Table of contents

Supplemental Methods

Tissue collection

Rice (*Kitaake* variety) was grown in soil in greenhouses under natural light. Plants were irrigated with deionized water twice a week and supplemented with fertilized water every other week. Gametes were isolated as described (Anderson et al. 2013; Li et al. 2019). Briefly, ovaries were dissected from pre-anthesis flowers. A transverse cut was made at the middle region of the ovary. The lower part of the cut ovary was gently pushed by an acupuncture needle under a phase inverted microscope. Once the egg cell floated out from the ovary incision, it was captured by a fine capillary in a volume of ~1 µl and frozen in liquid nitrogen. For small RNA, 35 – 50 cells were used for each biological replicate. For whole genome bisulfite sequencing (WGBS), 100 egg cells were used for each biological replicate. Six biological replicates were collected for egg cells and two for sperm cells. For sperm cell small RNA and WGBS libraries, around 50 panicles with mature flowers were used for each biological replicate. Eight biological replicates were collected for small RNA and one for WGBS. Ovaries were dissected from pre-anthesis flowers. For both small RNA and WGBS, five ovaries were pooled to make each biological replicate. Three biological replicates were collected for small RNA and six for WGBS. Bracts (lemma and palea) were collected from pre-anthesis flowers. For WGBS, three pairs of bracts were pooled to make each biological replicate and six biological replicates were collected. Seedling shoot segments were collected from 7-day-old water-germinated rice seeds for small RNA libraries, and one seedling was used for each of the four biological replicates. Individual mature endosperms and embryos were used for each biological replicate (three for each genotype). Dry seeds were soaked in 6% NaOH in water at 57°C for 8 min, and pericarps were removed with forceps. The endosperm and embryo were separated then ground to a powder with a mini pestle in a 2-ml Eppendorf tube. DNA was extracted from ovary, bract, endosperm, and embryo with a DNeasy Plant Mini Kit (Qiagen, 69104). Embryos and endosperm mutant for *drm2* were genotyped using the primers DRM2-Ri (TCTCACTACAAAGGCACCATAAAG) and DRM2-48F (CGAGGAGGAGGATGATACTAT) for the 48-bp deletion allele and primers DRM2-Ri and DRM2-52F (CGAGGAGGAGGATGATATG) for the allele 52-bp deletion allele.

CRISPR/Cas9 mutagenesis of *drm2* and genotyping

CRISPR-Cas9 was used to generate targeted mutations in the rice DRM2 gene (MSU: *LOC_Os03g02010*). Single-guide RNA (sgRNA) sequences, GGAGGAGGATGATACTAATT and GACAGGACTCCTCACTCTGA, were designed using the web tool https://www.genome.arizona.edu/crispr/ (Xie et al. 2014). CRISPR construct assembly was performed as described previously (Khanday et al. 2019). Rice transformation was performed at the UC Davis Plant Transformation Facility. One transgenic line carrying one in-frame 48-bp deletion and one frame-shift 52-bp deletion was selected for further analysis. Since homozygous rice *drm2* mutants were sterile [also previously reported (Moritoh et al. 2012)], we maintained the *drm2* mutation in a segregating population with the in-frame 48-bp deletion. The 48-bp allele was functional, as the plants carrying one or both 48-bp alleles were phenotypically indistinguishable from wild-type. Genotyping was performed using two forward primers, F48 (CGAGGAGGAGGATGATACTAT) and F52 (CGAGGAGGAGGATGATAT). Each specifically amplifies either the 48-bp or 52-bp alleles, with one reverse primer (TCTCACTACAAAGGCACCATAAAG). To genotype each sample, two separate PCR reactions were performed: F48 + R and F52 + R, at 59°C annealing temperature, 30 PCR cycles. Amplicon sizes were ~500-bp.

RNA extraction and small RNA library construction

RNA extractions were performed using Ambion RNaqueous Total RNA kit (AM1931), including an on-column DNase treatment using Qiagen DNase (79254). Total RNA was run on a Bioanalyzer to check for RNA integrity, using the eukaryotic total RNA-pico program. RNA input was around 30 ng total RNA for egg cells, 1 ng for sperm cells, 50 ng for ovaries, and 20 ng for seedlings. Small RNA libraries were made using the NEXTflex Small RNA-seq kit v3 (NOVA-5132-05), with the following modifications. Since RNA input was low, a 1/4 dilution of adapters was used. The 3' adapter ligation step was performed at 20°C overnight. Sperm libraries were amplified with 25 PCR cycles. All other libaries were amplified with 15 – 20 cycles, except one

of the four seedling replicates was amplified with 25 cycles. After amplification, libraries were run on a Bioanalyzer DNA High Sensitivity Assay. Libraries with a 130-bp peak (adapter dimer peak) constituting less than 10% of the 150-bp peak (expected small RNA peak) were ran on a 10% TBE-Acrylamide gel (100 V for 1 hr). Gels were stained and the area around 150 bp was excised and purified according to recommendations of the NEXTflex Small RNA-seq kit.

Mock egg isolations and qPCR quantification

Ovaries were dissected as during egg cell isolation. Rather than collecting an egg cell, about 1 µL of cell-free solution was collected into a microcentrifuge tube and immediately frozen in liquid nitrogen. Thirty collections were combined as a single replicate, and two independent replicates were collected. RNA extraction and library construction were performed as described above. A strong 150-bp band could be seen on a Bioanalyzer gel for a positive control ovary library, but no band for negative water control or mock samples (**Supplemental Fig. S2B**), indicating that the lack of a band is not due to failed library preparation, but lack of sufficient input RNA. To produce a DNA standard for qPCR absolute quantification, the following primers were used: P5_F (AATGATACGGCGACCACCGACATGACATTGACTATAAGGATGACG) and P7_R (CAAGCAGAAGACGGCATACGAGATCGAGGCCGATGCTATACTTT) using a plasmid template (Khanday et al. 2018). The target amplicon was 153bp and 49% CG content. The standard was PCR amplified, gel purified, and quantified using a Nanodrop spectrophotometer. qPCR was performed using SYBR Green Master Mix (Biorad 1725270), with Illumina P5 and P7 universal primers, AATGATACGGCGACCACCGA and CAAGCAGAAGACGGCATACGAGAT, with a 60°C annealing temperature, 30 second elongation time and 35 cycles. Serial dilutions of the standard (10^{-1} to 10^{-7}, ten-fold dilution each step, two technical reps each) were used to fit the standard curve (**Supplemental Fig. S2C**). A 1/10 dilution of each library (three technical reps each) was used as templates for qPCR in the same run. The number of molecules in each library was calculated using the standard curve (**Supplemental Fig. S2D**).

Genome annotations

The Os-Nipponbare-Reference-IRGSP-1.0 reference genome was used for all analyses (Kawahara et al. 2013). MSU7 Rice gene annotations were extracted from the all.gff genome annotation file downloaded from http://rice.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir/. Genes that were flagged as transposons were removed, leaving a set of 39,953 genes. Transposons were annotated using RepeatMasker version 4.0.05 (http://www.repeatmasker.org/), parameters as follows: "-gff -species rice -s -pa 8". miRNA annotations were downloaded from miRBase version 22 (Kozomara et al. 2019). To identify locations of the tandem repeat *CentO*, a consensus sequence (Zhang et al. 2013) was aligned to the genome using using Bowtie2 version 2.3.4.1 (Langmead and Salzberg 2012) parameters "--local -a -f". All alignments separated by 100 bp or fewer were merged using BEDTools merge (Quinlan and Hall 2010). Locations of 5S rRNA and tRNA repeats were identified in the same way as *CentO* but using the GenBank reference sequence KM036285.1 for 5S rRNA and a set of tRNA sequences from The tRNAscan-SE Genomic tRNA Database for tRNAs (http://gtrnadb.ucsc.edu/GtRNAdb2/genomes/eukaryota/Osati/Osati-tRNAs.fa.) (Chan and Lowe 2016). For NOR annotation, an 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S ribosomal RNA gene, internal transcribed spacer 2, and 26S ribosomal RNA gene complete sequence from GenBank (KM036285.1) was aligned to the genome with Bowtie2 version 2.3.4.1, parameters "--local --ma 1 --mp 24,8 --rdg 20,48 --rfg 20,48 -a f". All alignments separated by 100 bp or fewer were merged using BEDTools.

Small RNA sequencing analysis

Small RNA-seq reads were quality filtered and trimmed of adapters using cutadapt (Martin 2011), parameters "-q 20 -a TGGAATTCTCGGGTGCCAAGG -e .05 -O 5 --discard-untrimmed -m 28 -M 33". PCR duplicates were then removed using PRINSEQ, parameters "prinseq-lite.pl -fastq -out_format 3 -out_good -derep 1" (Schmieder and Edwards 2011). The four random nucleotides at each end were then removed using cutadapt "-u 4" followed by cutadapt "-u -4". Previously published small RNA libraries that did not include 4 random nucleotides at each end were processed similarly, but without removal of PCR duplicates. Reads were aligned to

the genome with BWA-backtrack (version 0.7.15) (Li and Durbin 2009), parameters "aln -t 8 -l 10." A single mapped position was kept per input read, regardless of the possibility of mapping to multiple locations. Except where indicated otherwise, multi-mapping reads were included in all analyses. Locations of 21-nt and 24-nt phasiRNA loci were identified using PHASIS version 3.3 (https://www.biorxiv.org/content/10.1101/158832v1). The phasdetect module was run on small RNA reads of each length from 20 to 25-nt separately, followed by the phasemerge module, parameters "-mode merge -pval 1e-5" with each length. In addition to the small RNA reads produced in this study, reads from the spikelet developmental series (Fei et al. 2016), and anther/ovary developmental series (Li et al. 2017) were included for identifying phasiRNA loci. Only 21 and 24-nt lengths produced detectable phasiRNA loci (2364 21-nt phasiRNA loci and 68 24-nt phasiRNA loci, **Supplemental Table S4**). The complete set of read alignments was compared with miRNA, phasiRNA, tRNA, 5S rRNA, and NOR RNA loci in the genome, and all reads that aligned by at least 90% with any of these was categorized as such using BEDTools intersect. All other reads were categorized as siRNA reads and used for subsequent siRNA analyses (**Supplemental Table S5**). The uniquely mapping subset of siRNAs was defined by having MAPQ values of at least 20 using SAMtools (Li et al. 2009). For analysis of overlaps of siRNAs and *Gypsy* retrotransposons, the *CentO* centromeric tandem repeat, Terminal Inverted Repeat (TIR) DNA transposons, and 24-nt siRNA loci, only siRNAs that overlapped by at least 50% of their lengths were counted. *CACTA* elements were excluded from the TIR DNA transposons. Whole-genome small RNA heatmaps were made on 50-kb intervals using IGVtools (Thorvaldsdóttir et al. 2013). For better visualization of midrange values, heatmap intensity was maxed out at 1.25X coverage (per 10 million 24-nt siRNAs). To identify representative 24-nt siRNA loci, reads alignments were subsampled then combined from each sperm cell sample, from each egg cell sample, and from each seedling shoot sample to get as equal a representation as possible from each sample and a final combined number of 2 million in each using SAMtools view –s followed by SAMtools merge. The genome was divided into non-overlapping 100-bp loci, and siRNAs were counted for each locus using BEDTools coverage. All loci that had at least three mapping 24-nt siRNAs that spanned at least a third (34 bp) of the 100 bp were categorized as 24-nt siRNA loci. siRNAs that mapped to the intersection of two 100-bp loci contributed toward the siRNA count for both loci. Adjacent 100-bp loci that both qualified as 24-nt siRNA loci were not merged, but instead counted as individual 24-nt siRNA loci. 5' nucleotide frequencies were calculated with FastQC, version 0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

miRNA expression analyses

miRNA expression data were organized into a matrix, with each row as an individual miRNA and each column as a library (**Supplemental Dataset 1**). R package EdgeR was used to analyze miRNA expression (McCarthy et al. 2012). Individual miRNA counts were normalized by total mapped small RNAs and filtered for >1 counts per million reads (CPM) in at least three libraries. Libraries were then further normalized by the TMM method (Robinson and Oshlack 2010), as recommended by the EdgeR package. Differential expression analyses were performed under log2FC > 1 and FDR < 0.05 cutoffs. Differential expressing miRNAs were visualized under counts per million miRNAs. Principal component analyses were performed using log-transformed CPM values. Clustering analyses (**Supplemental Fig S1** and **Supplemental Table S3**) were performed using hierarchical clustering, and assignment of miRNA into clusters was done using the Dynamic Tree Cut R package (Langfelder et al. 2008).

