The Plant Cell, Vol. 31: 444–464, February 2019, www.plantcell.org © 2019 ASPB.

Meiocyte-Specific and AtSPO11-1–Dependent Small RNAs and Their Association with Meiotic Gene Expression and Recombination[OPEN]

Jiyue Huang,[a,b,1] Cong Wang,[a,1] Haifeng Wang,[c] Pingli Lu,[a] Binglian Zheng,[a] Hong Ma,[a,d] Gregory P. Copenhaver,[b] and Yingxiang Wang[a,2]

[a] State Key Laboratory of Genetic Engineering and Collaborative Innovation Center of Genetics and Development, Ministry of Education Key Laboratory of Biodiversity Sciences and Ecological Engineering, Institute of Plant Biology, School of Life Sciences, Fudan University, Shanghai 200433, China

[b] University of North Carolina at Chapel Hill Department of Biology and the Integrative Program for Biological and Genome Sciences, Chapel Hill, North Carolina 27599–3280

[c] State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi Key Lab of Sugarcane Biology, College of Agriculture, Guangxi University, 530005, Nanning, Guangxi, China

[d] Department of Biology, the Pennsylvania State University, University Park, Pennsylvania 16802

ORCID IDs: 0000-0002-7185-4201 (J.H.); 0000-0002-1647-1072 (C.W.); 0000-0001-7213-927X (H.W.); 0000-0002-3941-9955 (P.L.); 0000-0001-5650-5348 (B.Z.); 0000-0001-8717-4422 (H.M.); 0000-0002-7962-3862 (G.P.C.); 0000-0001-6085-5615 (Y.W.)

Meiotic recombination ensures accurate chromosome segregation and results in genetic diversity in sexually reproducing eukaryotes. Over the last few decades, the genetic regulation of meiotic recombination has been extensively studied in many organisms. However, the role of endogenous meiocyte-specific small RNAs (ms-sRNAs; 21–24 nucleotide [nt]) and their involvement in meiotic recombination are unclear. Here, we sequenced the total small RNA (sRNA) and messenger RNA populations from meiocytes and leaves of wild type Arabidopsis (*Arabidopsis thaliana*) and meiocytes of *spo11-1*, a mutant defective in double-strand break formation, and we discovered 2,409 ms-sRNA clusters, 1,660 of which areSPORULATION 11-1 (AtSPO11-1)–dependent. Unlike mitotic small interfering RNAs that are enriched in intergenic regions and associated with gene silencing, ms-sRNAs are significantly enriched in genic regions and exhibit a positive correlation with genes that are preferentially expressed in meiocytes (i.e. Arabidopsis *SKP1-LIKE1* and *RAD51*), in a fashion unrelated to DNA methylation. We also found that AtSPO11-1–dependent sRNAs have distinct characteristics compared with ms-sRNAs and tend to be associated with two known types of meiotic recombination hotspot motifs (i.e. CTT-repeat and A-rich motifs). These results reveal different meiotic and mitotic sRNA landscapes and provide new insights into how sRNAs relate to gene expression in meiocytes and meiotic recombination.

INTRODUCTION

Meiosis is a specialized cell division that involves two rounds of chromosome segregation after a single round of premeiotic DNA replication, which serves to generate haploid germ cells. Unlike mitosis, which faithfully transmits genetic information from mother cells to daughter cells, in most eukaryotes meiosis produces cells that are not genetically identical due to the action of meiotic recombination, which shuffles chromosome segments between homologous chromosomes (homologs). The molecular mechanisms of meiotic recombination have been intensively studied in several organisms, including fungi, animals, and plants (Szostak et al., 1983; Wang et al., 2013; Hunter, 2015; Mercier et al., 2015; Wang and Copenhaver, 2018). According to the classic double-strand-break (DSB) repair model for meiotic recombination (Szostak et al., 1983), initiation of meiotic recombination starts with programmed DNA DSBs generated along homologs by SPORULATION 11 (SPO11), which is a subunit of the topoisomerase-VI–like complex resembling the archaeal topoisomerase (Keeney et al., 1997; de Massy, 2013; Robert et al., 2016; Vrielynck et al., 2016). DSB ends are resected by the MEIOTIC RECOMBINATION 11/RADIATION SENSITIVE 50/NIJMEGEN BREAKAGE SYNDROME 1 (MRE11-RAD50-NBS1) complex to form 3′ single-stranded DNA tails (Neale et al., 2005), which are protected by the single-stranded DNA-binding protein Replication Protein A (RPA) in *Saccharomyces cerevisiae* (Soustelle et al., 2002). However, some Arabidopsis (*Arabidopsis thaliana*) homologs, including RPA1a, function in a later step of meiotic recombination, probably by facilitating second-end capture (Osman et al., 2009; Li et al., 2013). The DNA binding ATPases RAD51 (Shinohara et al., 1992) and DMC1 (Bishop et al., 1992) then replace RPA and facilitate invasion of the 3′ tail into a non-sister chromatid, displacing one of its strands and creating a D-loop, to form a stable recombination intermediate known as a single-end invasion (SEI; Hunter and Kleckner, 2001). SEIs can go on to form another intermediate called a “double Holliday

[1] These authors contributed equally to this paper.

[2] Address correspondence to yx_wang@fudan.edu.cn.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Yingxiang Wang (yx_wang@fudan.edu.cn).

[OPEN]Articles can be viewed without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.18.00511

junction," which can be resolved to form a crossover (CO) via the "synapsis initiation complex" protein–dependent CO-interference sensitive pathway or the MMS AND UV SENSITIVE 81-dependent CO-interference insensitive pathway (Hunter, 2015). Alternatively, SEIs can be processed by the synthesis-dependent strand annealing pathway to form non-COs (McMahill et al., 2007). Over the last few decades, molecular genetic studies have identified ~30 genes required for CO formation in Arabidopsis (Mercier et al., 2015; Wang and Copenhaver, 2018).

In addition to genetic analysis, high throughput sequencing technologies have advanced our understanding of meiotic recombination (Qi et al., 2009; Brunschwig et al., 2012; Lu et al., 2012; Wijnker et al., 2013). These include the ability to examine the genome-wide distribution of COs in multiple species. Most COs occur in narrow regions called "hotspots." Recent studies have linked genomic and epigenetic features, including DNA methylation, nucleosome positioning, and histone modification, to shaping the meiotic hotspot landscape (Choi et al., 2013, 2018; Underwood et al., 2018). In Arabidopsis, DNA methylation is found at cytosine residues in three DNA sequence contexts: primary CG methylation, and non-CG methylation at CHG and CHH (H equals A, T, or C) sites. Loss of CG methylation causes redistribution of COs with an increase in euchromatic regions and a decrease in heterochromatic regions, but it does not change the total number of COs per meiosis (Melamed-Bessudo and Levy, 2012; Mirouze et al., 2012; Yelina et al., 2012, 2015). In addition, loss of non-CG methylation activates pericentromeric COs, possibly by increasing DSB formation near centromeres (Underwood et al., 2018). Other studies have shown that transcriptionally active chromatin contexts, such as low nucleosome density regions, influence the activity of meiotic recombination in yeast and plants (Pan et al., 2011; Choi et al., 2013). Transcriptionally active chromatin is marked by a histone H2A variant, H2A.Z, whose distribution is positively correlated with CO hotspots at gene promoters in Arabidopsis (Choi et al., 2013). H2A.Z also promotes initiation of meiotic recombination in fission yeast (Yamada et al., 2018). Consistent with these observations, most COs in maize (*Zea mays*) are also found in the genic region (Rodgers-Melnick et al., 2015). The distribution of the meiotic recombination hotspots is also correlated with the distribution of histone H3 Lys-4 trimethylation (H3K4me3) in yeast and mammals (Borde et al., 2009; Baudat et al., 2010). In addition, in *S. cerevisiae*, a protein required for DSB formation, Mer2, physically interacts with Spp1, which is a subunit of the H3K4 methylase Set1/COMPASS complex (Acquaviva et al., 2013; Sommermeyer et al., 2013), suggesting a molecular mechanism for the epigenetic regulation of meiotic recombination. In mammals, H3K4me3 is deposited by a SET-domain protein called PR/SET DOMAIN 9, which uses its zinc finger domain to recognize specific DNA motifs and determines the distribution of H3K4me3 marks (Baudat et al., 2010). Plants do not have a PR/SET DOMAIN 9 homolog, but H3K4me3 is associated with CO hotspots in Arabidopsis and maize (Choi et al., 2013; He et al., 2017). By contrast, H3K9me2, a marker of heterochromatin, exerts a repressive effect on both meiotic DSBs and CO formation (Underwood et al., 2018).

Another important epigenetic modification in the genome is small interfering RNA (siRNA)-guided RNA-dependent DNA methylation (RdDM). In plants, siRNAs are a subset of small RNAs (sRNAs), which typically are 24- or 21- to 22–nt long, and are produced from double-stranded RNA precursors. Because siRNAs are processed from longer RNA precursors, they can be mapped back to the genome as clusters. siRNAs mainly function in transcriptional gene silencing of viral DNA, transgenes, and transposable elements (TEs; Borges and Martienssen, 2015). Intriguingly, mitotic DSBs generated by external DNA-damaging agents are able to induce the expression of novel endogenous sRNAs in plants, Drosophila, and human cells (Francia et al., 2012; Michalik et al., 2012; Wei et al., 2012), suggesting the existence of DNA repair-associated sRNAs; however, the functions of these sRNAs are not clear. It is not known whether meiotic SPO11-dependent DSBs can also trigger sRNA production in meiocytes and, if so, what role they might play during meiosis.

To investigate the presence and possible roles of sRNAs, we performed deep sequencing of total sRNAs, ranging from 15 to 30 nt in length, from leaves and meiocytes isolated from wild-type Arabidopsis (Col-0) and meiocytes of the *spo11-1-1*and *pol iv* single mutants, and *dcl2/3/4* triple mutant. We then examined the characteristics and genomic distributions of these sRNAs and found that approximately one-third are meiocyte-specific (referred to here as "ms-sRNAs"). By contrast to the heterochromatic siRNAs expressed in leaves, ms-sRNAs exhibit an obvious enrichment in non-TE–associated coding sequences, and are significantly correlated with positive gene expression in meiocytes. Sequence analysis of AtSPO11-1–dependent ms-sRNAs identified two previously reported meiotic recombination-associated motifs: a CTT-repeat motif associated with genic regions and an A-rich motif associated with gene promoters. In total, we identified 2,409 ms-sRNA clusters, 1,660 of which are AtSPO11-1–dependent and have distinct features with regard to either gene expression or meiotic recombination. These discoveries significantly broaden our understanding of sRNAs and their potential role in meiotic recombination.

RESULTS

Characteristics of sRNAs in Arabidopsis Meiocytes

To gain a genome-wide view of the sRNAs expressed during meiosis, we first isolated Arabidopsis meiocytes using a microcapillary procedure (Chen et al., 2010; Yang et al., 2011; Wang et al., 2014). We sequenced six sRNA libraries: two from wild-type meiocytes, two from 3-week-old wild-type leaves, and two from *spo11-1-1* (Grelon et al., 2001) meiocytes (Supplemental Table 1). Each pair of biological replicates had a >0.9 correlation coefficient (Supplemental Table 2). Thirteen to thirty-five million raw reads were retrieved for each library, resulting in up to 3 million mapped reads with 22% to 66% unique reads per library (Supplemental Table 1). We filtered out all annotated noncoding sRNAs including micro RNAs (miRNAs) and then examined the read distributions. Both wild-type and *spo11-1-1* meiocytes have a major peak at 23 nt and a smaller peak at 24 nt (Figure 1A). Leaf sRNAs, by comparison, have a major peak at 24 nt and a smaller peak at 23 nt. Previous studies demonstrate that 24-nt heterochromatic siRNAs are the most abundant in somatic cells, and that their biogenesis is

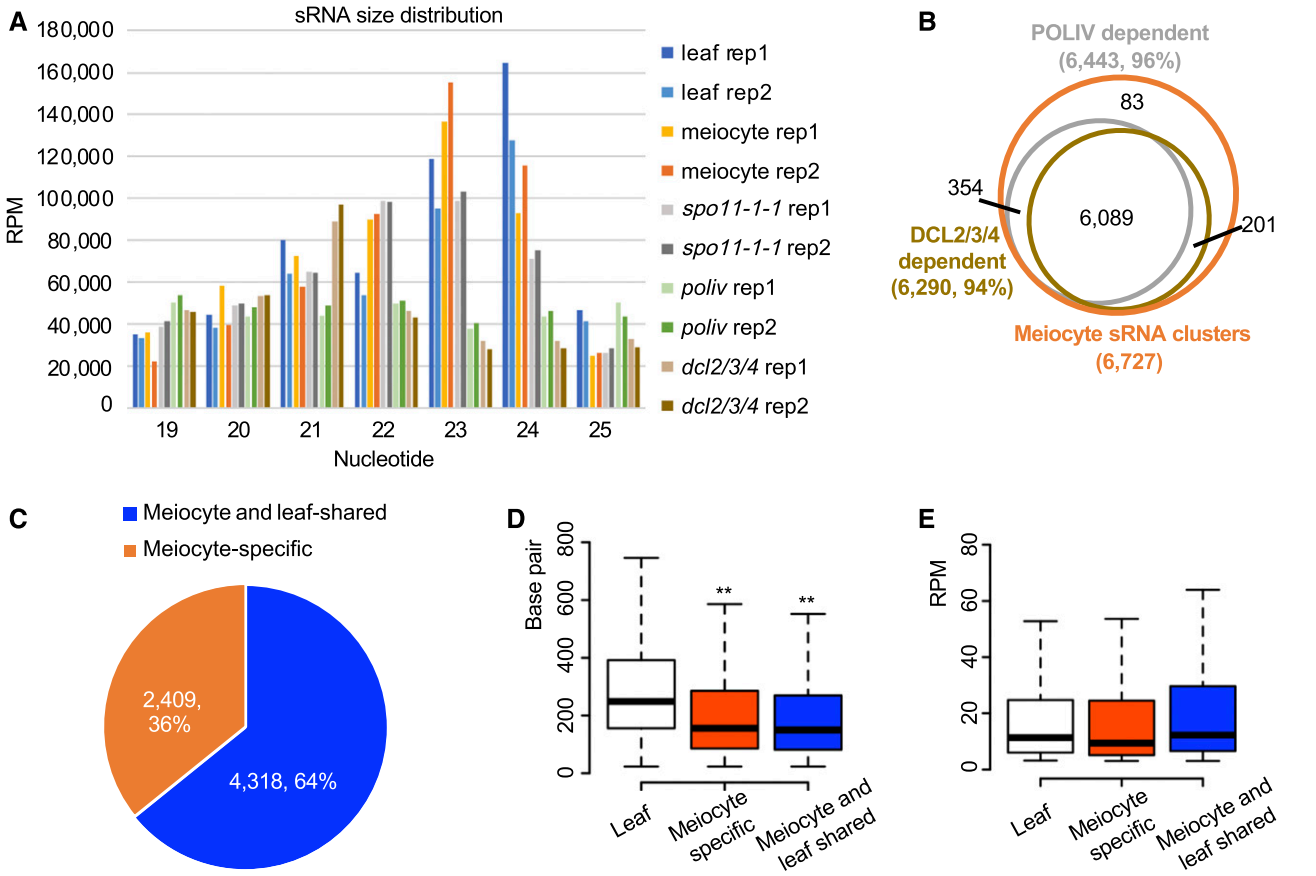

Figure 1. Characteristics of Meiocyte sRNA Clusters.

