S1 Appendix

S1 Methods

Plant materials and growth conditions

Plant materials included *Arabidopsis thaliana* wild-type (Col-0), T-DNA insertion lines for *nrpd1* (aka, *nrpd1a-4*, ABRC, CS66151), *rdr2-1* (ABRC, CS66076), *agl40* (SALK107011), and *agl91* (SALK042875). All plants were grown under an illumination cycle of 16-h day and 8-h night at 22ºC (day) and 20ºC (night). For crosses, flowers in floral stage 12C (1) were emasculated and manually pollinated 24 to 48 hours after emasculation. Seeds were collected at designated days after pollination (DAP) for β-Glucuronidase (GUS) staining, functional assays, fixation for LCM, or whole seed RNA extraction.

Laser capture microdissection and library preparation

Siliques from crossed flowers at 6 DAP were collected between 14:00 and 17:00 to minimize possible circadian effects, sliced along the replum to partially expose the locules and seeds within, and cut longitudinally into three parts. The silique portions and seeds contained within were then fixed overnight in 3:1 ethanol:acetic acid at 4º C. LCM samples were processed as previously described (2) or via a rapid microwave fixation process utilizing a Pelco Biowave 34700 equipped with Cold Spot (Ted Pella). The microwave fixation protocol used an ethanol dehydration series, followed by butanol and paraffin infiltration (3).

The resulting paraffin blocks were sectioned at 7μm with serial sections placed on PEN-foil slides (Zeiss), which were then dipped twice in Xylenes to remove the paraffin and air dried. LCM of seed coat and endosperm was carried out using a PALM Microbeam (P.A.L.M Microlaser Technologies, München, Germany). Briefly, without the collector in position, any embryo present was cut around and removed to prevent contamination of seed coat and endosperm samples, followed by collection of endosperm into adhesive cap collection tubes (Zeiss, Oberkochen, Germany). When possible, all available endosperm was collected from individual sections, and the section consisted a combination of micropylar, peripheral and chalazal endosperm. After the embryo has been removed and the endosperm had been collected, the remaining seed coat, comprising the distal and chalazal seed coat regions, was collected (Supplementary Fig. 1).

Approximately 500-600 individual microdissected endosperm or seed coat sections were used per replicate. Total RNA was isolated using the RNAqueous-Micro Kit (Invitrogen, Carlsbad, California). 50 ng of total RNA was used to generate libraries using the TruSeq Small RNA Library Kit and Indices Set C (Illumina, San Diego, California). Libraries were size selected on 6% TBE-PAGE gels and checked for quality on a BioAnalyzer 2100 using an RNA Pico chip (Agilent, Santa Clara, California).

The *nrpd1*X*nrpd1* (*n*X*n*) sample was prepared and sequenced as described in the previously published paper (4).

Sequencing and bioinformatic analysis

Sample libraries from LCM samples were multiplexed and sequenced on a HiSeq4000 using either 100 or 150 base single-ended reads to a depth of over 50 million reads per sample. All mapping of reads was done using Bowtie (5), requiring perfect alignment while allowing individual reads to map to multiple locations. Reads were clipped to remove adapter sequences and then reads which aligned to TAIR10 rRNA sequences were discarded. This processing resulted in a minimum of 25 million reads per sample. The size ranges of reads for between 20

and 25-nt are shown in Supplementary Fig. 4, and show 21 and 24-nt reads as the most common class of sequenced small RNAs.

Sequence reads with 21-nt and 24-nt from each sample were split into separate files for further analysis. These reads were mapped to the TAIR10 genome. Mapped reads were given a weighted score consisting of the number of identical reads at a position divided by the number of distinct positions in the genome those reads mapped. The genome was then divided into 100-bp bins and the weighted scores for reads contained within the bins were added together to give a set of bin-scores. In order to normalize the bin-scores between samples, 21-nt reads were mapped to known *A. thaliana* mature miRNA sequences in miRBase and count data determined. As miRNA distributions are relatively unaffected by the mutations of RdDM pathway genes, scaling factors were determined using the calcNormFactors function of the edgeR library (6) in R (http://www.R-project.org/). These scaling factors were multiplied with the 100-bp bin scores to give the normalized scores used in subsequent analyses.

The scores for 100-bp bins between biological replicates were compared and correlation coefficients were calculated (Supplementary Fig. 2). In order to determine differentially expressed loci, scores for 100-bp windows were compared between tissues or genotypes from one set of biological replicates. A 4-fold difference in the signal counts was used to identify differentially expressed tiles between conditions. All possible combinations of differential expression between a particular tile were considered to determine a particular class of locus (ECec, ECEc, etc.) and are summarized in Supplementary Fig. 8B. The same analysis was performed for the second set of biological replicates. To remove spurious loci from 100-bp bins with low signals, a minimum signal cutoff score of 20 was used, as it reflected a consensus inflection point in the number of differentially expressed loci identified for all loci types (Supplementary Fig. 4). Next, the results from the two sets of replicates were compared, and only those that were assigned to the same class of locus in both replicates were used. Finally, 100-bp bins were consolidated if they were adjacent to or within 100-bp of another bin of the same class in order to give a final set of loci for each class and associated genome coordinates. In order to determine the raw number sites with 24-nt small RNAs regardless of differential abundance or class, the number of 100-bp bins with replicated averaged signals above the minimum cutoff value was counted.

In order to differentiate between endosperm specific (EcEc) loci that were *NRPD1*-depenendent or *NRPD1*-independent, we examined 24-nt siRNA reads from *nrpd1*X*nrpd1* seed sample (6 DAP) normalized against the LCM reciprocal cross samples. EcEc loci with the count value exceeding the minimum expression threshold in the *nrpd1*X*nrpd1* were considered to be *NRPD1*-independendent, and the rest were counted as *NRPD1*-dependent.

For comparison with vegetative H3K9me2 levels, ChIP-Chip data was from GSE12383 (7) was used. H3K9me2 loci were identified as at least two adjacent tiles with an intensity z-score ≥ 0.2, as described previously (7). Genome coordinates were then updated to reflect the TAIR10 genome using the NCBI Genome Remapping Service (http://www.ncbi.nlm.nih.gov/genome/tools/remap). Small RNA distribution and H3K9me2 levels was displayed in Integrative Genomics Viewer (IGV) (http://www.broadinstitute.org/igv/) (8, 9). Overlap and distance calculations between loci and genes, TEs, DMRs performed using "Genomic Ranges" package in R (10).

Pilot studies using manually dissected embryos, endosperm and seed coat

Approximately ~3,000 seeds of ~60 siliques (~50 seeds per silique) at 6 DAP were used to manually dissect each region for extracting total RNA. Embryos were rinsed three times with

Sorbitol solution (0.3M) before RNA extraction. Small RNA libraries were prepared as previously described (4). In brief, an aliquot of total RNA (10 µg) was resolved in a 15% urea-polyacrylamide gel, and the fraction of 18-30-nt small RNAs was recovered. Purified small RNAs were ligated to 5' and 3' RNA oligo adapters and reverse transcribed to produce first-strand cDNAs. PCR-amplified cDNAs were sequenced using Illumina HiSeq2000.

Small RNA sequences were trimmed for 3' adapters, collapsed into unique sequences and mapped to *A. thaliana* genome (TAIR10, November 2010 release) using CASHX with perfect match option (http://carringtonlab.org/resources/cashx) and according to the previous protocol (4). The sequences from chloroplast, mitochondrial, tRNAs and snRNAs were excluded from the analysis. Small RNA reads were normalized by dividing the total number of reads of a library by 10 millions. Multiple-hit read was assigned equally to each locus and divided by the number of hits in the genome. siRNA loci were identified using a Python script as regions containing at least 20 distinct reads, each < 200-nt apart. The siRNA loci were then combined to a merged locus set based on their ranges in each library. siRNA loci from endosperm, embryo and seed coat were then compared and removed if they overlapped to generate a set of embryo, endosperm and seed coat specific loci.

This dataset was used only for the comparisons between LCM datasets and other manually dissected samples (Fig. S16) because of a potential contamination between seed coat and embryo samples. Dendgrogram and bootstrap values generated using the Pvclust package (11) in R.

Principal Component Analysis (PCA)

PCA was performed using the prcomp function in R on binned and normalized siRNA data. Bins which did not pass the minimum expression filter were excluded from the analysis. PC 1 and 2 accounted for 87.4% of the variance in the data.

Dosage Effects

To check the effect of *NRPD1* dosage on expression levels of siRNA loci, we generated MA plots using the maPlot function of edge R (Fig. S9). Signal in the endosperm of Col x *nrpd1* and *nrpd1* x Col crosses were compared. The endosperm has a 2:1 ratio of functional *NRPD1*, if there is a dosage effect of *NRPD1* on siRNA loci, the expression levels would be expected to have a 2-fold ($\log_2$ 1) difference. The expression bias of ECec loci was significantly higher, with the median equal to a 111-fold difference. The seed coat siRNAs (eCec) also showed a high degree of bias, while the non-maternally biased loci (EcEc and ECEc) have distributions centered on a 1:1 ratio.

Methylation Analysis

Bisulfite sequencing data were obtained from GSE15922, which provided fractional CG, CHG and CHH methylation levels (12). The genomic coordinates and methylation levels were split into separate files for each methylation context and were displayed in the IGV as separate tracks with a scale of a 10-kb window centered on the genes of interest, *AGL40* and *AGL91*.

Gene Ontology (GO) Enrichment Analysis

Gene lists were analyzed for GO enrichement using the agriGO tool (http://bioinfo.cau.edu.cn/agriGO/). Enrichment was checked against a background TAIR10 genome loci using the hypergeometric distribution and correcting for multiple testing using Hochberg FDR < 0.05.

