



# DNA sequence properties that predict susceptibility to epiallelic switching

Marco Catoni<sup>1,2</sup> , Jayne Griffiths<sup>1</sup>, Claude Becker<sup>3</sup>, Nicolae Radu Zabet<sup>1,‡</sup>, Carlos Bayon<sup>1</sup>, Mélanie Dapp<sup>2</sup>, Michal Lieberman-Lazarovich<sup>2,†</sup> , Detlef Weigel<sup>3</sup> & Jerzy Paszkowski<sup>1,2,\*</sup>

## Abstract

Transgenerationally heritable epialleles are defined by the stable propagation of alternative transcriptional states through mitotic and meiotic cell cycles. Given that the propagation of DNA methylation at CpG sites, mediated in *Arabidopsis* by MET1, plays a central role in epigenetic inheritance, we examined genomewide DNA methylation in partial and complete loss-of-function *met1* mutants. We interpreted the data in relation to transgenerational epiallelic stability, which allowed us to classify chromosomal targets of epigenetic regulation into (i) single copy and methylated exclusively at CpGs, readily forming epialleles, and (ii) transposon-derived, methylated at all cytosines, which may or may not form epialleles. We provide evidence that DNA sequence features such as density of CpGs and genomic repetitiveness of the loci predispose their susceptibility to epiallelic switching. The importance and predictive power of these genetic features were confirmed by analyses of common epialleles in natural *Arabidopsis* accessions, epigenetic recombinant inbred lines (epiRILs) and also verified in rice.

**Keywords** DNA methylation; epialleles; epigenetic; transcriptional silencing

**Subject Categories** Chromatin, Epigenetics, Genomics & Functional Genomics; Plant Biology

**DOI** 10.15252/embj.201695602 | Received 9 September 2016 | Revised 9 December 2016 | Accepted 9 December 2016 | Published online 9 January 2017

**The EMBO Journal (2017) 36: 617–628**

## Introduction

Transgenerational epigenetic inheritance has been well documented in plants; however, the primary competence that leads to the formation of alternative epialleles at only certain loci is not well understood. It has been reported that maintenance of mCpG (methylated CpG) patterns by MET1 and the chromatin remodeler DDM1 is central for transgenerational epigenetic inheritance (Lippman *et al*, 2003; Reinders *et al*, 2009; Teixeira *et al*, 2009). Importantly,

inactivation of MET1 or DDM1 results in the loss of mCpGs, which is not easily corrected after re-introduction of MET1 and DDM1, although remethylation occurs at certain loci (Reinders *et al*, 2009; Teixeira *et al*, 2009). Thus, chromosomal targets of CpG methylation were divided into two broad categories: (i) those that can form two distinct epigenetic states (epialleles) that are maintained over generations in the presence of all epigenetic activities of the wild type, and (ii) those that revert to only one prevalent epigenetic state (reversible) and are thus not able to form heritable epialleles (Reinders *et al*, 2009; Teixeira *et al*, 2009). Although epiallelic reversion was associated with RNA-directed DNA methylation (RdDM) (Teixeira *et al*, 2009), the primary determinants underlying differences between the chromosomal targets of epigenetic regulation in susceptibility to epiallelic switching remained unknown. Here, we compare the whole-genome distribution of DNA methylation in *Arabidopsis* plants carrying either a weak or strong allele of MET1. Although the partial loss-of-function allele *met1-1* reduces CpG methylation levels to ~25% of wild type (Kankel *et al*, 2003), this causes only minor developmental defects. In contrast, the null allele *met1-3* causes an almost complete loss of mCpGs and is semilethal (Mathieu *et al*, 2007). The methylation remaining in *met1-1* thus identifies a particularly important subset of mCpG sites, offering a unique opportunity to understand the function of mCpG methylation. We find that loci forming stable epialleles are similarly affected in the two mutants, while epigenetically reversible loci are affected differently in *met1-1* and *met1-3*. These observations allow inferences about the molecular mechanisms of epigenetic transgenerational inheritance at these distinct classes of loci and the formulation of genetic and epigenetic features determining and thus predicting the capacity of chromosomal targets to form stable epialleles.

## Results

### Comparison of methylomes in different *met1* mutants

We compared whole-genome transcriptomes and methylomes between wild-type Col-0 plants, *met1-1* and *met1-3*, and also the *F2*

1 The Sainsbury Laboratory, University of Cambridge, Cambridge, UK

2 Department of Plant Biology, University of Geneva, Geneva, Switzerland

3 Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany

\*Corresponding author. Tel: +44 1223 761159; E-mail: jerzy.paszkowski@slcu.cam.ac.uk

†Present address: The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, The Hebrew University, Rehovot, Israel

‡Present address: School of Biological Sciences, University of Essex, Colchester, UK

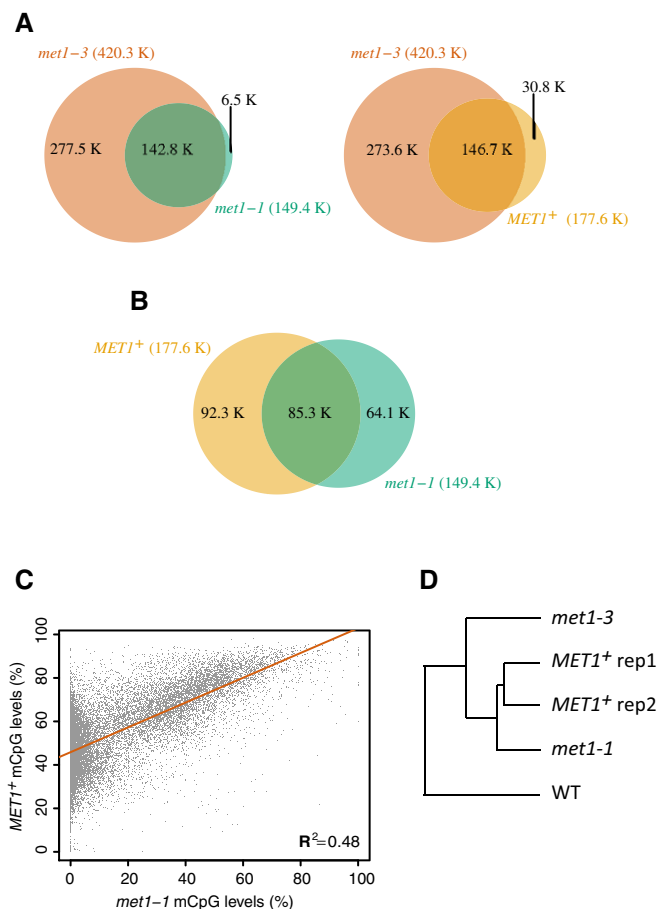
[The copyright line of this article was changed on 3 February 2017 after first online publication.]

progeny of a hybrid derived from a cross between *met1-3* and wild type. The  $F_2$  plants were genotyped, and only individuals with homozygous wild-type alleles of *MET1* were analysed. These  $MET1^+$   $F_2$  (hereafter  $MET1^+$ ) siblings had inherited half of their genomes from a *met1-3* grandparent, except for the region on chromosome 5 around *MET1*. For simplicity, we excluded chromosome 5 from subsequent analyses. The  $MET1^+$  segregants had on average 57% of wild-type mCpGs that indicated a certain degree of *de novo* methylation, at DNA sequences likely inherited from the *met1* grandparent (Appendix Fig S1C, Appendix Table S1).

Next, we screened for differentially methylated regions (DMRs) in pairwise comparisons with *met1-3*, *met1-1* and  $MET1^+$  plants using wild type as the common denominator. DMRs in *met1-3* included almost all DMRs of both *met1-1* and  $MET1^+$  plants (Fig 1A and Appendix Fig S2). Interestingly, DMRs of *met1-1* and  $MET1^+$  segregants overlapped in more than 85,000 commonly methylated cytosines ( $P$ -value = 0.001), representing 57 and 48% of *met1-1* and  $MET1^+$  DMRs, respectively (Fig 1B). There was a significant correlation (Spearman  $R^2 = 0.48$ , Pearson  $R^2 = 0.63$ ) of methylation distribution between *met1-1* and  $MET1^+$  (Fig 1C), which was further confirmed by hierarchical clustering of mCpGs in 200-bp non-overlapping genomic windows (tiles) (Fig 1D). These results suggested that certain methylation patterns in *met1-1* and  $MET1^+$  are associated with particular loci, possibly due to intrinsic characteristics of DNA sequences of the affected loci themselves. This hypothesis was tested in subsequent experiments.

