

ARTICLE

DOI: 10.1038/s41467-018-03289-7

OPEN

TET-mediated epimutagenesis of the *Arabidopsis thaliana* methylome

Lexiang Ji¹, William T. Jordan², Xiuling Shi², Lulu Hu^{3,4,5}, Chuan He^{3,4,5}  & Robert J. Schmitz²

DNA methylation in the promoters of plant genes sometimes leads to transcriptional repression, and the loss of DNA methylation in methyltransferase mutants results in altered gene expression and severe developmental defects. However, many cases of naturally occurring DNA methylation variations have been reported, whereby altered expression of differentially methylated genes is responsible for agronomically important traits. The ability to manipulate plant methylomes to generate epigenetically distinct individuals could be invaluable for breeding and research purposes. Here, we describe “epimutagenesis,” a method to rapidly generate DNA methylation variation through random demethylation of the *Arabidopsis thaliana* genome. This method involves the expression of a human ten-eleven translocation (TET) enzyme, and results in widespread hypomethylation that can be inherited to subsequent generations, mimicking mutants in the maintenance of DNA methyltransferase *met1*. Application of epimutagenesis to agriculturally significant plants may result in differential expression of alleles normally silenced by DNA methylation, uncovering previously hidden phenotypic variations.

¹Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA. ²Department of Genetics, University of Georgia, Athens, GA 30602, USA.

³Department of Chemistry, University of Chicago, Chicago, IL 60637, USA. ⁴Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA. ⁵Howard Hughes Medical Institute, University of Chicago, Chicago, IL 60637, USA. These authors contributed equally: Lexiang Ji and William T. Jordan. Correspondence and requests for materials should be addressed to R.J.S. (email: schmitz@uga.edu)

Our ability to develop novel beneficial crop traits has significantly improved over the last 100 years, although the ability to maintain this trajectory is limited by allelic diversity. While genetic variation has been heavily exploited for crop improvement, utility of epigenetic variation has yet to be efficiently implemented. Epigenetic variation arises not from a change in the DNA sequence, but by changes in modifications to DNA such as cytosine methylation. This variation can result in the emergence of novel and stably inherited phenotypes, as well as unique patterns of gene expression.

In plant genomes, cytosine methylation occurs at three major sequence contexts: CG, CHG, and CHH (where H = A, C, or T)¹. Methylation at these different contexts is coordinated by distinct maintenance mechanisms during DNA replication. The methylation of DNA in all three contexts is essential for transcriptional silencing of transposons, repeat sequences, and certain genes. Genes regulated by this mechanism are stably repressed throughout the soma and represent an untapped source of hidden genetic variation if transcriptionally re-activated, as revealed from pioneering studies in the model plant *A. thaliana*^{2–4}. However, the impact of this variation is not observed in wild-type plants, as genes silenced by DNA methylation are not expressed. This novel source of genetic variation was uncovered by creating epigenetic recombinant inbred lines (epiRILs) from crosses between a wild-type individual and a mutant defective in maintenance of DNA methylation^{2–4}. EpiRILs, while genetically wild type, contain mosaic DNA methylomes dependent on chromosomal inheritance patterns, as DNA methylation is meiotically inherited in *A. thaliana*^{2, 5–8}. Phenotypic characterization of epiRILs has revealed extensive morphological variation with respect to traits such as flowering time, root length, and resistance to bacterial infection^{2–4}. The morphological variation generated by the creation of epiRILs has revealed extensive hidden genetic variation in plant genomes that can be observed due to expression of newly unmethylated regions. However, the creation of epiRILs requires one founding parent to be a null mutant in the maintenance DNA methylation pathway. Unfortunately, unlike in *A. thaliana*, the loss of DNA methylation maintenance activity often results in lethality in crops^{9, 10}. Therefore, novel methodologies are required to realize the potential of these hidden epialleles in crop genomes.

Epimutagenesis is an alternative method to generate epiRILs. Instead of relying on the genome-wide demethylation of one of the two founding parents, epimutagenesis introduces random methylation variation via the introduction of a transgene. Here, we describe a novel epimutagenesis approach in *A. thaliana* using a human ten–eleven translocation methylcytosine dioxygenase 1 (TET1)^{11–14}, which catalyzes the conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC). Although TET enzymes and their primary product 5hmC are not found in plant genomes¹⁵, ectopic expression of a human TET enzyme resulted in widespread DNA demethylation and induced phenotypic variation in *A. thaliana*.

Results

Overexpressing TET1 in *Arabidopsis* hypomethylates the genome. Transgenic *A. thaliana* plants were generated expressing the catalytic domain (residues 1418–2136) of the human TET1 protein (hTET1cd) under the control of the CaMV35S promoter. To assess the impact of hTET1cd expression on the *A. thaliana* methylome, whole-genome bisulfite sequencing (WGBS) was performed on two independently derived transgenic plants (35S:TET1-1 and 35S:TET1-2; Supplementary Table 1). WGBS revealed a global reduction of CG methylation from 18.2% in two wild-type individuals to 8.9% in 35S:TET1-1 and 6.9% in 35S:

TET1-2 (compared to 0.5% in *met1-3*). The effects of hTET1cd expression on CHG and CHH methylation were not as severe compared to CG methylation (Fig. 1a). Importantly, different degrees of CG hypomethylation were observed in different independent transgenic plants. This result has important implications for epimutagenesis in economically and agriculturally significant plant species, as it appears feasible to control the degree of DNA hypomethylation by screening for plants with desired levels of demethylation. Taken together, these results show that the expression of hTET1cd results in intermediate CG methylation levels when compared to wild-type and *met1* individuals.

The primary product of TET1 oxidation is 5hmC, which is indistinguishable from 5mC by WGBS. We therefore performed Tet-assisted bisulfite sequencing (TAB-seq) to profile 5hmC levels in 35S:TET1 plants¹⁶. No detectable levels of 5hmC were found in the transgenic lines assayed (Supplementary Fig. 1a, b). Thus, the widespread loss of CG DNA methylation observed may result from a failure to maintain methylation at CG sites that possess 5hmC, or through active removal of 5hmC or further oxidized products via the base excision repair pathway.

To better understand the effects of hTET1cd expression, we determined changes in the *A. thaliana* methylome at the chromosomal and local levels. Plotting methylation levels across all five chromosomes revealed a strong depletion of CG methylation at the pericentromeric region (Fig. 1b). CG hypomethylation occurred at both gene body methylated (gbM) and select RNA-directed DNA methylated (RdDM) loci (Fig. 1c, d). To further quantify the observed hypomethylation, metaplots were created for genes and transposons, respectively (Fig. 1e, f and Supplementary Fig. 1c–f). A strong reduction of mCG and a mild reduction of mCHG/mCHH were observed at both genes and transposons. On average, 97.9% of gbM genes and 56.7% of methylated transposons (where these regions have at least 50% mCG in wild type) lost at least half of their CG methylation in epimutagenized lines (Supplementary Data 1 and 2). Collectively, these results indicate that hypomethylation was more severe in genes than transposons, possibly the result of de novo methylation by the RdDM pathway, which is primarily active at transposons.

