







# The CLASSY family controls tissue-specific DNA methylation patterns in Arabidopsis

Ming Zhou<sup>1,2</sup>, Ceyda Coruh<sup>2</sup>, Guanghui Xu<sup>2</sup> , Laura M. Martins<sup>2</sup>, Clara Bourbousse<sup>2,4</sup> ,  
Alice Lambolez<sup>2,5,6</sup>  & Julie A. Law<sup>2,3</sup>  

DNA methylation shapes the epigenetic landscape of the genome, plays critical roles in regulating gene expression, and ensures transposon silencing. As is evidenced by the numerous defects associated with aberrant DNA methylation landscapes, establishing proper tissue-specific methylation patterns is critical. Yet, how such differences arise remains a largely open question in both plants and animals. Here we demonstrate that CLASSY1-4 (CLSY1-4), four locus-specific regulators of DNA methylation, also control tissue-specific methylation patterns, with the most striking pattern observed in ovules where CLSY3 and CLSY4 control DNA methylation at loci with a highly conserved DNA motif. On a more global scale, we demonstrate that specific *clsy* mutants are sufficient to shift the epigenetic landscape between tissues. Together, these findings reveal substantial epigenetic diversity between tissues and assign these changes to specific CLSY proteins, elucidating how locus-specific targeting combined with tissue-specific expression enables the CLSYs to generate epigenetic diversity during plant development.

<sup>1</sup>State Key Laboratory of Plant Physiology and Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou 310058, China. <sup>2</sup>Plant Molecular and Cellular Biology Laboratory, Salk Institute for Biological Studies, La Jolla 92037, USA. <sup>3</sup>Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA. <sup>4</sup>Present address: Institut de biologie de l'École normale supérieure (IBENS), École normale supérieure, CNRS, INSERM, PSL Université Paris, 75005 Paris, France. <sup>5</sup>Present address: RIKEN Center for Sustainable Resource Science, 1-7-22 Suehiro-cho, Tsurumi, Yokohama, Kanagawa 230-0045, Japan. <sup>6</sup>Present address: Department of Biological Sciences, Faculty of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan. email: [jlaw@salk.edu](mailto:jlaw@salk.edu)

Global DNA methylation patterns reflect a balance between pathways controlling the establishment, maintenance, and removal of this otherwise stable chromatin modification. Modulation within these pathways thus affords the opportunity to generate the unique DNA methylation patterns observed between somatic or reproductive tissues in plants<sup>1,2</sup> and mammals<sup>3–5</sup>. For example, during plant reproduction, both male and female tissues (the vegetative nucleus and the central cell, respectively) are hypomethylated due to reduced maintenance and increased demethylation activities<sup>2</sup>. In somatic plant tissues, the differences in DNA methylation patterns are less well characterized. Nonetheless, key roles for the RNA-directed DNA methylation (RdDM) pathway, which controls the establishment of DNA methylation via 24nt-siRNAs, are already emerging. Several recent studies in *Arabidopsis thaliana* (*Arabidopsis*)<sup>6–10</sup>, rice<sup>11,12</sup>, *Brassica rapa*<sup>13</sup>, and soybean<sup>14</sup> have compared the epigenomes of different somatic tissues and/or cell types. All of these studies implicate the RdDM pathway in the establishment of specific DNA methylation patterns, including several<sup>12,13</sup> that describe a highly expressed set of 24nt-siRNA producing loci initially designated as siren (small-interfering RNA in endosperm) loci based on their tissue of discovery<sup>12</sup>. In addition, many of these studies suggest that there may be more differences in methylation patterns between plant tissues than previously envisioned<sup>1,15</sup>. However, in each of these cases, how the RdDM pathway is regulated to generate the observed tissue- or cell type-specific DNA methylation patterns remains unclear.

Much of our understanding of how DNA methylation is established and maintained stems from the characterization of proteins discovered using genetic and biochemical approaches. These studies have revealed multiple interconnected methylation and demethylation pathways that now provide a framework for understanding how different DNA methylation patterns are generated. Briefly, the RdDM pathway utilizes the sequence information encoded within two types of non-coding RNAs (24nt-siRNAs generated by RNA POLYMERASE-IV (Pol-IV) and long intergenic transcripts generated by Pol-V) to establish DNA methylation in all contexts (CG, CHG, and CHH; H = A, T or C) at cognate sequences throughout the genome<sup>16,17</sup>. Once established, DNA methylation in each context is maintained via largely distinct pathways involving self-reinforcing loops between DNA and/or histone methylation readers and writers<sup>18,19</sup>. CG methylation is maintained by METHYLTRANSFERASE 1 (MET1) in connection with the VARIATION IN METHYLATION (VIM1-3) family of DNA-methylation readers<sup>20–22</sup>. CHG methylation is maintained by CHROMOMETHYLTRANSFERASE 3 (CMT3) in connection with three SU(VAR)3-9 HOMOLOG proteins (SUVH4-6) that read CHG methylation and write H3K9 methylation<sup>20,23–27</sup>. Finally, CHH methylation at short, euchromatic transposons is maintained primarily by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) in connection with SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1), an H3K9me reader<sup>28,29</sup>, while CHH methylation at long, pericentromeric transposons is primarily maintained by CMT2 in connection with SUVH4-6<sup>30,31</sup>. All contexts of DNA methylation are also subject to demethylation by methyl-cytosine glycosylases<sup>32</sup>. Together, these methylation and demethylation pathways constitute a dynamic system to control DNA methylation patterns.

Given the central role of 24nt-siRNAs in controlling where methylation is deposited, understanding how Pol-IV is regulated has been of keen interest and has already provided insights into how methylation patterns are modulated<sup>19,33,34</sup>. Initial investigation into the composition of the Pol-IV complex led to the identification of several co-purifying accessory factors<sup>35,36</sup>, including four related putative chromatin remodeling factors, CLASSY (CLSY) 1–4<sup>35</sup>. Recently, we demonstrated that the

CLSYs act as locus-specific regulators of the RdDM pathway by facilitating Pol-IV recruitment to distinct genomic targets in connection with different chromatin modifications<sup>37</sup>. CLSY1 and CLSY2 are required for the association of SHH1 with the Pol-IV complex, linking Pol-IV targeting by these CLSYs to H3K9me2, whereas targeting by CLSY3 and CLSY4 relies on CG methylation, but not H3K9me2<sup>37</sup>.

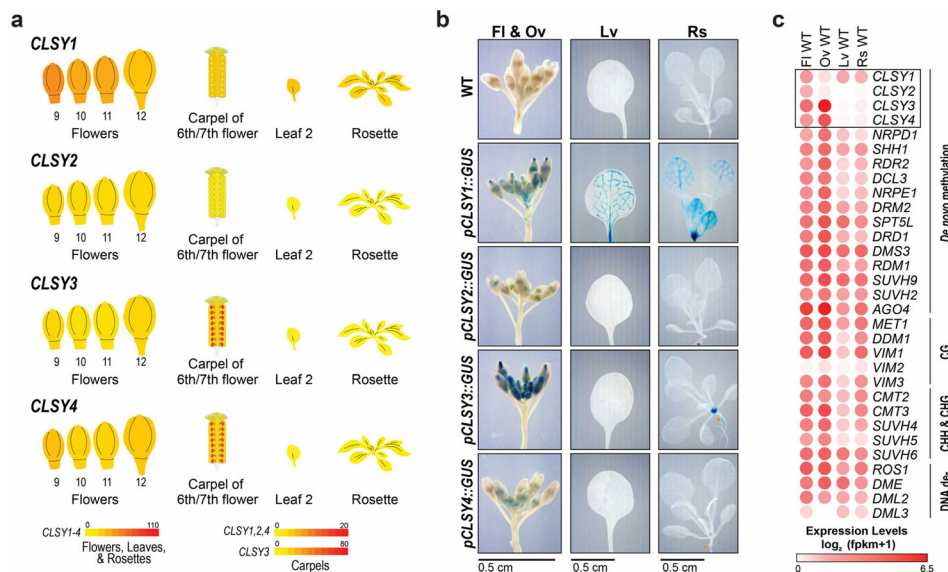
Here we describe tissue-specific roles for the four CLSY proteins. First, we show that tissues with different CLSY expression levels show distinct DNA methylation patterns and attribute these differences to the RdDM pathway. Next, we demonstrate that, depending on the tissue, different combinations of CLSYs, or even individual CLSY family members, control global 24nt-siRNA and DNA methylation patterns. For example, CLSY1 is required for the production of the vast majority of 24nt-siRNAs in leaf tissue, while CLSY3 is required for the production of high levels of 24nt-siRNAs at a few hundred siren loci that dominate the small RNA landscape in ovules and are enriched for a robust DNA sequence motif. Finally, we found that while *clsy3,4*-dependent loci show more variance across tissues than *clsy1,2*-dependent loci, either of these double mutants is sufficient to shift the 24nt-siRNA landscape from one tissue to nearly the wild-type landscape of another. Together, these findings establish the CLSYs as major, tissue-specific regulators of plant epigenomes, providing mechanistic insight into how diverse DNA methylation patterns are generated during plant development.

## Results

**The CLSYs are unique in their degree of tissue-specific expression.** As a first step in determining whether the CLSYs play a role in shaping tissue-specific patterns of DNA methylation, their expression patterns were assessed during *Arabidopsis* development. Based on previously published data available in the ePlant database<sup>38</sup>, (Supplementary Fig. 1 and Fig. 1a), four tissues [flower buds (Fl; stage 12 and younger), ovules (Ov; mature, but unfertilized), 1<sup>st</sup> and 2<sup>nd</sup> true leaves (Lv; 25 day old plants), and rosettes (Rs; 15 days old, no roots)] were selected for further characterization using GUS reporters driven by the CLSY promoters (Fig. 1b) and mRNA sequencing (mRNA-seq) (Supplementary Data 1 and Fig. 1c). For the GUS reporters, the CLSY promoter regions were vetted in constructs driving the expression of the CLSY genes, which complemented the 24nt-siRNA defects at the *clsy*-dependent loci identified in Zhou et al.<sup>37</sup> (Supplementary Fig. 2). Together the mRNA-seq and GUS reporters showed expression of all four CLSY genes in flower buds, though they are enriched in different tissues, and showed strong expression of CLSY3 and to a lesser extent CLSY4 in mature, but unfertilized ovules; in addition, they also showed expression of CLSY1 in both the 1<sup>st</sup> and 2<sup>nd</sup> true leaves of 25 day old plants and the rosettes of 15 day old plants (Fig. 1). Beyond confirming the tissue-specific expression patterns of the CLSYs, analysis of the mRNA-seq data revealed that these genes show the most diverse expression patterns when compared to the rest of the RdDM machinery, the main genes required for the maintenance of DNA methylation in the different sequence contexts, or the four demethylases (Fig. 1c). These findings demonstrate that the CLSYs are excellent candidates for regulating tissue-specific DNA methylation patterns during plant development.

## Tissues that express different CLSYs have distinct epigenomes.

Although DNA methylation patterns have been described for *Arabidopsis* flower, leaf, and rosette tissues individually, these patterns have not been compared to each other, and no DNA methylation data is available for ovules. Thus, to compare the epigenomes of these tissues, and assess the contribution of the



**Fig. 1** Tissue-specific expression patterns of the *CLSY* genes. **a** Relative expression levels of the four *CLSY* genes in select tissues from two ePlant expression viewers<sup>71–73</sup> (see also Supplementary Fig. 1a, b). Here, all tissues and organs are labeled as in the ePlant viewers<sup>38</sup> and the cartoon images are not to scale. **b** Images showing GUS expression levels from wild-type (WT) plants or plants harboring GUS reporters driven by the promoters of the four *CLSY*s (*pCLSY1-4::GUS*) in the following tissues: flower buds (FI; stage 12 and younger), unfertilized ovules (Ov), 1st and 2nd true leaves (Lv) from 25 day old plants, and rosettes (Rs; no roots) from 15 day old plants. For each set of images, a scale bar is shown below and the individual image presented is representative of two or more independent experiments. **c** Heatmap showing the expression levels of the main genes required for the establishment, maintenance, and removal of DNA methylation based on mRNA-seq data (Supplementary Data 1) from same tissues described in panel **b**. The *CLSY* genes are surrounded by a black box.

RdDM pathway in shaping these landscapes, small RNA-seq (smRNA-seq) (Supplementary Data 2) and MethylC sequencing (methylC-seq) (Supplementary Data 3) data sets were generated in parallel from both wild-type samples and RdDM pathway mutants (*e.g. nrpd1*; hereafter named *pol-iv*).

