Supplemental Figure Legends

Figure S1. DNA methylation in mutant lines, related to Figure 1. (A) Genomic hypomethylation in mutant lines shown as the percentage of wild-type methylation. Methyltransferase mutant lines with the strongest methylation loss in a given context are denoted by asterisks. Mean genomic methylation levels can be found in Table S1. (B) Patterns of TE DNA methylation (CG, CHG and CHH) in wild-type and four independent *cmt2* alleles. *Arabidopsis* TEs were aligned at the 5′ end or the 3′ end, and average methylation levels for all cytosines within each 100-bp interval are plotted. The dashed lines represent the points of alignment. The *cmt2-3* allele is used for *cmt2* analyses throughout the manuscript. (C) Distribution of TEs by size (upper panels) and type (lower panels) across *Arabidopsis* chromosomes. (D) LOWESS fit of the DNA methylation distribution averaged in 100-kb bins across chromosome 1 in wild-type and indicated mutants. (E) LOWESS fit of the TE DNA methylation distribution calculated in 50-bp windows across chromosome 1 in wild-type and indicated mutants. Windows with fractional CG, CHG and CHH methylation of at least 0.5, 0.3 and 0.1, respectively, in at least one of the samples are shown. Note that DNA methylation in the pericentric chromocenter (ChC) preferentially requires DDM1 (blue trace) and methylation on the arms preferentially requires DRD1 (black trace). Also note the strong synergistic CHH hypomethylation of the ChC in *ddm1drd1* (pink trace). (F) DNA methylation in wild-type and indicated mutants was averaged specifically in long TEs (> 4 kb) as described for (B). Note the loss of CHH methylation at TE edges in *drm2* (black trace), but not in *cmt2* (blue trace), and the converse loss of CHH methylation from TE bodies in *cmt2* but not in *drm2*.

Figure S2. Correlations between DNA methylation dependence and chromatin features, related to Figure 2. (A) Box plots of nucleosome occupancy from this study (red) and from (Chodavarapu et al., 2010) (blue) of 50-bp windows within TEs of different sizes. (B) Box plots of DDM1- and DRD1-mediated TE CHH methylation within 50-bp windows with sRNA level of 6 counts and different levels of nucleosome counts (left to right: 0 to 6, 7 to 16, 17 to 34 and 34 to 76), H3K9me2 (left to right: log2 of -1 to 1.2, 1.2 to 1.9, 1.9 to 2.5, and 2.5 to 3.9), and H3K4me2 (left to right: log2 of -4 to -2, -2 to -1.5, -1.5 to -0.7, and -0.7 to 1). (C) Box plots of DRD1-mediated TE methylation within 50-bp windows with different levels of sRNA (left to right: 0, 1 to 2, 3 to 6, 7 to 15, and over 15). (D) Spearman correlation coefficients between DRD1-mediated TE DNA methylation (*ddm1* DNA methylation minus *ddm1drd1* DNA methylation) and chromatin features within 50-bp windows with sRNA level of 6 counts.

Figure S3. DNA methylation phenotypes of *h1* plants, related to Figure 3. (A) Genomic structure of *H1.1* and *H1.2* and their T-DNA insertions (triangles). Black lines represent UTRs, kinked lines represent introns, and boxes represent exons. (B-C) qPCR analysis (B) and Western blot (C) of wild-type (WT), *h1.1* (only in Western), *h1.2* (only in Western), and *h1*. The antibody against H1.1 is apparently specific, whereas the antibody against H1.2 cross-reacts with H1.1, as evidenced by consistently stronger signal in *h1.2* compared to *h1*. The signal with the anti-H1.2 antibody is also consistently higher in *h1.1* compared to WT, suggesting that the *H1.2* gene is up-regulated in *h1.1*. (D) M-spline curve fits of absolute DNA methylation levels in 50-bp windows plotted against H3K9me2. (E) DNA methylation of indicated euchromatic TEs, heterochromatic TEs, and genes in wild type, *h1*, *ddm1*, and two biological replicates of *h1ddm1*. Genes and TEs oriented 5′ to 3′ and 3′ to 5′ are shown above and below the line, respectively.

Figure S4. CG and CHG methylation of four TE families, related to Figure 4. CG and CHG methylation in wild-type and indicated mutants in four types of TEs, aligned as in Figure 4.

Figure S5. Synergistic TE demethylation and activation in *ddm1drd1* and *ddm1rdr2*, related to Figure 5. (A) Box plots of absolute fractional CG and CHG demethylation of 50-bp windows in TEs that are at least 32-fold overexpressed either in *drd1* or in *ddm1*. (B) DNA sequencing coverage (log2(reads in mutant/reads in wild-type)), DNA methylation and RNA levels of a section of chromosome 2 encompassing the DNA transposon *CACTA1* (*AT2TE20205*). The sequence coverage is indicative of *CACTA1* copy number relative to wild type.

