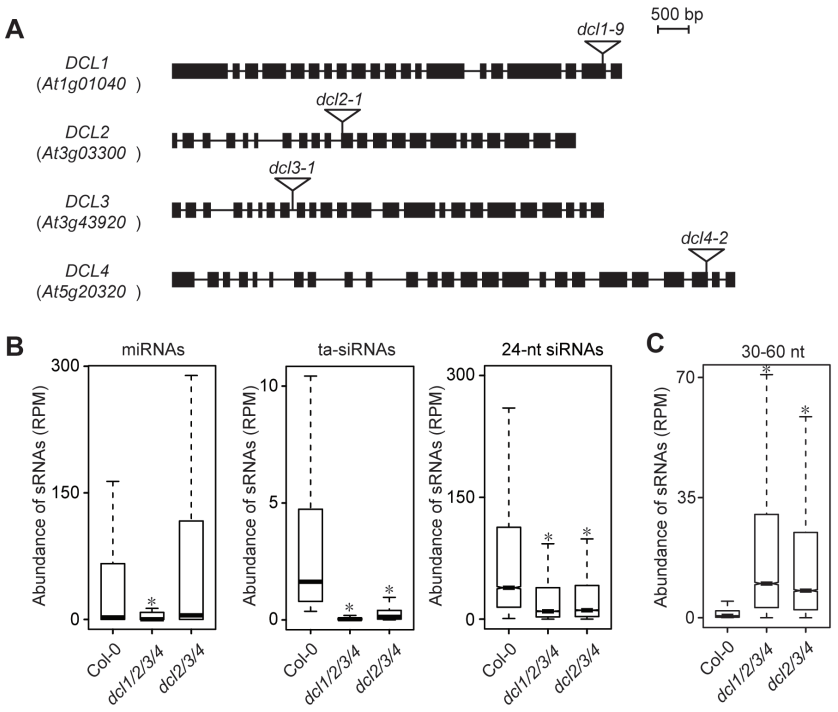

Figure S1, Identification of sidRNAs, Related to Figure 1.

(A) Diagrams showing the gene structures of *DCLs*. The positions of T-DNA insertions are indicated.

(B) Box plots of the levels of miRNAs, ta-siRNAs and P4siRNAs in the indicated plants.

(C) Box plots of the levels of 30-60-nt sRNAs generated from sidRNA loci in the indicated plants. Asterisks indicate a significant difference between Col-0 and the mutants ($P < 10^{-15}$, Mann–Whitney U test).

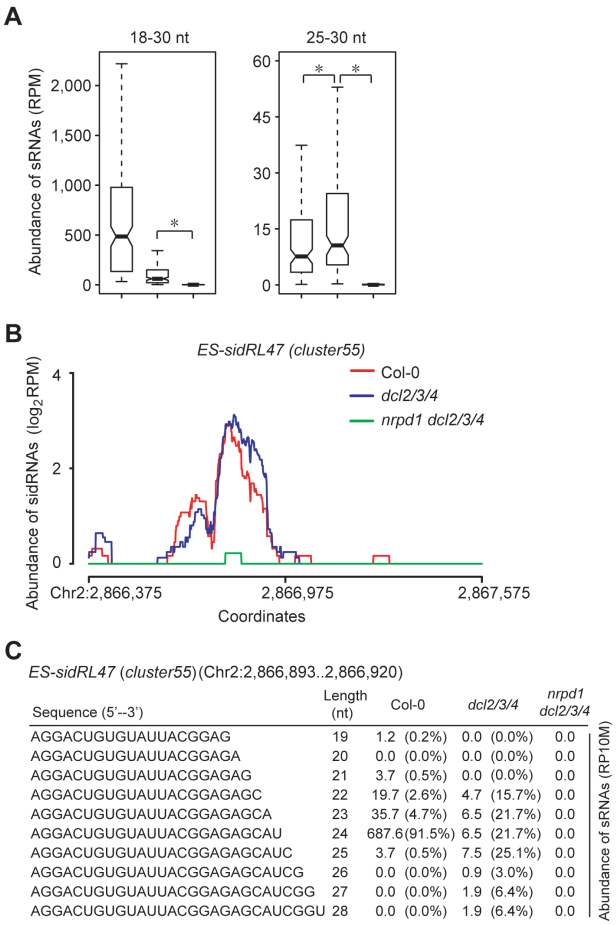

Figure S2, Identification of Endosperm-Specific sidRNAs, Related to Figure 1.