Preparation of WGBS libraries

PBAT libraries were prepared using Pico Methyl-Seq Library Prep Kits (Zymo D5456). Sperm and egg cell isolates of approximately 100 cells each were diluted in 200 µl of 10 mM Tris, pH 8 in a 1.5-ml tube, then centrifuged for 10 minutes at 16,000 x G at 4°C. The supernatant was removed except for 9 µl at the bottom of the tube. The 9 µl was pipetted up and down 10 times, then transferred to a 0.2-ml tube. 10 µl M-Digestion Buffer and 1 ml Proteinase K (Zymo D3001-2-5) were added and incubated for 20 minutes at 50°C. For pollen vegetative cells, 12 µl was retained after the initial dilution in Tris buffer, and 13 µl Zymo M-Digestion Buffer and 1 ml Zymo Proteinase K (20 mg/mL) were added. Following incubation for 20 minutes at 50°C, pollen vegetative cells were centrifuged again for 5 minutes at 16,873 x G, then 20 µl of supernatant transferred to a new tube. Subsequent steps were as directed in the Pico Methyl-Seq Library Prep kit protocol, version 1.2.0,

with a 16-minute incubation in L-Desulphonation buffer and 5:1 ratios of DNA binding buffer in DNA purification steps. Conventional WGBS libraries from endosperm, embryo, ovary, and bract were prepared using the methylC-seq method (Urich et al. 2015).

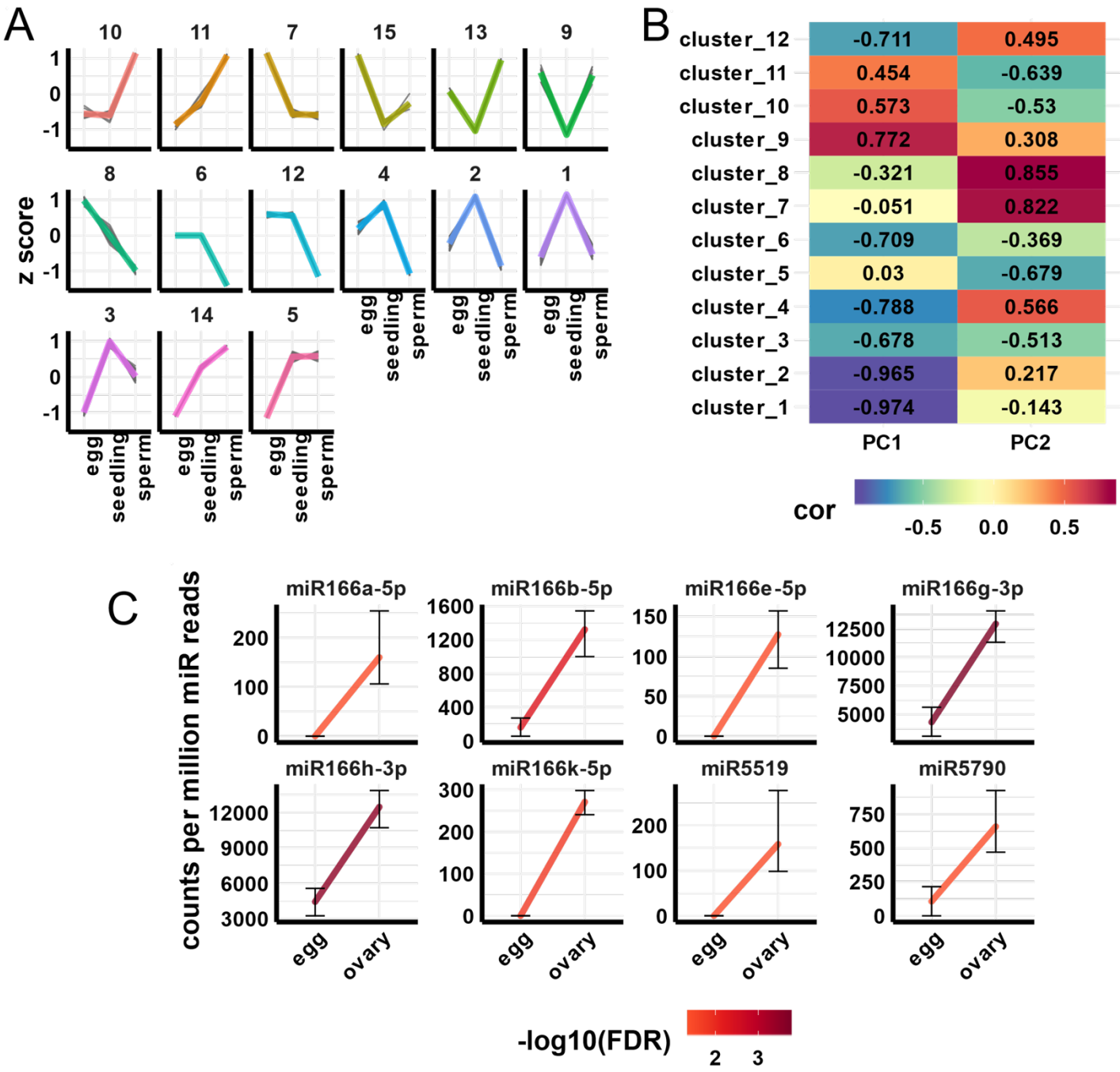
WGBS analysis

MethylC-seq and PBAT reads were quality filtered and trimmed of adapters using cutadapt (Martin 2011), parameters as follows: "-q 20 -a AGATCGGAAGAGC -e .1 -O 1 -m 50". PBAT reads were aligned to the genome with BS-Seeker2 (version 2.1.5) (Guo et al. 2013) with parameters as follows: "--aligner=bowtie2 --bt2--end-to-end -m 1 -t Y -s 5 -e 100" methylC-seq reads were aligned similarly, but BS-Seeker2 parameters were modified to "-m 1 --aligner=bowtie2". For all reads except paired-end, previously published reads (Tan et al. 2016), PCR duplicates were removed prior to alignment with the BS-Seeker2 FilterReads.py module. DMRs were identified with CGmapTools version 0.1.1 (Guo et al. 2018). Biological replicates were first merged using the CGmapTools mergelist tosingle module, then the set of cytosines covered by reads in both samples were identified using the CGmapTools intersect module. Methylation comparisons were made using the CGmapTools dmr module, parameters "-c 3 -C 50 -s 100 -S 100 -n 5". DMRs were selected based on four criteria: at least five measured cytosines in the region; a P-value less than 0.001; absolute difference in methylation proportion of greater than 0.25 (wild-type value minus *drm2* mutant value), and a relative difference in methylation proportion of less than 0.3 (wild-type value divided by *drm2* mutant value).

mRNA expression analysis

Previously published mRNA reads (Anderson et al. 2013; Anderson et al. 2017) were aligned to the genome using Tophat, version 2.0.13 (Kim et al. 2013), parameters as follows: "--read-realign-edit-dist=0 --min-intron-length=15 --max-intron-length=20000 --max-multihits=1 --microexon-search --library-type=fr-unstranded --b2-very-sensitive". For uniquely mapping reads only, "--prefilter-multihits" was also included. The number of reads that overlapped with genomic features was counted using BEDTools intersect, requiring that half of each read's mapped length overlapped with a feature using the "-f .5" parameter to be counted.

All R-scripts for statistical analyses and data visualization can be found at https://github.com/cxli233/gamete_smRNA_revision and as Supplemental Code.

