

## A female-specific RdDM-associated element guides de novo DNA methylation during rice gametophyte development

Taehoon Kim<sup>1</sup>, Marcio F. R. Resende Jr.<sup>2</sup>, Meixia Zhao<sup>3</sup>, Kevin Begcy<sup>1\*</sup>

<sup>1</sup>Environmental Horticulture Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, Florida 32611, USA

<sup>2</sup>Horticultural Science Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, Florida 32611, USA

<sup>3</sup>Microbiology and Cell Science Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, Florida 32611, USA

\*Correspondence: Kevin Begcy (kbegcy.padilla@ufl.edu)

### Abstract

DNA methylation is crucial for regulating gene expression and silencing transposons. While the spatiotemporal specificity of DNA methylation controls organ formation and development, the molecular mechanisms underlying locus- or tissue-specific DNA methylation remain unclear. Previously, locus- or tissue-specific DNA methylation was thought to depend solely on the recognition of histone methylations. Here we show that high expression of 24-nucleotide small interfering RNAs (24-nt siRNAs) leads to incremental CHH methylation at specific loci via the RNA-directed DNA methylation (RdDM) pathway during female gametophyte development in rice. We also identify a DNA motif, the Female-specific RdDM-associated Element (FRE), which guides methylation to these loci. Furthermore, we show that the female-preferentially expressed protein OsCLSY3 directly interacts with FRE to regulate CHH methylation specifically in female reproductive tissues. These findings uncover a novel molecular mechanism necessary for tissue- and locus-specific de novo DNA methylation that might be conserved across diverse organisms including animals.

## Introduction

The addition of a methyl group on the 5-position carbon of cytosine (5mC), known as DNA methylation, is extensively reconfigured over the course of development in plants and animals.<sup>1–3</sup> The establishment and maintenance of methylation marks is highly regulated. In mammals, DNA methylation is restricted to the symmetric CG context, although non-CG methylation is also observed in embryonic stem cells and nervous system.<sup>4–6</sup> In plants, DNA methylation occurs in the contexts of CG, CHG and CHH, where H corresponds to adenine (A), thymine (T), or cytosine (C). CG and CHG methylation are maintained by METHYLTRANSFERASE1 (MET1) and CHROMOMETHYLASE3 (CMT3), respectively.<sup>7,8</sup> CHH methylation is maintained by CMT2 or DOMAINS REARRANGED METHYLASE 2 (DRM2), a de novo DNA methyltransferase.<sup>9,10</sup> DRM2 functions in the RNA-directed DNA methylation (RdDM) pathway, which is primarily involved in de novo DNA methylation. The RdDM pathway produces 24-nucleotide small interfering RNAs (siRNAs), that are used by the de novo methyltransferase DRM2 to de novo methylate cytosines of adjacent DNA in a sequence-specific manner.<sup>10</sup> The 24-nt siRNA production can be initiated by the transcription of 30- to 45-nt single-stranded RNAs (ssRNAs) by a plant-specific RNA polymerase complex, RNA polymerase IV (Pol IV).<sup>11,12</sup> These ssRNAs are reverse-transcribed into double-stranded RNAs (dsRNAs) by RNA-directed RNA polymerase 2 (RDR2) and subsequently cleaved into 24-nt fragments by Dicer-like 3 (DCL3).<sup>13</sup> The resulting 24-nt siRNAs are loaded onto Argonaute 4 (AGO4) or AGO6 proteins to form an AGO-siRNA duplex, which facilitates sequence-specific de novo DNA methylation.<sup>14–16</sup>

Tissue-specific dynamics of DNA methylation is believed to serve as a key epigenetic control to regulate reproductive developmental transitions in both mammals and plants. For example, during embryonic development in mammals, active demethylation of specific genomic regions continues to occur, while somatic tissue genomes remain hypermethylated.<sup>17,18</sup> In rice, *OsRDR2*, which is required for 24-nt siRNA production and CHH methylation deposition, has been shown to play a crucial role in the proper development of both male and female reproductive organs.<sup>19</sup> Similarly, mutation of the maize *RDR2* homolog, *mop1*, disrupted CHH methylation, leading to an imbalance of

meiotic recombination between chromosomal arms and pericentromeric regions,<sup>20</sup> highlighting the critical role of the RdDM pathway in plant reproduction. Interestingly, distinct from vegetative and male reproductive tissue, female reproductive tissues (such as ovary, ovule, or egg cells) were reported to show a high accumulation of 24-nt siRNAs at a limited number of genomic regions in several plant species including rice, Arabidopsis, and *Brassica rapa*.<sup>21–24</sup> However, the mechanism underlying this female-specific accumulation of 24-nt siRNAs remains elusive. Given the analogous enrichment of PIWI-interacting RNAs (piRNAs) during female germline development in mammals,<sup>25,26</sup> understanding the tissue-specific regulation of small RNA expression provides deep insights underlying the epigenetic landscape during female reproduction in both plants and animals.

In Arabidopsis, four CLASSY (CLSY) proteins have been shown to regulate tissue- and locus-specific *de novo* DNA methylation through their distinct expression pattern across different tissues by recruiting Pol IV complex to different genomic regions.<sup>23,27,28</sup> The siRNA production requires the recruitment of the Pol IV complex to the heterochromatin through its interaction with CLSY proteins and SAWADEE homeodomain homolog 1 (SHH1).<sup>27,29,30</sup> SHH1 specifically recognizes dimethylation of histone H3 lysine9 (H3K9me2) and unmethylated H3 lysine4 (H3K4me0),<sup>30</sup> facilitating its interaction with CLSY proteins, which in turn recruits Pol IV to the RdDM target loci.<sup>28</sup> While CLSY1 directly interacts with SHH1 *in vitro*,<sup>29</sup> CLSY1 and CLSY2 are required for the interaction of SHH1 to the Pol IV complex *in vivo*,<sup>28</sup> highlighting that locus-specific DNA methylation relies on histone methylation recognition. However, CLSY3 and CLSY4 recruit Pol IV in an SHH1-independent manner,<sup>28</sup> raising the question of whether locus-specific RdDM regulation can occur through alternative mechanisms that bypass histone methylation. Here we uncover a histone methylation-independent mechanism for tissue- and locus-specific RdDM regulation through the direct interaction between a CLSY protein and a *cis*-regulatory element DNA motif. First, we identify female gametophyte-developmental differentially methylated regions (FG-dDMRs), where CHH methylation regions are substantially increasing during female gametophyte (FG) development in rice. We then demonstrate that this CHH methylation in FG-dDMRs is regulated in a female-specific manner via the RdDM pathway. Further analysis reveals that the female-specific

expression of 24-nt siRNAs is strongly associated with a *cis*-regulatory element, the Female-specific RdDM-associated Element (FRE). Finally, we demonstrate that the female-preferentially expressed protein OsCLSY3 directly interacts with FRE, enabling female-specific and locus-specific RdDM regulation during FG development. Together, these findings propose a novel molecular model for tissue- and locus-specific epigenetic regulation, providing new mechanistic insights into the coordination of DNA methylation and small RNA pathways during plant development.

## Results

### ***Dynamic CHH methylation during female gametophyte development***

To understand the dynamics of DNA methylation during rice FG development, we developed a staging method that allowed us to isolate individual stages during rice female gametophyte development (**Figure 1A**). We identified six developmental stages corresponding to megasporocyte (megaspore mother cell) stage (FG1), meiotic stages (FG2 and FG3), serial mitotic stages (FG4 and FG5), and mature embryo sac stage (FG6) containing three antipodal cells (AC), two synergid cells (Syn), a central cell (CC) consisting of two polar nuclei, and an egg cell (EC) (**Figure 1B**). Whole genome enzymatic methylation sequencing (EM-seq) experiments on each FG developmental stage (FG1 to FG6) and flag leaves (FL) used as a vegetative tissue revealed that genome-wide CG and CHG methylation levels did not differ among ovaries at different stages and FL, except for a lower CHG methylation in FL compared to ovaries (**Figure S1A; Table S1**). CG and CHG methylation peak in the centromeric regions of each chromosome. In contrast, CHH methylation did not show significant higher levels in those regions. Given that DNA methylation is important for gene expressions and transposable element (TE) silencing,<sup>1</sup> we explored the methylation levels in the 2-kb upstream and downstream regions of rice genes and TE bodies (**Figures S1B–S1H**). Although high CG methylation levels in gene bodies correlated with high transcriptional levels, CHG and CHH methylation did not show significant association with gene expression (**Figures S1D–S1F**). Noteworthy, CHG methylation levels in ovaries were higher than in FL. Interestingly, the methylation levels in the CHH context showed dynamic changes near TEs across different FG stages (**Figures S1G and S1H**).

Notably, the top 25% distribution of the CHH methylation (above upper hinges of boxplots) exhibited a gradual increase from FG1 to FG6 (**Figure S1H**). Taken together our results show that CHH methylation levels gradually increase in TEs during FG development in rice.

Given the gradual increase in CHH methylation during FG development (**Figure S1H**), we hypothesized that genomic regions with an increased pattern would be hypomethylated in immature FG stages and hypermethylated in mature FG stages compared to the average of other stages (**Figure S1K**). To test this hypothesis, we calculated the fractional difference by comparing the methylation levels of each FG stage to the average methylation levels of the remaining stages within a 50-bp window across the entire genome (**Figures 1C, S1I, and S1J**). Regions were defined as hypermethylated or hypomethylated if the methylation level at a given stage was higher or lower, respectively, than the average methylation level across all other developmental stages. To monitor the gradual increase of CHH methylation, we focused on the CHH hypomethylated regions in FG1 compared to the average of the remaining FG stages. While CG methylation patterns of these regions were constant across FG stages, CHG methylation of these regions were slightly hypomethylated in FG1 and hypermethylated in FG6 (**Figures S1I and S1J**). Most interestingly, CHH methylation levels in these genomic regions were hypomethylated in FG2, centered in FG3 and FG4, and hypermethylated in FG5 and FG6 (**Figure 1C**). These results further confirm that CHH methylation undergoes a gradual increase during FG development.

### ***Increasing CHH methylation in female gametophyte-developmental differentially methylated regions (DMRs)***

To identify genomic regions showing this gradual increase, we identified DMRs for all three CG, CHG, and CHH contexts by comparing one stage to the average of the other stages (**Figures 1D, S1L, and S1N**).

Within CG and CHG contexts, fewer than 100 hypomethylated and hypermethylated DMRs (hypo- and hyper-DMRs) were identified in comparisons among all FG stages, except for the comparison of FG1 to the other FG stages in the CG context, where 551 hypo-DMRs were identified (**Figures S1L and S1N**). Notably, in the CHH methylation

context, hypo-DMRs were identified in FG1 and FG2, totaling 2,073 and 3,934, respectively. CHH hyper-DMRs were found in FG5 and FG6 accounting for 1,053, and 5,109 regions, respectively (**Figure 1D**). Since genomic regions showing gradual increase of CHH methylation showed hypomethylation in early FG stages and hypermethylation in later FG stages (**Figure S1K**), we analyzed common DMRs among hypo-DMRs (FG1 and FG2) and hyper-DMRs regions (FG5 and FG6) and identified 711 overlapping DMRs present in at least 3 FG stages. We refer to these regions as female gametophyte-developmental DMRs (FG-dDMRs) (**Figure 1E**). In FG-dDMRs, CG methylation levels were constant across FG development, while CHG showed a slight but not significant increase between subsequent developmental stages. In contrast, CHH methylation in FG-dDMRs showed a significant increase from FG1 to FG6 (**Figure 1F**). The distribution of CG, CHG, and CHH methylation demonstrated that the DNA methylation levels in FG-dDMRs were generally higher in ovaries compared to FL. To determine the genomic elements where FG-dDMRs are located, we calculated the overlap between FG-dDMRs and either genes or TEs. A total of 63.71% of FG-dDMRs are located within genes and their flanking regions, including 2-kb upstream regions (31.22%), 2-kb downstream (29.54%), and introns (15.19%). Moreover, the majority of FG-dDMRs (86.78%) were located within TEs (**Figure 1G**). In particular, terminal inverted repeat (TIR) DNA transposons, including *hAT*, *PIF-Harbinger*, and *Tc1-Mariner*, as well as *Ty1-copia* long terminal repeat (LTR) retrotransposon, were highly enriched in FG-dDMRs (**Figures 1H–1L**). These results indicate the potential involvement of increasing CHH methylation in regulating both gene expression and TE silencing during FG development.

### ***RdDM-dependent and RdDM-independent mechanisms of CHH methylation in FG-dDMRs***

De novo DNA methylation via RdDM pathway and maintenance of CHH methylation catalyzed by CMT2 are two known molecular pathways involved in CHH methylation in plants.<sup>3</sup> Given that the expression of 24-nt siRNAs is required for CHH methylation via the RdDM pathway, we performed small RNA sequencing (sRNA-seq) to determine whether the RdDM pathway contributes to the increasing levels of CHH methylation in

FG-dDMRs (**Table S1**). We first identified 24-nt siRNA clusters to analyze their expression patterns at each FG developmental stage and compared the cumulative expression distribution of each cluster. Based on the knee point (point of maximum curvature), we defined 24-nt siRNA clusters that represent most of the siRNA expression for each FG stage. We named these clusters FG1- to FG6-*siren*, given that “*siren*” has been used to describe a small number of loci responsible for the majority of siRNA expression in rice, Arabidopsis, and *B. rapa*.<sup>21–23</sup> FG stages exhibited lower numbers of *siren* loci with higher 24-nt siRNA expression compared to FL (**Figure 2A**). We further identified 389 common FG-*siren* loci across all FG stages. Interestingly, the majority of the FG6-*siren* loci (88.61%) accounted for the common FG-*siren* loci (**Figure 2B**).

To further investigate how the RdDM pathway contributes to CHH methylation in FG-dDMRs, we identified unique and common regions between FG-dDMRs and FG-*siren*. While 510 were unique to FG-dDMRs (hereafter named FG-dDMR only), 170 were unique to FG-*siren* loci (hereafter named FG-*siren* only). A total of 155 common FG-dDMR ∩ FG-*siren* belonged to both groups (**Figure 2C**). We then explored the 24-nt siRNA expression and CHH methylation levels in these unique and common regions during FG development (**Figures 2D–2K**). In the common FG-dDMR ∩ FG-*siren* regions, increasing patterns of 24-nt siRNA expression and CHH methylation levels were observed throughout FG development (**Figures 2D–2G and 2H**). Almost no 24-nt siRNAs were detected in those regions in FL (**Figures 2D, 2F, and 2H**). Methylation in the CG and CHG contexts remained stable or showed a moderate increase, respectively (**Figures S2A and S2B**). Taken together our results indicate that the increasing CHH methylation in FG-dDMR ∩ FG-*siren* regions during FG development is RdDM-dependent.

Interestingly, FG-dDMR only regions lacked 24-nt siRNA expression entirely (**Figures 2D, 2F, and 2I**) but exhibited marginal changes in CG methylation (**Figure S2A**), a moderate increase in CHG methylation (**Figure S2B**), and a significant increase in CHH methylation (**Figures 2E, 2G, and 2I**). These results indicate that the increasing pattern of CHH methylation in FG-dDMR only regions during FG development is independent of the RdDM pathway.

In FG-*siren* only regions, a similar 24-nt siRNA expression pattern to FG-dDMR ∩ FG-*siren* was observed (**Figures 2D, 2F, 2H, and 2J**). However, DNA methylation in these regions did not show notable increases during FG development (**Figures 2E, 2G, 2H, 2J, S2A, and S2B**).

To ensure that the observed 24-nt siRNA expression and CHH methylation changes in the identified regions were not artifacts, we generated shuffled genomic regions using bedtools shuffle.<sup>31</sup> The Shuffle regions showed no expression of 24-nt siRNAs, or changes in CG, CHG, or CHH methylation levels during FG development (**Figures 2F, 2G, 2K, S2A, and S2B**). These results validate that the sRNA expression and DNA methylation dynamics in FG-dDMR ∩ FG-*siren*, FG-dDMR only, and FG-*siren* only are locus-specific.