Plasmid construction

The plasmid *pAGL91:AGL91:GUS:3'TE* was generated using the following four steps. (1) A 2149-bp sequence, comprising a multiple cloning site (with 6 unique restriction sites), SV40 nuclear localization signal, restriction sites of *Pml*I and *Aat*II, and GUS coding sequence, was PCR-amplified using the plasmid DNA pFGUS2a (GenBank accession: KC920577) and synthetic oligos as templates. The amplicon was digested and inserted into the *Eco*RI/*Bam*HI of the plasmid vector pFAMIR (provided by Ramin Yadegari at University of Arizona). (2) A 964-bp sequence immediately downstream of the coding sequence of *AGL91* (At3g66656) was amplified from genomic DNA (Col-0). The sequence was digested and inserted into the *Bam*HI/*Xma*I site of the above plasmid. (3) The 537-bp coding sequence of *AGL91* was amplified from genomic DNA (Col-0). The sequence was digested and inserted into the *Pml*I/*Aat*II site of the same plasmid. (4) The 2103-bp sequence immediately upstream of the coding sequence of *AGL91* was amplified from genomic DNA (Col-0). The sequence was digested and inserted into the *Dra*III/*Rsr*II site of the above plasmid.

To generate plasmid *pAGL40:AGL40:GUS*, a 2473-bp sequence, comprising 1419-bp upstream regulatory sequence, 1041-bp coding region of *AGL40* and 13-bp linker sequence, was amplified from genomic DNA (Col-0). The sequence was digested and inserted into the *XhoI*/*Nco*I site of the plasmid vector pFGUS2a.

To generate the overexpression constructs, the 2059-bp regulatory sequence of *SUP16* (*At5g27880*) (13) was amplified from genomic DNA (Col-0). The amplicon was digested and inserted into the *Eco*RI/*Rsr*II site of pFAMIR. The coding sequences of *AGL40* (1044-bp) and *AGL91* (537-bp) were amplified and inserted into the *Aat*II/*Nco*I site of the above plasmid.

Quantitative RT-qPCR analysis

Total RNA (200 ng) from LCM dissection or whole seeds was used for was transcribed by reverse transcriptase using Omniscript (Qiagen) oligo-dT primers in reactions with a final volume of 20 µl. PCR was performed on a LightCycler 96 (Roche) using FastStart Universal SYBR Green Master mix (Roche) with three technical replicates. Relative quantification was carried using the ΔΔCt method using *PP2AA3* (At1g13320) amplification as a control. Endosperm and seed coat specific markers were obtained from the published data (2). Primers used were as follows: PP2AA3, forward TAACGTGGCCAAAATGATGC, and reverse: GTTCTCCACAACCGCTTGGT; At3g54890 forward: GCCATTGAGTTCTTAGCCATTGC, and reverse: ACAGAATCCTACAAACGCCAACA; At3g51820, forward: GCACTCTTACGCCAGATGTTGTTGTTCTAAC, and reverse: AAGCAACCAACGCCAACGCATAATAAGG; At1g08810, forward: ATACCTACCACAAAGAACGGACAATG, and reverse: ACGCTCTCATCCAACCCTCAAG; At1g72260, forward: AATTGTGAATGGAGCGTCGG, and reverse: TACGACACATGCACACACAC; At1g12010, forward: TCAAAGGGCTTAGGGCTCACACAGATG, and reverse: TGACAATGGAATGCTTGAGAGGAGGAACA; At4g36590, forward: CTTTTGGCCATCCAAGTGTT, and reverse: CAGCTGGTTGGTTTCATTCA; At3g6656, forward: CAAGACACAAACACGAAGCAA
, and reverse: GGAGAAAAGACAACAATACCAACC.

Transgenic plants

The AGO4:GFP reporter lines were constructed by inserting the GFP coding sequence at the *Apa*I site upstream of the FLAG sequence of the pAGO4:FLAG:AGO4 vector as described in the previous paper (14). This plasmid was transformed into the *ago4-3* (Col-0) mutant

background (WiscDSLox338A06), and a transgenic line containing a single insertion site was carried forward to homozygosity. Methylation-sensitive PCR analysis of the *AtSN1* locus has confirmed that the fusion protein largely complements the *ago4* mutant phenotype.

The remaining transgenic lines were produced as follows. Each plasmid construct was introduced into plants of the appropriate genotypes, namely, the wild-type *A. thaliana* (Col-0), *nrpd1*, *rdr2*, and *drm1/2* mutants through the floral dip method (15). T1 seeds (harvested from floral-dipped plants) were sterilized and germinated on MS agar plates containing 30 μM glufosinate ammonium (GoldBio, https://www.goldbio.com/) and 50 mg/L of cefotaxime (GoldBio). Glufosinate ammonium-resistant seedlings were transferred once to fresh plates containing the same selection reagents. Two-week-old T1 plants from 10 or more independently-derived transgenic events were transferred to soil. Developing T2 seeds (harvested from T1 plants) were harvested for GUS-staining or RNA extraction.

In each genetic cross, 3-5 independent transgenic lines showing strong GUS activities were selected and used for each genetic cross, and ~10 "hybrid" siliques per transgenic line were used for each assay.

The transgenic plants designated as follows: $Wt^{AGL91:GUS}$ and $nrpd1^{AGL91:GUS}$ were transgenic plants that expressed *AGL91:GUS* under the native promoter in the wild-type (Wt, Col-0) and *nrpd1* mutant, respectively. Likewise, $Wt^{AGL40:GUS}$ was transgenic plants that expressed *AGL40:GUS* under the native promoter in the wild-type (Wt, Col-0). Transgenic plants that expressed *AGL40* or *AGL91* under the strong endosperm-specific promoter *pSUP* (*At5g27880*) (13) was named $Wt^{SUP:AGL91}$ and $Wt^{SUP:GL40}$, respectively.

GUS staining and microscopy

The procedure for GUS staining was modified from a published protocol (16). In brief, seeds were removed from siliques and stained at 37°C for 24 hours in GUS staining solution [5 mM potassium ferricyanide, 5mM potassium ferrocyanide, 100mM sodium phosphate (pH 7.0), 0.05% Triton-X-100, 1mg mL^{-1} X-Gluc (GoldBio, https://www.goldbio.com/)]. Stained seeds were fixed in Carnoy's fixative (ethanol:glacial acetic acid, 3:1) for at least 3 hours, washed once with 90% ethanol, and kept in 70% ethanol for at least 24 hours. Immediately before observation, seeds were cleared in a clearing solution (chloral hydrate:glycerol:water, 8:1:2, wt/vol/vol). Pictures were taken using a compound light microscopy (Leica DM2500, Wetzlar, Germany) equipped with Normarski Optics.

Cleared seed image

Siliques containing seeds at the desired stage were examined under a dissecting microscope. Approximately 15 to 20 seeds were then removed from the silique and placed on the surface of a drop (1-2 cm diameter) of the clearing solution (chloral hydrate:glycerol:water, 8:1:2, wt/vol/vol) on a glass slide. After 1 hour, cleared seeds on the glass slide are visualized using a compound light microscope coupled with Nomarski optics (Leica DM2500, Wetzlar, Germany).

Confocal microscopy

6 DAP seeds were dissected out of siliques and mounted in water on microscope slides with coverslip. Images were taken using a LSM 710 confocal (Zeiss, Oberkochen, Germany) using 10x objective and illumination at 488 nm. Light was collected at 493-600 nm for GFP (green channel) and 609-700 nm for autofluorescence (red channel). Z-stacks were collected for each seed with images spaced 10 μm apart. To maintain comparability of GFP intensity between images, all microscope settings, including laser intensity, pinhole width and detector settings

were unchanged, with the exception of image rotation. Red and green channels were merged, and Z-stacks comprised of 20-25 images were used to generate animations through an individual seed, exported using ImageJ (http://imagej.nih.gov/ij/, 1997-2016) with the Fiji package (17).

Seed size and weight measurement

Seed length and width were measured using Image J software on the seed images taken at 6 DAP when endosperm is cellularized and seed size is fixed (18). Three biological replicates each containing 50 seed images were used to calculate means and standard error of the mean (s.e.m.) to test statistical significance. Seed weight was measured by weighing dried mature seeds on an analytical balance (Mettler Toledo AB54-S, Columbus, Ohio). Three biological replicates each containing ~500 seeds were used to calculate means and s.e.m. and to test the statistical significance. The unit of weight was converted to mg per 100 seeds. The seed-size increase was not statistically significant in the Wt^{AGL91} lines, probably due to the redundant functions of *AGLs* and/or promoter effects.

References

1. Christensen CA, King EJ, Jordan JR, & Drews GN (1997) Megagametogenesis in Arabidopsis wild type and the Gf mutant. (Translated from English) *Sex Plant Reprod* 10(1):49-64 (in English).
2. Belmonte MF*, et al.* (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. (Translated from eng) *Proc Natl Acad Sci USA* 110(5):E435-444 (in eng).
3. Takahashi H*, et al.* (2010) A method for obtaining high quality RNA from paraffin sections of plant tissues by laser microdissection. *J Plant Res* 123(6):807-813.
4. Lu J, Zhang C, Baulcombe DC, & Chen ZJ (2012) Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of Arabidopsis seeds. (Translated from eng) *Proc Natl Acad Sci USA* 109(14):5529-5534 (in eng).
5. Langmead B, Trapnell C, Pop M, & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. (Translated from eng) *Genome Biol* 10(3):R25 (in eng).
6. Robinson MD, McCarthy DJ, & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.
7. Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, & Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. (Translated from eng) *PLoS ONE* 3(9):e3156 (in eng).
8. Robinson JT*, et al.* (2011) Integrative genomics viewer. (Translated from eng) *Nature Biotechnology* 29(1):24-26 (in eng).
9. Thorvaldsdottir H, Robinson JT, & Mesirov JP (2012) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. (Translated from Eng) *Briefings in bioinformatics* (in Eng).
10. Lawrence M*, et al.* (2013) Software for Computing and Annotating Genomic Ranges. (Translated from English) *PLoS Computational Biology* 9(8) (in English).
11. Suzuki R & Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22(12):1540-1542.
12. Hsieh TF*, et al.* (2009) Genome-wide demethylation of Arabidopsis endosperm. *Science* 324(5933):1451-1454.
13. Wang D*, et al.* (2010) Identification of transcription-factor genes expressed in the Arabidopsis female gametophyte. (Translated from eng) *BMC plant biology* 10:110 (in eng).

