Supporting Information

Zhang et al. 10.1073/pnas.1300585110

SI Materials and Methods

Whole Genome Bisulfite Sequencing and Data Analysis. DNA was extracted from 2 g of 14-d-old seedlings and sent to Beijing Genomics Institute (Shenzhen, China) for bisulfite treatment, library preparation, and sequencing. For data analysis, adapter, and low-quality sequences ($q < 20$) were trimmed and clean reads were mapped to *Arabidopsis thaliana* TAIR 10 (10th release of the *Arabidopsis thaliana* genome sequence from the Arabidopsis Information Resource) genome using Bisulfite Sequence Mapping Program allowing two mismatches. Identification of differentially methylated regions (DMRs) was conducted according to ref. 1 with modifications. In brief, only cytosines with a depth of at least four in all libraries were retained for further analysis. DNA methylation level in every 200-bp window with a step size of 50 bp was compared between wild-type and mutant plants using Fisher's exact test with a P value cutoff of 0.05. The P values were then adjusted using the Benjamini–Hochberg method to control for false discovery rates. Windows with five or more differentially methylated cytosines (DMCs), defined as C with $P < 0.01$ in Fisher's exact test) and ≥1.5-fold change in DNA methylation level were retained and combined if gap size is no more than 100 bp to generate DMRs. Finally the lengths of DMRs were adjusted to be from the first mC to the last mC.

Small RNA Sequencing and Data Analysis. RNAs were extracted from 2-wk-old seedlings using TRIzol reagent as previously described (2). Analysis of small RNA (sRNA) data were conducted according to reference (3) with minor modifications. Briefly, after adapter sequences were trimmed, clean reads of size 18 nt–30 nt were mapped to the TAIR 10 genome and annotated structural RNAs (rRNAs, tRNAs, snRNAs, and snoRNAs) using Bowtie (4). Only sRNAs that had at least one perfect match to the genome and did not match structural RNAs were used for downstream analysis. Read counts were normalized to reads per 10 million (RPTM) according to total number of clean reads (excluding structural sRNAs) in the library and number of hits to the genome. Normalized expression levels of 24-nt sRNAs were summarized in nonoverlapping 500-nt windows over the nuclear genome and were compared between each mutant and the wild type. To remove lowly expressed regions, only 500-nt regions that had at least 200 RPTM combined expression in the mutant and the wild type were considered. The 500-nt regions that showed fivefold lower expression in the mutant were selected as sRNA-depleted regions.

Histone Peptide Array and in Vitro Pull-Down Assay. Histone peptide array was performed following a published protocol (5). For in vitro pull down between GST-fused proteins and histone peptides, 1 μg of GST-tagged protein and 0.5 μg of biotin-labeled histone peptide were added to 200 μL of assay buffer (50 mM Tris·Cl, pH 7.5 at 25 °C, 150 mM NaCl, 0.05% Nonidet P-40) and incubated at 4 °C for 2 h, followed by addition of streptavidin beads and incubated for another hour. The beads were then washed three times with the assay buffer at 4 °C and the bound GST protein were extracted using SDS/PAGE sample buffer. Anti-GST Western blot was used to detect the presence of GST-fused proteins in the elution.

Affinity Purification and Mass Spectrometry. The DNA-BINDING TRANSCRIPTION FACTOR 1 (DTF1) genomic sequence was fused in frame to 3× Flag tag and the CLASSY (CLSY) genomic sequence was fused in frame to 3× Myc tag and the chimeric genes were separately constructed into the pCambia1305 backbone and transformed into *Arabidopsis* plants. Three grams of flower tissues were harvested from transgenic plants and ground in liquid nitrogen. The proteins were extracted in 15 mL of lysis buffer [50 mM Tris (pH 7.6), 150 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 0.1% Nonidet P-40, 0.5 mM DTT, 1 mM PMSF, and 1 protease inhibitor mixture tablet/50 mL (Roche)]. After the proteins were extracted, anti-Flag M1 (Sigma; A4596) or anti-Myc agarose (Sigma; A7470) were added and incubated at 4 °C for 2–3 h. Thereafter, the resins were washed with the lysis buffer at least five times. The anti-Flag beads-associated proteins were eluted with 3× Flag peptide (Sigma; F 4799). The anti-Myc beads-associated proteins were eluted with 0.1 M ammonium hydroxide at pH 11.5. The eluted proteins were separated on SDS/PAGE and visualized by silver staining with the ProteoSilver Silver Stain kit (Sigma; PROT-SIL1). The protein bands were cut from gels and the proteins eluted from the cut bands were subjected to mass spectrometry analysis as described previously (6).