(A) Mappable sRNA size distribution from all samples examined. Each sample has two biological replicates. Adapters were trimmed from sRNA reads, and low quality reads and reads that mapped to ribosomal RNAs, transfer RNAs, small nRNAs, small nucleolar RNAs, miRNAs, and other annotated small nuclear RNAs were filtered out.

(B) The majority of meiocyte sRNAs are dependent on RNA Polymerase IV (POLIV) and DCL2/3/4.

(C) Thirty-six percent of sRNA clusters found in meiocytes are meiocyte-specific. ms-sRNA clusters are defined as the clusters having a <40% overlap rate or do not overlap with the sRNA clusters in leaves.

(D) Meiocyte sRNA clusters are significantly shorter than leaf sRNA clusters (**P value < 0.01). sRNA cluster length in leaves was compared with ms-sRNA cluster length and meiocyte/leaf-shared sRNA cluster length (both P value < 2.2e-16; Student's t test).

(E) No significant difference in sRNA abundance was observed between leaf and meiocyte clusters. Student's t test was used to compare RPM in leaf versus meiocyte-specific (P value = 0.35) and leaf versus meiocyte/leaf-shared (P value = 0.18).

mainly dependent on RNA polymerase IV and Dicer proteins (Matzke and Mosher, 2014). To determine if the meiocyte sRNAs we observed are also RNA polymerase IV- and Dicer-dependent, we sequenced sRNA from *poliv* and *dcl2/3/4* triple mutants, with two replicates (Supplemental Table 1), and found that the majority of the 23- and 24-nt meiocyte sRNAs are abolished in either *poliv* or *dcl2/dcl3/dcl4* triple mutants (Figure 1A). This suggests that 23-nt meiocyte sRNAs are likely nonspecific decay products from the 24-nt sRNAs. Because 23- and 24-nt sRNAs account for ~60% of the total sRNAs in meiocytes (Figure 1A), the subsequent analyses are focused on 23- and 24-nt sRNAs.

Clusters of sRNAs were identified by aligning 23–24-nt reads to the Arabidopsis genome (The Arabidopsis Information Resource 10 [TAIR10] release), which resulted in 14,352 and 10,666 sRNA clusters from the two meiocyte libraries and 16,692 and 16,329 sRNA clusters from the leaf libraries (Table 1), indicating that sRNAs in meiocytes are less abundant than those in leaves. Further analyses showed that 7,942 meiocyte clusters occurred in both library replicates with at least 1-nt overlap and 13,871 leaf clusters replicated using the same criteria, indicating a high degree of concordance between the replicates for each tissue/cell types. For all subsequent analyses, we used a conservative criterion of a minimum overlap of 60% (more than 60% of one sRNA cluster region in replicate 1 can be covered by one sRNA cluster region in replicate 2) between replicates to define a core set of 6,727 meiocyte sRNA clusters and 11,753 leaf sRNA clusters (Table 1). For the 6,727 meiocyte sRNA clusters, 96% and 94% are dependent on RNA polymerase IV and DCL proteins, respectively (Figure 1B), indicating that meiocyte sRNAs are likely a kind of siRNA and their biogenesis also requires RNA polymerase IV and

Table 1. sRNA Clusters in Different Genotypes

Genotype	Cluster No.	Shared cluster No.
Meiocyte_rep1	14,352	6,727
Meiocyte_rep2	10,666	
Leaf rep1	16,692	11,753
Leaf rep2	16,329	
spo11-1-1 rep1	13,818	4,756
spo11-1-1 rep2	14,876	
poliv rep1	19,134	2,997
poliv rep2	19,631	
dcl2,3,4 rep1	4,013	658
dcl2,3,4 rep2	3,366	

By ShortStack, ≥3 RPM, overlapping > 60%

DCLs. However, before the biogenesis and mechanism of action of these meiotic sRNAs are fully described, it would be premature to speculate that they are a distinct sRNA species. For the purpose of this article, we will only refer to them as "meiocyte sRNAs."

Comparing leaf and meiocyte clusters revealed that 36% of the meiocyte sRNA clusters (2,409) were absent in the leaf cluster set (Figure 1C; Supplemental Table 3). We defined these as meiocyte-specific sRNAs (ms-sRNAs). Correspondingly, 64% of the meiocyte sRNA clusters (4,318) were shared between meiocytes and leaves (Figure 1C). We also compared the length and the abundance among three groups of sRNA clusters: meiocyte-specific, meiocyte/leaf-shared, and leaf. The abundance of sRNA cluster is defined with Reads Per Million mapped reads (RPM). Median lengths of both the meiocyte-specific (156 basepairs, bp) and the meiocyte/leaf-shared sRNA clusters (150 bp) were significantly shorter than that of the leaf sRNA clusters (245 bp; Student's *t* test, both *P* values < 2.2e-16; Figure 1D). By contrast, sRNA abundances were similar among all three cluster sets (Student's *t* test, *P* value = 0.17 between leaf and meiocyte-specific, and *P* value = 0.08 between leaf and meiocyte/leaf-shared; Figure 1E).

ms-sRNA Clusters Tend To Localize with Coding Sequences Rather than TEs

Previous analysis of sRNAs (15–30 nt) from Arabidopsis showed that RNA polymerase IV-dependent 24-nt siRNAs are the most abundant sRNA category in somatic tissues (Borges and Martienssen, 2015). These sRNAs mainly function as intermediates to guide DNA methylation in heterochromatic regions, where TEs are enriched in the Arabidopsis genome, and are thus called "siRNAs" (Kasschau et al., 2007; Law et al., 2013). We analyzed the genomic distribution of meiocyte and leaf sRNAs by plotting leaf, meiocyte-specific, and meiocyte/leaf-shared sRNAs using 10-kb windows. In leaves, most of the clusters are located in centromeric regions (51%; consistent with their role in promoting DNA methylation and silencing TEs) and several high abundance peaks scattered along the chromosome arms (Figure 2A). This pattern is similar to previous results (Kasschau et al., 2007). However, both meiocyte-specific and meiocyte/leaf-shared sRNAs showed a distinct distribution compared with that of leaf sRNAs (Figures 2B and 2C)). Many more ms-sRNA peaks were observed along the chromosome arms (Figure 2B), and significantly fewer ms-sRNA (45%) and meiocyte/leaf-shared sRNA clusters (44%) were found in the centromeric regions (Figure 2D; Supplemental Table 4). A moderate correlation could be drawn between the leaf and the meiocyte/leaf-shared sRNAs (Pearson's $r = 0.66$, P value = 0.025), but not between the leaf and ms-sRNAs (Pearson's $r = 0.02$, P value = 0.036).

We also compared the position of these sRNA clusters to other genomic features, including promoters, 5′ untranslated regions (UTRs), coding DNA sequences (CDSs), introns, 3′ UTRs, intergenic regions, TEs, pseudogenes, and non-coding RNAs (ncRNAs). The results showed that in leaves, 50% of the clusters map to TE regions, consistent with previously reported observations (Kasschau et al., 2007; Law et al., 2013; Figure 2E; Supplemental Figure 1B; Supplemental Table 4). In contrast, TE-derived reads decreased to 42% for the meiocyte/leaf-shared sRNA clusters and 36% for the ms-sRNA clusters. Moreover, only 5% of leaf and meiocyte/leaf-shared sRNA clusters are derived from CDSs, while 17% of ms-sRNA clusters are derived from CDSs (Figure 2E; Supplemental Table 4). The fractions of sRNA clusters that were derived from 5′ UTR, 3′ UTR, ncRNA, pseudogenes, intergenic regions, introns, and promoter regions did not show any obvious differences among the groups (Figure 2E). Taken together, ms-sRNA clusters tended to localize within coding sequences and are less frequently associated with TEs, suggesting that they represent a previously undescribed set of ms-sRNAs with different characteristics compared with those found in leaves.

In the Arabidopsis genome (TAIR10 release), there are 31,189 annotated TEs, which are categorized into 18 super families, with Helitron (34 families; belonging to RC type, which are hypothesized to transpose through a rolling circle replication method (Thomas and Pritham, 2015), MuDR (70 families; belong to DNA type TE), and Gypsy (32 families; belonging toLong Terminal Repeat [LTR] type of retrotransposons) among the most abundant. Aligning sRNAs against these sequences indicated that 1,062 TEs have ms-sRNAs, 2,632 TEs correspond to meiocyte/leaf-shared sRNA clusters, and 4,880 TEs match leaf-specific sRNAs. Among these sRNA-associated TEs, Helitron (563/1,042), MuDR (235/692), and Gypsy (167/183) are the most abundant TEs associated with both ms-sRNA and meiocyte/leaf-shared sRNA clusters, similar to those among TEs associated with leaf-specific sRNAs (Figure 3). More specifically, Helitron elements are overrepresented among TEs associated with ms-sRNA clusters, compared with TEs corresponding to leaf-specific and meiocyte/leaf-shared sRNA clusters, but not to all TEs in Arabidopsis (Supplemental Table 4). By contrast, MuDR, Long Interspersed Nuclear Element (LINE)/L1, and LTR/Copia elements are underrepresented in TEs associated with ms-sRNA clusters, compared with TEs associated with leaf-specific sRNA clusters (Supplemental Table 4). Whereas LTR/Copia and LTR/Gypsy-related sRNAs are underrepresented in meiocyte/leaf-shared sRNA clusters (Supplemental Table 4). These results imply that compared with leaf-specific or meiocyte/leaf-shared sRNAs, ms-sRNAs each showed greater association with different Helitron families and lesser association with other TEs.

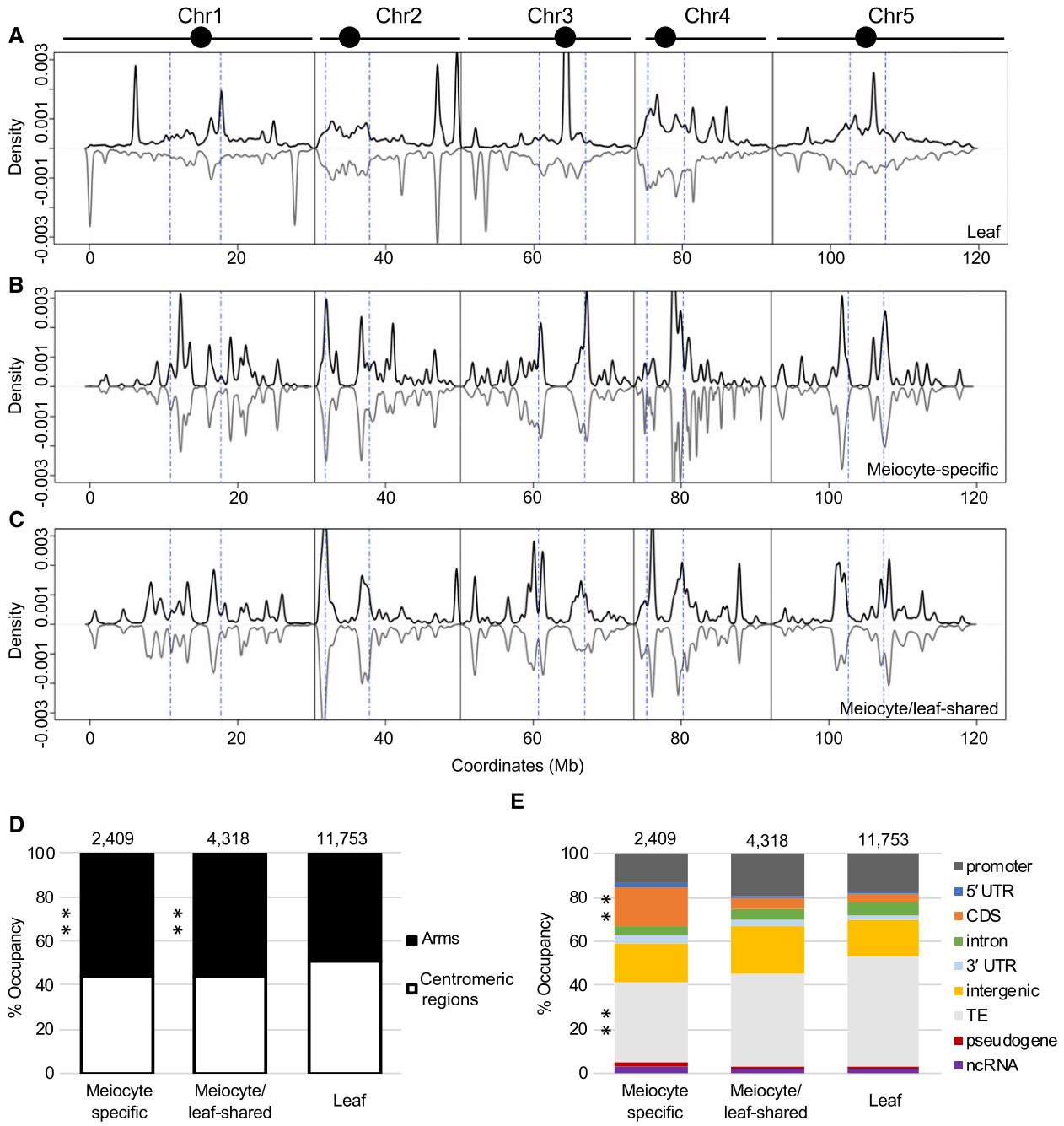

Figure 2. ms-sRNAs Tend To Localize with Genic Regions Rather than TEs.

(A) Density plot of the distribution of leaf sRNAs on the five Arabidopsis chromosomes (black bars; circles represent centromeres). sRNA density on the sense strand shows a positive value (plotted in black), while sRNA density on the antisense strand shows a negative value (plotted in gray). Chromosomes are portioned into arm regions and centromeric regions (blue dot-dashed lines).

(B) Density plot of the distribution of ms-sRNAs on the five Arabidopsis chromosomes.

(C) Density plot of the distribution of meiocyte/leaf-shared sRNAs on the five Arabidopsis chromosomes.

(D) ms-sRNA clusters are enriched on chromosome arms (**P value < 0.01). Pearson's χ^2 tests were used to compare the percentage of reads mapped chromosome arms versus centromeric regions between leaf and ms-sRNA clusters (χ^2 = 30.54, P value = 3.28e-08) and between leaf and meiocyte/leaf-shared sRNA clusters (χ^2 = 51.36, P value = 7.68e-13).

(E) ms-sRNA clusters are significantly enriched in genic regions and underrepresented in TEs (**P value < 0.01). Pearson's χ^2 tests were used to determine whether leaf and the ms-sRNA clusters are differentially enriched at several classes of genomic features. Significant differences were observed at CDSs (χ^2 = 588.18, P value < 2.2e-16) and TEs (χ^2 = 79.20, P value < 2.2e-16).

