Metaplot depicting Pol II occupancy over genes having at least 10% of their length with annotated repeats normalised to the H3 ChIP. Numbers in grey describe the number of loci taken for analyses. Box-violin plots shows the difference in enrichment in *kd* compared to WT over the described sets of loci. The Y-axis is scaled to inverse sine hyperbolic function of enrichment values. Mann-Whitney U test, * p-value < 0.0001.



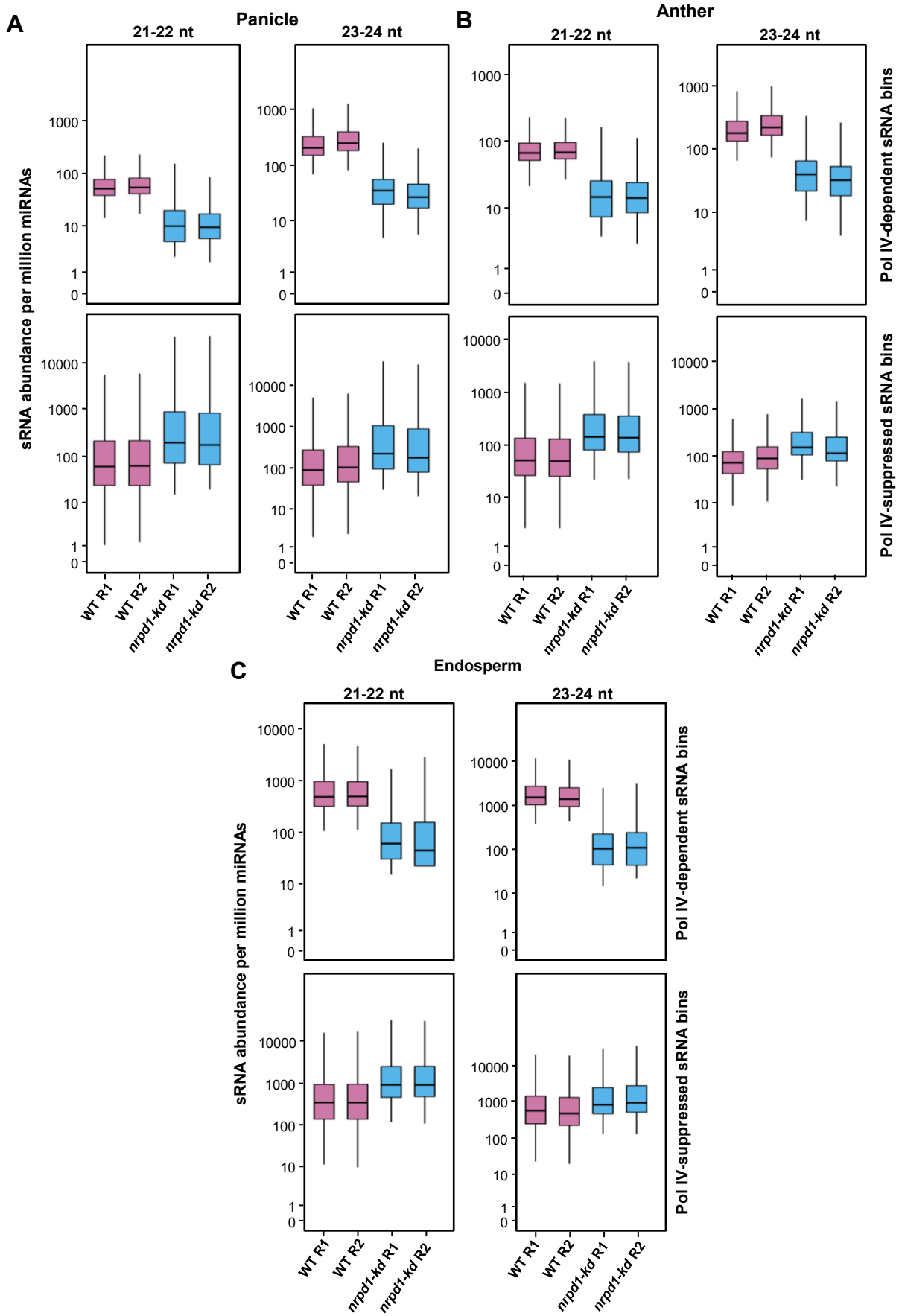
Supplemental Figure S12. Pol IV complex suppresses peculiar sRNA production from several loci. *(A)* Volcano plots showing levels of deregulation of sRNAs over clusters identified by shortstack from panicle, anther and endosperm sRNA datasets. The clusters were identified after subcategorizing into 21-22 nt and 23-24 nt size classes. Clusters showing statistically significant difference of 4-folds relative to WT are highlighted. The number of up- and downregulated clusters are mentioned with arrows. *(B)* Point density plots of the volcano plots depicted in *(A)* showing the clustering of DE sRNA clusters. *(C)* MA plots of sRNA clusters showing significant (red, P-value < 0.05) differentially expressed clusters. *(D)* Stacked bar plots showing normalised abundance of rice sRNAs of different sizes from Pol IV-suppressed and dependent bins displaying the 5' nucleotide bias. *(E)* IGV screenshots of sRNA-seq tracks of Pol IV-suppressed loci 1 and 2 with expression (RPM) range in square brackets. These probes against represented sRNAs are used in Fig. 3B and 3C.



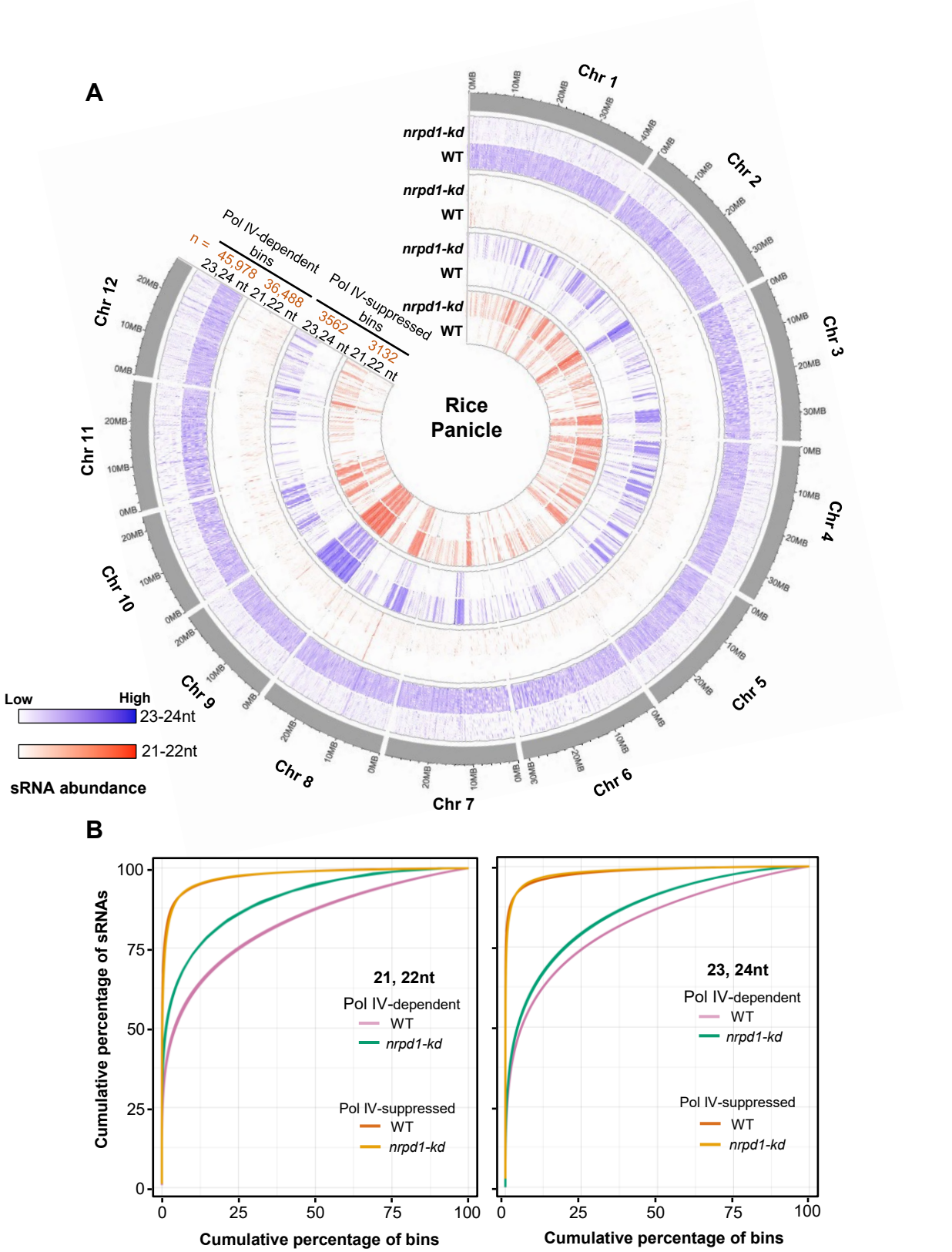
Supplemental Figure S13. Pol IV-dependent and Pol IV-suppressed sRNA clusters are of similar length. *(A)* Venn plots showing sRNA size enrichment of Pol IV-dependent and suppressed clusters. Clusters are classified as 21-22nt (clusters with more than 65% sRNAs in 21-22nt size class), 23-24nt (clusters with more than 65% sRNAs in 23-24nt size class) and mixed clusters (all the other clusters with mixed size profile of sRNAs). *(B)* Density distribution plots of the sizes of shortstack identified sRNA clusters sub-categorized into Pol IV-suppressed and dependent clusters. Central line within the peak denotes the median length of the clusters and other two lines represent the interquartile range. Median length is denoted in the plots as M.