1. Ausin I, et al. (2012) INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in Arabidopsis. *Proc Nat Acad Sci USA* 109(22):8374–8381.
2. Liu J, et al. (2011) An atypical component of RNA-directed DNA methylation machinery has both DNA methylation-dependent and -independent roles in locus-specific transcriptional gene silencing. *Cell Res* 21(12):1691–1700.
3. Lee TF, et al. (2012) RNA polymerase V-dependent small RNAs in Arabidopsis originate from small, intergenic loci including most SINE repeats. *Epigenetics* 7(7):781–795.
4. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
5. Qian W, et al. (2012) A histone acetyltransferase regulates active DNA demethylation in Arabidopsis. *Science* 336(6087):1445–1448.
6. Zhang C-J, et al. (2012) IDN2 and its paralogs form a complex required for RNA-directed DNA methylation. *PLoS Genet* 8(5):e1002693.

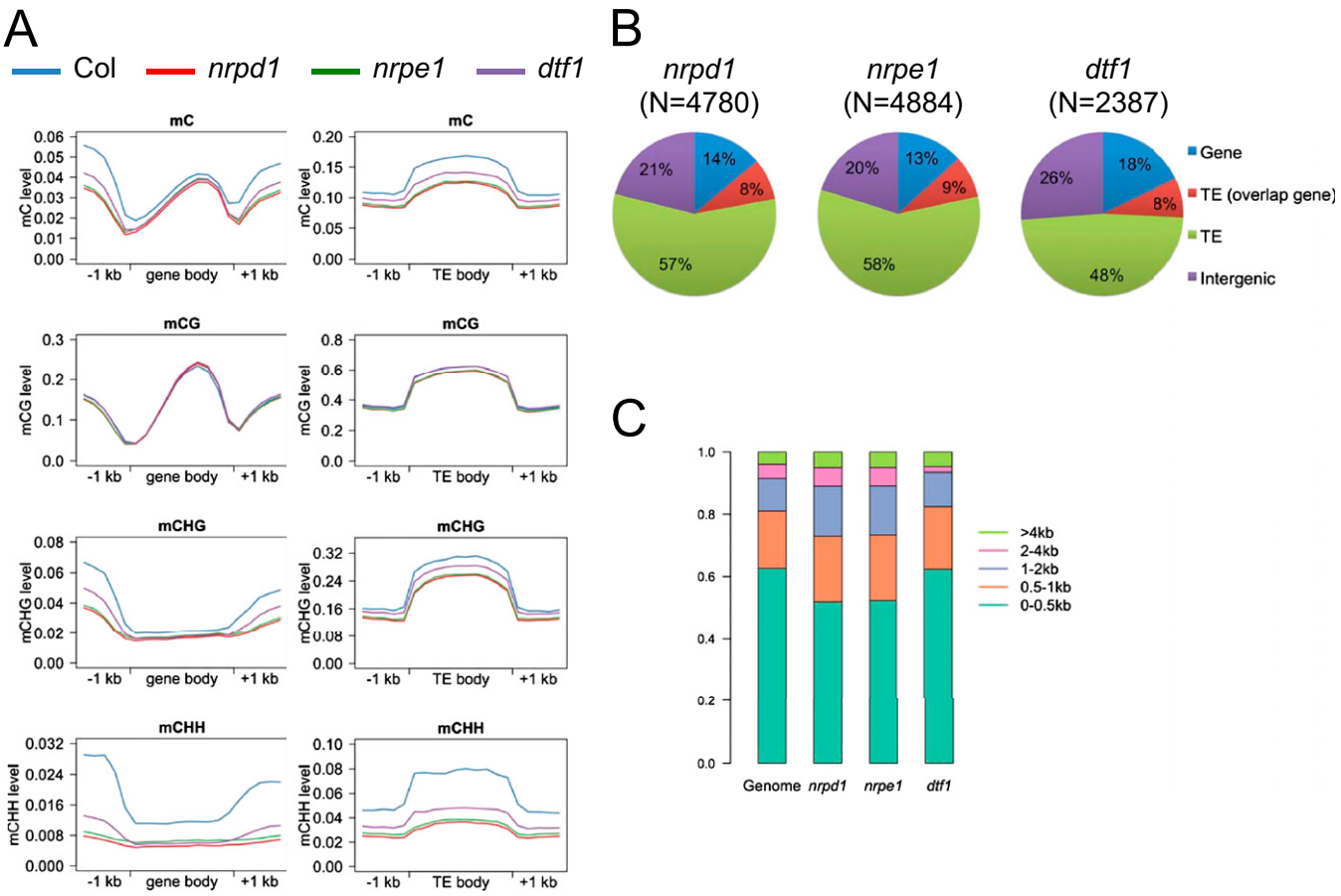

Fig. S1. Effects of *dtf1* on gene and TE methylation and characterization of DTF1-dependent hypo DMRs. (*A*) Average methylation levels in gene and TE bodies. (*B*) Distribution of DMRs among TEs, TEs that overlap with genes, genes, and intergenic regions. (*C*) Percentage of hypo DMRs that overlap with TEs in different size groups.

