MATERIALS and METHODS

Design of TALEN target sites

mRNA sequences of *B. belcheri* genes were retrieved from our transcriptome dataset based on their homology to the published *B. floridae* query sequences using local Blastn program (Altschul et al., 1990), and the exon-intron structure of a gene was determined by mapping its mRNA sequence to the corresponding *B. belcheri* genomic region (http://mosas.sysu.edu.cn/genome/index.php). TALEN forward and reverse target sites for each gene were designed using the following parameters: 1) nucleotide T was at position 0, and 2) both spacer length and nucleotides that bound to forward or reverse TALEN proteins were about 16 bp. Additional preference was also given to target sites: 1) close to the 5′- terminus of the gene to induce premature translation stop mutations earlier in the coding sequence, 2) conserved in *B. becheri* and/or *B. floridae* in order to minimize the interference of polymorphism, and 3) including a restriction endonuclease site near the center of the spacer for convenient mutation detection and estimation of mutation ratio. The details of TALENs used in the present study are provided in Table 1.

TALEN vector assembly

Three TALEN backbone vectors (pCS2-PEAS/PERR (Huang et al., 2011), pCS2-ELD/KKR (Lei et al., 2012) and GoldyTALEN (Bedell et al., 2012) (Fig. S2) were used for constructing two target sites of amphioxus *Pax1/9* gene (Table 1). RVD repeat modules for pCS2-PEAS/PERR were described in the reference and assembled using the ‘unit assembly’ method (Huang et al., 2011), and RVD repeat modules for pCS2-ELD/KKR and GoldyTALEN were derived from Cermark et al. and assembled using the Golden Gate assembly method (Cermak et al., 2011). GoldyTALEN backbone plasmid and RVD repeats for Golden Gate assembly were purchased from Addgene Co. (USA), and plasmids used in the ‘unit assembly’ method were kindly provided by Prof. Bo Zhang (Peking University, China). For convenience, the three TALEN systems using the GoldyTALEN, pCS2-ELD/KKR and pCS2-PEAS/PERR backbone vectors were respectively referred as Goldy, HZ and BZ hereafter. Assembled TALEN vectors were verified using double restriction enzyme digestion and DNA sequencing as described previously (Huang et al., 2011). After confirming the efficiency of the Goldy system in mutation induction, we further assembled eight new pairs of TALEN vectors targeting 7 additional amphioxus genes using the system (Table 1).

In vitro synthesis of TALEN mRNAs

In-house-assembled TALEN vectors were linearized with *Not* I (for HZ and BZ system) or with *Sac* I (for Goldy system). The linearized vectors were purified using phenol/chloroform extraction and applied as the

templates for *in vitro* TALEN mRNA synthesis using either SP6 (for pCS2-PEAS/PERR and pCS2-ELD/KKR) or T3 (for GoldyTALEN) mMACHINE kits (Ambion, USA). The transcribed mRNAs were precipitated using LiCl according to kit manual, and then diluted in desired concentrations in nuclease-free water (Promega).

Microinjection and analysis of somatic mutations

The cultivation of adult amphioxus (*B. belcheri*) and collection of mature gametes were carried out as described previously (Li et al., 2012, 2013). Microinjection on amphioxus unfertilized eggs (1-2 pl per egg) was conducted following a previous established procedure (Liu et al., 2013). The injected embryos were incubated at 25°C in the dark until desired stages (Liu et al., 2013). Optimum TALEN transcript concentrations were established by microinjecting amphioxus embryos with several doses of Pax3/7-Fw1/Rv1 mRNAs (50 ng/μL, 260 ng/μL, 600 ng/μL and 1.3 μg/μL). Meanwhile, the eggs injected with 200 mmol/L KCl served as a control to monitor gamete quality. For each experimental group, three batches of unfertilized eggs (600-700 for each) were microinjected. All microinjected embryos were scored under a dissecting microscope (SZX2-ILLK, Olympus) as 'normal', 'abnormal' or 'dead' at the larval stage (~2 days after fertilization). Thereafter, unless indicated otherwise, 1.3 μg/μL was chosen as the optimal concentration for all *in vitro* synthesized TALEN transcripts. At late neurula stage (~24 hours post fertilization), about 200 microinjected embryos were selected for genomic DNA extraction using DNeasy® Blood & Tissue Kit (Qiagen, Germany). DNA segments containing a TALEN-targeted genome region were amplified using the genomic DNA as templates and then digested with proper restriction enzymes to determine mutation ratios by comparing band intensities between undigested (due to the loss of RE sites caused by mutation) and digested (wild type) bands. PCR products were also cloned into pGEMT-easy vector (Promega) for sequencing to further verify the TALEN-induced mutations. For each TALEN-treated embryo, about 20 amplicon-containing colonies were sequenced. Primers used in the present study are listed in Table S1.