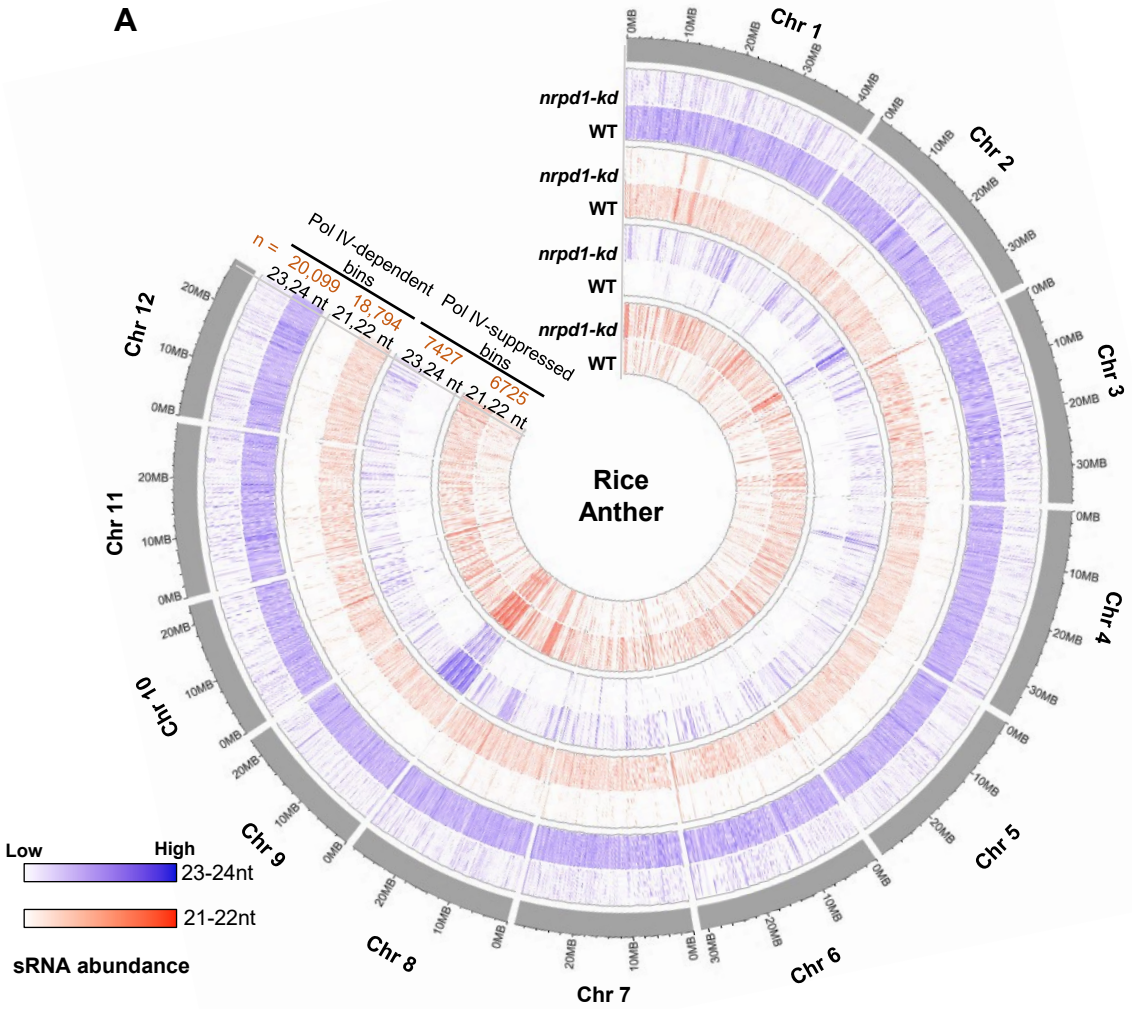
Supplemental Figure S14. Pol IV-dependent and Pol IV-suppressed sRNA bins are distinct in producing sRNAs with specific sizes. *(A-C)* Upset plots describing the numbers and overlap of differentially expressed sRNA bins categorized as Pol IV-suppressed and dependent bins in panicle *(A)*, anther *(B)* and endosperm *(C)*.

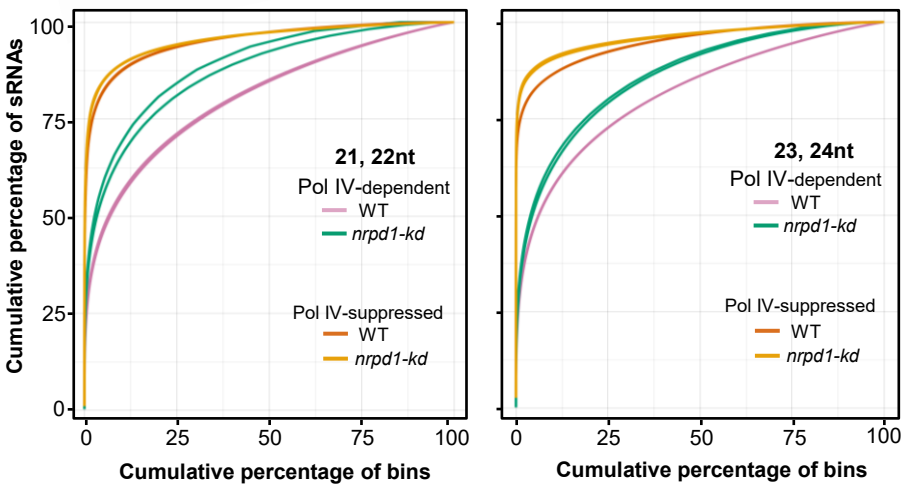


Supplemental Figure S15. Pol IV-dependent and Pol IV-suppressed sRNA bins are not due to oversampling and library normalisation. *(A-C)* Boxplots describing the relative abundance of sRNAs normalised to net raw abundance of miRProf identified miRNAs in Pol IV-suppressed and dependent bins from panicle *(A)*, anther *(B)* and endosperm *(C)*. The Y-axis is scaled to inverse sine hyperbolic function of values.

