The interplay between DNA methylation and gene expression in *Arabidopsis thaliana*

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard-Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Thanvi Srikant

aus Coimbatore, Indien

Tübingen

2021

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard-Karls Universität Tübingen

Tag der mündlichen Qualifikation:	19.11.2021
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Detlef Weigel
2. Berichterstatter:	Prof. Dr. Dirk Schwarzer

ACKNOWLEDGEMENTS

Firstly, I would like to thank Detlef Weigel for being the best PhD supervisor that any student could ask for. I will be ever thankful for his belief in my capabilities, his constant guidance, and for how he actively inspires me to continue my academic pursuits. Rebecca Schwab and Wei Yuan have been instrumental for my progress and growth in the lab. They are truly role models for women in science - their zeal for perfection and ability to efficiently handle a multitude of responsibilities always motivates me to do better. I would like to thank Anjar Wibowo, Hajk-Georg Drost and Chang Liu for all their help and advice, and for inspiring me to continue working with epigenetics. I extend my thanks to the other members of my Thesis advisory committee, Prof. Dirk Schwarzer and Dr. Martin Bayer.

I have been fortunate to find some very good friends during my PhD - Bridgit Waithaka-Vasilljevic, Cristina Barragan, Sergio Latorre, Effie Symeonidi, Adrian Contreras, Alejandra Duque, Ulrich Lutz and Miriam Lucke. Their constant enthusiasm for life and for science have kept me cheerful during these last four years. Both past- and present members of the Weigel lab have nurtured a healthy and happy lab environment, where everyone helps and learns from each other, but most importantly, puts their best foot forward to make hilarious Christmas videos! I would especially like to mention Derek Lundberg, for being an amazing bench partner and for inspiring me with his innovative thinking. I am thankful for meeting and learning from Wanyan Xi during my first tryst with the Weigel lab, someone who has always been encouraging towards my scientific progress.

Finally, I would not have come this far if it were not for my family. I especially thank my mother for her perpetual positivity and prayers, my father for trusting my decisions and always standing by my side, and my brother for making me appreciate the small joys in life. I am truly grateful to Prateek, for being the perfect partner both in life and in science, and for cooking delicious Indian food after my long experiment marathons. I thank my thatha, paati, Gma, my in-laws and my extended family for all their warmth and kindness. I feel privileged to have two aunts who are both successful female academics and have encouraged me to pursue research from my early years. A special thanks to Popsi athhai for ensuring that I stay balanced even during chaotic times and for being a constant cheerleader for my academic career. My closest friends from India have always had my back through thick and thin, and I thank them for making me laugh when I needed it the most. Lastly, I would like to thank all my teachers for encouraging me to strive for both excellence and happiness.

TABLE OF CONTENTS

1. **ZUSAMMENFASSUNG**……..**5-6**
2. **SUMMARY**……**7-8**
3. **PUBLICATIONS**……..**9**
4. **INTRODUCTION**
   - **4.1.** Epigenetic modifications in plants……**10-11**
   - **4.2.** DNA methylation and its establishment……**11**
     - **4.2.1.** *de novo* methylation pathways……**12-14**
     - **4.2.2.** Maintenance of methylation……**14-15**
   - **4.3.** DNA demethylation……**15-16**
   - **4.4.** Gene expression regulation by DNA methylation
     - **4.4.1.** Promoter methylation……..**16**
     - **4.4.2.** Gene-body methylation……..**17**
     - **4.4.3.** Methylation at transposable elements……**17-19**
   - **4.5.** Methylation patterns during developmental transitions……....**19-22**
   - **4.6.** Plant epialleles and their formation……..**22-24**
   - **4.7.** MET1-dependent epigenetic control in *A. thaliana*.……..**24-26**
   - **4.8.** Chromatin accessibility landscape in plants
     - **4.8.1.** Interplay with DNA methylation and other epigenetic features…..**27-28**
     - **4.8.2.** Identification of *cis* regulatory elements in accessible chromatin..**28-29**
     - **4.8.3.** Stress response facilitated by accessible chromatin……...**29-30**
   - **4.9.** The model plant *Arabidopsis thaliana* and its natural genetic and epigenetic diversity
     - **4.9.1.** Genetic variation……....**30-31**
     - **4.9.2.** Methylome variation……..**31-32**
     - **4.9.3.** Transcriptome, transcription factor binding and chromatin accessibility variation……...**32-33**
     - **4.9.4.** Genetic basis of epigenetic variation……..**33-34**
5. **PROJECT AIMS….**……...**35**
6. **CHAPTER ONE :** Position-dependent effects of cytosine methylation on *FWA* expression in *Arabidopsis thaliana*
   - **6.1.** Abstract……..**36**
   - **6.2.** Author Contributions……**37**
7. **CHAPTER TWO :** Accession-specific and shared responses to genome wide hypomethylation in *Arabidopsis thaliana*
   - **7.1.** Abstract……..**38**
   - **7.2.** Author Contributions……**39**
8. **DISCUSSION**
   - **8.1.** Fine-tuning of gene expression by methylation may not be position-dependent……**40-41**
   - **8.2.** A large-scale approach: Which genes in the genome rely on methylation for their regulation and what are the different ways they can be regulated? ……..**41-42**
   - **8.3.** The challenges of propagating *met1* mutants : similar but different parent-origin effects?……**42-44**

8.4. The molecular consequences of genome-wide hypomethylation: Epigenomic imbalance…….**44-45**
8.5. TEs are upregulated in *met1* mutants and can mobilize over generations……..**45-47**
8.6. Epigenetic regulation at Non-TE genes is complex and distinct from TE-genes…..……...**47-49**
8.7. Investigating global and accession-specific gene-regulatory patterns in Non-TE-DEGs……**49-50**
8.8. Non-TE-DEGs comprise many known epialleles which can affect mutant phenotypes……..**51-53**
8.9. MET1 is required to maintain stable chromatin architecture……...**53-56**
9. CONCLUSION AND PERSPECTIVE……..**56-58**
10. REFERENCES….……...**59-73**
11. APPENDIX
I. Position-dependent effects of cytosine methylation on *FWA* expression in *Arabidopsis thaliana*
II. Accession-specific and shared responses to genome wide hypomethylation in *Arabidopsis thaliana*

ZUSAMMENFASSUNG

Pflanzen müssen sich trotz ihrer Unbeweglichkeit häufig an eine sich verändernde Umwelt anpassen und verfügen daher über ein breites Spektrum an Regulierungsmechanismen, um die Genexpression unter ungünstigen Bedingungen fein abzustimmen. Epigenetische Veränderungen wie kovalente Modifikationen der DNA oder der mit ihr verbundenen Histonproteine sowie die optimale Verpackung des Chromatins innerhalb der Zelle können die Genexpression stark beeinflussen, ohne die zugrunde liegende genetische Sequenz zu verändern. Die DNA-Cytosin-Methylierung, bei der ein Methylrest an das Cytosin-Nukleotid angehängt wird, ist eine gut untersuchte epigenetische Markierung, die in mehreren Blütenpflanzen vorkommt und dafür bekannt ist, dass sie Veränderungen der Genexpression an vielen genomischen Loci hervorruft. In Pflanzen tritt die DNA-Methylierung in drei verschiedenen Nukleotidkontexten auf - CG, CHG und CHH (wobei H = A, T, C) - und ist vorwiegend in heterochromatischen Regionen mit sich wiederholende Sequenzen inaktiven Transposons zu finden. In dieser Dissertation habe ich die Modellpflanze *Arabidopsis thaliana* untersucht um herauszufinden, wie DNA-Methylierung die Genexpression direkt und indirekt regulieren kann, wenn sie in *cis* zu Genen auftritt, und wie sie die Genom- und Epigenomstabilität beeinflussen kann.

In meinem ersten Projekt führte ich künstlich eine unterschiedliche Methylierung in verschiedenen Promotorregionen eines Gens ein, um zu untersuchen, ob die nachgeschaltete Genexpression davon beeinflusst wird. Als Beispiel diente mir das bekannte Epiallel des *FWA* Gens - ein Allel, bei dem die normalen Methylierungsmarkierungen im Promotor fehlen und dadurch die Expression des Gens aktivieren. Ich habe transgene Linien erzeugt, die das *fwa* Epiallel tragen, wobei die Methylierung auf drei verschiedene Regionen des *FWA* Promotors zielte, und in diesen Pflanzen die induzierte Methylierung, die veränderte Genexpression und den Blühzeitpunkt gemessen. Ich fand heraus, dass die Methylierung an Tandem-Sequenzwiederholungen in der Nähe des Gens die größten Auswirkung auf die nachgeschaltete Genexpression hatte, während die Auswirkung minimal war, wenn die Methylierung weiter entfernt von dem Gen stattfand.

Als nächstes untersuchte ich die Folgen einer genomweiten Hypomethylierung. Ich erzeugte CRISPR-Cas9-Knockouts des *MET1* Gens, das für eine DNA-Methyltransferase kodiert, die die CG-Methylierung katalysiert, in 18 natürlichen Akzessionen von *A. thaliana*. Die Mutanten wiesen starke phänotypische Defekte auf und wurden mit entsprechenden Wildtyp-Linien verglichen, um die unterschiedliche Methylierung, Chromatin-Zugänglichkeit

und Genexpression mittels Bisulfit-Sequenzierung, ATAC-Sequenzierung und RNA-Sequenzierung zu untersuchen. Erstens stellte ich fest, dass alle *met1* Mutanten sehr wenige genomweite CG-Methylierung haben, wie von der Inaktivierung von *MET1* zu erwarten war. Darüber hinaus wiesen diese Mutanten große Veränderungen in der Chromatinzugänglichkeit und in der Genexpression auf, die sich jedoch quantitativ zwischen den verschiedenen Akzessionen unterschieden. Während viele Gene, die mit Transposons assoziiert sind, in *met1* Mutanten aktiviert wurden, traten die größten Unterschiede in der Genexpression zwischen den Akzessionen bei Protein kodierenden Genen auf, von denen viele bekannte Epiallele haben. Schließlich entdeckte ich mehrere epigenetische Zustände derselben Gene in verschiedenen Akzessionen, was auf komplexe und einzigartige Regulierungsmechanismen im Zusammenhang mit der DNA-Methylierung hinweist.

Zusammengenommen zeigen diese Ergebnisse, dass epigenetische Muster oft eng miteinander verknüpft sind und dass eine Störung der Methylierung ausreicht, um mehrere direkte und indirekte nachgeschaltete Effekte zu katalysieren, die den Phänotyp der Pflanze bestimmen. Letztlich kann das Verständnis der epigenetischen Mechanismen, die für die verschiedenen Akzessionen einzigartig sind, erklären, wie sich natürliche Populationen unabhängig voneinander entwickelt haben, um eine optimale genetische und epigenetische Stabilität in ihren verschiedenen natürlichen Lebensräumen zu erhalten.

SUMMARY

Plants often need to adapt to changing environments despite their immobility, and employ a wide range of regulatory controls to fine-tune gene expression under adverse conditions. Epigenetic changes such as covalent modifications to DNA or their associated histone proteins, and optimal packaging of chromatin within the cell can strongly influence gene expression without changing the underlying genetic sequence. DNA cytosine methylation, which refers to the addition of a methyl residue to cytosine nucleotide, is a well-studied epigenetic mark that is prevalent in several flowering plants and is known to induce gene expression changes at many genomic loci. In plants, DNA methylation occurs in three different nucleotide contexts - CG, CHG and CHH (where H = A,T,C) and is predominantly found at heterochromatic regions that carry repeat sequences and silenced transposable elements. In this dissertation, I studied the model plant species *Arabidopsis thaliana* to investigate how DNA methylation can directly and indirectly regulate gene expression when it occurs in *cis* to genes, and how it can affect genome and epigenome stability.

In my first project, I artificially introduced differential methylation in various promoter regions of a gene, to investigate whether downstream gene expression would be affected. I used the well-known *fwa* epiallele as an example - a locus where methylation marks are absent in the promoter, and thereby activate the expression of the *FWA* gene. I generated transgenic lines carrying the *fwa* epiallele with methylation targeted at three different regions of the *FWA* promoter and measured the induced methylation, altered gene expression and the flowering phenotype associated with the gene. I found that methylation at tandem-repeats in proximity to the gene exerted the greatest impact on downstream gene expression, as opposed to minimal impact when methylation was targeted further upstream.

Next, I investigated the consequences of genome-wide hypomethylation. I generated CRISPR-Cas9 knockouts of the *MET1* gene (which encodes a DNA methyltransferase catalyzing CG methylation) in 18 natural accessions of *A. thaliana*. These mutant lines showed severe phenotypic defects and were compared with respective wild-type lines to examine differential methylation, chromatin accessibility and gene expression using bisulfite sequencing, ATAC sequencing and RNA sequencing. Firstly, I found that all *met1* mutant lines exhibited very low levels of genome-wide CG methylation resulting from the inactivation of MET1. Furthermore, these mutants exhibited large changes in their chromatin accessibility and gene expression profiles, which quantitatively varied across different accessions. While many genes associated with transposable elements were activated in

met1 mutants, the largest variation in gene expression between accessions occurred at protein-coding genes, many of which included known epialleles. Finally, I uncovered multiple epigenetic states of the same genes in different accessions, indicating complex and unique regulatory mechanisms associated with DNA methylation.

Together, these results show that epigenetic patterns are often tightly linked to each other, and that perturbation of methylation is sufficient to catalyse multiple direct and indirect downstream effects which determine plant phenotype. Ultimately, understanding epigenetic mechanisms unique to different accessions can explain how natural populations have independently evolved to maintain optimal genetic and epigenetic stability in their diverse natural habitats.

PUBLICATIONS

Pre-prints and advanced manuscripts included in this dissertation:

Srikant, T., Wibowo, A., Schwab, R. & Weigel, D. Position-dependent effects of cytosine methylation on FWA expression in *Arabidopsis thaliana*. *bioRxiv* 774281 (2019) doi:10.1101/774281.

Srikant, T., Yuan W., Schwab R., Berendzen, B., Weigel, D. Accession-specific and shared responses to genome-wide hypomethylation in *Arabidopsis thaliana.* (In preparation)

Publications not included in this dissertation:

Monroe, J. G., **Srikant, T.**, Carbonell-Bejerano, P., Becker, C., Lensink, M., Exposito-Alonso, M., Klein, M., Hildebrandt, J., Neumann, M., Kliebenstein, D., Weng, M., Imbert, E., Ågren, J., Rutter, M. T., Fenster, C. B. Weigel, D. Mutation bias reflects natural selection in *Arabidopsis thaliana. Nature* (2021). (in press)

Srikant, T. & Drost, H.-G. How stress facilitates phenotypic innovation through epigenetic diversity. *Front. Plant Sci.* **11**, 606800 (2021)

Srikant, T. & Wibowo, A. T. The underlying nature of epigenetic variation: origin, establishment, and regulatory function of plant epialleles. *Intern. J. Mol. Sci.* **22**, 8618 (2021)

Advanced manuscripts not included in this dissertation:

Karasov, T. L., Neumann, M., Shirsekar, G., Monroe, J.G., PATHODOPSIS Team*, Weigel, D. & Schwab, R., Drought selects not only on *Arabidopsis* populations but also their microbiomes. *Barragan, C., Bezrukov, I., González, A. H., Hildebrandt, J., Kersten, S., Lang, P., Latorre, S., Lucke, M.,,Habring, A., Friedemann, C., Paul, F. ,Lundberg, D., Lutz, U., Rabanal, F., Regalado, J., **Srikant, T.,** Waithaka, B., Wibowo, A., Yuan, W.

Yuan, W., Beitel, F., **Srikant, T.,** Bezrukov, I., Schäfer, S., Kraft, R., Weigel, D. Pervasive under-dominance in gene expression systematically explains hybrid vigor of biomass in *Arabidopsis thaliana**.***

INTRODUCTION

4.1. Epigenetic modifications in plants

While DNA encodes the primary genetic information in the cell, covalent modifications to both DNA and the protein components that package DNA into chromatin can shape cell identity. These modifications, referred to as 'epigenetic', play crucial roles in gene transcription, independent of their underlying genetic sequences. The concept of epigenetics goes back to Conrad Waddington [1].

Plants are masters of epigenetic regulation, as their genomes carry almost all of the epigenetic machinery identified in eukaryotes. Being immobile unlike animals, and facing therefore more adverse environmental changes than animals, they rely on complex gene regulatory mechanisms for short-term acclimation and long-term adaptation, thereby maximizing their reliance on epigenetic modifications. One notable example of an epigenetic modification is DNA 5-methylcytosine (^{5}mC) methylation, referring to the presence of a methyl residue at the 5' C position in cytosine nucleotides. ^{5}mC DNA methylation is well-known to alter gene transcriptional activity in plants [2]. While ^{5}mC methylation has been widely studied in many eukaryotes, recent work has identified the presence of ^{4}mC (4'-N-methyl-cytosine) in the early land plant *Marchantia polymorpha* [3], a mark previously found only in prokaryotes [4].

Within the cell, several layers of compaction are involved in chromatin formation. The fundamental repeat unit of chromatin is the nucleosome, which consists of approximately 147 bp of DNA wrapped around a core of eight histone proteins, H2A, H2B, H3 and H4, each occurring twice. Another histone protein, H1, functions as a linker histone and binds to short stretches of linker DNA between nucleosomes [5]. Post-translational modifications of histones, such as methylation, acetylation, phosphorylation, ubiquitination, biotinylation and ADP-ribosylation occurring largely at lysine and arginine residues, are also notable epigenetic modifications, each with a unique regulatory potential, as identified in *Arabidopsis thaliana* [6,7]. The most well-studied histone epigenetic marks are histone H3K4 methylation (H3K4me1, H3K4me2 and H3K4me3), histone H3K27 tri-methylation (H3K27me3) and histone H3K9 dimethylation (H3K9me2) [8].

Furthermore, sequence divergent and structurally diverse histone proteins occur as natural 'histone variants' across eukaryotes, and these have evolved to co-occur with distinct

chromatin states (histone modifications), thus determining heritability of epigenetic marks, or assisting in cell differentiation [9]. The concerted effect of all these epigenetic marks can impact chromatin compaction (demonstrated in studies such as [10,11]), and control which regions of the DNA (euchromatin or heterochromatin) are more or less accessible to transcription factors for activating genes (for example, results from [12–14]). Ultimately, these modifications can have both a localised and a global effect on gene transcription, not only during various stages of plant development, but also under stress, thereby facilitating adaptive evolution.

4.2. DNA methylation and its establishment

The regulatory role of DNA methylation was first suggested by [15,16], upon observing DNA methylation marks occurring largely at palindromic repeat sequences in inactive X chromosomes of female mice. Subsequent studies in animal systems identified CpG nucleotide motifs as principal sites carrying methylation [17], eventually leading to the discovery of methylation being semi-conservatively transmitted during replication [18,19].

In plants, ^{5}mC methylation patterns occur in three different contexts determined by the underlying nucleotide sequence - CG, CHG and CHH (where H= A, T, C) . Although these methylation marks are distributed along the chromosome arms, they are found primarily at transposable elements, centromeric repeats or arrays of 5S or 45S rRNA gene repeats that are largely found in centromeric or pericentromeric regions [20]. Heterochromatin regions in *A. thaliana* are enriched with transposable elements, hence are generally characterized by high levels of methylation. Cytosine methylation also accumulates at some differentially-regulated promoters of protein-coding genes, and within the protein-coding regions of highly expressed genes [21]. Recent studies have shown that methylation at specific positions within transcription factor binding motifs in the *A. thaliana* genome may directly affect the interaction between the transcription factor and genomic DNA *in vitro* [22].

C5 methyltransferases catalyse DNA methylation in plants. They are classified into four main families on the basis of their linear protein domain arrangement: DRMs (Domains-rearranged methyltransferases) are involved in the *de novo* methylation (RdDM) pathway, while methyltransferases (METs) and chromomethyltransferases (CMTs) are involved in maintenance of methylation, and DNMT2s (DNA methyltransferase homologue-2) can methylate aspartic acid transfer RNA (tRNA Asp)[23]. Irrespective of the family they belong to, all C5 Methyltransferases in plants carry out two roles - the first being the recognition of a target DNA sequence and the second the transfer of a methyl group

from the cofactor S-adenosyl-l-methionine to the fifth carbon (C5) of the pyrimidine ring of the cytosine nucleotide. While the variable N-terminal domain of the protein carries out the recognition function, the C-terminal domain catalyses the methyl transfer [24,25].

4.2.1. *de novo* methylation pathways

Cytosine methylation marks in the developing plant are established by small RNAs that target DNA sequences that are homologous to them, a phenomenon known as RNA-directed DNA methylation [26–28] **(Figure 1)**. In the canonical RNA-directed DNA methylation (RdDM) pathway [2,29–31], RNA POLYMERASE IV (Pol IV) initiates transcription at the silenced locus and the transcript is copied by RDR2 (RNA-DEPENDENT RNA POLYMERASE 2) to create double-stranded RNA (dsRNA). The dsRNAs are subsequently cleaved into 24-nucleotide long short interfering RNAs (siRNAs) by DICER-LIKE PROTEIN 3 (DCL3). These siRNAs are loaded onto ARGONAUTE (AGO) proteins such as AGO4 and AGO6, and pair with complementary scaffold RNAs. Scaffold RNAs are the nascent transcripts synthesised by RNA POLYMERASE V (Pol V) from the locus which is targeted for silencing. The DOMAINS-REARRANGED METHYLTRANSFERASE 2 (DRM2) protein interacts with AGO4, and catalyses *de novo* DNA methylation in a sequence-independent manner, in *cis* to the scaffold RNA [32] . RNA-DIRECTED DNA METHYLATION 1 (RDM1) may assist AGO4 and DRM2 in this methylation, and may also bind to single stranded methylated DNA [33].

Non-coding RNAs (ncRNAs) transcribed by Pol V remain on the chromatin to serve as scaffold RNAs. NRPE1 (the largest subunit of Pol V) and RDM3 (a Pol V-associated putative transcription elongation factor), along with proteins including those of the INVOLVED IN DE NOVO 2 (IDN2)–IDN2 PARALOGUE (IDP) complex, are also involved facilitating and stabilizing the siRNA- scaffold RNA interaction in the RdDM pathway (as reviewed in [34]).

Pre-existing chromatin modifications also affect the recruitment of Pol IV and Pol V to the RdDM target loci. The SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1), which binds dimethylated histone H3 lysine 9 (H3K9me2) recruits Pol IV [35,36]. The SNF2 DOMAIN-CONTAINING PROTEIN CLASSY 1 (CLSY1), a chromatin-remodelling protein associated with Pol IV, interacts with SHH1 [36,37]. In recent years, the study of other members of the CLASSY family proteins has shown that CLSY1 and CLSY2 are both required for the Pol IV-SHH1 interaction through H3K9me2, while CLSY3 and CLSY4 carry out a similar role but through recognition of DNA methylation marks. Furthermore, differential abundance of the CLSY1-4 proteins in various plant tissues can determine their methylation levels at various loci [38,39].

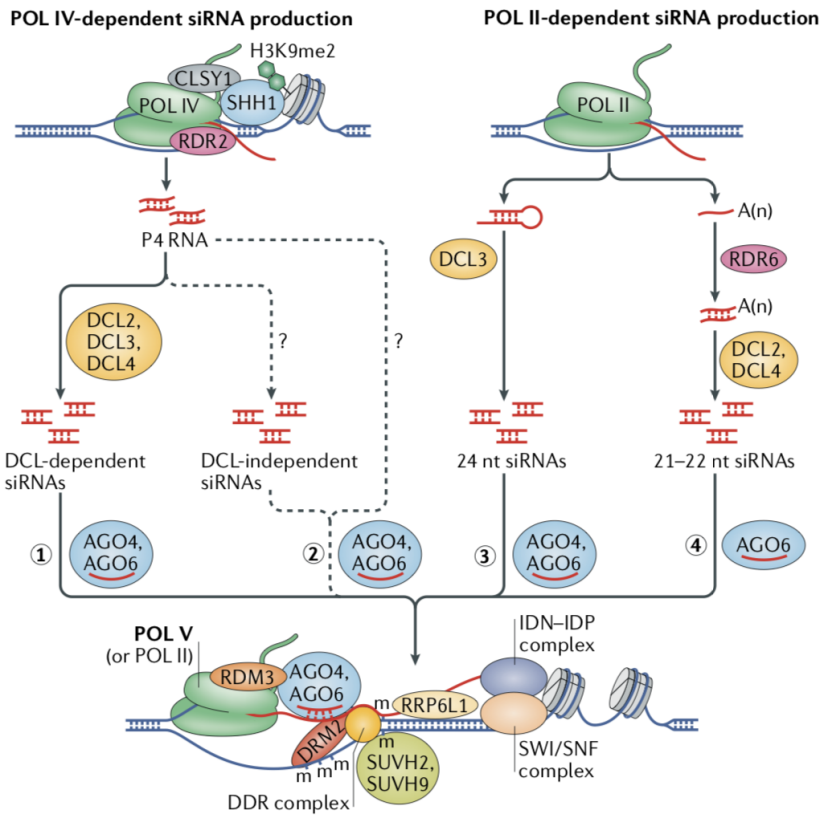


Figure 1: RdDM pathways in *A. thaliana* (Fig. 1 from [34]**). (1)** canonical RNA-directed DNA methylation (RdDM) pathway **(2)** DCL-independent RdDM pathway which may be mediated by DCL-independent siRNAs or directly by Pol IV -generated non-coding RNAs (P4 RNAs) **(3)** POL II- mediated RdDM pathway which produces 24-nucleotide siRNAs and scaffold RNAs **(4)** POL II - RDR6 RdDM pathway which produce precursors of 21-nucleotide or 22-nucleotide siRNAs. AGO4 and AGO6 carry the siRNAs to complementary scaffold RNAs produced by Pol V. A cohort of proteins stabilize this interaction, in addition to retaining the scaffold RNAs in the chromatin.

For the Pol V–chromatin association, the so-called DDR complex plays a crucial role. This complex recruits two proteins that belong to the SU(VAR) 3-9 histone methyltransferase family but lack the corresponding activity. These proteins are SUVH2 and SUVH9, and their SRA (SET and RING finger-associated) domains are necessary for the recognition of methylated cytosines. Thus, it has been proposed that these proteins recognize pre-existing DNA methylation and recruit Pol V (as reviewed in [34]).

Apart from the canonical Pol IV and Pol V RdDM pathway, Pol II-mediated transcription can also generate 24-nucleotide siRNAs in some loci through a non-canonical pathway. In certain cases, Pol II can also recruit Pol IV and Pol V [40]. Pol II interacts with a different cohort of AGO proteins compared to Pol V, and catalyses RdDM along with RDR6

(RNA-DEPENDENT RNA POLYMERASE 6). RDR6 copies Pol II transcripts to make dsRNAs, followed by DCL2/DCL4-mediated generation of 21- or 22-nucleotide siRNAs. The Pol II - RDR6 pathway is preferred over the Pol IV - RDR2 pathway only in certain regions of the genome such as *trans*-acting siRNA genes and transcriptionally active transposable elements. However, 24-nucleotide siRNAs are the most abundant in the *A. thaliana* genome (as reviewed in [34]). Although these siRNAs can be reduced to negligible levels in *dcl* quadruple mutants, DNA methylation still remains in two-third of siRNA target sites [41,42], suggesting the presence of DCL3-independent RdDM pathways.

4.2.2. Maintenance of methylation

DRM2 activity can catalyze methylation in all sequence contexts (CG,CHG,CHH) through the *de novo* methylation pathways in the growing plant - yet, cells in each tissue could have distinct methylation profiles. To ensure that these methylation marks, especially at key loci, are faithfully transmitted to the progeny, plants employ various maintenance methylation pathways.

The heritability of pre-exisiting methylation patterns from parent to progeny is facilitated particularly at CG and CHG sites, which provide a symmetry for cytosine methylation on mother strands to be copied to the complementary daughter strands. Methylation can also be maintained after DNA repair by using the non-damaged strand as a template.

In plant genomes, the CG-context constituting the highest proportion of all cytosine methylated sites [43] is maintained by the METHYLTRANSFERASE 1 (MET1) enzyme. The *A. thaliana* MET1 enzyme is orthologous to DNA (cytosine-5)-methyltransferase 1 (DNMT1) which maintains methylation in mammalian cells, despite varying from DNMT1 in its N-terminal domain structure [44]. *MET1* in *A. thaliana* is one among four genes in a multigene family that may have arisen due to ancestral duplications - *MET1, METIIa, METIIb and METIII*. While all these genes have conserved intron positions, *METIII* encodes a truncated protein and *METIIa* and *METIIb* transcripts are expressed in levels 10,000 fold lower than *METI* in vegetative tissue [45]. Hemi-methylated CG dinucleotides are recognized by MET1 after DNA replication and copied in the cytosines of the daughter strand. The VARIANT IN METHYLATION (VIM) proteins are proposed to play a role in the recruitment of MET1 to DNA, similar to the recruitment of DNMT1 by the E3 ubiquitin-protein ligase UHRF1 [46,47].

CHG methylation in *A. thaliana* is largely maintained by CMT3 (CHROMOMETHYLASE 3) and by CMT2 (CHROMOMETHYLASE 2) in a few regions. The bromo-adjacent homology (BAH) and chromo domains of the CMT3 homolog in maize were shown to bind to the histone methylation mark H3K9me2. Loss of the SU(VAR)3–9 HOMOLOG 4 (SUVH4)/ KRYPTONITE (KYP) proteins which establish H3K9me2, or CMT3, dramatically reduces CHG methylation genome-wide. KYP proteins additionally contain an SRA domain which specifically recognizes CHG methylation. However, mechanisms of a possible crosstalk between KYP and CMT3 in maintaining methylation is yet to be understood (as reviewed by Law and Jacobsen, 2011).

The DRM2 and CMT2 proteins are involved in establishing CHH methylation, and the former carries out its role through the RdDM pathway. The regions targeted by DRM2 are RdDM target loci that include repeat sequences and short transposon elements. CMT2 preferentially targets histone H1-containing heterochromatin, where RdDM is inhibited. DMR2 and CMT2 can also establish methylation in other sequence contexts (as reviewed by Zhang et al. 2018).

4.3. DNA demethylation

Passive DNA demethylation can occur when methyltransferase activity is lost or when there is a lack of methyl donors after DNA replication (as reviewed in [34]). Methylation can also be actively removed, and this is catalysed by a group of enzymes, with the initiator referred to as the DNA demethylase. The 5-mC DNA glycosylases- apurinic/apyrimidinic lyases in plants employ a base-excision repair pathway for demethylation, by direct recognition and removal of methylated cytosines [48–50].

In *Arabidopsis*, REPRESSOR OF SILENCING 1 (ROS1), DEMETER (DME), DEMETER-LIKE PROTEIN 2 (DML2) and DML3 belong to the family of demethylases and can excise 5-mC from all cytosine sequence contexts [49,51–53]. While DME is expressed only in companion cells of the female and male gametes, the other enzymes are active in all vegetative tissues [53,54]. In general, DME targets euchromatic, AT-rich transposable elements, thereby affecting genes in their proximity [54–57]. DME is particularly involved in the demethylation of imprinted genes during embryonic development [58] and demethylation in male gamete companion cells [59–61].

ROS1 mainly functions in preventing the spreading of methylation from transposable elements to nearby genes, and also targets many RdDM loci [62]. Targets of ROS1 are also positively associated with the histone marks H3K18 and H3K27me3, but depleted of

H3K27me and H3K9me2, suggesting a crosstalk between DNA demethylation and chromatin modifications in regulating genes. Both *met1* mutants and mutants defective in RdDM pathway components show reduced expression of *ROS1 [32,63–67]*, indicative of tight coordination between DNA methylation and active DNA demethylation mechanisms. Furthermore, the *ROS1* gene carries a 39 bp sequence in its promoter which is methylated in wild-type plants and functions as a 'methylstat' sequence which recognizes and maintains methylation levels throughout the genome [67,68]. Active DNA demethylation is also critical for fruit ripening in tomato (*S. lycopersicum*), where the DML2 glycosylase catalyses demethylation of thousands of loci involved in fruit ripening [69,70].

4.4. Gene expression regulation by DNA methylation

4.4.1. Promoter methylation

In most promoter regions of plant genes, DNA methylation acts as an inhibitor of gene transcription. Conversely, demethylation of certain promoters results in the activation of gene expression. Studies on embryonic stem cells have found that promoter DNA methylation can inhibit binding of transcriptional activators, promote binding of transcriptional repressors, or differentially attract specific repressive or permissive histone modifications to repress gene transcription [71,72]. In *A. thaliana*, *in vitro* studies using DNA affinity purification sequencing (DAP-seq) have revealed that 72% of 327 transcription factors (TFs) tested were inhibited from binding methylated sites, 24% of TFs showed weak sensitivity to methylation in their binding affinity, while 4% preferentially bound to methylated motifs [22].

Approximately 5% of all genes in *A. thaliana* are methylated in their promoter regions [31,73] and therefore the majority of genes is not dependent on methylation for their transcriptional regulation. However, this can vary from species to species; crop plants with larger genomes for example, are enriched for transposable elements interspersed between genes and therefore a larger fraction of genes may have promoter methylation to silence these elements. Consequently, the loss of methylation in such genomes may result in more severe developmental defects than in *A. thaliana*. [69]

Although the exact mechanisms by which promoter methylation activates gene transcription is still unclear, such methylation patterns are commonly observed over regions containing transposable elements, repeat sequences or the binding sites of transcription machinery. In

some genomic regions, transposon methylation can also spread to flanking regions over time [74], thereby occurring at gene promoters.

4.4.2. Gene-body methylation

Gene body methylation (gbM) refers to the exclusive presence of CG methylation within the transcribed region of genes, and depletion of such methylation near transcriptional start and stop sites [21,73,75]. It has been observed among angiosperm genomes that gbM genes are generally constitutively expressed and are longer than unmethylated genes. Reduced gbM is known to occur as a consequence of reduced levels of the histone variant H3.3, and is associated to an increase in distribution of the linker histone H1 within the gene body, which restricts access to DNA methyltransferases and therefore impacts transcriptional activity [76]. gbM has also been observed in intron regions that contain repeats or transposon sequences, such as in the *IBM1* locus (Wang et al. 2013)

While gbM is completely absent in the angiosperms *Eutrema salsugineum* and *Conringia planisiliqua,* this has been attributed to the absence of CMT3 in their respective genomes. While CMT3 primarily methylates DNA substrates in the CHG context, it is proposed that when the histone methylation mark H3K9me2 is not actively removed from gene bodies (by histone demethylases such as IBM1), CMT3 is recruited thus depositing CHG methylation, and enabling methylation of CG and CHH sites. Consequently, this CMT3-triggered methylation can spread throughout the gene body [77]

The extent to which gbM can determine the transcription of genes is still debated. Some studies suggest that gbM could occur as a byproduct of perturbations to chromatin homeostasis near genes [78], due to several works demonstrating the absence of a functional relationship between gbM and expression changes [43,77,79,80]. However, another study has shown that natural gbM variation can influence drought and heat tolerance and flowering time, demonstrating that gbM can potentially shape phenotypic diversity in plants independent of genetic variation [81].

4.4.3. Methylation at transposable elements

Transposons or “jumping genes” [82], today mostly called transposable elements (TEs), are considered parasitic, selfish sequence elements that are capable of relocating across the genome and often inserting new copies of themselves. In plants and most other eukaryotes,

TEs are silenced epigenetically by the host to prevent the insertional disruption of protein-coding genes and other functional sequences by these elements. Upon epigenetic repression, inactivated TEs may retain their coding potential for mobilization, but cannot produce the required proteins [83]. Based on whether the mobilization of TEs requires an RNA or a DNA intermediate, they are classified as retrotransposons (Class I) and DNA transposons (Class II) respectively [83].

TEs can cause gene-disruption not only by insertion of their copies, but also through imprecise excision. However, they may also provide certain benefits to the genome, such as the creation of new regulatory elements near promoters, or the building of regulatory networks that have a concerted effect on the expression of several genes. Furthermore, the presence of TEs or TE remnants may also serve as alternative transcription start sites or transcript isoforms. Lastly, the methylation marks carried by these TEs can spread to flanking regions, thereby generating new epigenetic regulatory mechanisms [84,85].

The pericentromeric heterochromatin and other repeat regions across the *A. thaliana* genome show high levels of DNA methylation in all cytosine contexts [73,86]. Incidentally, these regions also harbour a large number of TEs, where short transposons are primarily methylated in the asymmetric CHH context by the components of the RdDM pathway, and long transposon methylation is primarily catalyzed by CMT2 and also depends on the remodeler DDM1 [87,88]. *Arabidopsis thaliana* is distinct from other plant species in its mobilome landscape (collection of TE families with recent mobilization activities) due its relatively compact genome size (125 Mb) and small number of TE families [89]. The maize genome for example, is larger (2.3 Gb) and is densely populated by TEs. Islands of methylated cytosines in the CHH context that often occur near euchromatic genes in the maize genome can lose methylation, leading to CG and CHG hypomethylation and transcriptional activation in proximal transposable elements, thereby determining the boundaries between euchromatin and heterochromatin [90]. In tomato genomes, chromatin is clearly separated into large repeat-rich pericentric domains, gene-rich domains and an intermediate domain, with Class I TEs predominant in repeat-rich regions and Class II TEs in repeat-poor regions [91,92].

When methylation is absent over TEs, they are de-repressed and thereby activated for mobilization in the genome, transposing themselves and their copies to disrupt protein-coding genes. High TE activity is often observed in mutant lines which have reduced methylation, especially in the predominant CG context, and can have dramatic phenotypic consequences. For example, *A. thaliana* plants lacking the SWI/SNF chromatin remodeler

DDM1 (*ddm1)*, exhibit genome-wide hypomethylation and consequently activate ATGP3 (a *GYSPY* family retrotransposon) upon inbreeding [93]. Such hypomethylation mutant lines can be crossed to wild-type plants, and their progeny that inherit the wild-type alleles are inbred over several generations to generate epigenetic recombinant inbred lines (epiRILs). *ddm1*-derived epiRILs in *A. thaliana* can widely vary between individuals in the locations of their differentially methylated regions, resulting in spontaneous TE activation following burst-like patterns [94]. Similarly, when mutants of the MET1 methyltransferase (*met1*) are inbred for more than eight generations, the activation and re-insertion of the *Evadé* (*EVD*) retrotransposon family into genes such as *BRI1*, *LFY* and *VAR2* can result in developmental defects [95].

TE mobilization can be initiated both by artificial means (using methylation mutant lines) as well as natural environmental cues. One of the best examples of environmental stress-induced TE transposition is demonstrated in the *A. thaliana nrpd1* mutants, defective in RdDM methylation. Seedlings of *nrpd1* mutants exposed to heat stress release the silencing of multiple ONSEN retrotransposon family elements that can make copies and re-insert themselves, preferentially into genic regions to have a local effect on gene regulation [96,97]. Overexpression of a *ROS1* homolog (*DNG701*) in rice also leads to hypomethylation (due to excess of active demethylation), resulting in the expression and mobilization of the Tos17 retrotransposon [98].

Although TE mobilization can have a detrimental impact in the plant genome by inserting near genes and promoter regions, there is increasing evidence to suggest that this phenomenon can also drive evolutionary adaptation [99,100].

4.5. Methylation patterns during developmental transitions

While DNA methylation patterns tightly regulate gene expression in several tissues and cell types during a plant's life cycle, one of the most crucial roles occurs during embryogenesis. In flowering plants, sexual reproduction processes begin with the differentiation of the megaspore mother cell (MMC) and the microspore mother cells (MiMC) in the female and male floral organs respectively. Both MMC and MiMC give rise to the haploid megaspore and microspore through meiosis, during which they experience large scale chromatin changes. Recent studies on resolving the epigenetic state of these progenitor cells has shown that methylation levels in the CHH context fluctuate for the MMC, while the same marks are very low for MiMC [101] .

Subsequently, male microspores undergo gametogenesis to form male gametophytes (pollen grains), which further differentiate into the sperm cell and the vegetative cell, each with a distinct methylome **(Figure 2)**. While the sperm cell forms pollen, the male gamete companion cell (vegetative cell) eventually forms the pollen tube. The vegetative cell is hypomethylated in the CG and CHG context at many loci, concurrent with the downregulation of DDM1 activity [57,60]. As a consequence, heterochromatic TEs are transcribed and activated in vegetative cells [102]. Not only are these TEs suspected to be mobile, but there is accumulating evidence supporting the notion that RdDM-associated siRNAs are able to target TE loci in the sperm cell by accumulating epigenetically-activated or 'easiRNAs' derived from the vegetative cell [59–61].

Epigenetic changes during development also occur during female gametogenesis **(Figure 2)**. The female haploid functional megaspore undergoes meiosis three times to form an eight-celled female gametophyte surrounded by the maternal integument. On reaching maturity the gametophyte contains two gametes, a diploid central cell and a haploid egg cell that fuse with two sperm cells to develop into the endosperm and the embryo respectively. Within the central cell, the demethylase DME (DEMETER GLYCOSYLASE) is highly active, selectively removing methylation marks from certain loci, resulting in several downstream consequences. A majority of DME targets belong to the group of genes that contain TE-derived *cis* elements in proximity to their promoters. This hypomethylation also triggers the production of small RNAs, which diffuse from the central cell to the egg cell to eventually silence TEs in the developing embryo [55,57,103] .

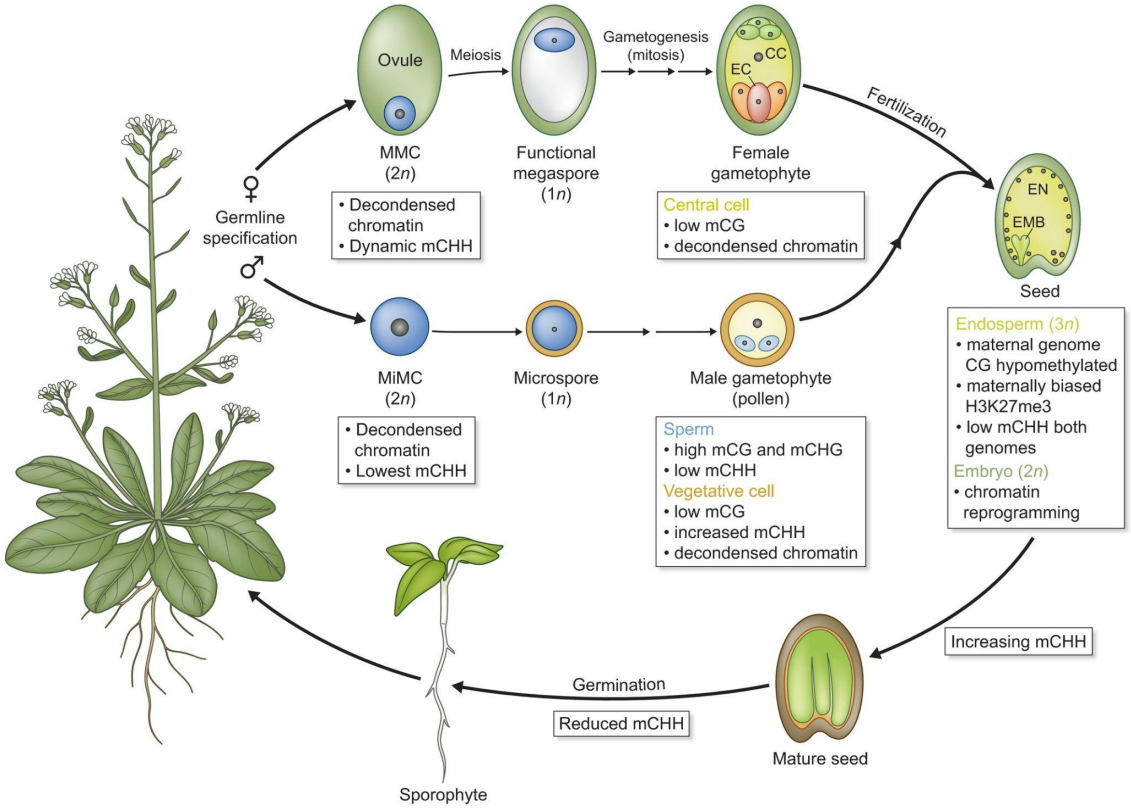


Figure 2 : Epigenetic dynamics during flowering plant reproduction. Fig. 1 from [101]. Haploid megaspore and microspore cells are shown in dark blue arising from progenitor cells MMC (megaspore mother cell) and MiMC (microspore mother cell) respectively. The central cell (CC), egg cell (EC) , embryo (EMB) and endosperm (EN) are annotated. mCG, mCHG and mCHH refer to CG, CHG and CHH methylation respectively.

Recent work has shown that the CLASSY family of DNA-binding proteins (CLSY1, 2, 3 and 4), can determine methylome state in flower buds, ovules, young leaves and rosette tissue [39]. Interestingly, CLSY3 along with CLSY4 is primarily required for ovule methylation, but CLSY3 also aids in the production of nurse-cell induced small RNAs (niRNA) in tapetal cells that surround the meiocyte, and are eventually transported to the male meiocyte for TE-silencing [104].

Double fertilization of both the egg cell and the central cell with a sperm cell, gives rise to the embryo (diploid) and the endosperm (triploid) of the seed respectively. Post fertilization, the triploid endosperm carries two copies of the DME-demethylated maternal genome, thereby activating the expression of certain genes sensitive to a change in methylation ('imprinted epialleles') [105–108] **(Figure 3A)**. The paternal allele counterparts of these genes are often found to be marked with hypermethylation and H3K4me3 histone methylation, further

supporting their repressed state. Paternal imprinting can also occur by the establishment of methylation marks in the maternal allele of some genes, thereby inducing the expression of paternal alleles. Notable examples of maternally expressed genes (MEGs) are *FWA, MEA, FIS2 [49,52,58,109]*, while *HDG3* is a paternally expressed gene (PEG) [55]. The deposition of repressive histone marks such as H3K27me3 by POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) can suppress the paternal alleles of MEGs or maternal alleles of PEGs [110–113].

MET1 is suppressed in the central cell during female gametogenesis, which could additionally contribute to hypomethylation of maternal alleles in MEGs [114,115]. It has been found that imprinting is subject to a parent-of-origin effect, whereby the difference in methylation levels of both parental alleles determines the direction of imprinting, i.e. whether it will be a MEG or a PEG [116].

4.6. Plant epialleles and their formation

Plant phenotype and underlying gene expression profiles can be modulated through dynamic changes in chromatin properties and DNA methylation. These changes can arise as epialleles, epigenetic variations that are fixed at certain loci and inherited stably across generations [117]. In plants, epialleles are often associated with regions of differential methylation (called Differentially Methylated Regions or DMRs) [118]. Some DMRs are known to affect the expression of overlapping, proximal, or distantly located genes, making them an important source of variation and phenotypic plasticity, especially in the absence of genetic variation [34]. However, most DMRs have not been assigned any biological function despite their proximity to genes.

Imprinted epialleles, as previously discussed, can be generated during reproductive development in plants, where parental allele-specific methylation marks are removed, initiating gene transcription **(Figure 3A)**. Heritable changes of methylation at single cytosine positions can also arise spontaneously in plants **(Figure 3B)**. A methylome analysis for *A. thaliana* mutation accumulation lines grown for 30 generations revealed that epimutations occur at a much higher rate compared to genetic mutations: ~10^{-4} bp^{-1} compared to ~10^{-9} bp^{-1} per sexual generation [119–122]. Such spontaneous methylation changes can lead to the formation of stable and heritable epialleles, ("pure epialleles"), although occurring at a very low frequency [123]. Many naturally occurring epialleles with diverse biological roles

have been identified in *A. thaliana*, tomato (*Solanum lycopersicum*), melon (*Cucumis melo*), *Linaria vulgaris*, rice (*Oryza sativa*) and maize (*Zea mays*) [124].

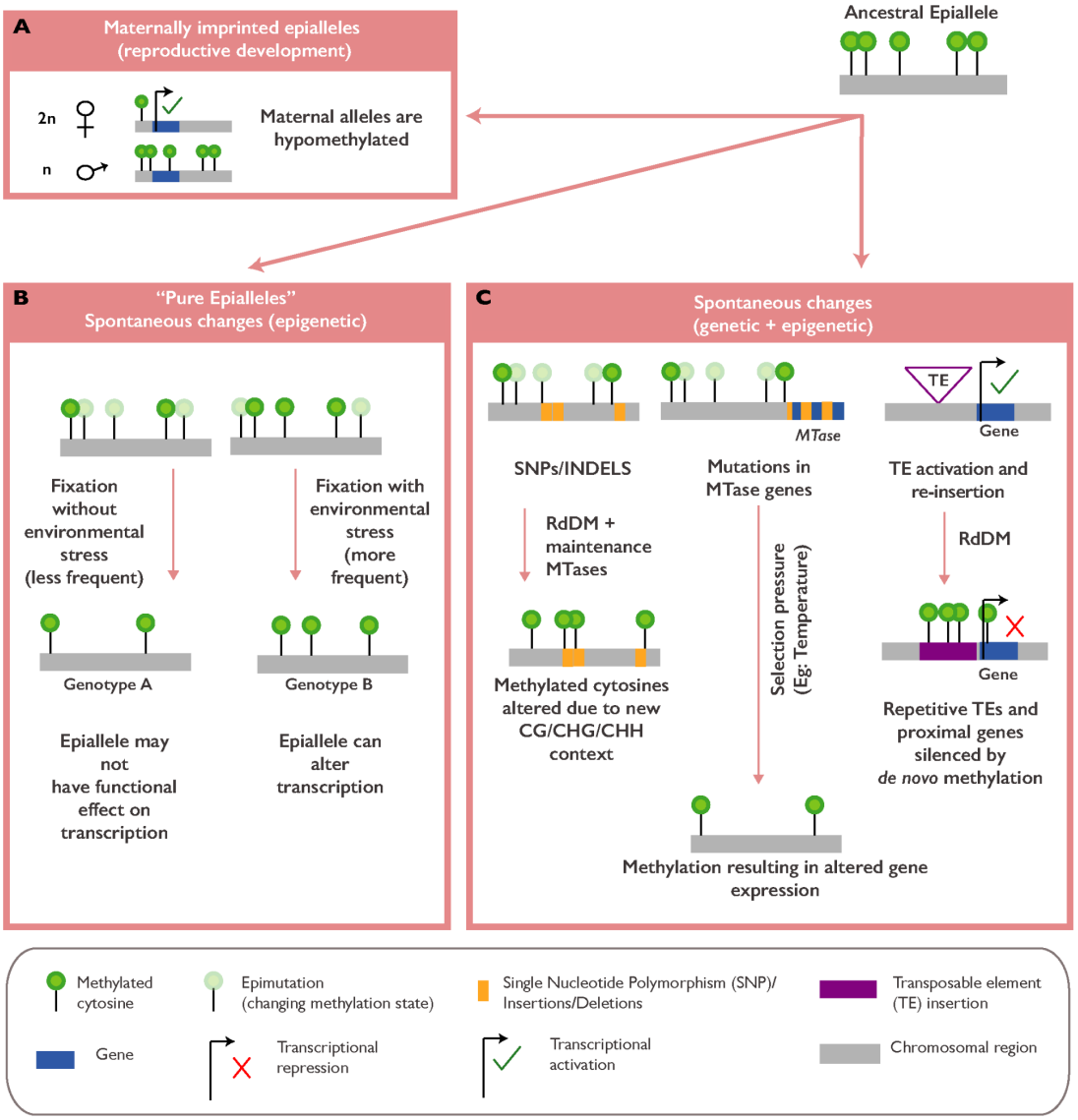


Figure 3 : The origins of natural epialleles [124]. (**A**) Imprinted epialleles can be generated during the reproductive development in wild-type (WT) plants where parental allele-specific methylation marks are removed, initiating gene transcription. (**B**) Spontaneous methylation changes can occur independent of the underlying genetic variation ("pure epialleles") and may also be subjected to evolutionary fixation. (**C**) Mutations in protein-coding genes, including genes involved in the maintenance of DNA methylation (methyltransferases or "MTases") and the mobilization of transposable elements (TEs), can also recruit methylation marks, which result in the formation of epialleles, as they can influence the proximal gene transcription. Together, these newly acquired genetic and epigenetic changes may provide an adaptive advantage for the plant.

Single nucleotide polymorphisms, insertion and deletion mutations in protein-coding genes, may affect the position of cytosines, thereby altering cytosine methylation patterns across the genome. Natural *A. thaliana* accessions, which are characterized by ample genetic and phenotypic variation, have highly variable methylation in 78% of the methylated sites, with 22,060 DMRs identified across 1,107 accessions, showing the abundance of naturally occurring epialleles at the intraspecies level [79].

Epialleles can arise from mutations in methyltransferase genes ("MTases")**(Figure 3C)**, possibly altering their activity. Variation in non-CG methylation in natural *A.thaliana* accessions is strongly associated with genetic variation of genes involved in the RdDM pathway, such as *CMT 2, AGO1 and AGO2*, and *NUCLEAR RNA POLYMERASE D1B (NRPD1B)* [79,125–127], while variation at CG gene body methylation is associated with variation in *MET1[79]*.

