# Supplemental Methods

**DNA affinity pull-down assay**

*Nuclear protein extraction from Arabidopsis cell culture*

*Arabidopsis thaliana* Landsberg erecta (Ler) ecotype cell suspensions derived from root callus was cultured in growth media (1× Linsmaier and Skoog basal salts with minimum organics, 3% (w/v) sucrose, 0.5 mg/L naphthalene acetic acid and 0.05 mg/L kinetin, adjusted to pH 5.7 with KOH) at 25°C under constant darkness and 130 rpm orbital shaking. Cultures were maintained by inoculating 20 mL of 7-day-old cells into 100 mL of fresh media in 250 mL Erlenmeyer flasks. Cultured cells were pelleted for 10 min at 500 rpm and protoplasted by enzymatic digestion in Enzyme Buffer (0.4 M Mannitol, 3% Sucrose, 8 mM $CaCl_2$, 1% Cellulase, and 0.5% Macerozyme) for 4 hours. The protoplast solution was then filtered through 70 µM nylon mesh and centrifuged for 10 min at 26 ×g after addition of 1 volume Mannitol/W5 Buffer (0.4 M Mannitol and 0.2× W5 where 1× W5 is 5 mM Glucose, 154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, and 1.5 mM MES, pH adjusted to 5.6 with 0.1 M KOH). Pelleted protoplasts were washed twice with Mannitol/Mg Buffer (0.4 M Mannitol, 0.1% MES and 15 mM $MgCl_2$, pH adjusted to 5.6 with 0.1 M KOH), resuspending the protoplasts by gentle rocking and centrifuging for 10 min at 26 × g. Pelleted protoplasts were resuspended in Nuclei Isolation Buffer (NIB: 20% Glycerol, 10 mM Tris-HCl pH7.5, 10 mM $MgCl_2$, 10 mM KCl, 0.5% Triton X-100, 10 mM β-Mercaptoethanol and 1× EDTA-free complete protease inhibitors), filtered through two sheets of miracloth and pelleted at 1500 × g for 20 min. Pellets were resuspended in NIB and centrifuged at 800 x g for 15 min. Nuclei were resuspended in Nuclei Resuspension Buffer (50% Glycerol, 50 mM HEPES-KOH pH 7.6, 100 mM NaCl, 5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT, and 1× EDTA-free complete protease inhibitors) for visualization under fluorescence microscope by DAPI staining, or nuclear proteins were extracted by resuspending the nuclei in Nuclei Extraction Buffer (20 mM Tris-HCl (pH 7.9), 420 mM KCl and 1.5 mM $MgCl_2$, 0.5 mM DTT, and 1× EDTA-free complete protease inhibitors), filtering the extract through a 12-gauge needle followed by centrifugation to pellet the membrane and debris. Protein concentration in the supernatant was measured with the Bio-Rad Protein Assay, establishing a linear regression with a range of bovine serum albumin standards.

*Probes design*

DNA probes were designed for each sequence context following the model of (Spruijt et al. 2013). For each possible 5 bp motif containing a unique cytosine in each context (i.e. in the CG context, all WCGWW motifs, where W is A or T), the number of sites in the *Arabidopsis* genome were evaluated, grouped by genomic feature, and the average methylation of the cytosine was calculated from Col-0 WT bisulfite sequencing data (Secco et al. 2015). The motifs found with higher frequency in more highly methylated annotated sequences (genes and TEs) were selected. The motifs were repeated 4 times on the probe and flanked by unmethylated sequences to create the 33-bp oligos (Table 1).

DNA probes (Table 1) were synthesized as single-stranded oligonucleotides by IDT, to be paired as follows: CG1 and CG2 constitute the methylated CG probe, and CG3 and CG4 the unmethylated CG control probe; CHG1 and CHG2 constitute the methylated CHG probe, and CHG3 and CHG4 the unmethylated CHG control probe; CHH1 and CHH2 constitute the CHH methylated probe, and CHH3 and CHH2 the unmethylated control probe.

DNA affinity pull-down assays were then performed following an adapted version of (Spruijt et al. 2013). Oligos were paired by incubating 20 µg of each oligo in NEB buffer at 95°C for 5 min and gradually cooled-down overnight. Then, 10 µL of streptadivin sepharose high-performance beads (GE Healthcare) were used to immobilize the DNA probes, verifying that the probes were indeed immobilized on the beads by agarose gel electrophoresis. Three technical replicates were performed for each probe. For each replicate, 450 µg of nuclear protein extract was incubated with the DNA probes, in presence of 10 µg of dAdT as competitor DNA, rotating for 90 min. Beads were washed twice with PBS buffer and proteins bound to the probes were digested with trypsin. Tryptic peptides were eluted in Peptide Elution Buffer (100 mM Tris-HCl pH 7.5, 2 M Urea, and 10 mM DTT) and loaded on C18 Stage-Tips pre-activated by methanol, washed once

with Solution B (0.1% Formic acid, 80% Acetonitrile) and washed twice with Solution A (0.1% Formic acid). Samples were then washed once with Solution A.

**Table 1. DNA probe sequences and modifications**

| | |
|---|---|
| CG1 | 5'-Biotin-GATGATGTmCGATTmCGATTmCGATTmCGATATGATG-3' |
| CG2 | 5'-CATCATATmCGAATmCGAATmCGAATmCGACATCATC-3' |
| CG3 | 5'-Biotin-GATGATGTCGATTCGATTCGATTCGATATGATG-3' |
| CG4 | 5'-CATCATATCGAATCGAATCGAATCGACATCATC-3' |
| CHG1 | 5'-Biotin-GATGATGTmCTGATmCTGATmCTGATmCTGAATGATG-3' |
| CHG2 | 5'-CATCATTmCAGATmCAGATmCAGATmCAGACATCATC-3' |
| CHG3 | 5'-Biotin-GATGATGTCTGATCTGATCTGATCTGAATGATG-3' |
| CHG4 | 5'-CATCATTCAGATCAGATCAGATCAGACATCATC-3' |
| CHH1 | 5'-Biotin-GATGATGTmCAATTmCAATTmCAATTmCAATATGATG-3' |
| CHH2 | 5'-CATCATATTGAATTGAATTGAATTGACATCATC-3' |
| CHH3 | 5'-Biotin- GATGATGTCAATTCAATTCAATTCAATATGATG-3' |

*LC-MS/MS measurements and statistical analysis*
Peptides were eluted from Stage-Tips in Solution B, after which the acetonitrile was evaporated using a vacuum concentrator. The samples were resuspended in 12 µl of Solution A, of which 5 µl were loaded onto a 30 cm column packed with 1.8 µm Reprosil-Pur C18-AQ (Dr Maisch GmbH). A 114 min gradient of acetonitrile (7%-32%), followed by washes at 50% then 90% acetonitrile, was employed to elute the peptides from the column, using an Easy-nLC 1000 system (Thermo Fisher Scientific). Eluted peptides were sprayed directly into an LTQ-Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific). Scans were collected in data-dependent top-speed mode of a 3-second cycle with dynamic exclusion set at 60 sec, for 140 min of total data collection.

Measured peptides were searched against the UniProt *Arabidopsis thaliana* proteome (version 2014-09-03) with MaxQuant version 1.5.1.0 (Cox and Mann 2008). Default settings were used, with additional options for 'match between runs', 'label free quantification (LFQ)' and 'intensity based absolute quantification (iBAQ)' enabled. Data were analysed with Perseus version 1.4.0.0 and figures were generated using in-house R scripts (Supplemental Code). Reverse and contaminant hits were removed, and the LFQ columns were transformed into $\log_2$ scale. Resulting data were filtered for proteins with three valid values in at least one of the samples, after which missing values were imputed by semi-random, low values (width = 0.3, shift = 1.8). To generate volcano plots between two samples, statistical outliers were determined using a two-tailed *t*-test with a permutation-based false discovery rate (FDR). The following different FDR and s0 (similar to a minimal fold-change) cut-offs were used to limit the number of proteins enriched in each context: CG: FDR=0.005, s0=3.0; CHG: FDR=0.025, s0=5.0; CHH: FDR=0.05, s0=4.0. To perform hierarchical clustering, an ANOVA was executed with all the samples, also with a permutation-based FDR. Insignificant proteins were discarded, then the median of each row was subtracted to get deviations from the median. Statistical analysis and original plots were generated with Perseus software (Tyanova et al. 2016). R statistical software (R Core Team 2021) was used to generate the final plots.

**DNA Affinity Purification Sequencing (DAP-seq)**
DAP-seq has been performed according to the original protocol (Bartlett et al. 2017). Genomic DNA libraries were created using genomic DNA extracted from 2-week old Col-0 seedlings grown on ½ Murashige and Skoog media supplemented with 1% sucrose with DNeasy Plant Mini kit (Qiagen). Plasmids containing the proteins of interest in the pIX-HALO backbone were ordered from ABRC, and an empty pIX-HALO plasmid was used as negative control (Table 2).

**Table 2. Plasmids containing the proteins of interest in the pIX-HALO backbone**

| | |
|---|---|
| MBD1 | HALO_TF-13_H08 |
| MBD2 | HALO_TF-18_D10 |
| MBD4 | HALO_TF-18_H12 |
| MBD5 | HALO_TF-13_B06 |
| MBD6 | HALO_TF-19_D09 |
| SUVH1 | HALO_TF-20_B11 |
| SUVH3 | HALO_TF-19_D07 |
| HDG11 | HALO_SFI_15-D12 |
| Empty | CD3-1742 |

Proteins were expressed using the TnT Coupled Wheat Germ Extract System with SP6 promoter (Promega). Protein expression was confirmed by Western Blot. After incubation of the proteins with DAP and ampDAP libraries, DAP-seq and ampDAP-seq libraries were pooled, separated by agarose gel electrophoresis, and the 200 to 400 bp range was purified by gel extraction using Isolate II PCR and Gel Kit (Bioline). Libraries were sequenced (101 or 121 bp single-end) on a HiSeq 1500 instrument (Illumina) according to the manufacturer's instructions. Raw sequencing data were then de-multiplexed with bcl2fastq software (Illumina).