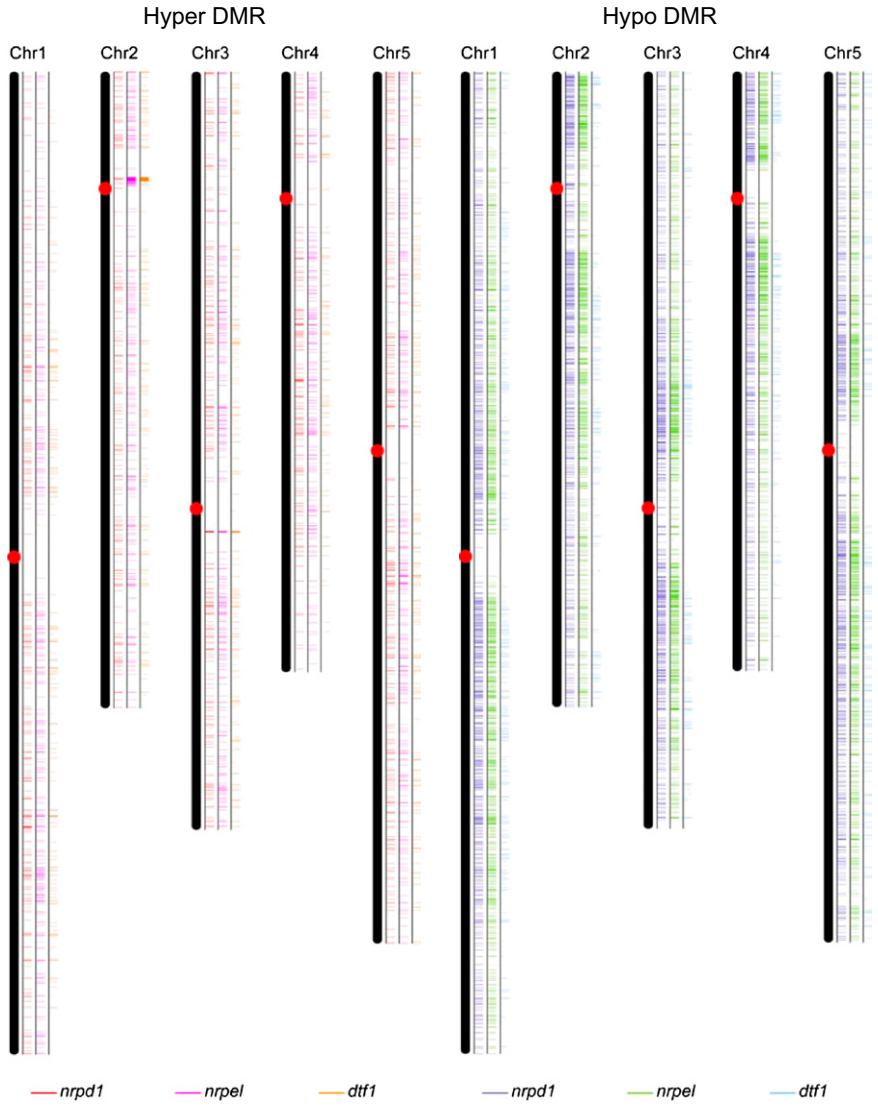

Fig. S2. Distribution of hypo and hyper DMRs in *Arabidopsis* chromosomes 1–5. Centromeres are indicated by red dots.

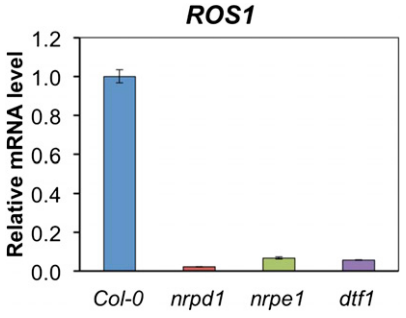

Fig. S3. *ROS1* transcript level as measured by real-time PCR. ACT7 was used as the internal control and transcript level was normalized relative to the wild-type Col-0.