Western blot analysis

Antibodies specific for Flag or HA tags in the pCS2-PEAS and pCS2-PERR vectors (BZ system), and for Flag tags in the pCS2-ELD and pCS2-KKR vectors (HZ system) (Fig. S2) were used to determine whether the TALEN mRNAs were translated into proteins in amphioxus embryos. Transcripts synthesized from a pair of TALEN vectors (DrAmmcr1-Fw1/Rv1) constructed using the BZ system showed robust activity in mutagenizing zebrafish *Ammecr1* gene (the mutation ratio is about 70%). We adopted the proteins extracted

from DrAmmcr1-Fw1/Rv1-treated zebrafish embryos as a positive control for detecting TALEN protein expression in amphioxus embryos. Approximately 400 TALEN-treated neurula embryos were collected and lysed in RIPA Lysis buffer (Beyotime, China). Mouse anti-Flag (Sigma, USA, in 1:2500 dilution) and rat anti-HA (Roche, Switzerland, in 1:2500 dilution) were adopted as primary antibodies and HRP-conjugated goat anti-mouse or -rat antibodies (Proteintech, China, in 1:5000 dilution) as the secondary in the analysis. Signals were detected after incubating the samples in diaminobenzidine at room temperature.

Large segmental deletions in amphioxus

Two TALENs targeting amphioxus *Pax1/9* and *Pax2/5/8* genes (Fig. S5A) were used to evaluate whether TALEN could induce large segmental deletion in amphioxus. Bra-Fw1/Rv1 and Bra-Fw2/Rv2 were also used in this experiment. These two TALEN pairs were designed to target the conserved sequences of two closely linked amphioxus *Bra1* and *Bra2* genes (Fig. S5A). Therefore, TALEN proteins translated from Bra-Fw1/Rv1 or Bra-Fw2/Rv2 were able to target two separate sites in the two genes (Fig. S5A). The total concentration of TALEN mRNAs used in this experiment was 1.3 μg/μL (625 ng/μL for each TALEN). Genomic DNAs were extracted from the TALEN-treated embryos at late neurula stage (~24h post-fertilization) described above. Deletions were confirmed by PCR using primers spanning the targeted loci, and then, were further verified *via* DNA sequencing of the amplicons. The deletion ratio was estimated according to Gupta's description (Gupta et al., 2013) by comparing the intensity of a deletion band with that of either a full-length genomic amplicon (for *Pax1/9* locus) or a length-comparable amplicon from unmodified genome using a third primer (for *Pax2/5/8* and two *Bra* loci).

Homology-directed repair (HDR) in amphioxus

Efficiency of homology-directed repair (HDR) is positively related to the mutation ratio of TALEN used. Therefore, we selected two pairs of TALENs (Pax3/7-Fw1/Rv1 and Pax2/5/8-Fw1/Rv1), which were of relatively high efficiency in mutation induction, in subsequent experiments. For each locus, two distinct ssDNA oligonucleotides were synthesized. They were designed to include a modified *LoxP* (*mLoxP*) site (5′-ATAACTTCGTATAGCATACATTATAGCAATTTAT-3′) in the center and 25 bp homology arms on two sides (Fig. S6A and B). The sequences of 25 bp were chosen since the frequency of HDR decreased with the length increase of homology arm (Bedell et al., 2012). The injection solutions were prepared as described above but contained an equal amount (650 ng/μL) of each TALEN mRNAs and ssDNA oligonucleotides. Genomic DNA was extracted from the injected embryos as described above. *mLoxP* insertion was determined by PCR using a forward gene-specific primer (Pax2/5/8-TALEN-PCR-F1 or

Pax3/7-TALEN-PCR-F1) and a mLoxP-R reverse primer. Genomic fragments flanking the target sites were also amplified using the gene-specific primers (Pax2/5/8-TALEN-PCR-F1 & Pax2/5/8-TALEN-PCR-R2 or Pax3/7-TALEN-PCR-F1 & Pax3/7-TALEN-PCR-R1) for insertion efficacy evaluation under the following steps: 1) cloning the amplicon into pGEMT-easy vector (Promega), 2) colony screening using gene-specific primers (Pax2/5/8-TALEN-PCR-F1 & Pax2/5/8-TALEN-PCR-R2 or Pax3/7-TALEN-PCR-F1 & Pax3/7-TALEN-PCR-R1), 3) re-screening above colonies using the same forward primer (Pax2/5/8-TALEN-PCR-F1 or Pax3/7-TALEN-PCR-F1) and a mLoxP-R reverse primer, and 4) estimating the insertion frequency by comparing number of positive clones acquired in step 2 to that acquired in step 3. The primers used here were listed in Table S1.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., 1990. Basic local alignment search tool. J. Mol. Biol. 215**,** 403-410.