**Plant material**

*Collection of single and higher order mutants*

All plants were *Arabidopsis thaliana* accession Col-0, grown on soil at 22°C in 16h light/8h dark cycles. For SUVH1, SUVH3, MBD1, MBD2, and MBD6, existing T-DNA lines were ordered from the Arabidopsis Biological Resource Center (Supplemental Table S4). For MBD4 and MBD5, single mutants were created by CRISPR-Cas9 mutagenesis. Guide RNAs were designed for each target using CRISPRdirect (Naito et al. 2015) and further analyzed with E-CRISPR (Heigwer et al. 2014). The cassettes containing two gRNAs for each mutant pair were ordered as gBlocks (IDT) and were inserted into pHEE2E backbone, provided by Dr Qi-Jun Chen (Wang et al. 2015), by restriction-ligation using BsaI (NEB). WT plants were transformed by Agrobacterium-mediated T-DNA insertion using the floral dipping procedure (Clough and Bent 1998). Plants transformed with the CRISPR cassette were selected by resistance to hygromycin. Genomic DNA was extracted from leaf tissue (Edwards et al. 1991). For T-DNA lines, genotyping was performed via PCR amplification of the inserted T-DNA sequence. The primers used for genotyping PCRs were designed with the iSect tool from the Salk Institute Genomic Analysis Laboratory and are listed in (Supplemental Table S7). Plants homozygous for the insertion were selected for subsequent experiments.

For CRISPR-Cas9 mutated plants, homozygous plants were determined by the presence of a single trace including a mutation after Sanger sequencing (primers used are listed in Supplemental Table S7). Only homozygous T2 plants that were selected for the absence of the CRISPR cassette by PCR were selected for subsequent experiments. Higher order mutants were generated by crossing the wanted mutations (Supplemental Table S4).

*Complementation of mutants with epitope-tagged proteins under endogenous promoter*

For each candidate, the single mutant line from T-DNA or CRISPR-Cas9 described above was complemented with an epitope-tagged version of the respective protein driven by their endogenous promoter. Promoters were defined by incorporating the closest DNase I hypersensitivity (DHS) sites upstream of the transcription start site, from DHS data generated from 7-day old seedlings in Col-0 background (Sullivan et al. 2014). 3'-UTRs were included for MBD6, as a DHS site was also present downstream, near the transcription termination site. Promoters and 3' UTRs were amplified from Col-0 genomic DNA using the primers listed in (Supplemental Table S3) with Q5 Hot Start High-fidelity DNA

polymerase (NEB). For MBD5 and SUVH3, Gibson Assembly was performed to insert these fragments into the backbone. For the others, PmeI (or SacI) and AvrII restriction sites were incorporated at the 5' end of the primers to allow for restriction-ligation of the promoters into plasmids containing the SHH tag (2×Strep-HA-6×His), generously provided by Dr. Dmitri Nusinow. Insert and entry plasmids were digested with PmeI (or SacI) and AvrII enzymes (NEB). Generated fragments were gel extracted, ligated and purified ligation products were electroporated in TOP10 *Escherichia coli* competent cells. Transformed bacteria were selected by resistance to spectinomycin and plasmid sequences were confirmed by Sanger sequencing. For MBD6, 3'-UTRs were then incorporated through a similar process using EcoRV restriction sites. Final plasmids were inserted into GV3101 *Agrobacterium tumefaciens* by heat-shock for plant transformation. T1 plants were grown on soil and transformants selected by resistance to glufosinate. Presence of both the original mutation and the complementation plasmid was confirmed by PCR using primers listed in Supplemental Table S7. To assess the expression level of the complemented plasmids, RNA was extracted from a population of T2 seedlings grown on ½ Murashige and Skoog media supplemented with 1% sucrose for two weeks with TRIzol Reagent (ThermoFisher Scientific). RNA was treated with RQ1 DNase (Promega), purified and converted to cDNA with sensiFAST cDNA Synthesis Kit (Bioline), using the provided mix of random hexamers and anchored Oligo(dT) primers. qPCR reactions were performed with KAPA SYBR FAST qPCR Kit (Sigma-Aldrich) on a LightCycler 480 instrument (Roche). The qPCR primers used for this study are listed in Supplemental Table 7.

**Chromatin immunoprecipitation sequencing (ChIP-seq)**

Three to eight grams of seedlings grown on ½ Murashige and Skoog media supplemented with 1% sucrose were harvested 14-days after germination induction and crosslinked in Crosslinking Buffer (10mM HEPES-NaOH, pH 7.4; 1% Formaldehyde) by drawing vacuum for 10 min, twice. Seedlings were transferred to Quenching Buffer (10mM HEPES-NAOH, pH 7.4; 200 nM Glycine) and vacuum applied for 10 min. After 3 washes in $ddH_2O$, seedlings were frozen in liquid $N_2$ and stored at -80°C. Tissue was grounded to fine powder using a mortar and pestle in liquid $N_2$ and resuspended in Buffer A (10mM Tris-HCl, pH 8.0; 400mM Sucrose; 5mM β-Mercaptoethanol; 1× Protease inhibitors), rotating for 10 min. Samples were kept at 4°C from this step up to the reverse crosslinking step, including during the centrifugation steps. After filtering through a sheet of miracloth, samples were centrifuged for 20 min at 2,880 × g. Pellets were resuspended in Buffer B (10 mM Tris-HCl, pH 8.0; 250 mM Sucrose; 10 mM $MgCl_2$; 1% Triton X-100; 5 mM β-Mercaptoethanol; 1× Protease inhibitors) and centrifuged for 10 min at 12,000 × g. Pellets were then resuspended in Buffer C (10 mM Tris-HCl, pH8.0; 1.7 M Sucrose; 2 mM $MgCl_2$; 0.15% Triton X-100; 5 mM B-Mercaptoethanol; 1× Protease inhibitors) and overlaid on Buffer C cushion before centrifuging for 1h at 16,000 × g. Pellets were washed twice in Wash Buffer (10 mM Tris-HCl, pH 8.0; 200 mM NaCl; 1 mM EDTA, pH 8.0; 0.5 mM EGTA, pH 8.0; 1× Protease inhibitors) by resuspending in the buffer and centrifuging for 5 min at 4,000 × g. Final pellets were resuspended in Shearing Buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA, pH 8.0; 0.1% SDS; 1× Protease inhibitors) and sonicated with Covaris S2 (12 min, duty cycle 5%, intensity 4, 200 cycles per burst, peak power 140). Triton X-100 and NaCl were added to the samples to reach a final concentration of 1% and 150mM, respectively. Chromatin was then cleared of debris by centrifuging for 10 min at 10,000 × g.

Chromatin (50 µL; 5%) was put aside to be used as input control. Then 4 µg of primary antibodies (αHA antibody: Biolegend, #901502; αH2A.Z antibody: abcam, ab4174; αH2AK121ub antibody: Cell Signaling Technology, 8240S; αH3K4me1 antibody: abcam, ab8895; αH3K4me2 antibody: abcam, ab32356; αH3K4me3 antibody: abcam, ab8580; αH3K36me3 antibody: abcam, ab9050; αH3K27me3 antibody: abcam, ab6002; αH3K27ac antibody: abcam, ab4729; αH3K9me2 antibody: abcam, ab1220; αH3 antibody: abcam, ab1791) were added to 1 mL of chromatin and incubated slowly rotating overnight. For each sample, 25 µL Protein G beads (Life Technologies) and 25 µL Dynabeads M-280 sheep anti-mouse IgG (Thermo Fischer Scientific) were equilibrated by washing three times with IP Buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA, pH 8.0; 0.1% SDS; 1% Triton X-100; 150 mM NaCl) prior to add the chromatin, and samples were incubated rotating for 90 min. Protein/DNA complexes bound to the beads were washed twice with Low Salt Wash Buffer (20 mM Hepes-KOH, pH 7.9; 2 mM EDTA; 0.1% SDS; 1% Triton X-100; 150 mM NaCl),

twice with High Salt Wash Buffer (20 mM Hepes-KOH, pH 7.9; 2 mM EDTA; 0.1% SDS; 1% Triton X-100; 500 mM NaCl), once with LiCl Wash Buffer (100 mM Tris-HCl, pH 7.5; 0.5 M LiCl; 1% NP-40; 1% Sodium Deoxycholate) and once with TE/10 Buffer (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA). Immunoprecipitated samples were resuspended in 49 µL Proteinase K Digestion Buffer (20 mM HEPES, pH 7.9; 1 mM EDTA; 0.5% SDS) and 1 µL of 20 mg/mL Proteinase K was added to each sample. In parallel, 2 µL of 10% SDS and 1 µL of 20 mg/mL Proteinase K was added to the input samples. Both types of samples were then incubated at 50°C for 15 min. DNA/protein complexes eluted from the beads were then transferred to a new tube. Subsequently 3 µL 5M NaCl and 0.5 µL 100 mg/mL RNase A were added to immunoprecipitated and input samples and were incubated overnight at 65°C shaking at 1,000 rpm to reverse the cross-linking. The next day, an additional 1.5 µL 20 mg/mL Proteinase K were added to all samples and samples were incubated for 60 min at 50°C. DNA was finally purified with AMPure beads and 15 µL of each sample were recovered. Input DNA (2 ng) and immunoprecipitated samples (5 µL) were processed according to Rubicon thruPLEX DNA-seq kit, using half reactions per sample. Libraries were sequenced on a HiSeq 1500 or NextSeq 500 instrument (Illumina) according to the manufacturer's instructions. Raw sequencing data were then de-multiplexed with bcl2fastq software (Illumina).

**Affinity purification coupled to mass spectrometry**

Three to eight grams of seedlings grown on ½ Murashige and Skoog media supplemented with 1% sucrose were harvested 14-days after germination induction, frozen in liquid $N_2$ and stored at -80°C. Tissue was grounded to fine powder using a mortar and pestle in liquid $N_2$ and resuspended in Extraction Buffer (2 M Hexylene glycol; 20 mM PIPES-KOH, pH 7.0; 10 mM $MgCl_2$; 5 mM β-Mercaptoethanol), rotating for 10 min. Samples were kept at 4°C from this step until the trypsin digestion, including during the centrifugation steps. After filtering through a sheet of miracloth, Triton X-100 was added in small aliquots until the final concentration reached 1%. Samples were overlaid on a density gradient of 30% and 80% Percoll Buffer (2 M Hexylene glycol; 5 mM PIPES-KOH, pH 7.0; 10 mM $MgCl_2$; 1% Triton X-100; 5 mM β-Mercaptoethanol; 30 or 80% Percoll) and centrifuged for 30 min at 2,000 × g. Nuclei were collected from the 30/80% percoll interphase, underlaid with 30% Percoll Buffer and centrifuged for 10 min at 2,000 × g. Pelleted nuclei were resuspended in SII Buffer (100 mM Sodium Phosphate, pH 8.0; 150 mM NaCl; 5 mM EDTA; 5 mM EGTA; 0.1% Triton X-100; 1× Protease inhibitors; 1× Phosphatase inhibitors; 1 mM PMSF; 50 µM MG-132), and membranes were disrupted by sonication (40% power, 1s on/1s off, for 20s in total, repeated 3 times per sample). Samples were clarified by two successive centrifugation steps at 14,000 × g for 10 min.