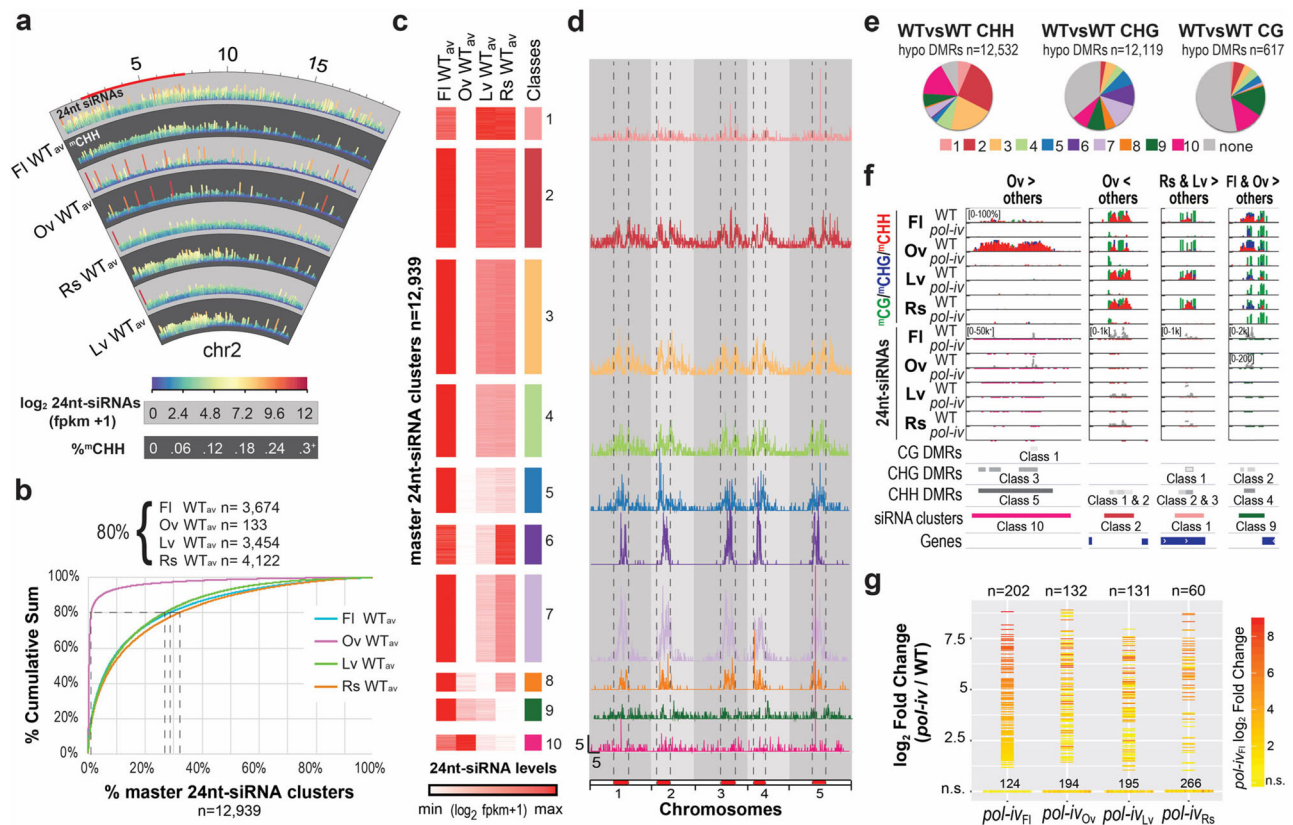
Initial comparisons of 24nt-siRNA levels on a chromosomal scale revealed distinct profiles amongst the four tissues and confirmed a strong dependency on *pol-iv* across all tissues (Fig. 2a and Supplementary Fig. 3a). Relative to flowers, the patterns in leaf and rosette tissues differ most in the chromosome arms and edges of the pericentromeric heterochromatin, where 24nt-siRNA levels are lower (Fig. 2a and Supplementary Fig. 3a). However, the most striking difference was observed in ovules, where 24nt-siRNA levels are globally low with the exception of a small number of high expressing regions. For CHH methylation, which is known to be controlled by both the RdDM and CMT2 pathways<sup>30,31</sup>, the differences between flower, leaf, and rosette tissues in both the wild-type and *pol-iv* mutants are less evident at this scale (Fig. 2a and Supplementary Fig. 3a). However, for ovules, the CHH methylation profiles were strongly *pol-iv*-dependent and highly correlated with 24nt-siRNA levels (Fig. 2a and Supplementary Fig. 3a), demonstrating a major role for the RdDM pathway in this tissue. Taken together, these data demonstrate that epigenetic features associated with the RdDM pathway differ between flower, ovule, leaf, and rosette tissues to a degree that is apparent even on a genome-wide scale.

To compare the profiles of these tissues in more detail, 24nt-siRNA clusters were first identified based on all three wild-type replicates from smRNA-seq experiments using ovule, leaf, and rosette samples, as was done previously for flowers<sup>37</sup> (Supplementary Data 4). The data from all four tissues were then merged to generate a list of “master 24nt-siRNA clusters” ( $n = 12,939$ ) representing all the clusters amongst the four tissues (Supplementary Data 5). These 24nt-siRNA clusters were then grouped into 10 classes based on their expression levels and comparisons

between the expression patterns and genomic distributions of each class revealed several tissue-specific trends (Fig. 2c, d and Supplementary Data 6). First, in ovules, the levels of 24nt-siRNAs were low in all classes except 8–10, which show increasing expression levels, respectively (Fig. 2c). Indeed, the top 133 expressing clusters in ovules, which correspond well with the most abundant 24nt-siRNA regions in the chromosome views (see red and dark orange bars in Fig. 2a and Supplementary Fig. 3a), account for 80% of all the 24nt-siRNAs (Fig. 2b and Supplementary Data 5). Given the similarity of these ovule loci to those described in rice<sup>12</sup> and Brassica<sup>13</sup>, which were distinguished from other 24nt-siRNA producing loci based on their high expression from a small number of loci, they are hereafter referred to by their initial designation as siren loci<sup>12</sup>. Second, the expression profiles in Class 1, Class 2, and to a lesser extent Class 3, are similar in flower, rosette, and leaf tissues (Fig. 2c). These clusters, which correspond to nearly half of all 24nt-siRNA clusters (47%), appear to be similarly acted upon by Pol-IV across diverse tissues, which is consistent with their relative enrichment in the chromosome arms (Fig. 2d). Third, the expression levels in Classes 4–8 are much lower in leaf tissue compared to rosette and/or flower tissues (Fig. 2c), revealing that 24nt-siRNA production in pericentromeric heterochromatin (Fig. 2d) is less robust in leaves. Finally, while flowers show strong expression across all classes, leaf and rosette tissues, like ovules, show low expression in specific classes, namely Class 8 for leaf and Class 9 for rosette tissues (Fig. 2c). Together, these analyses reveal that tissue-specific differences in 24nt-siRNA patterns vary based on chromosomal locations and range from being essentially absent in a single tissue to being highly dynamic across tissues, suggesting a tunable means to control 24nt-siRNA patterns.

To determine the extent to which the observed tissue-specific differences in 24nt-siRNA levels contribute to altered DNA methylation patterns, differentially methylated regions (DMRs) between tissues were first determined in 100 bp bins across the





**Fig. 2** Tissues with different *CLSY* expression patterns have distinct epigenetic landscapes. **a** Circular view of chromosome 2 (chr2) showing the patterns of 24nt-siRNAs (light grey) and CHH methylation (dark grey) in 5 kb bins based on the average (av) expression of three WT controls (WT<sub>av</sub>) from each tissue. The color scales are as indicated and the tracks are labeled every 5 Mb, with the pericentromeric heterochromatin, as designated in Yelina et al.<sup>69</sup>, marked in red. For CHH methylation, between 1 and 3 bins, depending on the tissue, had values > 0.3 (0.3<sup>+</sup>), but were capped at 0.3 to facilitate genome-wide visualization. **b** Cumulative sum plot based on the WT<sub>av</sub> expression levels for each tissue, showing the percentage and number (n) of 24nt-siRNA clusters (x-axis) required to reach 80% of the fpkm-normalized 24nt-siRNA reads (y-axis) in each tissue. **c** Heatmap showing 10 classes of 24nt-siRNA clusters based on the WT<sub>av</sub> expression levels for each tissue at the 12,939 master clusters. **d** Locations of the 24nt-siRNA clusters described in **c** along the 5 chromosomes. The pericentromeric regions are colored red and denoted by dashed lines. Scale bars for x-axis (Mb) and y-axis (clusters/100 kb bin) are indicated in the lower left corner of the Class 10 dataset. **e** Pie charts showing the proportions of hypo DMRs in each context that overlap with the 24nt-siRNA cluster classes colored as in **c**. **f** Screenshots of representative sites showing differential methylation between tissues. DNA methylation levels are shown as bar graphs while the 24nt-siRNAs are shown as point graphs, where the levels of 24nt-siRNAs are plotted and the regions containing 24nt-siRNAs are also underlined below the track in colors corresponding to the cluster class. Track scales, in brackets, apply to all lower tracks of the same tissue and methylation in the CG, CHG, and CHH contexts are shown in green, blue, and red, respectively. DMRs, colored based on the classes in Supplementary Fig. 4a, and 24nt-siRNA clusters, colored based on the classes in panel **c**, are indicated below. The > and < symbols in the headers refer to relative DNA methylation levels. **g** Plot showing the expression levels [ $\log_2$  FC (*pol-iv*/WT)] of the 326 *pol-iv*-upregulated transcripts represented as horizontal lines ( $\log_2$  FC  $\geq 1$  and FDR  $\leq 0.05$ ) or dots (not significantly upregulated; n.s.) in the four tissues, all colored based on the *pol-iv* expression level in flowers.

whole genome and then mapped onto the 24nt-siRNA landscape. Notably, this method required the identification of DMRs in all pairwise combinations of wild-type data sets (e.g. 3 wild-type ovule datasets versus 3 wild-type flower datasets = 9 pairwise combinations), and required 10%, 20%, or 40% changes for CHH, CHG and CG methylation, respectively (Supplementary Data 7). Thus, rather than capturing the full range of epigenetic differences, it reveals the most robust tissue-specific differences in methylation between each pair of tissues. When combined, these analyses yielded a total of 12,532 CHH and 12,119 CHG DMRs that were hypo methylated between at least two tissues (e.g. regions less methylated in ovules versus flowers or less methylated in leaves versus rosettes, etc., were determined in a recursive manner and then combined to generate a unique list of DMRs for each context), but only 617 hypo CG DMRs (Supplementary Data 7; “WTvsWT all tissues”). The majority of these DMRs overlap with 24nt-siRNA clusters (Fig. 2e) and,

when clustered by their methylation levels, these DMRs show tissue-specific differences that are most apparent at loci regulated by *pol-iv* (Supplementary Fig. 4a and Supplementary Data 8). Thus, as exemplified by the regions shown in Fig. 2f, these analyses not only reveal that DNA methylation patterns differ significantly between these tissues at thousands of loci, especially for non-CG methylation, but they also demonstrate that these differences are largely attributable to the RdDM pathway.

Finally, to assess the role of the RdDM pathway in facilitating the silencing of genes and transposons in a tissue-specific manner, differentially expressed loci in *pol-iv* mutants were identified using RNA-seq data from flower<sup>37</sup>, ovule, rosette and leaf tissues and combined to generate a list of *pol-iv*-dependent loci amongst all tissues (Supplementary Data 9 and 10). To visualize and compare the expression levels of these loci between tissues, a profile plot was generated where significantly upregulated loci ( $n = 326$ ;  $\log_2$  Fold Change (FC)  $\geq 1$  and FDR  $\leq 0.05$ ;

Supplementary Data 10) are shown as bars, colored relative to the *pol-iv* levels in flowers, while those not passing these thresholds are shown at the bottom as dots (Fig. 2g). While there are some loci that are upregulated across all tissues (red/orange lines), there are others that are most strongly up in flowers, but are not significantly up in the other tissues (orange/red dots at the bottom of the plots) and vice versa. As was previously shown in flowers<sup>37</sup>, the vast majority of upregulated loci in all these tissues show proximal decreases in DNA methylation and 24nt-siRNA levels (Supplementary Fig. 5a), consistent with their misregulation being a direct effect of an altered epigenome. Similar analyses were conducted on *pol-iv*-dependent downregulated loci. Although fewer in number ( $n = 213$ ; Supplementary Fig. 5b), and overall less associated with proximal changes in DNA methylation and 24nt-siRNA levels, those with associations tended to lose, rather than gain, epigenetic signals (Supplementary Fig. 5c). This pattern is similar to the previously described scenario for the *ROSI* gene<sup>32</sup> and is consistent with observations that promoter CHH methylation is positively correlated with gene expression in rice<sup>39</sup> and maize<sup>40</sup>, suggesting that promotion of gene expression by DNA methylation may be more common than previously appreciated. Together, these findings demonstrate that tissue-specific differences in gene regulation and transposon silencing also rely on the RdDM pathway, highlighting major roles for this pathway in shaping the epigenomes of flower, ovule, leaf, and rosette tissues by controlling patterns of 24nt-siRNAs, DNA methylation, and gene expression.