Figure S6. Distinct features of euchromatic and heterochromatic genes, related to Figure 6. (A) LOWESS fit of the distribution of euchromatic and heterochromatic genes summed in 100-kb windows across chromosome 1. (B) Box plots of averaged CHH methylation in TEs, euchromatic genes (mCG < 0.6), and heterochromatic genes (mCG > 0.6). (C) Kernel density plots of methylation differences between *ddm1* and wild-type (blue trace), and between *h1* and wild-type (red trace; positive numbers indicate greater methylation in the mutant line). (D) Distribution of CHG methylation in representative genes *AT1G76580* and *AT5G07920*. (E) Patterns of genic DNA methylation (CG, CHG and CHH) in wild-type and *ibm1drd1*, plotted as in Figure 6D. (F) DNA methylation and RNA levels of a section of chromosome 3 encompassing the *IBM1* gene (AT3G07610). The long *IBM1* transcript (IBM1-L) is downregulated in *met1*, but not in *ddm1* or *drd1*.

Figure S7. Genes regulated by DNA methylation, related to Figure 7. (A) Number of significantly upregulated and downregulated genes in the indicated mutants. (B) CHH methylation and RNA levels near *APUM9* (*AT1G35730*) – hypomethylation in the promoter of *APUM9* in *drd1* and *ddm1drd1* is associated with upregulation of *APUM9*. (C) DDM1 and RdDM collaborate to repress *FWA* expression. sRNA, H3K9me2, DNA methylation, and RNA levels near the *FWA* gene (*AT4G2553*0).

	Nuclear					Chloroplast
Sample	Mean coverage	C	CG	CHG	CHH	CHH
WT roots	13.69	6.68%	25.82%	8.05%	2.70%	0.16%
WT shoots	3.08	6.48%	25.99%	8.76%	2.49%	0.05%
drd1-7	22.61	5.78%	25.51%	7.19%	1.82%	0.04%
rdr2-1	39.76	5.61%	25.16%	7.16%	1.80%	0.20%
ddm1-2	13.50	3.32%	10.85%	3.48%	1.84%	0.20%
ddm1drd1	10.42	2.08%	10.79%	1.65%	0.45%	0.02%
ddm1rdr2	85.37	2.01%	10.95%	1.74%	0.46%	0.11%
met1-6	15.50	2.28%	1.02%	6.78%	1.59%	0.05%
cmt3-12	6.11	4.96%	26.30%	1.12%	1.87%	0.18%
cmt2-3	3.44	5.20%	25.94%	7.76%	0.75%	0.16%
cmt2-4	9.15	5.42%	26.67%	8.56%	0.991%	0.34%
cmt2-5	7.61	5.45%	26.18%	8.82%	1.05%	0.35%
cmt2-6	8.15	5.41%	26.49%	8.88%	1.01%	0.35%
drm2-3	4.65	5.66%	26.09%	7.55%	1.61%	0.17%
ibm1drd1	75.49	6.1%	25.4%	11.2%	1.6%	0.11%
H1;DDM1	26.85	6.0%	23.1%	8.2%	2.6%	0.6%
H1;*ddm1*	22.09	3.2%	12.6%	3.1%	1.6%	0.4%
h1;ddm1 replicate 1	17.27	4.5%	18.3%	6.1%	1.8%	0.1%
h1;ddm1 replicate 2	18.47	4.5%	18.2%	5.9%	1.8%	0.3%
h1;DDM1	26.29	5.5%	22.3%	8.6%	1.9%	0.1%
H1 roots	11.82	6.54%	26.50%	8.39%	2.61%	0.22%
h1 roots	8.89	6.89%	27.12%	10.69%	2.54%	0.13%

Table S1. Mean genomic coverage and DNA methylation for wild-type and mutant samples, related to Figure 1. *drd1, rdr2, ddm1,* and *met1* samples are from roots; *ddm1drd1, ddm1rdr2, cmt3, cmt2, drm2, ibm1, ibm1drd1,* H1DDM1, H1*ddm1, h1ddm1* (both replicates), and *h1*DDM1 samples are from 2-3 week shoots. Chloroplast CHH methylation is a measure of cytosine non-conversion and other errors.

	cmt2	sRNA	*drd1*	*rdr2*	*drm2*
sRNA	0.115				
drd1	-0.25	-0.057			
rdr2	-0.229	-0.042	0.7128		
drm2	-0.141	-0.003	0.5181	0.493	
ddm1	0.4022	0.1053	-0.111	-0.136	-0.018

Table S2. Pearson correlation coefficients between CHH methylation in mutants and sRNA, related to Figure 2. Pearson correlation coefficients were calculated between CHH methylation in the indicated mutants and sRNA in 50-bp windows that correspond to TEs.

