Title

Chromatin Regulates Bipartite-Classified Small RNA Expression to Maintain Epigenome Homeostasis in Arabidopsis

Authors

Ranjith K. Papareddy[1], Katalin Páldi[1,2], Subramanian Paulraj[1,2], Ping Kao[1], Michael D. Nodine[1,*]

Affiliations

[1]Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna Biocenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria

[2]These authors contributed equally

[*]Corresponding author

Author email addresses

R.K.P., ranjith.papareddy@gmi.oeaw.ac.at; K.P., katalin.paldi@gmi.oeaw.ac.at; S.P., subramanian.paulraj@gmi.oeaw.ac.at; P.K., ping.kao@gmi.oeaw.ac.at; M.D.N., michael.nodine@gmi.oeaw.ac.at

Abstract

Eukaryotic genomes are partitioned into euchromatic and heterochromatic domains to regulate gene expression and other fundamental cellular processes. However, chromatin is dynamic during growth and development, and must be properly re-established after its decondensation. Small interfering RNAs (siRNAs) promote heterochromatin formation in eukaryotes, but little is known about how chromatin regulates siRNA transcription. We demonstrated that thousands of transposable elements (TEs) produce exceptionally high levels of siRNAs in *Arabidopsis thaliana* embryos. Depending on whether they are located in euchromatic or heterochromatic regions of the genome, bipartite-classified TEs generate siRNAs throughout embryogenesis according to two distinct patterns. siRNAs are transcribed in embryos and required to direct the re-establishment of DNA methylation on TEs from which they are derived in the new generation. Decondensed chromatin also permits the production of 24-nt siRNAs from heterochromatic TEs during post-embryogenesis, and siRNA production from bipartite-classified TEs is controlled by their chromatin states. Decondensation of heterochromatin in response to developmental, and perhaps environmental, cues promotes the transcription and function of siRNAs in plants. Our results indicate that chromatin-mediated siRNA transcription provides a cell-autonomous homeostatic control mechanism to reconstitute pre-existing chromatin states during growth and development including those that ensure silencing of TEs in the future germ line.

Keywords: Small RNAs, DNA methylation, Chromatin, Epigenetics, Linker Histone H1, Plant embryogenesis, RNAi, Transposable elements

Introduction

Eukaryotic genomes are partitioned into euchromatic and heterochromatic domains [1,2]. Euchromatic regions are enriched for genes and provide a transcriptionally permissive state. Heterochromatic regions are densely packed, or condensed, regions of the genome that are typically transcriptionally quiescent and characterized by highly repetitive DNA and transposable elements [3,4]. Heterochromatin formation is promoted by various pathways including those affecting covalent modifications of histones, which package DNA into nucleosomes, as well as cytosine methylation [5–7]. Chromatin states are re-established after fertilization of egg and sperm in diverse animals [8–12], and histone reprogramming has also been observed in plants shortly after fertilization [13,14]. However, little is known about how heterochromatin-promoting pathways respond to labile chromatin states shortly after fertilization, and help re-establish euchromatic and heterochromatic states in animals and plants.

Small RNA-based pathways promote heterochromatin formation and associated transcriptional silencing in animals, fungi and plants [15]. Plants employ 24-nucleotide (nt) small interfering RNAs (siRNAs) to help promote silencing of repetitive elements including TEs, and thus prevent DNA mutations caused by TE mobilization. During canonical RNA-directed DNA Methylation (RdDM), RNA Polymerase IV (Pol IV) is recruited to target loci and generates transcripts that are co-transcriptionally converted into double-stranded RNAs by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) [16–21]. The resulting double-stranded RNAs are then processed into 23-nt/24-nt RNA duplexes by the DICER-LIKE 3 (DCL3) endoribonuclease [22]. The 24-nt strand of the duplex binds to ARGONAUTE 4 (AGO4), and guides AGO4 and associated proteins to siRNA-complementary sites contained within noncoding RNAs produced by RNA Polymerase V [23,24]. This effectively recruits DOMAINS REARRANGED METHYLTRANSFERASES 1/2 (DRM1/2) to target loci and results in the de novo methylation of cytosines in the CG, CHG and CHH contexts (where H ≠ G) [25,26]. Because CG and CHG methylation are maintained independently of siRNAs by DNA METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE 3 (CMT3), respectively, CHH methylation is a good indicator of RdDM

3

activities [20]. Nevertheless, CHH methylation can also be maintained independently of siRNAs by CMT2 in post-embryonic tissues, and this occurs mostly on the bodies of long TEs that are densely packed in nucleosomes and inaccessible to DRM2 [27].

Although sRNAs promote heterochromatin formation in diverse eukaryotic species, little is known about how chromatin states regulate small RNA production in animals and plants. Plant gametes and associated companion cells that support the gametes are contained within multicellular haploid gametophytes [28]. Current models propose that the large-scale chromatin decondensation observed in terminally differentiated companion cells facilitates TE transcription, and the resulting transcripts serve as substrates for RDR1/6 and DCLs to generate 20-22-nt siRNAs that move into gametes and stabilize TE silencing [29–33]. The endosperm and embryo are products of double-fertilization, and it has also been suggested that hypomethylation of endosperm promotes the production of siRNAs, which then move into the embryo to mediate TE methylation [31,34]. RdDM is dynamic during *Arabidopsis thaliana* (Arabidopsis) embryogenesis, and is required for the progressive methylation of a few target loci [35], and late-staged embryos are CHH hyper-methylated compared to other tissues [36–38], the dynamics of embryonic RdDM have not been reported genome-wide in plants due to the difficulty in generating genome-wide profiles of siRNAs and methylomes from early embryos, which are small and deeply embedded within maternal seed tissues. More generally, it is virtually unknown how chromatin states promote the de novo production of sRNAs that may promote the cell-autonomous re-establishment of silenced chromatin states of TEs including in the future germ-line.

Results

Embryos Are Enriched for Transposon-Derived Small RNAs

We recently developed a low-input small RNA sequencing (sRNA-seq) method to profile small RNA dynamics during eight stages of Arabidopsis embryogenesis, as well as floral buds and leaves [39] (Fig. 1a and Additional file 1). We previously focused on the ~21-nt microRNA class of small RNAs involved in post-transcriptional regulation [39], but noticed that the vast majority of TE-derived small RNAs were 24-nt long and highly enriched in embryos compared to floral bud or leaf tissues (Fig. 1b and Fig. S1a). Small RNAs were detected from 28,087 TEs and the highest amounts from several families including Gypsy, MuDR and En-Spm were detected during mid-embryogenesis. The levels of TE-derived 24-nt sRNAs were highly correlated among biological replicates from floral bud, embryonic and leaf tissues indicating stage- and tissue-specific sRNA populations (Fig. 1c and Fig. S1b). Moreover, principal component analysis revealed that 76.5% and 17.5% of the variation in TE-derived 24-nt siRNAs was accounted for by principal components 1 and 2, respectively. Principal component 1 distinctly separated post-embryonic and mature green embryo stages from pre-maturation embryonic stages; whereas, principal component 2 stratified the pre-maturation embryonic samples according to developmental stage. Libraries prepared from 50, 5, 1 or 0.5 ng of RNA isolated from bent cotyledon embryos clustered together with the biological replicates generated from ≥500 ng of bent cotyledon RNA (Fig. 1c and Fig. S1b). This indicated that the vast majority of variation observed in 24-nt embryonic and post-embryonic siRNA populations was biological rather than technical.

Small interfering RNAs involved in RdDM are typically 24-nt long and begin with a 5′ adenine [40]. Accordingly, adenosines were the dominant first base of 24-nt sRNAs in embryos (Fig. S1c), and the levels of embryonic 24-nt sRNAs were most highly correlated with the levels of transcripts encoding key canonical RdDM components such as RDR2, DCL3 and AGO4, which were also enriched in developing embryos (Fig. 1d,e) [20,41]. Therefore, canonical 24-nt siRNAs are highly enriched in embryos, exhibit distinct developmental dynamics, and upon maturation become similar to post-embryonic siRNA populations.

Small RNAs From Euchromatic and Heterochromatic Transposons Exhibit Distinct Developmental Dynamics

To examine the temporal dynamics of embryonic siRNAs mapping to TEs in more detail, we used *mclust* [42] to define eight unique clusters of the 31,189 TAIR10-annotated TEs based on their 24-nt siRNA levels (Fig. S2a,b). At least 2 sRNA-seq reads per million genome-matching reads were detected for 11,845 TEs and these were grouped into either class A (6,116 TEs; 19.6% of total) or class B (5,729 TEs; 18.4% of total) based on the dynamics of their corresponding siRNAs between floral buds, developing embryos and leaves (Fig. 2a,b, Additional file 2 and Fig. S2b). The remaining 19,344 TEs (62.0% of total) were considered siRNA-depleted (Fig. 2a: Fig. S2b,c), and served as negative controls. Small interfering RNAs from class A TEs had low levels in preglobular embryos that were increased at the globular stage, and were stable until sharply increasing at the mature green stage and then remained at high levels in leaves and floral buds (Fig. 2a,b). In contrast, class B TEs produced large amounts of siRNAs already in preglobular embryos, then continued to gradually increase through mid-embryogenesis and were strongly reduced in mature embryos, leaves and floral buds (Fig. 2a,b).

Class A and B TEs have distinct features. Class A TEs are short, and dispersed along pericentromeric and euchromatic regions of chromosomes (Fig. 2c-e and Fig. S2d). In contrast, class B TEs are generally longer, concentrated in heterochromatic centromeres, and especially in embryos, siRNAs are generated from throughout whole TEs (Fig. 2c-e and Fig. S2d-e). The sizes and genomic locations of class A and B TEs are characteristic of TEs respectively methylated by either the RdDM or CMT2 pathways in post-embryonic tissues [26,27,43]. Indeed, class A TEs were enriched for TEs with reduced methylation in RdDM-defective *drm1/2* mutant leaves including HAT, SINE, SADHU and other short TE families; whereas, class B TEs were enriched for TEs with reduced methylation in *cmt2* mutant leaves including MuDR, En-Spm and long terminal repeat families such as Gypsy and Copia (Fig. 2F and Fig. S2e,f). Altogether, we identified two distinct classes of TEs based on the levels of siRNAs they produce during development: euchromatic TEs (i.e. class A) that

6

progressively generate siRNAs and are methylated by the siRNA-dependent RdDM pathway in post-embryonic tissues, and heterochromatic TEs (i.e. class B) that produce very large amounts of siRNAs during embryogenesis prior to maturation, and are methylated independently of siRNAs in post-embryonic tissues.