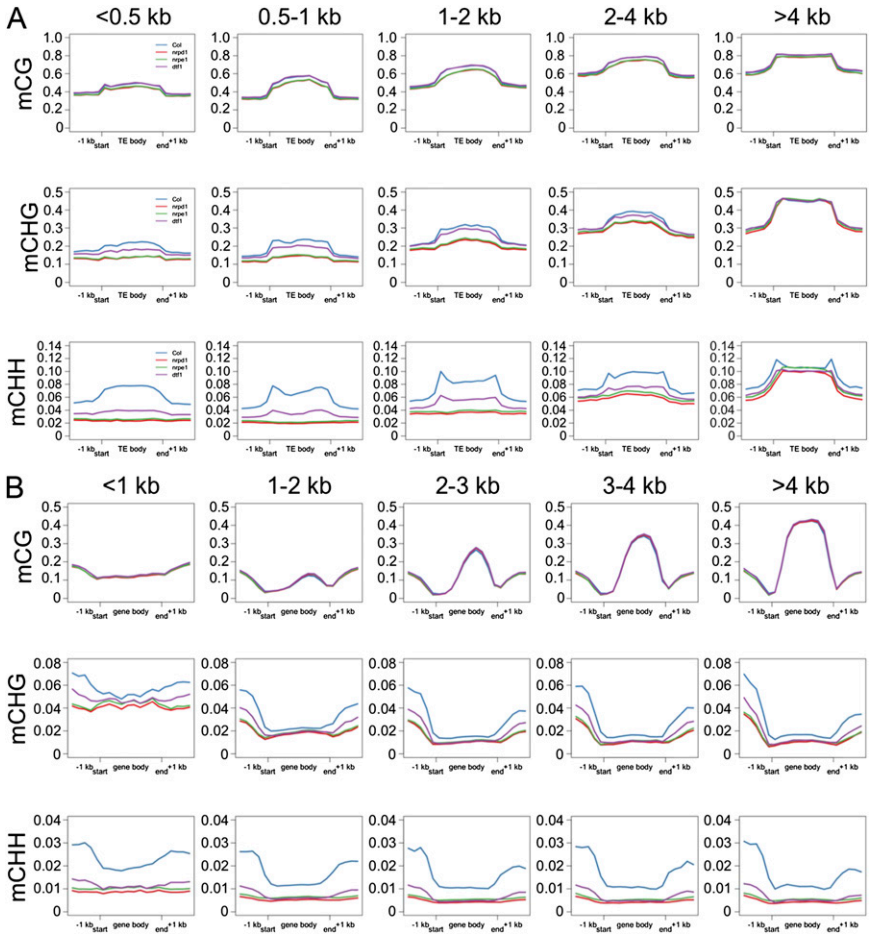

Fig. S4. Average DNA methylation levels in the TE (*A*) and gene (*B*) bodies. Each gene or TE was aligned from start to end and divided into 10 equal parts. DNA methylation level was calculated for each part and then averaged over all of the TEs or genes. Only cytosines with four or more times coverage were used for this analysis.

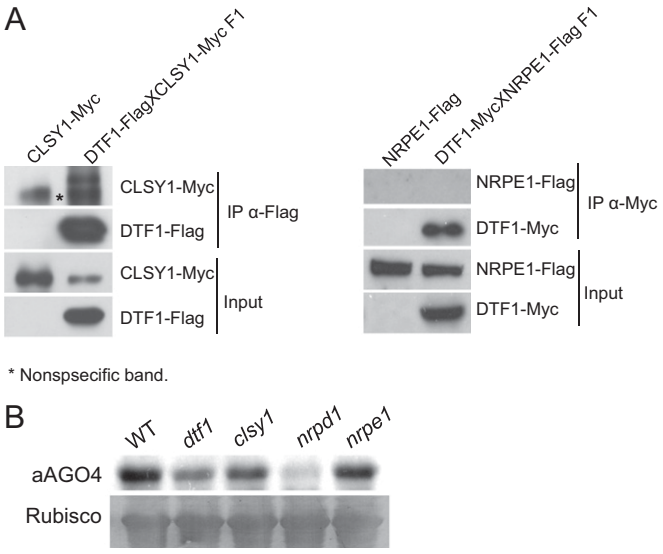

Fig. S5. DTF1 associates with Pol IV but not Pol V in vivo. (*A*) DTF1 interacts with CLSY1 but not NRPE1 in vivo. Coimmunoprecipitation assay between epitope-tagged DTF1 and CLSY1 or NRPE1. (*B*) The AGO4 protein level is reduced in *dtf1*. Anti-AGO4 Western blot was performed using total proteins extracted from the indicated genotypes. Ponceau S stained Rubisco largest subunit band was used as loading control.