#### ***Female-specific control of CHH methylation in genomic regions intersecting FG-dDMR and FG-siren***

The lack of 24-nt siRNA expression and low CHH methylation in FL compared to developing ovaries within the same genomic regions (**Figures 2D–2K**) led us to hypothesize that DNA methylation in these regions is not only locus-specific but also tissue-specific. To test this hypothesis, we compared the spatial specificity of DNA methylation patterns in the identified regions (FG-dDMR ∩ FG-*siren*, FG-dDMR only, and FG-*siren* only) across various tissue and cell types, including seedling, panicle, stamen, egg cell, central cell, embryo and endosperm.<sup>19,32,33</sup> CG and CHG methylation showed minimal differences among seedling, flag leaves, panicle, stamen and ovary/pistil (**Figures S2C–S2H**). However, central cell, embryo, and endosperm showed lower CG and CHG methylation in all regions (FG-dDMR ∩ FG-*siren*, FG-dDMR only, and FG-*siren* only, and Shuffle). In addition, hypomethylation was observed in central cell and endosperm compared to egg cell and embryo, respectively (**Figures S2C–S2H**), which is consistent with previous observations of local CG hypomethylation and global CHG hypomethylation in central cell and endosperm of rice.<sup>32,33</sup> Consistent with our findings, CHH methylation showed the most divergent patterns across different tissues. Notably, in the FG-dDMR ∩ FG-*siren* regions, high CHH methylation levels were found exclusively in female reproduction-related tissue or cell

types, whereas CHH methylation in other tissues remained marginal (**Figures 3A–3C**). A similar pattern was observed in FG-*siren* only regions, although CHH methylation levels were generally lower than those in FG-dDMR  $\cap$  FG-*siren* (**Figures 3A and 3B**). Interestingly, FG-dDMR only regions showed generally high CHH methylation across both vegetative and reproductive tissues, with methylation levels being higher in reproductive tissues (panicle, stamen, ovary/pistil, egg cell, and central cell) compared to vegetative tissues (seedling and flag leaves) (**Figures 3A, 3B, and 3D**). Similarly to CHG methylation, global CHH hypomethylation was observed in central cell and endosperm compared to egg cell and embryo, respectively (**Figures 3A and 3B**). Furthermore, given the critical role of RNA-dependent RNA polymerase 2 (RDR2) in the conversion of single-stranded RNA into double-stranded RNA to produce siRNAs in the RdDM pathway,<sup>19</sup> we compared the methylation levels of pistils from *osrdr2* mutants. The *OsRDR2* null mutation resulted in an almost complete loss of CHH methylation in all genomic regions (**Figures 3A–3D**). Collectively, our results demonstrate that CHH methylation in the FG-dDMR  $\cap$  FG-*siren* regions is mediated by the RdDM pathway and is specifically associated with female reproductive development.

### ***Female-specific RdDM-associated Element (FRE) controls the female-specific RdDM pathway***

To search for potential regulators of CHH methylation and 24-nt siRNA expression in a female-specific manner, we conducted a STREME (Sensitive, Thorough, Rapid, Enriched Motif Elicitation) analysis<sup>34</sup> to identify *cis*-regulatory elements enriched in the 2-kb flanking regions of FG-dDMR and FG-*siren* compared to random genomic regions. We identified 24 DNA motifs associated with FG-dDMRs and 10 with FG-*siren* loci (**Figure S3A**). Subsequently, we examined the enrichment of these motifs in the four identified regions (FG-dDMR  $\cap$  FG-*siren*, FG-dDMR only, and FG-*siren* only, and Shuffle). Additionally, we analyzed the overlap between the discovered motifs and TE families to remove motifs originated from TE-associated sequences (**Figures S3B–S3D**). Among the 34 identified motifs, a DNA sequence (FG-*siren*\_01: GGGGGCNNNNGCCCCC) was significantly enriched in FG-dDMR  $\cap$  FG-*siren* and FG-*siren* only, with minimal overlap with any TE family (**Figures 4A, 4B and S3A–S3D**).

Given its unique enrichment in genomic regions associated with 24-nt siRNA expression in a female-specific manner, we named this motif FRE (Female-specific RdDM-associated Element) (**Figures 4A, 4B and S3A–S3D**).

To assess how variations in FRE sequence influence 24-nt siRNA expression, we analyzed FRE occurrences in the genome based on the number of mismatches from core sequence. We identified 65 occurrences with 0 mismatches, 1,082 with 1 mismatch, 7,153 with 2 mismatches, and 7,345 with 3 mismatches from the core sequence. We confirmed that high 24-nt siRNA expression levels in FG-dDMR  $\cap$  FG-*siren* and FG-*siren* only occur in the presence of FRE sequences, with higher enrichment of the FRE sequence with no mismatch in FG-dDMR  $\cap$  FG-*siren* compared to FG-*siren* only (**Figure 4C**). We then explored whether regions adjacent to the FRE motif showed high 24-nt siRNA expression in a tissue-specific manner. Interestingly, we detected high expression in ovaries and pistil, moderate levels in panicles, minimal expression in stamen, and almost no expression in seedling, flag leaf, embryo, and endosperm (**Figures 4D–4F, S3E, S3I, and S3J**). In these flanking regions of FRE, tissues with high levels of 24-nt siRNA expression also showed elevated CHH methylation levels, whereas CG and CHG methylation were rarely and moderately affected by 24-nt siRNA, respectively (**Figures S3F–S3J**). Taken together, our results indicate that FRE modulates locus-specific 24-nt siRNA production in the RdDM pathway during FG development.

### ***OsCLSY3 directly interacts with FRE to promote female-specific de novo DNA methylation***

In Arabidopsis, CLASSY proteins (CLSY) are known to regulate de novo DNA methylation in a locus- and tissue-specific manner via 24-nt siRNAs.<sup>23,28</sup> To investigate whether CLSY genes are involved in the female-specific de novo DNA methylation, we first investigated the expression profiles of the three OsCLSY genes across various tissues. Among them, OsCLSY1 and OsCLSY4 were broadly expressed in both vegetative and reproductive tissues. Interestingly, OsCLSY3 showed a preferentially high expression in female-reproductive tissues (ovules, ovaries, egg cell), moderate expression in seeds and flowers, minimal expression in sperm cells, and no detectable

expression in vegetative tissues (flag leaf, seedlings, meristems, stems, leaves, and root) (**Figure 5A**).

Given that tissue-specific regulation can be mediated by *cis*-regulatory elements,<sup>35</sup> and a few DNA motifs have been proposed to regulate 24-nt siRNA production in the RdDM pathway in plants,<sup>23,36</sup> we hypothesized that OsCLSY3 interacts with FRE to control FG-siren loci in a female-specific manner via the RdDM pathway. To test this, chromatin immunoprecipitation and sequencing (ChIP-seq) data from panicles of transgenic lines expressing OsCLSY3 under its native promoter with two different C-terminal tags (GFP and 2xMYC) was analyzed.<sup>36</sup> While ChIP-seq data from OsCLSY3-GFP lines were comparable to the input controls (**Figure S4A**) and thus not informative (**Figures S4B–S4F**), OsCLSY3 was enriched at FRE regions of FG-dDMR  $\cap$  FG-*siren* and FG-*siren* only loci but not in FG-dDMR only or Shuffle regions (**Figures 5B–5D, S4D, and S4E**). These results indicate that OsCLSY3 may directly or indirectly interact with FRE to induce 24-nt siRNA expression.

To test whether OsCLSY3 directly interacts with FRE, we first performed *in silico* prediction of the interactions using AlphaFold3 (AF3).<sup>37</sup> To ensure reliability of multi-molecular prediction, we used a truncated OsCLSY3 (870-1,439) containing SNF2, N-terminal, and helicase, C-terminal-like domains with high prediction confidences (**Figures S5A and S5B**). We tested two FRE-containing probes (Probe1: **GGGGGCCGACGCC**CCCC and Probe2: **GGGGGCCTTG**CCCC) and a non-target probe (ATATCCACCCCATGCT), which is a sequence found near Probe2 that does not contain the FRE motif sequence (**Figure 5D**). AF3 predicted no significant interaction between OsCLSY3 and probes lacking FRE sequence ( $0.51 \leq pTM$  (predicted template modeling)  $\leq 0.67$ ;  $0.45 \leq ipTM$  (interface pTM)  $\leq 0.56$ ). In contrast, the interaction with FRE-containing probes were predicted to be similar to the native structure ( $0.80 \leq pTM \leq 0.86$ ), with high accuracy of the predicted relative position ( $0.70 \leq ipTM \leq 0.81$ ) (**Figures 5E and S5C–S5E**).

To confirm these predicted interactions between OsCLSY3 and FRE, we conducted electrophoretic mobility shift assays (EMSA) using recombinant proteins (**Figure 5F**). In Arabidopsis, CLSY1 and CLSY2 regulate the RdDM pathway in an H3K9me (histone H3 lysine9 methylation)-dependent manner, while CLSY3 and CLSY4 function

independently of H3K9me.<sup>28</sup> Given that OsCLSY3 and OsCLSY4 are phylogenetically related to Arabidopsis CLSY3 and CLSY4 (**Figure S5F**), we tested both rice CLSYs for interaction with FRE. As AF3 predicted that OsCLSY3 requires two functional domains (SNF2, N-terminal and helicase, C-terminal-like) to interact with FRE (**Figures 5E and S5B–S5E**), recombinant proteins containing the same domains, OsCLSY4 (850-1,445) and OsCLSY3 (870-1,439) were used in EMSA assays (**Figure S5G**). Interestingly, we only detected interaction between OsCLSY3 and FRE-containing probes (**Figures 5F and S5I**). No interaction was detected between FRE probes and OsCLSY4 (**Figure S5I**). Overall, our results demonstrate that the female-specific OsCLSY3 protein directly interacts with FRE to regulate the RdDM pathway in a female-specific (tissue-specific) and FG-*siren*-specific (locus-specific) manner in rice.

## Discussion

Our study establishes a comprehensive framework for exploring epigenetic dynamics during developmental progression within a single tissue, providing unique insights unattainable through cross-tissue investigations. Epigenetic dynamics fundamentally differ from transcriptomic or proteomic dynamics because epigenetic marks are heritable across generations,<sup>38,39</sup> unlike the transient nature of RNA or protein expressions. While comparisons of epigenetic marks across different tissues can identify what marks differ among subjective tissues, they fail to capture when and how these marks are established or diminished during development. By focusing on gradual developmental changes in DNA methylation within a single tissue, we captured the temporal dynamics of epigenetic modifications. Specifically, we identified genomic regions exhibiting incremental CHH methylation throughout female reproduction, mediated via both RdDM-dependent and RdDM-independent mechanisms (**Figure 2**). Notably, in *osrdr2* mutants,<sup>19</sup> CHH methylation is absent in pistils and various vegetative and reproductive tissues across the genome (**Figure 3**), indicating that even regions showing RdDM-independent CHH methylation during FG development rely on the RdDM pathway at earlier developmental stages. These findings underscore the effectiveness of our framework in revealing dynamic epigenetic patterns during tissue

development, offering invaluable insights that are difficult to obtain from comparisons among different tissues.

Unlike somatic cells, female and male germ cells undergo meiosis followed by mitotic division to generate differentiated egg, central, and sperm cells in plants.<sup>40,41</sup> These processes are dynamically regulated by the RNA-directed DNA methylation (RdDM) pathway to ensure normal development.<sup>19</sup> In *Arabidopsis*, histone methylation at H3 lysine 9 (H3K9me) is required for locus-specific regulation by CLSY1 and CLSY2, together with SHH1, and the Pol-IV complex.<sup>28,30</sup> Similar associations between de novo DNA methylation and histone modifications have been observed in other species, including humans.<sup>42–44</sup> Our study identified a novel mechanism for de novo DNA methylation that does not rely on histone modifications, but instead uses a DNA motif (FRE) through direct interaction with OsCLSY3 to regulate CHH methylation in a tissue- and locus-specific manner. Overall, our findings not only shed light on the developmental dynamics of epigenetic regulation but also uncover a long-sought *cis*-regulatory element that guides the RdDM pathway enabling locus- and tissue-specific control of DNA methylation patterns.

In several plant species, including *Arabidopsis*, rice, and *B. rapa*, unique loci with high levels of 24-nt siRNAs, known as *siren* loci have been observed in female reproductive tissues.<sup>21,23,24,45</sup> However, the mechanisms underlying the abundant expression of 24-nt siRNAs in these tissues is largely unclear. We found that these *siren* loci (FG-*siren*) display dynamic expression patterns during ovary development, specifically controlling CHH methylation via the RdDM pathway (**Figure 2**). In *Arabidopsis*, the tissue-specific expression of *CLASSY* genes controls tissue-specific patterns of DNA methylation.<sup>23</sup> The interaction of SHH1 with CLSY1 that recruits Pol IV to targeted loci results in locus-specific DNA methylation patterns via the RdDM pathway.<sup>28,29</sup> Although CLSY3 and CLSY4 are preferentially expressed in *Arabidopsis* ovules, they regulate distinct loci.<sup>23</sup> In rice, three *CLASSY* genes exhibit different tissue-specific expression patterns (**Figure 5A**). OsCLSY3 is almost exclusively expressed in female reproductive tissues, while OsCLSY1 and OsCLSY4 are expressed across various tissues at different developmental stages (**Figure 5A**). This suggests a developmental female-specific control of OsCLSY3 and a more constitutive one for OsCLSY4. Mechanistically, we

demonstrate that 24-nt siRNA production at FG-*siren* loci relies on OsCLSY3 (**Figures 5B–5D, 5F, S4D–S4F, and S5I**), similarly to findings in *Arabidopsis*<sup>28</sup>, highlighting the conserved nature of this process. Although DNA motifs have been proposed to be involved in the recruitment of CLSY proteins,<sup>23,36</sup> this had not been directly demonstrated. Using ChIP-seq and EMSA experiments, we show that locus- and tissue-specific patterns of DNA methylation are regulated through the direct interaction of OsCLSY3 with a highly conserved DNA female-specific RdDM-associated Element (FRE), which guides the de novo CHH methylation via the RdDM pathway. Collectively, our findings demonstrate that OsCLSY3-mediated regulation is essential for establishing the epigenetic landscape during female reproduction.

Despite divergent patterns of epigenetic regulation, the conserved nature of DNA methylation across diverse organisms is evident. Similar to the RdDM pathway in plants, the association between PIWI-interacting RNAs (piRNAs) and PIWI proteins in mammals directs de novo DNA methylation of TEs during germline development.<sup>46,47</sup> In addition, DNA methylation in mammals occurs in the H3K9me-enriched and H3K4me-deficient genomic regions.<sup>42,43</sup> These analogies highlight the conserved mechanisms underlying epigenetic regulation across species. Our study provides invaluable insights into the crosstalk among different epigenetic regulators during development.

Furthermore, our findings open a window to fine-tune DNA methylation patterns in a developmental, locus- and tissue-specific manner. This approach could help revert or prevent deleterious epigenetic alterations related to abnormal development or diseases in plants and mammals.

Our study provides a novel female-specific RdDM mechanism by direct interaction between a *cis*-regulatory element FRE and a SNF2-related putative chromatin remodeling protein OsCLSY3. We demonstrate their interaction, both *in vivo* via ChIP-seq and *in vitro* via EMSA (**Figure 5**). However, given that (1) OsCLSY3 was predicted to interact with the backbone of DNA (**Figures 5 and S5**), and (2) SNF2-related chromatin remodelers are known to recognize specific DNA conformation rather than specific DNA sequence,<sup>48,49</sup> future studies are needed to explore how OsCLSY3 interacts with FRE in a sequence-specific manner.