Embryonic Methylome Dynamics

During RdDM, 24-nt siRNAs are loaded onto ARGONAUTE proteins and serve as sequence-specific guides for the recruitment of methyltransferases to target loci. siRNA-directed methylation of TEs contributes to their transcriptional silencing and immobilization, and limits their mutagenic potential [44–47]. To investigate the functions of embryonic 24-nt siRNAs, we adapted a whole-genome bisulfite sequencing approach called methylC-seq [48] to profile methylomes at single-base resolution from the low amounts of DNA available from early Arabidopsis embryos (see Methods) (Table S1). Comparisons of methylomes generated with 0.1, 0.5, 1 or ~4 ng of genomic DNA isolated from bent cotyledon embryos had nearly identical cytosine methylation levels indicating that there was low variability of this method when using different amounts of input DNA (Fig. S3a). Therefore, we used this robust low-input methylC-seq method to profile methylomes from 8-cell/16-cell (preglobular; 3 days after pollination [DAP]), early heart (4 DAP) and bent-cotyledon (8 DAP) embryos, as well as leaves and floral buds. We compared these datasets with publicly available methylomes generated from sperm [32], or late-staged embryos from early torpedo [49], mid-torpedo to early maturation [31], or mature green [37] stages. Because methylation of cytosines in the CHH context (mCHH, where H ≠ G) is a hallmark of siRNA-directed DNA methylation [20], we focused on CHH methylation. More specifically, the Arabidopsis genome was divided into 50-kb bins and the mean-weighted CHH methylation rates from reproductive, embryonic and vegetative tissues were calculated for each bin (Fig. S3b). Consistent with previous studies, CHH methylation gradually increased during embryogenesis until peaking at the maturation stage and then decreased in leaves (Fig. S3b) [35,37,38]. As expected, CHH methylation was most prominent at pericentromeric and centromeric regions densely populated with euchromatic and heterochromatic TEs that

7

were enriched for siRNAs during embryogenesis (Fig. S3b,c). Relative to sperm, euchromatic TEs had low CHH methylation levels in early embryos that increased during embryogenesis and were hypermethylated relative to leaves in 8 DAP bent-cotyledon staged embryos (Fig. 3a). In contrast, CHH methylation was barely detectable from heterochromatic TEs in sperm, but then increased in early embryos and became hypermethylated relative to leaves by 6 DAP in early torpedo staged embryos (Fig 3b). Therefore, CHH methylation is established on both euchromatic and heterochromatic TEs during early embryogenesis, and TEs become hypermethylated relative to post-embryonic tissues at late stages of embryogenesis.

To examine embryonic methylation dynamics in more detail, we identified significant differentially methylated regions (DMRs) by pairwise comparisons between six embryonic stages (preglobular, early heart, early torpedo, bent cotyledon, late torpedo-to-early mature green and mature green) (see Methods). We found 21,361 embryonic CHH DMRs with a median size of approximately 100-bp (Additional file 3 : Fig. S3d). Consistent with the genome-wide CHH methylation dynamics described above, embryonic CHH DMRs were hypomethylated in preglobular embryos, progressively methylated until embryo maturation, and then were sharply reduced in leaves and floral buds (Fig. 3c). Moreover, these embryonic DMRs were observed on various TE families especially Gypsy and Copia LTR retrotransposons, as well as MuDR and En-Spm family DNA TEs (Fig. 3d,e). The enrichment of mCHH DMRs across various TE families was generally consistent with the corresponding levels of embryonic 24-nt siRNAs (Fig. 3e and Fig. S2f). To examine the relationships between siRNAs and methylation further, we quantified 24-nt siRNA levels on mCHH DMRs across embryogenesis. Compared to randomized controls, CHH DMRs were highly enriched for 24-nt siRNAs (Fig. 3f). siRNAs overlapping euchromatic TE DMRs were lowest in early embryos, peaked at maturation and were also abundant in floral buds and leaves. In contrast, siRNAs overlapping heterochromatic TE DMRs were highly abundant at early-to-middle stages of embryogenesis and then strongly reduced in late-embryonic stages, leaves and floral buds (Fig. 3g). Therefore, large-scale changes of DNA methylation occur on both euchromatic and heterochromatic TEs during embryogenesis, and are associated with 24-nt siRNAs.

8

Notably, both euchromatic and heterochromatic TEs were hypermethylated in mature embryos relative to earlier stages and post-embryonic tissues (Fig. 3a-c and Fig. S3b). Because the proportion of embryonic tissue composed of cotyledons also increases during embryo development, we tested whether the progressively increasing levels of embryonic CHH methylation could be merely due to increased proportions of cotyledon tissues in embryos as they develop. Namely, we dissected cotyledon and non-cotyledon tissues from bent-cotyledon staged embryos, and profiled their methylomes (Fig. 3h). Both cotyledon and non-cotyledon tissues were similarly hypermethylated on DMRs indicating that hypermethylation occurs throughout late-staged embryos and is not confined to the terminally differentiated cotyledons (Fig. 3i). Interestingly, the CHH hypermethylation observed in mature embryos resembled the hypermethylation reported in root columella cells and pollen vegetative nuclei [30,50]. Similar to these specific cell-types, mature embryos have also exited the cell-cycle, which is further supported by increased and decreased levels of transcripts encoding negative and positive regulators of the cell-cycle, respectively (Fig. S3e) [51]. Therefore, CHH hypermethylation of mature embryos, as well as root columella cells and vegetative nuclei, is associated with cell cycle dormancy or exit.

Small RNA-Directed Methylation of Transposons During Embryogenesis

To test whether 24-nt siRNAs derived from euchromatic and heterochromatic TEs were necessary for progressive TE methylation during embryogenesis, we performed methylC-seq on 24-nt siRNA-deficient early heart (4 DAP) and bent-cotyledon embryos (8 DAP), as well as leaves and floral buds (Table S1). *NRPD1A* encodes the largest subunit of RNA polymerase IV (PolIV) [16], and accordingly 24-nt siRNAs overlapping euchromatic and heterochromatic TEs were nearly eliminated in *nrpd1a-3* mutants (Fig. 4a,c). Nearly all euchromatic TEs were completely hypomethylated in *nrpd1a* embryonic and post-embryonic tissues (Fig. 4b). Because the 5,729 heterochromatic TEs classified based on their embryonic siRNA dynamics also included short TEs methylated by the RdDM pathway in post-embryonic tissues (Fig. S2e-f), we partitioned

9

heterochromatic TEs into either short (≤723 bp), medium (724-2,114 bp) or long (>2,114), and examined their methylation levels in *nrpd1a* tissues (Fig. 4d, Fig. S4a-c). Consistent with previous observations from post-embryonic tissues [26,27,43], CHH methylation of short and medium TEs was significantly reduced compared to wild type in all *nrpd1a* tissues tested including embryos (Fig. S4c). In contrast, long heterochromatic TEs were globally unaffected in *nrpd1a* post-embryonic tissues, but significantly hypomethylated in *nrpd1a* mutant embryos relative to wild type (Fig. 4d). Methylation of long heterochromatic TEs is thus partially dependent on siRNAs in embryonic, but not post-embryonic tissues.

Long heterochromatic TEs are methylated by CMT2 in post-embryonic tissues and their highly condensed chromatin states were proposed to inhibit siRNA-directed DRM2-mediated methylation [26,27,43,52]. CHH methylation can be classified as CWA (W = A or T) or non-CWA. DRM2 methylates CWA and non-CWA sites, and CMT2 preferentially methylates CWA nucleotides [53–55]. Only the edges of long heterochromatic TEs were hypomethylated in non-CWA contexts of post-embryonic *nrpd1a* mutant tissues relative to wild type (Fig. 4f,h). However, both edges and bodies of long heterochromatic TEs were hypomethylated in all CHH contexts in early heart and especially bent-cotyledon *nrpd1a* embryos compared to wild type (Fig. 4e,g). Therefore, 24-nt siRNAs from euchromatic TEs increase during embryogenesis and direct progressive TE methylation; whereas, 24-nt siRNAs from throughout heterochromatic TEs accumulate to high levels in embryos and are required for their full methylation.

Chromatin Regulates Small RNA Transcription

Heterochromatin prevents access to de novo methyltransferases [27,56], and thus it may also impede PolIV access and resulting siRNA transcription. Consistent with relatively low transcript levels of heterochromatin-promoting factors in early embryos [57], reduction of DAPI-stained chromocenters and enlarged nuclei in zygotes compared to somatic tissues indicated that zygotic chromatin is decondensed (Fig. 5a,b and Fig. S5a). A marked increase in zygotic nucleoli size, as well as the co-localization of hundreds of 5S

rRNAs and heterochromatic TEs in centromeric regions (Copenhaver, 1999; Simon et al., 2018), further suggested, that heterochromatic TEs could be decondensed in early embryos (Fig. 5b,c) Therefore, post-fertilization heterochromatin decondensation, potentially associated with rRNA production, may permit PolIV accessibility to heterochromatic TEs and corresponding transcription of 24-nt siRNA precursors soon after fertilization. We then used a more experimentally tractable post-embryonic tissue to test how chromatin may generally regulate siRNA production in plants.

Because linker histone 1 (H1) inhibits RNA polymerases binding to chromatin [58,59], and its depletion results in the loss of chromocenters and chromatin compaction [60–62], we tested whether decreased H1 levels during post-embryogenesis were sufficient to increase siRNA biogenesis throughout long heterochromatic TEs similar to what we observed in embryos. That is, we performed sRNA-seq on leaves carrying null mutations in the two expressed *H1.1* and *H1.2* isoforms (i.e. *h1.1-1/h1.2-1* or *h1* mutants), and found that 24-nt siRNAs from heterochromatic TEs were significantly increased by more than 3.7-fold, and were predominantly derived from TE bodies (Fig. 5d: Additional file 1: Fig. S5a). Strikingly, hierarchical clustering of TE-derived 24-nt siRNAs demonstrated that *h1* leaf siRNA populations were more similar to siRNA populations from wild-type mature embryos instead of leaves indicating that H1 depletion was sufficient to induce an embryo-like siRNA population in a post-embryonic tissue (Fig. S5c).

To investigate the effects of H1 depletion on TE-derived siRNAs in more detail, we ranked euchromatic and heterochromatic TEs based on their 24-nt siRNA levels in *h1* mutant leaves, and found that increased 24-nt siRNA levels from long heterochromatic TEs in *h1* mutants were positively correlated with TE length, CHH methylation and H1 occupancy (Fig. 5e-g and Fig. S5d). Because TEs with increased 24-nt siRNAs in *h1* mutants were also enriched for HISTONE 3 LYSINE 9 di-methylation (H3K9me2) (Fig. 5h), we tested whether loss of H3K9me2 can also affect heterochromatic siRNA production. We examined the levels of siRNAs from long heterochromatic TEs in leaves triple mutant for *SU(VAR)3-9 HOMOLOG 4/5/6* histone methyltransferases (*suvh4/5/6*), which are deficient in H3K9me2 levels (Table S1) [43]. Long heterochromatic TEs produced only

1.2-fold more 24-nt siRNAs in *suvh4/5/6* leaves compared to wild-type, and the global 24-nt siRNA populations were similar to wild type (Fig. S5b,c). These results suggest that reducing heterochromatin, rather than H3K9me2 marks associated with heterochromatin, is sufficient for siRNA production from long heterochromatic TEs. Accordingly, H1 promotes nucleosome occupancy on heterochromatic regions [62,63], and we found that heterochromatic TEs were enriched for H1 (Fig. 5g and Fig. S5d) and had reduced nucleosome occupancy in *h1* mutants (Fig. S5f). Our results indicate that depletion of H1 is sufficient to decrease nucleosome occupancy of heterochromatic TEs, as well as increase corresponding siRNA and CHH methylation levels.

Because 24-nt siRNAs were enriched on the bodies of long heterochromatic TEs in *h1* mutants with reduced chromatin compaction (Fig. 5d and Fig. S5b), we next examined the relationships between nucleosome occupancy and 24-nt siRNA levels of long heterochromatic TEs in wild-type leaves. We employed an iterative *k*-means clustering approach to generate three groups of long heterochromatic TEs based on their nucleosome occupancy using publicly available micrococcal nuclease sequencing data [62] (Fig. 5i). Group 1 comprised 1,098 TEs (56.7% of total) that had high densities of nucleosomes and were devoid of 24-nt siRNAs throughout their lengths (Fig. 5i,j). Group 2 contained 641 TEs (33.1% of total), and had low and high nucleosome occupancy over the edges and bodies, respectively, and were enriched for 24-nt siRNAs only on the edges (Fig. 5i,j). Group 3 consisted of only 196 TEs (10.2% of total), and had very low nucleosome levels, but abundant 24-nt siRNAs, on both their edges and bodies similar to euchromatic TEs (Fig. 5i,j). Altogether these results suggest that increased nucleosome occupancy restricts RNA PolIV activity, and thus chromatin states alone appear to explain 24-nt siRNA production from TEs.