### Loci methylated only in CpGs switch between stable epigenetic states

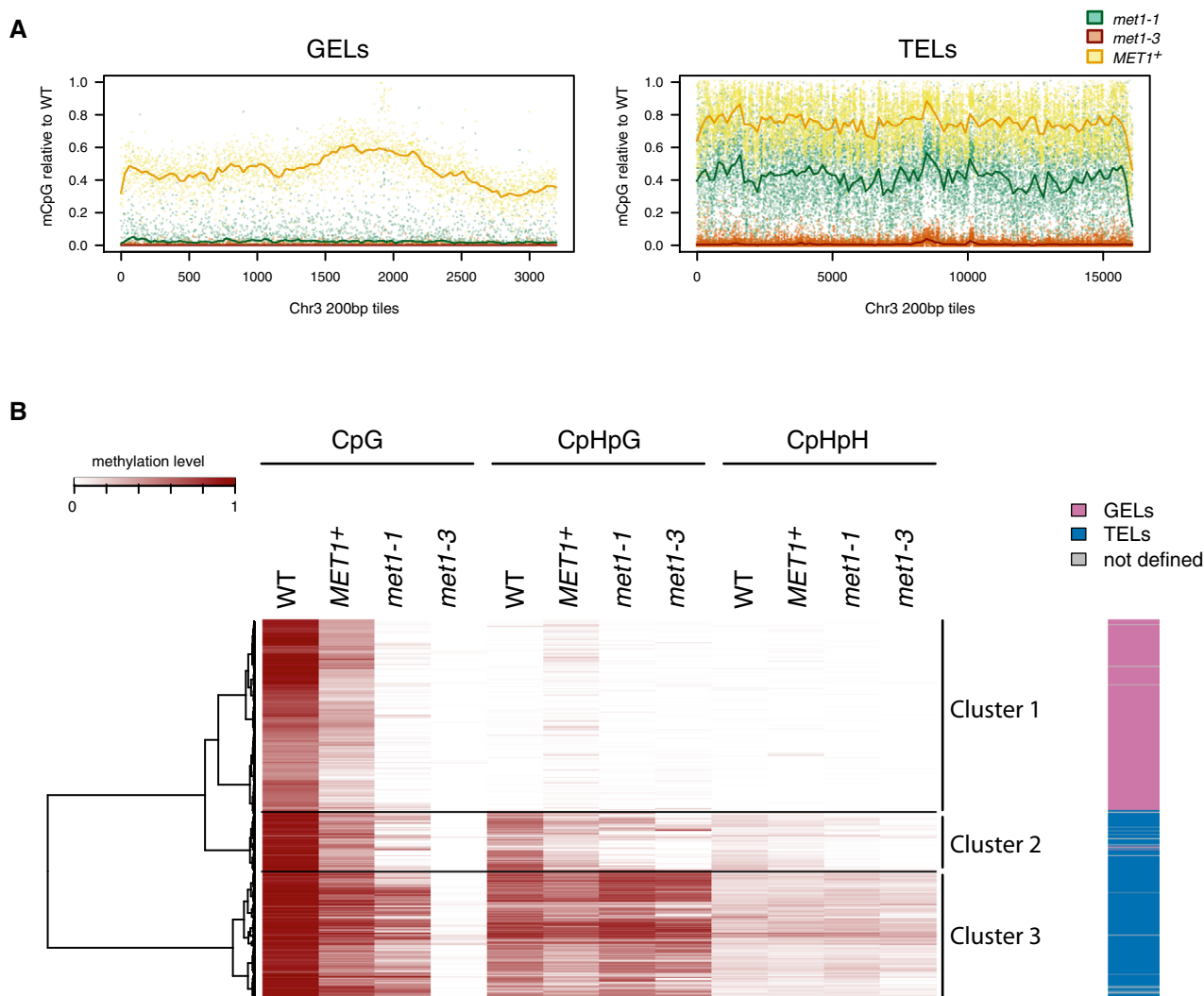
It has been proposed that targets of DNA methylation in plants are of two distinct types (Saze & Kakutani, 2011): (i) gene body methylation (hereafter referred to as gene body-like or GEL), present mostly in coding regions of expressed genes and consisting exclusively of mCpGs, and (ii) transposable element methylation (hereafter referred to as transposable element-like or TEL), found predominantly at transposons and chromosomal repeats and affecting CpGs but also non-CpGs. Here, GELs were defined as gene models in which averaged methylation of cytosines in CpG context was above 5%, but < 5% in non-CpG contexts and TELs as loci with methylation of both mCpGs and non-CpGs over 5% (Appendix Table S2). This produced totals of 11,746 GELs and 4,743 TELs. As expected, 98% of GELs were annotated as genes, while 73% of TELs were annotated as transposable elements (Appendix Fig S3) and 22% as genes (Appendix Fig S3), indicating that transposon-like methylation is also associated with a subset of protein-coding genes. We also annotated this way 200-bp genomic tiles containing at least 5% of averaged CpG methylation. We then examined GEL and TEL methylation in *met1-3*, *met1-1* and  $MET1^+$ . In *met1-3*, the two types of methylation were equally erased (Appendix Fig S4). In contrast, although in *met1-1* methylation at GELs was uniformly lost, methylation losses were not uniform across TELs, with many predominantly losing and others predominantly maintaining DNA methylation (Fig 2A and Appendix Fig S5). Interestingly,  $MET1^+$  segregants displayed a methylation pattern very similar to *met1-1*. Analyses of GELs in the  $MET1^+$  segregants revealed an average methylation of 50%, suggesting ubiquitous maintenance of the mid-parental methylation levels (Fig 2A and



**Figure 1. Patterns of CpG methylation in *met1-1* and  $MET1^+$  plants.**

- A Venn diagrams of genomewide hypomethylated CpG DMRs (chromosomes 1–4) of *met1-3* compared with control Col-0 plants and later compared with DMRs of *met1-1* and  $MET1^+$ . For *met1-1* and  $MET1^+$  plants, DMRs were considered with a relative difference to wild type of  $\geq 80$  and  $\geq 40\%$ , respectively. Overlapping areas represent number of cytosines in shared portions of DMRs.
- B Venn diagram of genomewide hypomethylated CpG DMRs (chromosomes 1–4) of *met1-1* ( $\geq 80\%$  difference) and combined (union of DMRs)  $MET1^+$  plant pools ( $\geq 40\%$  difference). Permutation test with 1,000 iterations using *regioner* package (Gel et al, 2016) confirmed a statistically significant overlap ( $P$ -value = 0.001).
- C Scatter plot of CpG DMRs (chromosomes 1–4) of *met1-1* and union of DMRs  $MET1^+$  plants (DMR criteria as in A and B). Linear regression—red line. The correlation coefficient is indicated in the bottom right corner of the plot.
- D Hierarchical clustering of CpG methylation in genomic tiles of wild-type (WT), *met1-1*, *met1-3* and  $MET1^+$  plants (chromosomes 1–4). The methylation of  $MET1^+$  was adjusted by subtracting 50% of the methylation derived from wild type. Clustering based on randomly selected genomic tiles (10%) ( $n = 13,668$ ) representing  $\geq 50\%$  CpG-methylated in wild type were used for analyses.

Appendix Figs S4 and S5), and TELs in the  $MET1^+$  segregants revealed an average methylation of 70% relative to wild type, implying substantial remethylation above the mid-parental level (Fig 2A and Appendix Figs S4 and S5). Notably, the patterns of methylation at particular TELs in *met1-1* and  $MET1^+$  were overlapping (Fig 2A and Appendix Fig S5), indicating that TELs have



**Figure 2. Distribution of CpG methylation at GELs and TELs in wild type, *met1-1*, *met1-3* and *MET1+*.**

**A** Distribution of CpG-methylation levels at 200-bp tiles ( $\geq 80\%$  CpG-methylated in wild type) along chromosome 3 at GELs or TELs (Appendix Table S2). Lines were generated using lowest smooth R with a span of  $1/50$ . Legend to colours assigned to particular genotypes is above the right graph. *MET1+* showed is mean value of two replicates. Corresponding graphs for chromosomes 2 and 4 are in Appendix Fig S5.

**B** Heatmap R of DNA methylation levels at 200-bp tiles (10% randomly selected on chromosome 1, CpG-methylated in wild type  $\geq 40\%$ ,  $n = 11,094$ ). Sequence context of methylation, genotypes and methylation levels are indicated above the figure. Tiles were ordered according to hierarchical clustering (hclust, R software environment) and the main clusters are delineated by horizontal lines and explained at the right of the figure and in the main text. The Euclidean distance dendrogram is presented left of the figure. *MET1+* sample is derived from the second replicate. Column on the right illustrates distribution of GELs and TELs as they are defined in Appendix Table S2 and Appendix Fig S3.

similar remethylation capacities in the two genotypes or, in the case of *met1-1*, similar resistance to methylation loss. The tiling array transcriptome mapping of *met1-1*, *met1-3* and wild type revealed that residual methylation in *met1-1* was also reflected by transcriptional silencing of these TELs (Appendix Fig S6).

Using 200-bp tiles, we further classified CpG-methylated regions according to their association with methylation of non-CpGs and performed hierarchical clustering that defined three main clusters (Fig 2B). The cluster 1 contained almost exclusively GELs, displaying the greatest methylation deficiency in *MET1+* segregants relative to other clusters, while cluster 2 and cluster 3 identified TELs with different degree of methylation in *MET1+* and *met1-1* (Fig 2B and Appendix Fig S7). Further comparison of the global CpG

methylation levels of GELs and TELs in *met1-3*, *met1-1* and *MET1+* in relation to wild type indicated that in *met1-3*, GELs and TELs both lacked mCpGs, while in *met1-1* and *MET1+* TELs had retained or regained significant levels of CpG methylation, respectively (Fig 2B and Appendix Fig S4).