The presence or absence of TEs and repeat-rich sequences can also contribute to epiallele formation over evolutionary time, by affecting the transcription of genes in their proximity **(Figure 3C)**. TE variants are strongly associated with methylation level of overlapping or proximal regions, suggesting that TE insertions may exert a *cis*-regulatory effect on methylation [128,129]. A recent study revealed that TE mobilization rates across different *A. thaliana* accessions are associated with *NUCLEAR RNA POLYMERASE E1 (NRPE1)* genetic variation. *NRPE1* encodes the largest subunit of RNA Pol V, a main component of RdDM and non-CG methylation. This suggests that non-CG methylation variation across *Arabidopsis* accessions is strongly influenced by *NRPE1* genetic variation and TE transposition [99], and therefore can determine epiallele formation. Epialleles can also be induced upon biotic and abiotic stresses, upon propagation of epigenetic recombinant inbred lines (epiRILs), through clonal propagation and by transgenic approaches for targeted epi-mutagenesis [124].

4.7. MET1-dependent epigenetic control in *A. thaliana*

Among the first studies to understand the function of the MET1 maintenance methyltransferase in plants were those by Finnegan et. al [130] and Ronemus et al. [131], who generated knockdowns of *MET1* in *A. thaliana* accessions C24 and Col-0. In both studies, *met1* mutant lines displayed multiple phenotypic abnormalities **(Figure 4A)** attributed to varying levels of reduced methylation (10% to 100% of wild-type parental methylation levels), which were also meiotically inherited. Floral homeotic genes *SUPERMAN* and *AGAMOUS* were ectopically hypermethylated (largely in the asymmetric context) in *met1* partial loss of function mutants and in *MET1* antisense lines [130,132].

Subsequently, *met1-1* mutants generated by Kankel et al. [133], which exhibited 70% reduction in CG methylation levels, were found to have phenotypes such as thick inflorescence meristems, delayed juvenile to adult leaf transition, and delayed flowering. This was among the first studies to link MET1-induced methylation changes to gene expression changes, andit led to the discovery of the *FWA* epiallele which causes a late-flowering phenotype due to its ectopic expression in *met1* mutants.

It was known at the time that another demethylation mutant in *A. thaliana*, *ddm1*, harbored highly reduced methylation levels at centromeric repeats and transposon-associated sequences [134,135]. Therefore, when comparing *met1* mutants to *ddm1*, Kankel et al. (2003) found that the ability to restore methylation was greater in *met1* lines, where methylation was only knocked out in the CG context, than in *ddm1* lines, where all three contexts of methylation were wiped out.

The most severe knockout of *MET1* so far has been observed in *met1-3* mutant lines, where the catalytic domain of the enzyme houses a 7.1 kb T-DNA insertion, thereby disrupting gene function and resulting in a null allele [136]. Similar to what had been observed in *ddm1* lines, 180-bp centromeric repeat sequences in these *met1-3* mutants were completely demethylated. Certain pericentromeric repeat sequences were also shown to be reactivated in both heterozygous and homozygous individuals in this mutant lineage. This study also showed that MET1 could catalyze passive demethylation already at the gametophytic stage of cells, thereby impacting the epigenetic landscape of the progeny based on the parental line carrying the loss of function *MET1* allele.

The absence of CG methylation in the *met1-3* strain is known to co-occur with the depletion of H3K9me2 histone methylation in chromocenters of first generation nuclei [137]**,** although the signals reappear in subsequent generations [63]. Furthermore, such progressive alterations in nuclear architecture were seen to be correlated with increased frequency of phenotypic abnormalities with inbreeding (dwarfism, late-flowering, altered flower morphology and reduced fertility) **(Figure 4B, 4C)**. After the first generation of CG hypomethylation, *met1-3* plants also appeared to gain ectopic CHH methylation at heterochromatic loci in successive generations, suggesting that MET1-induced CG methylation may have an indirect effect on controlling levels of non-CG methylation [63].

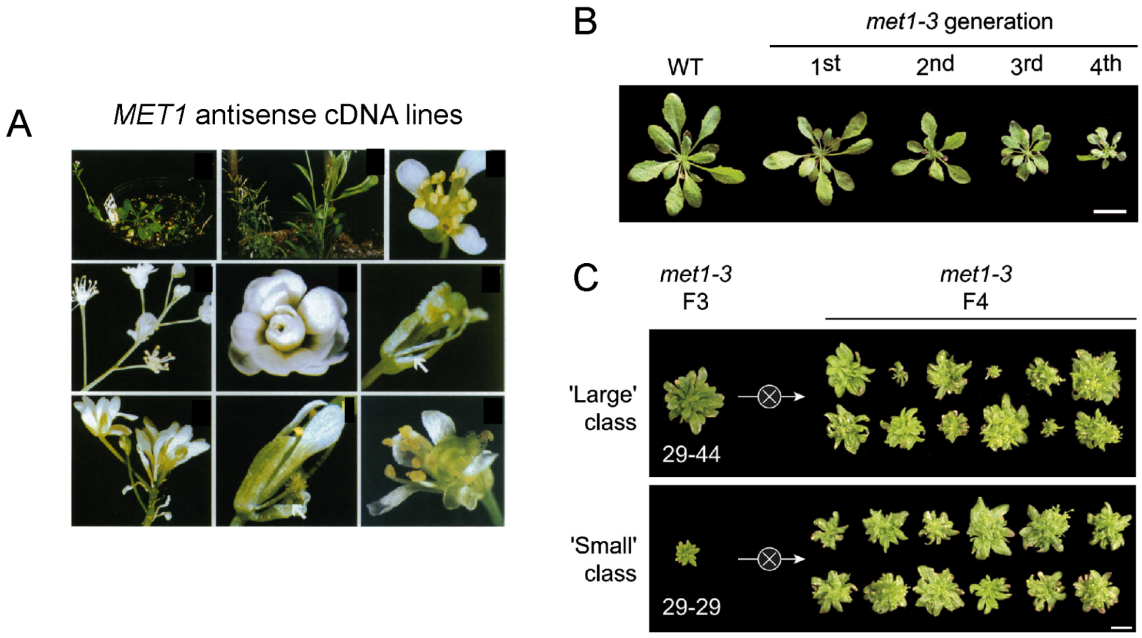


Figure 4: Phenotypic consequences of *MET1* inactivation. (**A**) Plant and flower phenotypes in *A. thaliana* lines carrying an antisense cDNA construct for *MET1* knockdown (Fig. 3 from [130]). (**B**) Increased phenotypic abnormalities with inbreeding of *met1-3* mutants (Fig. 2A from [63]). (**C**) Phenotypes of fourth generation homozygous *met1-3* mutants derived from two phenotypically different siblings (Fig. 2B from [63]). Scale bars for (**B**) and (**C**) represent 1.5 cm.

Subsequently, genome-wide analyses in *met1-3* mutants corroborated the previous results by examining genes that were marked with increased H3K9me2 [138]. Interestingly, several genes targeted by Polycomb Group proteins (which are chromatin modifiers) in *met1-3* showed a significant loss of H3K27me3, and replacement of this mark by H3K9me2. It was concluded that both these histone methylation marks were redistributed in *met1-3* mutants in a mutually-exclusive manner, with the ability to replace one another.

Mutants in which MET1 is overexpressed also carry novel epialleles, and show various phenotypic abnormalities such as reduction in primary root length, increase in secondary roots and a delay in bolting. Interestingly, a subset of genes with altered expression in *MET1* overexpression lines also overlap with differentially expressed genes in *met1* mutants, while other genes overlap with differentially expressed genes in *ddm1* and *hda6* mutants [139].

4.8. Chromatin accessibility landscape in plants

4.8.1. Interplay with DNA methylation and other epigenetic features

Within the cell, the organization of chromatin renders various genomic regions differentially accessible to regulatory proteins and transcription factors, often influenced by local epigenetic modifications such as histone modifications, the presence of histone variants, and the alteration of nucleosomal positions by chromatin-remodelling complexes [140].

One of the techniques used for chromatin structure profiling is MNase-seq, where enzymatic digestion of DNA by a micrococcal nuclease is combined with next generation sequencing [141–143]. The endonuclease enzyme cleaves and digests nucleosome-free DNA in the chromatin, and the remaining intact nucleosomal regions are amplified and sequenced, thereby generating a map of nucleosomal positions in the genome. MNase-seq combined with DNA methylation profiles in *A. thaliana* revealed that nucleosomal DNA exhibited higher methylation levels than flanking DNA sequences [144]. While being relatively evenly distributed across euchromatic chromosome arms, nucleosome content was significantly enriched in pericentromeric heterochromatin, which is also rich in transposons and marked by dense DNA methylation [86]. From the same nucleosome sequencing dataset, WW dinucleotides (where W=A or T) and SS dinucleotides (where S=G or C) were found to periodically occur at 10 bp intervals, with the SS dinucleotides being 5 bp out of phase with WW. Interestingly, WW and SS dinucleotides are favoured at sites where the minor groove of DNA faces towards and away from the histone core, respectively [144], suggesting that these nucleotides determine how DNA can optimally bend [145,146]. Furthermore, DNA methylation patterns in the CG, CHG and CHH context also show a 10 base pair periodicity (which is also observed in human genomes), and is specifically enriched in nucleosome-bound DNA, suggesting that nucleosomal DNA may act as a substrate for DNA methyltransferases *in vivo* [144].

By integrating mononucleosome-sequencing data in *Arabidopsis* and rice with transcriptome data, several authors have been able to conclude that promoters of highly expressed genes are associated with a larger number of nucleosome-depleted regions than lowly expressed genes [147–149].

DNase-seq is also used for assaying chromatin architecture, and is specifically used to identify genomic regions of open chromatin [150]. This technique involves the digestion of

intact nuclei with the non-specific endonuclease DNase I, that preferentially cleaves open chromatin sites, followed by whole-genome sequencing of the digested ends to identify DNase I hypersensitivity sites. In plants, this assay has enabled the identification of transcription-factor binding footprints and *cis* regulatory elements [151–153]. Since the nucleosome core may physically hinder access of DNaseI, the identification of DNaseI hypersensitive sites using DNase-seq has been successful for locating open chromatin regions in different tissues of rice and *A. thaliana* [151,153,154].

4.8.2. Identification of *cis* regulatory elements in accessible chromatin

ATAC-seq (Assay for Transposase-Accessible Chromatin) [155,156] is an efficient method for assaying chromatin accessibility, being less labor intensive than MNase-seq and DNase-seq. In ATAC-seq, nuclei are treated with an engineered transposase which can simultaneously cleave DNA in open chromatin and insert sequencing adapters so that the cleaved fragments can be amplified with PCR and sequenced to generate a whole-genome library of accessible chromatin regions.

ATAC-seq analyses of DNA from root tip nuclei of *Arabidopsis*, the legume *Medicago truncatula*, tomato (*Solanum lycopersicum*) and rice (*Oryza sativa*) show that the genomic distribution of transposase hypersensitive sites is highly similar across all four species [157]. Although all four species vary in their genome size, >75% of their accessible sites consistently occur outside transcribed regions and largely within 3kb upstream of transcription start sites. The tomato genome in particular is known to have a high density of transposase hypersensitive sites in a larger proportion of genes compared to other species, possibly attributed to its density of gene-proximal long-terminal repeat retrotransposons, that may have a gene-regulatory role [158]. Accessible regions may influence gene expression for different cell types, as seen in tomato [159], and sperm cells [160] and root nuclei of *A. thaliana* [161]

In maize, approximately 32.5% of all accessible chromatin regions (ACR) are distal to genes (greater than 2 kb). These distal ACRs are GC rich, show low sequence diversity and contain multiple TF binding sites. Many of these distal ACRs can form chromatin loops with genes in *cis*, and ACRs that are flanked by histone acetylation marks are preferentially looped three dimensionally to highly expressed genes. In addition to these characteristics, ACR sequence elements can function as transcriptional enhancers, and overlap with previously hypothesized fine-mapped *cis* regulatory elements, making such ACRs the primary genomic regions for housing *cis* regulatory elements in the maize B73 genome [10].

A deeper analysis of the ACRs in this study revealed that nearly 20% of all distal ACRs carry 348 non-redundant instances of TE insertions [162]. These ACR sequences harbor TE insertions that show considerable variation across maize genotypes W22, Mo17 and PH207, and the presence/absence of these TEs can determine their corresponding chromatin accessibility state. When comparing the B73 and PH207 genomes, it was found that only 37% of TE insertions within ACRs exhibited a gain of DNA methylation in the haplotype with the TE insertion, and some other TEs flanking ACRs exhibited methylation gains preferentially on one side of the TE. Overall, only a subset of the TE insertions that lie within or in proximity to ACRs carry DNA methylation gains may be linked to differential chromatin accessibility.

ACRs have enabled the identification of tissue-specific developmental ontogenies in maize using single-cell ATAC-seq experiments. Marand et al. [13] have described 92 distinct chromatin accessibility states across 52 known cell types in *Zea mays*. Among the gene-distal ACRs identified, 30% overlap with LTR transposons and were depleted of DNA methylation. Together with data on their TF binding dynamics and co-accessible chromatin interactions at single-cell resolution, these accessible regions have posed as sites of *cis* regulatory elements, functioning as a robust substitute of gene expression analysis at the single-cell level.

Chromatin architecture and 3D genome organization can be particularly crucial for polyploid genome stability, tightly regulating transcription in the duplicated sub-genomes. This has been demonstrated in hexaploid wheat (an allopolyploid evolved through two rounds of genome doubling), where a comprehensive epigenetic map of three subgenomes shows that intrachromosomal interactions (in topologically associated domains), chromatin accessibility and histone methylation marks determine gene expression profiles [163]. These results demonstrate how 'transcription factories' involving multiple genes can be regulated by the presence of condensed chromatin in repetitive regions, which bring together distal genes in physical proximity.

4.8.3. Stress response facilitated by accessible chromatin

While accessible chromatin may regulate gene expression in wild plants, it may especially be instrumental for facilitating response to stresses. In response to submergence (flooding) stress, rice and *Medicago truncatula* show a preference for the opening of chromatin, in comparison to three tomato species [12]. Genes that were highly up-regulated under submergence were also marked with increased accessibility upstream and downstream of

their transcripts. Upregulated genes were also enriched for four significant TF binding motifs in their promoters and accessible regions, two of the motifs functioning in hypoxia response [12].

Accessible chromatin regions affecting gene transcription have also been identified under chilling stress in tea plants (*Camellia sinensis*) [164], cold stress in *Brachypodium distachyon*, foxtail millet (*Setaria italica*) and sorghum (*Sorghum bicolor*) [165], and in rice cultivars exposed to high temperatures, water deficit and agricultural envrionments [166].

4.9. The model plant *Arabidopsis thaliana* and its natural genetic and epigenetic diversity

4.9.1. Genetic variation

Thale cress, *Arabidopsis thaliana*, belongs to the eudicotyledonous group of angiosperm vascular plants, in the *Brassicaceae* family. Having a short generation time and the ability to self-fertilize in addition to exhibiting large phenotypic diversity in wild strains, it has proven to be an apt model organism for understanding plant physiology, development and adaptation to natural habitats. A reference genome sequence for *A. thaliana* was the first nuclear genome of a flowering plant to be published [167]. Although the evolutionary origins of *A. thaliana* may be from Africa [168], many of its natural populations are currently found in Eurasia [169] and North America [170,171]. Locally, *A. thaliana* populations are often found in dry, rocky and shallow soil or in nutrient-poor meadow and forest habitats [172].

The collection and characterization of these intra-specific variants in *A. thaliana* prompted several efforts to identify genome-wide polymorphisms. Over the years, genomic and transcriptomic analyses of large collections of *A. thaliana* accessions have been carried out [173–176], with the largest being the worldwide collection of 1,135 accessions [169]. The 1001 Genomes collection has provided an enormous resource of SNPs and polymorphisms from short-read sequencing data, for linking genetic variation with phenotypic variation. These genomes show an average of one genetic variant for every 10 bp of the single-copy genome, thereby enabling genome-wide association studies (GWAS) for several phenotypic traits (for example, flowering, response to bacterial elicitors, etc. [177]).

In examining the population structure of these accessions, it was found that the 1,135 accessions could be divided into a relict group (26 accessions with extreme pair-wise divergences) and a 'non-relics' group (the remaining 1,109 accessions). Based on a model

of isolation-by-distance, where genetic distance between accessions is largely explained by geographical distance, relicts are relatively stationary compared to non-relicts that have rapidly expanded and found often in both agricultural and urban areas [169].

Ancestral *A. thaliana* may have diverged from its outcrossing parent population, and in the process, a subpopulation may have migrated into Africa by 1.2 - 0.8 Mya (million years ago). These populations further split later, at 120 - 90 kya (thousands years ago), into the major clades found in Africa today - Morocco, Levant and sub-Saharan Africa. Subsequently, at around 40 kya, some of these populations migrated back into Eurasia [168]. Although not native to the continent, *A. thaliana* populations are also found in North America and have been recently shown to exhibit widespread genetic admixture with Eurasian and African populations, suggesting multiple independent introductions associated with human colonization patterns [170].

4.9.2. Methylome variation

Increasing evidence suggests that the immobility of plants has especially favoured the accumulation of epigenetic marks such as DNA methylation to survive and adapt under changing environments. Apart from being induced by strong environmental stresses, DNA methylation may also change spontaneously, or as a result of genetic variation in transposable element architecture, chromosomal rearrangements, and mutations in methyltransferase genes.

Analysis of the methylation landscape in 1,028 *A. thaliana* natural accessions revealed the presence of 22,060 differentially methylated regions (DMRs) that represented 45 Mb (38%) of the reference genome [79]. It was found that gbM genes (genes which have CG methylation within gene-bodies, with a depletion of methylation at transcription start sites and transcription termination sites) were highly variable across accessions. The methylation state (hyper- or hypomethylation) at gbM sites roughly separates the accessions also based on their geographic origins. For some accessions such as Cvi-0 and UKID116, gbM appears to be dispensable for genome stability, as their hypomethylated state at gbM sites does not affect their overall transcriptomes in comparison with other accessions such as Col-0 (which is moderately methylated) and Bak-5 (which is hypermethylated).

Unlike genes marked with gbM, TE-like methylated (teM) genes that exhibit CHG or CHH and/or CG methylation, show less variation in their methylation state across accessions (i.e. were less poly-epiallelic), suggesting that teM epialleles may have been more recently

established during evolution than gbM epialleles. The relict accessions were found to have a large number of teM singletons, where methylation states differed only in one accession compared to others. Poly-epiallelic genes also harbored a larger number of non-synonymous genetic mutations compared to genes that did not vary in their methylation state. Interestingly, TEs were enriched within 500bp in proximity to poly-epiallelic genes that were teM in Col-0, suggesting that the methylation variation may have arisen from the spreading of RdDM targeting the TEs [79].

CHH methylation at TEs was found to be positively correlated with temperature, based on methylation data from accessions grown at two different temperatures [79,127]. GWAS on TE methylation across accessions identified several peaks (genomic regions where structural genetic variation is causal for methylation variation), overlapping/proximal to candidate genes such as *CMT2*, *AGO9* (involved in siRNA silencing), *AGO1* and *NRPD1B* . Additionally, GWAS with relaxed parameters could identify more candidate loci such as *MBD3 (METHYL-CPG-BINDING DOMAIN 3)* for CMT2-dependent CHH methylation and a strong association at *MET1* for gbM signatures [79].

4.9.3. Transcriptome, transcription-factor binding and chromatin accessibility variation

Transcriptomes from 727 natural accessions of *A. thaliana* have been analysed, and have enabled the identification of differentially expressed genes among various admixture groups separated by their genetic distances [79]. Genes that vary highly between these groups of accessions were enriched for biotic and temperature responses, as expected from their growth in diverse natural habitats. Furthermore, groups of differentially methylated genomic bins also harbor binding sites for distinct TF families including the heat shock response factor HSFA6B, the meristem formation TF CUC2, and MYB-related transcription factors. Integrating methylome and transcriptome data along with *in vitro* methylation-binding data obtained through DAP-seq [22], identified a general pattern in 352 TFs where the strength of TF inhibition by methylation was tightly associated with the depletion of its binding sites in differentially methylated loci across accessions. Recent studies have mined the above transcriptome dataset more thoroughly, identifying large accession-specific variation in gene expression associated with methylation marks in gene bodies, further emphasizing the regulatory links between DNA methylation and gene expression [78,81].

Conserved non-regulatory sequences (CNSs) which are highly identical at the inter-species level [178] are also known to contain transcription factor binding sites [178,179] and

colocalize with accessible chromatin in plant genomes [10,180]. CNSs identified in 30 accessions of *A. thaliana* show considerable presence-absence variation and positional genetic variation, harboring motifs of stress responsive transcription factor families [181]. Interestingly, in 18 such accessions, approximately 14% of CNSs overlap with ATAC-seq peaks of chromatin accessibility, with positional variants being strongly associated with accessible chromatin regions. Additionally, genes proximal to CNSs overlapping accessible chromatin are enriched for gene ontology terms associated with response to abiotic stresses and phytohormones. Furthermore, the number of CNSs associated with a gene in a given accession appear to influence gene expression, with an increasing number of CNSs correlated with gene upregulation and vice versa [181]. ATAC-seq has also been carried out for *A. thaliana* accessions with high quality genome assemblies, identifying >50,000 accessible regions, of which 15% are differentially accessible between five accessions (Bay-0, Bur-0, Col-0, Est-1 and Tsu-1) and largely occur independently of genetic and epigenetic variation [182]. Most of these differentially accessible regions occur in proximity to differentially expressed genes, which maintain their gene expression and accessibility state even in the presence of *cis* structural variants.

4.9.4. Genetic basis of epigenetic variation

Although methylation diversity may arise from spontaneous methylation changes at several loci for evolutionary adaptation [183,184], GWAS results have shown that such epigenetic variation could also arise from genetic variation at key *trans* -acting loci. In *A. thaliana* accessions, these include structural variants at *CMT2* that are causal for genome-wide CHH methylation variation and temperature adaptation [125–127]), while variants of *CMT2*, *NRPE1*, *AGO1* and *AGO9* are causal for CHH methylation variation at TEs [79,126,127].

However, genetic variants may also include large insertions and deletions, especially of transposable elements. In fact, such variation in mobilomes (transposable element landscape) [128] is tightly linked to methylome variation. Indeed, analysis of differential transposition activity in seven Class I and Class II TE families across 211 *A. thaliana* accessions was found to arise from genetic variation at the *MET2a* locus that encodes the *MET2* methyltransferase. However, improved pipelines to detect TE insertion polymorphisms in 1047 accessions have enabled the identification of *NRPE1* as the genetic basis of TE mobilization [99]. Additionally, modelling adaptive trajectories of transposon mobilization in these accessions have helped identify bio-climatic variables such as seasonality of precipitation and diurnal temperature range as determinants of differential transposition activity. Presence-absence variants in the genome (including Non-TE genes) have also

been identified in an extended collection of *A. thaliana* genomes, including African accessions, demonstrating that such variants can be crucial for determining phenotypes such as vernalization response for flowering, drought and heat tolerance [185].

PROJECT AIMS

DNA methylation as an epigenetic mark has been well-characterized as being causal for gene expression changes in both plants and animals. However, it remains to be clearly understood why methylation state can determine gene expression only at a subset of loci in the genome, to what extent its effect is direct, and whether it involves an interplay with other epigenetic factors.

For my PhD thesis, I chose to examine the epigenetic role of DNA methylation in the model plant *Arabidopsis thaliana*. The availability of previously characterized epigenetic mutants, and diverse natural accessions in the *A. thaliana* germplasm collection enabled me to ask the following questions:

1) Does variation in methylation patterns have consequences on the expression of adjacent genes?
2) Does the absence of methylation have similar effects on all genes?
3) Is methylation linked to chromatin accessibility, and can these two epigenetic factors have a concerted effect on gene expression changes?
4) For each gene, is the relationship between methylation, chromatin accessibility and gene expression conserved across different genetic backgrounds?

To answer the above questions, I carried out the following projects which are presented as two chapters of this dissertation:

1. Re-introduction of methylation at various promoter regions in a hypomethylated *Arabidopsis thaliana FWA* epiallele to examine its consequences on gene expression (Chapter One)

2. Large-scale genome-wide hypomethylation in diverse natural accessions of *Arabidopsis thaliana* using CRISPR mutagenesis, to study the consequences on gene expression changes and chromatin accessibility (Chapter Two)

CHAPTER ONE:

Position-dependent effects of cytosine methylation on *FWA* expression in *Arabidopsis thaliana*

Citation: Srikant, T., Wibowo, A., Schwab, R. & Weigel, D. Position-dependent effects of cytosine methylation on FWA expression in *Arabidopsis thaliana*. *bioRxiv* 774281 (2019) doi:10.1101/774281.

Abstract :

Gene expression can be modulated by epigenetic modifications to chromatin, and variants of the same locus distinguished by fixed, heritable epigenetic differences are known as epialleles. DNA methylation at cytosines is a prominent epigenetic modification, particularly in plant genomes, that can modulate gene expression. There are several examples where epialleles are associated with differentially methylated regions that affect the expression of overlapping or close-by genes. However, there are also many differentially methylated regions that have not been assigned a biological function despite their proximity to genes. We investigated the positional importance of DNA methylation at the *FWA (FLOWERING WAGENINGEN)* locus in *Arabidopsis thaliana*, a paradigm for stable epialleles. We show that cytosine methylation can be established not only over the well-characterized *SINE*-derived repeat elements that overlap with the transcription start site, but also in more distal promoter regions. *FWA* silencing, however, is most effective when methylation covers the transcription start site.

Author Contributions:

Author	Author position	Scientific ideas %	Data generation %	Analysis and interpretation %	Paper writing %
Thanvi Srikant	1	55	100	60	60
Anjar Wibowo	2	15		10	5
Rebecca Schwab	3	5		10	10
Detlef Weigel*	4	25		20	25

Title of paper : Position-dependent effects of cytosine methylation on *FWA* expression in *Arabidopsis thaliana*

Status in publication process : Published as a pre-print in biorXiv

* Corresponding author

CHAPTER TWO:
Accession-specific and shared responses to genome-wide hypomethylation in *Arabidopsis thaliana*

Abstract :

Epigenetic marks including cytosine DNA methylation are important drivers of gene regulation and environmental adaptation in plants. In *Arabidopsis thaliana*, genome-wide DNA methylation marks differ substantially between wild accessions, and some of these differences have been linked to geographical origin. Genome-wide methylation in the CG context is catalysed by the MET1 methyltransferase, often in proximity to genes and transposable elements (TEs). We generated knockouts of *MET1* in 18 early-flowering *A. thaliana* accessions, to uncover how CG methylation interacts with genetic background in regulating the epigenome, the transcriptome, and phenotypic diversity. Homozygous *met1* mutants suffered from several developmental defects such as dwarfism and delayed flowering compared to their wild-type parents, in addition to accession-specific abnormalities in rosette leaf architecture, silique morphology and fertility. Inactivation of *MET1* reduces CG methylation to 0.1 - 0.5% in all accessions, and alters chromatin accessibility at several thousands of loci. The epigenetic reprogramming leads to altered gene expression and the activation of TEs that are unique to each accession. Taken together, the results underscore how methylation and methylation-induced chromatin accessibility changes can be drivers of transcriptional activity, and thus may facilitate adaptive diversification.

Author Contributions:

Author	Author position	Scientific ideas %	Data generation %	Analysis and interpretation %	Paper writing %
Thanvi Srikant	1	75	85	60	57.5
Wei Yuan	2	5	5	7.5	7.5
Rebecca Schwab	3	10	5	7.5	10
Kenneth Berendzen	4		5		
Detlef Weigel*	5	10		25	25

Title of paper : Accession-specific and shared responses to genome-wide hypomethylation in *Arabidopsis thaliana*

Status in publication process : Advanced manuscript, awaiting journal submission

* Corresponding author

DISCUSSION

When faced with challenges, living organisms can employ adaptive strategies by manipulating a diverse network of genes in the cell. For plants, understanding how such genes can be regulated to drive molecular and phenotypic changes is of prime importance in engineering new varieties that can tolerate climate change, and sustain the growing population. In this thesis, I focus on DNA methylation in plants and how its altered patterns in the genome can have an impact on gene expression and ultimately, plant phenotype.

8.1. Fine-tuning of gene expression by methylation may not be position-dependent

My first aim was to identify whether methylation changes can impact genes as a function of their distance to genes. Many epialleles (loci where methylation changes can influence gene expression levels) in *A. thaliana* exhibit methylation marks in their promoters, in close proximity to their transcription start site, or within their gene bodies. In some cases, these marks occur over repetitive sequences or remnants of transposable elements that have been integrated into the genome over evolution.

I took the example of a well-known epiallele at the *FWA* locus to ask whether methylation marks beyond known regulatory regions can affect its expression. I generated several transgenic lines carrying methylation in three different regions of the *FWA* promoter, and observed the extent to which downstream gene transcription could be silenced. I found that the strongest effects on gene downregulation occurred when methylation was established in repeat regions overlapping the transcription start site (TSS) of the gene. The effects of this downregulation could also be observed phenotypically, by the rescue of delayed flowering phenotype exhibited by the parental *fwa-1* epimutant lines. However, methylation further upstream of these repeat sequences could only minimally reduce gene expression, and this appeared to be insufficient in magnitude to observe significant phenotypic changes. I learned from these experiments that the relative position of methylation marks to genes may not necessarily determine their control of gene regulation. The nature of the underlying sequence (such as the presence of repeat sequences, transposable elements, transcription factor binding motifs, high GC content, etc.) may be more important for attracting methylation marks, which evolve as regulatory elements for genes lying in *cis* to their position [184]. Another possibility was that methylation in certain regions could be co-localized with other epigenetic factors, thereby having a concerted effect on gene expression [10–12]. Since I did not examine other chromatin marks (histone methylation, chromatin accessibility, etc.) in the

parental lines used for my experiments (*fwa*-1 epimutants) at the time, the above hypotheses could not be further investigated.

8.2. A large-scale approach: Which genes in the genome rely on methylation for their regulation and what are the different ways they can be regulated?

The *FWA* epiallele was only used as an example locus for my experiments, but the above results could certainly not be generalized for all other epialleles in the genome. For a better understanding of how methylation could affect gene expression at a genome-wide level, I aimed to generate mutant lines with highly reduced methylation throughout the genome. Since a majority of methylation marks in *A. thaliana* occur in the CG context, I carried out CRISPR/Cas9 mutagenesis to knockout the CG DNA methyltransferase gene *MET1* (*AT5G49160*) in 18 accessions of *A. thaliana* which naturally grow in diverse geographical habitats around the world **(Figure 5)**.

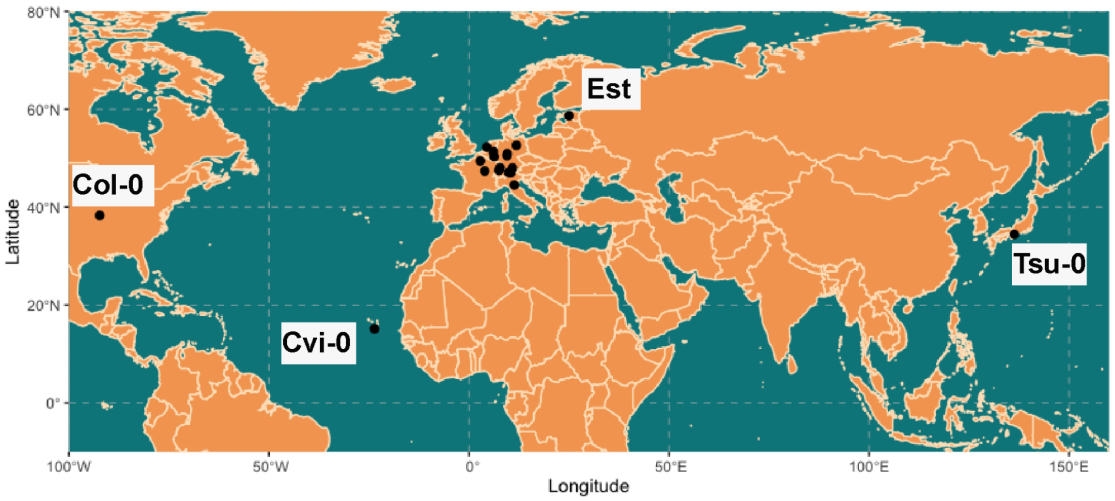


Figure 5: Geographical diversity of the 18 accessions chosen to generate *met1* knockouts. A majority of accessions are from Europe, and four accessions that are highly geographically separated from others are labelled (accessions shown as black dots).

met1 knockout mutant lines have been previously generated in *A. thaliana* [63,133,136], but their study has been limited to the reference accession Columbia (Col-0). The advantages of creating such mutants in multiple accessions for my project were manifold: firstly, I could examine how the same genes could be differently regulated by methylation under different genetic backgrounds (since each accession has distinct genetic variants, including both structural polymorphisms and transposable elements [89,99,169]). Extensive natural diversity at the methylome and transcriptome level for more than 700 accessions has been described in *A. thaliana* [79], but whether methylation is a cause or a co-occurrence for

differential gene expression was yet unclear. By making hypomethylated mutants in a subset of accessions, I could not only observe if the same genes are differently regulated in each accession, but also whether methylation can regulate the expression of a distinct group of genes in every accession. Secondly, I could also investigate whether natural variation at the chromatin level co-occurs with variation observed in methylation and gene expression. There was sufficient evidence to demonstrate that heterochromatic regions marked by dense methylation in *A. thaliana* often co-occurred with chromatin marks that were characteristic of gene-silencing, such as H3K9me2 and H3K27me3 [63,137,138], also suggesting that methylation could be associated with inaccessibility of condensed chromatin. The CRISPR *met1* mutants generated as part of my project would help identify whether the absence of methylation patterns are sufficient to drive changes in chromatin architecture (chromatin accessibility in particular) and if the interplay between methylation, gene expression and accessibility varied from accession to accession.

8.3. The challenges of propagating *met1* mutants: similar but different parent-origin effects?

While generating homozygous mutants of *met1*, I noticed striking phenotypic variation between mutants in different accessions. Common phenotypes that were found for *met1* homozygotes in all accessions were dwarfed stature, delayed flowering time and impaired fertility. The differences between accession-specific phenotypes, on the other hand, largely arose in the architecture of their rosette leaves (abnormal length, thickness, curling or size), or the extent of their sterility (normal siliques, abnormal siliques with limited seeds or empty siliques). It was interesting to note that many of these phenotypes were only observed upon inbreeding in previous *met1* mutants of the Col-0 reference accession [63], indicating that my CRISPR knockouts were more severe, and non-reference accessions may vary in their response to MET1 inactivation.

Homozygous *met1* mutants, however, were rare in the segregating population of plants derived from a heterozygous parental line. This rarity in identifying homozygotes was surprisingly consistent in all 18 accessions, and often strongly deviated from Mendelian segregation ratios (where one could expect to find at least one homozygous mutant among four segregating lines). By carrying out amplicon-sequencing of segregating individual lines, I found that segregation distortion was indeed prevalent across all the accessions in my study, with homozygous mutants being as rare as 1-2% of the population in some accessions, up to 15-18% in other accessions. Biologically, this segregation distortion

against homozygous plants could arise because of pre-fertilization defects such as non-viable gametes, non-random meiotic segregation of gametes, post-fertilization defects or germination defects [186]. Previous work on *met1-3* mutants (which carry a T-DNA insertion disrupting the *MET1* gene), identified frequent seed abortion in siliques of heterozygous plants, which significantly varied in degree between individual mutant lines. Consequently, it was observed in T2 segregating populations that the number of heterozygotes was 32% lower than expected, and in homozygotes, 94% lower. From reciprocal crosses, it was found that paternal transmission of the mutated allele was 20% lower, suggesting that inactive MET1 may impact the paternal genome by impeding successful male gametogenesis [136]. A similar observation was made by Mathieu et al., who reported that only 2% of expected numbers of homozygotes can be recovered from self-fertilized heterozygous *met1-3* parents [63].

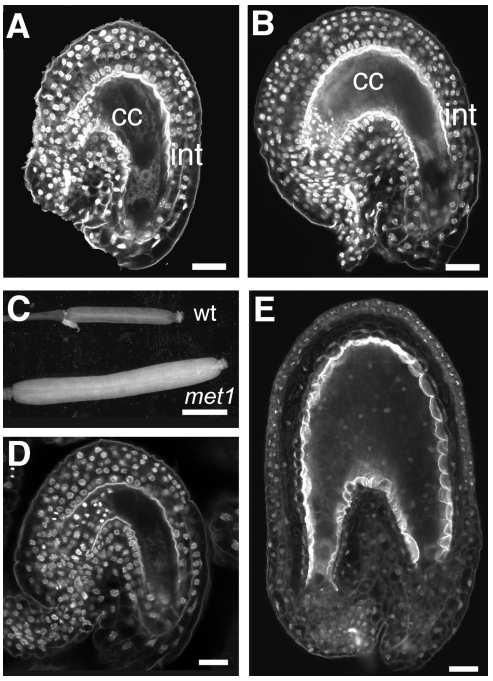


Figure 6: Ovule integument is impacted by maternal effect in *met1/met1* homozygous plants (Fig. 3 from [187]**). (A)** Mature wild-type ovule before fertilization exhibits a central cell (cc) and integument (int) layers surrounding it. **(B)** Confocal section of a *met1/met1* ovule before fertilization. **(C)** Siliques of WT and *met1/met1* plants; *met1/met1* elongate without fertilization. **(D)** Unfertilized WT ovule after pollen emasculation shows a collapsed central cell . **(E)** Unfertilized ovule in *met1/met1* after pollen emasculation shows a seed-like structure. Scale bars represent 20 µm (A, B, D and E) and 1.5 mm (**C**).

The parental inheritance of a fully functional MET1 can impact seed size in *A. thaliana*, as demonstrated by initial observations in *MET1* knockdown lines, where maternal inheritance results in larger seed size and paternal inheritance results in smaller seeds [130,188,189]. When ovules of a wild-type plant are fertilized with pollen heterozygous for the *met1-3* null

mutant allele, the seeds generated are significantly smaller in size compared to seeds obtained from wild-type crosses [187]. Furthermore, a similar experiment carried out by crosses of wild-type pollen with ovules carrying heterozygous *met1-3* mutant alleles revealed that all seeds from this cross were larger in size than a wild-type cross, irrespective of whether they carried a mutant or a wild-type *met1* allele, suggesting that the uniform effects may arise from maternal tissues that nourish the seeds or the maternal seed integument. Microscopic observations in *met1-3* homozygous seeds identified that these integument cells indeed exhibited 50% more cell growth, resulting in elongated and non-viable seeds. Together, these results show that MET1 may affect the integument, independently of female gametogenesis, but also impact seed size through paternal gametogenesis. An increase in seed size has also been observed during inbreeding of heterozygous *met1-3* plants, further suggesting that MET1-induced methylation may affect seeds before the onset of meiosis [187]. Based on these studies, I suspect that the fertility and segregation defects in my *met1* mutants may originate from the gametophytic level, which in future studies can be verified by assaying pollen viability in heterozygous *met1*, genotyping and microscopy of mutant and wild-type pollen, performing pollen competition assays, or by carrying out reciprocal crosses between mutants and wild-types in various accessions for examining embryo and seed morphology.

8.4. The molecular consequences of genome-wide hypomethylation: epigenomic imbalance

Having identified homozygous mutants in all accessions, I next investigated the changes in the methylome, gene expression and chromatin accessibility using whole-genome sequencing techniques. I sampled adult rosette leaves to examine their methylome (Bisulfite-sequencing), transcriptome (RNA-sequencing) and chromatin architecture (ATAC-sequencing). These sequencing libraries were parallely generated for the corresponding wild-type plants in all accessions. I found that *met1* mutants were largely hypomethylated in the CG- context, showing 0.1- 0.5% genome-wide CG methylation for all accessions, confirming that the CRISPR mutagenesis had been successful in the knockout of *MET1*. Genomic regions where methylation patterns differed in all or a subset of *met1* mutants and wild-types were identified as differentially methylated regions ('DMR's). As expected from the knockout of MET1 function, a majority of all DMRs (2388, 84%) were CG-DMRs. The absence of MET1 is also known to indirectly affect non-CG methylation [190], and hence I could also identify 350 CHG-DMRs and 1023 CHH DMRs in my dataset.

Approximately half of all CG-DMRs overlapped in position with a transposable element, while 60% occurred within protein coding genes, suggesting that CG-DMRs could also overlap both TEs and genes in some cases. Next, I examined accessible chromatin regions (ACRs) in all mutants and wild-types, and observed that *met1* mutants exhibited an overall increase in accessibility levels compared to wild-types, with large variations between accessions in the degree of their accessibility. To interpret how such an epigenome imbalance can impact gene expression in each accession and thereby result in phenotypic variation between mutants, I identified differentially expressed genes (DEGs) between mutants and wild-types for each accession. I observed that transposable-element genes (TE-DEGs) were consistently upregulated and that the number and expression of other protein-coding genes (Non-TE-DEGs) highly varied across accessions.

When comparing *met1* mutants and their wild-type parents, TE sequences in all accessions were marked by high chromatin accessibility and low methylation levels in *met1* mutants, as observed in previous studies [133,138,191]. However, when protein-coding genes were examined, the most distinct increases in chromatin accessibility and reduction in methylation was observed only among DEGs. Furthermore, these patterns not only varied between TE-DEGs and Non-TE-DEGs, but also differed from accession to accession, which I further investigated.

8.5. TEs are upregulated in *met1* mutants and can mobilize over generations

TE-DEGs in *met1* mutants of all 18 accessions were largely found near pericentromeric regions, and highly upregulated. In numbers, TE-DEGs differed moderately across accessions, with Bl-1 having the lowest number (648) and Tscha-1 having the highest (1,030). Incidentally, Tscha-1 also exhibited the highest number of down-regulated TE-DEGs (32). TE-DEGs were also more common between accessions than Non-TE-DEGs: of 291 DEGs that were universal to all 18 accessions, 276 (95%) were TE-DEGs. This is surprising, since each accession is known to harbor different TE families over evolution [89,99], but it appears that the loss of genome-wide methylation triggers many homologous genes across different accessions, pointing to some conserved mechanisms that silence these genes.

When examining the distribution of TE superfamilies among TE-DEGs of all accessions, the Class II DNA transposon superfamily En-Spm was found to be overrepresented in *met1* mutants. The DNA transposon *CACTA*, a member of the En-Spm family, is known to mobilize in several epigenetic mutants with unstable methylomes (*ddm1, cmt3-7,*

cmt3-7/met1-1) [134,192,193]. These elements are also transcriptionally upregulated in *met1-1* mutants of Col-0, but do not mobilize [192,194]. In the collection of *met1* mutants generated for my thesis, I found that the gene associated with *CACTA1*, AT2G12210, was differentially expressed in 12 out of 18 accessions studied, including Col-0, although I did not examine their mobilization. In future work, it would be interesting to examine whether the differential epigenetic landscape of *CACTA* in the first generation of homozygosity determines its mobilization potential in successive generations.

Genomes of epigenetic recombinant inbred lines (epiRILs) derived from crosses between *met1-3* mutants and wild-type plants are also known to carry mobilized *CACTA* elements in a substantial proportion (28%) of inbred lines [195]. Although this has so far not been linked to altered transcription at other epigenetic regulators, it provides evidence to show that epi-heterozygous states in parental lines can trigger stochastic alternative epigenetic regulatory mechanisms.

Another TE that is known to mobilize in *met1-3*- derived epiRILs is *Evadé* (*EVD*), a *Ty1/copia* long terminal repeat (LTR) retrotransposon encoded by the *AT5G17125* locus and belonging to the *ATCOPIA93* family [95]. The mobilization of *EVD* was first discovered in plants lacking male flower organs, a phenotypic defect arising from the disruption of the *LEAFY* (*LFY*) gene by this element. Other inbred lines were also seen to have *EVD* insertions in the *BRASSINOSTEROID INSENSITIVE 1* (*BRI1*) and *VARIEGATED 2* (*VAR2*) genes. *EVD* (which occurs as two copies in WT Col-0 *A. thaliana*) is transcriptionally silenced by CG methylation in WTs but lacks methylation in other contexts. However, phenotypically different epiRIL lines containing mobilized *EVD* elements carry both *EVD* epialleles, suggesting that methylation patterns for *EVD* silencing may vary from line to line even after its initial mobilization.

Interestingly, *EVD* is transcribed in first generation *met1*-3 homozygous plants, but does not mobilize until the second generation of inbreeding [95]. *EVD* can also be mobilized in epiRIL lines without causing any obvious phenotypic defects, since its transcripts specifically accumulate in the adaxial subepidermal (L2) layer of cotyledons, post the globular stage of embryonic development and until seed dormancy [196]. On generating extensive genetic diversity by transposition in each line, *EVD* can reach upto 40 copies per genome, beyond which can be re-silenced by *de novo* methylation pathways.

EVD was also identified as a TE-DEG in my collection of *met1* mutants, albeit only in 9 accessions of 18. Surprisingly, it was not found to be expressed in *met1* mutants of the Col-0

accession as previously identified, possibly due to differences in residual methylation arising from varying degrees of MET1 inactivation in each mutant background. Other TEs known to be mobilized resulting from hypomethylation in *met1* mutants include the *Hiun* (*Hi*) element from the *VANDAL21* family of DNA transposons [197] and the *ATCOPIA21* element *AT5TE65370 [198]*, both of which were identified as TE-DEGs among my *met1* mutant collection. *Hi* elements can also cause trans-demethylation of other TEs belonging to the same family, both in *met1* mutants and their F_1 epi-hybrids with *MET1* wild-type plants [198].

8.6. Epigenetic regulation at Non-TE genes is complex and distinct from TE-genes

Non-TE-DEGs in my *met1* mutant collection ranged in number from 278 in Com-1 up to 3,409 in Est. This substantial variation in the number of Non-TE-DEGs was striking, since it points to different accessions varying in sensitivity to loss of genome-wide methylation. *met1* mutants in Col-0, the reference accession used for most studies in *A. thaliana*, featured 1,044 Non-TE-DEGs, suggesting that previous work may have mostly overlooked the majority of potentially epigenetically regulated genes.

Moreover, unlike TE-DEGs, Non-TE-DEGs were both up- or down-regulated in *met1* mutants, consistent with several genes being positively regulated by DNA methylation [66,67,81]. When Non-TE-DEGs from all accessions were examined together, they showed a wide range of CG-methylation reduction (between 0% to -100%) in *met1* mutants, both for CG-DMRs occurring in the gene-body and in *cis*. Moreover, some genes that showed highly reduced methylation in *met1* mutants also increased in expression levels, suggesting that these genes require methylation for their silencing in the wild-type state.

In most angiosperms, dense CG methylation can also be commonly observed in gene bodies of constitutively expressed genes and is termed as gene-body methylation (gbM)[75]. Other genes can exhibit TE-like methylation (teM), being enriched in CHG and CHH methylation, and can favour gene silencing when co-occurring with the histone methylation mark H3K9me2. Among 725 natural accessions of *A. thaliana*, 9 - 20% (average about 17%) of all genes are characterized by gbM, and are often marked by distinct genic features such as gene length, expression levels, and nucleotide sequence properties [78]. The number of gbM genes in each accession has also been found to be negatively correlated with CHG methylation levels in the genome, suggesting a trade-off between methylation at genic loci and methylation at heterochromatin loci that are largely marked with CHG

methylation. Several studies also propose that gbM genes in many accessions can exist as teM genes in other accessions, resulting from a continuous spectrum of chromatin states at these loci [78,79,199].

However, an alternate school of thought considers gbM and teM to be two independent phenomena [81]. Shahzad et al. re-examined methylome and transcriptome data from 928 *A. thaliana* accessions [79] and found that gbM and teM methylation states often tend to occur in different genes, and that the numbers of gbM and teM genes across accessions show a weak negative correlation. Additionally, the authors found that gbM has a weak positive effect on gene transcription, in contrast to teM which has a strong negative effect, as also demonstrated from previous work [79]. These correlations remain even after eliminating the effect of genetic variation in these loci, suggesting that intragenic methylation can indeed affect transcription. Furthermore, variation in gbM at certain loci could also explain phenotypic variation for flowering time and plant fitness under two different rainfall conditions.

The collection of *met1* mutants generated as part of my thesis opens the possibility to examine the consequences of losing gbM and teM in different accessions. I examined Non-TE-DEGs from all accessions, and found that genes which were highly expressed and heavily CG-methylated in wild-type accessions ('gbM-like' genes) were largely downregulated in *met1* mutants. Interestingly, not all of these gbM-like genes were shared between the 18 accessions. Often, a gbM-like gene in one accession was marked by a different methylation state in another accession (lower CG methylation level), and showed minimal change in expression or downregulation. Yet, these genes were very rarely up-regulated, indicating that they did not have any teM-like characteristics. Although I did not examine non-CG methylation differences in my *met1* mutants in detail, I looked for genes that were highly CG-methylated, but lowly expressed in wild-type accessions ('CG-teM-like' genes). Interestingly, CG teM-like genes among the Non-TE-DEGs were always upregulated in *met1* mutants, and showed a very similar trend compared to TE-DEGs. Although these teM genes also differed across accessions, and were found to be upregulated to different extents, there were some accessions such as Cvi-0 and Est where many CG-teM-like genes were downregulated in *met1* despite not changing their methylation state. These observations point to accession-specific mechanisms that may rely on different epigenetic factors apart from methylation for regulating gene expression. Most importantly, no gbM-like genes (across all accessions) were common to CG-teM-like genes, suggesting that both these phenomena may be independent of each other, at least in the CG context. However, the presence of variable methylation and gene expression in orthologous genes across

accessions shows that there may be other mechanisms that establish a gene-regulatory gradient.

A large majority of Non-TE-DEGs exhibited accessible chromatin regions in *cis*, and accessibility changes in these genes showed a weak positive association with expression changes in *met1* mutants. On evaluating methylation, gene expression and accessibility together in *met1* mutants, an important observation made was the role of the wild-type epigenetic state in determining the degree of epigenetic change in the *met1* mutant. Wild-type methylation levels are strongly correlated with the methylation reduction level in mutants, across all accessions. It appears that loss of MET1 erases methylation marks completely and therefore proportionally to the level of methylation present in the wild-type (for every accession). Wild-type expression and accessibility levels also showed a weak negative correlation with expression change and accessibility change in *met1* mutants, respectively. This suggests a strong inter-dependence between methylation, gene expression and accessibility, where the wild-type states are optimally suited for maintaining epigenetic stability.

8.7. Investigating global and accession-specific gene-regulatory patterns in Non-TE-DEGs

To investigate epigenetic regulatory patterns in my dataset, I focused on Non-TE-DEGs, since they qualitatively (gene identity) and quantitatively (differential gene expression levels) differed to a higher degree across accessions than TE-DEGs. I investigated 392 Non-TE-DEGs and and 951 Non-TE genes (that were not differentially expressed), for which both CG-DMRs and ACRs were found in *cis*. In both groups of genes, methylation at CG-DMRs was reduced, with a roughly equal tendency to increase or decrease in their accessibility in *met1* mutants. Genes where CG methylation levels were reduced by 75% or more appeared to have a greater tendency to increase in accessibility in *met1* mutants. Additionally, I observed that genes did not separate into any distinct clusters (of methylation and accessibility changes) based on their expression differences. In other words, differential gene expression (low or high) could arise from multiple epigenetic states. Yet, a small group of DEGs that showed high increase in gene expression in *met1* mutants, were also marked by high increases in accessibility and highly reduced methylation. These DEGs (collectively the "High" cluster), when compared across accessions, were distributed in many other epigenetic states. For example, a "High" cluster gene in one accession, could show reduced accessibility, reduced expression and/or lower levels of CG methylation reduction in another

accession. These results once again point to various other unknown epigenetic mechanisms at play that can influence gene expression changes, in addition to CG methylation and chromatin accessibility.

I combined 1,343 genes (392 Non-TE-DEGs and 951 Non-TE genes) and found that both gene expression differences and accessibility differences exhibited a non-linear relationship with methylation differences in this set. However, there was a weak positive correlation between accessibility and gene expression. Interestingly, the relationship between accessibility and expression also showed minor variations from accession-to-accession, with genes in Bu-0, Aa-0 and Tsu-0 *met1* mutants exhibiting larger reductions in accessibility compared to other accessions. One interpretation of this observation is that methylation changes may have indirect effects in influencing either chromatin architecture or gene expression, which in turn can influence each other.

Next, I asked whether the observed epigenetic patterns were diverse due to variation in the number of methylated or non-methylated genes in every accession. I discovered that Cvi-0 exhibited the highest fraction of hypomethylated genes in the wild-type state, and therefore also the highest fraction with minimal reduction in methylation levels (since methylation cannot be further reduced drastically if already hypomethylated). This observation was consistent with previous findings that Cvi-0 has a lower genome-wide methylation level compared to other accessions of *A. thaliana* [79], which may also influence its chromatin properties [200].

