

ARTICLE

DOI: 10.1038/s41467-017-02219-3

OPEN

# Stable transgenerational epigenetic inheritance requires a DNA methylation-sensing circuit

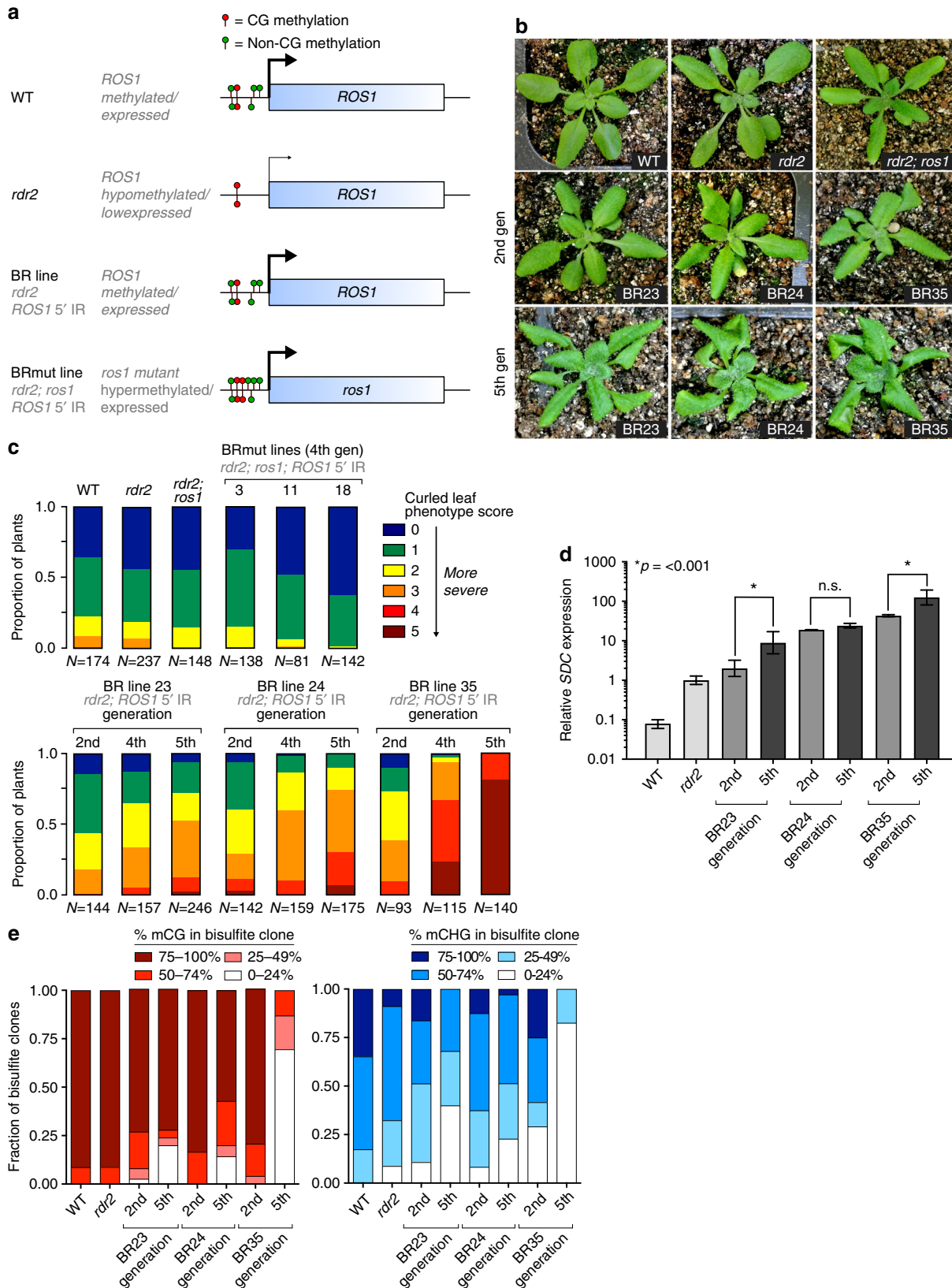
Ben P. Williams<sup>1</sup> & Mary Gehring <sup>1,2</sup>

Epigenetic states are stably propagated in eukaryotes. In plants, DNA methylation patterns are faithfully inherited over many generations but it is unknown how the dynamic activities of cytosine DNA methyltransferases and 5-methylcytosine DNA glycosylases interact to maintain epigenetic homeostasis. Here we show that a methylation-sensing gene regulatory circuit centered on a 5-methylcytosine DNA glycosylase gene is required for long-term epigenetic fidelity in *Arabidopsis*. Disrupting this circuit causes widespread methylation losses and abnormal phenotypes that progressively worsen over generations. In heterochromatin, these losses are counteracted such that methylation returns to a normal level over four generations. However, thousands of loci in euchromatin progressively lose DNA methylation between generations and remain unmethylated. We conclude that an actively maintained equilibrium between methylation and demethylation activities is required to ensure long-term stable inheritance of epigenetic information.

