

Received Date : 20-Jun-2011

Revised Date : 10-Aug-2011

Accepted Date : 11-Aug-2011

Article type : Technical Advance

## **Title page**

**Title: Localized egg-cell expression of effector proteins for targeted modification of the *Arabidopsis* genome**

**Authors:** Liron Even-Faitelson\*, Aviva Samach\*, Cathy Melamed-Bessudo, Naomi Avivi-Ragolsky, Avraham A. Levy

**Institution:** Department of Plant Sciences, The Weizmann Institute of Science, Rehovot, Israel.

**Corresponding Author:** Avraham A. Levy - Department of Plant Sciences, The Weizmann Institute of Science, Rehovot 76100, Israel. Telephone: 972-8-9342734. Fax: 972-89344181

E-mail: avi.levy@weizmann.ac.il

\* These authors contributed equally to this work.

**Running Title: Egg-cell protein expression for targeted genetic changes**

**Keywords:** Targeted mutagenesis, Gene targeting, Arabidopsis, Egg-cell enhancer, Zinc finger nuclease, RAD54

This is an Accepted Article that has been peer-reviewed and approved for publication in the *The Plant Journal*, but has yet to undergo copy-editing and proof correction. Please cite this article as an "Accepted Article"; doi: 10.1111/j.1365-313X.2011.04741.x

## Summary

Targeted modification of the genome is an important genetic tool, which can be achieved via homologous, non-homologous or site-specific recombination. Although numerous efforts have been made, such a tool does not exist for routine applications in plants. This work describes a simple and useful method for targeted mutagenesis or gene targeting, tailored to floral dip transformation in *Arabidopsis*, by means of specific protein expression in the egg-cell. Proteins stably or transiently expressed under the egg-apparatus specific enhancer (EASE) were successfully localized to the area of the egg-cell. Moreover, a zinc finger nuclease expressed under the EASE induced targeted mutagenesis. Mutations obtained under the EASE control corresponded to genetically independent events that took place specifically in the germline. In addition, RAD54 expression under the EASE led to an approximate tenfold increase in gene targeting efficiency, when compared to wild type plants. EASE-controlled gene expression provides a method for precise engineering of the *Arabidopsis* genome through temporally and spatially controlled protein expression. This system can be implemented as a useful method both in basic research in *Arabidopsis*, as well as in optimization of tools for targeted genetic modifications in crop plants.

## Introduction

A variety of methods have been developed to allow for precise modifications of genomes. These include: (i) gene targeting (GT) performed via targeted integration of exogenous DNA by means of homologous recombination (HR), (ii) site-specific recombination at a pre-introduced site, e.g. Cre-Lox recombination, (iii) DNA cleavage executed by engineered endonucleases, resulting in error-prone, non-homologous end-joining and subsequent targeted gene mutagenesis (TGM), and (iv) oligo-mediated mutagenesis. These techniques enable precise modifications, such as gene knockouts, mutations, or insertions, at specific genomic loci. They are routinely implemented in yeast (Rothstein, 1991) and mice (Bogue, 2003), deeming them advantageous model organisms. The model species for plant research is *Arabidopsis thaliana*: its genome has been sequenced (Arabidopsis Genome Initiative, 2000), there are many mutant lines (Alonso et al., 2003) and it is fairly easy to transform. Its transformation does not require tissue culture, but is typically performed by floral dipping (Clough and Bent, 1998; Zhang et al., 2006). Budding plants are dipped in an *Agrobacterium tumefaciens*-containing solution and are transformed by the genetically engineered bacterial transfer DNA (T-DNA). Transformed cells, which successfully transmit the T-DNA to the next generation, are those which stem from the female megaspore, including the egg cell (Desfeux et al., 2000). T-DNA integration into the plant genome occurs in a non-homologous, random manner (Kim et al., 2007; Tzfira et al., 2004); it is very difficult to achieve precise targeted integration, and even more so, homologous integration, in this model organism. Thus, development of an efficient method allowing for targeted genetic modifications in *Arabidopsis*, remains an important goal for plant research per se, as well as for applications in crop plants.

To date, few groups have successfully performed TGM or GT in *Arabidopsis*. Most of the reported TGMs in *Arabidopsis* have been obtained by use of engineered zinc finger nucleases (ZFNs) (de Pater et al., 2009; Lloyd et al., 2005; Osakabe et al., 2010; Tovkach et al., 2009;

Zhang et al., 2010), while more recently, via transcription activator-like effector nucleases (TALENs) (Boch et al., 2009; Cermak et al., 2011; Mahfouz et al., 2011). In principle, these nucleases can be engineered to cut any site along the genomic DNA, resulting in the formation of a double strand break (DSB). Repair of this double strand break by the non-homologous end-joining (NHEJ) mechanism, usually leads to small insertions/deletions (Gorbunova and Levy, 1997), hence resulting in TGM (Urnov et al., 2010; Weinthal et al., 2010). Moreover, it has been shown in plants that DSBs induce HR-driven repair (Puchta et al., 1993). Such repair mechanism can be used for making precise insertions or mutations, providing the appropriate template for HR is present. A recently published example of endogenous GT executed via an HR-driven repair mechanism mediated by ZFN DSB formation involved tobacco *Acetolactate synthase* gene targeting, yielding herbicide resistance, and showed GT frequencies of over 2% in tobacco protoplasts (Townsend et al., 2009). Similarly, Shukla *et al.* performed GT of the endogenous IPK1 gene in *Zea mays*, conferring herbicide tolerance and decreased phytate levels, two agronomically relevant outcomes (Shukla et al., 2009). Expression of HR-promoting proteins, such as the yeast RAD54, have also been applied to enhance GT frequencies in plants (Shaked et al., 2005). The efficiency and success of the noted GT and TGM methods largely depends on the spatial and temporal expression of the GT- or TGM-promoting proteins for the genetic modification to occur and pass to the next generation.

We hereby describe an approach whereby GT- and TGM-promoting proteins are expressed in a spatiotemporal-specific pattern that coincides with the use of the floral-dip transformation method. For this purpose, we developed a system in which the GT/TGM-promoting proteins are specifically expressed in the *Arabidopsis* egg cell, the target of *Agrobacterium* transformation with the floral-dip method, by exploiting the egg-apparatus specific enhancer (EASE), described by Yang, et al. (Yang et al., 2005). EASE is a characterized 77 bp enhancer sequence, which efficiently controls specific gene expression in the egg apparatus of *Arabidopsis* (Yang et al.,

2005). The work described here shows that ZFN and RAD54 proteins, expressed downstream to the EASE, catalyze TGM and GT at the desired time and location, resulting in induced, germline-restricted genetic modifications, which were, in turn, transmitted to the next generation.