14. Havecker ER*, et al.* (2010) The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* 22(2):321-334.
15. Bent AF & Clough SJ (1998) *Agrobacterium* germ-line transformation: Transfromation of *Arabidopsis* without tissue culture. *Plant Molecular Biology Manual*, eds Gelvin SB & Schilperoort RA (Kluwer Academic Pub., Dordrecht, The Netherlands), Vol Section B, pp 1-14.
16. Bemer M, Heijmans K, Airoldi C, Davies B, & Angenent GC (2010) An atlas of type I MADS box gene expression during female gametophyte and seed development in Arabidopsis. (Translated from eng) *Plant Physiol* 154(1):287-300 (in eng).
17. Schindelin J*, et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9(7):676-682.
18. Meinke D (1994) Seed development in *Arabidopsis thaliana*. *Arabidopsis*, eds Meyerowitz EM & Somerville C (Cold Spring Harbor Laboratory, New York), pp 253-295.

List of supplemental Figures, Datasets and Movies

Fig. S1. Laser-capture microdissection (LCM) of endosperm and seed coat tissues
Fig. S2. Analysis of correlation of 24-nt siRNAs between biological replicates
Fig. S3. Principal Component Analysis (PCA) of small RNA profiles
Fig. S4. Small RNA size distributions and *AGO4:GFP:AGO4* localization patterns
Fig. S5. Minimum expression cutoff values for locus classes
Fig. S6. Specificity of gene expression in LCM samples
Fig. S7. Heatmap of siRNA expression levels in four groups
Fig. S8. Four groups of siRNA loci and their patterns in seed sub-regions
Fig. S9. Fold changes of siRNAs in four locus groups
Fig. S10. Relationship between siRNA loci and genomic and epigenomic features
Fig. S11: Spatial-temporal *AGL40* expression is dependent on maternal *NRPD1* or *RDR2*
Fig. S12. *NRPD1*-dependency of EcEc loci
Fig. S13. Heatmap showing spatial-temporal transcript levels of RdDM pathway (top) and non-RdDM pathway (bottom) genes
Fig. S14. Methylation of *AGL40* and *AGL91* loci
Fig. S15. Increased expression levels of endogenous *AGL91* and *AGL40* when the maternal *NRPD1* is mutated
Fig. S16. Comparison of small RNA profiles in the LCM and manually dissected samples

Dataset S1. List of 24-nt loci in the endosperm and seed coat of reciprocal crosses with the *nrpd1* mutant.
Dataset S2. Group 1 siRNA loci (ECec)
Dataset S3. Group 2 siRNA loci (eCec)
Dataset S4. Group 3 siRNA loci (ECEc)
Dataset S5. Group 4 siRNA loci (EcEc)
Dataset S6. List of siRNA-associated genes in four locus groups

Movie S1. pAGO4:GPP:AGO4 expression in transgenics (*ago4* mutant). Movie of confocal images in developing seeds showing GFP:AGO4 expression patterns in the *ago4* mutant (*AGO4:GFP*X*AGO4:GFP*).

Movie S2. pAGO4:GPP:AGO4 expression in *AGO4:GFP*XWt plants. Movie of confocal images in developing seeds showing GFP:AGO4 expression patterns in the *AGO4:GFP*XWt plants.

Movie S3. pAGO4:GPP:AGO4 expression in WtX*AGO4:GFP* plants. Movie of confocal images in developing seeds showing GFP:AGO4 expression patterns in the WtX*AGO4:GFP* plants.

Movie S4. Confocal images in the WtXWt plants. Movie of confocal images in developing seeds showing no GFP:AGO4 expression in the WtXWt plants.

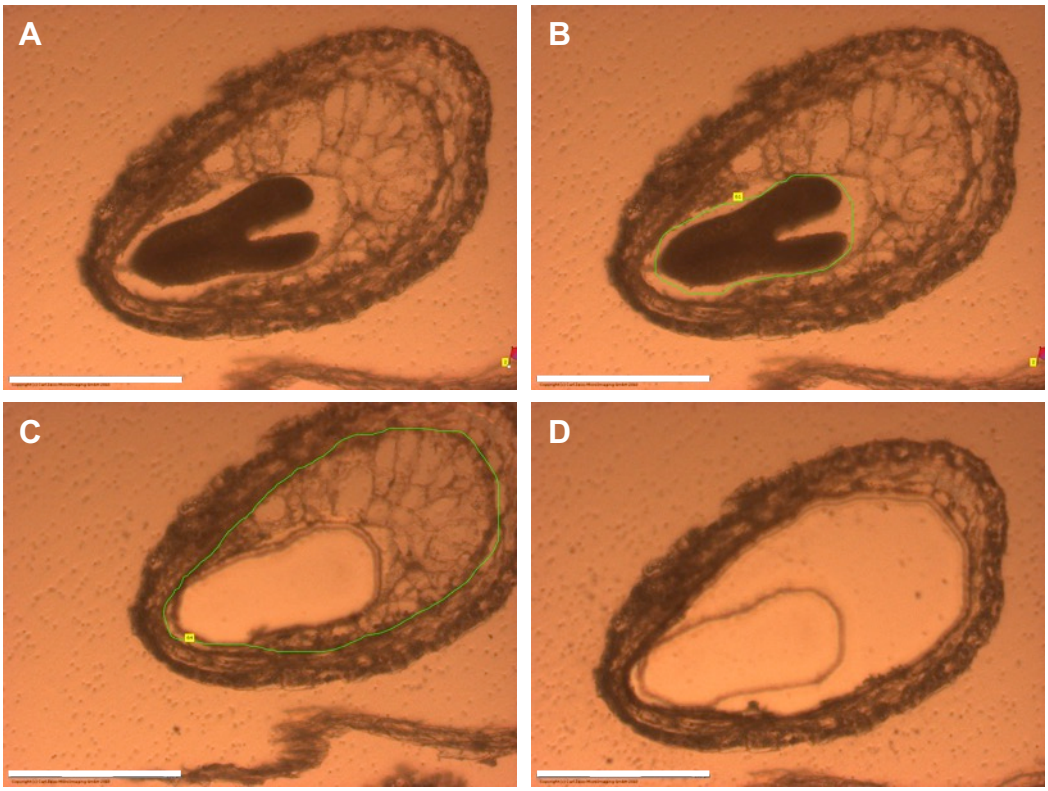


Fig. S1. Laser-capture microdissection (LCM) of endosperm and seed coat tissues. Examples of collection order of seed regions to minimize tissue contamination.
(*A*) Sectioned *Arabidopsis* seed containing linear cotyledon-stage embryo.
(*B*) Example of laser-cutting path (green) to remove the embryo.
(*C*) Example of the laser path to be used for collecting the endosperm after the embryo has been removed.
(*D*) Once the endosperm has been collected, the remaining seed coat tissue was harvested by cutting the surrounding PEN-foil membrane. Scale bar = 100 μm.

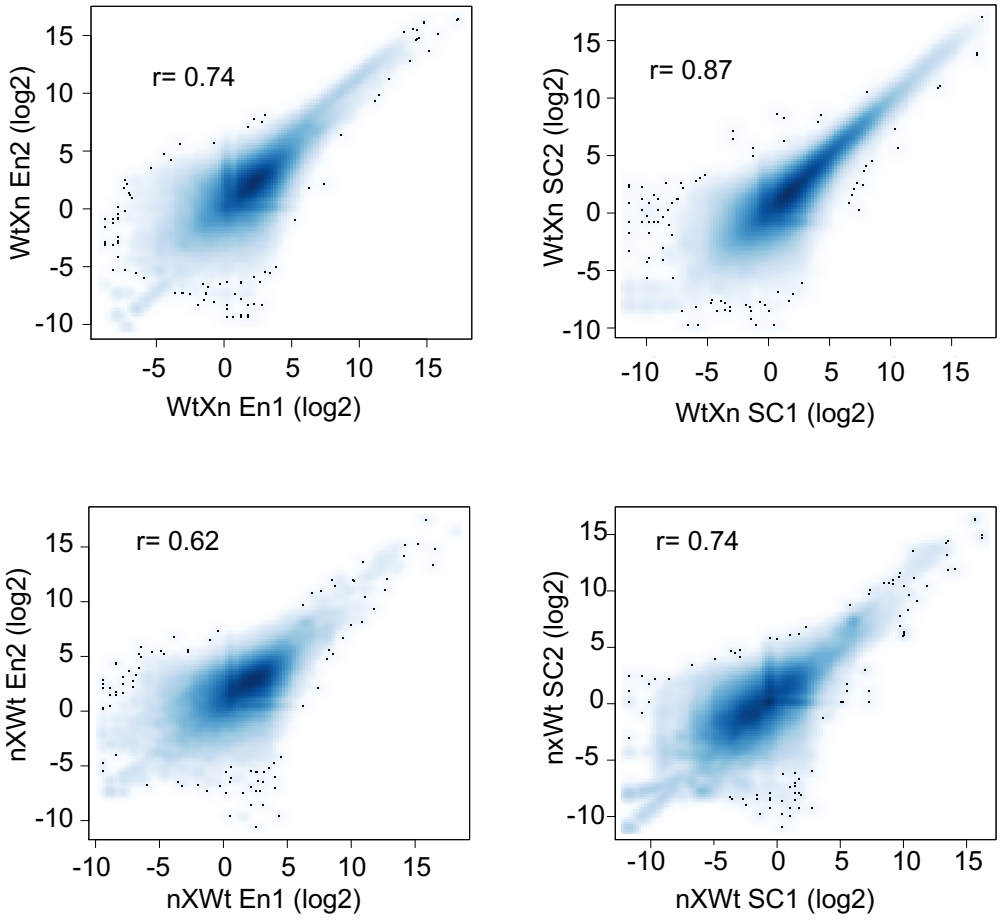


Fig. S2. Analysis of correlation of 24-nt siRNAs between biological replicates. Density scatterplot of 24-nt siRNA abundance scores from 100-bp bins showing correlation coefficients between two replicates (1 and 2) in the seed coat (SC) and endosperm (En) of reciprocal crosses (*n*XWt and WtX*n*) between *nrpd1* (*n*) and Wt plants.