In my dataset, Cvi-0 *met1* mutants showed changes in accessibility and minor changes in gene expression, even in genes that are hypomethylated in wild-type. Incidentally, genes which are highly methylated in wild-type Cvi-0 also show minor changes in accessibility and gene expression in *met1* mutants. This is in contrast with other accessions that show more drastic expression and accessibility changes for the same genes. Together, these results indicate that the epigenome architecture of Cvi-0 may be more refractory to methylation changes than other accessions, but nevertheless can still be perturbed by genome-wide hypomethylation.

8.8. Non-TE-DEGs comprise known epialleles which can affect mutant phenotypes

Having identified many Non-TE-DEGs across accessions, I looked for genes among these which were previously known to be regulated by DNA methylation, and could affect plant phenotypes upon their mis-regulation. Among all Non-TE-DEGs, only 15 genes were common to each of the 18 accession contrasts. Three of these genes (*FWA, SDC, AT1G59930*) are maternally imprinted and are strongly expressed in siliques of wild-type plants [201,202]. *FWA* and *SDC* were previously reported to be expressed in *met1* mutants in Col-0, but show striking gene expression changes in other accessions; for example, up to almost 1,000-fold expression change of *SDC* in Nok-3 *met1* mutants, compared to ~30-fold change in Col-0. Ectopic expression (spatial/temporal expression that differs from native state) of *FWA* and *SDC* can potentially explain some of the common phenotypes in my collection of *met1* mutants, such as delayed flowering, dwarfism and leaf curling [203][204,205].

The appearance of an indeterminate floral phenotype in some second generation homozygous *met1* mutant lines of the Tscha-1, Tsu-0 and Com-1 prompted me to investigate flower developmental genes that could be differentially regulated by methylation. Indeed, when a Tscha-1 *met1* second-generation mutant exhibiting such a phenotype was subjected to bisulfite sequencing, I found that the *AGAMOUS* (*AT4G18960*) gene was hypermethylated compared to the wild-type parents. This gene was also a Non-TE-DEG that was only expressed in Tscha-1 and expressed up to ~15-fold fold higher in *met1* mutants compared to the wild-type parents, showing partial methylation and expression also in first generation homozygous mutants. In line with these findings, *AGAMOUS* has been previously found to be expressed in leaves of *met1* knockdown lines in the accession C24 [130].

Some second generation *met1* mutants of Tsu-0 produced flower-like structures composed of leaf tissue, a phenotype similar to that of genetic knockout of several *SEP* genes, encoding MADS-box TFs that control the identity of reproductive organs [206–209]. Although the genomes of mutant lines with this specific phenotype were not sequenced, I found that all four known *SEP* genes (*SEP1-4*) were differentially expressed in some of my *met1* mutants, suggesting that these genes could be regulated by differential methylation in certain accessions. The *AGAMOUS* and *SEP* gene, along with the Non-TE-DEGs *AT1G59930* and *AT5G35120* (which are common to all accessions), all belong to the

MADS-box family of TFs that control key reproductive developmental pathways. The appearance of differential methylation at these loci, accompanied by differential gene expression in *met1* mutants, exemplifies the importance of MET1 in the tight regulation of plant development.

The *clark kent-3* hypermethylated epiallele of *SUPERMAN* (*AT3G23130*) gene is associated with strongly reduced *SUP* expression, resulting in the formation of increased stamens[210]. However, the *SUPERMAN* gene is highly expressed only in the stamen primordia during floral meristem development. Since I examined only rosette leaves in this work, I did not detect *SUPERMAN* transcripts in any of my RNA-seq datasets. Yet, I found that *SUPERMAN* was indeed hypermethylated in *met1* mutants of many accessions. Both *AGAMOUS* and *SUPERMAN* are known to be hypermethylated in previous *met1* mutants [132,138,210,211], and hypermethylated epialleles of *SUPERMAN* have also been found to occur in 12 natural *A. thaliana* accessions [212]. Furthermore, phenotypes caused by these natural epialleles of *SUPERMAN* such as curly and tetra-carpellary siliques, were also observed in my collection of *met1* mutants, and could possibly be identified by sequencing floral tissues in the mutants.

Other Non-TE-DEGs in my *met1* mutant collection that have been previously identified as epialleles include *PAI1* (*AT1G07780*) and *PAI2* (*AT5G05590*), which are involved encoding the tryptophan enzyme phosphoribosylanthranilate isomerase [213]; *ATFOLT1* (*AT5G66380*), that can cause hybrid incompatibility between accessions [214]; *IBM1* (*AT3G07610*), which encodes a histone H3K9 demethylase and is stably inherited in a hypomethylated state in *met1* epigenetic recombinant inbred lines (epiRILs) [215]; and the *ROS1* DNA demethylase (*AT2G36490*)[216], which can function as an epigenetic rheostat [67].

Notably, I also identified the gene *AT4G16890* among Non-TE-DEGs, which is regulated by the *bal* epiallele and can induce the activation of several pathogen response genes. *A. thaliana* plants homozygous for the *bal* epigenetic variant have been derived from another epigenetic mutant, *ddm1*, and exhibit severe dwarfism arising from the hyper-immune responses [217]. *met1* mutants of the Nok-3 accession which exhibited the highest differential expression of *AT4G16890* and were also severely dwarfed and were phenotypically similar to *bal* lines. In addition, gene ontology enrichment analyses of Non-TE-DEGs identified several genes involved in pathogen response, including the *PATHOGENESIS-RELATED GENE 1* (*PR1, AT2G14610*) involved in systemic acquired resistance [218].

I compared the transcript accumulation (transformed read counts) of the above-mentioned epialleles across all my RNA-seq libraries and observed that each epiallele was differentially regulated in a different subset of accessions **(Figure 7).**

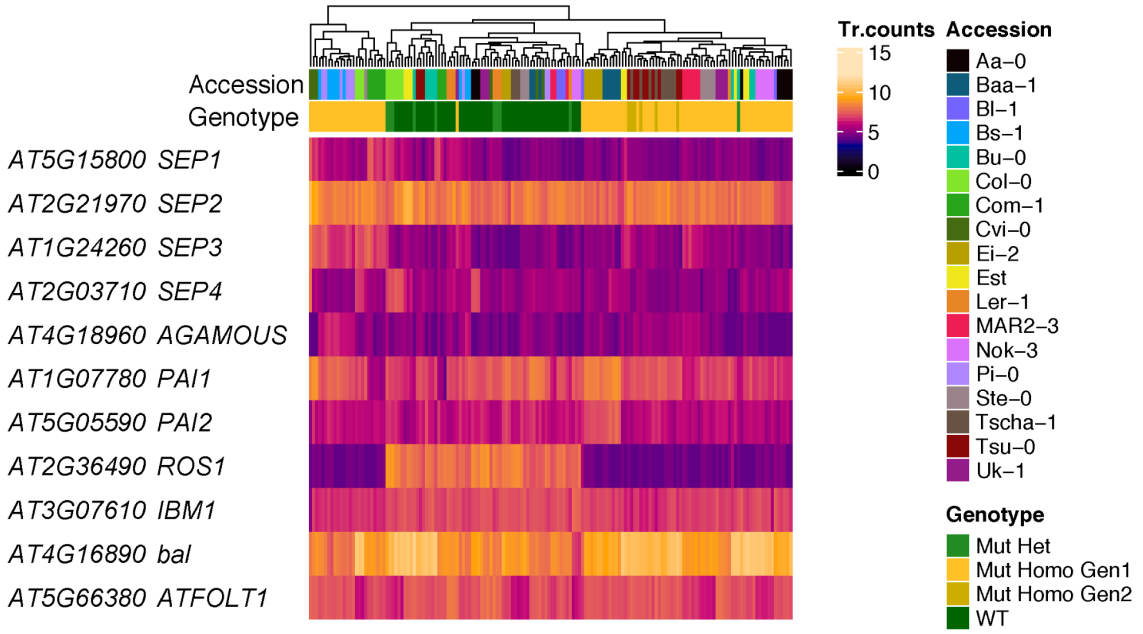


Figure 7: Variation in transcript counts across 158 RNA-seq samples for 11 Non-TE-DEGs that overlap known epialleles. The heatmap represents transformed read counts of 11 genes (represented as rows), with columns representing samples colored by accession-of-origin and genotype.

8.9. MET1 is required to maintain stable chromatin architecture

One of the major findings of this dissertation was how MET1 affects chromatin architecture and the degree to which this can differ across accessions. In general, the loss of MET1 and the CG methylation marks it establishes in the genome results in an overall increase in genome-wide chromatin accessibility. Previous work on the reference accession Col-0 has shown that euchromatic regions are enriched for accessibility compared to heterochromatic regions where accessibility is reduced, both observations being negatively correlated with the methylation levels in these regions [10,191]. Among 18 mutant lines carrying altered genome-wide methylation in Col-0, the most drastic changes in chromatin accessibility are observed in mutants that have reduced CG methylation, such as the *met1*, *fwa-1* and *ddm1* [191]. On the other hand, mutants with reduced *de novo* methylation marks (mostly CHH)

are least impacted at the chromatin level. The gain of CHG methylation in gene bodies of *met1* mutants is also suspected to be a determinant of reduced chromatin accessibility in these regions [191,219]. The redistribution of non-CG methylation marks in my *met1* mutants have currently not been explored in detail, but future work in this area may complement analyses on CG methylation, chromatin architecture and gene expression generated from this study.

In my collection of *met1* mutants, I was able to examine the relationship between methylation and chromatin accessibility in 17 accessions in addition to the reference accession Col-0, observing that CG methylation reduction is not always associated with an increase in chromatin accessibility. When all overlapping CG-DMRs and ACRs were examined, I observed that a reduction in CG methylation can result in both an increase or a decrease in accessibility, but the most dramatic increases in accessibility occured when CG methylation is highly reduced. This suggests that most regions with moderate or minimal methylation levels in wild-type genomes often go hand-in-hand with an optimal accessibility state, therefore showing only minor accessibility changes when CG methylation is lost; while highly methylated regions suffer the greatest impact on their accessibility since they may be more dependent on the co-existence of both epigenetic factors.

For TE-genes alone, the epigenetic patterns appear more uniform, since most of these genes showed similar heterochromatic silencing by methylation in the wild-type state. Therefore, almost all of the differentially expressed TE-genes (TE-DEGs) showed increased accessibility in *met1* mutants. Similar to the observations made by Zhong et al. [191], I found that there were also some TE genes that changed in accessibility, but did not exhibit differential gene expression in the *met1* mutant background, suggesting that a change in accessibility does not always guarantee a change in gene expression.

Non-TE-DEGs on the other hand, which showed a larger spectrum of gene expression changes than TE-DEGs, exhibited a stronger association with accessibility changes. In fact, many such genes often had reduced accessibility despite having reduced CG methylation. Considering two extreme ranges of methylation, accessibility and gene expression changes and various combinations of their co-occurrence, I found that Non-TE genes across 17 accessions could be classified in 8 different epigenetic states. This suggests a complex interplay between methylation, accessibility and gene expression that may vary in epigenetic regulation for every gene, and also depend on other epigenetic factors. For example, highly accessible heterochromatic regions in Col-0 *met1* show increased long-range interactions

than other regions, suggesting that loss of MET1 may directly and indirectly impact three-dimensional chromatin organization [191].

An example of epigenetic regulatory complexity is the *FWA* locus, which I had initially examined for the positional effects of DNA methylation on downstream gene expression (Chapter One). Artificial restoration of methylation marks in the hypomethylated *FWA* locus of *met1* and *fwa* mutants in the Col-0 accession was previously known to silence the gene, but has recently been shown to also reduce chromatin accessibility [191]. In my collection of *met1* mutants, I was able to examine the same locus not only in different accessions, but also in their *met1* mutant backgrounds. *FWA* expression changes and accessibility changes higher in *met1* mutants compared to wild-types for all accessions, yet showed variations that did not indicate a clear association between expression and chromatin accessibility **(Figure 8A)**. Methylation (all contexts), on the other hand, was consistently absent in all *met1* mutants, although wild-type methylation patterns largely varied **(Figure 8B)**. This example once again suggests that the differences in methylation level and patterns from the wild-type state may be the driving force behind differential chromatin accessibility and gene expression. This hypothesis further supports my inferences from the first project (Chapter One), where the nature of methylation and underlying sequence variation could be more influential in altering gene expression than the positional occurence of methylation patterns.

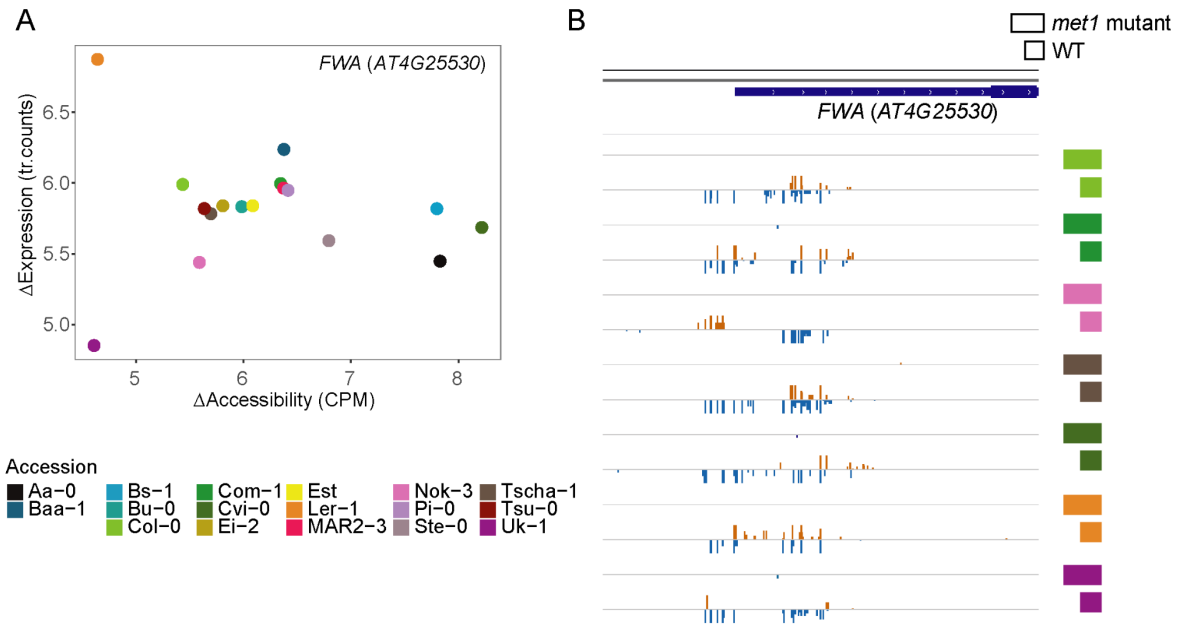


Figure 8 : Epigenetic variation at the *FWA* locus across 17 accessions. (**A**) Scatter-plot of accessibility differences vs expression differences of homozygous mutants in 17

accessions. (**B**) Browser screenshot of methylated cytosines (all-contexts) in *met1* mutants and wild-type parents of seven accessions.

My project provides some of the first evidence of quantitative variation in chromatin accessibility at the intra-species level in plants, in two different methylation backgrounds (*met1* and wild-type). Previously, the examination of leaf mesophyll nuclei from 21 accessions showed that chromatin compaction levels are associated with a latitudinal gradient across accessions, with Cvi-0 and Ler having the most and least dispersed chromocenters in their nuclei, respectively [200]. It was found that genetic variation at the *PHY-B* and *HDA6* genes was causal for the chromatin compaction differences, suggesting that histone acetylation (*HDA6*) and light-response mechanisms (*PHYB*) could be strong drivers of epigenetic adaptation in these accessions. Genetic variation could also occur as a result of TE mobilization and variation in TE copy number [89], thereby impacting either local or global epigenetic landscape. High-quality genome assemblies have enabled the identification of presence/absence variants in each accession [99,182,185], which can be important for accurately determining chromatin accessibility differences (therefore distinguishing between inaccessible chromatin and natural deletions)[182]. TEs can also mobilize in *met1* mutants, and future studies integrating the results of this project with structural variation, and intra-specific variation in other epigenetic marks can provide a better understanding of genome-wide epigenetic mechanisms.

CONCLUSION AND PERSPECTIVE

DNA methylation, which is known for its potential to regulate gene expression, can occur in diverse genome-wide patterns, even at the intra-species level, enabling adaptive responses to the environment. The broad aim of this thesis was to investigate both the local and genome-wide effects of DNA methylation in regulating gene expression in *A. thaliana*. My first project involved the artificial introduction of methylation marks at the promoter of the *FWA* gene, to ask whether proximity of methylation to the gene could impact gene expression changes. I found that upstream methylation exerts only minimal effects on gene expression, but genetic and epigenetic features of the promoter sequences carrying the methylation marks may have a greater influence on gene expression, especially if they are proximal to the gene body. My second and main project focused on the generation and molecular analyses of *met1* mutants in 18 diverse *A. thaliana* accessions, thereby simultaneously creating epi-mutations at thousands of loci in the genome. I found that the genome-wide loss of CG methylation, the most predominant context of methylation in plants, results in epigenomic instability across accessions. Almost all *met1* mutants had altered

chromatin accessibility at many loci and varying gene expression levels, including activated TEs, which are usually silenced in wild-type plants. The intersection of differential methylation, accessibility and gene expression brings to light multiple gene regulatory epigenetic states that occur differently in each accession. The response to methylation loss in each accession is highly dependent on its native epigenetic state in the wild-type background, which has evolved to optimally control the expression of several genes. Finally, the analysis of *met1* mutants in non-reference accessions demonstrates that the epigenetic flexibility of the genome is greater than previously thought, and helps uncover novel methylation-sensitive genes, apart from novel epi-allelic variants that cause multiple phenotypic defects.

The appearance of differentially methylated regions and associated epigenomic changes in first generation *met1* mutants is only an immediate response to the absence of MET1. However, it would be interesting to identify methylation, accessibility and expression states that are heritably transmitted over several generations, and examine revertant regions that exhibit unstable epigenetic marks. Even during the life cycle of the plant, methylation can vary in somatic cells of different tissues. Therefore, determining the factors that ensure meiotic transmission of "beneficial" methylation states in gametophytic tissue is essential. Moreover, these novel epialleles can be tested for their response to stress, and their interaction with homologous genes upon hybridization of divergent accessions.

In aggregate, my work increases our knowledge of how methylation can tightly balance gene regulation at the cell by intricate associations with chromatin architecture, and that destabilization of methylation catalyzes a domino-effect by altering the epigenome and also the gene expression of other epigenetic players. Studying the regulatory capacity of methylation changes and its interplay with other chromatin features such as histone methylation, histone variants, small RNAs and GC content, is essential for generating plants that can withstand strong environmental pressures.

In carrying out the above experiments, I was able to uncover the profound diversity in epigenetic and epigenomic regulatory landscapes for several accessions, extending our knowledge beyond study of the reference accession Col-0. While genetic polymorphisms have been identified for more than a thousand accessions, there remains to be complete sequence information for each accession, limiting our research to the singular Col-0 reference genome. The availability of high quality reference genomes for individual accessions will greatly improve the results from this dissertation, by enabling the investigation of large-scale structural variants and their influence on the epigenetic

landscape. Furthermore, such genomes would help disentangle the effects of genetic and epigenetic variants on gene regulation, or even be pivotal for understanding how both factors may have co-evolved.

Nevertheless, the availability of *met1* mutants in different genetic backgrounds opens possibilities for discovering novel genetic and epigenetic variants in natural populations and examining their trans-generational stability. Ultimately, similar epigenetic variants in crop species can be harnessed at a large scale for increasing plant fitness and yield under changing climates (abiotic stresses) , pathogen infections or herbivory (biotic stresses).

REFERENCES

1. Waddington CH. Epigenetics and evolution. Symp Soc Exp Biol. 1953. pp. 186–199.

2. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11: 204.

3. Liu Y, Vickers M, Dolan L, Nakajima K, Feng X. Extensive N4 cytosine methylation is essential for Marchantia sperm function. bioRxiv. 2021. Available: https://www.biorxiv.org/content/10.1101/2021.02.12.428880v1.abstract

4. Beaulaurier J, Schadt EE, Fang G. Deciphering bacterial epigenomes using modern sequencing technologies. Nat Rev Genet. 2019;20: 157–172.

5. Raghuram N, Carrero G, Th'ng J, Hendzel MJ. Molecular dynamics of histone H1. Biochem Cell Biol. 2009;87: 189–206.

6. Zhang K, Sridhar VV, Zhu J, Kapoor A, Zhu J-K. Distinctive core histone post-translational modification patterns in Arabidopsis thaliana. PLoS One. 2007;2: e1210.

7. Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, et al. Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res. 2004;32: 6511–6518.

8. Feng S, Jacobsen SE. Epigenetic modifications in plants: an evolutionary perspective. Curr Opin Plant Biol. 2011;14: 179–186.

9. Borg M, Jiang D, Berger F. Histone variants take center stage in shaping the epigenome. Curr Opin Plant Biol. 2021;61: 101991.

10. Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, et al. Widespread long-range cis-regulatory elements in the maize genome. Nat Plants. 2019. doi:10.1038/s41477-019-0547-0

11. Montgomery SA, Tanizawa Y, Galik B, Wang N, Ito T, Mochizuki T, et al. Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. Curr Biol. 2020;30: 573–588.e7.

12. Reynoso MA, Kajala K, Bajic M, West DA, Pauluzzi G, Yao AI, et al. Evolutionary flexibility in flooding response circuitry in angiosperms. Science. 2019;365: 1291–1295.

13. Marand AP, Chen Z, Gallavotti A, Schmitz RJ. A cis-regulatory atlas in maize at single-cell resolution. Cell. 2021;184: 3041–3055.e21.

14. Tannenbaum M, Sarusi-Portuguez A, Krispil R, Schwartz M, Loza O, Benichou JIC, et al. Regulatory chromatin landscape in Arabidopsis thaliana roots uncovered by coupling INTACT and ATAC-seq. Plant Methods. 2018;14: 113.

15. Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet. 1975;14: 9–25.

16. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science. 1975;187: 226–232.

17. Doskočil J, Šorm F. Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. Biochimica et Biophysica Acta (BBA) - Specialized Section on Nucleic Acids and Related Subjects. 1962;55: 953–959.

18. Bird AP, Southern EM. Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis. J Mol Biol. 1978;118: 27–47.

19. Bird AP. Use of restriction enzymes to study eukaryotic DNA methylation: II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern. J Mol Biol. 1978;118: 49–60.

20. Pikaard CS, Mittelsten Scheid O. Epigenetic regulation in plants. Cold Spring Harb Perspect Biol. 2014;6: a019315.

21. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet. 2007;39: 61–69.

22. O'Malley RC, Huang S-SC, Song L, Lewsey MG, Bartlett A, Nery JR, et al. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell. 2016;166: 1598.

23. Pavlopoulou A, Kossida S. Plant cytosine-5 DNA methyltransferases: structure, function, and molecular evolution. Genomics. 2007;90: 530–541.

24. Pósfai J, Bhagwat AS, Pósfai G, Roberts RJ. Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res. 1989;17: 2421–2435.

25. Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, et al. The DNA (cytosine-5) methyltransferases. Nucleic Acids Res. 1994;22: 1–10.

26. Wassenegger M, Heimes S, Riedel L, Sänger HL. RNA-directed de novo methylation of genomic sequences in plants. Cell. 1994;76: 567–576.

27. Henderson IR, Jacobsen SE. Epigenetic inheritance in plants. Nature. 2007;447: 418–424.

28. Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJM. RNA-mediated chromatin-based silencing in plants. Curr Opin Cell Biol. 2009;21: 367–376.

29. Zhang H, Zhu J-K. RNA-directed DNA methylation. Curr Opin Plant Biol. 2011;14: 142–147.

30. Pikaard CS, Haag JR, Pontes OMF, Blevins T, Cocklin R. A Transcription Fork Model for Pol IV and Pol V–Dependent RNA-Directed DNA Methylation. Cold Spring Harb Symp Quant Biol. 2012;77: 205–212.

31. Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet. 2014;15: 394–408.

32. Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, et al. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell. 2014;157: 1050–1060.

33. Gao Z, Liu H-L, Daxinger L, Pontes O, He X, Qian W, et al. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. Nature. 2010;465: 106–109.

34. Zhang H, Lang Z, Zhu J-K. Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol. 2018;19: 489–506.

35. Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature. 2013;498: 385–389.

36. Zhang H, Ma Z-Y, Zeng L, Tanaka K, Zhang C-J, Ma J, et al. DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proc Natl Acad Sci U S A. 2013;110: 8290–8295.

37. Smith LM, Pontes O, Searle I, Yelina N, Yousafzai FK, Herr AJ, et al. An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. Plant Cell. 2007;19: 1507–1521.

38. Zhou M, Palanca AMS, Law JA. Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. Nat Genet. 2018;50: 865–873.

39. Zhou M, Coruh C, Xu G, Bourbousse C, Lambolez A, Law JA. The CLASSY family controls tissue-specific DNA methylation patterns in Arabidopsis. bioRxiv. 2021. p. 2021.01.23.427869. doi:10.1101/2021.01.23.427869

40. Zheng B, Wang Z, Li S, Yu B, Liu J-Y, Chen X. Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes Dev. 2009;23: 2850–2860.

41. Yang D-L, Zhang G, Tang K, Li J, Yang L, Huang H, et al. Dicer-independent RNA-directed DNA methylation in Arabidopsis. Cell Res. 2016;26: 1264.

42. Ye R, Chen Z, Lian B, Rowley MJ, Xia N, Chai J, et al. A Dicer-Independent Route for Biogenesis of siRNAs that Direct DNA Methylation in Arabidopsis. Mol Cell. 2016;61: 222–235.

43. Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, et al. Widespread natural variation of DNA methylation within angiosperms. Genome Biol. 2016;17: 194.

44. Finnegan EJ, Dennis ES. Isolation and identification by sequence homology of a putative cytosine methyltransferase from Arabidopsis thaliana. Nucleic Acids Res. 1993;21: 2383–2388.

45. Genger RK, Kovac KA, Dennis ES, Peacock WJ, Finnegan EJ. Multiple DNA methyltransferase genes in Arabidopsis thaliana. Plant Mol Biol. 1999;41: 269–278.

46. Woo HR, Pontes O, Pikaard CS, Richards EJ. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes Dev. 2007;21: 267–277.

47. Woo HR, Dittmer TA, Richards EJ. Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. PLoS Genet. 2008;4: e1000156.

48. Gong Z, Morales-Ruiz T, Ariza RR, Roldán-Arjona T, David L, Zhu JK. ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell. 2002;111: 803–814.

49. Gehring M, Huh JH, Hsieh T-F, Penterman J, Choi Y, Harada JJ, et al. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell. 2006;124: 495–506.

50. Ortega-Galisteo AP, Morales-Ruiz T, Ariza RR, Roldán-Arjona T. Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. Plant Mol Biol. 2008;67: 671–681.

51. Agius F, Kapoor A, Zhu J-K. Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. Proceedings of the National Academy of Sciences. 2006. pp. 11796–11801. doi:10.1073/pnas.0603563103

52. Morales-Ruiz T, Ortega-Galisteo AP, Ponferrada-Marín MI, Martínez-Macías MI, Ariza RR, Roldán-Arjona T. DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. Proc Natl Acad Sci U S A. 2006;103: 6853–6858.

53. Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL. DNA demethylation in the Arabidopsis genome. Proceedings of the National Academy of Sciences. 2007. pp. 6752–6757. doi:10.1073/pnas.0701861104

54. Huh JH, Bauer MJ, Hsieh T-F, Fischer RL. Cellular programming of plant gene imprinting. Cell. 2008;132: 735–744.

55. Gehring M, Bubb KL, Henikoff S. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science. 2009;324: 1447–1451.

56. Hsieh T-F, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, et al. Genome-Wide Demethylation of Arabidopsis Endosperm. Science. 2009 [cited 29 Sep 2021]. Available: https://science.sciencemag.org/content/324/5933/1451.abstract

57. Ibarra CA, Feng X, Schoft VK, Hsieh T-F, Uzawa R, Rodrigues JA, et al. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science. 2012;337: 1360–1364.

58. Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, et al. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. Cell. 2002;110: 33–42.

59. Borges F, Parent J-S, van Ex F, Wolff P, Martínez G, Köhler C, et al. Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in Arabidopsis. Nat Genet. 2018;50: 186–192.

60. Slotkin RK, Vaughn M, Borges F, Tanurdzić M, Becker JD, Feijó JA, et al. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell. 2009;136: 461–472.

61. Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, et al. miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. Nature. 2014;508: 411–415.

62. Tang K, Lang Z, Zhang H, Zhu J-K. The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. Nat Plants. 2016;2: 16169.

63. Mathieu O, Reinders J, Caikovski M, Smathajitt C, Paszkowski J. Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. Cell. 2007;130: 851–862.

64. Du J, Johnson LM, Groth M, Feng S, Hale CJ, Li S, et al. Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. Mol Cell. 2014;55: 495–504.

65. Li X, Qian W, Zhao Y, Wang C, Shen J, Zhu J-K, et al. Antisilencing role of the

RNA-directed DNA methylation pathway and a histone acetyltransferase in Arabidopsis. Proc Natl Acad Sci U S A. 2012;109: 11425–11430.

66. Lei M, Zhang H, Julian R, Tang K, Xie S, Zhu J-K. Regulatory link between DNA methylation and active demethylation in Arabidopsis. Proc Natl Acad Sci U S A. 2015;112: 3553–3557.

67. Williams BP, Pignatta D, Henikoff S, Gehring M. Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. PLoS Genet. 2015;11: e1005142.

68. Lei M, Zhang H, Julian R, Tang K, Xie S, Zhu J-K. Regulatory link between DNA methylation and active demethylation in Arabidopsis. Proc Natl Acad Sci U S A. 2015;112: 3553–3557.

69. Lang Z, Wang Y, Tang K, Tang D, Datsenka T, Cheng J, et al. Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. Proc Natl Acad Sci U S A. 2017;114: E4511–E4519.

70. Liu R, How-Kit A, Stammitti L, Teyssier E, Rolin D, Mortain-Bertrand A, et al. A DEMETER-like DNA demethylase governs tomato fruit ripening. Proc Natl Acad Sci U S A. 2015;112: 10804–10809.

71. Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. Nat Rev Genet. 2016;17: 551–565.

72. Domcke S, Bardet AF, Adrian Ginno P, Hartl D, Burger L, Schübeler D. Competition between DNA methylation and transcription factors determines binding of NRF1. Nature. 2015;528: 575–579.

73. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW-L, Chen H, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell. 2006;126: 1189–1201.

74. Ahmed I, Sarazin A, Bowler C, Colot V, Quesneville H. Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. Nucleic Acids Res. 2011;39: 6919–6931.

75. Tran RK, Henikoff JG, Zilberman D, Ditt RF, Jacobsen SE, Henikoff S. DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes. Curr Biol. 2005;15: 154–159.

76. Wollmann H, Stroud H, Yelagandula R, Tarutani Y, Jiang D, Jing L, et al. The histone H3 variant H3.3 regulates gene body DNA methylation in Arabidopsis thaliana. Genome Biol. 2017;18: 94.

77. Bewick AJ, Ji L, Niederhuth CE, Willing E-M, Hofmeister BT, Shi X, et al. On the origin and evolutionary consequences of gene body DNA methylation. Proc Natl Acad Sci U S A. 2016;113: 9111–9116.

78. Zhang Y, Wendte JM, Ji L, Schmitz RJ. Natural variation in DNA methylation homeostasis and the emergence of epialleles. Proc Natl Acad Sci U S A. 2020;117: 4874–4884.

79. Kawakatsu T, Huang S-SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, et al. Epigenomic

Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell. 2016;166: 492–505.

80. Seymour DK, Koenig D, Hagmann J, Becker C, Weigel D. Evolution of DNA Methylation Patterns in the Brassicaceae is Driven by Differences in Genome Organization. PLoS Genetics. 2014. p. e1004785. doi:10.1371/journal.pgen.1004785

81. Shahzad Z, Moore JD, Zilberman D. Gene body methylation mediates epigenetic inheritance of plant traits. bioRxiv. 2021. Available: https://www.biorxiv.org/content/10.1101/2021.03.15.435374v1.abstract

82. McClintock B. The origin and behavior of mutable loci in maize. Proceedings of the National Academy of. 1950. Available: https://www.pnas.org/content/36/6/344.short

83. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet. 2007;8: 272–285.

84. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol. 2018;19: 199.

85. Hirsch CD, Springer NM. Transposable element influences on gene expression in plants. Biochim Biophys Acta Gene Regul Mech. 2017;1860: 157–165.

86. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature. 2008;452: 215–219.

87. Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol. 2014;21: 64–72.

88. Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell. 2013;153: 193–205.

89. Quadrana L, Bortolini Silveira A, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, et al. The Arabidopsis thaliana mobilome and its impact at the species level. Elife. 2016;5. doi:10.7554/eLife.15716

90. Li Q, Gent JI, Zynda G, Song J, Makarevitch I, Hirsch CD, et al. RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. Proc Natl Acad Sci U S A. 2015;112: 14728–14733.

91. Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. Nature. 2012;485: 635–641.

92. Jouffroy O, Saha S, Mueller L, Quesneville H, Maumus F. Comprehensive repeatome annotation reveals strong potential impact of repetitive elements on tomato ripening. BMC Genomics. 2016;17: 624.

93. Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T. Bursts of retrotransposition reproduced in Arabidopsis. Nature. 2009;461: 423–426.

94. Quadrana L, Etcheverry M, Gilly A, Caillieux E, Madoui M-A, Guy J, et al. Transposition favors the generation of large effect mutations that may facilitate rapid adaption. Nat Commun. 2019;10: 3421.

95. Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, et al.

Selective epigenetic control of retrotransposition in Arabidopsis. Nature. 2009;461: 427–430.

96. Kobayashi A, Kato A, Ito H. The effects of heat induction and the siRNA biogenesis pathway on the transgenerational transposition of ONSEN, a copia-like retrotransposon in Arabidopsis thaliana. Plant and Cell. 2012. Available: https://academic.oup.com/pcp/article-abstract/53/5/824/1943070

97. Gaubert H, Sanchez DH, Drost H-G, Paszkowski J. Developmental Restriction of Retrotransposition Activated in Arabidopsis by Environmental Stress. Genetics. 2017;207: 813–821.

98. La H, Ding B, Mishra GP, Zhou B, Yang H, Bellizzi M del R, et al. A 5-methylcytosine DNA glycosylase/lyase demethylates the retrotransposon Tos17 and promotes its transposition in rice. Proc Natl Acad Sci U S A. 2011;108: 15498–15503.

99. Baduel P, Leduque B, Ignace A, Gy I, Gil J Jr, Loudet O, et al. Genetic and environmental modulation of transposition shapes the evolutionary potential of Arabidopsis thaliana. Genome Biol. 2021;22: 138.

100. Baduel P, Quadrana L, Hunter B, Bomblies K, Colot V. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. Nat Commun. 2019;10: 5818.

101. Gehring M. Epigenetic dynamics during flowering plant reproduction: evidence for reprogramming? New Phytol. 2019;224: 91–96.

102. Hsieh P-H, He S, Buttress T, Gao H, Couchman M, Fischer RL, et al. Arabidopsis male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues. Proc Natl Acad Sci U S A. 2016;113: 15132–15137.

103. Park K, Kim MY, Vickers M, Park J-S, Hyun Y, Okamoto T, et al. DNA demethylation is initiated in the central cells of Arabidopsis and rice. Proc Natl Acad Sci U S A. 2016;113: 15138–15143.

104. Long J, Walker J, She W, Aldridge B, Gao H, Deans S, et al. Nurse cell-derived small RNAs define paternal epigenetic inheritance in Arabidopsis. bioRxiv. 2021. p. 2021.01.25.428150. doi:10.1101/2021.01.25.428150

105. Zhang M, Xie S, Dong X, Zhao X, Zeng B, Chen J, et al. Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. Genome Res. 2014;24: 167–176.

106. Klosinska M, Picard CL, Gehring M. Conserved imprinting associated with unique epigenetic signatures in the Arabidopsis genus. Nat Plants. 2016;2: 16145.

107. Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, et al. Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. Proc Natl Acad Sci U S A. 2013;110: 7934–7939.

108. Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M. Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. eLife. 2014. doi:10.7554/elife.03198

109. Choi Y, Harada JJ, Goldberg RB, Fischer RL. An invariant aspartic acid in the DNA

glycosylase domain of DEMETER is necessary for transcriptional activation of the imprinted MEDEA gene. Proc Natl Acad Sci U S A. 2004;101: 7481–7486.

110. Köhler C, Page DR, Gagliardini V, Grossniklaus U. The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nat Genet. 2005;37: 28–30.

111. Baroux C, Gagliardini V, Page DR, Grossniklaus U. Dynamic regulatory interactions of Polycomb group genes: MEDEA autoregulation is required for imprinted gene expression in Arabidopsis. Genes Dev. 2006;20: 1081–1086.

112. Jullien PE, Katz A, Oliva M, Ohad N, Berger F. Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in Arabidopsis. Curr Biol. 2006;16: 486–492.

113. Makarevich G, Leroy O, Akinci U, Schubert D, Clarenz O, Goodrich J, et al. Different Polycomb group complexes regulate common target genes in Arabidopsis. EMBO Rep. 2006;7: 947–952.

114. Jullien PE, Kinoshita T, Ohad N, Berger F. Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. Plant Cell. 2006;18: 1360–1372.

115. Jullien PE, Mosquna A, Ingouff M, Sakata T, Ohad N, Berger F. Retinoblastoma and its binding partner MSI1 control imprinting in Arabidopsis. PLoS Biol. 2008;6: e194.

116. Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M. Correction: Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. Elife. 2015;4. doi:10.7554/eLife.08658

117. Eichten SR, Schmitz RJ, Springer NM. Epigenetics: Beyond Chromatin Modifications and Complex Genetic Regulation. Plant Physiol. 2014;165: 933–947.

118. Kalisz S, Purugganan MD. Epialleles via DNA methylation: consequences for plant evolution. Trends Ecol Evol. 2004;19: 309–314.

119. Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, et al. Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature. 2011;480: 245–249.

120. Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, et al. Transgenerational epigenetic instability is a source of novel methylation variants. Science. 2011;334: 369–373.

121. Van Der Graaf A, Wardenaar R. Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. Proceedings of the. 2015. Available: https://www.pnas.org/content/112/21/6676.short

122. Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, et al. The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. Science. 2010;327: 92–94.

123. Hofmeister BT, Lee K, Rohr NA, Hall DW, Schmitz RJ. 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. 2017;18: 155.

124. Srikant T, Wibowo AT. The Underlying Nature of Epigenetic Variation: Origin,

Establishment, and Regulatory Function of Plant Epialleles. International Journal of Molecular Sciences. 2021. p. 8618. doi:10.3390/ijms22168618

125. Shen X, De Jonge J, Forsberg SKG, Pettersson ME, Sheng Z, Hennig L, et al. Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. PLoS Genet. 2014;10: e1004842.

126. Sasaki E, Kawakatsu T, Ecker JR, Nordborg M. Common alleles of CMT2 and NRPE1 are major determinants of CHH methylation variation in Arabidopsis thaliana. PLoS Genet. 2019;15: e1008492.

127. Dubin MJ, Zhang P, Meng D, Remigereau M-S, Osborne EJ, Paolo Casale F, et al. DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. Elife. 2015;4: e05255.

128. Quadrana L, Silveira AB, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, et al. The Arabidopsis thaliana mobilome and its impact at the species level. eLife. 2016. doi:10.7554/elife.15716

129. Stuart T, Eichten SR, Cahn J, Karpievitch YV, Borevitz JO, Lister R. Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. Elife. 2016;5. doi:10.7554/eLife.20777

130. Finnegan EJ, Peacock WJ, Dennis ES. Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. Proc Natl Acad Sci U S A. 1996;93: 8449–8454.

131. Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL. Demethylation-induced developmental pleiotropy in Arabidopsis. Science. 1996;273: 654–657.

132. Jacobsen SE, Sakai H, Jean Finnegan E, Cao X, Meyerowitz EM. Ectopic hypermethylation of flower-specific genes in Arabidopsis. Current Biology. 2000. pp. 179–186. doi:10.1016/s0960-9822(00)00324-9

133. Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, et al. Arabidopsis MET1 cytosine methyltransferase mutants. Genetics. 2003;163: 1109–1122.

134. Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T. Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature. 2001;411: 212–214.

135. Vongs A, Kakutani T, Martienssen RA, Richards EJ. Arabidopsis thaliana DNA methylation mutants. Science. 1993;260: 1926–1928.

136. Saze H, Mittelsten Scheid O, Paszkowski J. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nat Genet. 2003;34: 65–69.

137. Tariq M, Saze H, Probst AV, Lichota J, Habu Y, Paszkowski J. Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proc Natl Acad Sci U S A. 2003;100: 8823–8827.

138. Deleris A, Stroud H, Bernatavichute Y, Johnson E, Klein G, Schubert D, et al. Loss of the DNA Methyltransferase MET1 Induces H3K9 Hypermethylation at PcG Target Genes and Redistribution of H3K27 Trimethylation to Transposons in Arabidopsis

thaliana. PLoS Genetics. 2012. p. e1003062. doi:10.1371/journal.pgen.1003062

139. Brocklehurst S, Watson M, Carr IM, Out S, Heidmann I, Meyer P. Induction of epigenetic variation in Arabidopsis by over-expression of DNA METHYLTRANSFERASE1 (MET1). PLOS ONE. 2018. p. e0192170. doi:10.1371/journal.pone.0192170

140. Shu H, Wildhaber T, Siretskiy A, Gruissem W, Hennig L. Distinct modes of DNA accessibility in plant chromatin. Nat Commun. 2012;3: 1281.

141. Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, Altschuler SJ, et al. Genome-scale identification of nucleosome positions in S. cerevisiae. Science. 2005;309: 626–630.

142. Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. Epigenome characterization at single base-pair resolution. Proc Natl Acad Sci U S A. 2011;108: 18318–18323.

143. Zentner GE, Henikoff S. High-resolution digital profiling of the epigenome. Nat Rev Genet. 2014;15: 814–827.

144. Chodavarapu RK, Feng S, Bernatavichute YV, Chen P-Y, Stroud H, Yu Y, et al. Relationship between nucleosome positioning and DNA methylation. Nature. 2010;466: 388–392.

145. Widom J. Role of DNA sequence in nucleosome stability and dynamics. Q Rev Biophys. 2001;34: 269–324.

146. Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, et al. A genomic code for nucleosome positioning. Nature. 2006;442: 772–778.

147. Zhang T, Zhang W, Jiang J. Genome-Wide Nucleosome Occupancy and Positioning and Their Impact on Gene Expression and Evolution in Plants. Plant Physiol. 2015;168: 1406–1416.

148. Wu Y, Zhang W, Jiang J. Genome-wide nucleosome positioning is orchestrated by genomic regions associated with DNase I hypersensitivity in rice. PLoS Genet. 2014;10: e1004378.

149. Li G, Liu S, Wang J, He J, Huang H, Zhang Y, et al. ISWI proteins participate in the genome-wide nucleosome distribution in Arabidopsis. Plant J. 2014;78: 706–714.

150. Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc. 2010;2010: db.prot5384.

151. Zhang W, Zhang T, Wu Y, Jiang J. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. Plant Cell. 2012;24: 2719–2731.

152. Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, et al. High-resolution mapping of open chromatin in the rice genome. Genome Res. 2012;22: 151–162.

153. Pajoro A, Madrigal P, Muiño JM, Matus JT, Jin J, Mecchia MA, et al. Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. Genome Biol. 2014;15: R41.

154. Zhu B, Zhang W, Zhang T, Liu B, Jiang J. Genome-Wide Prediction and Validation of Intergenic Enhancers in Arabidopsis Using Open Chromatin Signatures. Plant Cell. 2015;27: 2415–2426.

155. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10: 1213–1218.

156. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015;109: 21.29.1–9.

157. Maher KA, Bajic M, Kajala K, Reynoso M, Pauluzzi G, West DA, et al. Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules. Plant Cell. 2018;30: 15–36.

158. Xu Y, Du J. Young but not relatively old retrotransposons are preferentially located in gene-rich euchromatic regions in tomato (Solanum lycopersicum) plants. Plant J. 2014;80: 582–591.

159. Kajala K, Shaar-Moshe L, Alex Mason G, Gouran M, Rodriguez-Medina J, Kawa D, et al. Innovation, conservation and repurposing of gene function in plant root cell type development. bioRxiv. 2020. p. 2020.04.09.017285. doi:10.1101/2020.04.09.017285

160. Borg M, Papareddy RK, Dombey R, Axelsson E, Nodine MD, Twell D, et al. Epigenetic reprogramming rewires transcription during the alternation of generations in Arabidopsis. Elife. 2021;10. doi:10.7554/eLife.61894

161. Dorrity MW, Alexandre CM, Hamm MO, Vigil A-L, Fields S, Queitsch C, et al. The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution. Nat Commun. 2021;12: 3334.

162. Noshay JM, Marand AP, Anderson SN, Zhou P. Cis-regulatory elements within TEs can influence expression of nearby maize genes. BioRxiv. 2020. Available: https://www.biorxiv.org/content/10.1101/2020.05.20.107169v1.abstract

163. Concia L, Veluchamy A, Ramirez-Prado JS, Martin-Ramirez A, Huang Y, Perez M, et al. Wheat chromatin architecture is organized in genome territories and transcription factories. Genome Biol. 2020;21: 104.

164. Wang P, Jin S, Chen X, Wu L, Zheng Y, Yue C, et al. Chromatin accessibility and translational landscapes of tea plants under chilling stress. Hortic Res. 2021;8: 96.

165. Han J, Wang P, Wang Q, Lin Q, Chen Z, Yu G, et al. Genome-Wide Characterization of DNase I-Hypersensitive Sites and Cold Response Regulatory Landscapes in Grasses. Plant Cell. 2020;32: 2457–2473.

166. Wilkins O, Hafemeister C, Plessis A, Holloway-Phillips M-M, Pham GM, Nicotra AB, et al. EGRINs (Environmental Gene Regulatory Influence Networks) in Rice That Function in the Response to Water Deficit, High Temperature, and Agricultural Environments. Plant Cell. 2016;28: 2365–2384.

167. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. 2000;408: 796–815.

168. Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, et al.

African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. Proc Natl Acad Sci U S A. 2017;114: 5213–5218.

169. 1001 Genomes Consortium. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell. 2016;166: 481–491.

170. Shirsekar G, Devos J, Latorre SM, Blaha A, Dias MQ. Fine-scale Population Structure of North American Arabidopsis thaliana Reveals Multiple Sources of Introduction from Across Eurasia. bioRxiv. 2021. Available: https://www.biorxiv.org/content/10.1101/2021.01.22.427575v1.abstract

171. Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, et al. The Scale of Population Structure in Arabidopsis thaliana. PLoS Genetics. 2010. p. e1000843. doi:10.1371/journal.pgen.1000843

172. Krämer U. Planting molecular functions in an ecological context with Arabidopsis thaliana. Elife. 2015;4. doi:10.7554/eLife.06100

173. Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, et al. Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet. 2011;43: 956–963.

174. Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, et al. Multiple reference genomes and transcriptomes for Arabidopsis thaliana. Nature. 2011;477: 419–423.

175. Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, Platzer A, et al. Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nat Genet. 2013;45: 884–890.

176. Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, et al. Patterns of population epigenomic diversity. Nature. 2013. pp. 193–198. doi:10.1038/nature11968

177. Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, et al. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature. 2010;465: 627–631.

178. Velde JV de, Van de Velde J, Van Bel M, Vaneechoutte D, Vandepoele K. A Collection of Conserved Noncoding Sequences to Study Gene Regulation in Flowering Plants. Plant Physiology. 2016. pp. 2586–2598. doi:10.1104/pp.16.00821

179. Burgess D, Freeling M. The most deeply conserved noncoding sequences in plants serve similar functions to those in vertebrates despite large differences in evolutionary rates. Plant Cell. 2014;26: 946–961.

180. Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ. The prevalence, evolution and chromatin signatures of plant regulatory elements. Nat Plants. 2019. doi:10.1038/s41477-019-0548-z

181. Yocca AE, Lu Z, Schmitz RJ, Freeling M, Edger PP. Evolution of Conserved Noncoding Sequences in Arabidopsis thaliana. Mol Biol Evol. 2021;38: 2692–2703.

182. Alexandre CM, Urton JR, Jean-Baptiste K, Huddleston J, Dorrity MW, Cuperus JT, et al. Complex Relationships between Chromatin Accessibility, Sequence Divergence, and Gene Expression in Arabidopsis thaliana. Mol Biol Evol. 2018;35: 837–854.

183. Schmitz RJ, Ecker JR. Epigenetic and epigenomic variation in Arabidopsis thaliana.

Trends Plant Sci. 2012. Available: https://www.sciencedirect.com/science/article/pii/S1360138512000027

184. Weigel D, Colot V. Epialleles in plant evolution. Genome Biol. 2012;13: 249.

185. Göktay M, Fulgione A, Hancock AM. A New Catalog of Structural Variants in 1,301 A. thaliana Lines from Africa, Eurasia, and North America Reveals a Signature of Balancing Selection at Defense Response Genes. Mol Biol Evol. 2021;38: 1498–1511.

186. Seymour DK, Chae E, Arioz BI, Koenig D, Weigel D. Transmission ratio distortion is frequent in Arabidopsis thaliana controlled crosses. Heredity . 2019;122: 294–304.

187. FitzGerald J, Luo M, Chaudhury A, Berger F. DNA methylation causes predominant maternal controls of plant embryo growth. PLoS One. 2008;3: e2298.

188. Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A. Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. Proc Natl Acad Sci U S A. 2000;97: 10637–10642.

189. Garcia D, Saingery V, Chambrier P, Mayer U, Jürgens G, Berger F. Arabidopsis haiku mutants reveal new controls of seed size by endosperm. Plant Physiol. 2003;131: 1661–1670.

190. Chan SW-L, Henderson IR, Jacobsen SE. Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet. 2005;6: 351–360.

191. Zhong Z, Feng S, Duttke SH, Potok ME, Zhang Y, Gallego-Bartolomé J, et al. DNA methylation-linked chromatin accessibility affects genomic architecture in Arabidopsis. Proc Natl Acad Sci U S A. 2021;118. doi: 10.1073/pnas.2023347118

192. Kato M, Miura A, Bender J, Jacobsen SE, Kakutani T. Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. Curr Biol. 2003;13: 421–426.

193. Singer T, Yordan C, Martienssen RA. Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). Genes Dev. 2001;15: 591–602.

194. Le NT, Harukawa Y, Miura S, Boer D, Kawabe A, Saze H. Epigenetic regulation of spurious transcription initiation in Arabidopsis. Nat Commun. 2020;11: 3224.

195. Reinders J, Wulff BBH, Mirouze M, Marí-Ordóñez A, Dapp M, Rozhon W, et al. Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes Dev. 2009;23: 939–950.

196. Marí-Ordóñez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O. Reconstructing de novo silencing of an active plant retrotransposon. Nat Genet. 2013;45: 1029–1039.

197. Hosaka A, Saito R, Takashima K, Sasaki T, Fu Y, Kawabe A, et al. Evolution of sequence-specific anti-silencing systems in Arabidopsis. Nat Commun. 2017;8: 2161.

198. Rigal M, Becker C, Pélissier T, Pogorelcnik R, Devos J, Ikeda Y, et al. Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in Arabidopsis thaliana F1 epihybrids. Proc Natl Acad Sci U S A. 2016;113: E2083–92.

199. Inagaki S, Kakutani T. What triggers differential DNA methylation of genes and TEs: contribution of body methylation? Cold Spring Harb Symp Quant Biol. 2012;77: 155–160.

200. Tessadori F, van Zanten M, Pavlova P, Clifton R, Pontvianne F, Snoek LB, et al. Phytochrome B and histone deacetylase 6 control light-induced chromatin compaction in Arabidopsis thaliana. PLoS Genet. 2009;5: e1000638.

201. Vu TM, Nakamura M, Calarco JP, Susaki D, Lim PQ, Kinoshita T, et al. RNA-directed DNA methylation regulates parental genomic imprinting at several loci in Arabidopsis. Development. 2013;140: 2953–2960.

202. Hsieh T-F, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, et al. Regulation of imprinted gene expression in Arabidopsis endosperm. Proc Natl Acad Sci U S A. 2011;108: 1755–1762.

203. Henderson IR, Jacobsen SE. Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. Genes Dev. 2008;22: 1597–1606.

204. Soppe WJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, et al. The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol Cell. 2000;6: 791–802.

205. Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJJ, Koornneef M, et al. Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. Plant J. 2007;49: 38–45.

206. Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. B and C floral organ identity functions require SEPALLATA MADS-box genes. Nature. 2000;405: 200–203.

207. Pelaz S, Tapia-López R, Alvarez-Buylla ER, Yanofsky MF. Conversion of leaves into petals in Arabidopsis. Curr Biol. 2001;11: 182–184.

208. Honma T, Goto K. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature. 2001. pp. 525–529. doi:10.1038/35054083

209. Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. Curr Biol. 2004;14: 1935–1940.

210. Jacobsen SE, Meyerowitz EM. Hypermethylated SUPERMAN epigenetic alleles in arabidopsis. Science. 1997;277: 1100–1103.

