

Epigenetic Alterations at Genomic Loci Modified by Gene Targeting in *Arabidopsis thaliana*

Michal Lieberman-Lazarovich^{1,2*}, Cathy Melamed-Bessudo², Sylvia de Pater³, Avraham A. Levy²

1 Department of Plant Biology, University of Geneva, Geneva, Switzerland, **2** Department of Plant Sciences, The Weizmann Institute of Science, Rehovot, Israel, **3** Department of Molecular and Developmental Genetics, Institute of Biology, Leiden University, Leiden, The Netherlands

Abstract

Gene Targeting (GT) is the integration of an introduced vector into a specific chromosomal site, via homologous recombination. It is considered an effective tool for precise genome editing, with far-reaching implications in biological research and biotechnology, and is widely used in mice, with the potential of becoming routine in many species. Nevertheless, the epigenetic status of the targeted allele remains largely unexplored. Using GT-modified lines of the model plant *Arabidopsis thaliana*, we show that the DNA methylation profile of the targeted locus is changed following GT. This effect is non-directional as methylation can be either completely lost, maintained with minor alterations or show instability in the generations subsequent to GT. As DNA methylation is known to be involved in several cellular processes, GT-related alterations may result in unexpected or even unnoticed perturbations. Our analysis shows that GT may be used as a new tool for generating epialleles, for example, to study the role of gene body methylation. In addition, the analysis of DNA methylation at the targeted locus may be utilized to investigate the mechanism of GT, many aspects of which are still unknown.

Citation: Lieberman-Lazarovich M, Melamed-Bessudo C, de Pater S, Levy AA (2013) Epigenetic Alterations at Genomic Loci Modified by Gene Targeting in *Arabidopsis thaliana*. PLoS ONE 8(12): e85383. doi:10.1371/journal.pone.0085383

Editor: Xiaoyu Zhang, University of Georgia, United States of America

Received: September 29, 2013; **Accepted:** November 26, 2013; **Published:** December 26, 2013

Copyright: © 2013 Lieberman-Lazarovich et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the European Union FP6 TAGIP (LSHG-CT-2005-018785) and FP7 RECBREED (KBBE-2008-227190) grants (URL: http://cordis.europa.eu/home_en.html). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: michal.lieberman@unige.ch

Introduction

The ability to modify genomes in a precise manner provides a powerful tool for both basic and applied research [1,2]. It can be used to study gene function, to perform gene therapy or to improve crop plants. Several methods have been developed to allow for precise modifications of genomes. These include gene targeting (GT), DNA cleavage and mutagenesis by custom-designed endonucleases, site-specific recombination (e.g. Cre-Lox recombination), and oligo-mediated mutagenesis. These techniques enable targeted modifications, such as gene knock-outs, gene replacement, mutations or insertions and are routinely implemented in model organisms such as yeast [3] and mice [4].

It was shown, for example, that the prolonged clotting times in a mouse model of haemophilia B can be corrected by in vivo GT [5], raising the possibility of using this strategy for the treatment of genetic disease. Another study demonstrated an efficient GT procedure in *Nannochloropsis* sp., a fast-growing, unicellular algae capable of accumulating large amounts of oil [6]. In higher plants, the low frequency of site-directed homologous recombination hindered GT from becoming

routine. The frequency of GT events detected in early studies was in the range of 10^{-3} to 10^{-6} [7]. The finding that DNA double-strand breaks (DSBs) induce homologous-recombination-driven repair in plants [8] was exploited for enhancing GT by 2-3 orders of magnitude, using custom-designed nucleases such as Zinc Finger Nucleases (ZFNs) [9,10] Transcription Activator-Like Effector Nucleases (TALENs) [11] and the recently Clustered, Regularly Interspaced, Short Palindromic Repeats (CRISPR)-associated protein (Cas)[12]. It is therefore possible that GT will also become routine in plants in the near future, facilitating efforts to deal with challenges such as food and fuel security. Regardless of the technology employed, precise engineering of genomes, via GT, is assumed to enable controlled expression of the introduced gene as a result of its insertion into the original genomic context. However, the introduced DNA is devoid of the epigenetic modifications which have been established as critical in determining chromatin structure and gene expression patterns [13-16]. Thus, GT might bear an impact on the epigenetic status of the targeted allele, a hypothesis which was largely unexplored. To test this, we assessed the effect of GT on the pattern of DNA cytosine