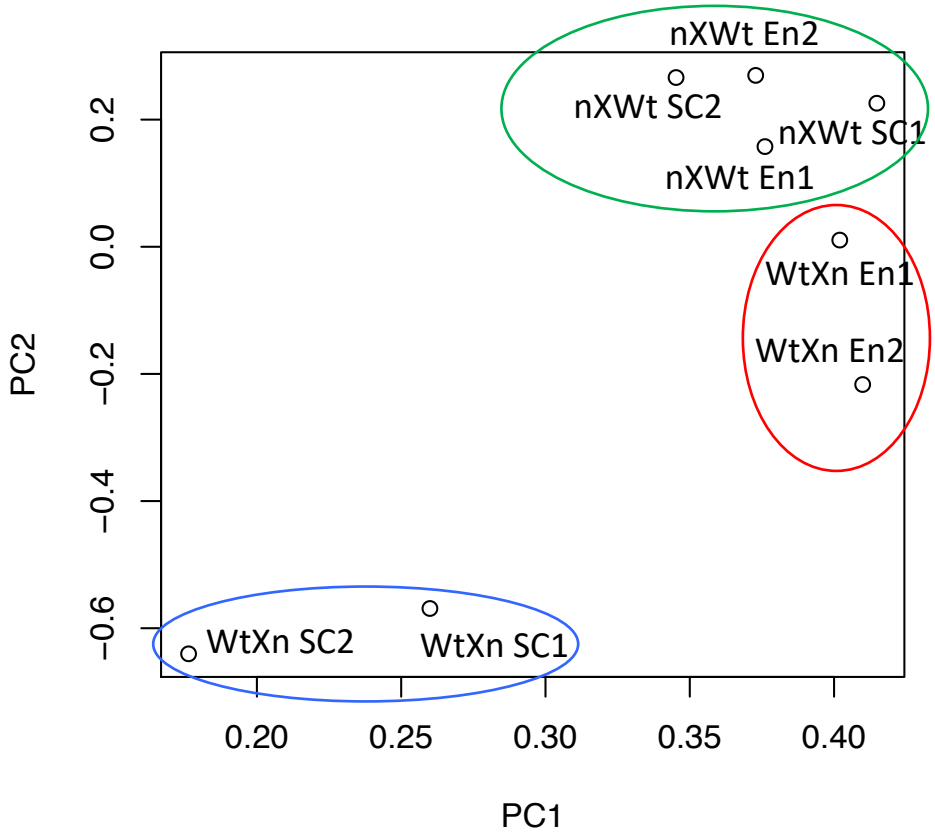


Fig. S3. Principal Component Analysis (PCA) of small RNA profiles.
PCA revealed separation between the endosperm (En1 and En2 replicates, red) and seed coat (SC1 and SC2 replicates, blue) in the Wild type X *nrpd1* (WtXn) cross. In the reciprocal cross lacking maternal *nrpd1* activity (nXWt), the loss of 24-nt siRNA resulted in a close clustering amongst both the endosperm and seed coat samples (green). Biological replicates displayed slight separation, which was likely caused by the modifications made to the fixation method after the first set of sample preparation.

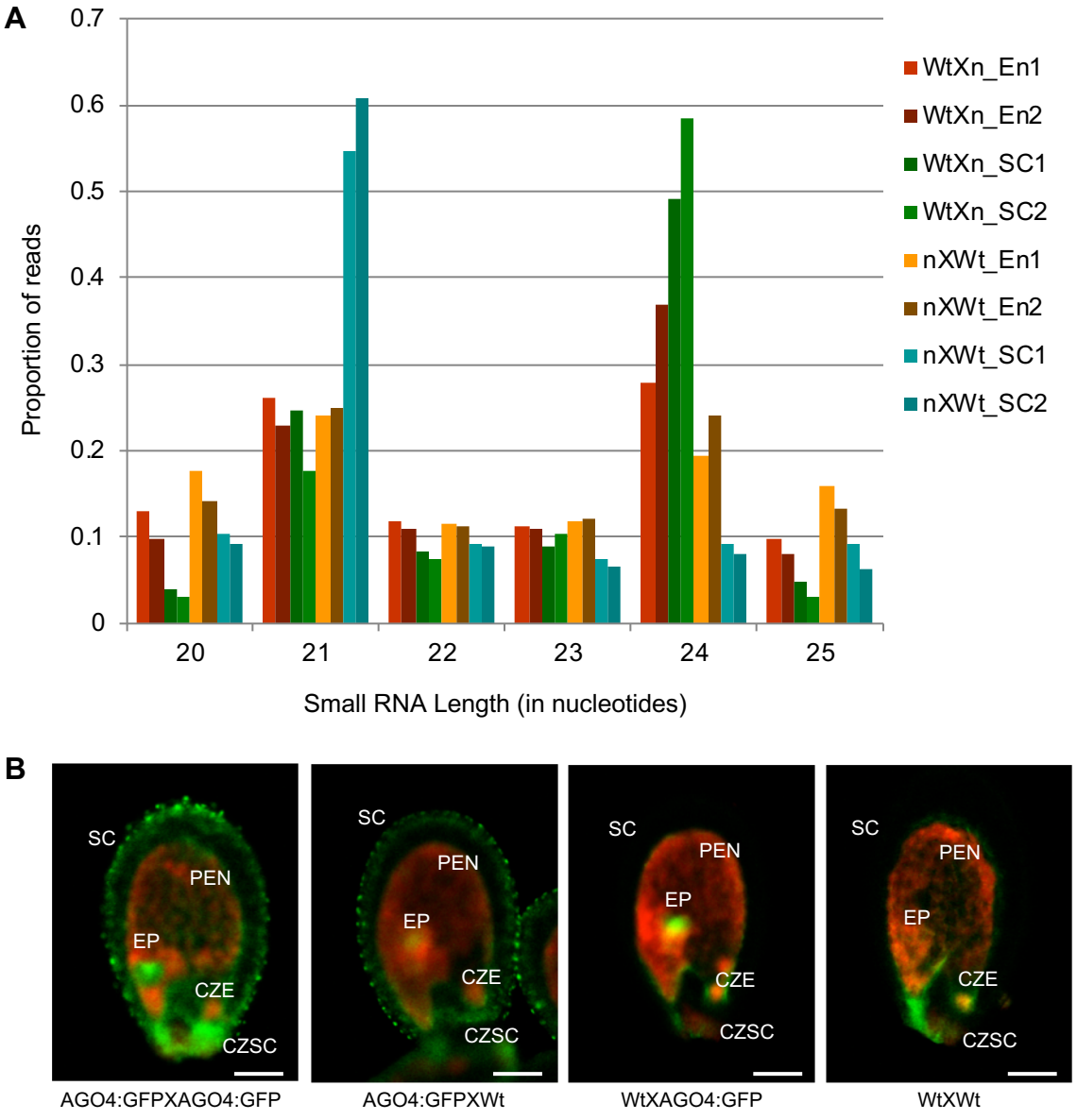


Fig. S4. Small RNA size distributions and *AGO4:GFP:AGO4* localization patterns.
(*A*) Size distribution of 20-25-nt small RNA in found in LCM sample replicates. 21-nt and 24-nt make up the largest categories sequenced. Loss of maternal *NRPD1* results in dramatic reduction in the 24-nt siRNA found in the seed coat, with substantially more modest effects in the endosperm. Wt: Wild type (Col-0); n: *nrpd1*; En: endosperm; SC: seed coat; 1, 2: different replicates.
(*B*) Confocal images of seeds (6 DAP) displaying AGO4:GFP expression patterns in the embryo proper (EP), seed coat (SC), peripheral endosperm (PEN), chalazal endosperm (CZE), and chalazal seed coat (CZSC). GFP in green, chlorophyll autofluorescence in red. Scale bars=100 μm.

