CORRECTED 16 June 2021

science.sciencemag.org/cgi/content/full/science.abg6130/DC1

Supplementary Materials for

MBD5 and MBD6 couple DNA methylation to gene silencing through the J-domain protein SILENZIO

Lucia Ichino, Brandon A. Boone, Luke Strauskulage, C. Jake Harris, Gundeep Kaur, Matthew A. Gladstone, Maverick Tan, Suhua Feng, Yasaman Jami-Alahmadi, Sascha H. Duttke, James A. Wohlschlegel, Xiaodong Cheng, Sy Redding, Steven E. Jacobsen*

*Corresponding author. Email: jacobsen@ucla.edu

Published 3 June 2021 on *Science* First Release
DOI: 10.1126/science.abg6130

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at science.sciencemag.org/cgi/content/full/science.abg6130/DC1)

Table S1 (Excel)
MDAR Reproducibility Checklist (PDF)

CORRECTION: Two minor issues were found in the original supplementary materials PDF. In Fig. S6B, the title of the plot was illegible, probably because of an issue that occurred during the conversion from .docx to .pdf, and Fig. S6A was missing a *y*-axis label. These issues have now been fixed. Also, all of the figures have been replaced with higher-resolution versions.

Materials and Methods

Plant materials and growth conditions

All plants used in this study were in the Columbia-0 ecotype (Col-0) and were grown on soil in a greenhouse under long-day conditions (16h light / 8h dark).

The following mutant lines were purchased from Arabidopsis Biological Resource Center (ABRC): *mbd5* (SAILseq_750_A09.1) (*13*), *mbd6* (SALK_043927), *sln* (gene id: AT5G37380 (*26*, *27*, *33*), mutant line: SALK_090484), *met1-3* (CS16394). The *mbd5 mbd6* T-DNA double mutant was generated by crossing *mbd5* (SAILseq_750_A09.1) to *mbd6* (SALK_043927). The *fwa* epiallele was previously generated from a *met1-3* segregating population (*fwa-4*) (*29*).

Generation of CRISPR lines

The *mbd5 mbd6* CRISPR line was generated using the previously published pYAO::hSpCas9 system (*34*). A CRISPR guide for the MBD5 gene (ACCGGAGAACCCGGCTACTC) was cloned in the AtU6-26-sgRNA cassette by overlapping PCR with primer tails containing the guide sequence. The PCR product was cloned into the SpeI site of the pYAO::hSpCas9 destination plasmid by In-Fusion (Takara, 639650). The final vector was electroporated into AGLO agrobacteria and transformed in *mbd6* mutant plants by agrobacterium-mediated floral dipping. T1 plants were selected on ½ MS agar plates with hygromycin B and were genotyped by sanger sequencing of the PCR amplified genomic region surrounding the MBD5 guide target. Several lines containing mutations around the PAM site were identified. We selected a few lines and grew the T2 progeny on soil. The T2 plants were first genotyped to find the ones that had segregated out the pYAO::hSpCas9 transgene. The Cas9 negative plants were genotyped by sanger sequencing of the MBD5 gene, and one plant was found containing a single base insertion (A) 3bp before the PAM site in one allele. The progeny of this plant (T3 population) was grown on a hygromycin plate to confirm the absence of the pYAO::hSpCas9 transgene and genotyped to identify homozygous *mbd5* mutant plants. The T4 progeny of an *mbd5* homozygous mutant plant (containing homozygous T-DNA mutations in *MBD6*) was used for RNA-seq.

Cloning and generation of transgenic lines

The MBD5, MBD6, and SLN -FLAG lines used for ChIP-seq and IP-MS were generated by cloning the genes from genomic DNA into pENTR/D-TOPO vectors (Thermo Fisher), including their endogenous promoter and introns, until the last base before the STOP codon. The MBD5 gene was cloned starting from 1094 bp before the TSS, MBD6 from 294 bp before the TSS and SLN from 2351 bp before the TSS. The point mutations in MBD5 (R1=R46A and R2=R69A), MBD6 (R1=R92A and R2=R115A), and SLN (H94Q) were generated by amplifying the genes in two or three overlapping PCR products containing the mutations in the primer sequence and cloning them in a PCR amplified linear pENTR vector by In-Fusion (Takara, 639650). All constructs were verified by sanger sequencing. The genes were then transferred via a Gateway LR Clonase II reaction (Invitrogen, 11791020) into a pEG302 based binary destination vector including a C-terminal 3xFLAG epitope tag. The binary vectors were electroporated into AGL0 agrobacteria that were used for plant transformation by agrobacterium-mediated floral dipping. The FLAG-tagged transgenes were transformed in their respective T-DNA mutant backgrounds, or in *mbd5 mbd6* T-DNA or CRISPR double mutant plants for the complementation RT-qPCR shown in figure S10. T1 transgenic plants were selected with hygromycin B on ½ MS agar medium or with Basta (Glufosinate) directly on soil.

For the ChIP-seq experiment shown in Figure 3 and S14 and IP-MS experiment in Figure S17, single insertion homozygous T3 lines of the FLAG-tagged genes (MBD5, MBD6, or SLN) in their respective mutant backgrounds were crossed to either *sln* or *mbd5 mbd6* T-DNA mutant plants. The F2 populations were genotyped to identify plants that inherited two wild-type copies or two mutant copies of the indicated gene(s), and one copy of the FLAG transgene. The progeny of the selected plants (F3 generation) was used for the ChIP-seq or IP-MS experiment after verifying by western blot that transgene silencing had not occurred in any of the lines.

The *pUBQ10::ZF108-SLN* lines were generated by cloning into pENTR vectors the *SLN* open reading frame (from the ATG until the STOP codon) using a cDNA library as PCR template. The *SLN* coding sequence was transferred into a pMDC123 destination vector encoding an N-terminal ZF108-3xFLAG tag driven by the *UBQ10* promoter (*30*). The construct was transformed into *fwa* epiallele plants as described above. Flowering time was scored by counting the total number of rosette and cauline leaves in segregating T2 populations. The investigator who counted the leaves was blinded to the genotyping that was performed to identify the null segregants. Col0 and *fwa* control plants were grown side by side.

Protein purification

MBD5 and MBD6 proteins were expressed in BL21 (DE3) *E. coli*. from a pET28a plasmid containing an N-terminal MBP tag and a C-terminal 6xHIS Tag. Overnight seed cultures were inoculated into 0.5 or 1 liters of LB media and grown at 37°C to an optical density of ~0.6. Protein expression was induced with 1mM IPTG and grown overnight (>12 hours) at 16°C. Pellets were collected by centrifugation at 1200 g, resuspended in 25 mL of lysis buffer (25 mM Tris pH 7.5, 150 mM NaCl, 1% Triton X-100, 5 mg lysozyme, 50 mM Imidazole, and 1 Roche Protease Inhibitor Cocktail tablet), and lysed for ~45 minutes on ice. Lysates were sonicated for intervals of 5 seconds on and 5 seconds off until the lysate was homogenous. Lysates were cleared by centrifugation at 30,000g for 20 minutes at 4°C. The cleared lysates were incubated with Ni-NTA resin (Thermo Scientific, 88222) for 1 hour at 4°C. The resin was collected by centrifugation at 200g and washed with 3x5 column volumes (CVs) of wash buffer (25 mM Tris pH 7.5, 150 mM NaCl, 75 mM imidazole). Bound protein was eluted with 5 CVs of elution buffer (25 mM Tris pH 7.5, 150 mM NaCl, 300 mM Imidazole), and dialyzed overnight at 4°C into protein buffer (25 mM Tris pH 7.5, 150 mM NaCl) to remove imidazole. Proteins were then concentrated using spin columns (Millipore Sigma, UFC903024) prior to use in binding and single molecule assays.

Fluorescence Polarization (FP) Assay

The FP assays were performed as previously described with some variations (*9*). The following oligonucleotides were obtained from IDT:

Methylation Context	Direction	Sequence
mCG	Sense	FAM-TTCTTG/**iMe-dC**/**G**ATTCAACATTCAGCTCTT/**iMe-dC**/**G**ATTCAACATTCAGCTCTT/**iMe-dC**/**G**ATTCAACATTCAGCTCTT/**iMe-dC**/**G**TT
	Antisense	AA/**iMe-dC**/**G**AAGAGCTGAATGTTGAAT/**iMe-dC**/**G**AAGAGCTGAATGTTGAAT/**iMe-dC**/**G**AAGAGCTGAATGTTGAAT/**iMe-dC**/**G**CAAGAA

uCG	Sense	FAM-TTCTTG**CG**ATTCAACATTCAGCTCTT**CG**ATTCAACATTCAGCTCTT**CG**ATTCAACATTCAGCTCTT**CG**TT
	Antisense	AA**CG**AAGAGCTGAATGTTGAAT**CG**AAGAGCTGAATGTTGAAT**CG**AAGAGCTGAATGTTGAAT**CG**AAGAGCTGAATGTTGAAT**CG**CAAGAA
mCHG	Sense	FAM-TTCTTG/**iMe-dC**/**AG**TTCAACATTCAGCTCTT/**iMe-dC**/**TG**ATCAACATTCAGCTCTT/**iMe-dC**/**CG**ATTAACATTCAGCTCTT/**iMe-dC**/**AG**ATTCAACATTCACTCTT/**iMe-dC**/**TG**T
	Antisense	A/**iMe-dC**/**AG**AAGAGTGAATGTTGAAT/**iMe-dC**/**TG**AAGAGCTGAATGTTAAT/**iMe-dC**/**GG**AAGAGCTGAATGTTGAT/**iMe-dC**/**AG**AAGAGCTGAATGTTGAA/**iMe-dC**/**TG**CAAGAA
uCHG	Sense	FAM-TTCTTG**CAG**TTCAACATTCAGCTCTT**CTG**ATCAACATTCAGCTCTT**CCG**ATTAACATTCAGCTCTT**CAG**ATTCAACATTCACTCTT**CTG**T
	Antisense	A**CAG**AAGAGTGAATGTTGAAT**CTG**AAGAGCTGAATGTTAAT**CGG**AAGAGCTGAATGTTGAT**CAG**AAGAGCTGAATGTTGAA**CTG**CAAGAA
mCHH	Sense	FAM-TTCTTG/**iMe-dC**/**AA**ATTCAACATTCACTCTG/**iMe-dC**/**TA**ATCAACATTCAGCTCTG/**iMe-dC**/**AT**GTCAACATTCAGCTCTG/**iMe-dC**/**TT**ACAACATTGGTTCGTCTG/**iMe-dC**/**CA**
	Antisense	TGG/**iMe-dC**/AGACGAACCAATGTTGTAAG/**iMe-dC**/AGAGCTGAATGTTGACATG/**iMe-dC**/AGAGCTGAATGTTGATTAG/**iMe-dC**/AGAGTGAATGTTGAATTTG/**iMe-dC**/AAGAA
uCHH	Sense	FAM-TTCTTG**CAA**ATTCAACATTCACTCTG**CTA**ATCAACATTCAGCTCTG**CAT**GTCAACATTCAGCTCTG**CTT**ACAACATTGGTTCGTCTG**CCA**
	Antisense	TGGCAGACGAACCAATGTTGTAAGCAGAGCTGAATGTTGACATGCAGAGCTGAATGTTGATTAGCAGAGTGAATGTTGAATTTGCAAGAA

The sense and antisense oligonucleotides were resuspended in water to a final concentration of 100 µM. Annealing reactions were created by mixing 10 µL of sense oligonucleotide, 10 µL of antisense oligonucleotide, 4 µL of 10x DNA annealing buffer (100mM Tris-HCL pH 7.5, 1M NaCl, 10mM EDTA), and 16 µL of water. DNA annealing was accomplished by heating the annealing reactions to 95°C for 5 minutes using a PCR machine and then decreasing the temperature by 1°C per minute until 25°C was reached. The samples were run on a 1.5 % agarose gel to verify the annealing.

For FP assays, proteins were serially diluted 1.8 fold in protein buffer with a final 10 nM concentration of annealed, FAM-labelled DNA, in 384-well low binding capacity plates

(Corning, 3575). Final NaCl concentration was 150 mM for MBD6 binding assays and 75 mM for MBD5 binding assays. The protein-DNA mixtures were incubated on ice for 30 minutes. Anisotropy measurements were obtained from polarization readings measured at an excitation wavelength of 485 nM and an emission wavelength of 535 nM, using a Tecan microplate reader. Anisotropy values measured from wells containing only 10 nM DNA (DNA only controls) were subtracted from anisotropy values measured from wells containing protein bound samples to account for background anisotropy signal. Scatter plots of anisotropy versus protein concentration were plotted using GraphPad Prism software and binding models were fit to the data using non-linear regressions of specific binding with a Hill slope. To create fraction bound plots for datasets containing saturated binding of DNA (plateau is reached), all anisotropy values were divided by the highest binding maximum (Bmax) anisotropy value of either methylated or unmethylated DNA calculated from non-linear regression models. If neither methylated or unmethylated DNA curves reached plateaus, the anisotropy values were divided by the maximum anisotropy value measured with either methylated or unmethylated DNA within the dataset, to create fraction bound values. Final scatterplots of fraction bound versus protein concentration were then created in GraphPad Prism software to estimate equilibrium dissociation constants (Kd) of MBD proteins for DNA methylation contexts. Binding curves without saturated binding could not be modeled and were labeled as non-determinable (N.D.).

Labeling Proteins with Fluorescent Dyes

MBP-MBD6-6xHIS protein was labeled with Cy3-NHS-Ester fluorescent dyes by mixing 1:1 molar ratio of dye to protein in protein buffer (25 mM Tris pH 7.5, and 150 mM NaCl) overnight (>12 hours) at 4°C. To remove excess dye samples were dialyzed in protein buffer for at least 4 hours and then concentrated using spin columns (Millipore Sigma, UFC903024). Flowthrough from the concentration was checked for Cy3 dye to confirm no presence of excess dye.

DNA curtains assay

All curtains experiments were performed using DNA from bacteriophage λ (λ-DNA) (NEB Catalog # N3011). For methylated DNA curtains, λ-DNA was methylated at all CpGs using M.SssI (NEB Catalog # M0226). The reaction was set up according to NEB's protocol for methylation of genomic DNA with their suggested adjustments to push methylation to completion (4 hours at 37°C). Afterwards, the reaction was moved to 65ºC to heat-inactivate the methyltransferase. During this incubation, an excess of 12nt DNA oligomers (5'-AGGTCGCCGCCC-3' and 5'-GGGCGGCGACCT-3') were added to prevent the single-stranded cohesive ends of λ-DNA from recircularizing. The oligo targeting the 5'-end of λ-DNA contains a 3'-biotin to enable tethering of the DNA during curtain experiments. The entire reaction was cooled down slowly to room temperature before adding T4 DNA ligase (NEB Catalog # M0202) and incubating overnight at room temperature for ligation. The biotinylated λ-DNA was then purified using size exclusion chromatography. The entire reaction was run through sephacryl S-1000 SF resin (GE Product # 17047601) equilibrated in TE150 (10mM Tris-HCl, pH=7.4, 1mM EDTA, and 150mM NaCl). Fractions were collected and stored at 4ºC. Full methylation of all CpGs was verified using NEBNext Enzymatic Methyl-seq (EM-seq) (NEB Catalog # E7120), following manufacturer's instructions.

DNA curtains were performed as previously described with minimal adjustments (*35*). Briefly, chrome diffusion barriers were nanofabricated on quartz microscope slides with holes

drilled through them to enable use of microfluidics. Flowcells were created on nanofabricated slides by creating a small chamber using double-sided tape and coverslips, which were baked on using a vacuum oven. On the back side of the slide, NanoPorts (IDEX Part # N-333) were hot glued over the holes to attach tubing for microfluidics. To begin an experiment, a flowcell was equilibrated in lipids buffer (10mM Tris-HCl, pH 7.5 and 100mM NaCl). Then a lipid bilayer composed of DOPC, PEG-2000 DOPE, and biotinylated DOPE lipids (Avanti Polar Lipids 850375, 880130, and 870273) was deposited on the surface of the slide. Streptavidin, diluted in blocking buffer (0.5mg/mL BSA, 0.5% Pluronic F-127, 25mM Tris-HCl pH 7.5, 150mM NaCl, 1mM MgCl, and 1mM DTT), was injected into the flowcell to functionalize the biotinylated lipid heads. λ-DNA molecules were tethered to individual lipid heads via a biotin-streptavidin-biotin linkage and visualized using imaging buffer (blocking buffer supplemented with 10pM YOYO-1, ThermoFisher Y3601). Buffer flow was used to align λ-DNA molecules at diffusion barriers and keep DNA in an extended conformation while imaging. MBD6-Cy3 was diluted to 10nM in imaging buffer, injected into the flowcell and allowed to bind to the DNA under constant buffer flow.

Images were captured using a custom built, prism-type total internal reflection fluorescence (TIRF) microscope with an Andor EMCCD camera. Analysis was performed in Fiji and Python using code explained here (https://github.com/ReddingLab/Ichino_et_al_Science2021). Briefly, protein and DNA channels were aligned using offsets determined by the differential position of TetraSpeck Microspheres (0.1um, Invitrogen, REF T-7279) in each channel. Next, individual DNA molecules were identified in an image based on YOYO-1 signal, which was used to determine the beginning, center, and end of individual DNAs. These bounds were used to identify MBD6-Cy3 binding events on DNA. Two-dimensional gaussians were fit to each binding event to determine its sub-pixel position. The relative base pair position of each binding event was estimated based on its distance from the end of the DNA using the average extension of the DNA molecule. 1000bp bins were used to plot the distribution of MBD6-Cy3 binding events along with the distribution of mCG density on DNA curtains. The mCG density was calculated as the sum of methylation percentages per 1000bp bin, based on the results of the EM-seq.