211. Kishimoto N, Sakai H, Jackson J, Jacobsen SE, Meyerowitz EM, Dennis ES, et al. Plant Molecular Biology. 2001. pp. 171–183. doi:10.1023/a:1010636222327

212. Bondada R, Somasundaram S, Marimuthu MP, Badarudeen MA, Puthiyaveedu VK, Maruthachalam R. Natural epialleles of Arabidopsis SUPERMAN display superwoman phenotypes. Commun Biol. 2020;3: 772.

213. Bender J, Fink GR. Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. Cell. 1995;83: 725–734.

214. Durand S, Bouché N, Perez Strand E, Loudet O, Camilleri C. Rapid establishment of genetic incompatibility through natural epigenetic variation. Curr Biol. 2012;22: 326–331.

215. Rigal M, Kevei Z, Pélissier T, Mathieu O. DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns. EMBO J. 2012;31: 2981–2993.

216. Julian R, Tang K, Xie S, Zhu JK. Regulatory link between DNA methylation and active demethylation in Arabidopsis. Proceedings of the. 2015. Available: https://www.pnas.org/content/112/11/3553.short

217. Stokes TL, Kunkel BN, Richards EJ. Epigenetic variation in Arabidopsis disease resistance. Genes Dev. 2002;16: 171–182.

218. Glazebrook J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annu Rev Phytopathol. 2005;43: 205–227.

219. Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell. 2013;152: 352–364.

Position-dependent effects of cytosine methylation on *FWA* expression in *Arabidopsis thaliana*

Thanvi Srikant[1], Anjar Wibowo[1,2], Rebecca Schwab[1] and Detlef Weigel[1]

[1]Max Planck Institute for Developmental Biology, Max-Planck-Ring 9, 72076 Tübingen, Germany
[2]Faculty of Science and Technology, Airlangga University, Kampus C, Mulyorejo, Surabaya City, East Java 60115, Indonesia

ABSTRACT

Gene expression can be modulated by epigenetic modifications to chromatin, and variants of the same locus distinguished by fixed, heritable epigenetic differences are known as epialleles. DNA methylation at cytosines is a prominent epigenetic modification, particularly in plant genomes, that can modulate gene expression. There are several examples where epialleles are associated with differentially methylated regions that affect the expression of overlapping or close-by genes. However, there are also many differentially methylated regions that have not been assigned a biological function despite their proximity to genes. We investigated the positional importance of DNA methylation at the *FWA (FLOWERING WAGENINGEN)* locus in *Arabidopsis thaliana*, a paradigm for stable epialleles. We show that cytosine methylation can be established not only over the well-characterized *SINE*-derived repeat elements that overlap with the transcription start site, but also in more distal promoter regions. *FWA* silencing, however, is most effective when methylation covers the transcription start site.

INTRODUCTION

Methylation of cytosine nucleotides in DNA is a prominent epigenetic mark in plant and animal genomes (1). It is found mostly over transposons and repeat elements, consistent with its primary function in silencing their transcriptional activity in association with methylation at lysine 9 of histone 3 (H3K9) (2). Mutants with defects in *METHYLTRANSFERASE 1* (*MET1*), encoding the major methyltransferase maintaining cytosine methylation in *Arabidopsis thaliana,* display various phenotypic abnormalities such as delayed flowering, dwarfism, and sterility, with increasing severity during successive rounds of inbreeding (3, 4). Genomes of *met1* mutants are largely hypomethylated at CG dinucleotides that usually inherit cytosine methylation faithfully during DNA replication through MET1 activity (5). Phenotypic abnormalities do, however, not require whole-genome changes in cytosine methylation, as a recent study describes how hypomethylation at a few select loci is sufficient to establish quantitative resistance to a pathogenic oomycete, *Hyaloperonospora arabidopsidis (Hpa)* (6).

Transposable elements and repeats in the *A. thaliana* genome are mostly confined to heterochromatic regions, such as pericentromeres and telomeres. Others are distributed in euchromatic regions, and their proximity to protein-coding genes has been associated with constitutive or induced silencing of these proximal genes upon changes in cytosine methylation (7, 8). There are several examples of loci with alternative states of cytosine methylation and associated gene expression; the variants are known as epialleles. One example comes from the *SDC* (*SUPPRESSOR OF ddc*) locus with a direct tandem repeat in its promoter. When methylation at the repeat is absent, *SDC* is expressed, resulting in a dwarfed phenotype (9). Epialleles can also form following stress exposure; treatment with the 22-amino-acid peptide flg22, an immune-response inducing fragment of the bacterial flagellin protein, causes differential methylation of helitron-derived repeats lying within a 3 kb promoter region of the defense gene *RESISTANCE METHYLATED GENE 1* (*RMG1*), and ensuing activation of *RMG1* expression (10). Furthermore, cytosine methylation has been shown to modify gene expression when located farther away from the gene body, such as in the case of *FLOWERING LOCUS T* (*FT*), where methylation on two enhancers located 5 kb upstream and 1 kb downstream of the gene

can repress transcriptional activity (11). This was shown by experimentally targeting cytosine methylation to these enhancers using Inverted Repeat-Hairpins (IR-Hairpins), which lead to the downregulation of *FT* expression and delayed flowering.

While these examples provide substantial evidence for methylation-dependent transcriptional changes, the requirements for cytosine methylation to exert this function is not well understood, as not all cytosine methylation, even when densely focused in methylated regions, triggers silencing of adjacent genes (12).

To address such functional differences, we investigated the promoter of the well-characterized *Arabidopsis thaliana fwa*-1 epiallele. The *FWA* (*FLOWERING WAGENINGEN*) locus (At4G25530) harbors two sets of tandem repeats originating from a *SINE3* retrotransposon (13). These repeats overlap the promoter and the *FWA* transcribed sequence, and are covered by dense CG methylation in wild-type plants. Throughout vegetative development, *FWA* is transcriptionally inactive, and activated only in the female gametophyte and endosperm by maternal imprinting, when DNA methylation is erased (14). Methylation at the repeats is also absent in *fwa*-1 epimutants, where the gene is constitutively active, which results in late flowering (15–18).

It is known that the presence of cytosine methylation at a specific position, the *SINE*-derived repeat elements, imposes transcriptional silencing on the entire locus. We made use of the unmethylated promoter in the *fwa*-1 epiallele and asked the reverse, whether *FWA* silencing can be triggered only by cytosine methylation at these repeats, or whether it can be similarly induced when cytosine methylation is artificially directed to other, non-repetitive, promoter elements.

RESULTS

Hairpins directing methylation to the *FWA* promoter

We chose three regions of 100 or 200 basepairs (bp) in length, located approximately 100, 500 and 700 bp upstream of the *FWA* transcription start site (Table 1 and Figure 1), and generated inverted repeat (IR)-hairpins (19), intended to introduce cytosine methylation at these regions that are otherwise unmethylated in *fwa-1* mutants.

For each region, we generated two IR-hairpins, differing in the orientation of sense and antisense sequences within the hairpin (Table 1, Figure 1, Methods).

Table 1 : IR-Hairpin constructs targeting the *FWA* promoter

Construct	Length (bp)	Orientation of 5' arm target sequence relative to *FWA* transcription	Distance from *FWA* transcription start site
1A	100	sense	-122 to -22 (Overlap with short tandem repeats)
1B	100	antisense	-122 to -22 (Overlap with short tandem repeats)
2A	100	sense	-659 to -559
2B	100	antisense	-659 to -559
3A	200	sense	-936 to -736
3B	200	antisense	-936 to -736

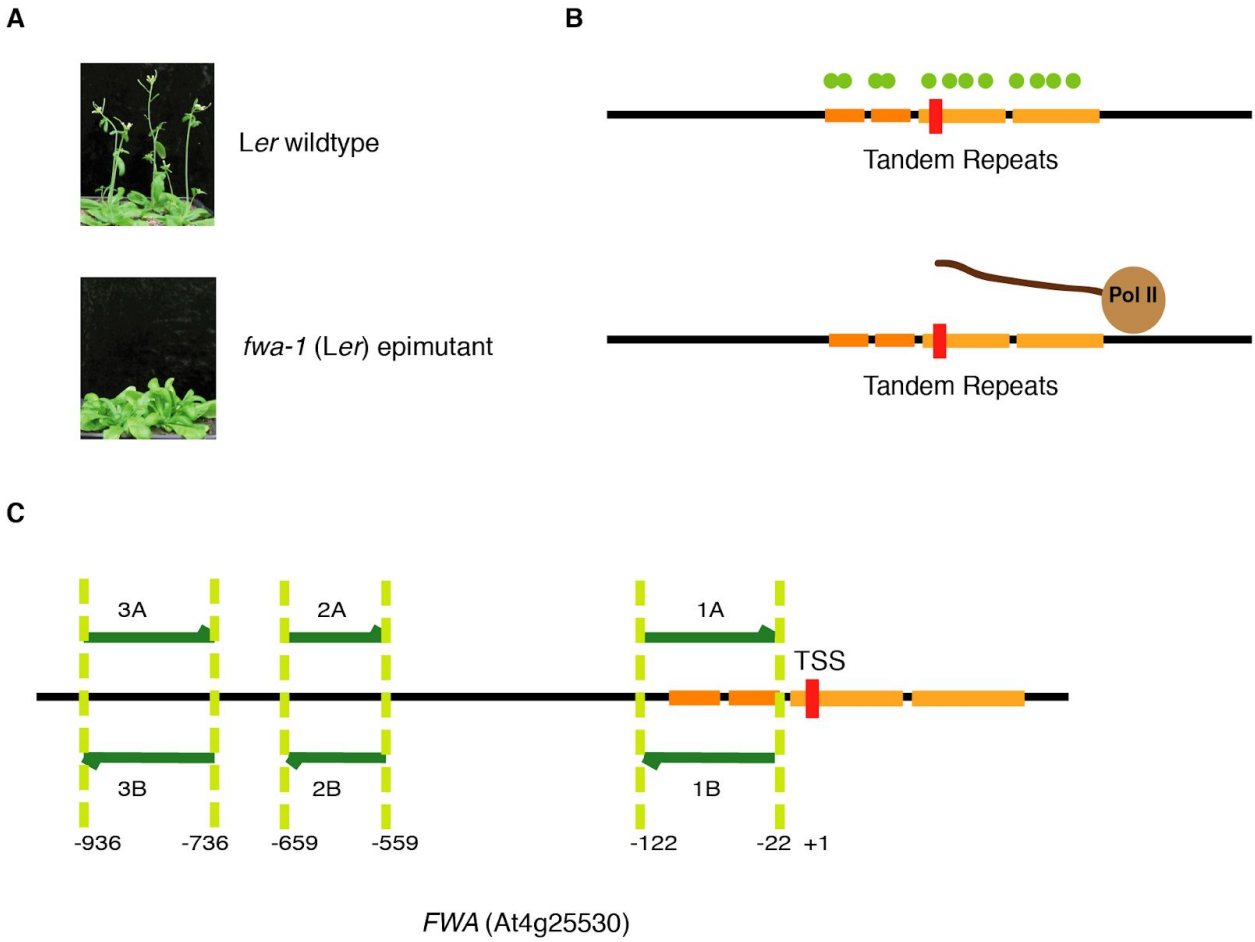


Figure 1. Methylation tiling in the *FWA* promoter

The diagram illustrates the *FWA* locus and the regions targeted for methylation tiling. (A) Late flowering phenotype of *fwa-1* (B) Release of methylation in the tandem repeats which overlap the TSS activates the *FWA* transcript (C) Three Regions in the *FWA* promoter chosen for hairpin-induced targeted methylation. Red line indicates the transcription start site (TSS) and the orange horizontal bars indicate the tandem repeats.

Region-specific methylation effects on *FWA* activity

We transformed late-flowering *fwa-1* mutants with IR-hairpin transgenes as described above (Table 1), and identified homozygous insertion lines to monitor flowering time as a proxy for *FWA* activity. Corroborating previous work (15–18, 20), hairpins derived from sequences overlapping the repeat elements surrounding the transcription start site of *FWA* (lines 1A and 1B) accelerated flowering. This was different for hairpins derived from more distal promoter

sequences (lines 2A, 2B, 3A and 3B), where we did not observe obvious flowering differences compared to parental *fwa-1* individuals (Figure 2 and Figure 3).

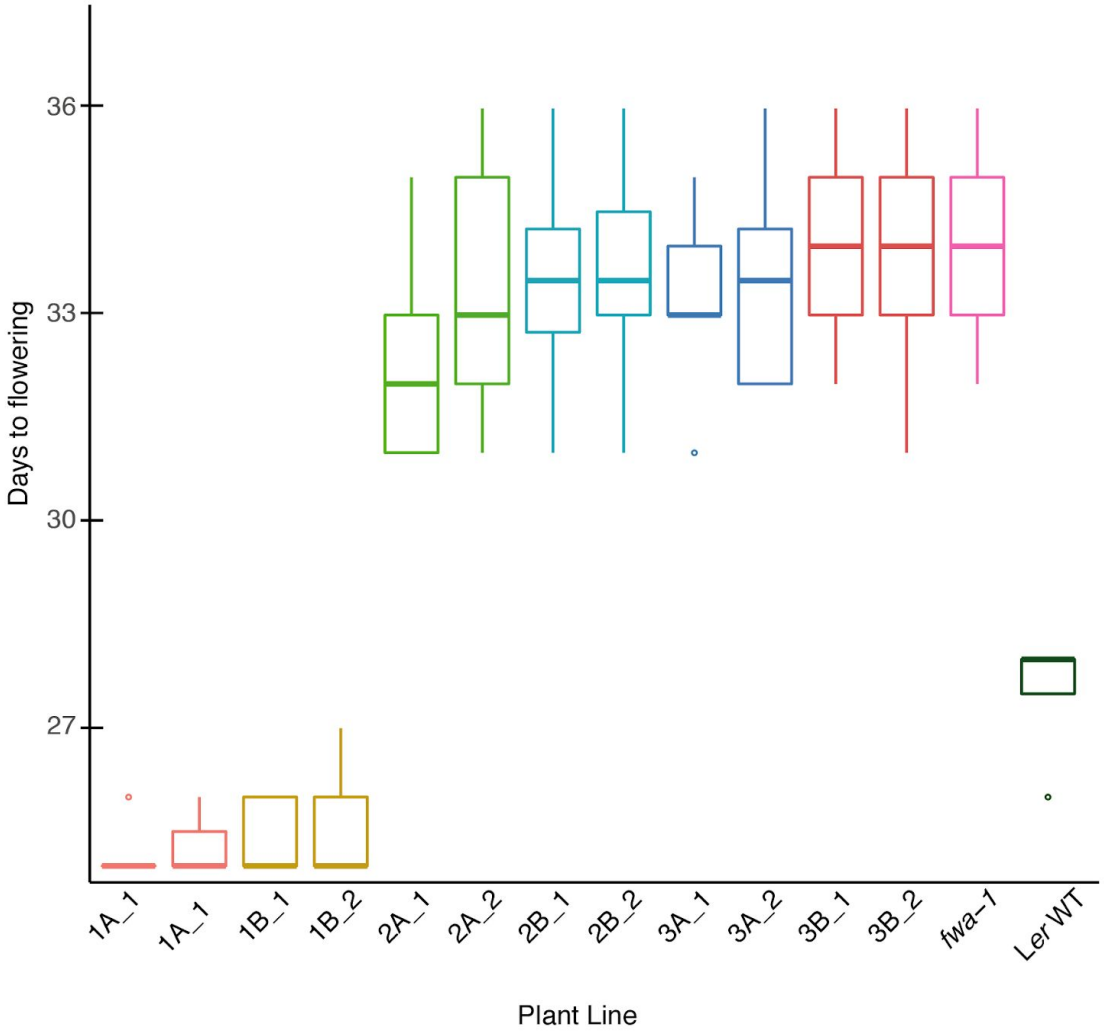


Figure 2. Flowering time in various transgenic lines

Box plots showing days to flowering in L*er* WT, *fwa-1* and various transgenic lines in the *fwa*-1 background (with 10 randomized replicates for each line).

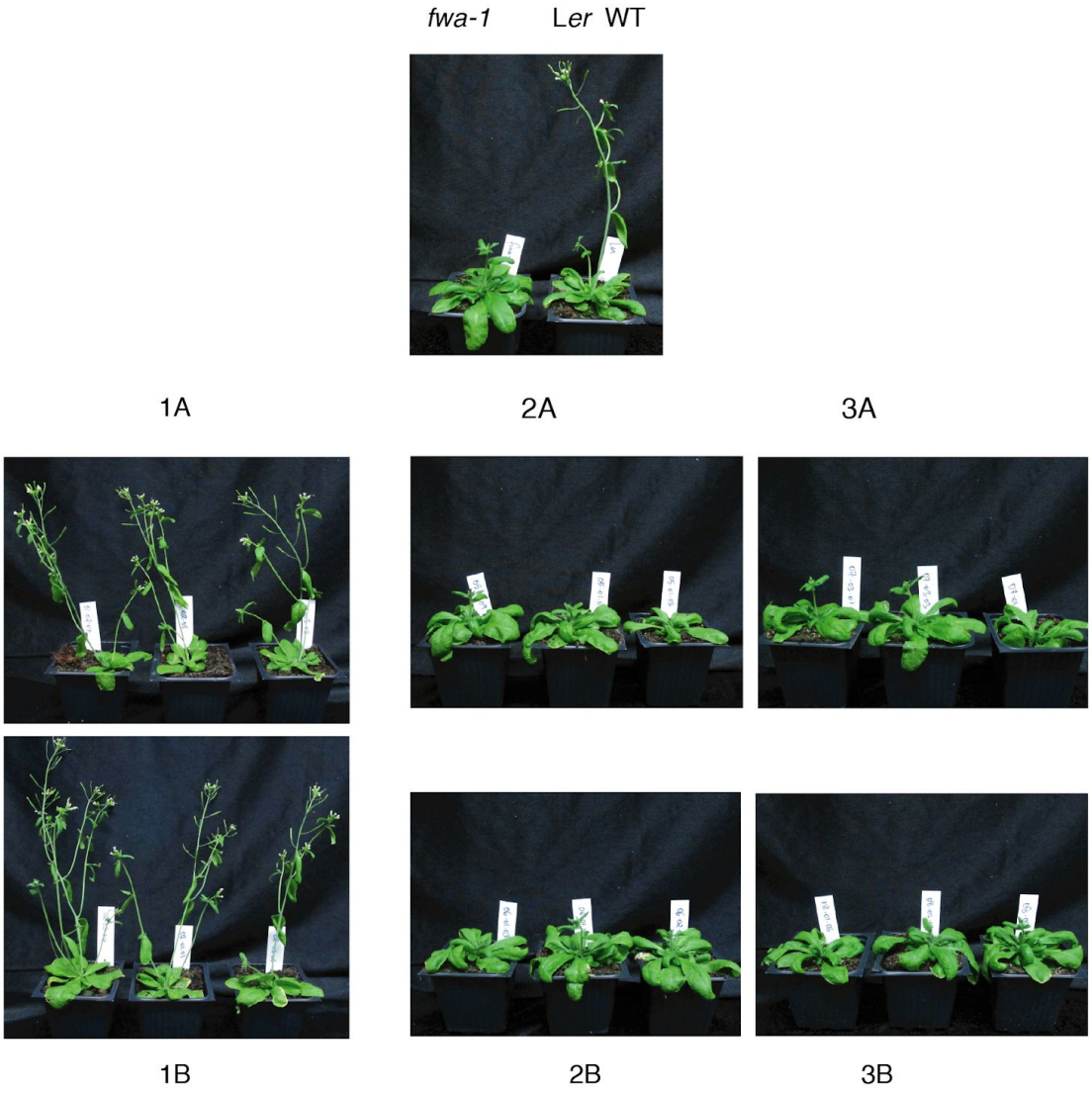


Figure 3. Phenotypes of transgenic lines

Transgenic lines and control plants at 28 days after germination.

Consistent with these observations, *FWA* transcripts in lines 1A and 1B accumulated to levels much lower than in parental *fwa*-1, and more similar to those in early-flowering L*er* wild-type plants, indicative of effective gene silencing (Figure 4). Similarly, lines in which flowering time was unaltered showed less reduction (lines 2A, 2B, 3A) or no reduction (line 3B) of *FWA* mRNA accumulation (Figure 4).

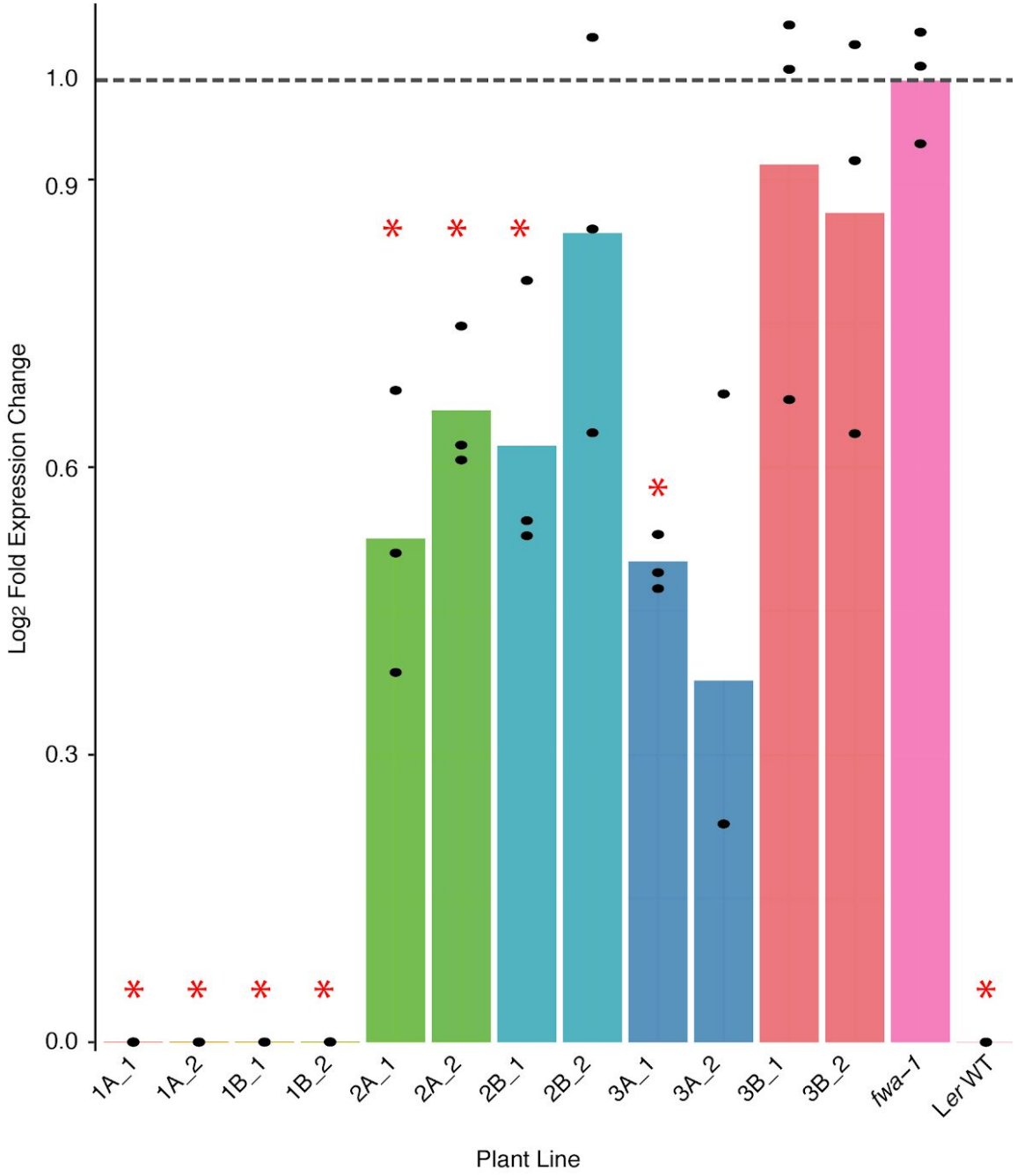


Figure 4. *FWA* mRNA accumulation in plants expressing IR-hairpins

Expression of *FWA* in L*er* WT , *fwa-1* and various transgenic lines carrying the hairpin-construct in sense and antisense orientations. Gene expression was measured by RT-qPCR. Bars show mean $\log_2$-fold change compared to *fwa-1*, with black dots representing pooled biological replicates. The red asterisks represent statistical significance ($p<0.05$ for a *Student's t-test*).

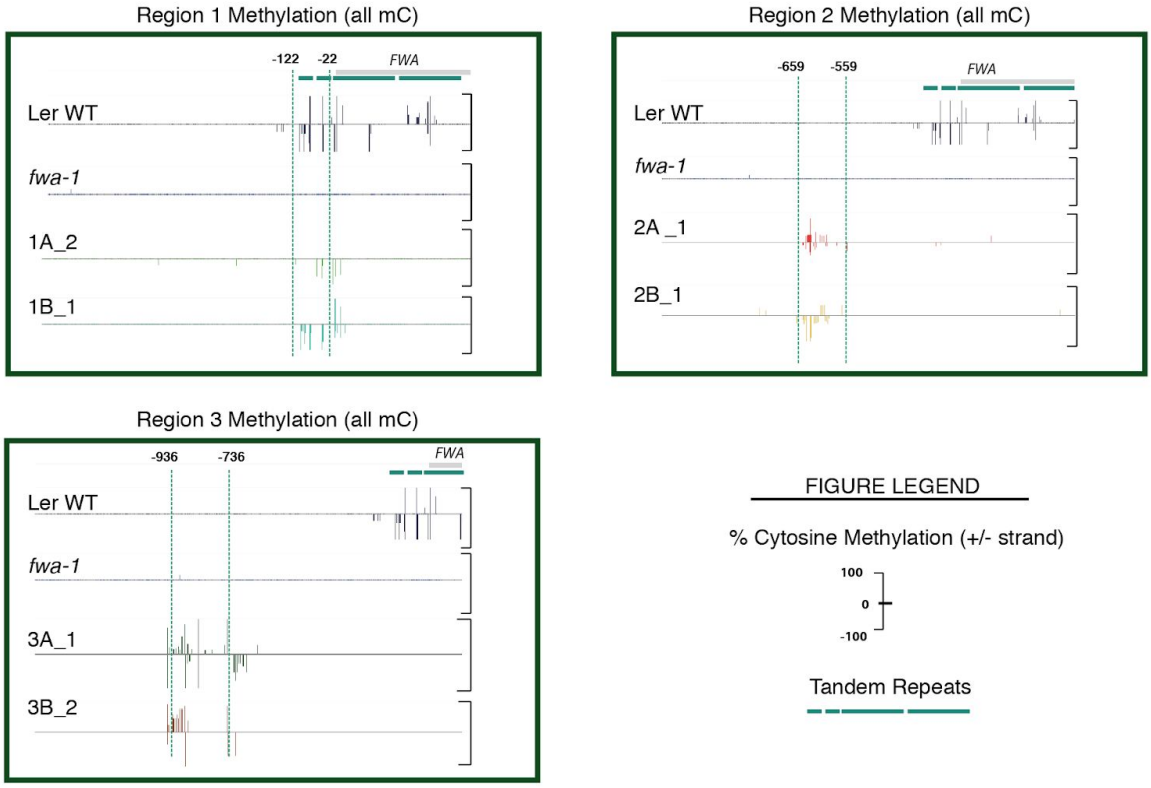


Figure 5. Methylation profile of targeted regions in transgenic lines

Hairpin-induced methylation for various transgenic lines. Each panel shows the methylation status of the targeted region in the L*er* WT, *fwa-1* and transgenic lines carrying the hairpin-construct in sense and antisense orientations. Green dotted lines demarcate the positions of each region. The vertical lines represent percentage methylation at every cytosine in both strands.

Establishment of cytosine methylation at distal promoter regions

As IR-hairpins derived from more distal regions of the *FWA* promoter did not induce sufficient *FWA* silencing to alter the timing of flowering, we asked whether cytosine methylation had been effectively established at the targeted sequences using whole-genome bisulfite sequencing. As shown in Figure 5, all six transgenes efficiently introduced methylation in the targeted regions to comparable levels. The effect of the transgene on methylation was stable over at least three generations, and indicates that cytosine methylation at this locus can be introduced

not only at the *SINE*-derived repeat elements. We further conclude that methylation at more distal promoter elements was less effective in *FWA* silencing, which may be due to the distance from the transcription start site.

DISCUSSION

Cytosine methylation can be altered at many genomic regions under environmental stress. For example, phosphate starvation in rice and *Arabidopsis thaliana* results in the generation of hypermethylated differentially methylated regions (DMRs), many of which are found over transposable elements that are close to genes induced upon such nutrient starvation (21). On the contrary, some studies report that stress-induced methylation and gene expression are not necessarily correlated with one another. In *A. thaliana*, most genomic regions exhibiting differential methylation under drought are not associated with expression changes of adjacent genes, with only 2 of 468 drought-response genes being linked to drought DMRs (22). These studies support the observation that methylation changes often flank protein-coding genes, yet only a few of them have been associated with a biological function. Several epialleles characterized in *A. thaliana* are marked by differential cytosine methylation in transposon-derived repetitive elements located proximal to the transcription start site (for example, *SDC*, (9) ; *RMG1*, (10); *HDG3*, (23)). This pattern also holds true for the *FWA* locus, where *SINE* -derived tandem repeats that overlap with the transcription start site are heavily methylated and silence the gene.

Transcription is initiated in the proximity of DNA elements providing a platform for the recruitment and assembly of polymerase II-containing complexes. Their binding is modulated by enhancers and repressors, which themselves may recognize distal sequence elements. It is likely that cytosine methylation over elements of polymerase docking may directly or indirectly hinder effective protein recruitment. When located in regions distal to the transcription start site, methylation-based silencing may be achieved by the inhibition of transcriptional enhancers (24), and by altering short or long-distance chromatin interactions (25), thus affecting

accessibility to the transcription machinery. Such a mechanism may possibly account for our observations, that methylation targeted in Regions 2 and 3, located more than 500 bp upstream of the transcription start site, only moderately impacts *FWA* transcript accumulation.

On examining publicly available data from the PlantDHS Browser (26), we observed that the 1-kb promoter region immediately upstream of the *FWA* transcription start, including all three regions targeted for methylation, has a uniform state of chromatin accessibility (Figure 6), not only when in the silent wild-type state, but also in the unsilenced state in *ddm*-1 mutants, which express *FWA* due to hypomethylation of the tandem repeats (27). Therefore, it may be expected that one would see a similar transcriptional readout upon modulating different regions within the same inaccessible sequence. At least in the *FWA* locus, this is not the case. Our results indicate that, while IR hairpin-induced methylation can be successfully introduced in the distal *FWA* promoter, this methylation can induce only moderate downregulation of *FWA* transcription, in contrast to very potent effects when methylating elements more proximal to the transcription start site.

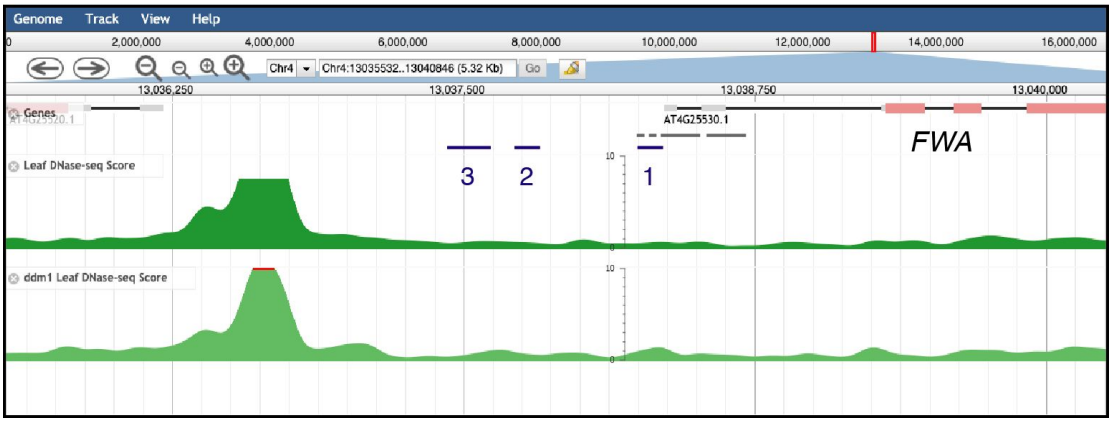


Figure 6. Chromatin conformation at the *FWA* locus

Screenshot from the PlantDHS browser (http://plantdhs.org/) showing the DNaseI-seq profile of the *FWA* locus in leaf tissue of Col-0 wild-type (top panel) and *ddm-1* mutant (bottom panel) backgrounds. Grey bars below the *FWA* gene annotation indicate the tandem repeats; blue bars indicate the three regions targeted for methylation in this study.

It will be interesting to examine the chromatin state and conformation of the *FWA* promoter in our transgenic lines, to understand whether they can explain the differential transcriptional downregulation that we observed.

This study was focused on the *FWA* locus; further investigation into differential promoter methylation at other genomic loci is required to dissect the mechanism behind methylation-dependent control on downstream transcription processes, and ultimately uncover its adaptive and biological function.

METHODS

Cloning of inverted-repeat hairpin (IR-Hairpin) constructs

Three 100 bp - 200 bp regions in the *FWA* promoter were chosen for IR-hairpin construction (Table 1 and Supplementary Table S1), and cloned in two ways: All 'A' constructs carry sequences in sense orientation relative to the open-reading frame of the *FWA* locus in the hairpin 5' arm, while the 'B' constructs are reversed.

Hairpins, including flanking attB gateway sites were synthesised (GeneArt), PCR- amplified and transferred to pDONR207 (Invitrogen) using BP Clonase II . Recombinant entry clones were subjected to LR Clonase II reaction with the pJawohl-ACT2 destination vector (28) and introduced into *E.coli* DH5α (Invitrogen) cells by heat-shock transformation. Colonies carrying the hairpin construct in the correct orientation were verified by Sanger sequencing (oligonucleotide sequence provided in Supplementary Table S1).

Recombinant plasmids (listed in Supplementary Table S4) were introduced into *Agrobacterium tumefaciens* strain GV3101(pMP90RK) (29) by electro-transformation and grown in selective LB medium. *Arabidopsis thaliana fwa-1* mutants (16) were transformed with a floral dip protocol (30). Transgenic seeds were selected with 1% BASTA (Sigma-Aldrich).

RT-qPCR

Homozygous third-generation transformants (T_3) derived from three independent parent lines were sown on half-strength MS (Duchefa Biochemie) plates with sucrose and grown under long day (16h light/8h dark) conditions at 23 °C for 10 days in a Percival chamber (Model CU -36L5, CLF Plant Climatics GmbH, Germany). For each line, 3 biological replicates of 20 pooled seedlings were subjected to RNA extraction (based on the LogSpin method, (31)). cDNA synthesis was carried out with an equimolar concentration of an Oligo(dT)18 primer and an *FWA* gene-specific primer (Supplementary Table S2) using the RevertAid First Strand cDNA

synthesis kit (Thermo Fisher Scientific), followed by qPCR of a 120bp region within the *FWA* gene body (primer sequences provided in Supplementary Table S2). The housekeeping gene *ACTIN2* (*At3G18780*) was used as a control gene for the experiment (primer sequences provided in Supplementary Table S2).

Plant growth and flowering time analyses

Seeds were sterilised by treatment with chlorine gas for 4 hours, followed by stratification in the dark at 4°C for 2 days in 0.1% agar. All plants were grown in controlled growth chambers at 23 °C, long day conditions (16h light/8h dark) with 65% relative humidity under 110 to 140 µmol m^{-2} s^{-1} light provided by Philips GreenPower TLED modules (Philips Lighting GmbH, Hamburg, Germany) with a mixture of 2:1 DR/W LB (deep red/white mixture with ca. 15% blue) and W HB (white with ca. 25% blue), respectively and watered at 2-day intervals.

Ten plants belonging to each transgene were grown on soil in a randomized–block design, to reduce position effects in the growth chamber. Flowering time was recorded when the primary inflorescence meristem was approximately 1cm in height.

Bisulfite library preparation and sequencing

Genomic DNA was isolated using the DNeasy Plant Mini Kit (Qiagen) from pools of twenty 10-day old seedlings grown on the same MS- Agar plates as seedlings used for RT-qPCR. 100ng of genomic DNA was used to prepare Bisulfite libraries with the TruSeq Nano kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions, with the modifications used in (12). The libraries were sequenced with a paired-end mode, at 125 million reads/library using an Illumina HiSeq3000 instrument.

Processing of sequenced bisulfite libraries

Raw sequencing reads were aligned using Bismark (default parameters; (32) and mapped to the *A. thaliana* (Landsberg *erecta*) reference genome. Bam files generated after deduplication of reads were processed for identifying methylated cytosines using pipelines as previously

described in (33). Methylated cytosines in the *FWA* locus were loaded onto the EPIC-CoGe browser (34) for visualisation.

Author contributions

Study design: T.S., A.W., D.W. Experimental work and data analysis: T.S. Data interpretation: T.S., A.W., R.S., D.W. Drafting of initial manuscript: T.S. Editing and finalizing of manuscript: T.S., A.W., R.S., D.W.

Acknowledgments

We thank Claude Becker (Gregor Mendel Institute, Vienna) for providing access to the pipeline for DMR calling, and Isaac Rodriguez (Gregor Mendel Institute, Vienna) for his expertise. The pJawohl-ACT2 and pDONR207 plasmid vectors were kindly shared by José Gutierrez-Marcos (University of Warwick). This study was supported by Marie Sklodowska-Curie Fellowship 751204-H2020-MSCA-IF-2016 (A.W.), DFG (ERA-CAPS AUREATE), and the Max Planck Society.

REFERENCES

1. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11:204.

2. Johnson LM, et al. (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr Biol* 17(4):379–384.

3. Kankel MW, et al. (2003) Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics* 163(3):1109–1122.

4. Mathieu O, Reinders J, Caikovski M, Smathajitt C, Paszkowski J (2007) Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell* 130(5):851–862.

5. Finnegan EJ, Dennis ES (1993) Isolation and identification by sequence homology of a putative cytosine methyltransferase from Arabidopsis thaliana. *Nucleic Acids Res* 21(10):2383–2388.

6. Furci L, et al. (2019) Identification and characterisation of hypomethylated DNA loci controlling quantitative resistance in Arabidopsis. *Elife* 8. doi:10.7554/eLife.40655.

7. Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8(4):272–285.

8. Quadrana L, et al. (2016) The Arabidopsis thaliana mobilome and its impact at the species level. *Elife* 5. doi:10.7554/eLife.15716.

9. Henderson IR, Jacobsen SE (2008) Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev* 22(12):1597–1606.

10. Fischer RL, Voinnet O, Navarro L (2013) Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. *Proceedings of the*. Available at: https://www.pnas.org/content/110/6/2389.short.

11. Zicola J, Liu L, Tänzler P, Turck F (2019) Targeted DNA methylation represses two enhancers of FLOWERING LOCUS T in Arabidopsis thaliana. *Nat Plants* 5(3):300–307.

12. Wibowo A, et al. (2018) Partial maintenance of organ-specific epigenetic marks during plant asexual reproduction leads to heritable phenotypic variation. *Proc Natl Acad Sci U S A* 115(39):E9145–E9152.

13. Lippman Z, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430(6998):471–476.

14. Kinoshita T, et al. (2004) One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science* 303(5657):521–523.

15. Koornneef M, Hanhart CJ, van der Veen JH (1991) A genetic and physiological analysis of late flowering mutants in Arabidopsis thaliana. *Mol Gen Genet* 229(1):57–66.

16. Soppe WJ, et al. (2000) The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell* 6(4):791–802.

17. Chan SW, Zhang X, Bernatavichute YV, Jacobsen SE (2006) Two-Step Recruitment of RNA-Directed DNA Methylation to Tandem Repeats. *PLoS Biology* 4(11):e363.

18. Kinoshita Y, et al. (2007) Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. *Plant J* 49(1):38–45.

19. Mette MF, Aufsatz W, van der Winden J, Matzke MA, Matzke AJM (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J* 19(19):5194–5201.

20. Johnson LM, et al. (2014) SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* 507(7490):124–128.

21. Secco D, et al. (2015) Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *Elife* 4. doi:10.7554/eLife.09343.

22. Van Dooren TJM, et al. (2018) Mild drought induces phenotypic and DNA methylation plasticity but no transgenerational effects in Arabidopsis. *bioRxiv*:370320.

23. Pignatta D, Novitzky K, Satyaki PRV, Gehring M (2018) A variably imprinted epiallele impacts seed development. *PLoS Genet* 14(11):e1007469.

24. O'Malley RC, et al. (2016) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* 166(6):1598.

25. Crevillén P, Sonmez C, Wu Z, Dean C (2013) A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. *EMBO J* 32(1):140–148.

26. Zhang T, Marand AP, Jiang J (2016) PlantDHS: a database for DNase I hypersensitive sites in plants. *Nucleic Acids Res* 44(D1):D1148–53.

27. Kakutani T (1997) Genetic characterization of late-flowering traits induced by DNA hypomethylation mutation in Arabidopsis thaliana. *Plant J* 12(6):1447–1451.

28. Wibowo A, et al. (2016) Hyperosmotic stress memory in Arabidopsis is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. *Elife* 5. doi:10.7554/eLife.13546.

29. Koncz C, Schell J (1986) The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. *Mol Gen Genet* 204(3):383–396.

30. Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* 16(6):735–743.

31. Yaffe H, et al. (2012) LogSpin: a simple, economical and fast method for RNA isolation from infected or healthy plants and other eukaryotic tissues. *BMC Res Notes* 5:45.

32. Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27(11):1571–1572.

33. Hagmann J, et al. (2015) Century-scale methylome stability in a recently diverged Arabidopsis thaliana lineage. *PLoS Genet* 11(1):e1004920.

34. Nelson ADL, Haug-Baltzell AK, Davey S, Gregory BD, Lyons E (2018) EPIC-CoGe: managing and analyzing genomic data. *Bioinformatics* 34(15):2651–2653.

SUPPLEMENTARY DATA

Table S1. Primer sequences for IR-hairpin cloning (FP forward primer, RP reverse primer)

Sequence (5'-->3')	Purpose
ATACGACTCACTATAGGGCGAA	FP for amplification of attB-target sequence from GeneART plasmid
ATGTTAATGCAGCTGGCACGAC	RP for amplification of attB-target sequence from GeneART plasmid
TTGATGCCTGGCAGTTCCCTA	FP for Colony PCR of entry clones in pDONR207
CATCAGAGATTTTGAGACACGGG	RP for Colony PCR of entry clones in pDONR207
GACAATGGTACCGGTATGGT	ACT2 promoter Sequencing primer (pJAW-ACT2 destination clones)
TCGCGTTAACGCTAGCATGGATCTC	FP for pDONR207 entry clone sequencing
GTAACATCAGAGATTTTGAGACAC	RP for pDONR207 entry clone sequencing
GGGGACAAGTTTGTACAAAAAAGCAGGCT	attBL
ACCCAGCTTTCTTGTACAAAGTGGTCCCC	attBR

Table S2. Primer sequence for RT-q-PCR analysis

Sequence (5'-->3')	Purpose
CTTTGGTACCAGCGGAGA	cDNA synthesis primer for *FWA*
TACATTGGGAGAAGTGGACTGT	FP for RT-qPCR *FWA*
GGGAGGTTTTGTGACATTGTTG	RP for RT-qPCR *FWA*
GCCATCCAAGCTGTTCTCTC	FP for RT-qPCR *ACTIN2*
GCTCGTAGTCAACAGCAACAA	RP for RT-qPCR *ACTIN2*

Table S3. Regions targeted for methylation [Sequences in 5'-->3 orientation]

Region 1A	GAGTTATGGGCCGAAGCCCATACATCTTTCCGTCGAGAATCTCAT ATATTCTTTATCGAAGCCCATACATCTTTCCGTCGAGAATCTCATA TATACCTTA
Region 1B	TAAGGTATATATGAGATTCTCGACGGAAAGATGTATGGGCTTCGA TAAAGAATAATGAGATTCTCGACGGAAAGATGTATGGGCTTCGGC CCATAACTC
Region 2A	CCTCATCTGCGCTTATAAATAAGGCAAAGCAACTAGAAAAGATTA AAACCAAAACCAAAACAAAAAACTAGTTAAGACCCTGATTTTGTTT CATAGGTAC
Region 2B	GTACCTATGAAACAAAATCAGGGTCTTAACTAGTTTTTTGTTTTGG TTTTGGTTTTAATCTTTTCTAGTTGCTTTGCCTTATTTATAAGCGCA GATGAGG
Region 3A	ATTTTTTCTATCATTTCATATCATTGTAACTATAAATTTTCGTAAAT AGACCTTTAGTGTTAATACAATAGATTTTTATTAATTTTATATCGGA TTTTGTTTAAAAAAGAAAAACCATAGGATGGATGATGATTGGTACT TATAAGATTGTAATTGGGTATTTTTGGATTGTTACCACCATTACAA AGCTATTAACAGAG
Region 3B	CTCTGTTAATAGCTTTGTAATGGTGGTAACAATCCAAAAATACCCA ATTACAATCTTATAAGTACCAATCATCATCCATCCTATGGTTTTTC TTTTTTAAACAAAATCCGATATAAAATTAATAAAAATCTATTGTATT AACACTAAAGGTCTATTTACGAAAATTTATAGTTACAATGATATGA AATGATAGAAAAAAT

Table S4. Recombinant plasmids generated for this study

Plasmid	Purpose
pTS001	pJAW-ACT2 Destination vector targeting methylation in Region 1A
pTS002	pJAW-ACT2 Destination vector targeting methylation in Region 1B
pTS003	pJAW-ACT2 Destination vector targeting methylation in Region 2A
pTS004	pJAW-ACT2 Destination vector targeting methylation in Region 2B
pTS005	pJAW-ACT2 Destination vector targeting methylation in Region 3A
pTS006	pJAW-ACT2 Destination vector targeting methylation in Region 3B

Accession-specific and shared responses to genome-wide hypomethylation in *Arabidopsis thaliana*

Thanvi Srikant[1], Wei Yuan[1], Rebecca Schwab[1], Kenneth Berendzen[2] and Detlef Weigel[1]

[1]Max Planck Institute for Developmental Biology, Tübingen, Germany
[2]Plant Transformation and Flow Cytometry Facility, ZMBP, University of Tübingen, Tübingen, Germany

ABSTRACT

Epigenetic marks including cytosine DNA methylation are important drivers of gene regulation and environmental adaptation in plants. In *Arabidopsis thaliana*, genome-wide DNA methylation marks differ substantially between wild accessions, and some of these differences have been linked to geographical origin. Genome-wide methylation in the CG context is catalysed by the MET1 methyltransferase, often in proximity to genes and transposable elements (TEs). We generated knockouts of *MET1* in 18 early-flowering *A. thaliana* accessions, to uncover how CG methylation interacts with genetic background in regulating the epigenome, the transcriptome, and phenotypic diversity. Homozygous *met1* mutants suffered from several developmental defects such as dwarfism and delayed flowering compared to their wild-type parents, in addition to accession-specific abnormalities in rosette leaf architecture, silique morphology and fertility. Inactivation of *MET1* reduces CG methylation to 0.1 - 0.5% in all accessions, and alters chromatin accessibility at several thousands of loci. The epigenetic reprogramming leads to altered gene expression and the activation of TEs that are unique to each accession. Taken together, the results underscore how methylation and methylation-induced chromatin accessibility changes can be drivers of transcriptional activity, and thus may facilitate adaptive diversification.

INTRODUCTION

In eukaryotes, gene expression can be fine-tuned by epigenetic changes such as modifications to the DNA, histone proteins, or changes in chromatin architecture. DNA methylation is established and maintained by a cohort of methyltransferases, including MET1/DNMT1 (DNA METHYLTRANSFERASE 1), which semi-conservatively copies methylation marks in the 'CG' context from the template to the daughter strand. In plants, MET1 is the principal enzyme for establishing methylation marks in newly replicating cells, especially during the various stages of fertilization and embryogenesis[1,2].

The very first studies of MET1 function in *A. thaliana* already showed that partially reduced activity of MET1 can affect the genome-wide distribution of methylation, often causing phenotypic abnormalities due to epialleles near developmental genes[3,4]. These effects are amplified in the EMS-induced *A. thaliana met1-1* mutant[5] and the T-DNA insertion mutant *met1-3*[6], where methylation is highly reduced genome-wide, particularly at pericentromeric heterochromatin. This reduction in CG methylation can trigger ectopic methylation by *de novo* pathways[7,8], affect the distribution of histone methylation marks[9–11] alter chromatin accessibility and long-range chromatin interactions [12].

Naturally inbred lines (accessions) of *A. thaliana* that have grown under diverse ecological conditions widely vary in their genetic composition, with an average of one single nucleotide polymorphism (SNP) every 200 basepairs of the genome, when comparing pairs of accessions from different parts of the range[13]. Furthemore, these accessions also vary substantially in their methylome (DNA methylation landscape), transcriptome[14] and mobilome (transposable element landscape)[15,16].

Large-scale structural variation and methylome variation at transposable elements (TEs) in *A. thaliana* accessions are known to be influenced by genetic variation at methylation-associated enzymes[14,16–18], suggesting that methylation mechanisms may have co-evolved with TE mobilization patterns. However, methylome variation is also highly prevalent in genic regions, functioning as a storehouse of epialleles[19,20], some of which can impact key developmental processes and fitness under new environments [19]. All of these findings suggest that methylome variation may have evolved to maintain genome integrity, but it remains unexplored whether such variation can also determine differential epigenome stability at the intra-species level.

In this work, we generated loss-of-function mutants of the *MET1* gene in 18 accessions of *A. thaliana* and studied how their transcriptome and chromatin architecture are differentially altered upon genome-wide CG hypomethylation. This epigenetic perturbation results in the differential transcription of many genes, with numbers ranging from hundreds to thousands across accessions. Approximately 40% of all differentially expressed genes are associated with TEs, but it is non-TE associated protein-coding genes that show the largest variation in number and expression levels across accessions. The same genes can exist in different epigenetic states, both in wild-type and in *met1* mutant backgrounds. We show that the absence of MET1 may affect key developmental abnormalities and also impede reproduction. Our results provide first evidence of intra-species epigenetic complexity in *A. thaliana* and can help identify novel gene regulatory mechanisms that shape adaptive phenotypes.

RESULTS

met1 mutants in 18 natural *A. thaliana* accessions have distinct phenotypes and different degrees of segregation distortion

We used CRISPR-Cas9 mutagenesis[21] to target the *MET1* gene(AT5G49160) of 18 early-flowering accessions of *A. thaliana*, creating frameshift mutations in exon 7 **(Supplementary tables S1,S2,S4;, Supplementary Figure S4a)**, which encodes part of the cytosine binding domain. For each accession, transgene-free T_2 mutant lines were selected. After one or two rounds of propagation of heterozygous individuals, the first generation of homozygous plants, either from the T_3 or T_4 generation, was used for phenotyping and genotyping. Compared to wild-type individuals from the respective parental accessions, mutants were dwarfed, flowered late and had altered rosette leaf architecture, although to different extents in each accession **(Figure 1a)**. Homozygous mutants also suffered from a range of silique abnormalities **(Figure 1b-d)**, in some cases affecting their fertility, with greatly diminished seed set.

Consistent with reduced transmission of *met1* alleles through the germline[6,22], homozygous mutants were underrepresented in the progeny of heterozygous parents. To accurately estimate the extent of this segregation distortion, we grew a maximum of 96 segregating progeny each from heterozygous parent lines (1-2 independently derived mutant lines per accession) and genotyped them individually using high-throughput amplicon-sequencing of the *MET1* locus. As expected from visual examination, genotyping results showed that *met1* mutants of all 18 accessions indeed exhibited segregation distortion, with homozygotes being represented below the Mendelian proportion of recessive inheritance of 25% **(Figure 2a)**. However, the extent of proportion of homozygotes varied from accession to accession, with the lowest being in Aa-0 (2%; pChi.sq < 2e-05) and the highest being in Uk-1 (18%; pChi.sq <0.1). Incidentally, Aa-0 homozygotes were among the most dwarfed in stature, and often did not survive until the onset of flowering.

While examining the *MET1* locus in individual *met1* mutant plants, we observed some unexpected genotypes in individuals belonging to seven of the 32 lines analysed. In three *met1* lines (Bs-1 Line 2, Bu-0 Line 2 and Ste-0 Line 2), we identified two mutant alleles. Bs-1 Line 2 progeny not only occurred in two homozygous states, but also in a bi-allelic state, suggesting that they likely originated from a bi-allelic parent. One of the two homozygotes was apparently more likely to survive, with both homozygotes also being phenotypically distinct from each other **(Figure 2b)**. For our downstream analyses, we examined homozygous individuals carrying the rarer allele among the two. For *met1* mutants of Bu-0 Line 2 **(Figure 2e)** and Ste-0 Line 2, the possibility of bi-allelic progenitors were ruled out due to the occurrence of heterozygous and tri-allelic genotypes including the second mutant allele.