TET1-mediated DNA demethylation mimics *met1* mutants. An analysis of differentially methylated regions (DMRs) was then carried out to assess the genome-wide impact of hTET1cd expression. A total of 56,283 CG DMRs ranging in size from 6–20,286 base pairs (bp) were identified (Fig. 1g and Supplementary Data 3). Of these, 38.7% were located in intergenic sequences, 53.7% overlapped with genes, and 7.6% were located in promoter regions (≤ 1 kb upstream of a gene). As also seen in *met1* mutants, the predominant effect of hTET1cd expression is CG hypomethylation (12,641 and 20,601 DMRs lost more than 50% mCG in 35S:TET1-1 and 35S:TET1-2, respectively; no region gained more than 50% mCG). However, the extent of CG methylation loss caused by hTETcd expression is lower than in *met1*: 31.8 Mb of the genome significantly lost CG methylation in *met1*, whereas 9.9 Mb and 18.0 Mb were lost in 35S:TET1-1 and 35S:TET1-2, respectively.

Previous studies of the *met1* methylome have revealed a loss of mCHG/mCHH methylation in a subset of CG-hypomethylated regions¹⁷. At these loci, DNA methylation is stably lost, in contrast to regions where DNA methylation is re-established by de novo methylation pathways. These loci are ideal targets of epimutagenesis, as the co-existence of all three types of methylation is more frequently correlated with transcriptional repression of genes than CG methylation alone. This, coupled

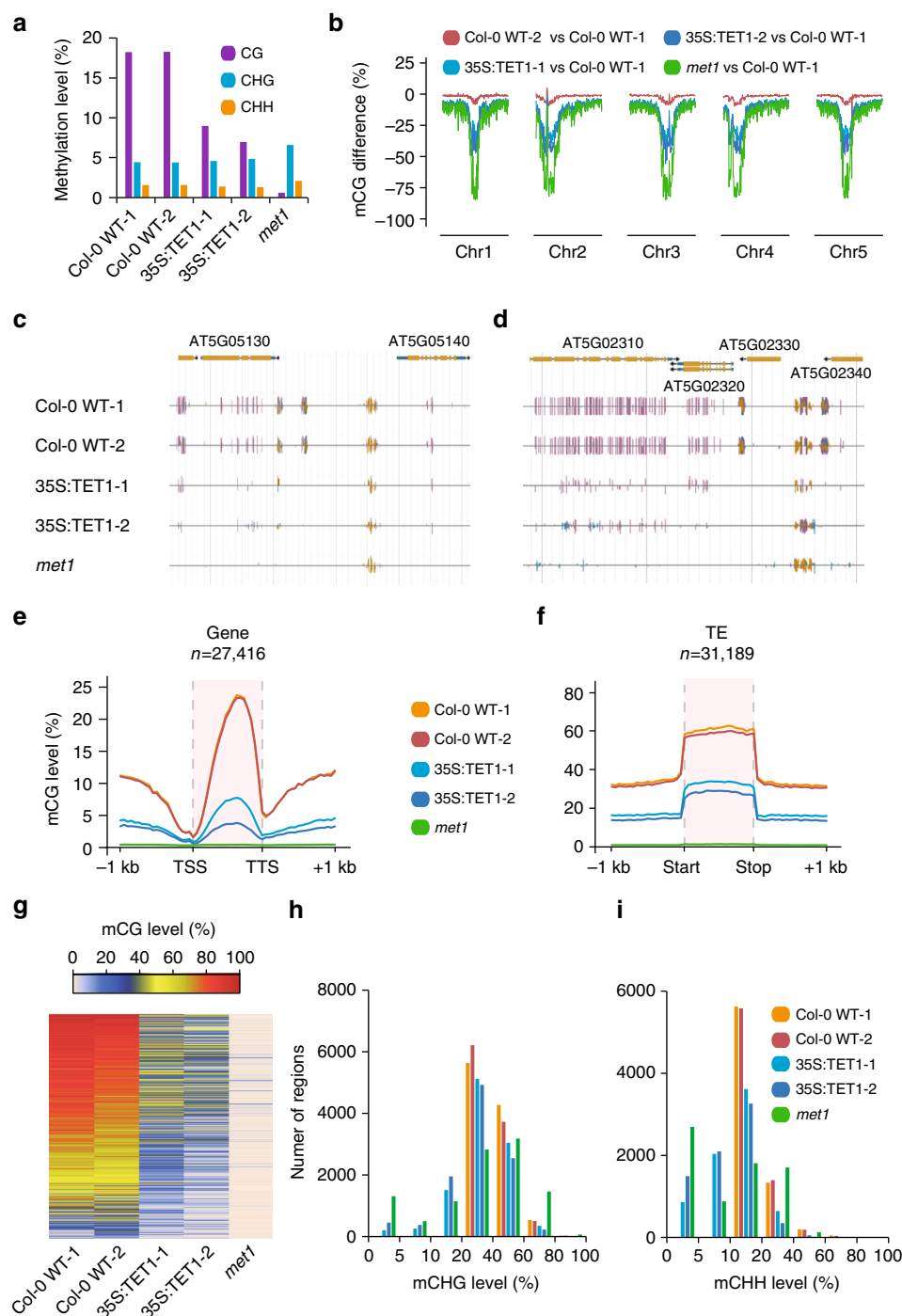


Fig. 1 Overexpression of hTET1cd-induced global CG demethylation in *A. thaliana*. **a** Bar graph of global methylation levels in two Col-0 WT replicates, two 35S:TET1 transgenic individuals, and *met1*. **b** Metaplot of CG methylation levels (100 kb windows) across five *A. thaliana* chromosomes. Methylation level differences were defined relative to Col-0 WT-1, and Col-0 WT-2 was used to assess background interference. Genome browser view of methylome profile of two regions (**c**, **d**) of the *A. thaliana* genome (purple vertical lines = CG methylation, blue vertical lines = CHG methylation, and gold vertical lines = CHH methylation). Metagene plots of CG methylation level across **e** gene bodies and **f** transposable elements. **g** Heat map of CG methylation level of CG DMRs. Bar plots of CHG (**h**) and CHH (**i**) methylation levels of CG DMRs that possess non-CG methylation in wild-type individuals

with the long-term stability of hypomethylation, may facilitate inherited transcriptional changes. An analysis of the interdependence of the loss of CG methylation on non-CG methylation levels revealed that 39.7 Kb and 931.5 Kb of CHG methylated sequences lost significant amounts of methylation in two

independent epimutagenized lines, compared to 4.0 Mb of sequence in *met1* mutants. A similar analysis for the loss of CHH methylation revealed losses of 23.3 Kb and 492.5 Kb in epimutagenized individuals, compared to 1.1 Mb lost in *met1* mutants. Of the 56,283 identified CG DMRs, 10,491 overlapped