Three technical replicates were performed under all conditions. Error bars were calculated by bootstrap analysis. Bootstrapping was done by generating 300 sample data sets with the same depth as the measured data (n=15811) to calculate 95% confidence intervals. The correlation between MBD6-Cy3 binding and mCG density was determined using Pearson's correlation coefficient.

Structural modeling

The amino acid sequences of known *Arabidopsis thaliana* MBD proteins (MBD1 to MBD13) were retrieved from UNIPROT database (*36*). The closest structural homolog for each MBD protein in Protein Data Bank (www.rcsb.org/pdb) was identified using the BLAST search. The templates were selected for each MBD protein and the homology models were built using the Phyre3D webserver (*36*). The built homology models were analyzed for their confidence score. The models with high confidence scores were selected and further docked onto the available crystal structure(s) of MBD-DNA complexes (retrieved from the PDB). The residues of the MBD protein potentially involved in the recognition of the methylated CpG base pairs were identified. The structure figures were generated using PyMol (Delano Scientific, LLC.). Based on the structural models, all amino acid sequences of MBD proteins were aligned manually.

Based on the presence of the conserved amino acid residues (R-D and R-E pairs), *A. thaliana* MBD proteins were further sub-classified in four different categories: Group-I (MBD2, 4, 5, 6, 7 and 12), Group -II (MBD1 and 3), Group-III (MBD8, 10, 11 and 13) and Group-IV (MBD9).

Phylogenetic analysis

To identify the SLN orthologs we performed BLASTp (*37*) searches with standard parameters, using the SLN full-length sequence as query, against the proteomes of eight species representative of monocots, dicots, mosses and ferns (*Glycine max, Solanum lycopersicum, Zea mays, Triticum aestivum, Oryza sativa, Marchantia polymorpha, Physcomitrium patens, Adiantum capillus-veneris*). For each species we picked the top hit based on E-value and we retrieved the corresponding full-length protein sequence from National Center for Biotechnology Information (NCBI) GenBank (*38*). We then performed a multiple sequence alignment using Clustal Omega with standard parameters (*39*). The figure displays the region of the alignment that covers the entire length of SLN.

When we performed BLASTp searches of SLN excluding the plant kingdom we were only able to detect hits corresponding to the J-domain, while no significant similarity was found with the rest of the protein. Therefore, we concluded that SLN doesn't have any clear orthologs outside of plants.

RT-qPCR

All RT-qPCR experiments were performed on unopened floral bud tissue, except for the *pUBQ10::ZF108-SLN* experiments, which were done on leaf tissue from 3 to 4 week-old plants. The MBD5-FLAG complementation experiment was performed in the T2 generation while the MBD6-FLAG complementation experiment was done in T1 (Figure S10).

RNA was extracted using the Zymo Direct-zol RNA MiniPrep kit (Zymo Research, R2052), with in column DNAse I treatment. Approximately 400 ng of total RNA was reverse transcribed into cDNA with Superscript III First Strand Synthesis Supermix (Invitrogen, 18080400), using random examers. The qPCR was performed with iQ SYBR Green Supermix (Bio-Rad, 1708882) with the Agilent Technologies Mx3005p qPCR System (Stratagene). 0.5 µl of cDNA were used for each 20 µl reaction with technical duplicates or triplicates for each primer pair. The housekeeping gene ISOPENTENYL PYROPHOSPHATE DIMETHYLALLYL PYROPHOSPHATE ISOMERASE 2 (IPP2) was used as control. The data was analyzed with the DDCt method, comparing to the average of the reference genotype (indicated in the y-axis of each graph or in the legend). The statistical analysis was performed in RStudio using the function *stat_compare_means* from the *ggpubr* R package (two-sided, unpaired, two sample t-test).

List of primers used for RT-qPCR:

Primer name	Sequence
FWA RT-qPCR Fw	TTAGATCCAAAGGAGTATCAAAG
FWA RT-qPCR Rev	CTTTGGTACCAGCGGAGA
IPP2 RT-qPCR Fw	GTATGAGTTGCTTCTCCAGCAAAG
IPP2 RT-qPCR Rev	GAGGATGGCTGCAACAAGTGT

McrBC assay

DNA was extracted from leaf tissue using the cetyl trimethylammonium bromide (CTAB) method and treated with RNase A (Qiagen). 250 ng of DNA were digested with the McrBC restriction enzyme (NEB, M0272) for 4 h at 37 °C in 10 µl of reaction volume, following the manufacturer's instructions. At the same time, an equal amount of DNA was incubated in the digestion buffer without enzyme (mock control). 1 µl of digestion mix or mock sample was used as template for RT-qPCR. The experiment was conducted with technical triplicates. The data was analyzed with the DCt method (average of the McrBC digested Ct values minus the average of the mock digested Ct values). The statistical analysis was performed in RStudio using the function *stat_compare_means* from the *ggpubr* R package (two-sided, unpaired, two sample t-test).

List of primers used for McrBC assay:

Primer name	Sequence
FWA McrBC Fw	TTGGGTTTAGTGTTTACTTG
FWA McrBC Rev	GAATGTTGAATGGGATAAGGTA

Western blot

For the western blots of the ChIP-seq samples, 20 µl of ChIP input chromatin were mixed to an equal volume of 2X SDS buffer and boiled for 5 minutes at 95°C. In all other cases, a leaf punch was harvested from individual plants and immediately frozen in liquid nitrogen. The tissue was ground, resuspended in 2X SDS buffer, and boiled for 5 minutes at 95°C. The protein samples were loaded on a 4 to 12% Bis-Tris gel (NuPage, Invitrogen WG1402BX10) and ran in MOPS buffer. The gels were blotted on PVDF membranes with the iBlot 2 Dry Blotting System (Thermo Fisher Scientific, 88518), blocked with 5% milk in TBS-T (Tris-Buffered Saline with 0.1% Tween20) and incubated with HRP conjugated anti-FLAG antibody (Sigma-Aldrich, A8592). The H3 loading control was detected using an anti-H3 antibody (Abcam, ab1791) and a secondary anti-Rabbit HRP (Thermo Fisher Scientific, 31460). The blot was imaged with ECL method. In the case of Figure S15A, the blot was simultaneously incubated with anti-H3 (Abcam, ab1791) and Anti-FLAG M2 (Sigma, F1804) primary antibodies, and then with IRDye 800CW and 680RD secondary antibodies (LI-COR). The blot was imaged on a LI-COR Odyssey CLx instrument.

Immunoprecipitation Mass Spectrometry (IP-MS)

The IP-MS experiments were performed as previously described (*40*) with minor variations. We used either T2 or T3 plants of the indicated FLAG-tagged lines and Col0 controls grown side by side. The MBD5 and MBD6 IP in the *sln* mutant background was done in F3 as explained in the paragraph "Cloning and generation of transgenic lines". For each sample, 8 to 10 g of inflorescences were harvested and immediately frozen in liquid nitrogen. The tissue was ground in liquid nitrogen and resuspended in 25 mL ice-cold IP buffer (50 mM Tris·HCl pH 8.0, 150 mM NaCl, 5 mM EDTA, 10% glycerol, 0.1% Tergitol, 0.5 mM DTT, and cOmplete EDTA-free Protease Inhibitor Cocktail [Roche]). For the SLN IP, we used an IP buffer containing 20% glycerol instead of 10%, because we noticed improved detection of interactors. The samples were filtered through one layer of miracloth and further disrupted with a Dounce homogenizer,

before centrifugation for 10 min at 4°C at 20,000 g. In the second experiments of SLN and SLN^{H94Q} IP (see Table S1), we included here a step of benzonase treatment to digest the DNA and avoid co-precipitation of proteins because of binding to the same chromatin fragments. We added 25 µl of benzonase (EMD Millipore 700664-3) to each supernatant and incubated them for 30 minutes at room temperature. The treatment eliminated many chromatin contaminants like histones, but it did not change the conclusions made in Figure 4.

After the centrifugation (or the benzonase treatment), the supernatant was incubated with 200 µL of M2 magnetic FLAG-beads (SIGMA, M8823) for 2 hours rotating at 4°C. The beads were washed five times in ice-cold IP buffer for 5 min rotating at 4°C, and immunoprecipitated proteins were eluted two times with 300 µl of 250 µg/mL 3X-FLAG peptides (SIGMA, F4799) for 15 min at 37°C in a thermomixer (1000 rpm). The eluted protein complexes were precipitated overnight with 20% trichloroacetic acid (TCA), the pellets were washed two times with ice-cold acetone and dried.

The protein pellets were resuspended in 50 µl of digestion buffer composed of 8 M urea in 100 mM Tris at pH of 8.5. The samples were reduced, alkylated, using 5 mM Tris (2-carboxyethyl) phosphine and 10 mM iodoacetamide, respectively, and digested after addition of lys-C and trypsin proteinases at 37°C. The digested samples were quenched by the addition of formic acid to 5% (v./v.) final concentration. Finally, each sample was desalted via C18 tips (Thermo Scientific, 87784) and reconstituted in 5% formic acid before analyzed by LC-MS/MS.

Tryptic peptide mixtures were fractionated online using a fused silica with 75 µm inner diameter and 25 cm long, packed in-house with bulk C18 reversed phase resin (particle size, 1.9 µm; pore size, 100 Å; Dr. Maisch GmbH). The peptides were separated using a Dionex Ultimate 3000 UHPLC system (Thermo Fisher Scientific). The 140-min water–acetonitrile gradient was delivered at a flow rate of 300 nL/min. Peptides were ionized, entered the Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific), and analyzed by tandem mass spectrometry

Data was acquired using a Data-Dependent Acquisition (DDA) method at an MS1 resolution of 120,000, followed by sequential MS2 scans at resolution of 15,000 to utilize the remainder of the 3 second cycle time.

Data analysis including peptide and protein identification was performed using the ProLuCID search against TAIR Arabidopsis reference proteome followed by filtering of peptide-to-spectrum matches (PSMs) by DTASelect using spectrum level FDR of less than 1%.

The heatmaps were generated with the *pheatmap* R package.

Bisulfite PCR (BS-PCR)

BS-PCR experiments were performed as previously described (*30*). DNA was extracted from leaf tissue using the cetyl trimethylammonium bromide (CTAB) method and treated with RNase A (Qiagen). The DNA was bisulfite treated using the Epitect Bisulfite Conversion kit (QIAGEN). The conversion reaction in the thermocycler was performed twice to increase the efficiency. The converted DNA was amplified with the primers indicated below, using the Pfu Turbo Cx DNA polymerase (Agilent). Different PCR products for the same sample were pooled and purified with AMPure beads (Beckman Coulter). Libraries were generated using the Kapa HyperPrep DNA kit (KR0961) and the TruSeq DNA UD adapters (IDT for Illumina).

List of primers used for BS-PCR:

Primer name	Sequence
FWA region 1 BS-PCR Fw	TCATATAAAAAAAAAATTAAATTTCATTTCACAATAACCATT
FWA region 1 BS-PCR Rev	GTATGGGYTTYGATAAAGAATATATGAGATTYT
FWA region 2 BS-PCR Fw	CTCATATATACCTTATCCCATTCAACATTCATA
FWA region 2 BS-PCR Rev	AAGATYTGATATTTGGYTGGAAAAAAYAATAATAAT
FWA region 3 BS-PCR Fw	CRCTCTTTATCCCATTCAACATTCATAC
FWA region 3 BS-PCR Rev	TTTGGTTGAAAAAAATAATAAAAATTTGATTGTYAGTAT

poly(A) and total RNA sequencing (RNA-seq)

For most of the RNA-seq experiments, RNA was extracted from unopened floral bud tissue harvested from an individual plant for each sample. In the case of the *pUBQ10::ZF108-SLN* RNA-seq experiment, leaf tissue from four primary T1 transformants and from four control *fwa* plants were used. RNA extraction was done with the Zymo Direct-zol RNA MiniPrep kit (Zymo Research, R2052), with in column DNase I digestion. For poly(A) RNA-seq, libraries were generated using the TruSeq Stranded mRNA Library Prep Kit (Illumina, 20020594), with 1 µg of RNA as input, following the manufacturer's instructions. For total RNA-seq we used the Zymo-Seq RiboFree Total RNA Library Kit (Zymo Research, R3000), with 2 µg of RNA as input.

Global run-on sequencing (GRO-seq)

The GRO-seq protocol employed in this study is not specific for Pol II but instead detects all RNA polymerases engaged in transcription.

Approximately 10 g of inflorescence tissue was collected from wild-type or *mbd5 mbd6* T-DNA mutant plants and immediately placed in 200 ml of ice-cold grinding buffer (300 mM sucrose, 20 mM Tris, pH 8.0, 5 mM $MgCl_2$, 5 mM KCl, 0.2% Triton X-100, 5 mM β-mercaptoethanol, and 35% glycerol). The samples were ground with an OMNI International General Laboratory Homogenizer on ice until well homogenized, filtered twice through miracloth and once through a 40 µm cell strainer into 50 ml conical tubes. The samples were spun down for 10 minutes at 5,250g, nuclei pellets were combined and washed in 25 ml of grinding buffer using a Dounce homogenizer. The wash step was repeated once more before resuspending the nuclei in 1 ml of freezing buffer (50 mM Tris, pH 8.0, 5 mM $MgCl_2$, 20% glycerol, and 5 mM β-mercaptoethanol) and freezing them in liquid nitrogen.

Approximately 5×10^6 nuclei were run-on in 3×NRO-reaction buffer with 20 nM limiting cytidine triphosphate (CTP) at RT as described (*41*). Reactions were stopped by addition of 750 µl TRIzol LS (Fisher Scientific), and RNA was purified according to the manufacturer's manual. RNA pellets were resuspended in 18 µl ddH_2O + 0.05% Tween (dH2O+T), 2 µl 10x fragmentation buffer [100 mM $ZnCl_2$, 10 mMTris-HCl (pH 7.5)] added and incubated at 70°C for 15 minutes. Fragmentation was stopped by addition of 2.5 µl of 100 mM EDTA and 500 µl of ice cold GRO binding buffer [0.25× saline-sodium-phosphate-EDTA buffer (SSPE), 0.05% (vol/vol) Tween, 37.5 mM NaCl, 1 mM EDTA]. BrdU enrichment was performed as described (*42*). For RNA end-repair and decapping by the pyrophosphohydrolase RppH (NEB), RNA

pellets were dissolved in 8 µl TET [10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.05% Tween20] by vigorous vortexing, heated to 70°C for 2 minutes and quick chilled on ice. After a quick spin, 22 µl Repair MM [3 µl 10x PNK buffer, 15.5 µl dH_2O+T, 0.5 µl, SUPERase-In RNase Inhibitor (10 U), 2 µl PNK (20U), 1 µl RppH (5U)] was added, mixed by flicking and incubated at 37°C for 1 hour. To ensure phosphorylated 5'ends, 0.5 µl 100 mM ATP was subsequently added and the reactions incubated for another 45 minutes at 37°C. Following a second IP, sRNA libraries were prepared using the NEB sRNA kit (E7330) as described by the manufacturer. Libraries were amplified for 15 cycles, and size selected on an 8% PAGE/TBE gels to 150–250 base pairs. DNA eluted from gel slices was precipitated with ethanol and resuspended in Tris buffer (Qiagen) before sequencing.

Bisulfite sequencing (BS-seq)

Three bunches of unopened floral bud tissue were collected from individual plants (two plants per genotype as biological replicates). DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, 69106). 200 ng of DNA per sample were sheared with the Covaris S2 instrument to an average fragment size of 200bp (Duty Cycle = 10%, Intensity = 5, Cycles per Burst = 200, Treatment Time = 120 seconds). A-tailing and adapter ligation was performed with the Kapa Hyper Prep Kit (KR0961) using Illumina methylated adapters (TruSeq DNA Sgl index Set A or B). Bisulfite conversion of the DNA was carried on with the EpiTect Bisulfite Kit (Qiagen, 59104). The conversion reaction in the thermocycler was performed twice to increase the efficiency. The converted library was PCR amplified using myTaq HS mix (Bioline) with 12 PCR cycles, purified with Ampure XP beads (Beckman Coulter) and sequenced.

Chromatin immunoprecipitation sequencing (ChIP-seq)

All the experiments shown in the paper were done using FLAG-tagged transgenic lines. The lot of anti-FLAG antibody used in this study generated non-specific signal in the no-FLAG negative control (FLAG hyperchippable sites). We noticed that anti-MYC ChIP-seq performed on *pMBD5::MBD5-9XMYC* and *pMBD6::MBD6-9XMYC* lines generated much cleaner data (although the conclusions made in the paper were unchanged). For this reason, we uploaded the MYC ChIP-seq datasets to GEO (accession: GSE165095).

The experiments reported in the manuscript were performed in T2, on pools of four independent lines per genotype. Col0 plants were grown alongside and used as negative control (no-FLAG control). The experiment shown in Figure 3 was done in F3 (as explained in "Cloning and generation of transgenic lines").