During genotyping-by-sequencing, we also noticed individuals carrying an unequal ratio of reads representing the wild-type and mutant alleles of *MET1.* We assigned these individuals the genotype of "skewed heterozygote", and discovered that they were substantially represented (29% - 56%) in the segregants of four lines (Bu-0 *met1* Line 1, Pi-0 *met1* Line 1, Col-0 *met1* Line 2, Bl-1 *met1* Line 1). Since heterozygotes of these four lines also exhibited phenotypic variation **(Figure 2c, Supplementary Figure S4)**, we suspected that ploidy variation may be a likely

explanation. We found that both Bu-0 *met1* mutants (Line 1 and Line 2) and the wild-type Bu-0 plants used in this study were tetraploid **(Figure 2d)**. However, other lines exhibiting skewed heterozygosity were derived from diploid parents **(Supplementary Figure S4)**, awaiting a better explanation of phenotypic variation in these lines.

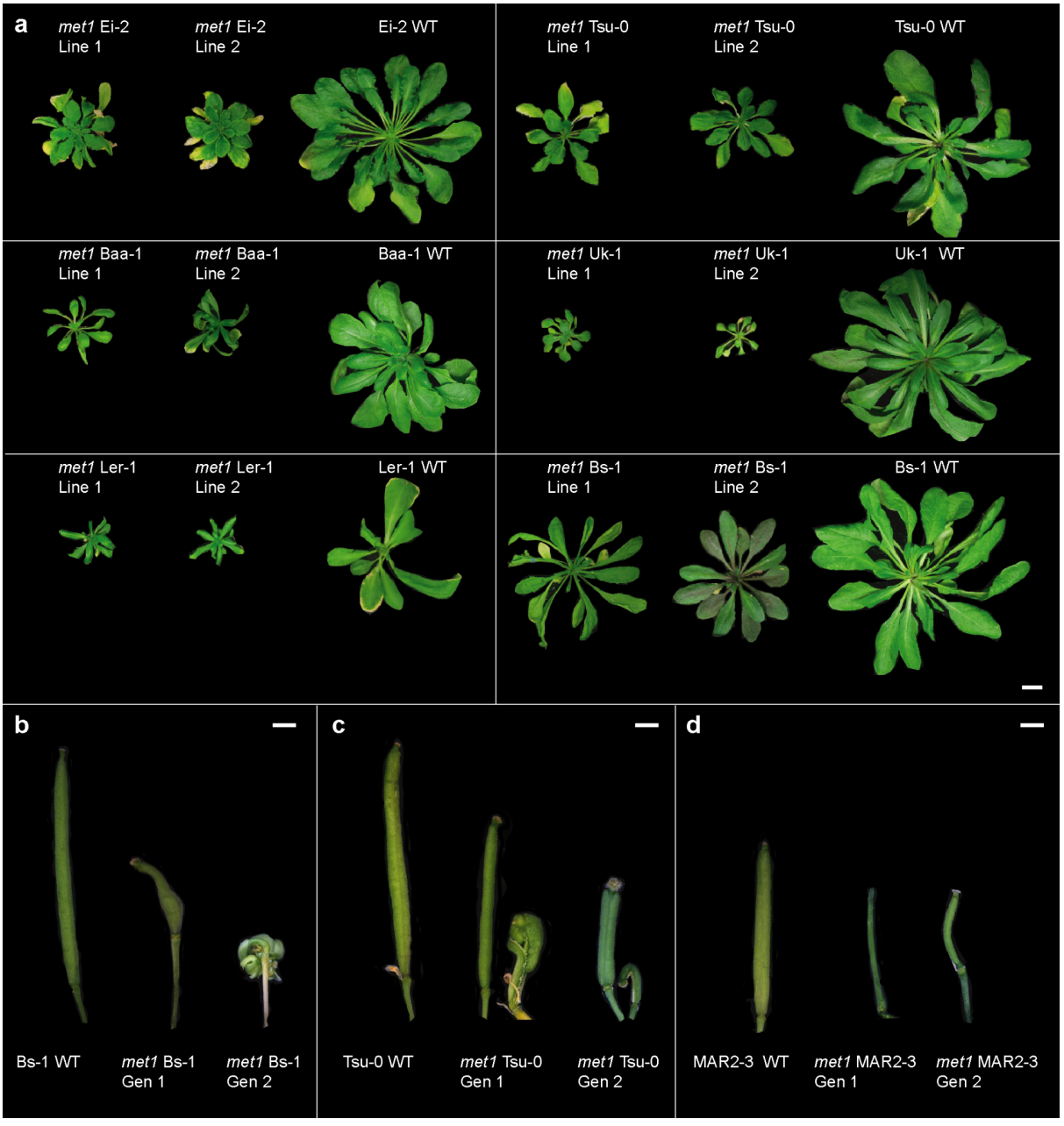


Figure 1: Rosette and silique phenotypes of *met1* mutants in various accessions. (a)Representative images of two independently derived mutant lines and a wild-type (WT) plant for six accessions at six weeks post germination; scale bar denotes 1 cm. Silique phenotypes of WT, first generation and second generation mutants from **(b)** Bs-1, **(c)** Tsu-0 and **(d)** MAR2-3 respectively. Scale bars for **(b)**, **(c)** and **(d)** represent 1 mm.

Despite the distorted segregation for homozygous individuals, we were able to analyse homozygous *met1* mutants for almost all accessions, except Bl-1. Heterozygous Bl-1 *met1* individuals also showed morphological defects, albeit not as extreme as Bl-1 *met1* homozygotes. Because homozygous individuals were also rare in Col-0, Ler-1 and Bu-0, we included in our analyses a few heterozygous individuals from these three accessions as well.

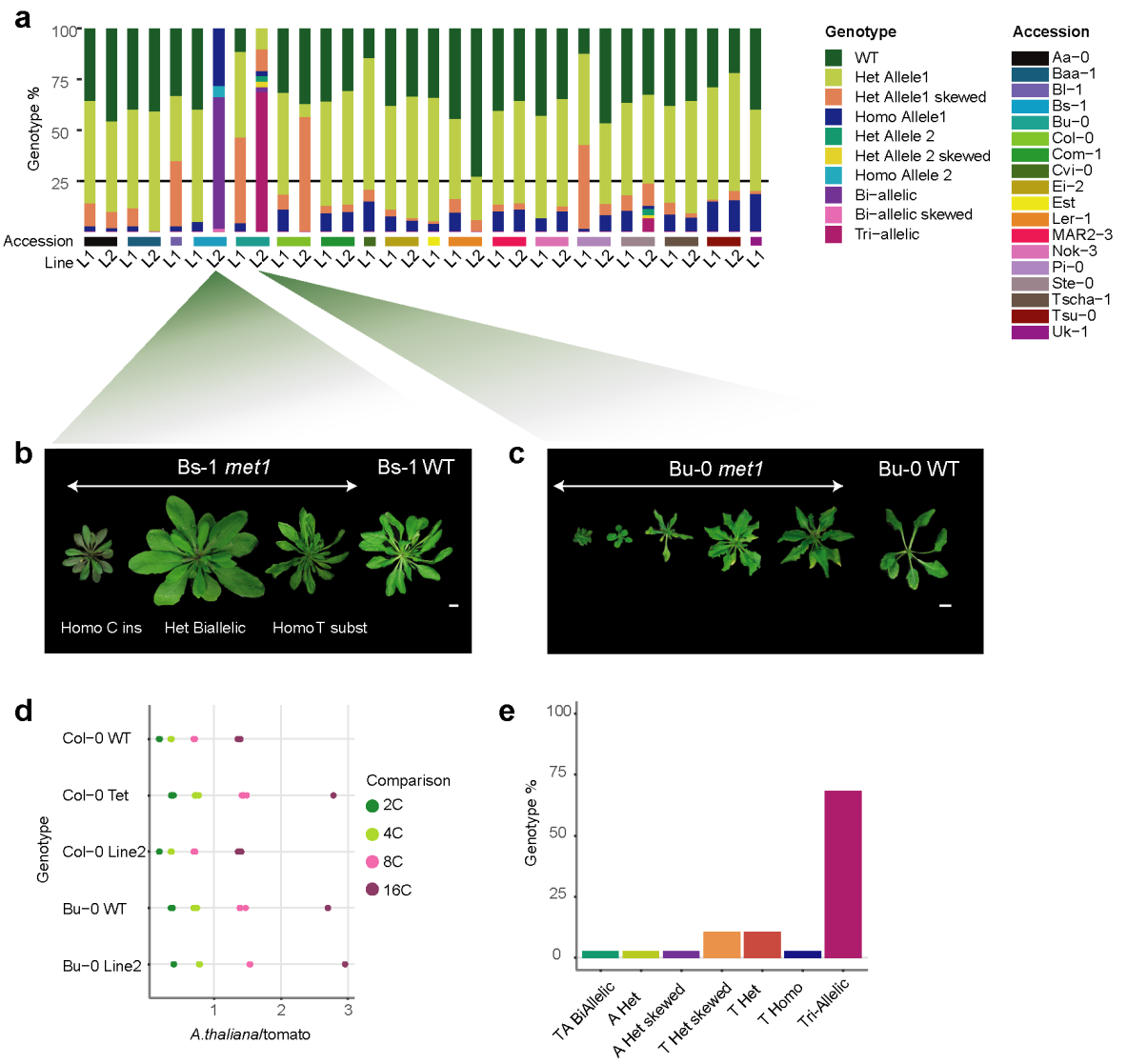


Figure 2: Segregation distortion in *met1* mutants. (a) Barplot showing proportions of various genotypes in progeny of 32 heterozygous lines. Horizontal bars indicate accessions, below which the line names are denoted. **(b)** Bi-allelic phenotypes of *met1* Bs-1 Line2 progeny. Bar = 1 cm. **(c)** Phenotypic diversity in Bu-0 polyploid *met1* mutants, Bar =1 cm. **(d)** Endopolyploidy peak position ratios (from flow cytometry profiles) in Bu-0 and Col-0 lines ('Col-Tet' indicates a Col-0 tetraploid control line) relative to the tomato internal standard. **(e)** Proportion of various segregating genotypes in *met1* Bu-0 Line 2 progeny.

To study the effects of loss of *MET1* activity on DNA methylation, chromatin accessibility and gene expression, we sampled rosette leaves from 25-day old mutant and corresponding wild-type plants. It is known that epigenetic states across the genome can diverge in different lineages of *met1* mutants over several generations[5,7]. We wanted to ensure that we could directly link chromatin state and gene expression, and therefore produced paired BS-seq, ATAC-seq and RNA-seq libraries from leaf tissue of the same plant rosettes. Three or four leaves were collected and processed for preparation of an ATAC-seq library, and the remaining leaves were homogenized and then split for preparation of BS-seq and RNA-seq libraries. All accessions were represented by three wild-type individuals and three individuals from one mutant line. For 16 accessions, we additionally included three individuals from a second, independent mutant line.

Finally, to obtain first insights into progressive changes at later generations of homozygosity, we analyzed three Tsu-0 and Tscha-1 T_4 individuals descended in both cases from the same T_3 homozygous parent. Overall, we obtained 72 BS-seq, 158 ATAC-seq and 158 RNA-seq libraries that passed quality control. The reads from all accessions were mapped against the TAIR10 reference genome, and we used TAIR10 annotations to compare genomic features.

Increased chromatin accessibility and reduced DNA methylation over TEs in *met1* mutants

Genome-wide DNA methylation levels in the CG context were drastically reduced to 0.1 to 0.5% for all homozygous *met1* mutant lines, indicating that the mutations had rendered *MET1* inactive in all accessions **(Supplementary Figure S7a).** A major role of *MET1* is in the transcriptional silencing of TEs, by establishing and maintaining CG methylation[23,24]. A combined analysis of 73 methylomes of *met1* mutants and their wild-type parents identified 2,388 differentially methylated regions (DMRs) in the CG context, 350 in the CHG context and 1,023 in the CHH context. As expected from the knockout of *MET1*, CG-DMRs showed the most striking differences between *met1* mutants and wild-type parents across all accessions, with *met1* mutants being highly hypomethylated **(Figure 3a, 3b)**. Among all CG-DMRs, 50% directly overlapped in position with a TE, while 60% overlapped protein-coding genes (including genes associated with TEs). **(Supplementary Figure S7).** A majority of CHH-DMRs (70%) also occurred at TE sequences, while >80% of CHG-DMRs occurred in protein-coding genes. Although all *met1* mutants had negligible residual CG methylation at CG-DMRs, the extent to which methylation was altered differed across accessions, consistent with accessions having unique methylomes in the presence of functional MET1[14]. Cvi-0 and Tscha-1 *met1* mutants had the lowest and highest CG methylation changes at DMRs, respectively, and only 20 CG-DMRs showed similar methylation changes across all accessions.

Since the starting methylation landscape can differ substantially between accessions [14], we asked whether this strong reduction in methylation in different accessions differentially impacts chromatin architecture in each accession. To this end, we examined our ATAC-seq data in 158 samples from *met1* mutants and wild-type plants to identify accessible chromatin regions (ACRs) in the genome. Treating each biological replicate independently, we generated a consensus set of 34,966 ACRs **(Supplementary Figure S8a)**. UMAP visualization of scaled accessibility in consensus ACRs revealed that wild-type plants as a group were distinct from *met1* mutants as a group **(Figure 3c)**. Not only different wild-type, but also different *met1* individuals from the same accession clustered close to each other, with wild-type plants from different accessions being more similar to each other than *met1* mutants from different accessions. As a corollary, when comparing how much wild-type and *met1* individuals of each accession differed from each other, it was clear that the effects of *met1* mutations varied across accessions, with the smallest difference between wild-type plants and *met1* mutants in Cvi-0. This variation is not explained by *met1* mutants having many more ACRs, and therefore being more likely to vary between accessions.

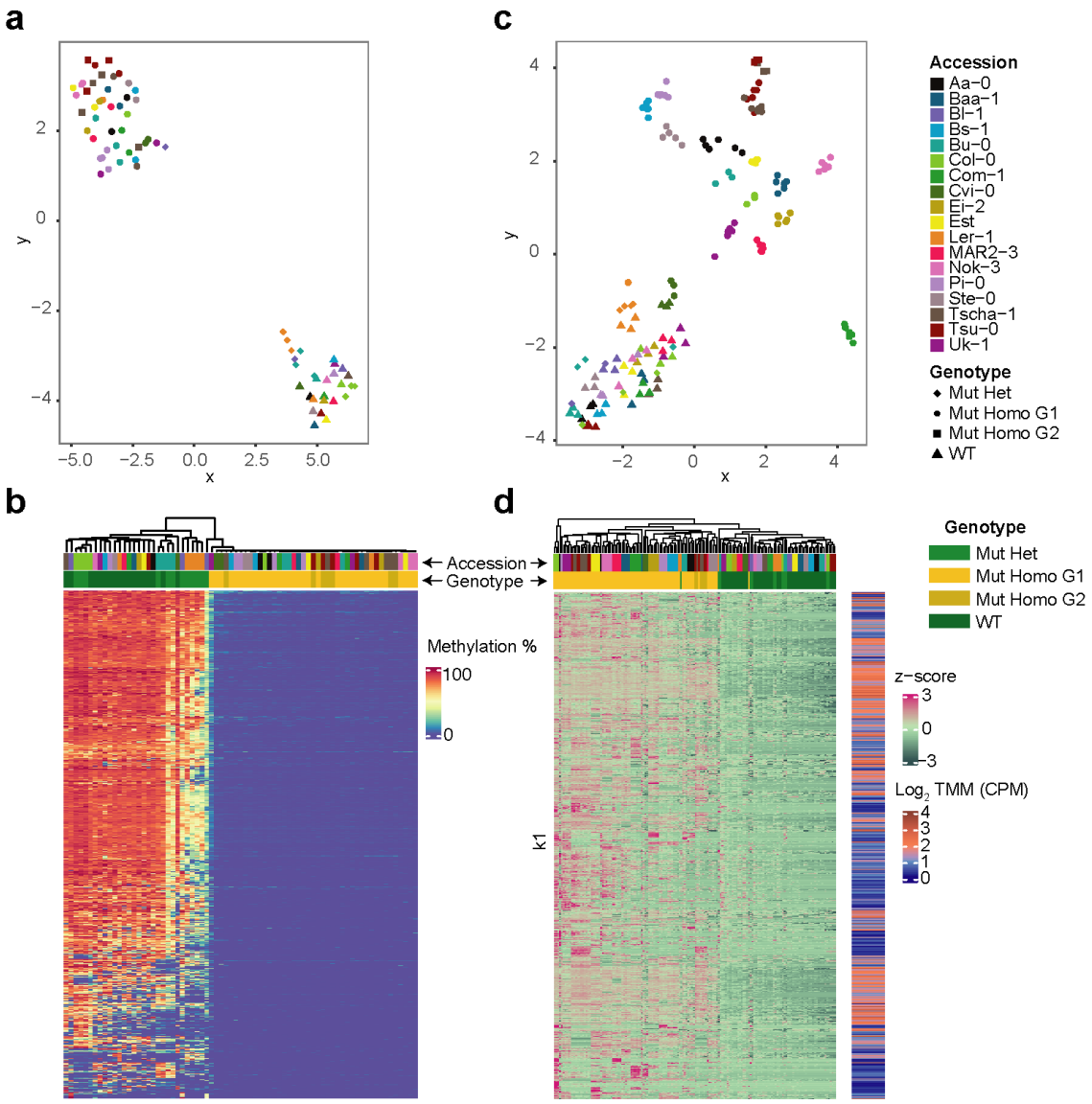


Figure 3: *met1* mutants have reduced CG methylation and increased chromatin accessibility. (a) UMAP visualization and **(b)** heatmap of CG methylation levels in 73 samples (18 wild-type plants and 55 mutants) across 749 CG-DMRs (from a total of 2,388 CG-DMRs). **(c)** Chromatin accessibility in $\log_2$(CPM) of 34,966 ACRs measured across 158 samples (54 wild-type plants and 104 mutants) visualized by UMAP. **(d)** z-scaled values of 5,560 consensus ACRs from k-group 1 as a heatmap, with mean accessibility indicated on the right. Methylation levels in **(b)** are represented as %CG methylation in CG-DMRs and accessibility levels in **(d)** are represented as TMM normalized values in counts per million (CPM). Genotypes represented are wild-types ('WT'), heterozygous *met1* mutants ('Mut Het'), first generation homozygous *met1* mutants ('Mut Homo G1') and second generation homozygous *met1* mutants ('Mut Homo G2').

Similarities between the 34,966 ACRs clustered samples into 5 k-groups according to shared variation across accessions. We next examined the variation in scaled accessibility levels across each ACR. Mutant lines had generally higher accessibility than wild-type plants; this difference was particularly stark in group k1 with 5,610 ACRs **(Figure 3d)**.

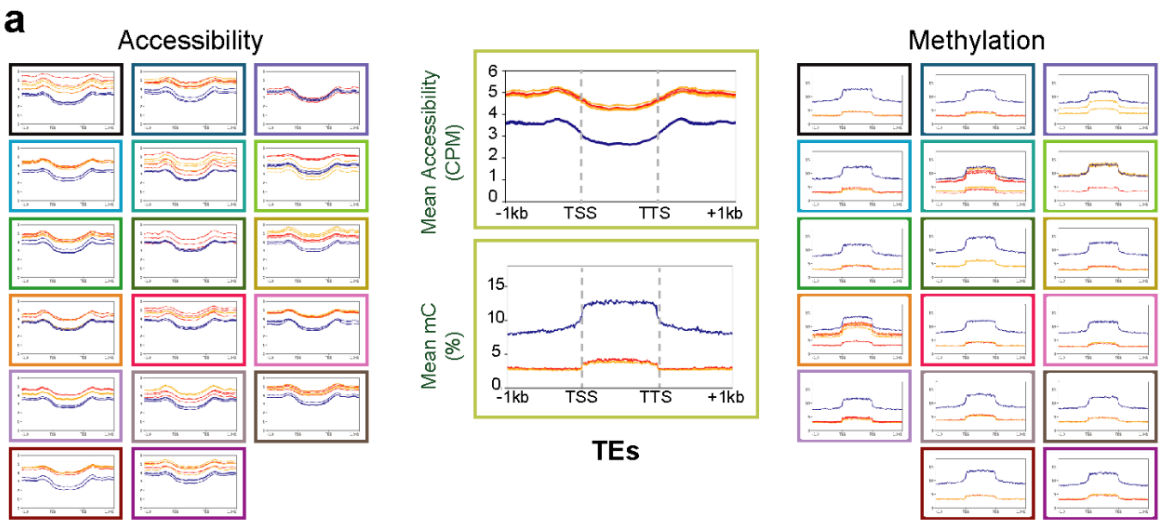
To evaluate if accessions had similar position-specific epigenetic profiles, we compared metaplots of mean chromatin accessibility and methylation in all contexts for each sample. The metaplots were generated by averaging over TEs (31,189) **(Figure 4a)** and protein-coding genes that were not associated with TEs (27,206)**(Figure 4b)** annotated in the TAIR10 reference genome in 10 bp non-overlapping bins and including 1 kb upstream and downstream sequences.

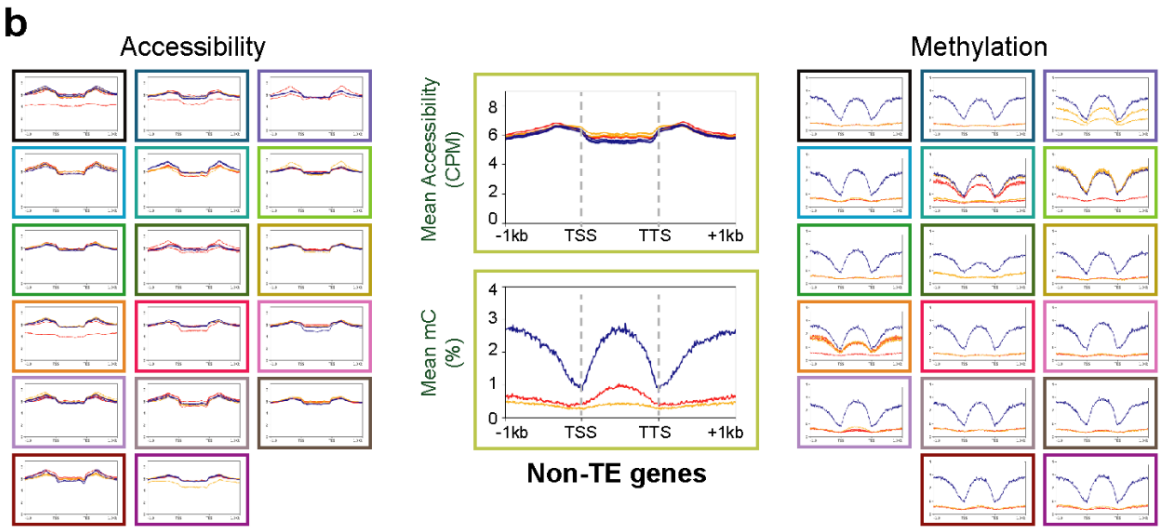
In wild-type plants of all accessions, methylation over the body of TE sequences was higher than in flanking regions, consistent with established knowledge of TEs being highly methylated [25,26]. In homozygous *met1* mutants, methylation was overall reduced, but low levels remained over both TEs and gene bodies **(Figure 4a , 4b, right panels)**. As expected, most heterozygotes were closer to wild-type plants than to homozygous mutants. In Col-0, heterozygotes were the least different from wild-type plants.

We observed that chromatin accessibility patterns in *met1* mutants mirrored cytosine methylation levels, where increase in accessibility was often accompanied by decrease in methylation. Mean accessibility in *met1* mutants from all accessions was profoundly increased across TEs **(Figure 4a, left panel)**, while differences in accessibility over all genes (excluding TE-associated genes) were small **(Figure 4b, left panel)**. This is consistent with the notion that TEs are primary targets of MET1 methyltransferase activity, with chromatin over all TEs being more sensitive to methylation loss than the bodies of all Non-TE protein-coding genes. Increased accessibility and reduced cytosine methylation in all *met1* mutants often co-occur at TEs, but not consistently in other protein-coding genes, as described by quantitative analyses in later sections.

A large fraction of TE genes become expressed in *met1* mutants

To investigate which *met1*-induced gene expression changes might cause the abnormal mutant phenotypes, we examined the transcriptomes of 104 mutant and 54 wild-type samples from all 18 accessions. A UMAP visualization of RNA-seq read counts across 21,657 genes revealed two distinct clusters representing mutant and wild-type plants **(Figure 5a)**. Similar to chromatin accessibility profiles, expression levels varied much more between accessions within the group of mutant samples than among wild-type samples. For two accessions, Tscha-1 and Tsu-0, we had also examined expression profiles of second-generation homozygous mutants, finding that there were no obvious changes relative to the first generation. Mutants of the Cvi-0 accession were notable, because they differed little in overall chromatin accessibility from their wild-type parents **(Figure 3a)**, but had very different expression profiles **(Figure 5a)**. This indicates that expression changes due to hypomethylation are not necessarily mediated by stark differences in chromatin accessibility.


a
Accessibility
Mean Accessibility (CPM)
Mean mC (%)
-1kb
TSS
TTS
+1kb
TEs
Methylation


b
Accessibility
Mean Accessibility (CPM)
Mean mC (%)
-1kb
TSS
TTS
+1kb
Non-TE genes
Methylation


Aa-0
Baa-1
Bl-1
Bs-1
Bu-0
Col-0
Com-1
Cvi-0
Ei-2
Est
Ler-1
MAR2-3
Nok-3
Pi-0
Ste-0
Tscha-1
Tsu-0
Uk-1
wild-type
met1 mutant Line 1
met1 mutant Line 2

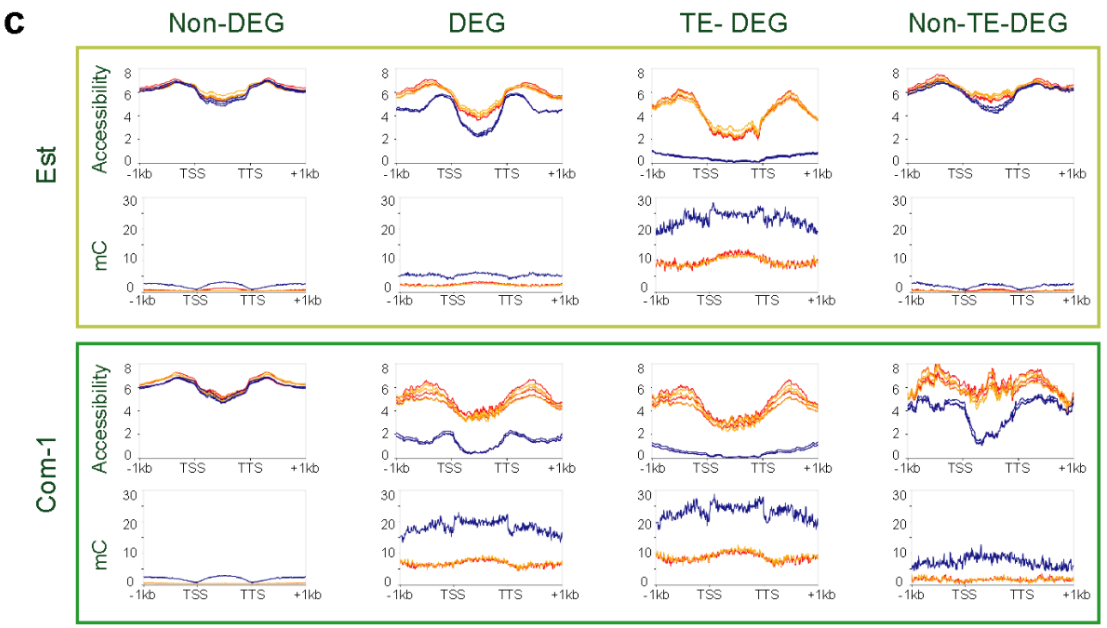

c
Non-DEG
DEG
TE- DEG
Non-TE-DEG
Est
Com-1
Accessibility
mC
-1kb
TSS
TTS
+1kb

Figure 4: *met1* mutants have more accessible chromatin and are hypomethylated over TEs and genes. Metaplots of mean chromatin accessibility across all **(a)** TAIR10 TEs and **(b)** Non-TE protein-coding genes in 18 accessions. Boxes are color-coded by accession, and data are colored by genotype. **(c)** Metaplots of accessibility and methylation for Est-1 and Com-1, across Non-DEGs, DEGs, TE-DEGs and Non-TE-DEGs. TSS and TTS denote transcription start site and transcription termination site respectively. Methylation levels are represented as % cytosine methylation (all-contexts) and accessibility levels are represented as TMM normalized values in counts per million (CPM).

From the 21,657 genes analysed, we were first interested in genes whose expression levels changed across all mutant and wild-type plants. We identified 3,479 ($p<=0.01$, $|\log_2$ fold change$|>=1$) differentially expressed genes (DEGs), when contrasting all mutants from the 18 accessions with all wild-type samples from the 18 accessions. Given that a main role of *MET1* is to prevent inappropriate expression of TEs[27], we first asked whether TE transcripts accumulated in our *met1* mutants. We therefore examined so-called TE-associated genes; these are genes encoding products required for TE mobility such as helicases, integrases, or transposases[28]. Approximately 42% (1,466) of all DEGs originated from TEs (TE-DEGs), and almost all of these (except 5 genes) were upregulated in mutants compared to wild-type plants **(Figure 5b)**. This observation is consistent with previous findings that many TE-sequences that are marked by methylation in wild-type plants can be activated upon hypomethylation in Col-0 *met1* mutants [25,29,30] and in their inbred lineages [27,31]. Among the 2,013 Non-TE-DEGs, only about two thirds were upregulated in the mutants, with 728 genes being downregulated. TE-DEGs were enriched in pericentromeric regions and were very highly expressed in mutants, often 1,000 fold higher than in wild-type plants **(Figure 5e)**. In contrast, Non-TE-DEGs were more evenly distributed over chromosome arms, and their expression changes were more moderate. Among the 2,013 Non-TE-DEGs, a series of GO terms was enriched, many relating to abiotic and biotic stress and stimulus response **(Supplemental Fig. 13)**.

Accession-specific DEGs are distributed across the genome and show large variation in gene expression changes

Although we were able to identify DEGs by contrasting all mutants with all wild-type samples, not all DEGs were uniformly induced or repressed in different accessions upon inactivation of *MET1*. An example of a random subset of 30 DEGs is shown in **Supplemental Figure S9d**.

Having noted that DEGs do not necessarily behave uniformly across accessions, we also contrasted *met1* mutants and wild-type samples for each of the 18 accessions. We first investigated the chromosomal distribution of such DEGs in each accession with accession-specific ideograms of all five chromosomes **(Supplementary Figure S10)**. In agreement with *MET1* silencing of heterochromatic TEs, pericentromeric and centromeric regions were consistently enriched for upregulated TE-DEGs in all accessions. Although the TE superfamilies in closest proximity to TE-DEGs were similarly distributed for *met1* mutants of all accessions, the superfamily that was most likely to be activated in *met1* mutants was the Class II DNA transposon family En/Spm **(Supplementary Figure S9e).** CACTA elements of the En/Spm superfamily are transcribed in Col-0 *met1* mutants[32,33], although transposition events are not detectable[32].

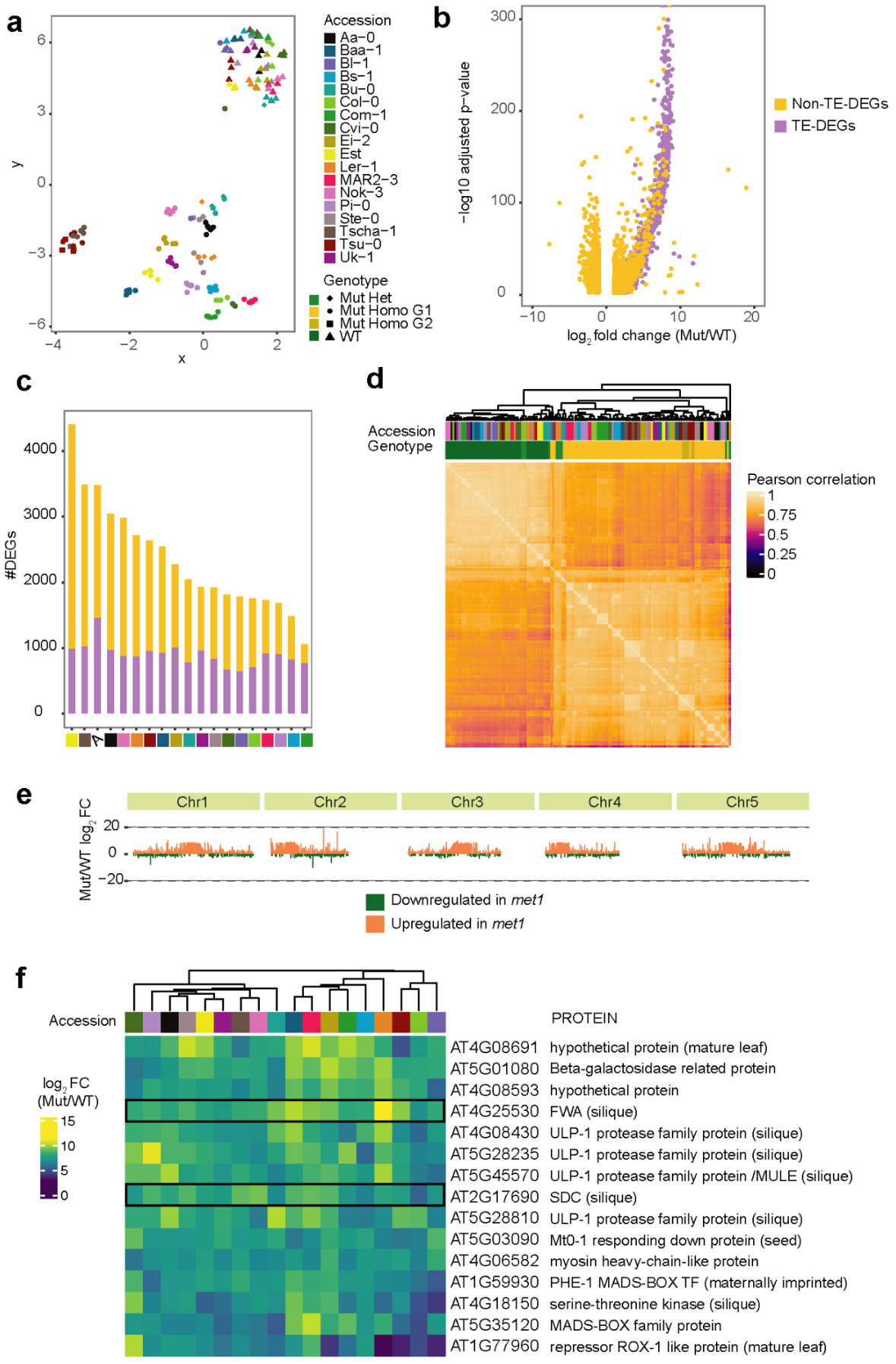

a
Accession
Aa−0
Baa−1
Bl−1
Bs−1
Bu−0
Col−0
Com−1
Cvi−0
Ei−2
Est
Ler−1
MAR2−3
Nok−3
Pi−0
Ste−0
Tscha−1
Tsu−0
Uk−1
Genotype
Mut Het
Mut Homo G1
Mut Homo G2
WT
x
y
b
−log10 adjusted p−value
log2 fold change (Mut/WT)
Non-TE-DEGs
TE-DEGs
c
#DEGs
d
Accession
Genotype
Pearson correlation
1
0.75
0.5
0.25
0
e
Chr1
Chr2
Chr3
Chr4
Chr5
Mut/WT log2 FC
Downregulated in met1
Upregulated in met1
f
Accession
log2 FC (Mut/WT)
PROTEIN
AT4G08691 hypothetical protein (mature leaf)
AT5G01080 Beta-galactosidase related protein
AT4G08593 hypothetical protein
AT4G25530 FWA (silique)
AT4G08430 ULP-1 protease family protein (silique)
AT5G28235 ULP-1 protease family protein (silique)
AT5G45570 ULP-1 protease family protein /MULE (silique)
AT2G17690 SDC (silique)
AT5G28810 ULP-1 protease family protein (silique)
AT5G03090 Mt0-1 responding down protein (seed)
AT4G06582 myosin heavy-chain-like protein
AT1G59930 PHE-1 MADS-BOX TF (maternally imprinted)
AT4G18150 serine-threonine kinase (silique)
AT5G35120 MADS-BOX family protein
AT1G77960 repressor ROX-1 like protein (mature leaf)

Figure 5: Accession-specific DEGs in *met1* mutants. (a) UMAP visualization of transformed RNA-seq counts from 158 samples (104 *met1* mutants and 54 wild-type plants) across 21,657 genes. Colors indicate accessions, and shapes indicate genotype. **(b)** Volcano plot of 3,479 DEGs identified in a contrast between all mutant samples (104) and all wild-type samples (54). DEGs were defined as genes which showed $\log_2$ fold change >1 and p.adj <=0.01. Purple colored dots represent DEGs that are TE-associated (TE-DEGs), and yellow colored dots are Non-TE-associated DEGs (Non-TE-DEGs). **(c)** DEGs identified in 18 accession-specific contrasts, and an all-mutants-against-all-wild-type contrast (denoted by 'A). **(d)** Correlation heatmap of 7,132 consensus DEGs from all accessions. **(e)** Karyogram indicating DEGs (all-mutants-against-all-wild-type contrast) along five *A. thaliana* chromosomes, and their $\log_2$ fold change (Mut/WT). Upregulated DEGs are colored orange and downregulated DEGs green. **(f)** Heatmap of $\log_2$ fold changes (Mut/WT) in 15 universal Non-TE-DEGs across all contrasts; the *FWA* and *SDC* genes are highlighted with black outlines. TAIR10 gene names, encoded proteins and the tissue they are known to be preferentially expressed in wild type are denoted for every row. Genotypes represented are wild-types ('WT'), heterozygous *met1* mutants ('Mut Het'), first generation homozygous *met1* mutants ('Mut Homo G1') and second generation homozygous *met1* mutants ('Mut Homo G2').

Non-TE-DEGs, which were similarly likely to be up- or down-regulated **(Supplementary Figure S11b)**, were distributed throughout the chromosome arms. In contrast to TE-DEGs, the number of Non-TE-DEGs varied substantially between accessions **(Figure 5c)**, which in turn explained why some accessions had many more downregulated DEGs than others **(Supplementary Figure S11a).** The number of Non-TE-DEGs was as low as 278 in Com-1, and as high as 3,409 in Est-1. Est-1 *met1* mutants appeared to be particularly sensitive to hypomethylation in Non-TE genes, which accounted for 77% of DEGs in this accession. The converse was true for Com-1, where only 26% of DEGs were Non-TE genes. That there was much more variation in the number of Non-TE-DEGs across accessions than in the number of TE-DEGs potentially points to different mechanisms that cause changes in TE- and Non-TE genes in *met1* mutants.

To begin to investigate what was behind the large variation in number of Non-TE-DEGs in the different accessions, we first looked at the number of shared Non-TE-DEGs between different groups of accessions **(Supplementary Figure S11d)**. We found that Est-1 had the largest number of Non-TE-DEGs, 983, not found in any other contrast. Com-1, which had the fewest Non-TE-DEGs among all contrasts, also had the fewest genes, six, not observed in any other contrast. In addition to these six accession-specific genes, there were 272 Non-TE-DEGs in Com-1 that were found in at least one other contrast.

Next, we identified DEGs common to all 19 comparisons. i.e., found in all 18 accession-specific contrasts and the contrast between all mutants and all wild-type plants, to study genes that were universally altered in expression among *met1* mutants of all accessions. We found 291 universal DEGs, all but 15 (5%) of which were TE-DEGs. There was a wide range in expression profiles between accessions, suggesting that similar genes between accessions can be highly variable in the absence of functional MET1 **(Supplementary Figure S9b)**. For example, Ler-1 was among the accessions with the highest expression levels of universal DEGs in *met1* mutants (up to 5,500 fold), while universal DEGs showed the least change in Col-0 *met1* mutants (up to 1,075 fold).

Among the remaining 15 universal Non-TE-DEGs **(Figure 5f)**, three genes are known to be strongly expressed in siliques and to be maternally imprinted (*FWA, SDC, AT1G59930*)[34,35], in

addition to five other genes that are also strongly expressed in wild-type siliques. Ectopic expression of *FWA* associated with hypomethylated tandem repeats overlapping the transcription start causes late flowering in *met1* and *fwa-1* mutants[36,37]. Indeed, all our *met1* mutants were late flowering compared to their wild-type parents, likely due to the effect of the hypomethylated *FWA* epiallele. *SDC* is another locus that becomes activated upon hypomethylation of short tandem repeats in its promoter, in both *met1* mutants and, even more strongly, in *drm1 drm2 cmt3* triple mutants. When *SDC* is activated, plants exhibit dwarfism and leaf curling **(Supplementary Figure S12)**[38,35,38].

To ask whether the large expression differences at DEGs could arise from epigenetic variation, we first re-generated metaplots of chromatin accessibility and DNA methylation in two accessions (Est-1 and Com-1) to separately examine Non-DEGs (genes that were not differentially expressed), DEGs, TE-DEGs and Non-TE-DEGs **(Figure 4c)**. In both accessions, we found that DEGs were marked by increased chromatin accessibility and reduced DNA methylation in *met1* mutants. Yet, the accession-specific differences for both *met1* mutants and wild-type plants were highest at Non-TE-DEGs, especially in chromatin accessibility. This variation in gene regulation between accessions is striking, and likely reflects variation in the number of genes compared (278 Non-TE-DEGs in Com-1 and 3409 Non-TE-DEGs in Est). These observations provide further evidence to support that gene regulatory patterns across accessions may be more variable at Non-TE genes.

Non-TE-DEGs are marked by varied methylation levels in *cis* and reflect parental epigenetic state

Having observed that differential gene expression could be accompanied by epigenetic changes, we next examined DEGs across all accessions to quantitatively assess their DNA methylation and gene expression levels. We generated a consensus set of 7,132 DEGs from all 18 accessions (see Methods, **Figure 5d**), which we classified as TE-DEGs (1,401) and Non-TE-DEGs (5,731), and identified the nearest CG-DMR to each DEG **(Supplementary Figure S14)**. We chose to focus on CG-DMRs, and therefore CG-methylation, to simplify our analyses, also because these regions suffered the most direct consequences of MET1 inactivation.

Only a small fraction of DEGs (21% of TE-DEGs and 7% of Non-TE-DEGs) contained CG-DMRs in close proximity (either overlapping gene body or within 1.5 kb of the gene). Conversely, most CG-DMRs were distally positioned to DEGs, with only 28% and 21% of CG-DMRs found in proximity to Non-TE-DEGs and TE-DEGs. respectively. For DEGs with proximal CG-DMRs, we plotted methylation differences relative to wild-type against gene expression changes for *met1* mutants across all 18 accessions.

Since TEs are generally known to be silenced in wild-type plants, we anticipated that the loss of CG methylation would have similar effects in activating their expression. Indeed, most TE-DEGs in *met1* mutants were marked by substantial hypomethylation (difference in methylation < -50%), both within their gene bodies **(Supplementary Figure 19e)** and in *cis* **(Supplementary Figure S20e).** This methylation loss was accompanied by high increases in gene expression, although not to the same extent in every accession.

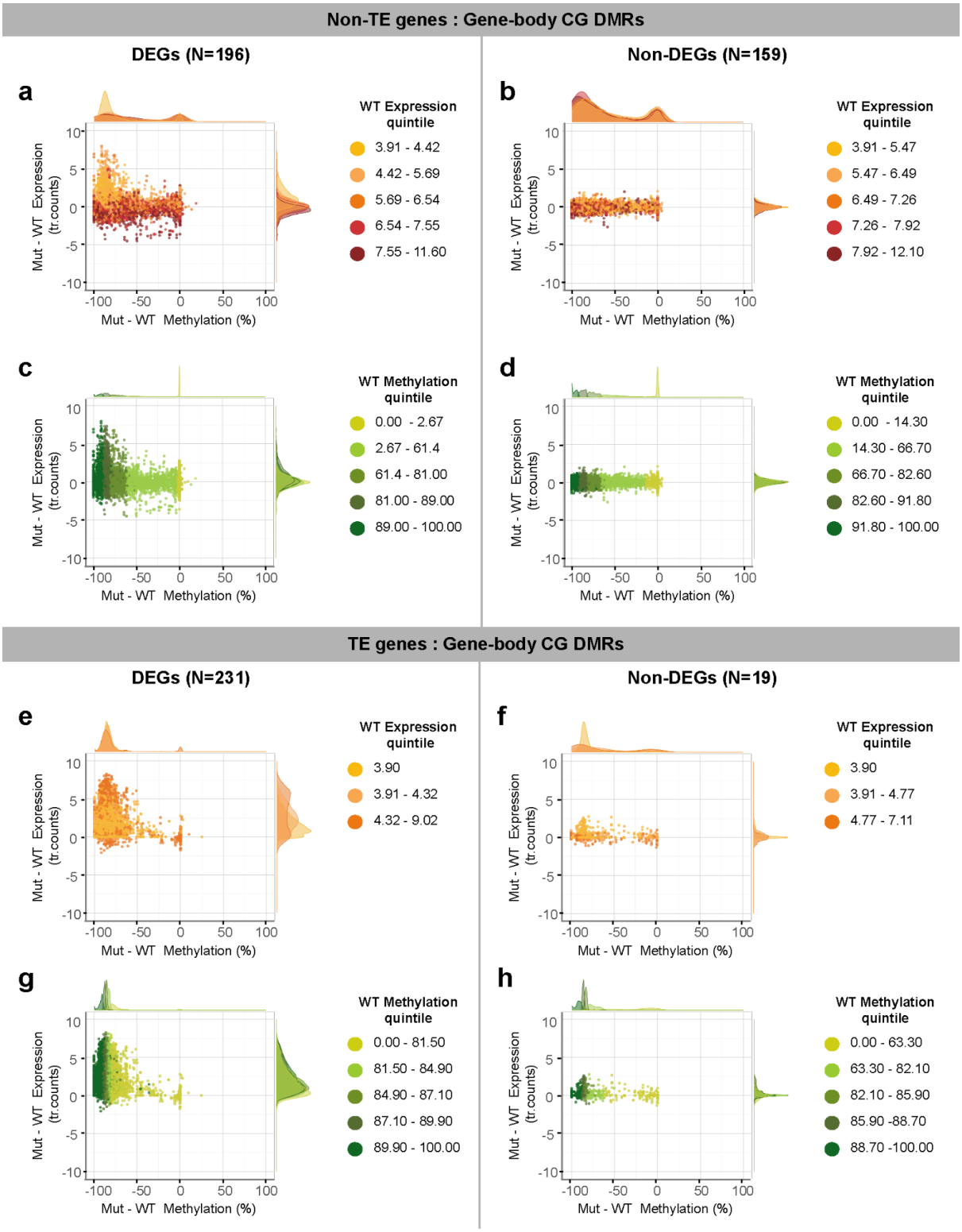


Fig 6: CG-DMRs in gene-bodies of Non-TE-DEGs (a-d) and TE-DEGs (e-h). Scatter plots showing differences in CG methylation between *met1* mutants and wildtype plants against differences in gene expression. Dots are colored by wild-type expression quintiles **(a,b,e,f)** and wild-type methylation quintiles **(c,d,g,h)** with x- and y-axis density distributions of each expression/methylation quintile. Expression levels are represented as transformed read counts and methylation levels are represented as % CG methylation in CG-DMRs.

Non-TE-DEGs, which varied greatly in number and expression levels across accessions, also exhibited a wide gradient of methylation differences relative to wild-type (ranging from -100% to 10%) both in gene bodies **(Supplementary Figure 19a)** and in *cis* **(Supplementary Figure S20a).** A subset of genes that were highly upregulated among Non-TE-DEGs also exhibited highly reduced CG methylation (differences in methylation < -75%) in *met1* mutants. However, there were also many genes that showed negligible methylation differences in *met1* mutants.

Knowing that wild-type accessions can show substantial variation in their methylomes[14], we suspected that genes with minimal methylation differences were likely to exist in a hypomethylated state in their wild-type parental lines. Therefore, we asked whether parental methylation and expression state can determine the epigenetic and transcriptional plasticity of genes in *met1* mutants. We observed that wild-type methylation states of CG-DMRs near DEGs were highly correlated with methylation differences in *met1* mutants (up to R_{sq}=1 for Non-TE-DEGs). This suggests that genes flanked by heavy methylation in wild-type plants are more likely to show the largest reductions in methylation levels in the absence of MET1 **(Figure 6c)**. In contrast, genes that were poorly expressed in wild-type plants were more prone to being highly upregulated in *met1* mutants **(Figure 6a)**. For all of our analyses with DEGs, we also carried out similar analyses with randomly sampled Non-DEGs (or genes that were not differentially expressed) as a control **(Figure 6b, 6d, 6f, 6h)**. We found that the correlation between wild-type methylation state and methylation differences in the *met1* mutants remained **(Figure 6d, 6h)**, suggesting that this effect was independent of gene expression changes.

In wild accessions, many constitutively active genes are marked by gene-body CG methylation (gbM)[39] and the methylation levels of these genes are known to vary in tandem with their differential expression across *A. thaliana* accessions[19,20] We also examined gbM-like genes among Non-TE-DEGs (Methods) in our dataset to find that they were mostly downregulated in *met1* mutants, but were not shared across all accessions **(Supplementary Figure S21).** Incidentally, genes that were highly methylated and lowly expressed in wild-types (exhibiting TE methylation characteristics in the CG context, or 'CG teM-like' genes) were always upregulated in *met1* mutants and showed high similarities with TE-DEGs **(Supplementary Figure S22)**. Surprisingly, there were no genes common to both gbM-like and CG teM-like gene groups, suggesting different epigenetic phenomena underlying their regulation.

Non-TE-DEGs can exhibit multiple chromatin accessibility states

There were many more ACRs than DMRs in a consensus set compared across all of our mutant and wild-type samples. When we looked for positional overlaps of ACRs relative to DEGs, we found that a majority of DEGs, over 90%, had ACRs in *cis* and/or within their gene body (which we collectively refer to as '*cis* ACRs').

TE-DEGs, which were almost always upregulated in *met1* mutants, were marked by a roughly proportional increase in chromatin accessibility overall **(Figure 7e, 7g)**, as we had expected from earlier results **(Figure 4c)**. However, this pattern was not as straightforward for Non-TE-DEGs, where an increase in accessibility could be associated with, but did not necessitate an increase in gene expression. Similarly, some down-regulated genes showed an increase in accessibility, and others a decrease. Although a common finding in many previous

studies has been that accessible chromatin can favour gene transcription by facilitating transcription-factor binding[40,41], there has also been evidence that inaccessible regions can occur in some long, highly transcribed genes[42].

Non-TE-DEGs with low expression levels and low *cis* accessibility in wild type were more prone to increased expression and accessibility in *met1* mutants, compared to other genes **(Figure 7c, 7g)**. Many genes with highly reduced accessibility in *met1* mutants were also highly accessible in the wild-type state. However, it appeared that hypomethylation in *met1* mutants did not consistently affect accessibility in all other genes. Together, these observations indicate that perturbation of the methylation state can destabilize chromatin architecture, but chromatin accessibility changes need not always be accompanied by proportional changes in gene expression.

MET1 can have direct and indirect effects on transcription at Non-TE genes

Since Non-TE-DEGs appeared to be more variable across accessions in their differential gene expression **(Figure 8b)**, differential chromatin accessibility **(Figure 6a,6c)** and differential CG methylation **(Figure 7a,7c)** in *met1* mutants, we focused on these genes to better understand epigenetic plasticity in our *met1* mutants. Among a total of 5,731 consensus Non-TE-DEGs, a majority (87%) carried accessible chromatin regions in *cis*, but lacked a *cis* CG-DMR ('*cis*' here including the gene-body).

Some DEGs (361) exhibited neither an ACR nor a CG-DMR in their proximity**,** suggesting that the loss of CG-methylation may have triggered other epigenetic mechanisms to indirectly affect expression of these genes. Only 27 DEGs exhibited *cis* CG-DMRs without ACRs in their proximity. We examined a few genes from each of these categories **(Figure 8a)** and found that the relationship between gene expression changes, CG methylation changes and accessibility changes could substantially vary across *met1* mutants in different accessions.