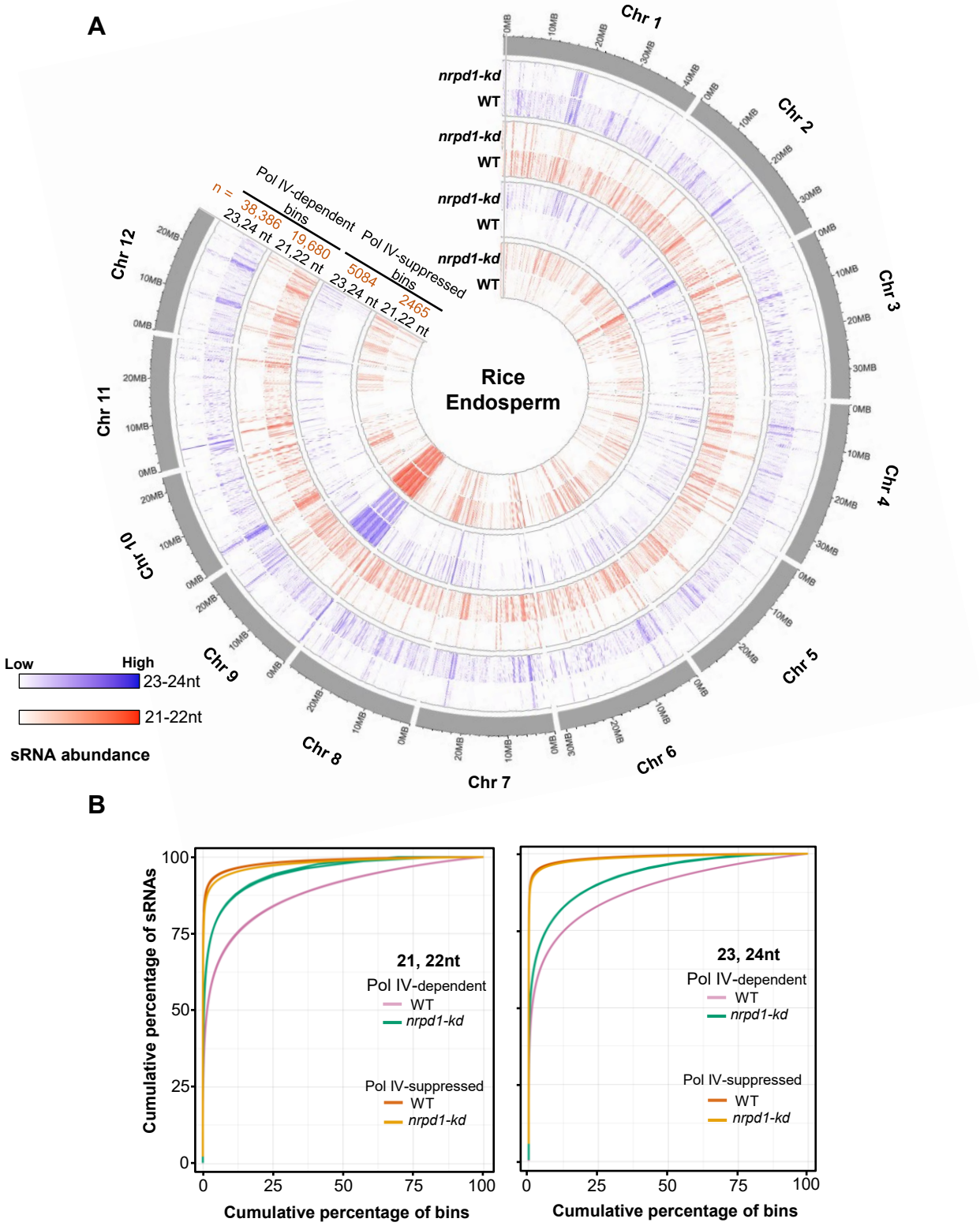


Supplemental Figure S16. Pol IV-suppressed sRNAs from panicle are non-uniformly distributed across the genome. *(A)* Circos plot showing the normalised abundance and distribution of sRNAs in 100 bp windows categorised as Pol IV-dependent and Pol IV-suppressed bins across 12 chromosomes. The abundance values of replicates are merged and displayed as 21,22nt and 23,24nt tracks. The number of bins identified for each category is labelled as n. *(B)* Percentage cumulative sum plots for 21,22nt and 23,24nt size classes of sRNAs from panicle subclassified into Pol IV-suppressed and dependent sRNA bins. The replicate lines are identically coloured.