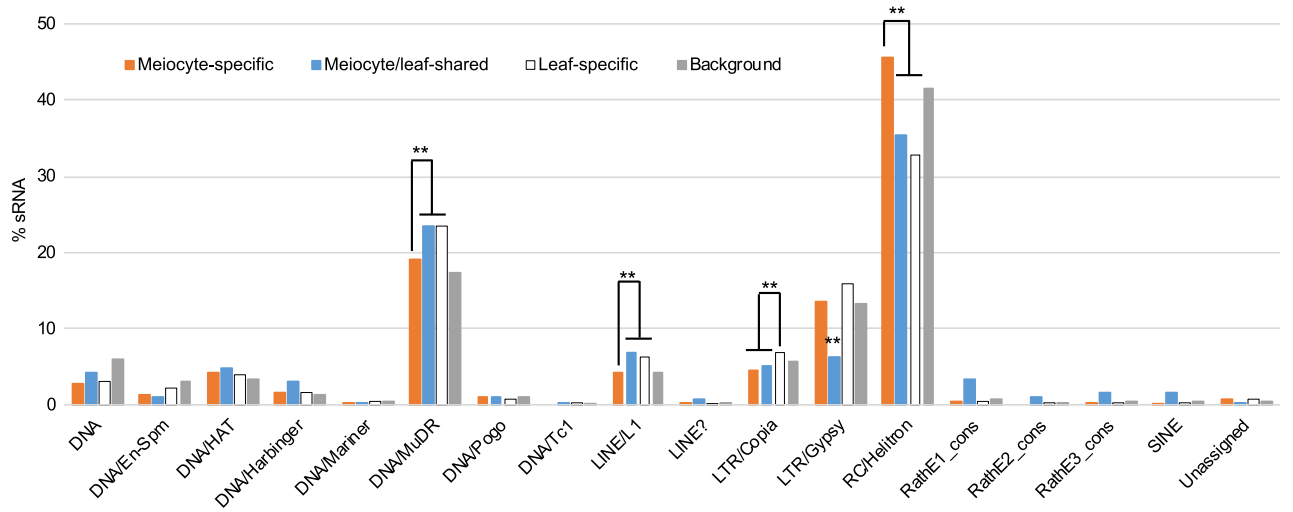

Figure 3. Meiocyte sRNA Occupancy among TE Superfamilies.

Distribution of meiocyte-specific, meiocyte/leaf-shared, and leaf-specific sRNA clusters among TE superfamilies. **P value < 0.01 in Pearson's χ^2 tests.

ms-sRNAs Are Positively Associated with Gene Expression in Meiocytes

Previous studies have shown that plant siRNAs are preferentially associated with heterochromatic regions and inhibit both TE activity and gene expression (Borges and Martienssen, 2015). To determine the relationship between the euchromatin-enriched ms-sRNAs and gene expression, we sequenced messenger RNA (mRNA) from the same samples used for small RNA sequencing (sRNA-Seq; Supplemental Table 1) and obtained 18,953 and 18,889 expressed genes with at least three reads detected from meiocytes and leaves, respectively (Trapnell et al., 2012). Each kind of tissue/cells has two biological replicates with a correlation coefficient > 0.9 (Supplemental Figure 2; Supplemental Table 5). We mapped sRNAs to regions 5 kb upstream and downstream of the genes expressed in leaves and meiocytes. The results showed that, as expected, sRNAs from leaves map less frequently to the gene body region of relatively highly expressed genes compared with genes with reduced expression (including repressed, silenced, and those with low levels; Figure 4A), consistent with their role in gene silencing (Borges and Martienssen, 2015). Meiocyte/leaf-shared sRNAs possessed similar distribution patterns to that of the leaf sRNAs (Figure 4B). Surprisingly, ms-sRNAs map more frequently to gene body regions of highly expressed genes than others (Figure 4C). These observations suggest a positive correlation of ms-sRNAs with meiocyte-expressed genes.

To verify the observed positive correlation between gene-body–associated ms-sRNAs and gene expression in meiocytes (Figure 4C), we calculated the percentage of sRNAs associated with an expressed gene in three intervals: upstream of transcription start site (TSS), gene body (from TSS to transcription termination site, TTS), and downstream of TTS. In leaves, ~45% of the sRNAs found upstream of the TSS or downstream of the TTS are associated with expressed genes (Figure 4D). This percentage decreases to 15% among sRNAs within the gene body (Figure 4D), indicating most of the leaf sRNAs in gene body regions are associated with unexpressed genes. A similar pattern was also observed for the meiocyte/leaf-shared sRNAs (Figure 4D). By contrast, ~50% of ms-sRNAs upstream of the TSS or downstream of the TTS are associated with expressed genes (Figure 4D). Strikingly, >50% of the ms-sRNAs within the gene body are associated with expressed genes (Figure 4D). Considering the similar number of expressed genes in leaves and meiocytes (18,889 and 18,953, respectively), this strongly supports a positive correlation between ms-sRNAs and gene expression.

To further explore the relationship between sRNAs and gene expression, we searched for genes (TSS to TTS) within 2 kb of a sRNA and defined them as "sRNA-associated genes." This yielded 867 ms-sRNAs, 830 meiocyte/leaf-shared, and 2,329 leaf sRNA-associated genes. Approximately 16% (371) of leaf sRNA-associated genes and 21% (175) of meiocyte/leaf-shared sRNA-associated genes are expressed (Figure 4E). However, 54% (470) of the ms-sRNA–associated genes are expressed ($\chi^2 = 475.51$, P value < 2.2e-16; Figure 4E). Among the three groups, ms-sRNA–associated genes have a median gene expression value of 22.4 reads per kilobase of exon model per million mapped reads; while both meiocyte/leaf-shared and leaf sRNA-associated genes showed significantly lower values of 1.0 and 1.7, respectively (both P value < 2.2e-16 by Mann-Whitney tests; Figure 4F). In addition, among those 470 ms-sRNA–associated genes expressed in meiocytes, 198 are differentially expressed (Q value < 0.05, $\log^2$ [fold change] < −1 or > 1) between meiocytes and leaves, with the majority (149) being up-regulated ($\log^2$ [fold change] > 1) and 49 being down-regulated ($\log^2$ [fold change] < −1; Figure 4G). However, both meiocyte/leaf-shared and leaf sRNA-associated genes show the opposite trend. Of the 175 meiocyte/leaf-shared sRNA-associated genes expressed in meiocytes, 73 are differentially expressed between meiocytes and leaves (Q value <0.05, $\log^2$ [fold change] < −1 or > 1) with 20 up-regulated ($\log^2$ [fold change] > 1) and 53 down-regulated ($\log^2$ [fold change] < −1; Figure 4H). Of the 371 leaf-expressed genes

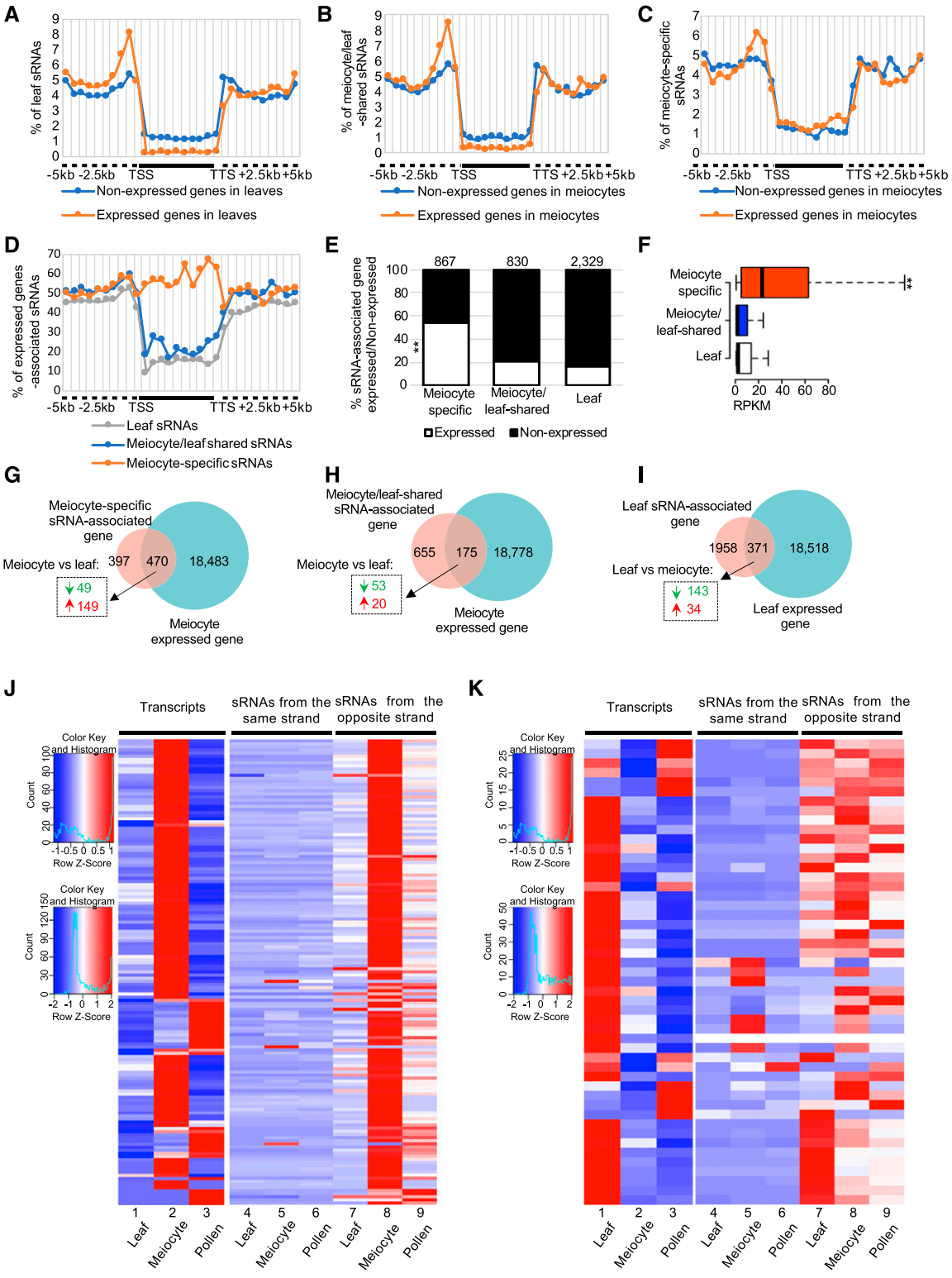

Figure 4. ms-sRNAs Are Positively Associated with Gene Expression in Meiocytes.

(A) Leaf sRNAs are preferentially located at unexpressed genes rather than expressed genes, which are partitioned into up- and downstream regions (dashed line) and gene bodies (solid line) between TSS and TTS.

(B) Meiocyte/leaf-shared sRNAs are preferentially located at unexpressed genes rather than expressed genes.

associated with leaf sRNA clusters, 177 are differentially expressed between meiocytes and leaves (Q value < 0.05, $\log^2$ [fold change] < −1 or > 1) with 34 up-regulated ($\log^2$ [fold change] > 1) and 143 down-regulated ($\log^2$ [fold change] < −1; Figure 4I). Together, these observations indicate that unlike sRNAs in somatic cells, which are associated with gene silencing, ms-sRNAs are associated with transcriptionally active genes that are up-regulated in meiocytes compared with leaf tissue.

To investigate whether the genes associated with ms-sRNAs show meiotic-specific gene expression, we analyzed pollen-derived mRNA and sRNA data from a public database (see "Methods" for details) and compared sRNA and gene expression variation among leaves, meiocytes, and pollen grains. In addition, because strand specificity was well preserved within the library preparation, we could tell whether sRNA clusters and their associated gene transcripts were from the same strand, to preclude the possibility of general mRNA degradation. We focused on the 149 up-regulated genes associated with the ms-sRNA clusters described earlier. All the highly meiotically expressed genes (Figure 4J, column 2) were consistent with a higher abundance of sRNA in meiocytes compared with leaves (Figure 4J, columns 7 and 8). At the pollen grains stage, 78% (116) of the genes have lower expression levels in pollen (Figure 4J, column 3). At the same time, ~91% (136) of the genes have lower sRNA abundance compared with meiocytes (Figure 4J, columns 8 and 9). We also analyzed the sRNA reads on the same strand of the transcribed genes, and observed no meaningful changes (Figure 4J, columns 4 to 6). In contrast, for the 49 down-regulated genes associated with ms-sRNA, 43% (21) of genes have lower sRNA abundance in meiocytes than in leaves (Figure 4K, columns 7 and 8). At the pollen grains stage, only 33% (16) of the genes have higher expression levels in pollen (Figure 4K, column 3) and 39% (19) of the genes have higher sRNA abundance compared with meiocytes (Figure 4K, columns 8 and 9). These findings demonstrate that ms-sRNAs specifically correlated with meiocyte-specific, up-regulated gene expression during reproductive development.

ms-sRNA–Targeted Gene Expression Is Independent of DNA Methylation

In plants, the RdDM pathway, which requires both Domains Rearranged Methyltransferase2 and 24-nt siRNAs, can mediate de novo DNA methylation in all three contexts (Law and Jacobsen, 2010). Recently, Walker and colleagues reported a distinct DNA methylation pattern in Arabidopsis meiocytes that shows higher CG/CHG and lower CHH methylation (Walker et al., 2018). To investigate whether these differentially methylated regions (DMRs) are correlated with the ms-sRNA–associated genes we identified, we compared DMRs between meiocytes and leaves and identified 362 hyper-DMRs versus 257 hypo-DMRs in CG contexts, 3,648 hyper-DMRs versus 226 hypo-DMRs in CHG contexts, and 884 hyper-DMRs versus 18,447 hypo-DMRs in CHH contexts, consistent with previous reports indicating more hyper-CG/CHG and hypo-CHH DMRs in meiocytes (Walker et al., 2018). We then compared those DMRs and meiocyte-specifically expressed genes and found no significant positive correlation between gene expression and gene body CHH methylation (Figure 5A) or gene body CHG methylation (Figure 5B). Intriguingly, gene body CG methylation is positively correlated with gene expression (Figure 5C), consistent with previous studies (Tran et al., 2005; Wang et al., 2015). These results suggest that the sRNAs associated with genes specifically up-regulated in meiocytes might not trigger canonical RdDM-mediated de novo CHH methylation in meiocytes. We also analyzed the correlation of ms-sRNAs with gene body CG methylation. For the 362 hypermethylated CG sites, only eight (2%) correspond to ms-sRNA clusters, and hypermethylated CHG sites are similarly sparse (61/3,648, 2%). By contrast, hypermethylated CHH sites are considerably more abundant (168 of 884, 19%). Taken together, these results provided evidence that sRNAs in meiocytes are similar to those in somatic cells in that they mainly target nonsymmetric CHH sites instead of symmetric sites. However, unlike the case in somatic cells, ms-sRNAs are positively correlated with gene expression and lack corresponding DNA methylation patterns typically associated with the RdDM pathway.

Gene Ontology Annotation of Up-Regulated ms-sRNA–Associated Genes

To learn more about possible functions of the ms-sRNA–associated meiotically expressed genes, we conducted gene ontology (GO) enrichment analysis (Mi et al., 2017) of the 149 up-regulated meiotic genes with mapped ms-sRNAs and found that

Figure 4. (continued).

(C) ms-sRNAs are preferentially located at expressed genes rather than unexpressed genes.
(D) Over half of ms-sRNAs from gene body regions are associated with expressed genes.
(E) A significantly high proportion of ms-sRNA–associated genes are expressed. $^{**}P$ value < 2.2e-16, $\chi^2 = 475.51$.
(F) ms-sRNA–associated genes have a significantly higher average gene expression value. $^{**}P$ value < 2.2e-16, Mann-Whitney test.
(G) Seventy-five percent of the differentially expressed genes (dashed box) among the ms-sRNA-associated genes are up-regulated (red arrow, Q value < 0.05, $\log^2$ [fold change] > 1), as opposed to down-regulated (green arrow, Q value < 0.05, $\log^2$ [fold change] < −1).
(H) Twenty-seven percent of the differentially expressed genes (dashed box) among the meiocyte/leaf-shared sRNA–associated genes are up-regulated.
(I) Nineteen percent of differentially expressed genes (dashed box) are up-regulated genes among leaf sRNA-associated genes.
(J) Expression pattern of ms-sRNAs associated with the 149 up-regulated genes identified in **(G)**, in pollen, meiocytes, and leaves, divided into sRNAs from sense (fourth to sixth columns) and antisense strands (seventh to ninth columns) with a comparison to the expression pattern of the primary transcript in the same tissues (first to third columns). Left top legend for the mRNA data; left bottom legend for the sRNA data.
(K) Expression pattern of ms-sRNAs associated with the 49 down-regulated genes identified in **(G)**, in pollen, meiocytes, and leaves, divided into sRNAs from sense (fourth to sixth columns) and antisense strands (seventh to ninth columns) with a comparison to the expression pattern of the primary transcript in the same tissues (first to third columns). Left-top legend for the mRNA data; left-bottom legend for the sRNA data.