<sup>1</sup>Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA. <sup>2</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Correspondence and requests for materials should be addressed to M.G. (email: [mgehring@wi.mit.edu](mailto:mgehring@wi.mit.edu))

In plants and animals, DNA methylation is important for transcriptional repression of repetitive elements<sup>1,2</sup>, silencing of recombination at repetitive regions<sup>3-5</sup>, and establishment of genomic imprinting<sup>6</sup>. Disrupted DNA methylation patterns in human cells are associated with multiple cancers<sup>7</sup>. Plant DNA methylation patterns are stably inherited over many

generations<sup>8,9</sup>, and can persist relatively unchanged in reproductively isolated populations<sup>10</sup>. Symmetric CG methylation is faithfully copied to the daughter strand after DNA replication by the methyltransferase MET1<sup>1</sup>, whereas methylation of CHG and CHH sites (where H is any base other than G) is established and maintained by the enzymes CMT3 and CMT2, respectively<sup>11</sup>. All



sequence contexts (CG, CHG, and CHH) can be methylated de novo by the cytosine DNA methyltransferase DRM2, which is recruited to DNA by the RNA-directed DNA methylation pathway (RdDM)<sup>1</sup>. Methylated bases in all sequence contexts can also be actively removed by 5-methylcytosine DNA glycosylases through base excision repair<sup>12</sup>. We recently discovered that the expression of the *Arabidopsis* 5-methylcytosine DNA glycosylase gene *ROS1* is activated by the RdDM pathway and repressed by DNA demethylation by the *ROS1* enzyme<sup>13</sup>. These pathways converge at a short 220 bp sequence in the *ROS1* 5' region<sup>13,14</sup>, such that *ROS1* transcriptional output is quantitatively coupled to the methylation level of this sequence. This coupling is conserved in *Arabidopsis lyrata* and may also operate in distantly related maize<sup>13</sup>, suggesting it may have an important role in regulating DNA methylation homeostasis. We predicted that methylation-sensitive expression of *ROS1* functions as an epigenetic rheostat for the genome<sup>13</sup>.

In this study, we demonstrate that the methylation-sensitive expression of *ROS1*, which couples DNA demethylation activity to the methylation state of the genome, functions to ensure stable epigenetic inheritance. Restoring *ROS1* expression to wild-type levels in a methylation mutant causes abnormal phenotypes that worsen over multiple generations. This also causes widespread methylation losses across the genome. Methylation losses worsen over multiple generations in euchromatic regions, but are progressively reversed in heterochromatin. We conclude that the regulation of *ROS1* by methylation and demethylation pathways functions as a methylation-sensing circuit that buffers against fluctuations or instability in the maintenance of DNA methylation patterns over long periods of time.

## Results

**Abnormal phenotypes in an RdDM mutant that expresses *ROS1*.** To test the biological significance of the *ROS1* regulatory circuit, we engineered *Arabidopsis rdr2* mutant plants so that the expression of *ROS1* was uncoupled from the methylation status of the genome. *rdr2* mutants do not produce the double stranded RNAs that initiate canonical RdDM, and are characterized by reduced CHH methylation and 10-fold lower *ROS1* expression. We restored *ROS1* expression to wild-type levels in *rdr2* mutants by remethylating the short 220 bp sequence within the endogenous *ROS1* 5' region (Fig. 1a; Supplementary Fig. 1)<sup>13</sup>. We termed these plants Broken Rheostat (BR) lines because expression of *ROS1* is uncoupled from the genome-wide methylation state. To control for possible *ROS1*-independent effects, we also remethylated the locus in an *rdr2; ros1* double mutant background, which we term BRmut lines (Fig. 1a; Supplementary Fig. 1). We observed that a fraction of BR line plants exhibited an abnormal curled leaf phenotype compared to the parental *rdr2* mutant, which does not display overt morphological differences compared to WT (Fig. 1b, c). Strikingly, both the penetrance and the severity of the curled leaf phenotype increased over multiple generations—even BR lines that initially lacked a phenotype

exhibited a portion of plants with curled leaves in later generations (Fig. 1b, c). By the fifth generation, the majority of individuals from multiple BR lines exhibited curled or severely curled leaves. Conversely, no abnormal phenotypes were apparent in BRmut control lines, indicating that the curled leaf phenotype is dependent on *ROS1* (Fig. 1c).

The triple DNA methyltransferase mutant *drm1; drm2; cmt3* (*ddc*) has reduced CHG and CHH methylation and exhibits a curled leaf phenotype similar to what we observed, although there is no variability in the frequency or extent of *ddc* leaf curling<sup>15</sup>. Curled leaves in *ddc* mutants are caused by expression of *SUPPRESSOR OF DDC* (*SDC*), a gene silenced in leaves of wild-type plants by methylation of 5' tandem repeats by the chromomethyltransferase CMT3 and by RdDM<sup>15</sup>. To determine whether the abnormal leaf morphology in BR lines was correlated with hypomethylation and expression of the *SDC* locus, we performed bisulfite-PCR and sequencing on the *SDC* 5' region from individual second generation BR line plants that exhibited curled leaves (Supplementary Fig. 2). Consistent with their phenotype, *SDC* CG and CHG methylation was reduced in BR line plants, but not in BRmut line plants, and *SDC* expression was increased between 5- and 100-fold (Supplementary Fig. 2). Because the leaf phenotype of BR lines was variable within and between multiple generations (Fig. 1c), we hypothesized that in this genetic background *SDC* is an unstable epiallele that progressively loses methylation and silencing over multiple generations. To test this hypothesis, RNA and DNA were extracted from a population of pooled plants from second and fifth generation BR lines and used for RT-qPCR and bisulfite-PCR. We observed a significant increase in the abundance of *SDC* transcripts in the fifth generation compared to the second generation for two of three BR lines (Fig. 1d). In addition, the *SDC* locus lost methylation over generational time—the frequency of bisulfite sequencing clones with reduced CG and CHG methylation was higher in the fifth generation compared to the second generation for all three BR lines (Fig. 1e; Supplementary Fig. 2). These changes were not correlated with differences in *ROS1* expression, which remained steady at wild-type levels across generations in the BR lines (Supplementary Fig. 1).