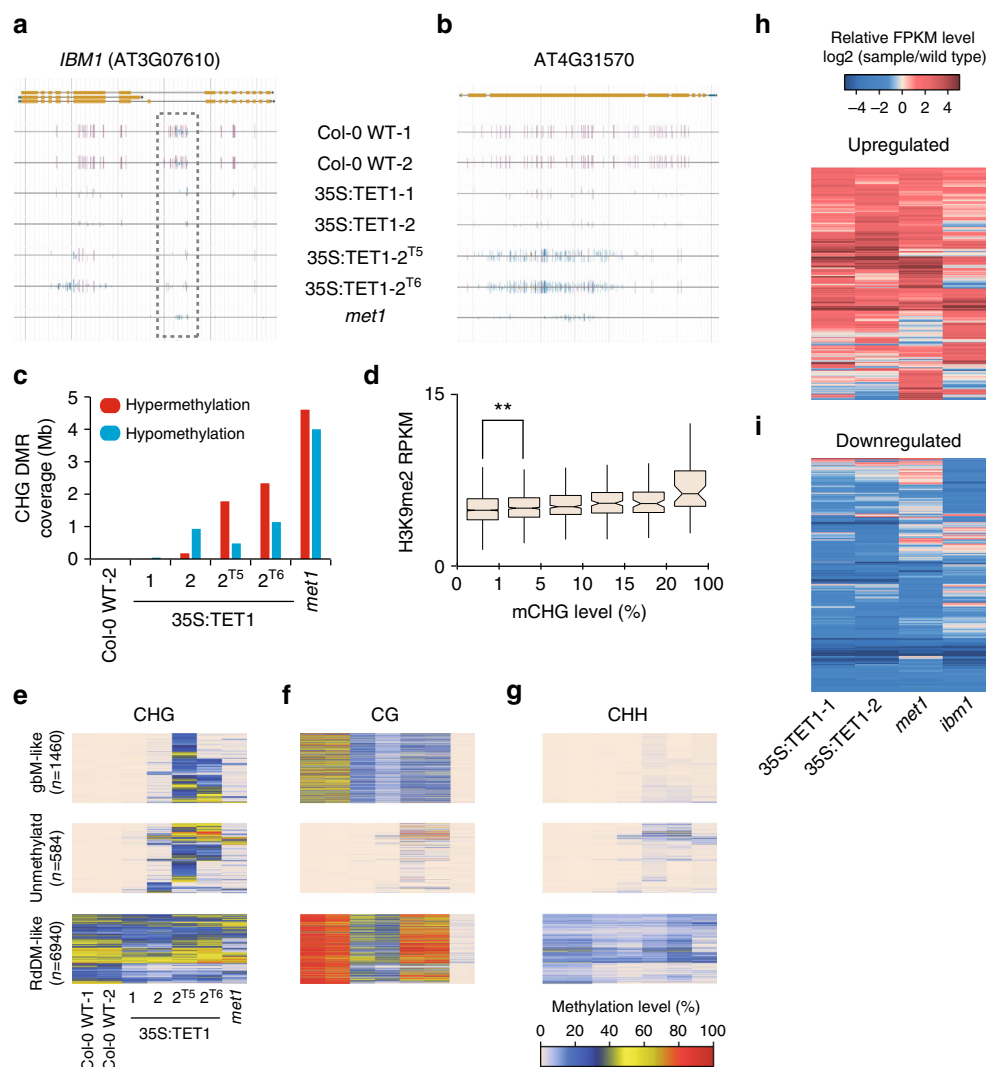


Fig. 2 Global fluctuation of CHG methylation in 35S:TET1 plants. **a** Genome browser view of *IBM1* (AT3G07610) in Col-0 WT, four 35S:TET1 transgenic plants, and *met1*. A decrease in CG methylation from coding regions was accompanied by an increase in non-CG methylation. Both CG and non-CG methylation were lost from the large intron (purple vertical lines = CG methylation, blue vertical lines = CHG methylation, and gold vertical lines = CHH methylation). **b** Genome browser view of a representative CHG hypermethylated region. **c** The amount of the genome affected by differential CHG methylation. These DMRs were defined relative to Col-0 WT-1, as Col-0 WT-2 DMRs were used to assess background interference. **d** Boxplot of H3K9me2 distribution in gBM loci (***t*-test, *p* value < 0.01). **e** Heat map of CHG methylation displaying CHG DMRs. Corresponding CG and CHH methylation levels are shown in **f** and **g**. Heat maps showing log₂ transformed FPKM profiles of upregulated genes (**h**) and downregulated genes (**i**) in two 35S:TET1 transgenic individuals, *met1*, and *ibm1* mutants relative to WT

regions that contained at least 20% CHG methylation and 7214 overlapped regions that contained at least 10% CHH methylation in wild-type individuals (Supplementary Data 3). To determine how many of these regions are susceptible to losing non-CG methylation if CG methylation is first depleted, we created a frequency distribution of mCHG and mCHH levels in wild-type and epimutagenized individuals (Fig. 1h, i). In total, 2341 and 3447 regions lost more than 10% CHG methylation in 35S:TET1-1 and 35S:TET1-2, respectively, whereas 2475 and 3379 regions lost more than 5% CHH methylation in 35S:TET1-1 and 35S:TET1-2, respectively. Regions that are susceptible to losses of CG and non-CG methylation in lines expressing hTET1cd share a substantial overlap with regions that lose non-CG methylation in *met1* (Supplementary Fig. 1g, h). In total, 1708 (73.0%) and 2386 (69.2%) regions that have lost more than 10% mCHG in 35S:

TET1-1 and 35S:TET1-2 have reduced levels in *met1*, whereas 2013 (81.3%) and 2563 (75.9%) regions that have lost more than 5% mCHH in 35S:TET1-1 and 35S:TET1-2 have reduced levels in *met1*. As crop genomes have a greater number of loci targeted for silencing by CG, CHG, and CHH methylation when compared to *A. thaliana*, ectopic expression of hTET1cd is likely a viable approach for the creation epiRILs¹⁸.

TET1-mediated variation of CHG methylation. Mutations in *met1* also result in hypermethylation of CHG sites in gene bodies due to the loss of methylation in the seventh intron of the histone 3 lysine 9 (H3K9) demethylase, increase in bonai methylation 1 (*IBM1*)^{19–22}. This results in alternative splicing of *IBM1*, ultimately producing a non-functional gene product (*IBM1-S*), which results in ectopic accumulation of di-methylation of H3K9

(H3K9me2) throughout the genome²³. As in *met1*, the seventh intron of *IBM1* was hypomethylated in 35S:TET1-1, 35S:TET1-2, and an additional two lines, 35S:TET1-2^{T5} and 35S:TET1-2^{T6}, which were propagated for an additional two and three generations, respectively (Fig. 2a). The increased abundance of *IBM1-S* transcript was confirmed by RT-qPCR in line 35S:TET1-2^{T6} (Supplementary Fig. 2a), leading to CHG hypermethylation at gbM loci (Fig. 2b). Further quantitative analysis revealed extensive variation in genome-wide gains and losses of CHG methylation in these two lines, ~1.8 Mb and 2.3 Mb of additional CHG methylation, respectively (Fig. 2c and Supplementary Fig. 2b, c). To test the impact of a reduction in functional *IBM1* on H3K9me2, we performed chromatin immunoprecipitation (ChIP) against H3K9me2 in 35S:TET1-2^{T6}, which revealed a subtle increase in H3K9me2 in gbM loci that possessed CHG hypermethylation (Fig. 2d).