## Materials and Methods

### ***Plant Material and Growth Conditions***

The rice (*Oryza sativa*) japonica cultivar Nipponbare was utilized in this study. Rice seeds were dehulled and surface sterilized using 70% ethanol for one minute followed by 1.5% sodium hypochlorite for 30 minutes. Sterilized seeds were germinated on 1/2 Murashige and Skoog (MS) medium supplemented with 3.0% (w/v) sucrose and 0.3% (w/v) Gelrite and incubated in an AR-36 growth chamber (Percival Scientific) under continuous light condition (4000 lux) at 28°C. After ten days, rice seedlings were transplanted to a custom potting mixture: 5 portions of Pro-Mix HP Mycorrhizae (Premier Tech Horticulture) and 2 portions of Sun Gro® Peat Moss Grower Grade White (Sun Gro Horticulture). The seedlings were grown in the greenhouse (28°C day/ 25°C night) under natural light condition (~14 h sunlight/day, 29.65° N latitude) in Gainesville, Florida, United States.

### ***Defining FG Developmental Stages and Sample Collection***

During the reproductive stage, immature rice spikelets were collected. The spikelets were carefully opened, and pistils were extracted using sharp forceps. To study developmental dynamics during FG development, we categorized the developmental stages based on the pistil lengths, with pistils measuring 0.8-, 1.2-, 1.6-, 2.0-, 2.4-, and 2.6-mm defined as FG1, FG2, FG3, FG4, FG5, and FG6 stages, respectively. We placed the pistils on an ice-cold cooling cartridge, examined them using a dissecting microscope Leica MSV269 (Leica), and then measured the lengths using LAS X microscope software (Leica).

For WE-CLSM (whole-mount Eosin B-staining confocal laser scanning microscopy), intact pistils were collected into FAA solution (37% formaldehyde : acetic acid : ethanol : distilled water = 5 : 6 : 44 : 45). For EM-seq (enzymatic methylation sequencing), RNA-seq (RNA sequencing), and sRNA-seq (small RNA sequencing) analyses, each pistil was dissected into ovary and stigma/style using a razor blade (along the dotted lines as shown in **Figure 1A**), and only ovaries were collected in liquid nitrogen. Flag leaves from the same developmental age were also collected as vegetative controls and frozen in liquid nitrogen.

## **WE-CLSM**

Pistils collected in FAA solution were fixed for 24 hours, washed with 50% ethanol for 30 minutes, and stored in 70% ethanol at 4°C for long-term storage.

For hydration, pistils were incubated sequentially in 50% ethanol, 30% ethanol, and distilled water for 30 minutes each. Subsequently, we replaced the solution with 2% aluminum potassium sulfate and incubated for 30 minutes to allow the eosin B dye to enter the ovule readily. To stain cell granules and nucleoli, samples were incubated in 10 mg/L eosin B for 16 hours. The remaining dye was removed by incubating the pistils in 2% aluminum potassium sulfate for 30 minutes. After rinsing the samples briefly with distilled water three times, pistils were dehydrated with a graded series of ethanol solutions (30%, 50%, 70%, 90%, 100%, 100%, and 100%) for 20 minutes each. The dehydrated samples were treated with 1:1 mixture of ethanol and methyl salicylate and then cleared with a pure methyl salicylate for 1 hour each.

For imaging, cleared samples were mounted on a slide glass with the plane constituted by the two stigmas positioned perpendicular to the plane of the slide glass. Cellular status of pistils was observed using a confocal laser scanning microscope (IX81-DSU, Olympus, Tokyo, Japan) coupled with a TRITC filter (excitation wavelength 532–554 nm; emission wavelength 570–613 nm).

## ***EM-Seq Library Preparation and Mapping***

For enzymatic methylation sequencing (EM-seq) analysis, more than 35 ovaries were pooled per replicate, and two biological replicates per sample were used. Genomic DNA was extracted from ovaries and flag leaves using the DNeasy Plant Mini Kit (Qiagen). Methyl-seq libraries were prepared using NEBNext® Enzymatic Methyl-seq Kit (NEB) and subsequently sequenced using the NovaSeq6000 platform (Illumina) at the Interdisciplinary Center for Biotechnology Research at the University of Florida.

Raw reads were assessed using FastQC v0.11.7.<sup>50</sup> Paired-end reads were trimmed with Trimmomatic v0.39 (LEADING:3, TRAILING:3, SLIDINGWINDOW:4:20, MINLEN:36).<sup>51</sup> and mapped to the rice genome MSUv7<sup>52</sup> using Bismark v0.22.3 (-l 50 - N 1)<sup>53</sup> with bowtie2 mapper.<sup>54</sup> PCR duplicates were removed using

deduplicate\_bismark tool under Bismark.<sup>53</sup> Subsequently, methylated and unmethylated cytosines were counted by employing bismark\_methylation\_extractor, bismark2bedGraph, and coverage2cytosine tools under Bismark.<sup>53</sup> Only cytosines having at least four mapped reads were retained for further analyses. Data from publicly available sources for embryo, endosperm,<sup>32</sup> seedling, panicle, stamen, and pistil (Nipponbare and osrdr2 mutants)<sup>19</sup> were retrieved then processed in the same manner as delineated above. For egg cell and central cell,<sup>33</sup> available GFF files containing methylated and unmethylated counts for individual cytosines were used. General statistics for EM-seq analysis are shown in **Table S1**.

### **DNA Methylation Data Analyses and Visualization**

Genome-wide DNA methylation levels in CG, CHG, and CHH contexts were calculated within 500-kb sliding windows with 100-kb shifts. The 500-kb window containing the highest density of centromeric repeats in each chromosome was designated as the centromeric region. For metaplots, CG, CHG, and CHH methylation levels were calculated within 100-bp bins spanning 2-kb upstream and downstream regions of genes/TEs and 20 proportional bins across gene/TE bodies. Boxplots were used to display DNA methylation rates in CG, CHG, and CHH sequence contexts for each genomic region, such as gene or TE bodies. DNA methylation levels were visualized using the ggplot2 package<sup>55</sup> in RStudio.

For density plots, fractional differences in DNA methylation levels for CG, CHG, and CHH contexts were calculated within 50-bp windows. The 50-bp windows included in the analysis were defined as containing at least 20 informative cytosines in both samples and having the methylation rates of at least one sample greater than 0.5, 0.2, and 0.1 for CG, CHG, and CHH methylation, respectively. The fractional differences comparing one FG stage to the other five stages were calculated using the following formula:

$$\text{Fractional difference} = \%mC_{FGN} - \frac{\sum_{k=1}^6 \%mC_{FGk} - \%mC_{FGN}}{5} \quad (\text{where } 1 \leq N \leq 6)$$

Kernel density plots were generated using the ggplot2 package in RStudio.<sup>55</sup>

### **DMR Analysis**

Differentially methylated regions (DMRs) were identified using metilene v0.2.8.<sup>56</sup> For each sequence context (CG, CHG, or CHH), bedGraph files were generated by calculating methylation rates for each cytosine with more than four mapped reads. By processing these bedGraph files, we produced input files using “metilene\_input.pl” under metilene.<sup>56</sup> Subsequently, metilene was performed to compare a certain FG stage to all other FG stages with default parameters except for allowing a maximum distance of 150-nt between two cytosine within a DMR (-M 150 -A group\_A -b group\_B). Using “metilene\_output.pl” under metilene<sup>56</sup>, candidate DMRs were filtered with the following parameters: (1) at least 10 cytosine sites within a DMR, (2) the mean methylation difference > 10%, and (3) Bonferroni-adjusted p-value < 0.05. Female gametophyte-developmental DMRs (FG-dDMRs) were defined as overlaps among hypo-CHH DMRs in FG1 and FG2, and hyper-CHH DMRs in FG5 and FG6. Venn diagram was created using venn module of Intervene v0.6.2.<sup>57</sup>

### ***Enrichment of Genomic Elements in FG-dDMRs***

The overlap between FG-dDMRs and genomic features, including 2-kb upstream and downstream regions, 5' and 3' untranslated regions, coding sequences, intronic regions of genes, and TEs, was assessed using bedtools (v2.30.0)<sup>31</sup> intersect. To examine the enrichment of TE families overlapping with FG-dDMRs, the observed overlap was compared with the expected overlap, calculated based on their genomic occupancy.

### ***sRNA-Seq Library Preparation and Mapping***

For sRNA-seq analysis, each replicate consisted of more than 50 ovaries, with two biological replicates per sample. Total RNA was extracted from ovaries and flag leaves using the RNeasy Plant Mini Kit (Qiagen), coupled with on-column DNase I treatment using RNase-free DNase Set (Qiagen). Libraries were prepared using the NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (NEB). Library products were purified using 8% polyacrylamide gel, and fragments corresponding to 140–160 bp (sRNA plus 3' and 5' adapters) were excised. Sequencing was conducted on a NovaSeq6000 platform at Novogene.

Single-end 50-bp reads were trimmed and filtered using TrimGalore v0.6.10 with default parameters except for a minimum length of 10 bp (--length 10). Cleaned reads were mapped to the rice Nipponbare reference genome MSUv7<sup>52</sup> using Bowtie v1.2.3<sup>58</sup> with no mismatches allowed (-v 0). Uniquely mapped reads were retained (-m 1). Reads mapped to rRNA, tRNA, snRNA, snoRNA (based on RNACentral v22<sup>59</sup>), and miRNA (based on miRbase v22<sup>60</sup>) were removed, and the remaining reads were considered as siRNA reads. The uniquely mapped 24-nt siRNA reads were utilized for further analyses.

Publicly available sRNA data from seedling, panicle, stamen, and pistil (Nipponbare and osrdr2 mutants)<sup>19</sup> were retrieved then processed in the same manner as delineated above. For embryo and endosperm,<sup>45</sup> raw reads were trimmed with TrimGalore with the default parameters except for a minimum length of 10 bp and a manual adapter (--length 10 -a CTGTAGGCACCA) and subsequently processed as described above. General statistics for sRNA-seq analysis are shown in **Table S1**.

### ***Identification of 24-nt siRNA Expressing Loci***

To identify 24-nt siRNA clusters, BAM files from replicates of the same sample were merged using samtools (v1.18)<sup>61</sup> merge. Merged BAM files containing 24-nt reads were converted to BED files using bedtools<sup>31</sup> bamtobed. Clusters were identified using bedtools<sup>31</sup> merge (-d 75), and read counts per cluster were counted with bedtools<sup>31</sup> coverage. Clusters with reads per million (RPM) > 10 were retained for further analyses. We used “*siren loci*” to indicate 24-nt siRNA clusters with abundant expression as previously utilized.<sup>21–23</sup> To identify *siren loci*, cumulative sums of 24-nt siRNA levels were calculated by sorting clusters in descending order based on their expression levels. We plotted a cumulative sum plot of 24-nt siRNAs using the ggplot2 package,<sup>55</sup> and *siren loci* were identified as 24-nt siRNA clusters at the knee point using the kneedle package<sup>62</sup> in RStudio.

### ***Defining FG-dDMR ∩ FG-siren, FG-dDMR only, FG-siren only, and Shuffle Regions***

To identify overlaps among genomic regions, (1) a master BED file was generated by merging individual BED files using bedtools<sup>31</sup> merge, (2) genome coordinates for

individual BED files were revised using bedtools<sup>31</sup> intersect (-a master BED -b individual BED -wa), and (3) unique genome coordinates in individual BED files were retained.

Venn diagrams or UpSet plots were generated using the venn or upset module in Intervene (v0.6.2),<sup>57</sup> respectively.

As a control, shuffled genomic regions were generated using bedtools<sup>31</sup> shuffle with 835 genomic regions (combining 155 FG-dDMR  $\cap$  FG-siren, 510 FG-dDMR only, and 170 FG-siren only, as depicted in **Figure 2C**) as input. Each shuffled region was constrained to maintain the same length on the same chromosome as each genomic region in input (-chrom).

### ***Visualization of 24-nt siRNA Expression and DNA Methylation Levels***

The 24-nt siRNA expression and DNA methylation levels surrounding regions of interest were visualized using deeptools v3.5.2.<sup>63</sup> For 24-nt siRNAs, bigWig files were generated from BAM files using deeptools<sup>63</sup> bamCoverage (--binSize 1 --normalizeUsing CPM) with 1-bp bins and CPM normalization, where the CPM (counts per million) is equivalent to the RPM. For methylation levels, bedGraph files were converted to bigWig format utilizing the bedGraphToBigWig tool.<sup>64</sup>

The resultant bigWig files for 24-nt siRNA and DNA methylation levels were used to generate a matrix using deeptools<sup>63</sup> computeMatrix (reference-point -a 2000 -b 2000 --binSize 200 --referencePoint center). Heatmaps were plotted using deeptools<sup>63</sup> plotHeatmap. To illustrate 24-nt siRNA expression and DNA methylation levels of representative genomic regions, bigWig files were visualized in Integrative Genomics Viewer (IGV) v2.15.2.<sup>65</sup>

For boxplots, 24-nt siRNA read counts mapped to regions of interest (FG-dDMR  $\cap$  FG-siren, FG-dDMR only, FG-siren only, and Shuffle) were determined using the featureCounts command<sup>66</sup> from Subread v2.0.3.  $\log_{10}(1+\text{RPM})$ -normalized expression levels were visualized using the ggplot2 package<sup>55</sup> in RStudio.

### ***Discovery of Enriched Motifs in FG-dDMRs and FG-siren***

Differentially enriched DNA motifs were identified using STREME<sup>34</sup> under MEME Suite v5.4.1.<sup>67</sup> Parameters include a motif width of 8 to 20 and E-value lower than 0.05, e

determined by Fisher's exact test (--minw 8 --maxw 20 --thresh 0.05 --eval) by comparing the 2-kb flanking regions of FG-dDMRs or FG-siren to randomly selected genomic regions. The occurrences of discovered motifs against the rice genome were assessed using FIMO<sup>68</sup> with the P-value under 0.0001 (--thresh 1.0E-4) under MEME Suite.<sup>67</sup>

BED files for the 1- and 2-kb flanking regions of FG-dDMR ∩ FG-siren, FG-dDMR only, FG-siren only, and Shuffle regions were generated using bedtools(Jühling et al., 2016)<sup>31</sup> slop (-b 1000/2000). Subsequently, bedtools<sup>31</sup> intersect was used to determine motif occurrences in these regions. The frequencies of motif-containing regions were visualized using the ggplot2 package<sup>55</sup> in RStudio. To confirm the motifs were not derived from TE sequences, the same procedure was applied for 0-, 1-, and 2-kb flanking regions of different TE families .

### ***Association between FRE and either 24-nt siRNA expression or DNA methylation***

All FRE motifs found in the rice genome using FIMO<sup>68</sup> were categorized based on the number of mismatches within the FRE core sequence (GGGGCNNNNGCCCC). The occurrences of FRE within the 2-kb flanking regions of FG-dDMR ∩ FG-siren, FG-dDMR only, and FG-siren only regions were determined using bedtools<sup>31</sup> intersect. Heatmaps were generated using the ggplot2 package<sup>55</sup> in RStudio to display log<sub>10</sub>(1+RPM)-normalized 24-nt siRNA expression levels along with the FRE occurrences by the number of mismatches within the FRE core.

For 24-nt siRNA expression and DNA methylation levels in the 2-kb flanking regions of FRE motifs categorized by the number of mismatches within the FRE core, bigWig files were used to generate a matrix using deeptools<sup>63</sup> computeMatrix (reference-point -a 2000 -b 2000 --binSize 10 --referencePoint center). Heatmaps were plotted using deeptools<sup>63</sup> plotProfile (--plotType heatmap).

### ***mRNA-Seq Library Preparation and Mapping***

For mRNA-seq analysis, each replicate consisted of more than 50 ovaries, with three biological replicates per sample. Total RNA was extracted from ovaries and flag leaves using the RNeasy Plant Mini Kit (Qiagen), coupled with on-column DNase treatment

using RNase-free DNase I (Qiagen). mRNA was selected using the NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB), and libraries were prepared with the NEBNext® Ultra™ II RNA Library Prep Kit for Illumina® (NEB). Sequencing was performed on the NovaSeq6000 platform (Illumina) at the Interdisciplinary Center for Biotechnology Research at the University of Florida.