## Results

### *Proteins expressed downstream to the EASE are localized to the egg cell*

Two EASE-regulated, egg-cell-localized protein expression systems were established in *Arabidopsis* for egg-cell specific expression of genome-modifying proteins (both are variations of the original system developed by Yang et al. (Yang et al., 2005)). The first system combines the EASE and the trans-activation system (Moore et al., 1998), in which a driver transgene, *EASE:LhG4*, produces a chimeric transcription factor in the egg-cell, that activates the responder transgene, *Op:mRFP* or *Op:ScRAD54* (*Op* stands for the Lac operator sequence that is bound by LhG4). Microscopic analysis of the mRFP reporter expression in EASE>>mRFP plants confirmed the egg-cell specific expression pattern of the EASE in the *Arabidopsis* ovule (Figure 1). An *EASE:LhG4* line showing strong mRFP expression was then selected for further crossing with *Op:ScRAD54* lines.

The second system positioned the gene of interest along the same binary vector as and directly downstream to the EASE (Figure 2a, *EASE:EGFP*). *Arabidopsis* plants were transformed with *EASE:EGFP* using the floral dip method. Ovules were dissected and visualized 2 and 6 days following transformation. At two days post-transformation, only one of 48 examined ovules showed weak expression of EGFP localized to the egg cell. Six days following transformation, 7% of the 59 ovules analyzed contained EGFP expression localized to the egg cell (Figure 2). The seeds from these plants were collected and *EASE:EGFP* T-DNA integration was assessed by measuring resistance to glufosinate (Basta). Only 0.44% of the 14400 seeds tested contained integrated *EASE:EGFP* sequences, indicating a significant difference between the number of

ovules demonstrating transient expression and the number that underwent stable genomic integration of the T-DNA. These results indicate that EASE can be applied for transient gene expression in the egg cell. Transient expression, as seen by the *EASE:EGFP* reporter gene, can be detected within ~ 2 days of transformation, and increases further over the following four days. Fluorescent proteins expressed under the EASE via the trans-activation and transient expression systems, localized to the region of the egg-cell (Figures 1 and 2), indicating the efficacy of both approaches in targeting protein expression to the egg cell.

### ***Developmental timing of ZFN activity and TGM frequency under the EASE***

In order to determine the timing of expression of genes regulated by EASE, we modified a system used by Tovkach *et al.* for the assessment of ZFN efficiency *in planta*. The system is based on NHEJ repair of a DSB made in a *GUS* gene (Tovkach *et al.*, 2009) mutated to contain a QQR-ZFN (Kim *et al.*, 1997) binding site immediately after the start ATG, followed by a stop codon and then another QQR-ZFN binding site (Figure 3a). Certain indels, resulting from the repair of the DSB, will lead to an in-frame amendment of the *GUS* gene and allow for its expression. For our assay, we transformed *Arabidopsis* plants with T-DNA containing a mutated *GUS* cassette (pRCS2.[KAN][QQR-TS\*::GUS]) (Figure 3a). The QQR-ZFN was chosen for testing the EASE system, because it is a well characterized ZFN (Bibikova *et al.*, 2001; Kim *et al.*, 1997) that has been shown previously to be functional *in planta*. Two stable lines with a single locus of the mutated *GUS* cassette were transformed with T-DNA bearing *QQR-ZFN* downstream to the EASE (*EASE:QQR-ZFN*) and hygromycin resistance. X-Gluc staining allowed for monitoring of the precise timing of the DNA cleavage performed by the EASE-driven QQR-ZFN, thereby suggesting the timing of the nuclease and repair activities. The interpretation of the outcomes of the assay is as described by Bechtold *et al.* (Bechtold *et al.*, 2003). In short, an entire blue

seedling and a blue endosperm (i.e. blue seed coat) indicate a very early cleavage and repair event, transpiring before the division of the gametophyte into the 8 nuclei stage (Figure 2a). An entire blue seedling lacking a blue seed coat indicates a cleavage event that occurred in either the egg cell or the very early zygotic stage. Finally, a “chimeric” seedling containing blue sectors indicates somatic cleavage events that occurred following a number of embryo divisions.

The first experiment was carried out on ~15,000 seedlings originating from seeds that were collected from plants bearing the mutated *GUS* and transiently transformed with *EASE:QQR-ZFN*. X-Gluc staining of four days-old seedlings revealed no blue staining. The second experiment was performed on second generation seedlings that originated from plants that were selected on kanamycin and hygromycin, indicating the presence of the mutated *GUS* cassette and the *EASE:QQR-ZFN* cassette, respectively. A total of ~9000 seedlings containing both the mutated *GUS* cassette and the genome integrated *EASE:QQR-ZFN*, were tested. In this case, seven seedlings contained blue staining throughout the entire seedling, with no staining of the seed coat (Figure 3b). No other staining patterns were observed, indicating that the QQR-ZFN under the EASE is active in the egg cell or the very early zygote, at early stages of development.

In order to confirm that the seven blue seedlings resulted from independent cleavage events, we performed PCR directly on the seedlings, using primers complementary to the QQR-ZFN break-site surrounding region, followed by sequencing (Figure 3c). Six of the seven seedlings had unique indels, indicating that they originated from independent mutagenesis events, each of which allowed for restoration of the reading frame of the *GUS* gene.

The rate of mutagenesis evaluated via the GUS staining assay, was of 7 out of 9000 seedlings, namely  $7.8 \times 10^{-4}$ . However, this method of assessment results in an underestimate of the actual TGM rate, as it only monitors those events that led to GUS expression. PCR using primers complementary to the QQR-ZFN break-site surrounding region was performed on randomly

selected second generation seedlings, in order to allow for more accurate determination of the rate of EASE-regulated, QQR-ZFN-driven targeted mutagenesis (Figure 4a). The PCR products were then incubated with the *DdeI* restriction enzyme, which has a unique recognition sequence within the QQR-ZFN break-site (Figure 4a and 4c). In cases of QQR-ZFN-driven DSB repair, indels abolishing the *DdeI* restriction site are formed (Figure 4b – see lane under arrow). This mode of analysis yielded one of 366 seedlings with a nullified *DdeI* restriction site, corresponding to a mutagenesis rate of  $2.7 \times 10^{-3}$ . While the present sample size is too small for performing a statistical analysis, the calculated mutagenesis rate is approximately three-fold that determined by means of the GUS staining assay. These findings compare favorably with the expectation that only one third of the QQR-ZFN-induced indels will result in a valid reading frame and to GUS expression. The PCR based assay still underestimates the actual rate of TGM, as it does not consider large deletions which could either obstruct PCR performance or result in small, undetectable PCR products. Additionally, break events later undergoing precise repair, are not detectable using this assay.