Bedell, V. M., Wang, Y., Campbell, J. M., Poshusta, T. L., Starker, C. G., Krug, R. G., 2nd, Tan, W., Penheiter, S. G., Ma, A. C., Leung, A. Y., Fahrenkrug, S. C., Carlson, D. F., Voytas, D. F., Clark, K. J., Essner, J. J., Ekker, S. C., 2012. *In vivo* genome editing using a high-efficiency TALEN system. Nature 491**,** 114-118.

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J., Voytas, D. F., 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 39**,** e82.

Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S., Zhang, B., 2011. Heritable gene targeting in zebrafish using customized TALENs. Nat. Biotechnol. 29**,** 699-700.

Lei, Y., Guo, X., Liu, Y., Cao, Y., Deng, Y., Chen, X., Cheng, C. H., Dawid, I. B., Chen, Y., Zhao, H., 2012. Efficient targeted gene disruption in *Xenopus* embryos using engineered transcription activator-like effector nucleases (TALENs). Proc. Natl. Acad. Sci. USA 109**,** 17484-17489.

Li, G., Yang, X., Shu, Z., Chen, X., Wang, Y., 2012. Consecutive spawnings of Chinese amphioxus, *Branchiostoma belcheri*, in captivity. PLoS ONE 7**,** e50838.

Li, G., Shu, Z. H., Wang, Y. Q., 2013. Year-round reproduction and induced spawning of Chinese amphioxus, *Branchiostoma belcheri*, in laboratory. PLoS ONE 8**,** e75461.

Liu, J., Li, C., Yu, Z., Huang, P., Wu, H., Wei, C., Zhu, N., Shen, Y., Chen, Y., Zhang, B., Deng, W. M., Jiao, R., 2013. Efficient and specific modifications of the *Drosophila* genome by means of an easy TALEN strategy. J. Genet. Genomics 39**,** 209-215.

Table S1.

PCR primers and synthesized *mLoxP* used in the present study.

Primer name	**Primer sequence (3'→5')**
Pax1/9-TALEN-PCR-F1	GAGCAAACATTTGGGGAGGTG
Pax1/9-TALEN-PCR-R1	CTGCTTGTTCGTTAGCGTTGAT
Pax2/5/8-TALEN-PCR-F0	AGCTAGAGAAAACAGAGGAAGTC
Pax2/5/8-TALEN-PCR-R0	CACAGCCGTGGGACACTCGT
Pax2/5/8-TALEN-PCR-F1	AATCAAGCCGGGTGTTATCG
Pax2/5/8-TALEN-PCR-R2	CTGTTTATAGAGCTGACACTGGG
Pax3/7-TALEN-PCR-F1	CACGAGACTTGCGTAGTTTGTT
Pax3/7-TALEN-PCR-R1	CGGAGAATGCGTGAGATGGAC
Pax4/6-TALEN-PCR-F2	GGGTTGACTCGGTCTTGCTG
Pax4/6-TALEN-PCR-R0.5	TTGGCATACACTACCTTCCCAG
Bra2-TALEN-PCR-F1	CTGGGAGAAGTTCAAGTCTCTC
Bra1-TALEN-PCR-F1	CATCGGTGTGAAGTGCGA
Bra-TALEN-PCR-R1	TGAGTTTGTTGGTGAGTTTGAC
FGF8/17/18-TALEN-PCR-F1	ACCAAAACAAGTGTGGCTTGCAT
FGF8/17/18-TALEN-PCR-R1	GGGGAGGCAACTGACGACACC
En-TALEN-PCR-F1	ATGGCTAACAGTAGCGCGG
En-TALEN-PCR-R1	CAGATCATGGGCTGGGGCAC
	TACAAGCGACAGAACCCCACCATGTATAACTTCGTATAG
Pax2/5/8-mLoxP-ssDNA1	CATACATTATAGCAATTTATTCGCATGGGAGATCAGAGA
	CAGGCT
	TACAAGCGACAGAACCCCACCATGTATAACTTCGTATAG
Pax2/5/8-mLoxP-ssDNA2	CATACATTATAGCAATTTATAGGCATTTGTGACAACGAC
	ACAGTT
	GATCGTGGAGATGGCCGCCGCCGGGATAACTTCGTATAG
Pax3/7-mLoxP-ssDNA1	CATACATTATAGCAATTTATATCCGGCCCTGCGTCATCA
	GCCGCC
	GATCGTGGAGATGGCCGCCGCCGGGATAACTTCGTATAG
Pax3/7-mLoxP-ssDNA2	CATACATTATAGCAATTTATCGTGTCGCACGGCTGCGTC
	TCCAAG
mLoxP-R	ATAAATTGCTATAATGTATGCTATACGAAGT