The experimental procedure was performed as previously described (*43*) with some variations. 2 or 2.5 g of inflorescences were used for each ChIP sample. Samples were ground in liquid nitrogen and crosslinked for 10 or 12 minutes at room temperature in Nuclei Isolation buffer (50mM Hepes pH8, 1M sucrose, 5mM KCl, 5mM MgCl2, 0.6% triton X-100, 0.4 mM PMSF, 5 mM benzamidine, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)) containing 1% formaldehyde. In the case of the SLN-FLAG ChIP shown in Figure 3, the samples were first crosslinked with 1.5 mM ethylene glycol bis(succinimidyl succinate) (EGS) for 10 min, and subsequently with 1% formaldehyde for 10 more minutes. The crosslinking reactions were stopped with 125 mM glycine. Crosslinked nuclei were filtered through one layer of miracloth, washed with Extraction Buffer 2, centrifuged through a layer of Extraction Buffer 3 and lysed with Nuclei lysis buffer (buffer compositions are described in the published protocol (*43*)). Chromatin was sheared using a Bioruptor Plus (Diagenode) (20 cycles of 30 seconds on

and 30 seconds off) and immunoprecipitated overnight at 4°C with 3.5 µl of Anti-FLAG M2 (Sigma, F1804). 25 µl each of Protein A and Protein G magnetic Dynabeads (Invitrogen) were added to each sample and incubated for 2 more hours at 4°C. The immunoprecipitated chromatin was washed with Low Salt (2X), High Salt, LiCl and TE buffers, for 5 minutes each at 4°C (buffer compositions are described in the published protocol (*43*)) and eluted twice with 250 µl of elution buffer (1% SDS and 0.1 M $NaHCO_3$) in a thermomixer at 65°C and 1000 rpm, 20 minutes for each elution. Reverse crosslinking was performed overnight at 65°C in 0.2 M NaCl, and proteins were degraded by proteinase K treatment at 45°C for 5h. The DNA fragments were purified using phenol:chloroform and ethanol precipitated overnight at -20°C. Libraries were prepared using the Ovation Ultra Low System V2 1-16 kit (NuGEN, 0344NB-A01) following the manufacturer's instructions, with 15 cycles of PCR.

The H3K9me2 ChIP-seq dataset shown in Figure S11B was publicly available (*44*).

DNA Affinity Purification sequencing (DAP-seq)

The DAP-seq experiment was performed as previously described for the analysis of transcription factors (*21*) with minor variations. First, The MBD5 and MBD6 genes were cloned from cDNA into the pENTR/D-TOPO vector (Thermo Fisher) from the ATG until the last codon before the STOP codon. The genes were transferred into the pIX-HALO destination vector (available from ABRC) via a Gateway LR Clonase II reaction. A control pENTR-GUS vector from the Gateway LR Clonase Enzyme Mix (Thermo Fisher, 11791-043) was also transferred into pIX-HALO and used as negative control for DAP-seq.

The DNA library preparation was prepared as follows: 5 ug of genomic DNA were purified with the DNeasy Plant Mini Kit (Qiagen, 69106) from either Col-0 unopened flower buds or *met1-3* leaf tissue (note that the same *met1-3* DNA used for DAP-seq was also analyzed by BS-seq to obtain accurate information on the methylation status of this DNA). The DNA was diluted with water into 55 µl aliquots containing 1000 ug of DNA each and sheared with the Covaris S2 instrument to an average fragment size of 200 bp (Duty Cycle = 10%, Intensity = 5, Cycles per Burst = 200, Treatment time = 120 seconds). 50 µl of sheared DNA underwent steps 1 to 3 of library construction using the KAPA Hyper Prep kit (KR0961) (End Repair and A-tailing, Adapter Ligation, Post-ligation Cleanup). The adapter ligation was performed using 5 µl of Y adapters, which were annealed as described in the published protocol (*21*). The Post-ligation Cleanup was performed with 1X AMPure beads (Beckman Coulter) instead of 0.8X. The libraries were eluted in 20 µl of QIAGEN EB buffer, pooled according to genotype, and quantified with the Qubit instrument.

The rest of the DAP-seq protocol was performed as described (*21*) using 1000ng of plasmid DNA for each TNT expression reaction, and with either 15, 30, or 100 ng of genomic DNA, as indicated in the figures. We found that the amount of genomic DNA used did not affect the result in an appreciable way. The final PCR amplification of the pulled-down DNA was performed using the Fusion High-Fidelity PCR Master Mix with HF Buffer (F531S) in a final volume of 60 ul, with the following cycling parameters: 95°C for 2 min, 20 cycles of [98°C for 15 sec, 60°C for 30 sec, 72°C for 1 min], 72°C for 10 min. Each sample was individually run on agarose gel and the smear between 200 and 400 bp was cut out, purified and sequenced.

Bioinformatic analyses

BS-seq data analysis

Raw reads were aligned to the TAIR10 genome using Bismark (*45*), with default settings. Reads with three or more consecutive methylated CHH sites were considered as unconverted and were removed. Bismark was used to obtain the methylation ratio for each cytosine and to generate the per-position DNA methylation tracks. The methylation metaplots were generated using ViewBS (*46*) and the genome-wide profiles were made in Rstudio (v 1.3.959) with custom scripts.

BS-PCR data analysis

The BS-PCR data analysis was performed as described in Gallego-Bartolomé 2019 (*30*). The three regions of the *FWA* promoter are: Region 1 (chr4: 13038143-13038272); Region 2 (chr4: 13038356- 13038499); Region3 (chr4: 13038568-13038695).

ChIP-seq and DAP-seq data analysis

ChIP-seq and DAP-seq datasets were analyzed with the same pipeline. The raw reads were filtered based on quality score and trimmed to remove Illumina adapters using Trim Galore (Babraham Institute). Filtered reads were mapped to the Arabidopsis reference genome (TAIR10) with Bowtie2 (*47*) with default parameters. PCR duplicates were removed using MarkDuplicates.jar (picard-tools suite, Broad Institute). Genome browser tracks for visualization purposes were generated using deeptools (v 3.0.2) bamCoverage (*48*) with the options --normalizeUsing RPKM and --binSize 10. To obtain tracks normalized over the no-FLAG control or the GUS control, we used deeptools bamCompare (*48*) with either the "ratio" option or the "log2" option.

ChIP-seq peaks were called with MACS2 (v 2.1.0) (*49*) using an FDR cutoff of 0.01. The no-FLAG wild-type control sample was used as reference. The lot of anti-FLAG antibody used for some of the ChIP-seq experiments generated a considerable amount of non-specific signal (peaks detected in the no-FLAG control). Therefore, for some of the analyses we called peaks on the no-FLAG control alone using MACS2 and we blacklisted those regions of the genome to avoid confounding results caused by FLAG "hyperchippable" sites. The heatmaps were generated using deeptools (*48*) plotHeatmap. The genome-wide correlation scatterplots were obtained with deeptools (*48*) multiBigwigSummary and plotCorrelation, or with the *ggpointdensity* package in RStudio (v 1.3.959).

The wild-type H3K9me2 ChIP-seq data shown in Figure S11B was reanalyzed from a publicly available dataset (GSM3130575) (*44*). The data was normalized over the total H3 ChIP-seq control (GSM3130576).

Methylation density analysis

The methylation density analysis was performed with an in-house script (available upon request). To calculate wild-type methylation density values, we used a BS-seq dataset previously published from our lab, obtained on wild-type (Col-0) flower bud tissue (*9*) because of its deep sequencing coverage. The *met1-3* BS-seq data instead was generated from the same DNA used for DAP-seq. Briefly, the two cytosines of individual CG sites were combined to obtain an average percentage value of methylation for each CG site. In the case of CHG and CHH instead each cytosine was considered individually. The methylation percentage of each cytosine was defined as the number of unconverted Cs divided by the sum of converted and unconverted Cs and multiplied by 100. We then calculated the sum of the methylation percentages in 400 bp bins, for each sequence context. This value of methylation density takes into account both the

number of methylated sites existing in a given bin and the methylation percentage of each site. For each sequence context, the bins that contained one or more cytosines with a coverage of less than 4 reads were discarded. The ChIP-seq or DAP-seq enrichment per 400bp bin was calculated with deeptools bamCompare (*48*). The regression analyses were done in Rstudio (v 1.3.959) using the R packages *ggplot* for the loess curves or *ggpointdensity* for the scatterplots. The genome-wide loess curves for DAP-seq were calculated based on a subset of the entire genome, including only one 400 bp bin every five bins.

To identify the three different genomic regions shown in Figure S6 (GbM, heterochromatin, and RdDM) we employed the following methods: we defined GbM genes with stringent parameters, as genes containing an average CG methylation percentage higher than 20% and average CHG and CHH methylation percentages lower than 0.5% (total of 4244 genes). To define GbM bins we identified the bins that intersected both a GbM gene and a *met1-3* hypoCG high confidence differentially methylated region (hcDMR). This was done to eliminate the regions of the genes that do not contain any methylated CG site. The *met1-3* hypoCG hcDMRs are available from Zhang et al. 2018 (*50*) (dataset: GSM981031). Heterochromatin was defined as bins intersecting a *cmt2* hypoCHH hcDMR, given that CHH methylation is lost in the *cmt2* mutant exclusively in heterochromatic genomic regions enriched in H3K9 methylation, that don't overlap with RdDM sites, but contain CG methylation as well (*25*). The *cmt2* hypoCHH hcDMRs are available from Zhang et al. 2018 (*50*) (dataset: GSM981002). Lastly, RdDM sites were defined as bins overlapping Pol V ChIP-seq peaks (the Pol V data is publicly available (*51*)).

RNA-seq data analysis

The RNA-seq reads were filtered based on quality score and trimmed to remove Illumina adapters using Trim Galore (Babraham Institute). The filtered reads were mapped to the Arabidopsis reference genome (TAIR10) using STAR (*52*), allowing 5% of mismatches (--outFilterMismatchNoverReadLmax 0.05) and unique mapping (--outFilterMultimapNmax 1). Coverage tracks for visualization in the genome browser were generated using deeptools 3.0.2 (*48*) bamCoverage with the options --normalizeUsing RPKM and --binSize 10. The number of reads mapping to genes or transposable elements were determined using HTseq (*53*) with the option --mode=union. The gene counts were used to perform the differential gene expression analysis using the R package DEseq2 (*54*) (cutoff for significance: padj <0.05). In the figures, "transcripts" indicates non-overlapping genes, TE-genes and transposons (Araport11 annotations). The transcripts per million (TPM) values were estimated using Kallisto version 0.46.0 (*55*). In Figure S11 we selected the genes and TEs containing promoter methylation by identifying the loci with an mCG ratio higher than 0.5 in a region of 400bp around the TSS.

The RNA-seq data for *drm1 drm2* and *cmt2 cmt3* mutants and wild-type controls were obtained from GEO (accession n. GSE51304, samples SRR1005385, SRR1005386, SRR1005387, SRR1005388, SRR1005393, SRR1005394) (*25*). The fastq files were downloaded and reanalyzed with the pipeline described above.

The boxplots, MAplots, scatterplots, and barplots were generated using the R package *ggplot*. The heatmap of RNA-seq and methylation data was generated using the R package *ComplexHeatmap* (*56*). For Figures 2A and 2C, we eliminated from the list of differentially expressed transcripts *MBD5*, *MBD6*, and the loci known to show hypervariable DNA methylation (Figure S19), which are genes that lose or acquire methylation over time due to

spontaneous transgenerational epigenetic variation (*57*). This list was generated as described below.

Identification of hypervariable bins and genes.

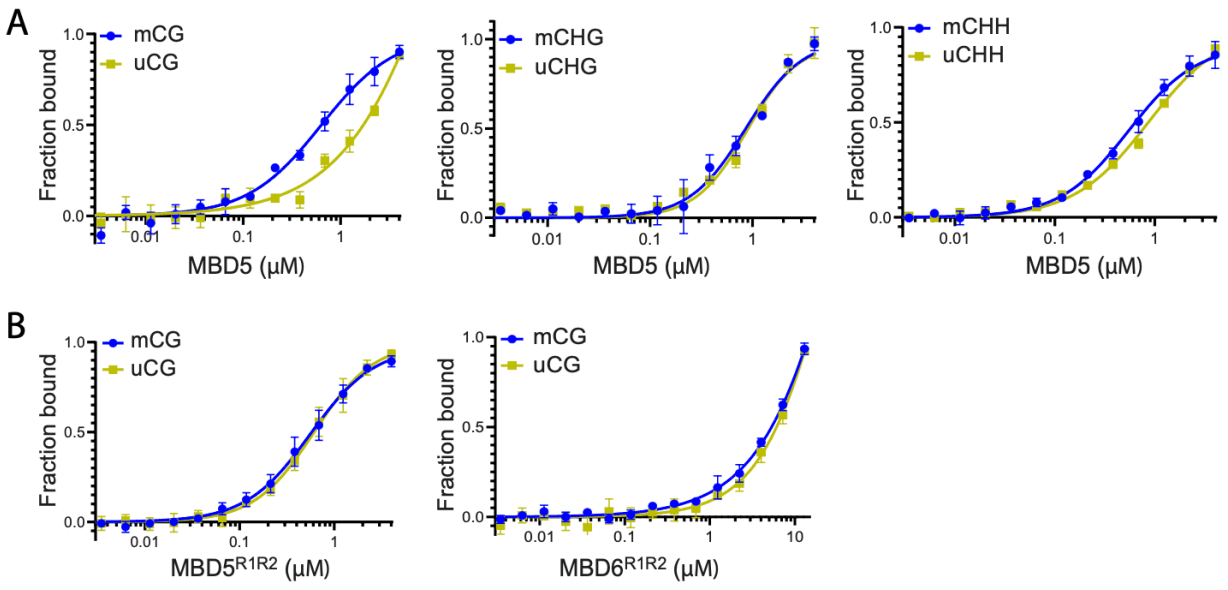
We first identified methylation hypervariable regions based on two published datasets (*50*, *57*). We downloaded the dataset containing the 54 wild-type methylomes by Zhang et al. (https://github.com/yu-z/hcDMR_caller) (*50*) and for each CG methylated bin (100 bp bins with a CG methylation ratio higher than 0.1 in at least one sample) we calculated the standard deviation of the CG methylation level among all the available samples. Note that a large number of bins had missing data and therefore could not be analyzed. The standard deviations of the analyzed bins were distributed around a median value of 0.009, but they displayed a positive skew, corresponding to bins with unusually high variability (Figure S19B). Therefore, we defined as hypervariable the bins that had a standard deviation higher than 0.2 (n=4554). We then merged this list with the set of 72 C-DMRs that were previously defined in Schmitz et al. 2011 (*57*) as regions of spontaneous methylation variation across generations. Next, we compiled a list of potential hypervariable genes by identifying the genes that have a methylation hypervariable bin in a 1 kb region around their TSS (+/- 500 bp). Of note, many of the hypervariable bins correspond to GbM, therefore they are not likely to regulate the expression of the corresponding genes. The list of genes and the list of hypervariable bins are available in Gene Expression Omnibus database (accession: GSE165095).

Analysis of ZF108 off-target sites.

The analysis of the *pUBQ10::ZF108-SLN* RNA-seq experiment shown in Figure S16 was performed with a custom script available upon request. The RNA-seq data of *pUBQ10::ZF108-Ypet* lines, *pUBQ10::ZF108-DNAJ1* lines and their relative controls were obtained from GEO (accession n. GSE108414, samples SRR7727954 to SRR7727961 and SRR6410869 to SRR6410880) (*9*). The fastq files were downloaded and reanalyzed with the pipeline described above. We used a previously published list of ZF108 peaks (n=4951) (*9*). For each gene, we annotated the distance between its TSS and the closest ZF108 peak summit using bedtools *closest* (*58*), which takes into account whether the peak is upstream or downstream of TSS. We then divided all genes in distance groups and calculated the observed over expected ratio of either up or downregulated genes in each group as follows: $[Obs/Exp]_d = (m_d/n_d)/(M/N)$ where m_d is the number of up or downregulated genes in distance group d, n_d is the total number of genes in group d, M is the total number of up or downregulated genes and N is the total number of detected genes (we did not count genes that were not expressed). The p-values of the hypergeometric tests were calculated with *phyper* in R (using the lower.tail=FALSE option).

GRO-seq data analysis

Raw reads were filtered, trimmed and aligned as described for the ChIP-seq data. PCR duplicates were removed as described for the ChIP-seq data. Genome browser tracks for visualization purposes were generated using deeptools (v 3.0.2) (*48*) bamCoverage with the options --normalizeUsing RPKM and --binSize 10. Normalized tracks (log2 fold change of *mbd5 mbd6* over wild-type) were obtained with deeptools bamCompare, with the options --normalizeUsing RPKM and --binSize 10. The log2 fold change enrichment at gene bodies was obtained with deeptools multiBigwigSummary BED-file.



C

Protein	oligo	Kd (μM)	95% CI	Bmax	Max anisotropy
MBD6	mCG	0.163	0.126-0.220	**80.41**	
	uCG	N.D.	N.D.		
MBD6	mCHH	N.D.	N.D.		
	uCHH	N.D.	N.D.		
MBD6	mCHG	N.D.	N.D.		
	uCHG	N.D.	N.D.		
MBD5	mCG	0.635	0.424-1.280	**34.02**	
	uCG	N.D.	N.D.		
MBD5	mCHH	0.545	0.457-0.664	104.7	
	uCHH	0.862	0.740-1.028	**115.1**	
MBD5	mCHG	0.858	0.701-1.071	**74.6**	
	uCHG	0.888	0.757-1.046	73.12	
MBD5R1R2	mCG	0.557	0.440-0.765	84.16	
	uCG	0.614	0.470-0.910	**86.19**	
MBD6R1R2	mCG	N.D.	N.D.		**77**
	uCG	N.D.	N.D.		73

Figure S1: Fluorescence polarization (FP) assays show *in vitro* affinity of MBD5 and MBD6 for CG methylation

A,B) Binding curves of wild-type MBD5 (A) or mutant MBD5^{R1R2} or MBD6^{R1R2} (B) with DNA oligos either methylated (m) or unmethylated (u) in the three different sequence contexts (CG, CHG, CHH) measured by FP assays. Data shown are the averages of three independent experiments and error bars represent the standard error of the mean (SEM). The FP assays demonstrate the specificity of MBD5 binding to mCG DNA (Kd of 0.163 μM) versus uCG DNA (non-specific binding). This specificity is lost when conserved arginine residues are mutated to alanine (MBD5^{R1R2}). C) Table of equilibrium dissociation constant (Kd) values, 95% confidence intervals (CI), binding maximums (Bmax), and max anisotropy values. Bolded Bmax or max anisotropy values were used to normalize anisotropy measurements to create fraction bound plots. Colors of boxes represent data analyzed in the same FP plot.