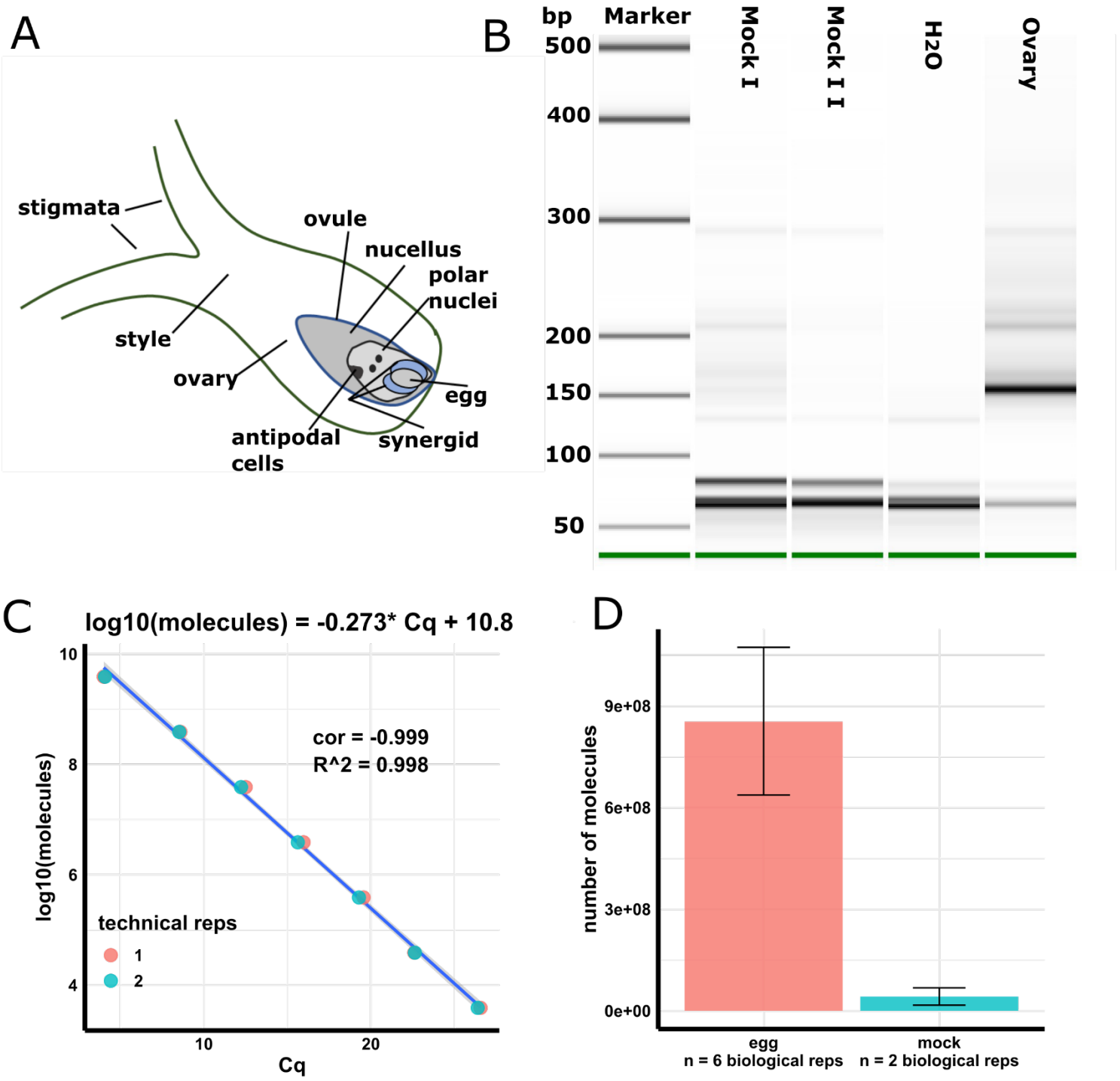


Supplemental Figure 1: Differential miRNA expression and clustering

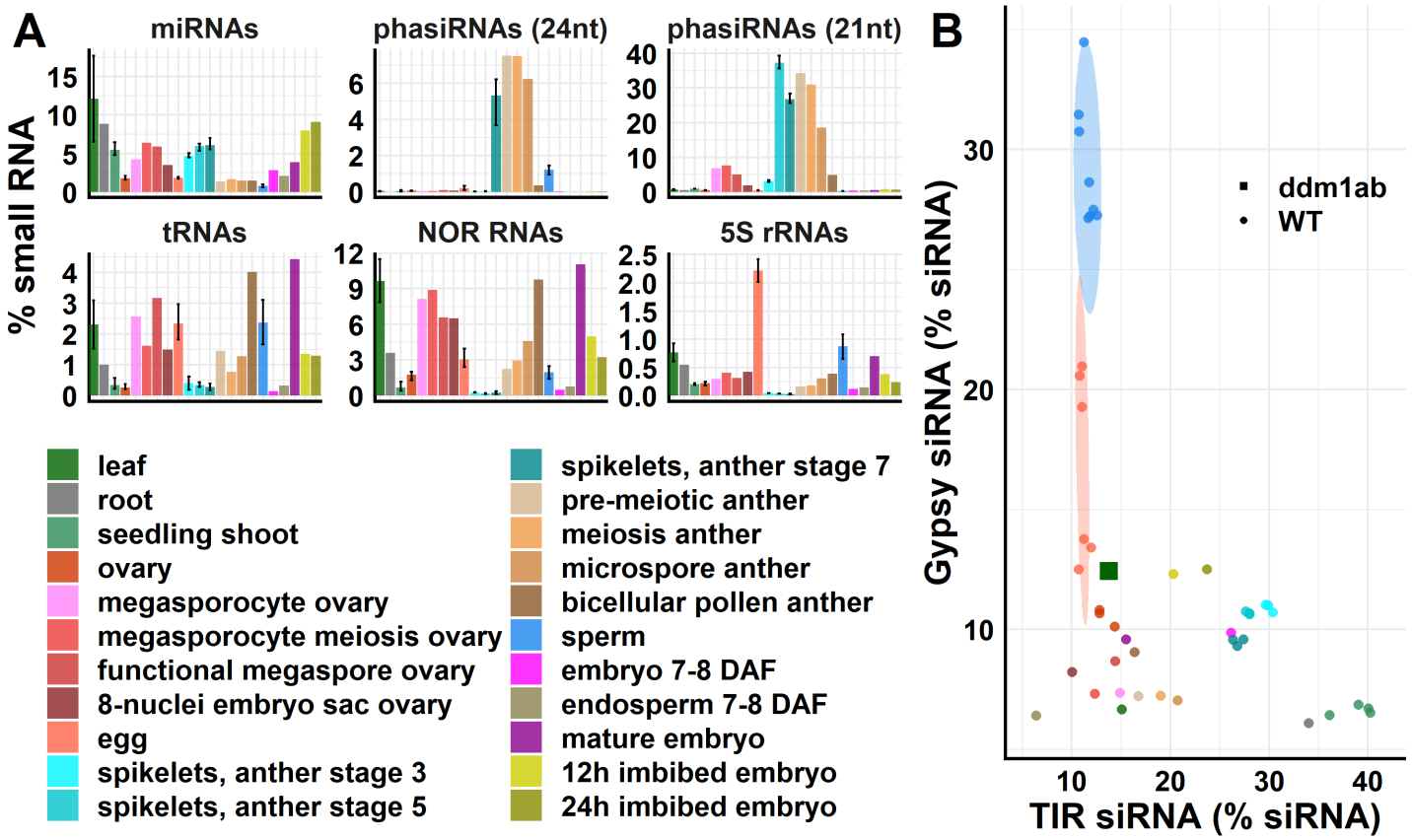
A: Clusters of miRNAs based on expression in egg cell, sperm cell, and seedling shoot. miRNAs were clustered by hierarchical clustering of z-scores. The z-score of each miRNA was calculated relative to the mean and standard deviation of each miRNA across tissues. Colored lines are the average of each cluster. Grey lines are individual miRNAs. See also **Supplemental Table S3**.

B: "Cor" indicates correlation of miRNA clusters to each principal component (PC) axis (see also **Fig 1A**). Color and text reflect Pearson's correlation coefficient.

C: Representative differentially expressed miRNAs between egg cell and ovary. Differential expression was determined by 2-fold change and FDR (false discovery rate) < 0.05 cutoffs. Y-values are relative to the total number of miRNA reads in each sample. Error bars are 95% confidence intervals. Color of line reflects -log10 of FDR.



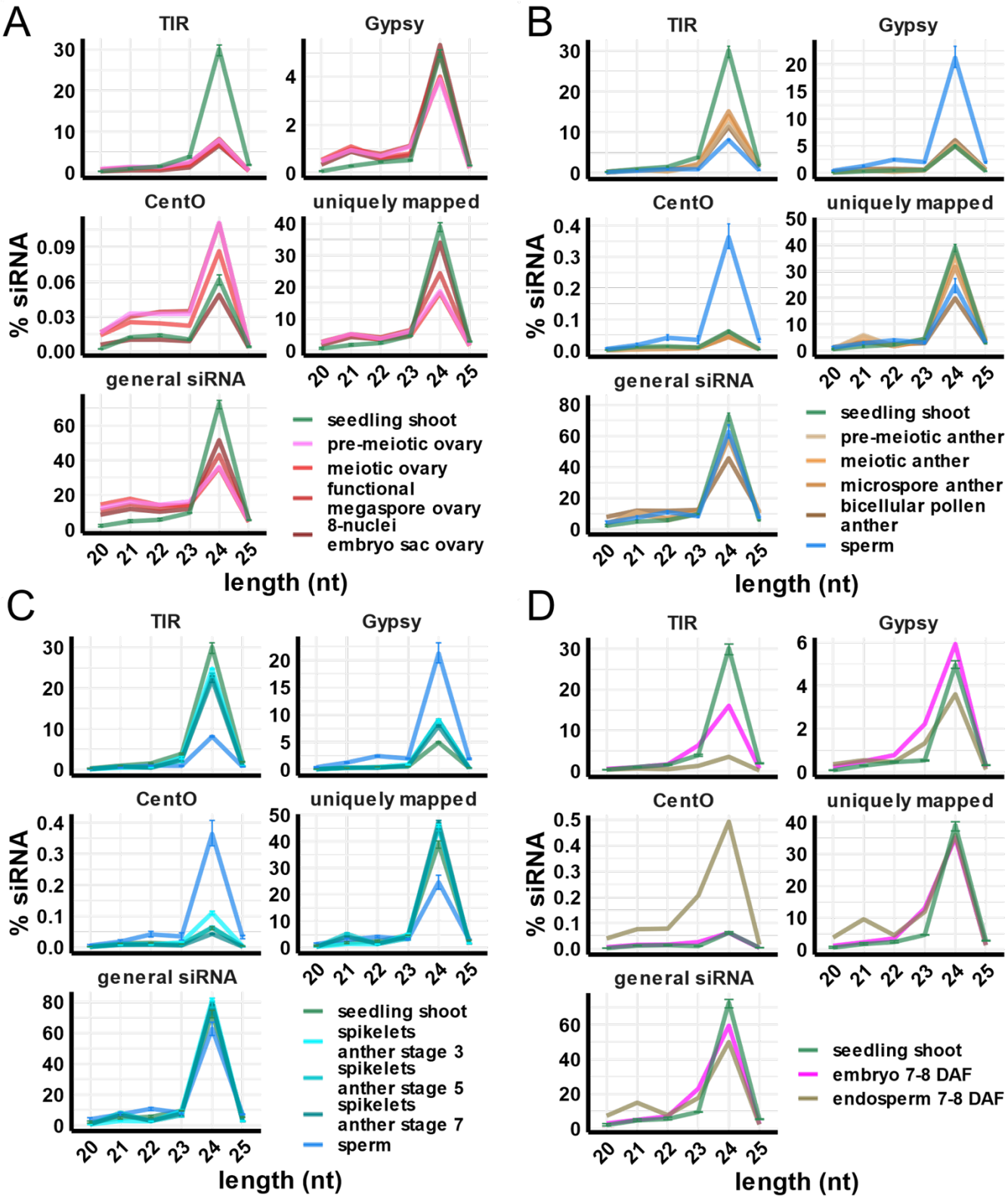
Supplemental Figure 2: Mock egg cell isolations and qPCR quantification of small RNA libraries
A: Scheme of rice ovary, redrawn from Li et al (2019).
B: Bioanalyzer gel image of mock libraries. Expected library product size is about 150 bp.
C: Standard curve for qPCR quantification.
D: Number of molecules in PCR-amplified egg cell libraries and mock libraries.



Supplemental Figure 3: Small RNA compositions across tissues

A: small RNA compositions across sample types, as in **Fig 1C**. Y-axis values are relative to the total number of reads that mapped to the genome. For samples with two or more biological replicates, error bars are 95% confidence intervals.

B: Scatter plot showing TIR siRNAs on X-axis and Gypsy siRNAs on Y-axis, each measured as percent of the total siRNAs in each library. Egg and sperm cell replicates formed unique clusters, highlighted by pink and blue eclipse, respectively. Color code as in **A**. Published data sources: leaf (Tan et al. 2018), root (Shin et al. 2018), ovary and anther (Li et al. 2017); spikelet (Fei et al. 2016), 7-8 DAF embryo and endosperm (Rodrigues et al. 2013), and mature and imbibed embryo (He et al. 2015).



Supplemental Figure 4: siRNA abundance by length and category

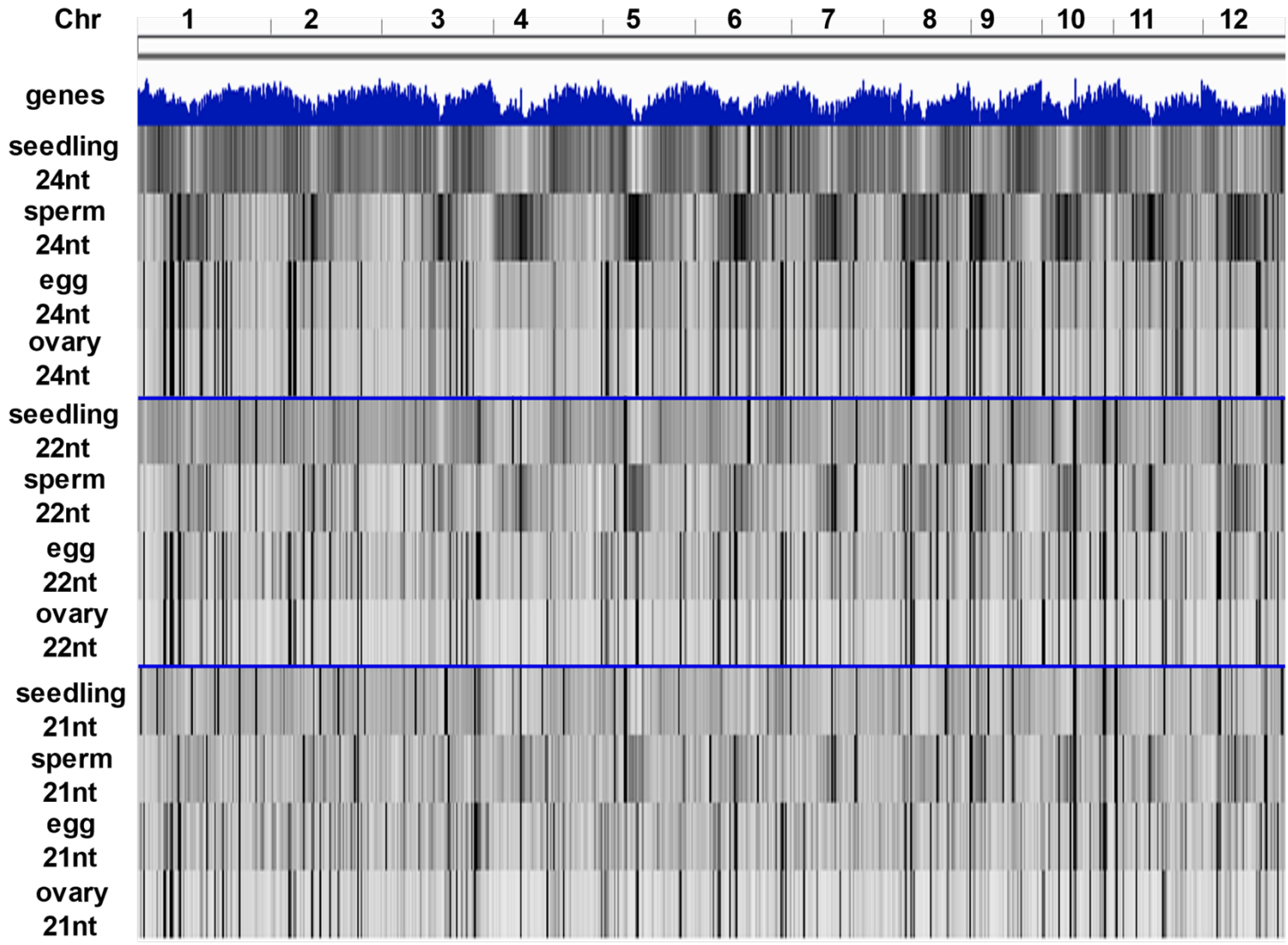
Y-axis values are the number of siRNAs for each length normalized by the total number of siRNAs. Error bars are 95% confidence intervals.