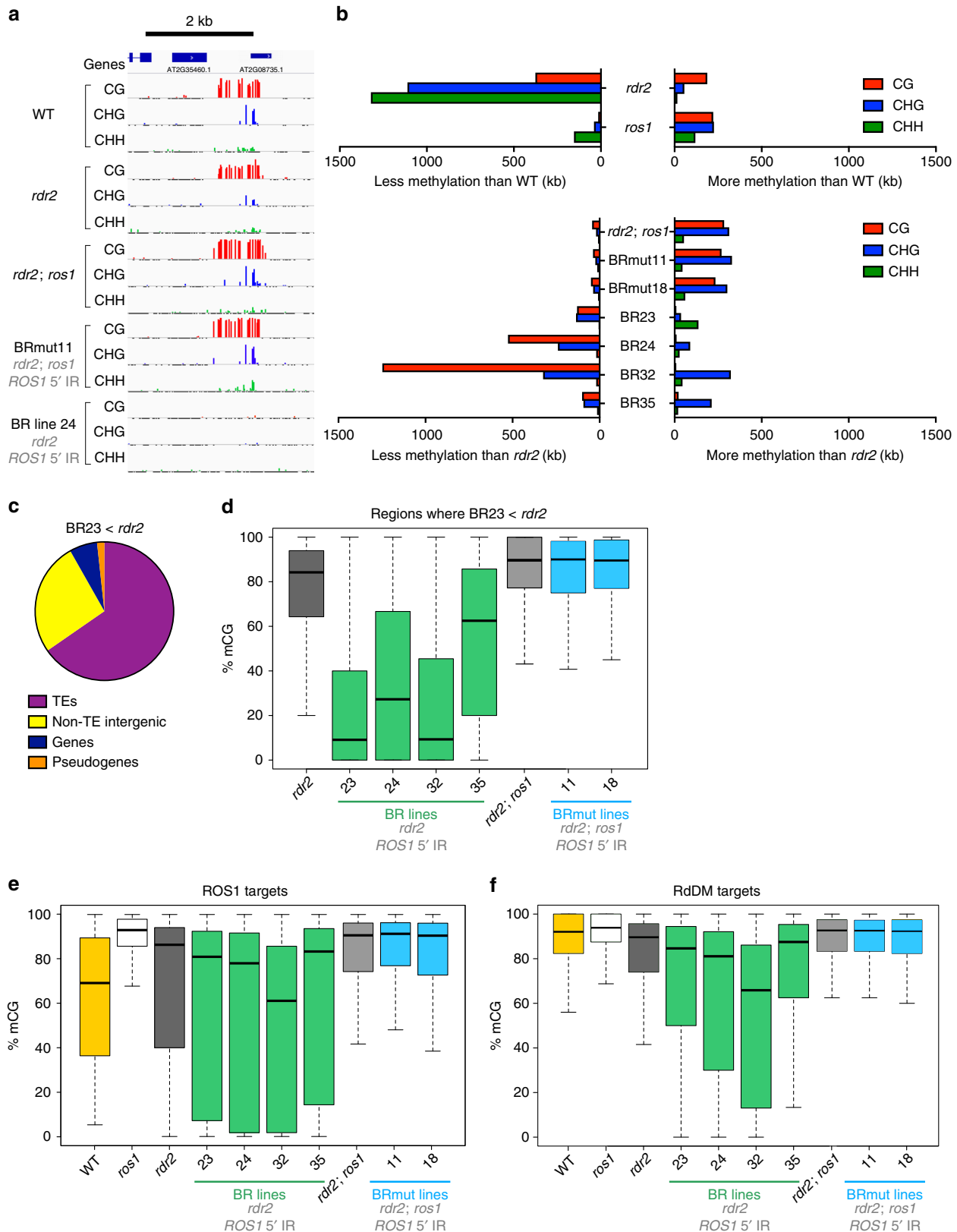
These results suggested two important hypotheses, which we proceeded to test using whole-genome approaches: First, DNA methylation patterns are not stably maintained when *ROS1* expression is uncoupled from the methylation state of the genome and second, the effect of disrupting equilibrium between methylation and demethylation activities cumulatively increases over generational time.

**Uncoupling *ROS1* from the methylome causes methylation losses.** We performed whole-genome bisulfite sequencing on individual fourth generation plants of varying phenotypes from four independent BR lines and two BRmut control lines, along with WT, *rdr2*, *ros1*, and *rdr2; ros1* mutants (Supplementary Table 1). A methylation score was computed within overlapping

**Fig. 1** Abnormal phenotypes develop over multiple generations after uncoupling *ROS1* expression from genomic methylation levels. **a** Schematic of the *ROS1* 5' methylation state and gene expression level in wild-type (WT), *rdr2* mutants, broken rheostat (BR) lines, and control broken rheostat mutant (BRmut) lines. *ROS1* 5' IR denotes an inverted repeat (IR) transgene complementary to the *ROS1* methylated 5' region. **b** Representative images of 3-week-old WT (Col-0), *rdr2* and *rdr2; ros1* mutants, as well as second and fifth generation plants from three independent, self-fertilized broken rheostat (BR) lines. **c** Proportion of plants with a curled leaf phenotype, represented by a phenotype score ranging from 0 (no curled leaves) to 5 (all leaves severely curled). Individual plants of median phenotype score for the population were used as parents for the subsequent generation. *N* represents the number of plants examined for each genotype/generation. **d** RT-qPCR quantification of *SDC* transcript levels in RNA from three independent pooled populations of WT, *rdr2*, second, and fifth generation BR line plants. 15–20 plants were combined for each biological replicate. Error bars represent one standard deviation of the mean from three biological replicates. \**p* ≤ 0.001 fifth generation compared to second generation (two-tailed *t*-test). **e** Bisulfite-PCR sequencing of the tandem repeats 5' of *SDC* from DNA extracted from the same pooled tissue used for RT-qPCR as shown in (d). Methylation data are shown as the fraction of bisulfite clones that are methylated at different levels. Between 24–37 clones were sequenced for each genotype/generation