(A) Box plots of levels of sRNAs of the indicated sizes produced from endosperm-specific sidRNA loci in Col-0, *dcl2*/*3*/*4* and *dcl1*/*2*/*3*/*4*. Asterisks indicate a significant difference between Col-0 and the mutants ($P < 10^{-15}$, Mann–Whitney *U* test).

(B) Density of sidRNAs mapped to the *ES-sidRL47 (cluster55)* locus in the indicated plants.

(C) A representative group of laddered sRNAs produced from *ES-sidRL47 (cluster55)* locus. Numbers represent the abundance (RPM) of sRNAs with different lengths ranging from 19 to 28 nt. Values in the parentheses indicate the percentages of sRNAs with different lengths.

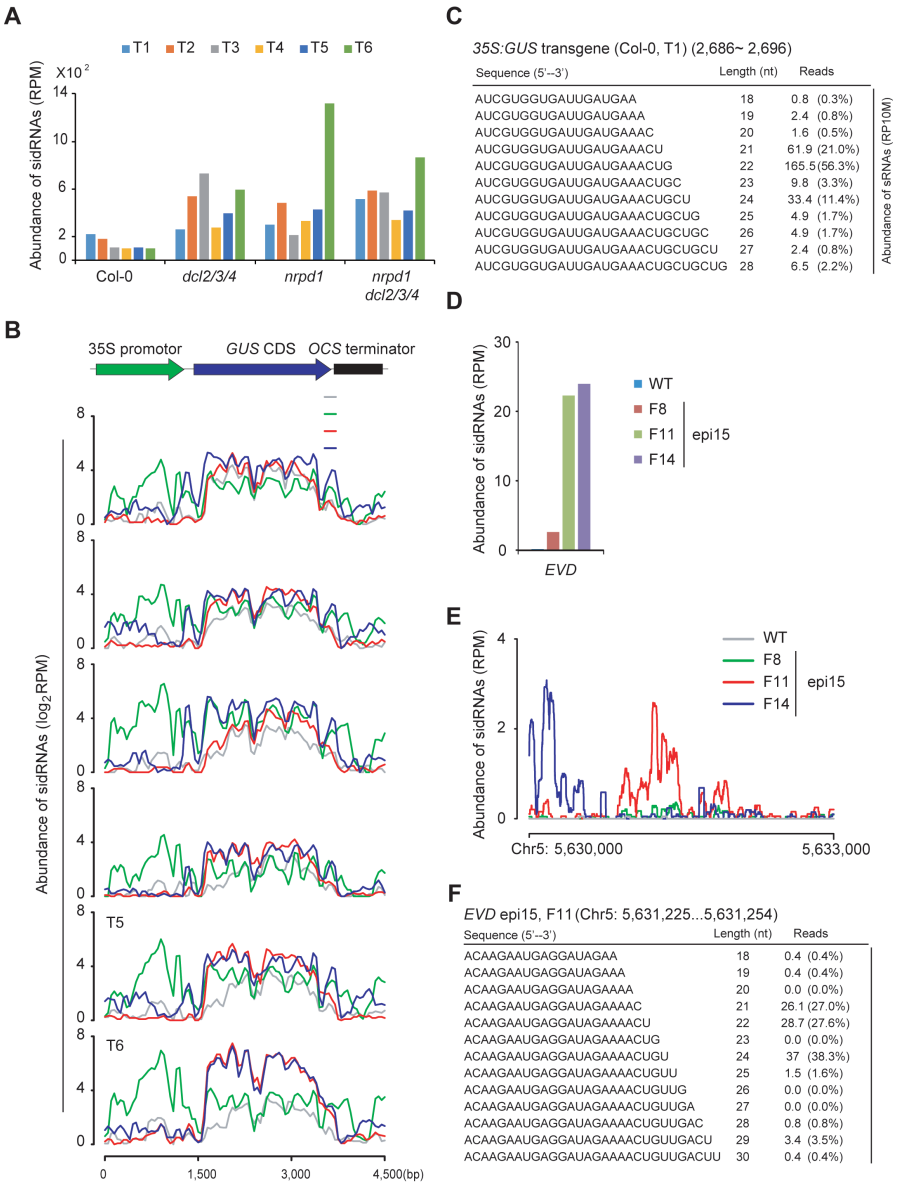

Figure S3, Detection of sidRNAs from Transgenic Loci and Transposons Undergoing *de novo* Silencing, Related to Figure 1.