For the experimental Set1, IBA MagStrep "type3" XT beads (Fisher Biotech) were washed twice in SII Buffer before adding the protein extract and were then incubated rotating for 60 min. The beads were washed twice in SII Buffer and three times with Strep-to-His Buffer (100 mM $Na_2HPO_4/NaH_2PO_4$, pH 8.0; 150 mM NaCl; 0.05% Triton X-100). Protein complexes were eluted twice with Strep Elution Buffer (100 mM $Na_2HPO_4/NaH_2PO_4$, pH 8.0; 150 mM NaCl; 0.05% Triton X-100; 50 mM Biotin) by incubating rotating for 10 min. Dynabeads His-Tag Isolation and Pulldown (Thermo Fisher Scientific) were washed twice in Strep-to-His Buffer before adding the eluted proteins and incubating while rotating for 20 min. The beads were washed twice with Strep-to-His Buffer and three times with 25 mM ammonium bicarbonate before being snap-frozen in liquid $N_2$ and stored at -80°C. Beads were resuspended in 50 µL Protein Elution Buffer (100 mM Tris-HCl, pH 7.5; 1 M Urea; 10 mM DTT) and incubated for 20 min, shaking. 5 µL 0.55 M Iodoacetamide were added to each sample and the tubes were incubated for 10 min, shaking in the dark. 2.5 µL of trypsin (0.4 µg/µL in 0.01% TFA) were added to each sample and the tubes were incubated for 2h, shaking. After transferring the eluted peptides to a new tube by immobilizing the beads with a magnet, 50 µL were added to the beads for a second elution, after incubating for 5 min shaking. Then 1 µL of trypsin was added to the combined eluates and the tubes were incubated while shaking overnight. Peptides were loaded on C18 resin MicroSpin Columns Silica C18 (The Nest Group) which were pre-activated by methanol, washed once with Buffer B (0.1% formic acid, 80% acetonitrile) and washed twice with Buffer A (0.1% formic acid). Samples were then washed once with Buffer A and eluted from the Stage-Tips in Buffer B, after which the acetonitrile was evaporated using a vacuum concentrator. Samples were resuspended in

30 µl of 5% (v/v) acetonitrile/0.1% (v/v) formic acid before online reversed phase nanoflow (EASY-Spray HPLC column 50 cm × 75 µm ID) ESI coupled to an Orbitrap Fusion (ThermoFisher). Gradients were 5–35% (v/v) acetonitrile in 0.1% (v/v) formic acid (250 nl min−1) formed by a Dionex UltiMate 3000 series HPLC (ThermoFisher) over 120 min. Spectra were acquired in data dependent mode with a 120k resolution survey scan from 300–1500 m/z followed by selection of the eight most abundant doubly or triply charged ions for MS/MS analysis. Ions were dynamically excluded for 60 sec. Raw data was converted to mzML format by msconvert (3.0.9992) before spectral matching against the Araport11 peptide sequence database by CometMS (2016.01 rev. 2). Results were processed through the Trans-Proteomic Pipeline (5.0) tools peptide and protein prophet. Proteins with a probability above 0.9 were accepted for further analysis.
For the experimental Set2, single strep IP protein pull-downs using 200 µg input lysate were also performed. We note that the MS/MS ID rates for these protein pull-downs are very low, for reasons we have not been able to decipher.

**Whole genome bisulfite sequencing (MethylC-seq)**
Genomic DNA was extracted from 2-week old seedlings with DNeasy Plant Mini Kit (Qiagen) and further purified with Isolate II Gel and PCR Clean-up Kit (Bioline). Libraries were then generated from fragmented DNA (Covaris, 250 bp) with NxSeq AmpFREE Low DNA Library Kits (Lucigen). After size selection by AMPureXP beads, bisulfite conversion was performed with EZ DNA Methylation Gold kit (Zymo Research). Libraries were amplified with Kapa HiFi HotStart Uracil+ ReadyMix (Roche) and purified with AMPureXP beads. MethylC-seq libraries were sequenced on HiSeq 1500 platform (Illumina) using single-end 100 bp format, according to the manufacturer's instructions.

**High-throughput RNA sequencing (RNA-seq)**
Three biological replicates were performed in parallel for all genotypes from populations of 2-week old seedlings grown on ½ Murashige and Skoog media supplemented with 1% sucrose. Total RNA was extracted with RNeasy Plant Mini Kit (Qiagen) and treated with RQ1 DNase (Promega). Libraries were then generated with TruSeq Stranded Total RNA Library Prep Kit (Illumina), after depletion of ribosomal RNAs with Ribo-Zero rRNA Removal Kit Plant (Illumina). RNA-seq libraries were sequenced on a NextSeq 500 (Illumina) using paired-end 42 bp format, according to the manufacturer's instructions.

**Global run-on sequencing (GRO-seq)**
Global Run-On sequencing (GRO-seq) and 5'-GRO-seq were performed on 10-20g of 2-week old seedlings (~10 plates), according to the published protocol (Hetzel et al. 2016).

**Data analysis**
*mC reader ChIP-seq and (amp)DAP-seq*
Illumina adapters were trimmed from the raw data using cutadapt (Martin 2011) with default parameters. Remaining reads were mapped to the *Arabidopsis* TAIR10 genome reference using Bowtie 2 in default end-to-end mode (Langmead and Salzberg 2012). Resulting files were converted to sorted and indexed BAM files using the SAMtools suite (Li et al. 2009). Peaks were called with the version 2 of MACS (Zhang et al. 2008) by comparing each immunoprecipitated sample to its input, and bigWig files were generated after normalization to input, including for a WT sample (Col-0 plant with no tagged protein). For DAP and ampDAP-seq, the samples were compared to the DNA control library before and after amplification, respectively.
For each ChIP sample, peaks with 10% reciprocal overlap with the peaks called in WT were discarded, and to increase stringency, only the remaining top 80% based on local fold-change constitute the final set of peaks for each sample. These sets were merged into a single file by merging 50-bp overlapping peaks to generate a list of unique bound loci, using `bedmap -echo-map-range -echo-map-id-uniq` (Neph et al. 2012) and BEDtools merge (Quinlan and Hall). These loci were annotated with annotatePeaks.pl from the Homer suite (Heinz et al. 2010), and then intersected with TAIR10 annotated transposable elements and transposable element genes with BEDtools intersect. The deepTools suite

(Ramírez et al. 2016) was used to convert BAM files to bigWig files and to plot the heatmaps and profiles. The other figures were generated using R code based on the ggplot2 package (Wickham 2016) and UpSetR package (Conway et al. 2017). The homer suite was used to identify motifs under the ChIP-seq peaks (Heinz et al. 2010). Read mapping statistics can be found in Supplemental Table S8.
For controls (Fig 2A-C), MBD1 peaks were randomly shuffled in the genome with `bedtools shuffle -noOverlapping` to create random regions of comparable size distribution.
To generate the clusters in Fig 2, the signal of all mC readers were plotted as a heatmap, centered on the peaks from each mC reader individually, and splitting them in up to 4 clusters by kmeans. The clusters that appeared visually bound by the same set of proteins were then combined, and the peaks within each cluster were merged into non-overlapping regions, which generated the five clusters with distinct binding profiles. For Fig 3, the unique list of protein-coding genes (Fig 3A) or TE and TE genes (Fig 3B) that intersected at least one of the peaks in each cluster from Fig 2 were selected for plotting.

*Histone modification ChIP-seq*
Illumina adapters were trimmed from the raw data using cutadapt (Martin 2011) with default parameters. Remaining reads were mapped to the *Arabidopsis* TAIR10 genome using Bowtie 2 in default end-to-end mode (Langmead and Salzberg 2012). Resulting files were converted to sorted and indexed BAM files using the SAMtools suite (Li et al. 2009). Peaks were called with the version 2 of MACS (Zhang et al. 2008) by comparing each immunoprecipitated sample to its input. Peaks called for each histone modification from each genotype (WT and mC reader mutants) were merged with the following command: `cat WT_Mark.narrowPeak MutantGenotype_Mark.narrowPeak | sort -k1,1 -k2,2n | bedtools merge`, with mark being the histone modification in question, and MutantGenotype being whatever mutants were used to study alterations in that histone modification. Then edgeR (Robinson et al. 2010; Nikolayeva and Robinson 2014) was used to test for differential binding between WT and each mC mutant for each histone modification from ChIP-seq data, as previously done (Kim et al. 2023), using the merged histone peak files as a reference. Read mapping statistics can be found in Supplemental Table S8.

*MethylC-seq*
Raw sequencing data were base-called and de-multiplexed with the bcl2fastq software (Illumina). FASTQ files were mapped to the TAIR10 genome (previously processed to a 3-letter genome reference) with BS-Seeker2 (Guo et al. 2013). Processing by BS-Seeker2 includes the removal of sequencing adapters. Methylation was called with BS-Seeker2. DNA methylation heatmaps over genes and transposable elements were generated with the deepTools suite (Ramírez et al. 2016), by generating bigWig files for each sample and computing matrix with BED containing TAIR10 genic regions. Differentially methylated regions (DMRs) were identified by HOME (Srivastava et al. 2019), available on GitHub (https://github.com/Akanksha2511/HOME). DMRs were identified in CG and non-CG contexts, requiring a change in methylation levels of at least 0.1 over the whole region and an average coverage for all cytosines greater than three for both mutant and control. All annotated genes in TAIR10 were split into 3 categories based on their DNA methylation levels: genes with >2% mCHG and mCHH were classified as TE-like methylation; in the remaining list, genes with >5% mCG were labeled as gbM genes; and the rest are unmethylated genes (Supplemental Table S9). Read mapping statistics can be found in Supplemental Table S8.

