TET-mediated epimutagenesis of the *Arabidopsis thaliana* methylome

Lexiang Ji[a,*], William T. Jordan[b,*], Xiuling Shi[b], Lulu Hu[c], Chuan He[c], Robert J. Schmitz[b]

[a]Institute of Bioinformatics, University of Georgia, Athens, GA, 30602, USA.

[b]Department of Genetics, University of Georgia, Athens, GA, 30602, USA.

[c]Department of Chemistry, Department of Biochemistry and Molecular Biology, and Institute for Biophysical Dynamics, Howard Hughes Medical Institute, University of Chicago, Chicago, IL 60637, USA

*Indicates these authors contributed equally to this work

Corresponding author:

Robert J. Schmitz

University of Georgia

120 East Green Street, Athens, GA 30602

Telephone: (706) 542-1887

Fax: (706) 542-3910

E-mail: schmitz@uga.edu

Keywords: DNA methylation, TET, epigenetics, epigenomics, epimutagenesis

DNA methylation in the promoters of plant genes often leads to transcriptional repression, and the wholesale removal of DNA methylation in methyltransferase mutants results in severe gene expression and developmental defects. However, many cases of naturally-occurring DNA methylation variations have been reported, where the differential expression of differentially methylated genes are responsible for agronomically important traits. The ability to manipulate plant methylomes to generate populations of epigenetically distinct plants could provide invaluable resources for breeding and research. Here we describe a novel "epimutagenesis" method to rapidly generate methylation variations 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 and the redistribution of heterochromatin, mimicking mutants in the maintenance DNA methyltransferase *met1*. Application of TET-mediated epimutagenesis to agriculturally significant plants may result in differential expression of alleles typically silenced by DNA methylation, uncovering previously hidden traits.

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 DNA methylation that can result in stably inherited changes of both gene expression and phenotypes.

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*[2,5-7]. 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 that 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[8,9]. 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 funding parents, epimutagenesis introduces random methylation variations. Here we describe a novel epimutagenesis approach in *A. thaliana* using a human Ten-eleven translocation (TET1) methylcytosine dioxygenase[10-12], which catalyzes the conversion of 5-methylcytosine (5mC) to 5-

hydroxymethylcytosine (5hmC). Although TET enzymes or their primary product, 5hmC, are not found in plant genomes[13], ectopic expression of a human TET enzyme resulted in widespread DNA demethylation, redistribution of heterochromatin and induced phenotypic variations in *A. thaliana*.

RESULTS

Overexpressing Tet1 in Arabidopsis hypomethylates the genome

Transgenic *A. thaliana* plants were generated expressing the catalytic domain 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**). The results revealed a global reduction of CG methylation levels 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 *A. thaliana* CHG and CHH methylation were not as strong as it was for 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 activity. Taken together, these results showed that the expression of hTET1cd resulted 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 plants[14]. No detectable levels of 5hmC were found in either of 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 and d**). To further quantify the observed hypomethylation, metaplots were created for genes and transposons, respectively (**Fig. 1e and f, Supplementary Fig. 1c-f**). Strong reduction of mCG and 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) have lost at least half of their CG methylation in epimutagenized lines. Collectively, these results indicate 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. 56,283 CG DMRs ranging in size from 6 - 20,286 base pairs (bp) were identified (**Fig. 1g**). Of these, 38.7% were located in intergenic sequences, 53.7% overlapped with genes and 7.6% were located in promoter regions (1kb upstream of a gene). Like the *met1* mutant, 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 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 regions. 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 with the long-term stability of hypomethylation, may facilitate inheritable 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 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 regions that contained at least 20% CHG methylation and 7,214 overlapped regions that contained at least 10% CHH methylation in wild-type individuals. 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 and i**). 2,341 and 3,447 regions lost more than 10% CHG methylation in 35S:TET1-1 and 35S:TET1-2, respectively, whereas 2,475 and 3,379 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 and h)**. 1,708 (73.0%) and 2,386 (69.2%) regions that have lost more than 10% mCHG in 35S:TET1-1 and

35S:TET1-2 have reduced levels in *met1*, whereas 2,013 (81.6%) and 2,563 (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 epiRILs[15].

Tet1-mediated redistribution of heterochromatin

Mutations in *met1* also leads to hypermethylation of CHG sites in gene bodies due to the loss of CG methylation in the 7th intron of the histone 3 lysine 9 (H3K9) demethylase, *INCREASED IN BONSAI METHYLATION 1* (*IBM1*)[16-19]. This results in alternative splicing of *IBM1,* ultimately producing a non-functional gene product and resulting in ectopic accumulation of di-methylation of H3K9 (H3K9me2) throughout the genome. As in *met1*, the 7th intron of *IBM1* was hypomethylated in 35S:TET1-1, 35S:TET1-2 and 35S:TET1-2.5, a line that was propagated for an additional two generations (**Fig. 2a**). A subtle increase in global CHG methylation was observed in 35S:TET1-1 and 35S:TET1-2 (**Fig. 1a**). However, extensive variation in genome-wide gains and losses of CHG methylation was observed in line 35S:TET1-2.5 as it displayed approximately 1.8 Mb of additional CHG methylation (**Fig. 2b**). This indicates that the expression of hTET1cd over a longer period of time not only leads to loss of methylation, but also results in redistribution of DNA methylation genome-wide.