Extended Experimental Procedures

Phylogenetic analysis. The catalytic domains of the indicated CMT methyltransferase proteins were aligned using MUSCLE v3.7. All alignment files were checked, modified and converted to NEXUS files using MacClade v4.06. Phylogenetic trees were inferred using MrBayes v3.1.2. Two independent Markov chain Monte Carlo (MCMC) runs of four chains using the default fixed Poisson model were started from independent random trees, and were carried through four million generations, with trees sampled every 100th generation. Convergence was confirmed by checking that the standard deviations of split frequencies were <0.01, that the log probabilities of the data given the parameter values fluctuated within narrow limits, that the PSRF (potential scale reduction factor) = 1.000, and by examining the plot of the generation versus the log probability of the data. The first 7,500 stored trees from each run were discarded and the remaining 32,501 trees were used to construct the consensus tree. Final trees were checked and graphically presented in FigTree v1.2.2 (http://tree.bio.ed.ac.uk/software/figtree). Trees were made from the highly conserved amino acid blocks by excluding weakly conserved N- and C-termini along with unique inserts, yielding a final dataset of 487 characters.

qPCR. We used three biological replicates of *h1* and wild-type roots. RNA was extracted using RNeasy Plant Mini Kit (Qiagen). cDNA libraries were generated using RevertAid First Strand cDNA Synthesis Kit (Thermo Sci) and the primers spanning the exons were used to determine the transcription level.

Coomassie staining and Western blot. Antibodies were produced by Pro-Sci (San Diego, CA) against synthetic peptides, then affinity purified using the appropriate epitopes as described previously (Ascenzi and Gantt, 1999). Total protein from roots was prepared as previously described (Tsugama et al., 2011) with minor modifications. The roots were ground into a fine powder and lysed in a solution containing 10 mM EDTA, 0.12M Tris-HCl, 4% SDS, 10% betamercapto ethanol, 5% glycerol and 0.005% bromophenol blue. Proteins were separated in an Any-KD mini-PROTEAN TGX precast gel. For coomassie staining, the gel was stained in a colloidal coomassie solution. For Western blotting, proteins were transferred onto nitrocellulose in a mini-PROTEAN tetra blotting module according to the manufacturer's instructions (Bio-Rad). Binding of antibodies was performed using 1:5,000 dilution of primary antibody and 1:10,000 dilution of HRP-conjugated goat anti-rabbit IgG secondary antibody (Thermo Sci). Signal was detected using SuperSignal West Pico Chemiluminescent Substrate (Thermo Sci).

BS-seq. About 500 ng of genomic DNA was isolated from roots or seedlings, fragmented by sonication, end repaired and ligated to custom-synthesized methylated adapters (Eurofins MWG Operon) according to the manufacturer's (Illumina) instructions for gDNA library construction. Adaptor-ligated libraries were subjected to two successive treatments of sodium bisulfite conversion using the EpiTect Bisulfite kit (Qiagen) as outlined in the manufacturer's instructions. The bisulfite-converted libraries were then amplified by PCR using the following conditions: 2.5 U of ExTaq DNA polymerase (Takara Bio), 5 µl of 10X Extaq

- C 60 sec. The enriched libraries were either gel-purified (~300 bp band) or purified with solid phase reversible immobilization (SPRI) method using AM-Pure beads (Beckman Coulter) prior to quantification

with a Bioanalyzer (Agilent). Sequencing on the Illumina platform was performed at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley.

RNA-seq. Total RNA samples from 3-week-old roots were isolated using the RNeasy mini kit (Qiagen #74106) including on-column DNase treatment. mRNA was purified from 10-50 µg of total RNA by either (i) two cycles of poly-A enrichment using the Oligotex kit (Qiagen #72022), followed by a rRNA removal step using the RiboMinus Eukaryote Kit for RNA-Seq (Invitrogen #A1083702), or by (ii) using the Ribo-Zero rRNA Removal Kit (Plant Leaf). Precipitated mRNA samples were eluted with 9 µl of RNase free water and fragmented with 1 µl of 10X fragmentation buffer (Ambion, #AM8740) at 70 C. Reactions were stopped after 5 minutes by adding 1 µl Stop buffer and RNA was purified by ethanol precipitation. cDNA was synthesized from 100-300 ng of mRNA using SuperScript III reverse transcriptase (Invitrogen #18080-051). Double-stranded DNA was synthesized according to the instructions for the SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen). DNA was cleaned with a QIAquick PCR spin column (Qiagen, #28106), sequencing adapters were ligated according to the Illumina protocol and library was amplified by 18 cycles of PCR using Phusion DNA polymerase (NEB, #F-530). Bands around 300 bp were gel-purified and cloned for validation. Traditional sequencing confirmed that the libraries were properly constructed, showing high percentage of mRNA over rRNA. The libraries were sequenced at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley to generate single ends (SE) 36-50 base reads.