#### ***Expression of ScRAD54 in the egg-cell enhances GT events***

We tested the effect of egg-cell specific ScRAD54 expression on GT, using the *Cru3* GT assay, which was described in detail in Shaked *et. al*, 2005 (Shaked et al., 2005). In the present study, the *Cru3* GT assay was modified by replacing both the *GFP* with *mRFP*, fused to nucleotides 37-1217 of the *Cru3* ORF, and the *NOS* terminator with the *OCS* terminator. The terminator was switched due to a more convenient cloning procedure. Successful *Cru3* GT events result in easily detectable red fluorescent seeds. An additional assay alteration incorporated in the present model, involved the use of plants expressing EASE-regulated ScRAD54, rather than relying on the 35S promoter. Driver lines, expressing the LhG4 activator under the EASE, were selected for

Op:ScRAD54 expression, based on the intensity and specificity of expression of a Op:mRFP reporter transactivated by the same driver.

GT rates in *EASE>>ScRAD54 Arabidopsis* lines were compared to those of wild type *Arabidopsis* plants (WT) (Figure 5). In WT, only three GT events were isolated from the ~6,320,000 seeds screened in 27 independent experiments. The rate of GT in these experiments, namely the ratio of homologous versus non-homologous integration, was  $3.45 \times 10^{-5}$ , e.g. 3 GT events out of ~63,200 transformed seeds (see GT rate calculation in the Experimental Procedures). The *EASE>>ScRAD54* lines produced six GT events isolated from the ~2,400,000 seeds that were screened and collected from three independent lines. The calculated GT rate out of an estimated 24 thousand transformed seeds in these experiments is  $2.5 \times 10^{-4}$ . These results indicate that specific expression of RAD54 in the egg cell is associated with a statistically significant ~10 fold increase in GT rate ( $P(\chi^2) = 0.0002$ ).

Upon further analysis, we found that all three GT events detected in WT plants, were non-precise targeting events. These three events showed precise copying of the 5' genomic *Cru3* region upstream of the 5' region of homology, but vector integration via non-homologous end-joining at the 3' junction. Southern blot analysis of their progeny indicated that the WT allele was not modified in two of the events; the vector underwent ectopic gene targeting by copying the 5' *Cru3* region and integrating at an ectopic genomic location, without replacing the WT allele. In the third event the WT allele was changed, but gene replacement was not precise. The six GT events originating from the *EASE>>ScRAD54* lines included one true and precise GT event, as implicated from sequencing of both 5' and 3' junctions between the vector and the *Cru3* genomic locus. Moreover, Southern blot analysis revealed that the WT allele was lost in 25% of the progeny of this specific event, and the band size of the targeted allele indicated precise gene replacement (Figure 6c). In parallel, five non-precise GT events were obtained, all of which yielded the expected 5' junction, but not the 3' junction (Figure 6a) when analyzed by PCR. Four

of these non-precise GT events showed no change of the WT allele in their progeny, thus suggesting ectopic events (Figure S1). In the fifth non-precise GT event, the WT allele was changed, but not as expected for homologous gene replacement. These results show that while in WT plants no true GT events were detected, in EASE>>ScRAD54 plants we obtained a rate of  $4.17 \times 10^{-5}$  bona fide GT.

## Discussion

Currently, the ability to perform targeted modifications of the *Arabidopsis* genome is quite limited. This work describes a system designed to improve existing methods used for TGM and GT in *Arabidopsis*. Egg-cell specific expression of TGM- and GT-promoting proteins regulated by the EASE element (Yang et al., 2005), allowed for localized protein activity in the transformed germinal tissue when using the floral dipping transformation method (Clough and Bent, 1998). The present study shows that egg-cell-localized expression can be achieved by transient or stable transformation of a reporter protein regulated in *cis* or *trans* by the EASE. Further evidence of the specificity of EASE and of its utility in the context of targeted genetic manipulations, comes from the genetic data shown in Figure 3 confirming the highly accurate catalytic activity of the EASE:QQR-ZFN in the female megaspore. However, transient expression of the QQR-ZFN did not facilitate increased TGM, although earlier works in tobacco and maize reported stimulation of GT following transient ZFN expression (Shukla et al., 2009; Townsend et al., 2009). Thus, gene targeting must be optimized for each combination of species, cell type and delivery method, and the absence of effect seen here may be due to properties of the EASE-driven transient expression in *Arabidopsis*, like egg cells expression, and/or *Agrobacterium*-mediated transient expression during floral dipping. Transient expression of a TGM or GT effector protein would have provided a more time-efficient system, bypassing the need to preselect stable transformants and eliminating

the need to remove the effector proteins after fulfilling their role. Therefore, it is worth pursuing the attempts to develop effective transient expression of ZFNs or of the new generation of TALENs proteins (Boch et al., 2009; Cermak et al., 2011). Conversely, stable expression of EASE:QQR-ZFN gave rise to a TGM rate of 0.27%, which fell in line with the range reported in previous ZFN-based systems reporting of TGM of either endogenous or non-endogenous targets in *Arabidopsis* plants (de Pater et al., 2009; Lloyd et al., 2005; Osakabe et al., 2010; Zhang et al., 2010). In three of these studies the ZFN sequence was placed under an inducible promoter in order to reduce the risk of ZFN-related toxicity. In this study, the EASE system eliminates the need for induction and restricts expression to the desired cell types, thereby avoiding ZFN toxicity in other tissues. An additional advantage of the EASE system is that every mutagenesis event obtained with the EASE system stemmed from an independent germinal event, generating a multiplicity of new alleles. It must also be stressed that the presented results were collected from randomly selected, EASE-QQR-ZFN-expressing lines. It is therefore likely that higher TGM rates can be reached with lines that are selected for having strong ZFN activity.