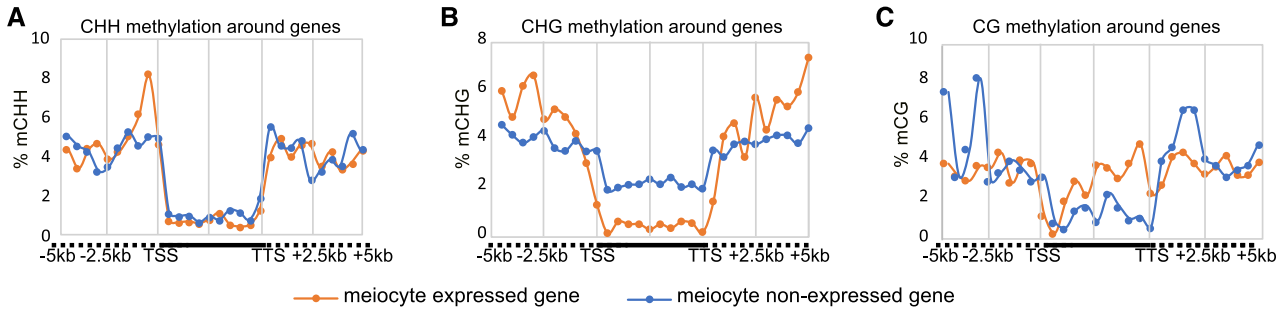

Figure 5. Distribution of CG, CHG, and CHH DNA Methylation Around Expressed Genes in Meiocytes.

(A) CHH methylation rate along ±5 kb regions of meiocyte expressed and nonexpressed genes.
(B) CHG methylation rate along ±5 kb regions of meiocyte expressed and nonexpressed genes.
(C) CG methylation rate along ±5 kb regions of meiocyte expressed and nonexpressed genes.

38 biological processes (BPs), six molecular functions (MFs), and 53 cellular components (CCs) are significantly enriched (false discovery rate *Q* value < 0.05; Supplemental Data 1). BP included two primary groups: One group is related to transcription/translation consisting of translational elongation (GO: 0006414, 5/145, *P* value = 1.7e-3), translation (GO: 0006412, 21/145, *P* value = 8.7e-9) and gene expression (GO: 0010467, 25/145, *P* value = 1.8e-4); the other group is related to stress response including response to oxidative stress (GO:0006979, 13/145, *P* value = 1.9e-4) and response to stress (GO:0006950, 34/145, *P* value = 1.2e-2; Figure 6A). The enriched MFs included translation elongation factor activity (GO: 0003746, 4/145, *P* value = 2.7e-2), ubiquitin protein ligase binding (GO: 0031625, 5/145, *P* value = 1.6e-2), and structural constituent of ribosome (GO: 0003735, 15/145, *P* value = 4.0e-6; Figure 6A). The CCs could be generally divided into three groups. The first group contains proteins related to translation: cytosolic ribosomes (GO: 0022626, 17/145, *P* value = 8.8e-10), ribosomal subunit (GO: 0044391, 16/145, *P* value = 6.1e-9), endoplasmic reticulum lumen (GO: 0005788, 5/145, *P* value = 1.3e-4), and the nucleolus (GO: 0005730, 14/145, *P* value = 1.2e-5). The second group is related to respiration of the mitochondrial respiratory chain complex I (GO: 0005747, 4/145, *P* value = 1.6e-2). The third group consists of protein-DNA complexes (GO: 0032993, 5/145, *P* value = 4.7e-3), chromatin (GO:0000785, 8/145, *P* value = 2.8e-4), and nucleosomes (GO: 0000786, 5/145, *P* value = 2.9e-4), which can be summarized as a chromatin-related group (Figure 6B) that includes genes required for maintaining chromatin structure during meiosis. We also compared our results with those of the 838 meiocyte-preferentially expressed genes as previously reported (Yang et al., 2011). However, only chromatin (GO:0000785, 13/694, *P* value = 1.7e-2) was enriched in meiocyte-preferentially expressed genes. Interestingly, we found that *RAD51*, one of the two essential recombinases required for meiotic recombination (Doutriaux et al., 1998), and *ARABIDOPSIS SKP1-LIKE1* (*ASK1*), a component of a putative E3 ubiquitin protein ligase required for normal homologous chromosome segregation (Yang et al., 1999), were both positively associated with ms-sRNAs (Figures 6C and 6D) This suggests that ms-sRNAs are correlated with gene expression in meiocytes (including the expression of certain genes involved in meiotic recombination), although the regulatory mechanism requires further studies.

Characteristics of AtSPO11-1–Dependent sRNAs

According to the DSB repair model, meiotic recombination starts with the programmed formation of SPO11-mediated DSBs (de Massy, 2013). Previous studies in plants, Drosophila, and human cells have shown that DSBs induced by mitotic DNA-damaging agents are able to generate novel endogenous sRNAs (Francia et al., 2012; Michalik et al., 2012; Wei et al., 2012). To test whether meiocyte sRNAs are also DSB-dependent, we sequenced the sRNAs from *spo11-1-1* meiocytes and compared them with wild-type meiocyte sRNAs. Only 4,756 sRNA clusters were obtained from the *spo11-1-1* meiocytes using the same criteria described for wild-type meiocyte and leaf (Table 1). Sixty-nine percent (4,668/6,727) of the wild-type meiocyte sRNA clusters are dependent on AtSPO11-1 (Figure 7A), which are absent from the *spo11-1-1* meiocytes. A comparison of the length of the sRNA clusters between the AtSPO11-1–dependent and AtSPO11-1–independent groups indicated that the median length of the AtSPO11-1–dependent clusters was significantly shorter than that of the AtSPO11-1–independent clusters (Student's *t* test, *P* value < 2.2e-16; Figure 7B; Supplemental Table 6), whereas the abundance of the AtSPO11-1–dependent sRNA clusters was similar to the meiocyte clusters in total (Figure 7C; Supplemental Table 6). AtSPO11-1–independent clusters are more abundant than either the AtSPO11-1–dependent or the meiocyte clusters (Student's *t* test, both *P* value < 2.2e-16; Figure 7C; Supplemental Table 6). We also analyzed the centromeric/chromosome arm distribution as well as different genomic feature distributions of the SPO11-1–dependent and –independent sRNA clusters. When compared with meiocyte sRNA clusters, AtSPO11-1–independent sRNA clusters have significantly more sRNAs on chromosome arms and intergenic regions and fewer sRNAs on TEs (Figures 7D and 7E; Supplemental Table 4). It indicates that AtSPO11-1–independent sRNA clusters are more similar to ms-sRNA clusters.

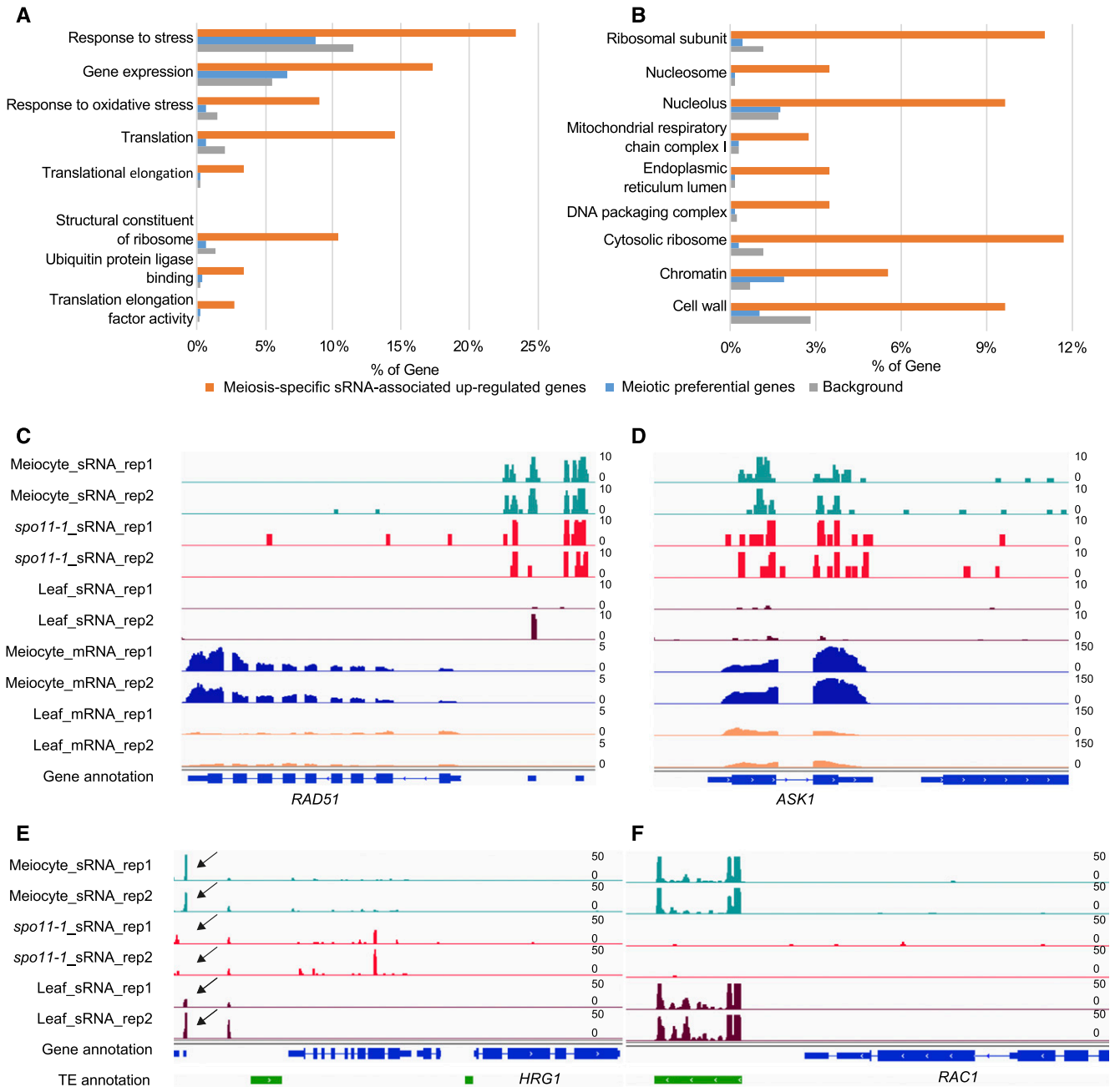

Figure 6. ms-sRNA–Associated Up-Regulated Genes in Meiocytes.

(A) Bar plots of the 149 ms-sRNA–associated up-regulated gene enriched GO terms from the categories of BP and MF.
(B) Bar plots of the 149 ms-sRNA–associated up-regulated gene enriched GO terms from the category of CC.
(C) Snapshot showing that meiocyte-specific and AtSPO11-1–independent sRNA clusters ahead of meiotic essential gene *RAD51*.
(D) Snapshot showing that meiocyte-specific and AtSPO11-1–independent sRNA clusters on meiotic essential gene *ASK1*.
(E) Snapshot showing AtSPO11-1–dependent sRNA clusters downstream of *HRG1*. Arrows indicate where the clusters are.
(F) Snapshot showing AtSPO11-1–dependent sRNA clusters downstream of *RAC1*.

The results described above show that ms-sRNAs are correlated with up-regulated genes in meiocytes. Similarly, among genes associated with both AtSPO11-1–dependent and –independent sRNA clusters, more genes were up-regulated than down-regulated (Supplemental Figure 3). We further conducted GO term-enrichment (Du et al., 2010) analyses of the up-regulated meiotic genes (Q value < 0.05, $\log^2$[fold change] > 1) from AtSPO11-1–dependent and –independent sRNA clusters and obtained 51 and 41 significantly enriched GO terms, respectively (false discovery rate Q value < 0.05; Supplemental Data 2). To

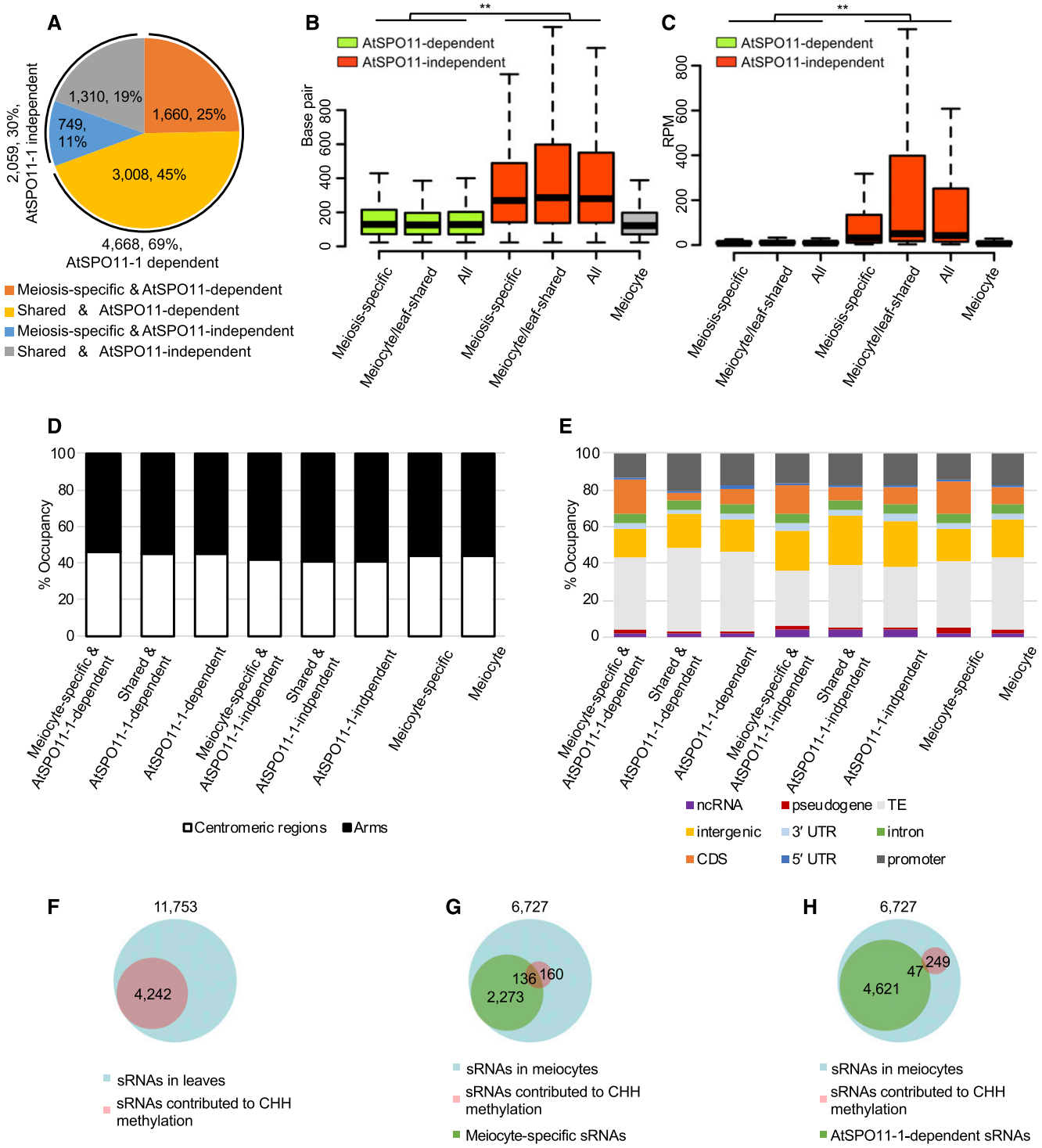

Figure 7. Characteristics of AtSPO11-1–Dependent sRNAs.