300 bp windows to identify differentially methylated regions based on the density of differentially methylated cytosines (see Methods). Overlapping windows were then merged to calculate the total length of DNA with significantly reduced or increased methylation in each sample. All four BR lines showed clear decreases in CG and CHG methylation compared to *rdr2*,

consistent with our findings at *SDC* (Fig. 2a, b). At some loci, CG and non-CG methylation was completely lost (Fig. 2a). In contrast, BRmut lines and *rdr2*; *ros1* double mutants exhibited slightly higher levels of CG and CHG methylation compared to *rdr2* single mutants (Fig. 2a, b). ROS1 is known to remove methylation from both TEs and gene proximal regions in wild-



type plants, many of which are also targeted by RdDM<sup>16</sup>. The majority of methylation decreases observed in BR lines were at transposable elements (TEs) and intergenic regions (Fig. 2c; Supplementary Fig. 3). Regions in each individual BR line that had reduced methylation compared to *rdr2* were also hypomethylated in all other BR lines (Fig. 2d; Supplementary Fig. 3), but not in BRmut lines. Thus, independent BR lines lost methylation at shared loci when the methylation-sensing circuit was disrupted. In addition, all four BR lines showed clearly reduced methylation at a subset of ROS1 targets (defined as 1366 loci that significantly gained methylation in any sequence context in *ros1* compared to WT) compared to *rdr2* (Fig. 2e), consistent with restoration of wild-type ROS1 expression levels in BR lines. There was also a clear reduction in CG methylation at a subset of RdDM targets (6947 loci that are CHH hypomethylated in *rdr2* mutants compared to WT) in all BR lines (Fig. 2f), but not in BRmut lines. Between 13 and 51% of RdDM targets lost more than 20% CG methylation in BR lines. We confirmed these methylation losses at high depth for two loci using bisulfite-PCR (Supplementary Fig. 3). ROS1 is therefore broadly active at a large number of RdDM targets. RdDM mutants like *rdr2* likely retain most CG methylation at these loci because ROS1 expression is downregulated by the rheostat mechanism. Restoring ROS1 expression to wild-type levels in *rdr2* upsets the homeostatic balance, tipping it towards demethylation and resulting in widespread methylation losses.

### Disrupting epigenetic homeostasis causes dynamic instability.

As BR lines exhibit unstable and progressively worsening phenotypes due to demethylation of *SDC* (Fig. 1), we sought to determine if methylation is gradually lost over multiple generations elsewhere in the genome. We performed whole-genome bisulfite sequencing for three self-fertilized BR lines on leaves from individual second, third, and fourth generation plants derived by single seed descent. Unexpectedly, at a genome-wide scale, there was the least methylation in second-generation BR plants in all sequence contexts (Fig. 3a; Supplementary Table 1; Supplementary Fig. 4). However, this effect was transient. Average genomic CG and non-CG methylation was partially restored in the third generation, and overall normal levels of methylation were present at most loci in the fourth generation (Fig. 3a; Supplementary Table 1; Supplementary Fig. 4). The reduction of methylation in the second generation occurred at thousands of discrete loci, which were most abundant near the centromeres (Fig. 3b, orange line). By contrast, we observed that the subset of methylation losses shared among all three generations (and thus likely inherited) were abundant across chromosome arms but reduced near the centromere (Fig. 3b, black line). This suggests that although second generation BR lines exhibit the majority of methylation losses around the centromere, losses outside of those regions are more likely to persist heritably in future generations. Consistent with this, chromatin states associated with promoters or with intergenic sequences enriched for the Polycomb mark H3K27me3 showed methylation losses that persisted across four generations (Supplementary Fig. 5), whereas methylation was completely restored over four generations in heterochromatin

(Supplementary Fig. 5), the predominant chromatin state around the centromere (Fig. 3b). To determine whether regions with inherited methylation losses behaved like *SDC* and lost methylation from one generation to the next, we identified regions with reduced methylation in the fourth generation compared to *rdr2* and examined their methylation history in previous generations. Similar to *SDC*, these loci showed a progressive loss of methylation over multiple generations in all three BR lines (Fig. 3c). A subset of genomic loci in chromosome arms therefore behaves in an opposite manner to the majority of the genome, and continues to lose methylation upon disruption of the methylation-sensing circuit.

The methylation dynamics we observed in regions were also present at the level of individual CG pairs (Fig. 3d–f; Supplementary Fig. 6). We performed k-means clustering (with  $k=4$ ) on individual CGs, comparing their methylation level in the BR lines to the parent *rdr2* genotype (Supplementary Fig. 6). We identified similar clusters in all three BR lines. One cluster (termed Type 1 CGs) behaved like methylcytosines at *SDC*, showing gradually or consistently reduced methylation compared to *rdr2* in all three generations (Fig. 3d). A second cluster (termed Type 2 CGs) behaved like the genomic average, losing methylation in the second generation, but regaining methylation over subsequent generations (Fig. 3d). Type 3 CGs exhibited increased methylation in BR lines compared to *rdr2* (Supplementary Fig. 6). However, the majority of these CGs were also hypermethylated in *rdr2*; *ros1* mutants and BRmut lines, suggesting they did not change as a result of ROS1 activity. Type 1 CGs were relatively rare around the centromere and were clustered in gene-rich chromosome arms (Fig. 3e; Supplementary Fig. 7), consistent with the location of regions with heritable methylation changes (Fig. 3b). Conversely, Type 2 CGs were abundant in the heterochromatic pericentromere, where CG methylation is most abundant in *rdr2* (Fig. 3e; Supplementary Fig. 7). The methylation restoration at Type 2 CGs, which we validated at high depth for two loci using bisulfite-PCR (Supplementary Fig. 8), occurred in all sequence contexts: CG and CHG (methylated by the enzymes MET1 and CMT3 in WT, respectively) and in CAA and CTA sequences (Supplementary Fig. 4), which are methylated by the enzyme CMT2<sup>17</sup>. Methylation restoration must occur independently of the canonical RdDM pathway because BR lines are *rdr2* mutants. This process is likely mediated by multiple non-RdDM pathways that methylate heterochromatic DNA<sup>18</sup>. Consistent with their targeting by non-RdDM pathways, Type 2 CGs were flanked by cytosines more highly methylated in all sequence contexts in the *rdr2* background than the cytosines flanking Type 1 CGs (Fig. 3f; Supplementary Fig. 7).