This work clearly demonstrates the potency of the EASE system in enhancement of GT in *Arabidopsis*. The GT system described here is a modification of the system described by Shaked *et al.* (Shaked et al., 2005), including improved seed selection by using mRFP as a reporter protein in place of GFP. Additionally, due to *RAD54* gene silencing when placed under the 35S promoter (Lieberman and Levy, unpublished results), gene activity monitoring, by means of fluorescent reporter proteins, was necessary. Although slightly lower than the rate described by Shaked *et al.* 2005, the enhancement of GT rate obtained with the EASE system was significant and provides the advantage of lowering the risk of silencing, as observed with the 35S promoter.

Overall, this work contributes to the optimization of TGM and GT procedures in *Arabidopsis*, in pursuit of development of efficient, routine and precise modification methods for the *Arabidopsis* genome. EASE-driven expression can be applied to a broad range of technologies involving the

expression of proteins, such as ZFNs, TALENs or other homologous recombination-enhancing proteins, together with the delivery of a GT vector via floral dipping. Moreover, lessons learnt on GT and TGM in *Arabidopsis*, can assist the development of similar approaches in crop plants.

## Experimental Procedures

### *Plant Growth and Transformation*

Plants were grown in a growth room with a controlled environment and maintained at 19°C, with 16 hr daylight. Transformation of *A. thaliana* (ecotype Columbia) was carried out by the floral dip method as previously described (Clough and Bent, 1998). *A. tumefaciens* cultures were grown overnight, in LB medium at 28°C, until they reached a stationary phase. Before each transformation experiment, cells were concentrated to an OD<sub>600</sub> of ≈1.8 in 5% sucrose, later supplemented with 0.5 X MS salts and Silwet L-77 (Momentive Performance Materials, OH, USA) that was added to a final concentration of 0.02%, before dipping.

### *Constructs*

All of the restriction enzymes used were purchased from New England Biolabs, MA, USA.

**GT vector:** The pGTmRFP gene-targeting vector was designed to contain the Cru box corresponding to the 1,210-bp genomic sequence at the 5' end of the *Cruciferin3* gene. The Cru box lacks the ATG initiation codon as well as the 36 bp downstream to it. The *mRFP* reporter was fused in-frame, downstream to the Cru box and immediately upstream to the transcription termination sequence of the *octopine synthase* gene (*OCS*). A 2,492-bp fragment identical to the genomic region downstream to the sequence of the Cru box in the *Cruciferin3* gene, of which 773 bp corresponded to the 3' end of the *Cruciferin3* gene (*ciferin*) followed by 1,719 bp of

downstream DNA, flanked the *mRFP* reporter. The cassette was cloned into the pART27 binary vector, which contains a gene that confers kanamycin resistance in plants (Data S1).

**EASE:LhG4 vector:** In this expression vector, *LhG4* is under the control of 35S minimal promoter plus EASE, arranged in four tandem repeats of its 77-bp sequence (Yang et al., 2005). In addition, the vector contains an *OCS* terminator at the 3' end of *LhG4*. This cassette was cloned into the pMLBart binary vector, which contains a gene conferring glufosinate (BASTA) resistance in plants.

**Op:mRFP vector:** In this vector, *mRFP* ORF is under the control of 35S minimal promoter plus the *Lac operator* arranged in seven tandem repeats (Moore et al., 1998). An *OCS* terminator was positioned at the 3' end of *mRFP*. This cassette was cloned into the pART27 binary vector, which contains a gene that confers kanamycin resistance in plants.

**Op:ScRAD54 vector:** In this vector, the *ScRAD54* ORF is under the control of 35S minimal promoter plus the *Lac operator* arranged in seven tandem repeats (Moore et al., 1998). In addition, an *OCS* terminator was positioned at the 3' end of the *ScRAD54*. This cassette was cloned into the pMLBart binary vector, which contains a gene conferring glufosinate (BASTA) resistance in plants.

**EASE:EGFP and EASE:QQR-ZFN vectors:** The *LhG4* gene in the EASE:LhG4 cassette was replaced by sequences encoding either *EGFP* (Clontech, CA, USA) or *QQR-ZFN* (originally from a plasmid kindly provided by Dr. Gary Drews, University of Utah) by subcloning with *XhoI* and *XbaI* restriction enzymes. The *EASE:EGFP* cassette was subcloned, using *NotI*, into the pMLBart binary vector, which contains a gene conferring glufosinate (BASTA) resistance in plants. The *EASE:QQR-ZFN* cassette was subcloned into the pCambia1300 binary vector (Cambia, Brisbane, Australia), using *NheI* and *NsiI* that cut the *EASE:QQR-ZFN*, and *PstI* and *XbaI*, which cut the pCambia1300. The final vector conferred hygromycin resistance in plants,

### ***Confocal Microscopy***

*Arabidopsis* ovules were collected from ovaries of flowers at one stage before anthesis. Ovules were kept in DDW and cellular localization was analyzed using a laser confocal microscope, Olympus IX81 FV1000 Spectral (Olympus, Hamburg, Germany), equipped with an objective lens FV1000 UPLAPO 40xO NA: 1.35. mRFP was detected using a Diode laser (ex: 559nm; em: 575-620nm). DAPI staining was detected using a Diode laser (ex: 405nm; em: 415-435nm). EGFP was detected using an Argon laser (ex: 488nm; em: 500-545nm).