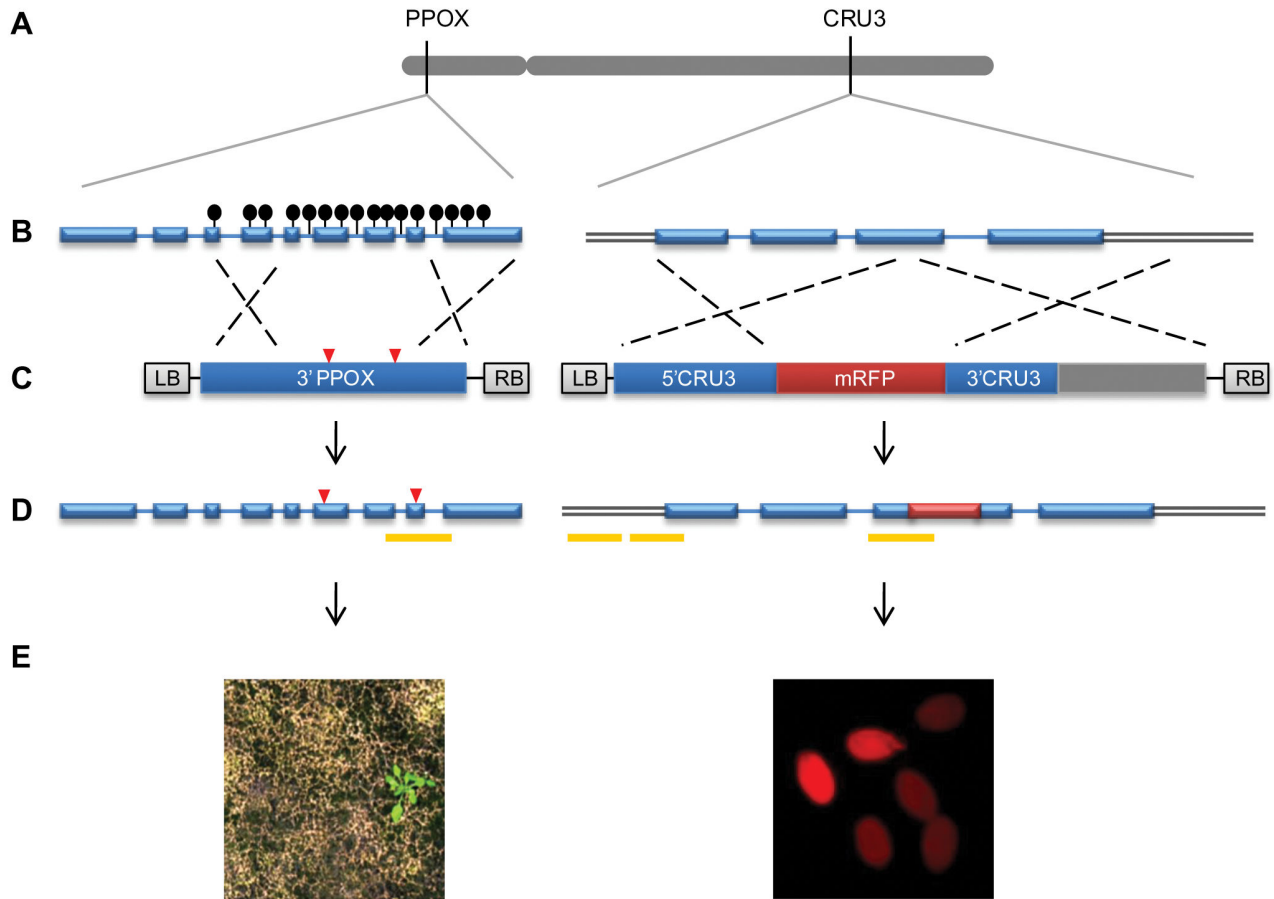


Figure 1. Genomic targets and gene targeting (GT) assays. (A) The *PPOX* and *CRUCIFERIN3* target loci are both located on chromosome 4 (in grey). (B) The target gene models with methylated cytosines, are shown as black lollipops. (C) The GT vectors are shown with indication of the left and right borders (LB and RB, respectively) of the *Agrobacterium tumefaciens* vector; red triangles indicate the two point mutations that confer herbicide resistance; mRFP (monomeric red fluorescence protein) was used as positive targeting marker; dashed lines indicate recombination between homologous genomic sequences. (D) The projected structure of the targeted loci following precise recombination with the targeting vectors is shown; Yellow bars correspond to the regions analyzed by bisulfite-sequencing. Blue rectangles, blue lines and double grey lines represent exons, introns and adjacent genomic regions, respectively. (E) *PPOX*-targeted plants are resistant to Butafenacil treatment (left) and *CRUCIFERIN3*-targeted seeds show red-fluorescence (right).

doi: 10.1371/journal.pone.0085383.g001

methylation, by analyzing two endogenous *Arabidopsis* target loci, both located on chromosome 4. The two loci differ in their methylation landscape, thus offering a contrasting perspective. The *PPOX* gene (At4g01690) is naturally enriched in CG genobody methylation, in a region that encompasses one of the two bases to be modified by the targeting vector (Figure 1). In contrast, the *CRUCIFERIN3* gene (At4g28520) fully lacks cytosine methylation in the WT background [17,18] (Figure 1). Using these two genomic targets, we could determine whether there is a change, of hypo- or hypermethylation at the targeted allele.

Materials and Methods

For the GT experiments, *Arabidopsis thaliana* plants (ecotype *Columbia*) were grown in controlled chambers at 21°C, with 16/8 hours of light/dark, respectively. The GT vector was transformed into inflorescences using the floral-dip method. Mature seeds were collected and screened for successful GT events. Genetic characterization of the GT lines was performed by PCR and Southern-blot assays. Methylation analysis of targeted lines was performed using the EpiTect kit (QIAGEN) for bisulfite conversion [19] of genomic DNA, according to the manufacturer's instructions with one modification: the number of incubation cycles was increased in

order to ensure full conversion. Genomic DNA was extracted with the DNeasy Plant Mini Kit (QIAGEN), according to the manufacturer's protocol. Converted DNA was used for amplification of the target and PCR products were cloned and Sanger-sequenced. The number of clones that were sequenced in each experiment is indicated in the relevant supplementary table. For the *PPOX* locus analysis, pools of 4-8 young *Arabidopsis* plantlets were used in each sample. For the *CRUCIFERIN3* and *PROHIBITIN1* methylation analyses mature plant tissue were sampled, by pooling 2-6 plants from each line. Conversion rate was estimated at 99 to 100%, by sequencing a fragment of the non-methylated *PHAVOLUTA* gene (At1g30490). Sequence analysis for calculating methylation fractions and for generating methylation plots was performed by the Kismeth software [20]. In order to independently determine the methylation patterns of the endogenous *PROHIBITIN1* and the copied *PROHIBITIN1*, we took an approach of size separation; the genomic DNA was digested with *PvuII*, which generates a fragment of around 8300bp at the endogenous *PROHIBITIN1* locus, and a 4000bp fragment at the duplicated gene. Digested DNA was extracted from the gel in two separate reactions, corresponding to the 8300bp and 4000bp fragments, and then used for bisulfite conversion and sequencing.

Results

Gene replacement of the endogenous *PPOX* locus, by a vector presenting one or two SNPs (Single Nucleotide Polymorphisms) that modify specific amino acids, confers herbicide resistance [21]. We analyzed three independent GT-derived lines by Southern blotting and sequencing, to confirm that gene targeting resulted in the precise replacement of the endogenous *PPOX* locus. These targeted lines did not contain randomly integrated copies of the vector in the genome (Figure S1), and were thus considered True Gene Targeting (TGT) events and consequently named TGT-1 [21], TGT-2, and TGT-3 (originally named C1 and T2, respectively [22]). Bisulfite-sequencing was performed on the CG-methylated region, encompassing the second SNP introduced by the GT vector (Figure 1), to enable comparison of the post-targeting methylation state of the targeted allele to that of the WT. In addition, methylation status in the progeny of the initial GT events was also evaluated. Two types of DNA methylation modifications were identified, that may have occurred at the targeted *PPOX* locus. The TGT-1 line, generated in the wild-type (WT) background (ecotype *Wassilewskija*, *Ws*), lost half of the total CG methylation in the examined region, reaching a level of 35.2%, compared to 69.7% in the wild-type *Ws* plant (Table S1). Methylation was virtually obliterated in the central region, encompassing six of the eleven CG sites as well as the second SNP introduced by the GT vector (Figure 2). This prominent pattern may possibly reflect the location and length of the conversion tract in the process of homology-mediated GT. The last CG site in this sequence, at position 320 (Figure 2), was methylated in TGT-1, but not in the WT (*Ws*). We found this specific site to be particularly unstable, as even in WT isolates of the same ecotype (*Columbia*), both full and lack of