RNA PolIV transcribes ~25-to-40-nt RNAs that are co-transcriptionally converted to double-stranded RNAs by RNA-dependent RNA Polymerases and rapidly processed into 23-nt/24-nt duplexes by DICER-LIKE (DCL) endoribonucleases [17,18,22]. These transient PolIV-dependent 24-nt siRNA precursors can be robustly detected in *dcl2/3/4* mutants [17,18], and thus 24-nt siRNA precursor levels indicate PolIV transcriptional activities. We performed sRNA-seq on *dcl2/3/4* flowers, globular embryos and leaves, and compared levels of

12

24-nt siRNA precursors from euchromatic and heterochromatic TEs (Table S1). Compared to leaves and flowers, we respectively detected 11-fold and 3.8-fold significantly more 24-nt precursors from heterochromatic TEs in *dcl2/3/4* early embryos (Fig. 5k). Importantly, the 24-nt siRNA precursors mostly originated from the bodies of long heterochromatic TEs in embryos, but were strongly reduced in floral buds and leaves (Fig. 5i). Together with the observations that 24-nt siRNAs were also enriched on the bodies of long heterochromatic TEs in early embryos and *h1* mutant leaves (Fig 2c,d), our results are consistent with a model whereby decompaction of heterochromatin in early embryos and *h1* mutant leaves permits PolIV access and transcriptional activities to produce 24-nt siRNAs.

Homeostasis of Transposon-Derived siRNAs

In contrast to heterochromatic TEs, we found 2.7-fold significantly less 24-nt siRNAs from euchromatic TEs in *h1* leaves compared to wild type, which was also associated with their CHH hypomethylation (Fig. 5d,f and Fig. S5b,d). Unlike heterochromatic TEs, euchromatic TEs were lowly enriched for H1 in wild type leaves and nucleosome occupancy was further reduced in *h1* mutants (Fig. 5g and Fig. S5d,f). Therefore, siRNA depletion from euchromatic TEs is likely an indirect consequence of sequestering PolIV to accessible heterochromatic TEs. Moreover, we observed the greatest enrichment of siRNAs derived from heterochromatic compared to euchromatic TEs at the preglobular stage of embryogenesis, which is the earliest post-fertilization sRNA-seq dataset available (Fig. 6a). This was reduced during mid-embryogenesis and then further decreased to almost post-embryonic levels during maturation (Fig. 6a). Together with our siRNA precursor analysis (Fig. 5k,l), this indicates that PolIV is more efficiently recruited to heterochromatic TEs compared to euchromatic TEs during the initial stages of embryogenesis. The enrichment of heterochromatic TE-derived 24-nt siRNAs in preglobular embryos surpassed what we observed in *h1* mutant leaves (Fig. 6a), suggesting that reduced nucleosome occupancy alone does not fully account for the extreme enrichment of heterochromatic siRNAs in preglobular embryos.

13

Based on available sRNA-seq datasets, euchromatic, but not heterochromatic, TE-derived 24-nt siRNAs were substantially reduced in mutants deficient in CG and CHH methylation (Fig. 6b and Fig. S6a) (Law et al., 2013; Lister et al., 2008; Stroud et al., 2014). Therefore, low CHH methylation in preglobular embryos (Fig. 3a,b), may reduce methylation-dependent feed-back loops that facilitate production of siRNAs from euchromatic TEs in preglobular stages. CLASSY (CLSY) chromatin remodeling factors also promote 24-nt siRNA production: CLSY 1/2 and CLSY 3/4 help recruit PolIV to euchromatic and heterochromatic regions, respectively (Fig. 6b) [19,21,43,64]. Dynamic chromatin states and corresponding establishment of methylation/CLSY-dependent transcription of siRNA precursors likely contribute to the unique siRNA populations observed in embryos.

Remarkably, the relative amounts of euchromatic and heterochromatic TE-derived siRNAs remained stable throughout post-embryonic development (Fig. 6a). Root columella and pollen vegetative cells are depleted for H1 [50,65], but were not depleted for euchromatic siRNAs (Fig. 6a). This may be due to these terminally differentiated cell-types being derived from a single mitotic division and thus they may retain the ability to recruit PolIV to euchromatic TEs. For example, these cell-types are CHH hypermethylated [30,50] and thus CHH methylation-dependent positive feedback loops on euchromatic TEs may counteract the loss of H1. Importantly, TE-derived siRNA populations in embryos were distinct from endosperm and seed-coat tissues (Fig. 6d). Consistent with chromatin states primarily regulating siRNA production, the endosperm has reduced cell division compared to embryos by 6 DAP [66]. Accordingly, endosperm siRNAs overlapped mostly edges, but not bodies, of heterochromatic TEs typical of other non-embryonic populations (Fig. 6d). Similar to pollen, but in contrast to other non-embryonic tissues, endosperm had similar levels of siRNAs from heterochromatic and euchromatic TEs (Fig. 6a). Because nearly equal siRNA levels from euchromatic and heterochromatic siRNAs were also observed in *met1* mutants (Fig. 6b and Fig. S6a), this balance may be due to loss of CG methylation-dependent euchromatic siRNA production in CG hypomethylated pollen vegetative nuclei and

14

endosperm [32]. Altogether our data indicate that the homeostasis of 24-nt siRNA production from euchromatic

and heterochromatic TEs are affected by dynamic chromatin states including, but not

restricted to, those associated with early embryogenesis.

15

Discussion

Although siRNAs direct faithful re-establishment of methylation genome-wide across generations [67], the dynamics of embryonic siRNAs and how they contribute to the nascent epigenome have not been reported. In this study, we demonstrated that thousands of TEs produce exceptionally high levels of 24-nt siRNAs in embryos and can be classified into two distinct groups based on their developmental dynamics (Fig 1,2). Euchromatic TEs produce gradually increasing amounts of siRNAs during embryogenesis eventually reaching post-embryonic levels, and are constitutively required to direct TE methylation throughout development (Fig 2-4). In contrast, heterochromatic TEs produce a burst of siRNAs soon after fertilization, and specifically during embryogenesis, to help establish TE methylation de novo, which is then maintained independent of siRNAs during post-embryogenesis (Fig 2-4) [26,27,43]. Interestingly, the levels of siRNAs from these euchromatic and heterochromatic bipartite-classified TEs are regulated according to their chromatin states (Fig 5,6). Decondensed chromatin permits transcription of 24-nt siRNAs, and this contributes to cell autonomous homeostatic control mechanisms that normalize chromatin states.

We propose a three-phase model for how chromatin states, and resulting siRNA dynamics, help shape the nascent epigenome (Fig 7). After fertilization, zygotic chromatin is decondensed and this appears to be associated with transcriptional activation of rRNAs, including hundreds of 5S rRNA loci that co-localize with heterochromatic TEs in near centromeric regions (Fig. 5a,c). Arabidopsis zygotes require de novo synthesis of gene products directly after fertilization [57,68], and genes involved in rRNA biogenesis produce high levels of transcripts in preglobular embryos relative to later stages [41]. In contrast to Arabidopsis, maternally donated proteins drive early embryogenesis in Xenopus and H1 dynamics mediate transcriptional activation of rRNA loci in oocytes and their silencing in somatic tissues [69,70]. In plants, decondensation of rRNA loci can also permit their transcription and concomitant pollen tube cell growth [71], and are likely required for the rapid cell divisions in early endosperm that also have enlarged nucleoli (Fig. 5a). Therefore, reduced heterochromatin in a variety of cell types, including those producing large amounts of protein such as early embryos, endosperm

16

and pollen vegetative cells, may be permissive for PolIV-mediated transcription of siRNA precursors from TEs that are typically in a deep heterochromatic state during other developmental phases (Fig 7; Phase 1). Consistent with decondensation of heterochromatin facilitating de novo production of sRNAs, PolIV-dependent 24-nt siRNAs were sharply increased throughout heterochromatic TEs in *h1* mutant leaves with reduced heterochromatin (Fig. 5d, Fig. S5d). Production of siRNAs from euchromatic TEs is delayed relative to those from heterochromatic TEs (Fig. 6a). Because euchromatic, but not heterochromatic, TEs require CHH methylation and CLSY1/2 chromatin remodelers to produce full siRNA levels (Fig. 6b) [19,21,43,64], both of which are gradually increased during embryogenesis (Fig. 3a-c and 6c), the developmental time-lag in euchromatic compared to heterochromatic siRNA production may also be partially due to the delay establishing methylation/CLSY-dependent positive feedback loops during early embryogenesis.

Embryos divide rapidly through the bent cotyledon stage [72] and dynamic chromatin condensation and decondensation associated with such increased cell division [73,74] likely allows access of PolIV to both heterochromatic and euchromatic TEs (Fig 7; Phase 2). Production of embryonic siRNAs from heterochromatic relative to euchromatic TEs is steady between the globular and bent cotyledon stages. Heterochromatic TE-derived siRNAs are rapidly reduced upon maturation when heterochromatin becomes highly condensed [75] and euchromatic domains containing genes encoding seed storage and oil body biogenesis proteins are transcriptionally activated [41] (Fig. 6a). As a consequence, PolIV access to heterochromatic TEs is likely greatly reduced and this results in more PolIV being readily available to produce siRNAs from euchromatic TEs (Fig 7; Phase 3). Consistently, we observed a burst of siRNAs from euchromatic TEs, and their associated hyper methylation, at the mature stage (Fig. 2a,b). Heterochromatic TEs also become hypermethylated at the mature stage, which appears to be largely independent of siRNAs, but rather dependent on CMT2 as is typical of subsequent post-embryonic development (Fig 4). Therefore, CHH hypermethylation throughout mature embryos may largely be a consequence of the rapid shift of chromatin states upon maturation.

Similar to *h1* mutants [62,63], heat-stress also causes reduced nucleosome occupancy and decondensed heterochromatin [76,77]. Moreover, we classified the heat-activated ONSEN/AT5TE15240 as a class B/heterochromatic TE based on its siRNA dynamics, and ONSEN transcription and transposition are greatly enhanced in RdDM-defective mutants [78]. Interestingly, 24-nt siRNAs were increased throughout the body of ONSEN TEs after heat-stress and this was further enhanced in rapidly dividing undifferentiated calli [79], which has increased chromatin accessibility in rice [79,80]. Based on these results, as well as our previous observation that heat-stress related genes are significantly enriched in preglobular embryos [41], we suggest that the chromatin dynamics caused by heat-stress and subsequent recovery are analogous to what we observed in early embryos. That is, heat-stress and fertilization may decrease nucleosome occupancy across heterochromatic TEs, which enables PolIV-mediated siRNA production and subsequent reconstitution of proper heterochromatin to help limit TE mobilization in the genome. Strong upregulation of TE-specific endo-siRNAs and piwi-interacting RNAs observed in H1-depleted Drosophila [81] suggests that H1-dependent regulation of chromatin states may also facilitate heterochromatic small RNA transcription in animals.