### ***EASE:QQR-ZFN activity timing assay***

A cassette including a mutated *GUS* gene with QQR-ZFN binding sites surrounding a stop codon, directly following the start ATG of the *GUS* gene, and conferring kanamycin resistance in plants (pRCS2[Kan][QQR-TS\*::GUS]) (Tovkach et al., 2009) (Figure 3a), was kindly provided by Dr. Tzvi Tzfira. This cassette was transformed into *Arabidopsis* plants, as described above, and plants were selected on 50 µg/ml kanamycin. Plants with a single loci T-DNA insertion were identified using Southern blotting, performed as described in Shaked et al. 2005, but with the following modifications: the genomic DNA was digested with *PvuII* and detection was carried out with a probe which detects the kanamycin resistance gene. The probe was prepared by PCR amplification with the following primers: nptII-Fwd: GGATTGCACGCAGGTTCTCC and nptII-Rev: TATTCGGCAAGCAGGCATCG, using the pART27 binary vector as template. The single loci lines (line #7 and line #15) were transformed with *Agrobacterium* containing the *pCambia1300-EASE-QQR-ZFN* cassette. To monitor the timeframe of EASE:QQR-ZFN activity, seeds were collected from plants that transiently expressed EASE:QQR-ZFN and germinated in the presence of kanamycin. In all, ~15,000 T1 seedlings, which were subjected to X-Gluc staining, were generated. In addition, 18 plants derived from these same plant lines, were selected on medium containing both kanamycin and hygromycin, to isolate those that were stably

transformed by both the *GUS mutant* and the *EASE:QQR-ZFN* constructs. Seeds were collected from these plants and germinated on medium containing both kanamycin and hygromycin. Approximately 9000 T2 seedlings, resistant to both antibiotics, were subjected to X-Gluc staining. For staining, seeds were sown on 0.5 X MS-Agar plates with the appropriate antibiotics, incubated for 72 hours at 4°C and then grown for 4 days at 26°C, under 16-hrs daylight conditions. The seedlings were then covered with ~6ml X-Gluc staining solution containing 1 mg/ml X-Gluc substrate (Gold Biotechnology, MO, USA) dissolved in a solution of 50 mM sodium phosphate buffer (pH 7.0), 0.1% Triton X-100, 0.1% dimethylsulphoxide, 5 mM potassium ferricyanide and 5 mM potassium ferrocyanide trihydrate (Sigma-Aldrich, Rehovot, Israel) and incubated for 24 hours at 37°C. Samples were then placed in 70% ethanol for destaining for another 48 hours. In all of the experiments, a plate with seeds collected from GUS<sup>+</sup> plants served as a positive control. Following complete destaining, the seedlings were visually examined and then further analyzed under a Nikon SMZ1500 stereomicroscope (Nikon, Tokyo, Japan). For sequencing of the QQR-ZFN break-site in the seedlings, a ~1 mm piece from a cotyledon or first leaf was placed in Terra<sup>TM</sup>-PCR Direct Polymerase Mix (Clontech, CA, USA) with primers surrounding the break-site (35SF forward primer 5'-GACCCTTCCTCTATATAAG-3' and GUSr reverse primer 5'-GGGATAGTCTGCCAGTTC-3'). The PCR was carried out according to the manufacturer's instructions, and resulted in a 680 bp product from the WT locus. The PCR fragments were sequenced at the Weizmann Institute DNA sequencing facility using the 35SF primer.

#### ***EASE:QQR-ZFN-targeted mutagenesis assay***

To analyze the targeted mutagenesis rate of *EASE:QQR-ZFN* in our system, ~400 seeds from 18 different plants of line #7, described above, that have the *EASE-QQR-ZFN* integrated into the genome, were selected on kanamycin and hygromycin. Three hundred and eighty seedlings were subjected to PCR using the Terra<sup>TM</sup>-PCR Direct Polymerase Mix (Clontech, CA, USA), as

described above. PCR reaction samples (5µl) were then restricted overnight at 37°C, using *DdeI* (New England Biolabs, MA, USA). The restriction reactions were separated on 1.7% agarose gels and bands were visualized. A PCR product, which showed an unrestricted band was sequenced to confirm that mutagenesis had occurred.

### ***GT assay***

Approximately 1000 WT plants and ~360 EASE>>ScRAD54 plants were transformed with the *pGTmRFP* vector by floral dipping (Zhang et al., 2006). Seeds were collected and red fluorescent seeds were selected under a Nikon SMZ1500 stereomicroscope (Nikon Instruments Inc.) adapted to the X-CITER 120PC Q light source system (Lumen Dynamics Group Inc.), which delivers a rich spectral fluorescence excitation and includes filter sets for Texas red (Ex 560/40, DM595 DCLP, Em 630/60). The frequency of *pGTmRFP* non-homologous integration was estimated from the transformation frequency: ~2000 seeds were plated from each GT experiment, and seedlings were selected for resistance to kanamycin. The GT rate was defined as the ratio of homologous versus non-homologous integration. It was calculated as the number of fluorescent seeds that underwent validation as GT events (see below) divided by the total number of seeds that were transformed.

In order to confirm precise homologous recombination events, we performed PCR reactions on genomic DNA that was extracted from 3-4 young leaves of plants that derived from red fluorescent seeds, as previously described (Shaked et al., 2005). The following primers were used to confirm the 5' *Cru3* GT junction: PRO\_Cru\_961\_F 5'-ACTGAAGTCCCCATGCAAACC-3' and mRFP\_ORF\_REV 5'-CCTTGATGACGTCCTCGGAG-3'. The expected size of the PCR product from this reaction is 1400 bp. The following primers were used to confirm the 3' *Cru3* GT junction: OCS\_3'\_Forw 5'-AGGACCGGCATGCAAGCTAG-3' and Cru\_down\_68981 5'-

GGCTTTAATGCGTGTGGTCA-3'. The expected size of the PCR product from this reaction is 2612 bp.

Southern blotting was used to confirm precise GT and replacement of the WT *Cruciferin3* allele.

The assay was carried out as previously described, but with a 1062 bp Cifgs probe generated by PCR on WT *Arabidopsis* (ecotype Columbia), using the following primers: Cif\_gs\_F 5'-TGACCAAACACGCATTAAAGCC-3' and Cif\_gs\_R 5'-CGAACGGTGATGGATCCACG-3'

The probe's recognition location is depicted in Figure 6a.

### Acknowledgements

We would like to thank Ueli Grossniklauss and Nir Ohad for providing us with useful advice on egg-cell-specific expression and Tzvi Tzfira for his fruitful discussions and generosity with reagents. We thank Vladimir Kiss for help with the confocal microscopy; Michal Kenan-Eichler for construction of the new RFP GT cassette; Michal Lieberman-Lazarovitch and members of the Levy lab for useful discussions. This work was supported by grants from BARD and by the EU-FP6 TAGIP and EU-FP7 RECBREED projects.

