Tandem Repeat Modification during Double-Strand Break Repair Induced by an Engineered TAL Effector Nuclease in Zebrafish Genome

Wanxu Huang[9], **Jianbo Zheng**[9], **Ying He, Chen Luo***

College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang, China

Abstract

Tandem repeats (TRs) are abundant and widely distributed in eukaryotic genomes. TRs are thought to have various functions in gene transcription, DNA methylation, nucleosome position and chromatin organization. Variation of repeat units in the genome is observed in association with a number of diseases, such as Fragile X Syndrome, Huntington's disease and Friedreich's ataxia. However, the underlying mechanisms involved are poorly understood, largely owing to the technical limitations in modification of TRs at definite sites in the genome *in vivo*. Transcription activator-like effector nucleases (TALENs) are widely used in recent years in gene targeting for their specific binding to target sequences when engineered *in vitro*. Here, we show that the repair of a double-strand break (DSB) induced by TALENs adjacent to a TR can produce serial types of mutations in the TR region. Sequencing analysis revealed that there are three types of mutations induced by the DSB repair, including indels only within the TR region or within the flanking TALEN target region or simultaneously within both regions. Therefore, desired TR mutant types can be conveniently obtained by using engineered TALENs. These results demonstrate that TALENs can serve as a convenient tool for modifying TRs in the genome in studying the functions of TRs.
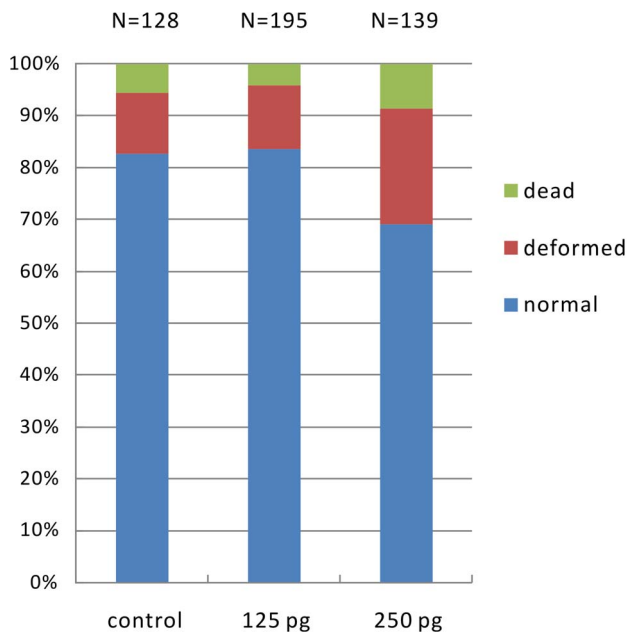
Introduction

Even years before the nucleotide sequencing methods were developed, a great mass of repeated sequences were detected in vertebrate genomes by denaturation-renaturation experiments [1]. Nowadays, whole-genome sequencing of various model organisms demonstrated that repetitive sequences are abundant and widely distributed in eukaryotic genomes. Typically, repeat sequences can be classified into two families: dispersed repeats and tandem repeats. Dispersed repeats are so called because of their interspersed distribution throughout the genome, such as transposons and gene paralogues; while repeat units in tandem repeated sequences are located next to other (i.e., in tandem). Each of the two families can be divided into subfamilies according to their sequence and distribution characteristics. When genomic DNA being separated by density-gradient centrifugation, tandem repeats (TRs) will be identified as satellite bands, so TRs are also named satellite DNA, among which, repeats with relatively large units (9 nucleotides long and above) are generally called minisatellites [2], and repeats with smaller units (from 1–8 nucleotides) are defined as microsatellites, also called short tandem repeats (STRs) or simple sequence repeats (SSRs) [3].