MNase-seq. Arabidopsis roots (1 g) were ground in liquid nitrogen, resuspended in 20 ml of HBM buffer (25mM Tris, pH 7.6, 0.44M Sucrose, 10 mM $MgCl_2$, 2 mM spermine and 0.1% Triton X-100), homogenized, filtered through miracloth, transferred to a 30 ml round bottom glass tube, centrifuged at 2000g (4 C) for 10 min and resuspended in 1 ml HBB buffer (25mM Tris, pH 7.6, 0.44M Sucrose, 10 mM $MgCl_2$ and 0.1% Triton X-100). Nuclei were further spun down at 200g, 4 C for 2 min and resuspended in 1 ml of TNE buffer (10 mM Tris, pH 8.0, 100 mM NaCl and 1 mM EDTA). MNase digestion was done with 4 ul of 1M $CaCl_2$ and 1 ul of diluted (1/20) MNase (200ul/ml; Sigma #N-3755) per 100 ul of pellet nuclei. Nuclei were then divided to several tubes and digestion was stopped at 45 sec intervals with 10mM EDTA. Digested nuclei were spun down at maximum speed for 5 min at 4 C, and released soluble nucleosomes were collected from the supernatant. Following RNase A and proteinase K digestion, DNA was purified using phenol/chloroform. Purified DNA samples were run on a 2% agarose gel, digested samples with most enriched intact mononucleosomes were chosen and bands corresponding to ~150 bp were cut and purified with a Gel Purification kit (Qiagen). Illumina libraries were constructed and sequenced at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley to generate paired ends (PE) 36 base reads.

Kernel density plots. We used data from sibling *h1* and wild-type roots because the sibling *h1* and wild-type seedlings used in Figure 3B-D and other *h1* analyses were segregated from a cross to heterozygous *ddm1* plants. Even though the *ddm1* line had been extensively backcrossed to wild type, this line contained some TEs that had been demethylated when *ddm1* was homozygous, and which did not recover methylation in backcrossed progeny (Teixeira et al., 2009). These segregating demethylated TEs complicated the density analysis by mimicking sequences that are specifically differentially methylated in *h1* plants.

Expression analysis. To identify genes and TEs differentially expressed in the mutants relative to the wild-type control samples, RNA-seq datasets were mapped to the TAIR cDNA and TE annotations and analyzed using the Bioconductor package edgeR (Robinson et al., 2010). This package uses an empirical Bayesian approach based upon the negative binomial distribution to model digital expression data. Before comparing expression between genotypes, we first imposed an expression value threshold, excluding genes or TEs that did not have at least one read per million reads in at least two libraries. When all annotated genes were considered, 70% passed this expression threshold, whereas when highly methylated genes (CG methylation greater than 0.6) were excluded, 74% of the genes passed this expression threshold. A negative binomial was fitted to the data, using parameters estimated from the RNA-seq data. The dispersion of individual genes and TEs was estimated using the quantile-adjusted conditional maximum likelihood method and an empirical Bayes strategy. Genes and TEs differentially expressed between genotypes were determined using the exact test for the negative binomial distribution (which has strong parallels to Fisher's exact test). The resulting p-values were adjusted using Benjamini and Hochberg's approach (Benjamini and Hochberg, 1995) to control the false discovery rate to below 5%.

Supplemental References

Ascenzi, R., and Gantt, J.S. (1999). Subnuclear distribution of the entire complement of linker histone variants in Arabidopsis thaliana. Chromosoma *108*, 345-355.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J., *et al.* (2010). Relationship between nucleosome positioning and DNA methylation. Nature *466*, 388-392.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Teixeira, F.K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M.F., *et al.* (2009). A role for RNAi in the selective correction of DNA methylation defects. Science *323*, 1600-1604.

Tsugama, D., Liu, S., and Takano, T. (2011). A rapid chemical method for lysing Arabidopsis cells for protein analysis. Plant Methods *7*, 22.