### Supporting Information

**Figure S1.** Southern blot of ectopic GT events from EASE>>ScRad54 plants.

**Data S1.** Sequence of *pGTmRFP* gene targeting vector.

## References

- Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers CC, Parker H, Prednis L, Ansari Y, Choy N, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmidt I, Guzman P, Aguilar-Henonin L, Schmid M, Weigel D, Carter DE, Marchand T, Risseuw E, Brogden D, Zeko A, Crosby WL, Berry CC and Ecker JR** (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**:653-657.
- Arabidopsis Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**:796-815.
- Bechtold N, Jolivet S, Voisin R and Pelletier G** (2003) The endosperm and the embryo of *Arabidopsis thaliana* are independently transformed through infiltration by *Agrobacterium tumefaciens*. *Transgenic Res* **12**:509-517.
- Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim YG and Chandrasegaran S** (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* **21**:289-297.
- Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A and Bonas U** (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**:1509-1512.
- Bogue CW** (2003) Genetic Models in Applied Physiology. Functional genomics in the mouse: powerful techniques for unraveling the basis of human development and disease. *J Appl Physiol* **94**:2502-2509.
- Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ and Voytas DF** (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39**: e82
- Clough SJ and Bent AF** (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**:735-743.
- de Pater S, Neuteboom LW, Pinas JE, Hooykaas PJ and van der Zaal BJ** (2009) ZFN-induced mutagenesis and gene-targeting in *Arabidopsis* through *Agrobacterium*-mediated floral dip transformation. *Plant Biotechnol J* **7**:821-835.
- Desfeux C, Clough SJ and Bent AF** (2000) Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation by the *Arabidopsis* floral-dip method. *Plant Physiol* **123**:895-904.
- Gorbunova V and Levy AA** (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic Acids Res* **25**:4650-4657.

- Kim SI, Veena and Gelvin SB** (2007) Genome-wide analysis of Agrobacterium T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. *Plant J* **51**:779-791.
- Kim YG, Shi Y, Berg JM and Chandrasegaran S** (1997) Site-specific cleavage of DNA-RNA hybrids by zinc finger/FokI cleavage domain fusions. *Gene* **203**:43-49.
- Lloyd A, Plaisier CL, Carroll D and Drews GN** (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. *Proc Natl Acad Sci U S A* **102**:2232-2237.
- Mahfouz MM, Li L, Shamimuzzaman M, Wibowo A, Fang X and Zhu JK** (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc Natl Acad Sci U S A* **108**:2623-2628.
- Moore I, Galweiler L, Grosskopf D, Schell J and Palme K** (1998) A transcription activation system for regulated gene expression in transgenic plants. *Proc Natl Acad Sci U S A* **95**:376-381.
- Osakabe K, Osakabe Y and Toki S** (2010) Site-directed mutagenesis in Arabidopsis using custom-designed zinc finger nucleases. *Proc Natl Acad Sci U S A* **107**:12034-12039.
- Puchta H, Dujon B and 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.
- Rothstein R** (1991) Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol* **194**:281-301.
- Shaked H, Melamed-Bessudo C and 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.
- Shukla VK, Doyon Y, Miller JC, DeKolver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X, Choi VM, Rock JM, Wu YY, Katibah GE, Zhifang G, McCaskill D, Simpson MA, Blakeslee B, Greenwalt SA, Butler HJ, Hinkley SJ, Zhang L, Rebar EJ, Gregory PD and Urnov FD** (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* **459**:437-441.
- Tovkach A, Zeevi V and Tzfira T** (2009) A toolbox and procedural notes for characterizing novel zinc finger nucleases for genome editing in plant cells. *Plant J* **57**:747-757.
- Townsend JA, Wright DA, Winfrey RJ, Fu F, Maeder ML, Joung JK and Voytas DF** (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* **459**:442-445.
- Tzfira T, Li J, Lacroix B and Citovsky V** (2004) Agrobacterium T-DNA integration: molecules and models. *Trends Genet* **20**:375-383.
- Urnov FD, Rebar EJ, Holmes MC, Zhang HS and Gregory PD** (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* **11**:636-646.

**Weinthal D, Tovkach A, Zeevi V and Tzfira T** (2010) Genome editing in plant cells by zinc finger nucleases. *Trends Plant Sci* **15**:308-321.

**Yadegari R and Drews GN** (2004) Female gametophyte development. *Plant Cell* **16** Suppl:S133-141.

**Yang W, Jefferson RA, Huttner E, Moore JM, Gagliano WB and Grossniklaus U** (2005) An egg apparatus-specific enhancer of Arabidopsis, identified by enhancer detection. *Plant Physiol* **139**:1421-1432.

**Zhang F, Maeder ML, Unger-Wallace E, Hoshaw JP, Reyon D, Christian M, Li X, Pierick CJ, Dobbs D, Peterson T, Joung JK and Voytas DF** (2010) High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. *Proc Natl Acad Sci U S A* **107**:12028-12033.

**Zhang X, Henriques R, Lin SS, Niu QW and Chua NH** (2006) Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nat Protoc* **1**:641-646.

## Figure Legends:

### Figure 1. A trans-activation system for expressing recombination proteins in the egg cell.

Confocal microscope images of an ovule at the flowering stage, before anthesis, collected from an EASE>>mRFP plant line. EASE>>mRFP denotes trans-activation of mRFP by the chimeric LhG4 activator expressed under the EASE enhancer. Panels from left to right: DIC; DIC, DAPI (blue) and mRFP (red) overlay; DIC and mRFP (red) overlay. High mRFP signal (red) is detected in the egg cell area. Scale bar, 20  $\mu$ m. The constructs used for the transactivation in this plant are shown below the picture and are described in detail in the Experimental Procedures.

### Figure 2. Cellular localization of EASE-EGFP following floral dip transformation.

a. Scheme of an *Arabidopsis* ovule at the 8-nuclei stage (adapted from (Yadegari and Drews, 2004)). The location of the egg cell and the synergid cell is indicated as ec and sc, respectively. The right section of panel A is a scheme of the construct used for EGFP expression under the EASE enhancer, with the Basta resistance gene (*Bar*). b. *Arabidopsis* 7-weeks old plants were transformed, using the floral dip method, with *Agrobacterium* carrying the *EASE-EGFP* construct. Six days following transformation, *Arabidopsis* ovules were dissected from ovaries of flowers, one stage before anthesis. Ovules were kept in DDW and cellular localization was analyzed using a laser confocal microscope, as described above. Panels from left to right: EGFP (green), DIC, and an overlay of DIC+EGFP (green). Scale bar, 20 $\mu$ m. c. Quantification of transformation efficiency in transient vs. stable transformation. *Arabidopsis* plants were transformed with the *EASE-EGFP* construct using the floral dip method and transient transformation efficiency was measured by counting the number of ovules exhibiting EGFP expression 6 days following transformation. Stable transformation efficiency was measured by counting seedlings from seeds of the same transformation, which were resistant to Basta

selection. The bar with an asterisk indicates a significant difference in the efficiency of transient versus stable transformation ( $\chi^2 = 51.24$  ( $p < 0.0001$ )).