To further characterize regions of differential CHG methylation, identified CHG DMRs in line 35S:TET1-2^{T5} were categorized into discrete groups based on their DNA methylation status in wild-type individuals. Of the 9917 CHG DMRs identified, 1460 were in loci that are defined as gbM in wild-type individuals, 584 were in unmethylated regions, and 6940 of them were in RdDM-like regions (Fig. 2e–g). Interestingly, in line 35S:TET1-2^{T5}, 1409 (96.5%) of the CHG DMRs in gbM-like loci gained CHG hypermethylation, whereas 2680 (38.6%) of the CHG DMRs in RdDM-like regions lost CHG, in contrast to 825 (11.9%) RdDM-like regions that gained CHG methylation. Lastly, there were 503 (86.1%) loci that are unmethylated in wild-type individuals that gain CHG methylation as well as CG and CHH methylation in the epimutagenized lines (Fig. 2e–g). These results reveal that methods for epimutagenesis can result in both losses and gains in DNA methylation genome wide.

To characterize the effect of hTET1cd-induced methylome changes on gene expression, we performed RNA-sequencing (RNA-seq) on leaf tissue of wild-type, 35S:TET1-1 and 35S:TET1-2. Compared to wild-type plants, 629 and 736 upregulated genes were identified in 35S:TET1-1 and 35S:TET1-2, respectively, with 176 and 260 genes overlapping with identified CG DMRs. A total of 1277 and 1428 downregulated genes were identified and 268 and 324 of them overlapped with CG DMRs. There was a high level of overlap in transcriptome changes seen in 35S:TET1-1 and 35S:TET1-2 compared to *met1* and *ibm1* (Fig. 2h, i). Of the genes upregulated in *met1*, 36.8 and 38.7% overlapped with upregulated genes in 35S:TET1-1 and 35S:TET1-2, respectively (Supplementary Fig. 2d). An even greater overlap was observed with downregulated genes in *met1*, as 60.1 and 65.2% overlapped with downregulated genes in 35S:TET1-1 and 35S:TET1-2, respectively (Supplementary Fig. 2e). These results reveal that hTET1cd expression in *A. thaliana* is a viable approach for accessing hidden sources of allelic variation by inducing expression variation.

TET1 expression leads to a delay in the floral transition. In the transgenic plants that were used for WGBS, we observed a delay in the developmental transition from vegetative to reproductive growth (Fig. 3a, b). We hypothesized that the observed late-flowering phenotype was associated with the demethylation of the *FLOWERING WAGENINGEN* (*FWA*) locus, as is observed in *met1* mutants^{24, 25}. A closer inspection of the DNA methylation status of this locus revealed that DNA methylation was completely abolished, as was methylation at adjacent CHG and CHH sites (Fig. 3c). As in *met1*, the loss of methylation at the *FWA* locus was associated with an increase in *FWA* expression (Fig. 3d), which is known to cause a delay in flowering by restricting the movement of the florigen signal, *FT*, to the shoot

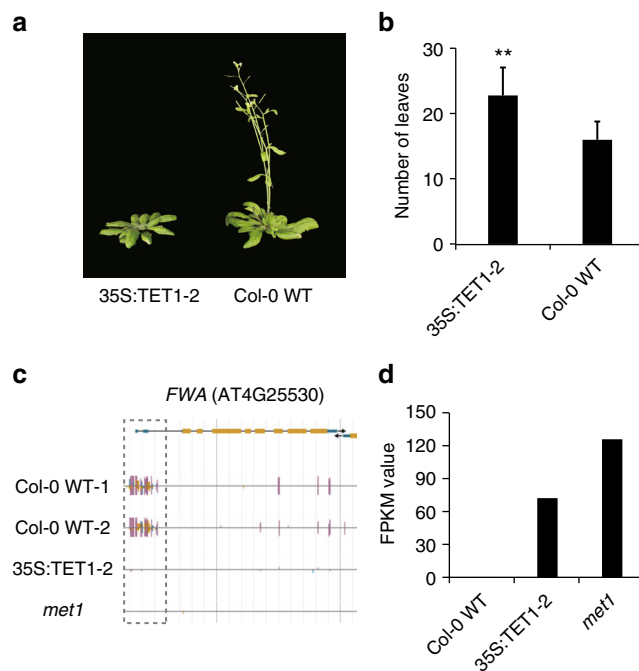


Fig. 3 35S:TET1 plants have a delayed flowering phenotype. **a** Photographs of one 35S:TET1-2 transgenic plant and Col-0 WT plant and **b** corresponding number of rosette leaves. Error bars indicate s.d. (**t-test, *p* value < 0.01). **c** Genome browser view of *FWA* (AT4G25530). Both CG and non-CG DNA methylation are depleted from the 5'UTR in 35S:TET1-2 plants (purple vertical lines = CG methylation, blue vertical lines = CHG methylation, and gold vertical lines = CHH methylation). **d** Expression level (FPKM) of *FWA*

apex²⁶. These results demonstrate that expression of hTET1cd leads to phenotypic variation by abolishing methylation at some regions in all sequence contexts (CG, CHG, and CHH sites).

TET1-mediated demethylation is transgenerationally inherited.

To assess the stability and inheritance of TET1-mediated demethylation, T1 individuals expressing 35S:TET1 were self-fertilized, allowing for the loss of the hTET1cd transgene due to allelic segregation. Unexpectedly, transgene-free 35S:TET1-1 T2 individuals (35S:TET1-1.3^{-TET1}, 35S:TET1-1.4^{-TET1}, and 35S:TET1-1.5^{-TET1}) exhibited a reversion to a normal flowering phenotype, and genome-wide methylation levels closely resembled that of wild-type individuals (Fig. 4a). WGBS on T2 individuals retaining the transgene revealed similar levels of CG methylation as the T1 parent (35S:TET1-1.1^{+TET1} and 35S:TET1-1.2^{+TET1}). The methylation level of CGs at genes and transposons revealed that demethylation of gbM loci was partially inherited in transgene-free T2 individuals, whereas active remethylation was found at transposons in the same individuals (Fig. 4b–f). These results indicate that an active process likely in the meristem and/or germline is counteracting the activity of the hTET1cd transgene²⁷. To quantify how many regions were susceptible to the loss of non-CG methylation, a DMR analysis was conducted for each 35S:TET1-1 T2 individual. A total of 655 and 659 CHG hypomethylated regions were identified in T2 lines retaining the transgene. In contrast, 211, 155 and 199 hypomethylated regions were identified in three transgene-free T2 individuals, respectively (Fig. 4g).