(A, B) Detection of sidRNAs derived from transgenic loci in T1-T6 generations of the indicated genotypes by deep sequencing analysis.

(C) A representative group of laddered sRNAs produced from transgenic loci in T1 generation in Col-0 background. Numbers represent the abundance (RPM) of sRNAs with different lengths ranging from 18 to 28 nt. Values in the parentheses indicate the percentages of sRNAs with different lengths.

(D, E) Detection of sidRNAs mapped to *EVD* locus in *epi15* F8, F11, F14, and wild-type (WT) plants by deep sequencing analysis.

(F) A representative group of laddered sRNAs produced from *EVD* locus in *epi15* F11 plants. Numbers represent the abundance (RPM) of sRNAs with different lengths ranging from 18 to 30 nt. Values in the parentheses indicate the percentages of sRNAs with different lengths.

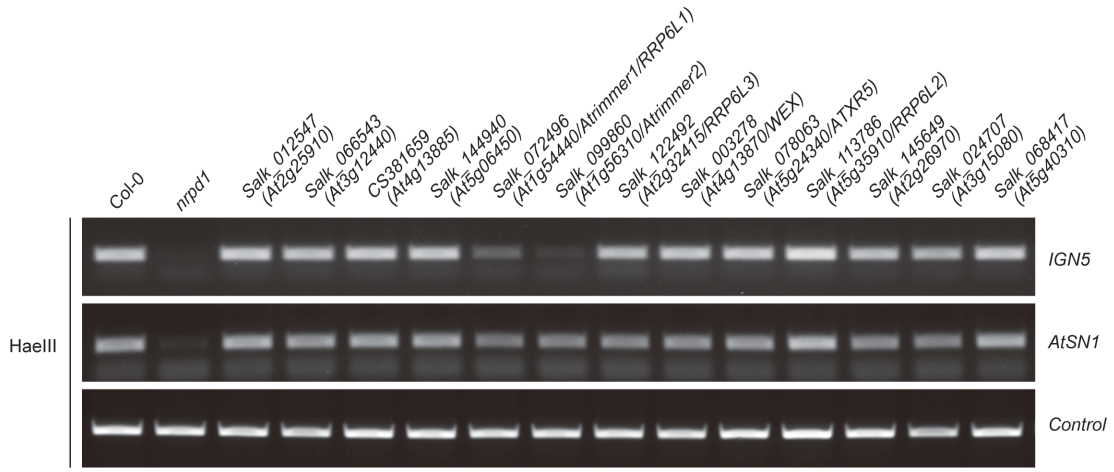

Figure S4, Identification of Distributive 3'-5' Exonucleases Required for DNA Methylation, Related to Figure 6.

DNA methylation levels at the *AtSN1* and *IGN5* loci in the indicated plants were measured by Chop-PCR. The methylation-sensitive restriction enzyme HaeIII was used to specifically cleave unmethylated DNA in CHH context. A fragment of *Actin* that lacks HaeIII sites was amplified as a control