*RNA-seq for differentially expressed genes and differential alternative splicing*
Raw sequencing data were de-multiplexed with bcl2fastq software (Illumina). FASTQ files were mapped to the TAIR10 genome using STAR with default parameters (Dobin et al. 2013). Reads underwent lightweight alignment using Salmon version 1.4.0 (Patro et al. 2017) to the TAIR10 transcriptome and reads from different transcripts originating from the same gene were combined using the R package tximport (Soneson et al. 2016) before being imported into the R package DESeq2 (Love et al. 2014) for normalisation and testing of differentially expressed genes (DEGs). For comparison of our *suvh1;3* differentially expressed

genes and *ros1* differentially expressed genes, RNA-seq samples from (Kim et al. 2019) were downloaded and analysed as described above. Read mapping statistics can be found in Supplemental Table S8.
To look for novel splicing changes that occurred within the mC reader mutants, the reads mapped with STAR (Dobin et al. 2013) were processed by StringTie and merged together (Pertea et al. 2015) into a master novel transcriptome comprising splicing events from TAIR10 and ones uniquely identified within this study. Then reads underwent lightweight alignment using Salmon version 1.4.0 (Patro et al. 2017) against the novel transcriptome from StringTie. Novel and known transcripts belonging to the same gene were analysed for splicing events by SUPPA2 (Trincado et al. 2018). Differential alternative splicing (DAS) was calculated for each event based on abundance of transcripts with and without inclusion of those events by SUPPA2 (Trincado et al. 2018). SUPPA2 abstracts away the problem of linking each read to a specific junction and takes in transcript level expression as transcripts per million (TPM) of inclusion vs inclusion and exclusion events to calculate PSI (Trincado et al. 2018). SUPPA2 allows for robust DAS analysis even from relatively few reads, according to tests with simulated data (Trincado et al. 2018). To confirm that this was robust with actual RNA-seq data, we subset our WT replicate 1 data down from 53 million paired-end reads to two mutually exclusive sets of 10 million and found very high correlation (Spearman's correlation coefficient = 0.997). This was repeated three more times, each yielding a correlation coefficient of between 0.996 and 0.998.

*GRO-seq*
Illumina adapters were trimmed from the raw data using cutadapt (Martin 2011) with settings for trimming of sRNA libraries that were used in GRO-seq: `-a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -m 10`. Ribosomal RNA reads were filtered from the BAM file with RSeQC Python package. Reads were then mapped to the TAIR10 genome using STAR with default parameters (Dobin et al. 2013). For 5' GRO-seq, HOMER with `-style tss` was used to define regions of transcriptional activity, meanwhile for standard GRO-seq, HOMER with `-style groseq` was used to define regions of transcriptional activity (Wang et al. 2011). Differential transcription of these transcribed regions was calculated by edgeR (Robinson et al. 2010; Nikolayeva and Robinson 2014), as done previously (Chae et al. 2015).

*Chromatin state analysis*
Chromatin states were generated with ChromHMM (Ernst and Kellis 2017), using the ChIP-seq, RNA-seq, DNA methylation and histone modifications in WT generated in this study. Analyses were done on the TAIR10 genome and annotation files from Araport11_Release_201606 (Araport11_GFF3_genes_transposons.201606.gff). For ChIP and RNA-seq, BAM files were binarized using the `ChromHMM.jar BinarizeBam` command. DNA methylation tracks were generated with `BinarizeBed`, keeping cytosines with more than 50% methylation levels. The binarized files were then merged with `MergeBinary` and the models were created with `LearnModel`. Models were compared with `Reorder` and `CompareModels`, and the 20-state model was chosen for representation. Enrichment plots were generated with ChromHMM.jar `OverlapEnrichment`. All details about the parameters, the model and the final genome segmentation are available in Supplemental Table S10.

*Affinity Purification-MS*
We filtered out the contaminant proteins listed in (Van Leene et al. 2015), since they are common contaminating proteins from TAP-MS experiments in *Arabidopsis*. One experiment was performed as a single replicate (Set 1) with tandem affinity purification (TAP) with Strep and His tags, and one experiment was performed in three replicates (Set 2) with affinity purification (AP) with just the Strep tag, each followed by mass spectrometry (MS) identification of peptide fragments. After MS, the bait proteins in each experiment were the most abundant protein by peptide count (Supplemental Table S5), as expected. We therefore normalised the peptides of other proteins as a percentage of those identified for the bait protein (Supplemental Table S5), to find the most enriched proteins, which are the most likely to be true interactors. For comparison to the literature, we extracted protein-protein interactors listed in the following publications

and then searched for all of these interactions that were supported by at least one peptide within either of our (T)AP-MS datasets: (Li et al. 2015, 2017; Harris et al. 2018; Zhao et al. 2019; Feng et al. 2021; Ichino et al. 2021; Zhou et al. 2021). The rules for inclusion in our network of mC reader protein-protein interactions were:

1. If above threshold (15%) in Set1 experiment and Set2 combined replicates.
2. If above a stringent threshold (35%) in combined Set2 replicates (and over 30% in at least 2/3 individual replicates).
3. If found in other published studies as an interactor and peptides were identified in either Set1 experiment or Set2 experiment.

# Supplemental References


Bartlett A, O'Malley RC, Huang S-SC, Galli M, Nery JR, Gallavotti A, Ecker JR. 2017. Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat Protoc* 12: 1659–1672.

Chae M, Danko CG, Kraus WL. 2015. groHMM: a computational tool for identifying unannotated and cell type-specific transcription units from global run-on sequencing data. *BMC Bioinformatics* 16: 222.

Clough SJ, Bent AF. 1998. Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* 16: 735–743.

Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33: 2938–2940.

Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367–1372.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.

Edwards K, Johnstone C, Thompson C. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res* 19: 1349.

Ernst J, Kellis M. 2017. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 12: 2478–2492.

Feng C, Cai X-W, Su Y-N, Li L, Chen S, He X-J. 2021. Arabidopsis RPD3-like histone deacetylases form multiple complexes involved in stress response. *J Genet Genomics* 48: 369–383.

Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, Chen P-Y, Pellegrini M. 2013. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* 14: 774.

Harris CJ, Scheibe M, Wongpalee SP, Liu W, Cornett EM, Vaughan RM, Li X, Chen W, Xue Y, Zhong Z, et al. 2018. A DNA methylation reader complex that enhances gene transcription. *Science* 362: 1182–1186.

Heigwer F, Kerr G, Boutros M. 2014. E-CRISP: fast CRISPR target site identification. *Nat Methods* 11: 122–123.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38: 576–589.

Hetzel J, Duttke SH, Benner C, Chory J. 2016. Nascent RNA sequencing reveals distinct features in plant transcription. *Proceedings of the National Academy of Sciences* 113: 12316–12321.

Ichino L, Boone BA, Strauskulage L, Harris CJ, Kaur G, Gladstone MA, Tan M, Feng S, Jami-Alahmadi Y, Duttke SH, et al. 2021. MBD5 and MBD6 couple DNA methylation to gene silencing through the J-domain protein SILENZIO. *Science*. http://dx.doi.org/10.1126/science.abg6130.

Kim H-M, Kang B, Park S, Park H, Kim CJ, Lee H, Yoo M, Kweon M-N, Im S-H, Kim TI, et al. 2023. Forkhead box protein D2 suppresses colorectal cancer by reprogramming enhancer interactions. *Nucleic Acids Res* 51: 6143–6155.

Kim J-S, Lim JY, Shin H, Kim B-G, Yoo S-D, Kim WT, Huh JH. 2019. ROS1-Dependent DNA Demethylation Is Required for ABA-Inducible NIC3 Expression. *Plant Physiol* 179: 1810–1821.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–359.

Li D, Palanca AMS, Won SY, Gao L, Feng Y, Vashisht AA, Liu L, Zhao Y, Liu X, Wu X, et al. 2017. The MBD7 complex promotes expression of methylated transgenes without significantly altering their

methylation status. *Elife* 6: e19893.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.
Li Q, Wang X, Sun H, Zeng J, Cao Z, Li Y, Qian W. 2015. Regulation of active DNA demethylation by a methyl-CpG-binding domain protein in Arabidopsis thaliana. *PLoS Genet* 11: e1005210.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17: 10–12.
Naito Y, Hino K, Bono H, Ui-Tei K. 2015. CRISPRdirect: Software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* 31: 1120–1123.
Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, et al. 2012. BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28: 1919–1920.
Nikolayeva O, Robinson MD. 2014. edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. *Methods Mol Biol* 1150: 45–79.
Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14: 417–419.
Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33: 290–295.
Quinlan AR, Hall IM. The BEDTools manual. *Genome*. http://waxmanlabvm.bu.edu/gracia/segex_files/www/DASR_2011/DASR_help/BEDTools-User-Manual.v4.pdf.
Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44: W160–5.
R Core Team. 2021. R: A Language and Environment for Statistical Computing. https://www.R-project.org/.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.
Secco D, Wang C, Shou H, Schultz MD, Chiarenza S, Nussaume L, Ecker JR, Whelan J, Lister R. 2015. Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *Elife* 4: e09343.
Soneson C, Love MI, Robinson MD. 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 4: 1521.
Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, Munzel M, Wagner M, Muller M, Khan F, et al. 2013. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152: 1146–1159.
Srivastava A, Karpievitch YV, Eichten SR, Borevitz JO, Lister R. 2019. HOME: a histogram based machine learning approach for effective identification of differentially methylated regions. *BMC Bioinformatics* 20: 253.
Sullivan AM, Arsovski AA, Lempe J, Bubb KL, Weirauch MT, Sabo PJ, Sandstrom R, Thurman RE, Neph S, Reynolds AP, et al. 2014. Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. *Cell Rep* 8: 2015–2030.
Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, Eyras E. 2018. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol* 19: 40.
Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13: 731–740.
Van Leene J, Eeckhout D, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Vercruysse L, Dedecker M, Verkest A, Vandepoele K, et al. 2015. An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes. *Nat Protoc* 10: 169–187.
Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, et al. 2011. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474: 390–394.
Wang Z-P, Xing H-L, Dong L, Zhang H-Y, Han C-Y, Wang X-C, Chen Q-J. 2015. Egg cell-specific

promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biol* 16: 144.