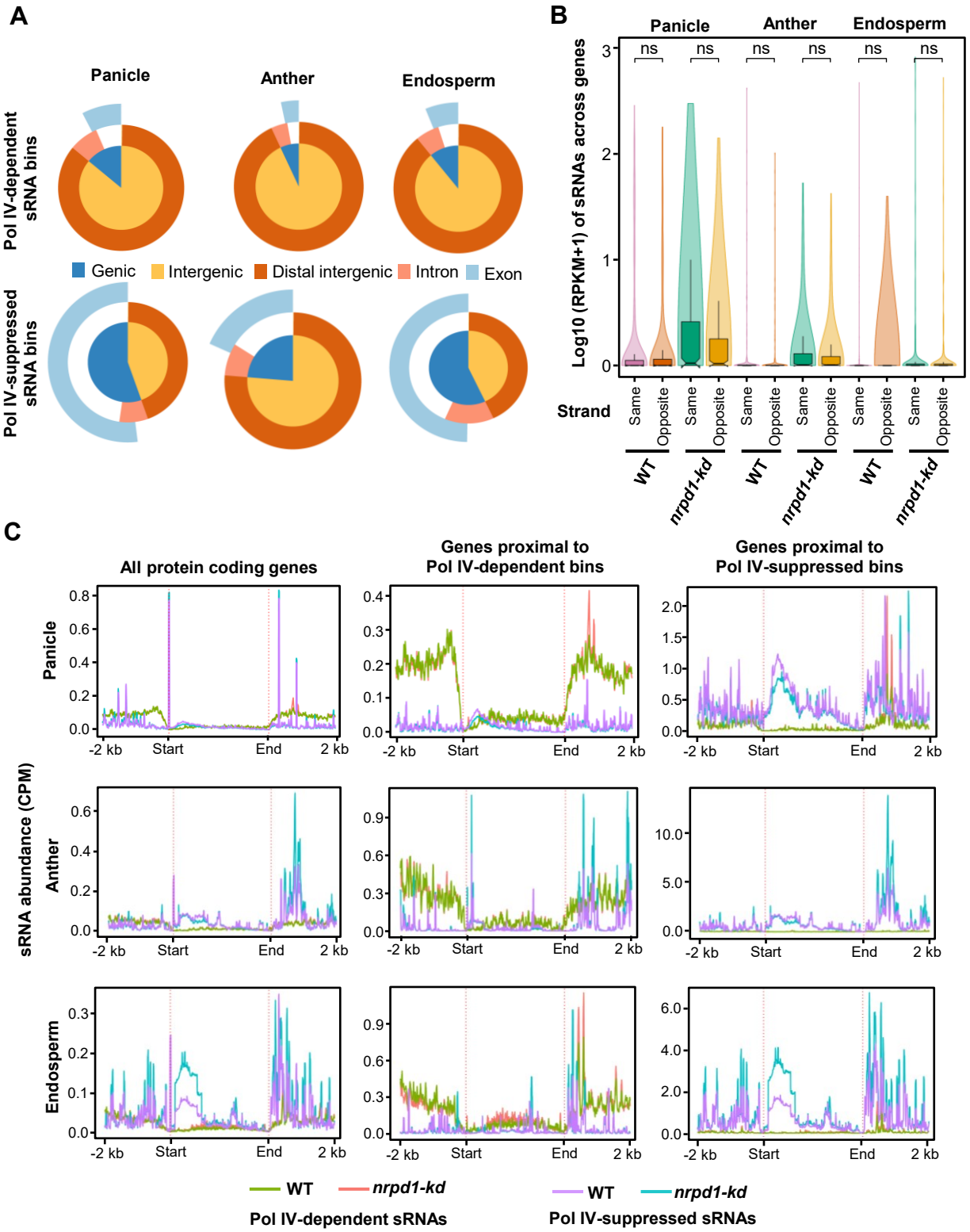




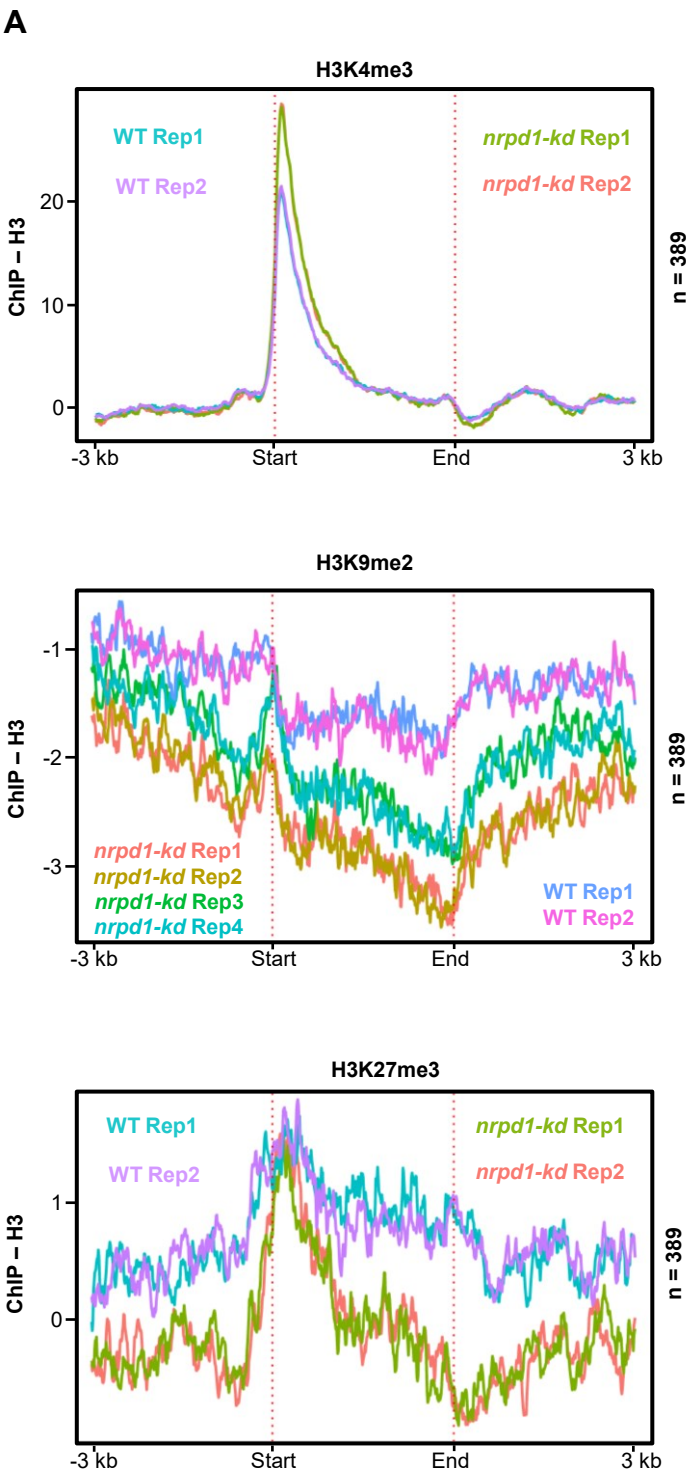
Supplemental Figure S17. Pol IV-suppressed sRNAs from anther are non-uniformly distributed across the genome. *(A)* Circos plot showing the normalised abundance and distribution of sRNAs in 100 bp windows categorised as Pol IV-dependent and Pol IV-suppressed bins across 12 chromosomes. The abundance values of replicates are merged and displayed as 21,22nt and 23,24nt tracks. The number of bins identified for each category is labelled as n. *(B)* Percentage cumulative sum plots for 21,22nt and 23,24nt size classes of sRNAs from anther subclassified into Pol IV-suppressed and dependent sRNA bins. The replicate lines are identically coloured.



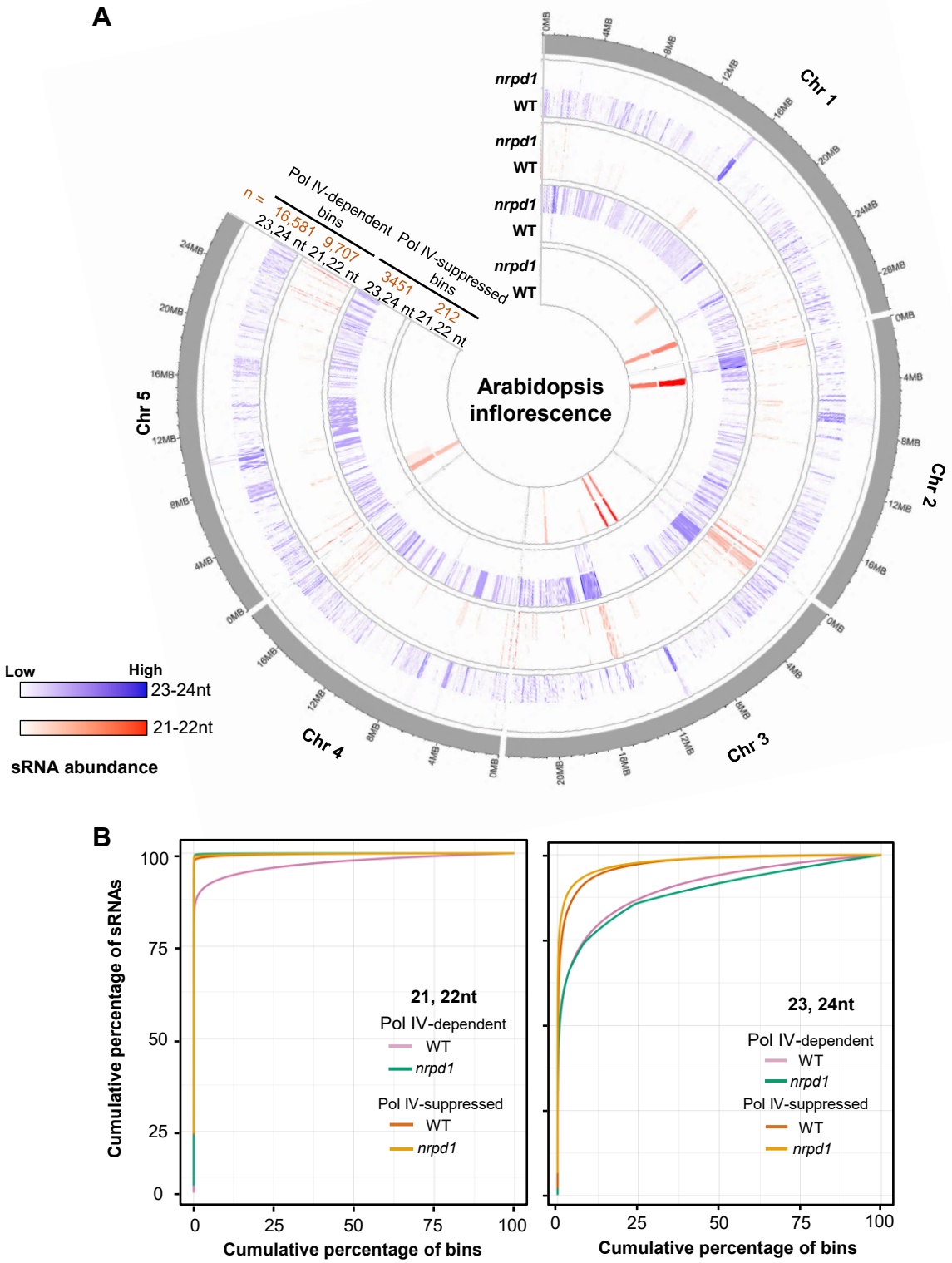
Supplemental Figure S18. Pol IV-suppressed sRNAs from endosperm are non-uniformly distributed across the genome. *(A)* Circos plot showing the normalised abundance and distribution of sRNAs in 100 bp windows categorised as Pol IV-dependent and suppressed bins across 12 chromosomes. The abundance values of replicates are merged and displayed as 21,22nt and 23,24nt tracks. The number of bins identified for each category is labelled as n. *(B)* Percentage cumulative sum plots for 21,22nt and 23,24nt size classes of sRNAs from endosperm subclassified into Pol IV-suppressed and dependent sRNA bins. The replicate lines are identically coloured.