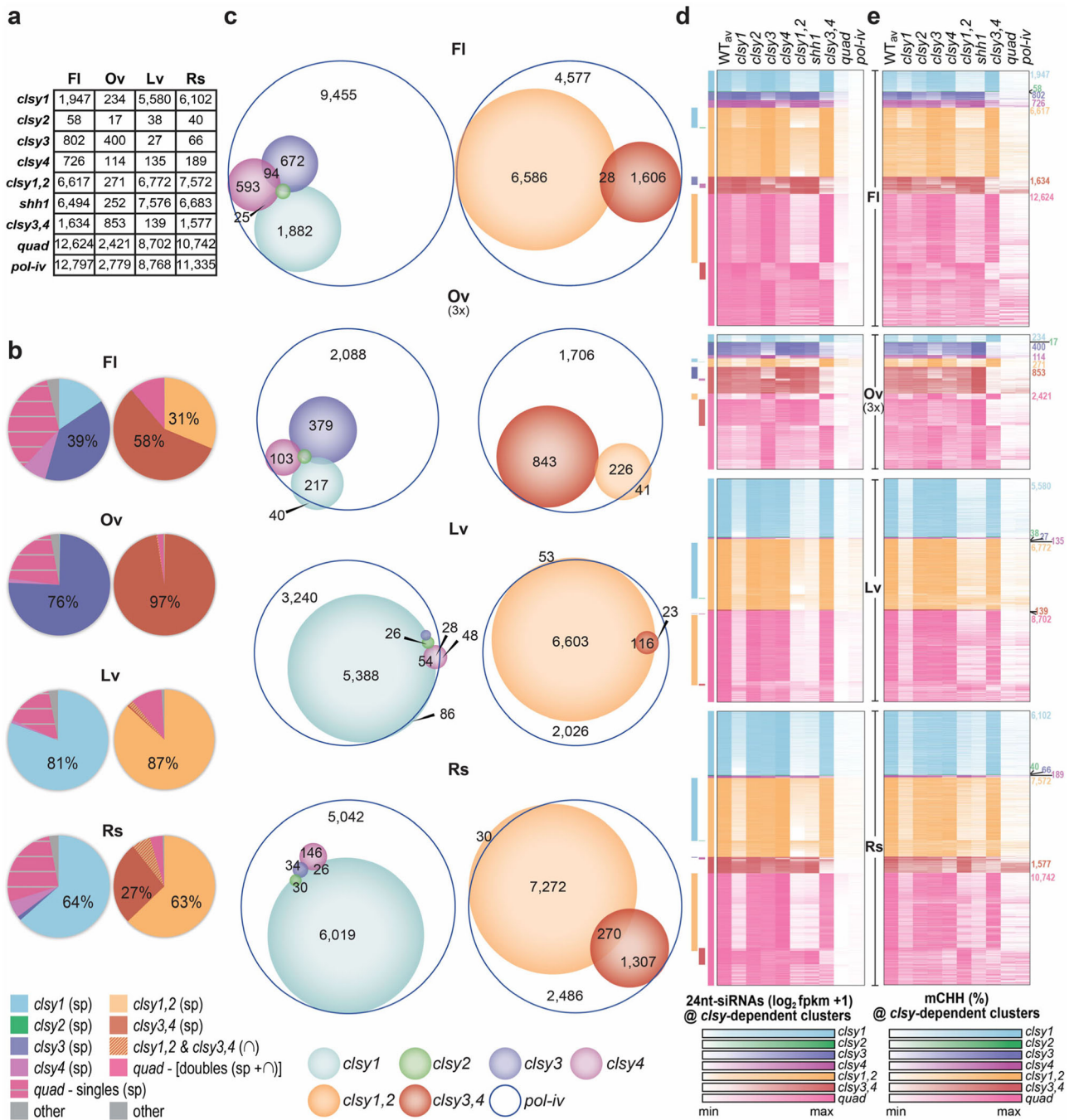
**The CLSYs control the epigenome in a locus- and tissue-specific manner.** To determine what roles the CLSY proteins play in shaping the epigenomes of the aforementioned tissues, differentially expressed 24nt-siRNA clusters and DMRs were identified in *clsy* single, double and quadruple mutants relative to tissue-matched wild-type controls at the master set of 24nt-siRNA clusters (Supplementary Data 11) or in 100 bp bins (Supplementary Data 12), respectively. The hypo CHH and, to a lesser extent, hypo CHG DMRs in each mutant and tissue were largely *pol-iv*-dependent (Supplementary Data 12), similar to what we observed in the 24nt-siRNA clusters (Supplementary Fig. 3a). Despite this dependence, many hypo CHH DMRs, especially in ovule and leaf tissues, did not overlap with reduced 24nt-siRNA clusters using our standard cutoff ( $\log_2 \text{FC} \leq -1$ ;  $\text{FDR} \leq 0.01$ ). Investigation into this discrepancy revealed that 24nt-siRNAs were largely present at these DMRs, but their expression was very low in wild-type samples making further reduction in mutant backgrounds difficult to resolve statistically. Thus, to capture a larger proportion of regions controlled by the RdDM pathway in each tissue, 24nt-siRNA clusters that overlapped with *pol-iv*-dependent hypo CHH DMRs, and were reduced by at least 25% compared to wild-type controls ( $\log_2 \text{FC} < -0.415$ ; Supplementary Fig. 6a), were added to the lists of differentially expressed clusters for each tissue and genotype (Supplementary Data 11 and Fig. 3a). These regions were then used to assess both 24nt-siRNA and DNA methylation patterns across tissues and genotypes (Fig. 3 and Supplementary Figs. 7–10).

As a percentage of wild-type 24nt-siRNA abundance (Fig. 3b), reduced 24nt-siRNA clusters (Fig. 3c) or hypo CHH DMRs (Supplementary Figs. 7–10), the four tissues show strikingly different degrees of dependence on the four CLSYs in a manner consistent with their observed expression patterns (Fig. 1). In terms of abundance, nearly all *pol-iv*-dependent 24nt-siRNAs in ovules reside in clusters reduced in *clsy3,4* mutants (Fig. 3b; 97%) while in leaves they reside in clusters reduced in *clsy1,2* mutants (Fig. 3b; 87%). This is in contrast to rosette and flower tissues where there is a more even reliance on these CLSY pairs, although

there is a skew towards loci reduced in the *clsy1,2* mutant for rosette tissue and the *clsy3,4* mutant for flower tissue (Fig. 3b). Notably, when compared to flowers, most of the differences in these tissues stem from an increased proportion of 24nt-siRNAs in clusters reduced in *clsy1* mutants in rosette (64%) and leaf (81%) tissues or by those reduced in *clsy3* mutants in ovule (76%) tissue (Fig. 3b), highlighting the impact even a single CLSY protein can have in controlling 24nt-siRNA populations in a tissue-specific manner.

In terms of 24nt-siRNA clusters and hypo CHH DMRs, their patterns are similarly diverse between tissues and are largely in agreement with the 24nt-siRNA abundance data. For example, in rosette and leaf tissues the largest numbers of reduced 24nt-siRNA clusters and hypo CHH DMRs are observed in the *clsy1* mutant and relatively few sites were affected in the *clsy2*, *clsy3* or *clsy4* mutants (Fig. 3c and Supplementary Figs. 7, 8), reinforcing the dominance of CLSY1 in shaping the epigenomes of leaf and rosette tissues. Here, the difference between tissues becomes more evident when comparing the effects of the *clsy* double mutants, as significantly more reduced 24nt-siRNA clusters and hypo CHH DMRs are observed in the *clsy3,4* double mutant in rosette versus leaf tissues. For ovules, the genetic dependency is the opposite, as the *clsy3* and *clsy3,4* mutants, rather than the *clsy1* and *clsy1,2* mutants, have the largest numbers of reduced 24nt-siRNA clusters and hypo CHH DMRs (Fig. 3c and Supplementary Fig. 9). Notably, for ovules, and to a lesser extent flowers, the number of reduced 24nt-siRNA clusters and hypo CHH DMRs in the *clsy1* and *clsy4* mutants is higher than expected based on the 24nt-siRNA abundance data. Therefore, in addition to the highly expressed 24nt-siRNA clusters controlled by CLSY3, more lowly expressed clusters controlled by CLSY1 and CLSY4 (Fig. 3c and Supplementary Figs. 9, 10) also play roles in shaping the 24nt-siRNA and DNA methylation landscapes in both these tissues. Finally, across all these tissues, CLSY2 has a minimal effect (Fig. 3b, c and Supplementary Figs. 7–10), suggesting that this CLSY may play a supporting role in general, or perhaps a more dominant role in tissues or cell types not yet profiled. Taken together, these findings establish the CLSY proteins as tissue-specific regulators of the RdDM pathway.

Beyond the newly discovered tissue-specific roles for the CLSY family, comparisons between 24nt-siRNA and DNA methylation levels across the four tissues demonstrate that several behaviors initially observed in flowers<sup>37</sup> appear to be conserved properties of the CLSY family. First, with the exception of leaf and rosette tissues that are dominated by the activity of CLSY1, clusters with reduced expression in the *clsy* single and double mutants are largely non-overlapping for all the other tissues (Fig. 3c). Furthermore, these reductions, especially for the double mutants, approach the levels observed in the *pol-iv* mutant and are less reduced, if at all, in the other *clsy* single mutants as compared to wild-type controls (Fig. 3d and Supplementary Figs. 7–10). Together, these findings demonstrate that controlling 24nt-siRNAs in a locus-specific manner is a common feature of the CLSY proteins across tissues. Second, similar genetic relationships between CLSY family members were observed across all tissues: the *clsy* doubles displayed larger 24nt-siRNA reductions and affect more loci than their single mutant counterparts, and in all cases the *clsy* quadruple mutant essentially phenocopies *pol-iv* (Fig. 3d and Supplementary Figs. 7–10), demonstrating that the CLSY family is responsible for controlling Pol-IV function across a diverse set of tissues. Third, across all tissues the *shh1* mutant phenocopies the *clsy1,2* double mutant, suggesting a conserved mechanism of targeting for these CLSYs irrespective of the tissue (Fig. 3d). Finally, for all combinations of *clsy* mutants, the loss of 24nt-siRNAs is associated with reduced CHH (and to a lesser extent CHG) methylation (Fig. 3e and Supplementary Figs. 7–10). Collectively, these findings demonstrate



**Fig. 3 The CLSYs regulate 24nt-siRNA and DNA methylation patterns in a tissue-specific manner.** **a** Table showing the numbers of downregulated 24nt-siRNA clusters in the indicated tissues and genotypes. **b** Pie charts showing the proportions of 24nt-siRNAs from all *pol-iv*-dependent clusters in a given tissue that are present in each *clsy*-dependent category based on a representative WT control. The charts are colored as indicated below and those on the left include *clsy* single and *clsy* quadruple categories while those on the right include *clsy* double and *clsy* quadruple categories. Here, “sp” indicates that only 24nt-siRNAs that reside in clusters specifically reduced in the indicated genotype are included (i.e. 24nt-siRNAs in clusters not reduced in any other single mutant (left) or double mutant (right)), “∩” indicates clusters that are shared between the indicated genotypes, and “-” indicates clusters subtracted from the *clsy* quadruple set of clusters. **c** Scaled Venn diagrams showing the relationships between reduced 24nt-siRNA clusters in *pol-iv* and the *clsy* mutants colored as indicated below. For readability, only overlaps >20 are labeled. A small number of overlaps is not shown due to spatial constraints. Unscaled Venn diagrams showing all the overlaps is presented in Supplementary Fig. 6b. **d** and **e** Heatmaps showing the 24nt-siRNA levels **d** or CHH methylation levels **e** for the WT<sub>av</sub> or the indicated mutants at the reduced 24nt-siRNA clusters identified for each tissue. The heatmaps for each *clsy*-dependent category are colored as indicated below. For **d**, the heatmaps for each tissue are ranked (high to low) for each *clsy*-dependent category and the overlaps between categories are indicated by the colored bars on the left. For **e**, the heatmaps are ordered as in **d**. The number of clusters in each category are indicated on the right. For **c**, **d**, and **e** the data for oviducts are scaled as 3x compared to the other tissues due to the reduced number of affected loci.



that CLSYs act as locus-specific regulators of the RdDM pathway in diverse tissues and control tissue-specific patterns of DNA methylation in a manner consistent with their expression profiles during plant development.

**CLSY3 targets Pol-IV to siren loci to control DNA methylation levels.** Given the dominant roles for CLSY3 and CLSY4 in ovules (Fig. 3), connections between these CLSYs and siren loci, which dominate the 24nt-siRNA landscape in ovules (see red and dark orange bars in Fig. 2a and Supplementary Fig. 3a), were investigated. Based on 24nt-siRNA cluster overlaps, siren loci are almost exclusively controlled by CLSY3 and CLSY4. Of the 133 siren loci, 132 overlap with *clsy3,4*-dependent 24nt-siRNA clusters and 106 overlap with *clsy3*-dependent clusters in ovules (Fig. 4a). In contrast, none of the *clsy4*-dependent 24nt-siRNA clusters overlap with siren loci, demonstrating that although there is some redundancy with CLSY4, CLSY3 is the primary CLSY acting at these sites. Consistent with these overlaps, 24nt-siRNA levels at siren loci in ovules and flowers are specifically reduced in *clsy3* and *clsy3,4* mutants, with the *clsy3,4* double mutant showing stronger reductions than observed in the *clsy3* single mutant (Fig. 4b). These analyses also further illustrate the tissue-specific expression of these clusters as their levels are significantly lower in flower tissue and are barely detectable in leaf and rosette tissues (Fig. 4b). Notably, the global pattern of 24nt-siRNA levels in ovule tissue is highly reproducible (Supplementary Fig. 11a), as are the regions identified as siren loci (Supplementary Fig. 11b) and the reductions in 24nt-siRNA levels displayed in the *clsy* mutants (Fig. 4b and Supplementary Fig. 11c, d). For DNA methylation, the patterns are also reproducible and the tissue-specificity is even more striking, as DNA methylation levels in all sequence contexts are reduced at siren loci in the *clsy3,4* double, *clsy* quadruple, and *pol-iv* mutants in ovules (Fig. 4c), but are already low in the leaf and rosette tissues and remain largely unaffected in the other RdDM mutants (Supplementary Figs. 11e, f, 12a). Interestingly, little to no reductions in DNA methylation are observed at siren loci in the *clsy3* single mutant in ovules (Fig. 4c and Supplementary Fig. 11e, f), suggesting that the 24nt-siRNAs remaining in this mutant are sufficient to maintain near wild-type levels of methylation. Although it cannot be excluded that some additional floral tissues produce 24nt-siRNAs at siren loci, these findings suggest that most, if not all, 24nt-siRNAs at these sites arise from ovules, that they are required to maintain DNA methylation in all sequence contexts, and that their production is most strongly dependent on CLSY3.