The quality of the raw reads was assessed using FastQC v0.11.7.<sup>50</sup> Paired-end reads were cleaned using Trimmomatic v0.39 (LEADING:3, TRAILING:3, SLIDINGWINDOW:4:20, MINLEN:36).<sup>51</sup> Trimmed reads were then aligned to the rice genome MSUv7<sup>52</sup> using Hisat2 v2.2.1.<sup>69</sup> Transcript counts were calculated using HTSeq-count v0.11.2<sup>70</sup> with the “union” resolution mode and excluding reads mapped to more than one gene (--mode union --nonunique none). Transcripts per million (TPM) values were calculated from the resultant count matrix for downstream analyses.

General statistics for RNA-seq analysis are summarized in **Table S1**.

### ***Transcriptional Levels of OsCLSY Genes***

For expression levels of OsCLSY genes in different tissues, TPM values for various tissue (ovule, ovary, egg, pollen, sperm, flower, panicle, seed, seedling, meristem, stem, leaf, and root) were retrieved from the CoNekT database.<sup>71</sup> These TPM values were combined with TPM values from ovaries and flag leaves obtained in this study. Log<sub>10</sub>(1+TPM)-normalized expression levels for OsCLSY1 (LOC\_Os07g49210), OsCLSY4 (LOC\_Os05g32610), and OsCLSY3 (LOC\_Os02g43460) were visualized as heatmaps using the ggplot2 package<sup>55</sup> in RStudio.

### ***ChIP-seq analysis of OsCLSY3 tagging lines***

ChIP-seq data of OsCLSY3 tagged with C-terminal epitopes (GFP and 2xMYC) in rice panicles<sup>36</sup> were analyzed as follow: first, to compare ChIP-seq signals across different tagging lines and input controls, principal component analysis (PCA) was performed using multiBigwigSummary (bins), followed by plotPCA with the --log2 option in deeptools.<sup>63</sup> The bigWig files were also visualized in IGV<sup>65</sup> to illustrate OsCLSY3 enrichment of representative genomic regions.

Final bigWig files, containing  $\log_2$ -transformed values of the ratio of the ChIP-seq signals to the input controls, were generated using deeptools<sup>63</sup> bigwigCompare (--operation log2 --binSize 5). For assessing OsCLSY3 enrichment within the 2-kb flanking regions of the genomic regions of interest (FG-dDMR  $\cap$  FG-*siren*, FG-dDMR only, FG-*siren* only, and Shuffle), a matrix was created using computeMatrix (scale-regions -a 2000 -b 2000 --binSize 10 -m 1000). To evaluate OsCLSY3 enrichment in the 2-kb flanking regions of FRE motifs categorized by the number of mismatches within the FRE core, another matrix was generated using computeMatrix (reference-point -a 2000 -b 2000 --binSize 10 --referencePoint center). The resulting matrix files were used to generate profile plots with plotProfile and heatmaps with plotHeatmap in deeptools.<sup>63</sup>

### ***Prediction of protein-DNA interaction using AlphaFold3***

The interaction between OsCLSY3 and DNA probes containing or lacking the FRE motif was predicted using AlphaFold3 (AF3).<sup>37</sup> Top-ranked prediction models for each DNA probe were visualized using PyMOL v3.0.5. A detailed list of probes used for AF3 predictions is provided in **Table S2**.

### **EMSA**

Recombinant proteins for both OsCLSY3 and OsCLSY4 were prepared by GenScript. Briefly, coding sequences corresponding to OsCLSY4 (850-1,445) and OsCLYS3 (870-1,439) were synthesized and cloned into pET-30a (+). Recombinant plasmids were transformed into *E. coli* strain BL21 (DE3), and protein expression was induced with 0.5 mM IPTG. Cells were lysed, inclusion bodies were washed, and proteins were purified using Ni-NTA resin (GenScript). Proteins were subsequently refolded in a buffer containing 50 mM Tris-HCl, 150 mM NaCl, 1 M L-Arginine, 10 % Glycerol, pH 8.0 (**Figure S5H**).

The EMSA assay was performed by Lifeasible. Briefly, 2 µg of recombinant proteins was incubated with 4 pmol of biotin-labeled probes with or without cold probes in EMSA/Gel-shift binding buffer (Beyotime) supplemented with 5 µM ATP and 5 µM MgCl<sub>2</sub> at 25°C for 30 minutes. Reaction samples were separated using 6% Native polyacrylamide gel electrophoresis (PAGE). After transfer to the PVDF membrane,

biotin-labeled probes were detected using HRP-conjugated streptavidin (Proteintech) and an ECL detection kit (Vigorous Biotech). Probes used for EMSA are listed in **Table S2**.

### Acknowledgements

We gratefully acknowledge Evert Varela for providing technical support of during the initial stages of this project. This work was supported by the Competitive Seed Grant Research Initiative (Grant No. 00129910) from the College of Agricultural and Life Sciences at the University of Florida to K.B.

### Author contributions

Conceptualization K.B.; Methodology, T.K., M.Z., and K.B.; Data curation, T.K.; Investigation, T.K.; Formal analysis, T.K. and K.B.; Visualization, T.K.; Writing – original draft, T.K. and K.B.; Writing – review & editing, T.K., M.F.R.R., M.Z., and K.B.; Funding acquisition, K.B.; Resources, M.F.R.R. and K.B.; Supervision, K.B.

## References

1. Zhang, H., Lang, Z. & Zhu, J.-K. Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol* **19**, 489–506 (2018).
2. Schmitz, R. J., Lewis, Z. A. & Goll, M. G. DNA Methylation: Shared and Divergent Features across Eukaryotes. *Trends in Genetics* **35**, 818–827 (2019).
3. Erdmann, R. M. & Picard, C. L. RNA-directed DNA Methylation. *PLoS Genet* **16**, e1009034 (2020).
4. Ramsahoye, B. H. et al. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5237–5242 (2000).
5. Chen, P.-Y., Feng, S., Joo, J. W. J., Jacobsen, S. E. & Pellegrini, M. A comparative analysis of DNA methylation across human embryonic stem cell lines. *Genome Biol* **12**, R62 (2011).
6. Jeong, H. et al. Evolution of DNA methylation in the human brain. *Nat Commun* **12**, 2021 (2021).
7. Kankel, M. W. et al. Arabidopsis MET1 Cytosine Methyltransferase Mutants. *Genetics* **163**, 1109–1122 (2003).
8. Bartee, L., Malagnac, F. & Bender, J. *Arabidopsis cmt3* chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes Dev.* **15**, 1753–1758 (2001).
9. Zemach, A. et al. The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. *Cell* **153**, 193–205 (2013).
10. Zhong, X. et al. Molecular Mechanism of Action of Plant DRM De Novo DNA Methyltransferases. *Cell* **157**, 1050–1060 (2014).
11. Herr, A. J., Jensen, M. B., Dalmay, T. & Baulcombe, D. C. RNA Polymerase IV Directs Silencing of Endogenous DNA. *Science* **308**, 118–120 (2005).
12. Onodera, Y. et al. Plant Nuclear RNA Polymerase IV Mediates siRNA and DNA Methylation-Dependent Heterochromatin Formation. *Cell* **120**, 613–622 (2005).
13. Xie, Z. et al. Genetic and Functional Diversification of Small RNA Pathways in Plants. *PLoS Biol* **2**, e104 (2004).

14. Zilberman, D., Cao, X. & Jacobsen, S. E. ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation. *Science* **299**, 716–719 (2003).
15. Havecker, E. R. *et al.* The *Arabidopsis* RNA-Directed DNA Methylation Argonautes Functionally Diverge Based on Their Expression and Interaction with Target Loci. *The Plant Cell* **22**, 321–334 (2010).
16. Qi, Y. *et al.* Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**, 1008–1012 (2006).
17. Bogdanović, O. *et al.* Active DNA demethylation at enhancers during the vertebrate phylotypic period. *Nat Genet* **48**, 417–426 (2016).
18. MacArthur, I. C. *et al.* Developmental DNA demethylation is a determinant of neural stem cell identity and gliogenic competence. *Sci. Adv.* **10**, eado5424 (2024).
19. Wang, L. *et al.* Reinforcement of CHH methylation through RNA-directed DNA methylation ensures sexual reproduction in rice. *Plant Physiology* **188**, 1189–1209 (2022).
20. Zhao, M. *et al.* The *mop1* mutation affects the recombination landscape in maize. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2009475118 (2021).
21. Grover, J. W. *et al.* Abundant expression of maternal siRNAs is a conserved feature of seed development. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 15305–15315 (2020).
22. Li, C. *et al.* Resetting of the 24-nt siRNA landscape in rice zygotes. *Genome Res.* **32**, 309–323 (2022).
23. Zhou, M. *et al.* The CLASSY family controls tissue-specific DNA methylation patterns in *Arabidopsis*. *Nat Commun* **13**, 244 (2022).
24. Li, C. *et al.* Genome-wide redistribution of 24-nt siRNAs in rice gametes. *Genome Res.* **30**, 173–184 (2020).
25. Klattenhoff, C. & Theurkauf, W. Biogenesis and germline functions of piRNAs. *Development* **135**, 3–9 (2008).
26. Williams, Z. *et al.* Discovery and Characterization of piRNAs in the Human Fetal Ovary. *Cell Reports* **13**, 854–863 (2015).
27. Law, J. A., Vashisht, A. A., Wohlschlegel, J. A. & Jacobsen, S. E. SHH1, a Homeodomain Protein Required for DNA Methylation, As Well As RDR2, RDM4,

- and Chromatin Remodeling Factors, Associate with RNA Polymerase IV. *PLoS Genet* **7**, e1002195 (2011).
28. Zhou, M., Palanca, A. M. S. & Law, J. A. Locus-specific control of the de novo DNA methylation pathway in *Arabidopsis* by the CLASSY family. *Nat Genet* **50**, 865–873 (2018).
29. Zhang, H. *et al.* DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8290–8295 (2013).
30. Law, J. A. *et al.* Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**, 385–389 (2013).
31. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
32. Zemach, A. *et al.* Local DNA hypomethylation activates genes in rice endosperm. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18729–18734 (2010).
33. Park, K. *et al.* DNA demethylation is initiated in the central cells of *Arabidopsis* and rice. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 15138–15143 (2016).
34. Bailey, T. L. STREME: accurate and versatile sequence motif discovery. *Bioinformatics* **37**, 2834–2840 (2021).
35. Marand, A. P., Eveland, A. L., Kaufmann, K. & Springer, N. M. *cis*-Regulatory Elements in Plant Development, Adaptation, and Evolution. *Annu. Rev. Plant Biol.* **74**, 111–137 (2023).
36. Pal, A. K., Gandhivel, V. H.-S., Nambiar, A. B. & Shivaprasad, P. V. Upstream regulator of genomic imprinting in rice endosperm is a small RNA-associated chromatin remodeler. *Nat Commun* **15**, 7807 (2024).
37. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
38. Paszkowski, J. & Grossniklaus, U. Selected aspects of transgenerational epigenetic inheritance and resetting in plants. *Current Opinion in Plant Biology* **14**, 195–203 (2011).
39. Schmitz, R. J. *et al.* Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. *Science* **334**, 369–373 (2011).

40. Drews, G. N. & Koltunow, A. M. G. The Female Gametophyte. *The Arabidopsis Book* **9**, e0155 (2011).
41. Borg, M., Brownfield, L. & Twell, D. Male gametophyte development: a molecular perspective. *Journal of Experimental Botany* **60**, 1465–1478 (2009).
42. Ooi, S. K. T. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).
43. Lehnertz, B. *et al.* Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology* **13**, 1192–1200 (2003).
44. Morselli, M. *et al.* In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *eLife* **4**, e06205 (2015).
45. Rodrigues, J. A. *et al.* Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7934–7939 (2013).
46. Thomson, T. & Lin, H. The Biogenesis and Function of PIWI Proteins and piRNAs: Progress and Prospect. *Annu. Rev. Cell Dev. Biol.* **25**, 355–376 (2009).
47. Zoch, A. *et al.* SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation. *Nature* **584**, 635–639 (2020).
48. Bieluszewski, T., Prakash, S., Roulé, T. & Wagner, D. The Role and Activity of SWI/SNF Chromatin Remodelers. *Annu. Rev. Plant Biol.* **74**, 139–163 (2023).
49. Liu, X., Li, M., Xia, X., Li, X. & Chen, Z. Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature* **544**, 440–445 (2017).
50. Andrew, S. FastQC: A quality control tool for high throughput sequence data. (2010).
51. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
52. Kawahara, Y. *et al.* Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 4 (2013).
53. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).

54. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
55. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer international publishing, Cham, 2016).
56. Jühling, F. *et al.* metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res.* **26**, 256–262 (2016).
57. Khan, A. & Mathelier, A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* **18**, 287 (2017).
58. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
59. RNACentral Consortium *et al.* RNACentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Research* **49**, D212–D220 (2021).
60. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Research* **47**, D155–D162 (2019).
61. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
62. Satopaa, V., Albrecht, J., Irwin, D. & Raghavan, B. Finding a ‘Kneedle’ in a Haystack: Detecting Knee Points in System Behavior. in *2011 31st International Conference on Distributed Computing Systems Workshops* 166–171 (IEEE, Minneapolis, MN, USA, 2011). doi:10.1109/ICDCSW.2011.20.
63. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165 (2016).
64. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
65. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178–192 (2013).

66. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
67. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res* **43**, W39–W49 (2015).
68. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
69. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).
70. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
71. Proost, S. & Mutwil, M. CoNekT: an open-source framework for comparative genomic and transcriptomic network analyses. *Nucleic Acids Research* **46**, W133–W140 (2018).

## Figure legends

**Figure 1. Incremental CHH methylation at specific loci during female gametophyte development.** **(A)** Developmental progression of rice pistils. Dotted red lines delimitate ovaries (below) and stigmas/ styles (above). FG1 to FG6 represent stage 1 to stage 6 during rice female gametophyte development. Scale bar: 0.5 mm. **(B)** Cellular status of the embryo sac during rice female gametophyte development. Scale bar: 40  $\mu$ m. **(C)** Density plots showing the frequency distribution of CHH methylation differences between a single FG stage and the average of all other stages. Red-shaded plots highlight 50-bp windows where the fractional difference for FG1 compared to other FG stages is lower than -0.1. **(D)** DMR analysis of CHH methylation comparing one FG stage to the average of all other stages. **(E)** Venn diagram depicting the overlap between FG1-hypo-DMRs, FG2-hypo-DMRs, FG5-hyper-DMRs, and FG6-hyper-DMRs. A total of 711 intersections involving at least three out of four DMR sets are classified as FG-dDMRs. **(F)** Boxplots showing CG, CHG, and CHH methylation levels in FG-dDMRs. **(G)** Bar charts displaying the count of genomic features overlapping with FG-dDMRs. The number of TE and non-TE elements with respect to the genic elements are shown. **(H)** Bar charts showing the observed overlap of TE families with FG-dDMRs compared to the expected overlap, calculated based on their genomic proportions (O/E, log2). Statistical significance was determined using a two-tailed Fisher's exact test (\*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ). Overrepresented TE families ( $O/E > 1.5$ ;  $P < 0.05$ ) are highlighted in blue. **(I-L)** Genome browser screenshots displaying CG, CHG, and CHH methylation levels at FG-dDMRs overlapping with overrepresented TE families: **(I)** *hAT*, **(J)** *Copia*, **(K)** *PIF Harbinger*, and **(L)** *Tc1-Mariner*.