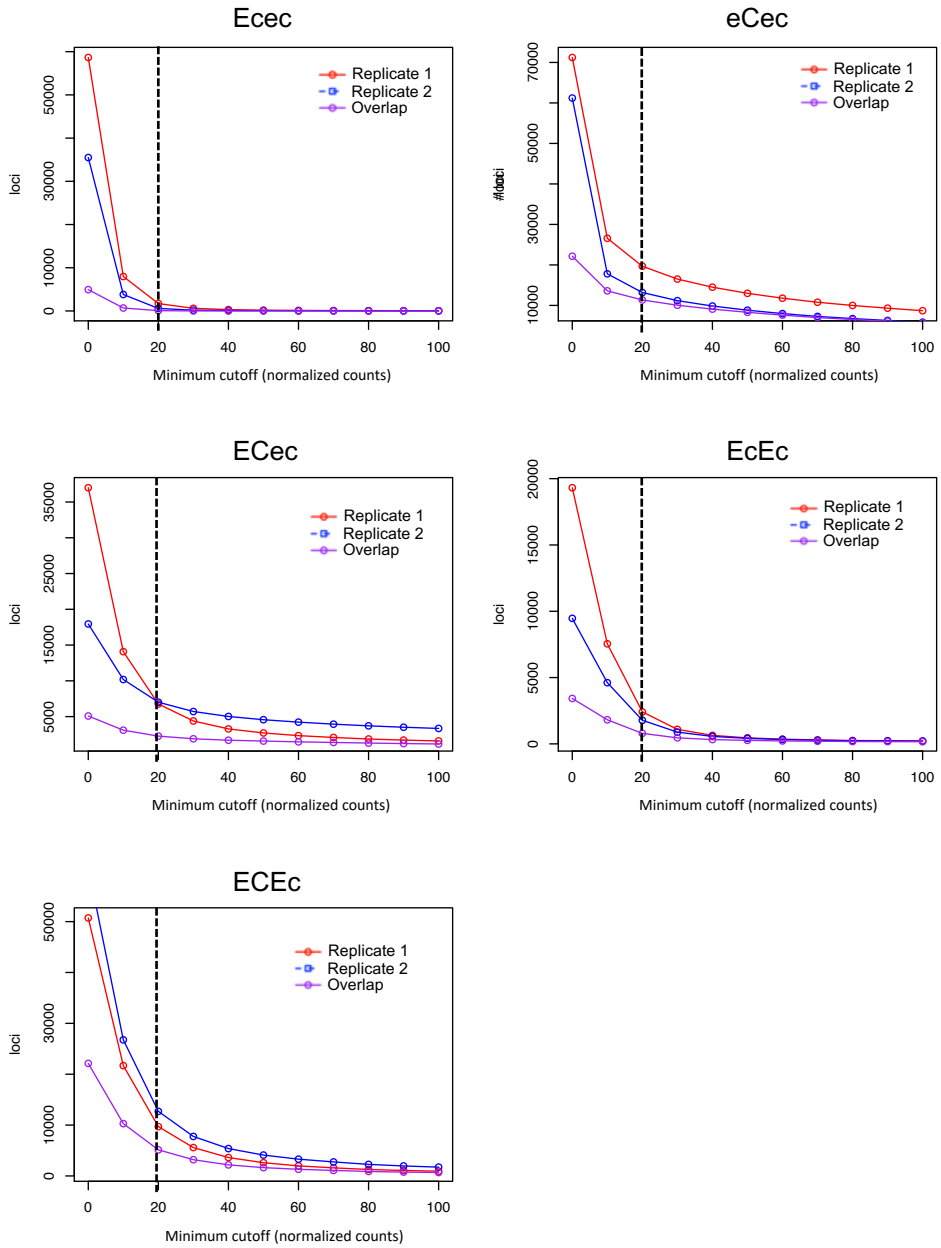


Fig. S5. Minimum expression cutoff values for locus classes.
Plots showing the number of the major loci types identified (Y-axis) in replicate 1 (red lines), replicate 2 (blue lines) and the overlap between the two (purple) using a set of cutoff values (normalized counts, X-axis). The cutoff value is shown by the dashed lines.

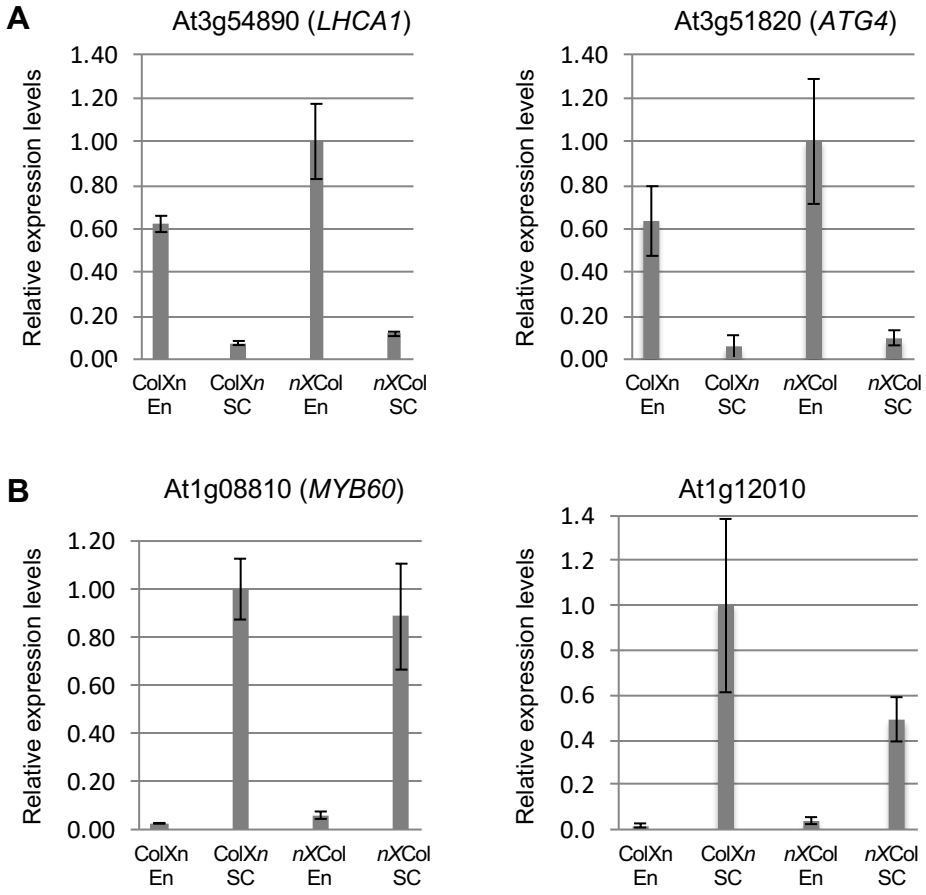


Fig. S6. Specificity of gene expression in LCM samples.
(*A*, *B*) Relative expression levels (determined by qRT-PCR) of the endosperm-specific marker At3g54890 (*LHCA1*) and At3g51820 (*ATG4*) (*A*) and seed coat-specific marker At1g08810 (*MYB60*) and At1g12010 (*B*) in the samples used for small RNA library construction. The markers were selected from the published data (Belmonte et al., 2013). n: *nrpd1*; Col: Columbia; En: endosperm; SC: seed coat. The results indicate low levels of contamination, if any, between the endosperm and seed coat samples.

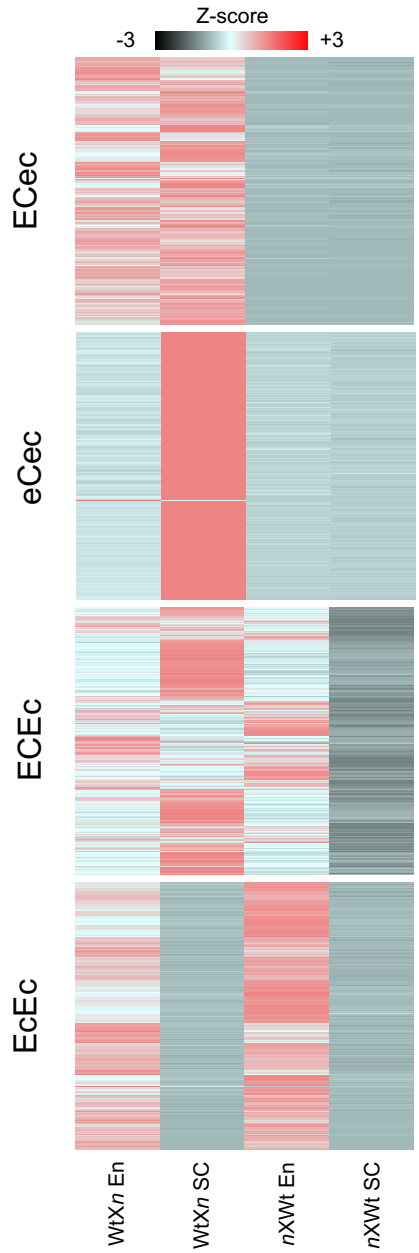
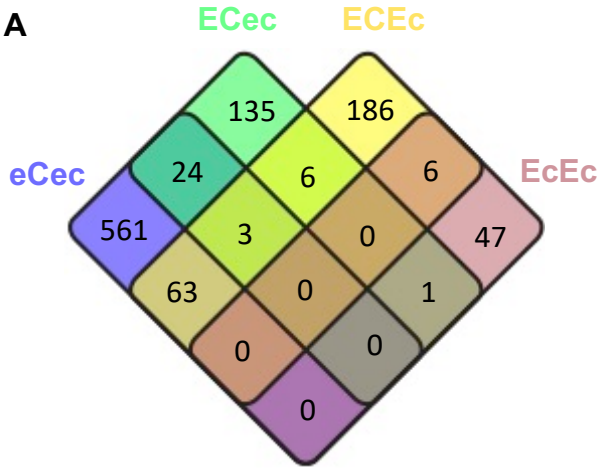


Fig. S7. Heatmap of siRNA expression levels in four groups. Heatmap showing expression profiles for siRNA loci in the endosperm and seed coat LCM samples from the reciprocal crosses. The four groups of siRNA loci show consistent expression profiles according to their classifications in the reciprocal crosses, with an exception of ECEc, which could be related to their expression in the endosperm, which is dependent on either maternal or paternal *NRPD1* .

A



B

siRNA type	No.
Ecec	71
eCec	**6496**
ecEc	216
eceC	11
EcEc	**625**
eCeC	3
ECec	**884**
eCEc	48
EceC	0
ecEC	0
ECEc	**3765**
eCEC	0
EcEC	16
ECeC	3

C

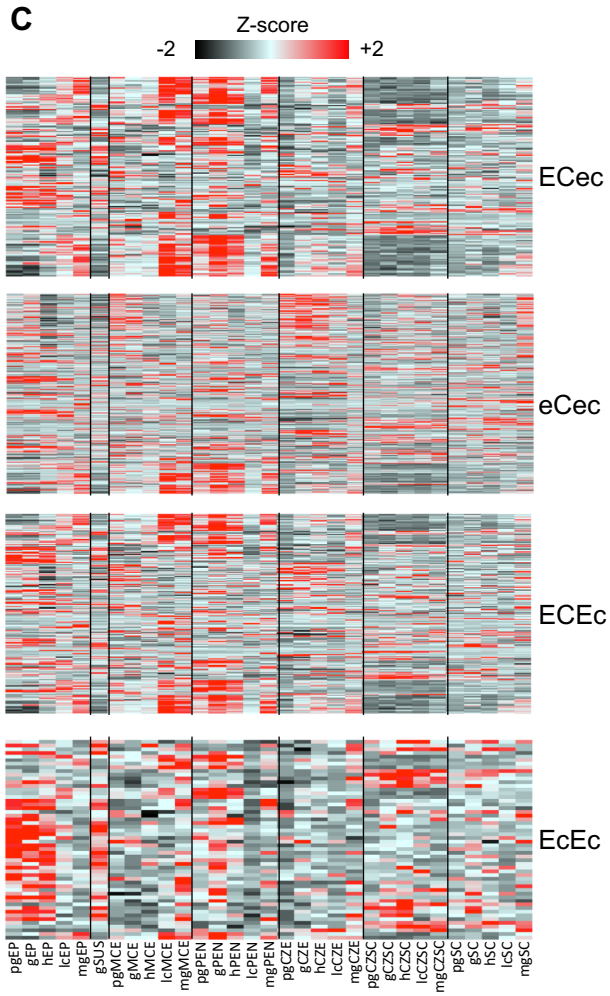


Fig. S8. Four groups of siRNA loci and their patterns in seed sub-regions.
(*A*) Four-way Venn-diagram showing genes with overlapping siRNA loci. Most genes with overlapping siRNA loci have only a single type of overlapping siRNA locus.
(*B*) All possible classes of siRNA enrichment in the endosperm and seed coat samples. The four dominant patterns discussed are highlighted in bold.
(*C*) Heatmap showing expression profile for genes with overlapping siRNA loci at various developmental stages (pg, preglobular; g, globular; h heart; lc, linear cotyledon; mg, mature green) and seed subregions (EP, embryo proper; SUS, suspensor; MCE, micropylar endosperm; PEN, peripheral endosperm; CZE, chalazal endosperm; SC, distal seed coat).