To determine whether CLSY3 acts directly at siren loci and whether it is required for Pol-IV recruitment to these sites, chromatin immunoprecipitation (ChIP)-seq experiments (Supplementary Data 13) were conducted using flower tissue from wild-type (Col) plants or *clsy3* mutant plants expressing a 3xFLAG-tagged version of CLSY3. This construct utilized the endogenous CLSY3 promoter (pC3) and was shown to complement the 24nt-siRNA defects observed in *clsy3* mutant flowers (pC3::CLSY3-3xFLAG; Supplementary Fig. 12b). From the resulting ChIP-seq data, 102 CLSY3 peaks were identified relative to the wild-type control (Col ChIP versus CLSY3-3xFLAG ChIP FC > 2.5 and FDR < 0.001; Supplementary Data 14), and these peaks were also enriched relative to an input control (Fig. 4d). Analyses of these CLSY3 peaks revealed the presence of a rare ( $n = 223$  occurrences genome-wide) and highly conserved motif under  $\sim 2/3$  of the peaks ( $n = 116$  occurrences), with a higher proportion and number of motifs identified at sites with the most CLSY3 enrichment (Fig. 4d and Supplementary Fig. 12c). In addition, despite the fact that the ChIP was conducted in flower tissue, these peaks were found to overlap with siren loci (94/102)

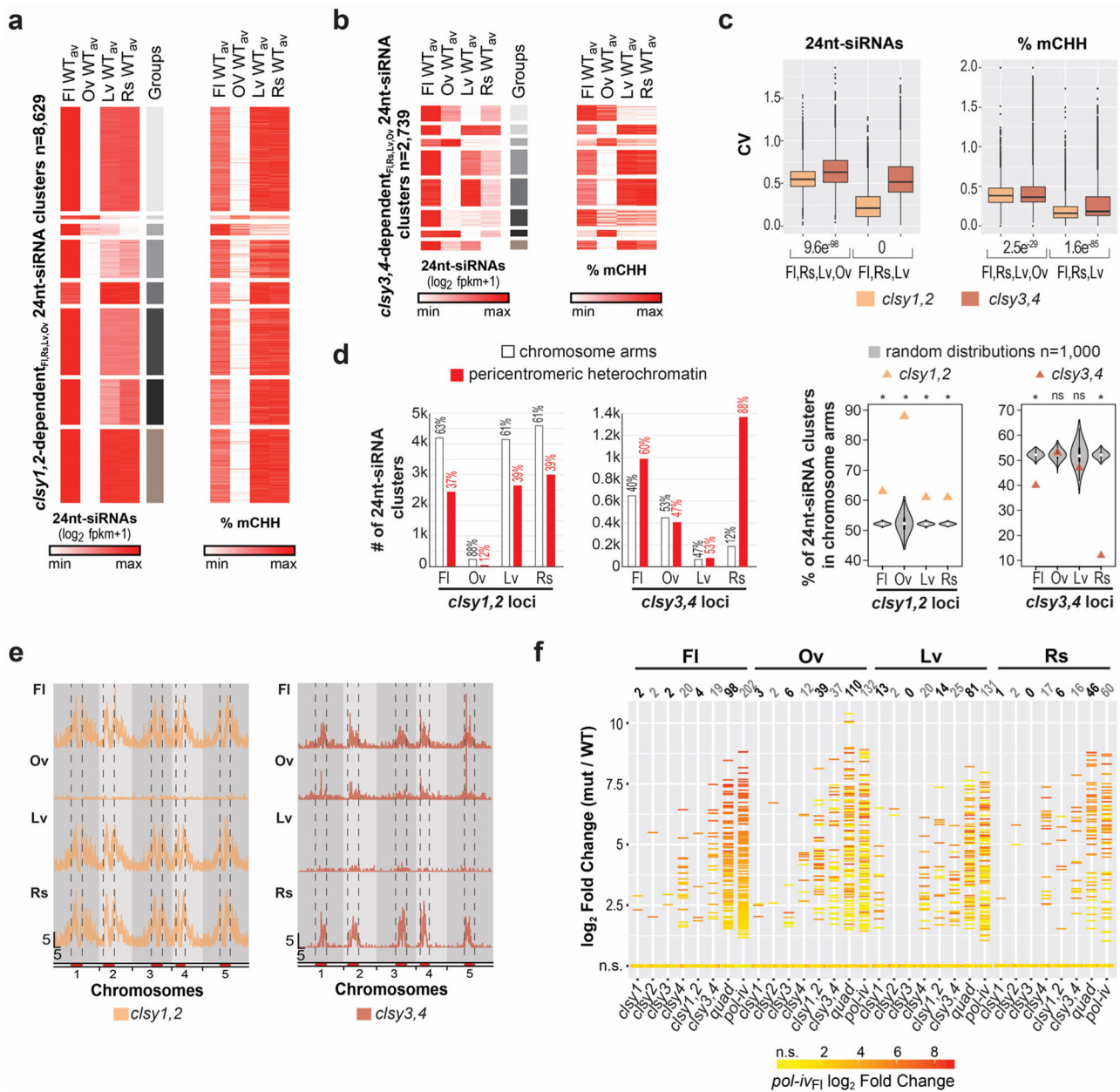
as well as *clsy3*- and *clsy3,4*-dependent 24nt-siRNA clusters from ovule tissue (74 and 98/102, respectively) (Fig. 4d and Supplementary Fig. 12d). Thus, while contributions from other flower tissues cannot be ruled out, these observations suggest that most of the CLSY3 ChIP signal from this complex tissue stems from CLSY3-bound siren loci in ovules. Finally, using previously published Pol-IV ChIP-seq data<sup>37</sup>, it was confirmed that CLSY3 and CLSY4, but not CLSY1 and CLSY2, are required for Pol-IV enrichment at these peaks (Fig. 4e). Together, these data demonstrate that CLSY3 is localized to siren loci, possibly through the direct or indirect recognition of a conserved sequence motif, and is required for the recruitment of Pol-IV to facilitate the production of very high levels of 24nt-siRNAs at siren loci specifically in ovules.

**Loci regulated by specific CLSY pairs behave distinctly across tissues.** To compare the behaviors of loci controlled by CLSY1 and CLSY2 versus CLSY3 and CLSY4, 24nt-siRNA clusters regulated by these pairs of CLSYs were identified across all tissues (*i.e.* *clsy1,2*- and *clsy3,4*-dependent<sub>Fl,Rs,Lv,Ov</sub> 24nt-siRNA clusters, respectively), and separated into 8 groups each based on the average wild-type expression levels at these sites (Fig. 5a, b). Across the four tissues, a higher degree of variation was observed between groups for the *clsy3,4*-dependent<sub>Fl,Rs,Lv,Ov</sub> clusters, which was further verified by calculating the coefficient of variation (CV) of these loci for all four tissues (Fl,Rs,Lv,Ov), as well as after excluding the ovule data (Fl,Rs,Lv; Fig. 5c). Assessment of CHH methylation levels at these same sets of 24nt-siRNA clusters revealed a similar pattern and again showed significantly more variation at sites controlled by CLSY3 and CLSY4 (Fig. 5a–c). In addition to variation in expression levels, the chromosomal distribution of *clsy3,4*-dependent<sub>Fl,Rs,Lv,Ov</sub> clusters are also more diverse (Fig. 5d, e and Supplementary Fig. 13). While *clsy1,2*-dependent<sub>Fl,Rs,Lv,Ov</sub> 24nt-siRNA clusters are enriched in the chromosome arms across all tissues, the *clsy3,4*-dependent<sub>Fl,Rs,Lv,Ov</sub> clusters shift from being enriched in pericentromeric heterochromatin in flower and rosette tissues to being more evenly distributed in ovules and leaves (permutation test  $p < 0.001$ ; Fig. 5d, e and Supplementary Fig. 13). Finally, transcriptome profiling in the *clsy* mutants across tissues revealed that the *clsy4* single and *clsy3,4* double mutants reactivate the largest numbers of *pol-iv*-dependent genes and TEs (Fig. 5f) but, like the other *clsy* single and double mutants, they suppress the expression of only a few loci (Supplementary Fig. 14a). As observed for the *pol-iv* mutant, the majority of the upregulated loci, as well as a subset of the downregulated loci, are associated with local ( $\pm 2$  kb) reductions in 24nt-siRNAs and/or non-CG DNA methylation across all tissues in the *clsy* mutants (Supplementary Fig. 14b), consistent with these expression changes being associated with defects in the RdDM pathway. Although it remains to be determined why loci controlled by CLSY3 and CLSY4 vary more in their distribution and their effects on 24nt-siRNA levels, DNA methylation patterns, and TE silencing, these data demonstrate that the relatively smaller set of sites controlled by these factors play a disproportionately large role in shaping the epigenomes across the four tissues analyzed.

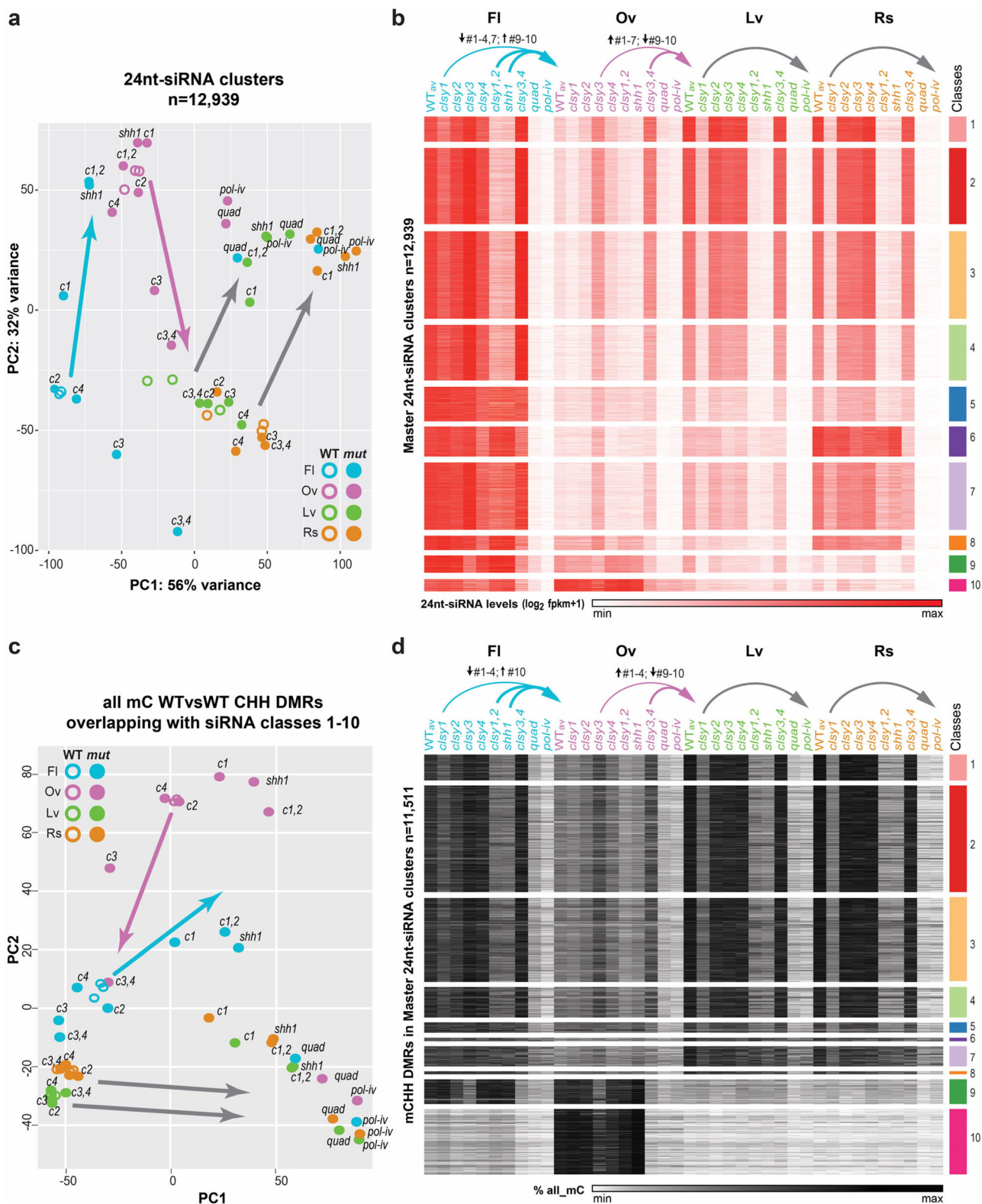
To investigate the relative contributions of the CLSYs in shaping the differences in the 24nt-siRNA and DNA methylation landscapes between tissues, the effect of each *clsy* mutant combination on the profiles at 24nt-siRNA clusters and WTvsWT CHH DMRs were determined on a genome-wide scale for each tissue. For 24nt-siRNAs, PCA analysis (Fig. 6a) and assessment of expression levels at the master set of 24nt-siRNA clusters (Fig. 6b) revealed major shifts in RdDM mutants. For example, in flower tissue, *shh1* or *clsy1,2* mutants show profiles most similar to wild-type samples from ovules, while in ovule tissue, *clsy3,4* mutants show profiles most similar to wild-type







**Fig. 5** CLSY3 and CLSY4 play a disproportionately strong role in shaping the epigenome. **a** and **b** Heatmaps showing the WT<sub>av</sub> levels of 24nt-siRNAs and DNA methylation for each tissue at the combined set of 24nt-siRNA clusters reduced in either the *clsy1,2* (**a**) or *clsy3,4* (**b**) mutants. The heatmaps were clustered (grey scale) into 8 groups based on their 24nt-siRNA levels and the CHH methylation levels at these regions are shown in the same order to the right. **c** Boxplots showing the coefficient of variation (CV) between the 24nt-siRNA expression levels (log<sub>2</sub> fpkm+1) or percent CHH methylation (% mCHH) at the regions shown in **a** and **b**. Comparisons were made including all four tissues (FI, Rs, Lv, and Ov) or just three tissues (Ov excluded). P-values based on two-sided student's t-tests are indicated for each *clsy* mutant pair below each graph. **d** Bar graphs showing the number of 24nt-siRNA clusters (1k = 1000) dependent on *clsy1,2* or *clsy3,4* that reside in the chromosome arms versus pericentromeric heterochromatin, as designated in Yelina et al.<sup>69</sup>. The percent of clusters in each category is indicated above. Violin plots showing the distribution for 1000 permutations of randomly selected 24nt-siRNA clusters from the master set of clusters (12,939). For the *clsy1,2* and *clsy3,4* loci each permutation included 8629 or 2739 clusters, respectively. For each tissue, the observed percentage of 24nt-siRNA clusters located in the chromosome arms are shown as triangles, with asterisks indicating significant deviations from the random distributions (\* for P-values < 0.001 and ns for P-values of 0.695 and 0.106 for the ovule and leaf data at *clsy3,4* loci, respectively). **e** Distributions of reduced 24nt-siRNA clusters in the *clsy1,2* or *clsy3,4* mutants along the 5 chromosomes. The pericentromeric heterochromatin is marked in red and denoted by vertical dashed lines. Scale bars for x-axis (Mb) and y-axis (clusters/100 kb bin) are indicated in the lower left corner and the tissue types are indicated in the upper left corner of each track. **f** Plot showing the expression levels [log<sub>2</sub> FC (mut/WT)] of the 326 *pol-iv*-upregulated transcripts represented as horizontal lines (log<sub>2</sub> FC ≥ 1.5 and FDR ≤ 0.05) or as dots (not significantly upregulated; n.s.) in the four tissues. The horizontal lines are colored based on the expression level in *pol-iv* from flower tissue and the number of DE loci are indicated above each column.