Reprogramming of heterochromatin during early embryogenesis occurs in diverse metazoa including flies, mammals, worms and zebrafish [8–12], and we observed that CHH methylation is also reprogrammed during early plant embryogenesis (Fig 3). For example, CHH methylation was essentially lost on the bodies of heterochromatic TEs in sperm and subsequently fully re-established by both siRNA-dependent and siRNA-independent pathways (Fig 4). Reduced heterochromatin in early animal embryos has been associated with increased developmental potential [82,83], and similar relationships have also been observed in plants including reprogramming associated with plant regeneration and heat-stress induced somatic embryogenesis [84–90]. However, decreased heterochromatin would also increase the risk of TE mobilization and resulting mutations, and this could be especially dangerous in plant zygotes because they are the precursors of all cell-types including the gametes. Our results indicate that embryos produce 24-nt siRNAs according to their chromatin states including those that are permissible for PolIV transcription soon after fertilization. These de

18

novo produced 24-nt siRNAs direct re-methylation of both euchromatic and heterochromatic TEs in the new generation. Therefore, decondensed chromatin permits transcription of early embryonic siRNAs to help promote cell-autonomous TE silencing. More generally, reduced heterochromatin due to sharp increases in rRNA production requirements during growth (e.g. early embryos, endosperm and pollen vegetative cells), and perhaps in response to external cues such as heat stress, enables the synthesis and functions of sRNAs that can help reconstitute proper chromatin states.

Methods

Plant material and growth conditions

All genotypes were in the Columbia-0 (Col-0) *Arabidopsis thaliana* background including *dcl2/3/4* mutants composed of *dcl2-1*, *dcl3-1* and *dcl4-2t* [95], *h1.1-1/h1.2-2 [27]*, *nrpd1a-3* [96], and *suv4/5/6* [97]. Plants were grown in a climate-controlled growth chamber at 20˚C to 22˚C under a 16-h light/8-h dark cycle with incandescent lights at 130 to 150 µmol/m^2/s.

Embryo isolation and nucleic acid extraction

Embryos were dissected from siliques either 3 days after pollination (DAP) (preglobular), 4 DAP (early heart/transition) or 8 DAP (bent cotyledon). Siliques were opened with forceps and seeds were collected in 2 ml Eppendorf tubes containing nuclease-free water, and kept on ice. Seeds were then crushed with pestles and embryos were selected under an inverted microscope using a microcapillary tube. Isolated embryos, as well cotyledon and non-cotyledon portions of bent-cotyledon embryos, were thoroughly and serially washed 4× with nuclease-free water, and stored at -80°C. RNA was isolated as previously described [39,98] and genomic DNA was isolated as follows. Genomic DNA was extracted from ≥50 embryos per stage, floral buds and leaves using *Quick*-DNA^{TM} Micro prep Kit (Zymo D3020) according to manufacturer recommendations.

Small RNA Profiling and Analysis

sRNA-seq libraries were generated as previously described [39], and together with publicly available datasets were analyzed in the same manner. For the model-based clustering of transposon-derived siRNAs, 24-nt siRNAs mapping to TEs annotated in the The Arabidopsis Information Resource 10 (TAIR10) genome [99] were partitioned into clusters using *Mclust* [42]. The mean reads per million genome-matching reads of 24-nt siRNAs mapping to TEs were calculated from biological triplicates of floral bud, embryonic and leaf samples

and used as input for *Mclust* to identify the optimal Gaussian mixture probability model. With a sequential increment of two components, we identified the VEV (Variable volume, Equal shape, Variable orientation) ellipsoidal distribution model to be optimal with the minimum number of components (i.e. eight) containing maximum Bayesian Information Criterion (BIC). Principal component analysis was performed with the R *prcomp* function using default parameters; heatmaps and metaplots of TE-derived siRNAs were generated with *deepTools [100].* Briefly, a matrix containing normalized 24-nt siRNA scores per genome regions for tissue types or genotypes were generated (*computeMatrix scale-regions -bs 5 -m 4000 -b 2000 -a 2000 --averageTypeBins mean*). The obtained matrix was used to generate heatmaps (*deepTools plotHeatmap*) or metaplots (*deepTools plotProfile*). For Fig. 5, regions without siRNA signals were removed and the remaining genomic regions were used to calculate matrices containing nucleosome signal and 24-nt siRNA levels. Hierarchical clustering of transposon-derived 24-nt siRNAs was performed by calculating Euclidean distances between samples and the distance matrix was subjected to the R function *hclust(*,"complete")*.

DNA Methylation Profiling

MethylC-seq libraries were generated as described [48] with the following modifications. Briefly, genomic DNA was subjected to bisulfite treatment for 200 minutes. Bisulfite-treated DNA was then preamplified for two cycles with primers (5′-CCCTACACGACGCTCTTCCGATCTNNNNNNN-3') containing random hexamers and purified using the Zymo DNA Clean and Concentrator kit. Adaptor primers (5′-CAGACGTGTGCTCTTCCGATCTNNNNNNN-3') were added to preamplified products and then amplified for 12 PCR cycles with indexing primers for Illumina sequencing. Methylome libraries were purified using Beckman Coulter AMPureXP DNA beads. Libraries quality checked for fragment length between 200-600bp were used for sequenced in single-read mode on an Illumina HiSeq2500 or Nextseq instrument.

21

DNA Methylation Analysis

Sequenced reads were trimmed using *Trim Galore* with default settings. In addition, the first six bases of each read were removed to exclude the random hexamer portion of the read used during the preamplification step. After quality filtering and adaptor trimming, bisulfite-converted reads were aligned against the TAIR10 genome [99] using *Bismark* (*bismark --non_directional -q --score-min L,0,-0.4*) [101]. BAM files containing clonal deduplicated and uniquely mapped reads were then used as input for the *Methylpy* software (https://bitbucket.org/schultzmattd/methylpy) to extract weighted methylation rates at each cytosine as previously described [102]. Bisulfite conversion rates were calculated using the unmethylated chloroplast genome or spiked-in unmethylated Lambda phage DNA controls **(**European Nucleotide Archive Accession Number J02459, Promega catalog number D1521).

Differentially methylated regions (DMRs) were defined using *Methylpy* as described [36]. Briefly, biological replicates were pooled and differentially methylated sites (DMSs) were identified by the root mean square tests with false discovery rates ≤ 0.01. Cytosine sites with ≥4 overlapping reads were retained for all samples except for preglobular in which DMSs with ≥3 overlapping reads were retained. Differentially methylated sites within 100-bp were collapsed into DMRs. CHH-DMRs were further filtered by discarding regions with < 4 DMSs and methylation difference < 20%. Using these parameters, DMRs were identified in all 10 pairwise combinations across embryonic samples (preglobular, early heart, early torpedo,bent cotyledon,mature green) and merged using the BEDtools *merge* function [103]. DMRs were used to calculate the weighted CHH methylation rate on all analyzed tissue types. CHH methylation metaplots for class A, B and siRNA-deficient TEs were plotted using the R library *Seqplots* [104]: Body, upstream and downstream regions of TEs were split into equal-sized bins, and the average weighted mCHH level for each bin was calculated and plotted.

Expansion Microscopy and DAPI quantification

The expansion microscopy technique [105] optimized for Arabidopsis seeds was conducted as previously described [68]. Anti-Fibrillarin antibody (ab4566, Abcam) and anti-alpha Tubulin antibody (ab89984, Abcam) were used in 1:500 dilution as primary antibodies. Goat Anti-Mouse IgG H&L (Alexa Fluor® 488) (ab150113, Abcam) and goat Anti-Chicken IgY H&L (Alexa Fluor® 555) (ab150170, Abcam) were used in 1:500 dilution as secondary antibodies. For each sample, a stack of nine images with 1 μm intervals were recorded by ZEISS LSM700 with 25× oil objective and ZEN software at 1024 × 1024 resolution in 8-bit. DAPI signals were excited by 405 nm laser and passed through SP490 filters. Alexa488 signals were excited by 488 nm laser and passed through BP490-635 filters. Alexa555 signals were excited by 555 nm laser and passed through 560-1000 nm filters. Pinhole sizes were kept as 1 airy unit for each color, and color channels were scanned separately. FIJI software was used for image processing and nuclear size quantification. Each stack of images was first Z-projected on maximum intensity and then the nuclear areas were determined based on DAPI signals. The zygotic nuclei were distinguished from the endosperm nuclei according to position and tubulin patterns.

Availability of data and materials

All sequencing data generated in this study are available at the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO, https://www.ncbi.nlm.nih.gov/geo/) under accession number GSEXXXXX. Publicly available next-generation sequencing data were downloaded from NCBI, GEO and are listed along with general mapping statistics in Additional File 1 (Table S1). The software code used for the sRNA, methylome and transcriptome analysis are available at github.com/Papareddy.

Acknowledgements

Funding

Author Contributions

R.K.P. and M.D.N. conceived the project; R.K.P. developed the methodology, implemented software used and performed formal analysis; R.K.P., S.P., K.P., P.K. and M.D.N. conducted the experiments; R.K.P. and M.D.N. wrote and edited the article; M.D.N. supervised the project and acquired funding.

Competing interests

The authors declare that they have no conflicts of interests.

24

References

1. Heitz E. Das heterochromatin der moose. Bornträger; 1928;

2. Grewal SIS, Jia S. Heterochromatin revisited. Nat Rev Genet. 2007;8:35–46.

3. Elgin SCR, Grewal SIS. Heterochromatin: silence is golden. Curr Biol. 2003;13:R895–8.

4. Huisinga KL, Brower-Toland B, Elgin SCR. The contradictory definitions of heterochromatin: transcription and silencing. Chromosoma. 2006;115:110–22.

5. Chan SW-L, Henderson IR, Jacobsen SE. Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet. 2005;6:351–60.

6. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011;21:381–95.

7. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet. 2019;20:207–20.

8. Laue K, Rajshekar S, Courtney AJ, Lewis ZA, Goll MG. The maternal to zygotic transition regulates genome-wide heterochromatin establishment in the zebrafish embryo [Internet]. Nature Communications. 2019. Available from: http://dx.doi.org/10.1038/s41467-019-09582-3

9. Wang C, Liu X, Gao Y, Yang L, Li C, Liu W, et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. Nat Cell Biol. 2018;20:620–31.

10. Rudolph T, Yonezawa M, Lein S, Heidrich K, Kubicek S, Schäfer C, et al. Heterochromatin formation in Drosophila is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. Mol Cell. 2007;26:103–15.

11. Mutlu B, Chen H-M, Moresco JJ, Orelo BD, Yang B, Gaspar JM, et al. Regulated nuclear accumulation of a histone methyltransferase times the onset of heterochromatin formation inC. elegansembryos [Internet]. Science Advances. 2018. p. eaat6224. Available from: http://dx.doi.org/10.1126/sciadv.aat6224

12. Ahmed K, Dehghani H, Rugg-Gunn P, Fussner E, Rossant J, Bazett-Jones DP. Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. PLoS One. 2010;5:e10531.

13. Ingouff M, Hamamura Y, Gourgues M, Higashiyama T, Berger F. Distinct dynamics of HISTONE3 variants between the two fertilization products in plants. Curr Biol. 2007;17:1032–7.

14. Ingouff M, Rademacher S, Holec S, Soljić L, Xin N, Readshaw A, et al. Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in Arabidopsis. Curr Biol. 2010;20:2137–43.

15. Martienssen R, Moazed D. RNAi and heterochromatin assembly. Cold Spring Harb Perspect Biol. 2015;7:a019323.

16. Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. RNA polymerase IV directs silencing of endogenous DNA. Science. 2005;308:118–20.

17. Zhai J, Bischof S, Wang H, Feng S, Lee T-F, Teng C, et al. A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis. Cell. 2015;163:445–55.

18. Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, et al. Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. Elife. 2015;4:e09591.

19. Zhou M, Palanca AMS, Law JA. Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. Nat Genet. 2018;50:865–73.

20. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11:204–20.

21. Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature. 2013;498:385–9.

22. Singh J, Mishra V, Wang F, Huang H-Y, Pikaard CS. Reaction Mechanisms of Pol IV, RDR2, and DCL3 Drive RNA Channeling in the siRNA-Directed DNA Methylation Pathway. Mol Cell. 2019;75:576–89.e5.