To determine if increased expression of hTET1cd in meristematic tissue would increase the likelihood of germline transmittance of demethylation to transgene-free progeny, transgenic *A. thaliana* plants were generated expressing a previously described

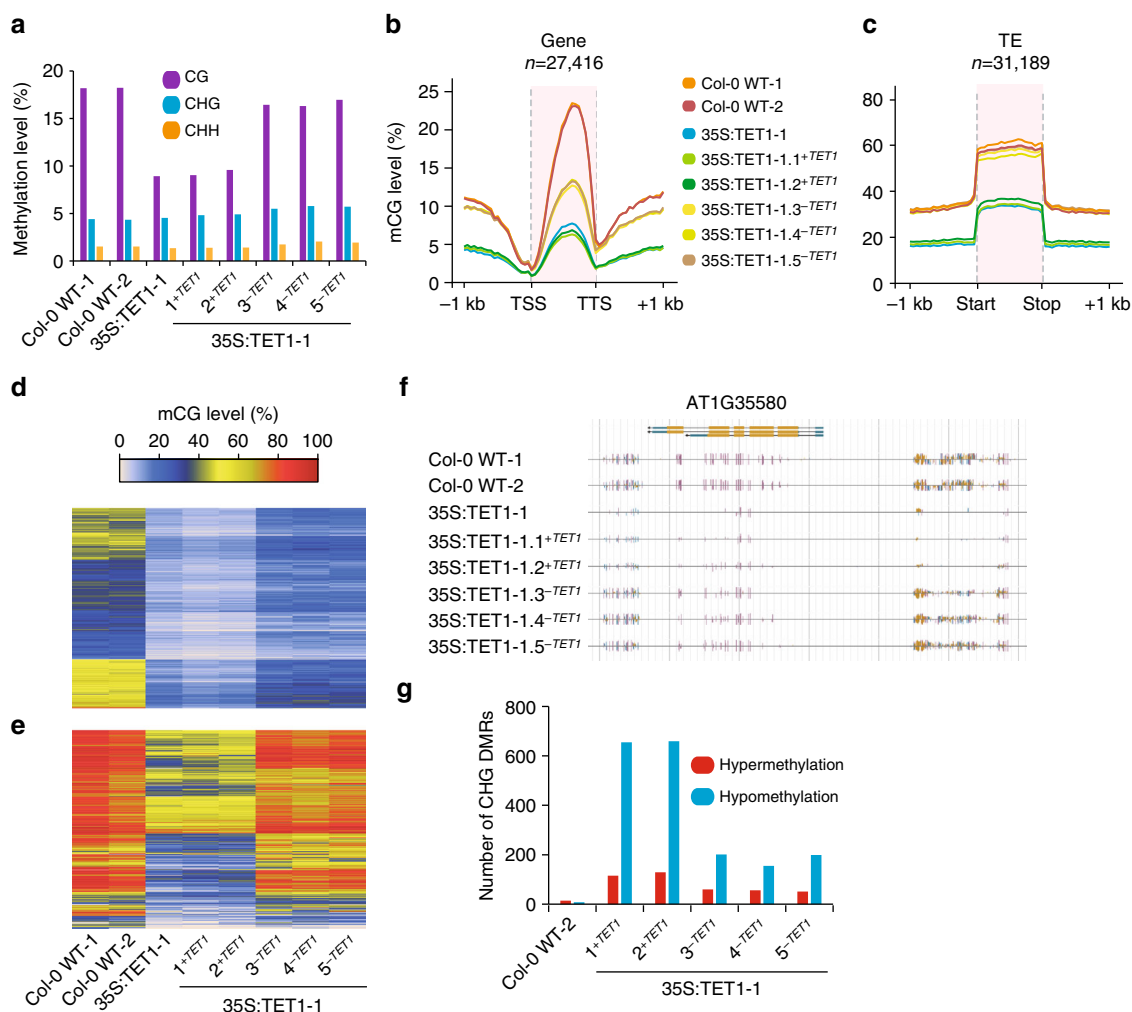


Fig. 4 Transgenerational demethylation profile of 35S:TET1 individuals. **a** Bar plots of global methylation levels in two Col-0 WT replicates and 35S:TET1-1 plants. Metagene plots of mCG level across **b** gene bodies and **c** transposable elements. Heat map of mCG level of **d** all gbM genes and **e** transposable elements with >20% mCG in wild type. **f** Genome browser view of a methylome profile of a representative region of the *A. thaliana* genome in 35S:TET1-1 individuals. **g** Number of identified CHG DMRs in 35S:TET1 T2 plants. These DMRs were defined relative to Col-0 WT-1, as Col-0 WT-2 DMRs were used to assess background

superfolder GFP (sfGFP) hTET1cd fusion, under control of the *A. thaliana* ACTIN 2 (ACT2) promoter (ACT2:sfGFP-hTET1cd), which is known to have activity in all tissues of juvenile plants, including meristematic tissue²⁸. Translation and nuclear localization of the sfGFP-TET1cd fusion protein was confirmed in young cotyledons using confocal microscopy (Supplementary Fig. 3a). T1 populations transformed with ACT2:sfGFP-hTET1cd exhibited a 27-fold increase in later flowering compared to 35S:hTET1cd, indicating high activity of the sfGFP-TET1cd fusion protein in *A. thaliana* (Fig. 5a). To assess the variation between lines, we performed WGBS on four independent ACT2:TET1 T1 lines (Supplementary Fig. 4a–c). Differential levels of global CG demethylation were observed in these four lines, further confirming that plants subjected to epimutagenesis can possess different degrees of demethylation. The expression of the sfGFP-hTET1cd transgene was also confirmed by RT-qPCR in select lines (Supplementary Fig. 4d). Subsequent DMR analysis revealed 68,260 CG DMRs, 9235 CHG DMRs, and 2793 CHH DMRs between these lines (Supplementary Data 4–6), a drastic increase in DMRs compared to those within siblings in 35S:TET1 lines (Fig. 4g).