**Fig. 6** *clsy* mutants shift the epigenetic landscapes between tissues. **a** and **c** Principal Component Analysis (PCA) plot of 24nt-siRNA or DNA methylation levels in all contexts (all mC) at the 12,939 master 24nt-siRNA clusters compiled from all four tissues (Supplementary Data 5) or at the WTvsWT CHH DMRs (Supplementary Data 7) that overlap with these clusters, respectively. WT samples (three replicates; open circles) and various RdDM mutants (filled circles) are colored by tissue, as indicated at the bottom right or top left. Mutants are abbreviated as follows, *c1* = *clsy1* etc. **b** and **d** Heatmaps showing 24nt-siRNA and DNA methylation levels in all sequence contexts (all mC) at reduced 24nt-siRNA clusters grouped by the ten classes defined in Fig. 2c (**b**) or at the WTvsWT CHH DMRs within these clusters (**d**) for the genotypes and tissues indicated above. The color scales and the numbers (*n*) of clusters or DMRs are as indicated on the x- and y-axes, respectively. In all panels, the arrows show mutants with patterns that resemble other WT tissues (colored) or single mutants that mimic *pol-iv* (grey). In **b** and **d**, the classes that are upregulated or downregulated are indicated above the colored arrows.

24nt-siRNA classes ( $n = 11,511/12,532$ ) revealed similar shifts (Fig. 6c, d; grey and colored arrows). Here, the main difference between the behaviors of the 24nt-siRNAs and DNA methylation profiles occurred at 24nt-siRNA classes enriched for pericentromeric loci (classes 5–8) where fewer DMRs were identified (Fig. 6b, d), in line with the known roles of the CMT2/3 pathways in maintaining DNA methylation in these regions<sup>30,31</sup>. For both the 24nt-siRNA and DNA methylation analyses, the overall changes in the epigenetic profiles for each genotype and tissue are quantified by class in Supplementary Fig. 15. While the effects of mis-expressing the *CLSYs* remains to be determined, these findings demonstrate that disruption of specific *CLSY* family members is sufficient to shift the global 24nt-siRNA and, to a lesser extent DNA methylation, landscapes between tissue types (e.g. between flowers and ovules) or to largely phenocopy loss of Pol-IV (e.g. *clsy1* in leaves), revealing the extent to which these tissue-specific regulators contribute to the observed epigenetic diversity.

## Discussion

Tissue- and cell type-specific patterns of DNA methylation have been identified in increasing detail across diverse eukaryotic organisms. However, the cellular processes and machinery responsible for generating such patterns have remained largely elusive. Here, our findings demonstrate that through a combination of tissue-specific expression and locus-specific targeting, the *CLSYs* control 24nt-siRNAs and DNA methylation patterns during plant development. While the differences in the epigenomes profiled are not solely regulated by RdDM, this pathway is responsible for the vast majority of robust DNA methylation changes between tissues (Fig. 2e and Supplementary Fig. 4a). Furthermore, our data links these changes to regulation of 24nt-siRNAs by the *CLSY* proteins (Fig. 3 and Supplementary Figs. 7–10), demonstrating they are major players shaping tissue-specific methylation patterns. Amongst the tissues profiled, we observed a range of different genetic dependencies on the *CLSYs*, with flower tissue showing the most even reliance across all four *CLSYs*, while ovules are most strongly dependent on *CLSY3* and *CLSY4*, and leaf and rosette tissues are strongly dependent on *CLSY1*. These tissue-specific differences and large effects on the epigenome match the observed expression patterns of the *CLSYs* (Fig. 1) and are consistent with the observation that specific *clsy* mutants are sufficient to shift the epigenetic landscapes between tissues (Fig. 6), respectively. Taken together, these findings allow us to pair, for the first time, the distinct epigenetic profiles observed in flower, ovule, leaf and rosette tissues with the responsible cellular machinery (i.e. specific *CLSY* proteins), providing substantial mechanistic insights into the processes that enable the generation of diverse methylation patterns during plant development and addressing a largely unanswered question in the field of epigenetics.

Given the complexity of the profiled tissues, our findings likely under-represent both the cell type specificity and the contribution of the *CLSYs* in controlling DNA methylation patterns. For example, in the flower, leaf, and rosette tissues, our GUS staining experiments demonstrate that only a subset of cell types in each tissue express detectable levels of a given *CLSY* reporter (Fig. 1b). Based on these observations, we hypothesize that the *CLSYs* likely control 24nt-siRNA and DNA methylation patterns on an even finer scale than reported here. We further hypothesize that more cell type-specific profiling will also reveal larger DNA methylation decreases in the *clsy* mutants, since potentially unaffected methylation patterns in neighboring cells, which are included in our current data, would be excluded from the percent methylation ratios. Perhaps the exception to this notion of under-

representation, is the data from ovules. Here we see more uniform GUS expression (Fig. 1b) and we see the best correlation between 24nt-siRNA and mCHH patterns on a global scale (Fig. 2a and Supplementary Fig. 3a). In moving forward, it will be important to test the aforementioned hypotheses and to determine whether the high degree of correlation in ovules is specific to this tissue or whether such correlations will be observed in other instances where more uniform cell types are able to be profiled. Nonetheless, our current data already demonstrate that by combining the activities and expression levels of the *CLSYs* in different ways, a variety of 24nt-siRNA and DNA methylation patterns can be generated. Furthermore, recent studies profiling epigenetic changes during Arabidopsis embryogenesis<sup>6</sup> and seed germination<sup>8,41</sup> both show dynamic *CLSY* expression patterns and profiling of male reproductive tissues shows a major role for *CLSY3* in controlling 24nt-siRNA levels in nurse cells<sup>10</sup>. Together with our data, these findings suggest that modulation of just these four factors can be leveraged to generate a plethora of different patterns of DNA methylation throughout plant development.

Undoubtedly, one of the most unique patterns of 24nt-siRNAs and DNA methylation was observed in ovules. In this tissue, the small RNA landscape is dominated by a small number of loci that produced high levels of 24nt-siRNAs (Fig. 2a, b, and Supplementary Fig. 3a). This distinct pattern of 24nt-siRNAs was initially observed in the endosperm of rice, where they were designated as siren loci<sup>12</sup>. Subsequently, the small RNA patterns leading to this designation have been detected in other plant species and have been shown to arise prior to formation of the endosperm, in ovule tissues<sup>13,42,43</sup>. Regarding their function, our work in Arabidopsis (Figs. 2a, 4c and Supplementary Fig. 3a), and work from Grover et al. in *B. rapa*<sup>13</sup>, demonstrate that, at least in these species, siren 24nt-siRNAs target high levels of DNA methylation via the RdDM pathway. Thus, by this metric, they behave similarly to other 24nt-siRNAs. However, given their locations near genes<sup>12,13</sup> and findings that mutations in Pol-IV components result in severe fertility defects in *B. rapa*<sup>44</sup> and rice<sup>45</sup>, and subtle defects in Arabidopsis<sup>44</sup>, it has been suggested that siren loci could play roles in regulating genes important for seed development and/or for controlling imprinting in the endosperm<sup>12,13</sup>. Although examination of the 24nt-siRNA, DNA methylation, and gene expression profiles at siren loci using our Arabidopsis ovule data did not show any clear candidates in support of these hypotheses, such roles may be more evident in species with stronger fertility defects. Alternatively, these 24nt-siRNAs could play additional roles outside of targeting methylation, akin to other reproduction-specific small RNAs like piwi-interacting RNAs (piRNAs), some of which target DNA methylation at transposons to facilitate gene silencing, while others fail to map to transposons and play regulatory roles that remain poorly understood<sup>46</sup>.

Regardless of their mode of action, the connections between siren loci and seed development, a process essential for crop yields, underscore the importance of understanding how the production of 24nt-siRNAs at these loci are regulated. Here, our findings provide several new mechanistic insights: First, we demonstrate that 24nt-siRNA production at siren loci relies mainly on *CLSY3*, with some contributions by *CLSY4*, such that these 24nt-siRNAs are reduced to *pol-iv* levels in *clsy3,4* double mutants (Fig. 4a, b). Second, we demonstrate that *CLSY3* and *CLSY4* behave differently in ovule tissue compared to flower tissue, as siren loci are distributed quite evenly across the Arabidopsis chromosomes in ovules (Fig. 2a and Supplementary Fig. 3a), while *CLSY3* and *CLSY4* targets in flower tissue are enriched in pericentromeric heterochromatin (Fig. 5d, e and Supplementary Fig. 13). Finally, our *CLSY3* ChIP-seq experiments led to the identification of a robust sequence motif at siren



loci (Fig. 4d). While it remains to be tested whether this motif alone, or in combination with ovule-specific proteins or chromatin marks, is involved in recruitment of CLSY3 and Pol-IV to siren loci, or alternatively, if it plays a more downstream role, perhaps in the regulation of neighboring genes, these findings suggest both genetic and epigenetic information are important for properly regulating siren loci.

Taken together, our characterization of the CLSY proteins suggests that to understand how they control DNA methylation patterns, it will be important to not only understand how they target Pol-IV in a locus-specific manner, but also how the CLSY genes themselves are regulated during plant development. Regarding the former, our previous findings demonstrated that CLSY1 and CLSY2 target Pol-IV in connection with H3K9 methylation via SHH1, while CLSY3 and CLSY4 target Pol-IV in connection with CG methylation via unknown mechanisms<sup>37</sup>. Here, we find that in all four tissues profiled, *clsy1,2* mutants phenocopy *shh1* mutants (Fig. 3d, e), and control loci enriched in the chromosome arms (Fig. 5d, e and Supplementary Fig. 13), suggesting a conserved mechanism of Pol-IV targeting by these CLSs across diverse tissues. However, the different patterns of loci controlled by CLSY3 and CLSY4 in the four tissues (Fig. 5), combined with the presence of a motif at the siren loci that does not contain a CG dinucleotide (Fig. 4d), suggests these CLSs may be targeted to chromatin via multiple pathways. If true, that would indicate that, within the CLSY family, we have identified both tissue-specific and tissue-independent mechanisms of epigenetic regulation. In contrast with these key insights into Pol-IV targeting, our understanding of how the CLSY genes are regulated during development, or possibly even in response to the environment, remains completely unknown and is a topic that will surely provide additional insights into the regulation of DNA methylation patterns in the future.

Finally, our findings raise the possibility that the CLSs could be leveraged to manipulate DNA methylation patterns in a tissue-specific manner to prevent or correct deleterious epigenetic defects, including those affecting crop yields<sup>47,48</sup>. Given the parallels between methylation systems in plants and mammals, and the role of DNA methylation in development and disease, our findings have broad implications for both agriculture and gene therapy.

## Methods

### Plant materials, growth conditions, and tissue collection

**Plant materials.** All plant materials used in this study were in the Columbia-0 (Col) ecotype. Previously characterized mutant lines include: *clsy1-7* (SALK\_018319)<sup>49,50</sup>, *clsy2-2* (SAIL\_484\_F03)<sup>37,51</sup>, *clsy3-1* (SALK\_040366)<sup>37,50</sup>, *clsy4-1* (SALK\_003876)<sup>37,50</sup>, *npr1-4* (SALK\_083051)<sup>50,52</sup> and *shh1-1* (SALK\_074540C)<sup>35,50</sup>.

**Growth conditions and tissue collection.** Growth conditions, tissue collection, and data for flower buds (Fl; stage 12 and younger) are described in Zhou et al.<sup>37</sup> (GEO accession: GSE99694). Growth conditions and tissue collection for 15 day old rosettes (Rs) and 25 day old leaves (Lv) for all genotypes were as follows: seeds were sterilized with bleach and plated on half strength Linsmaier and Skoog medium (1/2 LS; Caisson Labs, Cat# LSP03) supplemented with 1% agar. After three days stratification in the dark at 4 °C, plants were grown in a Percival growth chamber (GCU-36L5) under 16h-light and 8h-dark cycles at 22 °C for 7 days and were then transplanted onto soil in a Salk greenhouse with long day conditions. At day 15 and day 25 of total growth, rosette tissue (whole seedling without roots) or the first and second true leaves (without petioles) were collected from ~45 plants as the rosette (Rs) and leaf (Lv) samples, respectively. These samples were immediately frozen in liquid nitrogen, and kept at -80 °C until use. Growth conditions and tissue collection for mature, unfertilized ovules (Ov) were as follows: seeds were grown on soil in a Salk greenhouse with long day conditions and un-opened flower buds (stage 10-12) were dissected using tweezers to obtain carpels. The carpels were then fixed on a glass slide with double-sided tape and opened using a straight sewing pin under a dissection microscope. The ovules were collected with the vacuum-based micro-aspiration device as detailed below. After collection, ovules for RNA extraction were quickly mixed with 100  $\mu$ L Trizol (Invitrogen, Cat# 15596018), flash frozen in liquid nitrogen, and preserved at -80 °C until use, while ovules for

DNA preparation were flash frozen in liquid nitrogen and preserved at -80 °C without Trizol. For each genotype, >300 carpels were dissected and over 10,000 ovules were collected.