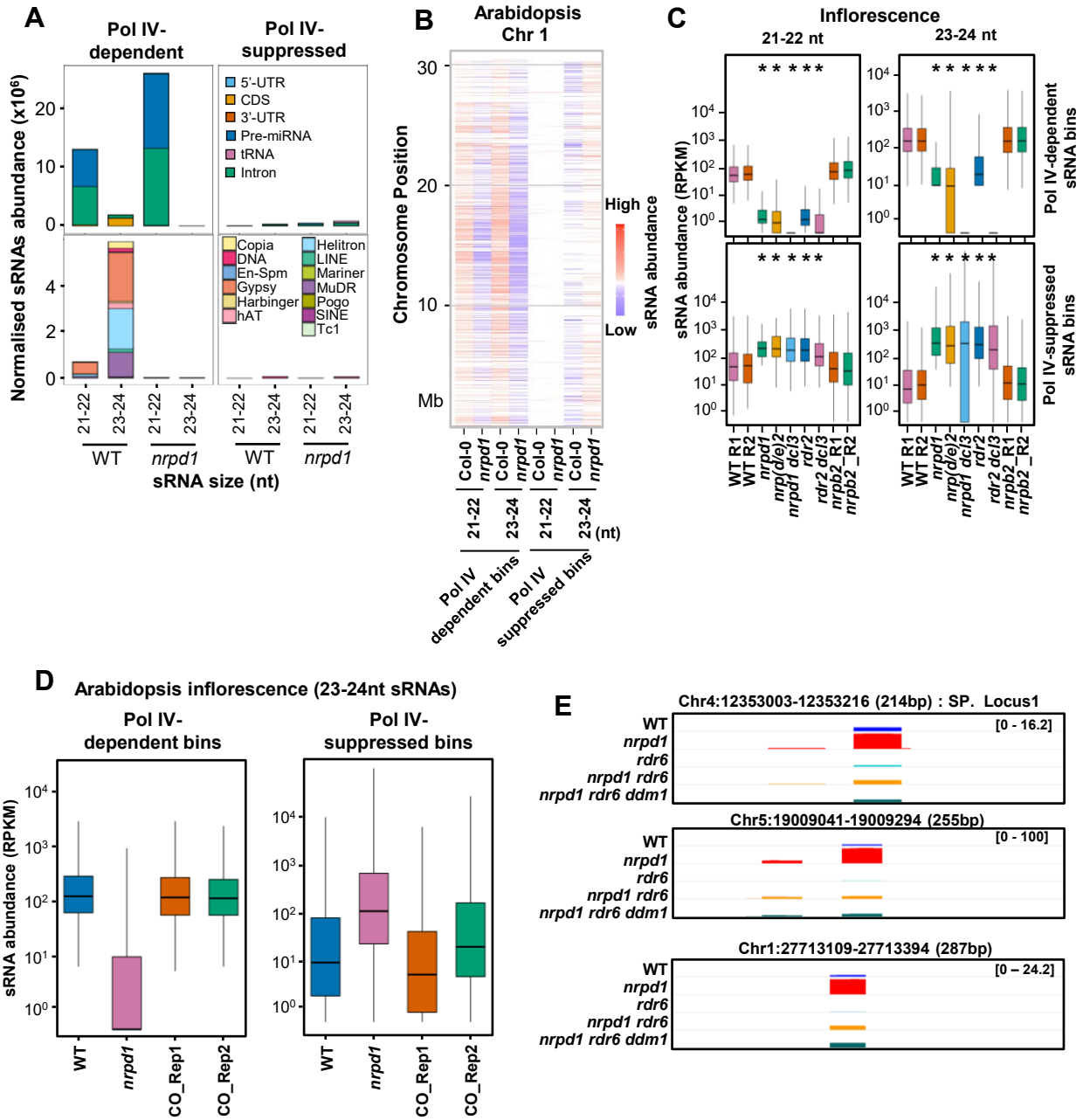
Supplemental Figure S19. Pol IV-suppressed sRNAs have significant overlap with coding regions. *(A)* Vennpie plots showing the overlap of the Pol IV-suppressed and dependent sRNA bins in major genomic features as identified by ChIPseekeR annotation. *(B)* Boxplots showing the distribution of Pol IV-suppressed sRNAs across genes overlapping with the Pol IV-suppressed bins. sRNAs are counted with respect to gene coding strand. Mann-Whitney U test was used for statistical testing. ns - not-significant (p-value > 0.05). The box shows the isigned rank nterquartile range and violin represent the distribution of datapoints. *(C)* Metaplots representing the coverage of sRNAs (Pol IV-dependent and suppressed sRNAs) across all the protein coding genes, genes proximal to dependent bins and genes proximal to suppressed bins identified in panicle, anther and endosperm.



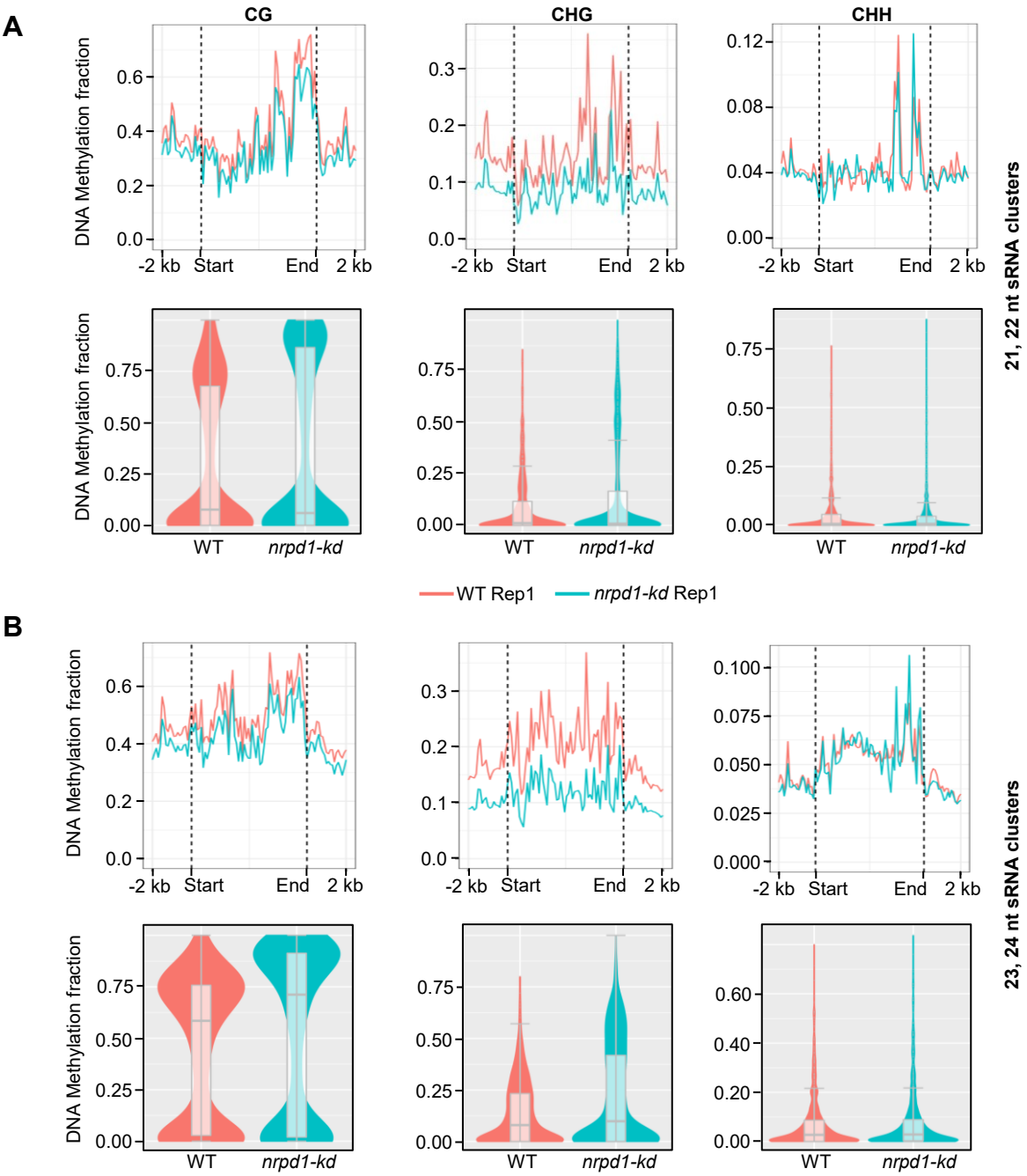
Supplemental Figure S20. Genes overlapping with suppressed sRNAs exhibit reduced silencing compensation by H3K27me3. *(A)* Metaplots describing the occupancy of histone H3 modifications H3K4me3, H3K9me2 and H3K27me3 normalised to total histone H3 occupancy plotted over the proteins overlapping with the suppressed sRNA bins (n = 389) from panicle.