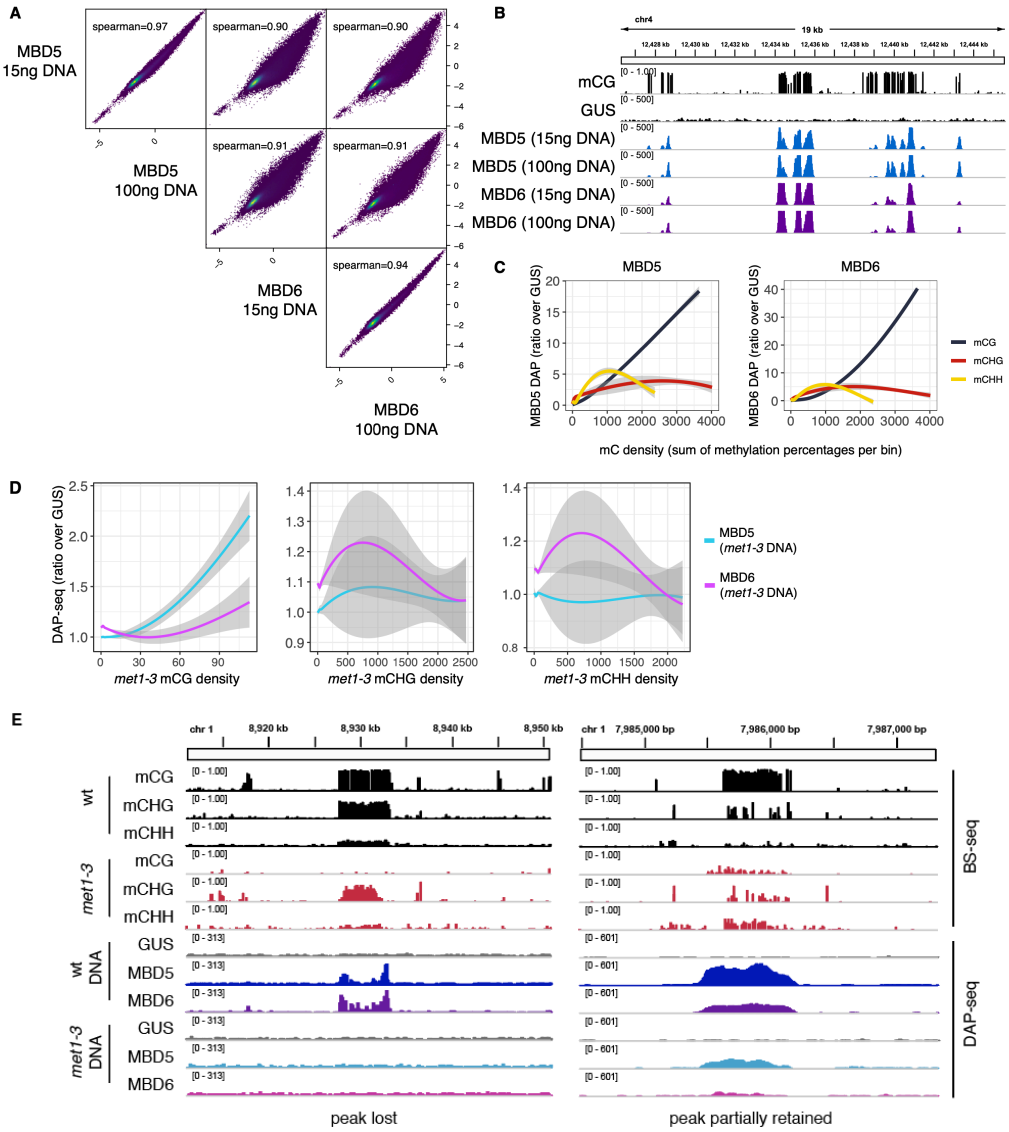


Figure S2: DAP-seq of MBD5 and MBD6 shows specificity for genomic regions methylated in the CG context.

A) Pairwise correlation scatterplots between MBD5 and MBD6 DAP-seq signal performed with either 15 ng or 100 ng of DNA. Each dot is a 500 bp bin of the genome. The signal is plotted as log2 fold change enrichment relative to GUS DAP-seq (negative control). MBD5 and MBD6 DAP-seq enrichment are strongly correlated (Spearman > 0.9). The amount of DNA used in the experiment has minimal effect on the result. B) Representative genome browser tracks showing MBD5 and MBD6 DAP-seq signal at regions of CG methylation. GUS DAP-seq is the negative control. C,D) Genome-wide regression analysis of MBD5 and MBD6 DAP-seq as a function of wild-type (C) or *met1-3* (D) methylation density, showing specificity for the CG sequence context. C,D) Shown are loess curves. Area shaded in grey: 95% confidence interval. E) Representative genome browser tracks showing a methylation patch in which CG (but not non-CG) methylation is completely absent in *met1-3* and MBD5 and MBD6 binding is lost (left) and a methylation patch with low levels of CG methylation in *met1-3* and residual binding of MBD5 and MBD6.

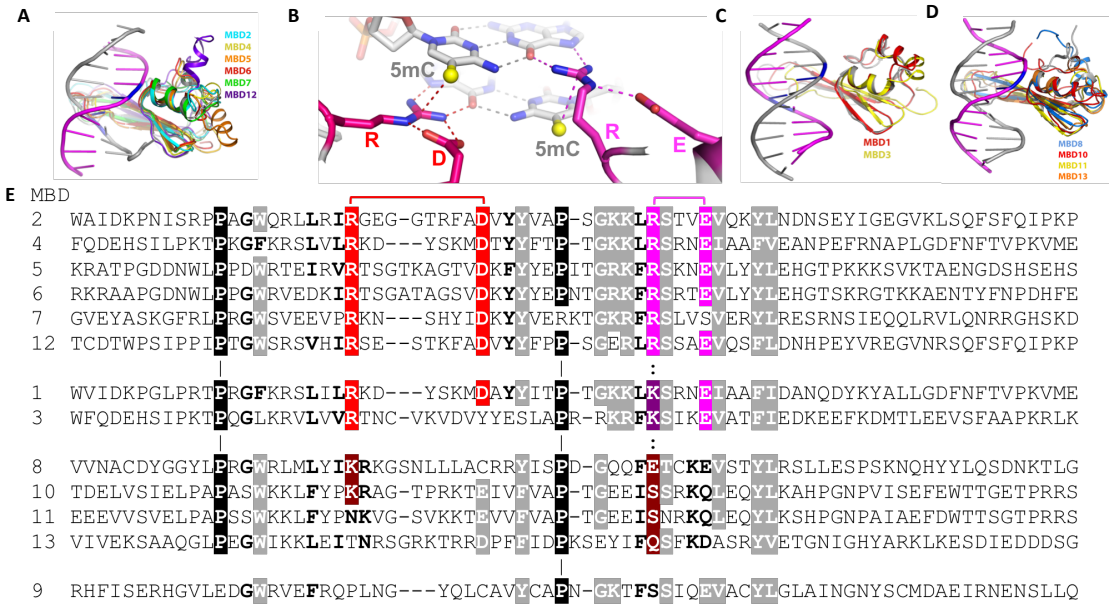


Figure S3. Homology models of Arabidopsis MBD proteins.

A) Structural superimposition of the generated homology models of Group I MBD proteins on to the available crystal structure of mouse MBD4 (PDB ID: 3VYQ). The DNA is shown in magenta and gray with 5mCpG base pairs in blue. B) Detailed view of the interactions between MBD and the methylated DNA interface. Protein interacting residues are labelled and depicted in red (R-D pair) and magenta (R-E pair). 5mCpG dinucleotides are shown in gray with the 5-position methyl group in yellow ball. The two arginine residues each make van der Waals contacts with the methyl group of the neighboring 5mC of the same DNA strand and form '5mC–Arg–G triads' (*19*). The dashed lines indicate the inter- and intra-molecular hydrogen bonds. C,D) Representation of the structural superimposition of the generated homology models of Group II (C) or Group III (D) MBD proteins on to the available solution structure of chicken MBD2 (PDB ID: 2KY8) (C) or crystal structure of human MBD2 (PDB ID: 6C1A) (D). E) Sequence alignment of *A. thaliana* MBD domains (MBD1-13) showing the conserved structural residues (prolines in white letters against black background), 5mCpG interacting residues (R-D and R-E pairs in white letters against red background). Additional conserved residues are in white letters against grey background. MBD9 deviates the most from the consensus and cannot be modeled as a reliable MBD domain.

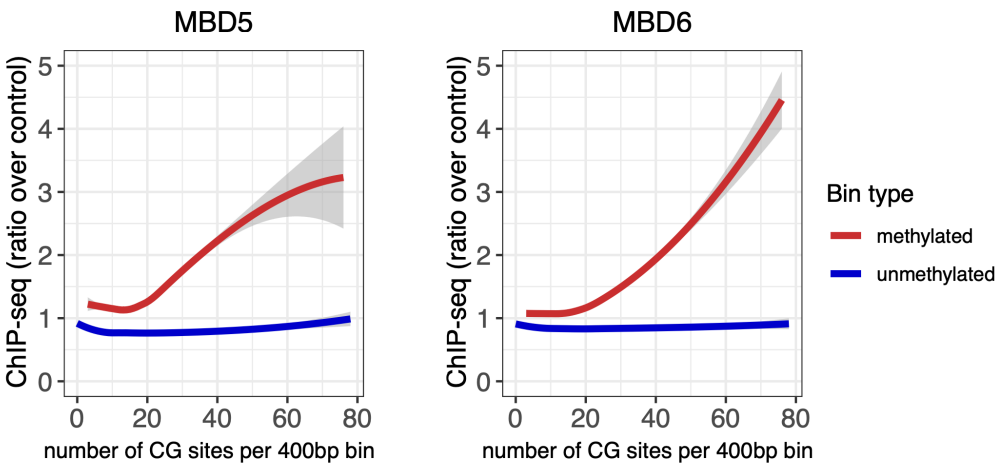


Figure S4: MBD5 and MBD6 ChIP-seq enrichment does not correlate with unmethylated CG density.

MBD5 and MBD6 ChIP-seq enrichment as a function of CG density (number of CG sites per bin) reveals a positive correlation in methylated genomic regions but not in unmethylated regions. Shown are trend lines calculated by locally weighted polynomial regression (loess curves) on 400 bp bins. Each bin was classified as either methylated or unmethylated based on WT CG methylation: 400 bp bins with an mCG density value (sum of CG methylation percentages) equal or higher than 100 were defined as methylated. The area shaded in grey indicates the 95% confidence interval.

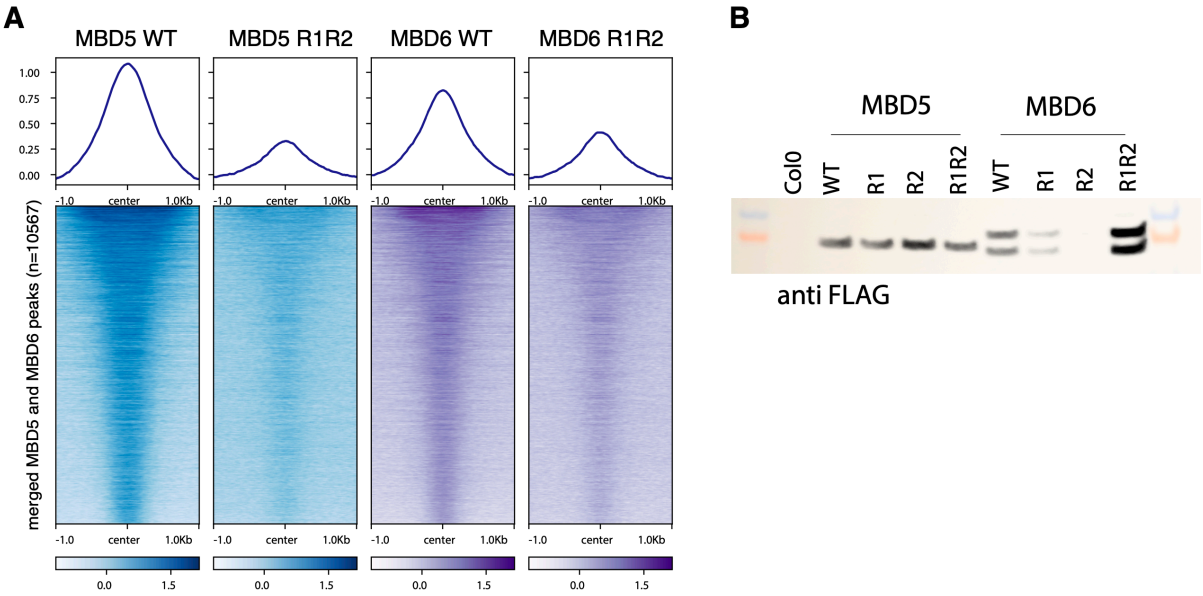


Figure S5: MBD5 and MBD6 are recruited to DNA by direct recognition of methylation.
A) Heatmap of MBD5 and MBD6 WT and R1R2 ChIP-seq data (log2 fold change over no-FLAG control), at MBD5 and MBD6 merged ChIP-seq peaks. Mutation of R1 and R2 to alanine largely affects the binding to chromatin. MBD5: R1=R46A and R2=R69A. MBD6: R1=R92A and R2=R115A. B) Western blot performed on the ChIP input, showing stability of the arginine mutant versions of MBD5 and MBD6. Arginine 1 (R1) and/or arginine 2 (R2) shown in figure 1F were mutated to alanine either individually or in combination. The R1 and R2 single residue mutants were not used for subsequent analysis by ChIP-seq because of poor expression levels for MBD6, likely due to transgene silencing.

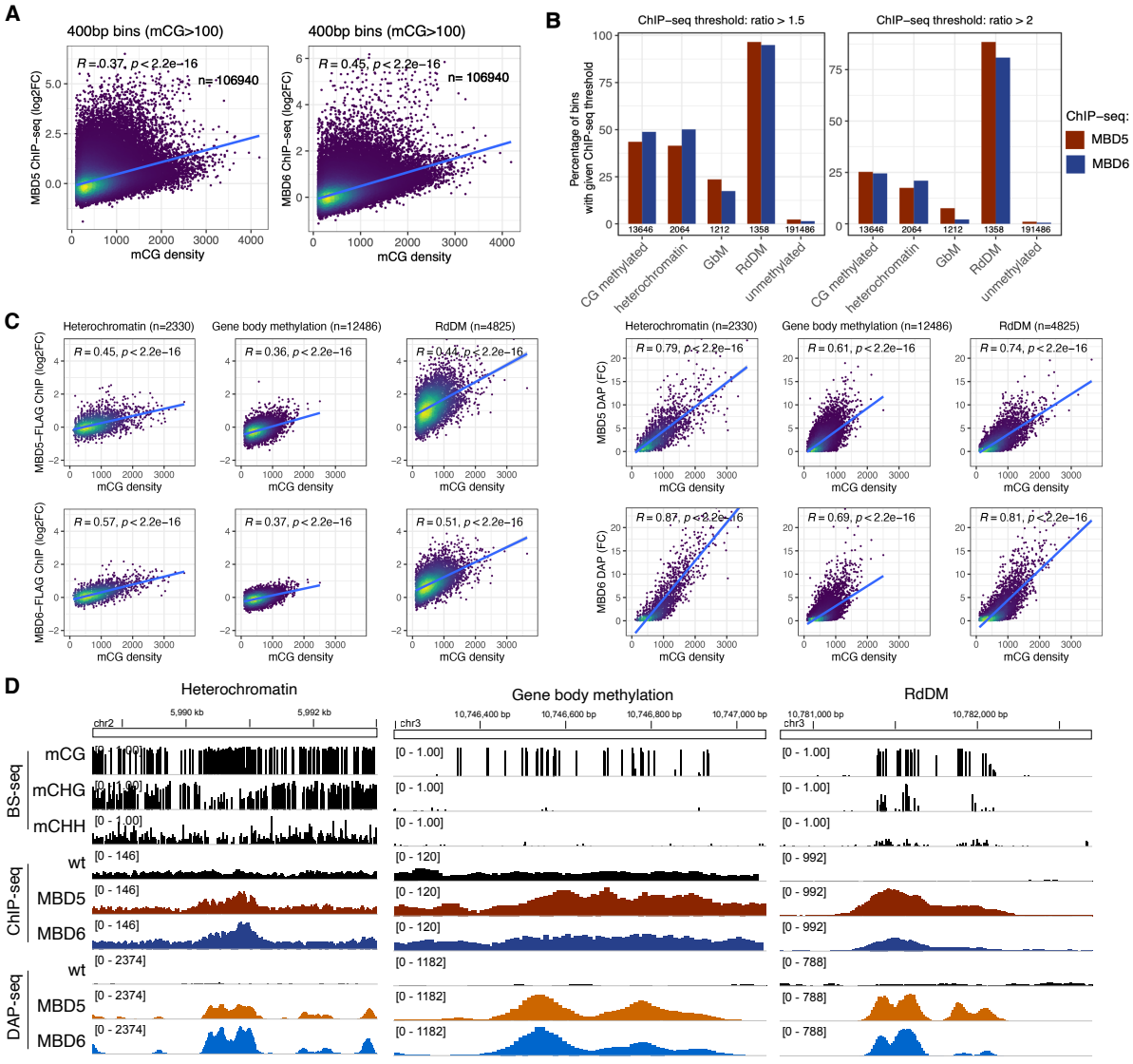


Figure S6: MBD5 and MBD6 correlate with CG methylation genome-wide, with higher enrichment at RdDM sites.