### Discussion

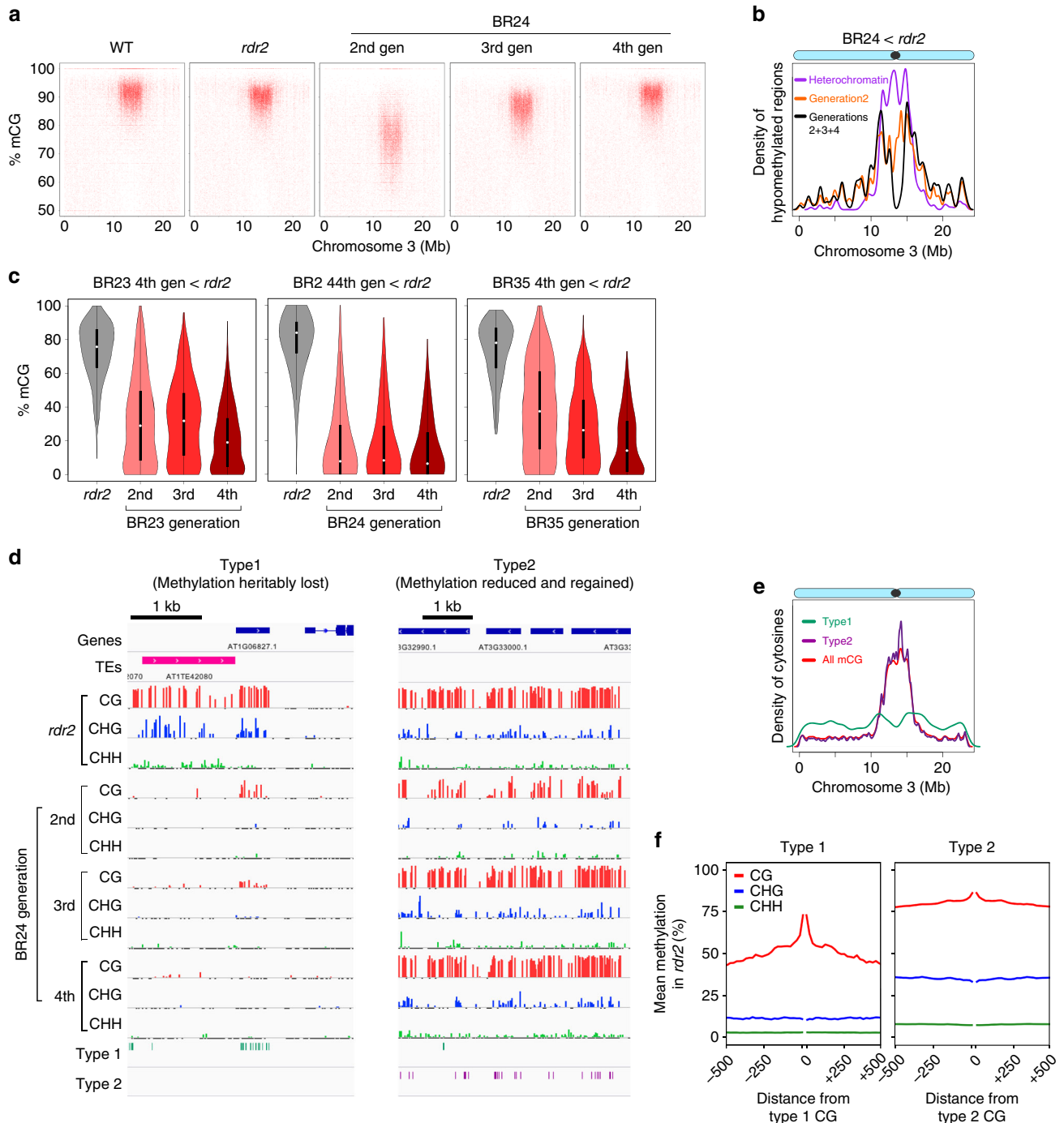
We have demonstrated that disrupting the methylation-sensing circuit integrated at ROS1 results in destabilization of epigenetic states. Our findings suggest that although RdDM pathway mutants, like *rdr2*, have lost substantial CHH methylation, they are in fact epigenetically stable because the ROS1-rheostat maintains homeostasis, simply at a lower genome-wide

**Fig. 2** Disrupting the methylation-sensing circuit causes genome-wide methylation losses. **a** A representative genome browser snapshot of whole-genome bisulfite sequencing data from the indicated genotypes. All BR and BRmut line data in this figure are from fourth generation plants. **b** The total sequence length (in kb) that has significantly gained or lost methylation compared to WT (*rdr2* and *ros1* (top)) or compared to *rdr2* (*rdr2*; *ros1*, BR and BRmut lines (bottom)). Significant methylation changes were determined for 300 bp windows by scoring the number of differentially methylated CG, CHG, or CHH context cytosines as a proportion of total cytosines in each window. **c** Proportion of regions in BR line 23 with reduced methylation compared to *rdr2* (in any sequence context) that overlap different genomic features. **d** Box and whisker plots of methylation levels of individual CGs at the most strongly hypomethylated regions from BR line 23. Hypickers represent 1.5 times the interquartile range. **e** Box and whisker plots of methylation levels for individual CGs overlapping ROS1 targets. **f** Box and whisker plots of methylation levels for individual CGs overlapping RdDM targets



methylation level than in the wild-type. Unstable methylation inheritance has only previously been documented in severe mutants that lose the majority of genomic DNA methylation and have widespread transposition, such as *met1*<sup>19</sup>. Breaking the

rheostat mechanism by restoring *ROS1* expression in *rdr2* creates an unstable state, resulting in abnormal phenotypes that are not observed in *rdr2*. The severity of these phenotypes worsens over generations due to the progressive loss of methylation at *SDC*, a