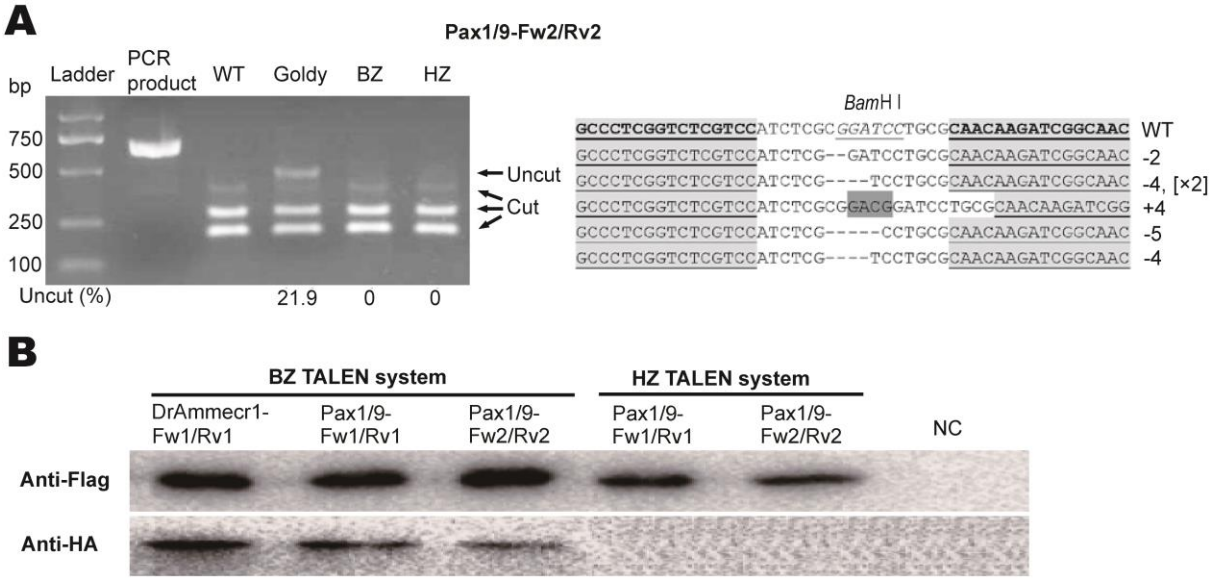

Fig. S1. Activities of three TALEN systems in amphioxus.

A: Activities of three TALEN systems at amphioxus *Pax1/9* locus targeted by TALEN pair Pax1/9-Fw2/Rv2. TALEN induced mutation ratios (estimated as percentages of uncut PCR products) are labeled under the gel image. Note that there are two *Bam*H I cutting sites in the amplicon of Pax1/9-Fw2/Rv2 targeted loci. **B:** Western blotting analysis of TALEN proteins using extracts from embryos microinjected with mRNAs of Pax1/9-Fw1/Rv1 or Pax1/9-Fw2/Rv2. A positive control using protein extracted from zebrafish embryos microinjected with a pair of BZ TALEN mRNAs (DrAmmecr1-Fw1/Rv1) that effectively induced mutations in zebrafish *Ammecr1* gene. Abbreviations: WT, wild type; Goldy, Goldy TALEN system; HZ, Hui Zhao's TALEN system; BZ, Bo Zhang's TALEN system; NC, negative control (no protein sample was loaded). Somatic cells for TALEN activity examination were all from late neurula (~24 hours post fertilization).