To further characterize regions of differential CHG methylation, identified CHG DMRs were categorized into discrete groups based on their DNA methylation status in wild-type individuals. Of the 9,917 CHG DMRs identified, 1,460 were in loci that are defined as gbM in wild-type individuals, 584 were in unmethylated regions, and 6,940 of them were in RdDM-like regions (**Fig. 2c-e**). Interestingly, in line 35S:TET1-2.5, 1,408 (96.4%) of the CHG DMRs in gbM-like loci gained CHG hypermethylation, whereas

2,156 (31.1%) of the CHG DMRs in RdDM-like regions lost CHG, in contrast to 595 (8.6%) RdDM-like regions that gained CHG methylation. Lastly, there were 502 (86.0%) loci that are unmethylated in wild-type individuals that gain CHG methylation as well as CG and CHH methylation in the epimutagenized lines (**Fig. 2c-e**). 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 plants and compared the results to *met1* and *ibm1* transcriptomes. Compared to wild-type plants, 629 and 736 up-regulated genes and 1,277 and 1,428 down-regulated genes were identified in 35S:TET1-1 and 35S:TET1-2, respectively. There was a high level of overlapping in transcriptome changes with 35S:TET1-1 and 35S:TET1-2 compared to *met1* and *ibm1* (**Fig. 2f and g**). Of the genes up-regulated in *met1,* 36.8% and 21.5% overlapped with 35S:TET1-1 and 35S:TET1-2, respectively. An even greater overlap was observed with down-regulated genes as 60.1% and 72.9% of down-regulated genes in 35S:TET1-1 and 35S:TET1-2 overlapped with down-regulated genes in *met1*, respectively. These results reveal that expressing the catalytic domain of human *TET1* in *A. thaliana* is a viable approach to access 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 growth to flowering (**Fig. 3a and b**). We hypothesized that the observed late-flowering phenotype was associated with the demethylation of the *FWA* (*FLOWERING WAGENINGEN*) locus, as is observed in *met1* mutants[20,21]. 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 apex[22]. 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).

DISCUSSION

The discovery that expression of the catalytic domain of the human TET1 protein in *A. thaliana* leads to widespread loss of CG methylation and redistribution of heterochromatin makes it possible to create epimutants without the need for methyltransferase mutants, which often causes lethality in crops. In addition to epimutagenesis, TET1cd could be used in combination with sequence-specific targeting machinery such as CRISPR-dCas9 to direct DNA demethylation in plant genomes, as has been demonstrated in mammalian systems[23-27]. 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 generation[28]. 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.

ACKNOWLEDGEMENTS

We thank Nathan Springer for comments and discussions on this study as well as the Georgia Genomics Facility 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.

AUTHOR CONTRIBUTIONS

R.J.S. conceived the project. L.J., W.T.J. and R.J.S. designed experiments. L.J., W.T.J., X.S., L.H., and C.H. performed 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.

COMPETING FINANCIAL INTERESTS

A provisional patent is pending for the development of this technology as it applies to epigenome engineering of plants

Figure 1. Overexpression of TET1 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 and 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.

Figure 2. Widespread redistribution of heterochromatin in 35S:TET1 plants. (a) Genome browser view of *IBM1* (AT3G07610) in Col-0 WT, three 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 was lost from the large intron (purple vertical lines = CG methylation, blue vertical lines = CHG methylation and gold vertical lines = CHH methylation). (b) 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. (c) Heat map of CHG methylation displaying CHG DMRs. Corresponding CG and CHH methylation levels are shown in (d) and (e). Heat maps showing $\log_2$ transformed FPKM profiles of up-regulated genes (f) and down-regulated genes (g) in two 35S:TET1 transgenic individuals, *met1* and *ibm1* mutants relative to WT.

Figure 3. 35S:TET1 plants have a delayed flowering phenotype. (a) Photographs of one 35S:TET1-1 transgenic plant and Col-0 WT plant and (b) corresponding number of rosette leaves. (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*.

METHODS

Synthesis and cloning of the human TET1 catalytic domain. A human TET1 catalytic domain (hTET1-CD) sequence was synthesized by GenScript, and moved to a plant transformation compatible vector (pMDC32) using LR clonase from Life Technologies per the manufacturer's instructions (Catalog #11791100).