A: Ovary developmental series (Li et al. 2017).

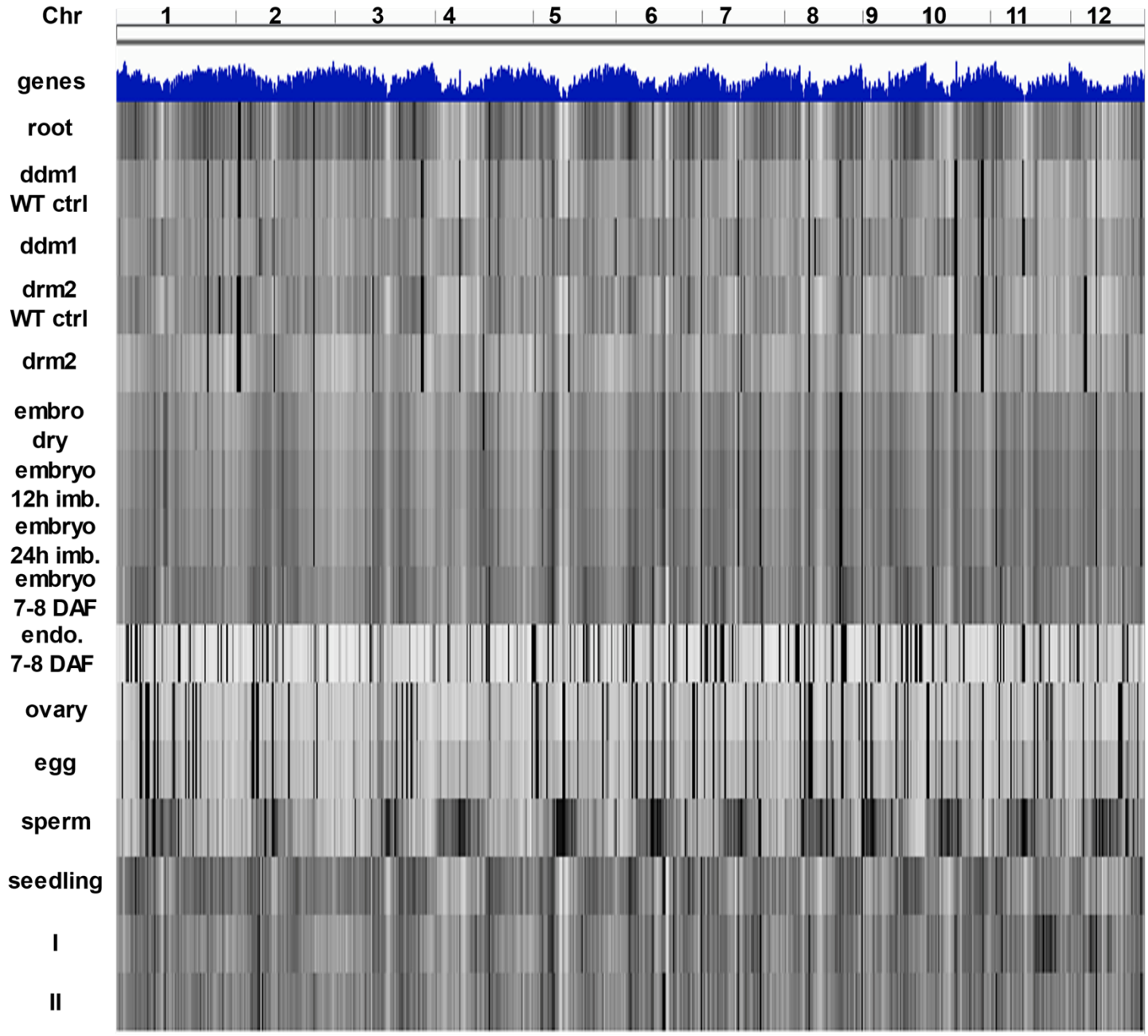
B: Anther developmental series (Li et al. 2017).

C: Spikelet developmental series (Fei et al. 2016).

D: Developing embryo and endosperm, 7 – 8 DAF (Rodrigues et al. 2013).



Supplemental Figure 5: Whole-genome heatmaps of 21-nt, 22-nt, and 24-nt siRNAs
Top track indicates gene density, all others siRNAs.

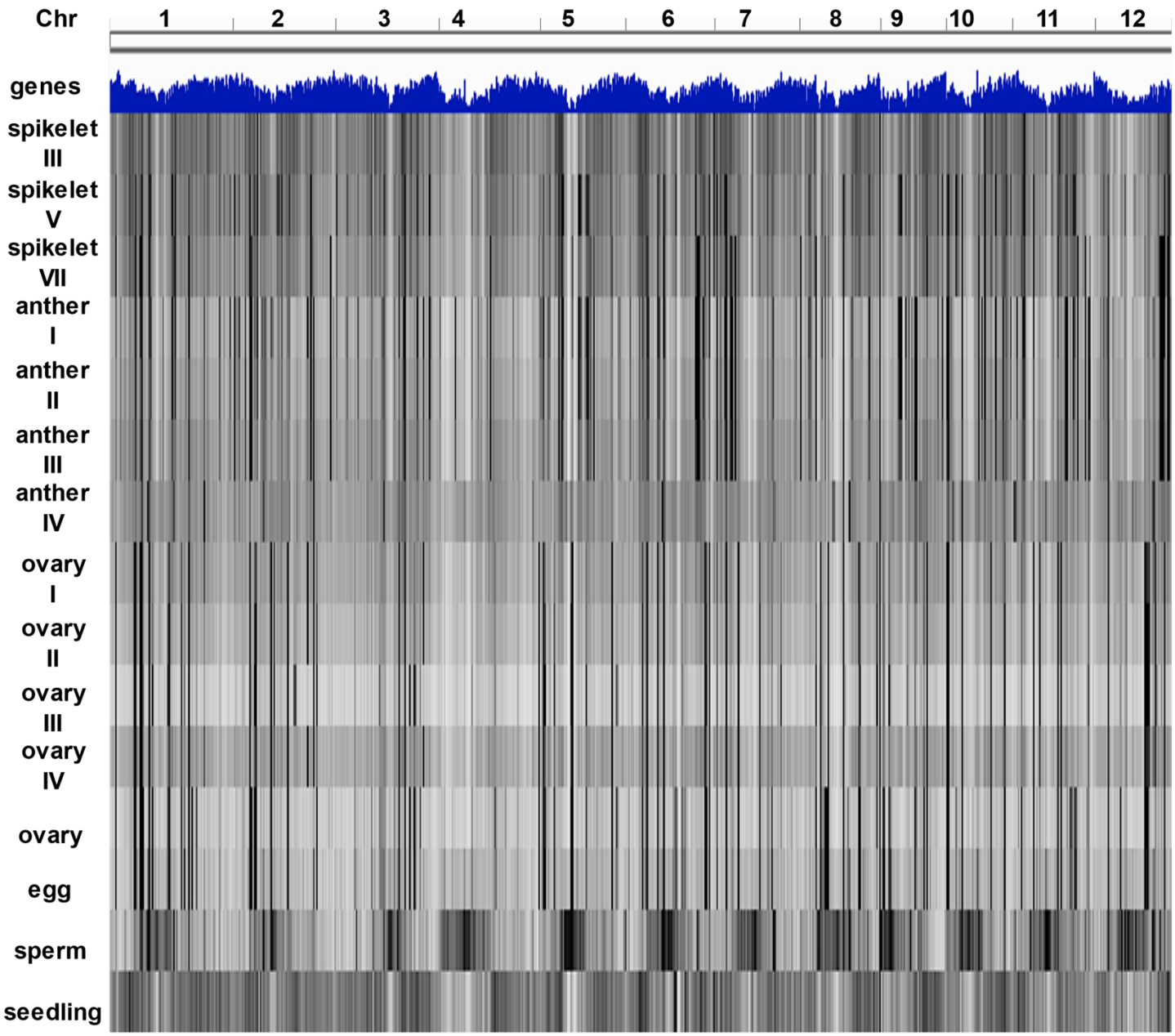


Supplemental Figure 6: Whole-genome heatmaps of vegetative tissues, embryo and endosperm 24-nt siRNAs

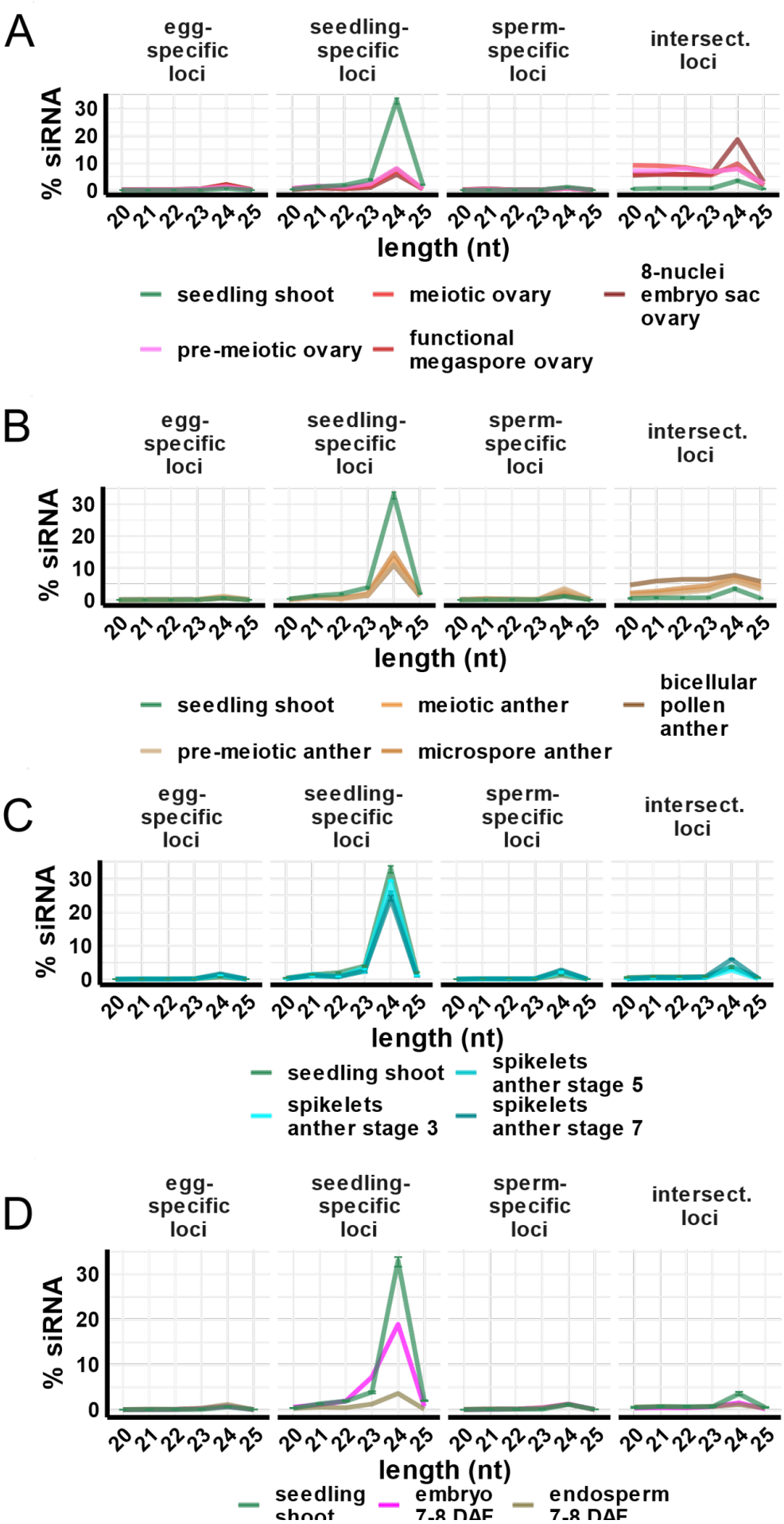
Top track indicates gene density, all others siRNAs. Published data sources: *drm2* and wildtype leaf (Tan et al. 2016), *ddm1* and wildtype leaf (Tan et al. 2018), root (Shin et al. 2018), 7-8 DAF embryo and endosperm (Rodrigues et al. 2013), and mature and imbibed embryo (He et al. 2015).