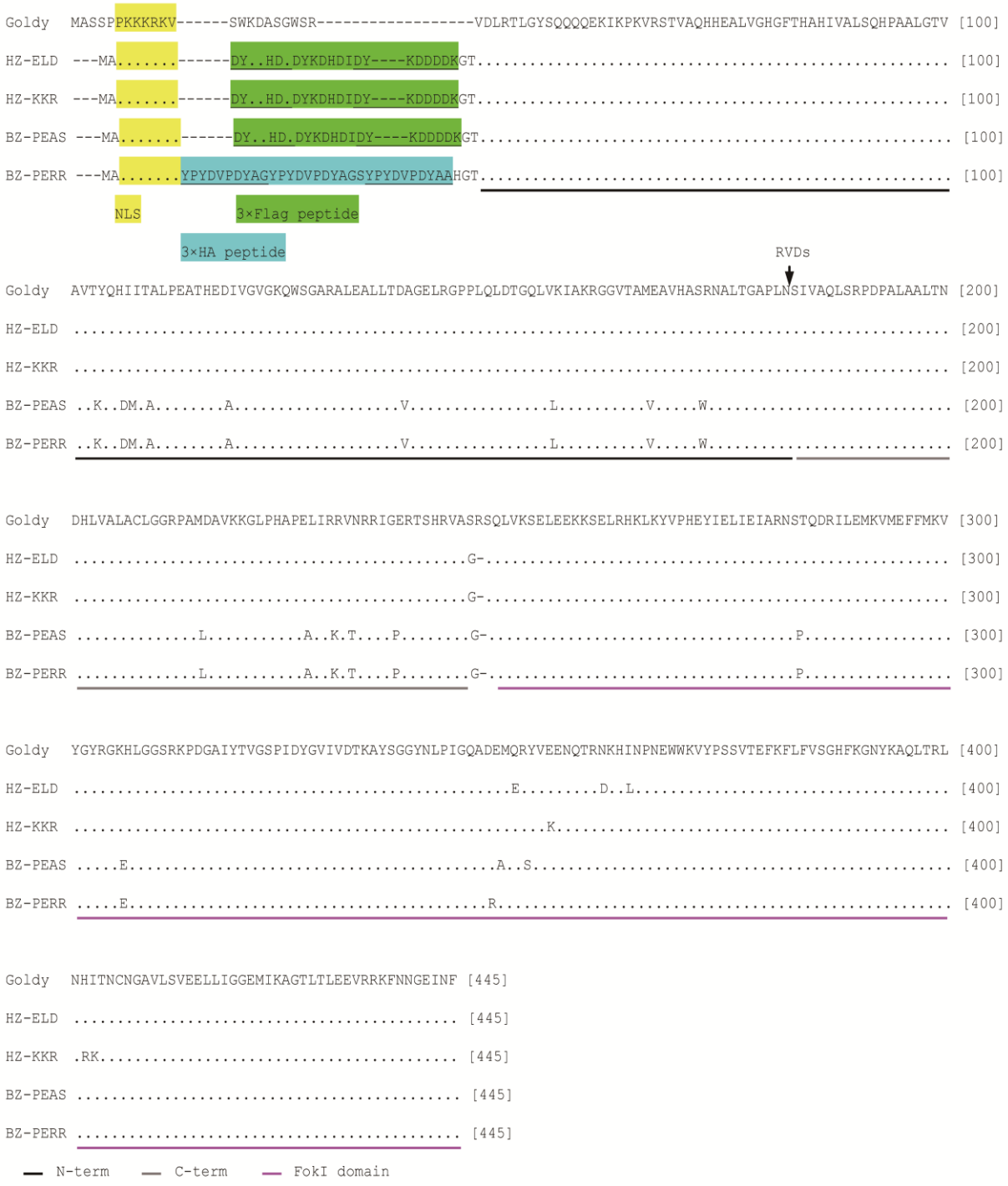

Fig. S2. Sequence alignment of the TALEN backbones used in the present study.

Sequence shaded in yellow is nuclear localization signal (NLS) peptide; sequence shaded in green is triple Flag peptide; and sequence shaded in light blue is triple HA peptide. Vertical arrow indicates where RVD modules are inserted. The last half RVD module in the BZ backbones is not included in the alignment.