23. Zilberman D, Cao X, Jacobsen SE. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science. 2003;299:716–9.

24. Wierzbicki AT, Haag JR, Pikaard CS. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell. 2008;135:635–48.

25. Cao X, Jacobsen SE. Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A. 2002;99 Suppl 4:16491–8.

26. Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell. 2013;152:352–64.

27. Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell. 2013;153:193–205.

28. Drews GN, Koltunow AMG. The female gametophyte. Arabidopsis Book. 2011;9:e0155.

29. Slotkin RK, Vaughn M, Borges F, Tanurdzić M, Becker JD, Feijó JA, et al. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell. 2009;136:461–72.

30. Calarco JP, Borges F, Donoghue MTA, Van Ex F, Jullien PE, Lopes T, et al. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell. 2012;151:194–205.

31. Hsieh T-F, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, et al. Genome-wide demethylation of Arabidopsis endosperm. Science. 2009;324:1451–4.

32. Ibarra CA, Feng X, Schoft VK, Hsieh T-F, Uzawa R, Rodrigues JA, et al. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science. 2012;337:1360–4.

33. Feng X, Zilberman D, Dickinson H. A conversation across generations: soma-germ cell crosstalk in plants.

Dev Cell. 2013;24:215–25.

34. Mosher RA, Melnyk CW. siRNAs and DNA methylation: seedy epigenetics. Trends Plant Sci. 2010;15:204–10.

35. Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F. DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. Curr Biol. 2012;22:1825–30.

36. Kawakatsu T, Nery JR, Castanon R, Ecker JR. Dynamic DNA methylation reconfiguration during seed development and germination. Genome Biol. 2017;18:171.

37. Bouyer D, Kramdi A, Kassam M, Heese M, Schnittger A, Roudier F, et al. DNA methylation dynamics during early plant life. Genome Biol. 2017;18:179.

38. Lin J-Y, Le BH, Chen M, Henry KF, Hur J, Hsieh T-F, et al. Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. Proc Natl Acad Sci U S A. 2017;114:E9730–9.

39. Plotnikova A, Kellner MJ, Schon MA, Mosiolek M, Nodine MD. MicroRNA Dynamics and Functions During Arabidopsis Embryogenesis. Plant Cell. 2019;31:2929–46.

40. Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, et al. Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell. 2008;133:116–27.

41. Hofmann F, Schon MA, Nodine MD. The embryonic transcriptome of Arabidopsis thaliana. Plant Reprod. 2019;32:77–91.

42. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. R J. 2016;8:289–317.

43. Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol. 2014;21:64–72.

44. Cao X, Jacobsen SE. Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Curr Biol. 2002;12:1138–44.

45. Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature. 2006;443:1008–12.

46. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet. 2009;41:630–4.

47. Bies-Etheve N, Pontier D, Lahmy S, Picart C, Vega D, Cooke R, et al. RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. EMBO Rep. John Wiley & Sons, Ltd; 2009;10:649–54.

48. Clark SJ, Smallwood SA, Lee HJ, Krueger F, Reik W, Kelsey G. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). Nat Protoc. 2017;12:534–47.

49. Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M. Correction: Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. Elife [Internet]. 2015;4. Available

from: http://dx.doi.org/10.7554/eLife.08658

50. Kawakatsu T, Stuart T, Valdes M, Breakfield N, Schmitz RJ, Nery JR, et al. Unique cell-type-specific patterns of DNA methylation in the root meristem. Nat Plants. 2016;2:16058.

51. Vandepoele K, Raes J, De Veylder L, Rouzé P, Rombauts S, Inzé D. Genome-wide analysis of core cell cycle genes in Arabidopsis. Plant Cell. 2002;14:903–16.

52. Lyons DB, Zilberman D. DDM1 and Lsh remodelers allow methylation of DNA wrapped in nucleosomes [Internet]. eLife. 2017. Available from: http://dx.doi.org/10.7554/elife.30674

53. Gouil Q, Baulcombe DC. DNA Methylation Signatures of the Plant Chromomethyltransferases. PLoS Genet. 2016;12:e1006526.

54. Li X, Harris CJ, Zhong Z, Chen W, Liu R, Jia B, et al. Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. Proc Natl Acad Sci U S A. 2018;115:E8793–802.

55. Wendte JM, Zhang Y, Ji L, Shi X, Hazarika RR, Shahryary Y, et al. Epimutations are associated with CHROMOMETHYLASE 3-induced de novo DNA methylation. Elife [Internet]. 2019;8. Available from: http://dx.doi.org/10.7554/eLife.47891

56. Schoft VK, Chumak N, Mosiolek M, Slusarz L, Komnenovic V, Brownfield L, et al. Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. EMBO Rep. John Wiley & Sons, Ltd; 2009;10:1015–21.

57. Zhao P, Zhou X, Shen K, Liu Z, Cheng T, Liu D, et al. Two-Step Maternal-to-Zygotic Transition with Two-Phase Parental Genome Contributions. Dev Cell. 2019;49:882–93.e5.

58. Russanova VR, Driscoll CT, Howard BH. Adenovirus type 2 preferentially stimulates polymerase III transcription of Alu elements by relieving repression: a potential role for chromatin. Mol Cell Biol. 1995;15:4282–90.

59. Krishnakumar R, Gamble MJ, Frizzell KM, Berrocal JG, Kininis M, Kraus WL. Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. Science. 2008;319:819–21.

60. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997;389:251–60.

61. Fyodorov DV, Zhou B-R, Skoultchi AI, Bai Y. Emerging roles of linker histones in regulating chromatin structure and function. Nat Rev Mol Cell Biol. 2018;19:192–206.

62. Rutowicz K, Lirski M, Mermaz B, Teano G, Schubert J, Mestiri I, et al. Linker histones are fine-scale chromatin architects modulating developmental decisions in Arabidopsis. Genome Biol. 2019;20:157.

63. Choi J, Lyons DB, Kim MY, Moore JD, Zilberman D. DNA Methylation and Histone H1 Jointly Repress Transposable Elements and Aberrant Intragenic Transcripts. Mol Cell [Internet]. 2019; Available from: http://dx.doi.org/10.1016/j.molcel.2019.10.011

64. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated

single-base resolution maps of the epigenome in Arabidopsis. Cell. 2008;133:523–36.

65. He S, Vickers M, Zhang J, Feng X. Natural depletion of histone H1 in sex cells causes DNA demethylation, heterochromatin decondensation and transposon activation [Internet]. eLife. 2019. Available from: http://dx.doi.org/10.7554/elife.42530

66. Boisnard-Lorig C, Colon-Carmona A, Bauch M, Hodge S, Doerner P, Bancharel E, et al. Dynamic analyses of the expression of the HISTONE::YFP fusion protein in arabidopsis show that syncytial endosperm is divided in mitotic domains. Plant Cell. 2001;13:495–509.

67. Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, et al. A Role for RNAi in the Selective Correction of DNA Methylation Defects [Internet]. Science. 2009. p. 1600–4. Available from: http://dx.doi.org/10.1126/science.1165313

68. Kao P, Nodine MD. Transcriptional Activation of Arabidopsis Zygotes Is Required for Initial Cell Divisions. Sci Rep. 2019;9:17159.

69. Bouvet P, Dimitrov S, Wolffe AP. Specific regulation of Xenopus chromosomal 5S rRNA gene transcription in vivo by histone H1. Genes Dev. 1994;8:1147–59.

70. Wolffe AP. Dominant and specific repression of Xenopus oocyte 5S RNA genes and satellite I DNA by histone H1. EMBO J. 1989;8:527–37.

71. Mérai Z, Chumak N, García-Aguilar M, Hsieh T-F, Nishimura T, Schoft VK, et al. The AAA-ATPase molecular chaperone Cdc48/p97 disassembles sumoylated centromeres, decondenses heterochromatin, and activates ribosomal RNA genes. Proc Natl Acad Sci U S A. 2014;111:16166–71.

72. Jenik PD, Jurkuta REJ, Barton MK. Interactions between the cell cycle and embryonic patterning in Arabidopsis uncovered by a mutation in DNA polymerase epsilon. Plant Cell. 2005;17:3362–77.

73. Sonneville R, Craig G, Labib K, Gartner A, Blow JJ. Both Chromosome Decondensation and Condensation Are Dependent on DNA Replication in C. elegans Embryos. Cell Rep. 2015;12:405–17.

74. Müller S, Almouzni G. Chromatin dynamics during the cell cycle at centromeres. Nat Rev Genet. 2017;18:192–208.

75. van Zanten M, Koini MA, Geyer R, Liu Y, Brambilla V, Bartels D, et al. Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation. Proc Natl Acad Sci U S A. 2011;108:20219–24.

76. Pecinka A, Dinh HQ, Baubec T, Rosa M, Lettner N, Mittelsten Scheid O. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. Plant Cell. 2010;22:3118–29.

77. Probst AV, Mittelsten Scheid O. Stress-induced structural changes in plant chromatin. Curr Opin Plant Biol. 2015;27:8–16.

78. Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature. 2011;472:115–9.

79. Masuta Y, Nozawa K, Takagi H, Yaegashi H, Tanaka K, Ito T, et al. Inducible Transposition of a

Heat-Activated Retrotransposon in Tissue Culture. Plant Cell Physiol. 2017;58:375–84.

80. Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, et al. High-resolution mapping of open chromatin in the rice genome. Genome Res. 2012;22:151–62.

81. Lu X, Wontakal SN, Kavi H, Kim BJ, Guzzardo PM, Emelyanov AV, et al. Drosophila H1 Regulates the Genetic Activity of Heterochromatin by Recruitment of Su(var)3-9. Science. American Association for the Advancement of Science; 2013;340:78–81.

82. Bošković A, Eid A, Pontabry J, Ishiuchi T, Spiegelhalter C, Raghu Ram EVS, et al. Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. Genes Dev. 2014;28:1042–7.

83. Becker JS, Nicetto D, Zaret KS. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends Genet. 2016;32:29–41.

84. She W, Grimanelli D, Rutowicz K, Whitehead MWJ, Puzio M, Kotliński M, et al. Chromatin reprogramming during the somatic-to-reproductive cell fate transition in plants [Internet]. Development. 2013. p. 4008–19. Available from: http://dx.doi.org/10.1242/dev.095034

85. Rosa S, Ntoukakis V, Ohmido N, Pendle A, Abranches R, Shaw P. Cell differentiation and development in Arabidopsis are associated with changes in histone dynamics at the single-cell level. Plant Cell. 2014;26:4821–33.

86. Ikeuchi M, Favero DS, Sakamoto Y, Iwase A, Coleman D, Rymen B, et al. Molecular Mechanisms of Plant Regeneration [Internet]. Annual Review of Plant Biology. 2019. p. 377–406. Available from: http://dx.doi.org/10.1146/annurev-arplant-050718-100434

87. Li H, Soriano M, Cordewener J, Muiño JM, Riksen T, Fukuoka H, et al. The histone deacetylase inhibitor trichostatin a promotes totipotency in the male gametophyte. Plant Cell. 2014;26:195–209.

88. Ikeda-Iwai M, Umehara M, Satoh S, Kamada H. Stress-induced somatic embryogenesis in vegetative tissues of Arabidopsis thaliana. Plant J. Wiley Online Library; 2003;34:107–14.

89. Kamada H, Tachikawa Y, Saitou T, Harada H. Heat stress induction of carrot somatic embryogenesis. Plant tissue culture letters. Japanese Society for Plant Cell and Molecular Biology; 1994;11:229–32.

90. Li W, Liu H, Cheng ZJ, Su YH, Han HN, Zhang Y, et al. DNA methylation and histone modifications regulate de novo shoot regeneration in Arabidopsis by modulating WUSCHEL expression and auxin signaling. PLoS Genet. 2011;7:e1002243.