I: 24-nt siRNAs from seedling shoot after removing all siRNAs that overlapped with seedling-specific 24-nt siRNA loci.

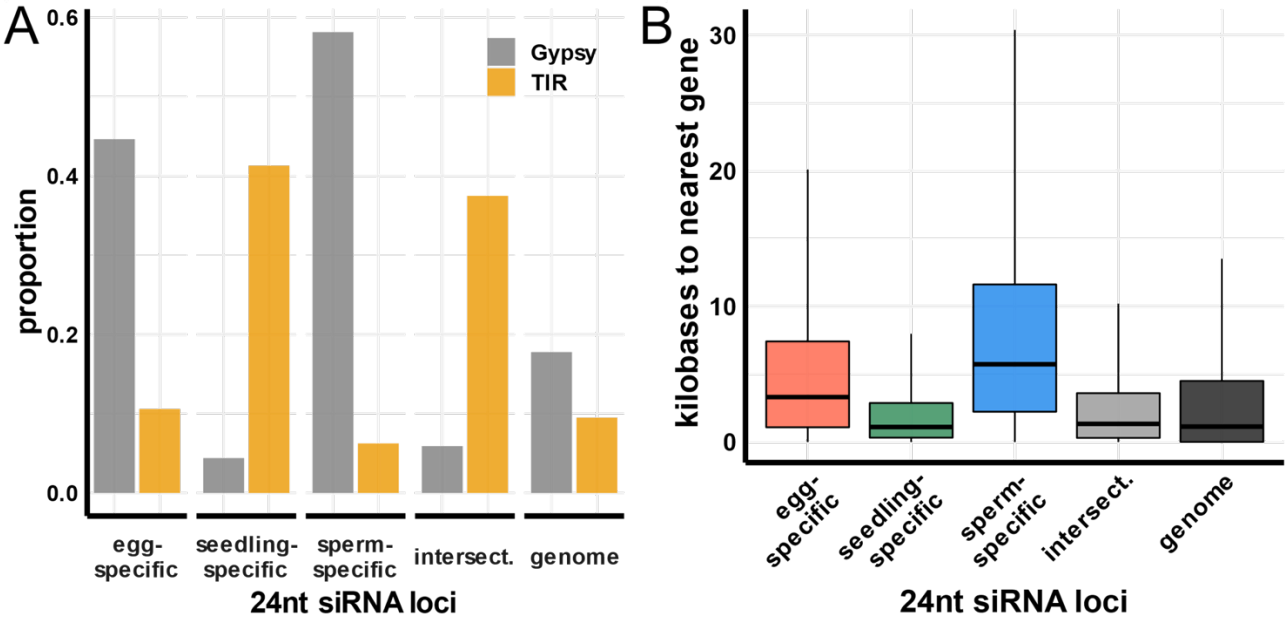
II: 24-nt siRNAs from seedling shoot after removing all siRNAs that overlapped with mCHH loci, the set of 100-bp loci in the genome with mCHH > 5% from leaf (Tan et al. 2016).



Supplemental Figure 7: Whole-genome heatmaps of reproductive tissue 24-nt siRNAs
Top track indicates gene density, all others siRNAs. Spikelet III, IV and VII: spikelet of anther developmental stages 3, 5 and 7, respectively (Fei et al. 2016); Anther I, II, III and IV: pre-meiotic anther; meiotic anther; microspore anther and bicellular pollen anther, respectively (Li et al. 2017);
Ovary I, II, III, IV: pre-meiotic ovary; meiotic ovary; functional megaspore ovary and 8-nuclei embryo sac ovary, respectively (Li et al. 2017).



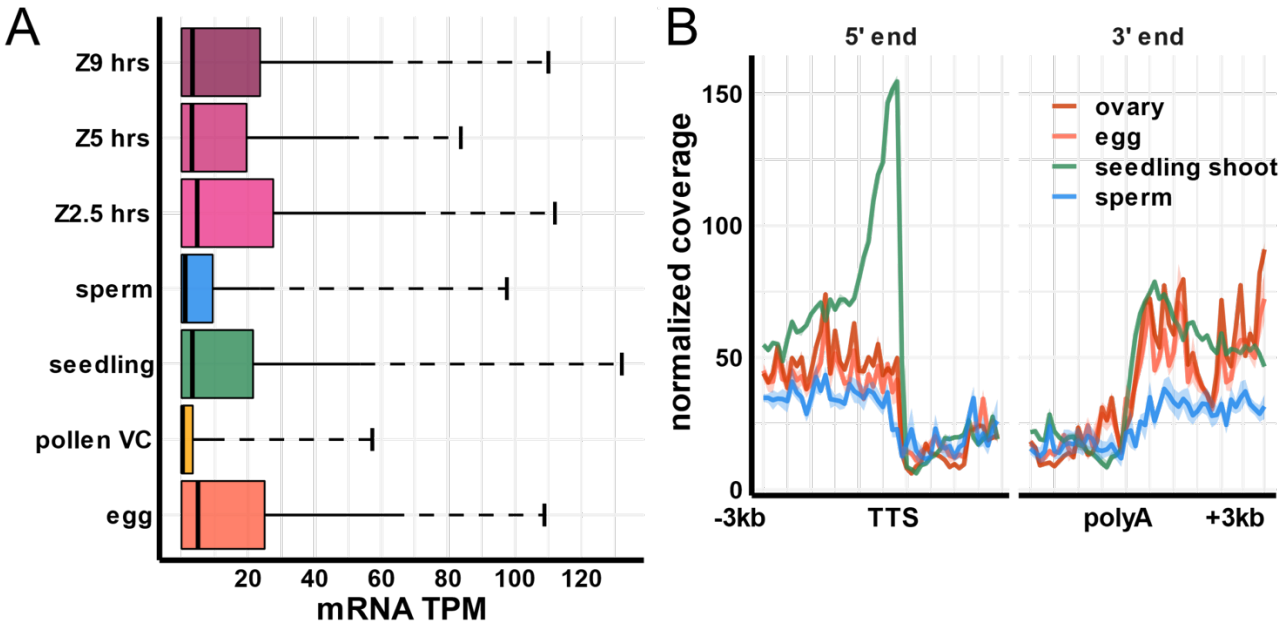
Supplemental Figure 8: siRNA abundance at 24-nt siRNA loci.
Y-values are the percent of siRNAs relative to total siRNAs. Error bars are 95% confidence intervals for sample types with more than one replicates, as in **Fig. 1D**. Data sources as indicated in **Fig. S4**.
A: Ovary developmental series.
B: Anther developmental series.
C: Spikelet developmental series.
D: Developing embryo and endosperm, 7 – 8 DAF.



Supplemental Figure 9: Overlaps of sample-specific siRNA loci with repeats and distances to nearest genes

A: Proportion of sample-specific or intersection 24-nt siRNA loci overlapping a *Gypsy* or TIR transposon.

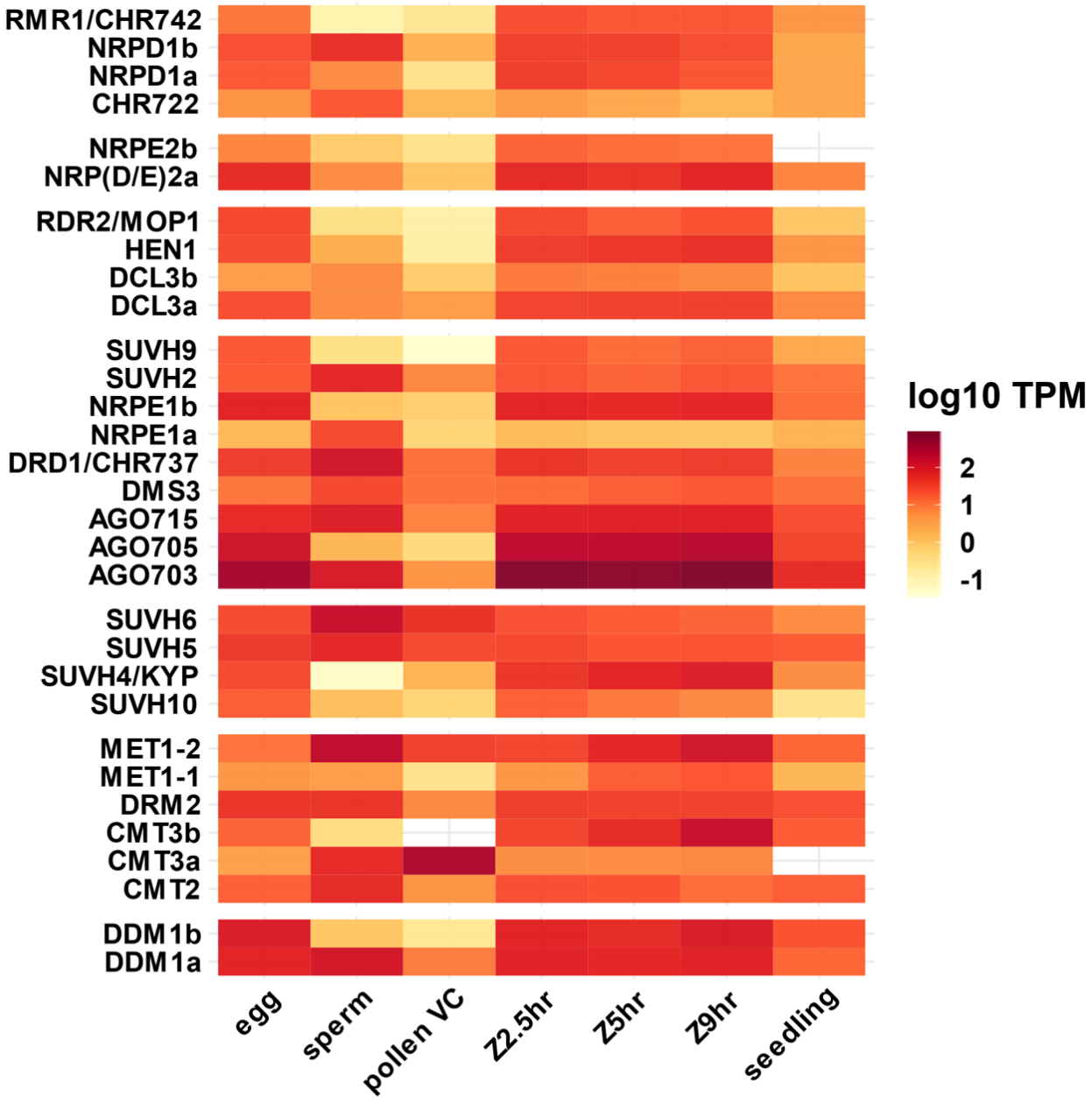
B: Distance in kilobases from 24-nt siRNA loci to the nearest gene. Boxes are interquartile range (IQR), upper whisker extends to upper quartile + 2 * IQR, lower whisker to zero.