methylation was seen (Table S2). Since other CG sites in this region showed a consistent methylation pattern in different WT isolates, position 320 can be considered an outlier. To address the question of whether the observed loss of methylation has an influence on gene expression, we carried out expression analysis of the *PPOX* gene using two different primer pairs. As presented in Figure S2, the *PPOX* transcript level did not differ between TGT-1 and the *Ws* line used as control. In contrast to the local depletion of methylation in the *PPOX* allele of the TGT-1 line, the overall methylation level was not dramatically changed in TGT-2 and TGT-3 [22], both in the *Columbia* background (Figure 3a). While the mean WT (*Columbia*) methylation level was 68.1%, the rates of CG methylation in the targeted allele of TGT-2 and TGT-3 were 66.4% and 74.4%, respectively (Table S1). Most CG sites (7 out of 11) presented methylation levels similar to WT, however, methylation at three sites (at positions 64, 155 and 194) was altered (Figure 3b) (Table S2). The CG dinucleotide at position 64 differed from the other two sites, presenting sharp increase in methylation only in the T3 generation of the TGT-2 line (Figure 3b). The CG sites at positions 155 and 194 exhibited altered methylation levels already in the T2 generation. Interestingly, this trend of change was further enhanced in the following generation (T3) (Figure 3b). At position 155, the WT methylation level was moderate (65%, Table S2). In the TGT-2 line, this position had lost some methylation in the first generation after GT, and completely lost all methylation in the T3 generation. In contrast, the TGT-3 line had gained methylation at this site, which rose to 90-100% in T3. In position 194, methylation was very low in WT (7%) but increased in the T2 generation of both TGT-2 and TGT-3 lines, and showed a further increase in T3, reaching methylation levels as high as 78-80% (Figure 3b). Non-CG methylation (i.e. CHG, CHH) was very low in this region (0 to 2.55% in WT, Table S1), a status that was maintained in the targeted lines, with the exception of the targeted allele of TGT-3, where CHG methylation was higher (Table S1). In order to confirm that changes in DNA methylation did not occur globally due to the transformation process, we have tested the methylation status of a non-target locus, which, similarly to the *PPOX* gene, presents CG body methylation in the WT background. As shown in Figure S3, no changes in CG body-methylation were observed for this locus (At2g33860) in all GT lines used in this study. In the TGT-1 line, a single CG site differs in methylation from the *Ws* WT line, however, it is unlikely that this is the result of GT, as the TGT-1 pattern is in agreement with the published methylome of the *Ws* ecotype [23]. In addition, this specific site is found to be unstable among various *Arabidopsis* ecotypes [23]. Gene targeting events at the *CRUCIFERIN3* locus, which lacks cytosine methylation, can be identified using a seed-fluorescent assay previously described [24,25] and shown in Figure 1. We analyzed methylation in plants derived from fluorescent seeds corresponding to two GT events. The first event gave rise to a precise gene targeting allele (CRU-TGT), where the WT locus was replaced by the GT vector and with no additional non-homologous integration in the background (Figure S4). Three regions were analyzed in WT, non-targeted sibling, and the homozygous CRU-TGT line, in T2 and/or T3 generations. One

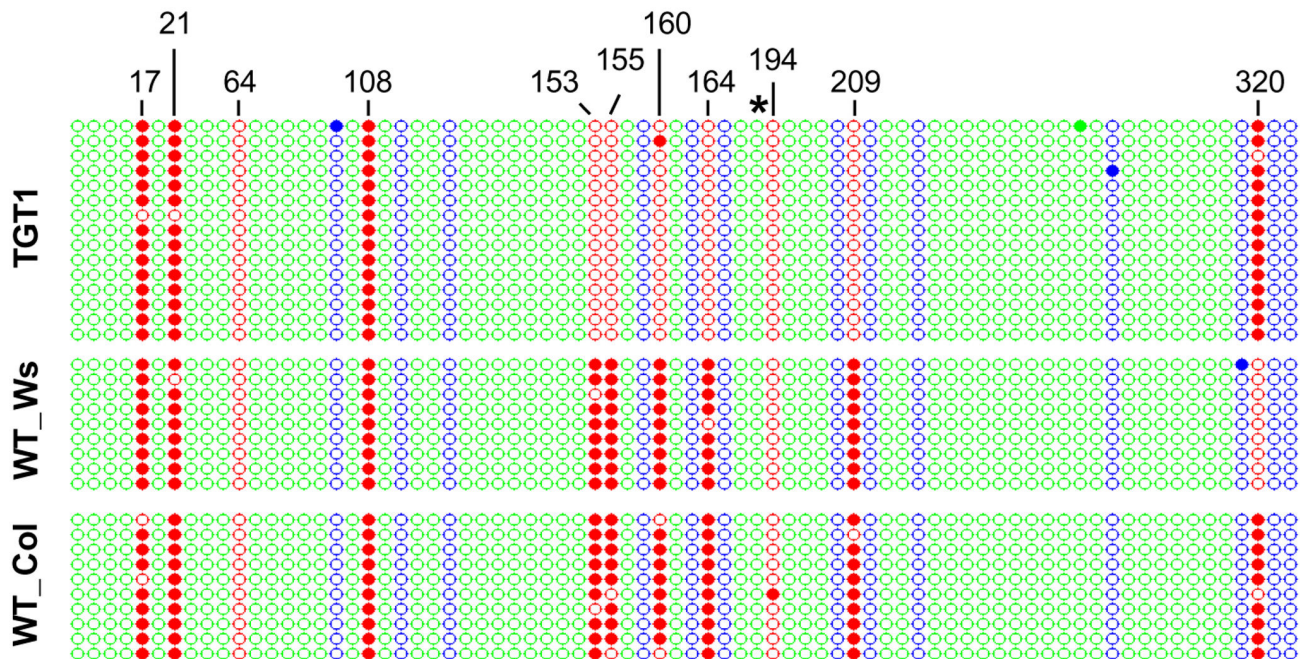


Figure 2. Loss of CG methylation at the targeted *PPOX* locus in the TGT-1 line. Cytosine methylation data obtained from bisulfite-sequencing of the *PPOX* fragment in the targeted line TGT-1 and two WT lines from ecotypes *Columbia* (*Col*) and *Wassilewskija* (*Ws*). Each circle represents a cytosine residue, either methylated (full circle) or non-methylated (empty circle). Cytosines are color-coded by their sequence context: red for CG, blue for CHG and green for CHH (H is C, T or A). Each row represents an independent clone. The numbers at the top of the columns indicate the position in base pairs, relative to the first base in this fragment. The second SNP introduced by the GT vector, is shown by an asterisk.

doi: 10.1371/journal.pone.0085383.g002

region was located upstream to the *CRUCIFERIN3* promoter, beyond the region of homology between the GT cassette and the genomic target, the second, was around the *CRUCIFERIN3* translation start site, a region of homology between the GT vector and the target, and the third, was at the junction between the *CRUCIFERIN3* gene and the introduced mRFP reporter gene (yellow bars in Figure 1). For all 3 regions, no alteration in GT-related cytosine methylation was observed, indicating that the occurrence of a precise GT did not change the methylation state of a target which was formerly unmethylated (Table S3). The second GT event was not precise; as frequently happens in GT experiments, the vector invades its homologous target, but instead of replacing the target, it dissociates and integrates at an ectopic location, often capturing target sequences, while the endogenous target is left intact [7]. Such events are thus referred to as Ectopic Gene Targeting (EGT). Detailed genetic characterization of this EGT event is presented in Figure S5. The methylation patterns of the *PROHIBITIN1* gene (*PHB1*, AT4G28510) that was duplicated upon this EGT event (Figure 4) were analyzed. In contrast to the unmethylated *CRUCIFERIN3* target, the upstream gene – *PHB1* – is naturally enriched with body CG methylation. Thus, in the EGT line, we could test the methylation state of the newly duplicated *PHB1* gene as well as that of its original template locus. To this end, we separately

isolated the endogenous *PHB1* and the duplicated copy for bisulfite-sequencing analysis (17). T2 and T3 generations of this EGT line were also analyzed, in order to test whether methylation of the duplicated copy is stable throughout generations. Interestingly, we found that cytosine methylation was altered following the GT-derived gene duplication at both the endogenous and the duplicated alleles, and further changed from T2 to T3 generation, in both CG and non-CG sites (Figure 4, Figure S6 and Tables S4 and S5). Increased methylation was observed at all CG dinucleotides analyzed in T2, with a most profound increase at position 18, which was unmethylated prior to targeting. In the subsequent generation (T3), methylation levels at position 18 reverted back to WT levels (0%). At the other CG sites, a decrease in methylation in the duplicated copy was observed, although not reconstructing the WT state as at position 18 (Figure 4, Figure S6 and Table S4). Although non-CG methylation is absent at the *PHB1* locus in WT, it appeared to be highly induced upon GT-derived duplication, as most (68.4%) non-CG sites were found to be 90-100% methylated in both the endogenous and the duplicated allele in the T2 generation. Interestingly, this methylation profile reverted back to the WT state already in T3, when only traces of non-CG methylation could be detected (Table S5).