#### Capacity of TELs to form transgenerationally stable epialleles is related to their genetic properties

The methylome and clustering analyses suggested that TELs are heterogeneous, consisting of at least two classes (Fig 2B). To define genetic and epigenetic features of TELs that result in differential methylation losses in *met1-1*, correlated with methylation levels

observed in *MET1*<sup>+</sup>, we sought DNA sequence properties of TELs that coincide with persistence or loss of CpG methylation in *met1-1*. We discovered that increases in the number of CpGs are associated with the propensity of TELs to lose CpG methylation (Fig 3A), suggesting a link between DNA sequence properties at methylated TELs and the formation of stably demethylated epialleles. We also found that presence of tandem repeats longer than 100 bp and increasing levels of their repetitiveness jointly correlate with maintenance of methylation in *met1-1* or thus remethylation in *MET1*<sup>+</sup> (Fig 3B). Therefore, intrinsic features of TEL DNA sequence are associated with the distribution of their methylation in *met1-1* and also with the previously described formation of transgenerationally stable methylation patterns (epialleles) in *MET1*<sup>+</sup> segregants (Lippman et al, 2003; Reinders et al, 2009).

To further characterize the epiallelic behaviour of TELs, we ranked TELs of clusters 2 and 3 (Fig 2B) according to their levels of methylation in *met1-1* and selected two contrasting subsets of TELs for further study (Fig 3C and Appendix Fig S8). In the first subset, *met1-1* retained < 5% of wild-type CpG methylation. As this subset resembled GELs in the capacity to form stable epialleles, we refer to them as epiallelic-TELs or E-TELs. More than 80% of CpG methylation was retained in *met1-1* in the second subset, correlating with regain of CpG methylation in *MET1*<sup>+</sup> and, thus, rapid reversal of the levels towards wild type. This subset of TELs was named reversible-TELs or R-TELs. It is important to bear in mind that GELs, E-TELs and R-TELs are equally depleted of CpG methylation in *met1-3* (Appendix Fig S4).

To directly compare methylation levels at all cytosines of E-TELs and R-TELs in wild-type, *met1-3*, *met1-1* and *MET1*<sup>+</sup> plants, we aligned their annotated sequences to construct plots for CpG, CpHpG and CpHpH methylation (Fig 3D). E-TELs displayed a complete loss of CpG methylation in both *met1* mutants and showed mid-parent levels of 50% of wild type in *MET1*<sup>+</sup> segregants, suggesting a lack of or minimal remethylation activity at these sequences by MET1, which is present already in *met1-3/MET1*<sup>+</sup> F1 hybrids and in *MET1*<sup>+</sup> segregants. In contrast, although R-TELs completely lost CpG methylation in *met1-3*, they reached ~80% of wild-type mCpGs in *MET1*<sup>+</sup> segregants (Fig 3D), supporting their active remethylation in the presence of MET1. Interestingly, although E-TELs and R-TELs do not show a relevant difference in their non-CpG methylation levels in wild-type plants (Appendix Fig S9), both CpHpG and CpHpH methylation were significantly reduced in both *met1* alleles at E-TELs but not at R-TELs (Fig 3D). Moreover, in *MET1*<sup>+</sup> plants, only E-TELs displayed a significant reduction in CpHpG methylation compared with the wild type (Fig 3D). Thus, non-CG methylation seems to persist at R-TELs in both *met1* mutant alleles but is depleted at E-TELs.

Methylation at CpHpGs is a part of the self-reinforcing regulatory loop with histone 3 dimethylation in lysine 9 (H3K9me2), and methylation at CpHpHs is maintained primarily by the RdDM pathway directed by small RNAs (Law & Jacobsen, 2010). Using published genomic data for H3K9me2 and small RNA (sRNA) distribution data available for wild type and *met1-3* (Lister et al, 2008; Deleris et al, 2012), we surveyed their distributions over R-TELs and E-TELs (Fig 3E and F). In wild-type plants, the levels of both H3K9me2 and the different classes of small RNAs were similar at R-TELs and E-TELs; however, in the *met1-3* mutant only R-TELs retained near wild-type levels of both H3K9me2 and sRNAs, while

E-TELs lose both (Fig 3E and F, and Appendix Fig S10). These results are consistent with previously observed association of sRNAs and non-CpG methylation with transgenerationally “remethylable” loci (Teixeira et al, 2009).

Next, we examined whether E-TELs and/or R-TELs form larger epigenetically coregulated domains or their local DNA sequences determine susceptibility to epigenetic switching. We found multiple examples of neighbourhoods of R-TELs and E-TELs (Fig EV1), consistent with the hypothesis that local features of DNA sequences can be used for the prediction of epigenetic reversibility (R-TELs) or ability to form stable epialleles (E-TELs) (Fig 3A and B). Certain transposon superfamilies are overrepresented among E-TELs or R-TELs (Appendix Fig S11), the most striking of which are Helitrons, which have the lowest CpG content and are present exclusively in R-TELs (Appendix Fig S11). Moreover, although most TE superfamilies were represented in E-TELs and R-TELs, certain TE families were enriched differentially among E-TELs and R-TELs (Appendix Tables S3 and S4, and Appendix Fig S12A), indicating that in addition to the sequence characteristics of a particular family, further sequence features, as delineated here, may explain the epigenetic properties of TEs. Certain transposon families could be clearly separated, consistent with E-TEL or R-TEL characteristics (Appendix Fig S12B), thus further reinforcing the hypothesis that DNA sequence composition, in combination with repetitiveness, can be used in defining likelihood of epiallelic properties of loci. Interestingly, one TE family (ATENSPM5) was represented in both E-TELs and R-TELs (Appendix Fig S12A). This unusual feature of ATENSPM5 appeared to exhibit a surprising duality of epigenetic regulation within this transposon by which one open-reading frame (ORF) behaves like an E-TEL and the other as an R-TEL (Fig EV2), again supporting the importance of very local DNA sequence determinants in differential epigenetic regulation.

### Rapid somatic remethylation and transcriptional silencing of R-TELs

To directly determine transgenerational epigenetic properties of GELs, E-TELs and R-TELs, and especially the remethylation timing of R-TELs, we backcrossed *met1-3*, which is in the Col-0 accession, to the wild-type Landsberg erecta (Ler) accession. This allowed the use of DNA sequence polymorphism to discriminate between alleles of the two parents in F1 hybrids and, thus, to separately examine their methylation levels. Parental methylation levels of GELs and E-TELs were maintained in F1 plants, displaying clear epiheterozygosity (Fig 4A). In contrast, *met1-3*-derived R-TELs underwent efficient remethylation, suggesting that *de novo* DNA methylation occurred as soon as functional MET1 became available (Fig 4B). Therefore, R-TELs seem different from “remethylable” loci where, in a study performed in the hypomethylated *ddm1* mutant, remethylation did not occur in the F1 (Teixeira et al, 2009). In addition, we examined in reciprocal backcrosses expression at R-TELs and E-TELs. The expression of E-TELs observed in *met1-1* was maintained in F1 plants, and R-TEL expression was efficiently silenced to the initial wild-type level, independent of the crossing direction (Fig 4C and D). This was true for loci annotated as transposons as well as for protein-coding genes with R-TELs or E-TELs in their promoters (Fig 4C and D).