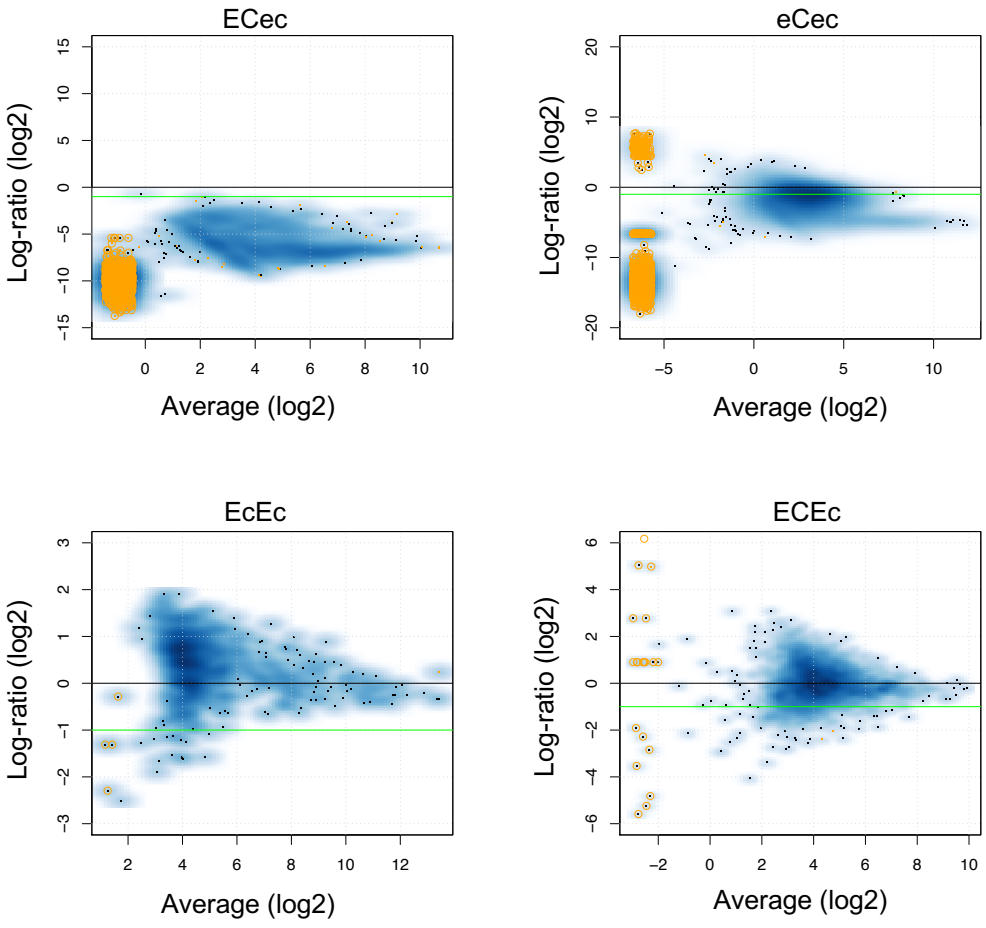


Fig. S9. Fold changes of siRNAs in four locus groups.
MA plots comparing the 24-nt fold changes in the endosperm between WtX*n* and *n*XWt for the four siRNA groups identified in this study. Green line represents an expected 2-fold bias for WtX*n* (where the endosperm has 2 copies of maternal *NRPD1* but a nonfunctional paternal allele) relative to *n*XWt (where the endosperm has 1 paternal *NRPD1* and nonfunctional maternal allele). Loci with the signals in only one of the two crosses shown as yellow smear plot at left of graph. None of the major classes of siRNA in the endosperm adhered to a 2-fold bias between WtX*n* and *n*XWt, showing either unbiased (EcEc, ECEc) or ~100-fold higher than the expected 2-fold difference (ECec and eCec), ruling out a possibility of these loci whose expression depends on the maternal *NRPD1* dosage in the endosperm.

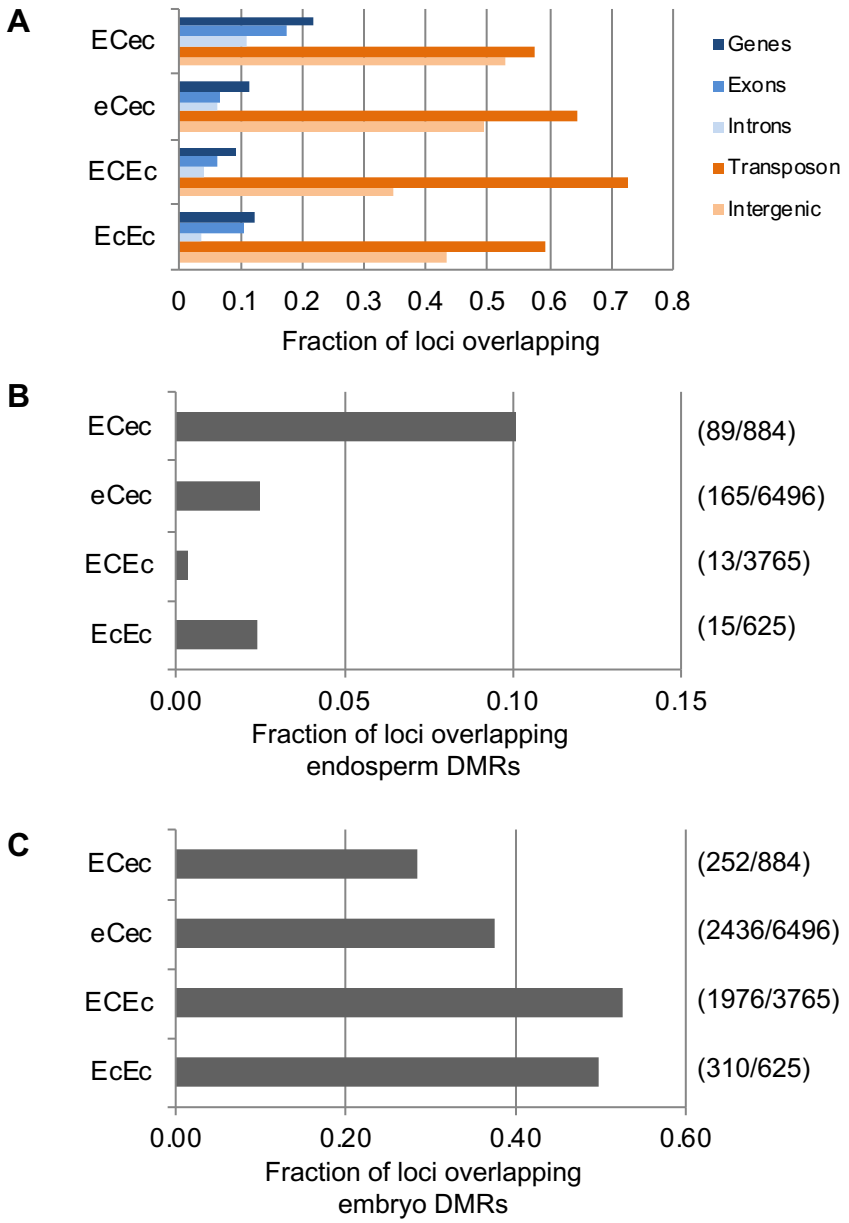


Fig. S10. Relationship between siRNA loci and genomic and epigenomic features.
(*A*) Overlap between siRNA loci and various genomic features. Note that any given siRNA locus (which are 100-bp windows) may overlap with more than one type of features, which makes cumulative fraction larger than 1. ECec loci show the most overlap genes that consist of TEs and TE fragments (Lu et al. 2012, PNAS), while the ECEc loci are most associated with transposons. (*B*, *C*) The fraction of siRNA loci that overlap with differentially CHH methylated regions (DMR) between the embryo and endosperm (Hsieh, 2009). The ECec class has substantially more overlap with regions that are more methylated in the endosperm than the embryo (*B*) compared to the other siRNA classes. No such relationship is seen in the regions with more methylation in the embryo than in the endosperm (*C*). Note that siRNA loci are based on 100bp windows, therefore overlapping with multiple genomic features is possible.