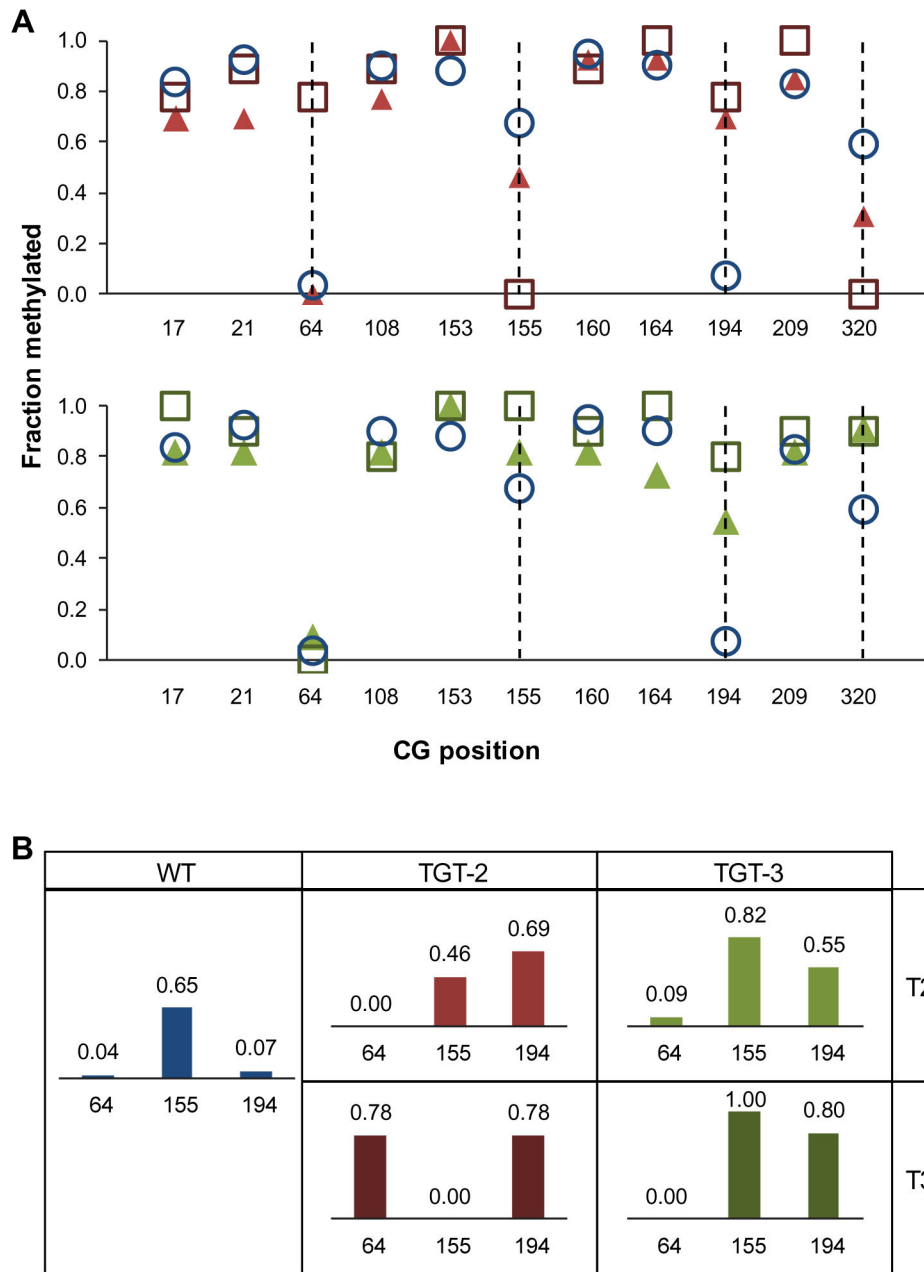


Figure 3. Methylation changes at the targeted *PPOX* locus in lines TGT-2 and TGT-3. (A) The methylation landscape in TGT-2 and TGT-3 lines (upper and lower graphs, respectively). Blue circles, triangles and squares correspond to WT, T2 and T3 generations, respectively. Dotted lines mark the CG positions at which the methylation level had changed relative to WT. (B) CG positions that lost methylation stability in the two targeted lines. CG position is denoted beneath each bar. The numbers above each bar denote the level of methylation as fraction.
doi: 10.1371/journal.pone.0085383.g003

Discussion

Taken together, our results demonstrate the possible impact of GT on the epigenetic landscape of the targeted locus. We found that several types of epigenetic alterations may occur at

the targeted locus, even when the alteration is genetically precise. GT can induce complete removal of methylation, in a limited region of the target, or cause loss of stability at certain methyl-CG sites, an effect that can be further enhanced in the following generation. The changes observed suggest that the