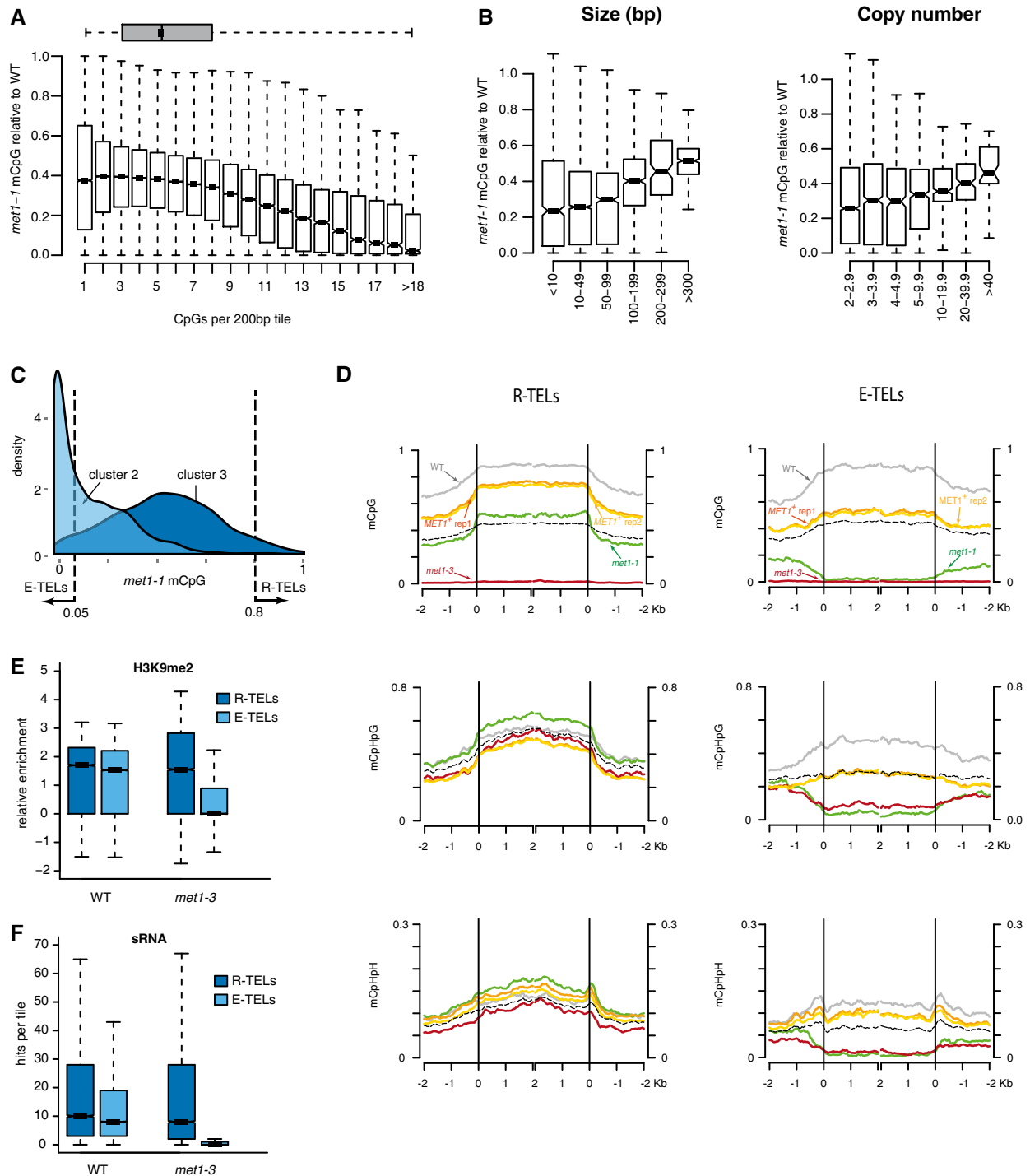


Figure 3.

Importantly, a backcross of *met1-1* to wild type brings hypomethylated and wild-type epialleles together in the F1, which opens the possibility of *trans* interactions between epialleles (Greaves *et al.*, 2012; Rigal *et al.*, 2016). To avoid such confounding epiallelic *trans* interactions, we introduced a *MET1* transgene into *met1-1* plants and examined R-TELs and E-TELs in two complemented transgenic lines with wild type-like *MET1* protein levels (Figs EV3 and EV4). Similar to what was observed in *met1/MET1*

F1 hybrids and in *MET1*<sup>+</sup> plants, R-TELs and E-TELs displayed their previously documented transcriptional attributes in both complemented transgenic lines (Fig 4E). Finally, we examined genomewide DNA methylation profiles of the two *MET1* complemented lines (Appendix Table S1 and Appendix Fig S1) and observed that transgenic *MET1*-dependent DNA remethylation occurs exclusively at R-TELs, and not at E-TELs or GELs (Figs 4F and EV4A and B). Likewise, epigenetic properties of certain TEs recorded in *MET1*<sup>+</sup> plants

**Figure 3. Genetic and epigenetic properties of TELs.**

- A Box plot of CpG methylation levels in *met1-1* relative to wild type (WT) in 200-bp tiles assigned to TELs (Appendix Table S2) and sorted according to the number of CpGs. The genomewide tile distribution is illustrated by a grey box plot at the top of the chart.
- B Box plot of CpG methylation levels in *met1-1* relative to wild type (WT) in tandem repeats sorted according to the size of the repeat units (left) or their copy number (right).
- C Density plot of CpG methylation level distribution in *met1-1* relative to wild type for 200-bp tiles assigned to cluster 2 and cluster 3 as defined in Fig 1B. Vertical dashed lines mark criteria for E-TEL and R-TEL selection which incorporate mostly tiles from cluster 2 and cluster 3, respectively.
- D Distribution of averaged DNA methylation levels in wild-type, *met1-1*, *met1-3* and *MET1*<sup>+</sup> plants at R-TELS and E-TELS. Identities of R-TEL and E-TEL are according to TAIR10 annotation (as defined in Appendix Fig S8) and aligned at the 5' end or the 3' ends (vertical lines). Average methylation levels for 50-bp sliding windows of 200 bp are plotted. R-TELS and E-TELS at chromosome 5 were excluded from the analyses. Black dashed line displays theoretical mid-parental methylation level between WT and *met1-3*.
- E Box plot of H3K9me2 relative enrichment calculated for R-TELS and E-TELS (200-bp tiles) in wild type (WT) and *met1-3* using processed data from Deleris *et al* (2012). The original normalized log<sub>2</sub> signal from ChIP analysis was averaged per 200-bp tile.
- F Box plot of small RNA levels at R-TELS and E-TELS in wild type and *met1-3*. Genomic distribution of small RNA from Lister *et al* (2008) was related to 200-bp tiles assigned to R-TELS and E-TELS and expressed as hits per tile.

Data information: In the graphs in (A, B, E and F) boxes delimit the the first and third quartiles. The horizontal lines represent the data medians. Whiskers delimit the lowest and the highest value within 1.5 of the interquartile range (IQR) of the lower and the upper quartile, respectively.

(Fig EV2) are consistent with those observed in the transgenic lines (Fig EV4C). Taken together, these results confirmed that distinct DNA sequence properties rather than trans-epiallelic interactions are associated with epiallelic stability or switching.

### Epiallelic properties during long-term, transgenerational inheritance

Finally, we wanted to know whether DNA sequence properties defining susceptibility to epigenetic switching have broader applicability. For that, we examined methylation of previously defined GELs, E-TELS and R-TELS in data from published experiments. For example, it has been demonstrated that progeny of heterozygous *met1-3* plants experience CpG methylation loss in the absence of MET1 during post-meiotic divisions of the haploid gametophytes (Saze *et al*, 2003). Therefore, it can be expected that inbreeding of *met1-3* heterozygous plants will result in gradual methylation losses. However, since MET1 would be present at each generation during somatic development, R-TELS would be subjected to remethylation, while GELs and E-TELS would remain hypomethylated. To test this prediction, we scored methylation levels of GELs, E-TELS and R-TELS in previously reported methylation profiles of plants propagated as heterozygous *met1-3* (*met1-3*<sup>+/-</sup>) and wild-type plants (*met1-3*<sup>+/+</sup>) segregating from these lines (Stroud *et al*, 2013). Consistent with our predictions, complete loss of CpG methylation at GELs and strong reduction in E-TELS were found in these datasets. In contrast, levels of methylation at R-TELS were similar to wild type (Fig 5A).

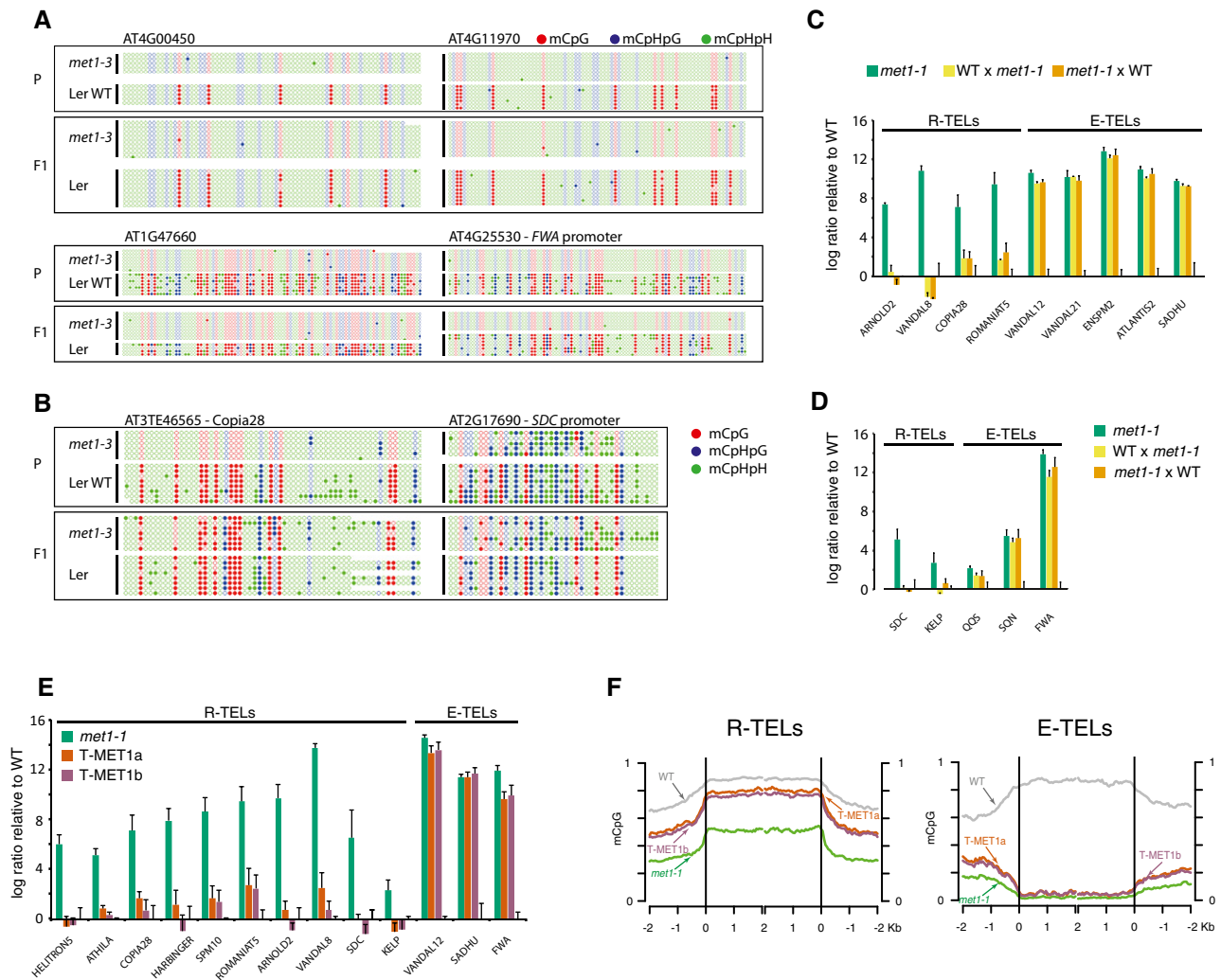
Next, we tested distribution of DMRs in R-TELS, GELs and E-TELS in the results of inbreeding experiments of *Arabidopsis* for 30 generations (Becker *et al*, 2011) and in natural *Arabidopsis* accessions (Schmitz *et al*, 2013). In both populations, we registered the occurrence of DMRs at GELs and E-TELS at a high frequency and at R-TELS at a very low frequency (Fig 5B). Thus, the revealed DNA sequence properties are predictive for patterns of transgenerational epigenetic inheritance in nature.