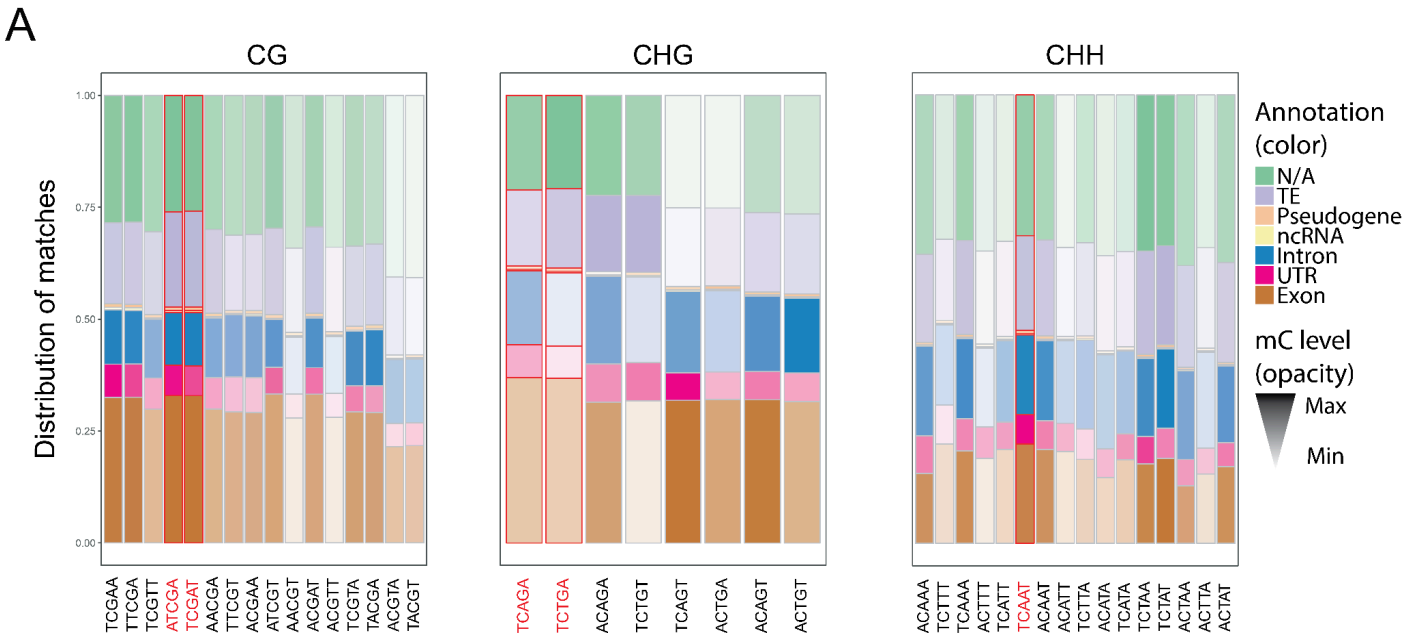
Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9: R137.

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

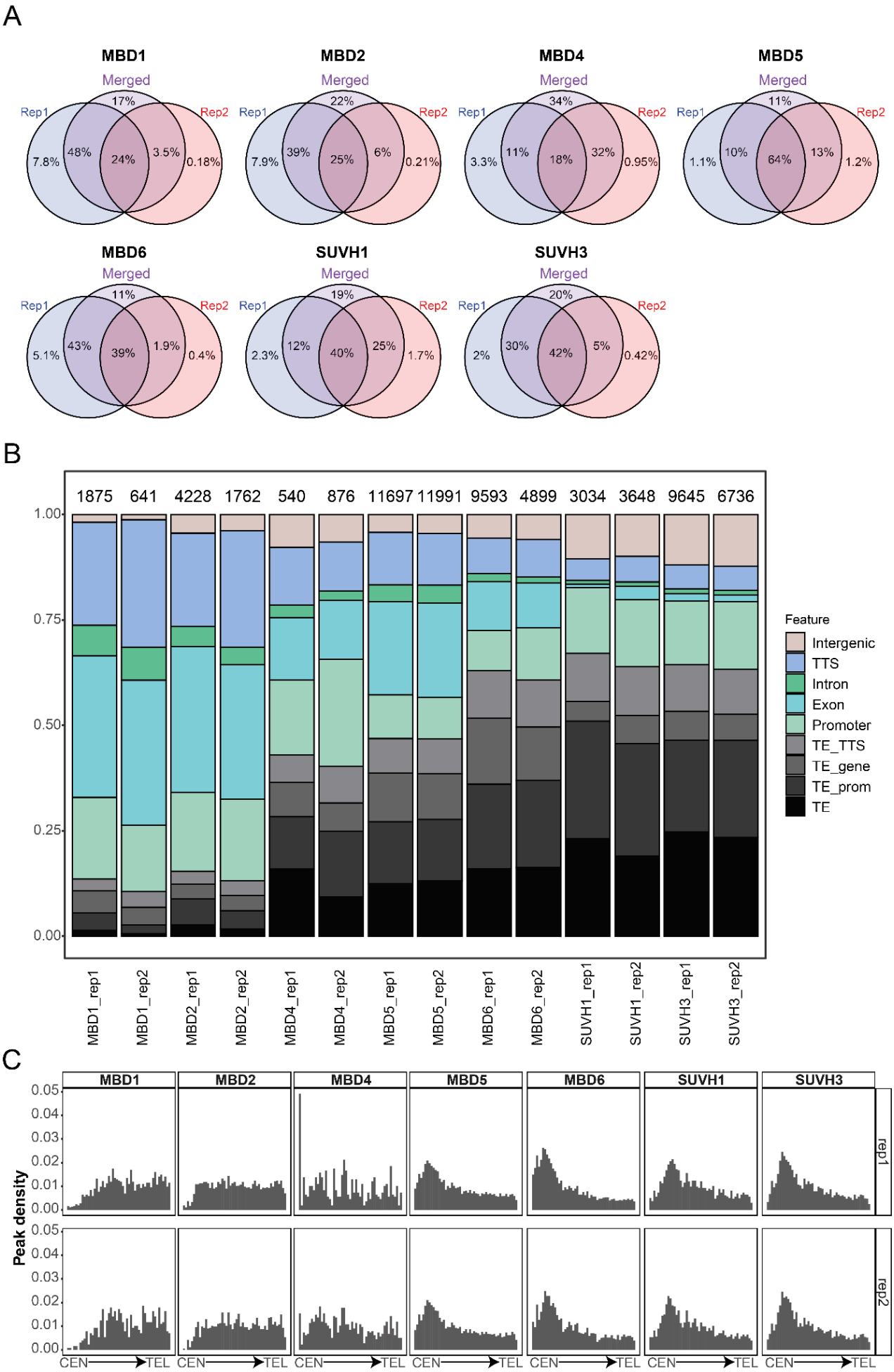
Zhou X, He J, Velanis CN, Zhu Y, He Y, Tang K, Zhu M, Graser L, de Leau E, Wang X, et al. 2021. A domesticated Harbinger transposase forms a complex with HDA6 and promotes histone H3 deacetylation at genes but not TEs in Arabidopsis. *J Integr Plant Biol* 63: 1462–1474.

# Supplemental Figures



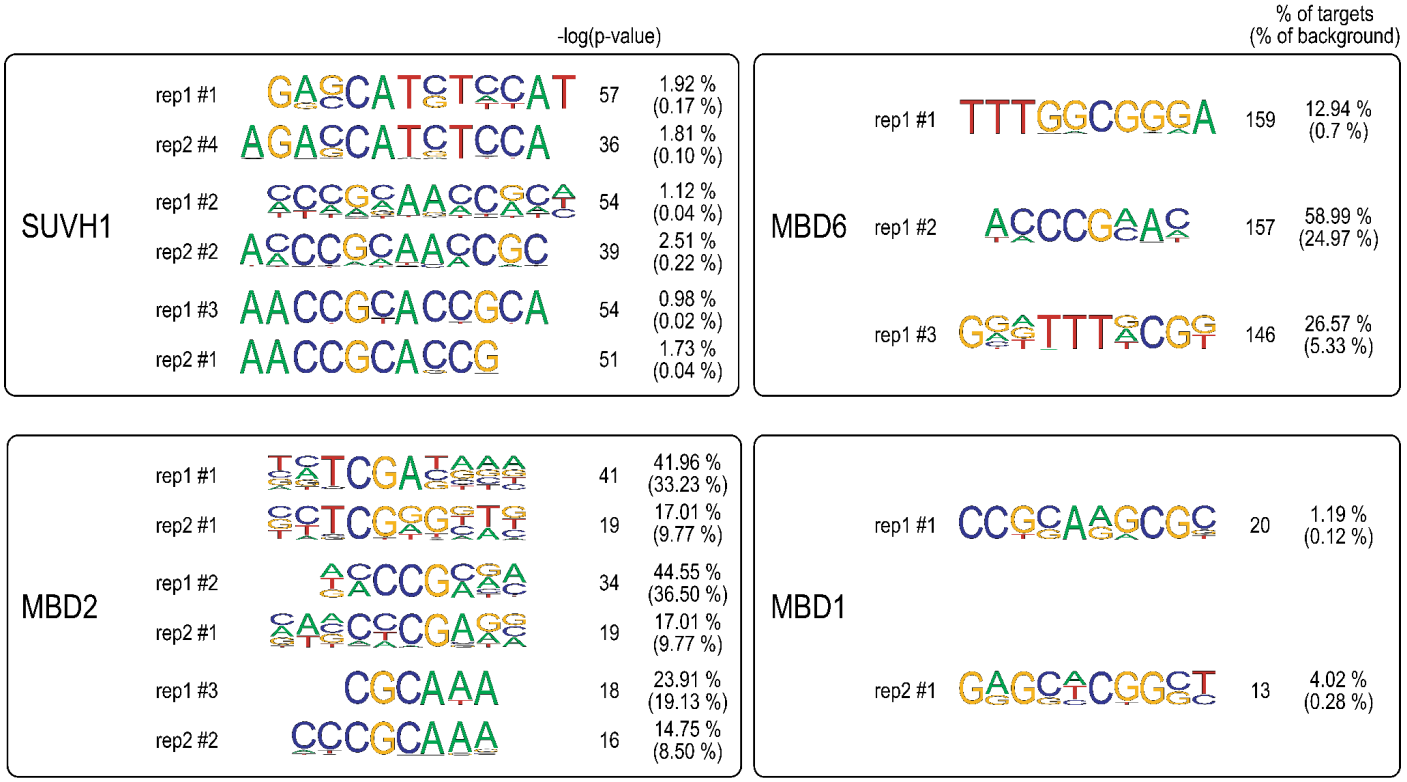

**Supplemental Figure S1: Design of the DNA probes used in the affinity pull-down of Arabidopsis mC readers.**

Distribution of matches of all the 5-bp motifs in Arabidopsis Col-0. The colour is representative of the annotated feature, the transparency is representative of the methylation level. For each feature, the motif with the highest methylation level is the darkest, the motif with the lowest methylation is the clearest. The chosen motifs are boxed and labelled in red.