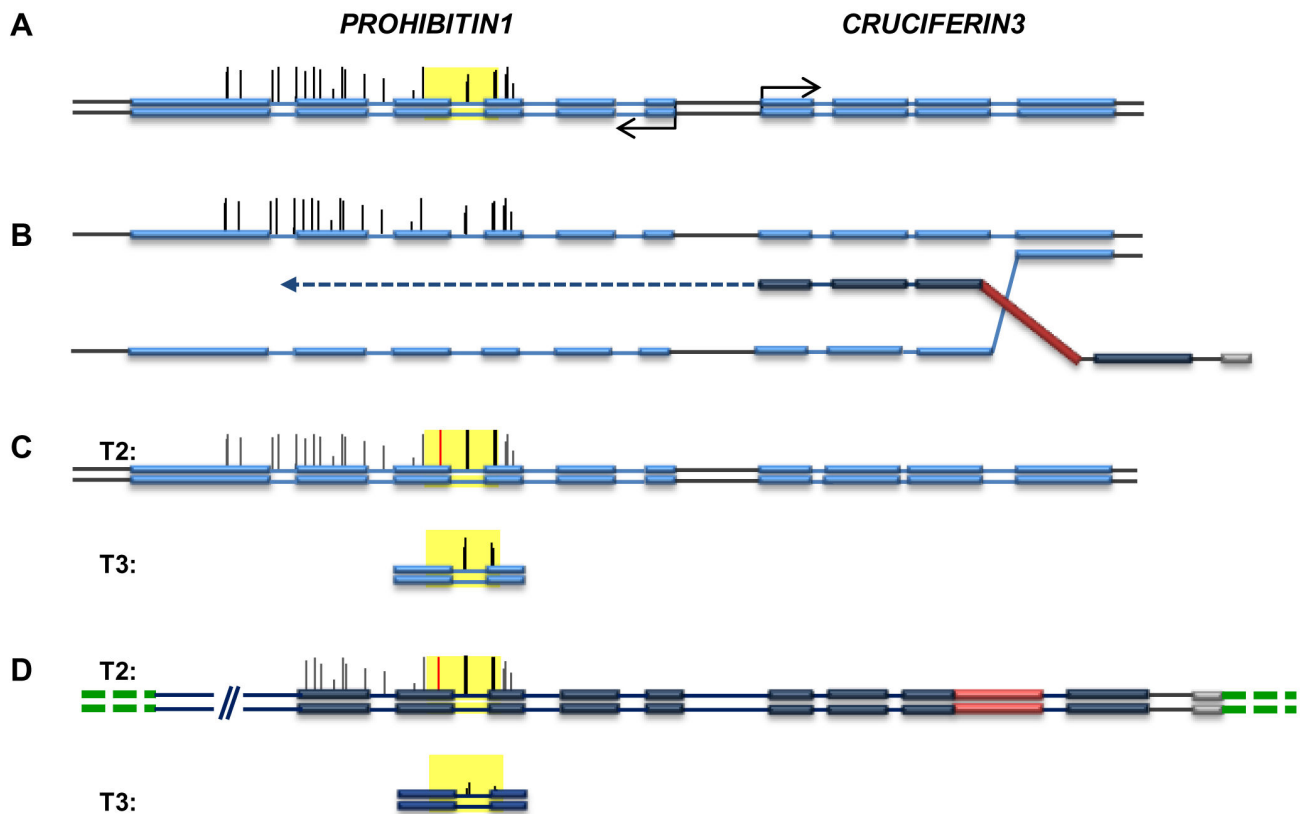


Figure 4. Ectopic Gene Targeting (EGT)-derived duplication of the *PROHIBITIN1* gene and subsequent changes in DNA methylation. (A) Schematic presentation of the WT *CRUCIFERIN3* locus, including the upstream *PROHIBITIN1* gene. Methylated cytosines are marked by vertical lines whose heights are proportional to the level of methylation (Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008). The region analyzed by bisulfite-sequencing is highlighted in yellow. (B) The GT vector, containing sequences homologous to the target (dark blue) as well as the mRFP fluorescent marker (red), invades the *CRUCIFERIN3* target and primes DNA synthesis (dark blue dotted arrow), copying the *PROHIBITIN1* sequence. (C) and (D) Genetic and epigenetic consequences of EGT: (C) Methylation status at the endogenous copy, in T2 (upper scheme) and T3 (lower scheme) generations after GT. The red vertical line marks the newly methylated CG at position 18. (D) The vector including the *PROHIBITIN1* fragment that had integrated into chromosome 1 (green dotted lines) creating gene duplication. The methylation status of this duplicated copy is shown, in T2 (upper scheme) and T3 (lower scheme) generations after GT. Grey bars: the vector RB (right border) sequences.

doi: 10.1371/journal.pone.0085383.g004

effect of GT on the methylation status of the targeted locus is not uniform or directional. The possible causes for the different effects on epigenetic alterations are not clear. One possibility is the different *Arabidopsis* ecotypes used for the initial GT experiments (i.e. Ws in the case of the TGT-1 line, and Col-0 for TGT-2 and 3 lines), both performed via floral-dip *Agrobacterium*-mediated transformation. It is known that different *Arabidopsis* ecotypes differ in their genome-wide methylation patterns, raising the question of whether differences in the mechanism maintaining DNA methylation exist as well, an issue that was never tested. If so, that could explain the different response to GT. It has been shown that different types of biotic stresses may cause epigenetic alterations [26], raising the possibility that the transformation event itself, carried out using bacterial infection, would trigger the observed changes in DNA methylation. To address that, we

tested the methylation status of a non-target locus, where we found no changes in methylation patterns (Figure S3), suggesting that the loss of methylation at the *PPOX* locus is indeed linked and caused by the GT process. In the case of the EGT line (Figure 4, Figure S6, Table S4 and Table S5), it is likely that the changes in DNA methylation were triggered by the duplication of the *PHB1* gene, similarly to methylation changes observed following genome duplication and polyploidization, through the involvement of siRNAs and the RdDM (RNA directed DNA Methylation) pathway. Transgene silencing is yet another well studied phenomenon where the presence of repeated sequences drives hypermethylation and silencing of both copies, mediated by RdDM. It is probable that GT-mediated demethylation, as shown here in plants, can also occur in mammalian systems, due to the conserved nature of homologous recombination mechanisms. Considering the

routine use of gene knockout in mammalian research this possibility should receive more attention. Indeed, earlier works have shown an alteration in allele imprinting as a result of targeted deletions or insertions [27,28], however, a detailed analysis of patterns of cytosine methylation was lacking. Note that the fate of the unmethylated DNA might differ between plants and mammals. While in mammals CG methylation can be regained de-novo, according to sequence determinants [29,30], in plants, de-novo methylation is usually of a non-CG type. Hence, loss of CG methylation in plants may result in accumulation of de-novo non-CG methylation [17,18] or, if the signal for de-novo methylation is lacking, then there is no compensation for the lost methylation and the region remains unmethylated. Epialleles have been shown to have potentially important phenotypic effects on development [31,32], in contributing to biodiversity [33] and in human disease, such as cancer [34]. In addition to its impact on gene expression and chromatin organization [13-16,35], DNA methylation was shown to be involved in essential cellular processes, such as meiotic recombination [36-38] and nucleosome positioning [39]. Therefore, the epigenetic status of a targeted locus should be taken into consideration in GT experiments, particularly when the target is methylated in the WT. In our case, we did not detect a change in expression of the *PPOX* gene, in seedlings of the TGT-1, 2, 3 lines (Figure S2). Nevertheless, since the exact role of gene body methylation is unclear, we cannot exclude an effect on another process and/or in a specific cell-type, developmental stage, or environmental condition. Targeted demethylation presents an ideal system to address the role of gene body methylation. The advances in technologies enhancing the precision and efficacy of gene targeting, such as Zinc Finger Nucleases [40], TALENs [11,41], and the more recent CRISPR [42-46], suggest that GT will become routine in several species in a near future, including plants [47-51]. In turn, the prospects of targeting genes for the sole purpose of changing their methylation status will rise. This may also provide a new tool for research or for applications in medicine and agriculture. Our results show that the changes in methylation provide new mechanistic insight into the GT

process. For example, obliteration of methylation patterns during precise GT, as with the *PPOX* TGT-1 allele, provides support for the double strand break (DSB) repair model [52], whereby the DSB in the target is followed by gap enlargement and repaired via strand invasion and copying of the unmethylated vector (Figure 5). The EGT event at the *CRUCIFERIN3* locus provides support for the GT model involving vector invasion into the target and initiation of DNA synthesis, as in the synthesis-dependent strand annealing model [53]. In this model, the vector is released from the target template, and inserted into a new location, thereby generating duplication of the target sequence as well as of its original methylation imprints (Figure 5). Remarkably, this type of duplication can induce epigenetic instability in both the genetically unaltered endogenous target region and in the new duplicated locus. The case of TGT-2 and TGT-3, where original methylation was maintained at most positions, can be explained by two different models, namely the model of strand assimilation [54] or a DSB repair model, where protruding ends are generated but without gap enlargement (Figure 5). In both cases, a template is present for methylation maintenance. Interestingly, we observed methylation instability at few CG sites of the region examined. Although the mechanism of this epigenetic instability is yet to be elucidated, it is reminiscent of the genetic instability reported during mitotic gene conversion in yeast [55] where mutations resulting from errors made by DNA polymerases seemed to be affected by the local sequence context. It is possible that the observed epigenetic instability reported here results from perturbations of the maintenance methyltransferase activity. Finally, we show (in TGT-1) that DNA methylation analysis of GT events provides novel insight into the analysis of the conversion tract predicted from the DSB repair model. Namely, by using methylation polymorphism, gap enlargement can be measured at a higher resolution than previously achieved, even in the absence of DNA polymorphism. We therefore propose that DNA methylation may be exploited in future experiments to better understand the mechanism(s) of GT in higher eukaryotes.