Extended Experimental Procedure

Plant Materials

The *dcl2-1 dcl3-1 dcl4-2* (*dcl2*/*3*/*4*), *nrpd1-3* (*nrpd1*), *rdr2-1* (*rdr2*) and *ago4-1* (*ago4*) mutants in the Col-0 ecotype were described previously (Henderson et al., 2006; Onodera et al., 2005; Wierzbicki et al., 2009; Xie et al., 2004). The quadruple mutant *dcl1-9 dcl2-1 dcl3-1 dcl4-2* (*dcl1*/*2*/*3*/*4*) was obtained by crossing *dcl2-1 dcl3-1 dcl4-2* with *dcl1-9 dcl2-5 dcl3-1 dcl4-2* (Blevins et al., 2011). The quadruple mutants *nrpd1 dcl2*/*3*/*4*, *rdr2 dcl2*/*3*/*4*, and *ago4 dcl2*/*3*/*4* were obtained by crossing *dcl2*/*3*/*4* with *nrpd1*, *rdr2,* and *ago4*, respectively. T-DNA insertion lines for genes encoding 3'-5' exonucleases in the Col-0 ecotype were obtained from the Arabidopsis Biological Resource Center (ABRC). All mutants were genotyped by PCR. The transgenic lines *NRPD1-FLAG* and *NRPD1(ASM)-FLAG* were described (Berger et al., 2009) and crossed into *nrpd1 dcl2*/*3*/*4*. The transgenic lines *GFP-AGO4* in Col-0 or *dcl2*/*3*/*4* background were described (Ye et al., 2012).

Constructs *pAGO4:GFP-AGO4 and pAGO4:GFP-AGO4*$^{YF/AA}$ were described previously (Ye et al., 2012) and transformed into the *ago4 dcl2*/*3*/*4* mutant using the floral-dip method (Clough and Bent, 1998). Positive transformants were identified through selection for antibiotic resistance and confirmed by Western blot. The *GUS* gene was PCR-amplified, cloned into pENTR/D-TOPO (Invitrogen) and introduced into pEarley-Gate202 (Earley et al., 2006) through LR recombination reaction. The resulting *p35S:GUS* construct was transformed into Col-0, *nrpd1*, *dcl2*/*3*/*4* and *nrpd1 dcl2*/*3*/*4*. At least 50 lines were obtained for each genotype. Propagation of the transgenic lines was performed by single-seed descent procedures. In every generation, ten flowers from each plant of the same genotype were harvested and pooled for further analysis.

sRNA Sequencing and Analysis

sRNAs of 18-30 nt or 30-60 nt were gel-purified on a 15% denaturing PAGE gel and subjected to library construction as described (Mi et al., 2008). A detailed protocol is available upon request. The libraries were single-end sequenced on an Illumina HiSeq2000 platform.

After removing adapters and low-quality reads, sRNA-seq reads were mapped to the *Arabidopsis* genome (TAIR10 version) with Bowtie (Langmead et al., 2009) allowing no mismatches, and the mapped reads were retained for further analyses.

sRNA reads from *dcl1/2/3/4* were used for identification of sidRNA loci. A sidRNA loci were annotated using the following criteria: 1) length ≥ 100 nt; 2) number of unique sRNAs ≥ 5, each separated from nearest neighbors by a maximum of 50 nt; 3) expression level ≥ 15 RPKM (reads per kilobase per million).

sRNA abundance was calculated as reads per million (RPM) or reads per 10 million (RP10M). To identify sidRNA loci differentially expressed in different tissues or genotypes, edgeR was applied to calculate the fold change and p-value. The resulting p-values were adjusted by Benjamini-Hochberg's approach to control false discovery rate (FDR). Those showing ≥ 3 fold change at a FDR < 0.01 were considered as differentially expressed loci.

Whole-genome Bisulfite Sequencing and Analysis

Paired-end bisulfite sequencing libraries were constructed using the NEXTflex Bisulfite-Seq kit (BIOO SCIENTIFIC) according to the manufacturer's instructions. About 2 μg of genomic DNA was sonicated to ~300 bp, end repaired and ligated to methylated adapters. Ligated products were subjected to bisulfite treatment using the EZ DNA Methylation-Gold kit (ZYMO Research) according to the manufacturer's instructions, and then PCR amplified for 15 cycles using EX Taq DNA polymerase (Takara). The amplified libraries were purified twice using Ampure XP purification beads (Beckman) prior to quantification with the Agilent 2100 Bioanalyzer. The libraries were paired-end sequenced on the Illumina HiSeq 2000 generating 2x100-mer reads.