**Figure 3. Developmental timing of EASE expression.**

a. Seedlings were grown from seeds of Arabidopsis plants containing: *top* – a single loci of the mutated GUS cassette (*uidA*), which includes a stop codon downstream to the CaMV 35S promoter (35SP), and situated in between two QQR-ZFN recognition sequences (QQR-TS), as well as the 35S terminator (35ST) downstream to the *GUS*. Downstream from the mutated *GUS* there is a kanamycin resistance gene (*NptII*); *bottom* – a QQR-ZFN cassette under EASE, with hygromycin resistance (*Hpt*). b. Seedlings were stained with X-Gluc staining solution, to detect NHEJ repair of the mutated *GUS* gene. A blue seedling represents a mutated seedling. c. Partial sequences of PCR fragments encompassing the QQR-ZFN break-site from seven blue seedlings, compared to WT. Underlined sequence - start ATG of the *uidA* gene; Bold sequence - *DdeI* restriction site; Italicized sequence - nucleotides inserted during NHEJ; Colon – represents a deleted nucleotide; Gray sequence - the QQR-TS.

**Figure 4. Expression of QQR-ZFN in the egg-cell leads to targeted mutagenesis.**

a. PCR was performed on genomic DNA extracted from randomly selected Arabidopsis seedlings of the same origin as those described in Figure 3. Arrows mark primers used for PCR. *DdeI* marks the site of restriction for this enzyme, used for mutagenesis analysis. b. PCR products were digested with *DdeI* and separated on agarose gel. The vertical arrow points to a lane with DNA from a seedling that was mutated by QQR-ZFN. c. Partial sequence of PCR fragment encompassing the QQR-ZFN break-site, generated from a randomly mutated seedling (lower sequence) compared to WT (upper sequence). Underlined sequence - start ATG of the *uidA* gene; Bold sequence - *DdeI* restriction site; Colon – represents a deleted nucleotide; Gray sequence - the QQR-TS.

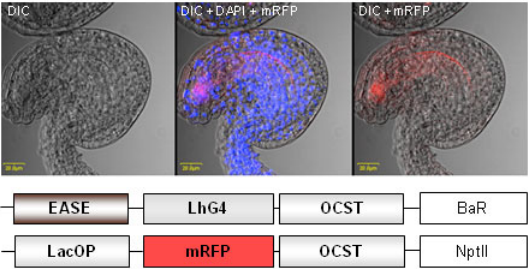
**Figure 5. Expression of ScRAD54 in the egg-cell enhances GT.**

Frequencies of Cru3 GT were compared between WT and EASE>>ScRAD54 lines. GT frequency was ~ 10 fold higher in EASE>>ScRAD54 than in WT. This difference shown by the bar with an asterisk was statistically significant as determined by the Chi-square test ( $P(\chi^2) = 0.0002$ ). The constructs used for the transactivation in EASE>>ScRAD54 plants are shown below the graph.

**Figure 6. Validation of Cru3 GT events originating from expression of ScRAD54 in the egg cell.**

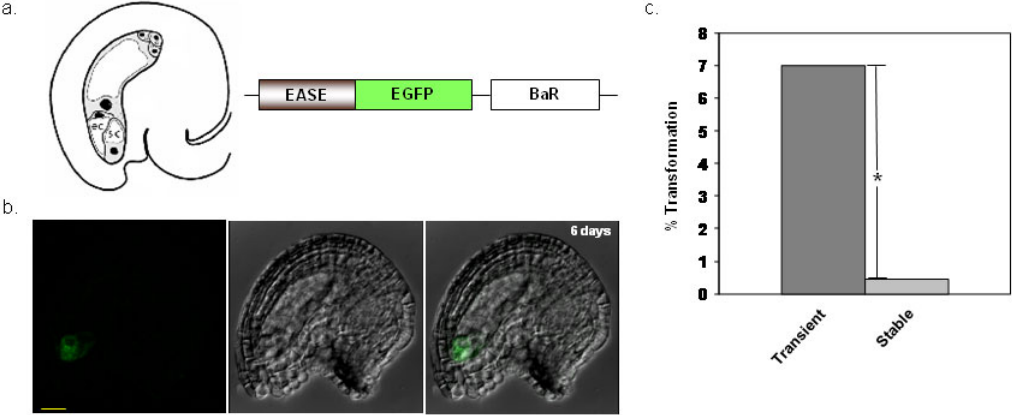
a. PCR fragments of genomic DNA from six EASE>>ScRAD54 GT events. 5' and 3' indicate PCR amplification of the 5' or 3' Cru3 junction between genomic DNA and the GT vector, respectively. One of the six events yielded the expected size for both 5' and 3' junctions (lane 1); the other five events yielded reactions at the 5' junction only (lanes 2-6). b. Scheme of the structure of the gene targeting vector, the WT allele and the expected targeted genomic locus. The T-DNA gene targeting vector includes the LB- T-DNA left border; RB- T-DNA right border and *NPT II*, the kanamycin resistance selection marker. The probe used in Southern blot analysis detects a 6000 bp fragment in the Cru3 WT allele, following digestion of genomic DNA with *HindIII*. WT allele - Cru3 WT allele; Promoter - Cru3 promoter; M - Cru3 ATG initiation codon; Probe - 1062 bp region used for Southern blot probing, located 78bp downstream to the genomic region that is included in the GT vector. In the targeted allele, digestion with *HindIII* and hybridization with the same probe, detects a 4600 bp fragment. Targeted allele - Cru3 allele that underwent HR with the GT vector. c. Southern blot of a true GT event progeny. The parent plant yielded both 5' and 3' junctions by PCR, as seen in panel A (lane 1). Genomic DNA was digested with *HindIII* and hybridized with the 1062bp probe described in panel b. WT- Cru3 WT allele, TGT - true gene targeting of Cru3 allele.