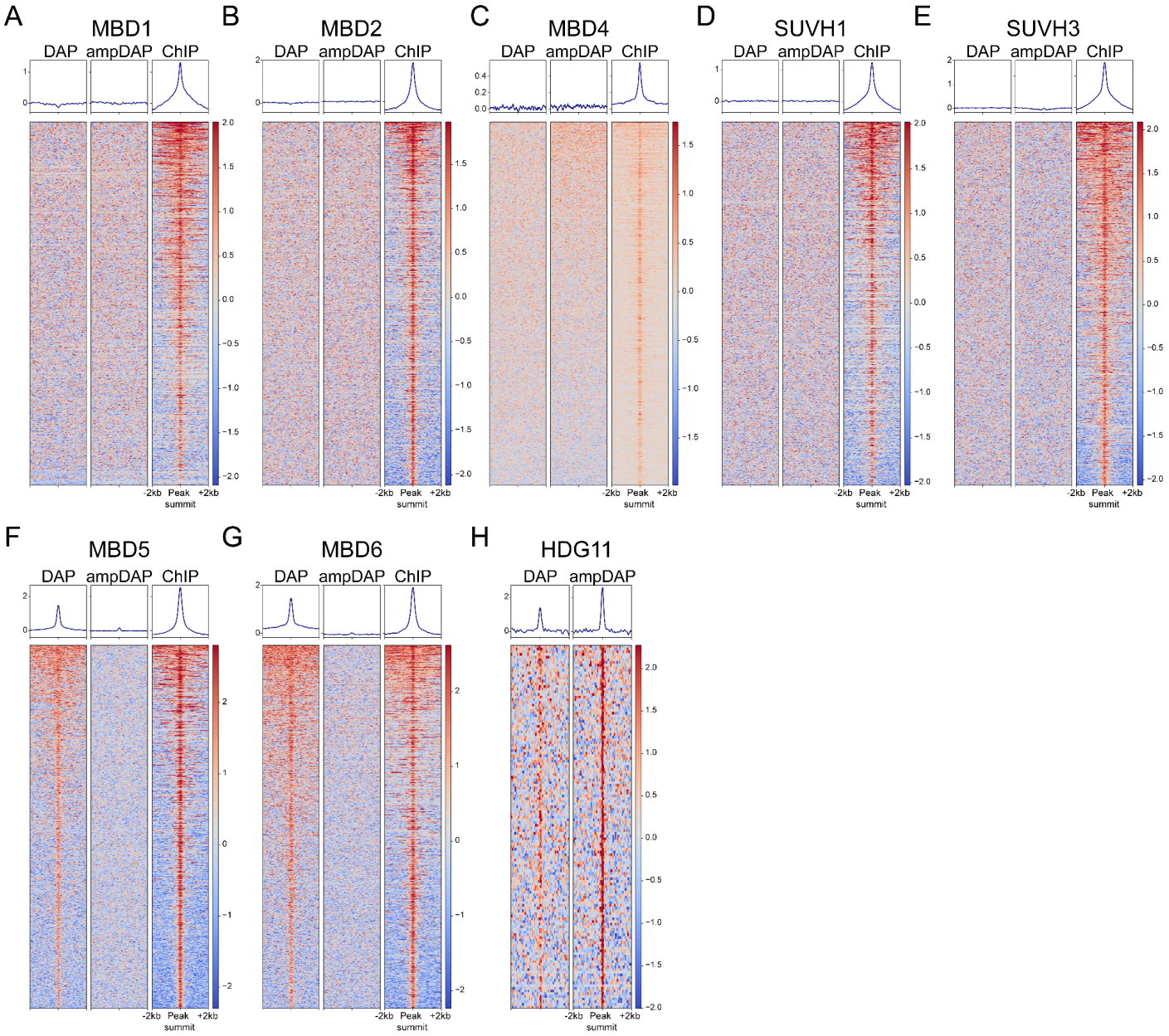



**Supplemental Figure S2: ChIP-seq biological replicates show reproducible results.**

**A)** Overlap between ChIP-seq peaks called for each individual replicate and when merged before calling peaks (as percentage of the total number of individual peaks). One replicate often contributes more to the total amount due to different expression levels of the tagged-protein, but very few peaks from each replicate are lost when merging the data, showing robustness of the signal.
**B)** Distribution of peaks in genomic features for each biological replicate, similar to Fig 2A. Total numbers of peaks for each sample are indicated at the top.
**C)** Density of mC reader ChIP-seq peak position along the Arabidopsis chromosomes, from centromere (CEN) to telomere (TEL), for each biological replicate, similar to Fig 2B.



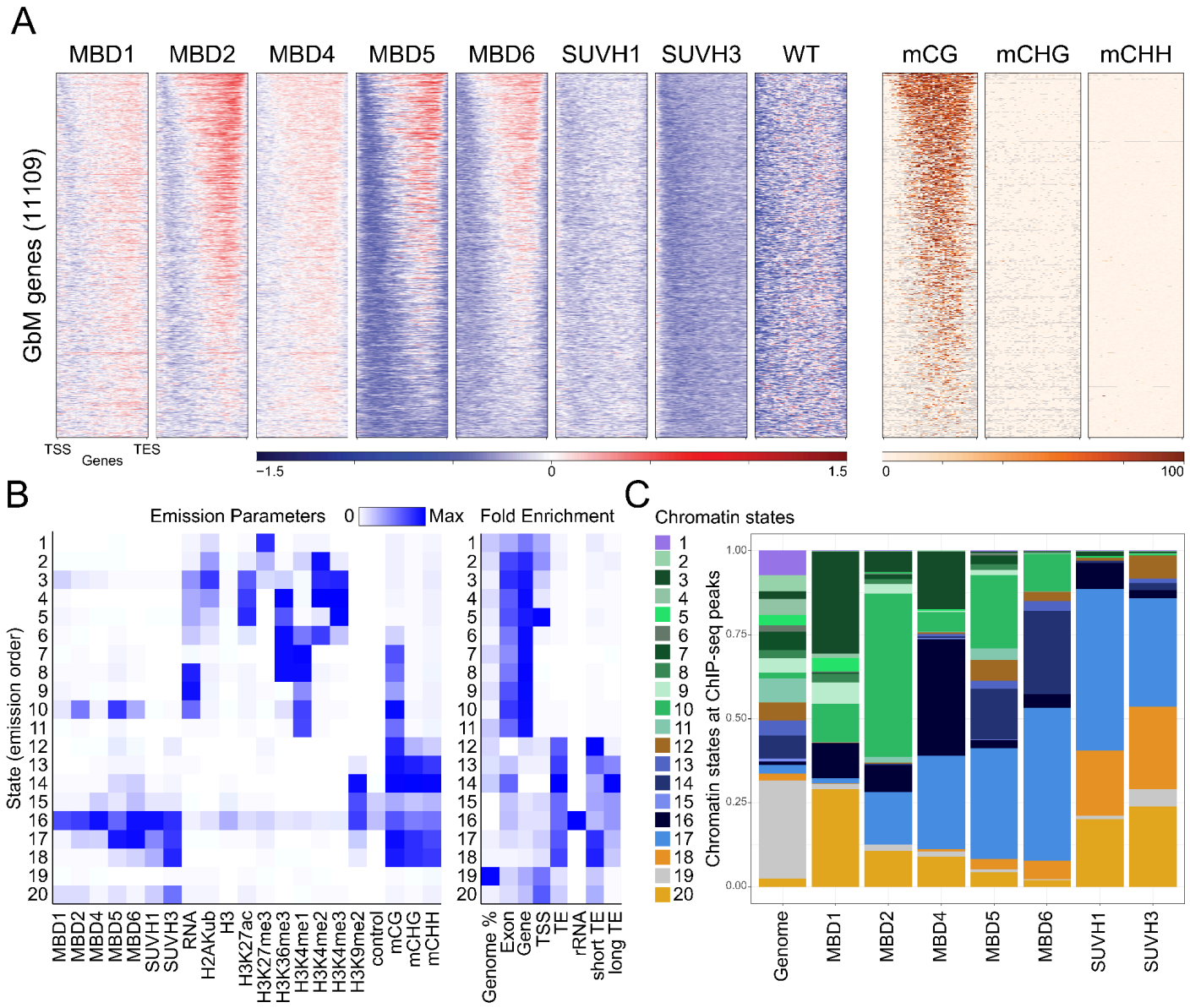

**Supplemental Figure S3: Top motifs over-represented under the peaks of each candidate.**
Peaks and motifs were identified with the Homer suite. According to Homer guidelines, only motifs with -log(p-value)>>50 can be considered real. No homology was found between MBD1 replicates, representative of a broader binding pattern, so only their respective first motifs were included.




**Supplemental Figure S4: Only MBD5 and MBD6 show enrichment in DAP-seq.**

**A-G)** Heatmaps centered on ChIP-seq peaks showing the absence of DAP-seq and ampDAP-seq signals for (A) MBD1, (B) MBD2, (C) MBD4, (D) SUVH1, (E) SUVH3, while high correlation between DAP-seq and ChIP-seq for (F) MBD5 and (G) MBD6.

**H)** HDG11, protein enriched in mCHG probes, showed higher enrichment in ampDAP-seq (unmethylated DNA) than DAP-seq, serving as a positive control for ampDAP-seq. The heatmaps are centered on peaks called in ampDAP-seq data.