It has been shown that certain hypomethylated loci require several generations for their remethylation, occurring stepwise at each plant generation (Teixeira *et al*, 2009). Therefore, it became interesting whether the genetic determinants of demethylated loci could predict even more accurately their epigenetic switching

properties, when remethylation is allowed for several generations. To address this, we determined DNA methylation patterns using bisulphite sequencing of three epigenetic recombinant inbred lines (epiRILs), which correspond to the eighth generation following introgression of wild-type *MET1* allele into *met1-3* mutant (Reinders *et al*, 2009). To compare directly these patterns with patterns formed due to the rapid remethylation occurring directly after introgression of *MET1* gene, we also examined recently published methylation data obtained for F1 hybrids between *met1-3* and wild-type Col-0 (Rigal *et al*, 2016). We scanned these *Arabidopsis* methylomes using 200-bp tiles evaluating their remethylation level, CpG content and mappability score, reflecting their repetitiveness (Derrien *et al*, 2012). To circumvent bias derived from the methylated DNA inherited from wild-type Col-0 during the initial cross for epiRILs, we analysed only tiles residing in genomic regions inherited from *met1-3* mutant. In agreement with the predictive DNA sequence properties, remethylation initiated in F1 hybrids (Fig 5C) and observed in epiRILs occurred mostly at sequences with low mappability (high repetitiveness) and low CpG density (Fig 5D and Appendix Fig S13). Remarkably, although the sequences with low mappability are most prone to transgenerational remethylation, however, elevated density of CpGs allows formation of stable epialleles for some of the repeated DNA (Fig 5D, last panel). To further examine the predictive power of the DNA sequences determining epigenetic switching, we asked whether sequences expected to remethylate (< 0.2 mappability and < 5 CpGs per tile) and those expected to form stable epialleles (> 0.6 mappability and > 12 CpGs per tile) would indeed acquire such properties in epiRILs. For that, we determined their epiallelic properties in three independent epiRILs (Fig 6A). Indeed, we observed that epigenetic switching characteristics experimentally documented in epiRILs were predicted accurately, with genomic tiles being very clearly remethylated or remaining stably demethylated according to their DNA sequence properties (Figs 6A and EV5).

### Genetic properties of epiallelic switches in rice

To test whether the DNA sequence rules for predicting epiallelic switching are universal and thus applicable to other plant species, we turned to methylation data from recently characterized *met1* mutants of rice (Hu *et al*, 2014). In this mutant, CpG methylation is



**Figure 4. Transgenerational epiallelic stability of GELs, E-TELS and R-TELS in backcrosses.**

A DNA methylation analyses at two GELs (AT4G00450 and AT4G11970) and two E-TELS (AT1G47660 and AT4G25530 *FWA* promoter) of parental strains (*met1-3* mutant, which is in the Col-0 accession and Landsberg erecta (Ler) accession (panels marked P) and F1 hybrid (panels marked F1) containing *met1-3*- and Ler-derived chromosomes, which can be distinguished by DNA polymorphism. Open circles represent unmethylated cytosine and closed circles methylated.

B DNA methylation analyses at two R-TELS (AT3TE46565-Copia28 and AT2G17690-*SDC* promoter). General marking as in (A).

C Transcript levels in *met1-1* and in reciprocal F1 hybrids of *met1-1* and wild type (colour code above the graphs) of transposons assigned to R-TEL or E-TEL. The values are log ratio relative to the wild type adjusted to 0 (baseline).

D Transcript levels in *met1-1* and in reciprocal F1 hybrids of *met1-1* and wild type of genes assigned to R-TEL or E-TEL.

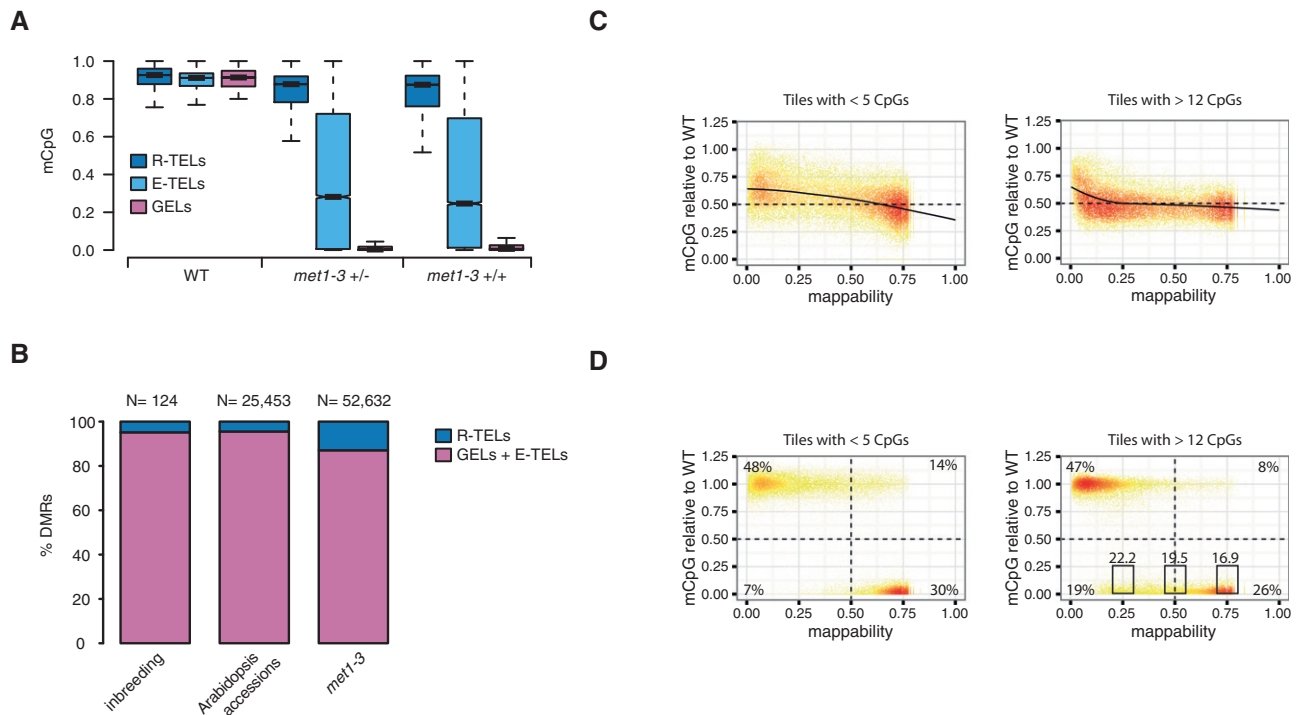
E Transcript levels in *met1-1* and in two independent transgenic lines with a reintroduced copy of the MET1 gene (T-MET1a and T-MET1b). Colour code is given next to the graphs. The identities of the tested loci are provided below the graph. General marking as in (C).

F Distribution of averaged CpG methylation levels in wild type (WT), *met1-1* and complemented *met1-1* lines (T-MET1a and T-MET1b) at R-TELS and E-TELS applied to TAIR10 annotation and aligned at the 5' end or the 3' ends (vertical lines). Average methylation levels for each 200-bp window are plotted. Chromosome 5 was excluded from the analyses. Corresponding plot for GELs is displayed on Fig EV4.