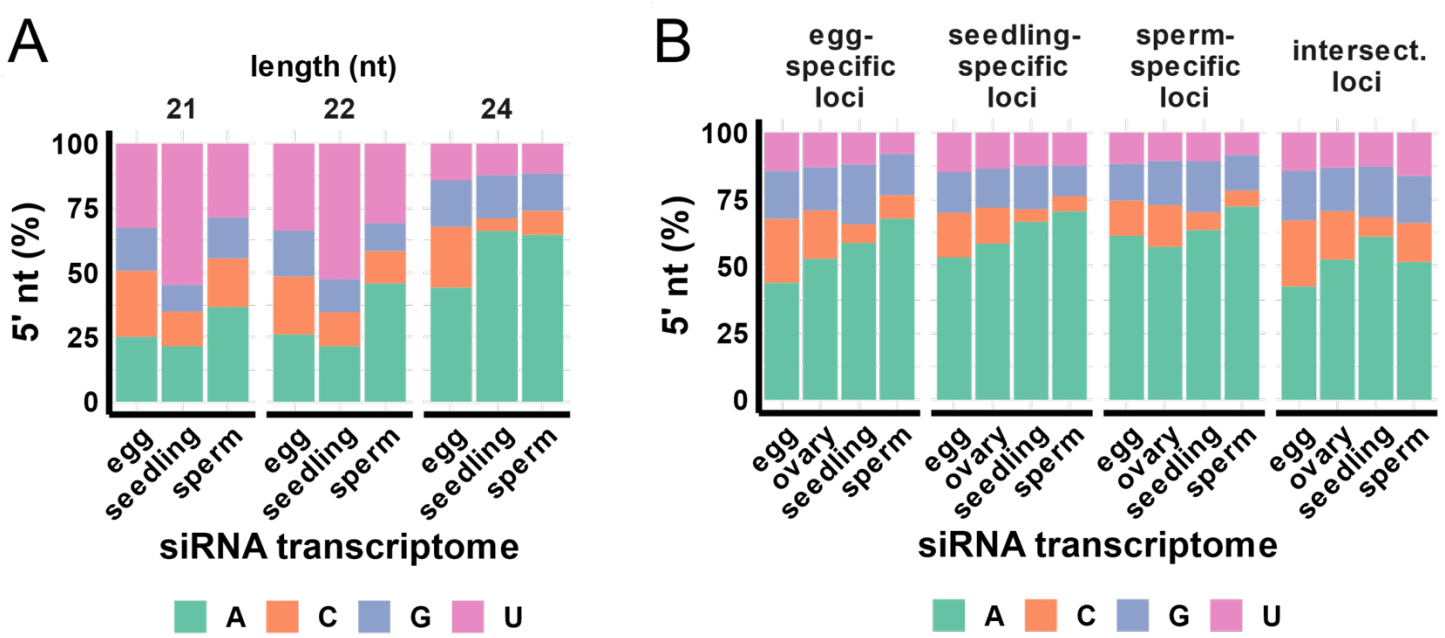
Supplemental Figure 10: Depletion of flanking 24-nt siRNAs for highly expressed sperm genes (>10 TPM)

A: Boxplot showing distribution of mRNA expression. Boxes are interquartile range (IQR), and whiskers extend to upper quartile + 1.5 * IQR. Dotted lines extend to 95th percentile. A TPM (transcripts per million) of 10 approximately corresponds to upper quartile in sperm.

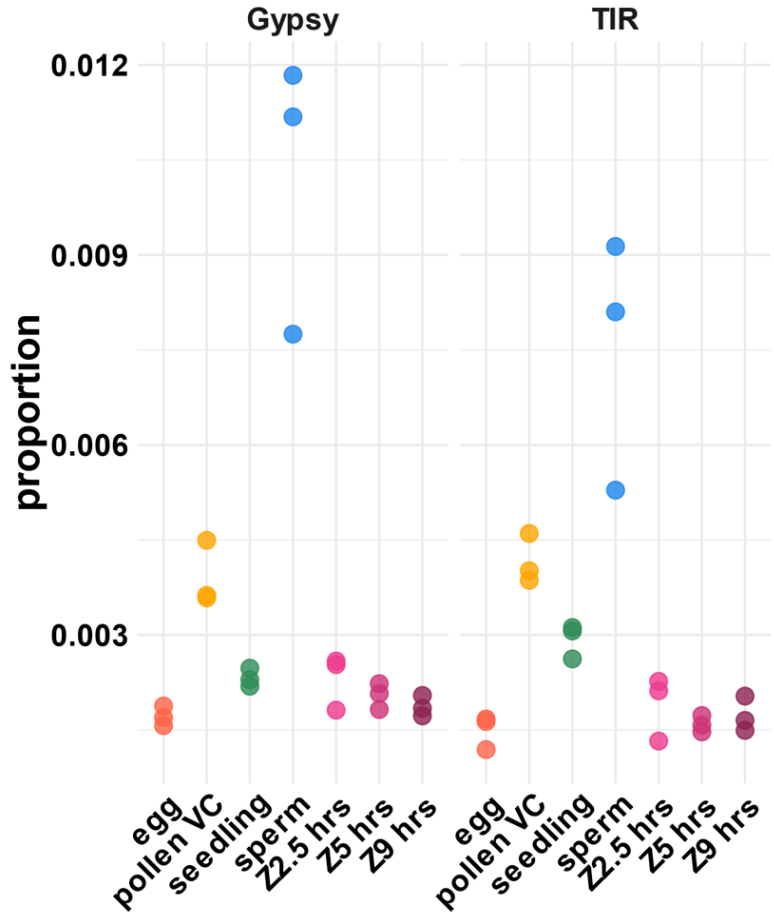
B: Metagene plot for 24-nt siRNAs for genes expressing at >10 TPM in sperm, as in **Fig 2B**. Plots indicate 24-nt siRNA coverage with 100-bp resolution from 3-kb upstream to 3-kb downstream of genes, normalized per 1000 total siRNAs. Tick marks indicate 500-bp intervals. TSS: Transcription start site; poly A: polyadenylation signal.



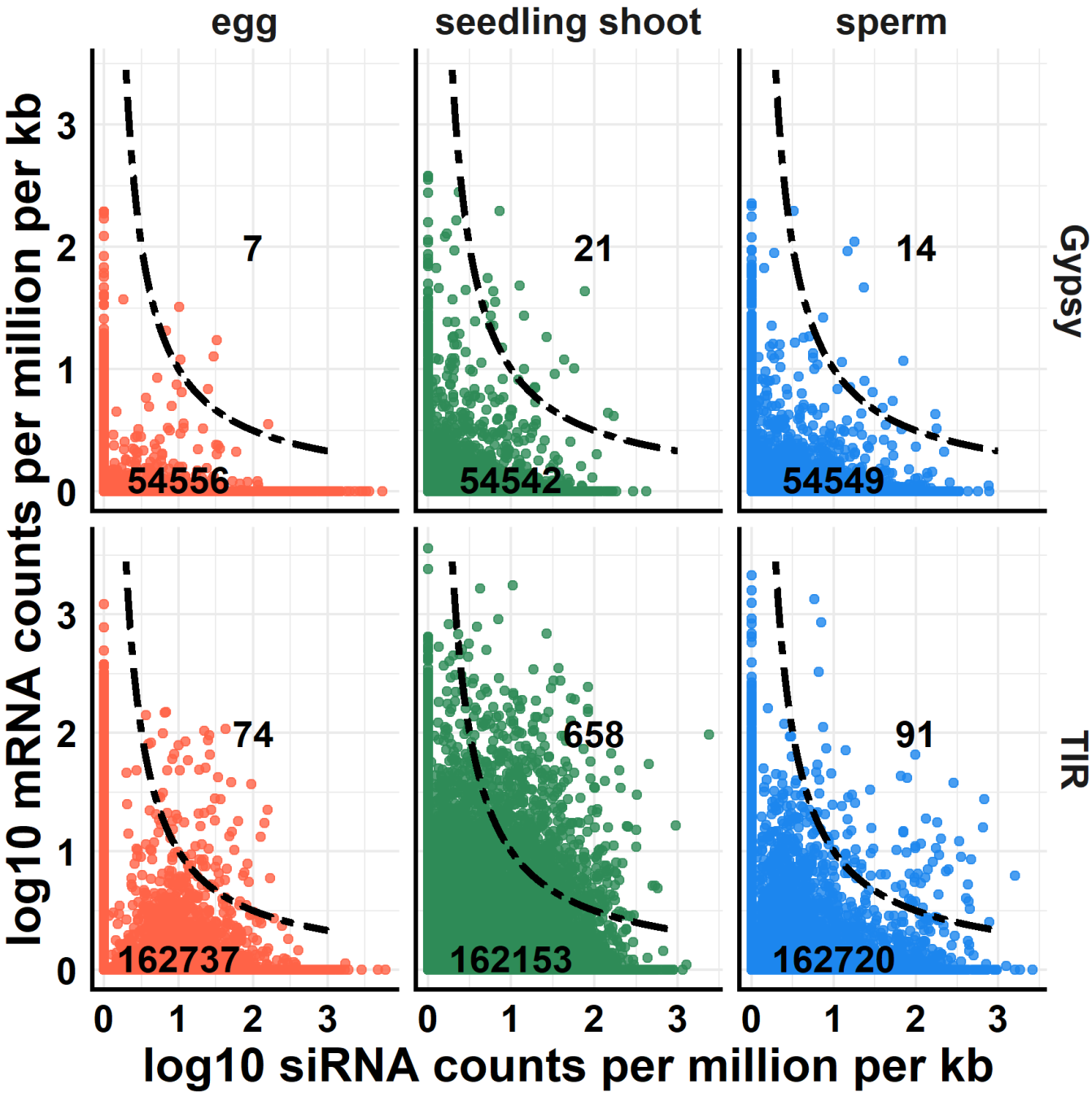
Supplemental Figure 11: Expression of RdDM and methylation related factors in gametes
Pollen VC: pollen vegetative cell; Z: zygote, 2.5hr, 5hr, and 9hr, at completion of karyogamy, nucleoli fusion and S-phase, respectively. Colors reflect transcripts per million values in log10 scale. mRNA data source: (Anderson et al. 2013; Anderson et al. 2017).



Supplemental Figure 12: siRNA 5' nucleotide preferences
A: The stacked bar charts show the percent of 24-nt siRNAs from the tissue-specific 24-nt siRNA loci categories that begin with each nucleotide. **B:** 5' nucleotide preference of 24-nt siRNAs across different siRNA loci.