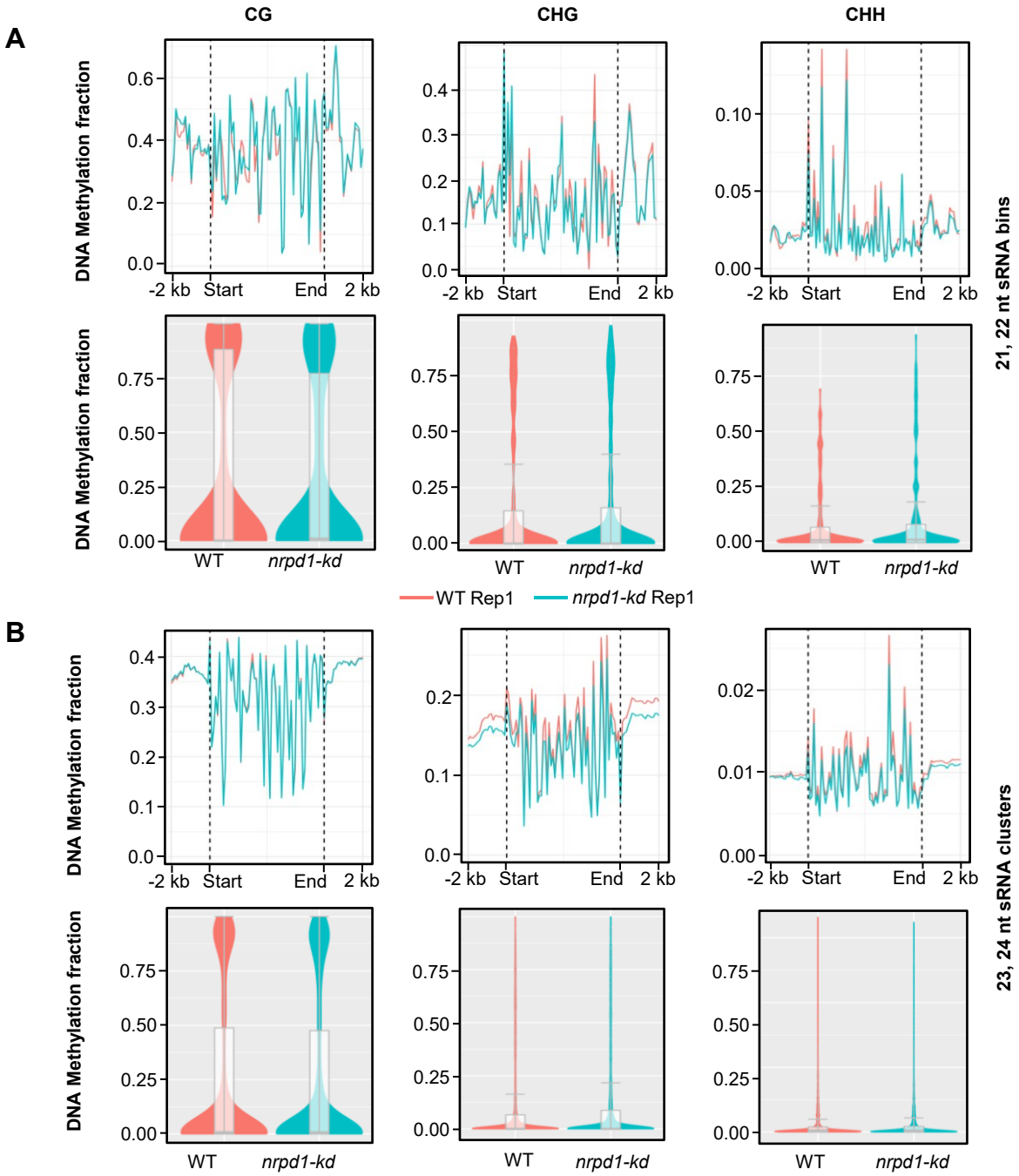
Supplemental Figure S21. Pol IV-suppressed sRNAs from *Arabidopsis* inflorescence are non-uniformly distributed across the genome similar to rice tissues. *(A)* Circos plot showing the normalised abundance and distribution of sRNAs in 100 bp windows categorised as Pol IV-dependent and suppressed bins across 5 chromosomes. The abundance values are displayed as 21,22nt and 23,24nt tracks. The number of bins identified for each category is labelled as n. The datasets are taken from from GSE61439. *(B)* Percentage cumulative sum plots for 21,22nt and 23, 24nt size classes of sRNAs from *Arabidopsis* inflorescence subclassified into Pol IV-suppressed and dependent sRNA bins.



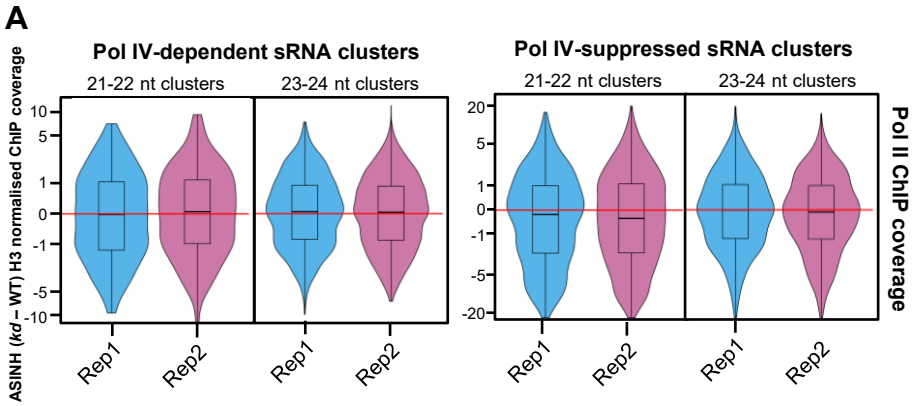
Supplemental Figure S22. Pol IV-suppressed sRNAs from Arabidopsis are characteristically similar to that of rice. *(A)* Stacked bar plots showing abundance of sRNAs from different features (coding features in the top and transposon categories in the bottom). *(B)* Heatmap showing the normalised abundance of sRNAs from Pol IV-dependent and suppressed bins identified in the same way as in rice over the *Arabidopsis* chromosome 1. The datasets from GSE61439 (inflorescence tissue) was analysed and sRNAs from differentially expressed bins were represented. *(C)* Boxplots showing the normalised sRNA abundance of Pol IV-suppressed and dependent bins from *Arabidopsis* seedlings tissue taken from GSE72993 and GSE98285. The Y-axis is scaled to inverse sine hyperbolic function of RPKM values. Mann Whitney U test. * p-value < 0.0001. *(D)* Boxplots showing the abundance of 23-24nt sRNAs in WT, *nrpd1* and NRPD1 complemented lines (CO) in the Pol IV-dependent and suppressed bins identified. Datasets taken from GSE140566. The Y-axis is scaled to inverse sine hyperbolic function of RPKM values. *(E)* IGV screenshots of sRNA tracks (16-50nt) from different genotypes represented in Fig. 5D.