**Figure 2. RdDM-dependent and RdDM-independent mechanisms of CHH methylation in FG-dDMRs.** **(A)** Cumulative sum plot illustrating the relationship between the percentage of 24-nt siRNA clusters (x-axis) and the cumulative sum of RPM-normalized expression (y-axis). *Siren* loci with high 24-nt siRNA expression are identified at the knee point of each graph. **(B)** UpSet plot showing intersections of *siren* loci of ovaries during female gametophyte development (FG1 to FG6). FG-*siren* loci

(highlighted in yellow) represent the common intersections across all FG stages. **(C)** Venn diagram showing the overlap between FG-dDMRs and FG-*siren* loci. **(D and E)** Heatmaps showing **(D)** RPM-normalized expression levels of 24-nt siRNAs and **(E)** CHH methylation levels in ovaries and flag leaves in the 2-kb flanking regions of FG-dDMR  $\cap$  FG-*siren*, FG-dDMR only, and FG-*siren* only regions. **(F and G)** Boxplots displaying the distribution of **(F)**  $\log_{10}(1+\text{RPM})$ -normalized expression levels of 24-nt siRNA and **(G)** CHH methylation levels of FG-dDMR  $\cap$  FG-*siren*, FG-dDMR only, FG-*siren* only, and Shuffle regions. **(H–K)** Genome browser screenshots displaying RPM-normalized 24-nt siRNA expression levels and CHH methylation levels for **(H)** FG-dDMR  $\cap$  FG-*siren*, **(I)** FG-dDMR only, **(J)** FG-*siren* only, and **(K)** Shuffle regions.

**Figure 3. Female-specific control of CHH methylation in genomic regions intersecting FG-dDMR and FG-*siren*.** **(A)** Boxplots showing CHH methylation levels across different tissues in rice. Different letters represent statistically different groups determined using the Kruskal-Wallis test followed by Dunn's test ( $P$  adjusted using the Benjamin-Hochberg procedure  $< 0.05$  was considered significantly different). **(B)** Heatmaps showing CHH methylation levels in different tissues within FG-dDMR  $\cap$  FG-*siren* and FG-dDMR only regions. **(C and D)** Genome browser screenshots displaying CHH methylation levels in different tissues of **(C)** FG-dDMR  $\cap$  FG-*siren* and **(D)** FG-dDMR only.

**Figure 4. Female-specific RdDM-associated Element (FRE) controls the female-specific RdDM pathway.** **(A)** Female-specific DNA motif sequence highly enriched in FG-*siren* loci. **(B)** Bar charts showing the enrichment of FRE in the 0-, 1-, and 2-kb flanking regions of FG-dDMR  $\cap$  FG-*siren*, FG-dDMR only, FG-*siren* only, and Shuffle regions. Statistical differences were determined using Fisher's exact test ( $P < 0.05$  was considered significant; N.S., not significant). **(C)** Heatmaps showing the expression levels of 24-nt siRNAs and motif presence categorized by the number of mismatches with the FRE core (GGGGCGNNNNNGCCCC) sequence in FG-dDMR  $\cap$  FG-*siren*, FG-dDMR only, and FG-*siren* only regions. **(D)** Heatmaps showing RPM-normalized expression levels of 24-nt siRNAs across different tissues within 2-kb flanking regions of

FRE motifs. Flanking regions were grouped by the number of mismatches with the FRE core sequence. **(E and F)** Genome browser screenshots displaying RPM-normalized 24-nt siRNA expression levels in different tissue for **(E)** FG-dDMR  $\cap$  FG-*siren* and **(F)** FG-*siren* only regions containing the FRE sequence. Each FRE motif is denoted with Roman numerals. The FRE core sequences are highlighted in bold, and mismatched bases within the FRE core are highlighted in red.

**Figure 5. OsCLSY3 directly interacts with FRE.** **(A)** Heatmap showing transcriptional levels of *CLSY* genes in rice. **(B and C)** Profile plots showing ChIP-seq signals of OsCLSY3-2xMYC within 2-kb flanking regions of **(B)** FG-dDMR  $\cap$  FG-*siren*, FG-dDMR only, FG-*siren* only, and Shuffle regions, and **(C)** FRE motifs grouped by the number of mismatches within the FRE core sequence. **(D)** Genome browser screenshots displaying RPM-normalized 24-nt siRNA expression and RPKM-normalized ChIP-seq signals of OsCLSY3 at two FG-dDMR  $\cap$  FG-*siren* regions. Each FRE motif is denoted with Roman numerals, and N represents a non-target sequence. Probe1, probe2, and non-target sequences were used in **Figures 5E, 5F, S5C–S5E, and S5I.** **(E)** AlphaFold3 predicted interactions between truncated OsCLSY3 and FRE-containing or non-target probes. SNF2 N-terminal domain, C-terminal helicase domain, FRE sequences, and non-FRE sequences are shown in maroon, green, aqua, and olive, respectively. **(F)** Electrophoretic mobility shift assay (EMSA) showing interactions between truncated OsCLSY3 and FRE-containing probes. The gradients of cold-probes represent 7.5-, 15-, and 30-fold amounts of biotin-labeled probes.

## Supplementary data

### Supplementary figure legends

**Figure S1. DNA methylation landscape during female gametophyte development in rice and its impact on transcriptional expression levels.** (A) DNA methylation levels across rice chromosomes during different FG stages. Flag leaf (FL) serves as a vegetative control tissue. Panels with different shades represent different chromosomes. Centromeric regions are indicated with dotted lines. (B) Metaplots of averaged methylation rates across 2-kb flanking regions for all rice genes in FG and flag leaf. (C) Boxplots of methylation rates in gene bodies. (D) Transcriptional levels of gene sets categorized by expression levels. Gene sets were defined based on the average TPM values for each sample (FG1 to FG6 and FL). The expression categories are as follows: No expression (TPM < 0.5), very low (up to the 2nd decile), low (up to the 4th decile), moderate (up to the 6th decile), high (up to the 8th decile), and very high (up to the 10th decile).  $\log_{10}(1+TPM)$ -normalized expression levels are visualized using boxplots. Gene counts per set are shown for each sample. (E) Metaplots of averaged methylation rates across 2-kb flanking regions of genes with varying expression levels. (F) Boxplots of methylation rates in gene body regions for gene sets with different expression levels. (G) Metaplots of averaged methylation rates across 2-kb flanking regions for all rice TEs in FG and FL. (H) Boxplots of methylation rates in TE bodies. (I and J) Density plots showing the frequency distribution of (I) CG and (J) CHG methylation differences between one FG stage and the average of all other stages. Red-shaded plots highlight the 50-bp windows shown in **Figure 1C**. (K) Schematic representation of CHH methylation levels during FG stages in regions with gradual increase in CHH methylation. (L and N) DMR analysis of (L) CG and (N) CHG methylation, comparing one FG stage to the average of all other stages.

**Figure S2. Tissue-specific DNA methylation patterns in RdDM-dependent and RdDM-independent FG-dDMRs.** (A and B) Boxplots displaying (A) CG and (B) CHG methylation levels in different region categories (FG-dDMR  $\cap$  FG-siren, FG-dDMR only,

FG-siren only, and Shuffle). **(C and D)** Boxplots of **(C)** CG and **(D)** CHG methylation levels across different tissue types in rice. Different letters represent statistically different groups determined by the Kruskal-Wallis test followed by Dunn's test ( $P$ -values adjusted using the Benjamin-Hochberg procedure  $< 0.05$  were considered significantly different). **(E and F)** Heatmaps of **(E)** CG and **(F)** CHG methylation levels in different tissue types. **(G and H)** Genome browser screenshots displaying CG, CHG, and CHH methylation levels in different tissue types at **(G)** FG-dDMR  $\cap$  FG-siren and **(H)** FG-dDMR only.

**Figure S3. Identification of FRE and its influence on 24-nt siRNA expression and DNA methylation.** **(A)** 34 DNA motifs identified by STREME analysis. **(B and C)** Heatmaps showing **(B)** the enrichment of the motifs in four identified genomic regions (FG-dDMR  $\cap$  FG-siren, FG-dDMR only, FG-siren only, and Shuffle) and **(C)**  $-\log_{10}$ -normalized P values from pairwise Fisher's exact tests among four genomic regions. DS, FG-dDMR  $\cap$  FG-siren; Do, FG-dDMR only; So, FG-siren only; Sh, Shuffle. **(D)** Heatmap showing the enrichment of the motifs in different TE families. **(E-H)** Heatmaps of RPM-normalized expression levels of 24-nt siRNAs **(E)** and CG **(F)**, CHG **(G)**, and CHH **(H)** methylation levels in the 2-kb flanking regions of FRE motifs. Flanking regions were grouped by the number of mismatches with the FRE core. **(I and J)** Genome browser screenshots displaying RPM-normalized 24-nt siRNA expression and methylation levels in **(I)** FG-dDMR  $\cap$  FG-siren and **(J)** FG-siren only regions containing FRE. Each FRE motif is denoted using Roman numerals. FRE core sequences are in bold, and mismatched bases are highlighted in red.

**Figure S4. ChIP-seq analysis of OsCLSY3.** **(A)** PCA plot among ChIP-seq samples from OsCLSY3-GFP and OsCLSY3-2xMYC lines and input controls. **(B and C)** Profile plots showing ChIP-seq signals of OsCLSY3-GFP within 2-kb flanking regions for **(B)** four genomic region categories (FG-dDMR  $\cap$  FG-siren, FG-dDMR only, FG-siren only, and Shuffle) and **(C)** FRE motifs grouped by the number of mismatches with the FRE core. **(D and E)** Heatmaps showing ChIP-seq signals of OsCLSY3-GFP and OsCLSY3-2xMYC lines within the 2-kb flanking regions of **(D)** four genomic region categories (FG-

dDMR  $\cap$  FG-*siren*, FG-dDMR only, FG-*siren* only, and Shuffle) and (E) FRE motifs grouped by the number of mismatches with the FRE core. (F) Genome browser screenshots displaying RPM-normalized 24-nt siRNA expression levels and CHH methylation levels across different tissue types and RPKM-normalized OsCLSY3 ChIP-seq signals at three FG-dDMR  $\cap$  FG-*siren* regions. Each FRE motif is denoted using Roman numerals. The FRE core sequences are highlighted in bold.

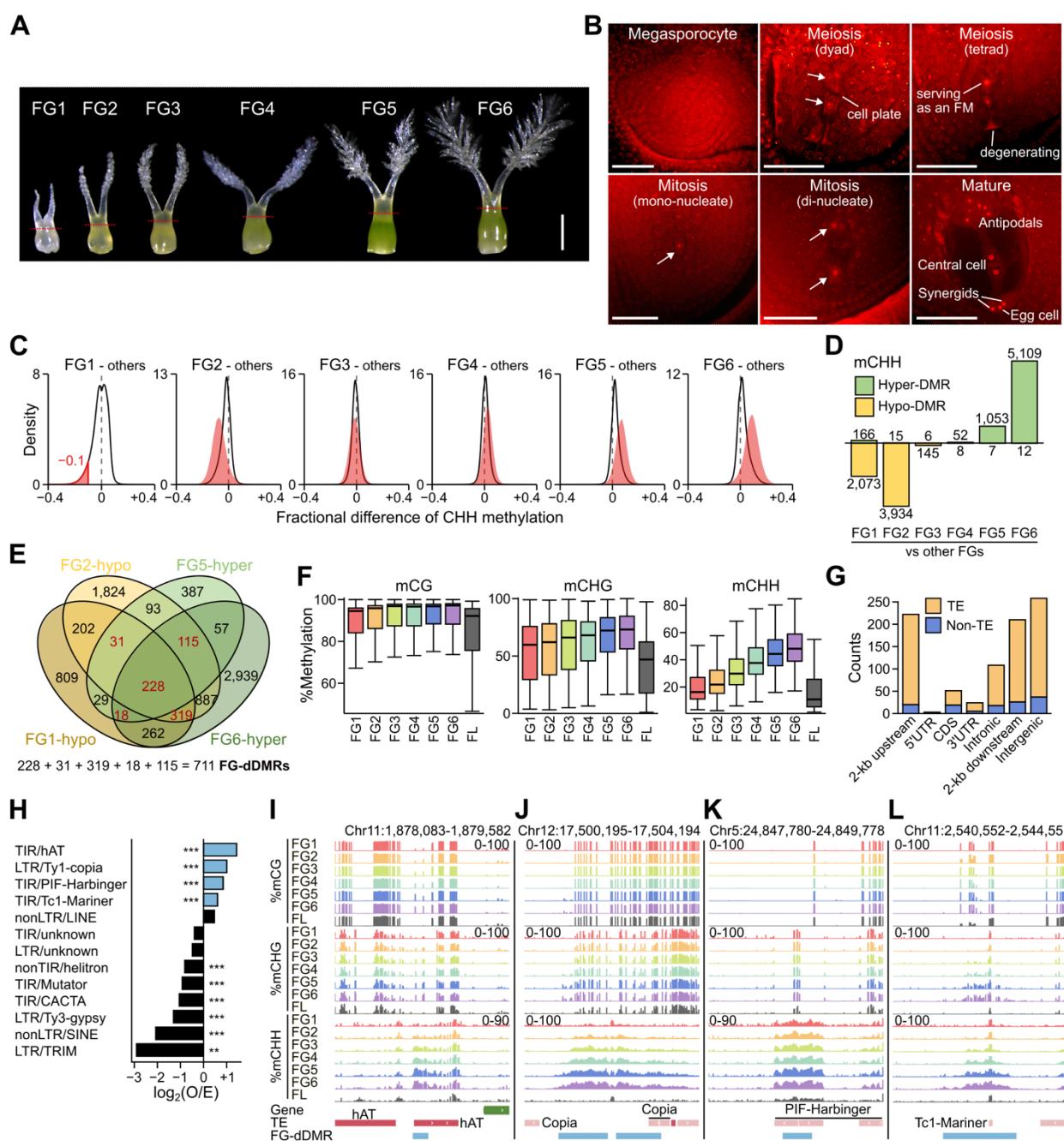
**Figure S5. Prediction and validation of the interaction between OsCLSY3 and FRE.** (A) AF3-predicted structure of OsCLSY3, colored based on pLDDT (prediction confidence; left panel) and functional domains (right panel). (B) Schematic representation of the OsCLSY3 protein with GC rates in 25-bp sliding windows and pLDDT score for each amino acid residue predicted using AF3. (C–E) AF3 predictions for FRE-containing (C) Probe1 and (D) Probe2, as well as (E) the Non-target, which does not contain FRE. The pTM (predicted template modeling) above 0.5 means the overall predicted fold for the complex might be similar to the true structure. The ipTM (interface pTM) measures the accuracy of the predicted relative positions of the subunits within the complex ( $ipTM > 0.8$ , confident high-quality prediction;  $0.6 \leq ipTM \leq 0.8$ , where predictions could be correct or incorrect;  $ipTM < 0.6$ , failed prediction). (F) Phylogenetic analysis of CLSY proteins in Arabidopsis, rice, and maize, adapted from Pal et al.<sup>36</sup> (G) Schematic representation of proteins used for EMSA. (H) Protein purification of 6His-tagged truncated OsCLSY4 and OsCLSY3. (I) EMSA testing the interaction between either truncated OsCLSY4 or OsCLSY3 and the FRE-containing Probe1. The gradients of cold-probes represent 3- and 30-fold excess amounts of biotin-labeled probes.