Bisulfite sequencing reads were mapped to *Arabidopsis* genome (TAIR10) by BRAT-BW (Harris et al., 2012), allowing up to three mismatches. Only uniquely mapped non-redundant reads were retained for further methylation analysis. Chloroplast bulk fractional methylation was used to monitor bisulfite conversion efficiency. Fractional DNA methylation levels were calculated by #C/(#C+#T). CHH DMRs (differentially methylated regions) were defined as previously described (Stroud et al., 2013; Zhang et al., 2013). DNA methylation levels in every 200 bp sliding window with 50 bp step size were compared in wild-type and mutants. Only windows with more than four informative cytosine (i.e. coverage ≥ 4) were retained. The windows with absolute methylation difference > 0.1 and Benjamini-Hochberg-adjusted FDR < 0.001 (Fisher's exact test) were considered as differentially methylated bins. Then, those differentially methylated bins separated by gaps within 100 bp were merged together as differentially methylated regions (DMRs). The final coordinates were further adjusted from the first methylated cytosine to the last methylated cytosine.

Analyses of Features of sidRNAs

To analyze whether sidRNAs have 5' monophosphate ends, sRNAs in total extracts and AGO4 immunoprecipitates prepared from Col-0 and *dcl2/3/4* were treated with Terminator exonuclease (Epicentre) following the manufacturer's standard procedure. A synthetic 21-nt spike-in RNA with 5' monophosphate was added in the reactions and used as a positive control. To examined whether the

sRNAs are single- or double-stranded, the sRNA samples were incubated with RNase T1 (Thermo SCIENTIFIC) at 37°C for 20 mins and *AtREP2* sRNAs were probed by Northern blot. Spike-in RNAs including a 50-bp double-stranded RNA with 2-nt 3' over-hang (ds) and a 50-nt single-stranded (ss) RNA were added in the reactions and used as controls for RNase T1 digestion. Reactions were stopped by phenol/chloroform extraction, and the RNAs were collected by ethanol precipitation. The resultant RNAs were subjected to Northern blot analysis using a probe to detect *AtREP2*-derived sRNAs.

Oligonucleotides

Primers and probes used in this study are listed below.

Oligo	Sequence (5'-3')
Probes for small RNA Northern blot	
anti-miR171	GATATTGGCGCGGCTCAATCA
anti-ta-siR255	TACGCTATGTTGGACTTAGAA
anti-AtREP2	GCGGGACGGGTTTGGCAGGACGTTACTTAAT
anti-SIMPLEHAT2	TGGGTTACCCATTTTGACACCCCTA
anti-AtSN1-1	CACCAACGTGTTGTTGGCCCAGTGGTAAATCTCTCAGATAGAGGTGCTGGATTCGAGACA
anti-AtSN1-2	CTCTCAGATAGAGGTGCTGGATTCGAGACACGTTGGGAAGGATCTCTTTTCCAAAGAAGA
anti-AtSN1-3	CGTTGGGAAGGATCTCTTTTCCAAAGAAGATGAATTTCTGGTATGGGTCCCGCCTCTGGG
anti-AtSN1-4	ATTTCTGGTATGGGTCCCGCCTCTGGGAGATGTAAGGCCTTTGGGCCTGAACTTCCAG
anti-At4TE27090	GTCTGAGAAGATTATGGGTTTCGGCCCAGAACCTCATGGATTTC
anti-IGN5-1	AGTTTTGTTGGCTATTGGGCCGAATAACAGCAAGTCCTTT
anti-IGN5-2	TCTGAGGTATTCCATAGCCCCTGATCCATCTTTCAATTAG
anti-sidRL2514	GATACGGGCTTTGCCCTCCATGCCCAATTCGGCCCAAACAC
anti-sidRL5049	TGGAGAAAAGATCTCTCCCAACGTACCTTGAACCCAGCATCT
Primers for bisulfite sequencing	
AtREP2-BiF	AYAYTTTTGTGGTATATTTGAAATG
AtREP2-BiR	CCAACCCRTACCRTAATAAAACTAT
SIMPLEHAT2-BiF	TAGGGGTGTTAAAATGGGTTAAAATT
SIMPLEHAT2-BiR	ATAAAAAAATTACGAATTTACTTTTTCTC
IGN5-BiF	GTTYYYGAGAAGAGTAGAAYAAATGYTAAAATGTATYATGYGGTT
IGN5-BiR	RRACTAARTCTTRTCRAACAARRACCCAACCATRTCCRCTTAAAAA
AtSN1-BiF	GTTGTATAAGTTTAGTTTTAATTTTAYGGATYAGTATTAATTT
AtSN1-BiR	CAATATACRATCCAAAAAACARTTATTAAAATAATATCTTAA
sidL5049-BiF	AGAGTAAGTTATTGGGGTTGGAGTG
sidL5049-BiR	CAATCATTAATARTCCAARTCTTTC
sidL2514-BiF	TGAGTGGTGATAGGTTYTGGAT
sidL2514-BiR	CCAAATCCTAAATCATRAAAAATTTTAA
sidL14057-BiF	TTAATTTGTYAYATTTAGYGTGTTT
sidL14057-BiR	AAATTACTCCRCTTACTCACTTAA
Primers for Chop-PCR	