91. Erdmann RM, Satyaki PRV, Klosinska M, Gehring M. A Small RNA Pathway Mediates Allelic Dosage in Endosperm. Cell Rep. 2017;21:3364–72.

92. Narsai R, Gouil Q, Secco D, Srivastava A, Karpievitch YV, Liew LC, et al. Extensive transcriptomic and epigenomic remodelling occurs during Arabidopsis thaliana germination. Genome Biol. 2017;18:172.

93. Breakfield NW, Corcoran DL, Petricka JJ, Shen J, Sae-Seaw J, Rubio-Somoza I, et al. High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in Arabidopsis. Genome Res. 2012;22:163–76.

94. Xu L, Hu Y, Cao Y, Li J, Ma L, Li Y, et al. An expression atlas of miRNAs in Arabidopsis thaliana. Sci China

Life Sci. 2018;61:178–89.

95. Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, et al. Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat Genet. 2006;38:721–5.

96. Onodera Y, Haag JR, Ream T, Costa Nunes P, Pontes O, Pikaard CS. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell. 2005;120:613–22.

97. Yelagandula R, Stroud H, Holec S, Zhou K, Feng S, Zhong X, et al. The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. Cell. 2014;158:98–109.

98. Lutzmayer S, Enugutti B, Nodine MD. Novel small RNA spike-in oligonucleotides enable absolute normalization of small RNA-Seq data. Sci Rep. 2017;7:5913.

99. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 2012;40:D1202–10.

100. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44:W160–5.

101. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27:1571–2.

102. Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. Nature. 2015;523:212–6.

103. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

104. Stempor P, Ahringer J. SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics. Wellcome Open Res. 2016;1:14.

105. Chen F, Tillberg PW, Boyden ES. Expansion microscopy [Internet]. Science. 2015. p. 543–8. Available from: http://dx.doi.org/10.1126/science.1260088

Main figures and legends
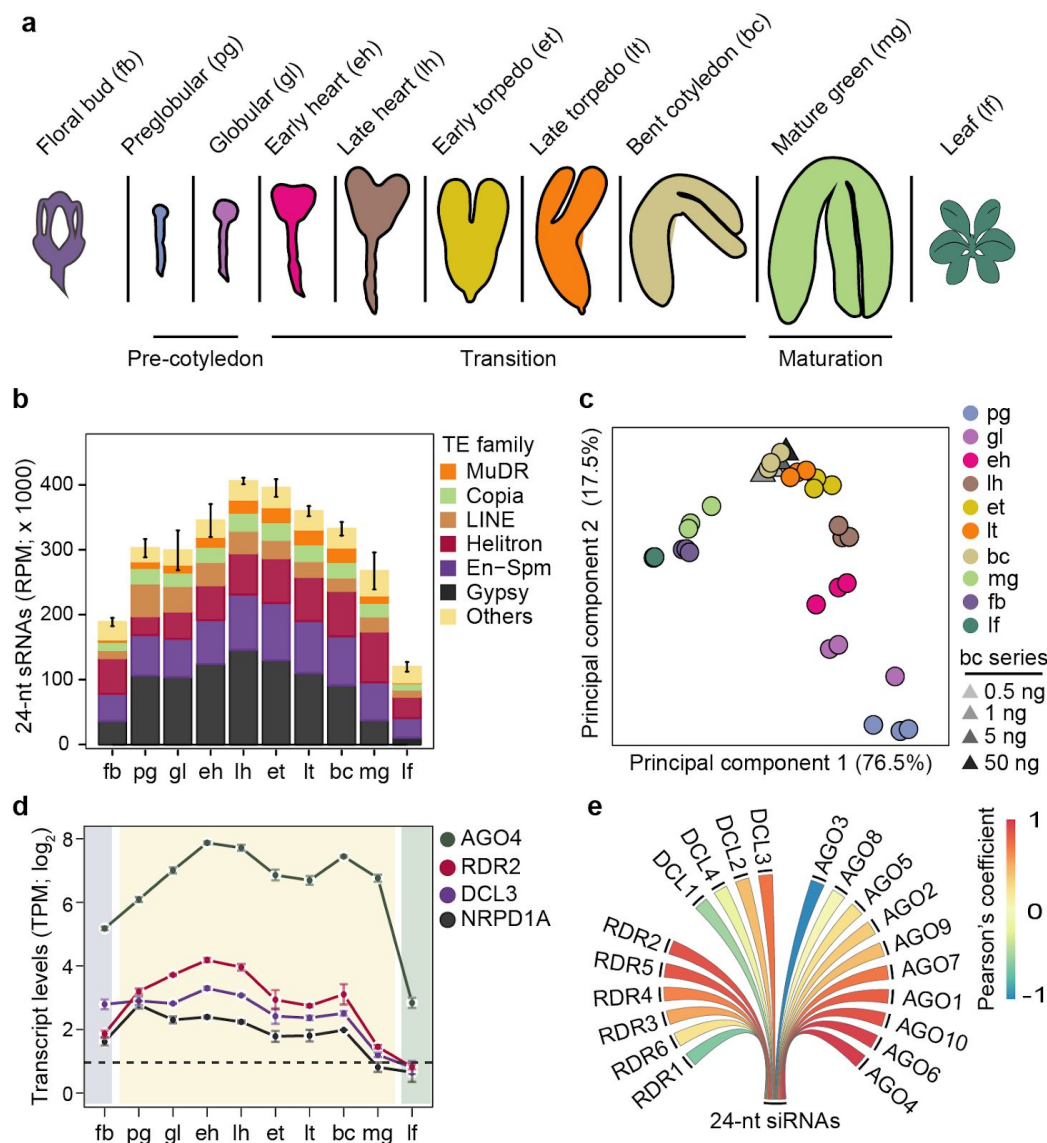
Fig 1. Embryos are enriched for transposon-derived small RNAs. **a** Schematic of embryonic stages and tissues previously used for small RNA profiling [39] and analyzed in this study. **b** Stacked bar chart representing the abundance of 24-nt sRNAs from various TE families in floral buds, embryos, and leaves (key). Error bars indicate standard error of mean TE-derived 24-nt siRNA levels from three biological replicates. RPM, sRNA-seq reads per million genome-matching reads. **c** Principal component analysis of 24-nt sRNAs mapping to TEs in floral buds, embryos, leaves and a dilution series of RNA isolated from bent cotyledon stage embryos (key). **d** Line chart illustrating transcript levels of 24-nt siRNA **(Fig 1**

continued) biogenesis factors in embryonic and post-embryonic tissues based on mRNA-seq [41]. Error bars represent standard error of mean transcripts from three biological replicates. TPM, transcripts per million. **e** Pearson's correlation coefficients between means of TE-derived 24-nt siRNAs and transcript levels of siRNA biogenesis and effector genes in floral buds, embryos and leaves. Pearson's correlation coefficient values are represented according to the key. See also Fig. S1.

Fig 2. Small RNAs from euchromatic and heterochromatic transposons exhibit distinct developmental dynamics. **a** Line graphs illustrating 24-nt siRNA levels from class A and B TEs in embryonic and post-embryonic tissues. Dashed lines represent the detection criteria used to select TEs yielding siRNAs (2 RPM, reads per million genome-matching reads). The number of TEs belonging to each class are indicated. Polygons represent the standard deviation of mean 24-nt siRNA levels. fb, floral buds; pg, preglobular; gl, globular; eh, early heart; lh, late heart; et, early torpedo; lt, late torpedo; bc, bent

(Fig 2 continued) cotyledon; mg, mature green; lf, leaf. **b** Heat map depicting 24-nt siRNA levels (RPM) from class A and B TEs across development. siRNA levels from ±2-kb of TEs are color-coded according to the key, and samples are labelled as in (A). Rows were ordered based on total 24-nt siRNA levels. **c** Line metaplot showing 24-nt siRNA levels from class A and B TEs. TEs were aligned on their 5′ ends, which are indicated by the grey vertical line. Samples are labeled according to the key and as in (A). **d** Circos plot representing densities of siRNA-deficient (-siRNAs; grey), class A (red) and class B (purple) TEs across the five Arabidopsis chromosomes (chr). Centromeres are indicated by dots. **e** Boxplot of TE lengths for either all annotated, siRNA-deficient, class A or class B TEs. Dashed red line represents median size of all annotated TEs. Thick horizontal bars indicate medians, and the top and bottom edges of the box indicate the 75th and 25th percentiles, respectively. kb, kilobases; *P* values < 0.0001 based on Mann-Whitney U test of differences between all annotated and TE classes are represented by *. **f** Stacked bar charts illustrating proportion of TEs that are CHH hypomethylated in *drm1/drm2* (RdDM; red), *cmt2* (CMT2; purple), both *drm1/drm2* and *cmt2* (both; yellow) or which were not methylated (unmethyl; black) in leaves [26]. See also Fig. S2.
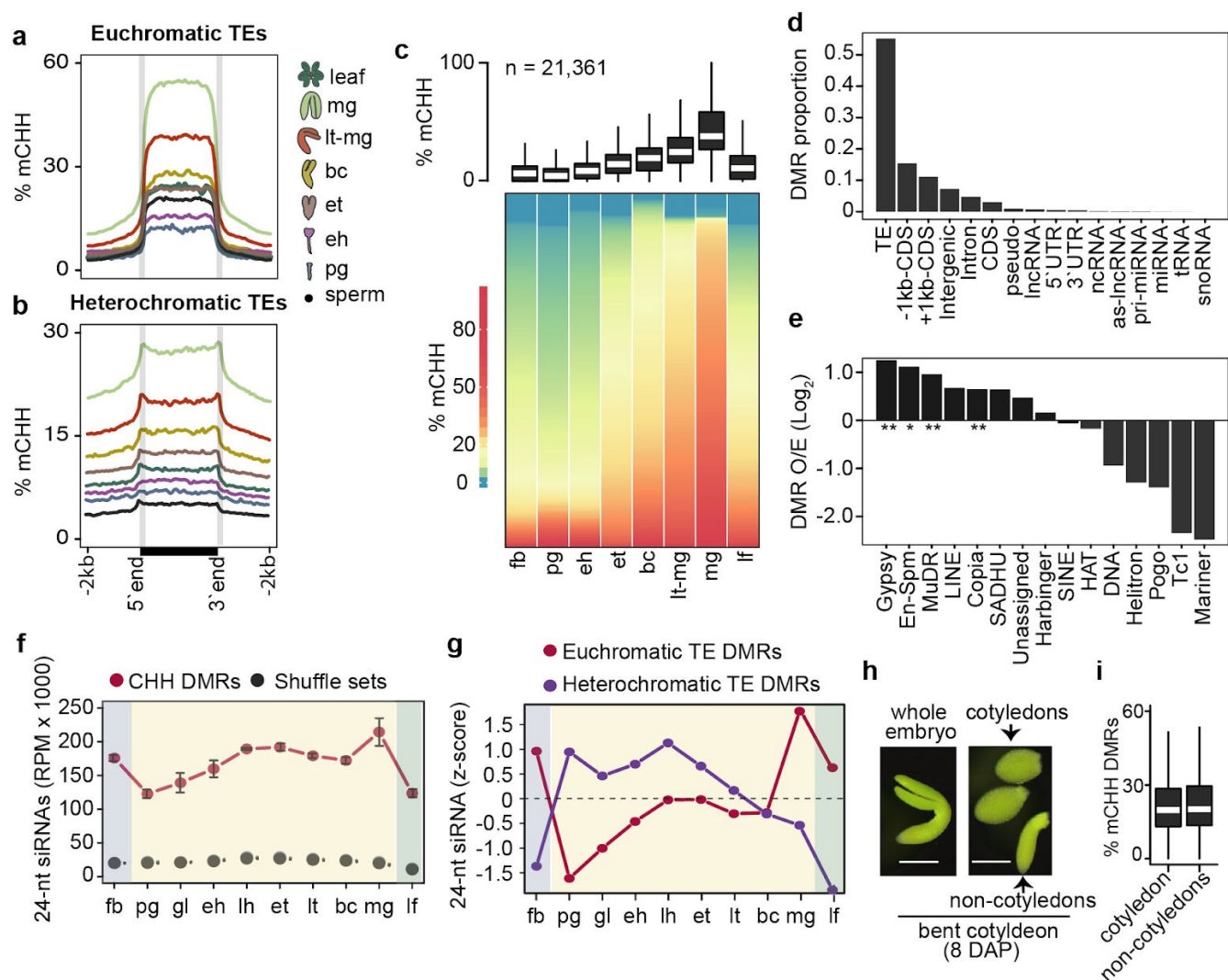
Fig 3. Embryonic methylome dynamics. **a** and **b** Metaplots of average weighted CHH methylation percentages across euchromatic (A) and heterochromatic (B) TEs in sperm, embryos and leaves (key). pg, preglobular; eh, early heart; et, early torpedo; bc, bent cotyledon; lt-mg, late torpedo-to-early mature green; mg, mature green. **c** Boxplot and heat map illustrating the percentage of CHH methylation across DMRs in flowers, embryos and leaves. The number of DMRs identified are indicated, and stages are labelled as in (a) and include floral buds (fb). Thick horizontal bars in the boxplot indicate medians, and the top and bottom edges of the box indicate the 75th and 25th percentiles, respectively. DMRs in heatmap were sorted by average methylation levels per column. **d** Proportion of genomic features overlapping DMRs.