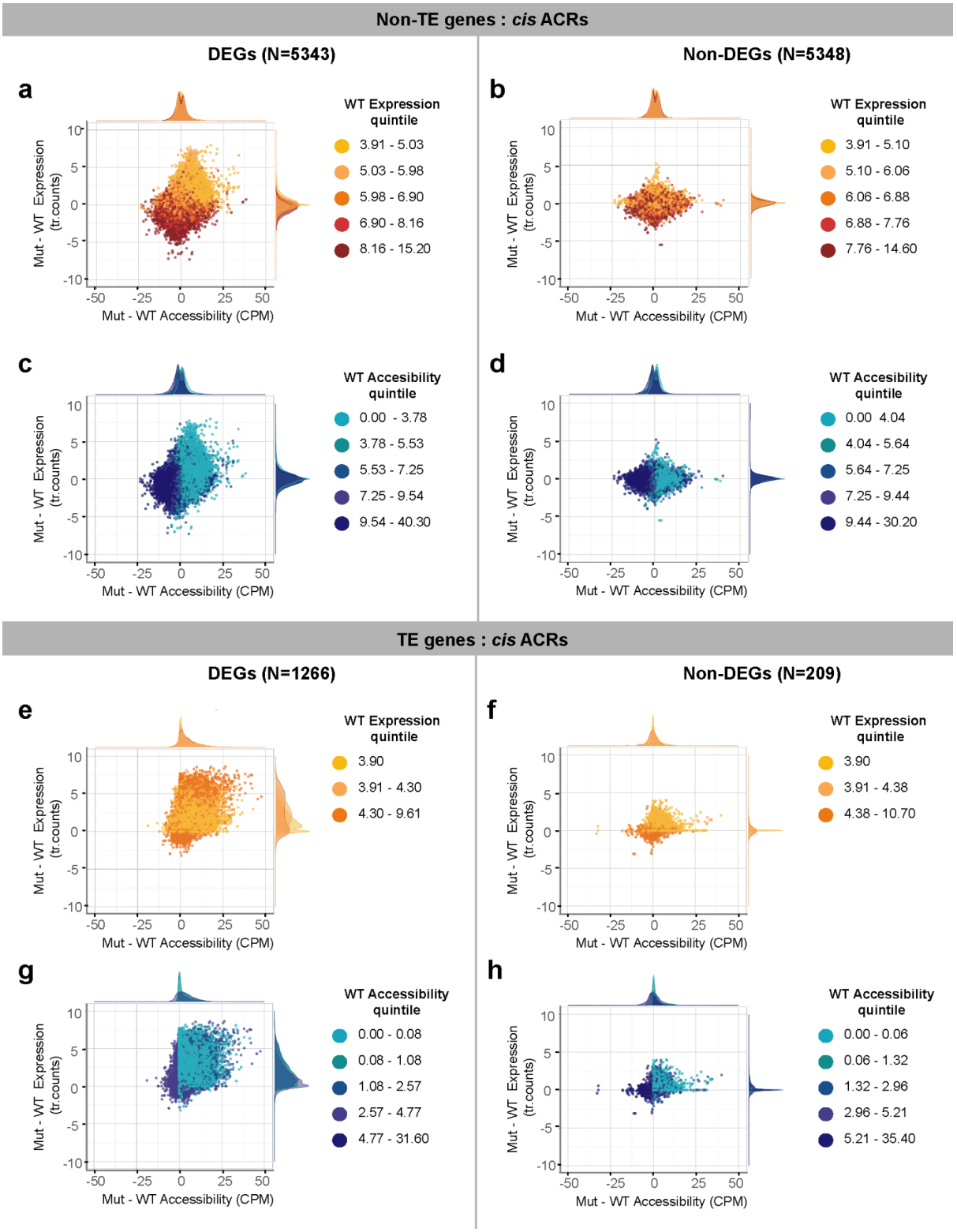


Fig 7: ACRs in *cis* to TE-DEGs (a-d) and Non-TE-DEGs (e-h). Scatter plots showing differences in chromatin accessibility between *met1* mutants and wild-type plants against differences in gene expression. Dots in the scatter plot are colored by wild-type expression quintiles **(a,b,e,f)** and wild-type accessibility quintiles **(c,d,g,h)** with x- and y-axis density distributions of each expression/accessibility quintile. Expression levels are represented as transformed read counts and accessibility levels are represented as TMM normalized values in counts per million (CPM).

We were most interested in Non-TE genes that carried both *cis* ACRs and CG-DMRs (392 DEGs and 951 Non-DEGs). In general, we found that genes with a high CG methylation reduction in *met1* mutants (ΔMethylation < -70%) had a tendency to show large increases or decreases in chromatin accessibility. Yet, for both DEGs **(Figure 8d)** and Non-DEGs **(Figure 8e)**, almost all genes were distributed in several combinations of methylation and accessibility states, and did not cluster by degree of expression change. This suggested not only the presence of multiple epigenetic states for different genes, but also the possibility of similar genes with different epigenetic states across accessions. However, one group of genes among DEGs stood out from the rest: this group was highly expressed in *met1* mutants, marked by high methylation reduction and also by highly increased accessibility **(Figure 8d)**.

We next examined all Non-TE genes together (1,343 genes comprising 392 DEGs and 951 Non-DEGs) for accession-specific epigenetic patterns. We first compared the methylation levels of these genes in the wild-type and mutant genotypes, and found that the accession Cvi-0 exhibited the highest fraction of genes that showed minimal reduction in methylation in *met1* mutants **(Supplementary Figure S25)**. This was explained by Cvi-0 having a higher fraction of lowly methylated genes in wild-type, compared to other accessions **(Supplementary Figure S25)**. However, both genes with low methylation changes **(Figure 8f)** and genes with high methylation changes **(Figure 8g)** in Cvi-0 changed in accessibility and expression, although not as largely as in other accessions. This suggests that genome-wide hypomethylation may alter epigenetic stability indirectly even in genes that do not change their methylation levels.

met1 mutants exhibit phenotypes characteristic of known epialleles

Among our *met1* mutants, we found well-known epialleles at loci as *FWA*[36] **(Supplementary Figure S25)** , *SDC*[38] **(Supplementary Figure S12)** , the *PAI* genes[43], *IBM1*[44], *SNC1* (similar to the *bal* variant)[45], *AG*[4] and *SUP* (similar to the *clark-kent* variant [46] **(Supplementary Figure S26)**, some of which have been observed in *met1* mutants of the accessions Col-0 and C24[3,47]. We observed different methylation patterns at these loci depending on the accession of origin, with considerable variation in chromatin accessibility and gene expression **(Supplementary Figure S25)** across accessions. As seen before[7], some epialleles arise in second-generation *met1* mutants **(Supplementary Figure S3)**, suggesting that epigenetic instability due to absence of MET1 continues during inbreeding.

Another notable example of a Non-TE-DEG includes the *ROS1* demethylase locus (*AT2G36490*)[48]. In accordance with its function as a methylation sensor[7],[49,50], *ROS1* was consistently downregulated in all *met1* mutants, with variation in the reduction of chromatin accessibility and gene expression across accessions and Cvi-0 showing the smallest changes **(Figure 8c)**. While methylation in all-contexts at the 'methylstat' sequence[49,50] in the *ROS1* promoter is absent in our *met1* mutants, mutants gain methylation further downstream in the gene body **(Figure 8c)**. These results show that genome stability may require tight epigenetic and epigenomic control even at master regulatory genes.

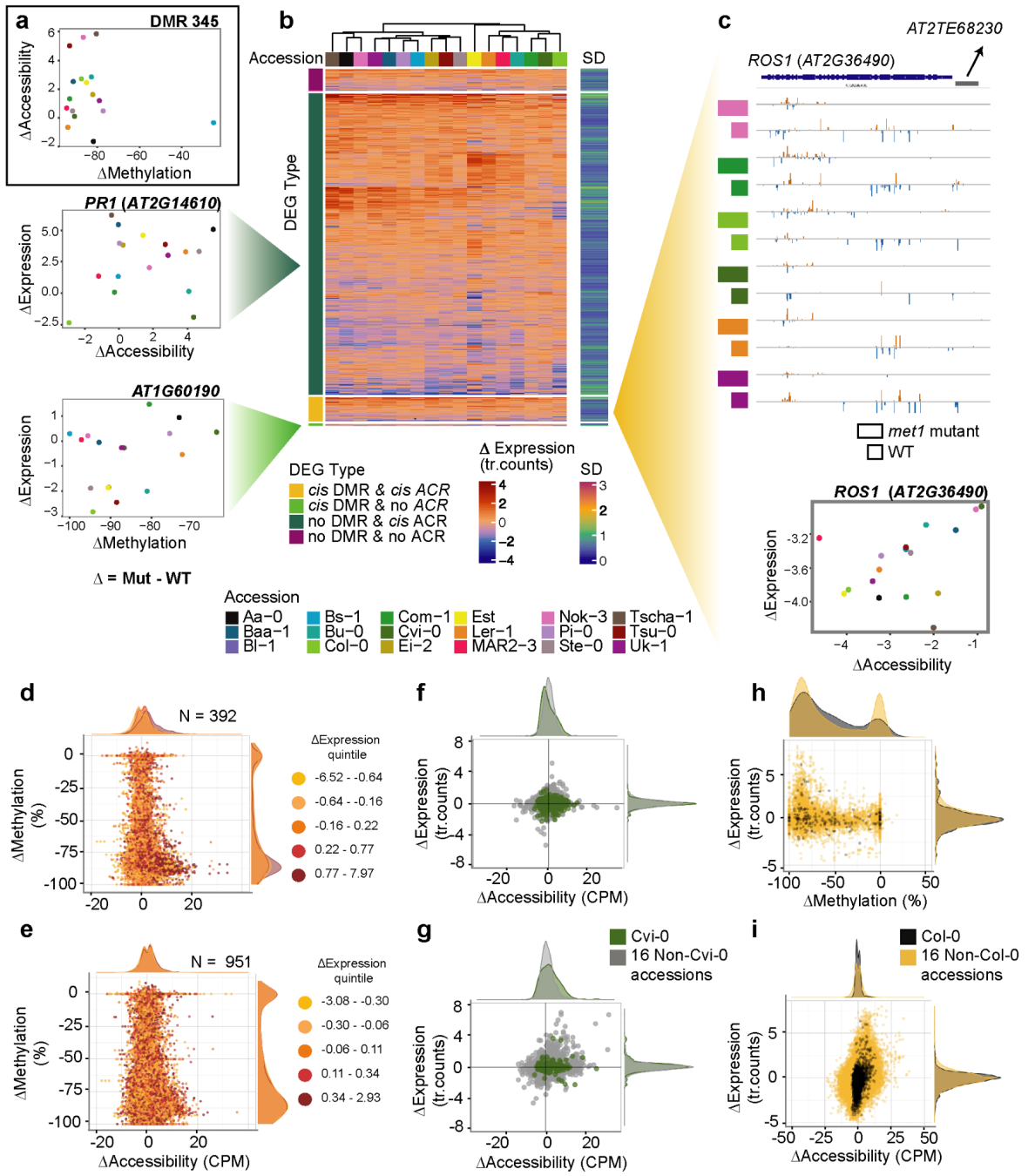


Fig 8: Non-TE-DEGs in *met1* mutants can exist in multiple epigenetic states across accessions. (a) Scatter plots showing the relationship between changes in methylation and accessibility at a representative DMR, DMR 345. **(b)** Heatmap of expression changes across 5,731 Non-TE-DEGs in 17 accessions, with an adjacent heatmap showing variance expressed as standard deviation (SD) across accessions, and scatterplots of changes in expression and accessibility in representative genes from two different DEG categories (based on overlap with *cis* CG-DMRs and ACRs). **(c)** Genome-browser view of methylated cytosines in all contexts and scatterplot of changes in expression and accessibility at *ROS1* locus. **(d)** Scatterplot of changes in accessibility and methylation in Non-TE-DEGs across 17 accessions. Colors and density distributions represent quintiles of expression changes. **(e)** Scatterplots similar to **(d)** for Non-TE genes. **(f)** Scatterplot of changes in accessibility and expression of lowly methylated genes in Cvi-0 (dark green) compared across the same genes in 16 other accessions (dark grey). **(g)** Scatterplot similar to **(f)** for highly methylated genes in Cvi-0. **(h)** Scatterplot of changes in methylation and expression in Non-TE-DEGs carrying gene-body CG-DMRs, colored by DEGs specific to Col-0 (black) against DEGs in other accessions (yellow). **(i)** Scatterplot of changes in accessibility and expression in Non-TE-DEGs carrying *cis* ACRs, colored by DEGs specific to

Col-0 (black) against DEGs in other accessions (yellow). Expression levels are represented as transformed read counts; accessibility levels as TMM values in counts per million (CPM), and methylation levels as % CG methylation.

DISCUSSION

Studying natural variation can be insightful for understanding the extent of genetic and epigenetic flexibility required for population-specific adaptation, especially in plants. In *A. thaliana*, there is plenty of evidence to demonstrate that the genome, methylome and transcriptome can largely vary even at the intra-specific level [13,14]. Furthermore, the intersection of genetics and epigenetics in natural accessions is best studied at TEs, whose silencing is determined by enzymes involved in DNA methylation[15,18]. Yet, DNA methylation marks are also prevalent throughout euchromatin, especially in proximity of protein-coding genes that are not associated with TEs. In our study, we show how the absence of methylation marks can simultaneously disturb the regulatory balance of thousands of such genes, and that it does so to remarkably different extent in accessions from geographically diverse origins. We generated *met1* mutants in 18 accessions of *A. thaliana* and found a wide range of phenotypic defects arising from highly reduced CG methylation, accompanied by substantial changes in both chromatin architecture and the transcriptome.

The absence of MET1 in *A. thaliana* has long been known to affect the chromatin landscape apart as an apparent consequence of its effects on methylation [7,9,10]. ATAC-seq analyses in *met1* mutants have also highlighted the prevalence of increased chromatin accessibility at several hypomethylated loci, which in turn can facilitate long-range chromatin interactions[12]. Previous studies, however, have been largely restricted to the reference accession Col-0, which only provides a singular view of the epigenetic landscape. In comparing two genotypes (wild-type plants and homozygous *met1* mutants) across 17 accessions, we show that CG methylation, gene expression and chromatin accessibility can widely vary beyond Col-0 levels, especially at genes that are not associated with TEs **(Figure 8h, 8i)**.

Knowing that methylation changes are likely to activate the expression of TE-genes[33], but may not consistently affect the expression of Non-TE genes, we separated our analyses of DEGs into two groups - TE-DEGs and Non-TE-DEGs. While the mobilization of TEs is known to disrupt genes and cause phenotypic defects in Col-0 *met1* mutants [27], this is often observed only upon inbreeding, since activation of TEs does not necessarily guarantee their mobilization [32]. Analyses of our *met1* mutants were largely restricted to first generation *met1* homozygotes - and while many TE-DEGs are upregulated in these mutants, the differential epigenetic regulation of highly variable Non-TE-DEGs across accessions may be more causal for the plethora of observed phenotypic defects.

When analysing a consensus set of DEGs from all accessions together, we always find that the consequences of MET1 inactivation greatly depend on the initial epigenetic state of the wild-type parent. Lowly expressed, inaccessible and highly CG methylated genes in wild-type plants show a tendency to have high expression, high accessibility and CG hypomethylation in their *met1* mutant derivatives. This pattern is observed for most TE-DEGs, but only for a small subset of Non-TE-DEGs.

Non-TE-DEGs show more variability in their methylation and accessibility changes in *met1* mutants, which is associated with more variation in gene expression changes, in either direction, i.e., both up- and down-regulation in *met1* mutants. Regions of increased accessibility in Col-0 *met1* mutants have been previously associated with multiple gene groups, classified by their expression changes[12]. In our study we find that this association can be multi-layered, varying by gene, genetic background, and the epigenetic background. For example, the study by Zhong and colleagues[12] demonstrated local changes in methylation were sufficient to alter chromatin accessibility at the *FWA* epiallele. Upon examining the same *FWA* locus in our homozygous *met1* mutants of 17 accessions, we find that the increases in accessibility can highly vary even when all *met1* mutants are similarly hypomethylated, although parental methylation levels differ **(Supplementary Figure S25)**. Additionally, the relationship between accessibility changes and expression changes is non-linear for this gene **(Supplementary Figure S25)**, suggesting that more complex processes may be involved in its regulation.

The *FWA* gene is only one example among a total of 11,675 unique DEGs in all of our *met1* mutants across 18 accessions (compared to 1,759 DEGs in Col-0 alone). That only 291 DEGs among these are universal across all accessions, shows the differential sensitivity of each genetic background to methylation-triggered gene regulation. DEGs include developmental genes, epigenetic regulatory genes such as *ROS1*, *IBM1* and *SUVH3*, many of which are well-known epialleles. Although we sampled rosette leaves for our analyses, many DEGs are natively expressed in other tissues of wild-type plants. Additionally, we find that both methylation and accessibility changes can also occur near genes that are not differentially expressed in rosette leaf tissue. It is possible that these epigenetic states may be poised for altering gene expression in other tissues[51], or during inbreeding.

A wide-range of silique abnormalities and distorted segregation ratios in our *met1* mutants indicates that the absence of MET1 function may be detrimental for gametogenesis, fertilization or post-zygotic development. Among the accessions used in our study, Ler-1 and Tsu-0 carry a microRNA haplotype which is known to impair silencing of TEs in male gametes (Ler-0 *MIR845* haplotype[52]). Furthermore, epigenetic variation across the accessions Col-0, Ler-1 and Cvi-0 can impact seed development due to differential imprinting[53,54]. For example, the comparatively higher fertility of our *met1* mutants in Cvi-0 could possibly be due to natural hypomethylation of the *HDG3* (*AT2G32370*) locus in wild-type Cvi-0 plants **(Supplementary Figure S26)**, thereby not causing dramatic effects in seed size as observed in mutants of other accessions. Future studies on the epigenetic and epigenomic landscape of gametophytic, embryonic and endosperm tissue of our *met1* mutants will shed light on how hypomethylation in somatic cells can impact fertilization processes.

The high number of altered chromatin accessibility regions identified in our *met1* mutants is in agreement with observations in Col-0 *met1*, *ddm-1* and *fwa-1* mutants, in which accessibility differences appear upon CG hypomethylation [12]. Accessible chromatin regions in several angiosperms are known to harbor conserved regulatory elements and transcription factor binding sites [40,41,55–58], and they can also be marked by reduced cytosine methylation[51,56,59–61]. The epigenetic profiles of our *met1* mutant collection can help characterize the potential plasticity of transcription factor binding[62] to various genomic regions in natural accessions.

Structural variation in the genome, especially over TEs and intergenic regions, has been well-studied in *A. thaliana* accessions [16,57,63] and can be important for preventing a 'reference bias' when evaluating chromatin accessibility differences[57,63]. It will be interesting to further investigate whether newly mobilized TEs in our *met1* mutants or associated protein-coding genes can 'carry' similar epigenetic states in all insertion sites[64], and if the excision sites retain residual epigenetic states from these elements. The availability of high-quality genomes at the accession-specific level, together with a large set of epigenetic mutants in the germplasm collection would greatly improve our understanding of epigenetic and epigenomic regulatory evolution.

METHODS

CRISPR/Cas9 knockout of *MET1* in 18 *A. thaliana* accessions

Using a plant molecular cloning toolbox[21], a supermodule destination binary vector carrying a plant-codon optimized *Cas9* driven by a *UBQ10* promoter was cloned with a single guide-RNA (gRNA) targeting the *A. thaliana MET1* (*AT5G49160*) gene. The gRNA was designed using the CRISPR design tool in *Benchling* (www.benchling.com) targeting a 20 bp region in exon 7 of *MET1* (**Supplementary table S1**), which is the same exon where previously described *met1-3* mutants are known to harbour a T-DNA insertion[6]. This exon is present in the catalytic domain of the protein and harbors a motif that is a binding site for cytosine nucleotide substrates [65]. Eighteen early-flowering *A. thaliana* accessions were transformed with the above CRISPR construct by *Agrobacterium*-mediated floral dipping[66], carried out twice with a 7-10 day interval. Seeds of primary transformants (T_1) were screened for the presence of the transgene by selecting for the mCherry fluorescence marker, and sown on soil. These T_1 plants were subjected to heat treatment cycles for enhancing Cas9 activity[67]. Genotyped lines carrying a mutation in the gRNA-target region were propagated to the T_2 generation after segregating the transgene (by selecting for non-mCherry seeds), followed by identification of lines carrying heritable heterozygous mutations. One to two heterozygous T_2 lines per accession were further propagated for another generation to identify homozygous plants in the segregating progeny.

Genotyping transformants and identification of homozygous mutants

First- and second-generation homozygous mutants were genotyped either using Sanger sequencing of a 649 bp PCR-amplicon, or by amplicon-sequencing of a 152 bp PCR-amplicon **(Supplementary Table S3)**, both covering the CRISPR guide-RNA target region. Most of the mutations identified in all mutants were single bp insertions or deletions that occurred within the first four bp from the 5' end of the target region, and disrupted the open reading frame due to frameshifting. Candidate homozygous mutants in segregating T_3 populations were first identified by visual phenotyping, followed by genotyping.

Plant growth conditions and tissue collection for large-scale sequencing

Seeds were sterilised by treatment with chlorine gas for 4 hours, followed by stratification in the dark at 4°C for 4 days in 0.1% agar. All plants were grown in controlled growth chambers at 23 °C, long day conditions (16 h light/8 h dark) with 65% relative humidity under 110 to 140 µmol m^{-2} s^{-1} light provided by Philips GreenPower TLED modules (Philips Lighting GmbH, Hamburg, Germany) with a mixture of 2:1 DR/W LB (deep red/white mixture with ca. 15% blue light) and W HB (white with ca. 25% blue light), respectively, and watered at 2-day intervals.

Since homozygous mutants from several accessions had reduced fertility and did not set sufficient seeds for further propagation, sampling for all sequencing experiments was carried out in the first homozygous generation. Segregating populations of T3 mutant plants were grown from 1-2 lines per accession, and homozygous individuals were marked by their distinct phenotype (as identified in previous growth experiments) and later confirmed by Sanger

sequencing of DNA used for BS-seq libraries. Wild-type plants from the same accessions were grown in parallel to all mutant lines.

At 25 days after germination, three homozygous individuals per T2 line per accession were collected as separate biological replicates, along with three wild-type individuals. Sampling involved the collection of three sets of rosette leaves from the same individual plant. One set of leaves was immediately frozen in liquid nitrogen containers (and subsequently at -80°C) for Bisulfite-sequencing analysis, and the second set similarly for RNA-sequencing. The third set of leaves (for ATAC-sequencing analysis) were subjected to syringe-infiltration with 0.1% formaldehyde in phosphate-buffer saline, followed by 0.125 M glycine in phosphate-buffer saline, washed with autoclaved water and dried before storage at -80°C. All tissue sampling and fixation was carried out within a 30 minute time window.

Bisulfite-seq library prep

Frozen leaf tissue from three biological replicates were mixed and grinded together. This powder was used for isolating genomic DNA using the DNeasy Plant Mini Kit (Qiagen). 100 ng of this genomic DNA was subsequently used to prepare Bisulfite libraries with the TruSeq Nano kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions, with the modifications used in ref.[68]. The libraries were sequenced in paired-end mode, with approximately 8.5 million 150 bp reads/library on an Illumina HiSeq3000 instrument.

Processing of BS-seq data and DMR calling

Raw BS-seq reads were aligned using Bismark with default parameters[69] and mapped to the *A. thaliana* (TAIR10) reference genome. The bisulfite conversion efficiency for each sample was estimated by evaluating the fraction of positions correctly called as unmethylated in the chloroplast genome. It was consistently above 99.6% in all samples. The mapping efficiency for all samples varied between 40 and 65%, with an average of 49% (Supplemental Table S2). Deduplicated bam files generated by Bismark were sorted and then processed using *Methylscore* (https://github.com/Computomics/MethylScore) in the *slurm* CBE cluster of the Vienna Biocenter (VBC), to identify DMRs (Differentially Methylated Regions) with the following parameters: DMR_MIN_C=10 (minimum 10 cytosines in each DMR), DMR_MIN_COV=3X (minimum 3X coverage in each cytosine), MR_FREQ_CHANGE=20 (at least 20% of samples showing a change in MR frequency to be tested as a candidate DMR), CLUSTER_MIN_METH_DIFF=20 (which sets a 20% cutoff for methylation difference between clusters in the CG, CHG and CHH contexts). All other parameters were based on default settings.

DMR coordinates were intersected with individual genome-wide cytosine methylation levels based on the *genome_matrix* file generated by Methylscore. A total of 2,836 AllC DMRs (in all 3 contexts), 2,388 CG DMRs, 350 CHG DMRs and 1,023 CHH DMRs were called across 73 samples (55 mutants and 18 wildtypes). In some cases, a DMR was enriched for more than one context, resulting in partial redundancies. Subsequently, the three context-specific DMRs (CG, CHG, CHH) were evaluated for context-specific average methylation levels by intersecting with sample-specific cytosine methylation data. This was achieved using the *bedtools* software (v2.26.0) with the following command:

bedtools map -a DMR_coordinates.bed -b methylated_cytosines_sampleX.bed -c 5 -o mean -nonamecheck -null "NA" -g TAIR10genomesize > DMR_Methavg_sampleX.bed

Although the DMR calling was performed with a three-read cutoff for each cytosine in a DMR, there remained some samples which did not have sufficient coverage. Therefore, we retained the same DMRs, but calculated average methylation by lowering the cutoff to 2 reads per cytosine. For downstream analysis, DMRs with a maximum of 7 NAs (insufficient coverage) out of 73 samples were retained. This resulted in 1,966 AllC DMRs, 1,569 CG DMRs, 207 CHG DMRs and 614 CHH DMRs which were used for intersecting with ACRs and DEGs.

Nuclei isolation for ATAC-seq

For ATAC-seq analyses, each biological replicates was processed individually. Fixed tissue was chopped finely with 500 µl of General Purpose buffer (GPB; 0.5 mM spermine•4HCl, 30 mM sodium citrate, 20 mM MOPS, 80 mM KCl, 20 mM NaCl, pH 7.0, and sterile filtered with 0.2 µm filter, followed by the addition of 0.5% of Triton-X-100 before usage). The slurry was filtered through one-layered Miracloth (pore size: 22-25 µm), followed by filtration through a cell-strainer (pore size: 40 µm) to collect nuclei. Approximately 20,000 DAPI stained nuclei were sorted using fluorescence-activated cell sorting (FACS) as two technical replicates. For samples from dwarfed mutant lines where leaf tissue was scarce, approximately 8,000 nuclei were sorted per technical replicate.

Fluorescence-activated nuclear sorting (FANS) for ATAC-seq

Liberated nuclei were sorted with a MoFlo XDP (Beckman Coulter) instrument outfitted with a 488 nm (elliptical focus, 100 mW) and a 375 nm (spherical focus, 35 mW) laser for scatter and DAPI emission, respectively. Nuclei were sorted with a 70 µm cytonozzle, sheath PBS [pH 7.0] at psi 30.5/30.0 sample/sheath, purify 1 drop, triggered off the DAPI emission (465/30nm). The 2C endoreduplicated population was identified as the first clear DAPI emitting population over scatter debris. DAPI emission was utilized to reduce further contaminating debris, followed by a clean-up utilizing 530/34 emission from the 488 nm laser. 488 nm channels: SSC (488/6), FL1 (520/34). 375 nm channels: FL8 (405/30), FL9 (465/30), FL10 (542/27). See **Supplementary Figure S5** for the gating scheme.

ATAC-seq library prep

Sorted nuclei were heated at 60°C for 5 minutes, followed by centrifugation at 4°C (1,000 g, 5 minutes). The supernatant was removed, and nuclei were resuspended with a transposition mix (1 µl homemade Tn5 transposase, 4 µl of 5X-TAPS-DMF buffer and 15 µl autoclaved water) followed by a 37°C treatment for 30 minutes. 200 µl SDS buffer and 8 µl 5 M NaCl were added to the reaction mixture, followed by 65°C treatment overnight. Nuclear fragments were then cleaned up using Zymo PCR column-purification (DNA Clean and Concentrator). 2 µl of eluted DNA was subjected to 14 PCR cycles, incorporating Illumina indices, followed by a 1.8:1 ratio clean-up using SPRI beads.

Genomic DNA libraries (10 ng input from the DNA extracts used for BS-seq-library prep) were prepared using a similar library prep protocol starting with Tn5 enzymatic digestion (0.5 µl homemade Tn5 transposase, 4 µl of 5X-TAPS-DMF buffer and autoclaved water made up to a final reaction volume of 20 µl including the DNA template). Digested gDNA was immediately column-purified, followed by PCR (2 µl of eluted DNA was used as template for 11 PCR cycles) incorporating Illumina indices, followed by a 1.6:1 ratio clean-up using SPRI beads.

Processing of ATAC-seq libraries and peak-calling

Libraries were sequenced on an Illumina HiSeq3000 instrument with 2 x 150bp paired-end reads. Each technical replicate derived from nuclei sorting was sequenced at approximately 7 million paired-end reads per library. The reads were aligned as two single-end files to the TAIR10 reference genome using *bowtie2* [default options], filtered for the SAM flags 0 and 16 (only reads mapped uniquely to the forward and reverse strands), converted separately to bam files. The bam files were then merged, sorted, and PCR duplicates were removed using *picardtools.* The sorted bam files were then merged with the corresponding sorted bam file of a second technical replicate (samtools merge --default options) to obtain a final average of 11 million mapped reads for each biological replicate.

Genomic DNA libraries were similarly aligned, with a final average depth of 4.5 million mapped reads per library.

Peak calling was carried out for each biological replicate using *MACS2* using the following parameters:

macs2 callpeak -t [ATACseqlibrary].bam -f BAM --nomodel --extsize 147 --keep-dup=all -g 1.35e8 -n [Output_Peaks] -B -q 0.01

After peak calling, every peakset was further filtered based on their respective q values in the MACS2 peaks.xls files, retaining peaks with q <= 0.001, thereby reducing the false positives when all 158 samples were subsequently tested together. This additional filtering step was carried out separately after MACS2 calling to minimise the effect on peak size based on the q-value.

Filtered peak files and .bam alignment files from a total of 158 samples (104 mutant samples plus 54 wild-type samples) were processed with the R package DiffBind to identify consensus peaks which overlapped in at least two out of three biological replicates per group, and represented peaks unique to at least one group (FDR <0.01). To normalize peak accessibility counts with the background probability of Tn5 integration biases in the genome, .bam files of the control gDNA libraries were also provided in the DiffBind samplesheet (thereby ensuring that peak counts were normalized to controls).

A total of 35,049 consensus peaks were identified, with accessibility scores in each peak per sample evaluated in counts per million (CPM) after TMM (Trimmed Mean of M-values) normalization. Except for three out of 158 ATAC-seq libraries, FRIP (Frequency of reads in peaks) scores relative to the Consensus peakset was between 0.2 - 0.31 for all samples,

reflecting the average representation of sample-specific peaks in the consensus dataset. After removing peaks which occurred in Chloroplast and Mitochondrial genomes, 34,966 peaks remained.

Metaplot generation

From 34,966 Consensus ATAC-peaks (pre-filtered), accessibility values were derived for each sample, and converted into bigWig files using the *bedGraphtobigWig* command (UCSC software). Similarly, cytosines in all-contexts and their corresponding methylation levels were derived for each Bisulfite library, and converted to bigWig files. Metaplots were generated using the deepTools (3.5.0) package (https://deeptools.readthedocs.io/en/develop/index.html), first with the *computeMatrix* command to evaluate the mean of the epigenetic factor tested (methylation in % or chromatin accessibility in CPM) within 1000bp upstream and downstream of a given set of reference regions (TAIR10 Transposable elements/TAIR10 Non-TE protein coding genes). The output bed file from this command was subsequently used to generate metaplots using the *plotProfile* function.

RNA extraction and RNA-seq library prep

RNA from each biological replicate was extracted individually using a column-based protocol adapted from ref.[70]. RNA quality was validated with the Nanodrop spectrophotometer, and normalized to 500 ng in a 50 µl volume. Normalized RNA was subsequently used for mRNA library prep using an in-house custom protocol adapted from Illumina's TruSeq library-prep, with details provided in ref. [71].

Mapping and identification of DEGs

RNA-seq libraries were sequenced at an average coverage of 8 million 150 bp single-end reads per library using HiSeq3000. Reads of the same sample from multiple sequencing lanes of the same flow cell were merged together, and 9 samples with total reads > 12.5 M were subsampled (using different seeds) to 80% using seqtk (v.2.0-r82-dirty, https://github.com/lh3/seqtk) with the following command:

seqtk sample -sX <merged_fastq> 0.80 > subsampled_output.fastq

All samples were aligned using *bowtie2* to the TAIR10 reference genome, prepared using the *rsem-prepare-reference* function of the RSEM software. Aligned bam files were sorted and indexed using *samtools* V1.9. Gene transcript counts for each sample were estimated using *rsem-calculate-expression*. From each sample, chloroplast genes, mitochondrial genes, and rDNA cluster genes were excluded from downstream analyses. Twelve genes with excessive read counts across all samples were also excluded.

Transcript counts per sample and corresponding metadata were then imported using the R packages "tximport" and "tximportData" for creating a DESeq object (R package "DESeq2"). The

DESeq function was applied to the object under the two-factor interaction model ~*genotype + accession + genotype:accession* (where *genotype ==* wild-type *or mutant*) using default parameters (nbinomWald test). To obtain DEGs between wild-type and mutant genotypes across all accessions, a contrast was performed set to *genotype*, retaining only those genes with a p< 0.01 and $|\log_2$ FoldChange| >1. For identifying accession-specific DEGs, similar contrasts were performed, nested within each accession.

A consensus set of 7,132 DEGs were derived from all 18 accession specific contrasts and the all-mutants-against-all-wild-type contrast, retaining only DEGs that occurred in at least two of all 19 contrasts. These were further classified as 1,401 TE-associated DEGs (TE-DEGs) and 5,731 Non-TE-DEGs based on the TAIR10 gene annotation (TE-genes defined as those with the term "transposable-element-gene").

Generation of consensus datasets

To understand how differentially expressed genes could be epigenetically regulated and simultaneously compare them across accessions, we generated consensus sets of DEGs, DMRs and ACRs (Methods) that best represented gene expression, methylation and chromatin accessibility variation across all genotypes and accessions.

Consensus DEGs were generated by including DEGs from a total of 19 contrasts - 18 pairwise contrasts (*met1* mutants vs wild-type plants) in each accession, and a contrast between all *met1* mutants and all wild-type plants. This set of DEGs was filtered to retain only DEGs that occurred in at least 2 out of the total 19 contrasts, to obtain a final set of 7,132 Consensus DEGs. These were further classified as 1401 TE-associated DEGs (TE-DEGs) and 5,731 Non-TE-DEGs (protein-coding). The metric for evaluating expression levels in DEGs was chosen as the variance stabilized transformed read counts (vsd counts) generated by the DESeq2 package for each of the 158 RNA-seq libraries. We refer to these as transformed read counts in all figures.

ATAC-peaks found across all samples (34,966) were also filtered in two steps. First, peaks that showed similar accessibility between both mutant lines relative to the WT line in each accession were retained. The second round of filtering retained peaks that showed an accessibility change between mutants and wild-type plants in at least two out of 18 accessions. This resulted in 31,223 filtered consensus ATAC-peaks. Chromatin accessibility levels were measured in counts per million (CPM) after TMM (trimmed mean of M-values) normalization generated by the DiffBind package.

From the complete set of context-specific DMRs identified, DMRs with a maximum of 7 NA values (insufficient coverage) out of 73 samples were retained. This resulted in 1,966 All-C-DMRs, 1,569 CG DMRs, 207 CHG DMRs and 614 CHH DMRs. Since we were primarily interested in understanding the effects of MET1 on methylation, we considered only CG-DMRs and All-C-DMRs for intersecting with other features. While both groups contained some redundant DMRs, the methylation level evaluated for each DMR was evaluated differently (for C-DMRs, methylation level evaluated only for CG context; for All-C-DMRs, methylation level evaluated for all Cs in CG, CHG, CHH contexts). Methylation levels in DMRs were measured as

mean methylation percentage of all context-specific and sample-specific cytosines within the assigned chromosomal region.

Generation of feature intersections between DEGs, DMRs and ACRs

Next, we examined the co-occurrence of DMRs with DEG positions. DMRs occurring over the gene-body of a DEG were called 'gene-body DMRs', while those that occurred within 1.5kb upstream or downstream of the TSS/TTS respectively were called 'cis DMRs'. Several DEGs had multiple DMRs associated with them, and therefore we retained only one DMR for each DEG, which showed the largest difference in methylation level between mutant and WT for each mutant genotype, thereby aiming to represent only the strongest methylation signals that could explain gene expression differences. To find the extent to which MET1-induced CG methylation could influence gene expression, and compare it with methylation in all contexts, we intersected DEGs (TE-DEGs and Non-TE-DEGs separately) with both CG-DMRs and All-C DMRs.

For each DEG, we measured changes in gene expression levels between *met1* mutants and wild-type plants, and corresponding changes in methylation levels of their closest DMRs.

Similarly, we analysed the ACRs closest to each DEG. Since a large majority of ACRs occurred in proximity to the transcription start site, we grouped all ACRs occurring either over the gene body or within 1.5 kb upstream or downstream of the TSS/TTS respectively, under a single 'cis' category. For each DEG, a single ACR which showed the largest difference in accessibility between mutant and WT for each mutant genotype, was retained.

gbM-like and CG teM-like genes

To follow conventional definitions of gbM, only Non-TE-DEGs with CG-DMRs overlapping the gene body were considered. This resulted in 196 DEGs across homozygous *met1* mutants in 17 accessions (since one accession, Bl-1, did not have homozygotes). We identified 51 gbM-like genes in this set, which had >80% methylation in the wild-type parent, and transformed expression counts in wild type >=7.55, which represented genes in the highest quintile range of the distribution of expression changes for all 196 genes.

Next, we identified 59 CG teM-like genes that exhibited >80% methylation in the wild-type state, and transformed expression counts in wild-type state being <=4.42, which represented genes in the lowest quintile range of the distribution of expression changes for all 196 genes).

Metadata (transformed expression counts and methylation levels in *met1* and mutants wild-type plants) for gbM-like and CG teM-like genes were then extracted from the total set of 196 genes (thereby representing the same genes in all accessions, irrespective of whether they were called gbM or CG teM) and used for visualisation.

Gene Ontology Enrichment and Visualisation

GO Enrichment was carried out using agriGO (http://bioinfo.cau.edu.cn/agriGO), with the Singular Enrichment Analysis (SEA) analysis tool and Arabidopsis (TAIR10) genemodel as the

reference. Graphical results of significant GO terms were generated in agriGO. The GO terms were further visualised with ReviGO (http://revigo.irb.hr).

Segregation distortion analyses

Experimental design: To accurately estimate the extent of this segregation distortion in mutants of various accessions, we grew a maximum of 96 segregating progeny from heterozygous parent lines (2 mutant lines per accession) and genotyped them individually using amplicon-sequencing of the *MET1* locus, amplifying a 150 bp region around the CRISPR/Cas9-induced frameshift mutations.

Amplicon-seq library prep, sequencing and genotyping: Amplicon-seq libraries were prepared according to the CRISPR-finder system[72], where amplicons from multiple 96-well plates can be pooled together for high-throughput sequencing by incorporating frameshifted primers and TruSeq adapters with 96 barcodes. The amplicons were designed as 152 bp sequences spanning the gRNA target site in *MET1* **(Supplementary Table S3)**. A total of 3,119 individual samples were sequenced at an average coverage of 12,000 reads per sample on a HiSeq3000 instrument with 2 x 150 bp paired-end reads.

Sequenced read pairs were first merged using *FLASH - Fast Length Adjustment of Short reads* (https://ccb.jhu.edu/software/FLASH/), followed by demultiplexing based on their plate-specific frameshifted primers (using the *usearch10 fastx_truncate* function).

Only samples that had >=80 reads were retained for downstream processing. For all samples within a plate (i.e., segregating progeny), amplicon-reads per individual were counted for the ratio of wild-type alleles to mutant alleles, to estimate whether the genotypes were Homozygous for the mutant allele (wild-type reads <= 15%), heterozygous (wild-type reads >= 42% or <= 58%) or wild type (wild-type reads >= 90%). A fourth genotypic classification, "skewed heterozygous" was made for individuals where the read ratio between the wild-type and mutant alleles were either 0.15 - 0.42 or 0.58 - 0.90 (i.e., if either one of the alleles were more represented than the other, but not approximately equal in counts).

For Bu-0 Line2, Ste-0 Line2 and Bs-1 Line 2 samples that had more than one mutant *MET1* allele, additional genotypic categories were specified: homozygous allele 2, heterozygous allele 2, skewed heterozygous allele 2, bi-allelic, skewed bi-allelic and tri-allelic.

Cytometric ploidy analysis

Cytometric determination of generative ploidy levels was conducted on a CytoFlex (BeckmanCoulter) outfitted with a 488 nm laser, 10 µL min^{-1} flow rate. Nuclei were freshly liberated by chopping into cold General-purpose Buffer [73], filtered through 40 µm mesh, and stained with 50 µg mL^{-1} propidium iodide and 50 µg mL^{-1} RNAse for 10 minutes at 20°C. The 2C endoreduplication population was identified as the first clear PI emitting population over scatter debris. The 2C nuclei of *Solanum lycopersicum* (var. Moneymaker) provided by the ZMBP Cultivation Facility or *Capsicum annuum* provided by Annett Strauss (ZMBP) were used as internal standards to determine the relative Arabidopsis generative ploidy levels.

SUPPLEMENTARY FIGURES

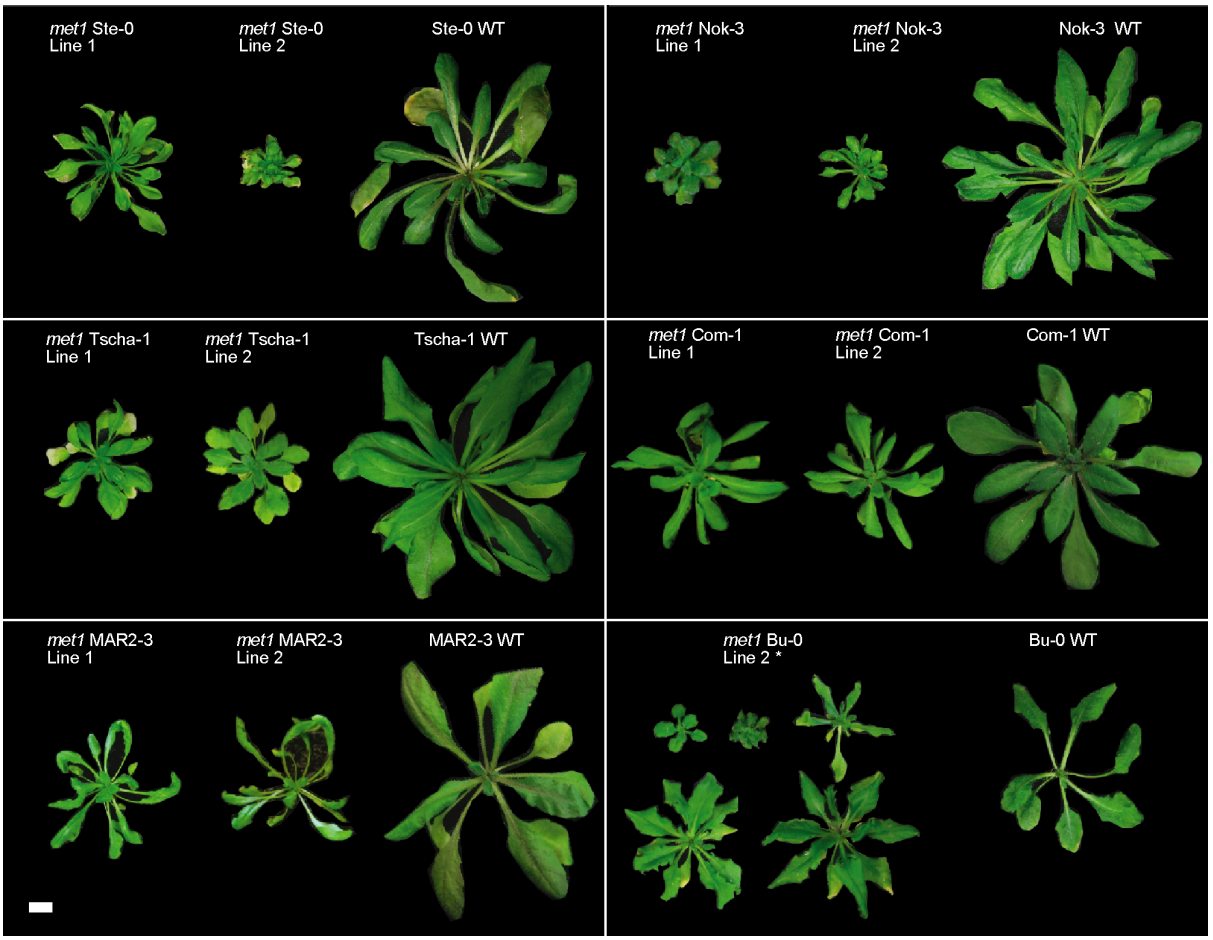


Figure S1: Rosettes of *met1* mutants in six accessions. Representative images of two independently derived mutant lines and a wild-type plant at six weeks after germination; scale bar denotes 1 cm. *met1* mutants of the accession Bu-0 are marked by a * since they were tetraploid and exhibited a wide range of phenotypes arising from different dosages of two mutant alleles and a wild-type allele.

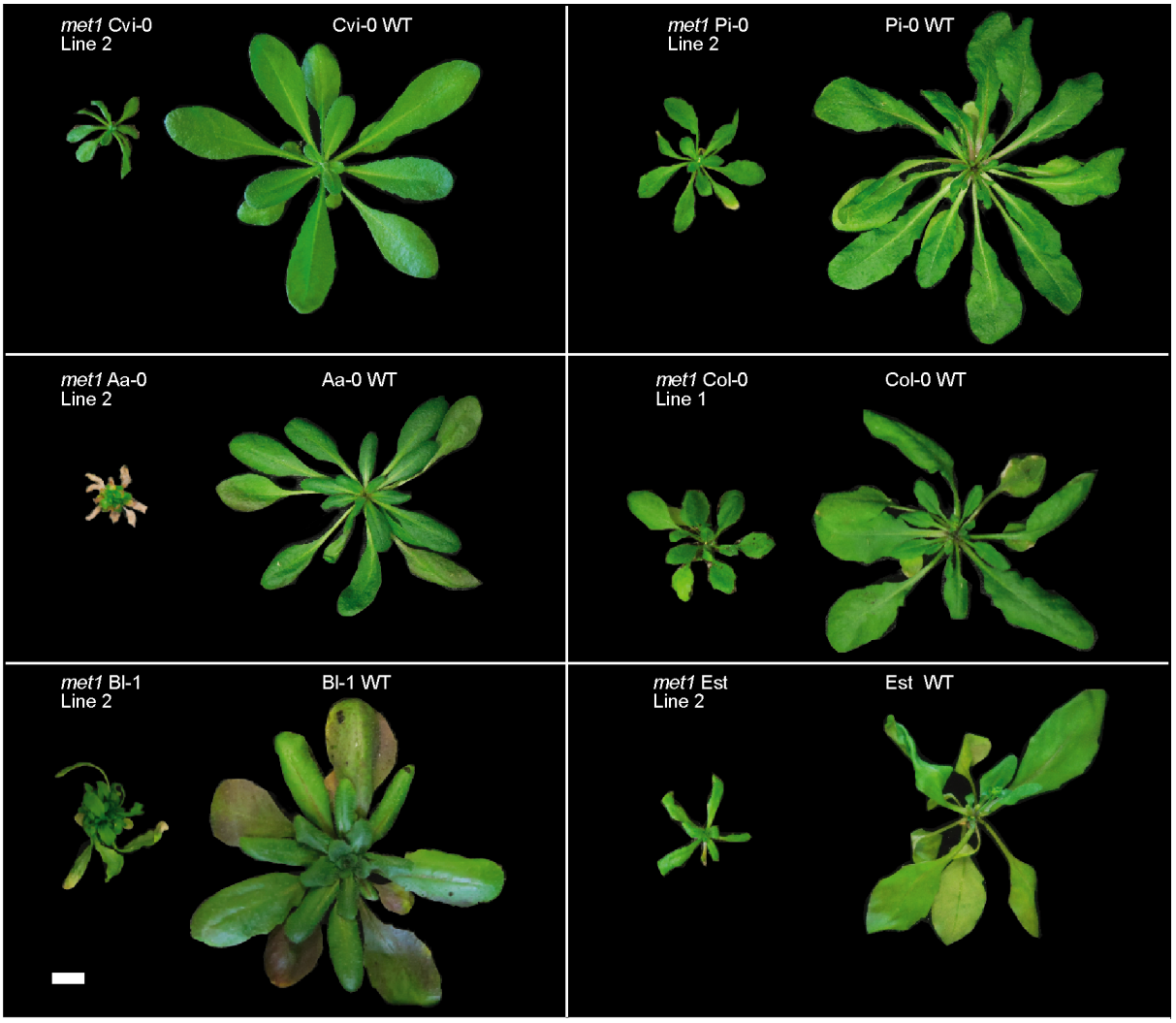


Figure S2: Rosettes of *met1* mutants in six accessions. Representative images of mutant and WT plants at six weeks after germination; scale bar denotes 1 cm.

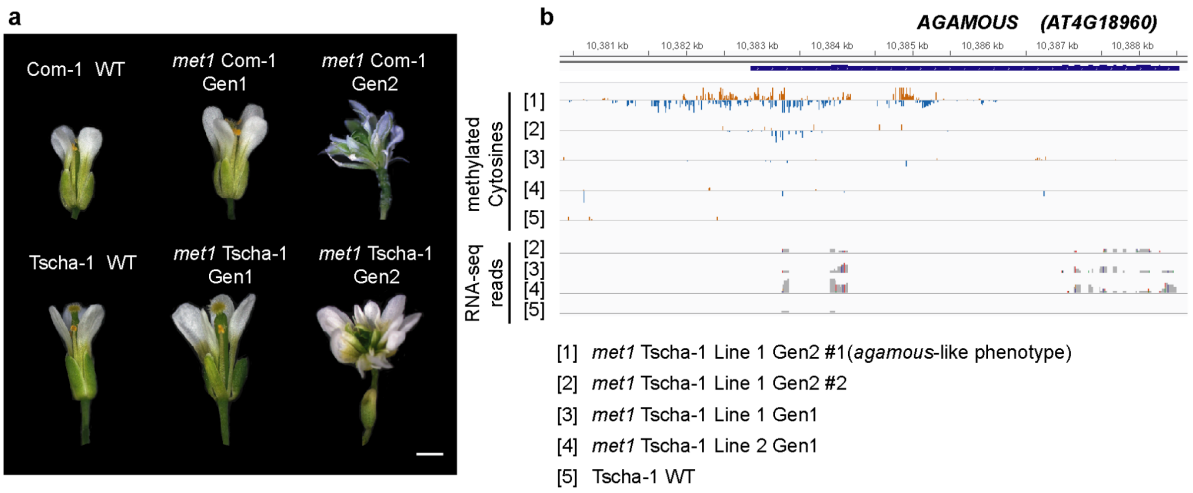


Figure S3: Transgenerational *ag-like* phenotypes in *met1* mutants of Com-1 and Tscha-1. A few *met1* mutant lines of the Tscha-1 and Com-1 accessions exhibited indeterminate flowers, a phenotype known to arise from genetic[74] and epigenetic inactivation[4] of the *AG* gene. BS-seq of one such line in Tscha-1 also showed an increase in methylation at this locus. **(a)** Flower phenotypes of wild-type and homozygous *met1* plants in two generations (Gen1 and Gen2). Scale bar denotes 1 mm. **(b)** Genome browser screenshot of methylated cytosines (all-contexts) and RNA-seq reads at the *AG* locus for various Tscha-1 lines.