Data information: Graphs in (C–E) represent mean values of three biological repetitions,  $\pm$  s.d. marked by black lines.

reduced by 76%, closely resembling levels in *Arabidopsis met1-1* (Appendix Fig S14A and Appendix Table S1). CpG methylation at rice GELs was lost efficiently and TELS were characterized by heterogeneous methylation losses (Appendix Fig S15A), which mirrored *Arabidopsis* results (Fig 2A and Appendix Fig S4). We then compared the DNA sequences of *Arabidopsis* and rice E-TELS and R-TELS (Fig 3C and Appendix Fig S8). The characteristics of DNA sequences (specific CpG frequencies) of rice E-TELS and R-TELS and

degrees of sequence repetitiveness (mappability) were found to be similar to those of *Arabidopsis* (Appendix Figs S14B and C, and S15B and C). Finally, we separated rice DNA sequences as potentially able to form stable epialleles and sequences likely regaining methylation using their genetic properties. As in *Arabidopsis*, epigenetic characteristics of the two groups of loci were predicted correctly on the basis of their DNA sequence composition and genomic repetitiveness (Figs 6B and EV5B).



**Figure 5. Epiallelic frequencies in *Arabidopsis*.**

- A** Box plot of CpG methylation levels in 200-bp genomic tiles sorted as R-TELS, E-TELS and GELs in wild type, inbred *met1-3* heterozygous mutant (*met1-3<sup>+/-</sup>*) and wild-type plants segregated from inbred *met1-3* heterozygous mutant (*met1-3<sup>+/+</sup>*) (Stroud *et al*, 2013). Only GEL tiles with on average CpG methylation  $\geq 80\%$  were considered for this analysis. Boxes delimit the the first and third quartiles, and whiskers the lowest and the highest value within 1.5 IQR of the lower and the upper quartile, respectively.
- B** Relative distribution of CpG DMRs (among sequences with  $\geq 60\%$  CpG methylation in wild type), found in *met1-3* second generation, in long-time inbreeding (Becker *et al*, 2011) and among *Arabidopsis* accessions (Schmitz *et al*, 2013), partitioned to R-TELS or GELs/E-TELS. Total number of considered DMRs ( $n$ ) for each experimental system is given above the graphs.
- C** Scatter density plots of 200-bp genomic tiles mappability and CpG methylation relative to wild type (WT) in F1 (backcross of *met1-3* with wild type). The degree of red indicates the dot density. Methylated tiles (more than 80% methylation in CpG context in wild type) with less than five CpGs ( $N = 36,381$ ) or with more than 12 CpGs ( $N = 29,980$ ) are separately plotted. The CpG methylation level was obtained by average of all individual CpG positions on the tile sequence with minimal coverage of three reads. Black dashed lines display theoretical mid-parental methylation level between WT and *met1-3*. The solid black lines represent the polynomial regression (R-loess).
- D** Scatter density plots of 200-bp genomic tiles mappability and CpG methylation relative to wild type (WT) in epi01 (eighth generation after MET1 re-introduction). General description is as in (C). Methylated tiles in portion of genome inherited from *met1-3* parent with < 5 CpGs ( $N = 23,883$ ) or with more than 12 CpGs ( $N = 20,226$ ) are separately plotted. The percentage of tiles in the section delimited by the dashed lines is given in each corner. The average of CpG count for tiles in the area delimited by the black squares is given above each square.

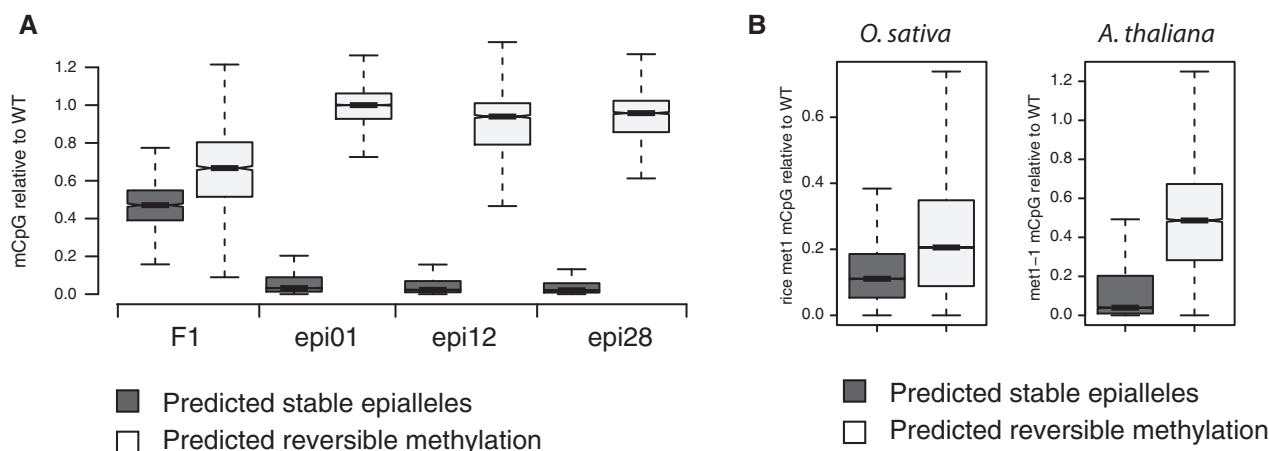
## Discussion

Development and fertility of *met1-3* plants, with completely erased CpG methylation, are severely compromised (Mathieu *et al*, 2007), while *met1-1* plants show no drastic phenotypic differences to the wild type, even though they lack three quarters of CpG methylation (Kankel *et al*, 2003). Our analyses revealed that specific subsets of loci are either strongly affected by the *met1-1* mutation, mimicking *met1-3*, or mostly retain their methylation (Fig 2 and Appendix Fig S5). Therefore, the developmental abnormalities in *met1-3* are probably associated with loci able to retain their methylation in *met1-1*. Most interestingly, these loci generally encode transposable elements with as yet unknown regulatory functions in plant development. Nevertheless, it is remarkable that they belong to chromosomal targets particularly recalcitrant to epigenetic switches and thus are found in only one epigenetic state. Therefore, it became intriguing to define the properties that secure such robust epigenetic stability.

It has often been discussed whether and, if so, to what extent DNA sequence itself impacts on epigenetic regulation (Schübeler, 2015). It has been known for many years that sequence redundancies in fungi attract DNA methylation, which occurs premeiotically and is thought to mark invading multicopy DNA (Rhoumim *et al*, 1992). More recent transgenic experiments, with site-directed insertion of various DNA fragments into a predefined chromosomal position of mammalian cells, provided evidence that the base composition of DNA sequence may attract or prevent DNA methylation according to CpG content and the occurrence of transcription factor binding sites (Krebs *et al*, 2014). However, molecular determinants that distinguish between chromosomal loci that rapidly revert to one prevalent epigenetic state and loci that form alternative transgenerationally heritable epiallelic states (epialleles) are largely unknown.

Here, we observed that gene body-like methylation in *Arabidopsis* is erased in both *met1-1* and *met1-3* mutants and the newly acquired





**Figure 6. Genetic features can predict most of epigenetic switches.**

**A** *Arabidopsis* genomic 200-bp tiles were grouped based on DNA sequence features in predicted epiallelic switches (epi), with mappability > 0.6 and CpG sites > 12; or in epigenetic reversible epiallele (rev), with mappability < 0.2 and CpGs < 5. The box plots represent DNA methylation relative to Col-0 (WT) in *met1-3* backcrossed to wild type (F1) or in three epiRILs (epi01, epi12 and epi28, respectively). For epiRILs, only the genome inherited from the *met1-3* parent was considered for the analysis. Only tiles methylated in wild type (> 5% average CpG methylation) are shown. The data in each group differ with a  $P$ -value <  $10^{-15}$  (one-sided Fisher's exact test).

**B** Box plot of CpG DNA methylation in epiallelic switches (epi) and epigenetic reversible epialleles (rev) relative to wild-type methylation (WT), predicted as described above for *Arabidopsis* and rice genome in respective *met1* mutants (*met1-1* for *Arabidopsis*). The data in the two groups differ with a  $P$ -value <  $10^{-15}$  (one-sided Fisher's exact test).

Data information: Boxes delimit the the first and third quartiles. The horizontal lines represent the data medians. Whiskers delimit the lowest and the highest value within 1.5 of the interquartile range (IQR) of the lower and the upper quartile, respectively.

hypomethylated states were transgenerationally transmitted and found in *MET1*<sup>+</sup> segregants (Figs 2 and 4). Although in rice the influence of gene body methylation on transcription seems to be significant (Rodrigues *et al*, 2013), in *Arabidopsis* changes in body methylation of genes do not significantly alter their transcriptional activities (Saze & Kakutani, 2011). Nevertheless, it is intriguing that body methylated genes are able to form and transgenerationally maintain two alternative epiallelic states. Under particular selective pressure, such epigenetic diversity might be of advantage. Indeed, we documented overrepresentation of gene body-associated DMRs during previous long-term inbreeding experiments (Becker *et al*, 2011) and found their preferential occurrence in natural *Arabidopsis* accessions (Schmitz *et al*, 2013; Fig 5B). This observation suggests that such alternative states occur in nature and, therefore, such epigenetic diversity is possibly beneficial. Remarkably, the distribution of gene body methylation is conserved between plant species (Takuno & Gaut, 2013), further reinforcing the potential importance of GELs.