(A) Seventy percent of the meiocyte sRNAs are AtSPO11-1–dependent (orange and yellow).

(B) AtSPO11-1–dependent sRNA clusters are significantly shorter than the AtSPO11-1–independent and meiocyte clusters (**P value < 0.01). Student's t test was used to compare the length of AtSPO11-1–dependent (green bars), AtSPO11-1–independent (red bars) and meiocyte (gray bar) sRNA clusters (P value < 2.2e-16).

discriminate whether AtSPO11-1–dependent and –independent sRNA-associated genes are associated with different GO terms, we did cross-comparison and the enriched GO terms from the 149 ms-sRNA–associated meiocyte up-regulated genes were used as a reference. Interestingly, we found that chromatin-related terms like protein-DNA complex (GO: 0,032,993) and nucleosome (GO: 0,000,786) were only enriched in the AtSPO11-1–dependent sRNA-associated genes. By contrast, structural-ribosome components and structural-molecule–related GO terms, that is, ribonucleoprotein complex (GO: 1,990,904) and ribosome (GO: 0,005,840), were only enriched in the AtSPO11-1–independent sRNA-associated genes (Supplemental Data 2).

Because previous studies demonstrated that both DSBs and COs tend to take place at promoter or genic regions (Choi et al., 2013, 2018), we speculated that AtSPO11-1–dependent sRNA-associated genes are potential sites for the occurrence of meiotic recombination. For example, defense genes *HRG1* and *RAC1* overlap meiotic recombination hotspots and have AtSPO11-1–dependent sRNA clusters nearby (Figures 6E and 6F). On the other hand, we found that the ms-sRNAs positively correlated with *RAD51* and *ASK1* are AtSPO11-1–independent (Figures 6C and 6D). Together, these results suggest that AtSPO11-1–dependent and –independent sRNAs have different roles in meiocytes.

Furthermore, we divided meiocyte sRNA clusters into four groups: 1,660 (25%) AtSPO11-1–dependent ms-sRNAs; 749 (11%) AtSPO11-1–independent ms-sRNAs; 3,008 (45%) AtSPO11-1–dependent meiocyte/leaf-shared sRNAs; and 1,310 (19%) AtSPO11-1–independent meiocyte/leaf-shared sRNAs (Figure 7A). No significant differences in cluster length or sRNA abundance were observed within either the AtSPO11-1–dependent or –independent groups (Figures 7B and 7C). When assessing chromosome regions and genomic features, we observed AtSPO11-1–dependent ms-sRNA clusters have the highest CDS occupancy (18%) as compared with the AtSPO11-1–independent ms-sRNA clusters (15%) and the other two meiocyte/leaf-shared sRNA clusters (5% for AtSPO11-1–dependent meiocyte/leaf-shared sRNA clusters, and 7% for AtSPO11-1–independent meiocyte/leaf-shared sRNA clusters; Figure 7E). The AtSPO11-1–dependent meiocyte/leaf-shared sRNA clusters have the highest promoter occupancy (20%) within the four groups (Figure 7E).

Comparing our data to bisulfite sequencing (BS-Seq) data from Arabidopsis meiocytes (Walker et al., 2018) showed that 34% (6,195/18,447) and 45% (401/884) of hyper-DMRs in a CHH context correspond to 36% (4,242/11,753) and 4% (296/6,727) sRNA clusters from leaves and meiocytes, respectively (Figures 7F and 7G). Intriguingly, of the 296 meiocyte sRNA clusters associated with hyper-DMRs in a CHH context, 46% (136) and 54% (160) are meiocyte-specific and meiocyte/leaf-shared sRNA clusters, respectively (Figure 7G), whereas only 16% (47) are AtSPO11-1–dependent (Figure 7H). As DMRs in the CHH context are generally associated with the RdDM pathways, these results imply that AtSPO11-1–dependent sRNAs are distinct to other sRNA groups and might be RdDM-independent.

AtSPO11-1–Dependent Meiocyte sRNAs Show Enrichment at Meiotic-Recombination–Associated DNA Motifs

After meiotic DSB formation, SPO11 remains covalently bound to the 5′ ends of the cleavage site (Keeney et al., 1997). To facilitate DSB repair, SPO11 and a covalently linked oligonucleotide of 20–40 nucleotides (called "SPO11-oligonucleotides") is released by endonucleases, including SPORULATION IN THE ABSENCE OF SPO ELEVEN 2 and MRE11 (Puizina et al., 2004; Neale et al., 2005; Uanschou et al., 2007). To investigate the relationship between AtSPO11-1–dependent sRNAs and meiotic recombination, we compared AtSPO11-1–dependent sRNA clusters with a previously published map of AtSPO11-oligonucleotides with ~6,000 hotspots genome-wide (Choi et al., 2018). We failed to observe a correlation between these groups (Pearson's $r = 0.07$, P value < 0.00001; Figure 8A). Similarly, AtSPO11-1–independent sRNAs did not correlate with AtSPO11-oligonucleotides either (Pearson's $r = 0.03$, P value < 0.00001; Supplemental Figure 4). One possible explanation that reconciles the existence of AtSPO11-1–dependent sRNAs with an apparent lack of correlation with AtSPO11-oligonucleotides is that the genome-wide absence of AtSPO11-derived DSBs in the *spo11-1-1* mutant changes the chromatin accessibility landscape broadly rather than at just potential DSB sites.

As an alternative strategy for detecting an association between COs and AtSPO11-1–dependent sRNAs, we looked for a relationship with CO-associated sequence motifs. Previous studies have shown that CO hotspots in Arabidopsis are enriched for CTT repeats and A-rich motifs, which preferentially locate in genic regions and promoter regions, respectively (Choi et al., 2013; Wijnker et al., 2013; Shilo et al., 2015). The results presented above indicates that AtSPO11-1–dependent ms-sRNA clusters were enriched the most in genic regions (Figure 7E). To determine whether the AtSPO11-1–dependent ms-sRNA clusters correlate with CO hotspots, we compared 407 gene-region–overlapping sRNA clusters from AtSPO11-1–dependent ms-sRNA clusters and searched for conserved DNA motifs. The hotspot-associated CTT repeat motif was the most significant result (E value = 2.7e-

Figure 7. (continued).

(C) AtSPO11-1–dependent sRNA clusters have significantly lower sRNA abundance than AtSPO11-1–independent clusters (**P value < 0.01). Student's t test was used to compare RPM of AtSPO11-1–dependent (green bars) and AtSPO11-1–independent (red bars) and meiocyte (gray bar) sRNA clusters (P value < 2.2e-16).

(D) The distribution of the eight different sRNA groups between arms (black) and centromeric regions (white).

(E) The distribution of the eight different sRNA groups among genomic features.

(F) Approximately 36% of the sRNAs in leaves are associated with hyper-DMRs in a CHH context.

(G) Approximately 46% of the meiocyte sRNA-associated hyper-DMRs in a CHH context are associated with ms-sRNAs.

(H) Only 16% of the meiocyte sRNA-associated hyper-DMRs in a CHH context are associated with AtSPO11-1–dependent sRNAs.

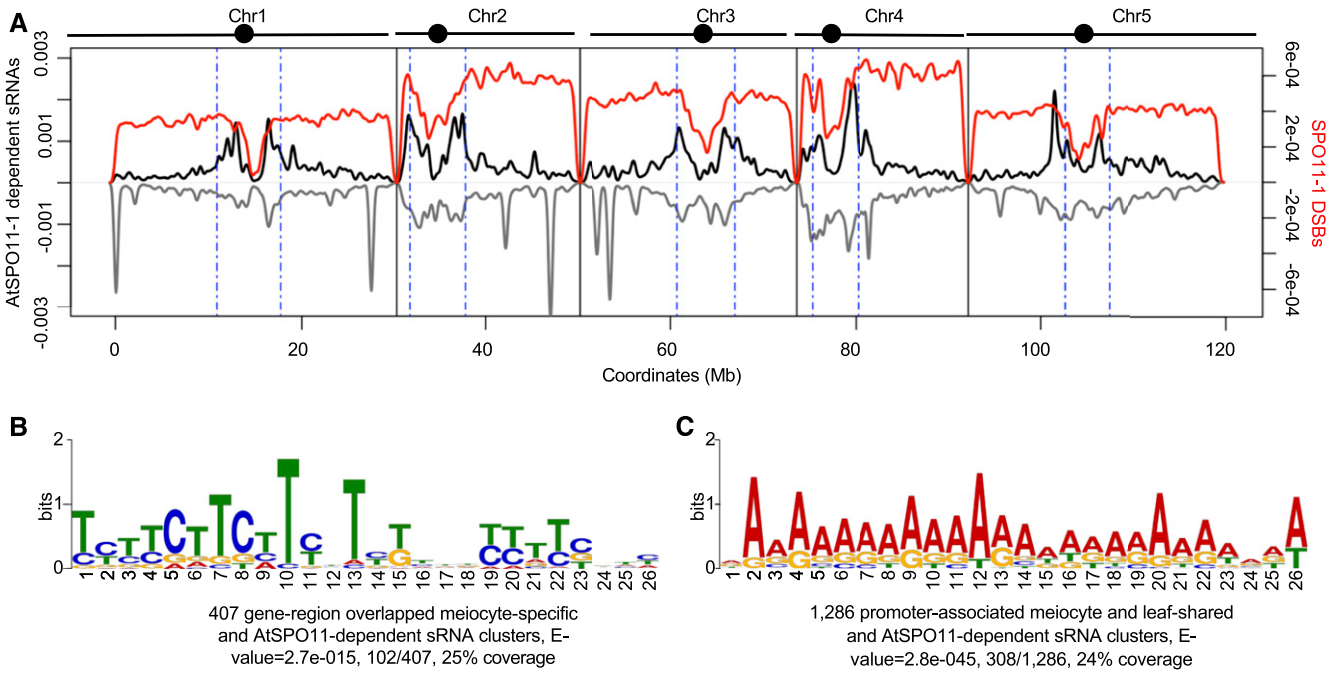

Figure 8. AtSPO11-1–Dependent sRNAs Show Enrichment at Meiotic CO-Associated DNA Motifs.

(A) Density plot of AtSPO11-1–dependent sRNAs and AtSPO11-1 oligos (Choi et al., 2018) on the five Arabidopsis chromosomes (black bars with circles marking centromeres). sRNA density on the sense strand shows a positive value (plotted in black), while sRNA density on the antisense strand shows a negative value (plotted in gray) per 10 kb. Chromosomes are portioned into arm regions and centromeric regions (green dashed lines). The sRNA density is compared to the density of AtSPO11-1 oligos per 10 kb (red curve).
(B) Meiotic recombination-associated CTT repeat motif is enriched with meiocyte-specific and AtSPO11-1–dependent sRNAs from gene body regions.
(C) Meiotic recombination-associated A-rich motif is enriched with meiocyte/leaf-shared and AtSPO11-1–dependent sRNAs from promoter regions.

015), with 25% coverage (Figure 8B). By contrast, no CTT repeat motifs were observed from the 194 gene-associated AtSPO11-1–independent ms-sRNA clusters, the gene-associated sRNA clusters from the meiocyte/leaf-shared group, or the leaf sRNA group. Additionally, we also identified two extended A-rich motifs (*E* value = 2.8e-045 and 1.2e-014) in 32% of the meiocyte/leaf-shared and the AtSPO11-1–dependent sRNA clusters that are located in gene promoter regions (Figure 8C; Supplemental Figure 5). Taken together, these results revealed that AtSPO11-1–dependent meiocyte sRNA clusters are associated with known CO hotspot motifs in promoter and genic regions.

AtSPO11-1–Dependent Meiocyte sRNAs Are Associated with Defense Gene-Related CO Hotspots

Previous studies showed that a number of meiotic recombination hotspots are associated with nucleotide-binding–site leucine-rich repeat (NBS-LRR) genes (Choi et al., 2016, 2018). Additionally, 73 of the 197 NBS-LRR genes in the Arabidopsis genome are close to DSB hotspots associated with transposons (Choi et al., 2018). Our RNA-Seq data showed that 124, and 85, of the 197 NBS-LRRs are expressed in leaves and meiocytes, respectively. Among the 85 meiocyte-expressed NBS-LRR genes, 79 are down-regulated in meiocytes compared with leaves. By comparing NBS-LRR genes with AtSPO11-1–dependent meiocyte sRNAs, 36 of 197 NBS-LRRs are within 2 kb of meiocyte sRNA clusters and 26 NBS-LRRs had both nearby DSB hotspots and AtSPO11-1–dependent meiocyte sRNA clusters. For example, defense genes overlapping meiotic recombination hotspots, such as *HRG1* and *RAC1*, have AtSPO11-1–dependent sRNA clusters at upstream or downstream of the genes (Figures 6E and 6F). In both cases, the sRNA clusters are from a DNA/MuDR or RC/Helitron DNA transposon, both of which correspond to DSB hotspots.