A) Methylome-wide scatterplots of MBD5 or MBD6 ChIP-seq as a function of CG methylation density (sum of CG methylation percentages per 400 bp bin). The figure includes all the 400 bp bins that have an mCG density value higher than 100. Blue line: linear trend. R and p-value: Spearman. B) Analysis of genome-wide binding preferences of MBD5 and MBD6. The bar plot indicates the percentage of bins containing a given threshold of MBD5 or MBD6 ChIP-seq enrichment [ratio over the no-FLAG control > 1.5 (left) or > 2 (right)]. "CG methylated" bins are all the 400 bp bins with an mCG density value higher than 1000. "Heterochromatin", "GbM", and "RdDM" are the bins with mCG density > 1000 and belonging to one of the three indicated genomic regions (see Methods for details). "Unmethylated" are all the bins with mCG density < 100. The total number of bins in each group is displayed at the bottom of the bars. C) Scatterplots of MBD5 or MBD6 ChIP-seq (left) or DAP-seq (right) enrichment as a function of CG methylation density, at bins overlapping the three genomic regions and containing mCG

density values higher than 100. In the ChIP-seq data the slope of the regression is higher at RdDM sites, while this difference is not observed in the DAP-seq data. Blue line: linear trend. Shaded area: 95% confidence interval. R and p-value: Spearman. D) Genome browser tracks of wild-type DNA methylation and MBD5/6 ChIP-seq and DAP-seq at a heterochromatic site, a gene body methylation site, and an RdDM site. Note that the MBD5/6 ChIP-seq signal to noise ratio is higher at the RdDM site.

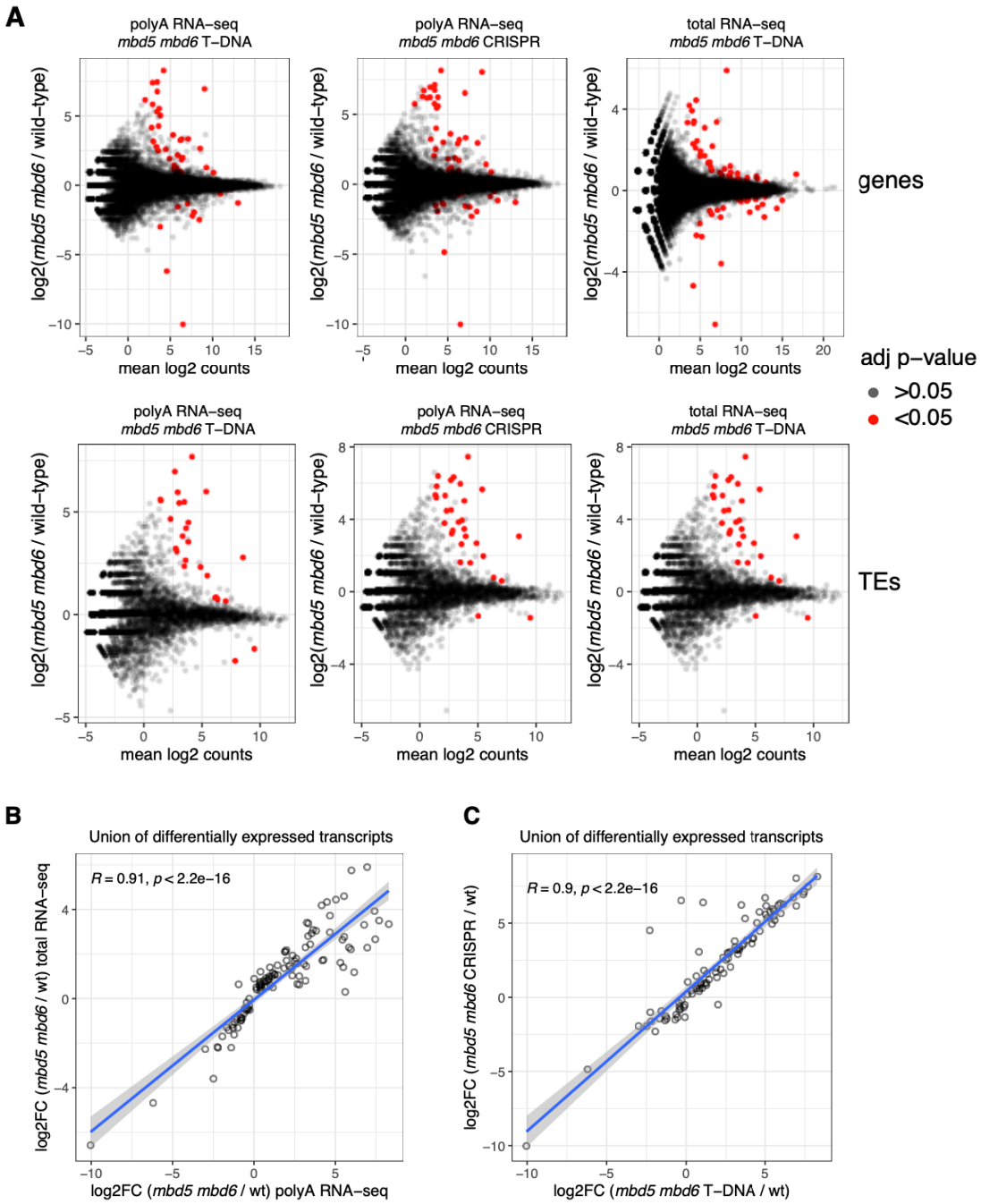


Figure S7: Differential gene expression analysis for *mbd5 mbd6* double mutants.

A) MA plots of differential gene expression analysis for poly(A) and total RNA-seq of *mbd5 mbd6* T-DNA and CRISPR mutants. Three biological replicates for each genotype were used. The top panels display the analysis performed with gene annotations (this includes TE-genes). The bottom panels display transposable elements only. B) Correlation between poly(A) RNA-seq and total RNA-seq for *mbd5 mbd6* T-DNA. C) Correlation between *mbd5 mbd6* T-DNA and CRISPR lines (poly(A) RNA-seq). B,C) R and p-value: Spearman. Blue line: linear model fit. Shaded area: 95% confidence interval. Transcripts include genes and transposons.

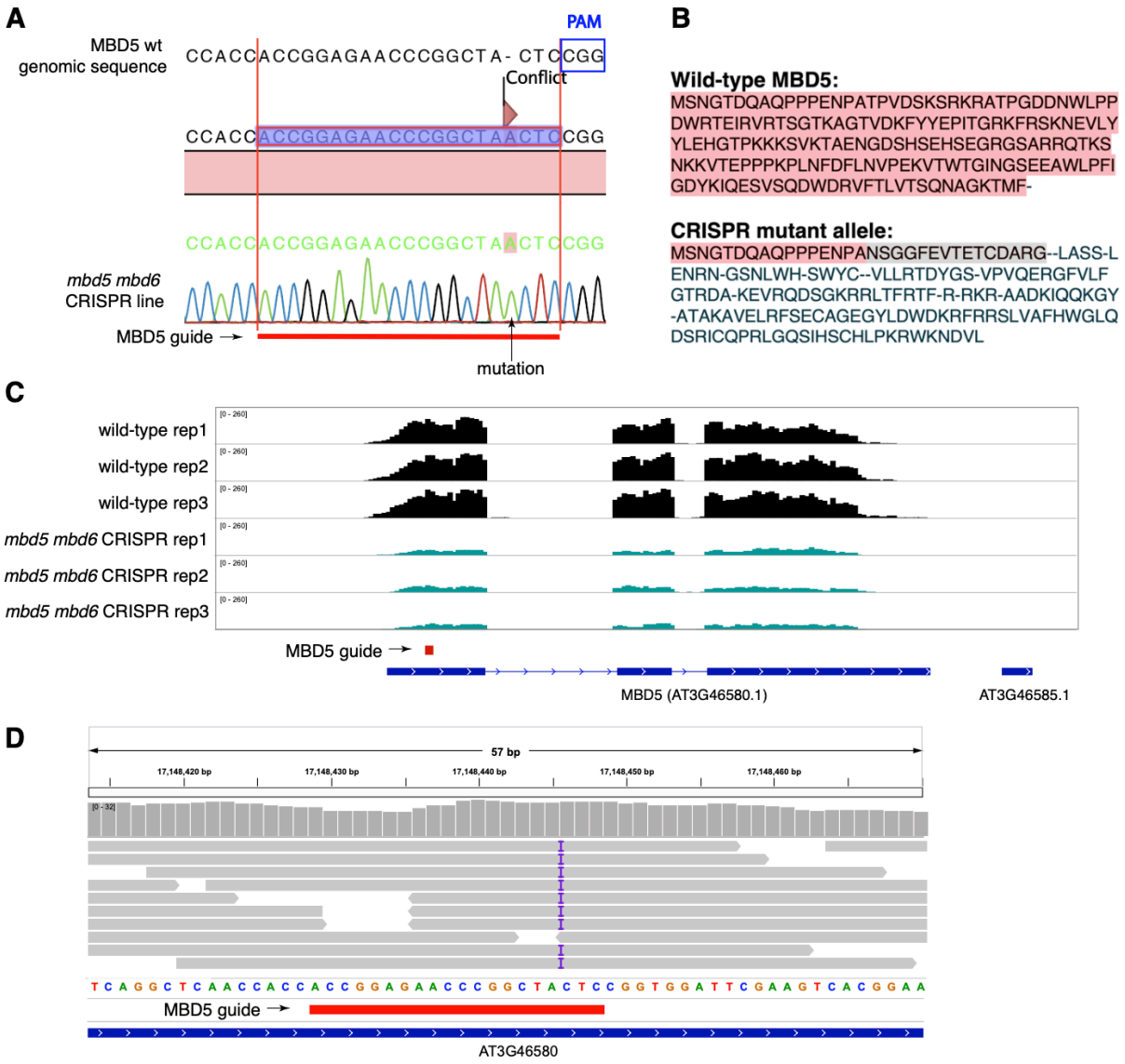


Figure S8: Generation of *mbd5* CRISPR mutant allele.

A) Sanger sequencing chromatogram of the MBD5 locus in the *mbd5 mbd6* CRISPR mutant line. The mutation that was generated is a single nucleotide insertion (adenine) 3bp before the PAM. B) Predicted amino acid sequence of the MBD5 WT and CRISPR mutant alleles. The CRISPR mutation generates a premature STOP codon after 31 amino acids. C) RNA-seq tracks showing nonsense-mediated decay of the *MBD5* mRNA. D) Representative BAM file of an RNA-seq track of *mbd5 mbd6* CRISPR showing the insertion 3 bp before the PAM.

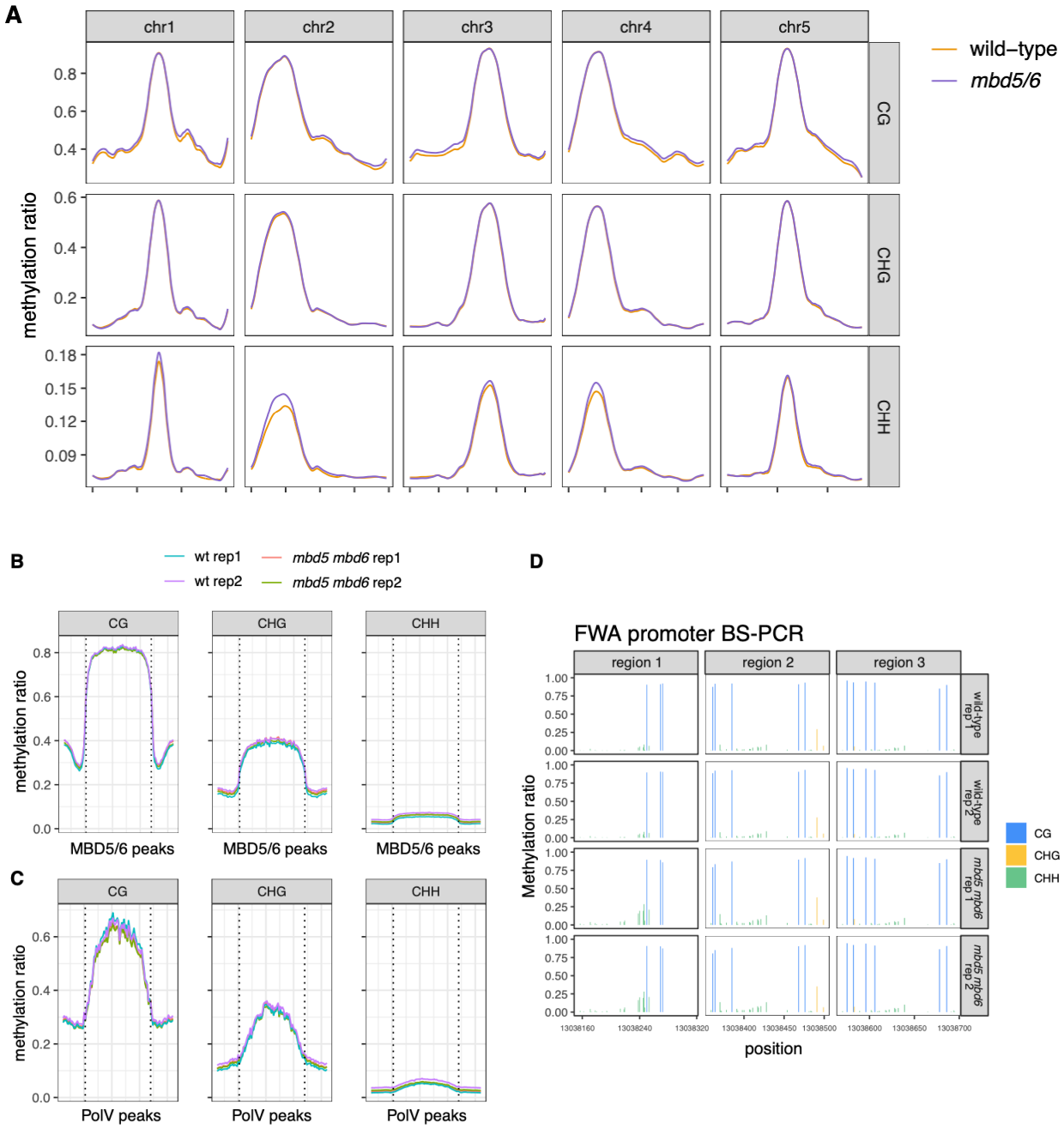


Figure S9: DNA methylation levels are not altered in *mbd5 mbd6* double mutant plants.
A) Genome-wide methylation profiles of wild-type and *mbd5 mbd6* BS-seq data. In this analysis two biological replicates for each genotype were merged. The data is smoothed with the R loess function (span = 0.05). B,C) Metaplots of methylation levels in two biological replicates each of wild-type and *mbd5 mbd6* plants at merged MBD5 and MBD6 ChIP-seq peaks (B) and at Pol V ChIP-seq peaks (C). D) BS-PCR of the *FWA* promoter, divided in three different regions.

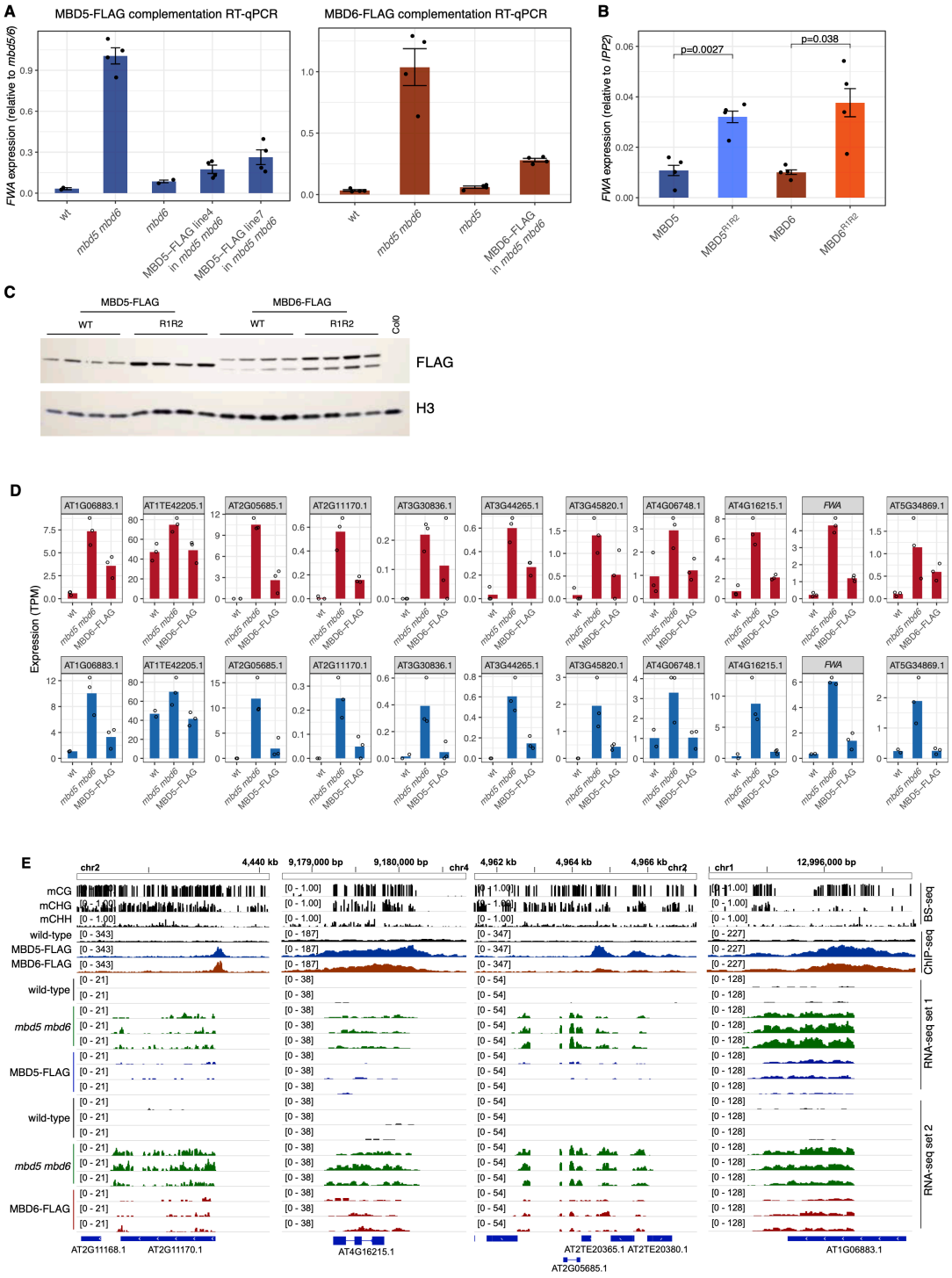


Figure S10: MBD5-FLAG and MBD6-FLAG complement the *mbd5 mbd6* derepression phenotype.