Notably, genetic determinants seem to provide only a general framework for epigenetic regulation. This framework seems to be rigid for GELs, since these sequences readily form stable epialleles but, considering their dynamics and stability, a broad spectrum of epiallelic switches seem to operate at TELs. Here, aiming at the clarification of the genetic and epigenetic determinants involved in such switches, we focussed on only two extremes of TELs, namely R-TELs and E-TELs. This selection allowed us to define general DNA sequence properties that predict both loci with protection of their given epigenetic state and loci susceptible to epigenetic shift. However, TELs include many loci with genetic and epigenetic properties residing somewhere between R- and E-TELs. During inbreeding of epigenetic hybrids derived from backcrosses of *Arabidopsis*

mutants deficient in DNA methylation, certain loci required several generations of inbreeding to regain methylation levels similar to wild-type plants (Teixeira *et al*, 2009). It is likely that such loci that only slowly regain their methylation have properties intermediate between R- and E-TELs. Remarkably, for the R-TELs studied here, the sexual cycle is not required for remethylation and resilencing. At these loci, the somatic DNA methylation activities seem to be fully responsible for the reversion of DNA methylation, illustrated by the somatic development of *F1* hybrids in which R-TELs rapidly regain their methylation. This regain could, however, be specific to the epigenetic composition of the *F1* hybrid, in which allelic R-TELs coexist in both methylated and demethylated forms. Therefore, as observed before (Reinders *et al*, 2009; Greaves *et al*, 2012), their epigenetic interaction may have been essential for rapid remethylation. To address this hypothesis, we provided missing *MET1* activity on a transgene directly introduced into *met1* plants. Also in this experimental set-up, R-TELs rapidly regained their silencing (Figs 4C and EV4). Thus, it can be concluded that their intrinsic properties and not inter-allelic epigenetic interactions indeed trigger rapid remethylation.

These properties resemble “silent locus identity”, a term coined to describe the status of transcriptionally activated loci in mutants deficient in RdDM components, which can be resilenced after restoration of the RdDM pathway (Blevins *et al*, 2014). However, “silent locus identity” can be lost at certain loci subjected to deficiency in histone deacetylase 6 (HDAC6). Loci with lost “silent identity” resemble GELs or E-TELs. It is intriguing that silent transposons in plants are also methylated in non-CpG context in addition to mCpGs and are associated with repressive histone marks and with the production of sRNAs. These properties, however, apply equally to R- and E-TELs. Therefore, it is remarkable in *met1-3*,

where all TELs completely lose their CpG methylation, that only in E-TELs is this accompanied by depletion of non-CG methylation, depletion of repressive histone marks and loss of sRNAs. Notably, plotting transposons belonging to R-TELs and E-TELs according to their genomic copy number and CpG frequency separated these two groups (Fig EV2 and Appendix Fig S11). This sequence-based separation suggests that genetic determinants control the epigenetic properties and thus their intrinsic “silent identity” properties. This is consistent with the observation that methylation transfer from methylated R-TEL loci to homologous demethylated R-TELs is not required for rapid remethylation.

That E-TELs might be stimulated to acquire DNA methylation when particular mechanisms involved in restoration of “silent locus identity” are locally reinforced is an interesting concept. Recent studies have revealed an alternative RdDM pathway involving 21- to 22-nt sRNAs and RNA-dependent RNA polymerase 6 (RDR6), known as RDR6-RdDM, that separates initiation of *de novo* DNA methylation and silencing from its maintenance (Nuthikattu *et al*, 2013; Bond & Baulcombe, 2015). It has been demonstrated that the lost DNA methylation at the *FWA* promoter, a typical and frequently studied E-TEL, can be reversed when additional 21- to 22-nt sRNAs are provided through a viral vector (Bond & Baulcombe, 2015). Thus, an activated epigenetic state of a demethylated E-TEL can be tilted back to silencing by forcing local RDR6-RdDM activity. This suggests that a threshold regulation is involved in epigenetic switches at E-TELs. In contrast, transposable element families naturally targeted by RDR6-RdDM, such as *AtREP10C* (part of the HELITRON superfamily) and *ROMANIAT5* (Nuthikattu *et al*, 2013), were consistently found among R-TELs in our analyses (Appendix Figs S11 and S12). This opens the possibility that TELs naturally able to produce RDR6-dependent sRNAs cannot stably switch to demethylated epialleles and would thus be in the R-TELs category. It is intriguing how these properties are mechanistically linked to the local abundance of CpG and genome-wide levels of sequence repetitiveness.

These DNA sequence-related predictions could possibly be applied to the observed restriction of active transposons. Recent studies of *de novo* silencing of LTR retrotransposon *EVADE* provide interesting illustrations of this possibility (Marí-Ordóñez *et al*, 2013). *EVADE* is transcriptionally activated in *met1-3* but remains inactive in *met1-1*; however, it loses 50% of CpG methylation and thus its properties reside between R- and E-TELs. During inbreeding, *EVADE* transposes with a sharp increase in copy number until it reaches a certain threshold, becomes methylated by RdDM and silenced. The increase in copy number moves *EVADE* into the R-TEL category; thus, it becomes an RdDM target and is stably hypermethylated.

Taken together, our results demonstrate that genetic determinants can be used to predict the epiallelic properties of plant chromosomal loci. Repetitiveness and relative scarcity of CpGs support maintenance “silent locus identity” and direct rapid reversion to one prevalent epigenetic state. In contrast, low copy number and high CpG content liberate the formation and support transgenerational stability of alternative epigenetic states.

## Materials and Methods

Supplementary Materials and Methods can be found in the Appendix.

### Plant growth and material

Plants used in this work were derived from *Arabidopsis thaliana* Columbia-0 lines, thirteenth- and second-generation homozygous *met1-1* (Kankel *et al*, 2003) and *met1-3* (Saze *et al*, 2003) have been used, respectively. The *met2,3,4* triple mutant was generated by crossing the lines SALK\_010893, SALK\_099592 and SALK\_098878, genotyping and selecting the segregating homozygous mutants through PCR. Seeds were stratified at 4°C for 4 days and plants grown in ½ MS 1.5% agar vertical plates or in soil, depending on the analysis. Plants were grown under long-day conditions (21°C, 16-h light, 8-h dark). EpiRILs have been obtained as described in Reinders *et al* (2009).

### Library preparation

Bisulphite-converted DNA libraries of *met1-1*, *met1-3*, two *MET1*<sup>+</sup> replicates, two transgenic complemented lines (T-MET1a and T-MET1b), and the epi01, epi12 and epi28 (Reinders *et al*, 2009), together with wild-type Col-0 control for genomic sequencing were performed starting from 0.5 to 1 µg of genomic DNA using the NEBNext DNA Sample Prep Reagent Set1 (New England Biolabs), following the Illumina Genomic Sample Prep Guide (Illumina), as described in Becker *et al* (2011). Libraries for RNA expression analysis were prepared in triplicate from 3 µg of total RNA, processed with the GeneChip Whole Transcript Amplified Double-Stranded Target Assay (Affymetrix), according to the manufacturer's protocol, to generate labelled cDNA for tiling microarray hybridization.

### Analysis of gene expression

The labelled cDNA was hybridized to the GeneChip *Arabidopsis* Tiling 1.0R array (Affymetrix) and scanned following the manufacturer's instructions. Tiling array hybridization data were processed with the R statistical software ([www.r-project.org](http://www.r-project.org)) and BioConductor ([www.bioconductor.org](http://www.bioconductor.org)) applying the chip definition file (CDF), kindly provided by Naouar *et al* (2009), as previously described (Yokthongwattana *et al*, 2010).

For real-time qRT-PCR analysis, total RNA (2 µg) was treated with RQ1 DNase (Promega) and reverse-transcribed with the SuperScript VILO cDNA Synthesis Kit (Invitrogen) according to the manufacturer's instructions. PCRs were carried out in triplicate using 10 ng of template cDNA, 200 nM target-specific primers (Appendix Table S5) and LightCycler 480 SYBR Green I Master (Roche) in the LightCycler 480 II detection system (Roche) in a volume of 10 µl.