We also asked whether AtSPO11-1–dependent sRNA clusters are associated with NBS-LRR loci on a larger scale. We compared the distribution of sRNA clusters from 5 kb upstream of TSSs to 5 kb downstream of TTSs for each NBS-LRR gene in meiocytes (wild type and *spo11-1-1*) and leaves. Thirty percent (59/197) of NBS-LRR genes have at least one sRNA cluster ±5 kb of the gene locus. Unlike the results described above indicating that sRNA clusters tend to associate with genic regions, NBS-LRR–associated sRNA clusters were more likely to occur outside of genes (Figure 9A). In addition, we found sRNA-enriched peaks around 1–1.5 kb behind the TTS preferentially in AtSPO11-1–dependent meiocyte sRNA clusters when compared with sRNA cluster distributions in leaves or *spo11-1-1* (Figure 9A). To avoid the bias caused by different data set sizes, we randomly sampled 1,500 sRNA clusters from each data set and mapped them proximal to NBS-LRRs, with 1,000 replications. The distribution of the sRNA clusters from each category was rarely altered (Figure 9B). To validate the specific sRNA distribution patterns proximal to NBS-LRRs, we searched for

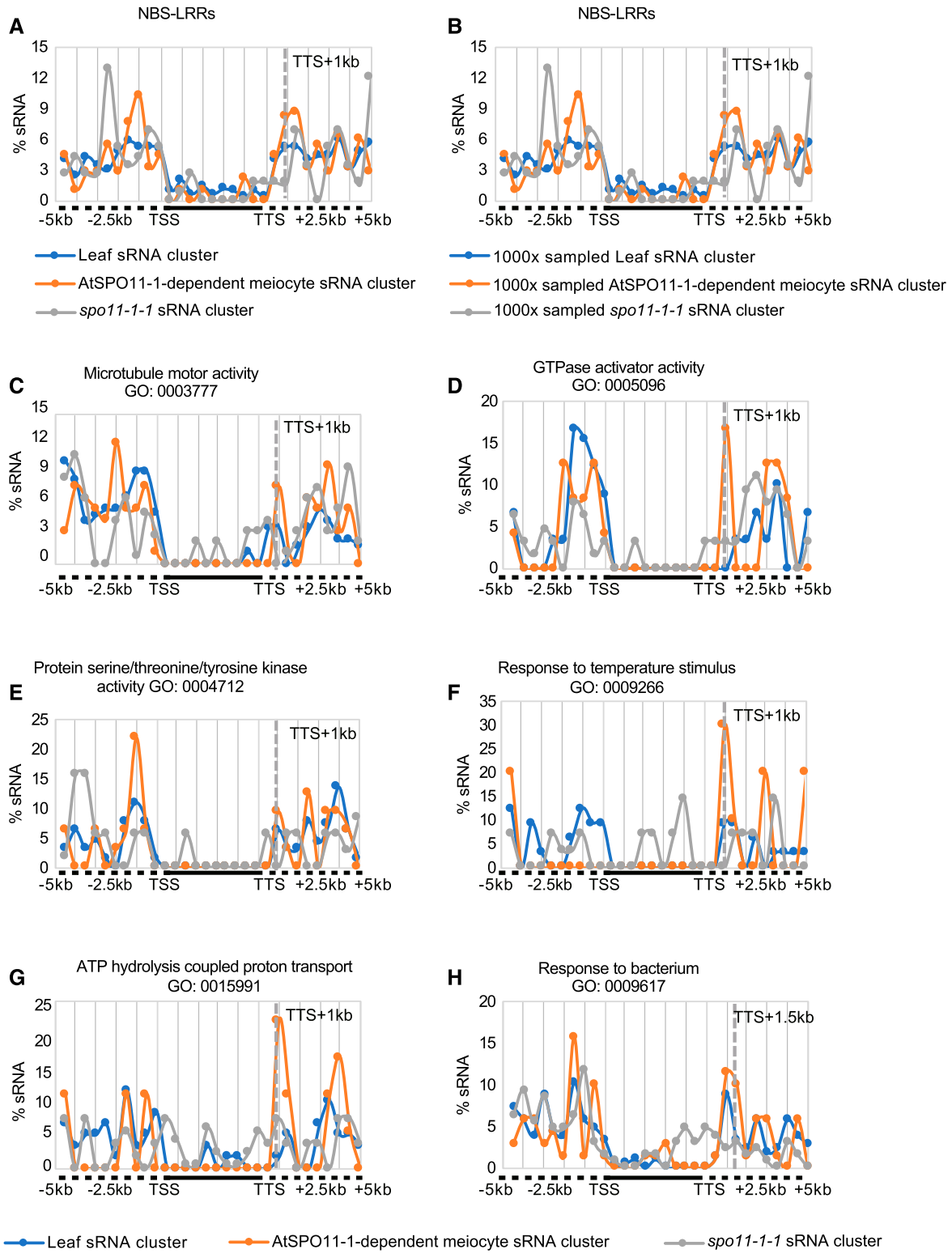

Figure 9. AtSPO11-1–Dependent Downstream sRNA Peaks Are Associated with NBS-LRRs and Six Other GO Terms.

(A) Distribution of leaf (blue), AtSPO11-1–dependent meiocyte (orange), and *spo11-1-1* (gray) sRNA clusters relative to 5-kb regions upstream and downstream (dotted lines) of 197 NBS-LRR gene bodies (black line) flanked by TSS and TTS shows an AtSPO11-1–dependent sRNA peak at 1 kb downstream of TTS.

similar patterns among 1,453 GO terms with 10 or more genes (from 6,813 GO terms annotated in TAIR10; see details in "Methods"). Twelve GO terms were found to have sRNA distribution patterns similar to that of the NBS-LRR genes, including microtubule motor activity (GO: 0,003,777), GTPase activator activity (GO: 0,005,096), protein Ser/Thr/Tyr kinase activity (GO: 0,004,712), response to temperature stimulus (GO: 0,009,266), ATP hydrolysis coupled proton transport (GO: 0,015,991), response to bacterium (GO: 0,009,617), and others (Figures 9C to 9H; Supplemental Figure 6). To determine whether the enrichment of downstream sRNA clusters implied *cis* regulation by potentially conserved DNA motifs, we used the software MEME (Bailey and Elkan, 1994) to search for shared DNA sequences among the DNA sequences of sRNA clusters that enriched downstream of interrogated GO term genes; no motifs reached statistical significance. Whereas ATGTCTCGYCSCC was found in 50% of the downstream regions from the genes coming from all 12 GO terms and the NBS-LRR genes (by MEME, *E* value = 4.3e-001, 10/20; Supplemental Figure 7), which is similar to the reported DNA CO motif CCN repeat (Shilo et al., 2015), it suggests that genes from the 12 GO terms might be close to potential CO hotspots due to the presence of this CCN-repeat–like DNA motif.

DISCUSSION

Distinct Features of Meiocyte-Specific and AtSPO11-1–Dependent sRNAs in Arabidopsis

We surveyed the sRNA population of Arabidopsis meiocytes using microcapillary isolation followed by deep sequencing. We found that the meiocyte sRNA population was distinct from the siRNA population in leaves in several aspects.

First, the genome-wide sRNA distributions were different (Figure 2; Supplemental Figure 1). In contrast to preferential association of leaf siRNAs with centromeric regions (Figure 2D; Supplemental Figure 1A; Zilberman et al., 2007), ms-sRNAs are associated with chromosome arms and coding sequences (Figures 2D and 2E; Supplemental Figure 1), similar to the recent findings in maize and sunflower (*Helianthus annuus*) meiocytes (Dukowic-Schulze et al., 2016; Flórez-Zapata et al., 2016), male and female germline cells of mice (Song et al., 2011; Stein et al., 2015), and *Caenorhabiditis elegans* (Ruby et al., 2006; Gu et al., 2009). The phenomenon observed in *C. elegans* that major endogenous sRNAs are complementary to mRNAs (Ruby et al., 2006; Gu et al., 2009), is similar to our findings (Figure 4J). Moreover, the significant difference in cluster length between leaf sRNAs and meiocyte sRNAs may reflect the fact that the former are more strongly correlated with repetitive sequences and TEs, while the latter are correlated with genes, suggesting that they are generated from different templates or genomic sites (Figure 1C).

Second, we found a significant positive correlation between gene-body–associated ms-sRNAs and increased gene expression in meiocytes (Figures 4F and G), but not for either leaf (Figures 4F and 4I) or meiocyte/leaf-shared sRNAs (Figures 4F and H); this distinct property of ms-sRNAs suggests that they may have a different role in fine tuning gene expression compared with other sRNAs (Matzke and Mosher, 2014). One possible explanation is that ms-sRNAs may function similarly to epigenetically activated siRNAs, which are generated from reactivated TEs (Creasey et al., 2014; He et al., 2015). ms-sRNAs may be the consequence of the up-regulation of meiotic genes that need to be subsequently suppressed in the next developmental stage (i.e. pollen). This hypothesis is further supported by the discovery that these ms-sRNAs display a strong temporal expression pattern (Figure 4K).

Third, we defined AtSPO11-1–dependent sRNAs in meiocytes, and found they are distinct from the AtSPO11-1–independent sRNAs in cluster length (Figure 7B) and abundance (Figure 7C). Also, because >60% of the AtSPO11-1–dependent sRNAs were meiocyte/leaf-shared sRNAs (Figure 7A), they are different from ms-sRNAs in terms of association with DNA methylation (Figures 7G and 7H). It is possible that AtSPO11-1 has a role in mitotic cells but this function is partially redundant with that of AtSPO11-3 (Stacey et al., 2006), which is specifically required for mitosis. This may potentially explain why SPO11-1 has an effect on leaf sRNAs.

Fourth, AtSPO11-1–dependent ms-sRNA clusters and the AtSPO11-1–dependent meiocyte/leaf-shared sRNA clusters are enriched at CTT repeat and A-rich meiotic CO-associated DNA motifs (Choi et al., 2013; Wijnker et al., 2013; Shilo et al., 2015), respectively (Figures 8B and 8C). No similar enrichment was found for loci of the AtSPO11-1–independent sRNAs. Moreover, the sequences from the sRNA peaks in meiocytes downstream of CO

Figure 9. (continued).

(B) Random sampling (1,000 times) of 1,500 leaf, AtSPO11-1–dependent meiocyte, and *spo11-1-1* sRNA clusters relative to 197 NBS-LRRs ±5 kb of flanking sequence.

(C) Distribution of leaf, AtSPO11-1–dependent meiocyte, and *spo11-1-1* sRNA clusters relative to genes ±5 kb of flanking sequence with the GO term microtubule motor activity (GO: 0,003,777) shows an AtSPO11-1–dependent sRNA peak at 1 kb downstream of TTS.

(D) Distribution of leaf, AtSPO11-1–dependent meiocyte, and *spo11-1-1* sRNA clusters relative to genes ±5 kb of flanking sequence with the GO term GTPase activator activity (GO: 0,005,096) shows an AtSPO11-1–dependent sRNA peak at 1 kb downstream of TTS.

(E) Distribution of leaf, AtSPO11-1–dependent meiocyte, and *spo11-1-1* sRNA clusters relative to genes ±5 kb of flanking sequence with the GO term protein Ser/Thr/Tyr kinase activity (GO: 0,004,712) shows an AtSPO11-1–dependent sRNA peak at 1 kb downstream of TTS.

(F) Distribution of leaf, AtSPO11-1–dependent meiocyte, and *spo11-1-1* sRNA clusters relative to genes ±5 kb of flanking sequence with the GO term response to temperature stimulus (GO: 0,009,266) shows an AtSPO11-1–dependent sRNA peak at 1 kb downstream of TTS.

(G) Distribution of leaf, AtSPO11-1–dependent meiocyte, and *spo11-1-1* sRNA clusters relative to genes ±5 kb of flanking sequence with the GO term ATP hydrolysis coupled proton transport (GO: 0,015,991) shows an AtSPO11-1–dependent sRNA peak at 1 and 1.5 kb downstream of TTS.

(H) Distribution of leaf, AtSPO11-1–dependent meiocyte, and *spo11-1-1* sRNA clusters relative to genes ±5 kb of flanking sequence with the GO term response to bacterium (GO: 0,009,617) shows an AtSPO11-1–dependent sRNA peak at 1.5 kb downstream of TTS.

hotspots-associated NBS-LRR genes also exhibit enrichment for a similar CCN-repeat motif (Supplemental Figure 7). Studies in somatic cells have already demonstrated a role for sRNAs in homology-dependent DNA repair in plant seedlings, Drosophila, and human cells (Francia et al., 2012; Michalik et al., 2012; Wei et al., 2012). Based on these findings and our observations, we propose that AtSPO11-1–dependent sRNAs tend to be associated with the open chromatin structure, which might favor meiotic recombination hotspots, but that further work needs to be done to elucidate the mechanistic relationship between the motifs, the AtSPO11-1–dependent sRNAs, and the regulation of meiotic recombination.

A Proposed Model for Meiocyte sRNAs in Meiotic Gene Expression and Recombination

Based on our results and recent studies, we propose a model for the role of meiocyte sRNAs in meiotic gene expression and recombination (Figure 10). For convenience, we used the leaf siRNAs as a reference (Figure 10, top) to better illustrate the commonalities and differences regarding meiocyte sRNAs (Figure 10, bottom). Somatic siRNAs suppress TE activity and silence genes in euchromatin or heterochromatin (Matzke and Mosher, 2014). In our model, we differentiate genes and TEs that are found in euchromatic and heterochromatic regions. In leaves, all TEs are suppressed by siRNA-guided DNA methylation and reinforced via DNA methylation in all three contexts in both euchromatin and heterochromatin, and gene silencing is also associated with distinct epigenetic markers, including CG and CHG DNA methylation. However, promoter-targeted siRNAs alone are insufficient for silencing. In fact, our results show that sRNA enrichment 500-bp upstream of a TSS is correlated with gene expression both in leaves and meiocytes (Figures 4Ato 4C), which is consistent with the correlation between tissue-specific CHH islands upstream of TSSs and elevated gene expression (Hsu et al., 2017). By contrast, approximately two-thirds of TEs from heterochromatin lack sRNA clusters in meiocytes (Table 1; Figures 2D and 2E), similar to a recent study showing that CHH methylation is significantly decreased in meiocytes compared with other tissues (Walker et al., 2018). However, primary CG methylation was unchanged (Walker et al., 2018), as was histone modification, thus preserving TE silencing. In euchromatin, some TEs such as Helitron and MuDR transposons are CO hotspots (Choi et al., 2018) and are associated with AtSPO11-1–dependent sRNAs (Figures 6E and 6F). Similar to previous studies that discovered an association between NBS-LRR genes and meiotic CO hotspots (Choi et al., 2016, 2018), our data detected 36 NBS-LRRs within 2 kb of

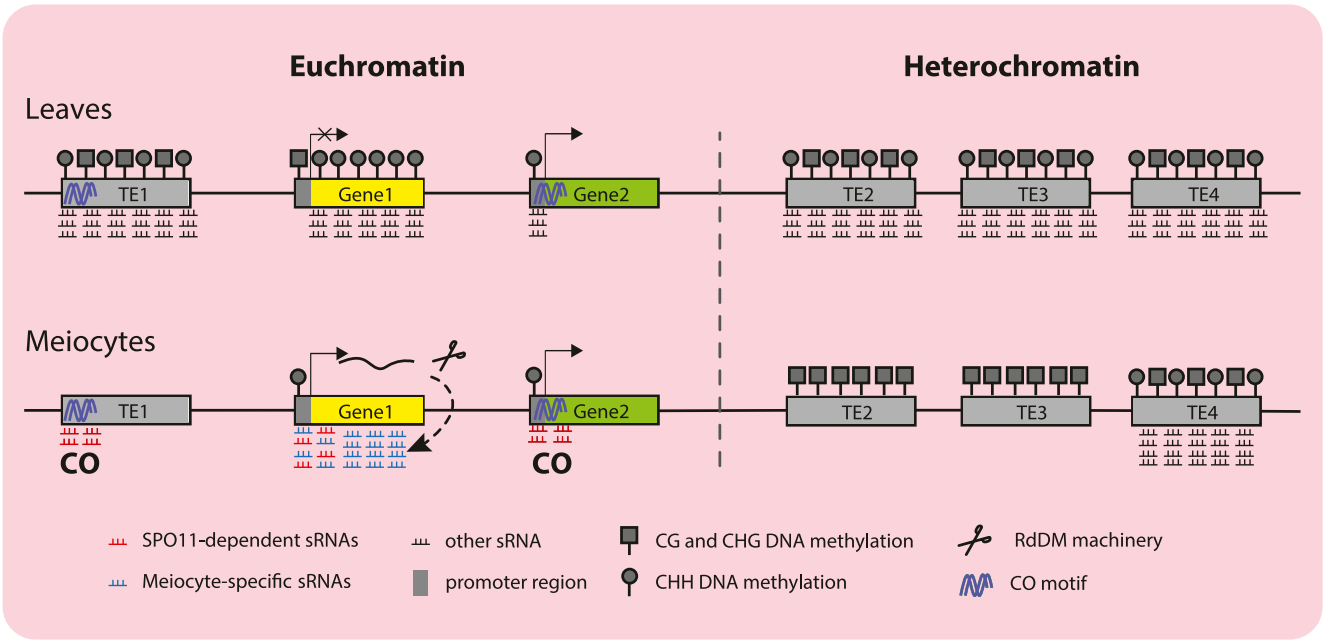

Figure 10. A Model for Meiocyte sRNA Function in Gene Expression and Meiotic Recombination.