TRs can distribute everywhere in the genome, including gene bodies, untranslated regions (UTRs) and promoters [4]. Although tandem repeat sequences were historically regarded as nonfunctional junk or selfish DNA [5,6], they have attracted great attention since early 1990s when several significant diseases were reported due to tandem repeats disorder, such as Fragile X Syndrome [7,8], Huntington's disease [9] and Friedreich's ataxia [10], and to date, more than 20 diseases identified are in association with abnormity of tandem repeats [11]. Thus, TRs have been suggested as a third category of genetic variation, besides of single nucleotide polymorphisms and copy number variations [12]. Apart from their roles in diseases, TRs are also thought to have various functions in gene transcription, DNA methylation, nucleosome positioning and chromatin organization [13–15]. However, the underlying mechanisms are poorly understood, largely owing to the technical limitations in modification of TRs at definite sites in the genome *in vivo*.

Transcription activator-like (TAL) effectors are sequence-specific DNA binding domain proteins identified from plant pathogenic bacteria *Xanthomonas* in recent years [16,17]. Synthetic genes encoding TAL effector nucleases (TALENs) can be constructed *in vitro* by fusing TAL effector with *Fok* I nuclease, so that TALENs can recognize specific DNA sequences (by TAL effector) followed by creating a double strand break (DBS) in the target site (by *Fok* I nuclease). In living cells, DSBs can be repaired in two pathways [18,19]: the non-homologous end-joining (NHEJ) and homologous recombination (HR). In the former pathway, the broken sites are simply rejoined in an error-prone fashion and hence usually leading to small insertions or deletions (indels) at the

Figure 1. A (TG)$_n$ sequence in the upstream region of *ntl* and TALEN targets design. (A) Position of a (TG)$_n$ sequence in the upstream region of zebrafish *ntl* and two target sites for designing TALENs. The arrow indicates the transcription start site (TSS) of *ntl*. The (TG)$_n$ region is showed in grey adjacent boxes, and the two designed TALEN targets are described below, in which red letters underlined represent the binding sites of left (L) and right (R) TALENs, respectively. All the TALEN target sites were designed with a preceding T at 5' terminal (showed in lowercase). (B) Alignment of (TG)$_n$ sequence at upstream region of *ntl* among zebrafish, bisexual diploid and unisexual polyploid goldfish. GF: bisexual diploid goldfish (*Carassius auratus*), PZ: unisexual polyploid goldfish (*Carassius auratus*, pengze), ZF: zebrafish. (C) Structure of TALEN fusion protein, which is composed of a N-terminal translocation domain (NT), a central DNA binding domain, and a C-terminal domain containing a nuclear location site (NLS) and followed by a *Fok* I nuclease. The TALEN DNA binding domain typically comprises a tandem array of 13–28 single repeat unit [22], each one consisting of 34 highly conserved residues, in which the residues at positions 12 and 13 are called repeat-variable di-residue (RVD). Different RVDs associate specifically with different nucleotides, with NI, NG, HD, and NN accounting for each of the four nucleotides A, T, C and G, respectively. The end of C-terminal repeat unit (showed in the short green box) generally contains only 20 amino acids and is therefore referred to as 'half-repeat', which indicates a RVD specifically recognizing the nucleotide T.
doi:10.1371/journal.pone.0084176.g001

DSB sites; while in the later pathway, the DNA surrounding the DSB site is replaced with a homologous template sequence. Here, we show that TALENs can serve as a convenient tool for modifying the TRs in the genome in studying the functions of TRs. This technique might be also a potential therapeutic approach for aforementioned genetic disorders resulted from TR instability.

Figure 2. Statistics of zebrafish embryos after one day of TALENs injection at different dosage. The number of embryos scored (N) is indicated at the top and the dosage is indicated at the bottom.
doi:10.1371/journal.pone.0084176.g002

Table 1. Efficiency and toxicity evaluation of the TALEN pairs *in vitro* by SSA assay.