IGN5-HaeIII-F	TCCCGAGAAGAGTAGAACAAATGCTAAAA
IGN5-HaeIII-R	CTGAGGTATTCCATAGCCCCTGATCC
AtSN1-HaeIII-F	ACCAACGTGCTGTTGGCCCAGTGGTAAATC
AtSN1-HaeIII-R	AAAATAAGTGGTGGTTGTACAAGC
ACT2-HaeIII-F	CGAGCAGGAGATGGAAACCTCAAA
ACT2-HaeIII-R	AAGAATGGAACCACCGATCCAGACA

Primers for AGO4-ChIP

ACTIN-ChIP-F	GAGAGATTCAGATGCCCAGAAGTC
ACTIN-ChIP-R	TGGATTCCAGCAGCTTCCA
AtREP2-ChIP-F	AATGGCTGCCACTGGATTTT
AtREP2-ChIP-R	TGGAGAGATTTTGGGAAGATTG
solo LTR-ChIP-F	GGATAGAGATGAATGATGGATAATGACA
solo LTR-ChIP-R	TTATTTTGATCAGTGTTATAAACCGGATA
IGN5-ChIP-F	AAGCCCAAACCATACACTAATAATCTAAT
IGN5-ChIP-R	CCGAATAACAGCAAGTCCTTTTAATA
SIMPLEHAT2-ChIP-F	TGATGGGTCAAATGAGTTGATGA
SIMPLEHAT2-ChIP-R	TTGGGTTAATTGGGTAAACCAT

Primers for genotyping

LBb1.3	ATTTTGCCGATTTCGGAAC
LB3	TAGCATCTGAATTTCATAACCAATCTCGATACAC
GABI-T8474	ATAATAACGCTGCGGACATCTACATTTT
dcl1(dcl1-9)_geno_R1	AGTAGAAGAACCGCAGCTGAATC
LB-pGV_CAF1_2(dcl1-9)	GACACACACATCATCTCATTGATGCTTGG
dcl2-1-genotyping-LP	TGAATCATCTGGAAGAGGTGG
dcl2-1-genotyping-RP	CTTCACAGGAGTTTTTGGCTG
dcl3-1-genotyping-LP	ACAGGTAACCTTGCCATGTTG
dcl3-1-genotyping-RP	TGGAAAAGTTTGCTACAACGG
dcl4-2-genotyping-P1	TCTCCATATTGACCATCATACTCATT
dcl4-2-genotyping-P2	GGCTGCACAGCTGATGATTACAA
dcl4-2-genotyping-P3	GCCGCTCGAGATCATCAGCAAAGGAAT
nrpd1a-3-genotyping-LP	TTTTGATCCCTTGATCACCTG
nrpd1a-3-genotyping-RP	CTGATGGTCCCGTTGAAGATA
rdr2-1-genotyping-P1	ACACATTAGGACTAACAAATTTACC
rdr2-1-genotyping-P2	ATGGTGTCAGAGACGACGACGAACCGATCAAC
rdr2-1-genotyping-P3	TAGCATCTGAATTTCATAACCAATCTCGATACAC
ago4-1-genotyping-F	TGACTGACAGCTGAAAATGGGATGTGGAT
ago4-1-genotyping-R	GCCACTCCCTAGAACTCACCACCTAAGTT
SALK_012547 LP	CTTCTGCCCAAGGCTAATATG