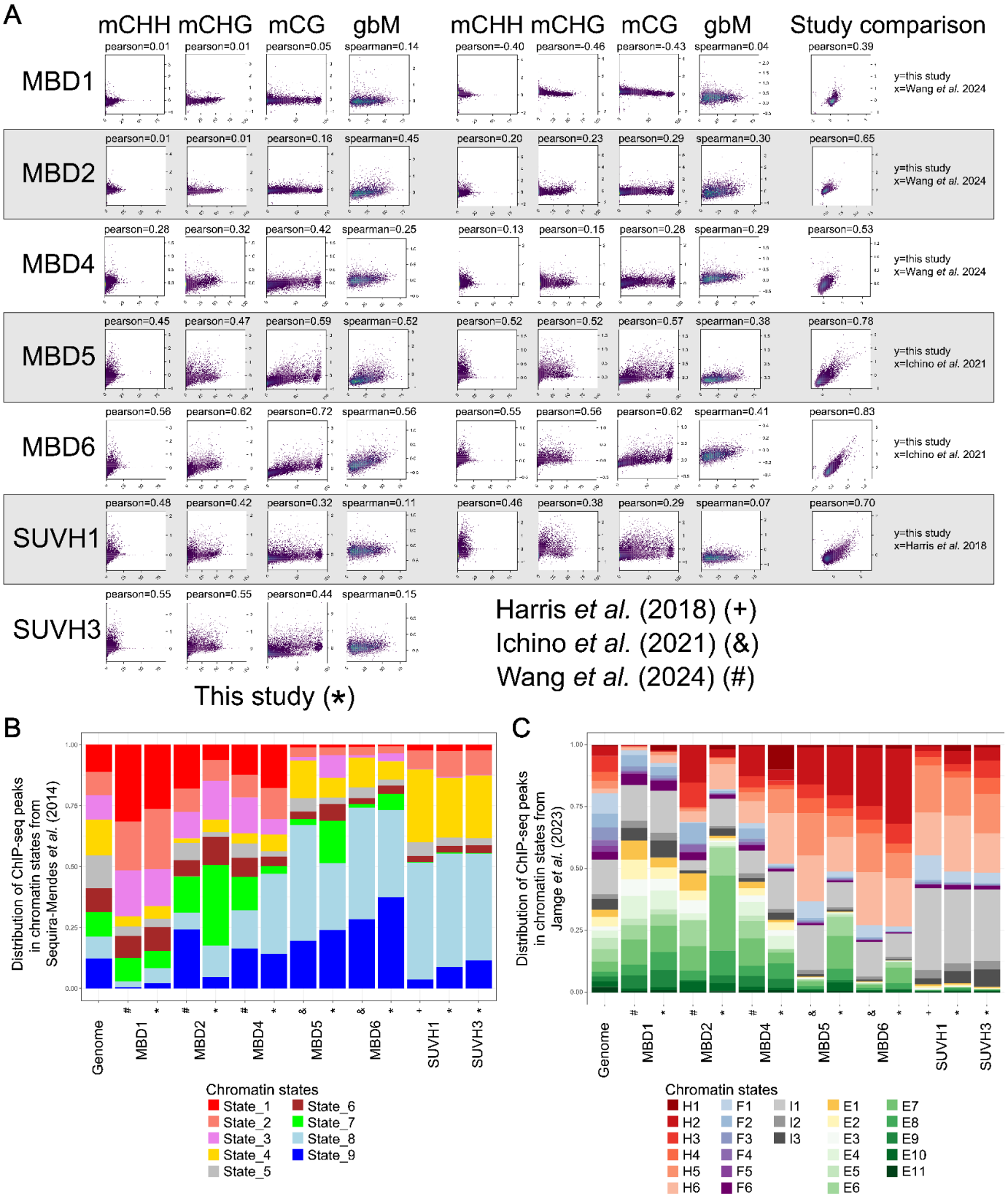



**Supplemental Figure S5: Chromatin contexts, and notably gene-body methylation, define mC reader binding.**

**A)** Heatmap showing enrichment of the mC reader candidates by ChIP-seq over all annotated genes with gene-body methylation (see Methods). MBD2, MBD5 and MBD6 show enrichment at the 5' end of genes, which colocalizes with mCG levels. Heatmap generated with Deeptools, by scaling all genes to 2 kb and plotting 50 bp bins.

**B)** Emission parameters and fold enrichment of annotations in a 20-state model built with chromHMM, highlighting the different chromatin marks defining mC readers binding. Of note, state 16: repeat and methylated regions, including rRNA, bound by all candidates and potentially representing ChIP-seq artifact; State 10 represents gbM genes, bound by MBD2, MBD5, and MBD6; State 3: unmethylated genes bound by MBD1; States 14 and 17: Heterochromatic loci with high mC (especially mCG) bound by MBD5 and MBD6; States 18: Loci with high mC and notably mCHH bound by SUVH1 and SUVH3. Control: ChIP-seq control in wild-type plants; Gene: protein-coding genes; TSS: transcription start site; TE: transposable element; rRNA: ribosomal RNA; short TE and long TE: transposable element less or more than 2kb long, respectively. For each column, the enrichment values were normalized, with darker blue color representing higher enrichment (see Supp. Table 10 for original values).

**C)** Distribution of loci bound by each mC reader (area under ChIP-seq peaks) in the different chromatin states defined in the chromHMM model in panel B.




**Supplemental Figure S6: Correlation between mC reader datasets from this study and previously published datasets from other tissues.**

**A)** Genome-wide correlation between ChIP-seq signal and methylation values in each sequence context in WT seedlings and when limiting to genes with gene body methylation (gbM), for the mC readers studied here, the previously published datasets, and the correlation between these studies.

**B)** Distribution of loci bound by each mC reader (area under ChIP-seq peaks) in the different chromatin states defined in Sequeira-Mendes *et al.* 2014, comparing the ChIP-seq data generated in this study and others, as labelled in **A**. While we refer the readers to the original publications for more details about each

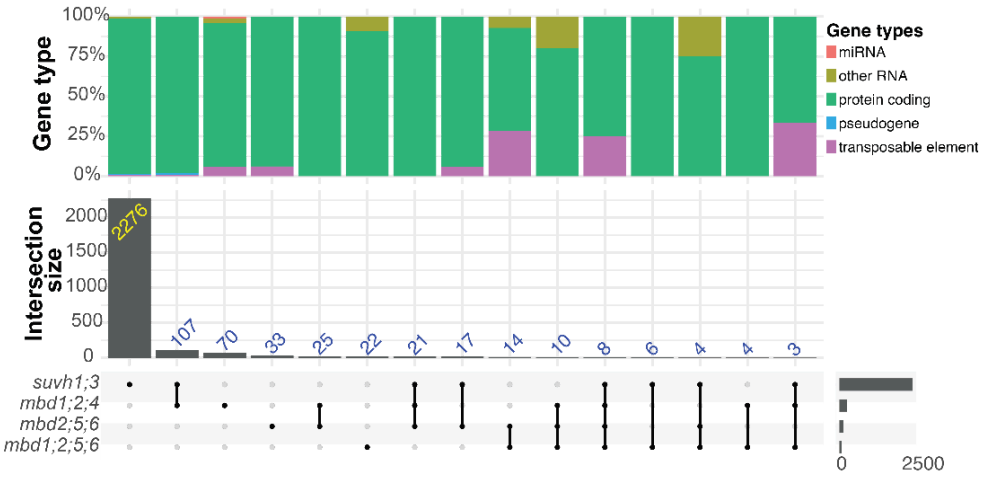
state, states 1, 2, 3, 6, and 7 correspond to expressed genes and their promoters, states 8 and 9 to heterochromatin, state 5 to regions regulated by Polycomb, and state 4 to intergenic regions. The main difference between states 1 and 2 is the presence of H3K27me3 in state 2, while the difference between state 1 and state 7 is the lack of H3K4me3 and the presence of mCG in state 7.

**C)** Distribution of loci bound by each mC reader (area under ChIP-seq peaks) in the different chromatin states defined in Jamge *et al.* 2023. While we refer the readers to the original publications for more details about each state, H=heterochromatin, F=Facultative heterochromatin, I=Intergenic and E=Euchromatin.

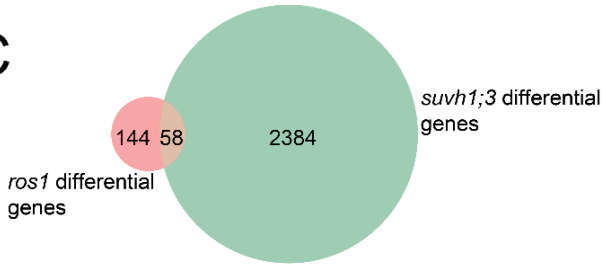
A

| HOME pairwise comparison | # CG DMRs |
|---|---|
| WT VS *mbd1* | 1098 |
| WT VS *mbd4* | 561 |
| WT VS *mbd6* | 1044 |
| WT VS *suvh1;3* | 1003 |
| WT VS *suvh1* | 1203 |
| WT VS *suvh3* | 1909 |
| WT VS *mbd1;2;4* | 1051 |
| WT VS *mbd1;2;5;6* | 1126 |
| WT VS *mbd2;5;6* | 1266 |
| WT VS *hdg11;12* | 1031 |
| WT VS *mbd1;2* | 875 |
| WT VS *mbd1;4* | 726 |
| WT VS *mbd2* | 1869 |
| WT VS *mbd5;6* | 815 |
| WT VS *mbd5* | 1020 |

B




C




D

| DMR WT vs: | ChIP bait | CG DMRs | mC reader ChIP | Overlap |
|---|---|---|---|---|
| *mbd1;2;4* | MBD1 | 1051 | 2069 | 28$^{NS}$ |
| | MBD2 | | 5311 | 135$^{NS}$ |
| | MBD4 | | 1544 | 47$^{NS}$ |
| *mbd1;2;5;6* | MBD1 | 1126 | 2069 | 39$^{NS}$ |
| | MBD2 | | 5311 | 179$^{NS}$ |
| | MBD5 | | 14531 | 458$^{NS}$ |
| | MBD6 | | 10249 | 370$^{NS}$ |
| *mbd2;5;6* | MBD2 | 1266 | 5311 | 194$^{NS}$ |
| | MBD5 | | 14531 | 528$^{NS}$ |
| | MBD6 | | 10249 | 412$^{NS}$ |
| *suvh1;3* | SUVH1 | 1003 | 4913 | 153$^{NS}$ |
| | SUVH3 | | 12164 | 242$^{NS}$ |