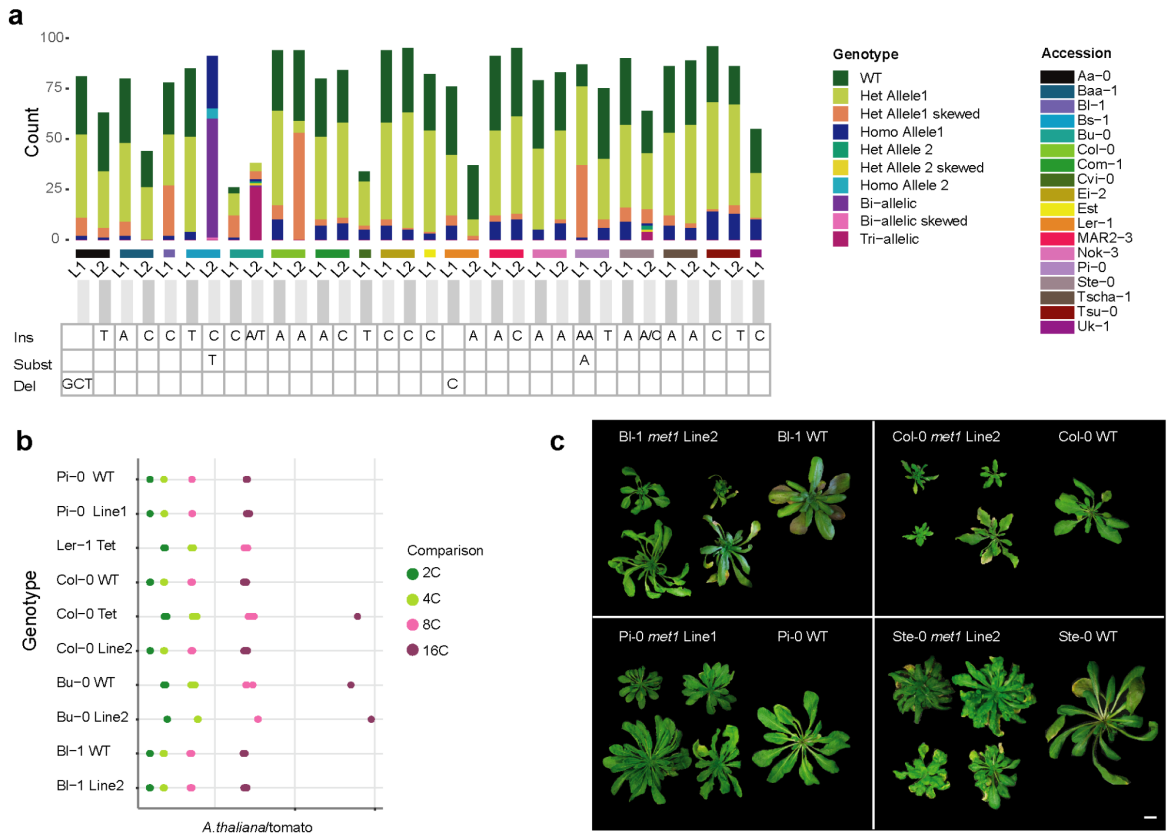


Figure S4: Segregation distortion in *met1* mutants and the presence of skewed heterozygous individuals. (a) Genotypes of segregating *met1* mutants representing sampled individuals and associated mutations for every line. **(b)** Scatter plot of endopolyploidy peak position ratios (from flow cytometry profiles) in candidate mutant lines and wild-type plants relative to the tomato internal standard. Col-0 and Ler-1 tetraploids ('Col-Tet' and 'Ler-Tet' respectively) were used as references for validating ploidy variation in candidate lines. **(c)** Phenotypic variation in heterozygous plants of Bl-1 Line 2, Col-0 Line 2, Pi-0 Line 1 and Ste-0 Line 2. Scale bar represents 1 cm.

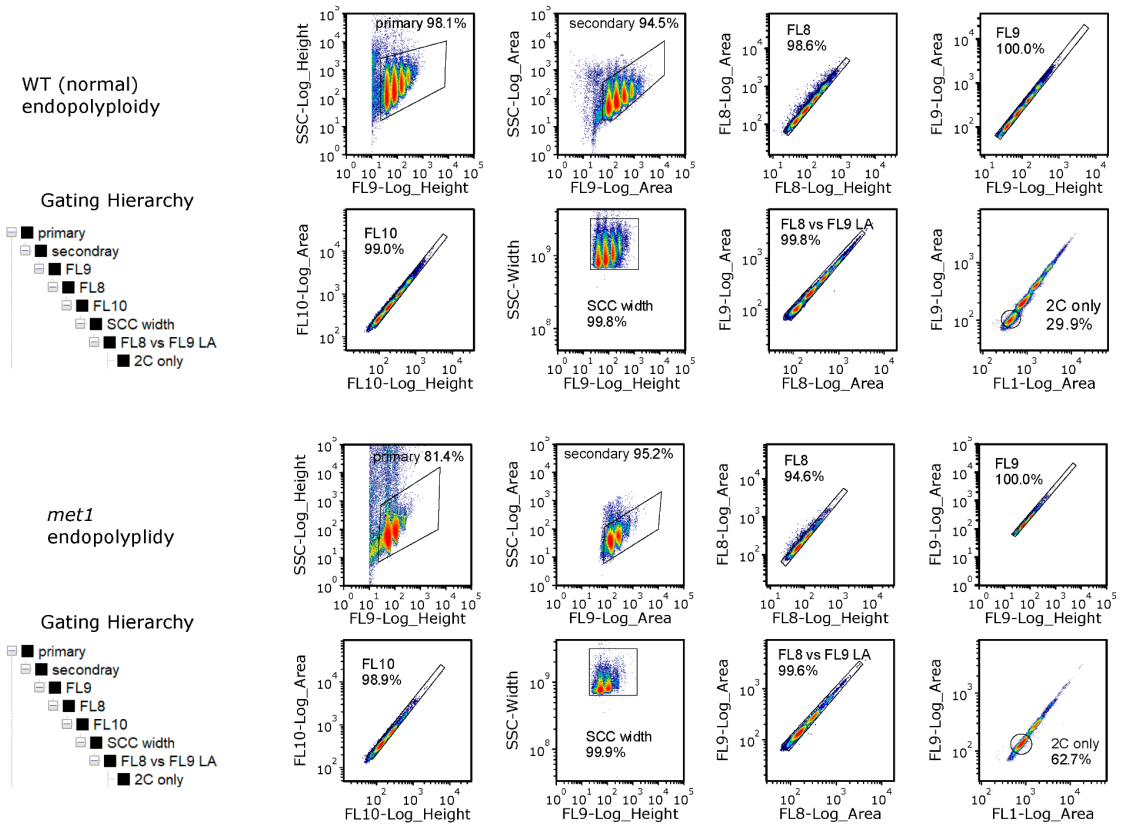


Figure S5: Reduced endoreduplication in *met1* mutants. Fluorescence-activated nuclear sorting and gating scheme of wild-type and *met1* nuclei isolated from the Nok-3 accession.

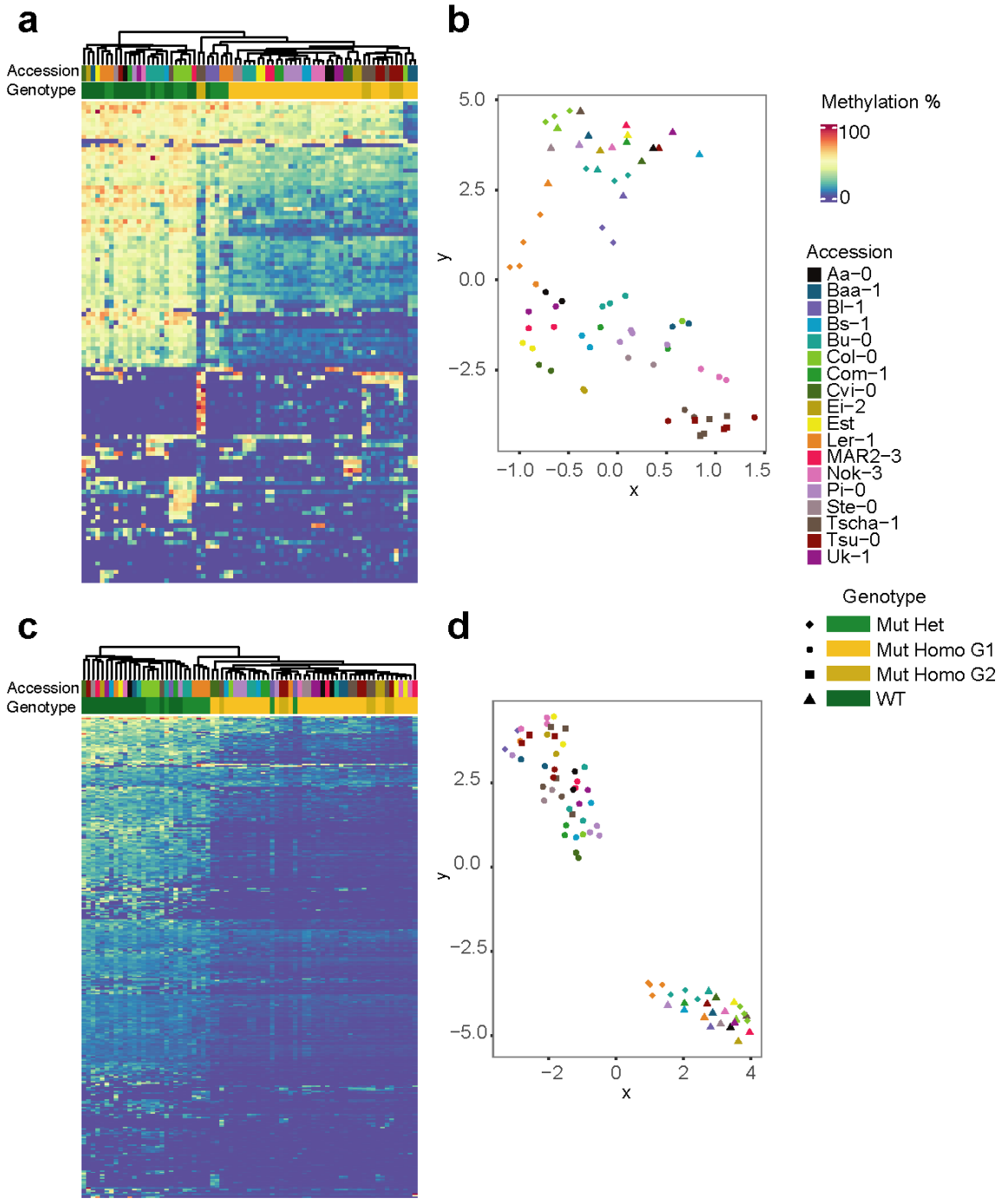


Figure S6: Differential methylation in non-CG contexts in *met1* mutants and wild-type individuals. (a) Heatmap and **(b)** UMAP visualization of CHG methylation levels in 73 samples (55 mutants and 18 wild-type plants) across 114 CHG-DMRs (from a total of 350 CHG-DMRs). **(c)** Heatmap and **(d)** UMAP visualization of CHH methylation levels in 73 samples (55 mutants and 18 wild-type plants) across 334 CHG-DMRs (from a total of 1,023 CHG-DMRs).

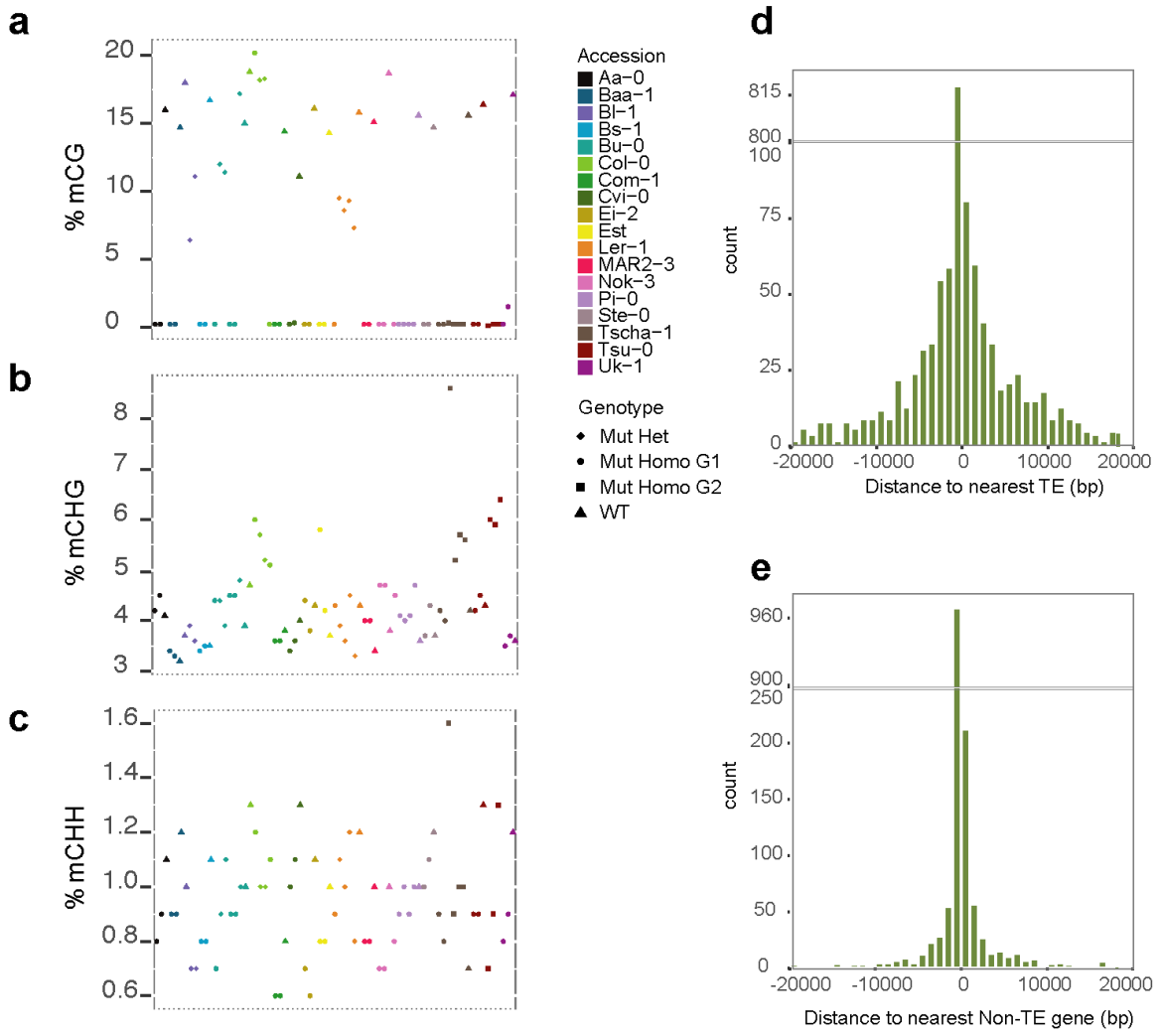


Figure S7: Altered genome-wide methylation levels in *met1* mutants compared to wild-type individuals. Genome-wide methylation levels of 73 samples (55 *met1* mutants and 18 wild-type plants) in the **(a)** CG , **(b)** CHG and **(c)** CHH contexts. Histogram showing distance of CG-DMRs to **(d)** nearest TE and **(e)** nearest Non-TE protein coding gene.

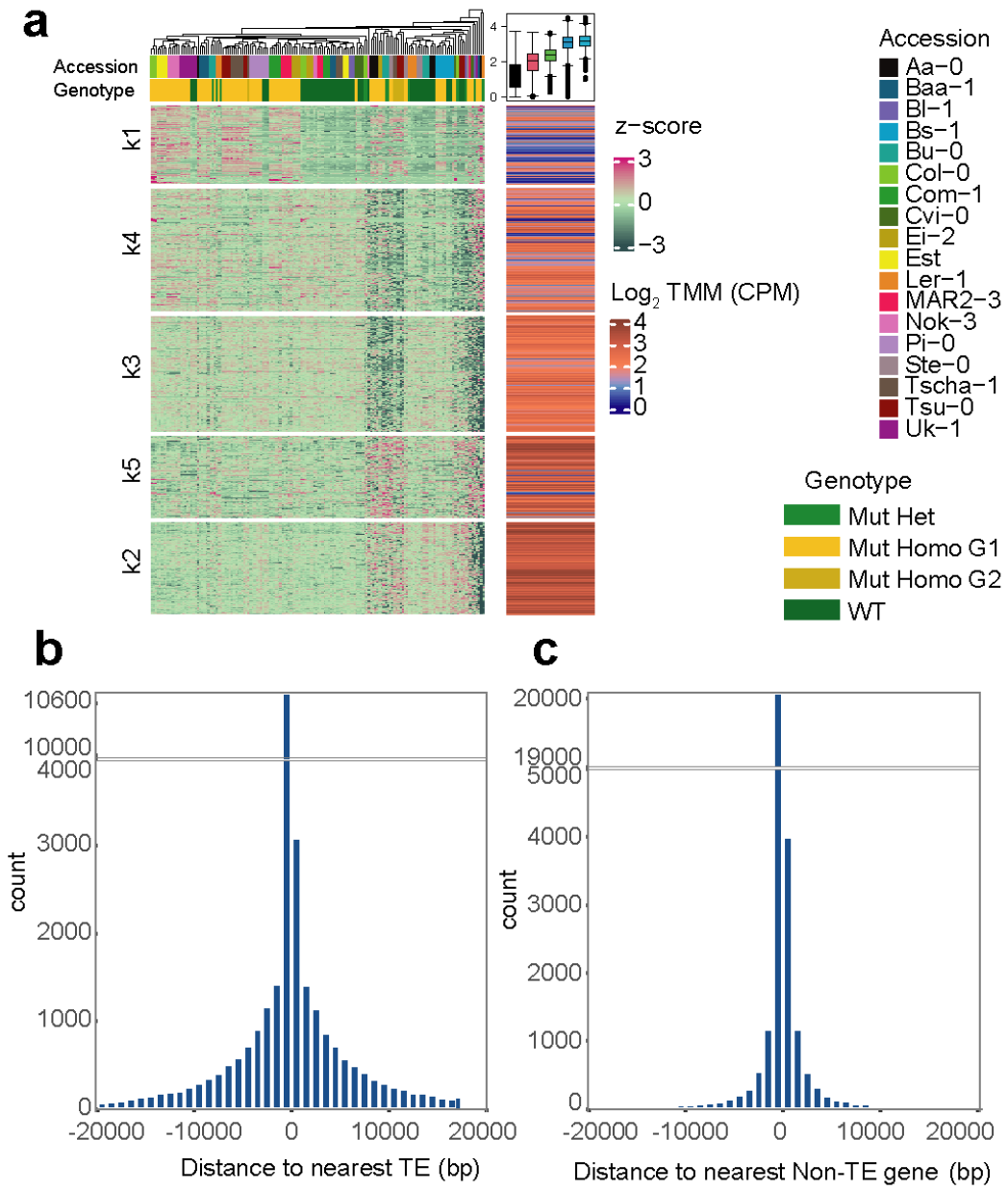


Figure S8: Altered genome-wide chromatin accessibility in *met1* mutants compared to wild-type individuals. (a) Heatmap of 35,049 z-scaled ACRs across 158 ATAC-seq libraries (104 *met1* mutant and 54 wild-type samples) and **(b)** histograms showing distance of ACRs to nearest TE and **(c)**nearest Non-TE protein coding gene.

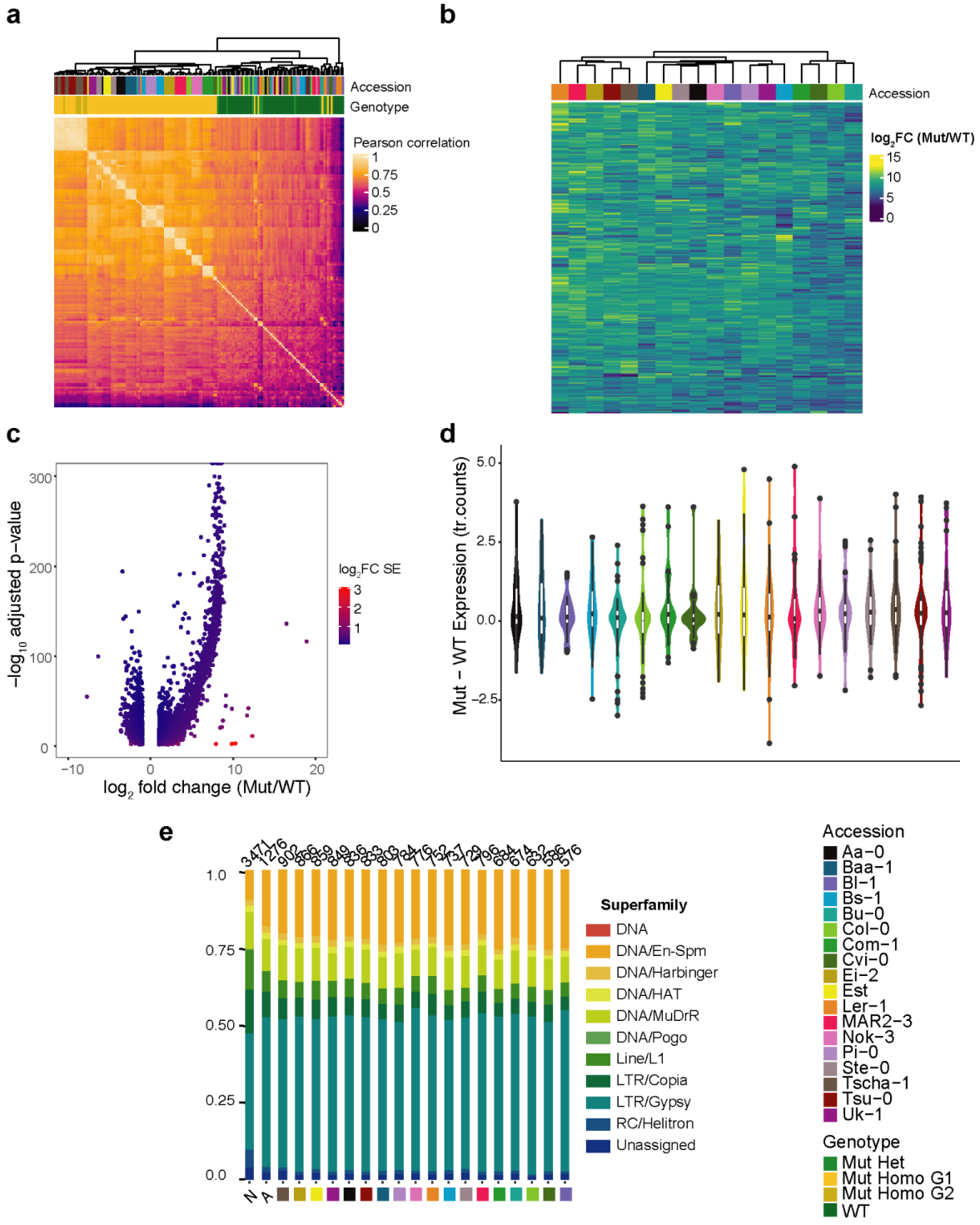


Figure S9: Accession-specific variation of differentially expressed genes (DEGs) in *met1* mutants. (a) Correlation between 291 universal DEGs across 158 RNA-seq libraries (104 *met1* mutant and 54 wild-type samples). **(b)** Heatmap showing $\log_2$ fold change in expression (Mut/WT) of 291 universal DEGs. **(c)** Volcano plot colored by standard error of $\log_2$ fold change in all-mutants-against-all-wild-type DEGs. **(d)** Thirty random genes examined for accession-specific variation in expression changes (measured as transformed read counts). **(e)**

Distribution of TE superfamilies in TE-DEGs across 19 contrasts, and a null distribution of all TE-genes (denoted by 'N'). Genotypes represented are wild-types ('WT'), heterozygous *met1* mutants ('Mut Het'), first generation homozygous *met1* mutants ('Mut Homo G1') and second generation homozygous *met1* mutants ('Mut Homo G2').

Figure S10: Accession-specific variation in the chromosomal distribution of up- and down-regulated DEGs. Ideograms showing chromosomal distribution of DEGs in 18 within-accession contrasts. Colors distinguish up- and downregulated DEGs.

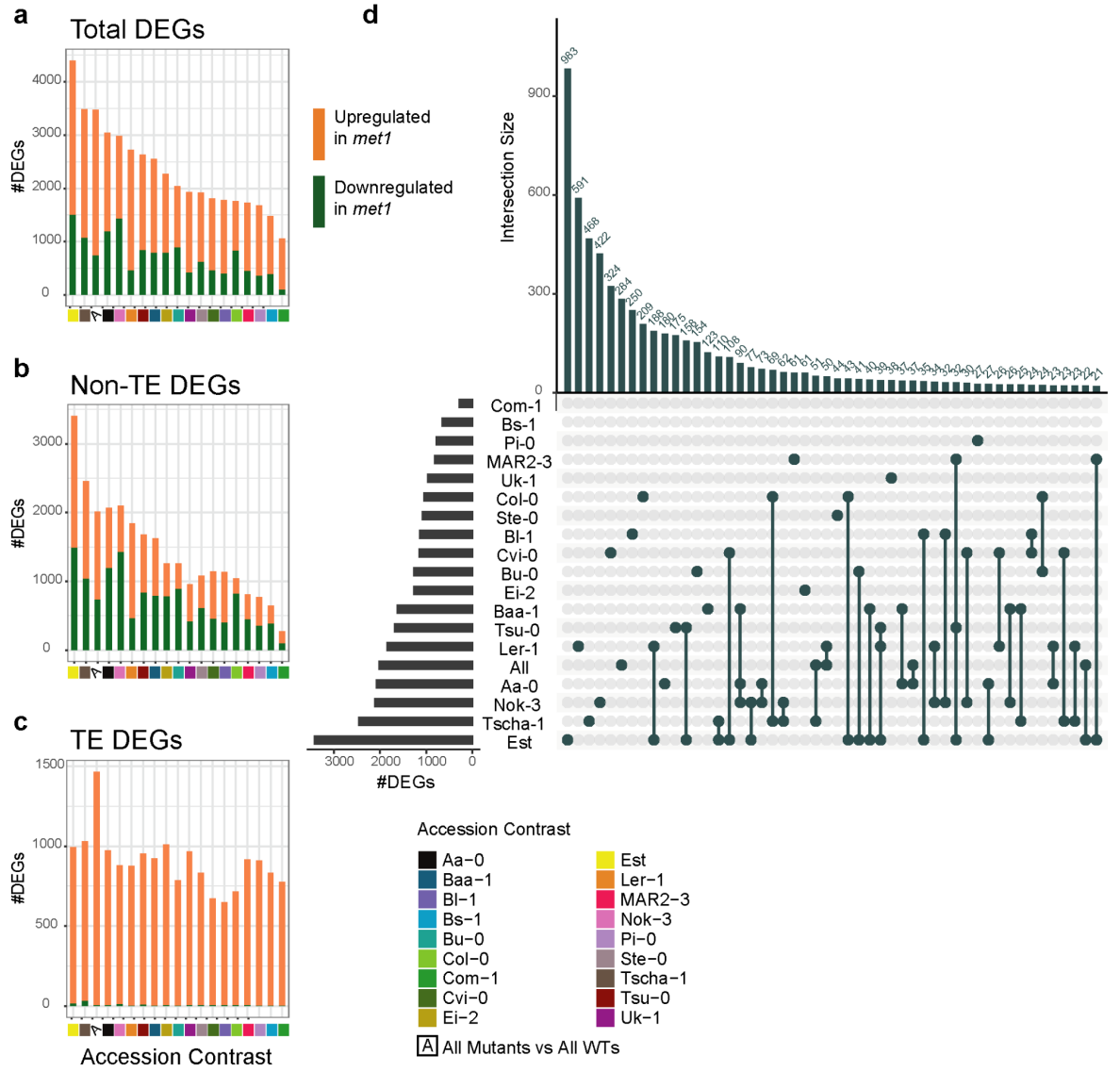


Figure S11: Comparisons and intersections between accession-specific DEGs. Variation in numbers of upregulated and downregulated **(a)** DEGs **(b)** Non-TE-DEGs **(c)** TE-DEGs across 19 accession-specific contrasts. **(d)** Upset plot of the top 50 intersections between Non-TE-DEGs in 19 accession-specific contrasts.

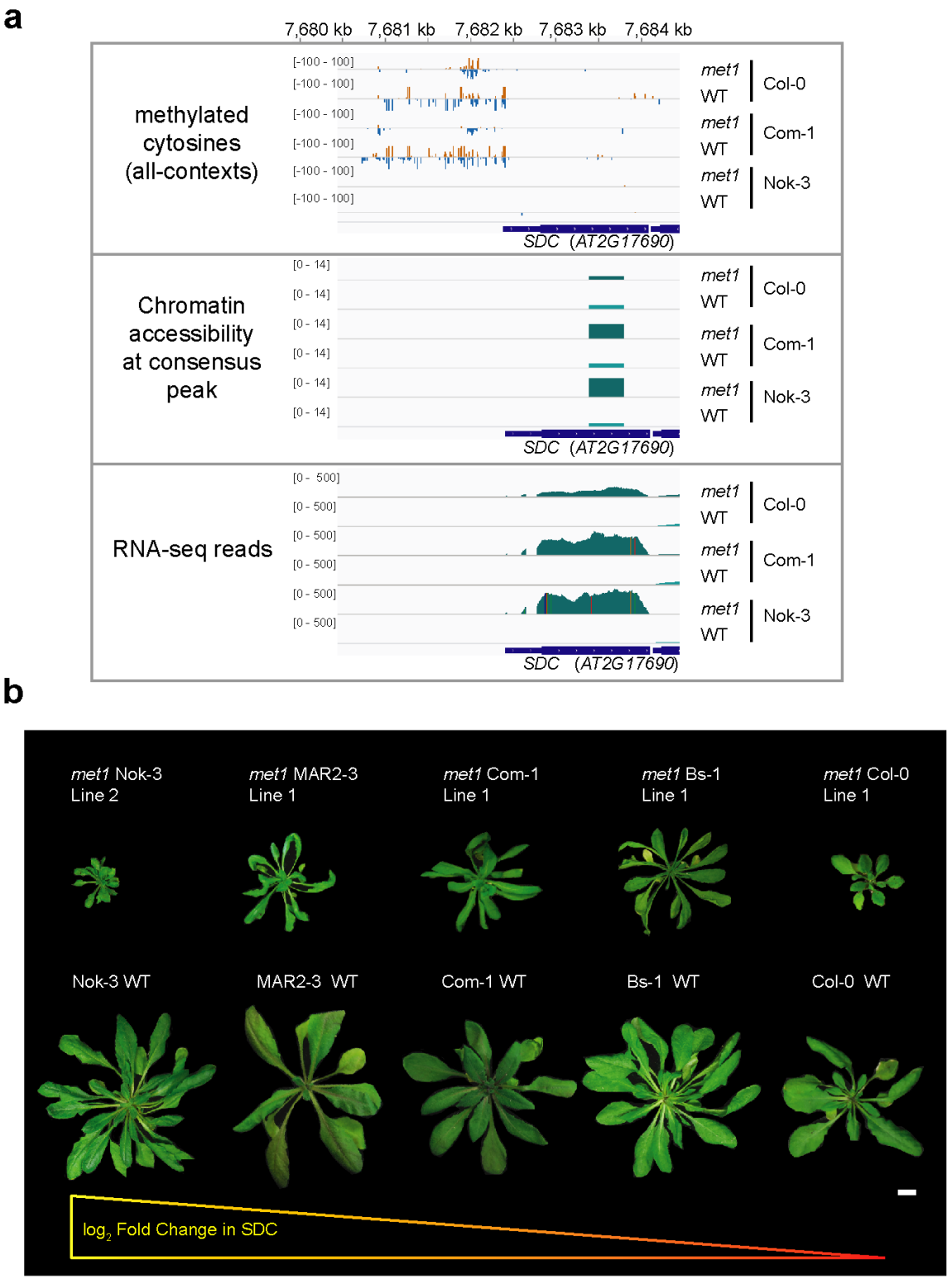


Figure S12: Epigenetic landscape at the *SDC* locus. (a) Genome browser screenshot of methylated cytosines, chromatin accessibility and RNA-seq read count in *met1* mutants and wild-types of three accessions, Col-0, Com-1 and Nok-3. **(b)** Rosettes of *met1* mutants and wild-type plants from five accessions ordered by $\log_2$ fold change in SDC expression.

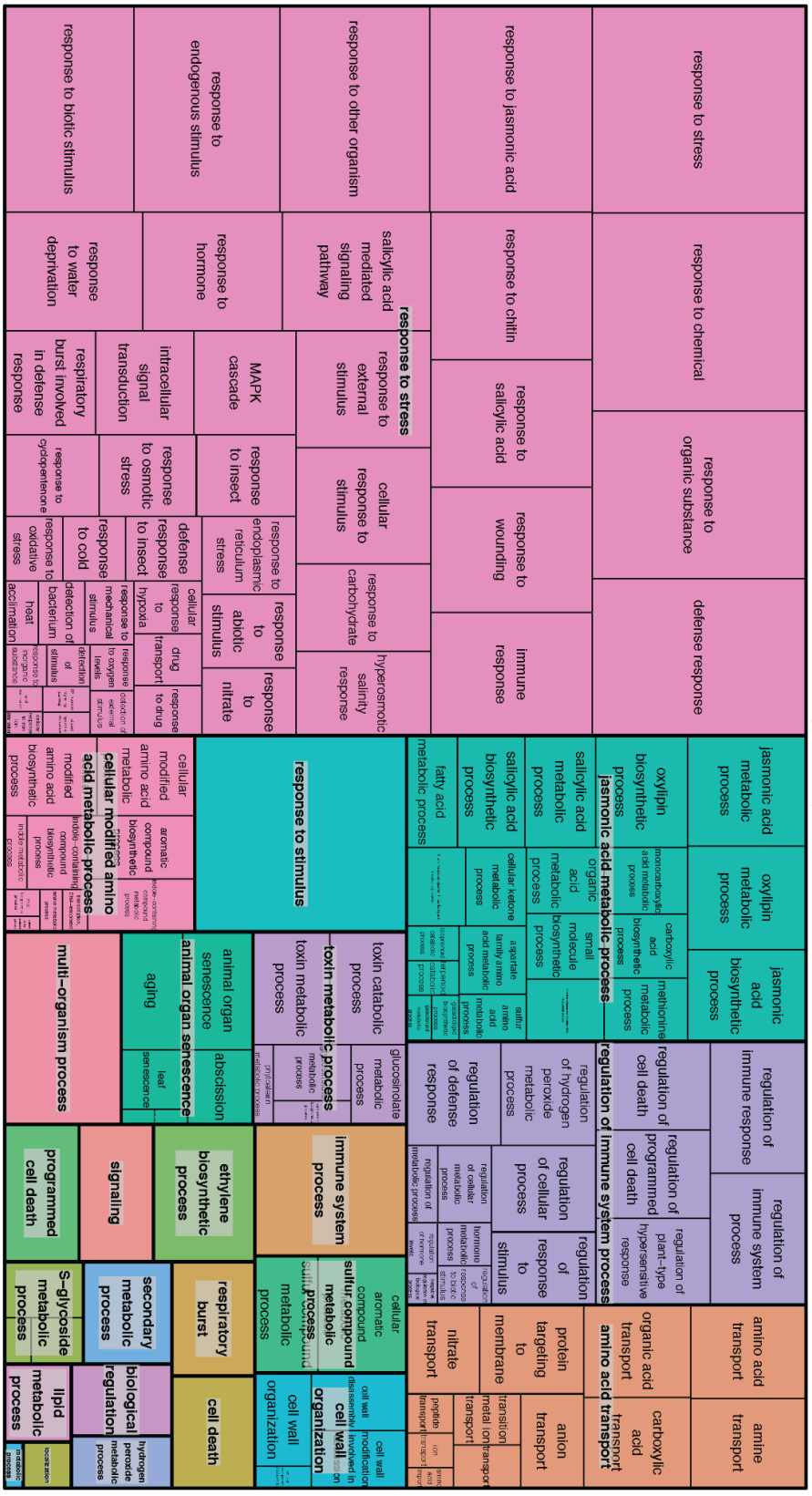


Figure S13: Gene Ontology enrichment for 2,013 Non-TE-DEGs identified in a contrast between all *met1* mutants and all wild-type samples.

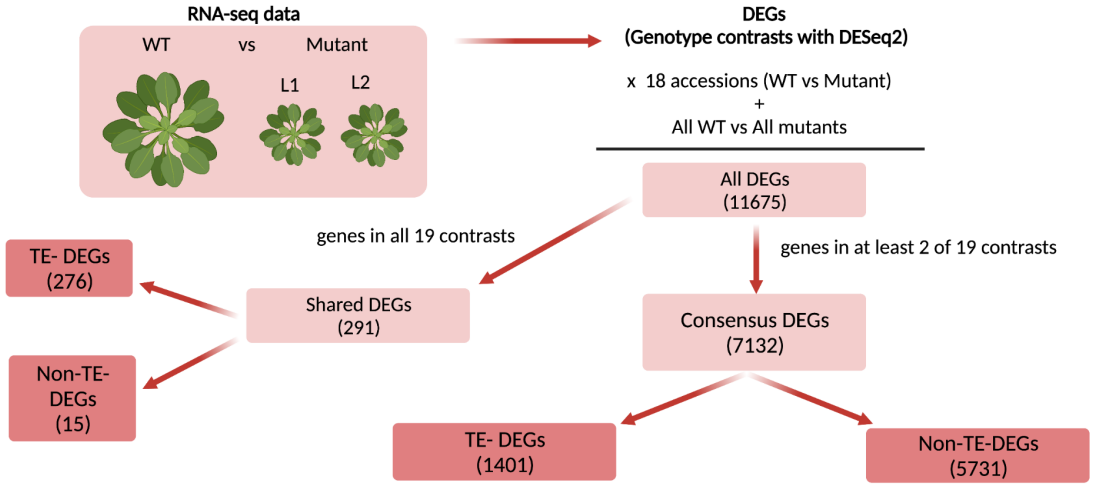


Figure S14: Diagram of generating consensus DEGs from RNA-seq data.

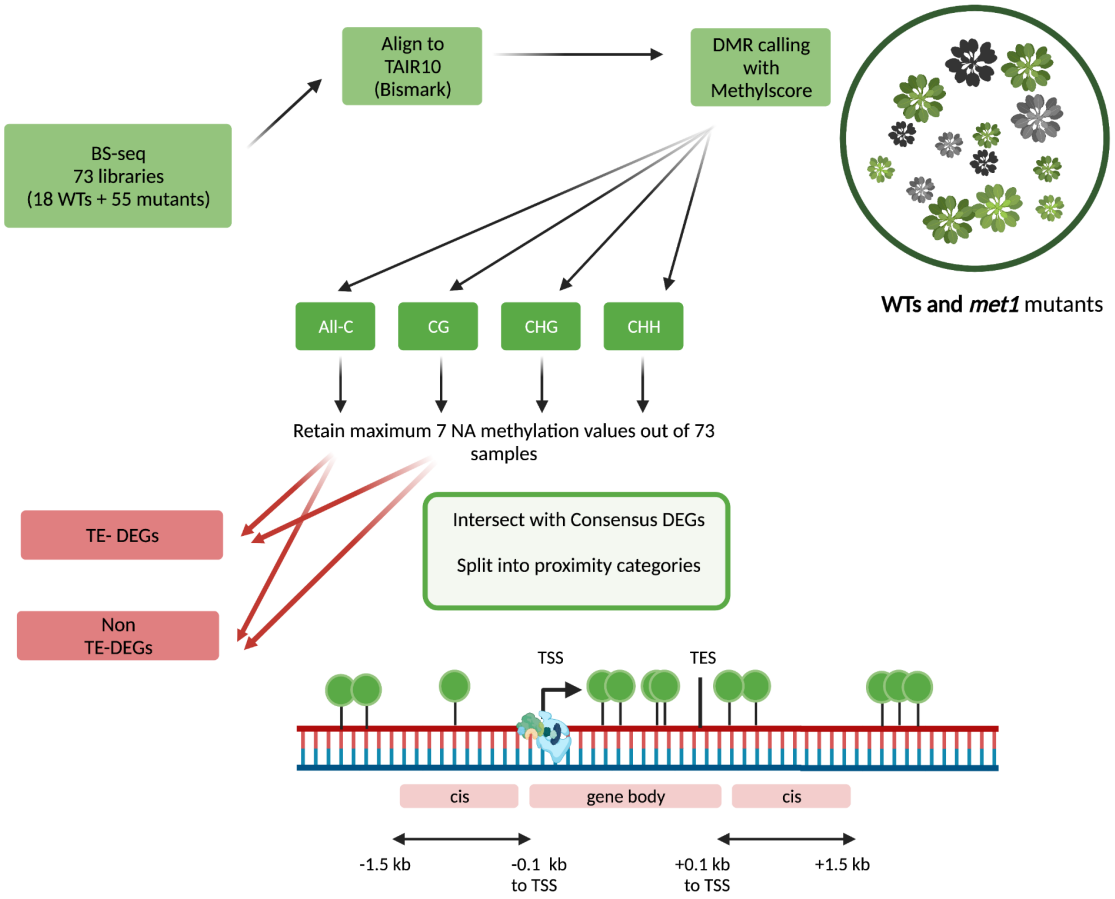


Figure S15: Diagram of generating DMRs from BS-seq data and intersections with consensus DEGs.

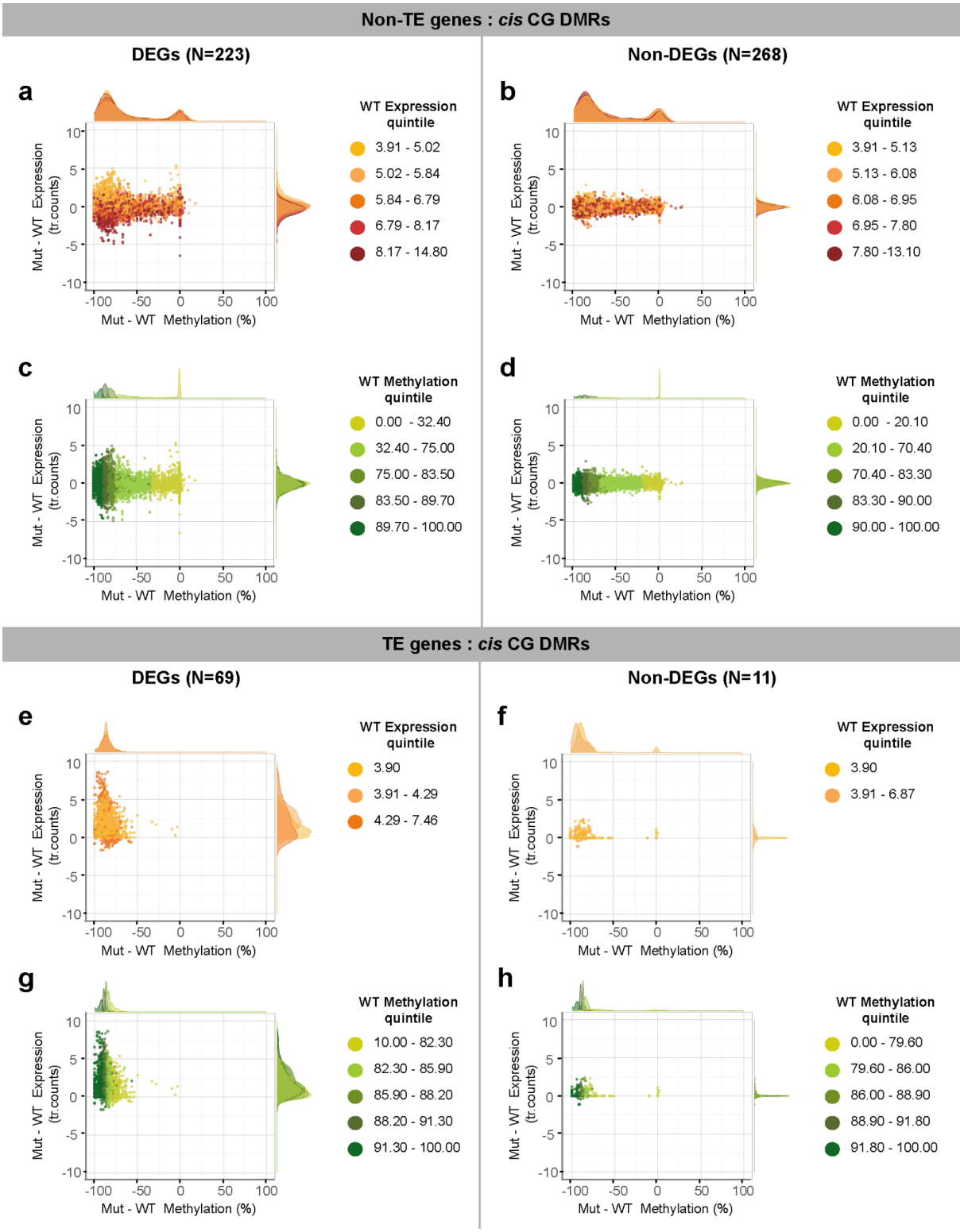


Figure S16: CG-DMRs in *cis* to Non-TE-DEGs (a-d) and TE-DEGs (e-h). Scatter plots showing differences in CG methylation between *met1* mutants and wild-type plants against difference in gene expression. Dots in the scatter plot are colored by wild-type expression quintiles **(a,b,e,f)** and wild-type methylation quintiles **(c,d,g,h)** with x- and y-axis density distributions of each expression/methylation quintile. Expression levels are represented as transformed read counts and methylation levels are represented as %CG methylation in CG-DMRs.

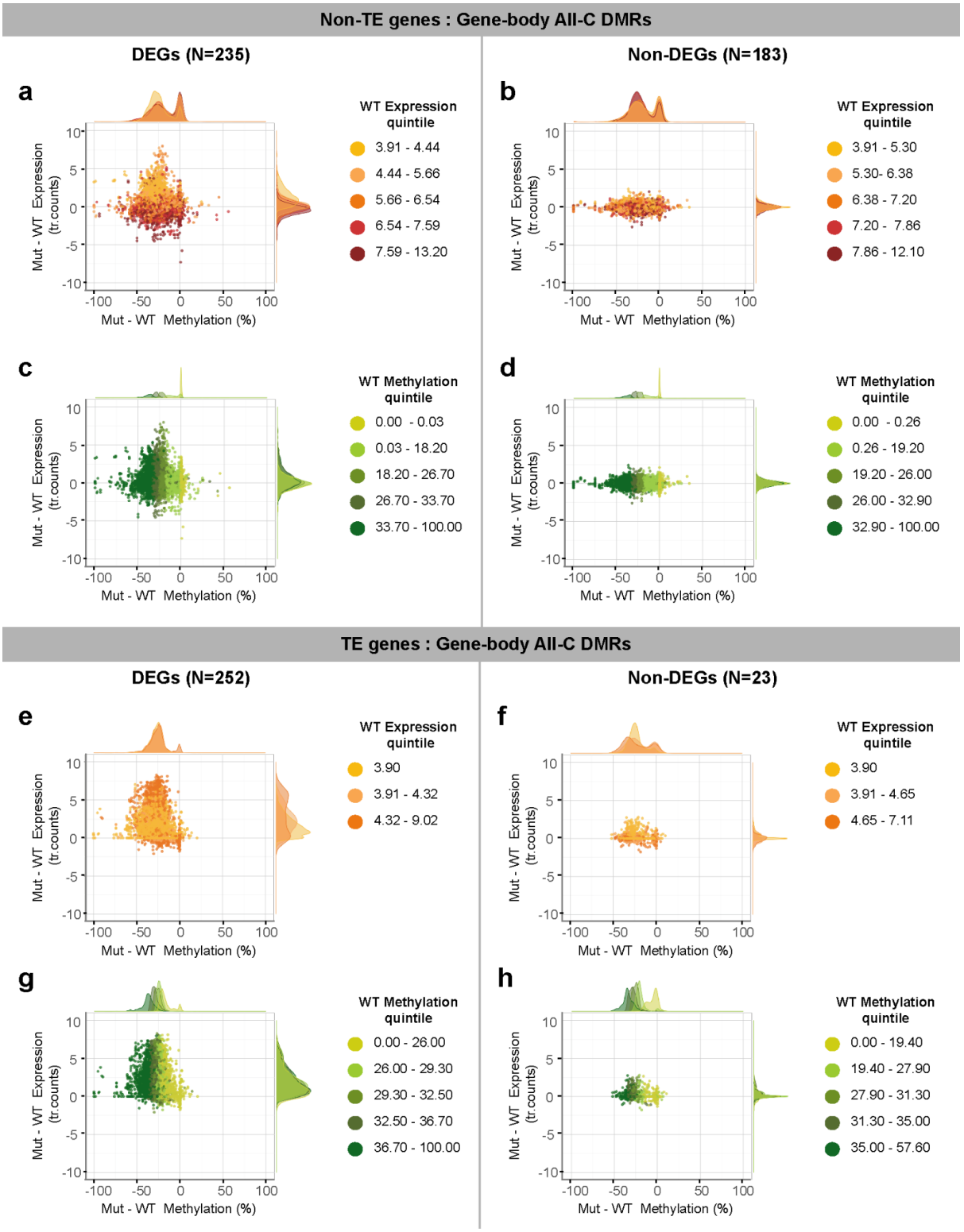


Figure S17: All-C-DMRs in gene-bodies of Non-TE-DEGs (a-d) and TE-DEGs (e-h). Scatter plots showing differences in methylation (all contexts) between *met1* mutants and wild-type plants against differences in gene expression. Dots in the scatter plot are colored by wild-type expression quintiles **(a,b,e,f)** and wild-type methylation quintiles **(c,d,g,h)** with x- and y-axis density distributions of each expression/methylation quintile. Expression levels are represented as transformed read counts and methylation levels are represented as % methylation in All-C-DMRs (methylation in all contexts).

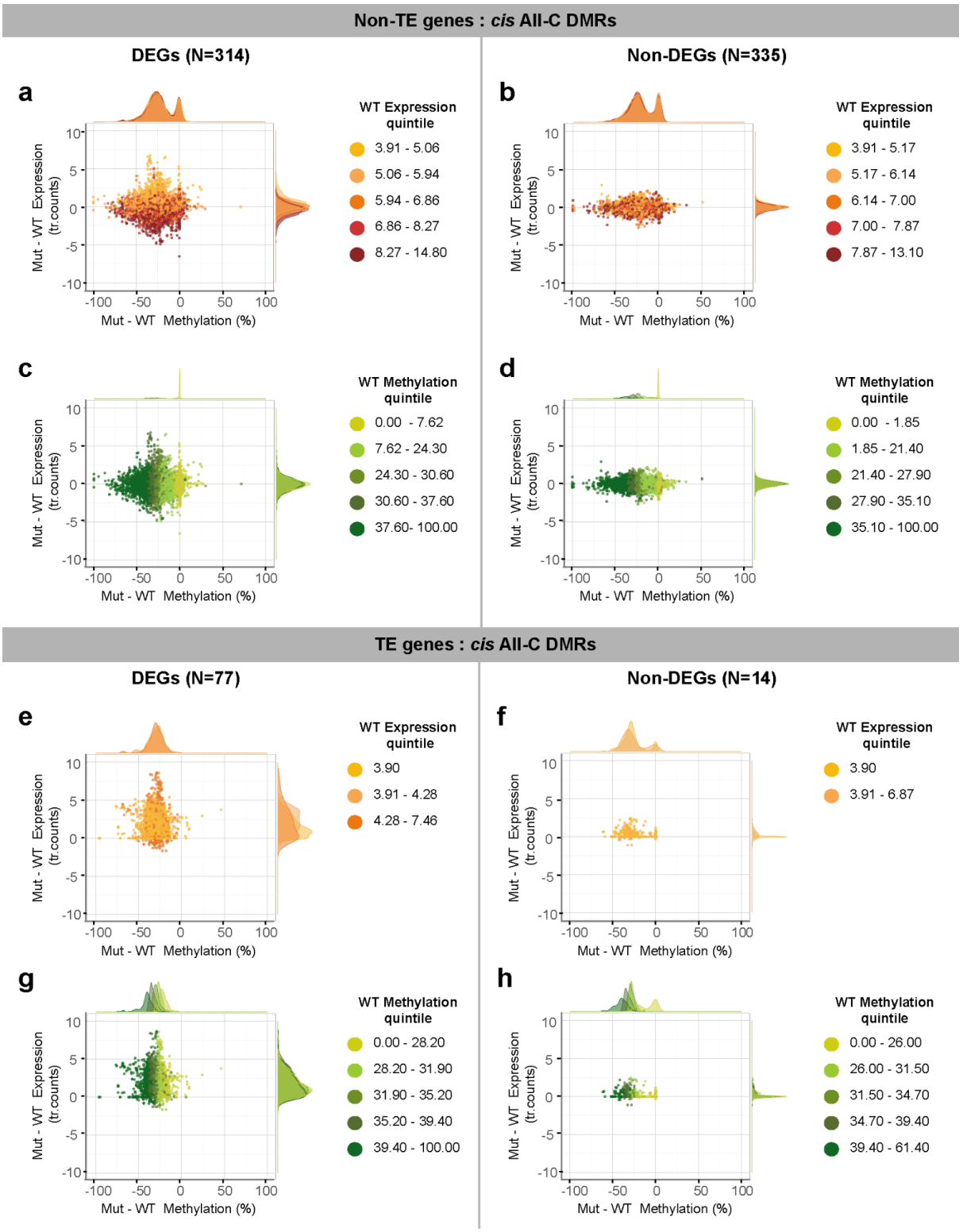


Figure S18: All-C-DMRs in *cis* to Non-TE-DEGs (a-d) and TE-DEGs (e-h). Scatter plots showing differences in methylation (all contexts) between *met1* mutants and wild-type plants against differences in gene expression. Dots in the scatter plot are colored by wild-type expression quintiles **(a,b,e,f)** and wild-type methylation quintiles **(c,d,g,h)** with x- and y-axis

density distributions of each expression/methylation quintile. Expression levels are represented as transformed read counts and methylation levels are represented as % methylation in All-C-DMRs (methylation in all contexts).