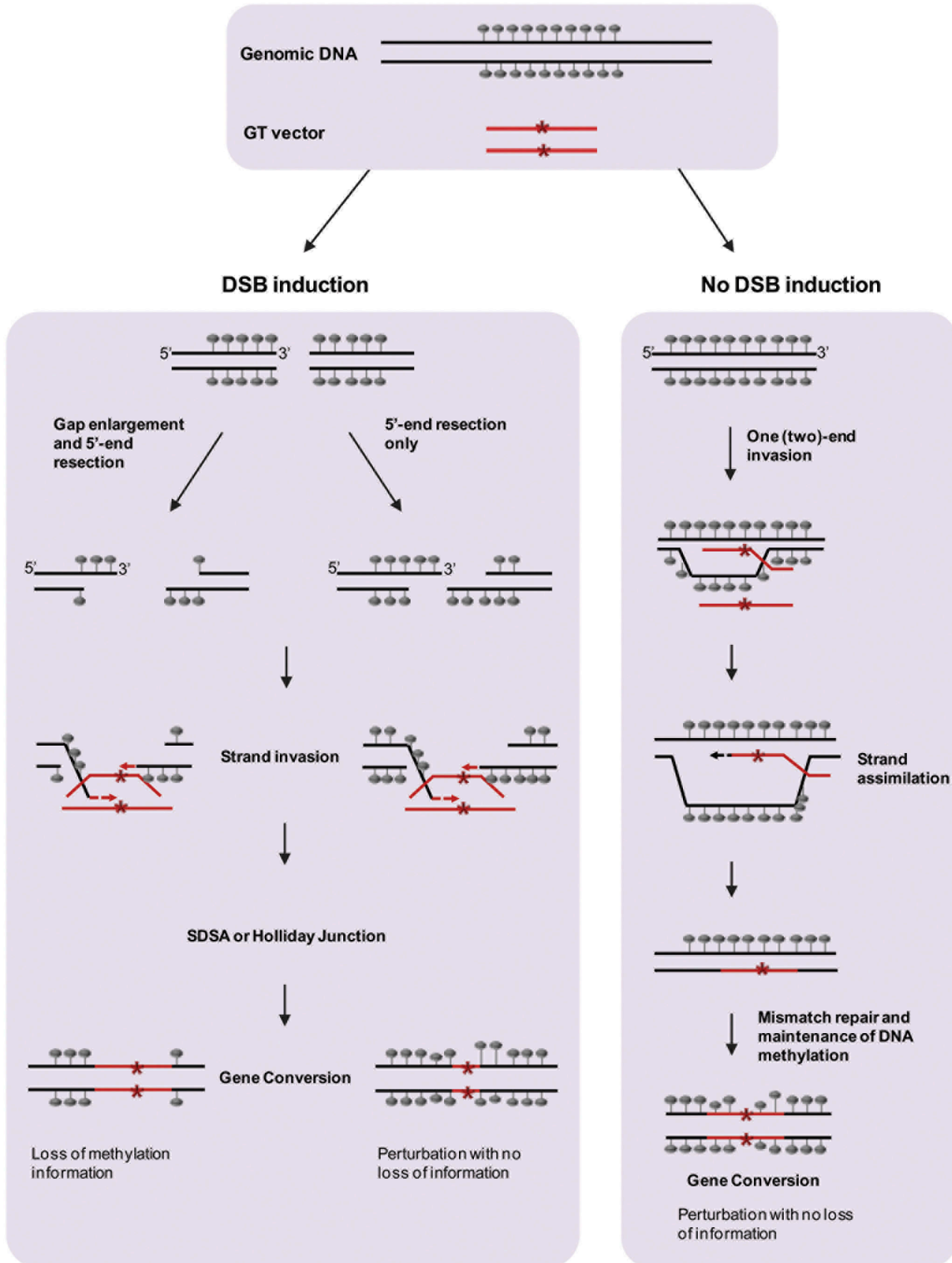


Figure 5. Mechanistic models for modification of methylation status of targeted alleles by GT. Methylated cytosines are indicated by grey lollipops. The GT vector is denoted by a red line, with a red asterisk illustrating the genetic perturbation introduced by the vector. If gap enlargement occurs following DSB induction (pathway on the left), the methylation template is lost and the targeted allele loses its methylation. However, if GT occurs via strand assimilation (pathway on the right) or via DSB induction but without gap enlargement (pathway in the middle), the methylation pattern of the WT can be restored upon maintenance methylation.

doi: 10.1371/journal.pone.0085383.g005

Supporting Information

Figure S1. Southern blot analysis of the *PPOX*-targeted lines. GT events at the *PPOX* locus were analyzed by Southern-blot hybridization using probe A [21]. Three plants of each line (different GT events or different generation) were analyzed. All plants of a given line gave the same result, therefore, only one plant per line is presented here as an example. As described previously for TGT-1 in the *Ws* background [21], using genomic DNA digested with *KpnI* and *NcoI*, the expected size for the GT allele is 2 Kb, whereas the WT allele and the T-DNA (if present) have the sizes of 11.6 Kb and 2.7 Kb, respectively. In the *Col-0* background (TGT-2 and TGT-3), expected sizes are 2 Kb, > 23 Kb and 4.2 Kb for the GT allele, WT allele and T-DNA, respectively. In this blot, we show that accurate GT was achieved and that no additional ectopic integration events occurred.
(TIF)

Figure S2. *PPOX* transcript level in the different GT lines. *PPOX* mRNA extracted from 7-day-old seedlings was measured by qPCR using two different primer pairs (**A**: Forward: 5'CGGGCTACGAAGGGCTAT3', Reverse: 5'ACCTCAATCGCGTTTCA3' and **B**: Forward: 5'GCCTCAAGCCATTCCTCA3', Reverse: 5'CTTCGTAGCCCGAAGACG3'). Error bars represent variation between two biological replicates (each replicate was measured twice). PDF2 (At1g13320) and SAND (At2g28390) were used as reference genes.
(TIF)

Figure S3. DNA methylation at the non-targeted locus At2g33860 is unchanged. Cytosine methylation data obtained from bisulfite-sequencing of a non-target locus, in all GT lines used in this study. Each circle represents a cytosine residue, either methylated (full circle) or non-methylated (empty circle). Cytosines are color-coded by their sequence context: red for CG, blue for CHG and green for CHH (H is C, T or A). Each row represents an independent clone. (**A**) Methylation pattern in the TGT-1 line, generated in the *Ws* background, and a *Ws*-WT control. (**B**) Methylation pattern in the TGT-2 and TGT-3 lines, generated in the *Col-0* background, compared to the respective *Col-0* WT control.
(TIF)