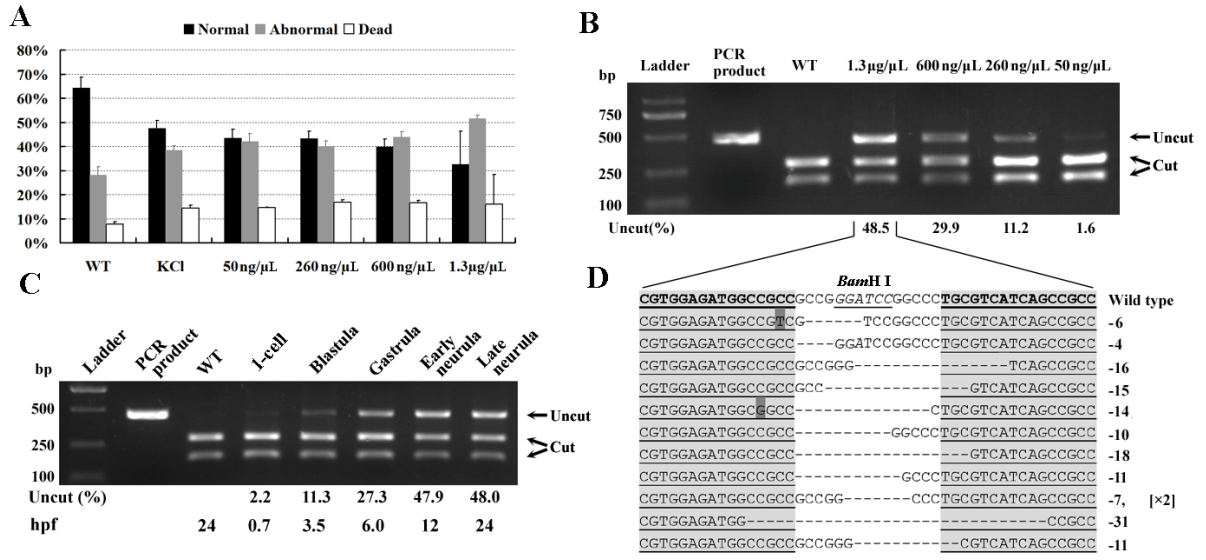

Fig. S3. Effects of different doses of Goldy TALEN treatments and mutation ratios in amphioxus embryos. **A:** Percentages of normal, abnormal and dead amphioxus embryos injected with mRNAs encoding an amphioxus *Pax3/7* Goldy TALEN (Pax3/7-Fw1/Rv1). As a control, same volume of KCl solution was also injected for evaluating the effect of microinjection and egg qualities. **B:** Dose-dependant activities of Goldy Pax3/7-Fw1/Rv1 in amphioxus embryos. TALEN induced mutation ratios estimated as percentages of uncut DNAs are labeled under the gel image. **C:** Activities of Goldy Pax3/7-Fw1/Rv1 at different developmental stages of amphioxus embryos. Percentages of uncut DNA (~mutation ratios) are shown under the gel image. Abbreviation: hpf, hour post fertilization. **D:** Goldy Pax3/7-Fw1/Rv1 targeted sequence (wild type) and the induced somatic mutations in amphioxus embryos. Restriction enzyme recognition sequences are underlined. Deletions (-) and single polymorphic sites (shaded in dark grey color) derived from embryos injected with 1-2 pL of 1.3 μg/μL TALEN mRNAs are listed.

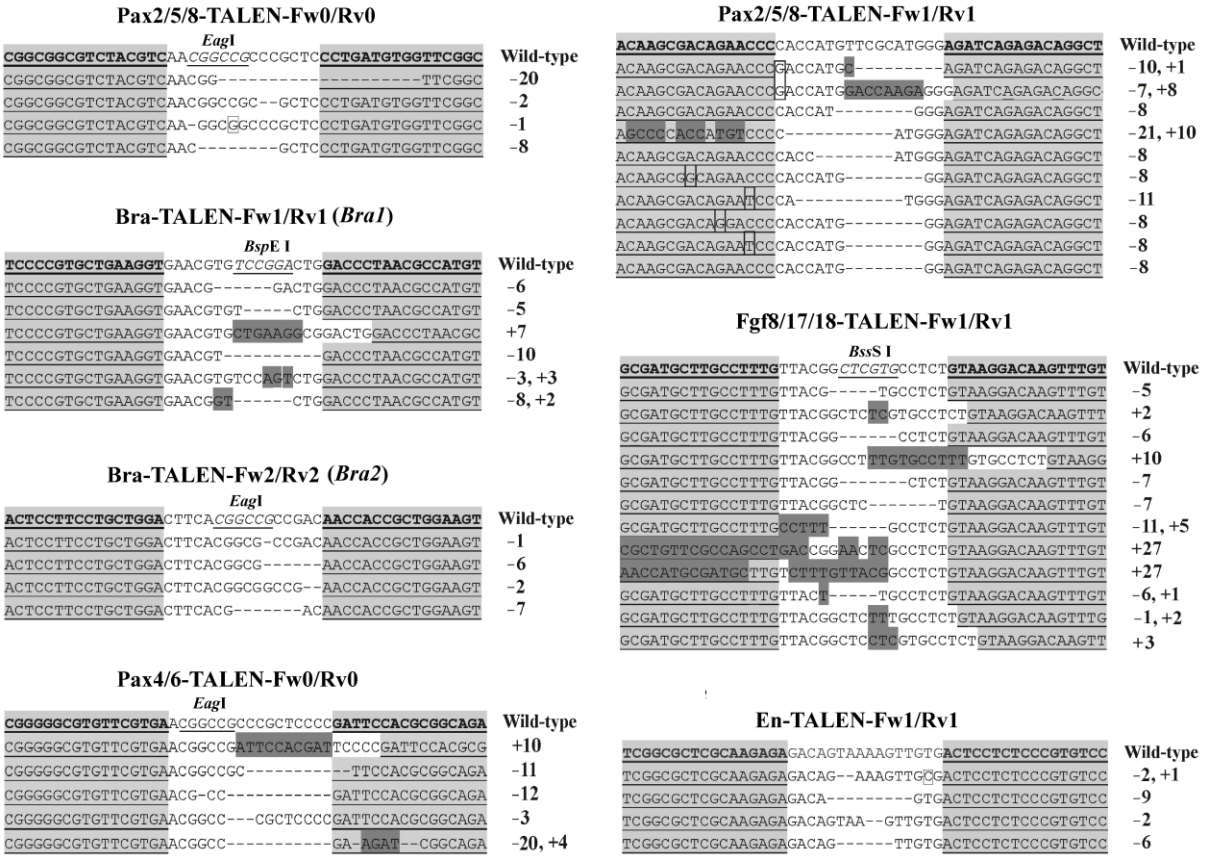

Fig. S4. Somatic mutations in six amphioxus genes induced by seven TALEN pairs. TALEN-binding sites are shown in bold fonts. Restriction enzyme recognition sequences are underlined. Deletions (-), insertions (shaded in dark grey color) and single polymorphic sites (in black boxes) are also listed.