## Supplementary tables

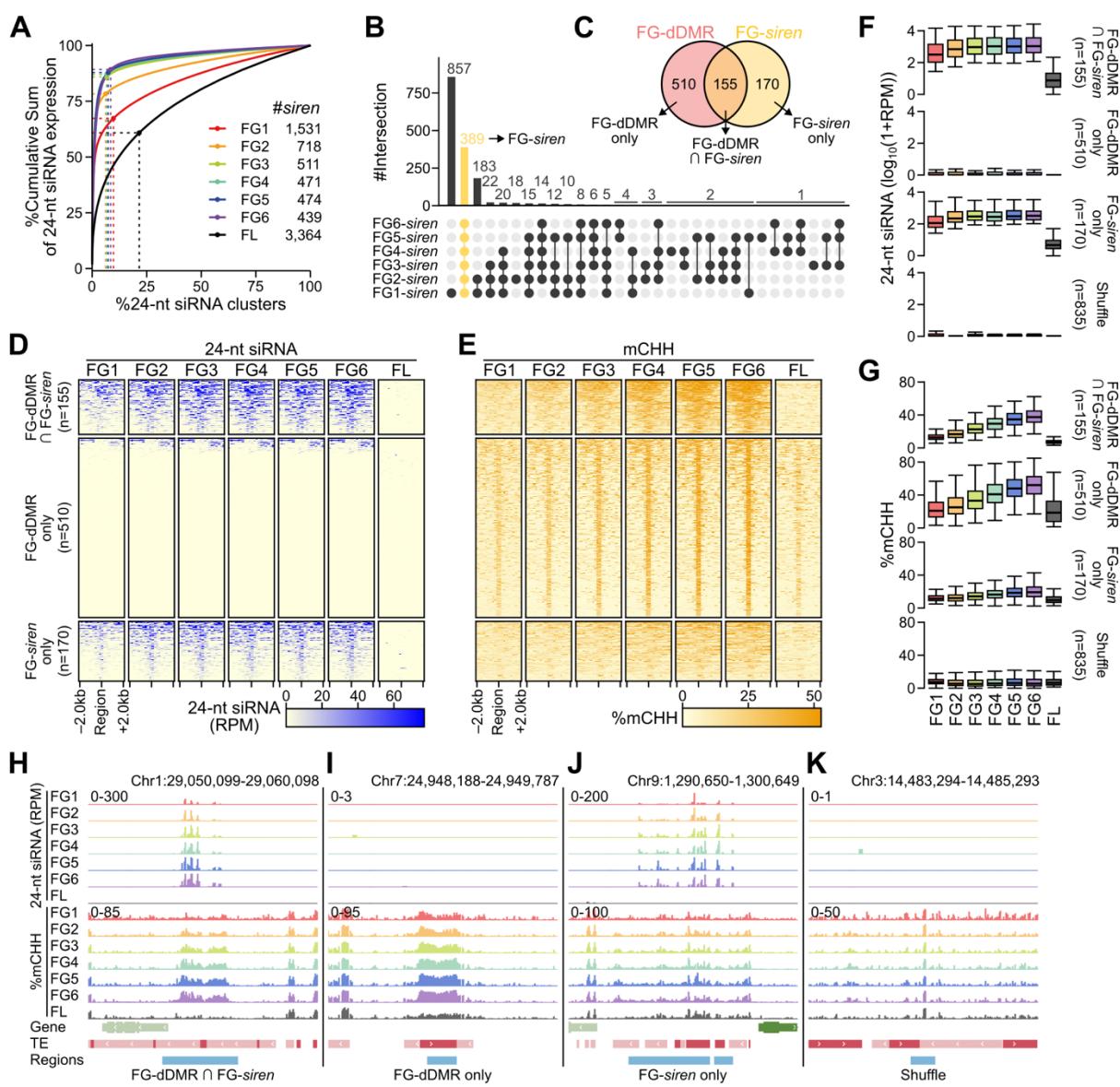
**Table S1.** Basic statistics of NGS datasets.