In leaves (top), TE1, TE2, TE3, and TE4 are predicted to be suppressed by siRNA-guided DNA methylation and reinforced via DNA methylation in all three contexts (CG and CHG, gray squares; and CHH, gray circles) in both euchromatin (left of dashed line) and heterochromatin (right of dashed line). Gene silencing (Gene1) is achieved by promoter-associated CG and CHG DNA methylation as well as siRNA-guided DNA methylation and is reinforced by gene body DNA CHH methylation. Expressed genes in both meiocytes and leaves (Gene2) are associated with sRNAs and CHH methylation upstream of the TSS. In meiocytes (bottom), many TEs (TE1, TE2, and TE3) are no longer associated with sRNAs. However, in heterochromatin (right), TE CG methylation (and histone methylation) is unchanged, and silencing is preserved. In euchromatin, some TEs (TE1) are potential meiotic recombination hotspots, which are associated with AtSPO11-1–dependent sRNAs. Genes that are up-regulated in meiocytes (Gene1) generate gene-body–associated sRNAs via the RNA processing machinery (scissors). Generation of gene-body–associated sRNAs, in turn, establishes promoter CHH methylation. COs tend to occur near transcribed genes, especially in their promoter regions (Gene2). DNA motifs associated with CO hotspots are also associated with AtSPO11-1–dependent sRNAs.

meiocyte sRNAs, and ~30 of these are found near CO hotspots, with associated transposons (Choi et al., 2018). In our model, genes that are up-regulated in meiocytes generate gene-body–associated sRNAs via the RNA processing machinery. Generation of gene-body–associated sRNAs, in turn, establishes CHH methylation in promoter regions. This idea is supported by the fact that the location of previously identified CO sites (Shilo et al., 2015) are correlated with the location of the genes expressed in meiocytes that we identified, indicating that COs tend to occur on transcribed gene bodies or promoter regions (χ^2 = 11.28, *P* value = 3.5e-3). Given that AtSPO11-1–dependent sRNAs are enriched with CO-associated DNA motifs in the gene body or promoter regions, we hypothesize that chromatin features around these sites are different between leaves and meiocytes. In leaves, these chromatin features can produce canonical siRNAs, whereas in meiocytes, AtSPO11-1–dependent sRNAs are associated with active genes that lack silencing-associated DNA methylation and meiotic CO hotspot motifs.

Unlike promoter-associated CG and CHG DNA methylation that suppresses gene expression, siRNAs and CHH islands (Hsu et al., 2017) ahead of TSSs show a positive correlation with gene expression in both leaves and meiocytes (Figure 10). We hypothesize that this phenomenon is similar to RNA-induced epigenetic gene activation (RNAa), in which sRNAs induce targeted gene expression, as first discovered in human cells (Li et al., 2006) and then in plants and *C. elegans* (Shibuya et al., 2009; Seth et al., 2013). The underlying mechanism of RNAa action at promoter regions involves the assembly of an RNA-induced-transcriptional activation complex, which interacts with RNA polymerase II to stimulate transcription (Portnoy et al., 2016). It is possible that RNA activation is a general and conserved mechanism involved in the regulation of gene expression. However, no RNAa in Arabidopsis has been reported. These results extend our understanding of how meiocyte sRNAs relate to meiotic gene expression and meiotic recombination.

METHODS

Plant Materials

The strains used in this study were Arabidopsis (*Arabidopsis thaliana*) *spo11-1-1* (Grelon et al., 2001), T-DNA insertional line SALK_128428 as the *poliv* mutant, SALK_064627 as *dcl2-1*, SALK_005512 as *dcl3-1*, GABI_160G05 as *dcl4-2t*, and wild-type Columbia (Col-0). Plants were grown in soil at 22°C with 16-h light (bulb type: Philips TLD 36W/865, with eight tubes) and 8-h dark.

sRNA-Seq and Analysis

Two biological replicates were taken for each tissue. The meiocyte collection procedure was described in Wang et al. (2014). In brief, 0.3- to 0.4-mm buds (stages 7 to 9) that were undergoing meiosis were excised from the inflorescence and then dissected under a dissection microscope to dissociate the light-green–colored anthers with needles (BD 27 G × 1/2′′). Anthers were transferred to a slide loaded with 20 μL of RNase-free H_2O containing Recombinant RNase inhibitor (Takara), which has a working concentration of 4U/μL. After collecting 20 to 40 anthers, we used 4.5-inch straight-end watchmaker forceps with sharp tips to release the worm-like mass of meiocytes (developing pollen sac). Approximately 300 meiocyte masses (equal to 12,000 to 15,000 cells) from anthers of ~10 to 15 plants were collected for each sample by mouth pipette using a micropipette connected to a home-assembled micro-capture system (Wang et al., 2014), transferred to a RNase free 2.0-mL tube containing ZR Plant RNA Miniprep Kit beads (Zymo Research), and immediately placed in liquid N. Two biological replicates were collected from plants in the same flat. The size of the meiocyte mass may vary in different genotypes. However, for the mutants we used, the meiocyte mass size is similar to that in wild type. Total RNA was extracted using Trizol (Thermo Fisher Scientific) using the standard protocol with the exception that samples were frozen in liquid N and thawed in a 37°C water bath at least four times for better homogeneity. The aqueous phase was transferred into a new tube, and 4 μL Dr. GenTLE Precipitation Carrier (Takara) per 400 μL volume was added. sRNA libraries were constructed using the TruSeq Small RNA Library Preparation Kit (Illumina), all starting with 1 μg total RNA. sRNA sequencing was performed via Illumina Hiseq 2000 with at least 20 million 1 × 50 single-end reads for each sample. Raw sequencing data were first trimmed using BBMap (version 35.85, Bushnell B., https://sourceforge.net/projects/bbmap/) to remove adapters and low-quality reads. Preprocessed data were mapped using Bowtie (1.1.2; Langmead et al., 2009) to discard all reads identical to known structural RNAs, including transfer RNAs, ribosomal RNAs, small nucleolar RNAs, small nuclear RNAs, micro RNAs, and the other RNAs listed in the Rfam database (Nawrocki et al., 2015). Remaining reads were then mapped and clustered against the Arabidopsis genome (TAIR10 release), with 0 mismatch using ShortStack 3.3 with the option "-mincov 1rpm -pad 75" (Axtell, 2013). An sRNA cluster was defined as reported in Axtell (2013). Briefly, filtered sRNAs were aligned onto the TAIR10 Arabidopsis genome allowing 0 mismatch. Continuously aligned regions with >3 RPM sRNAs were defined as "islands." Each island was extended 75 bp both upstream and downstream. The regions containing the islands were defined as "sRNA clusters." Statistical analyses via Mann-Whitney tests, box plots, chromosome distribution plots, and heatmap plotting were performed using Rstudio v0.98.501. The centromeric regions of Arabidopsis were as defined in Arabidopsis Genome Initiative (2000). Pie charts, histograms, and bar plots were all conducted in Microsoft Excel 2017.

Data correlations were calculated from BAM files using multiBamSummary in deepTools 2.5.2 (Ramírez et al., 2016). Because there were fewer ms-sRNAs and meiocyte/leaf-shared sRNAs compared with leaf sRNAs, we used the "multiBamSummary BED-file" to specify regions for conducting Pearson's product-moment correlation analyses.

Transcriptome Sequencing Analysis

Sample collection and total RNA extraction were performed as described above. RNA-Seq libraries were constructed using a TruSeq RNA Library Preparation Kit (Illumina) with 1-μg total RNA. Sequencing was performed via Illumina Hiseq 2000/3000 with at least 20 million of 2 × 100 pair-end reads for each sample. Adapter trimming was conducted using BBMap (version 35.85, Bushnell B., sourceforge.net/projects/bbmap/). The whole genome sequence and annotation were downloaded from TAIR10 (www.Arabidopsis.org/). Clean reads were mapped using TopHat2 (Trapnell et al., 2012) for further analysis.

Coupled sRNA and sRNA-Targeted Gene Expression Analysis and GO Analysis

Both differential gene expression and sRNA variation by gene loci were conducted using Cuffdiff 2.1.1 (Trapnell et al., 2012) with the same criteria of $\log^2$ fold change ≥ 1 or ≤ -1, *Q* value ≤ 0.05. The promoter region was defined as 500 bp ahead of the TSS of each gene. GO analysis was performed in the software R with the package clusterProfiler (Yu et al., 2012). All GO term enrichment results were simplified with the command "simplify" (x, cutoff = 0.7, by = "p.adjust", select_fun = min). Illustrations of the two meiotic gene loci *RAD51* and *ASK1* were plotted by the Integrative Genomics Viewer (IGV 2.3.68).

Distribution Analyses of Meiocyte sRNAs around Genes

The latest GO annotations, containing 6,813 GO terms in total (Berardini et al., 2004), were downloaded from TAIR (https://www.Arabidopsis.org/). To reach reliable association between ms-sRNAs and their potentially related gene functions, only 1,453 well-defined GO terms including 10 or more genes were used to calculate the sRNA cluster hit frequency per bin arbitrarily. Regions ±5 kb upstream or downstream of a gene were divided into sequential 500-bp bins; for gene regions, one bin was defined as one-tenth of the gene length from TSS to TTS. To find the GO term with "NBS-LRRs like sRNA cluster distribution pattern," we excluded the GO terms with <20 bins and composed a script looking for GO terms, which were: (1) for bins within genic regions, frequency differences between leaves and meiocytes were <50% of the frequency in leaves; (2) for non-genic regions, frequency difference between leaves and meiocytes at bins 21 to 24 (500 to 2,000 bp) were >50% of the frequency in leaves.

Meiotic Recombination-Associated DNA Motif Discovery

The sequences of sRNA clusters in each group were analyzed using MEME (4.11.1; Bailey and Elkan, 1994) with the default parameters and the following exceptions: -maxsize 1,000,000 -nmotifs 5 -minw 5 -maxw 30 -revcomp -minsites 20, to discover the enriched DNA motifs. Searching for known motifs with a newly discovered DNA motif was done by Tomtom (Gupta et al., 2007).

BS-Seq Data Analysis

We used public BS-Seq data from meiocyte and rosette leaves found in the Gene Expression Omnibus database with accession numbers GSE86583 (Walker et al., 2018) and GSE51304 (Stroud et al., 2014). We first mapped BS-Seq reads to the Arabidopsis reference genome using BSMAPv2.90 (Xi and Li, 2009), keeping only uniquely mapped reads. Only reads with <4 mismatches per 100-bp length were kept for further analysis. Methylation levels were estimated as weighted methylation levels (Schultz et al., 2012). DMRs were defined as in Ausin et al. (2016).

Accession Numbers

sRNA raw data were deposited at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under accession numbers SRR5209214 to SRR5209219. The pollen grain sRNA data for SRR1986037 were retrieved from the NCBI SRA (Martínez et al., 2016). SPO11-1-oligonucleotide data were retrieved from the ArrayExpress repository (https://www.ebi.ac.uk/arrayexpress/experiments/), hosted by the European Bioinformatics Institute, European Molecular Biology Laboratory, under accession number E-MTAB-5041. (Additional detailed information is presented in Supplemental Table 1.) Raw data of the transcriptome were deposited at the NCBI SRA under accession numbers SRR5209210 to SRR5209214. The pollen grain RNA-Seq data for SRR847501 and SRR847502 were retrieved from the NCBI SRA (Loraine et al., 2013; additional detailed information is presented in Supplemental Table 1). The datasets supporting the conclusions of this article are available in the NCBI SRA repository under the project name PRJNA510650 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA510650).

Supplemental Data

Supplemental Figure 1. Distributions of ms-sRNA clusters with a 100-bp designated cluster length.

Supplemental Figure 2. Correlation analyses of all mRNA-seq samples.

Supplemental Figure 3. AtSPO11-1–dependent/independent sRNAs-associated genes and their gene expression pattern among different tissues.

Supplemental Figure 4. Distribution of AtSPO11-1-independent sRNAs and AtSPO11-oligonucleotides in the arabidopsis genome.

Supplemental Figure 5. Meiocyte/leaf-shared AtSPO11-1-dependent meiocyte sRNA clusters are enriched for a short version of poly-A DNA motifs at promoter regions.

Supplemental Figure 6. sRNA distribution from the Six GO terms with the NBS-LRR gene-like sRNA distribution pattern.

Supplemental Figure 7. A meiotic recombination-associated CCN-repeat–like motif was enriched by downstream sRNAs from the 12 NBS-LRR like GO term genes.

Supplemental Table 1. sRNA- and mRNA-seq statistics.

Supplemental Table 2. Pearson correlation coefficient among all sRNA-seq samples.

Supplemental Table 3. sRNA cluster overlapping test showing one-third of sRNA clusters in meiocytes are meiocyte-specific.

Supplemental Table 4. χ^2 test table.

Supplemental Table 5. Pearson correlation coefficient among all mRNA-seq samples.

Supplemental Table 6. Features of the AtSPO11-1-dependent and -independent sRNA clusters.

Supplemental Data Set 1. GO terms from the 149 ms-sRNA–associated up-regulated genes.

Supplemental Data Set 2. Comparison of enriched GO terms from meiocyte up-regulated genes associated with AtSPO11-1-dependent sRNA clusters, AtSPO11-1-independent sRNA clusters, and ms-sRNA clusters, respectively.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (31570314, 31870293, and 31600246), Rijk Zwaan Zaadteelt en Zaadhandel B.V., and funds from Fudan University and The University of North Carolina at Chapel Hill.

We thank Dr. Jennifer L. Modliszewski from the Department of Biology, University of North Carolina at Chapel Hill for revising the article and Dr. Guodong Ren at Fudan University, Dr. Jixian Zhai from the Southern University of Science and Technology, and Dr. Chenjiang You from the University of California, Riverside for providing expert advice. We also thank Dr. Yijun Qi at Tsinghua University for providing us with the seeds of *pol iv* and *dcl2//3/4* mutants.

AUTHOR CONTRIBUTIONS

J.H., Y.W., C.W., and G.P.C. conceived and designed the experiments; J.H. analyzed the data; J.H. and C.W. performed most of the experiments; J.H. wrote the article with contributions from all the authors; Y.W., H.M., and G.C. supervised and supplemented the writing.

Received July 5, 2018; revised December 21, 2018; accepted January 18, 2019; published January 23, 2019.

REFERENCES

Acquaviva, L., Székvölgyi, L., Dichtl, B., Dichtl, B.S., de La Roche Saint André, C., Nicolas, A., and Géli, V. (2013). The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. Science **339:** 215–218.

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815.

Ausin, I., et al. (2016) DNA methylome of the 20-gigabase Norway spruce genome. Proc. Natl. Acad. Sci. USA **113:** E8106–E8113.

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

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. **2:** 28–36.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., and de Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science **327:** 836–840.

Berardini, T.Z., et al. (2004) Functional annotation of the *Arabidopsis* genome using controlled vocabularies. Plant Physiol. **135:** 745–755.

Bishop, D.K., Park, D., Xu, L., and Kleckner, N. (1992). *DMC1*: A meiosis-specific yeast homolog of *E. coli recA* required for recombination, synaptonemal complex formation, and cell cycle progression. Cell **69:** 439–456.