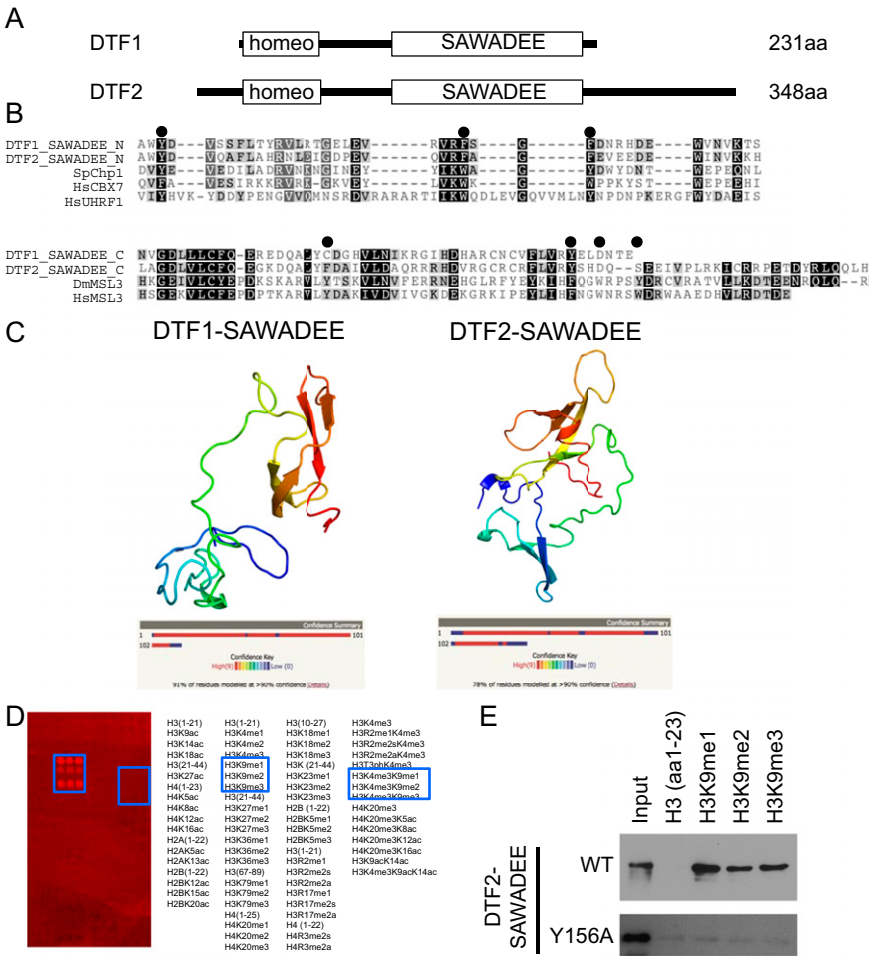

Fig. S6. Domain analyses for the DTF1/2 proteins. (*A*) Diagram shows the domain structures of DTF1 and DTF2. Domains were drawn relative to their position and length in the protein sequence. (*B*) Protein sequence alignment that shows the N- and C-terminal halves of the SAWADEE domain that aligns with proteins used as templates for structure prediction by the Phyre server. Black dots indicate aromatic residues that form binding pockets for methylated lysine/arginine in known structures. (*C*) 3D structures of SAWADEE domains predicted by the Phyre server. Confidence score for the prediction is indicated below the structure. (*D*) Histone peptide array for the GST–SAWADEE domain of DTF2. Each specific modified histone peptide is spotted in triplicate. Spots that contain H3K9 methylation are indicated by blue square boxes. (*E*) DTF2 SAWADEE domain specifically binds to H3K9 methylation. Pull-down assay was performed using GST-fused DTF2 SAWADEE domain and biotinylated histone H3 peptide. Results were visualized using anti-GST Western blot.