A) RT-qPCR analysis of *FWA* expression in wild-type plants, *mbd5* and *mbd6* single mutants, *mbd5 mbd6* T-DNA double mutants, and complementation lines of either MBD5-FLAG or MBD6-FLAG driven by their endogenous promoters, in the *mbd5 mbd6* T-DNA double mutant

background. B) The R1R2 mutant versions of MBD5 and MBD6 cannot complement *FWA* derepression in *mbd5 mbd6*. Shown are T1 lines of either wild-type or R1R2 FLAG-tagged proteins in the *mbd5 mbd6* CRISPR background. A,B) Each dot represents a biological replicate (individual plants). The data was normalized over the *IPP2* housekeeping control. Error bars: SEM. p-value: t-test. C) Western blot analysis showing expression levels of MBD5-FLAG and MBD6-FLAG wild-type and R1R2 mutants in the same plants that were used for the RT-qPCR in panel B. D) Bar plots showing the average expression level calculated by RNA-seq (transcripts per million [TPM]) of different genes and transposons derepressed in *mbd5 mbd6* and rescued by MBD5-FLAG or MBD6-FLAG. Each dot represents a biological replicate (individual plants). E) Genome browser shots showing four different transposons that are bound by MBD5 and MBD6 (ChIP-seq), derepressed in *mbd5 mbd6* and rescued by MBD5-FLAG or MBD6-FLAG (RNA-seq). B,D) The complementation lines are the same used for RT-qPCR in panel A.

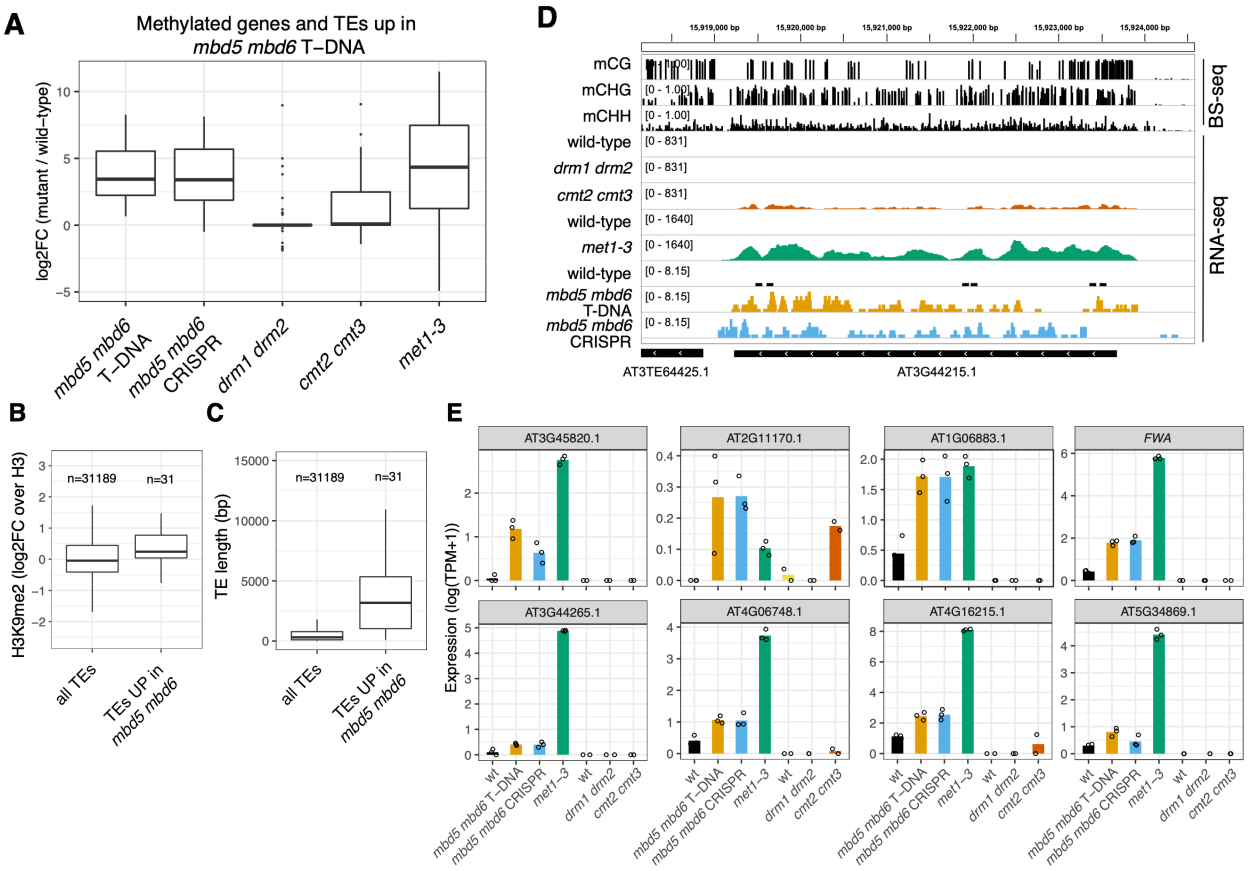


Figure S11: Comparison of the RNA-seq phenotype in *mbd5 mbd6* and in different methylation mutants.

A) Boxplot showing log2 fold changes in gene expression for the *mbd5 mbd6* T-DNA derepressed genes and TEs containing promoter methylation. B,C) Boxplot of wild-type H3K9me2 enrichment (B) or length (C) of all TEs (Araport11 annotations) and of the union of TEs derepressed in *mbd5 mbd6* T-DNA and CRISPR mutants (n=31). The H3K9me2 ChIP-seq dataset was reanalyzed from Li et al. (*44*). D) Genome browser track showing a representative TE derepressed in *mbd5 mbd6, cmt2 cmt3*, and *met1-3,* but not in *drm1 drm2*. E) Bar plots of average expression levels calculated by RNA-seq (transcripts per million [TPM]) for several high-confidence MBD5/6 targets. Each dot represents a biological replicate (individual plants). A,D,E) The *drm1 drm2* and the *cmt2 cmt3* RNA-seq data was reanalyzed from Stroud et al. (*25*).

	Species	Protein ID	aa range	Type
At	*Arabidopsis thaliana*	AT5G37380 (SLN)	1-431	Dicot
Gm	*Glycine max*	XP_003533259.1	1-420	Dicot
Sl	*Solanum lycopersicum*	XP_004234172.1	1-402	Dicot
Zm	*Zea mays*	PWZ45140.1	1-441	Monocot
Ta	*Triticum aestivum*	KAF7026567.1	1-444	Monocot
Os	*Oryza sativa*	EAY93887.1	1-434	Monocot
Mp	*Marchantia polymorpha*	PTQ49649 .1	1-598	Moss
Pp	*Physcomitrium patens*	XP_024401060.1	1-555	Moss
Ac	*Adiantum capillus-veneris*	MBC9845780.1	1-480	Fern

Figure S12: Conservation of SLN within the plant kingdom.

Sequence alignment of the closest homologue to SLN in 8 different plant species spanning monocots, dicots, mosses and ferns. The species names, protein IDs, and amino-acids (aa) range are indicated in the table. The full-length sequence of SLN is shown (aa 1-431).

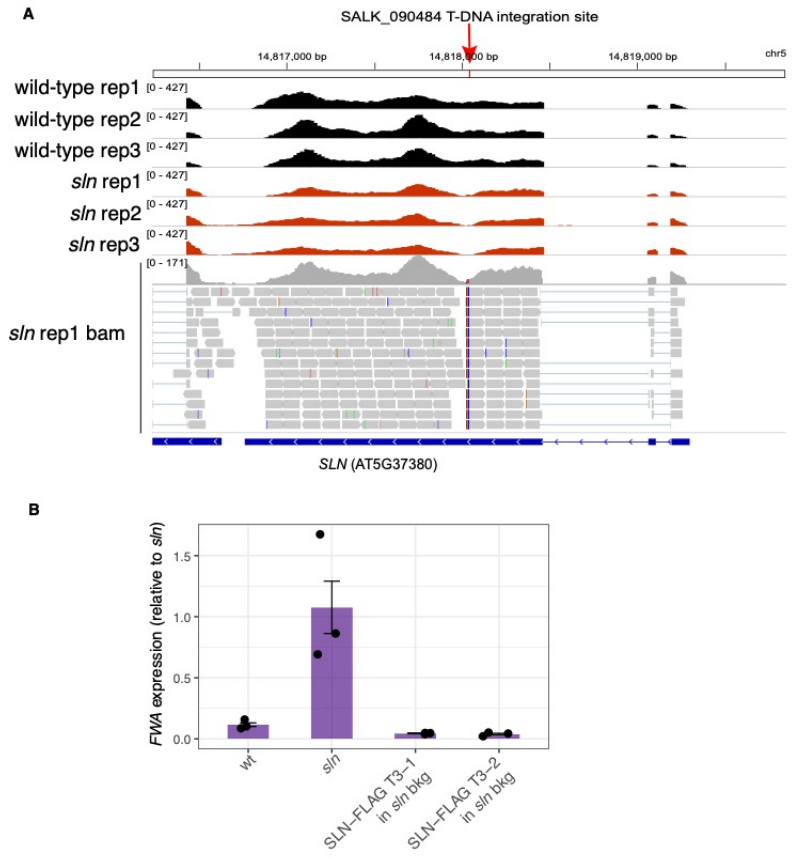


Figure S13: Validation of *sln* T-DNA mutant line.

A) RNA-seq tracks and one representative bam file showing the *SLN* locus in wild-type and *sln* mutant plants (SALK_090484). The red arrow indicates the T-DNA integration site, which lies in a coding region. B) RT-qPCR analysis of *FWA* expression in wild-type plants, *sln* mutants, and two complementation lines expressing SLN-FLAG driven by its endogenous promoters, in the *sln* mutant background. Both lines are single insertion homozygous lines, in the T3 generation. These lines were crossed to *mbd5 mbd6* to perform the ChIP-seq and IP-MS experiments (see Methods section).

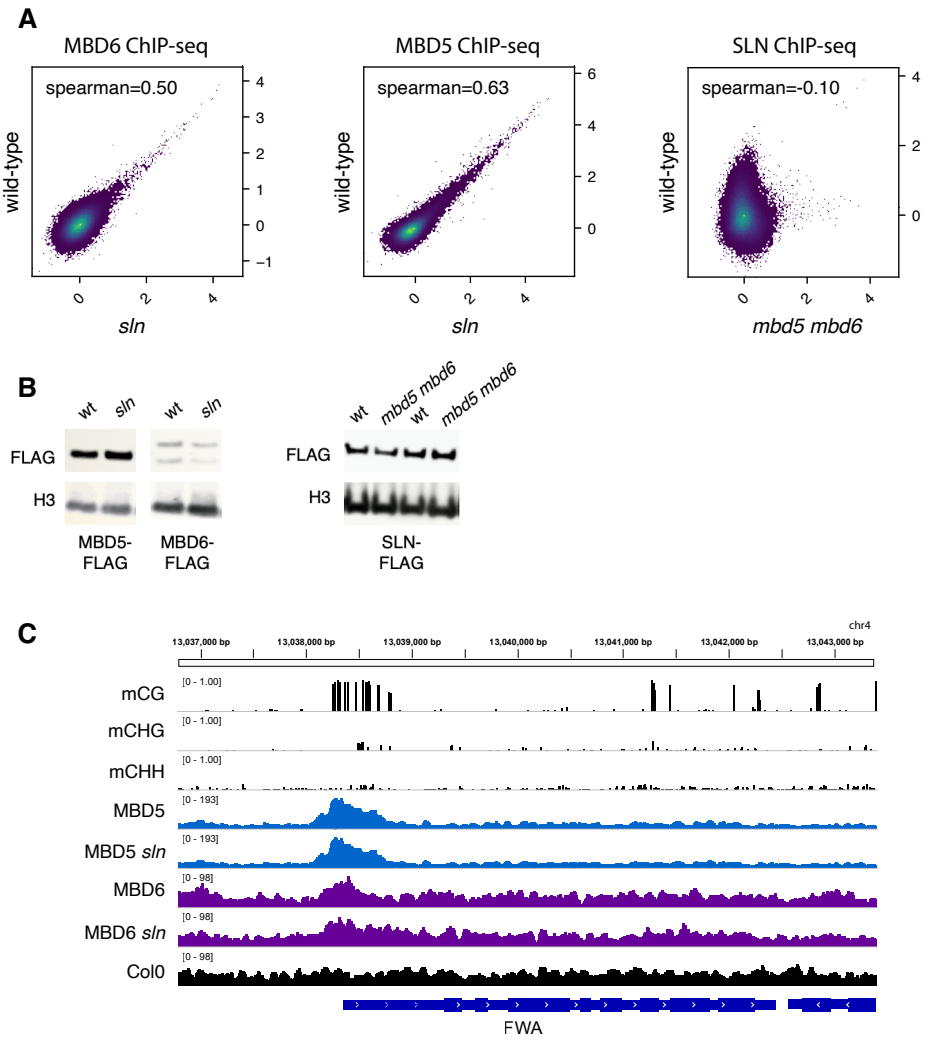


Figure S14: SLN is recruited to DNA by MBD5 and MBD6.

A) Genome-wide scatterplots of MBD5, MBD6 and SLN ChIP-seq signal in the different mutant backgrounds indicated in the axis labels. Each point represents a 500 bp bin, data is shown as log2 fold change over the no-FLAG control. The regions of the genome that were called as peaks in the no-FLAG control were blacklisted to remove hyperchippable sites from the scatterplot. B) Western blot analysis performed on the ChIP-seq input to show comparable expression levels of the FLAG transgenes in the different mutant backgrounds. C) Genome browser track showing that the binding of MBD5 and MBD6 to the *FWA* promoter is not affected in the *sln* mutant background.

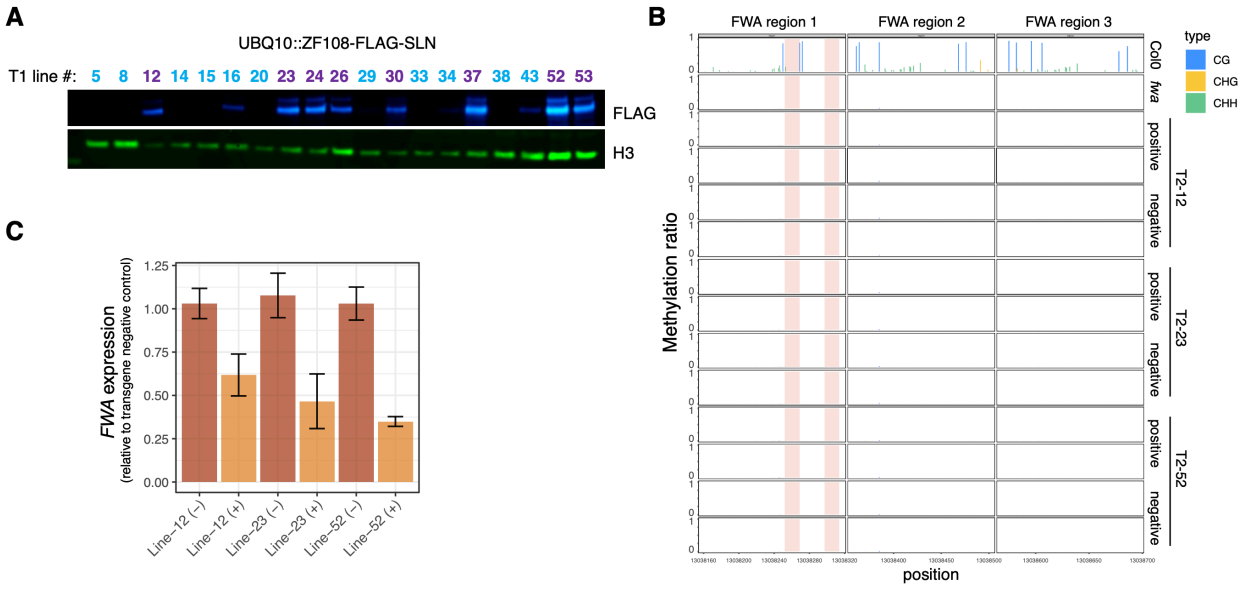


Figure S15: ZF108-SLN represses *FWA* in the absence of DNA methylation.