Figure S4. Genetic analysis of the CRU-TGT allele. (**A**) Schematic illustration of a precisely targeted *CRUCIFERIN3* locus (*CRU3*). P, promoter region. mRFP, the monomeric red fluorescence protein positive marker inserted by the GT vector. HindIII and PvuII restriction sites are marked by H and P, respectively. Fragments lengths correspond to the bands shown in Southern blots in c and d. Grey rectangle, probe. Black arrows indicate positions and directions of PCR primers used to confirm precise junctions at the targeted locus. (**B**) PCR results for the CRU-TGT line and two independent EGT lines. (**C**) Southern blot analysis of targeted T1 plants, using PvuII digestion, showing the TGT allele (5.7 Kb). (**D**) Southern blot analysis of T2 progeny of the CRU-TGT plant, using the

HindIII restriction enzyme. Analysis of only red-fluorescent seeds in T2 gave the expected 1 homozygous:2 heterozygous ratio ($p(\chi^2) = 0.73$).
(TIF)

Figure S5. Genetic analysis of the EGT allele at the *CRUCIFERIN3* locus. (**A**) Schematic illustration of the naïve *CRUCIFERIN3* locus (upper) and of the targeted locus (lower). P, promoter region. mRFP, the monomeric red fluorescence protein gene inserted by the GT vector. HindIII and PvuII restriction sites are marked by H and P, respectively. Fragment lengths correspond to the bands shown in Southern blots in b, c and d. Grey rectangles, probes. RB – the right border fragment of the GT vector that was found to be present in this EGT line. (**B**) Southern blot analysis of two EGT plants (T2 generation), using the “CRU” probe on HindIII-digested DNA. (**C**) Same membrane as in ‘b’ hybridized with the ‘CIFERIN’ probe, presenting the unexpected 4.2 Kb band. (**D**) Southern blot analysis of T3 progeny of the EGT line, using the PvuII restriction enzyme and the ‘CIFERIN’ probe. All plants tested had the WT allele.
(TIF)

Figure S6. CG methylation at the *PROHIBITIN1* gene. The level of methylation is shown for each CG position in the fragment analyzed, in T2 (**A**) and T3 (**B**) generations. Red lines represent the level of methylation in a WT line, whereas the bars represent methylation of three alleles tested: the duplicated allele (black), the endogenous allele (grey) and a non-targeted sibling allele (white).
(TIF)

Table S1. Average methylated fraction for each *PPOX*-targeted line.
(DOC)

Table S2. *PPOX* methylation fraction in different *Col-0* WT strains.
(DOC)

Table S3. Average methylated fraction at the CRU3 locus in the CRU-TGT line.
(DOC)

Table S4. Average CG-methylated fraction at the duplicated PHB1 locus in the EGT line.
(DOC)

Table S5. Non-CG-methylated fraction at the duplicated PHB1 locus in the EGT line.
(DOC)

Acknowledgements

We thank Prof. Jerzy Paszkowski for the TGT1 allele and for his support, Prof. Paul Hooykaas for the TGT2 and TGT3 alleles, and Dr. Aviva Samach, Dr. Liron Even-Faitelson, Dr.

Diego H. Sanchez and Dr. Mayumi Iwasaki for technical assistance.

Author Contributions

Conceived and designed the experiments: MLL AAL. Performed the experiments: MLL CMB. Analyzed the data:

MLL. Contributed reagents/materials/analysis tools: SP. Wrote the manuscript: MLL AAL.

References

- Capecchi MR (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* 6: 507-512. doi:10.1038/nrg1619. PubMed: 15931173.
- Mak TW (2007) Gene targeting in embryonic stem cells scores a knockout in Stockholm. *Cell* 131: 1027-1031. doi:10.1016/j.cell.2007.11.033. PubMed: 18083089.
- Rothstein R (1991) Targeting, disruption, replacement, and allele rescue - integrative dna transformation in yeast. *Method Enzymol* 194: 281-301. doi:10.1016/0076-6879(91)94022-5.
- Bogue CW (2003) Genetic models in applied physiology - Invited review: Functional genomics in the mouse: Powerful techniques for unraveling the basis of human development and disease. *J Appl Physiol* 94: 2502-2509. PubMed: 12736192.
- Li H, Haurigot V, Doyon Y, Li T, Wong SY et al. (2011) In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* 475: 217-U128. doi:10.1038/nature10177. PubMed: 21706032.
- Kilian O, Benemann CSE, Niyogi KK, Vick B (2011) High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proc Natl Acad Sci U S A* 108: 21265-21269. doi:10.1073/pnas.1105861108. PubMed: 22123974.
- Hanin M, Paszkowski J (2003) Plant genome modification by homologous recombination. *Curr Opin Plant Biol* 6: 157-162. PubMed: 12667873.
- Puchta H, Dujon B, Hohn B (1993) Homologous recombination in plant-cells is enhanced by in-vivo induction of double-strand breaks into dna by a site-specific endonuclease. *Nucleic Acids Res* 21: 5034-5040. doi: 10.1093/nar/21.22.5034. PubMed: 8255757.
- Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA et al. (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459: 437-U156. doi:10.1038/nature07992. PubMed: 19404259.
- Townsend JA, Wright DA, Winfrey RJ, Fu F, Maeder ML et al. (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* 459: 442-U161. doi:10.1038/nature07845. PubMed: 19404258.
- Zhang Y, Zhang F, Li X, Baller JA, Qi Y et al. (2013) Transcription Activator-Like Effector Nucleases Enable Efficient Plant Genome Engineering. *Plant Physiol* 161: 20-27. doi:10.1104/pp.112.205179.
- Carroll D (2013) Staying on target with CRISPR-Cas. *Nat Biotechnol* 31: 807-809. doi:10.1038/nbt.2684. PubMed: 24022156.
- Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* 39: 61-69. doi:10.1038/ng1929. PubMed: 17128275.
- Rigal M, Kevei Z, Pélissier T, Mathieu O (2012) DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns. *EMBO J* 31: 2981-2993. doi:10.1038/embj.2012.141. PubMed: 22580822.
- Rhee I, Bachman KE, Park BH, Jair KW, Yen RWC et al. (2002) DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* 416: 552-556. doi:10.1038/416552a. PubMed: 11932749.
- Deleris A, Stroud H, Bernatavichute Y, Johnson E, Klein G et al. (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-e1003062. doi:10.1371/journal.pgen.1003062. PubMed: 23209430.
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B et al. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452: 215-219. doi:10.1038/nature06745. PubMed: 18278030.
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC et al. (2008) Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell* 133: 523-536. doi:10.1016/j.cell.2008.03.029. PubMed: 18423832.
- Frommer M, McDonald LE, Millar DS, Collis CM, Watt F et al. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* 89: 1827-1831. doi:10.1073/pnas.89.5.1827. PubMed: 1542678.
- Gruntman E, Qi Y, Slotkin RK, Roeder T, Martienssen RA et al. (2008) Kismeth: Analyzer of plant methylation states through bisulfite sequencing. *Bmc Bioinformatics* 9: 371. doi:10.1186/1471-2105-9-371. PubMed: 18786255.
- Hanin M, Volrath S, Bogucki A, Briker M, Ward E et al. (2001) Gene targeting in Arabidopsis. *Plant J* 28: 671-677. PubMed: 11851913.
- de Pater S, Pinas JE, Hooykaas PJJ, van der Zaal BJ (2013) ZFN-mediated gene targeting of the Arabidopsis protoporphyrinogen oxidase gene through Agrobacterium-mediated floral dip transformation. *Plant Biotechnol J* 11: 510-515. doi:10.1111/pbi.12040. PubMed: 23279135.
- Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M et al. (2013) Patterns of population epigenomic diversity. *Nature* 495: 193-198. doi: 10.1038/nature11968. PubMed: 23467092.
- Even-Faitelson L, Samach A, Melamed-Bessudo C, Avivi-Ragolsky N, Levy AA (2011) Localized egg-cell expression of effector proteins for targeted modification of the Arabidopsis genome. *Plant J* 68: 929-937. doi:10.1111/j.1365-313X.2011.04741.x. PubMed: 21848915.
- Shaked H, Melamed-Bessudo C, Levy AA (2005) High-frequency gene targeting in Arabidopsis plants expressing the yeast RAD54 gene. *Proc Natl Acad Sci U S A* 102: 12265-12269. doi:10.1073/pnas.0502601102. PubMed: 16093317.
- Downen RH, Pelizzola M, Schmitz RJ, Lister R, Downen JM et al. (2012) Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci U S A* 109: E2183-E2191. doi:10.1073/pnas.1209329109. PubMed: 22733782.
- Ripoche MA, Kress C, Poirier F, Dandolo L (1997) Deletion of the H19 transcription unit reveals the existence of a putative imprinting control element. *Genes Dev* 11: 1596-1604. doi:10.1101/gad.11.12.1596. PubMed: 9203585.
- Tsai TF, Bressler J, Jiang YH, Beaudet AL (2003) Disruption of the genomic imprint in trans with homologous recombination at *Snrpn* in ES cells. *Genesis* 37: 151-161. doi:10.1002/gene.10237. PubMed: 14666508.
- Lienert F, Wirbelauer C, Som I, Dean A, Mohn F et al. (2011) Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* 43: 1091-1097. doi:10.1038/ng.946. PubMed: 21964573.
- Wienholz BL, Karetka MS, Moarefi AH, Gordon CA, Ginno PA et al. (2010) DNMT3L Modulates Significant and Distinct Flanking Sequence Preference for DNA Methylation by DNMT3A and DNMT3B In Vivo. *PLOS Genet* 6: e1001106.
- Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 401: 157-161. doi: 10.1038/43657. PubMed: 10490023.
- Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJJ et al. (2007) Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. *Plant J* 49: 38-45. doi:10.1111/j.1365-313X.2006.02936.x. PubMed: 17144899.
- Becker C, Hagmann J, Müller J, Koenig D, Stegle O et al. (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. *Nature* 480: 245-U127. doi:10.1038/nature10555. PubMed: 22057020.
- Hesson LB, Hitchins MP, Ward RL (2010) Epimutations and cancer predisposition: importance and mechanisms. *Curr Opin Genet Dev* 20: 290-298. doi:10.1016/j.gde.2010.02.005. PubMed: 20359882.
- Soppe WJJ, Jasencakova Z, Houben A, Kakutani T, Meister A et al. (2002) DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. *EMBO J* 21: 6549-6559. doi: 10.1093/embj/cdf657. PubMed: 12456661.
- Melamed-Bessudo C, Levy AA (2012) Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in