Ovules were collected in modified Eppendorf tubes using a vacuum-based micro-aspiration device engineered at the Salk Institute that consists of a collection tube and a vacuum pump as previously described in Sanchez-Leon, et al.<sup>53</sup>. The ovules were collected in a 1.5 mL microcentrifuge tube with the following lid modifications: (1) the hinge of the lid was cut and a 2 mm perforation as well as a 5 mm perforation were made in the center region of the lid with needles and scissors. (2) A ~5 cm (15  $\mu$ L capacity) glass capillary tube (Drummond Scientific, Cat# 1-1000-0150) was inserted in the 2 mm perforation such that the bottom of the tube reaches the 500  $\mu$ L mark when the lid is refastened to the Eppendorf tube. (3) A Handy Plastic Tubing (external diameter 5.0 mm and inner diameter 3.5 mm (Momok, Cat# 20171004)) was inserted in the 5 mm perforation such that the bottom of the tubing reached the 1000  $\mu$ L mark and the bottom end was sealed with a nylon mesh filter using all purpose Krazy glue. (4) Perforations on the lid were sealed with glue sticks using a 100 W Hot Glue Gun (Best, Cat# B073DCLK4). (5) The modified lid was then reattached to the Eppendorf tube and placed inside a 15 mL conical falcon tube (without a cap) that is filled to the 12–13 mL mark with ground dry ice. (6) The top of the plastic tubing in the lid was attached tightly to a vacuum pump capable of ensuring a steady pressure of 0.3 to 0.6 Kg cm<sup>-2</sup> (the force necessary to detach an ovule from the placental tissue). After an ovule collecting session, the lid of the tube was carefully removed, and a clean new lid was used to cap the Eppendorf tube. To avoid contamination, lids with the capillary and plastic tubes were carefully washed to reuse or replaced with new lids for subsequent sessions.

### Cloning and visualization of GUS reporter lines

**Cloning.** All primers are listed in Supplementary Data 15. For the GUS reporters, approximately 2.0–3.0 kb of the promoter regions from *CLSY1*, *CLSY2*, *CLSY3* and *CLSY4* were amplified from wild-type genomic DNA and cloned into the pDONR-P4P1R vector using the Gateway BP Clonase II Enzyme kit (Invitrogen, Cat# 11789020). These donor plasmids were recombined with a pDONR221 vector containing the GUS gene, a pDONR-P2RP3 vector with a mock fragment, and the pH7m34GW<sup>54</sup> destination vector, using the Gateway LR Clonase II kit (Invitrogen, CA, Cat# 11791020). For the four CLSY genes, their coding regions, either ending just upstream of the stop codon or including both the stop codon and 3'UTR region, were amplified from genomic DNA and cloned into the pDONR221 or pDONR-P2RP3 vectors, respectively, using the Gateway BP Clonase II Enzyme kit (Invitrogen, Cat# 11789020). To generate C- or N-terminally tagged CLSY constructs in the pH7m34GW<sup>54</sup> destination vector, the pDONR-P4P1R vectors containing the CLSY promoters were recombined with either the pDONR221 vectors containing the CLSY coding regions without their stop codons and the pDONR-P2RP3 vectors containing a 3x-FLAG-BRLP insert, or with the pDONR221 vectors containing the 3x-FLAG-BRLP insert and the pDONR-P2RP3 vectors containing the CLSY coding regions including their 3'-UTRs, respectively, using the Gateway LR Clonase II kit (Invitrogen, CA, Cat# 11791020). For the CLSY3, an additional tagged line was cloned into the pENTR/D-TOPO vector (Invitrogen, Cat# K240020) per manufacturer's instructions by first amplifying the CLSY3 gene, including its endogenous promoter region from genomic DNA. A carboxy-terminal 3x-FLAG-BLRP tag<sup>35</sup> was then inserted into a 3' Asc I site present in the pENTR/D-TOPO vector. The pENTR/D construct was then digested with the MluI restriction enzyme (NEB, Cat# R0198S) and recombined into a modified gateway destination vector harboring Hygromycin drug resistance as described in Law et al. 2011<sup>35</sup> using the Gateway LR Clonase kit (Invitrogen, CA, Cat# 11791019). The resulting CLSY constructs were all sequence verified, transformed into the AGLO Agrobacterium tumefaciens strain, and used to create Arabidopsis transgenic lines in the Col background for the GUS reporters or their respective *clsy3* mutant backgrounds for the complementation lines using floral dip method<sup>55</sup>. Primary transformants for each construct were selected on 1/2 LS medium with 25 mg/L Hygromycin and subsequent generations were screened to select for homozygous, single insert lines that were used for the GUS staining and ChIP experiments.

**GUS histochemical staining of seedlings, leaves and flower buds.** Rosettes and leaves from wild-type plants or plants harboring GUS reporters driven by the promoters of the four CLSs (*pCLSY1-4::GUS*) were placed in ice-cold 90% acetone in water for 1 h, vacuum infiltrated for 10 min at room temperature, and then fixed for 20–30 min also at room temperature. These tissues were then briefly rinsed in GUS staining buffer (50 mM sodium phosphate buffer, pH 7.2, 0.2% Triton X-100, 2 mM Potassium ferrocyanide, and 2 mM Potassium ferricyanide) and then vacuum infiltrated 2–3 times each for 10 min. The GUS staining buffer was then supplemented with fresh 2 mM X-Gluc (GoldBio, MO) and vacuum infiltrated again, 2–3 times each for 10 min. Samples were then incubated in the dark at 37 °C for ~15 h. Flowers from wild-type plants or plants harboring GUS reporters driven by the promoters of the four CLSs were fixed in 80% acetone for 20 min on ice, briefly rinsed in GUS staining buffer (without Potassium ferrocyanide and Potassium ferricyanide) and then vacuum infiltrated for 15 min in GUS staining buffer supplemented with 2 mM X-Gluc. These samples were then incubated in the dark at 37 °C for 16–18 h. The GUS enzymatic reaction was stopped and cleared by submerging tissues in a sequential series of ethanol washes (20, 35, 50, and 70%

ethanol) for 30 min each. The cleared tissues were treated in chloral hydrate solution (1.5 mM in 30% glycerol), observed under a stereo microscope (Fisher, Cat# 03000015), and photographed with a SeBaCam14C digital camera (Fisher, Cat # SEBACAM14C CMOS 5 V 14 mp CMOS, Laxco) using SeBaView digital imaging software (Laxco).

### mRNA-seq library preparation, mapping, and data analysis

**RNA isolation, mRNA-seq library construction, and sequencing.** Total RNA from rosette and leaf tissue was isolated using the Quick-RNA MiniPrep kit (Zymo Research, Cat# R1055), while total RNA from ovules was isolated using the Direct-zol RNA Microprep Kits (Zymo Research, Cat# R2061). 1.0 µg total RNA (0.1–0.5 µg for ovules) was used to generate mRNA-seq libraries using the NEBNext Ultra II RNA Library Prep Kit (New England Biolabs, Cat# E7775). Sera-Mag Magnetic SpeedBeads (Thermo Scientific, Cat# 65152105050250) were used for all size selection and clean-up steps and the resulting libraries were pooled and sequenced (single end 50 bp, SE50) on a HiSeq 2500 machine (Illumina).

**mRNA-seq mapping.** mRNA-seq reads were mapped to the TAIR10 reference genome using STAR (v2.5.0c)<sup>56</sup> with the following options: “--outFilterMismatchNmax 2”, to allow 2 mismatches, and “--outFilterMultimapNmax 1”, to include only uniquely mapped reads (Supplementary Data 1). The sorted bam files were then used to generate Tag Directories using the (Hypergeometric Optimization of Motif EnRichment) HOMER<sup>57</sup> makeTagDirectory script with the “-mis 2 and -unique” options and UCSC format tracks were generated using the makeUCSCfile script with the “tair10 -fragLength given and -style rna-seq” options.

**Quantification of expression levels.** To identify differentially expressed loci in each tissue and compare the expression of specific genes across tissues, raw read counts or normalized values for each sample were obtained using the HOMER<sup>57</sup> analyzeRepeats.pl script with the “-noadj” or “-fpkm” options, respectively, and a custom.gtf file (Supplementary Data 16). The.gtf file was generated based on the Araport11 annotation<sup>58</sup> and includes genes, transposons/repeats, as well as several novel transcripts that were characterized in Zhou et al.<sup>37</sup>. The resulting expression files were then filtered to report just the isoform with highest expression across all patterns (condense\_genes\_by\_highest\_sum\_exp\_header.py). The expression patterns [log<sub>2</sub> (fpkm + 1)] of genes required for DNA methylation were visualized for representative wild-type samples from each tissue using MORPHEUS (<https://software.broadinstitute.org/morpheus>) (Fig. 1c). Differentially expressed loci for each tissue (log<sub>2</sub> fold change |FC| ≥ 1 or 1.5| and FDR ≤ 0.05) were determined based on the raw read counts in R using RStudio with the DESeq2<sup>59</sup> package and default parameters including batch corrections for the ovule and rosette datasets that included samples from two different experiments.

**Analysis of pol-iv-dependent up and downregulated loci across tissues.** From the pol-iv up and downregulated loci in the four tissues, a list of 326 and 213 non-overlapping pol-iv regulated loci, respectively, were generated (e.g. if a gene and a TE annotation overlap, the most highly upregulated locus was selected) (Supplementary Data 10). Of these loci, those with significant upregulation (log<sub>2</sub> FC ≥ 1 and FDR ≤ 0.05) or downregulation (log<sub>2</sub> FC ≤ -1.5 and FDR ≤ 0.05) in pol-iv mutants from each tissue (Fig. 2g and Supplementary Fig. 5b, respectively) or across all tissues and mutants (Fig. 5f, Supplementary Fig. 14a, and Supplementary Data 9) were visualized in R using RStudio with the following parameters: the matrix was organized by tidy (gather), color-coded using tidy based on the FC value in the pol-iv data from flowers, and then visualized by ggplot2 (geom\_point). Loci outside the FC and FDR thresholds for each tissue were included as zero values in the matrix.

The pol-iv dependent changes in epigenetic features surrounding these loci were plotted using deepTools (v2.4.0)<sup>60</sup> (Supplementary Fig. 5a, c) as described in Zhou et al.<sup>37</sup>. Briefly, mRNA-seq data sets were compared to wild-type controls with the bamCompare tool using the “--ratio = log<sub>2</sub> --scaleFactorsMethod SES and -bs 10” options. smRNA-seq and MethylC-seq data sets were converted from wiggle (.wig) to bigwig format using the bedGraphToBigWig script and then compared to wild-type controls with the bigwigCompare tool using the “--ratio = log<sub>2</sub> and -bs 10” options or with the bigwigCompare tool using the “--ratio=subtract” options, respectively. The resulting bigwig files were used to calculate a matrix with the computeMatrix tool using the “scale-regions -a 2000 --regionBodyLength 2000 -b 2000 and -bs = 100” options and the data was plotted using the plotHeatmap tool and sorted based on the 24nt-siRNA values.

The associations between pol-iv-dependent differentially expressed loci and epigenetic features for all mutants and tissues were determined by padding the genes and TEs 2 kb up and downstream and identifying the tissue and genotype matched reduced 24nt-siRNA clusters and non-CG DMRs (hyper and hypo) in each region using the bedtools intersect with the -loj option. For each locus, the average log<sub>2</sub> fold change in expression across all overlapping reduced 24-nt siRNA clusters was calculated [Average (log<sub>2</sub> FC mut/WT)] along with the average difference in methylation levels across all overlapping DMRs [Average (% methylation mut - WT)]. These data were visualized as bubble plots using ggplot2 in RStudio (Supplementary Fig. 14b).

### ChIP-seq library preparation, mapping, and data analysis

**Chromatin immunoprecipitation, library construction, and sequencing.** A FLAG-tagged CLSY3 line, pCLSY3::CLSY3-3xFLAG in the cly3-1 mutant background, and non-transgenic wild-type plants were used for ChIP experiments. The ChIP was performed as previously described in Zhou et al.<sup>37</sup>. Briefly, for each genotype, ~2.0 g of un-opened flower buds (stage 12 and younger) was collected, ground in liquid nitrogen, and in vitro crosslinked with 1% formaldehyde (20 min at RT; Sigma, Cat# F8775). The chromatin was then fragmented to ~500 bp by sonication, incubated with anti-FLAG M2 Magnetic beads (4 °C for 2 h; Sigma, Cat# M8823), washed 5 times and then eluted with 3xFLAG peptide ([0.1 mg mL<sup>-1</sup>]; Sigma, Cat# F4799). The crosslinking was reversed overnight at 65 °C and purified DNA (Thermo Scientific, Cat# 17908) was used to generate libraries with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, Cat# 7645). The resulting libraries were sequenced (single end 50 bp, SE50) on a HiSeq 2500 machine (Illumina).

**ChIP-seq data mapping, peak calling, and data analysis.** The ChIP-seq data was aligned to TAIR10 reference genome using bowtie (v1.1.0)<sup>61</sup> and the “-m 1 -v 2 --all --best and --strata” options to allow 2 mismatches and including only uniquely mapping reads (Supplementary Data 13). TagDirectories and UCSC genome browser tracks from ChIP-seq data were generated with the makeTagDirectory and makeUCSCfile scripts from HOMER<sup>57</sup> using the “format sam -mis 2 and -unique” or “none -fragLength given and -norm 10000000” options, respectively. The 102 CLSY3 ChIP peaks were identified with the HOMER<sup>57</sup> findPeaks script using the “-style factor -region -L 2.5 -F 2.5 and -center” options (Supplementary Data 14). Overlaps between these peaks and siren loci, cly3-, and cly3,4-dependent 24nt-siRNA clusters were determined using BEDOPS<sup>62</sup> with the -e 1 option (Fig. 4a). Motifs within CLSY3 peaks (Fig. 4d and Supplementary Fig. 12c) were identified using the findMotifsGenome.pl script from HOMER<sup>57</sup> with the “-size given -len 6, 8, 10, 12, 18, 20 -p 10 -mis 4 -bits” options and a bed file showing the positions of these motifs and the number of motifs per peak was generated using the annotatePeaks.pl script from HOMER<sup>57</sup> with the “tair10 -m -mbed -matrix -nmotifs -multi” options (Supplementary Data 14). The total number of motif 1 occurrences genome-wide were determined using the scanMotifGenomeWide.pl script from HOMER<sup>57</sup> with the “tair10 -bed -keepAll” options. Metaplots showing enrichment of reads from ChIP-seq datasets and the distribution of motif 1 over the CLSY3 peaks (Fig. 4d, e) were generated using the annotatePeaks.pl script from HOMER<sup>57</sup> with the “tair10 -m -mbed -size 4000 -hist 100 -fpkm” options and plotted in Excel. Heatmaps showing enrichment of reads from ChIP-seq datasets (Fig. 4d, e) were generated using the annotatePeaks.pl script from HOMER<sup>57</sup> with the “none -size 4000 -hist 600 -ghist -fpkm” options and plotted in Morpheus. The Pol-IV ChIP-seq data sets in different cly3 mutants were downloaded from GSE99693<sup>37</sup>, and were mapped, analyzed, and visualized as described for the CLSY3 ChIP.