To further assess the inheritance of the demethylation pattern as a result of epimutagenesis, we selected T2 progeny of a late flowering ACT2:TET1-1.2^{TET1} individual containing and lacking the transgene. WGBS data from these two individuals was used to confirm the presence/absence of the sfGFP-hTET1cd transgene (Supplementary Fig. 3d, e). Individuals retaining and lacking the transgene due to allelic segregation both exhibited a late-flowering phenotype (31 and 28 leaves upon flowering in ACT2:TET1-1.1^{TET1} and ACT2:TET1-1.2^{TET1}, respectively), ectopic expression, and loss of DNA methylation of *FWA* (Supplementary Fig. 3b, c). WGBS on these two individuals revealed a reduction in CG methylation that was maintained and stably inherited irrespective of transgene presence, confirming the high activity of demethylation in ACT2:TET1 lines compared to 35S:TET1 lines (Fig. 5b–f). It is unclear exactly why the late-flowering phenotype was stably inherited in the T2 individuals without the transgene in the ACT2-driven lines versus the 35S-driven lines, although it is likely a combination of promoter strength and cell type specificity. For stable inheritance of demethylation and the late-flowering phenotype, TET1 activity would be required in the meristematic and/or germline cells. Collectively, these results demonstrate the stable inheritance of

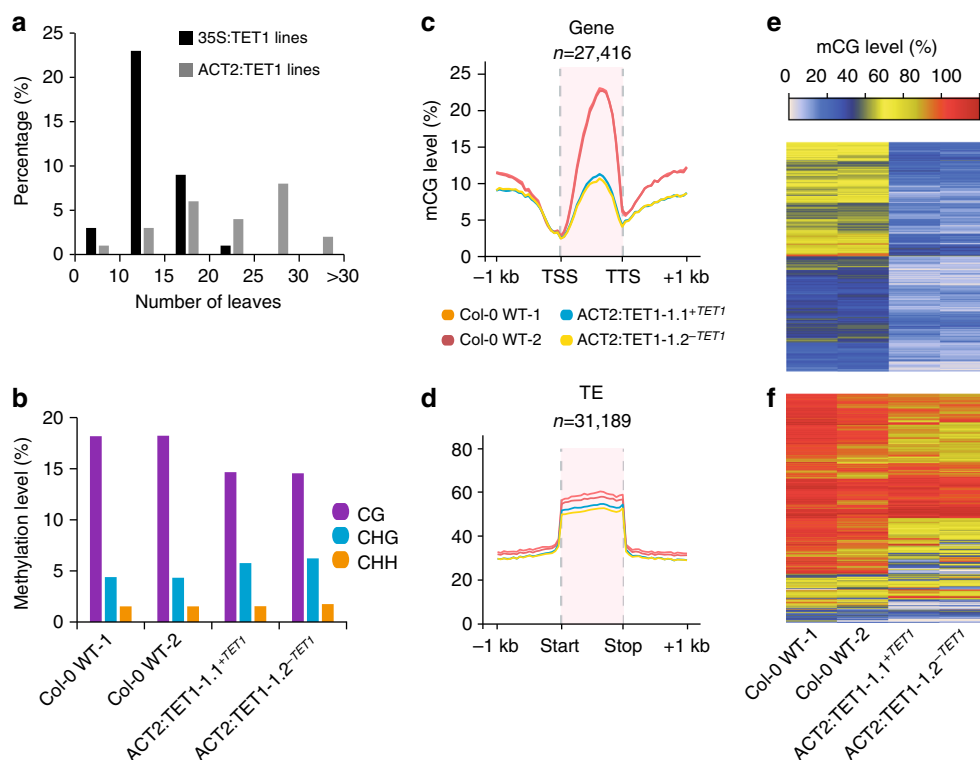


Fig. 5 Demethylation profile of ACT2:TET1 individuals. **a** Bar plot of number of rosette leaves in 35S:TET1 and ACT2:TET1 T1 individuals upon flowering. **b** Bar plot of global methylation levels in two Col-0 WT replicates and two ACT2:TET1-1 T2 plants. Metagene plots of mCG level across **c** gene bodies and **d** transposable elements. Heat map of mCG level of **e** all gbM genes and **f** transposable elements with >20% mCG in wild type

TET1-mediated demethylation and a delayed floral transition in the absence of the transgene.

Discussion

The discovery that expression of the catalytic domain of the human TET1 protein in *A. thaliana* leads to widespread loss of CG methylation enables the creation epimutants without the need for methyltransferase mutants, which often causes lethality in crops. In addition to epimutagenesis, hTET1cd could be used in combination with sequence-specific DNA-binding proteins such as dCas9 to direct DNA demethylation in plant genomes, as has been demonstrated in mammalian systems^{29–33}. The stable meiotic inheritance of DNA methylation states in flowering plant genomes provides a stark contrast to the inheritance of DNA methylation in mammalian genomes, where genome-wide erasure of DNA methylation and reprogramming occurs each generation³⁴. This property of flowering plant genomes makes them ideal targets of induced epialleles, as once a new methylation state occurs it is often inherited in subsequent generations. Application of epimutagenesis and the use of TET-mediated engineering of DNA methylation states in economically and agriculturally significant plant species will be an interesting area of future investigation.

Methods

Synthesis and cloning of the human TET1 catalytic domain. A human TET1 catalytic domain (hTET1cd) sequence (residues 1418–2136) was synthesized by GenScript, and moved to a plant transformation compatible vector (pMDC32) using LR clonase from Life Technologies as per the manufacturer's instructions (catalog #11791100). ACT2:sfGFP-hTET1cd was subcloned by Genscript in the pMDC32 vector background using the sfGFP-TET1cd fragment from Addgene plasmid #82561. The ACT2 promoter sequence was kindly provided by Dr. Richard Meagher.

Plant transformation and screening. The hTET1cd sequence in the pMDC32 vector was transformed into *Agrobacterium tumefaciens* strain C58C1 and plated on LB-agar supplemented with kanamycin (50 µg/mL), gentamicin (25 µg/mL), and rifampicin (50 µg/mL). A single kanamycin-resistant colony was selected and used to start a 250-mL culture in LB Broth Miller liquid media supplemented with gentamicin (25 µg/mL), kanamycin (50 µg/mL), and rifampicin (50 µg/mL), which was incubated for 2 days at 30 °C. Bacterial cells were pelleted by centrifugation at 4000 rpm for 30 min and the supernatant decanted. The remaining bacterial pellet was re-suspended in 200 mL of 5% sucrose with 0.05% Silwet L77. Plant transformation was performed using the floral dip method described by Clough and Bent³⁵. Seeds were collected upon senescence and transgenic plants were identified via selection on 1/2 LS plates supplemented with Hygromycin B (25 µg/mL). 35S:TET1-1 is a T1 individual, 35S:TET1-2 is a T3 plant, 35S:TET1-3 is a T4 plant. All transgenic individuals chosen for analysis contain independent insertions of hTET1cd and are not the result of single-seed decent unless otherwise noted.