A) Western blot showing expression levels of the transgene ZF108-SLN in the T1 lines analyzed in figure 4B. The T1 lines that were found to express high levels of the transgene are marked in purple, while the lines with low or no detectable expression are marked in light blue. B) BS-PCR analysis of DNA methylation levels at the *FWA* promoter of three independent T2 lines. Col0 and *fwa* epiallele plants are shown as control. For each T2 line, two transgene positive and two transgene negative plants were analyzed. None of the plants selected acquired DNA methylation at *FWA*. Shaded in pink are the binding sites of ZF108. C) RT-qPCR analysis of *FWA* expression in the three lines shown in panel B and in Figure 4C. The data was normalized over *IPP2* and is displayed as relative to the average signal of the null segregants (-) for each line. Values correspond to the average of 8 biological replicates with SEM as error bars.

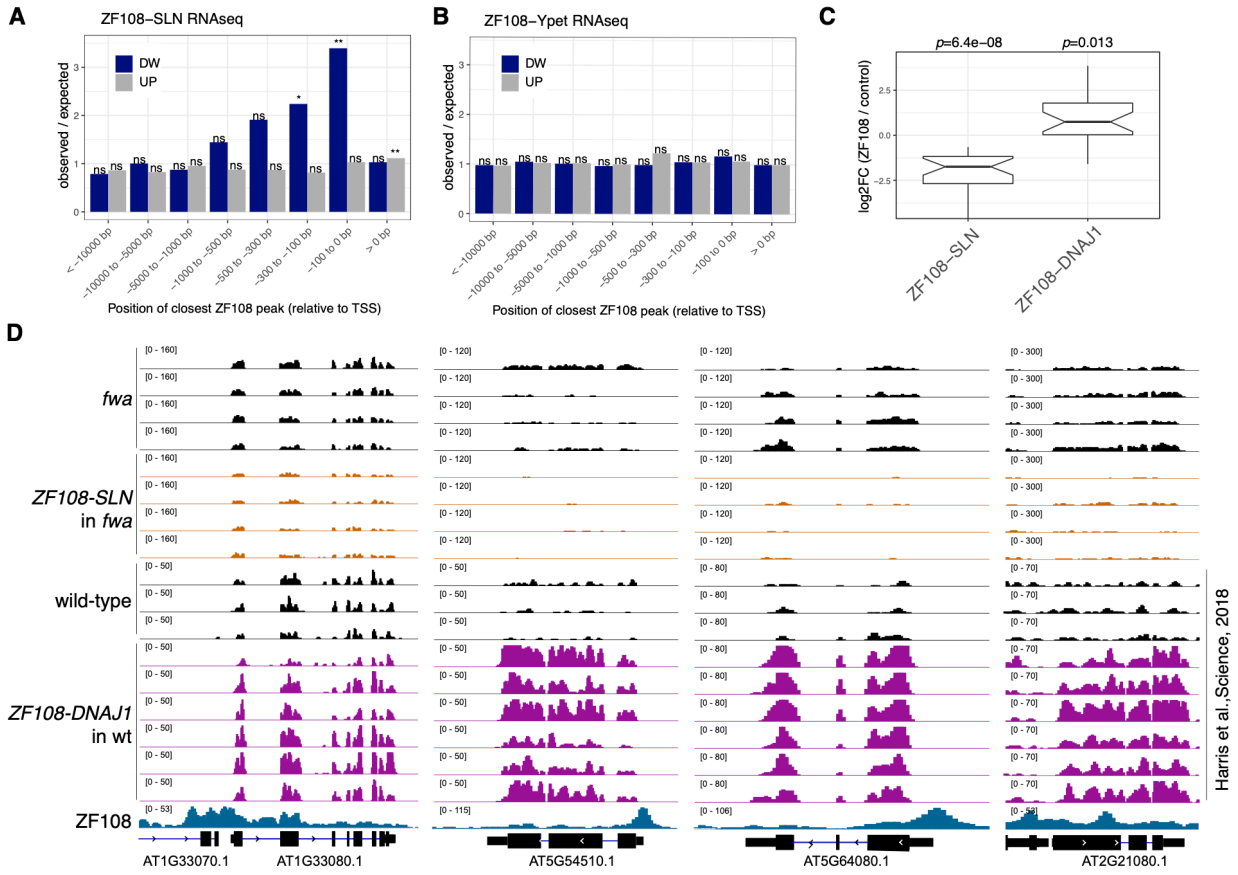


Figure S16: Ectopic recruitment of SLN causes silencing of ZF108-bound targets.

A,B) RNA-seq analysis of lines expressing ZF108-SLN or ZF108-Ypet driven by the *UBQ10* promoter. Genes that have a ZF108 peak close to their TSS are more likely to be downregulated in plants expressing ZF108-SLN (left) but not in control plants expressing ZF108-Ypet (right). The plot indicates the observed/expected ratio of the number of upregulated or downregulated genes among all the genes that have a ZF108 peak within a given distance from their TSS. Negative values in the x axis indicate positions upstream of the TSS. The ZF108-Ypet RNA-seq data is publicly available *(9)*. Hypergeometric test: *=p<0.05, **=p<0.01, "ns" = non-significant. C) Boxplot of expression changes at genes that have a ZF108 peak within 500 bp upstream of their TSS and that are downregulated in *pUBQ10::ZF108-SLN* lines (n=27). Publicly available RNA-seq data of *pUBQ10::ZF108-DNAJ1* lines *(9)* is shown as a control: the same genes are upregulated by ectopic recruitment of DNAJ1. p-value: one-sample t-test (μ=0). D) Genome browser tracks showing downregulation and upregulation respectively of four different ZF108 bound genes in the RNA-seq analysis of *pUBQ10::ZF108-SLN* and *pUBQ10::ZF108-DNAJ1* plants. Shown in blue is a published ZF108 ChIP-seq track (*9*).

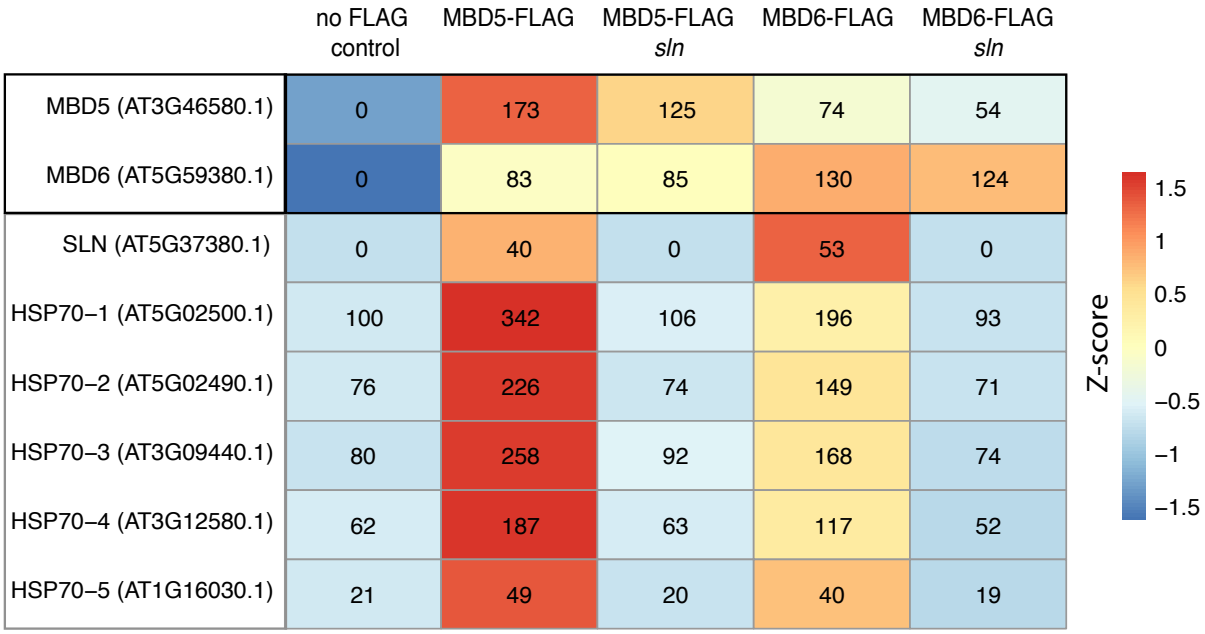


Figure S17: MBD5 and MBD6 interact with Hsp70 proteins only in the presence of SLN.
Immunoprecipitation mass spectrometry data of MBD5-FLAG and MBD6-FLAG in wild-type or *sln* mutant backgrounds. In the absence of SLN, MBD5 and MBD6 do not show enrichment of the HSP70 proteins. The numbers indicate spectral counts.

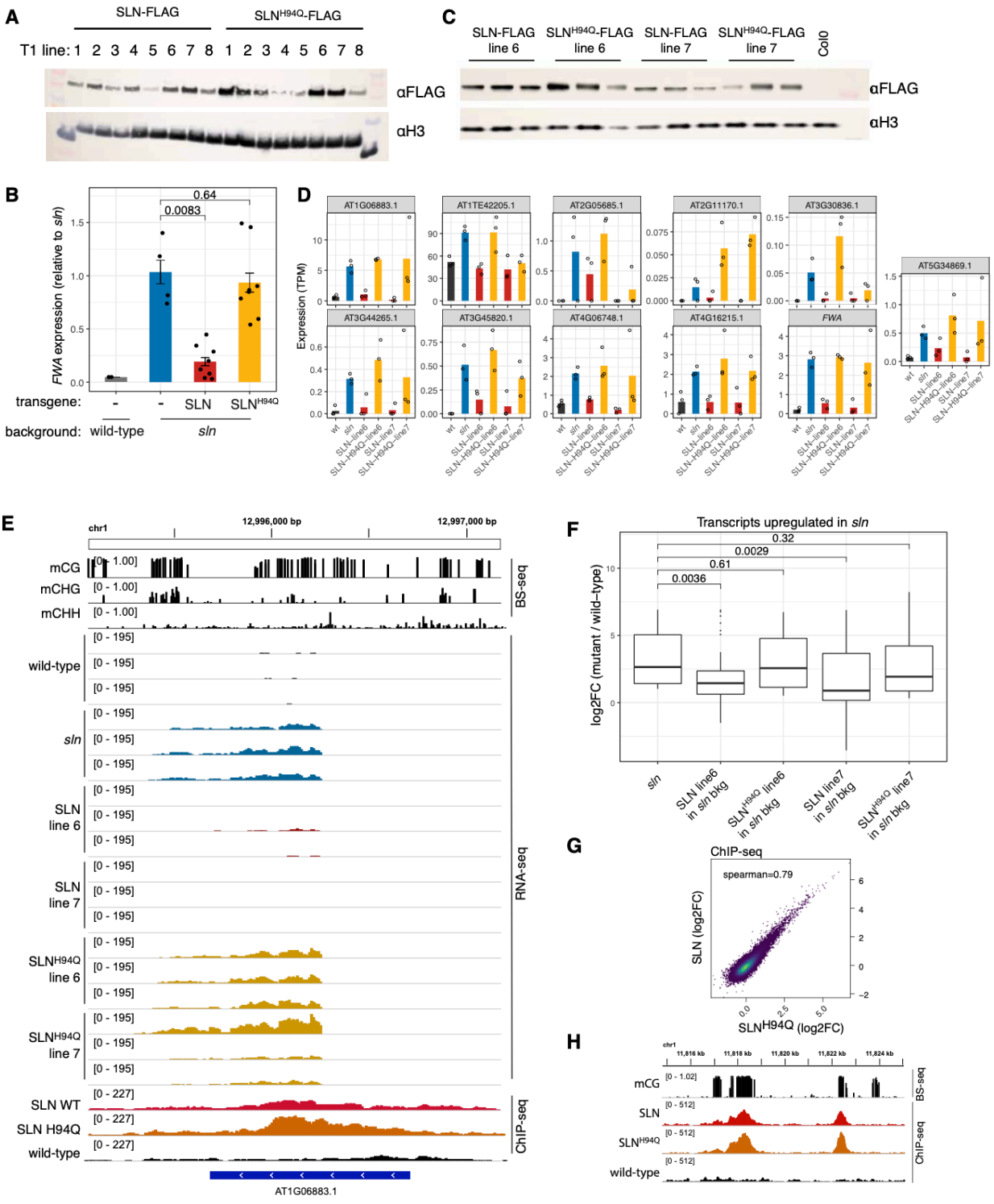


Figure S18: The HPD domain of SLN is required for gene silencing.

A) Western blot analysis shows stability of the SLN^{H94Q} mutant protein. The T1 plants analyzed are the same plants used for the RT-qPCR in Figure 4A. B) RT-qPCR analysis of *FWA* expression performed on the SLN and SLN^{H94Q} T2 lines used for IP-MS and ChIP-seq (Figure

4B,C). The data is normalized over the *IPP2* housekeeping control and is displayed as relative to *sln* mutant plants. Each dot represents a biological replicate (individual plants). Error bars: SEM. p-value: t-test. C) Western blot analysis of the FLAG-tagged SLN and SLN^{H94Q} expressing plants used for RNA-seq in panels D-F. D) Bar plots showing the average expression level calculated by RNA-seq (transcripts per million [TPM]) for the same genes and transposons derepressed in *mbd5 mbd6* (Figure S10D). Each dot represents a biological replicate (individual plants). SLN^{H94Q} fails to complement the transcription defect. E) Genome browser tracks showing a transposable element bound by SLN, derepressed in *sln*, complemented by wild-type SLN but not by SLN^{H94Q}. F) Boxplot showing the complementation of all genes and transposons derepressed in *sln*. p-value: t-test. G) Genome-wide correlation analysis between ChIP-seq enrichment of SLN and SLN^{H94Q} (data is shown as log2 fold change over the no-FLAG control). Each dot represents a 500 bp bin. The regions of the genome that were called as peaks in the no-FLAG control were blacklisted to remove hyperchippable sites from the scatterplot. H) Representative genome browser tracks showing SLN and SLN^{H94Q} ChIP-seq at patches of CG methylation. The point mutation does not affect the ability of SLN to bind chromatin.

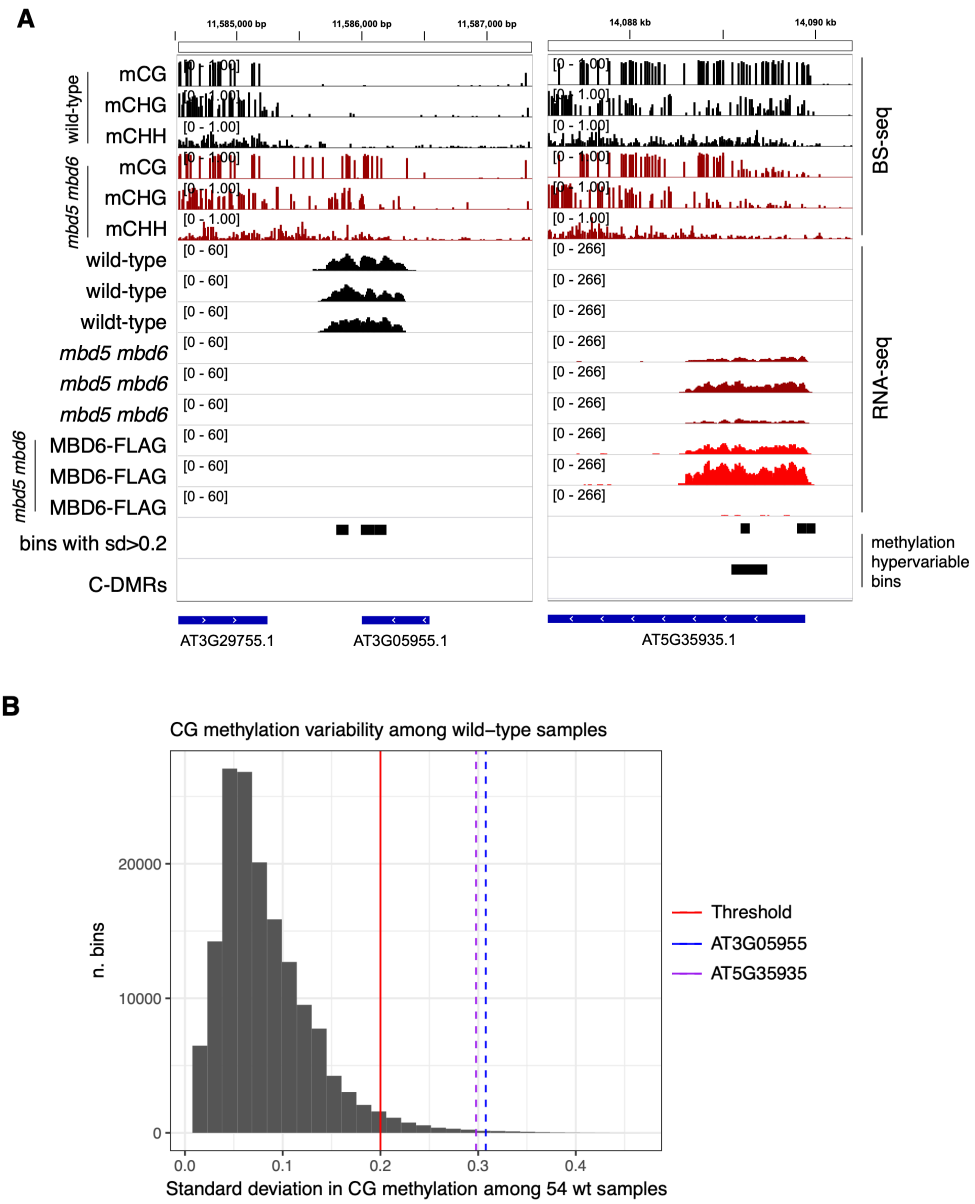


Figure S19: Identification of methylation hypervariable regions and genes.