Borde, V., Robine, N., Lin, W., Bonfils, S., Géli, V., and Nicolas, A. (2009). Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. EMBO J. **28:** 99–111.

Borges, F., and Martienssen, R.A. (2015). The expanding world of small RNAs in plants. Nat. Rev. Mol. Cell Biol. **16:** 727–741.

Brunschwig, H., Levi, L., Ben-David, E., Williams, R.W., Yakir, B., and Shifman, S. (2012). Fine-scale maps of recombination rates and hotspots in the mouse genome. Genetics **191:** 757–764.

Chen, C., Farmer, A.D., Langley, R.J., Mudge, J., Crow, J.A., May, G.D., Huntley, J., Smith, A.G., and Retzel, E.F. (2010). Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated Arabidopsis male meiocytes. BMC Plant Biol. **10:** 280.

Choi, K., et al. (2016) Recombination rate heterogeneity within Arabidopsis disease resistance genes. PLoS Genet. **12:** e1006179.

Choi, K., et al. (2018) Nucleosomes and DNA methylation shape meiotic DSB frequency in *Arabidopsis thaliana* transposons and gene regulatory regions. Genome Res. **28:** 532–546.

Choi, K., Zhao, X., Kelly, K.A., Venn, O., Higgins, J.D., Yelina, N.E., Hardcastle, T.J., Ziolkowski, P.A., Copenhaver, G.P., Franklin, F. C., McVean, G., and Henderson, I.R. (2013). Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. Nat. Genet. **45:** 1327–1336.

Creasey, K.M., Zhai, J., Borges, F., Van Ex, F., Regulski, M., Meyers, B.C., and Martienssen, R.A. (2014). miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. Nature **508:** 411–415.

de Massy, B. (2013). Initiation of meiotic recombination: How and where? Conservation and specificities among eukaryotes. Annu. Rev. Genet. **47:** 563–599.

Doutriaux, M.P., Couteau, F., Bergounioux, C., and White, C. (1998). Isolation and characterisation of the RAD51 and DMC1 homologs from *Arabidopsis thaliana*. Mol. Gen. Genet. **257:** 283–291.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: A GO analysis toolkit for the agricultural community. Nucleic Acids Res. **38:** W64-W70.

Dukowic-Schulze, S., Sundararajan, A., Ramaraj, T., Kianian, S., Pawlowski, W.P., Mudge, J., and Chen, C. (2016). Novel meiotic miRNAs and indications for a role of phasiRNAs in meiosis. Front. Plant Sci. **7:** 762.

Flórez-Zapata, N.M., Reyes-Valdés, M.H., and Martínez, O. (2016). Long non-coding RNAs are major contributors to transcriptome changes in sunflower meiocytes with different recombination rates. BMC Genomics **17:** 490.

Francia, S., Michelini, F., Saxena, A., Tang, D., de Hoon, M., Anelli, V., Mione, M., Carninci, P., and d'Adda di Fagagna, F. (2012). Site-specific DICER and DROSHA RNA products control the DNA-damage response. Nature **488:** 231–235.

Grelon, M., Vezon, D., Gendrot, G., and Pelletier, G. (2001). *AtSPO11-1* is necessary for efficient meiotic recombination in plants. EMBO J. **20:** 589–600.

Gu, W., et al. (2009) Distinct Argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. Mol. Cell **36:** 231–244.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol. **8:** R24.

He, Y., et al. (2017) Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize. Proc. Natl. Acad. Sci. USA **114:** 12231–12236.

He, H., Yang, T., Wu, W., and Zheng, B. (2015). Small RNAs in pollen. Sci. China Life Sci. **58:** 246–252.

Hsu, F.M., Yen, M.R., Wang, C.T., Lin, C.Y., Wang, C.R., and Chen, P.Y. (2017). Optimized reduced representation bisulfite sequencing reveals tissue-specific mCHH islands in maize. Epigenetics Chromatin **10:** 42.

Hunter, N. (2015). Meiotic recombination: The essence of heredity. Cold Spring Harb. Perspect. Biol. **7:** 7.

Hunter, N., and Kleckner, N. (2001). The single-end invasion: An asymmetric intermediate at the double-strand break to double-Holliday junction transition of meiotic recombination. Cell **106:** 59–70.

Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007). Genome-wide profiling and analysis of *Arabidopsis* siRNAs. PLoS Biol. **5:** e57.

Keeney, S., Giroux, C.N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell **88:** 375–384.

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 Biol. **10:** R25.

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

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

Li, L.C., Okino, S.T., Zhao, H., Pookot, D., Place, R.F., Urakami, S., Enokida, H., and Dahiya, R. (2006). Small dsRNAs induce transcriptional activation in human cells. Proc. Natl. Acad. Sci. USA **103:** 17337–17342.

Li, X., Chang, Y., Xin, X., Zhu, C., Li, X., Higgins, J.D., and Wu, C. (2013). Replication protein A2c coupled with replication protein A1c regulates crossover formation during meiosis in rice. Plant Cell **25:** 3885–3899.

Loraine, A.E., McCormick, S., Estrada, A., Patel, K., and Qin, P. (2013). RNA-Seq of Arabidopsis pollen uncovers novel transcription and alternative splicing. Plant Physiol. **162:** 1092–1109.

Lu, P., Han, X., Qi, J., Yang, J., Wijeratne, A.J., Li, T., and Ma, H. (2012). Analysis of Arabidopsis genome-wide variations before and after meiosis and meiotic recombination by resequencing Landsberg *erecta* and all four products of a single meiosis. Genome Res. **22:** 508–518.

Martínez, G., Panda, K., Köhler, C., and Slotkin, R.K. (2016). Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. Nat. Plants **2:** 16030.

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

McMahill, M.S., Sham, C.W., and Bishop, D.K. (2007). Synthesis-dependent strand annealing in meiosis. PLoS Biol. **5:** e299.

Melamed-Bessudo, C., and Levy, A.A. (2012). Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **109:** E981–E988.

Mercier, R., Mézard, C., Jenczewski, E., Macaisne, N., and Grelon, M. (2015). The molecular biology of meiosis in plants. Annu. Rev. Plant Biol. **66:** 297–327.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. **45** (D1): D183–D189.

Michalik, K.M., Böttcher, R., and Förstemann, K. (2012). A small RNA response at DNA ends in *Drosophila*. Nucleic Acids Res. **40:** 9596–9603.

Mirouze, M., Lieberman-Lazarovich, M., Aversano, R., Bucher, E., Nicolet, J., Reinders, J., and Paszkowski, J. (2012). Loss of DNA methylation affects the recombination landscape in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **109:** 5880–5885.

Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R. Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., Tate, J., and Finn, R.D. (2015). Rfam 12.0: Updates to the RNA families database. Nucleic Acids Res. **43:** D130–D137.

Neale, M.J., Pan, J., and Keeney, S. (2005). Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. Nature **436:** 1053–1057.

Osman, K., Sanchez-Moran, E., Mann, S.C., Jones, G.H., and Franklin, F.C.H. (2009). Replication protein A (AtRPA1a) is required for class I crossover formation but is dispensable for meiotic DNA break repair. EMBO J. **28:** 394–404.

Pan, J., Sasaki, M., Kniewel, R., Murakami, H., Blitzblau, H.G., Tischfield, S.E., Zhu, X., Neale, M.J., Jasin, M., Socci, N.D., Hochwagen, A., and Keeney, S. (2011). A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. Cell **144:** 719–731.

Portnoy, V., Lin, S.H., Li, K.H., Burlingame, A., Hu, Z.H., Li, H., and Li, L.C. (2016). saRNA-guided Ago2 targets the RITA complex to promoters to stimulate transcription. Cell Res. **26:** 320–335.

Puizina, J., Siroky, J., Mokros, P., Schweizer, D., and Riha, K. (2004). Mre11 deficiency in Arabidopsis is associated with chromosomal instability in somatic cells and Spo11-dependent genome fragmentation during meiosis. Plant Cell **16:** 1968–1978.

Qi, J., Wijeratne, A.J., Tomsho, L.P., Hu, Y., Schuster, S.C., and Ma, H. (2009). Characterization of meiotic crossovers and gene conversion by whole-genome sequencing in *Saccharomyces cerevisiae*. BMC Genomics **10:** 475.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. Nucleic Acids Res. **44** (W1): W160-W165.

Robert, T., Nore, A., Brun, C., Maffre, C., Crimi, B., Bourbon, H.M., and de Massy, B. (2016). The TopoVIB-Like protein family is required for meiotic DNA double-strand break formation. Science **351:** 943–949.

Rodgers-Melnick, E., Bradbury, P.J., Elshire, R.J., Glaubitz, J.C., Acharya, C.B., Mitchell, S.E., Li, C., Li, Y., and Buckler, E.S. (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. Proc. Natl. Acad. Sci. USA **112:** 3823–3828.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. Cell **127:** 1193–1207.

Schultz, M.D., Schmitz, R.J., and Ecker, J.R. (2012). "Leveling" the playing field for analyses of single-base resolution DNA methylomes. Trends Genet. **28:** 583–585.

Seth, M., Shirayama, M., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, C.C. (2013). The *C. elegans* CSR-1 Argonaute pathway counteracts epigenetic silencing to promote germline gene expression. Dev. Cell **27:** 656–663.

Shibuya, K., Fukushima, S., and Takatsuji, H. (2009). RNA-directed DNA methylation induces transcriptional activation in plants. Proc. Natl. Acad. Sci. USA **106:** 1660–1665.

Shilo, S., Melamed-Bessudo, C., Dorone, Y., Barkai, N., and Levy, A.A. (2015). DNA crossover motifs associated with epigenetic modifications delineate open chromatin regions in Arabidopsis. Plant Cell **27:** 2427–2436.

Shinohara, A., Ogawa, H., and Ogawa, T. (1992). Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. Cell **69:** 457–470.

Sommermeyer, V., Béneut, C., Chaplais, E., Serrentino, M.E., and Borde, V. (2013). Spp1, a member of the Set1 Complex, promotes meiotic DSB formation in promoters by tethering histone H3K4 methylation sites to chromosome axes. Mol. Cell **49:** 43–54.

Song, R., Hennig, G.W., Wu, Q., Jose, C., Zheng, H., and Yan, W. (2011). Male germ cells express abundant endogenous siRNAs. Proc. Natl. Acad. Sci. USA **108:** 13159–13164.

Soustelle, C., Vedel, M., Kolodner, R., and Nicolas, A. (2002). Replication protein A is required for meiotic recombination in *Saccharomyces cerevisiae*. Genetics **161:** 535–547.

Stacey, N.J., Kuromori, T., Azumi, Y., Roberts, G., Breuer, C., Wada, T., Maxwell, A., Roberts, K., and Sugimoto-Shirasu, K. (2006). Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. Plant J. **48:** 206–216.

Stein, P., Rozhkov, N.V., Li, F., Cárdenas, F.L., Davydenko, O., Vandivier, L.E., Gregory, B.D., Hannon, G.J., and Schultz, R.M. (2015). Essential role for endogenous siRNAs during meiosis in mouse oocytes. PLoS Genet. **11:** e1005013.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat. Struct. Mol. Biol. **21:** 64–72.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., and Stahl, F.W. (1983). The double-strand-break repair model for recombination. Cell **33:** 25–35.

Thomas, J., and Pritham, E.J. (2015). Helitrons, the eukaryotic rolling-circle transposable elements. Microbiol. Spectr. **3:** 3.

Tran, R.K., Henikoff, J.G., Zilberman, D., Ditt, R.F., Jacobsen, S.E., and Henikoff, S. (2005). DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes. Curr. Biol. **15:** 154–159.

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

Uanschou, C., Siwiec, T., Pedrosa-Harand, A., Kerzendorfer, C., Sanchez-Moran, E., Novatchkova, M., Akimcheva, S., Woglar, A., Klein, F., and Schlögelhofer, P. (2007). A novel plant gene essential for meiosis is related to the human CtIP and the yeast COM1/SAE2 gene. EMBO J. **26:** 5061–5070.

Underwood, C.J., Choi, K., Lambing, C., Zhao, X., Serra, H., Borges, F., Simorowski, J., Ernst, E., Jacob, Y., Henderson, I.R., and Martienssen, R.A. (2018). Epigenetic activation of meiotic recombination near *Arabidopsis thaliana* centromeres via loss of H3K9me2 and non-CG DNA methylation. Genome Res. **28:** 519–531.

Vrielynck, N., Chambon, A., Vezon, D., Pereira, L., Chelysheva, L., De Muyt, A., Mézard, C., Mayer, C., and Grelon, M. (2016). A DNA topoisomerase VI-like complex initiates meiotic recombination. Science **351:** 939–943.

Walker, J., Gao, H., Zhang, J., Aldridge, B., Vickers, M., Higgins, J. D., and Feng, X. (2018). Sexual-lineage–specific DNA methylation regulates meiosis in Arabidopsis. Nat. Genet. **50:** 130–137.

Wang, Y., and Copenhaver, G.P. (2018). Meiotic recombination: Mixing it up in plants. Annu. Rev. Plant Biol. **69:** 577–609.

Wang, H., Beyene, G., Zhai, J., Feng, S., Fahlgren, N., Taylor, N.J., Bart, R., Carrington, J.C., Jacobsen, S.E., and Ausin, I. (2015). CG gene body DNA methylation changes and evolution of duplicated genes in cassava. Proc. Natl. Acad. Sci. USA **112:** 13729–13734.

Wang, Y., Cheng, Z., and Ma, H. (2013). Meiosis: Interactions between homologous chromosomes. In Cell Biology, S. Assmann and and B. Liu, eds (New York, NY: Springer New York), pp. 1–34.

Wang, Y., Cheng, Z., Lu, P., Timofejeva, L., and Ma, H. (2014). Molecular cell biology of male meiotic chromosomes and isolation of male meiocytes in *Arabidopsis thaliana*. Methods Mol. Biol. **1110:** 217–230.

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Rendtlew Danielsen, J.M., Yang, Y.G., and Qi, Y. (2012). A role for small RNAs in DNA double-strand break repair. Cell **149:** 101–112.

Wijnker, E., et al. (2013) The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. eLife **2:** e01426.

Xi, Y., and Li, W. (2009). BSMAP: Whole genome bisulfite sequence MAPping program. BMC Bioinformatics **10:** 232.

Yamada, S., Kugou, K., Ding, D.Q., Fujita, Y., Hiraoka, Y., Murakami, H., Ohta, K., and Yamada, T. (2018). The histone variant H2A.Z promotes initiation of meiotic recombination in fission yeast. Nucleic Acids Res. **46:** 609–620.

Yang, H., Lu, P., Wang, Y., and Ma, H. (2011). The transcriptome landscape of *Arabidopsis* male meiocytes from high-throughput sequencing: The complexity and evolution of the meiotic process. Plant J. **65:** 503–516.

Yang, M., Hu, Y., Lodhi, M., McCombie, W.R., and Ma, H. (1999). The *Arabidopsis SKP1-LIKE1* gene is essential for male meiosis and may control homologue separation. Proc. Natl. Acad. Sci. USA **96:** 11416–11421.

Yelina, N.E., et al. (2012) Epigenetic remodeling of meiotic crossover frequency in *Arabidopsis thaliana* DNA methyltransferase mutants. PLoS Genet. **8:** e1002844.

Yelina, N.E., Lambing, C., Hardcastle, T.J., Zhao, X., Santos, B., and Henderson, I.R. (2015). DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in *Arabidopsis*. Genes Dev. **29:** 2183–2202.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS **16:** 284–287.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. **39:** 61–69.