Figure 1



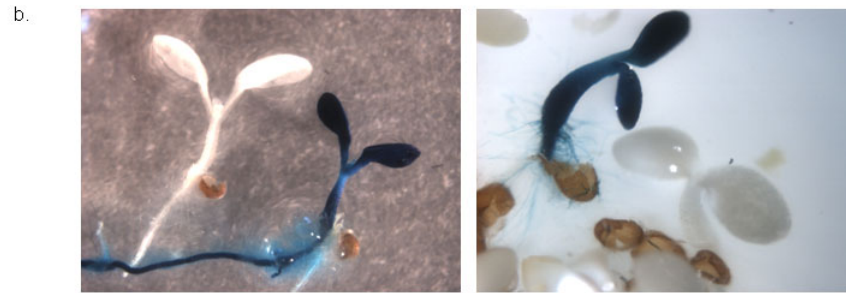
UNCORRECTED

Figure 2



UNCORRECTED

Figure 3



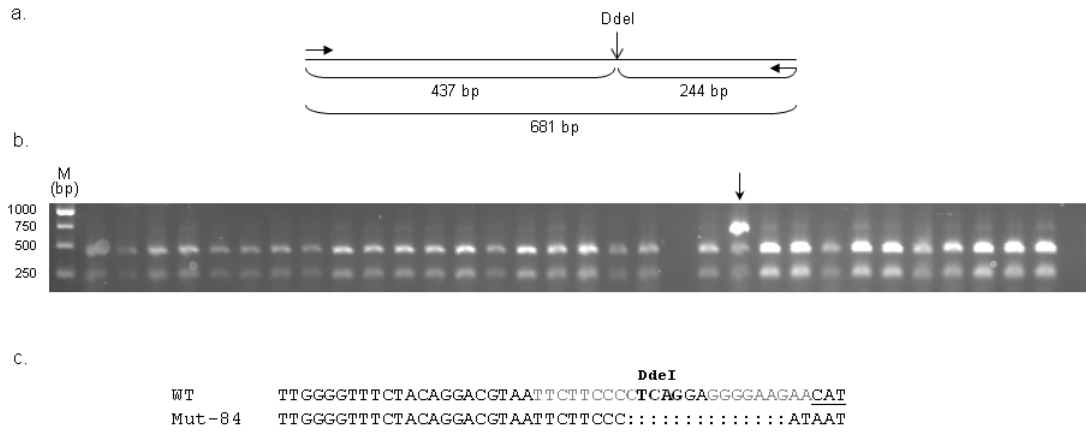
c.

	<b>DdeI</b>
WT	TTGGGGTTTCTACAGGACGTAATTCTTCCCCT <b>TCAG</b> GAGGGGAAGAACAT
Blue1	TTGGGGTTTCTACAGGACGTAATTCAC: : : : : AGGGGAAGAACAT
Blue2	TTGGGGTTTCTACAGGACGTAATTCTTCCCCTCGT: : : : : AAGAACAT
Blue3	TTGGGGTTTCTACAGGACGTAATTC: : : : : : : : : : : GAAGAACAT
Blue4	TTGGGGTTTCTACAGGACGTAATTCTTCCCCT: : : : : GAGGGGAAGAACAT
Blue5	TTGGGGTT: GGGGAAGAACAT
Blue6	TTGGGGTTTCTACAGGACGTAATTCTTCCCCT: : : : : GGGGAAGAACAT
Blue7	TTGGGGTTTCTACAGGACGTAATTTTCCCCT: : : : : GAGGGGAAGAACAT

UNCORRECTED



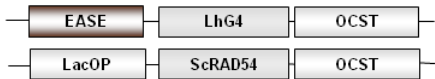
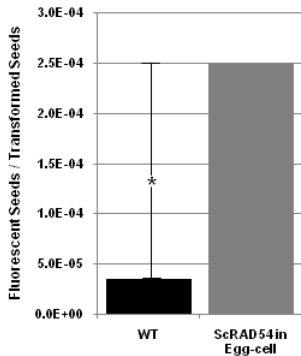
Figure 4



UNCORRECTED

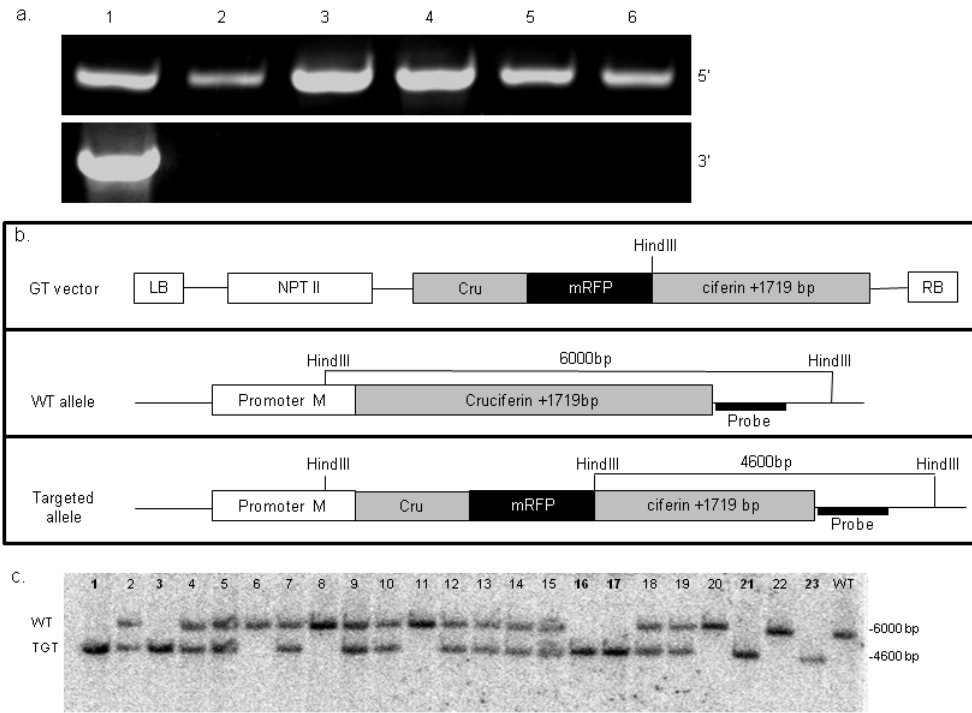
Figure 5

Frequency of Gene-targeting Events



UNCORRECTED

Figure 6



UNCORRECTED