A) Genome browser tracks of DNA methylation and RNA-seq in wild-type, *mbd5 mbd6* mutant, and MBD6-FLAG complementation lines in the *mbd5 mbd6* mutant background (same samples as in Figure S10). Shown are two examples of genes that overlap methylation hypervariable regions, are either down (left) or up (right) -regulated in *mbd5 mbd6,* and they are not complemented by MBD6-FLAG. The track named "bins with sd>0.2" contains hypervariable bins as defined in panel B (n=4554), the track named "C-DMRs" contains the hypervariable regions defined in Schmitz et al. 2011 (n=72) (*57*). B) Histogram showing the distribution of CG methylation variability (standard deviation) among 54 wild-type BS-seq samples (*50*). The red line indicates the threshold that was used to define hypervariable bins (see Methods for details). The blue and purple lines indicate the standard deviation values of the bins overlapping the two genes shown in panel A.

Additional data – Table S1

Excel file including all the IP-MS experiments performed in this study. Numbers indicate spectral counts. The samples that were processed in the same experiment are grouped by color. "NA" indicates proteins that were not detected in any of the samples belonging to a given experiment.

References and Notes

1. M. G. Goll, T. H. Bestor, Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481–514 (2005). doi:10.1146/annurev.biochem.74.010904.153721 Medline

2. A. Zemach, I. E. McDaniel, P. Silva, D. Zilberman, Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916–919 (2010). doi:10.1126/science.1186366 Medline

3. T. Baubec, R. Ivánek, F. Lienert, D. Schübeler, Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153**, 480–492 (2013). doi:10.1016/j.cell.2013.03.011 Medline

4. T. Shimbo, P. A. Wade, Proteins that read DNA methylation. *Adv. Exp. Med. Biol.* **945**, 303–320 (2016). doi:10.1007/978-3-319-43624-1_13 Medline

5. L. D. Boxer, W. Renthal, A. W. Greben, T. Whitwam, A. Silberfeld, H. Stroud, E. Li, M. G. Yang, B. Kinde, E. C. Griffith, B. Bonev, M. E. Greenberg, MeCP2 represses the rate of transcriptional initiation of highly methylated long genes. *Mol. Cell* **77**, 294–309.e9 (2020). doi:10.1016/j.molcel.2019.10.032 Medline

6. M. J. Lyst, R. Ekiert, D. H. Ebert, C. Merusi, J. Nowak, J. Selfridge, J. Guy, N. R. Kastan, N. D. Robinson, F. de Lima Alves, J. Rappsilber, M. E. Greenberg, A. Bird, Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **16**, 898–902 (2013). doi:10.1038/nn.3434 Medline

7. X. Nan, H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, A. Bird, Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998). doi:10.1038/30764 Medline

8. X. Zhang, J. Yazaki, A. Sundaresan, S. Cokus, S. W.-L. Chan, H. Chen, I. R. Henderson, P. Shinn, M. Pellegrini, S. E. Jacobsen, J. R. Ecker, Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**, 1189–1201 (2006). doi:10.1016/j.cell.2006.08.003 Medline

9. C. J. Harris, M. Scheibe, S. P. Wongpalee, W. Liu, E. M. Cornett, R. M. Vaughan, X. Li, W. Chen, Y. Xue, Z. Zhong, L. Yen, W. D. Barshop, S. Rayatpisheh, J. Gallego-Bartolome, M. Groth, Z. Wang, J. A. Wohlschlegel, J. Du, S. B. Rothbart, F. Butter, S. E. Jacobsen, A DNA methylation reader complex that enhances gene transcription. *Science* **362**, 1182–1186 (2018). doi:10.1126/science.aar7854 Medline

10. A. Zemach, G. Grafi, Characterization of *Arabidopsis thaliana* methyl-CpG-binding domain (MBD) proteins. *Plant J.* **34**, 565–572 (2003). doi:10.1046/j.1365-313X.2003.01756.x Medline

11. A. Berg, T. J. Meza, M. Mahić, T. Thorstensen, K. Kristiansen, R. B. Aalen, Ten members of the *Arabidopsis* gene family encoding methyl-CpG-binding domain proteins are transcriptionally active and at least one, AtMBD11, is crucial for normal development. *Nucleic Acids Res.* **31**, 5291–5304 (2003). doi:10.1093/nar/gkg735 Medline

12. N. M. Springer, S. M. Kaeppler, Evolutionary divergence of monocot and dicot methyl-CpG-binding domain proteins. *Plant Physiol.* **138**, 92–104 (2005). doi:10.1104/pp.105.060566 Medline

13. D. Li, A. M. S. Palanca, S. Y. Won, L. Gao, Y. Feng, A. A. Vashisht, L. Liu, Y. Zhao, X. Liu, X. Wu, S. Li, B. Le, Y. J. Kim, G. Yang, S. Li, J. Liu, J. A. Wohlschlegel, H. Guo, B. Mo, X. Chen, J. A. Law, The MBD7 complex promotes expression of methylated transgenes without significantly altering their methylation status. *eLife* **6**, e19893 (2017). doi:10.7554/eLife.19893 Medline

14. A. Zemach, Y. Li, B. Wayburn, H. Ben-Meir, V. Kiss, Y. Avivi, V. Kalchenko, S. E. Jacobsen, G. Grafi, DDM1 binds *Arabidopsis* methyl-CpG binding domain proteins and affects their subnuclear localization. *Plant Cell* **17**, 1549–1558 (2005). doi:10.1105/tpc.105.031567 Medline

15. F. Scebba, G. Bernacchia, M. De Bastiani, M. Evangelista, R. M. Cantoni, R. Cella, M. T. Locci, L. Pitto, *Arabidopsis* MBD proteins show different binding specificities and nuclear localization. *Plant Mol. Biol.* **53**, 715–731 (2003). doi:10.1023/B:PLAN.0000019118.56822.a9 Medline

16. M. Ito, A. Koike, N. Koizumi, H. Sano, Methylated DNA-binding proteins from *Arabidopsis*. *Plant Physiol.* **133**, 1747–1754 (2003). doi:10.1104/pp.103.026708 Medline

17. S. B. Preuss, P. Costa-Nunes, S. Tucker, O. Pontes, R. J. Lawrence, R. Mosher, K. D. Kasschau, J. C. Carrington, D. C. Baulcombe, W. Viegas, C. S. Pikaard, Multimegabase silencing in nucleolar dominance involves siRNA-directed DNA methylation and specific methylcytosine-binding proteins. *Mol. Cell* **32**, 673–684 (2008). doi:10.1016/j.molcel.2008.11.009 Medline

18. J. A. Law, S. E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010). doi:10.1038/nrg2719 Medline

19. Y. Liu, X. Zhang, R. M. Blumenthal, X. Cheng, A common mode of recognition for methylated CpG. *Trends Biochem. Sci.* **38**, 177–183 (2013). doi:10.1016/j.tibs.2012.12.005 Medline

20. H. W. Gabel, B. Kinde, H. Stroud, C. S. Gilbert, D. A. Harmin, N. R. Kastan, M. Hemberg, D. H. Ebert, M. E. Greenberg, Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89–93 (2015). doi:10.1038/nature14319 Medline

21. A. Bartlett, R. C. O'Malley, S. C. Huang, M. Galli, J. R. Nery, A. Gallavotti, J. R. Ecker, Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat. Protoc.* **12**, 1659–1672 (2017). doi:10.1038/nprot.2017.055 Medline

22. H. Stroud, M. V. C. Greenberg, S. Feng, Y. V. Bernatavichute, S. E. Jacobsen, Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**, 352–364 (2013). doi:10.1016/j.cell.2012.10.054 Medline

23. M. E. Potok, Y. Wang, L. Xu, Z. Zhong, W. Liu, S. Feng, B. Naranbaatar, S. Rayatpisheh, Z. Wang, J. A. Wohlschlegel, I. Ausin, S. E. Jacobsen, *Arabidopsis* SWR1-associated protein methyl-CpG-binding domain 9 is required for histone H2A.Z deposition. *Nat. Commun.* **10**, 3352 (2019). doi:10.1038/s41467-019-11291-w Medline

24. W. J. J. Soppe, S. E. Jacobsen, C. Alonso-Blanco, J. P. Jackson, T. Kakutani, M. Koornneef,

A. J. M. Peeters, The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* **6**, 791–802 (2000). doi:10.1016/S1097-2765(05)00090-0 Medline

25. H. Stroud, T. Do, J. Du, X. Zhong, S. Feng, L. Johnson, D. J. Patel, S. E. Jacobsen, Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat. Struct. Mol. Biol.* **21**, 64–72 (2014). doi:10.1038/nsmb.2735 Medline

26. A. Finka, R. U. H. Mattoo, P. Goloubinoff, Meta-analysis of heat- and chemically upregulated chaperone genes in plant and human cells. *Cell Stress Chaperones* **16**, 15–31 (2011). doi:10.1007/s12192-010-0216-8 Medline

27. V. B. V. Rajan, P. D'Silva, *Arabidopsis thaliana* J-class heat shock proteins: Cellular stress sensors. *Funct. Integr. Genomics* **9**, 433–446 (2009). doi:10.1007/s10142-009-0132-0 Medline

28. Q. Q. Zhao, R. N. Lin, L. Li, S. Chen, X. J. He, A methylated-DNA-binding complex required for plant development mediates transcriptional activation of promoter methylated genes. *J. Integr. Plant Biol.* **61**, 120–139 (2019). doi:10.1111/jipb.12767 Medline

29. L. M. Johnson, J. Du, C. J. Hale, S. Bischof, S. Feng, R. K. Chodavarapu, X. Zhong, G. Marson, M. Pellegrini, D. J. Segal, D. J. Patel, S. E. Jacobsen, SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507**, 124–128 (2014). doi:10.1038/nature12931 Medline

30. J. Gallego-Bartolomé, W. Liu, P. H. Kuo, S. Feng, B. Ghoshal, J. Gardiner, J. M.-C. Zhao, S. Y. Park, J. Chory, S. E. Jacobsen, Co-targeting RNA polymerases IV and V promotes efficient de novo DNA methylation in *Arabidopsis*. *Cell* **176**, 1068–1082.e19 (2019). doi:10.1016/j.cell.2019.01.029 Medline

31. L. Leng, Q. Liang, J. Jiang, C. Zhang, Y. Hao, X. Wang, W. Su, A subclass of HSP70s regulate development and abiotic stress responses in *Arabidopsis thaliana*. *J. Plant Res.* **130**, 349–363 (2017). doi:10.1007/s10265-016-0900-6 Medline

32. R. Rosenzweig, N. B. Nillegoda, M. P. Mayer, B. Bukau, The Hsp70 chaperone network. *Nat. Rev. Mol. Cell Biol.* **20**, 665–680 (2019). doi:10.1038/s41580-019-0133-3 Medline

33. P. Pulido, D. Leister, Novel DNAJ-related proteins in *Arabidopsis thaliana*. *New Phytol.* **217**, 480–490 (2018). doi:10.1111/nph.14827 Medline

34. L. Yan, S. Wei, Y. Wu, R. Hu, H. Li, W. Yang, Q. Xie, High-efficiency genome editing in *Arabidopsis* using YAO promoter-driven CRISPR/Cas9 system. *Mol. Plant* **8**, 1820–1823 (2015). doi:10.1016/j.molp.2015.10.004 Medline

35. E. C. Greene, S. Wind, T. Fazio, J. Gorman, M. L. Visnapuu, DNA curtains for high-throughput single-molecule optical imaging. *Methods Enzymol.* **472**, 293–315 (2010). doi:10.1016/S0076-6879(10)72006-1 Medline

36. UniProt Consortium, UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47** (D1), D506–D515 (2019). doi:10.1093/nar/gky1049 Medline

37. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990). doi:10.1016/S0022-2836(05)80360-2 Medline

38. R. Agarwala, T. Barrett, J. Beck, D. A. Benson, C. Bollin, E. Bolton, D. Bourexis, J. R. Brister, S. H. Bryant, K. Canese, M. Cavanaugh, C. Charowhas, K. Clark, I. Dondoshansky, M. Feolo, L. Fitzpatrick, K. Funk, L. Y. Geer, V. Gorelenkov, A. Graeff, W. Hlavina, B. Holmes, M. Johnson, B. Kattman, V. Khotomlianski, A. Kimchi, M. Kimelman, M. Kimura, P. Kitts, W. Klimke, A. Kotliarov, S. Krasnov, A. Kuznetsov, M. J. Landrum, D. Landsman, S. Lathrop, J. M. Lee, C. Leubsdorf, Z. Lu, T. L. Madden, A. Marchler-Bauer, A. Malheiro, P. Meric, I. Karsch-Mizrachi, A. Mnev, T. Murphy, R. Orris, J. Ostell, C. O'Sullivan, V. Palanigobu, A. R. Panchenko, L. Phan, B. Pierov, K. D. Pruitt, K. Rodarmer, E. W. Sayers, V. Schneider, C. L. Schoch, G. D. Schuler, S. T. Sherry, K. Siyan, A. Soboleva, V. Soussov, G. Starchenko, T. A. Tatusova, F. Thibaud-Nissen, K. Todorov, B. W. Trawick, D. Vakatov, M. Ward, E. Yaschenko, A. Zasypkin, K. Zbicz; NCBI Resource Coordinators, Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018). doi:10.1093/nar/gkx1095 Medline

39. F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn, R. Lopez, The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019). doi:10.1093/nar/gkz268 Medline

40. G. Moissiard, S. Bischof, D. Husmann, W. A. Pastor, C. J. Hale, L. Yen, H. Stroud, A. Papikian, A. A. Vashisht, J. A. Wohlschlegel, S. E. Jacobsen, Transcriptional gene silencing by *Arabidopsis* microrchidia homologues involves the formation of heteromers. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7474–7479 (2014). doi:10.1073/pnas.1406611111 Medline

41. J. Hetzel, S. H. Duttke, C. Benner, J. Chory, Nascent RNA sequencing reveals distinct features in plant transcription. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12316–12321 (2016). doi:10.1073/pnas.1603217113 Medline

42. V. M. Link, S. H. Duttke, H. B. Chun, I. R. Holtman, E. Westin, M. A. Hoeksema, Y. Abe, D. Skola, C. E. Romanoski, J. Tao, G. J. Fonseca, T. D. Troutman, N. J. Spann, T. Strid, M. Sakai, M. Yu, R. Hu, R. Fang, D. Metzler, B. Ren, C. K. Glass, Analysis of genetically diverse macrophages reveals local and domain-wide mechanisms that control transcription factor binding and function. *Cell* **173**, 1796–1809.e17 (2018). doi:10.1016/j.cell.2018.04.018 Medline

43. C. B. R. Villar, C. Köhler, Plant chromatin immunoprecipitation. *Methods Mol. Biol.* **655**, 401–411 (2010). doi:10.1007/978-1-60761-765-5_27 Medline

44. X. Li, C. J. Harris, Z. Zhong, W. Chen, R. Liu, B. Jia, Z. Wang, S. Li, S. E. Jacobsen, J. Du, 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.* **115**, E8793–E8802 (2018). doi:10.1073/pnas.1809841115 Medline

45. F. Krueger, S. R. Andrews, Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011). doi:10.1093/bioinformatics/btr167 Medline

46. X. Huang, S. Zhang, K. Li, J. Thimmapuram, S. Xie, ViewBS: A powerful toolkit for visualization of high-throughput bisulfite sequencing data. *Bioinformatics* **34**, 708–709

(2018). doi:10.1093/bioinformatics/btx633 Medline

47. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi:10.1038/nmeth.1923 Medline

48. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016). doi:10.1093/nar/gkw257 Medline

49. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nussbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008). doi:10.1186/gb-2008-9-9-r137 Medline

50. Y. Zhang, C. J. Harris, Q. Liu, W. Liu, I. Ausin, Y. Long, L. Xiao, L. Feng, X. Chen, Y. Xie, X. Chen, L. Zhan, S. Feng, J. J. Li, H. Wang, J. Zhai, S. E. Jacobsen, Large-scale comparative epigenomics reveals hierarchical regulation of non-CG methylation in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E1069–E1074 (2018). doi:10.1073/pnas.1716300115 Medline

51. W. Liu, S. H. Duttke, J. Hetzel, M. Groth, S. Feng, J. Gallego-Bartolome, Z. Zhong, H. Y. Kuo, Z. Wang, J. Zhai, J. Chory, S. E. Jacobsen, RNA-directed DNA methylation involves co-transcriptional small-RNA-guided slicing of polymerase V transcripts in *Arabidopsis*. *Nat. Plants* **4**, 181–188 (2018). doi:10.1038/s41477-017-0100-y Medline

52. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi:10.1093/bioinformatics/bts635 Medline

53. S. Anders, P. T. Pyl, W. Huber, HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015). doi:10.1093/bioinformatics/btu638 Medline

54. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi:10.1186/s13059-014-0550-8 Medline

55. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016). doi:10.1038/nbt.3519 Medline

56. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016). doi:10.1093/bioinformatics/btw313 Medline

57. R. J. Schmitz, M. D. Schultz, M. G. Lewsey, R. C. O'Malley, M. A. Urich, O. Libiger, N. J. Schork, J. R. Ecker, Transgenerational epigenetic instability is a source of novel methylation variants. *Science* **334**, 369–373 (2011). doi:10.1126/science.1212959 Medline

58. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). doi:10.1093/bioinformatics/btq033 Medline