**Fig. 3** Uncoupling methylation and demethylation activities causes dynamic methylome instability over multiple generations. **a** Scatterplots of average CG methylation levels in 300 bp windows across chromosome 3 in wild-type (WT) plants, *rdr2* mutants and three generations of BR line 24. **b** Kernel density plot showing regions that are hypomethylated in BR line 24 compared to *rdr2* on chromosome 3. The orange line denotes all regions hypomethylated in the second generation, whereas the black line shows the subset of regions that remain hypomethylated in the third and fourth generations. The purple line denotes the density of heterochromatin across chromosome 3. **c** Violin plots of average CG methylation levels at loci hypomethylated in fourth generation BR plants compared to *rdr2*. **d** Representative snapshots of two DNA methylation dynamics that occur over multiple generations when *ROS1* expression is restored in *rdr2* mutants. Type 1 CGs exhibit gradual or immediate losses of methylation that are inherited over multiple generations. Type 2 CGs are hypomethylated in the second generation but regain methylation over multiple generations. **e** Kernel density plots showing the distribution of Type 1 and Type 2 CGs from BR line 24 over chromosome 3. The red line denotes the density of all methylated CGs in *rdr2*. **f** Average methylation levels of the *rdr2* mutant in sequences proximal to Type 1 and Type 2 CGs. Methylation levels are calculated for sequences up to 500 bp upstream and downstream

phenomenon that is evident at many other loci (Fig. 3c). Yet, mechanisms exist to counteract unstable inheritance (Fig. 3a), restricting the degradation of methylation to the subset of the genome that is primarily methylated solely by the RdDM pathway in the wild-type<sup>20,21</sup>. Progressive remethylation of hypomethylated DNA over a few generations has been observed previously, but was proposed to be mediated by RdDM and did not occur when *RDR2* was mutated<sup>22–25</sup>. Our results demonstrate that the canonical RdDM pathway is in fact dispensable for the reversion of methylation losses in heterochromatin. This implies that other methyltransferases, perhaps CMT2, which is primarily active in heterochromatin<sup>11,18</sup>, are capable of reversing reduced methylation. Despite the operation of this apparent push-back mechanism, the methylation landscape does not return to the parental *rdr2* state, particularly in gene-rich regions, where methylation patterning becomes increasingly corrupted over time and results in DNA methylation losses that persist as long-term “epigenetic scars”. Thus, a tightly coupled equilibrium between demethylation and methylation pathways is required to preserve the transgenerational stability of the epigenome. It is likely that similar safeguards exist to preserve epigenetic inheritance over multiple cell divisions and generations in other eukaryotes.

## Methods

**Plant material.** All plants were grown in a Conviron BDR16 growth chamber with 16 h of light (120  $\mu$ Mol), 22 °C, and 50% humidity. The *rdr2* and *ros1* mutant alleles used in this study were *rdr2-1* and *ros1-7* as described<sup>13</sup> and are in the Col-0 background. *ros1-7* has a missense mutation in the DNA glycosylase catalytic domain (E956K). Wild-type and homozygous mutant plants used in this study were all derived from the self-fertilized progeny of an *rdr2*; *ros1* heterozygous parent. The *ROS1* 5' inverted repeat (IR) transgene used to generate BR lines and BRmut lines was previously described<sup>13</sup>. BR lines were generated by transforming *rdr2* mutants with the IR transgene by floral dipping, whereas BRmut lines were generated by floral dipping *rdr2*; *ros1* mutants. T<sub>1</sub> generation transgenic lines were selected on 0.5  $\times$  MS with 25  $\mu$ g/ml kanamycin and 25  $\mu$ g/ml hygromycin. T<sub>2</sub> transgenic lines were initially screened for single insertions by antibiotic selection. T<sub>2</sub> plants used in subsequent assays were grown directly on soil and the presence of the IR transgene was determined by PCR. Each BR or BRmut line is derived from an independent remethylation of *ROS1* and represents biological replicates.

**Plant phenotyping.** Three-week-old adult plants were imaged using a Nikon 1 J1 camera under ambient lighting. A phenotype score (ranging from 0 to 5) was designated to quantify the severity of curled leaf phenotypes in transgenic lines. All plant genotype names were blinded by an independent researcher to ensure unbiased assessment of phenotypes. Plants were manually assessed and scored according to the following criteria: 0 = no leaves show curled laminae; 1 = one or more leaves show curled laminae; 2 = more than half of leaves show some degree of curled laminae; 3 = more than half of leaves are curled and at least one leaf has severely curled laminae; 4 = More than half of leaves have severely curled laminae; 5 = all leaves have severely curled laminae.

Severely curled refers to leaves in which the adaxial surface has curled sufficiently to touch the abaxial surface (e.g., curvature of almost 360 degrees).

**Locus-specific bisulfite-PCR and sequencing.** DNA was extracted from 3-week-old leaves of individual BR line and BRmut line plants, as well as WT (Col-0), *rdr2* and *rdr2*; *ros1*. 200 ng of DNA was bisulfite converted using an Invitrogen MethylCode bisulfite conversion kit. PCR was performed using 2  $\mu$ l bisulfite converted DNA and ExTaq HS DNA polymerase (Clontech). Primers are listed in Supplementary Table 2. PCR cycling conditions were 95 °C for 3 min, 40 cycles of (95 °C for 15 s, 50 °C for 30 s, 67.5 °C for 30 s, 72 °C for 30 s), followed by 67.5 °C for 10 min. PCR products were gel extracted and cloned using a CloneJET PCR cloning kit (Thermo Scientific). Between 24 and 37 individual clones for each sample were amplified by colony PCR and sequenced using a T7 promoter primer. Sequenced clones were aligned using MUSCLE<sup>26</sup> and methylation of cytosines was calculated using Cymate<sup>27</sup>. For the analysis of *SDC* methylation in second (T<sub>2</sub>) and fifth (T<sub>5</sub>) generation BR lines, DNA was extracted from 2-week-old leaf tissue pooled from approximately 20 plants of each genotype/generation. Bisulfite conversion and PCR were performed as specified above. Pooled tissue was divided and the same tissue was used for the extraction of RNA to perform RT-qPCR, as specified below.