(Fig 3 continued) e Bar chart showing the enrichment of TE families observed overlapping DMRs relative to those expected based on their genomic proportions (O/E; $\log_2$). *P* values < 0.01 and < 0.05 based on Fisher's exact test are represented by ** and *. **f** Line chart of 24-nt siRNA levels overlapping CHH DMRs (red) compared to randomly selected genomic regions with equal sizes and quantities (grey) in floral buds (fb), embryos and leaves (lf). Embryo stages are labelled as in Fig. 1a. **g** Relative 24-nt siRNA levels (z-scores) on DMRs mapping to euchromatic (red) and heterochromatic (purple) TEs across development. Embryo stages are labelled as in Fig. 1a. **h** Representative image of a bent cotyledon staged embryo eight days after pollination (DAP) and dissected into cotyledon and non-cotyledon tissues for methylC-seq. Scale bar represents 0.25 millimeters. **i** Boxplot of CHH methylation percentages of DMRs defined in (c) for cotyledon and non-cotyledon tissues. Boxplots are as described in (b). See also Fig. S3.
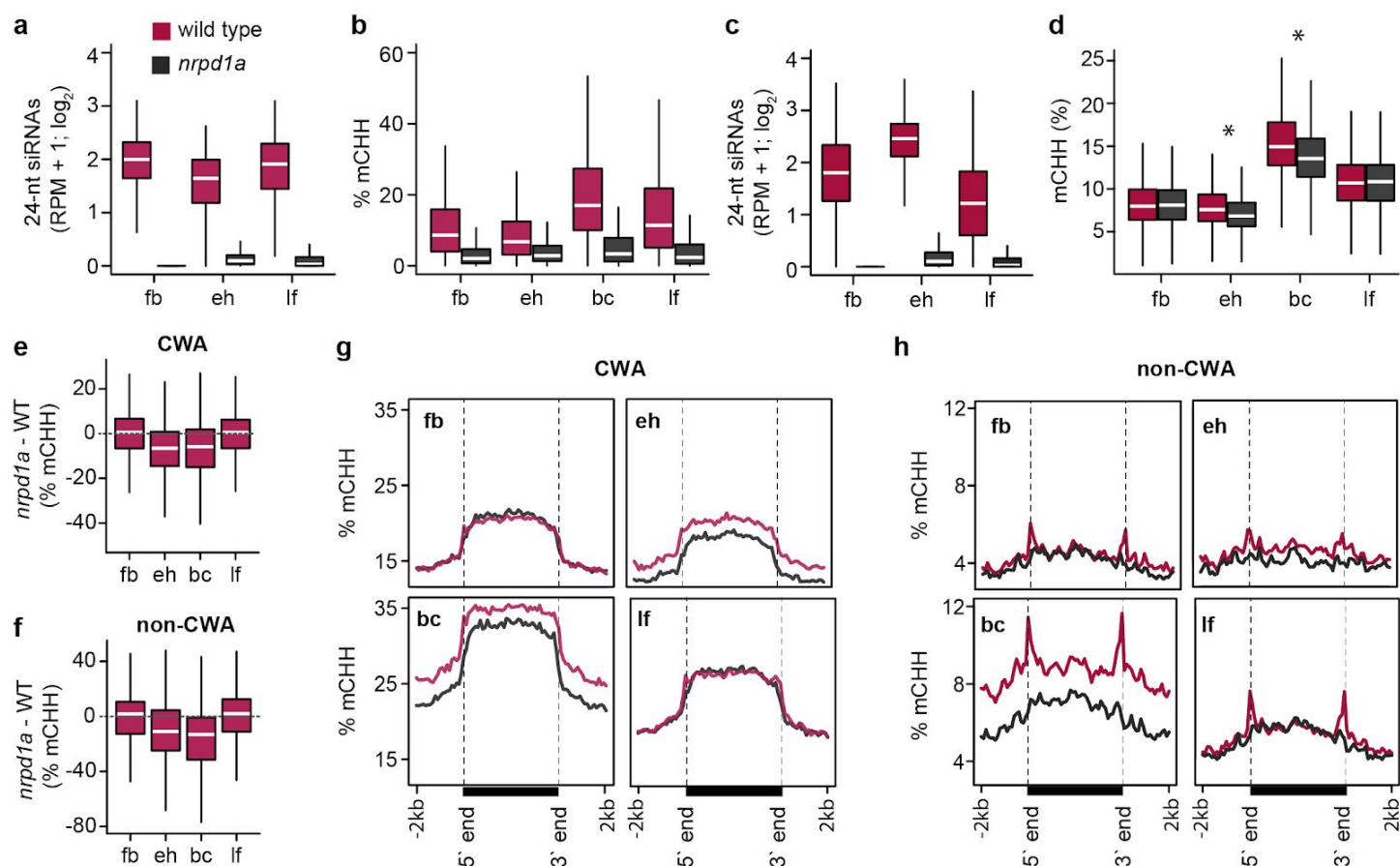
Fig 4. Small RNA-directed methylation of transposons during embryogenesis. **a** and **c** Boxplots of 24-nt siRNA levels in wild type and *nrpd1a* mutant floral buds (fb), early heart embryos (eh) and leaves (lf) derived from euchromatic (A) or heterochromatic (C) TEs. Thick horizontal bars indicate medians, and the top and bottom edges of the box indicate the 75th and 25th percentiles, respectively. **b** and **d** Boxplots of CHH methylation levels in wild type and *nrpd1a* mutant floral buds (fb), early heart embryos (eh), bent cotyledon embryos (bc) and leaves (lf) for euchromatic (b) or heterochromatic (d) TEs. *P* values < 0.0001 based on Mann-Whitney U test of methylation differences between wild-type and *nrpd1a* tissues are represented by * in (b). All differences displayed in (a-c**)** had *P* values < 0.0001 . **e** and **f** Boxplots of CHH methylation differences between *nrpd1a* and wild type tissues for long heterochromatic TEs in either CWA (e) or non-CWA (f) contexts. **g** and **h** Metaplots of average weighted CHH methylation percentages of long heterochromatic TEs in *nrpd1a* and wild type tissues in either CWA (g) or non-CWA (h) contexts. See also Fig. S4.
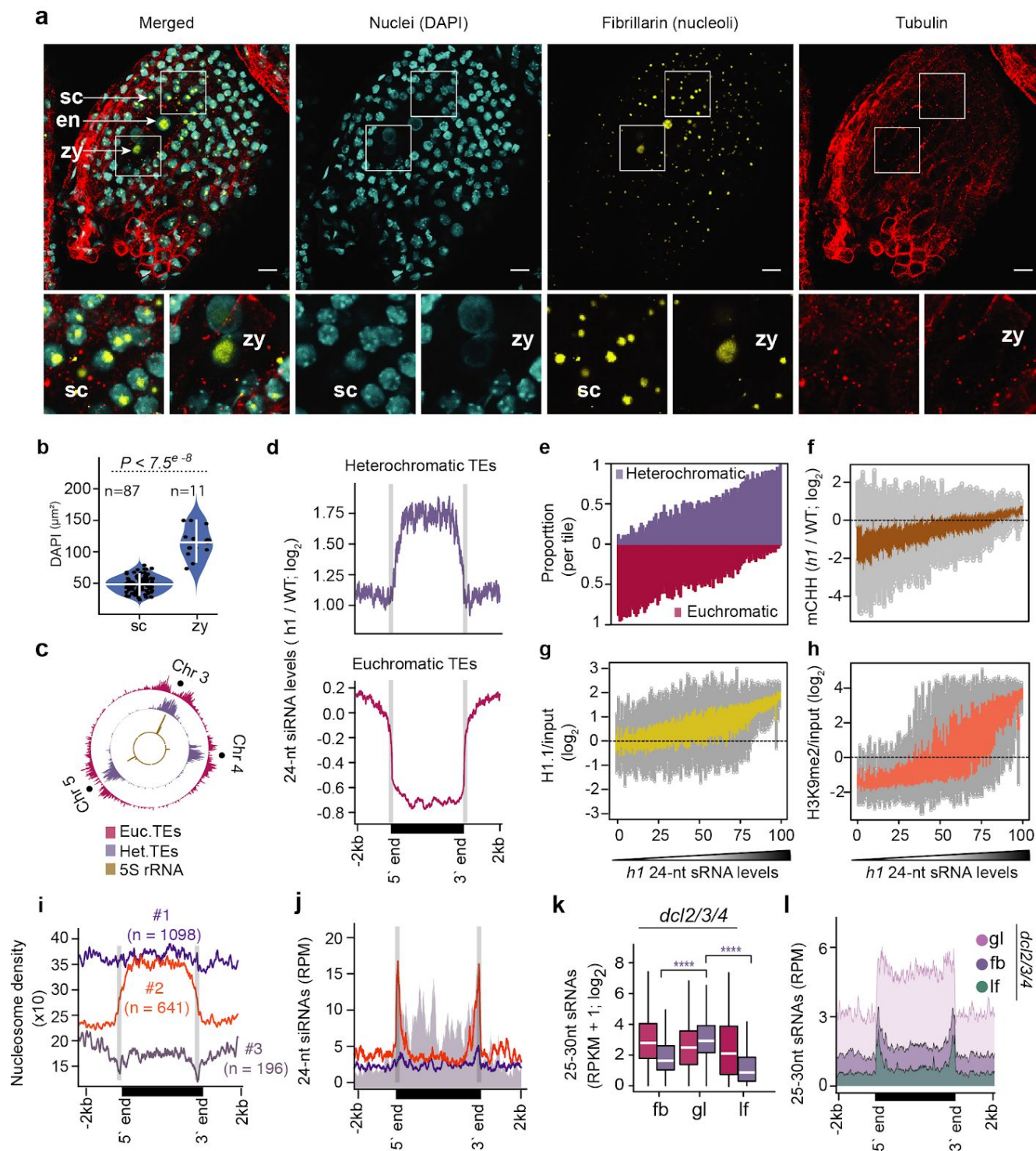
Fig 5. Chromatin regulates small RNA biogenesis. **a** Representative expansion microscopy images of DAPI-stained nuclei (cyan), fibrillarin (yellow) and tubulin (red) in zygote-containing seeds (*top*). Zoomed sections of seed coat and