DNA and RNA isolation. *A. thaliana* leaf tissue was flash-frozen and finely ground to a powder using a mortar and pestle. DNA extraction was carried out on all samples using the DNeasy Plant Mini Kit (Qiagen), and the DNA was sheared to ~200 bp by sonication. RNA was isolated from finely ground flash-frozen leaf tissue using Trizol (Thermo Scientific). For RT-qPCR, RNA was further treated with TURBO™ DNase (Thermo Scientific) according to the manufacturer's instructions. One microgram of RNA was subsequently reverse transcribed with M-MuLV reverse transcriptase according to the manufacturer's instructions (NEB). RT-qPCR was used to analyze cDNA populations using PP2AA-3 (AT1G13320) as an endogenous control, and was performed on a Roche LightCycler 480 instrument using SYBR Green detection chemistry. The genes assayed by this method were IBM1-S, IBM1-L, FWA and sfGFP-hTET1cd. Primers used for RT-qPCR were designed using PrimerQuest from Integrated DNA Technologies (www.idtdna.com/PrimerQuest/). Primer sequences used for RT-qPCR: PP2AA-3-F: 5'-AATGAGG-CAGAAGTTCGGATAG-3', PP2AA-3-R: 5'-CAGGGAAGAATGTCTGGATAG-3', ibm1s-F: 5'-TCTTTCTTCTAAGTCTGTCCATTCT-3', ibm1s-R: 5'-GTGACC-GATTAGGAATGGTATCT-3', ibm1L-F: 5'-CCGAAGCCAAAGTGGAGATA-3', ibm1L-R: 5'-CTTCCTCTTCCGTAGACTTCTTT-3', FWA-F: 5'-CAA-GATGGTGGAAAGGATGAGAA-3', FWA-R: 5'-CTCTGTTCCTCAGTGGGAT-GAG-3', sfGFP-hTET1cd-F: 5'-CAAAGATGACGGGACCTACAA-3', sfGFP-hTET1cd-R: 5'-GTACTCGAGTTTGTGTCCAAGA-3'.

Library construction. Genomic DNA libraries were prepared following the MethylC-seq protocol without use of the bisulfite conversion step. MethylC-seq libraries were prepared as previously described in ref. ³⁶. Briefly, genomic DNA was sonicated to 200 bp using a Covaris S-series focused ultrasonicator, and end-repaired using End-It DNA end-repair kit (Epicentre). End-repaired DNA was subjected to A-tailing using Klenow 3′–5′ exo[−] (NEB) and ligated to methylated adapters using T4 DNA ligase (NEB). Ligated DNA was subsequently bisulfite converted using the EZ DNA methylation-Gold kit as per the manufacturer's instructions and amplified using KAPA HiFi uracil + Readymix Polymerase. RNA-seq libraries were constructed using Illumina TruSeq Stranded RNA LT Kit (Illumina, San Diego, CA) following the manufacturer's instructions with limited modifications. The starting quantity of total RNA was adjusted to 1.3 µg, and all volumes were reduced to a third of the described quantity. TAB-seq libraries were prepared as previously described in ref. ¹⁶. Briefly, genomic DNA was glucosylated and oxidized using T4-βGT (NEB) and recombinant mTET1. Standard bisulfite treatment is then performed using the MethylCode Bisulfite Conversion kit (Thermo Fisher Scientific).

ChIP library preparation. Leaves were treated with formaldehyde to covalently link protein to DNA, washed several times with distilled water, patted dry, and ground into fine powder in liquid nitrogen. Chromatin was extracted with a series of extraction buffers and sonicated. The final chromatin solution was incubated overnight with anti-H3K9me2 antibody (Cell Signaling Technology, 9753S)-coated dynabeads protein A (Life Technologies, 10002D) to precipitate the immune complex. After a few washes, the immune complex was eluted and incubated at 65° in the presence of a high concentration of NaCl in a water-bath overnight. After degrading the proteins with proteinase K, DNA was recovered by phenol/chloroform/isoamyl alcohol extraction followed by ethanol precipitation. The DNA pellet was then dissolved in 30 µl of nuclease-free water.

Sequencing. Illumina sequencing was performed at the University of Georgia Genomics Facility using an Illumina NextSeq 500 instrument. For MethylC-seq and TAB-seq, raw reads were trimmed for adapters and preprocessed to remove low-quality reads using cutadapt 1.9.dev1³⁷. For RNA-seq and ChIP-seq, these processes were carried out by Trimmomatic v0.32³⁸. The mutant allele of *met1* is *met1-3* and methylome data was downloaded under accession GSE39901. The mutant allele of *ibm1* is *ibm1-6*.

MethylC-seq data processing. Qualified reads were aligned to the *A. thaliana* TAIR10 reference genome as described in ref. ³⁹. Chloroplast DNA (which is fully unmethylated) was used as a control to calculate the sodium bisulfite reaction non-conversion rate of unmodified cytosines. All conversion rates were >99% (Supplementary Table 1). The list of gbM genes used in this study was previously curated¹⁸. Heat maps were clustered by complete linkage method conducted by R (<https://www.r-project.org>). All methylation levels reported in all analyses are presented as differences in absolute values, including defining DMRs and calculating hyper/hypomethylated regions. The only exception is in the comparison of mCG loss between gbM, where we used a percentage difference.

RNA-seq data processing. Qualified reads were aligned to the *A. thaliana* TAIR10 reference genome using TopHat v2.0.13⁴⁰ (Supplementary Table 2). Gene expression values were computed using Cufflinks v2.2.1⁴¹. Genes determined to have at least twofold log₂ expression changes by Cufflinks and passed tests were identified as differentially expressed genes. The Col-0 wild-type transcriptomes were downloaded using data from accession GSE75071.

TAB-seq data processing. Qualified reads were aligned to the *A. thaliana* TAIR10 reference genome using Methylpy as described in ref. ³⁹. A modified lambda DNA sequence was used to assess the quality of prepared libraries. In the added lambda sequence, all non-CG cytosines are unmethylated and CG cytosines are methylated to 5mC. The “non-conversion” rate is used to measure the rate of non-CG cytosines failing to be converted to thymines after bisulfite treatment. The “5mC non-conversion” is used to estimate the 5mCG cytosines failing to be converted to thymines after TET treatment (Supplementary Table 3).

ChIP-seq data analysis. Qualified reads were aligned to the *A. thaliana* TAIR10 reference genome using Bowtie 1.1.1 with following parameters: bowtie -m 1 -v2 --best --strata --chunkmbs 1024 -S. Aligned reads were sorted using samtools v 1.2 and clonal duplicates were removed using SAMtools version 0.1.19 (Supplementary Table 4).

Metaplot analysis. For metaplot analyses, twenty 50 bp bins were created for both upstream and downstream regions of gene bodies/TEs. Gene bodies/TE regions were evenly divided into 20 bins. Weighted methylation levels were computed for each bin⁴².

DMR analysis. Identification of DMRs was performed as described in ref. ⁴³ and adjusted *p* value (Benjamini–Hochberg correction) 0.05 was adopted as the cutoff. Only DMRs with at least five DMSs (differential methylated sites) and a 10% absolute methylation level difference within each DMR were reported and used for subsequent analysis. For coverage calculations, each sample was combined with two Col-0 WT replicates to identify DMRs. Each sample was compared with both Col-0 WT replicates separately and for a DMR to be identified it must have been identified in both comparisons. Absolute methylation differences of ±(50% for CG, 10% for CHG and CHH) were defined as hyper/hypomethylation, respectively. DMRs overlapping regions with mCG ≥5%, mCHG and mCHH ≥1% in both two Col-0 WT replicates were defined as RdDM-like regions. DMRs overlapping regions with mCG ≥5%, mCHG and mCHH <1% in both two Col-0 WT replicates were defined as gbM regions. DMRs overlapping regions with all three contexts less methylated at less than 1% in both Col-0 WT replicates were defined as unmethylated regions. Overlap comparisons were performed using bedtools v2.26.0⁴⁴.