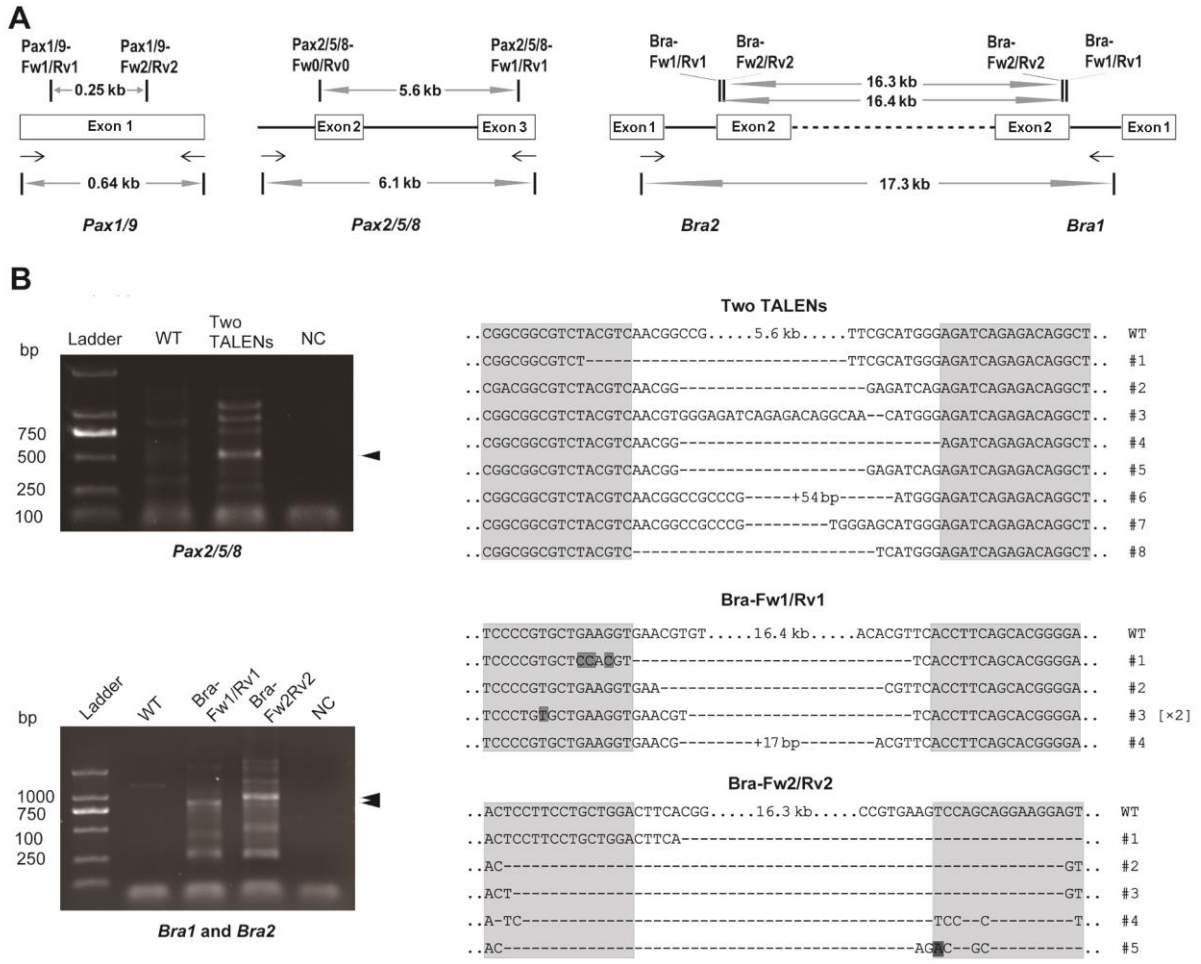

Fig. S5. Goldy TALEN induced large DNA fragment deletions in amphioxus somatic cells.