**Table S2.** Probes used in this study.

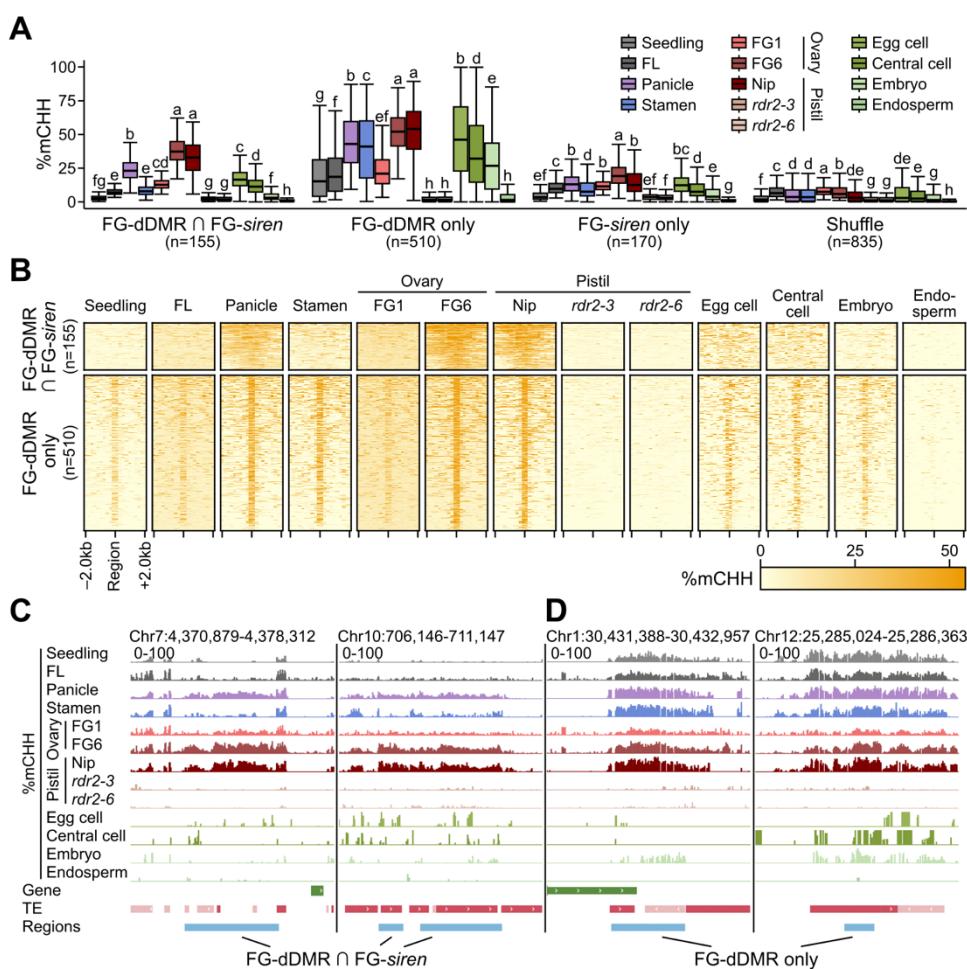
**Figure 1**



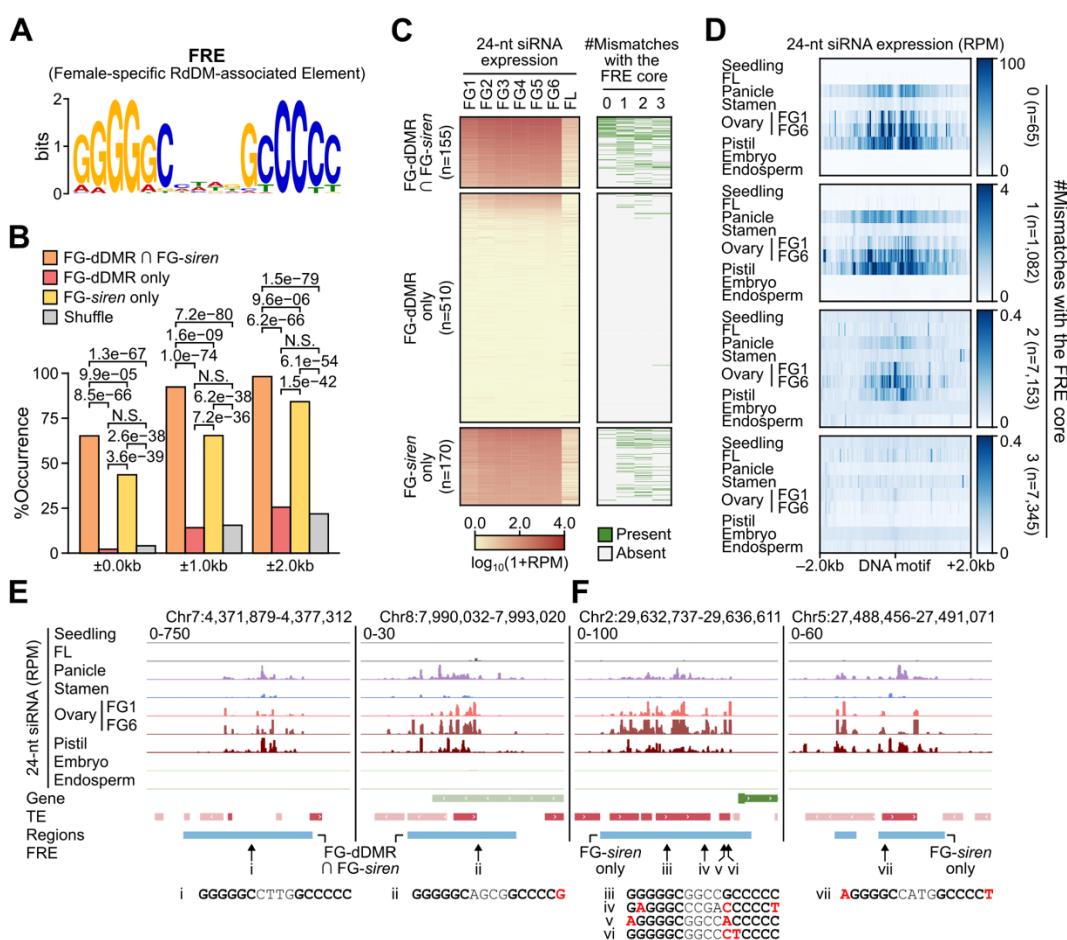
**Figure 2**



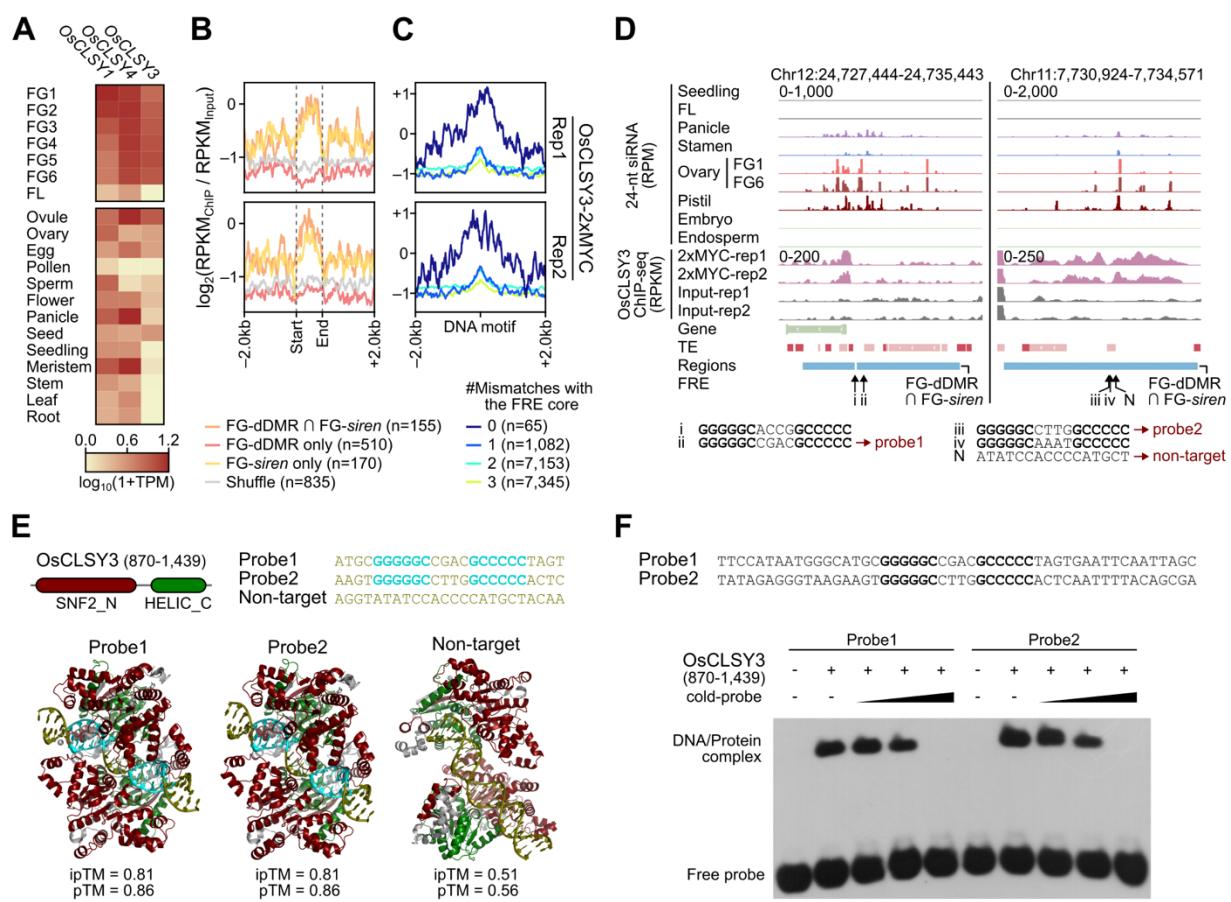
**Figure 3**



**Figure 4**

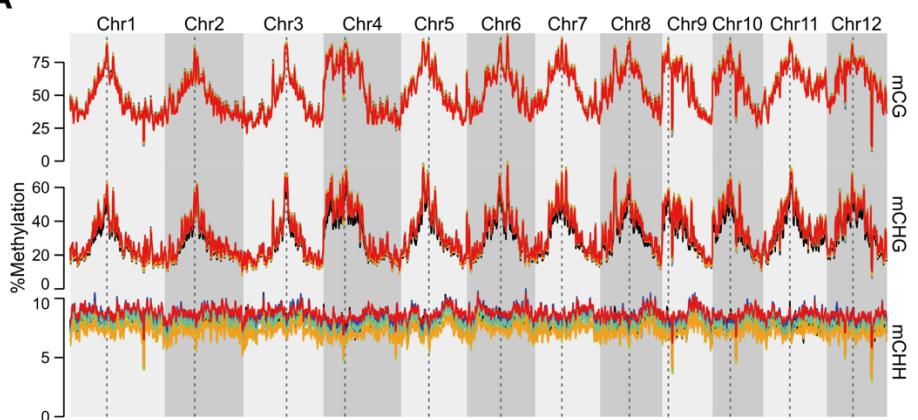


**Figure 5**

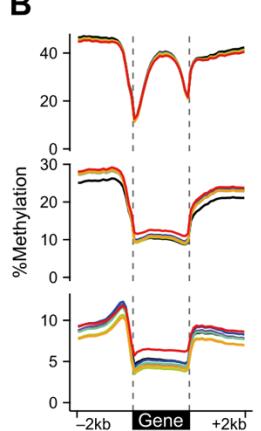


## Figure S1

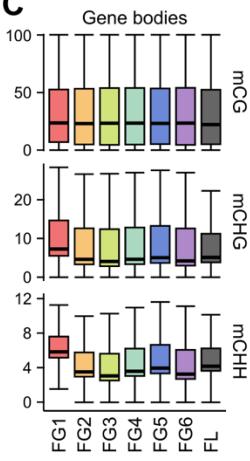
A



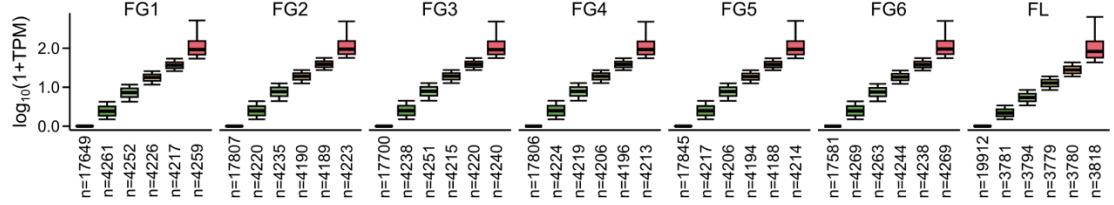
B



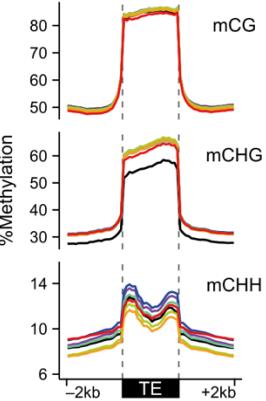
C



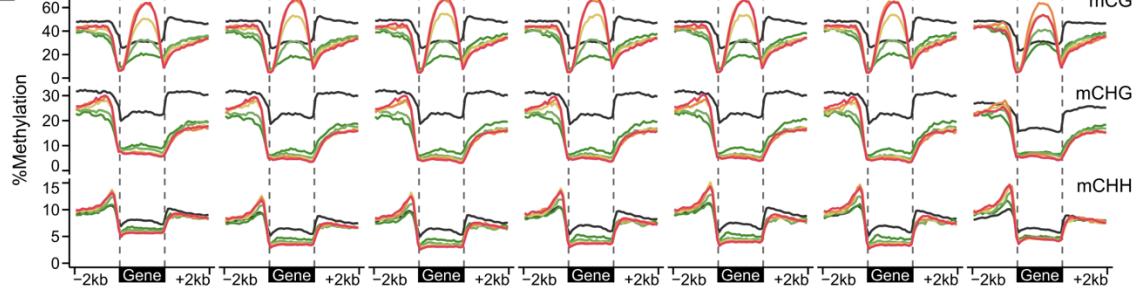
D



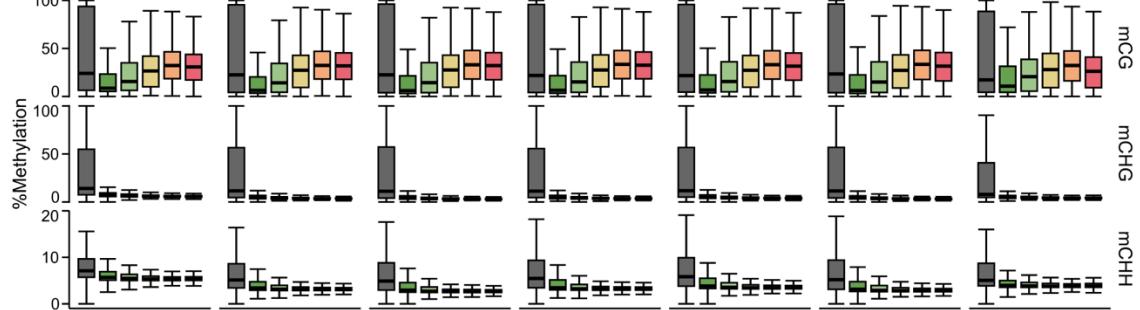
G



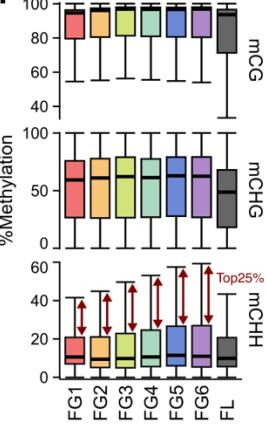
E



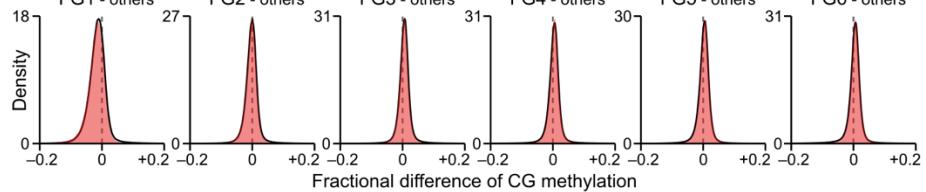
F



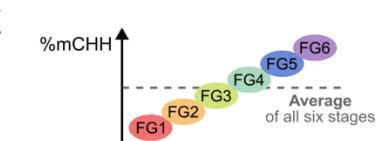
H



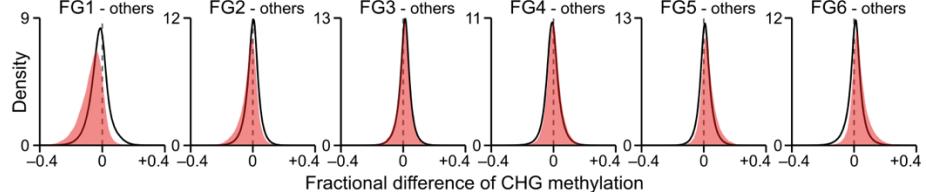
I



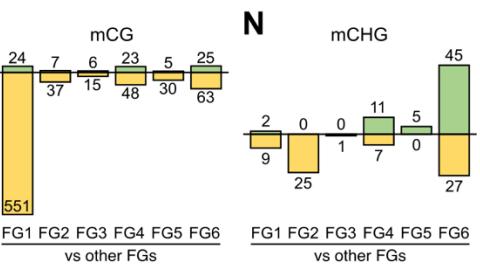
K



J



L



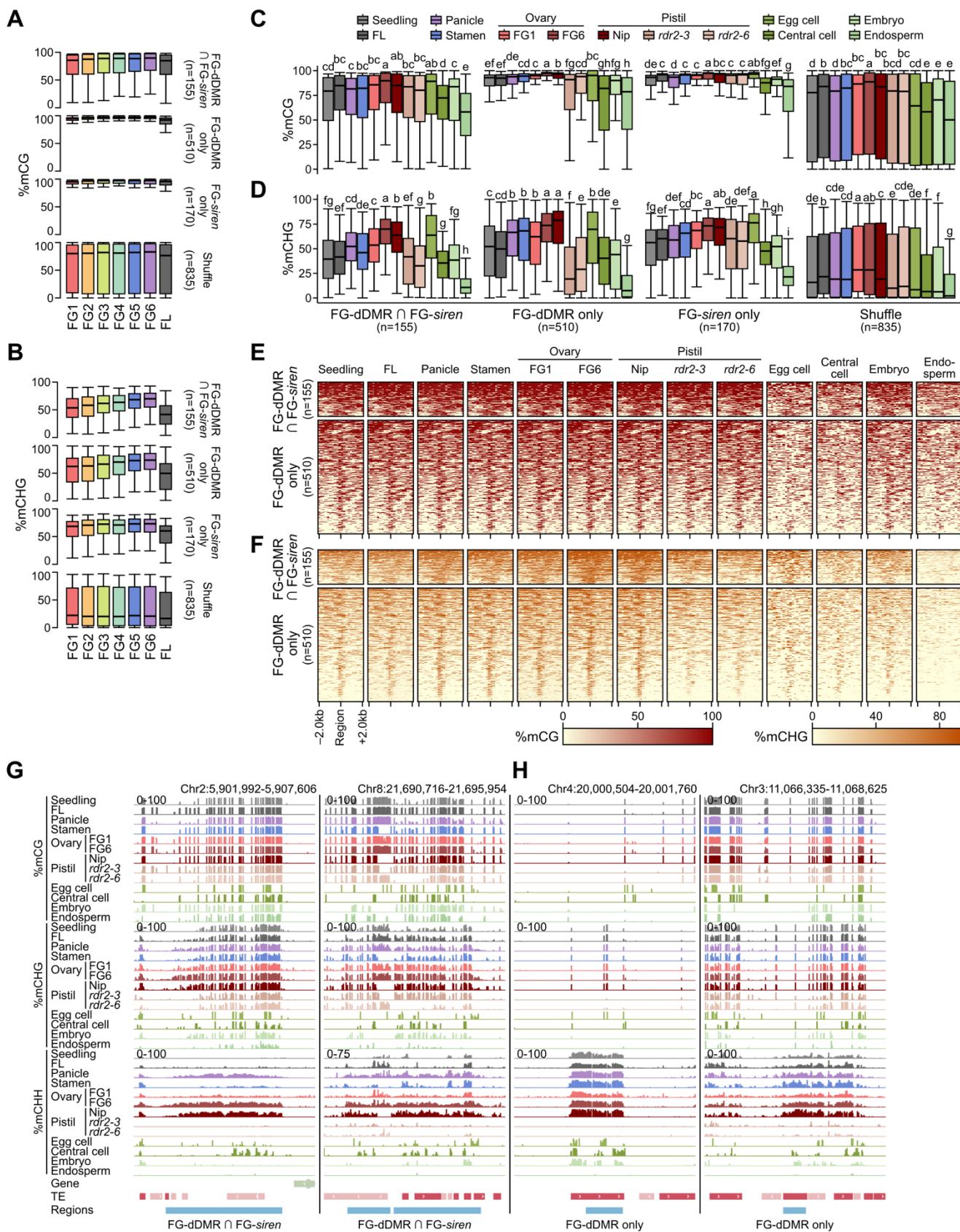
Key for Figures S1A, S1B, and S1G

FG1	FG3	FG5	FL
FG2	FG4	FG6	
—	—	—	—

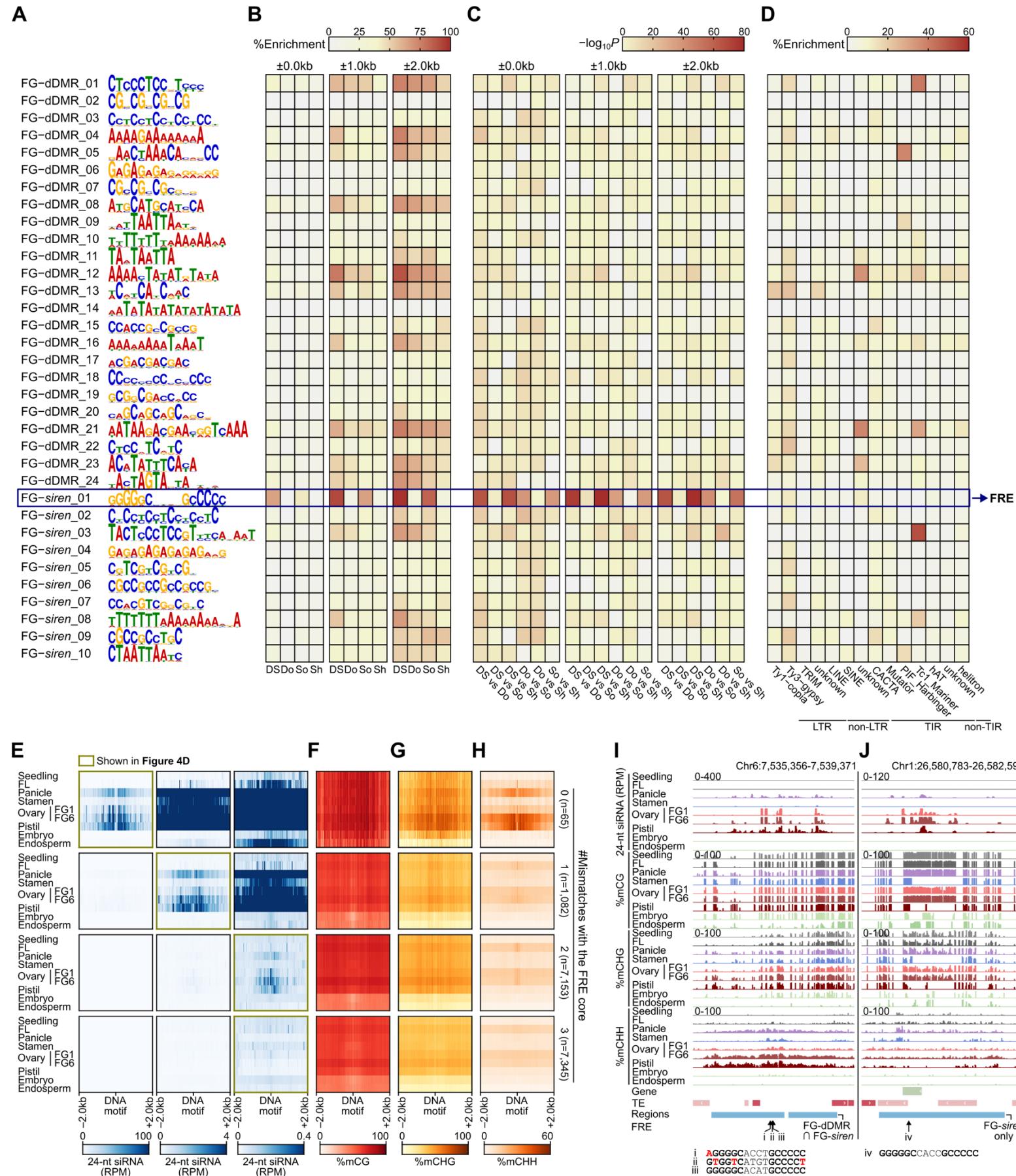
Key for Figures S1D–S1F

■ No expression	— Low	■ High
■ Very low	— Moderate	■ Very high

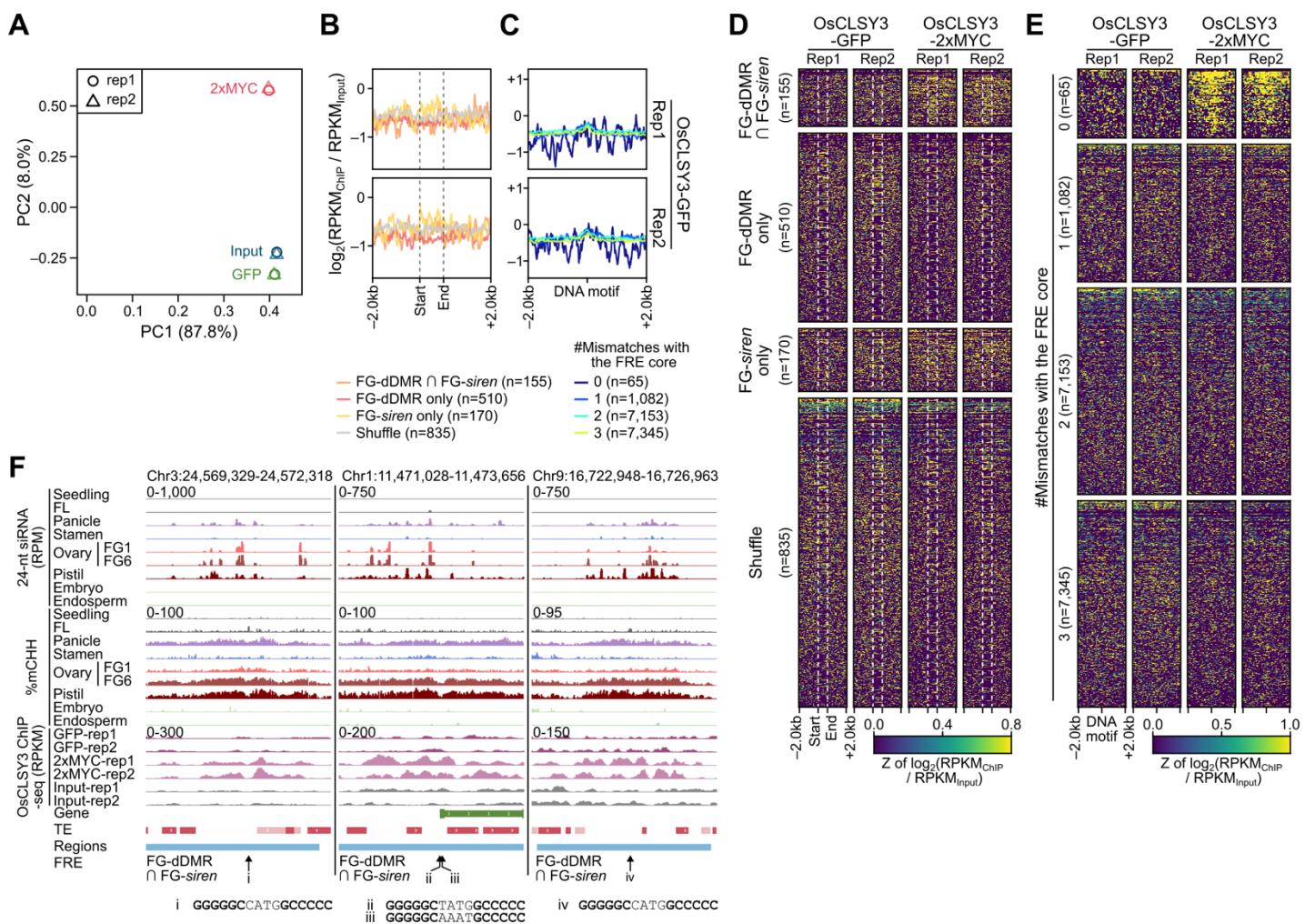
## Figure S2



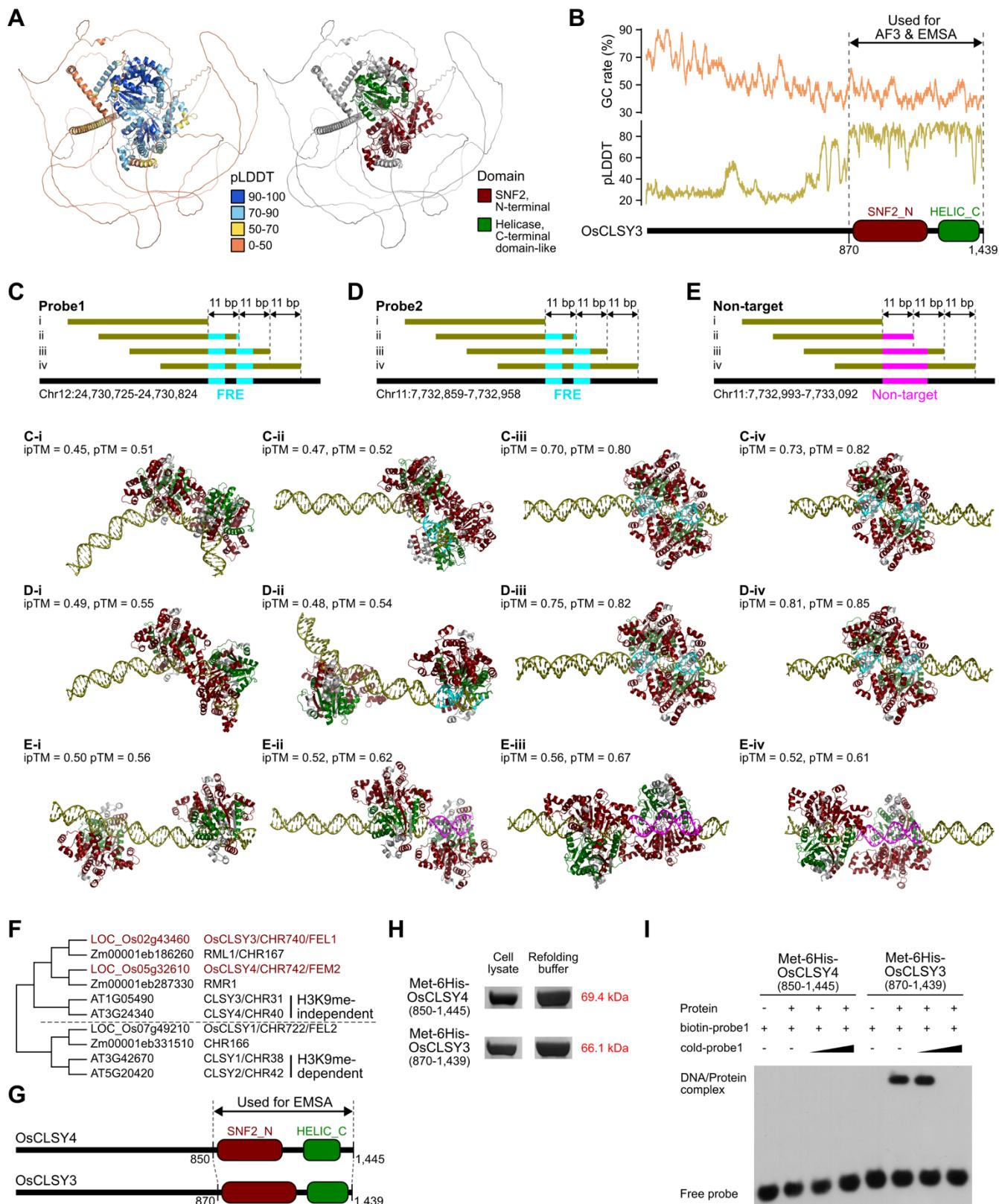
## Figure S3



## Figure S4



## Figure S5



**Table S1. Basic statistics of NGS datasets.**

(A) EM-seq analysis

Sample	Raw reads	Clean reads	Mapped reads	Mapped ratio	Deduplicated mapped reads
FG1rep1	201,705,724	189,478,412	124,055,126	65.47%	100,662,712
FG1rep2	197,040,689	184,962,464	118,629,907	64.14%	98,302,034
FG2rep1	161,898,510	150,796,049	96,343,428	63.89%	79,170,533
FG2rep2	180,819,375	170,326,475	110,775,452	65.04%	90,296,026
FG3rep1	197,086,799	184,333,979	121,108,349	65.70%	96,061,461
FG3rep2	171,135,451	160,469,535	105,701,372	65.87%	86,362,307
FG4rep1	172,281,939	162,381,256	102,255,974	62.97%	83,553,366
FG4rep2	189,146,375	177,301,160	111,161,322	62.70%	90,763,806
FG5rep1	182,961,341	171,031,339	106,480,601	62.26%	86,281,638
FG5rep2	157,844,691	147,533,135	90,925,300	61.63%	74,687,824
FG6rep1	176,061,556	164,865,978	103,017,429	62.49%	82,897,297
FG6rep2	172,587,127	161,412,931	101,507,737	62.89%	81,987,682
FLrep1	150,823,318	140,422,985	88,451,362	62.99%	72,301,079
FLrep2	172,387,236	160,817,717	103,305,286	64.24%	82,827,610
Seedlingrep1 <sup>a</sup>	113,738,486	113,606,330	88,011,802	77.47%	59,868,541
Seedlingrep2 <sup>a</sup>	89,310,616	89,200,407	68,870,510	77.21%	49,528,318
Paniclerep1 <sup>a</sup>	78,543,722	78,230,430	53,610,731	68.53%	47,326,950
Paniclerep2 <sup>a</sup>	84,316,534	84,261,819	56,390,082	66.92%	50,763,146
Stamen <sup>a</sup>	67,690,932	67,670,082	56,473,250	83.45%	42,126,345
Pistil_Nip <sup>a</sup>	67,417,309	67,392,749	54,981,418	81.58%	39,374,547
Pistil_osrdr2-3 <sup>a</sup>	67,555,642	67,531,715	55,477,835	82.15%	40,900,504
Pistil_osrdr2-6 <sup>a</sup>	67,635,731	67,613,398	53,725,512	79.46%	38,928,606
Embryo <sup>b</sup>	87,653,507	87,158,855	45,149,233	51.80%	34,820,514
Endosperm <sup>b</sup>	69,989,254	68,463,259	33,920,426	49.55%	28,454,388

<sup>a</sup>Wang et al., 2022<sup>19</sup>

<sup>b</sup>Zemach et al. 2010<sup>32</sup>

(B) sRNA-seq analysis

Sample	Raw reads	Clean reads	Mapped reads	Mapped ratio	Uniquely mapped reads	Uniquely mapped siRNA reads
FG1rep1	8,376,933	8,368,871	7,388,781	88.29%	1,894,816	1,672,032
FG1rep2	6,479,320	6,470,706	5,752,038	88.89%	1,328,803	1,206,493
FG2rep1	10,619,935	10,602,487	9,526,112	89.85%	2,490,941	2,331,922
FG3rep1	12,757,581	12,752,447	11,546,111	90.54%	4,012,298	3,860,447
FG4rep1	11,415,778	11,408,128	10,200,711	89.42%	4,054,435	3,864,641
FG4rep2	10,715,812	10,702,560	9,857,170	92.10%	3,004,973	2,902,440
FG5rep1	9,997,233	9,991,064	9,158,092	91.66%	2,947,157	2,737,779
FG5rep2	12,332,949	12,329,409	11,237,408	91.14%	4,171,290	3,988,792
FG6rep1	16,527,977	16,510,211	15,446,884	93.56%	3,621,224	3,382,962
FG6rep2	11,520,964	11,515,727	10,793,906	93.73%	2,680,229	2,446,920
FLrep1	9,098,086	9,093,866	6,922,716	76.13%	492,435	407,575
FLrep2	12,497,181	12,491,539	11,051,458	88.47%	763,838	662,389

Seedling-rep1 <sup>a</sup>	24,875,329	24,875,329	21,669,900	87.11%	10,435,683	9,067,513