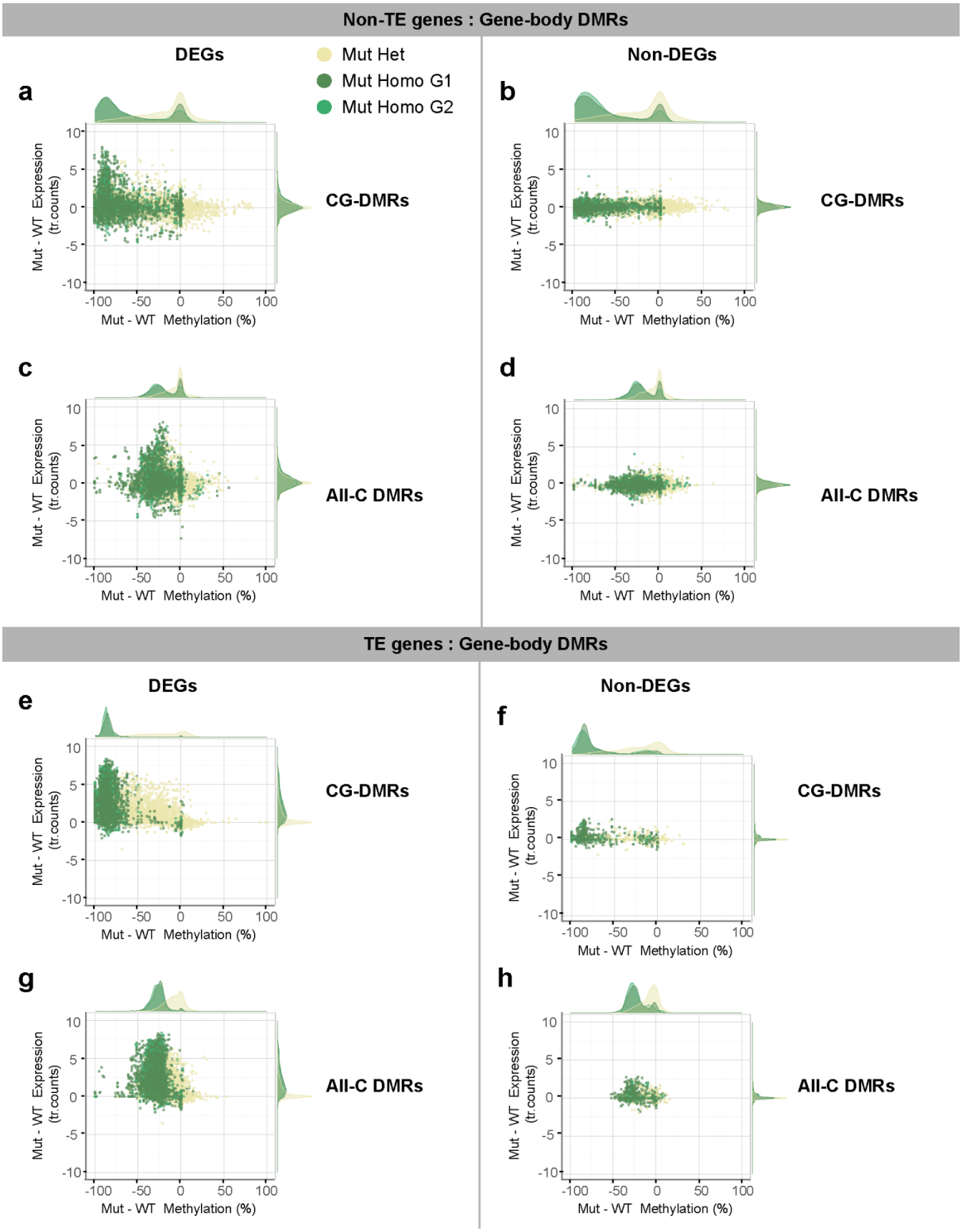


Figure S19: DMRs in gene-bodies of Non-TE-genes (a-d) and TE-genes (e-h). Scatter plots showing differences in CG methylation between *met1* mutants and wild-type plants against differences in gene expression. Dots in the scatter plot are colored by genotype of *met1* mutants; wild-types ('WT'), heterozygous *met1* mutants ('Mut Het'), first generation homozygous *met1* mutants ('Mut Homo G1') and second generation homozygous *met1* mutants ('Mut Homo G2') with x- and y-axis density distributions of each genotype. Expression levels are represented as

transformed read counts (Methods) and methylation levels are represented as % methylation (all-contexts for All-C-DMRs and CG for CG-DMRs).

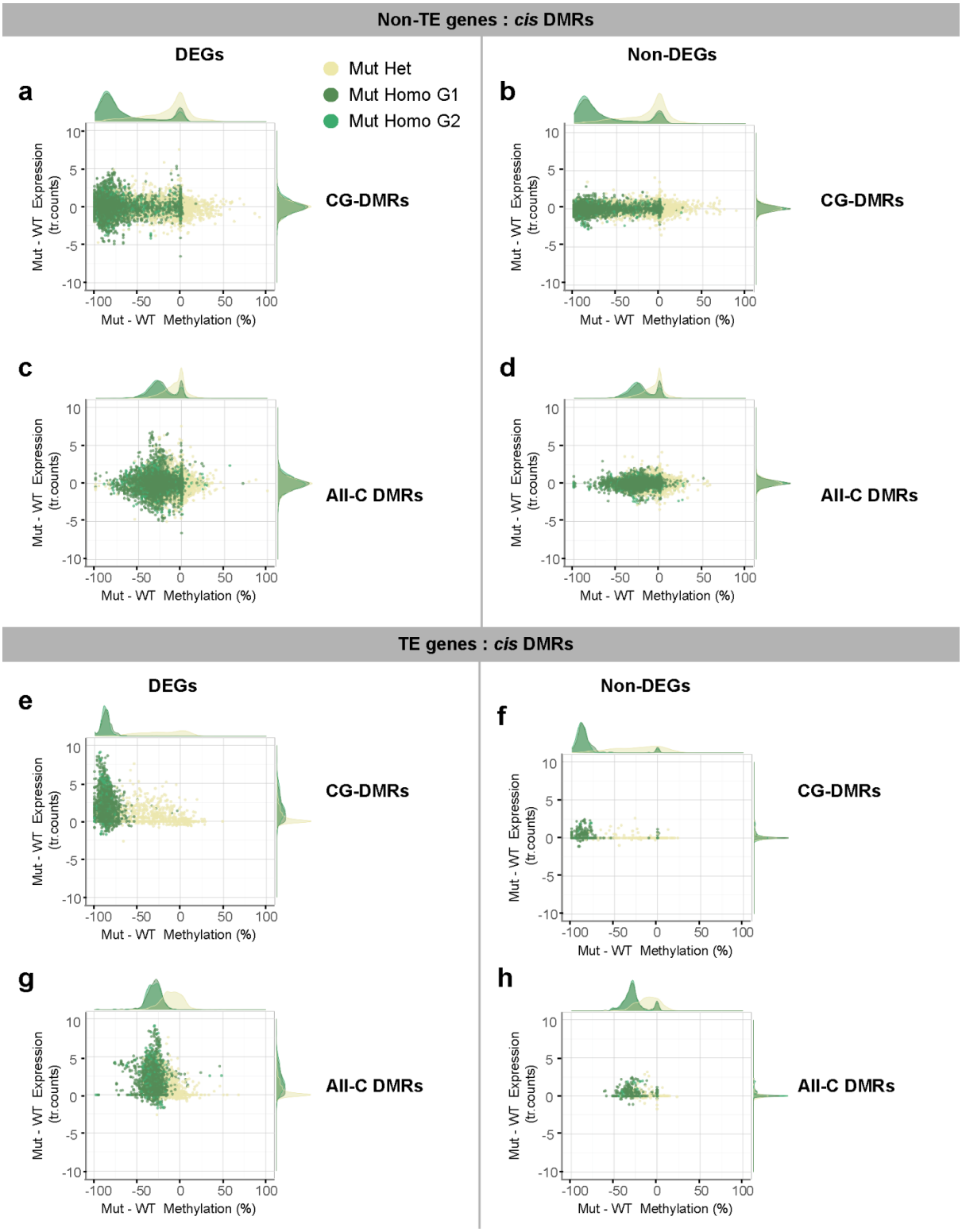


Figure S20: DMRs in *cis* to Non-TE-genes (a-d) and TE-genes (e-h). Scatter plots showing differences in CG methylation between *met1* mutants and wildtype plants against differences in gene expression. Dots in the scatter plot are colored by genotype of *met1* mutants ; wild-types ('WT'), heterozygous *met1* mutants ('Mut Het'), first generation homozygous *met1* mutants ('Mut Homo G1') and second generation homozygous *met1* mutants ('Mut Homo G2') with x- and y-axis density distributions of each genotype. Expression levels are represented as

transformed read counts and methylation levels are represented as % methylation (all contexts for All-C-DMRs and CG for CG-DMRs).

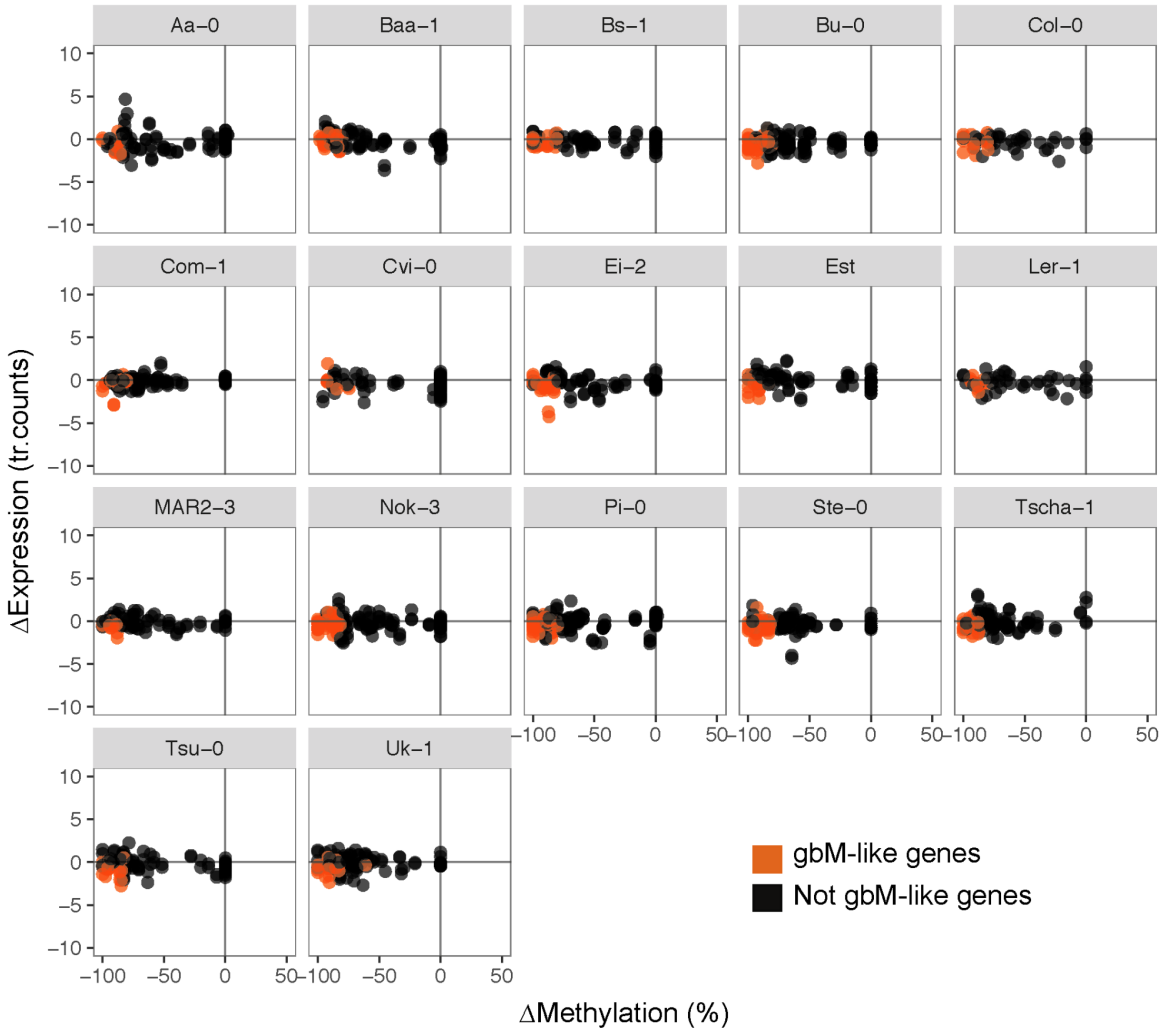


Figure S21: Methylation changes and associated gene expression changes for 51 gbM-like genes across 17 accessions. Orange colored dots represent gbM-like genes and black dots represent the same genes which are not gbM-like in other accessions. ΔMethylation represents *met1* - wild-type methylation, measured in % CG methylation level. ΔExpression represents *met1* - wild-type gene expression levels, measured in transformed read counts.

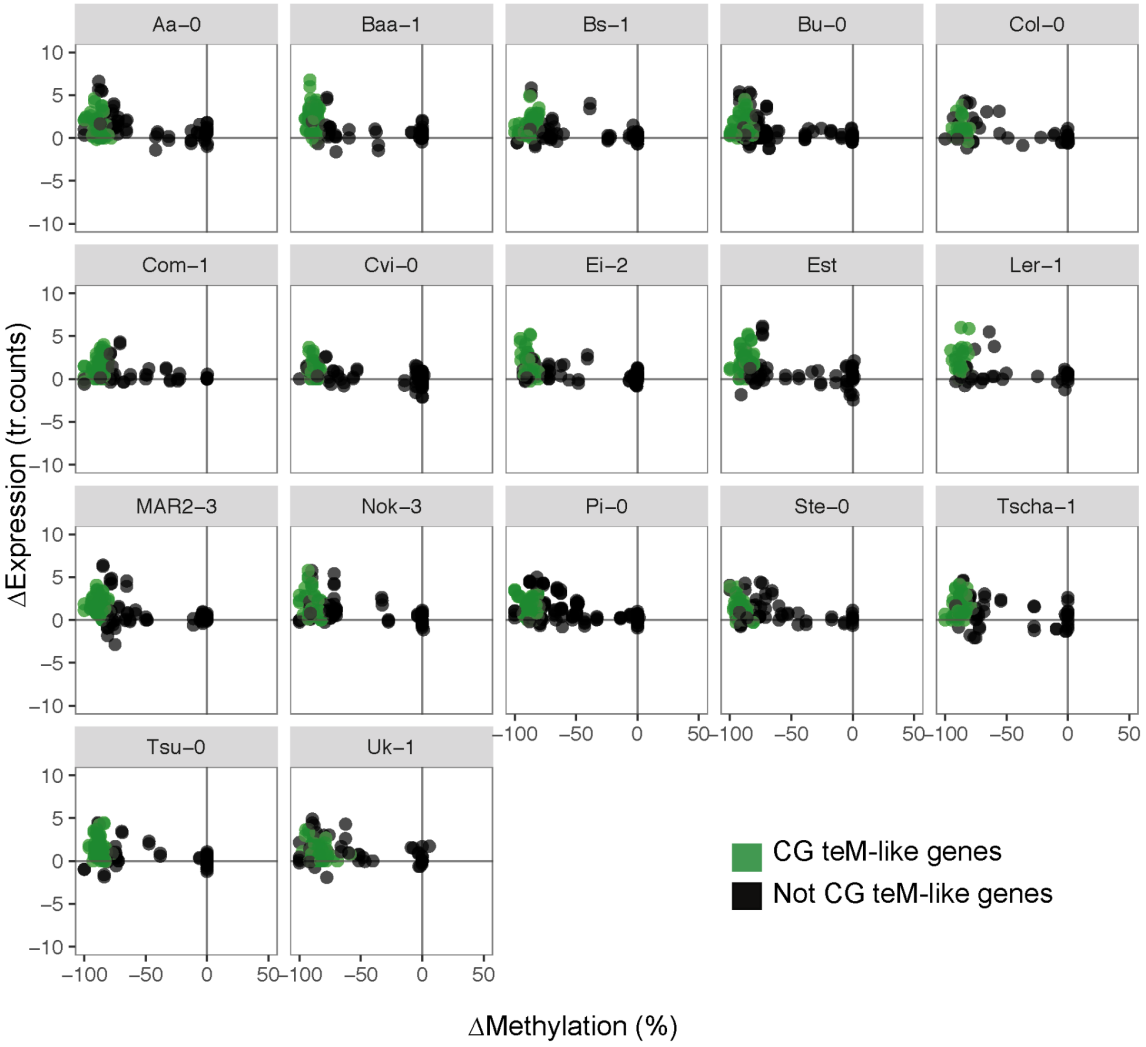


Figure S22: Methylation changes and associated gene expression changes for 59 CG teM-like genes across 17 accessions. Green colored dots represent CG teM-like genes and black dots represent the same genes which are not CG teM-like in other accessions. ΔMethylation represents *met1* - wild-type methylation, measured in % CG methylation level. ΔExpression represents *met1* - wild-type gene expression levels, measured in transformed read counts.

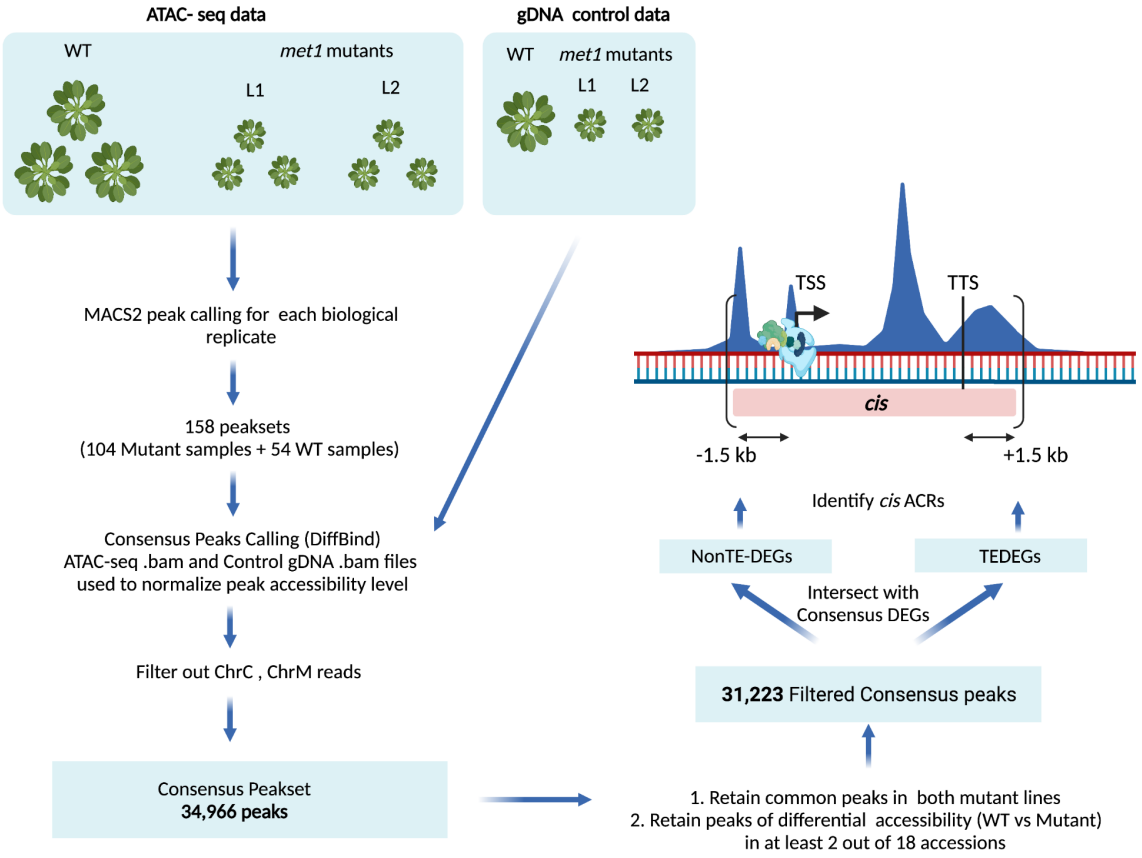


Figure S23: Diagram of ATAC-seq processing, generation of consensus peaks for ACRs and intersections with DEGs.

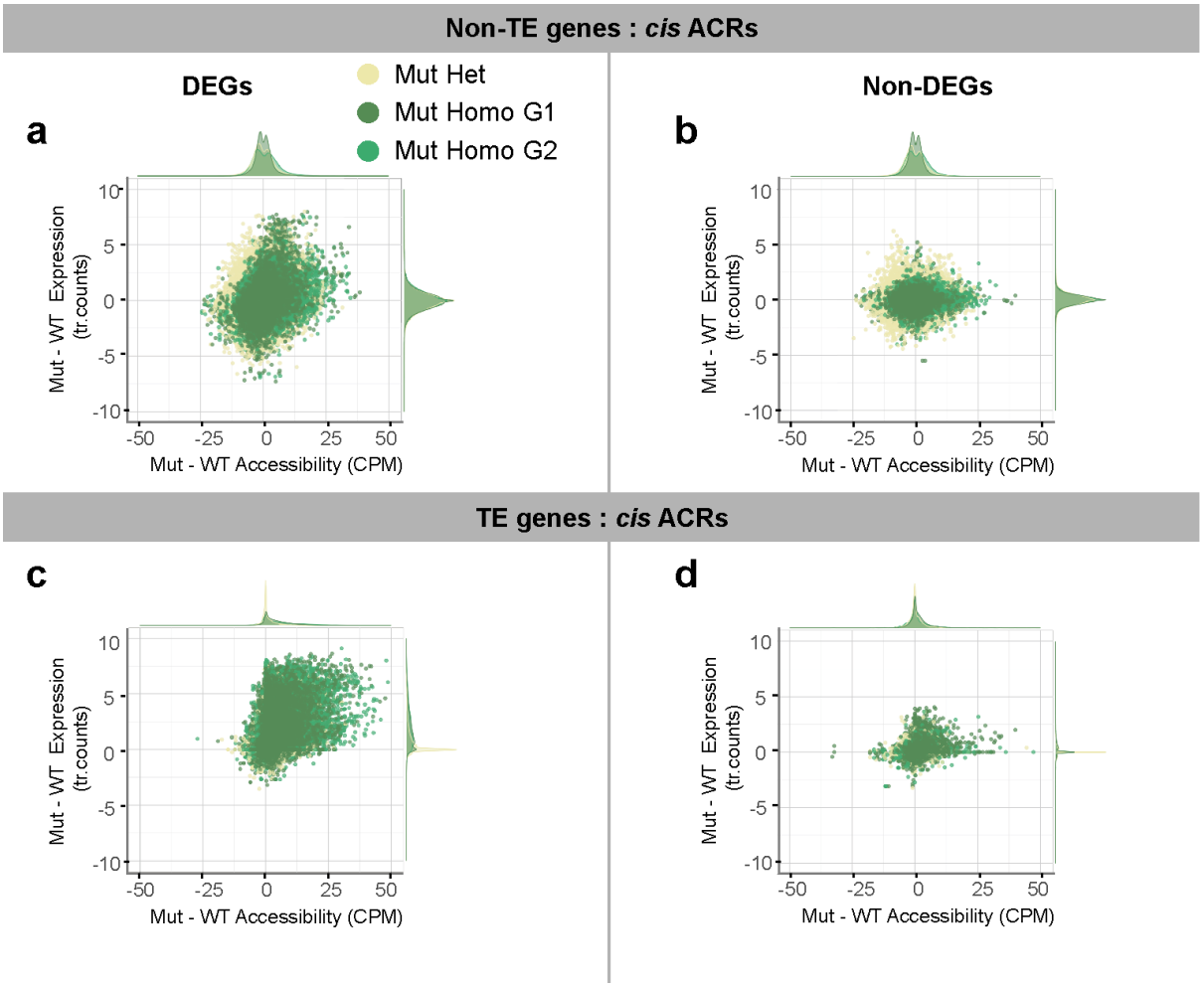


Figure S24: ACRs in *cis* to TE-genes (c,d) and Non-TE genes (a,b). Scatter plots showing difference in Chromatin accessibility between *met1* mutants and wild-type plants against differences in gene expression. Dots in the scatter plot are colored by genotype of *met1* mutants; wild-types ('WT'), heterozygous *met1* mutants ('Mut Het'), first generation homozygous *met1* mutants ('Mut Homo G1') and second generation homozygous *met1* mutants ('Mut Homo G2') with x- and y-axis density distributions of each genotype. Expression levels are represented as transformed read counts and accessibility levels are represented as TMM normalized values in counts per million (CPM).

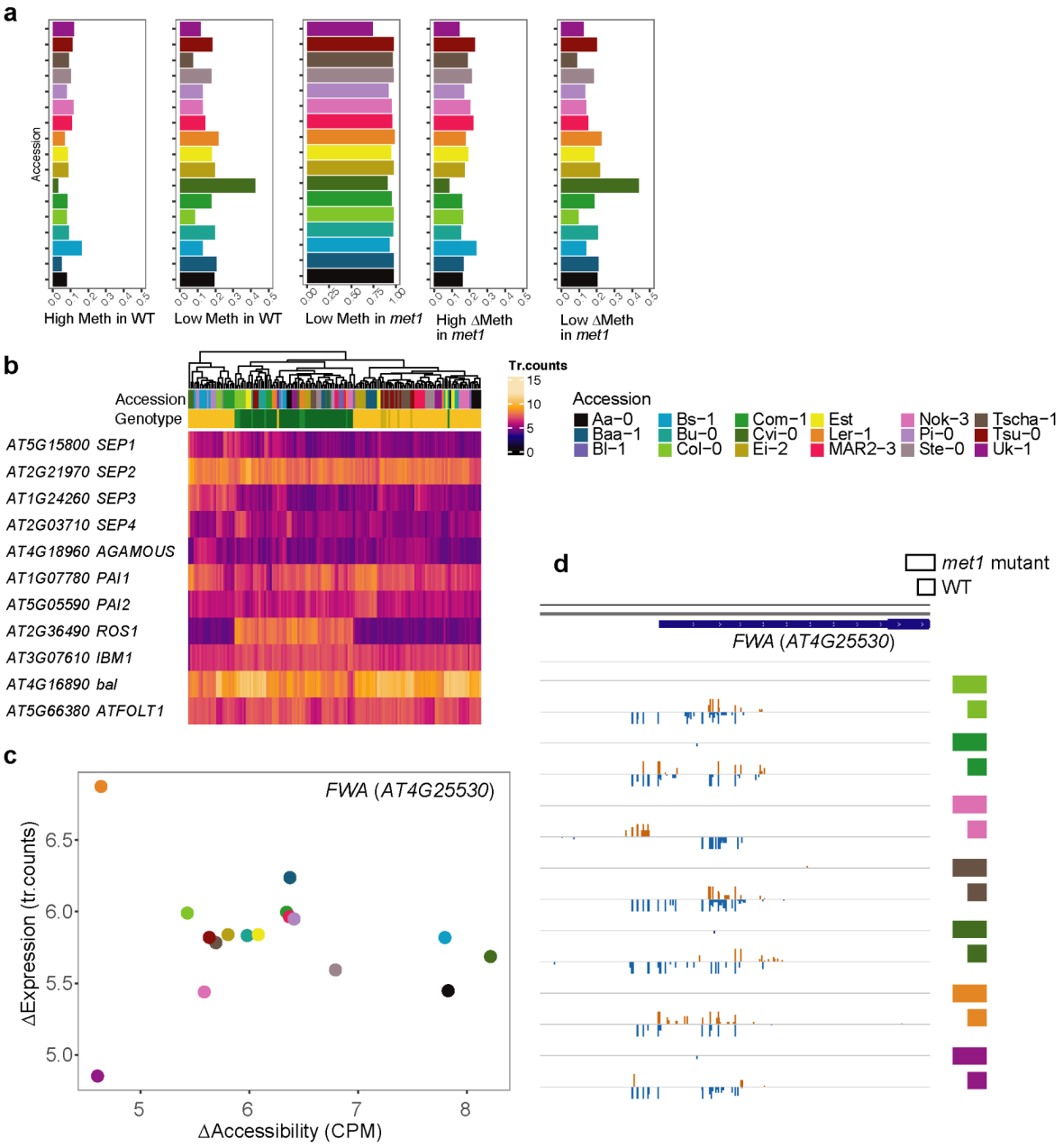


Figure S25: Variable epigenetic states of Non-TE genes across accessions. (a) Five panels showing fraction of genes in 17 accessions with (i) high CG methylation in wild type, (ii) low CG methylation in wild type, (iii) low CG methylation in *met1*, (iv) large methylation change in *met1*, (v) limited methylation change in *met1.* Colors represent different accessions. Low methylation is defined as CG methylation <= 10%, and high methylation is defined as CG methylation >= 90%. **(b)** Heatmap of transformed read counts at 11 epialleles across 158 RNA-seq libraries. The libraries are colored by accession-of-origin and genotype. **(c)** Scatterplot showing relationship between changes in accessibility and expression at the *FWA* locus. Accessibility is measured in counts per million (CPM) and expression is measured by transformed read counts. **(d)** Genome

browser screenshot of methylated cytosines (all contexts) at the *FWA* locus for *met1* mutants and wild-type plants of seven accessions.

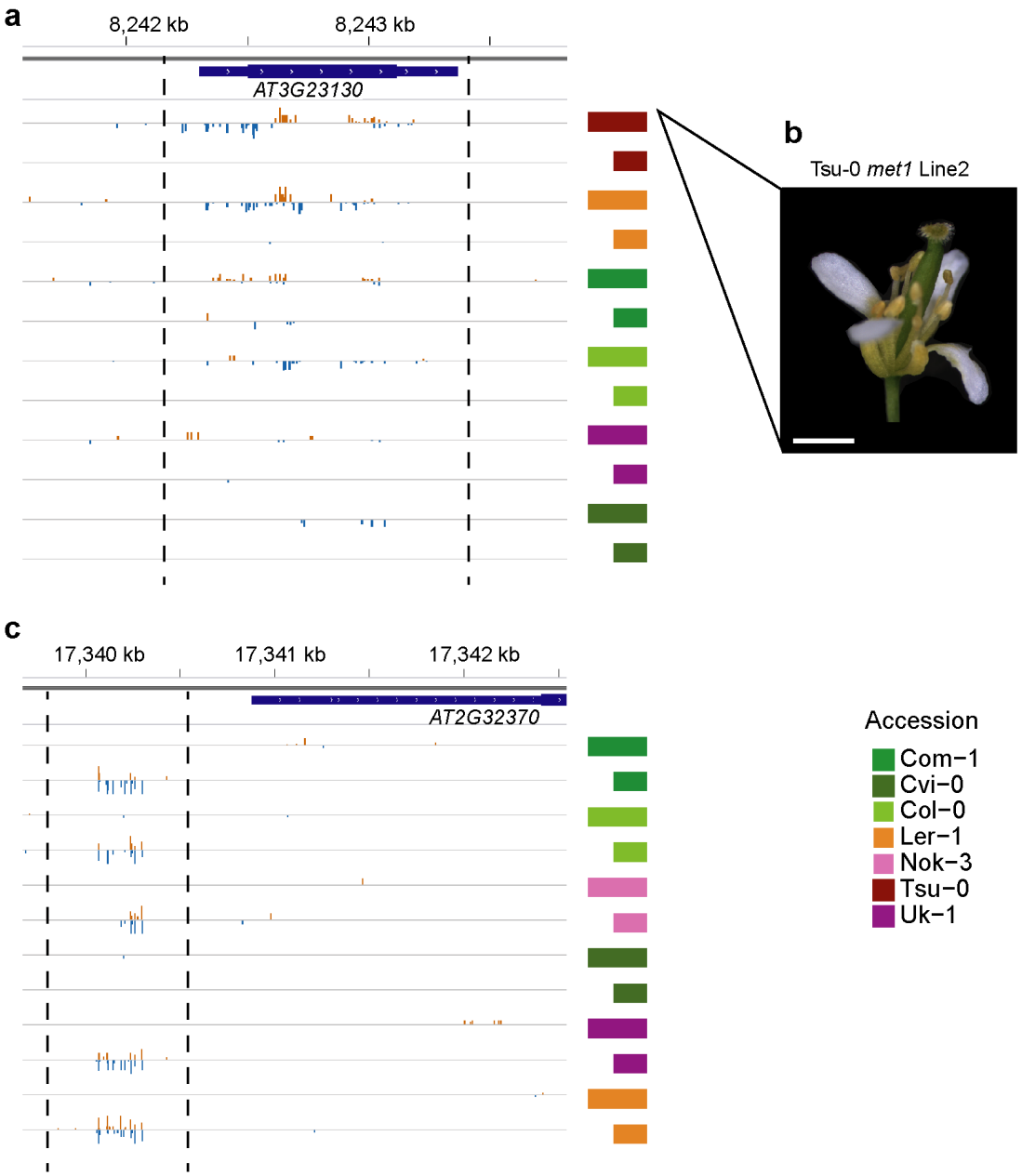


Figure S26: Variation in methylation levels at two epialleles in *met1* mutants and wild-type plants across a subset of accessions. (a) Genome browser view of methylated cytosines (all contexts) at the *SUP* (*AT3G23130*) locus. Gain of methylation in the gene body of *SUP* silences the gene and results in the formation of additional stamens[46]. **(b)** representative image of a Tsu-0 *met1* mutant flower with nine stamens. **(c)** Genome browser view of methylated cytosines (all contexts) at the *HDG3 (AT2G32370)* locus.

SUPPLEMENTAL TABLES

S1 : *MET1* (AT5G49160) CRISPR gRNA design

Target sequence	TTGCTGACTATTTCCGGCCA
Coordinates	Chr5: 19936752-19936771
Strand	Negative (-)
PAM	AGG

S2 : List of accessions used in this study

1001 ID	Accession name	Country	Latitude	Longitude	Group
6909	Col-0	USA	38,3	-92,3	germany
6932	Ler-1	GER	47,984	10,8719	admixed
6911	Cvi-0	CPV	15,1111	-23,6167	relict
159	MAR2-3	FRA	47,35	3,93333	western_europe
6915	Ei-2	GER	50,3	6,3	germany
6945	Nok-3	NED	52,24	4,45	germany
7000	Aa-0	GER	50,9167	9,57073	germany
7002	Baa-1	NED	51,3333	6,1	germany
7003	Bs-1	SUI	47,5	7,5	admixed
7025	Bl-1	ITA	44,5041	11,3396	central_europe
7036	Bu-0	GER	50,5	9,5	germany
7092	Com-1	FRA	49,416	2,823	western_europe
7298	Pi-0	AUT	47,04	10,51	central_europe
7346	Ste-0	GER	52,6058	11,8558	south_sweden
7372	Tscha-1	AUT	47,0748	9,9042	central_europe
7373	Tsu-0	JPN	34,43	136,31	admixed
7378	Uk-1	GER	48,0333	7,7667	admixed
7127	Est-1	EST	58,6656	24,9871	admixed

S3 : PCR Oligonucleotides for Genotyping

Primer sequences (5'--> 3')	Amplicon size	Purpose
TGTTGTGATTAATTGCAGGGCT AGCCCAAATGAAAGCTCGT	649 bp	Genotyping by Sanger sequencing
TCTTGGAGTAAAGTTCAGTGTGA TCGAGAAGGGAAGCCAAAGT	152 bp	Genotyping by amplicon sequencing

S4 : List of frameshift mutations in each *met1* mutant line

Frameshift mutation site (in *met1* mutants)	TTG[I]CTGACTATTTCCGGCCA (5'--> 3')
	[I] marks the preferred site of the mutations

Heterozygous parental line	**Accession**	**Genotyped Mutation**	**Homozygous Generation**	**Notes**
Line 1	Tscha-1	A insertion	first	
Line 2	Tscha-1	A insertion	first	
Line 2	Tscha-1	A insertion	second	
Line 3	Tscha-1	A insertion	second	mutant line with *ag*-like phenotype
Line 1	Tsu-0	C insertion	first	
Line 2	Tsu-0	T insertion	first	
Line 2	Tsu-0	T insertion	second	
Line 1	Com-1	A insertion	first	
Line 2	Com-1	C insertion	first	
Line 1	Est	A insertion	first	
Line 2	Est	C insertion	first	

Line 1	Bu-0	C insertion	first	tetraploid, skewed heterozygous plants
Line 2	Bu-0	T insertion / A insertion (tri-allelic)	first	tetraploid, skewed heterozygous plants
Line 1	Col-0	A insertion	first	
Line 2	Col-0	A insertion	first	skewed heterozygous plants
Line 1	MAR2-3	A insertion	first	
Line 2	MAR2-3	C insertion	first	
Line 2	Cvi-0	T insertion	first	
Line 1	Ei-2	C insertion	first	
Line 2	Ei-2	C insertion	first	
Line 1	Aa-0	GCT deletion	first	
Line 2	Aa-0	T insertion	first	
Line 1	Bs-1	T insertion	first	
Line 2	Bs-1	C insertion / T substitution (bi-allelic)	first	
Line 1	Pi-0	AA insertion, C-->A subst	first	skewed heterozygous plants
Line 2	Pi-0	T insertion	first	
Line 1	Ste-0	A insertion	first	
Line 2	Ste-0	C insertion / A insertion (tri-allelic)	first	skewed heterozygous plants
Line 1	Uk-1	C insertion	first	
Line 2	Uk-1	T insertion	first	

Line 1	Nok-3	A insertion	first	
Line 2	Nok-3	A insertion	first	
Line 2	Baa-1	A insertion	first	
Line 3	Baa-1	C insertion	first	
Line 2	Ler-1	C deletion	first	
Line 3	Ler-1	A insertion	first	
Line 2	Bl-1	C insertion	first	skewed heterozygous plants

EXTENDED METHODS

Generation of consensus ATAC-seq peaks using DiffBind:

The following commands were used in R using the package "DiffBind" (where 'dataset_Acc' refers to the samplesheet containing details of MACS2-called peaks from individual libraries)

```
library(DiffBind)
dataset_Acc <- dba(sampleSheet="./samplesheet_forDiffBind.csv")
dataset_Acc_consensus <- dba.peakset(dataset_Acc, consensus=-DBA_REPLICATE)
dataset_Acc_consensus<- dba(dataset_Acc_consensus,
mask=dataset_Acc_consensus$masks$Consensus, minOverlap=1)
consensus_peaks <-  dba.peakset(dataset_Acc_consensus, bRetrieve=TRUE)
dataset_Acc<-  dba.count(dataset_Acc,peaks=consensus_peaks,
score=DBA_SCORE_TMM_MINUS_FULL_CPM)
normCounts<-  dba.peakset(dataset_Acc, bRetrieve=TRUE, DataType=DBA_DATA_FRAME)
```

Correlations between methylome, transcriptome and chromatin accessibility

To intersect positions of each consensus DEG set (TE-DEG/Non-TE-DEG) with DMRs, we identified the five closest DMRs in each context (CG DMRs, All-C DMRs) to each DEG. Based on their proximity to the DEG, each of these hits were classified as 'gene-body' (distances within 100bp upstream of the TSS and 100bp downstream of the TTS), 'cis upstream/downstream' (within 1.5kb upstream/downstream of the gene body) and 'trans' if not falling within the above classes. All DEGs with DMR hits under the 'trans' category were filtered out. DEGs with unique or multiple 'gene-body' DMRs were classified as 'GB' , while DEGs with unique/multiple DMRs in gene-body and *cis* were classified as 'cis'. Subsequently, gene expression counts and methylation counts for the corresponding DMRs were extracted for these DEGs. Since many DEGs had multiple DMRs associated with them, we only retained DMRs which showed the

highest difference in methylation level between mutant and WT for each mutant genotype, thereby aiming to represent only the strongest methylation signals that could explain gene expression differences.

For intersecting consensus DEGs with ACRs, we used a similar approach, where the closest five ACRs to each DEG were identified. After filtering out DEGs with ACRs in 'trans', we retained all remaining DEGs (even with multiple 'gene-body' and 'cis' ACRs) in a single category called 'cis'. For each DEG, only ACRs with the highest difference in accessibility between mutant and WT for each mutant genotype were retained.

Intersects between DEGs and DMRs:

Consensus DEGs (from all accessions) were first split as 5731 Non-TEDEGs and 1401 TEDEGs. Next, each set was intersected with positions of consensus DMRs (generated from all samples) in two contexts, CG-DMRs and All-C -DMRs, using *bedtools closest*, to find the top 5 closest DMRs to each DEG. Based on their proximity to the DEG, each of these hits were classified as gene-body (closest distance >=-100 bp and <=100 bp), cis_upstream (closest distance <=-100 bp and >-1.5 kb), cis downstream (closest distance >=100 bp and <=1.5 kb) and trans (beyond 1.5 kb upstream or downstream).

#Example of command used (to find 5 closest hits of DMRs to DEGs)
bedtools closest -a Consensus_NonTEDEGs_coord.tab -b CG_DMRs_coord.bed -D a -k 5 > Consensus_NonTEDEGs_closestk5_CGDMRs.bed

All DMR hits under the 'trans' category were filtered out. Next, the number of duplicate DMR hits for each gene were counted. Each gene was further classified into the following categories: GB_unique (single DMR hit in gene body), GB_multi (multiple DMR hits in gene body), cis_upstream (single DMR hit in cis upstream), cis_downstream (single DMR hit in cis downstream), cis_multi (multiple cis DMRs) , GBandcis (multiple DMRs in genebody and cis).

Genes classified as GB_unique and GB_multi were pooled together in a "GB_all" category. All remaining categories were pooled as a "multiple_cis_regulatory" category. Gene expression counts were obtained for genes in each category (for 158 samples and reduced to 73 samples to match the BS-seq dataset. Similarly, methylation levels for the corresponding DMR hit in each gene was also obtained for 73 samples.

Next, each gene was scanned to examine duplicate DMR hits and their methylation values. For each of the 73 samples, the DMR carrying the highest difference between wild-type and mutant methylation levels was retained. These genes were subsequently plotted based on these methylation differences, gene expression differences.

Intersects between DEGs and ACRs :

Each set of consensus DEGs (TE-DEGs and Non-TE-DEGs) was intersected with positions of Consensus ACRs (generated from all samples) using *bedtools closest*, to find the top 5 closest ACRs to each DEG. Based on their proximity to the DEG, each of these hits were classified as

gene-body (closest distance >=-100 bp and <=100 bp), cis_upstream (closest distance <=-100 bp and >-1.5 kb), cis downstream (closest distance >=100 bp and <=1.5 kb) and trans (beyond 1.5 kb upstream or downstream).

#Example of command used (to find 5 closest hits of ACRs to DEGs)
bedtools closest -a Consensus_NonTEDEGs_coord.tab -b ACRs_coord.bed -D a -k 5 > Consensus_NonTEDEGs_closestk5_ACRs.bed

All ACR hits under the 'trans' category were filtered out. All other genes and corresponding ACR hits were pooled as a "multiple_cis_regulatory" category. Gene expression counts were obtained for genes in each category (for 158 ATAC-seq samples and reduced to 73 samples to match the Methylation dataset. Similarly, ATAC-seq accessibility levels (TMM) for the corresponding ACR hit in each gene was also obtained for 158 samples and reduced to 73 samples.

Next, each gene was scanned to examine duplicate ACR hits and their accessibility values. For each of the 73 samples, the ACR carrying the highest difference between wild-type and mutant accessibility levels was retained. These genes were subsequently plotted based on these accessibility differences, gene expression differences.

REFERENCES

1. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204 (2010).
2. Zhang, H., Lang, Z. & Zhu, J.-K. Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* 19, 489–506 (2018).
3. 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, 8449–8454 (1996).
4. Jacobsen, S. E., Sakai, H., Finnegan, E. J., Cao, X. & Meyerowitz, E. M. Ectopic hypermethylation of flower-specific genes in Arabidopsis. *Curr. Biol.* 10, 179–186 (2000).
5. Kankel, M. W. *et al.* Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics* 163, 1109–1122 (2003).
6. Saze, H., Mittelsten Scheid, O. & Paszkowski, J. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat. Genet.* 34, 65–69 (2003).
7. Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C. & Paszkowski, J. Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell* 130, 851–862 (2007).
8. Stroud, H., Greenberg, M. V. C., 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).
9. Tariq, M. *et al.* Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8823–8827 (2003).
10. Deleris, A. *et al.* Loss of the DNA Methyltransferase MET1 Induces H3K9 Hypermethylation at PcG Target Genes and Redistribution of H3K27 Trimethylation to Transposons in Arabidopsis thaliana. *PLoS Genetics* vol. 8 e1003062 (2012).

11. Soppe, W. J. J. *et al.* DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. *EMBO J.* 21, 6549–6559 (2002).
12. Zhong, Z. *et al.* DNA methylation-linked chromatin accessibility affects genomic architecture in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 118, (2021).
13. 1001 Genomes Consortium. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell* 166, 481–491 (2016).
14. Kawakatsu, T. *et al.* Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. *Cell* 166, 492–505 (2016).
15. Quadrana, L. *et al.* The Arabidopsis thaliana mobilome and its impact at the species level. *eLife* vol. 5 (2016).
16. Baduel, P. *et al.* Genetic and environmental modulation of transposition shapes the evolutionary potential of Arabidopsis thaliana. *Genome Biol.* 22, 138 (2021).
17. Dubin, M. J. *et al.* DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. *Elife* 4, e05255 (2015).
18. Sasaki, E., Kawakatsu, T., Ecker, J. R. & Nordborg, M. Common alleles of CMT2 and NRPE1 are major determinants of CHH methylation variation in Arabidopsis thaliana. *PLoS Genet.* 15, e1008492 (2019).
19. Shahzad, Z., Moore, J. D. & Zilberman, D. Gene body methylation mediates epigenetic inheritance of plant traits. *bioRxiv* (2021).
20. Zhang, Y., Wendte, J. M., Ji, L. & Schmitz, R. J. Natural variation in DNA methylation homeostasis and the emergence of epialleles. *Proc. Natl. Acad. Sci. U. S. A.* 117, 4874–4884 (2020).
21. Wu, R. *et al.* An efficient CRISPR vector toolbox for engineering large deletions in Arabidopsis thaliana. *Plant Methods* vol. 14 (2018).
22. FitzGerald, J., Luo, M., Chaudhury, A. & Berger, F. DNA methylation causes predominant maternal controls of plant embryo growth. *PLoS One* 3, e2298 (2008).

23. Lippman, Z. & Martienssen, R. The role of RNA interference in heterochromatic silencing. *Nature* 431, 364–370 (2004).

24. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285 (2007).

25. Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126, 1189–1201 (2006).

26. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* vol. 452 215–219 (2008).

27. Mirouze, M. *et al.* Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* 461, 427–430 (2009).

28. Lamesch, P. *et al.* The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–10 (2012).

29. Oberlin, S., Sarazin, A., Chevalier, C., Voinnet, O. & Marí-Ordóñez, A. A genome-wide transcriptome and translatome analysis of Arabidopsis transposons identifies a unique and conserved genome expression strategy for Ty1/Copia retroelements. *Genome Res.* 27, 1549–1562 (2017).

30. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133, 523–536 (2008).

31. Marí-Ordóñez, A. *et al.* Reconstructing de novo silencing of an active plant retrotransposon. *Nat. Genet.* 45, 1029–1039 (2013).

32. Kato, M., Miura, A., Bender, J., Jacobsen, S. E. & Kakutani, T. Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. *Curr. Biol.* 13, 421–426 (2003).

33. Le, N. T. *et al.* Epigenetic regulation of spurious transcription initiation in Arabidopsis. *Nat. Commun.* 11, 3224 (2020).

34. Vu, T. M. *et al.* RNA-directed DNA methylation regulates parental genomic imprinting at several loci in Arabidopsis. *Development* 140, 2953–2960 (2013).

35. Hsieh, T.-F. *et al.* Regulation of imprinted gene expression in Arabidopsis endosperm. *Proc. Natl. Acad. Sci. U. S. A.* 108, 1755–1762 (2011).

36. 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).

37. Kinoshita, Y. *et al.* Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. *Plant J.* 49, 38–45 (2007).

38. Henderson, I. R. & Jacobsen, S. E. Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev.* 22, 1597–1606 (2008).

39. Tran, R. K. *et al.* DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes. *Curr. Biol.* 15, 154–159 (2005).

40. Reynoso, M. A. *et al.* Evolutionary flexibility in flooding response circuitry in angiosperms. *Science* 365, 1291–1295 (2019).

41. Marand, A. P., Chen, Z., Gallavotti, A. & Schmitz, R. J. A cis-regulatory atlas in maize at single-cell resolution. *Cell* 184, 3041–3055.e21 (2021).

42. Shu, H., Wildhaber, T., Siretskiy, A., Gruissem, W. & Hennig, L. Distinct modes of DNA accessibility in plant chromatin. *Nat. Commun.* 3, 1281 (2012).

43. Bender, J. & Fink, G. R. Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. *Cell* 83, 725–734 (1995).

44. Rigal, M., Kevei, Z., Pélissier, T. & Mathieu, O. DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns. *EMBO J.* 31, 2981–2993 (2012).

45. Stokes, T. L., Kunkel, B. N. & Richards, E. J. Epigenetic variation in Arabidopsis disease resistance. *Genes Dev.* 16, 171–182 (2002).

46. Jacobsen, S. E. & Meyerowitz, E. M. Hypermethylated SUPERMAN epigenetic alleles in arabidopsis. *Science* 277, 1100–1103 (1997).

47. Ronemus, M. J., Galbiati, M., Ticknor, C., Chen, J. & Dellaporta, S. L. Demethylation-induced developmental pleiotropy in Arabidopsis. *Science* 273, 654–657 (1996).

48. Julian, R., Tang, K., Xie, S. & Zhu, J. K. Regulatory link between DNA methylation and active demethylation in Arabidopsis. *Proceedings of the* (2015).

49. Williams, B. P., Pignatta, D., Henikoff, S. & Gehring, M. Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. *PLoS Genet.* 11, e1005142 (2015).

50. Lei, M. *et al.* Regulatory link between DNA methylation and active demethylation in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 112, 3553–3557 (2015).

51. Zhou, P., Lu, Z. & Schmitz, R. J. Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. *Proceedings of the* (2020).

52. Borges, F. *et al.* Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in Arabidopsis. *Nat. Genet.* 50, 186–192 (2018).

53. Pignatta, D. *et al.* Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *eLife* vol. 3 (2014).

54. Pignatta, D., Novitzky, K., Satyaki, P. R. V. & Gehring, M. A variably imprinted epiallele impacts seed development. *PLoS Genet.* 14, e1007469 (2018).

55. Maher, K. A. *et al.* Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules. *Plant Cell* 30, 15–36 (2018).

56. Ricci, W. A. *et al.* Widespread long-range cis-regulatory elements in the maize genome. *Nat Plants* (2019) doi:10.1038/s41477-019-0547-0.

57. Yocca, A. E., Lu, Z., Schmitz, R. J., Freeling, M. & Edger, P. P. Evolution of Conserved Noncoding Sequences in Arabidopsis thaliana. *Mol. Biol. Evol.* 38, 2692–2703 (2021).

58. Tannenbaum, M. *et al.* Regulatory chromatin landscape in Arabidopsis thaliana roots uncovered by coupling INTACT and ATAC-seq. *Plant Methods* 14, 113 (2018).

59. Lu, Z. *et al.* The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat Plants* (2019) doi:10.1038/s41477-019-0548-z.

60. Hufford, M. B. *et al.* De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* 373, 655–662 (2021).

61. Zhou, C. *et al.* Accessible chromatin regions and their functional interrelations with gene transcription and epigenetic modifications in sorghum genome. *Plant Commun* 2, 100140 (2021).

62. O'Malley, R. C. *et al.* Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* 166, 1598 (2016).

63. Alexandre, C. M. *et al.* Complex Relationships between Chromatin Accessibility, Sequence Divergence, and Gene Expression in Arabidopsis thaliana. *Mol. Biol. Evol.* 35, 837–854 (2018).

64. Noshay, J. M., Marand, A. P., Anderson, S. N. & Zhou, P. Cis-regulatory elements within TEs can influence expression of nearby maize genes. *BioRxiv* (2020).

65. Pavlopoulou, A. & Kossida, S. Plant cytosine-5 DNA methyltransferases: structure, function, and molecular evolution. *Genomics* 90, 530–541 (2007).

66. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* 16, 735–743 (1998).

67. LeBlanc, C. *et al.* Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. *Plant J.* 93, 377–386 (2018).

68. Wibowo, A. *et al.* Partial maintenance of organ-specific epigenetic marks during plant asexual reproduction leads to heritable phenotypic variation. *Proc. Natl. Acad. Sci. U. S. A.* 115, E9145–E9152 (2018).

69. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572 (2011).

70. Yaffe, H. *et al.* LogSpin: a simple, economical and fast method for RNA isolation from infected or healthy plants and other eukaryotic tissues. *BMC Res. Notes* 5, 45 (2012).

71. Cambiagno, D. A. *et al.* HASTY modulates miRNA biogenesis by linking pri-miRNA transcription and processing. *Mol. Plant* 14, 426–439 (2021).

72. Symeonidi, E., Regalado, J., Schwab, R. & Weigel, D. CRISPR-finder: A high throughput and cost effective method for identifying successfully edited A. thaliana individuals. *bioRxiv* 2020.06.25.171538 (2020) doi:10.1101/2020.06.25.171538.

73. Loureiro, J., Rodriguez, E., Dolezel, J. & Santos, C. Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Ann. Bot.* 100, 875–888 (2007).

74. Yanofsky, M. F. *et al.* The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. *Nature* 346, 35–39 (1990).