A: Structure of four target loci. Two pairs of TALENs were used for deleting DNA fragments located in amphioxus *Pax1/9* and *Pax2/5/8* loci; whereas deleting of large DNA fragments (~16 kb) between two amphioxus *Bra* loci were induced by Bra-Fw1/Rv1 or Bra-Fw2/Rv2 because *Bra1* and *Bra2* genes face in opposite directions and the target sites of Bra-Fw1/Rv1 and Bra-Fw2/Rv2 were conserved in the two genes. Size of potential deleted segments (between two TALENs), primers used for amplifying the targeted loci (indicated by head to head arrows) and their intervening segmental sizes are shown. **B:** PCR products amplified from the genomic DNA extracted from injected (one or two TALENs) and un-injected (WT) embryos, and their sequences. No DNA template was added in negative control (NC). PCR products marked by black arrowheads would be obtained if the target sequences were deleted. The left half sequence of the first TALEN and the right half sequence of the second TALEN are shown in the first line (WT) in each alignment with TALEN site shaded in grey and spacer unmarked.

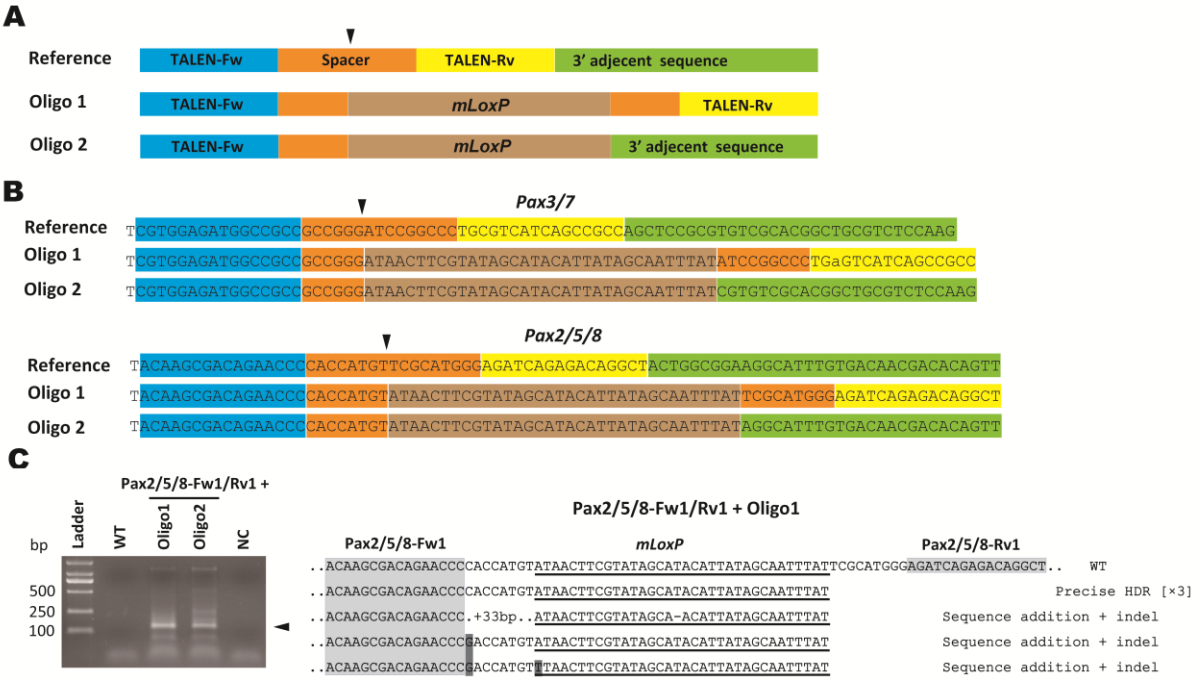

Fig. S6. Goldy TALEN mediated homology-directed repair (HDR) in amphioxus somatic cells.

A: A schematic drawing shows TALEN targets and two types of ssDNA oligonucleotides designed to introduce a *mLoxP* site at the target site. The left TALEN binding site (TALEN-Fw), spacer, right TALEN binding site (TALEN-Rv), 3′- adjacent sequence, and the *mLoxP* site are differently colored. Vertical arrowhead indicates putative TALEN cut site where the *mLoxP* will be inserted. **B:** ssDNA oligonucleotide sequences used for the amphioxus *Pax3/7* and *Pax2/5/8* loci. All sequences/sites are highlighted same as that in **A**. **C:** PCR products amplified from the genomic DNA of injected (Pax2/5/8-Fw1/Rv1 mRNAs + ssDNA oligonucleotides) and un-injected (WT) embryos and the induced insertion of *mLoxP* at the target site. TALEN binding sites are shaded in grey and the *mLoxP* site is underlined. PCR was conducted using a forward gene-specific primer and a reverse *mLoxP* primer. PCR amplicons as marked by black arrowheads will be obtained when HDR and *mLoxP* are inserted; while no PCR products will be obtained when no *mLoxP* is inserted. NC indicates a negative control without DNA template