**RT-qPCR.** For the data in Supplementary Figs. 1 and 2, RNA was extracted from 3-week-old leaves of individual T<sub>2</sub> BR line and BRmut line plants, as well as WT, *rdr2* and *rdr2*; *ros1* mutants. Genomic DNA was removed using amplification-grade

DNaseI (Invitrogen). cDNA was synthesized from 1  $\mu$ g RNA using Superscript II reverse transcriptase (Invitrogen) according to manufacturers' instructions, selecting for polyadenylated transcripts using an oligo-dT primer. qPCR was performed using Fast SYBR-Green PCR master mix (Applied Biosystems) and a StepOne Plus Real-Time PCR system (Applied Biosystems). Primers were designed as previously specified<sup>13</sup>. The reference gene AT1G58050 was used to normalize all reactions, as described<sup>28</sup> (Supplementary Table 2). All reactions were performed in technical triplicate, except for the data in Fig. 1d, which were performed in biological triplicates. For the quantification of *SDC* transcripts in T<sub>2</sub> and T<sub>5</sub> generation BR lines, RNA was extracted from pooled tissue from leaves of ~20 2-week-old plants of each genotype/generation. Three pooled samples were collected for each genotype to provide biological replicates.

**Whole-genome bisulfite sequencing.** Single generation samples (Fig. 2): DNA was extracted from 3-week-old whole rosettes of three independent T<sub>4</sub> BR lines, two independent T<sub>4</sub> BRmut lines, *ros1*, *rdr2*, and *rdr2*; *ros1* mutants and WT Col-0. Overall, 200 ng DNA was bisulfite converted using an Invitrogen MethylCode bisulfite conversion kit. Libraries were constructed using an Illumina TruSeq DNA methylation kit and sequenced on an Illumina HiSeq2000 using a 40  $\times$  40, 50  $\times$  50, 75  $\times$  75 or 100  $\times$  100 bp paired-end protocol at the Whitehead Institute Genome Technology Core (see Supplementary Table 1 for details). Read quality was assessed with FastQC and adapters were trimmed using Trim Galore (Babraham Bioinformatics). Reads were mapped to the TAIR10 genome using Bismark<sup>29</sup>, allowing for 1 mismatch for 40 bp and 50 bp reads and 2 mismatches for 75 and 100 bp reads. Bismark was also used to remove PCR duplicates, and the Bismark methylation extractor function was used to obtain methylation data from all mapped reads. Efficiency of bisulfite conversion was verified by quantifying the percentage of methylation mapped to the chloroplast (Supplementary Table 1). Multiple generation samples (Fig. 3): DNA was extracted from the fifth and sixth true leaves of individual T<sub>2</sub>, T<sub>3</sub>, and T<sub>4</sub> plants from three BR lines. T<sub>2</sub>, T<sub>3</sub>, and T<sub>4</sub> generation plants were derived from a single seed descent experiment (for example, the T<sub>3</sub> plants were the self-fertilized progeny of the sequenced T<sub>2</sub> plants). Bisulfite libraries were constructed and mapped as described above.

**Identifying differentially methylated regions.** To identify regions of the genome that were differentially methylated between two samples, a “window score” was calculated based on the number and density of differentially methylated cytosines within a given window. This approach was devised to minimize the false-positive discovery of differentially methylated regions in which only a small number of cytosines are differentially methylated. The following conditions were imposed: First, the genome was divided into 300 bp tiled windows that overlapped by 150 bp. Second, methylation data for cytosines in the CG context were averaged with the symmetric cytosine on the opposite strand. This resulted in one methylation value for each CG pair, so that cytosines in a symmetric CG dimer are not treated as independent. Third, only cytosines covered by 5 or more reads in both samples were considered. Fourth, cytosines were considered differentially methylated if their percentage methylation level differed by the following intervals: CG pairs: >30%, CHG: >25%, CHH: >20% (window scores are calculated separately for each methylation context). A methylation score was calculated by assigning +1 for every hypermethylated cytosine or CG pair and -1 for every hypomethylated cytosine or CG pair in a given window. This score was then multiplied by the proportion of cytosines in the window (with sufficient data for analysis) that were differentially methylated. The higher the score above 0, the greater the confidence a window is hypermethylated. The lower the score below 0, the greater the confidence a window is hypomethylated. This approach has several advantages over many conventionally used methods. First, two cytosines in a symmetric CG dimer are not treated as independent data points. Second, cytosines that are not well covered in both samples do not affect the window score. In the analysis presented in Figs. 2 and 3, hypermethylated regions were chosen as windows with a score of 3 or greater, and hypomethylated regions were chosen as windows with a score of -3 or lower. Bedtools intersect was used to determine the proportion of differentially methylated windows overlapping TEs, genes, non-TE intergenic regions and pseudogenes according to the Araport 11 genome annotation. To quantify the methylation state of individual CGs at *ROS1* and RdDM targets (Fig. 2e, f), a set of *ROS1* and RdDM target loci was defined based on their window score. *ROS1* targets were defined as merged windows with a score of 3 or greater in any cytosine context in *ros1* relative to WT (1366 loci identified). RdDM targets were defined as merged windows with a score of -3 or lower in the CHH context in *rdr2* compared to WT (6947 loci identified). Boxplots in Fig. 2e, f show the percentage average methylation value of all individual CGs covered by 5 or more reads that are located within *ROS1* or RdDM targets.