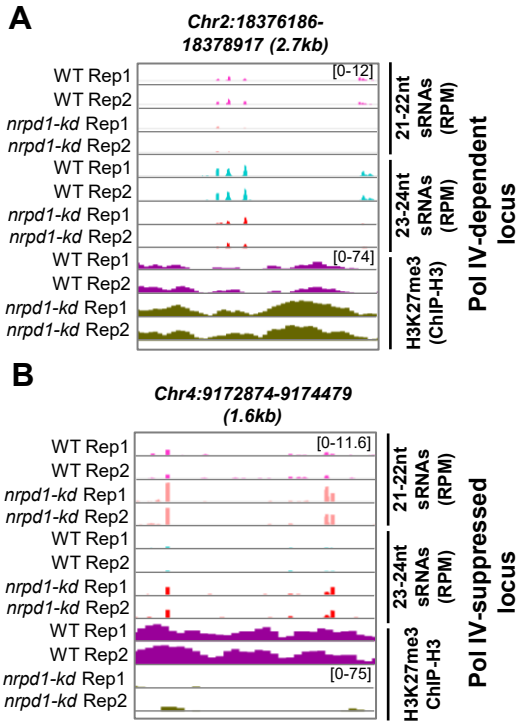
Supplemental Figure S23. Canonical RdDM does not potentiate Pol IV-suppressed sRNAs in rice. *(A-B)* Metaplots and box-violin plots depicting the level of DNA methylation in the panicle tissue over the 21,22nt *(A)* and 23,24nt *(B)* Pol IV-suppressed panicle sRNA clusters in CG, CHG and CHH contexts.



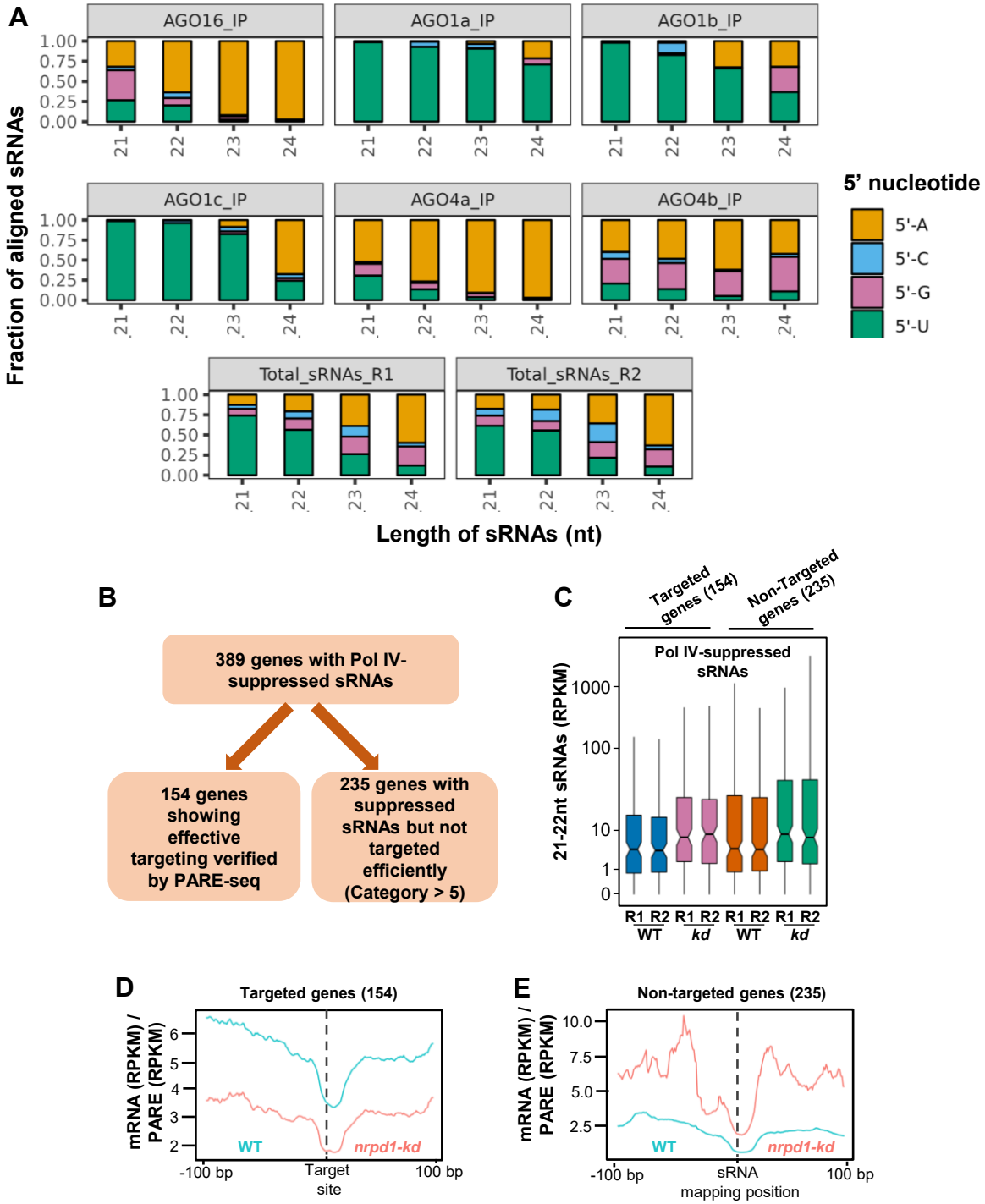
Supplemental Figure S24. Canonical RdDM does not potentiate Pol IV-suppressed sRNAs in *Arabidopsis*. *(A-B)* Metaplots and box-violin plots depicting the level of DNA methylation over the 21,22nt *(A)* and 23,24nt *(B)* Pol IV-suppressed panicle sRNA bins in CG, CHG and CHH contexts. The datasets were taken from GSE99689.



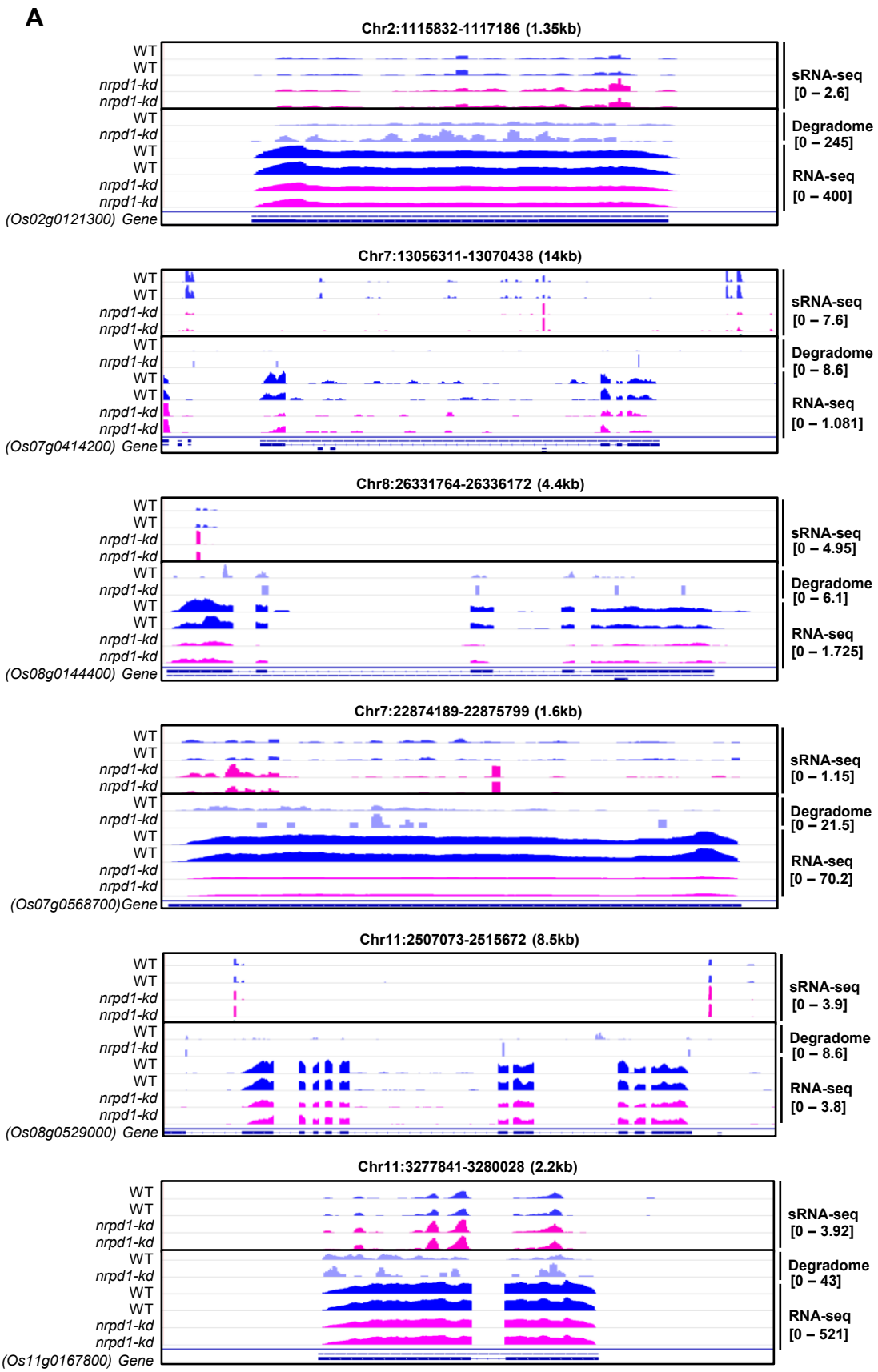
Supplemental Figure S25. Net Pol II occupancy over Pol IV-suppressed sRNA clusters does not change. *(A)* Box-violin plots showing the difference in Pol II coverage over the Pol IV-dependent and Pol IV-suppressed sRNA clusters size classified into 21-22 nt and 23-24 nt sRNAs. The Y-axis is scaled to inverse sine hyperbolic function of difference values normalised to H3.