### smRNA-seq library preparation, mapping, and cluster calling

**RNA isolation, smRNA-seq library construction, and sequencing.** Total RNA extraction and small RNA enrichment were performed as previously described in Zhou et al.<sup>37</sup>. Briefly, 2.0 µg of total RNA from rosette and leaf tissues, or ~0.5 µg of total RNA from ovules, were used for smRNA enrichment. The resulting small RNAs were used for library preparation with the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs, Cat# E7300) following the user’s manual. The final library products were further purified using an 8% polyacrylamide gel to excise 130–160nt products relative to the pBR322 DNA-MspI Digest ladder (New England Biolabs, Cat# E7323AA). The purified libraries were pooled and sequenced (single end 50 bp, SE50) on a HiSeq 2500 machine (Illumina).

**smRNA-seq mapping and 24nt-siRNA cluster calling.** The smRNA-seq data were mapped (Supplementary Data 2) and 24nt-siRNA clusters identified (Supplementary Data 4) as described in Zhou et al.<sup>37</sup>. Briefly, the adapters were removed from the de-multiplexed reads using cutadapt (v1.9.1)<sup>63</sup> and trimmed reads longer than 15nt were mapped to the TAIR10 genome using ShortStack (v3.8.1)<sup>64</sup> with the following options: --mismatches 1 and either --mmap f, to include multi-mapping reads, or --mmap n, to include only unique mapping reads. As described in Zhou et al.<sup>37</sup> only perfectly matched reads or reads with a single 3’ mismatch were used to call smRNA clusters using ShortStack<sup>64</sup> with the --mincov 20, --pad 100, --dicermin 21 and --dicermax 24 options, to generate Tag Directories using the makeTagDirectory script from the HOMER<sup>57</sup> with the -format sam -mis 1 and -keepAll options, and to make genome browser tracks specifically for the 24nt sized smRNAs using the makeUCSCfile script from HOMER<sup>57</sup> with the -fragLength 24 and -norm 10000000 options.

**Core, Master, and cly1,2- versus cly3,4-dependent<sub>FL,RS,LV,OV</sub> 24nt-siRNA cluster lists (Supplementary Data 5).** An inclusive list of 24nt-siRNA clusters present amongst the four tissues was generated by first identifying the core set of 24nt-siRNA clusters present in all three replicates for each tissue using the mergePeaks script from HOMER<sup>57</sup>, as was done for flower tissue in Zhou et al.<sup>37</sup>. These regions were then combined using the BEDOPS<sup>62</sup> --merge function to identify a non-redundant set of clusters based on all four tissues, hereafter referred to as the “master 24nt-siRNA clusters” list. Lists of clusters reduced in the cly1,2 or cly3,4 mutants across



all tissues were identified first using the BEDOPS<sup>62</sup> -u function with the clusters from each of the four tissues as input files, and then the sort -u function to remove duplicates.

**Differentially expressed (DE) 24nt-siRNA clusters analysis.** For differential expression analyses, 24nt-siRNA levels at each of the master 24nt-siRNA clusters were quantified for each tissue and genotype with the HOMER<sup>57</sup> annotatePeaks.pl script using the “-noadj, -size given and -len 1” options and differentially expressed 24nt-siRNA clusters compared to the three wild-type controls were identified using DESeq<sup>259</sup> ( $\log_2$  FC  $\leq 1$  and FDR  $\leq 0.01$ ) (Supplementary Data 11). To account for the extreme decreases in 24nt-siRNA counts in some mutants (*i.e.* *pol-iv*), the size factor estimates calculated by DESeq<sup>259</sup> for the wild-type samples from each tissue were used to adjust the size factor estimates for the other samples to normalize the data to mapped reads of all size classes rather than to just 24nt-siRNAs. For details, see Zhou et al.<sup>37</sup>.

For each mutant in each tissue, a list of 24nt-siRNA clusters that are either differentially expressed or overlapped with one or more *pol-iv*-dependent hypo CHH DMRs (see “DMR calling”) was generated as follows: first the BEDOPS<sup>62</sup> -n 1 function was used to find hypo CHH DMRs that do not overlap with DE clusters in the given mutant for each tissue and then the BEDOPS<sup>62</sup> -e 1 function was used to identify clusters from the master 24nt-siRNA clusters list overlapping the DMRs. These clusters were then filtered to only include those showing a  $\log_2$  FC  $< -0.415$  (~25% decrease), and finally the resulting 24nt-siRNA cluster list was combined with the DE clusters identified from the DESeq<sup>259</sup> analyses (Supplementary Data 11). 24nt-siRNA levels at the clusters included based on their overlaps with hypo CHH DMRs were determined with the HOMER<sup>57</sup> annotatePeaks.pl script using the “-noadj -size given -fpm and -len 1” options (Supplementary Fig. 6a).

### MethylC-seq library preparation, mapping, and DMR calling

**DNA isolation, methyl-seq library construction, and sequencing.** ~0.1 g of tissue from 15 day old rosettes, 25 day old leaves, or >5000 unfertilized ovules were collected for genomic DNA isolation using the DNeasy Plant Mini Kit (Qiagen, Cat# 69104). MethylC-seq libraries were generated using 2.0  $\mu$ g of DNA from rosettes and leaves, and 0.5–1.0  $\mu$ g of DNA from ovules as described in Li et al.<sup>65</sup>. The libraries were then pooled and sequenced (single end 50 bp, SE50) on a HiSeq 2500 machine (Illumina).

**MethylC-seq data processing.** MethylC-seq reads were mapped and processed using BS-Seeker<sup>266</sup> as described in Zhou et al.<sup>37</sup>. Briefly, reads were mapped with the bs\_seeker2-align.py script allowing 2 mismatches, clonal reads were removed using the MarkDuplicates function within picard tools (<http://broadinstitute.github.io/picard/>), and methylation levels at each cytosine covered by at least 4 reads were determined using the bs\_seeker2-call\_methylation.py script. For visualization of methylation levels in each sequence context or in all contexts together, wig files were generated using a custom perl script (BSseeker2\_2\_wiggleV2\_CG\_CHG\_CHH\_allmC.pl) based on the BS-Seeker2 CGmap output files. Information regarding the mapability, coverage, global percent CG, CHG, and CHH methylation levels, and non-conversion rates are presented in Supplementary Data 3.

**DMR calling.** DMRs were determined as described in Zhou et al.<sup>37</sup>. Briefly, DMRs in the CG, CHG or CHH contexts were identified in 100 bp non-overlapping bins based on either pair-wise comparisons between each mutant and three independent wild-type data sets, or between all three wild-type data sets from one tissue and all three wild-type datasets from another tissue. To be designated a DMR, the following criteria were required: (1) the 100 bp bins must contain  $\geq 4$  cytosines in the specified context, (2) the bins must have sufficient coverage in both genotypes being compared (*i.e.*  $\geq 4$  reads over the required 4 cytosines in the specific context), and (3) the bins must show a fold change of 40%, 20% or 10% methylation in the CG, CHG, and CHH contexts, respectively, with an adjusted *p* value of  $\leq 0.01$ . After these pairwise DMR calling, the final set of DMRs were determined for each mutant by taking the overlap between the DMRs called relative to each wild-type control (Supplementary Data 12). For the DMRs between tissues, a more stringent approach was taken, requiring the DMR to be present in across all 9 pair-wise comparisons to be included (*e.g.* “WT\_Rs\_CG\_hyperDMR\_vs\_FL” Supplementary Data 7). These wild-type (WT) versus WT DMRs were subsequently combined to generate a set of 100 bp hyper and hypo DMRs amongst all pairwise tissue comparisons (*e.g.* “WTvsWT all tissues hypo” DMRs; Supplementary Data 7 and Supplementary Fig. 4a).

**Methyl-cutting assay.** Genomic DNA was isolated from either flower, ovule, leaf, or rosette tissues using the CTAB method. 0.5  $\mu$ g of DNA was digested with the DNA methylation-sensitive restriction enzyme AclI (Thermo Scientific, Cat# FD0014) following the manufacturer’s instructions. PCR amplification of selected sites (site 32, site 61, site 75, and actin) was performed using either digested or undigested DNA as the templates using the primers listed in Supplementary Data 15 and resolved on a 1.5% agarose gel. In addition to the undigested DNA control, the *Actin* site, which lacks DNA methylation and the AclI recognition motif, was used as an internal loading control.

### 24nt-siRNA and DNA methylation data analyses and visualization

**Chromosome scale analysis of 24nt-siRNA and DNA methylation patterns.** The levels of 24nt-siRNAs or CHH methylation for the three wild-type controls and *pol-iv* mutants were determined for each tissue in 5 kb bins across the five Arabidopsis chromosomes with the HOMER<sup>57</sup> annotatePeaks.pl script using either the “-noadj -size given -len 1 and -ratio” or the “-noadj -size given -len 1 and -fpm” options, respectively. For 24nt-siRNAs, a pseudo count of one was added to all values prior to a  $\log_2$  transformation. After averaging the wild-type data, the 24nt-siRNA and methylation values were visualized as barplots in R using RStudio with the circularize package<sup>67</sup> (Fig. 2a and Supplementary Fig. 3a). For CHH methylation, between 1 and 3 bins, depending on the tissue, had values over 0.3%, but were capped at this value to facilitate visualization on a chromosome-wide scale.

**PCA analyses.** For Fig. 6a (PCA), raw values for 24nt-siRNA levels at the master clusters in the wild-type controls and *pol-iv* mutants for each tissue were calculated with the HOMER<sup>57</sup> annotatePeaks.pl script using the “-noadj, -size given and -len 1” options and the PCA was plotted in R using RStudio with the DESeq<sup>259</sup> and ggplot2 packages. For Fig. 6c, the WTvsWT all tissues hypo CHH DMRs (Supplementary Data 7) were filtered to include only DMRs overlapping with the 10 24nt-siRNA classes (Supplementary Data 6), resulting in 11,511/12,532 DMRs. For all cytosines (CG, CHG, and CHH) within these DMRs, the coverage and frequencies of C and T bases were calculated from the ATCGmap output files from BS-seeker2 for each genotype. This information was then formatted for use with methylkit and plotted in R using RStudio with the methylkit package<sup>68</sup> default options and a sd. threshold of 0.9.

**Cumulative sum analysis.** The cumulative sums of 24nt-siRNA levels were determined from the average values of the three wild-type replicates for each tissue at the master set of 24nt-siRNA clusters with the HOMER<sup>57</sup> annotatePeaks.pl script using the “-noadj -size given -len 1 and -fpm” options (Fig. 2b) or from a single wild-type control from a repeat ovule data set (Supplementary Fig. 11b). The top 133 clusters in ovules, which correspond to 80% of all 24nt-siRNAs, were designated as siren loci and are presented in Supplementary Data 5.

**Chromosomal distributions.** To visualize the chromosomal distributions of 24nt-siRNA clusters the TAIR10 genome was split into 100 kb bins and the number of 24nt-siRNA clusters from each class (1–10; Fig. 2d) or genotype (Fig. 5e and Supplementary Fig. 13a) in each bin was determined using the bedmap<sup>62</sup> count function requiring 1 bp overlap (--bp-ovr 1) and plotted along the five chromosomes with the pericentromeric regions as designated in Yelina et al.<sup>69</sup>. For the *clsy* mutants, the number and percent of clusters in each genomic region was also plotted in excel (Fig. 5d). For Fig. 5d, permutation tests were also performed by randomly selecting the same number of either *clsy1,2* or *clsy3,4*-dependent 24nt-siRNA clusters from the master set of 12,939 clusters 1,000 times for each tissue and then determining their chromosomal distribution as described above. The distributions of all 1000 permutation were visualized as violin plots in Fig. 5d and compared to the true *clsy1,2*-dependent and *clsy3,4*-dependent clusters distributions to determine *P*-values.

**Pie charts.** For Fig. 2e the fraction of DMRs (Supplementary Data 12) overlapping 24nt-siRNA clusters for each of the 10 classes (Supplementary Data 6), or none, were determined with BEDOPS<sup>62</sup> using the “-e 1” option. For Fig. 3b, the fraction of total 24nt-siRNAs present in specific categories were determined from raw read counts of a wild-type control from each tissue with the HOMER<sup>57</sup> annotatePeaks.pl script using the “-noadj -size given -noadj and -len 1” options.

**Venn diagrams.** Scaled Venn diagrams showing the overlaps between reduced 24nt-siRNA clusters (Supplementary Data 11) or hypo CHH DMRs (Supplementary Data 12) for the different sets of *clsy* single, double, and quadruple mutants were generated using VennMaster (version 0.38.2)<sup>70</sup> (Fig. 3c and Supplementary Figs. 7–10). Scaled Venn diagrams showing the overlaps between siren loci (Supplementary Data 5), reduced 24nt-siRNA clusters (Supplementary Data 11), and CLSY3 chIP peaks (BEDOPS<sup>62</sup> “-e 1” option; Supplementary Data 14) were also generated using VennMaster<sup>70</sup>. Unscaled Venn diagrams for 24nt-siRNA clusters (Supplementary Fig. 6b) were generated with the same data using venny (<https://bioinfogp.cnb.csic.es/tools/venny/index.html>).

**Coefficient of variation.** To assess the degree of variation in the levels of 24nt-siRNAs and CHH methylation between *clsy1,2*- and *clsy3,4*-dependent<sub>FL,Rs,Lv,Ov</sub> 24nt-siRNA clusters across tissues, the coefficient of variation (cv = standard deviation average for each cluster across the 4 tissues) was calculated and plotted as boxplots in R using RStudio and *P*-values were calculated using t-tests (Fig. 5c).