**k-means clustering.** Clustering was performed on symmetric CG pairs. For each BR line, only CGs covered by 5 or more reads in *rdr2* and in all three generations were considered (approximately one quarter of all possible cytosines met this coverage threshold). Methylation values for each BR line sample were subtracted from *rdr2* and the values of methylation differences were clustered by a *k*-means algorithm using Cluster 3.0 (Stanford University). A total of 1000 iterations were run and the iteration that produced the smallest within-cluster distances was

chosen. Clustering was performed with different values for  $k$  (4,5,6,8). Results obtained with  $k=4$  were chosen, as increasing  $k$  above 4 did not identify new variants. Methylation differences at the cytosines identified in each cluster were calculated between *rdm2* and *rdm2; ros1*, and between *rdm2* and BRmut lines, to control for changes to methylation independent of *ROS1*. The distance between the closest Type 1 or Type 2 cytosines was calculated using bedtools closest. The distance distributions expected by chance were calculated from five randomly selected datasets (of the same size as the Type 1 or Type 2 dataset) of cytosines methylated in *rdm2*. Kernel density plots of cytosines from different clusters were generated using R. Bedtools intersect was used to determine the proportion of cytosines in each cluster overlapping TEs, genes, non-TE intergenic regions and pseudogenes according to the Araport 11 genome annotation. The methylation profile of sequences surrounding Type 1 or Type 2 cytosines was generated as described<sup>30</sup>.

**Other analyses of bisulfite-seq data.** The methylation state of different chromatin states was determined using the chromatin states identified by Sequeira-Mendes et al.<sup>31</sup> Chromatin states 1 and 2 were combined to represent “Transcription start sites and promoters”, chromatin states 4 and 5 were combined to represent “Intergenic H3K27me3” and chromatin states 8 and 9 were combined to represent “heterochromatin”. Violin plots of the percent methylation of individual CGs (covered by 5 or more reads) were generated using R. The percentage methylation of different sequence contexts over whole chromosomes (Supplementary Fig. 4) was calculated in R using a moving average function, averaging methylation over 10 kb windows overlapping by 5 kb.

**Data availability.** All sequencing data is available in NCBI GEO under record GSE104240. The authors declare that all other data supporting the findings of this study are available within the manuscript and its supplementary files or are available from the corresponding author upon request.

Received: 31 July 2017 Accepted: 14 November 2017

Published online: 14 December 2017

## References

- Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).
- Miura, A. et al. Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature* **411**, 212–214 (2001).
- Gonzalo, S. et al. DNA methyltransferases control telomere length and telomere recombination in mammalian cells. *Nat. Cell Biol.* **8**, 416–424 (2006).
- Mirouze, M. et al. Loss of DNA methylation affects the recombination landscape in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **109**, 5880–5885 (2012).
- Yelina, N. E. et al. DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in *Arabidopsis*. *Genes Dev.* **29**, 2183–2202 (2015).
- Satyaki, P. R. V. & Gehring, M. DNA methylation and imprinting in plants: machinery and mechanisms. *Crit. Rev. Biochem. Mol. Biol.* **52**, 163–175 (2017).
- Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome—biological and translational implications. *Nat. Rev. Cancer* **11**, 726–734 (2011).
- Schmitz, R. J. et al. Transgenerational epigenetic instability is a source of novel methylation variants. *Science* **334**, 369–373 (2011).
- Becker, C. et al. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* **480**, 245–249 (2011).
- Hagmann, J. et al. Century-scale methylome stability in a recently diverged *Arabidopsis thaliana* lineage. *PLoS Genet.* **11**, e1004920 (2015).
- Stroud, H. et al. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat. Struct. Mol. Biol.* **21**, 64–72 (2014).
- Gehring, M., Reik, W. & Henikoff, S. DNA demethylation by DNA repair. *Trends Genet.* **25**, 82–90 (2009).
- 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).
- Lei, M. et al. Regulatory link between DNA methylation and active demethylation in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **112**, 3553–3557 (2015).
- 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).
- Penterman, J. et al. DNA demethylation in the *Arabidopsis* genome. *Proc. Natl. Acad. Sci. USA* **104**, 6752–6757 (2007).
- Gouil, Q. & Baulcombe, D. C. DNA methylation signatures of the plant chromomethyltransferases. *PLoS Genet.* **12**, e1006526 (2016).
- Zemach, A. et al. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205 (2013).
- Mathieu, O. et al. Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. *Cell* **130**, 851–862 (2007).
- Zhong, X. et al. DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat. Struct. Mol. Biol.* **19**, 870–875 (2012).
- Zheng, Q. et al. RNA polymerase V targets transcriptional silencing components to promoters of protein-coding genes. *Plant J.* **73**, 179–189 (2013).
- Chan, S. W.-L. et al. RNA silencing genes control de novo DNA methylation. *Science* **303**, 1336–1336 (2004).
- Chan, S. W.-L., Zhang, X., Bernatavichute, Y. V. & Jacobsen, S. E. Two-step recruitment of RNA-directed DNA methylation to tandem repeats. *PLoS Biol.* **4**, e363 (2006).
- Teixeira, F. K. et al. A role for RNAi in the selective correction of DNA methylation defects. *Science* **323**, 1600–1604 (2009).
- Wibowo, A. et al. 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**, e13546 (2016).
- Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
- Hetzl, J., Foerster, A. M., Raidl, G. & Scheid, O. M. CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J.* **51**, 526–536 (2007).
- Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W.-R. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* **139**, 5–17 (2005).
- Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
- Picard, C. L. & Gehring, M. Proximal methylation features associated with nonrandom changes in gene body methylation. *Genome Biol.* **18**, 73 (2017).
- Sequeira-Mendes, J. et al. The functional topography of the *Arabidopsis* genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* **26**, 2351–2366 (2014).

## Acknowledgements

We thank Katherine Novitzky and Elena Kingston for valuable assistance with locus-specific bisulfite-PCR and Colette L. Picard for advice on bioinformatics analysis. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award R01GM112851 to M. G.

## Author contributions

B.P.W. and M.G. conceived and designed experiments, B.P.W. performed experiments and analyzed data; B.P.W. and M.G. wrote the paper.


## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-017-02219-3>.

**Competing interests:** The authors declare that they have no competing financial interests.

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

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

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

© The Author(s) 2017