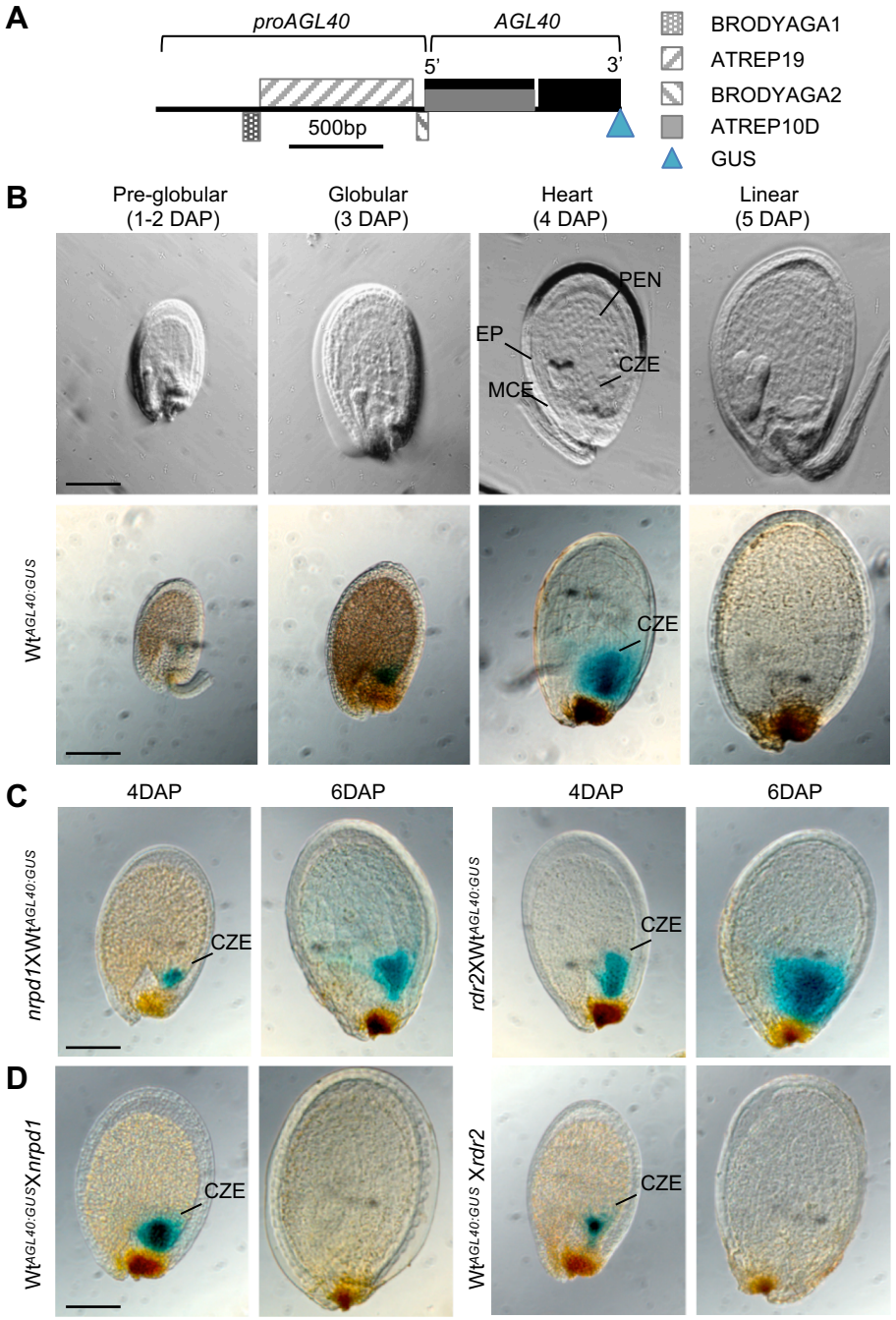
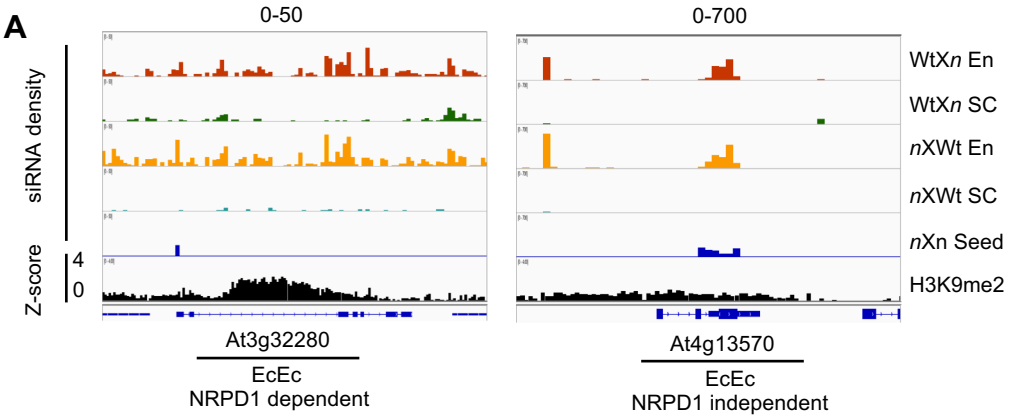


Fig. S11: Spatial-temporal *AGL40* expression is dependent on maternal *NRPD1* or *RDR2*.
(*A*) The genomic region of *AGL40* (AT4G36590) is surrounded by four short TEs, three in the promoter and one in the coding region. A GUS gene was fused in frame with the 3' end of the coding region driven by the native promoter.
(*B*) Cleared (upper) and GUS-stained seed images (lower) of the seeds in the wild-type transgenic Wt$^{AGL40:GUS}$ lines at different developmental stages (silenced at 5 DAP).
(*C*, *D*) GUS-stained seed images at 4 or 6 DAP in the reciprocal crosses of *nrpd1*XWt$^{AGL40:GUS}$ (*C*) and Wt$^{AGL40:GUS}$X*nrpd1* (*D*) in the left and in reciprocal crosses of *rdr2*$^{AGL40:GUS}$XWt (*C*) and Wt$^{AGL40:GUS}$X*rdr2* (*D*) in the right, respectively (n = 3 x100 seeds per cross). Note activation of *AGL40:GUS* at 6 DAP. Scale bars = 0.1 mm for all images. Locations of embryo (EP), chalazal endosperm (CZE), micropylar endosperm (MCE), peripheral endosperm (PEN) are marked at the heart stage. Scale bars = 0.1 mm for all images.

A



B

Genes with NRPD-independent EcEc loci

AGI ID	Type	Description
AT1G60840	pre_trna	pre-tRNA
AT1G78476	protein_coding	
AT1G78478	**mirna**	**MIR833a; miRNA**
AT2G23834	protein_coding	
AT2G24340	protein_coding	sequence-specific DNA binding transcription factors
AT2G24513	protein_coding	
AT2G25011	**mirna**	**MIR836a; miRNA**
AT4G08990	protein_coding	DNA (cytosine-5-)-methyltransferase family protein
AT4G13494	**mirna**	**MIR863a; miRNA**
AT4G13495	other_rna	other RNA
AT4G13564	**mirna**	**MIR841a; miRNA**
AT4G13570	protein_coding	histone H2A 4
AT4G14140	protein_coding	DNA methyltransferase 2
AT4G21050	protein_coding	Dof-type zinc finger domain-containing protein
AT5G13890	protein_coding	Family of unknown function (DUF716)
AT5G39693	**mirna**	**MIR869a; miRNA**
AT5G41765	protein_coding	DNA-binding storekeeper protein-related transcriptional regulator

C

GO term	Ontology	Description	Number in input list	Number in BG/Ref	p-value	FDR
GO:0090116	P	C-5 methylation of cytosine	2	4	4.50E-06	0.0004
GO:0032776	P	DNA methylation on cytosine	2	9	1.60E-05	0.00073

D

Locus Type	*NRPD1*-independent loci	(%)	*NRPD1*-dependent loci	(%)	Total
eCec	33	0.5	6463	99.5	6496
ECec	0	0.0	884	100.0	884
ECEc	4	0.1	3761	99.9	3765
EcEc	56	9.0	569	91.0	625

Fig. S12. *NRPD1*-dependency of EcEc loci.
(*A*) siRNA profiles from whole seeds derived from the *nrpd1* X *nrpd1* (*n*X*n*) cross were used to determine which EcEc loci that were *NRPD1*-dependent and which loci that were *NRPD1*-independent. The majority of loci (e.g., At3g32280) were *NRPD1*-dependent and had little or no siRNAs in the *n*X*n* samples (blue track, left), while some loci (e.g., At4g13570) were independent of *NRPD1* (blue track, right). Y-axis scale for each browser view is shown above the respective panel. Wt: wild-type; En: endpsperm; SC: seed coat. A track of H3K9me2 was also included for the analysis.
(*B*) List of genes with overlapping *NRPD1*-independent EcEc loci. miRNA loci bolded.
(*C*) Gene ontology enrichment for *NRPD1*-indepednent EcEc loci.
(*D*) Number and percentage of *NRPD1*-independent loci in four groups of siRNAs.

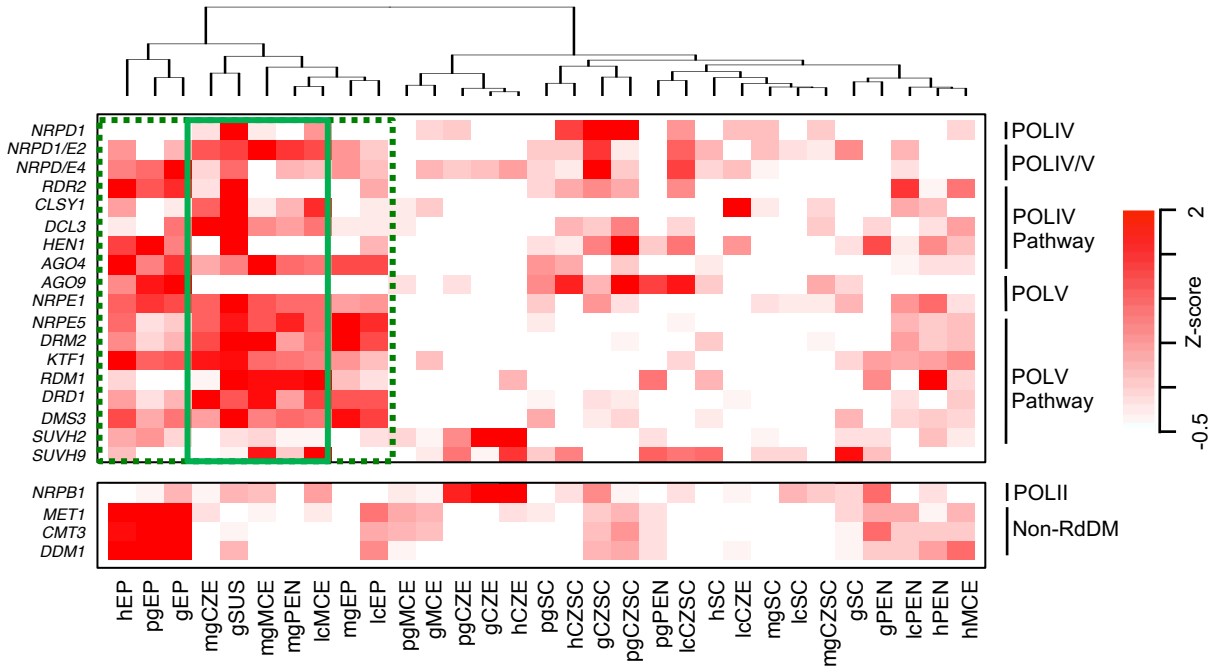


Fig. S13. Heatmap showing spatial-temporal transcript levels of RdDM pathway (top) and non-RdDM pathway (bottom) genes.