**Violin and boxplots.** For all violin and boxplots showing the expression levels of 24nt-siRNAs (Fig. 4b, Supplementary Figs. 6a, 7–10b, 11c, d, and 15a), the data was generated with the HOMER<sup>57</sup> annotatePeaks.pl script with the “-noadj -size given -fpm and -len 1” options and plotted in R using RStudio. For Supplementary Fig. 15a, the resulting data was then  $\log_2$  transformed after adding a pseudo count of one to each value. For all the violin and boxplots showing the percent



methylation levels (Fig. 4c, Supplementary Fig. 7–10b, d, 12a, and 15a), the data was generated with the HOMER<sup>57</sup> annotatePeaks.pl script using the “tair10 -size given -len 1 and -ratio” options and plotted in R using RStudio.

**Statistics and reproducibility.** For all boxplots, the boxes show the interquartile range (IQR), with the median shown as a black line, and the whiskers corresponding to 1.5 times the IQR. For all violin plots, the colored regions show the probability density distribution of the data, the black boxes show the IQR, with the median shown as a white dot, and the whiskers corresponding to 1.5 times the IQR.

**Heatmaps and data clustering.** For all heatmaps showing the expression levels of 24nt-siRNAs (Figs. 2c, 3d, 5a, b, 6b, and Supplementary Fig. 11a) the data was generated with the HOMER<sup>57</sup> annotatePeaks.pl script using the “tair10 -size given -fpkm and -len 1” options and plotted in Morpheus using the “log<sub>2</sub> + 1” adjustment with the coloring set as “relative color scheme” based on the min and max values for each row. For Fig. 2c the data was split into 10 classes (Supplementary Data 6) using the Morpheus “KMeans Clustering” tool by selecting the “cluster by rows” and “one minus Pearson correlation” options. For Fig. 3d, the data was ordered based on the overlaps in the reduced 24nt-siRNAs clusters (first for the singles, then doubles, then quadruple mutant clusters) and ranked high to low for each grouping (e.g. Fl *clsy1*-dependent clusters are ranked high to low based on the expression in the *clsy1* mutant). For Fig. 5a, b, the data was split into 8 groups using the Morpheus “KMeans Clustering” tool by selecting the “cluster by rows” and “one minus Pearson correlation” options.

For all heatmaps showing percent methylation levels (Figs. 3e, 6d, and Supplementary Fig. 4a), the data was generated with the HOMER<sup>57</sup> annotatePeaks.pl script using the “tair10 -size given -len 1 and -ratio options” and plotted in Morpheus with the coloring set as “relative color scheme” based on the min and max values for each row for Figs. 3e and 6d, or as a max percent methylation for Supplementary Fig. 4a. In Fig. 3e the heatmap is ordered exactly as in Fig. 3d and for Fig. 6d the DMRs are ordered based on the 24nt-siRNA classes they overlap with. In Supplementary Fig. 4a, the data was split into 8, 7, and 2 classes for the CHH, CHG, and CG contexts, respectively, using the Morpheus “KMeans Clustering” tool by selecting the “cluster by rows” and “one minus Pearson correlation” options (Supplementary Data 8).

**Volcano plots.** For the Volcano plots showing the complementation of 24nt-siRNAs defects in each *clsy* mutants, expression levels of 24nt-siRNAs were determined at either previously identified reduced 24nt-siRNA clusters<sup>37</sup> (Supplementary Fig. 2) or at the *clsy3*-dependent clusters identified in the current work (Supplementary Fig. 12b) using HOMER<sup>57</sup> annotatePeaks.pl script with the “noadj -size given and -len 1” options. For each complementation set, differentially expressed 24nt-siRNA clusters were identified relative to three wild-type controls using DESeq2<sup>59</sup> (log<sub>2</sub> FC ≤ -1 and FDR ≤ 0.01 for downregulation, log<sub>2</sub> FC ≥ 1 and FDR ≤ 0.01 for upregulation). As previously described, the size factor estimates calculated by DESeq2<sup>59</sup> for the wild-type samples from each tissue were used to adjust the size factor estimates for the other samples to normalize the data to mapped reads of all size classes rather than to just 24nt-siRNAs. The expression levels of the upregulated, downregulated, or unaffected 24nt-siRNA clusters were then visualized as Volcano plots in R using RStudio with the ggplot2 package.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Illumina sequencing data (smRNA-seq, MethylC-seq, mRNA-seq, and ChIP-seq) has been deposited in the NCBI Gene Expression Omnibus (GEO) and are accessible through the GEO series accession number [GSE165001](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165001). The raw data for Figs. 1c, 5d, and Supplementary Fig. 11f as well as the raw images for the gels in Supplementary Fig. 11e are provided as a Source Data file. Source data are provided with this paper.

## Code availability

All custom codes are available via the request to the corresponding author.

Received: 27 March 2021; Accepted: 6 December 2021;

Published online: 11 January 2022

## References

- Bartels, A. et al. Dynamic DNA methylation in plant growth and development. *Int. J. Mol. Sci.* **19**, 2144 (2018).
- Kawashima, T. & Berger, F. Epigenetic reprogramming in plant sexual reproduction. *Nat. Rev. Genet.* **15**, 613–624 (2014).
- Luo, C., Hajkova, P. & Ecker, J. R. Dynamic DNA methylation: In the right place at the right time. *Science* **361**, 1336–1340 (2018).

- Liu, H. et al. DNA methylation dynamics: identification and functional annotation. *Brief. Funct. Genom.* **15**, 470–484 (2016).
- Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).
- Papareddy, R. K. et al. Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in Arabidopsis. *Genome Biol.* **21**, 251 (2020).
- Kawakatsu, T. et al. Unique cell-type-specific patterns of DNA methylation in the root meristem. *Nat. Plants* **2**, 16058 (2016).
- Narsai, R. et al. Extensive transcriptomic and epigenomic remodelling occurs during Arabidopsis thaliana germination. *Genome Biol.* **18**, 172 (2017).
- Bouyer, D. et al. DNA methylation dynamics during early plant life. *Genome Biol.* **18**, 179 (2017).
- Long J. et al. Nurse cell-derived small RNAs define paternal epigenetic inheritance in Arabidopsis. *Science* **373**, abh0556 (2021).
- Zemach, A. et al. Local DNA hypomethylation activates genes in rice endosperm. *Proc. Natl. Acad. Sci.* **107**, 18729 (2010).
- 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.* **110**, 7934 (2013).
- Grover, J. W. et al. Abundant expression of maternal siRNAs is a conserved feature of seed development. *Proc. Natl. Acad. Sci. USA* **117**, 15305–15315 (2020).
- Song, Q.-X. et al. Genome-wide analysis of DNA methylation in soybean. *Mol. Plant* **6**, 1961–1974 (2013).
- Kim, K. D., El Baidouri, M. & Jackson, S. A. Accessing epigenetic variation in the plant methylome. *Brief. Funct. Genom.* **13**, 318–327 (2014).
- Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).
- Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408 (2014).
- Du, J. et al. Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Mol. Cell* **55**, 495–504 (2014).
- Zhou, M. & Law, J. A. RNA Pol IV and V in gene silencing: rebel polymerases evolving away from Pol II's rules. *Curr. Opin. Plant Biol.* **27**, 154–164 (2015).
- Johnson, L. M. et al. The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr. Biol.* **17**, 379–384 (2007).
- Woo, H. R., Dittmer, T. A. & Richards, E. J. Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. *PLoS Genet.* **4**, e1000156 (2008).
- Kim, J., Kim, J. H., Richards, E. J., Chung, K. M. & Woo, H. R. Arabidopsis VIM proteins regulate epigenetic silencing by modulating DNA methylation and histone modification in cooperation with MET1. *Mol. Plant* **7**, 1470–1485 (2014).
- Lindroth, A. M. et al. Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J.* **23**, 4286–4296 (2004).
- Rajakumara, E. et al. A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. *Genes Dev.* **25**, 137–152 (2011).
- Jackson, J. P. et al. Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. *Chromosoma* **112**, 308–315 (2004).
- Jackson, J. P., Lindroth, A. M., Cao, X. & Jacobsen, S. E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**, 556–560 (2002).
- Ebbs, M. L. & Bender, J. Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. *Plant Cell* **18**, 1166–1176 (2006).
- Law, J. A. et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**, 385–389 (2013).
- 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. USA* **110**, 8290–8295 (2013).
- Stroud, H. et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat. Struct. Mol. Biol.* **21**, 64–72 (2014).
- Zemach, A. et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205 (2013).
- Zhang, H., Lang, Z. & Zhu, J. K. Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **19**, 489–506 (2018).
- Zhang, S., Wu, X.-Q., Xie, H.-T., Zhao, S.-S. & Wu, J.-G. Multifaceted roles of RNA polymerase IV in plant growth and development. *J. Exp. Bot.* **71**, 5725–5732 (2020).

34. Wendte, J. M. & Pikaard, C. S. The RNAs of RNA-directed DNA methylation. *Biochim. Biophys. Acta* **1860**, 140–148 (2017).
35. 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).
36. Ream, T. S. et al. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol. Cell* **33**, 192–203 (2009).
37. 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).
38. Waese, J. et al. ePlant: visualizing and exploring multiple levels of data for hypothesis generation in plant biology. *Plant Cell* **29**, 1806–1821 (2017).
39. Wang, G. et al. Analysis of global methylome and gene expression during carbon reserve mobilization in stems under soil drying. *Plant Physiol.* **183**, 1809–1824 (2020).
40. Gent, J. I. et al. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res* **23**, 628–637 (2013).
41. Narsai, R., Law, S. R., Carrie, C., Xu, L. & Whelan, J. In-depth temporal transcriptome profiling reveals a crucial developmental switch with roles for RNA processing and organelle metabolism that are essential for germination in Arabidopsis. *Plant Physiol.* **157**, 1342–1362 (2011).
42. Li, X. et al. Analysis of small RNAs revealed differential expressions during pollen and embryo sac development in autotetraploid rice. *BMC Genom.* **18**, 129 (2017).
43. Li, C. et al. Genome-wide redistribution of 24-nt siRNAs in rice gametes. *Genome Res.* **30**, 173–184 (2020).
44. Grover, J. W. et al. Maternal components of RNA-directed DNA methylation are required for seed development in Brassica rapa. *Plant J.* **94**, 575–582 (2018).
45. Xu, L. et al. Regulation of rice tillering by RNA-directed DNA methylation at miniature inverted-repeat transposable elements. *Mol. Plant.* **13**, 851–863 (2020).
46. Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D. & Zamore, P. D. PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* **20**, 89–108 (2019).
47. Ong-Abdullah, M. et al. Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* **525**, 533–537 (2015).
48. Springer, N. M. & Schmitz, R. J. Exploiting induced and natural epigenetic variation for crop improvement. *Nat. Rev. Genet.* **18**, 563–575 (2017).
49. Dunoyer, P. et al. An endogenous, systemic RNAi pathway in plants. *EMBO J.* **29**, 1699–1712 (2010).
50. Alonso, J. M. et al. Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* **301**, 653–657 (2003).
51. Sessions, A. et al. A high-throughput Arabidopsis reverse genetics system. *Plant Cell* **14**, 2985–2994 (2002).
52. Herr, A. J., Jensen, M. B., Dalmay, T. & Baulcombe, D. C. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118–120 (2005).
53. Sánchez-León, N. et al. Transcriptional analysis of the Arabidopsis ovule by massively parallel signature sequencing. *J. Exp. Bot.* **63**, 3829–3842 (2012).
54. Karimi, M., Depicker, A. & Hilson, P. Recombinational cloning with plant gateway vectors. *Plant Physiol.* **145**, 1144–1154 (2007).
55. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* **16**, 735–743 (1998).
56. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
57. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
58. Cheng, C. Y. et al. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* **89**, 789–804 (2017).
59. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
60. Ramirez, F. et al. deepTools2: a next generation web server for deep-seq data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
61. 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).
62. Neph, S. et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
63. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* **17**, 10–12 (2011).
64. Johnson, N. R., Yeoh, J. M., Coruh, C. & Axtell, M. J. Improved placement of multi-mapping small RNAs. *G3 (Bethesda)* **6**, 2103–2111 (2016).
65. Li, D. et al. The MBD7 complex promotes expression of methylated transgenes without significantly altering their methylation status. *eLife* **e19893** (2017).
66. Guo, W. et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genom.* **14**, 774 (2013).
67. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. Circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
68. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
69. Yelina, N. E. et al. Epigenetic remodeling of meiotic crossover frequency in Arabidopsis thaliana DNA methyltransferase mutants. *PLoS Genet.* **8**, e1002844 (2012).
70. Kestler, H. A. et al. VennMaster: area-proportional Euler diagrams for functional GO analysis of microarrays. *BMC Bioinform.* **9**, 67 (2008).
71. Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D. & Penin, A. A. A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *Plant J.* **88**, 1058–1070 (2016).
72. Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y. & Nambara, E. Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed. *Plant J.* **41**, 697–709 (2005).
73. Schmid, M. et al. A gene expression map of Arabidopsis thaliana development. *Nat. Genet.* **37**, 501–506 (2005).

### Acknowledgements

We thank colleagues and lab members for helpful comments and discussions. This work was supported by the NIH (GM112966) and Hearst Foundation to J.L. M.Z. was supported by the 111 Project (Grant B14027), the Hundred-Talent Program of Zhejiang University, a Pioneer Fund Postdoctoral Award, and a postdoctoral fellowship from the Paul F. Glenn Center for Biology of Aging Research at the Salk Institute. This work was also supported by the NGS Core Facility and the Integrative Genomics and Bioinformatics Core Facility at the Salk Institute with funding from NIH-NCI CCSG: P30 014195, the Paul F. Glenn Center for Biology of Aging Research at the Salk Institute, the Chapman Foundation, and the Helmsley Charitable Trust.

### Author contributions

M.Z., C.C., and L.M.M. performed the experimental studies and M.Z., C.C., G.X., and J.A.L. analyzed the data including performing computational analyses. M.Z. and J.A.L. wrote the manuscript with helpful comments and suggestions from all authors. M.Z., C.B., A.L., and J.A.L. generated genetic materials and transgenic lines.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-27690-x>.

**Correspondence** and requests for materials should be addressed to Julie A. Law.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022