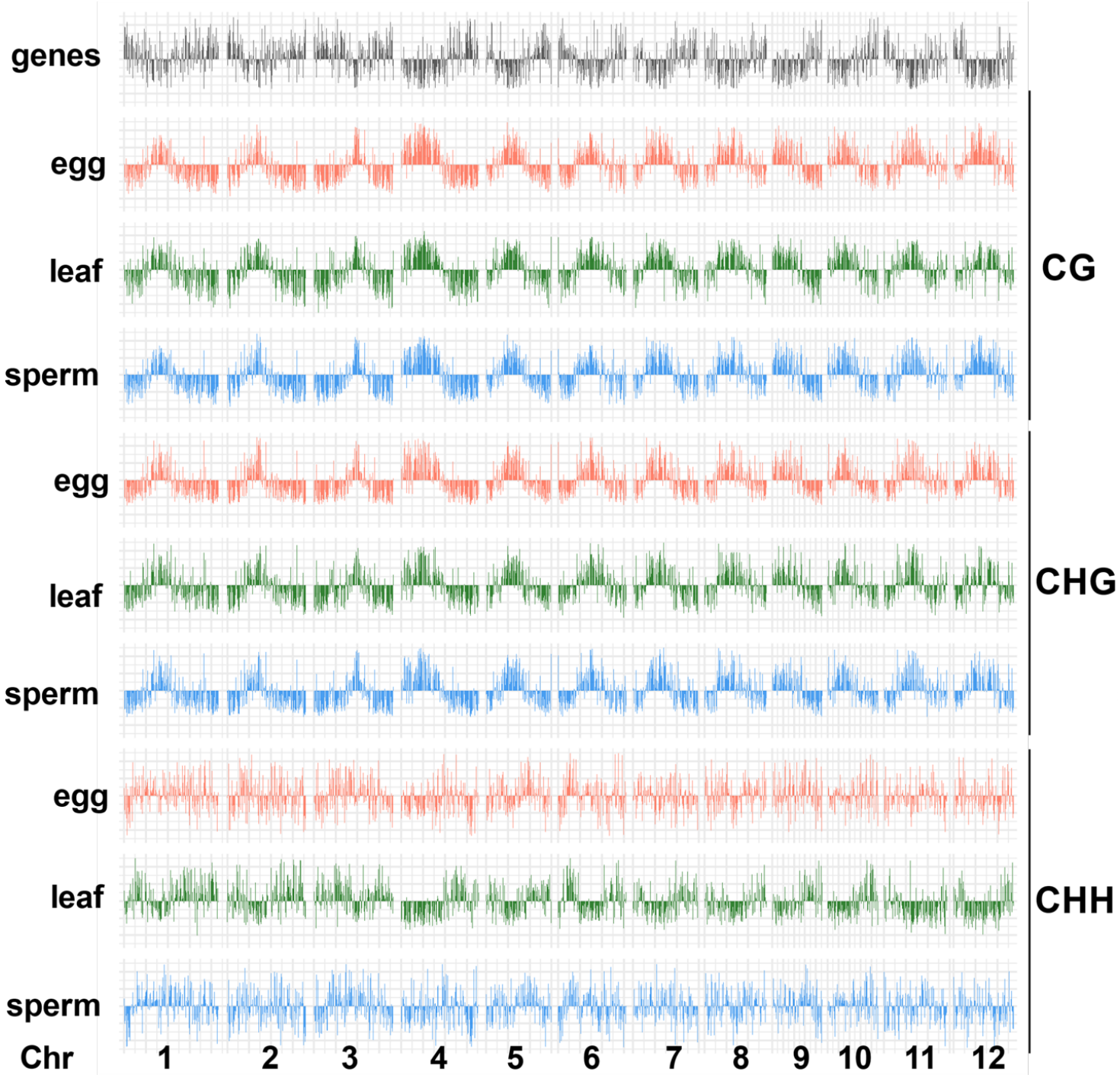


Supplemental Figure 13: Proportion of mRNA reads mapping to transposons in gametes
All reads were included in this analysis, uniquely and multi-mapping. Read counts were normalized by total number of mapping reads per library. Pollen VC: pollen vegetative cell; Z: zygote, 2.5hr, 5hr, and 9hr, at completion of karyogamy, nucleoli fusion and S-phase, respectively. mRNA data source: (Anderson et al. 2013; Anderson et al. 2017).

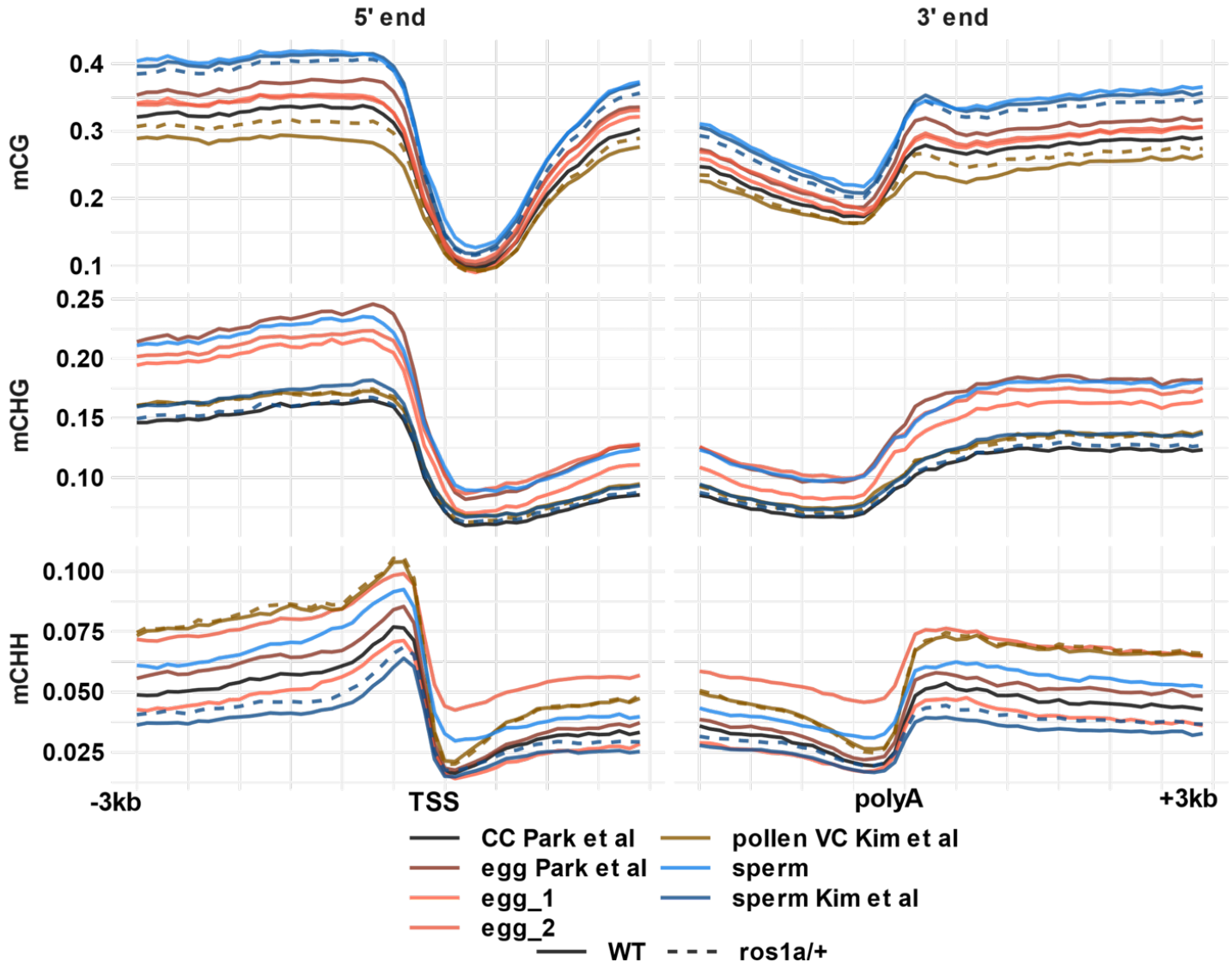


Supplemental Figure 14: mRNA read counts vs. 24-nt siRNA read counts for individual transposon copies

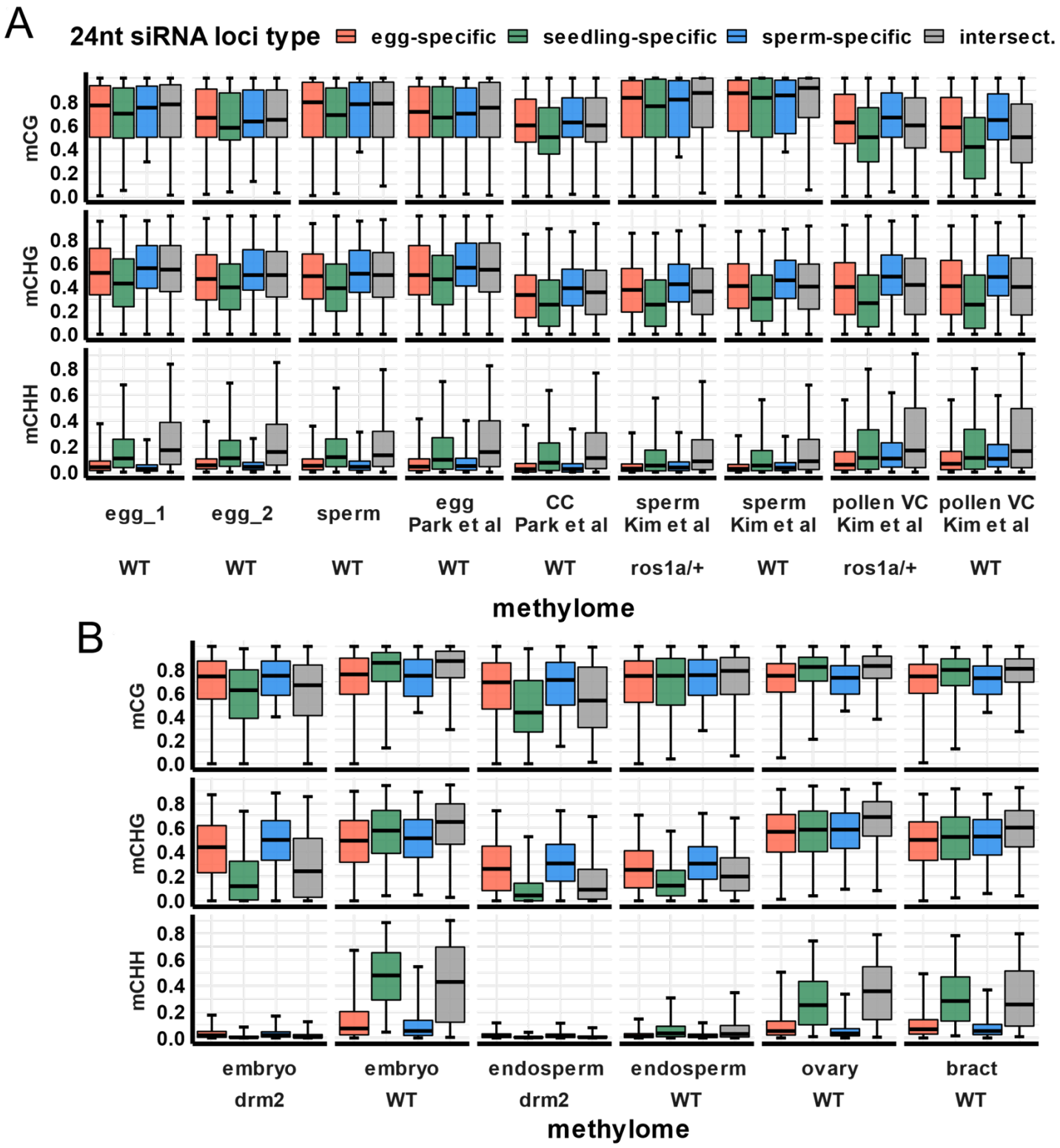
Only uniquely-mapping reads were included in this analysis. mRNA counts were normalized by total number of mapping reads and siRNA reads were normalized per million 24-nt siRNAs. Dotted line plots y = 1/x relationship. Numbers above and below the curve show number of transposon copies outside or inside the y = 1/x relationship, respectively. mRNA data source: (Anderson et al. 2013; Anderson et al. 2017).