### Sequencing and processing

Bisulphite-converted libraries were sequenced with 2 × 101-bp paired-end reads on an Illumina GAIIX or Illumina NextSeq 500 and HiSeq 2500 instrument. For image analysis and base calling, we used the Illumina OLB software version 1.8. The raw reads were trimmed using Trimmomatic (Bolger *et al*, 2014) in order to remove adapter sequences. Reads with an average quality value of at least 15 in a window of four nucleotides were trimmed from both ends. After trimming, reads shorter than 16 bases were discarded. The remaining sequences (on average 87% of raw reads) were aligned

against the *Arabidopsis thaliana* genome TAIR10 version using Bismark (Krueger & Andrews, 2011). Duplicated reads were collapsed into one read. Chloroplast sequences were used to estimate the bisulphite conversion (on average above 99%) (Appendix Table S1). To account for non-converted DNA, we applied a correction according to Lister *et al* (2013). The number of methylated reads were decreased as:  $m^* = \max(0, m - nc)$  (where  $m^*$  is the corrected number of methylated reads,  $m$  is the raw number of methylated reads,  $n$  is the total number of reads and  $c$  is the conversion rate). DMRs (differentially methylated regions) were defined comparing methylation in wild-type Col-0 with the other conditions analysed using the R package “DMRcaller” (Zabet & Tsang, 2015). We used “noise filter” method to compute CpG and CpHpG DMRs. Briefly, the “noise filter” method uses a triangular kernel to smooth the total number of reads and the total number of methylated reads. Note that the “noise filter” method uses the assumptions of BSmooth package (Hansen *et al*, 2012), namely that adjacent cytosines display correlated methylation. In particular, we used a window size of 172 nt, for CpG methylation, and 160 nt, for CpHpG methylation, to smooth the data and then we performed a score test at each position to determine the positions that display a statistically significant differences in methylation levels between the two conditions (note that using the score test leads to the same results as Fisher’s exact test, but was much faster to compute). At each position, we computed the *P*-value and adjusted for multiple testing using the Benjamini and Hochberg’s method (Benjamini & Hochberg, 1995) to control the false discovery and we discarded positions with FDR higher than 0.05. We further discarded the positions with less than four reads and the positions with differences in methylation levels lower than 0.4, in the case of CpG, and 0.2 in the case of CpHpG. Adjacent positions within 200 nt of each other were joined only if the resulting DMR displayed a minimal average number of reads per cytosine of 4 and statistically significant difference in methylation level (FDR lower than 0.05), which is at least 0.4 in the case of CpG and 0.2 in the case of CpHpG. For CpHpH DMRs, we used the “neighbouring” method, which performs the same algorithm as the “noise filter” method except that it does not perform data smoothing. In the case of CpHpHp DMRs, we considered only regions that display a minimal difference in the methylation level of 0.1 and a minimum size of 50 bp. Rice *met1* mutant and wild-type control sequencing data were obtained from (Hu *et al*, 2014) and re-analysed with the parameters described for *Arabidopsis*.

### Mappability

The mappability was computed on the TAIR10 *Arabidopsis* genome assembly or the ENSEMBL *Oriza sativa* IRGSP, with the gem-mappability tool from the Gem library (Derrien *et al*, 2012), with a kmer size of 20 bp and allowing a maximum of one mismatch. Gem output was converted to bed format and average mappability for 200-bp tile was calculated with the Bedtools package.

### Determination of parental inherited genome in epiRILs

In order to determine which part of the genome of each epiRILs was inherited from the *met1-3* parent, we took advantage of stable demethylated epialleles generated in epiRILs at body methylated

genes. A set of epi-markers were selected starting from the TAIR10 annotation based on the following criteria: (i) coding DNA (exons) covered with a minimum of five reads in both wild-type and *met1-3* samples; (ii) having at least three cytosines covered by reads; (iii) with an averaged methylation in CpG context above 50% in wild type; and (iv) with averaged wild-type methylation in both CpHpG and CpHpH below 1% or absent. For each of the 20,520 epi-markers obtained, we calculated the epi-status in each epiRIL with the following function:  $mCpG^{EPI} - mCpG^{MET}/mCpG^{WT} - mCpG^{MET}$ , where  $mCpG^{EPI}$  is the averaged CpG methylation in epiRILs,  $mCpG^{MET}$  is the averaged CpG methylation in *met1-3* and  $mCpG^{WT}$  is the averaged CpG methylation in wild type. Since CpG methylation in *met1-3* is virtually absent, an epi-status of 0 indicates absence of CpG methylation, while 1 indicates CpG methylation as wild type. The epi-status values across each chromosome were smoothed with the R function `smooth.spline` with a spar value of 0.6. All genome portions with a smoothed value above 0.8 or below 0.2 were considered WT-derived or *met1*-derived, respectively.

### Accession numbers

Sequencing and array data have been deposited in Gene Expression Omnibus under the accession number GSE89592.

**Expanded View** for this article is available online.

### Acknowledgements

This work was supported by European Research Council (EVOBREED) [322621]; Gatsby Fellowship [AT3273/GLE]; AENEAS; DFG [SFB 1101] and Max Planck Society [DFG SFB 1101].

### Author contributions

MC and JP participated in conception and design of the experiments; MC, MD, ML-L, CBe and CBa prepared nucleic acid libraries; CBe and CBa participated in genomewide sequencing; JG and MC participated in local DNA methylation profiles; MC performed all other experiments; MC, NRZ and CBe participated in analysis of the data; JP and DW contributed reagents/materials/analysis tools; MC and JP wrote the paper with the contribution of DW.

### Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, Weigel D (2011) Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* 480: 245–249
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57: 289–300
- Blevins T, Pontvianne F, Cocklin R, Podicheti R, Chandrasekhara C, Yerneni S, Braun C, Lee B, Rusch D, Mockaitis K, Tang H, Pikaard CS (2014) A two-step process for epigenetic inheritance in *Arabidopsis*. *Mol Cell* 54: 30–42
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30: 2114–2120

- Bond DM, Baulcombe DC (2015) Epigenetic transitions leading to heritable, RNA-mediated *de novo* silencing in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 112: 917–922
- Deleris A, Stroud H, Bernatavichute Y, Johnson E, Klein G, Schubert D, Jacobsen SE (2012) Loss of the DNA methyltransferase MET1 induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in *Arabidopsis thaliana*. *PLoS Genet* 8: e1003062
- Derrien T, Estellé J, Sola SM, Knowles DG, Raineri E, Guigó R, Ribeca P (2012) Fast computation and applications of genome mappability. *PLoS ONE* 7: e30377
- Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R (2016) RegioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* 32: 289–291
- Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, Dennis ES (2012) Trans chromosomal methylation in *Arabidopsis* hybrids. *Proc Natl Acad Sci USA* 109: 3570–3575
- Hansen KD, Langmead B, Irizarry RA (2012) BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* 13: R83
- Hu L, Li N, Xu C, Zhong S, Lin X, Yang J, Zhou T, Yuliang A, Wu Y, Chen Y-R, Cao X, Zemach A, Rustgi S, von Wettstein D, Liu B (2014) Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proc Natl Acad Sci USA* 111: 10642–10647
- Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddloh JA, Riddle NC, Verbsky ML, Richards EJ (2003) *Arabidopsis* MET1 cytosine methyltransferase mutants. *Genetics* 163: 1109–1122
- Krebs AR, Dessus-Babus S, Burger L, Schübeler D (2014) High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *eLife* 3: e04094
- Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27: 1571–1572
- Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11: 204–220
- Lippman Z, May B, Yordan C, Singer T, Martienssen R (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol* 1: e67
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133: 523–536
- Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghhighi FG, Sejnowski TJ et al (2013) Global epigenomic reconfiguration during mammalian brain development. *Science* 341: 1237905
- Marí-Ordóñez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O (2013) Reconstructing *de novo* silencing of an active plant retrotransposon. *Nat Genet* 45: 1029–1039
- Mathieu O, Reinders J, Čaikovski M, Smathajitt C, Paszkowski J (2007) Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. *Cell* 130: 851–862
- Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, Slotkin RK (2013) The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21–22 nucleotide small interfering RNAs. *Plant Physiol* 162: 116–131
- Reinders J, Wulff BBH, Mirouze M, Marí-Ordóñez A, Dapp M, Rozhon W, Bucher E, Theiler G, Paszkowski J (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev* 23: 939–950
- Rhounim L, Rossignol JL, Faugeron G (1992) Epimutation of repeated genes in *Ascobolus immersus*. *EMBO J* 11: 4451–4457
- Rigal M, Becker C, Pélissier T, Pogorelcnik R, Devos J, Ikeda Y, Weigel D, Mathieu O (2016) Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in *Arabidopsis thaliana* F1 epihybrids. *Proc Natl Acad Sci USA* 113: E2083–E2092
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. *Nat Biotechnol* 29: 24–26
- Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, Fischer RL, Zilberman D (2013) Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proc Natl Acad Sci USA* 110: 7934–7939
- Saze H, Scheid OM, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat Genet* 34: 65–69
- Saze H, Kakutani T (2011) Differentiation of epigenetic modifications between transposons and genes. *Curr Opin Plant Biol* 14: 81–87
- Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, Alix A, McCosh RB, Chen H, Schork NJ, Ecker JR (2013) Patterns of population epigenomic diversity. *Nature* 495: 193–198
- Schübeler D (2015) Function and information content of DNA methylation. *Nature* 517: 321–326
- Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* 152: 352–364
- Takuno S, Gaut BS (2013) Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proc Natl Acad Sci USA* 110: 1797–1802
- Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccard M, Ciaudo C, Cruaud C, Poulain J, Berdasco M, Fraga MF, Voinnet O, Wincker P, Esteller M, Colot V (2009) A role for RNAi in the selective correction of DNA methylation defects. *Science* 323: 1600–1604
- Yokthongwattana C, Bucher E, Čaikovski M, Vaillant I, Nicolet J, Scheid OM, Paszkowski J (2010) MOM1 and Pol-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing. *EMBO J* 29: 340–351
- Zabet N, Tsang J (2015) *DMRcaller: differentially methylated regions caller*. R package version 1.6.0. Bioconductor: [www.bioconductor.org](http://www.bioconductor.org)



**License:** This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.