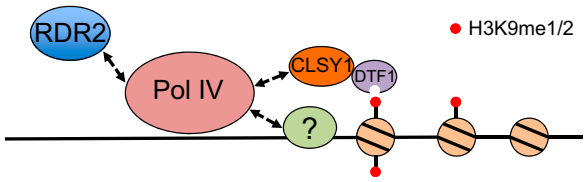

Fig. S7. Working model for DTF1 function in RdDM. Recruitment of Pol IV to many RdDM target loci may involve some unknown protein(s) (green) that bind DNA and DTF1 that binds to H3K9me1/2 and physically interacts with CLSY1. CLSY1 may help DTF1 binding to H3K9me1/2 nucleosomes as well as assist in Pol IV transcription of the chromatin DNA. CLSY1 and RDR2 also associate with Pol IV, although it is unclear whether the interaction is direct, as indicated by the dashed arrows in the model.

Dataset S1. Bisulfite sequencing statistics

Dataset S1

Dataset S2. List of DMR regions identified in *nrpd1*, *nrpe1*, and *dtf1*

Dataset S2

Dataset S3. Size distribution of small RNAs

Dataset S3

Dataset S4. List of differentially expressed siRNA clusters identified in *nrpd1*, *nrpe1*, and *dtf1*

Dataset S4

Dataset S5. Primer list

Dataset S5