SALK_012547 RP	TGATGCTGATTTTGCTGATTG
SALK_066543 LP	ATAAAGAGCCGGTGGAGAAAC
SALK_066543 RP	CGTCATCGTCAATGCTTAAGC
CS381695 LP	TGGTGATCAAGGGTGAAACTC
CS381695 RP	TCACAGAGTCACCATCAATTCC
SALK_144940 LP	CAGTCATCTTGAACTTTGGCC
SALK_144940 RP	TGTTTGGAAAATGATATGGGG
SALK_072496 LP	GAAAGGTAGATGAAGGGTGCC
SALK_072496 RP	GCTGCAGATACCAAAAAGCAG
SALK_099860 LP	GCTCTAACAGAGTCACAGCCG
SALK_099860 RP	GTTTTGCCGTTAAGAGGGAAG
SALK_122492 LP	AGTGGGCATTGATAAATGCAG
SALK_122492 RP	ACCATTCGAAGGTTTCTCTGC
SALK_003278 LP	CATATCTGGACAGTCGCAACC
SALK_003278 RP	TCCGACCGTACAAGCTACAAC
SALK_078063 LP	CGCAAAATCAAATCAAAGACAG
SALK_078063 RP	CACACGTGATCCAAACACAAG
SALK_113786 LP	ACTCACCACTTAATTGCACGC
SALK_113786 RP	ACCCATCATCTCCATCATCAG
SALK_145649 LP	ACTTTCTGCTCTCTTTTGCCC
SALK_145649 RP	CCAATGGCTTCTTCTTCTTCC
SALK_024707 LP	AACAATAGATCAATGCGACGG
SALK_024707 RP	GATTTCCCAGAGCAGGAAGAC
SALK_068417 LP	GATTAATGATTCGTATGGCGG
SALK_068417 RP	GTTGTTGCGTCCGAATATCAC
SALK_050279 LP	TGGTTGGATACGAACATTTCG
SALK_050279 RP	TCACAGTCGTGACTCGTGAAG

Supplemental Reference

Berger, F., Haag, J.R., Pontes, O., and Pikaard, C.S. (2009). Metal A and Metal B Sites of Nuclear RNA Polymerases Pol IV and Pol V Are Required for siRNA-Dependent DNA Methylation and Gene Silencing. PLoS ONE *4*, e4110.

Blevins, T., Rajeswaran, R., Aregger, M., Borah, B.K., Schepetilnikov, M., Baerlocher, L., Farinelli, L., Meins, F., Jr., Hohn, T., and Pooggin, M.M. (2011). Massive production of small RNAs from a non-coding region of Cauliflower mosaic virus in plant defense and viral counter-defense. Nucleic acids research *39*, 5003-5014.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. The Plant journal : for cell and molecular biology *16*, 735-743.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. The Plant journal : for cell and molecular biology *45*, 616-629.

Harris, E.Y., Ponts, N., Le Roch, K.G., and Lonardi, S. (2012). BRAT-BW: efficient and accurate mapping of bisulfite-treated reads. Bioinformatics *28*, 1795-1796.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nature Genetics *38*, 721-725.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C.*, et al.* (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell *133*, 116-127.

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.

Stroud, H., Greenberg, Maxim V.C., Feng, S., Bernatavichute, Yana V., and Jacobsen, Steven E. (2013). Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. Cell *152*, 352-364.

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

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS biology *2*, E104.

Ye, R., Wang, W., Iki, T., Liu, C., Wu, Y., Ishikawa, M., Zhou, X., and Qi, Y. (2012). Cytoplasmic Assembly and Selective Nuclear Import of Arabidopsis ARGONAUTE4/siRNA Complexes. Molecular Cell *46*, 859-870.

Zhang, H., Ma, Z.Y., Zeng, L., Tanaka, K., Zhang, C.J., Ma, J., Bai, G., Wang, P., Zhang, S.W., Liu, Z.W.*, et al.* (2013). DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proceedings of the National Academy of Sciences of the United States of America *110*, 8290-8295.