(Fig 5 continued) zygote marked in white squares (*bottom*). zy, zygote; en, endosperm; sc, seed coat. Scale bars represent 20 μm. **b** Violin and dot plot representing the DAPI area of seed coat (sc) and zygote (zy) nuclei. Horizontal and vertical white colored bars represent median and interquartile range, respectively. *P* values based on the Wilcoxon rank sum test of DAPI area differences between seed coat (sc) and zygote (zy) is shown. **c** Circos plot representing densities of euchromatic (red), heterochromatic (purple) TEs and 5S rRNA loci on chromosomes 3, 4 and 5. Centromeres are indicated by dots. **d** Metaplots of 24-nt siRNA levels in *h1* mutants relative to wild-type leaves for heterochromatic (top) or euchromatic (bottom) TEs. **e-h** TEs were divided into percentiles, ordered based on their 24-nt siRNA levels in *h1* mutants, and plotted according to proportion of euchromatic and heterochromatic TEs (**e**), CHH methylation levels in *h1* relative to wild type [27] (**f**), relative enrichment of H1.1(**g**) and relative enrichment of H3K9me2 (**h**) [63]. **i** and **j** Metaplots displaying nucleosome occupancy of long heterochromatic TE groups 1-3 according to MNase-seq datasets [62] (**i**) or 24-nt siRNA levels (**j**). Annotated 5′ and 3′ ends are labelled and indicated by vertical grey lines. **k** Boxplots of 25-30 nt precursor siRNA levels of euchromatic (red) and heterochromatic (purple) TEs in *dcl2/3/4* floral buds (fb), globular embryos (gl) and leaves (lf). Thick horizontal bars indicate medians, and the top and bottom edges of the box indicate the 75th and 25th percentiles, respectively. *P* values < 0.0001 based on Mann-Whitney U test of differences between globular and floral bud or leaf 25-to-30 nt siRNAs derived from heterochromatic TEs (purple) tissues are represented by ****. **l** Metaplots of 25-30 nt siRNA levels overlapping long heterochromatic TEs in *dcl2/3/4* floral buds (fb), globular embryos (gl) and leaves (lf) (key). Annotated 5′ and 3′ ends are labelled at the bottom. See also Fig. S5.
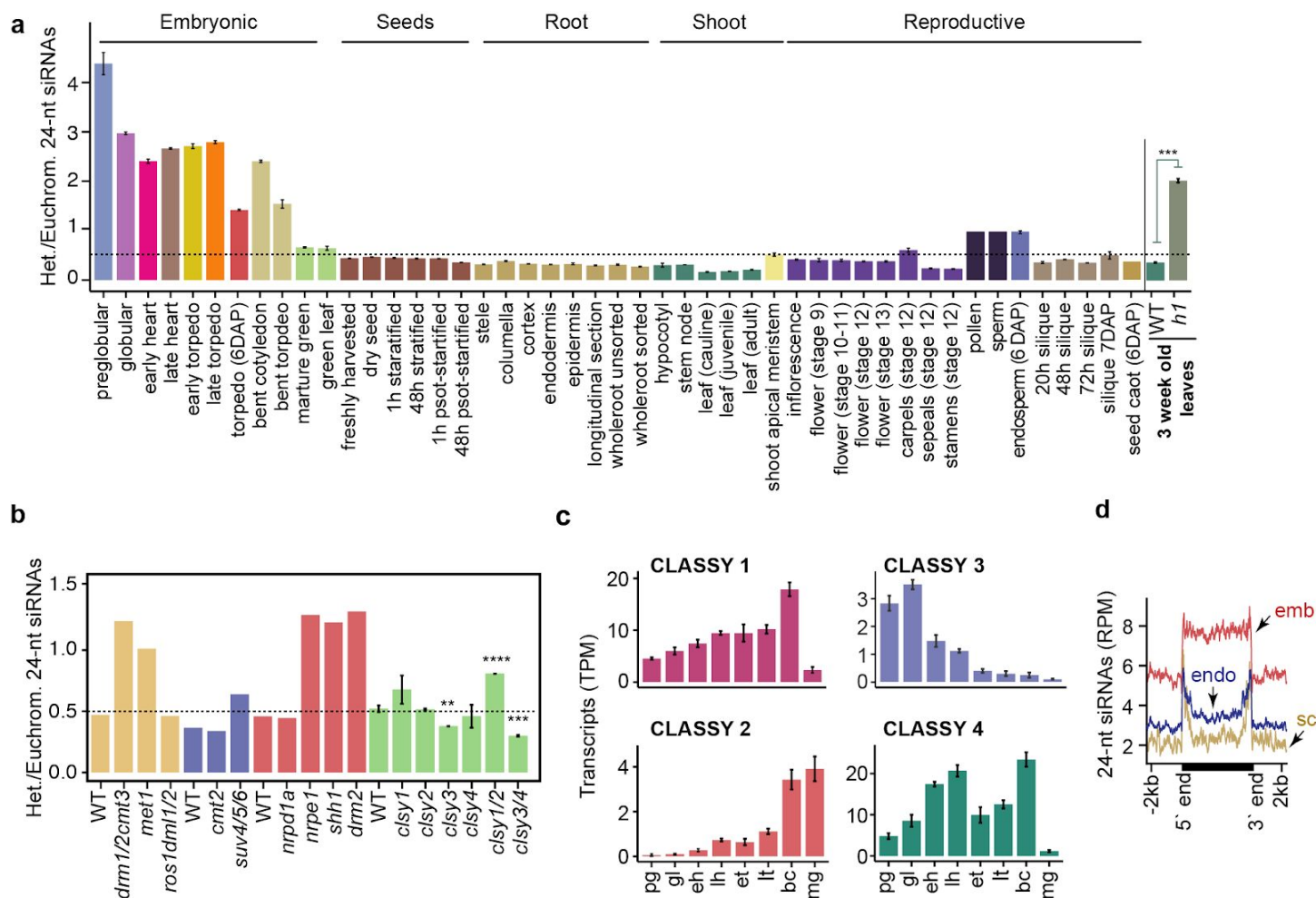
Fig 6. Homeostasis of transposon-derived siRNAs. **a** Bar chart illustrating the relative enrichment of length-normalized 24-nt siRNAs derived from heterochromatic relative to euchromatic TEs across development, and in *h1* mutant leaves. *P* values < 0.001 based on Student's *t*-test of heterochromatic vs euchromatic 24-nt siRNA enrichment differences between wild-type and *h1* tissues are represented by ***. Various tissues and cell-types from different phases of the plant life-cycle are labelled and include those from published datasets [29,39,91–94]. **b** Relative enrichments of 24-nt siRNAs derived from heterochromatic relative to euchromatic TEs in RdDM and related mutants. Wild type (WT), *drm1/2 cmt3*, *met1* and *ros1 dml1/2* [64]; WT, *cmt2* and *suv4/5/6* [43]; WT, *nrpd1a*, *nrpe1*, *shh1* and *drm2* [21]; WT, *clsy1*, *clsy2*, *clsy3*, *clsy4*, *clsy1/2* and *clsy3/4* [19]. *P* values < 0.0001, < 0.001 and < 0.01 based on Student's t-test of

(Fig 6 continued) heterochromatic vs euchromatic 24-nt siRNA enrichment differences between wild-type and *mutants* are represented by ****, *** and **. **c** CLASSY1, CLASSY2, CLASSY3 and CLASSY4 transcript levels (TPM) during embryogenesis. **d** Metaplots of 24-nt siRNA levels overlapping long heterochromatic TEs in 6 DAP embryo (red; emb), endosperm (blue; endo) or seed coat (yellow; sc) tissues based on [91]. See also Fig. S6.
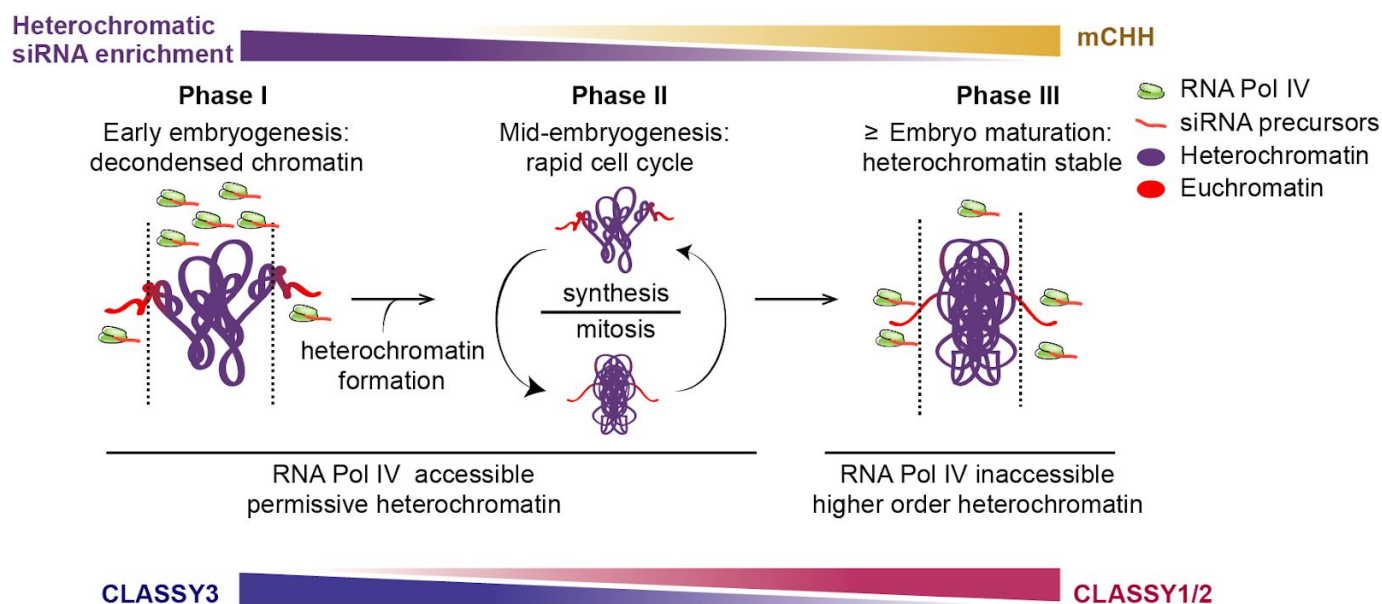
Fig 7. Three-phase model of how interplay between chromatin states and siRNAs shape the nascent epigenome. Heterochromatic siRNA enrichment, CHH methylation and CLSY levels are indicated. Symbols representing RNA PolIV, 24-nt siRNA precursors, euchromatin and heterochromatin are indicated in the key. See text in Discussion for details.

Supplementary Information

Supplemental Materials

Figure S1. Characteristics of embryonic 24-nt siRNAs and their similarities across samples.

Figure S2. siRNA dynamics and characteristics.

Figure S3. Benchmarking low-input methylC-seq, methylomes and cell-cycle transcripts.

Figure S4. Size-based partitioning of heterochromatic TEs and small RNA-directed methylation.

Figure S5. TE-derived siRNA accumulation and association with chromatin.

Figure S6. TE-derived siRNAs in methylation mutants

Supplemental Data 1. Datasets and General Mapping Statistics

Supplemental Data 2. Euchromatic and heterochromatic TE Classifications

Supplemental Data 3. Differentially Methylated Regions