Seedling-rep2 <sup>a</sup>	25,131,542	25,131,542	21,902,920	87.15%	9,003,575	7,960,481
Panicle-rep1 <sup>a</sup>	25,030,595	25,030,595	16,826,431	67.22%	6,122,121	5,752,858
Panicle-rep2 <sup>a</sup>	27,227,675	27,227,675	23,087,842	84.80%	8,910,300	8,151,884
Stamen <sup>a</sup>	21,045,563	21,045,563	16,266,169	77.29%	6,281,804	5,449,653
Pistil_Nip <sup>a</sup>	19,986,635	19,986,635	17,336,976	86.74%	8,459,326	8,247,440
Pistil_osrdr2-3 <sup>a</sup>	26,506,662	26,506,662	21,258,174	80.20%	4,793,548	3,764,110
Pistil_osrdr2-6 <sup>a</sup>	26,048,327	26,048,327	22,569,382	86.64%	3,989,207	3,161,748
Embryo <sup>b</sup>	26,233,129	25,408,968	21,099,082	83.04%	13,149,723	12,902,073
Endosperm <sup>b</sup>	22,605,850	20,587,987	16,683,200	81.03%	10,239,133	9,866,412

<sup>a</sup>Wang et al., 2022<sup>19</sup>

<sup>b</sup>Rodrigues et al. 2013<sup>45</sup>

### (C) mRNA-seq analysis

Sample	Raw reads	Clean reads	Mapped reads	Mapped ratio
FG1-rep1	47,755,217	42,991,639	42,600,011	99.09%
FG1-rep2	50,719,996	45,274,113	44,738,582	98.82%
FG1-rep3	51,496,840	46,475,959	46,012,910	99.00%
FG2-rep1	54,075,402	49,099,302	48,610,288	99.00%
FG2-rep2	51,767,009	46,782,721	46,308,183	98.99%
FG2-rep3	51,369,606	47,094,643	46,477,021	98.69%
FG3-rep1	55,352,130	49,947,514	49,154,624	98.41%
FG3-rep2	55,072,853	49,648,770	49,168,215	99.03%
FG3-rep3	50,923,899	46,167,540	45,625,854	98.83%
FG4-rep1	53,615,273	48,723,697	48,102,360	98.72%
FG4-rep2	54,813,797	49,908,086	49,261,362	98.70%
FG4-rep3	54,069,607	47,830,238	47,203,592	98.69%
FG5-rep1	50,790,301	46,748,581	46,352,125	99.15%
FG5-rep2	55,727,102	50,732,008	50,173,530	98.90%
FG5-rep3	49,204,070	44,913,001	44,518,632	99.12%
FG6-rep1	49,166,455	43,838,656	43,222,224	98.59%
FG6-rep2	59,577,899	54,676,809	54,116,572	98.98%
FG6-rep3	49,957,205	45,762,474	45,256,718	98.89%
FL-rep1	59,451,516	54,496,114	52,267,228	95.91%
FL-rep2	59,660,655	54,809,481	52,698,382	96.15%
FL-rep3	53,703,397	48,802,227	46,975,275	96.26%

**Table S2. Probes used in this study.**

Name	Sequence (5'-3')
<b>AF3 prediction</b>	
AF3_Probe1-short_F	ATGCGGGGGCCGACGCCCTAGT
AF3_Probe1-short_R	ACTAGGGGGCGTCGGCCCCCGCAT
AF3_Probe2-short_F	AAGTGGGGGCCTGGCCCCCACTC
AF3_Probe2-short_R	GAGTGGGGCCAAGGCCCACTT
AF3_Non-target-short_F	AGGTATATCCACCCATGCTACAA
AF3_Non-target-short_R	TTGTAGCATGGGTGGATATAACCT
AF3_Probe1-i_F	GTAAATTATAAAAAGTATATTTAATTAAATTCCATAATGGCATGC
AF3_Probe1-i_R	GCATGCCATTATGAAATTAAATTAAATACTTTTATAATTAC
AF3_Probe1-ii_F	AAAAGTATATTTAAATTAAATTCCATAATGGCATGCAGCG
AF3_Probe1-ii_R	CGTCGGCCCCCGCATGCCATTATGGAATTAAATTAAATACTTT
AF3_Probe1-iii_F	TTTAAATTAAATTCCATAATGGCATGCAGCGCTAGTGA
AF3_Probe1-iii_R	TCACTAGGGCGTCGGCCCCCGCATGCCATTATGGAATTAAATTAA
AF3_Probe1-iv_F	TTCCATAATGGCATGCAGCGCTAGTGAATTCAATTAGC
AF3_Probe1-iv_R	GCTAATTGAATTCACTAGGGCGTCGGCCCCCGCATGCCATTATGGAA
AF3_Probe2-i_F	AATCTCCTCGGTAAAGACACAGTATAGGGCGGATATAGAGGTAAGAAGT
AF3_Probe2-i_R	ACTTCTTACCCCTATATCGCCCCTACTGTGCTTACCGAGGAGATT
AF3_Probe2-ii_F	TAAGACACAGTATAGGGCGGATATAGAGGTAAGAAGTGGCTTGG
AF3_Probe2-ii_R	CCAAGGCCCACTTCTACCCCTATATCGCCCCTACTGTGCTTAA
AF3_Probe2-iii_F	ATAGGGCGGATATAGAGGTAAGAAGTGGCTTGGCCCCACTCAA
AF3_Probe2-iii_R	TTGAGTGGGGCCAAGGCCCACTTCTACCCCTATATCCGCCCCTAT
AF3_Probe2-iv_F	TATAGAGGTAAGAAGTGGGCCAGGCCACTCAATTACAGCGA
AF3_Probe2-iv_R	TCGCTGTAAATTGAGTGGGCCAGGCCACTTCTACCCCTATA
AF3_Non-target-i_F	GGTAGATCGTGGCCCCACTCTGCTACTTCTACATCCACCCCTAGGT
AF3_Non-target-i_R	ACCTAGGGCTGGATGTAGAAAGTAGCAAGAGTGGGCCAACGATCTACC
AF3_Non-target-ii_F	GGCCCCCACTCTGCTACTTCTACATCCACCCCTAGGTATATCCACCC
AF3_Non-target-ii_R	GGGGTGGATATACCTAGGGTGGATGTAGAAAGTAGCAAGAGTGGGCC
AF3_Non-target-iii_F	TTGCTACTTCTACATCCACCCCTAGGTATATCCACCCCATGCTACAACC
AF3_Non-target-iii_R	GGTTGTAGCATGGGTGGATATACCTAGGGTGGATGTAGAAAGTAGCAA
AF3_Non-target-iv_F	TACATCCACCCCTAGGTATATCCACCCCATGCTACAACCGTGACATGGTC
AF3_Non-target-iv_R	GACCATGTCACGGTTGTAGCATGGGTGGATATACCTAGGGTGGATGTA
<b>EMSA</b>	
EMSA_Probe1_F	TTCCATAATGGCATGCAGCGCTAGTGAATTCAATTAGC
EMSA_Probe1_R	GCTAATTGAATTCACTAGGGCGTCGGCCCCCGCATGCCATTATGGAA
EMSA_Probe2_F	TATAGAGGTAAGAAGTGGGCCAGGCCACTCAATTACAGCGA
EMSA_Probe2_R	TCGCTGTAAATTGAGTGGGCCAGGCCACTTCTACCCCTATA