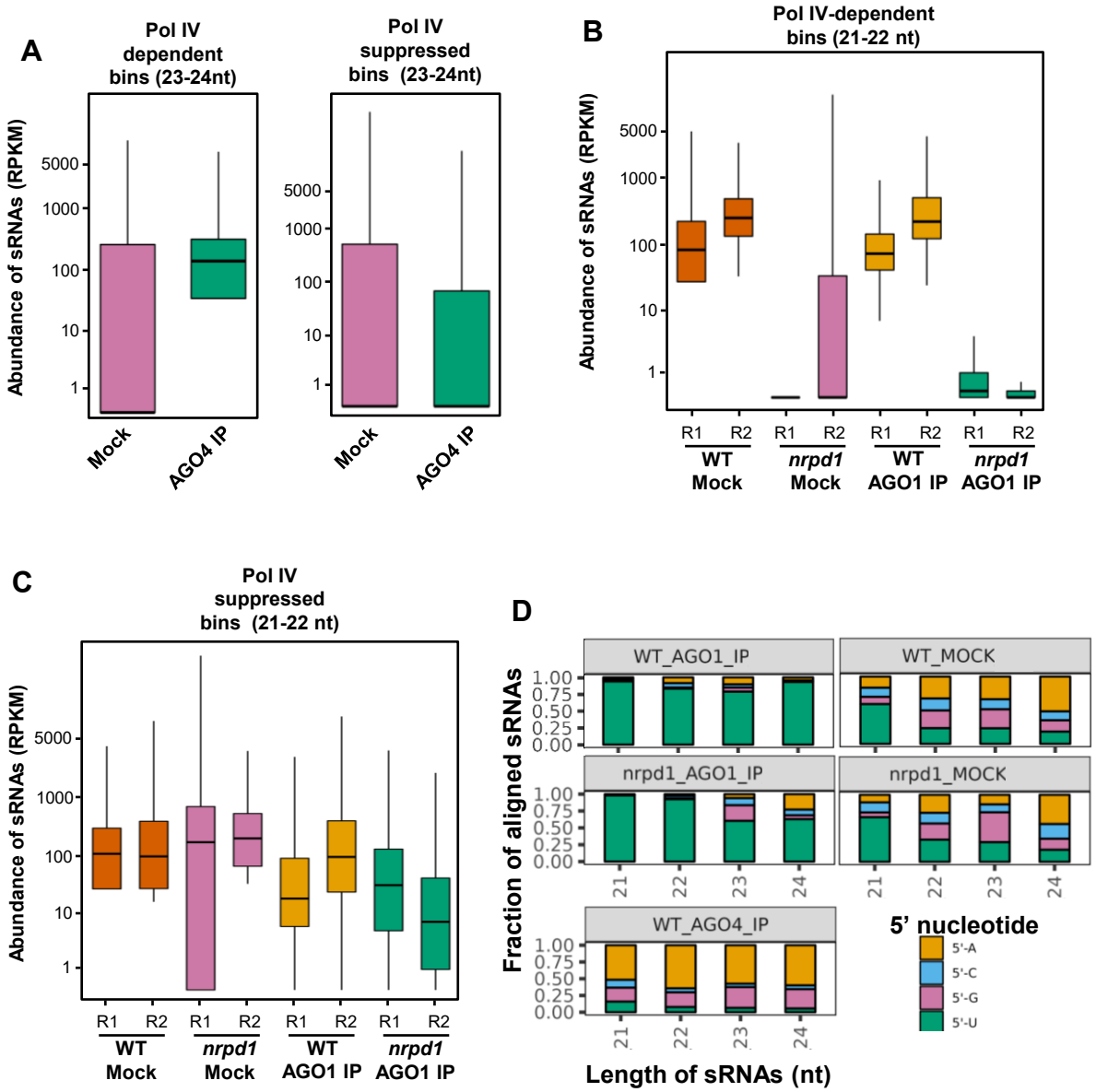
Supplemental Figure S26. Distinct H3K27me3 modifications at Pol IV-dependent and suppressed sRNA clusters. *(A-B)* Genome screenshots of representative Pol IV-dependent sRNA locus *(A)* and suppressed sRNA locus *(B)*.



Supplemental Figure S27. Analyses of rice AGO-IP and PARE datasets exhibits expected profiles. *(A)* Stacked bar-plots showing relative fraction of AGO1-IP sRNAs of different sizes (21-24nt) that aligned to the genome displaying their 5' nucleotide bias (Datasets: GSE18250). (*B)* Schematic showing number of genes with Pol IV-suppressed sRNAs and their targeting status. (*C)* Box-plots showing comparable abundance of Pol IV-suppressed sRNAs from targeted and non-targeted genes. *(D-E)* Metaplots showing the ratio of mRNA and PARE reads over the targeted genes *(D)* and non-targeted genes *(E)*. For the targeted genes, target site was centred and for non-targeted genes sRNA mapping positions were centred. Coverage over 100bp region on either side of centred location was plotted.



Supplemental Figure S28. Genes targeted by Pol IV-suppressed sRNAs show reduction in expression in *kd* lines. *(A)* IGV screenshots of Pol IV-suppressed sRNA containing genes with the degradome tag density and RNA-seq coverage. The targeted genes are denoted in the bottom track. The track data ranges are mentioned in square brackets.



Supplemental Figure S29. *Arabidopsis* Pol IV-suppressed sRNAs are not effectively loaded into AGO1 and unlikely to target genes. *(A)* Stacked bar plots showing normalised abundance of small RNAs of different sizes from Pol IV-suppressed and dependent bins displaying the 5' nucleotide abundance in both the categories. *(B)* Box plots showing the accumulation of 21-22nt sRNAs categorized into Pol IV-dependent and suppressed sRNA bins from WT inflorescence in the AGO1 immuno-precipitate fraction. Datasets were taken from GSE61439. The Y-axis is scaled to inverse sine hyperbolic function of RPKM values. *(C)* Box plots showing the accumulation of 21-22nt sRNAs categorized into Pol IV-dependent and suppressed sRNAs from WT and *nrpd1* inflorescence in the AGO1 immuno-precipitate fraction. Datasets were taken from GSE133618. The Y-axis is scaled to inverse sine hyperbolic function of RPKM values. *(D)* Stacked bar-plots showing relative fraction of sRNAs of different sizes (21-24nt) that aligned to the genome displaying their 5' nucleotide bias.