Plant transformation and screening. The hTET1-CD 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 (15 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 (15 µg/mL), which was incubated for two days at 30˚C. Bacterial cells were pelleted by centrifugation at 4,000 RPM for 30 minutes 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 Bent[29]. Seeds were harvested upon senescence and transgenic plants were identified via selection on ½ LS plates supplemented with Hygromycin B (15 µg/mL). 35S:TET1-1 is a T1 individual, whereas 35S:TET1-2 is a T3 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 sheered to approximately 200 bp by sonication. RNA was isolated from finely ground flash-frozen leaf tissue using Trizol (Thermo Scientific).

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[30]. 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[14].

Sequencing. Illumina sequencing was performed at the University of Georgia Genomics Facility using an Illumina NextSeq 500 instrument. Methylomes, and 5-hydroxymethylomes were sequenced to 150 bp whereas transcriptomes were sequenced to 75 bp. For MethylC-seq and TAB-seq, raw reads were trimmed for adapters and preprocessed to remove low quality reads using cutadapt 1.9.dev1[31]. For RNA-seq, these processes were carried out by Trimmomatic v0.32[32].

MethylC-seq data processing. Qualified reads were aligned to the *A. thaliana* TAIR10 reference genome as described in[33]. 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 curated[15]. 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 v2.0.13[34] (**Supplementary Table 2**). Gene expression values were computed using Cufflinks v2.2.1[35]. Genes determined to have at least 2-fold $\log_2$ expression changes by Cufflinks were identified as differentially expressed genes.

TAB-seq data processing. Qualified reads were aligned to the *A. thaliana* TAIR10 reference genome using Methylpy as described in[33]. The control 5mC modified lambda

DNA sequence was used to calculate the 5mC non-conversion rate upon TET and bisulfite treatment. Non-CG dinucleotide sites were used to compute the non-conversion rate of unmodified cytosines upon bisulfite treatment (**Supplementary Table 3**).

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 bin as described previously[36].

DMR analysis. Identification of DMRs was performed as described in[37]. Only DMRs with at least 5 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/hypo methylation, 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 v2.26.0[38].

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

References

1. Law, J.A. & Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**, 204-20 (2010).
2. Johannes, F. *et al.* Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet* **5**, e1000530 (2009).
3. Reinders, J. *et al.* Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. *Genes Dev* **23**, 939-50 (2009).
4. Cortijo, S. *et al.* Mapping the Epigenetic Basis of Complex Traits. *Science* (2014).
5. Bewick, A.J. *et al.* On the origin and evolutionary consequences of gene body DNA methylation. *Proc Natl Acad Sci U S A* **113**, 9111-6 (2016).
6. 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**, 125-49 (2014).
7. Reinders, J. *et al.* Genome-wide, high-resolution DNA methylation profiling using bisulfite-mediated cytosine conversion. *Genome Res* **18**, 469-76 (2008).
8. Li, Q. *et al.* Genetic Perturbation of the Maize Methylome. *Plant Cell* (2014).
9. 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 U S A* **111**, 10642-7 (2014).
10. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-5 (2009).
11. Pastor, W.A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394-7 (2011).
12. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-33 (2010).
13. 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).
14. Yu, M. *et al.* Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat Protoc* **7**, 2159-70 (2012).
15. Niederhuth, C.E. *et al.* Widespread natural variation of DNA methylation within angiosperms. *Genome Biol* **17**, 194 (2016).
16. Saze, H., Shiraishi, A., Miura, A. & Kakutani, T. Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana. *Science* **319**, 462-5 (2008).
17. Miura, A. *et al.* An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J* **28**, 1078-86 (2009).
18. 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**, 2981-93 (2012).
19. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**, 523-36 (2008).

20. 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).
21. Finnegan, E.J., Peacock, W.J. & Dennis, E.S. Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. *Proc Natl Acad Sci U S A* **93**, 8449-54 (1996).
22. 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**, 205-20 (2007).
23. Choudhury, S.R., Cui, Y., Lubecka, K., Stefanska, B. & Irudayaraj, J. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* (2016).
24. Maeder, M.L. *et al.* Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* **31**, 1137-42 (2013).
25. Vojta, A. *et al.* Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* (2016).
26. Mendenhall, E.M. *et al.* Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol* **31**, 1133-6 (2013).
27. Liu, X.S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167**, 233-247 e17 (2016).
28. Heard, E. & Martienssen, R.A. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* **157**, 95-109 (2014).
29. Clough, S.J. & Bent, A.F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* **16**, 735-43 (1998).
30. 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**, 475-83 (2015).
31. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, pp. 10-12 (2011).
32. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20 (2014).
33. Schmitz, R.J. *et al.* Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. *Genome Res* **23**, 1663-74 (2013).
34. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
35. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-5 (2010).
36. Schultz, M.D., Schmitz, R.J. & Ecker, J.R. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet* **28**, 583-5 (2012).
37. Schultz, M.D. *et al.* Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212-6 (2015).

38. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).