Samples represent distinct developmental stages (p, pre-globular; g, globular; h, heart; lc, linear cotyledon; mg, mature green) and sub-regions (EP, embryo proper; SUS, suspensor; MCE, micropylar endosperm; PEN, peripheral endosperm; CZE, chalazal endosperm; CZSC, chalazal seed coat; SC, distal seed coat). Boxes (in green) with dashed and solid lines highlight clusters of samples with the highest expression levels of RdDM pathway genes in the embryo at all stages and in the endosperm at later developmental stages, respectively.

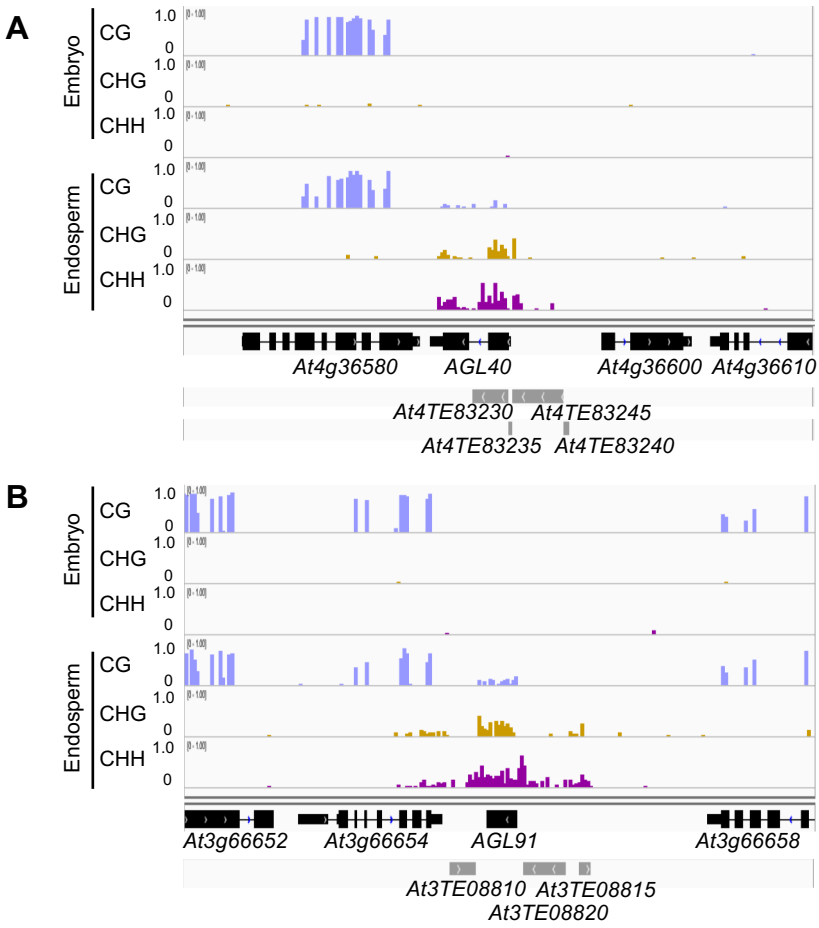


Fig. S14. Methylation of *AGL40* and *AGL91* loci.
(*A*, *B*) Fractional methylation of 10 kb genomic region centered on *AGL40* (*A*) and *AGL91* (*B*) using the published data (Hsieh et al., 2009). In the endosperm a substantial amount of methylation was present in the CG (lavender), CHG (gold), and CHH (purple) contexts, while little or no CHG and CHH methylation was found in the embryo. Genes shown in black and transposable elements in grey.

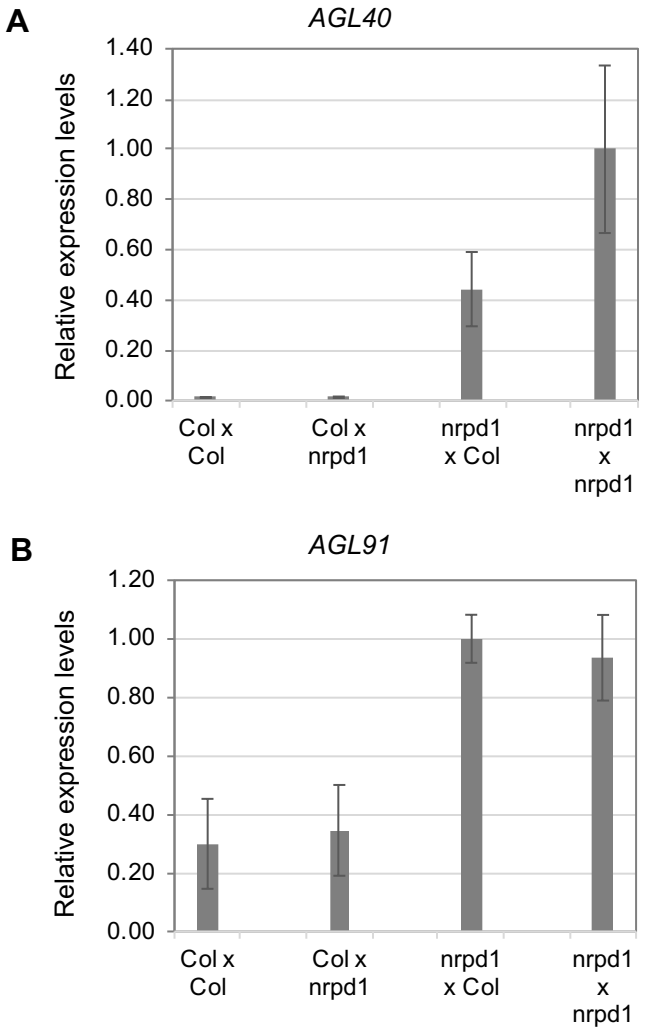


Fig. S15. Increased expression levels of endogenous *AGL91* and *AGL40* when the maternal *NRPD1* is mutated.
(*A*, *B*) Relative expression levels of endogenous *AGL4*0 (*A*) and *AGL91* (*B*) were elevated in the seeds of both *nrpd1*XCol and *nrpd1*X*nrpd1* crosses relative to the cross ColXCol or ColX*nrpd1*.

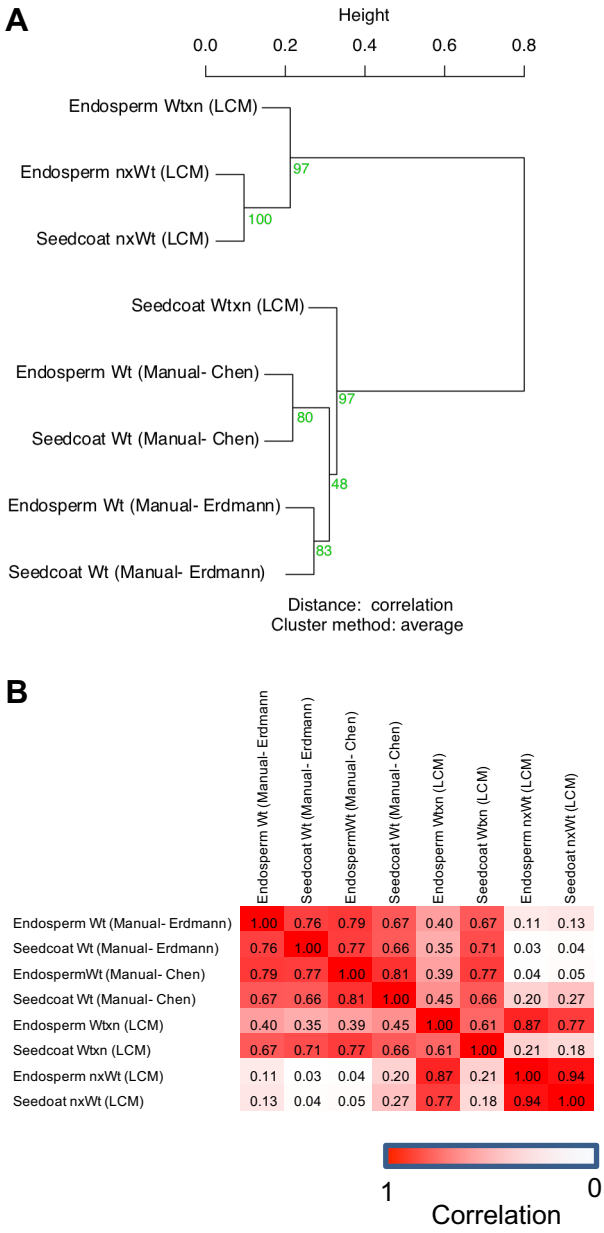


Fig. S16. Comparison of small RNA profiles in the LCM and manually dissected samples. (*A*) Dendrogram showing the relationship between the 24-nt sRNA profiles (100 bp windows) for LCM and manually dissected seed tissues. Bootstrap values are shown in green**.** Clustering reveals two major groups: the LCM group (top) consisting of the endosperm sample that is free of seed coat siRNAs in the WtX*n* cross and or absence of maternal siRNAs in the *n*XWt cross because of the *nrpd1* used as the maternal parental in the cross; the manually dissected group (bottom) representing some levels of contamination between endosperm and seed coat samples. Note the overlap between two two studies is small. (*B*) Heatmap of correlation coefficients of siRNAs between LCM and manually dissected seed samples.