Supplemental Figure 15: Genome-wide view of DNA methylation
Y-axis values are z-scores of average DNA methylation values on 50 kilobase intervals. Z-scores were calculated from mean and standard deviation of each sample type and each context. Leaf data source: (Tan et al. 2016).



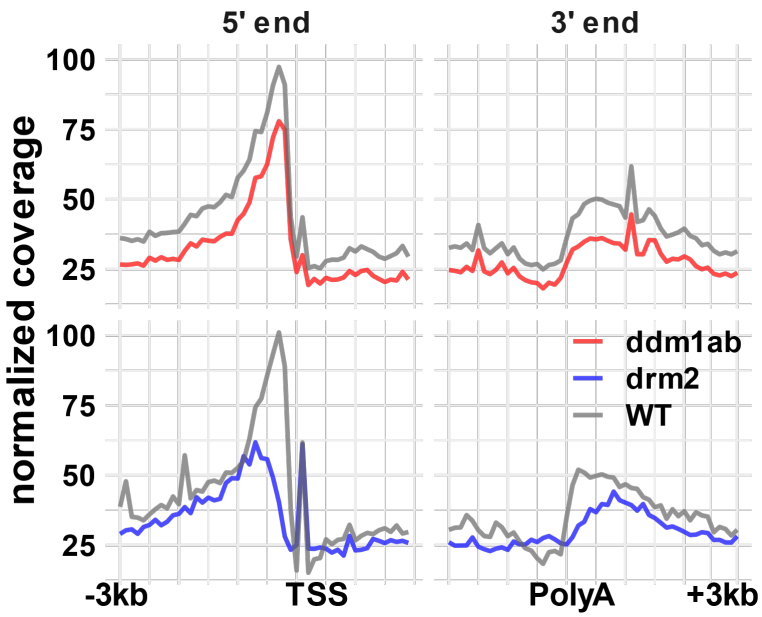
Supplemental Figure 16: Methylation metaplots of all PBAT libraries analyzed
Plots indicate average DNA methylation values over 100-bp intervals from 3-kb upstream to 3-kb downstream of genes. Tick marks indicate 500-bp intervals. DNA methylation is measured as the proportion methylated cytosines relative to total cytosines in each sequence context. TSS: Transcription start site; polyA: polyadenylation signal. CC, central cell (Park et al. 2016); egg cell (Park et al. 2016); sperm (Kim et al. 2019); and pollen VC, vegetative cell (Kim et al. 2019).



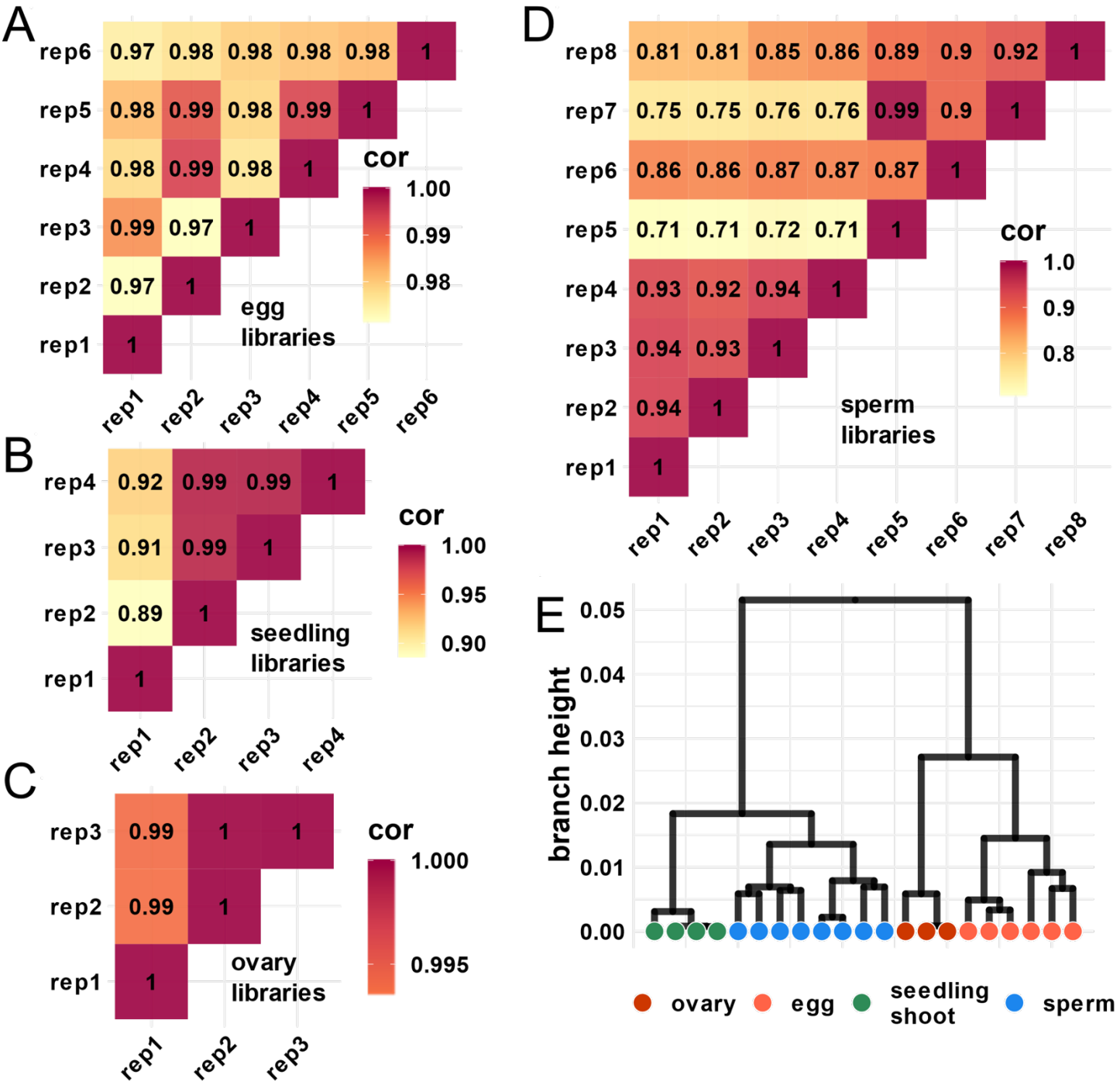
Supplemental Figure 17: DNA methylation of 24-nt siRNA loci

A: PBAT libraries. Center line is median methylation; box represents interquartile range; whiskers extend from 2.5th to 97.5th percentile.

B: Conventional libraries as in **A**. Bract: lemma and palea of rice florets. CC, central cell (Park et al. 2016); egg cell (Park et al. 2016); sperm (Kim et al. 2019); and pollen VC, vegetative cell (Kim et al. 2019).



Supplemental Figure 18: Metagene plot for 24-nt siRNAs in *ddm1* and *drm2* mutants Plots indicate 24-nt siRNA coverage with 100-bp resolution from 3-kb upstream to 3-kb downstream of genes, normalized per 1000 total siRNAs. Tick marks indicate 500-bp intervals. TSS: Transcription start site; polyA: polyadenylation signal. Data sources: (Tan et al. 2016; Tan et al. 2018).



Supplemental Figure 19: Correlation and clustering of 24-nt siRNA libraries
A: Pairwise Pearson's correlation coefficient between six egg cell libraries.
B: Pairwise Pearson's correlation coefficient between four seedling shoot libraries.
C: Pairwise Pearson's correlation coefficient between three ovary libraries.
D: Pairwise Pearson's correlation coefficient between eight sperm cell libraries.
E: Clustering of small RNA libraries based on hierarchical clustering.

Supplemental References

Anderson SN, Johnson CS, Chesnut J, Jones DS, Khanday I, Woodhouse M, Li C, Conrad LJ, Russell SD, Sundaresan V. 2017. The Zygotic Transition Is Initiated in Unicellular Plant Zygotes with Asymmetric Activation of Parental Genomes. *Dev Cell* **43**: 349-358.e344.

Anderson SN, Johnson CS, Jones DS, Conrad LJ, Gou X, Russell SD, Sundaresan V. 2013. Transcriptomes of isolated Oryza sativa gametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization. *Plant J* **76**: 729-741.

Chan PP, Lowe TM. 2016. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* **44**: D184-189.

Fei Q, Yang L, Liang W, Zhang D, Meyers BC. 2016. Dynamic changes of small RNAs in rice spikelet development reveal specialized reproductive phasiRNA pathways. *J Exp Bot* **67**: 6037-6049.

Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, Chen PY, Pellegrini M. 2013. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14**: 774.

Guo W, Zhu P, Pellegrini M, Zhang MQ, Wang X, Ni Z. 2018. CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* **34**: 381-387.

He D, Wang Q, Wang K, Yang P. 2015. Genome-Wide Dissection of the MicroRNA Expression Profile in Rice Embryo during Early Stages of Seed Germination. *PLoS One* **10**: e0145424.

Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S et al. 2013. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)* **6**: 4.

Khanday I, Skinner D, Yang B, Mercier R, Sundaresan V. 2019. A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. *Nature* **565**: 91-95.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.

Kim MY, Ono A, Scholten S, Kinoshita T, Zilberman D, Okamoto T, Fischer RL. 2019. DNA demethylation by ROS1a in rice vegetative cells promotes methylation in sperm. *Proc Natl Acad Sci U S A* doi:10.1073/pnas.1821435116.

Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Res* **47**: D155-D162.

Langfelder P, Zhang B, Horvath S. 2008. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**: 719-720.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Li C, Xu H, Russell SD, Sundaresan V. 2019. Step-by-step protocols for rice gamete isolation. *Plant Reprod* **32**: 5-13.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Li X, Shahid MQ, Xia J, Lu Z, Fang N, Wang L, Wu J, Chen Z, Liu X. 2017. Analysis of small RNAs revealed differential expressions during pollen and embryo sac development in autotetraploid rice. *BMC Genomics* **18**: 129.

Martin M. 2011. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnetjournal* **17**: 3.

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**: 4288-4297.

Moritoh S, Eun CH, Ono A, Asao H, Okano Y, Yamaguchi K, Shimatani Z, Koizumi A, 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.

Park K, Kim MY, Vickers M, Park JS, Hyun Y, Okamoto T, Zilberman D, Fischer RL, Feng X, Choi Y et al. 2016. DNA demethylation is initiated in the central cells of Arabidopsis and rice. *Proc Natl Acad Sci U S A* **113**: 15138-15143.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**: R25.

Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, Fischer RL, Zilberman D. 2013. Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proc Natl Acad Sci U S A* **110**: 7934-7939.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**: 863-864.

Shin SY, Jeong JS, Lim JY, Kim T, Park JH, Kim JK, Shin C. 2018. Transcriptomic analyses of rice (Oryza sativa) genes and non-coding RNAs under nitrogen starvation using multiple omics technologies. *BMC Genomics* **19**: 532.

Tan F, Lu Y, Jiang W, Wu T, Zhang R, Zhao Y, Zhou DX. 2018. DDM1 Represses Noncoding RNA Expression and RNA-Directed DNA Methylation in Heterochromatin. *Plant Physiol* **177**: 1187-1197.

Tan F, Zhou C, Zhou Q, Zhou S, Yang W, Zhao Y, Li G, Zhou DX. 2016. Analysis of Chromatin Regulators Reveals Specific Features of Rice DNA Methylation Pathways. *Plant Physiol* **171**: 2041-2054.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178-192.

Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. 2015. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat Protoc* **10**: 475-483.

Xie K, Zhang J, Yang Y. 2014. Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated genome editing in model plants and major crops. *Mol Plant* **7**: 923-926.

Zhang T, Talbert PB, Zhang W, Wu Y, Yang Z, Henikoff JG, Henikoff S, Jiang J. 2013. The CentO satellite confers translational and rotational phasing on cenH3 nucleosomes in rice centromeres. *Proc Natl Acad Sci U S A* **110**: E4875-4883.