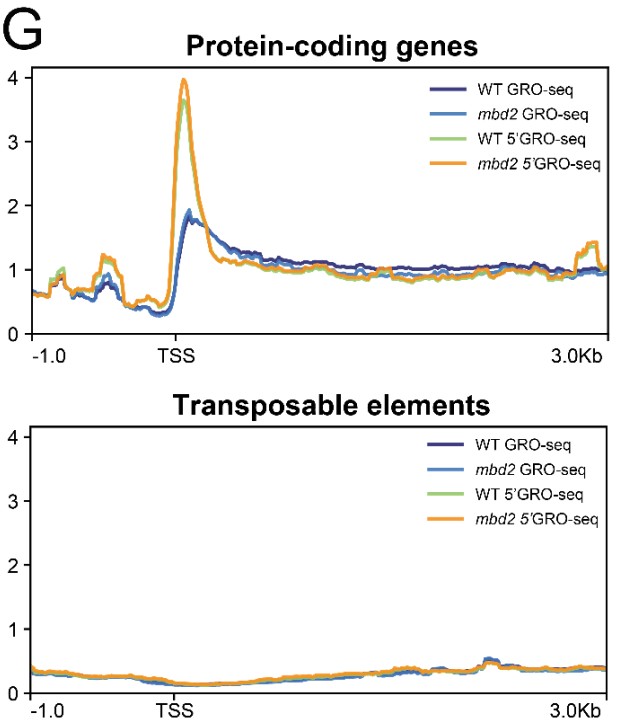
E

| Sample | CG DMRs | DEG (promoter) | Overlap |
|---|---|---|---|
| *mbd1;2;4* | 1051 | 248 | 2$^{NS}$ |
| *mbd1;2;5;6* | 1126 | 71 | 0$^{NS}$ |
| *mbd2;5;6* | 1266 | 132 | 4$^{NS}$ |
| *suvh1;3* | 1003 | 2442 | 24$^{NS}$ |

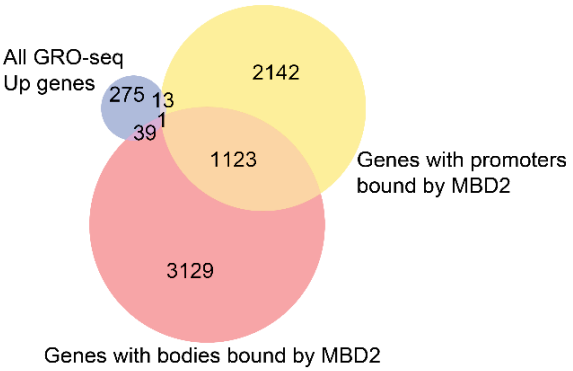
F

| Sample | CG DMRs | DEG (gene body) | Overlap |
|---|---|---|---|
| *mbd1;2;4* | 1051 | 248 | 7$^{NS}$ |
| *mbd1;2;5;6* | 1126 | 71 | 3$^{NS}$ |
| *mbd2;5;6* | 1266 | 132 | 8$^{NS}$ |
| *suvh1;3* | 1003 | 2442 | 38$^{NS}$ |

G




H




**Supplemental Figure S7: The DNA methylome and transcriptome of mC reader mutant plants is highly stable.**

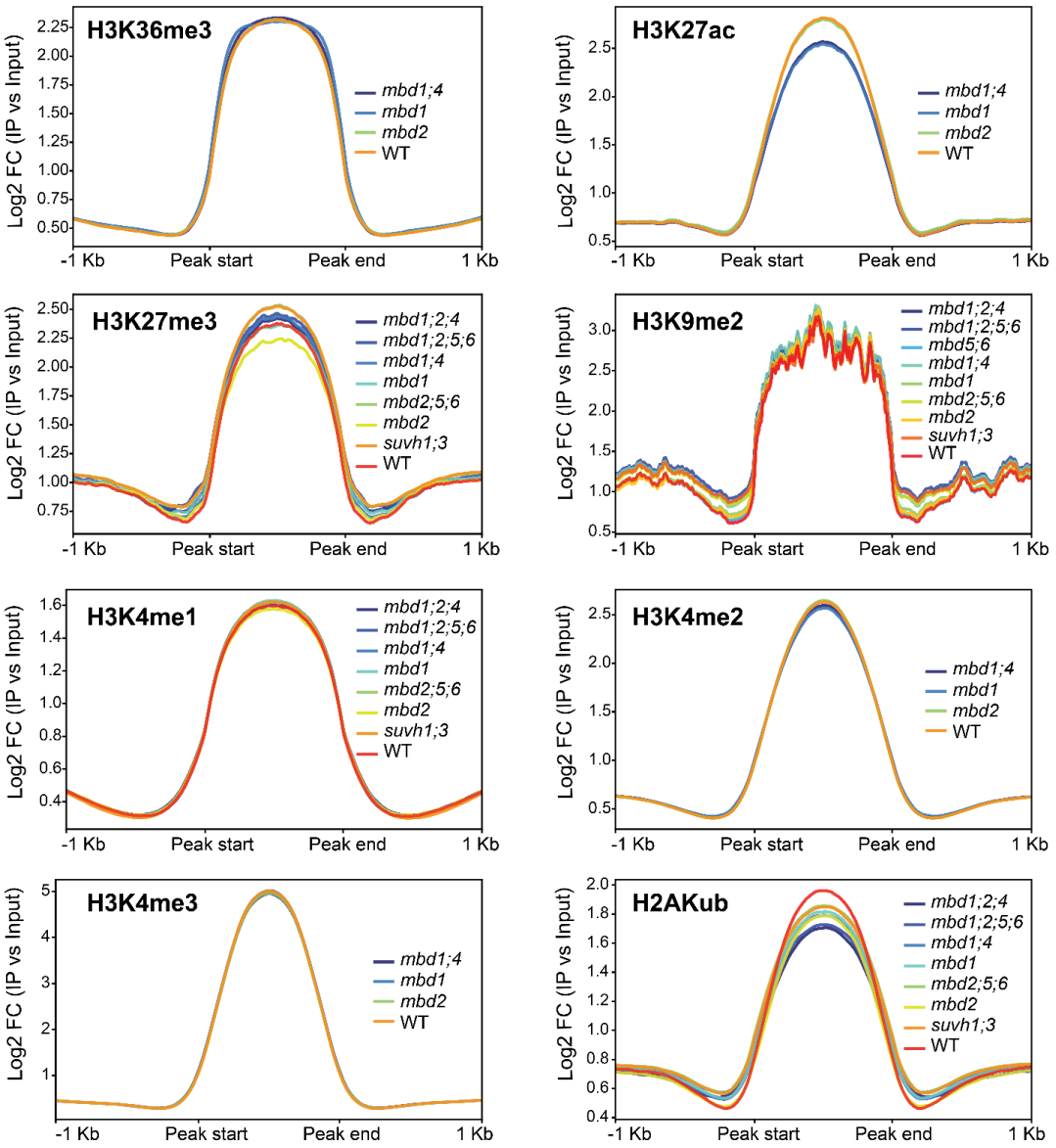
**A)** CG-DMRs identified in the various mC reader mutant seedlings in comparison to WT seedlings.

**B)** The number of differentially expressed genes identified in the higher-order mutants within this study and their overlap with other mutants.
**C)** Overlap of DEGs in *suvh1;3* (this study) and in *ros1* (Kim et al., 2019), hypergeometric test: *p*-value = $4.89 \times 10^{-21}$.
**D)** The overlap of CG-DMRs in higher order mutants with mC reader ChIP peaks. $^{NS}$ denotes non-significant hypergeometric statistical test of overlap.
**E)** The overlap of differentially expressed genes (DEGs) within each mutant with CG-DMRs identified within the higher order mutant that overlap the DEG's promoter. $^{NS}$ denotes non-significant hypergeometric statistical test of overlap.
**F)** The overlap of DEGs within each mutant with CG-DMRs identified within the same mutant that overlap the DEG's gene body. $^{NS}$ denotes non-significant hypergeometric statistical test of overlap.
**G)** GRO-seq signal over TAIR10 genes, split into protein-coding genes and transposable element genes.
**H)** Overlapping of the genes with increased transcription in either normal GRO-seq or 5' GRO-seq with MBD2-bound promoters (hypergeometric statistical test p=0.998) or the gene bodies (hypergeometric statistical test p=0.449).

A

| | WT vs: | *mbd1* | *mbd2* | *mbd1;4* | *mbd5;6* | *mbd1;2;4* | *mbd1;2;5;6* | *mbd2;5;6* | *suvh1;3* |
|---|---|---|---|---|---|---|---|---|---|
| **Heterochromatin mark** | H3K9me2 | 7/3 | 10/8 | 22/6 | 1/1 | 35/12 | 53/8 | 26/9 | 140/35 |
| **Polycomb marks** | H2AKub | 5/3 | 2/8 | 1/0 | NA | 33/59 | 356/352 | 146/190 | 176/227 |
| | H3K27me3 | 2/1 | 18/5 | 1/1 | NA | 13/24 | 52/48 | 18/55 | 21/73 |
| | H3K36me3 | 0/2 | 0/2 | 0/0 | NA | NA | NA | NA | NA |
| **Euchromatin marks** | H3K4me1 | 0/1 | 2/1 | 0/0 | NA | 2/0 | 4/0 | 0/0 | 0/0 |
| | H3K4me2 | 1/1 | 9/4 | 0/1 | NA | NA | NA | NA | NA |
| | H3K4me3 | 0/3 | 12/6 | 0/0 | NA | NA | NA | NA | NA |
| | H3K27ac | 3/2 | 13/4 | 0/1 | NA | NA | NA | NA | NA |

B




**Supplemental Figure S8: Mutations of mC readers does not impair the normal epigenome of histone modifications.**

**A)** Number of chromatin mark peaks that increase/decrease in each mC reader mutant. NA represents "not applicable" as that combination of mC reader mutant and histone modification/variant was not examined in this study. Purple labels indicate histone modifications that are generally found at repressed loci, while blue labels indicate histone modifications that are generally found at active loci.

**B)** Metaplots of histone modifications examined within this study. Peaks were identified by calling peaks of the histone modifications in each of the samples (WT and mutants) and then merging to find a unified set.