- heterochromatic regions in Arabidopsis. *Proc Natl Acad Sci U S A* 109: E981-E988. doi:10.1073/pnas.1120742109. PubMed: 22460791.
37. Mirouze M, Lieberman-Lazarovich M, Aversano R, Bucher E, Nicolet J et al. (2012) Loss of DNA methylation affects the recombination landscape in Arabidopsis. *Proc Natl Acad Sci U S A* 109: 5880-5885. doi:10.1073/pnas.1120841109. PubMed: 22451936.
 38. Yelina NE, Choi K, Chelysheva L, Macaulay M, de Snoo B et al. (2012) Epigenetic Remodeling of Meiotic Crossover Frequency in Arabidopsis thaliana DNA Methyltransferase Mutants. *PLoS Genet* 8: e1002844. PubMed: 22876192.
 39. Chodavarapu RK, Feng S, Bernatavichute YV, Chen P-Y, Stroud H et al. (2010) Relationship between nucleosome positioning and DNA methylation. *Nature* 466: 388-392. doi:10.1038/nature09147. PubMed: 20512117.
 40. Carroll D (2011) Genome Engineering With Zinc-Finger Nucleases. *Genetics* 188: 773-782. doi:10.1534/genetics.111.131433. PubMed: 21828278.
 41. Liu J, Li C, Yu Z, Huang P, Wu H et al. (2012) Efficient and Specific Modifications of the Drosophila Genome by Means of an Easy TALEN Strategy. *J Genet Genomics* 39: 209-215. PubMed: 22624882.
 42. Cong L, Ran FA, Cox D, Lin S, Barretto R et al. (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 339: 819-823. doi:10.1126/science.1231143. PubMed: 23287718.
 43. Mali P, Yang L, Esvelt KM, Aach J, Guell M et al. (2013) RNA-Guided Human Genome Engineering via Cas9. *Science* 339: 823-826. doi:10.1126/science.1232033. PubMed: 23287722.
 44. Chang N, Sun C, Gao L, Zhu D, Xu X et al. (2013) Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. *Cell Res* 23: 465-472. doi:10.1038/cr.2013.45. PubMed: 23528705.
 45. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ et al. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 31: 227-229. doi:10.1038/nbt.2501. PubMed: 23360964.
 46. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31: 233-239. doi:10.1038/nbt.2508. PubMed: 23360965.
 47. Jiang W, Zhou H, Bi H, Fromm M, Yang B et al. (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res* 41: e188. doi:10.1093/nar/gkt780. PubMed: 23999092.
 48. Li JF, Norville JE, Aach J, McCormack M, Zhang D et al. (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nat Biotechnol* 31: 688-691. doi:10.1038/nbt.2654. PubMed: 23929339.
 49. Miao J, Guo D, Zhang J, Huang Q, Qin G et al. (2013) Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res* 23: 1233-1236. doi:10.1038/cr.2013.123. PubMed: 23999856.
 50. Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31: 691-693. doi:10.1038/nbt.2655. PubMed: 23929340.
 51. Shan QW, Wang YP, Li J, Zhang Y, Chen KL et al. (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol* 31: 686-688. doi:10.1038/nbt.2650. PubMed: 23929338.
 52. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. *Cell* 33: 25-35. doi:10.1016/0092-8674(83)90331-8. PubMed: 6380756.
 53. Ho CK, Mazón G, Lam AF, Symington LS (2010) Mus81 and Yen1 Promote Reciprocal Exchange during Mitotic Recombination to Maintain Genome Integrity in Budding Yeast. *Mol Cell* 40: 988-1000. doi:10.1016/j.molcel.2010.11.016.
 54. Leung W-Y, Malkova A, Haber JE (1997) Gene targeting by linear duplex DNA frequently occurs by assimilation of a single strand that is subject to preferential mismatch correction. *Proc Natl Acad Sci U S A* 94: 6851-6856. doi:10.1073/pnas.94.13.6851. PubMed: 9192655.
 55. Hicks WM, Kim M, Haber JE (2010) Increased Mutagenesis and Unique Mutation Signature Associated with Mitotic Gene Conversion. *Science* 329: 82-85. doi:10.1126/science.1191125. PubMed: 20595613.