Data availability. The data generated from this study have been uploaded to Gene Expression Omnibus (GEO) database and can be retrieved through accession number GSE93024.

Received: 15 May 2017 Accepted: 2 February 2018

Published online: 01 March 2018

References

- Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).
- Johannes, F. et al. Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* **5**, e1000530 (2009).
- Reinders, J. et al. Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. *Genes Dev.* **23**, 939–950 (2009).
- Cortijo, S. et al. Mapping the epigenetic basis of complex traits. *Science* **343**, 1145–1148 (2014).
- Bewick, A. J. et al. On the origin and evolutionary consequences of gene body DNA methylation. *Proc. Natl Acad. Sci. USA* **113**, 9111–9116 (2016).
- Cortijo, S., Wardenaar, R., Colome-Tatche, M., Johannes, F. & Colot, V. Genome-wide analysis of DNA methylation in Arabidopsis using MeDIP-Chip. *Methods Mol. Biol.* **1112**, 125–149 (2014).
- Reinders, J. et al. Genome-wide, high-resolution DNA methylation profiling using bisulfite-mediated cytosine conversion. *Genome Res.* **18**, 469–476 (2008).
- Hofmeister, B. T., Lee, K., Rohr, N. A., Hall, D. W. & Schmitz, R. J. Stable inheritance of DNA methylation allows creation of epigenotype maps and the study of epiallele inheritance patterns in the absence of genetic variation. *Genome Biol.* **18**, 155 (2017).
- Li, Q. et al. Genetic perturbation of the maize methylome. *Plant Cell* **26**, 4602–4616 (2014).
- Hu, L. et al. Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proc. Natl Acad. Sci. USA* **111**, 10642–10647 (2014).
- Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
- Pastor, W. A. et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).
- Ito, S. et al. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
- Hollwey, E., Watson, M. & Meyer, P. Expression of the C-terminal domain of mammalian TET3 DNA dioxygenase in Arabidopsis thaliana induces heritable methylation changes at rDNA loci. *Adv. Biosci. Biotechnol.* **7**, 243–250 (2016).
- Erdmann, R. M., Souza, A. L., Clish, C. B. & Gehring, M. 5-hydroxymethylcytosine is not present in appreciable quantities in Arabidopsis DNA. *G3 (Bethesda)* **5**, 1–8 (2014).
- Yu, M. et al. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat. Protoc.* **7**, 2159–2170 (2012).
- Stroud, H., Greenberg, M. V., Feng, S., Bernatavichute, Y. V. & Jacobsen, S. E. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* **152**, 352–364 (2013).
- Niederhuth, C. E. et al. Widespread natural variation of DNA methylation within angiosperms. *Genome Biol.* **17**, 194 (2016).
- Saze, H., Shiraishi, A., Miura, A. & Kakutani, T. Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana. *Science* **319**, 462–465 (2008).

20. Miura, A. et al. An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J.* **28**, 1078–1086 (2009).
21. Rigal, M., Kevei, Z., Pelissier, T. & Mathieu, O. DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns. *EMBO J.* **31**, 2981–2993 (2012).
22. Lister, R. et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**, 523–536 (2008).
23. Rigal, M., Kevei, Z., Pelissier, T. & Mathieu, O. DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns. *EMBO J.* **31**, 2981–2993 (2012).
24. Soppe, W. J. et al. 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).
25. Finnegan, E. J., Peacock, W. J. & Dennis, E. S. Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. *Proc. Natl Acad. Sci. USA* **93**, 8449–8454 (1996).
26. Ikeda, Y., Kobayashi, Y., Yamaguchi, A., Abe, M. & Araki, T. Molecular basis of late-flowering phenotype caused by dominant epi-alleles of the FWA locus in Arabidopsis. *Plant Cell Physiol.* **48**, 205–220 (2007).
27. Baubec, T., Finke, A., Scheid, O. M. & Pecinka, A. Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. *EMBO Rep.* **15**, 446–452 (2014).
28. An, Y. Q. et al. Strong, constitutive expression of the Arabidopsis ACT2/ACT8 actin subclass in vegetative tissues. *Plant J.* **10**, 107–121 (1996).
29. Choudhury, S. R., Cui, Y., Lubecka, K., Stefanska, B. & Irudayaraj, J. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* **7**, 46545–46556 (2016).
30. Maeder, M. L. et al. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat. Biotechnol.* **31**, 1137–1142 (2013).
31. Vojta, A. et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* **44**, 5615–5628 (2016).
32. Mendenhall, E. M. et al. Locus-specific editing of histone modifications at endogenous enhancers. *Nat. Biotechnol.* **31**, 1133–1136 (2013).
33. Liu, X. S. et al. Editing DNA methylation in the mammalian genome. *Cell* **167**, 233–247.e217 (2016).
34. Heard, E. & Martienssen, R. A. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* **157**, 95–109 (2014).
35. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* **16**, 735–743 (1998).
36. Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J. & Ecker, J. R. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* **10**, 475–483 (2015).
37. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
38. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
39. Schmitz, R. J. et al. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. *Genome Res.* **23**, 1663–1674 (2013).
40. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
41. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
42. Schultz, M. D., Schmitz, R. J. & Ecker, J. R. ‘Leveling’ the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet.* **28**, 583–585 (2012).
43. Schultz, M. D. et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015).
44. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

Acknowledgements

We thank Nathan Springer for comments and discussions on this study as well as the Georgia Genomics & Bioinformatics Core and the Georgia Advanced Computing Resource Center for technical support. This work was supported by the National Science Foundation (MCB-1650331), by The Pew Charitable Trusts and by the Office of the Vice President of Research at UGA to R.J.S. C.H. was supported by the National Institutes of Health (NIH HG006827) and is an investigator of the Howard Hughes Medical Institute. W.T.J. was supported by a Scholars of Excellence Fellowship from the University of Georgia and National Institute of General Medical Sciences of the National Institutes of Health award number T32GM007103. The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institutes of Health.

Author contributions

R.J.S. conceived the project. L.J., W.T.J. and R.J.S. designed the experiments. L.J., W.T.J., X.S., L.H. and C.H. performed the research. L.J., W.T.J. and R.J.S. analyzed the data. R.J.S. wrote the paper. All authors discussed the results and commented on the manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-03289-7>.

Competing interests: R.J.S. has a provisional patent pending for the development of this technology as it applies to epigenome engineering of plants, and declares no other competing interests. The remaining authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018