	USTR TALEN		DSTR TALEN	
	untransfected	transfected	untransfected	transfected
Firefly Luciferase	549473.23	5032081.60	1069911.29	27217327.77
	549367.17	5040331.13	1041310.74	28502196.83
	558803.04	5038530.93	1036584.63	28666427.67
Renilla Luciferase	35837696.22	95180434.34	114368257.99	89138298.60
	37807751.30	96921199.88	111730895.80	91984880.33
	38025654.19	94727999.14	111413707.84	91213045.37
F/R	0.014843979	0.052682645	0.009326479	0.309859449
Efficiency index	3.549091952		33.22362528	

F/R = average Firefly luciferase readout/ average Renilla luciferase readout.
Efficiency index = TALEN (F/R)/ control (F/R).
doi:10.1371/journal.pone.0084176.t001

I nuclease functions as a dimer when used to make double-strand breaks and the length of the spacer can affect the specificity of TALENs, TALENs were designed in pairs that bind opposing DNA target sites separated by a spacer, and the length of the spacer can affect the specificity of the TALRNs The range of spacer length used for searching potential targets was between 12-18bp, within the suggested region of optimal activity for TALENs [22]. The lengths of the repeat arrays were from 15 to 21 units. The TALEN target sequences were chosen after a T and ended with a T.

Materials and Methods

Animals and Ethics Statement

Zebrafish (*Danio rerio*) used in our experiments is long-fin strain. Bisexual diploid goldfish (*C. auratus*) and unisexual polyploid goldfish (*C. auratus* pengze) were purchased from nearby farms and maintained in our laboratory in the breeding season. Artificial spawning and fertilization were performed as previously reported [20]. This study was approved by the Ethics Committee of Laboratory Animal Center of Zhejiang University (Zju201306-1-11-060).

Obtaining the upstream sequence of *ntl* in zebrafish and goldfish

The upstream sequence of zebrafish *no tail* (*ntl*), a decisive developmental regulatory gene, was obtained from the nucleotide database in EMBL. The upstream sequence of goldfish *ntl* was obtained after four rounds of genome walker using GenomeWalker™ Universal Kit (CloneTech, USA). The gene-specific primers (GSPs) used are listed in Table S1 (in File S1). Among which, GSPs for the third round of genome walker (GSP5 and GSP6) were designed according to the published sequence (GenBank accession NO. EU549781), and GSPs for the fourth round of genome walker (GSP7 and GSP8) were designed according to the result of the third round. Then, the totally four rounds of genome walker products were spliced and the entire fragment was validated by PCR with gene-specific primer pair (GF-ntl-promt-seq-S/AS) and sequence analysis.

TALEN targets design

A (TG)$_n$ repeat sequence far upstream region of zebrafish *ntl* was selected as a target for TALENs. A web-based tool called TAL Effector-Nucleotide Targeter 2.0 (TALE-NT 2.0; https://tale-nt.cac.cornell.edu) [21] was used to design TALEN targets. Since *Fok*

Construction of engineered TALENs

The engineered TAL effector repeats arrays were assembled according to the designed targets, and we adopted the "unit assembly" method described by Huang et al. [23]. The units of repeats arrays were supplied by the manufacture (CWBIO, Beijing), and the procedure of assembly is just as reported formerly [23]. To construct engineered TALENs, the assembled TAL effector repeats arrays were double digested from their original vectors by *Spe* I and *Nhe* I (TaKaRa, Japan) and then cloned into pCS2-*Fok* I plasmids (CWBIO, Beijing).

Efficiency evaluation of the TALEN pairs

The efficiencies of the TALEN pairs in inducing DBS and single strand annealing (SSA) recombination were evaluated *in vitro* by Luciferase SSA recombination assay [23,24]. The luciferase SSA reporter (pSSA-luciferase) is composed of a CMV promoter and two homologous luciferase coding fragment separated by a stop codon and an inserted TALEN target sequence. Primers used for cloning TALEN target sequences into pSSA-luciferase are listed in Table S2 (in File S1). 100 ng of each TALEN pair were co-transfected with 50ng corresponding pSSA-luciferase into HEK293T cells (CWBIO, Beijing) in 24-well plates using TurboFect™ *in vitro* Transfection Reagent (Fermentas, Canada). 10 ng of Renilla luciferase driven by the β–actin promoter in co-transfection with 100 ng of corresponding TALEN pair was used as the toxic reference to judge the toxicity of designed TALEN pairs. For each sample, the test and control reactions were run in triplicate. The cells were harvested 24 hours after transfection and lysed using Luciferase Cell Lysis Buffer (NEB, USA). The relative luciferase activity was then detected by Dual-Luciferase Reporter®
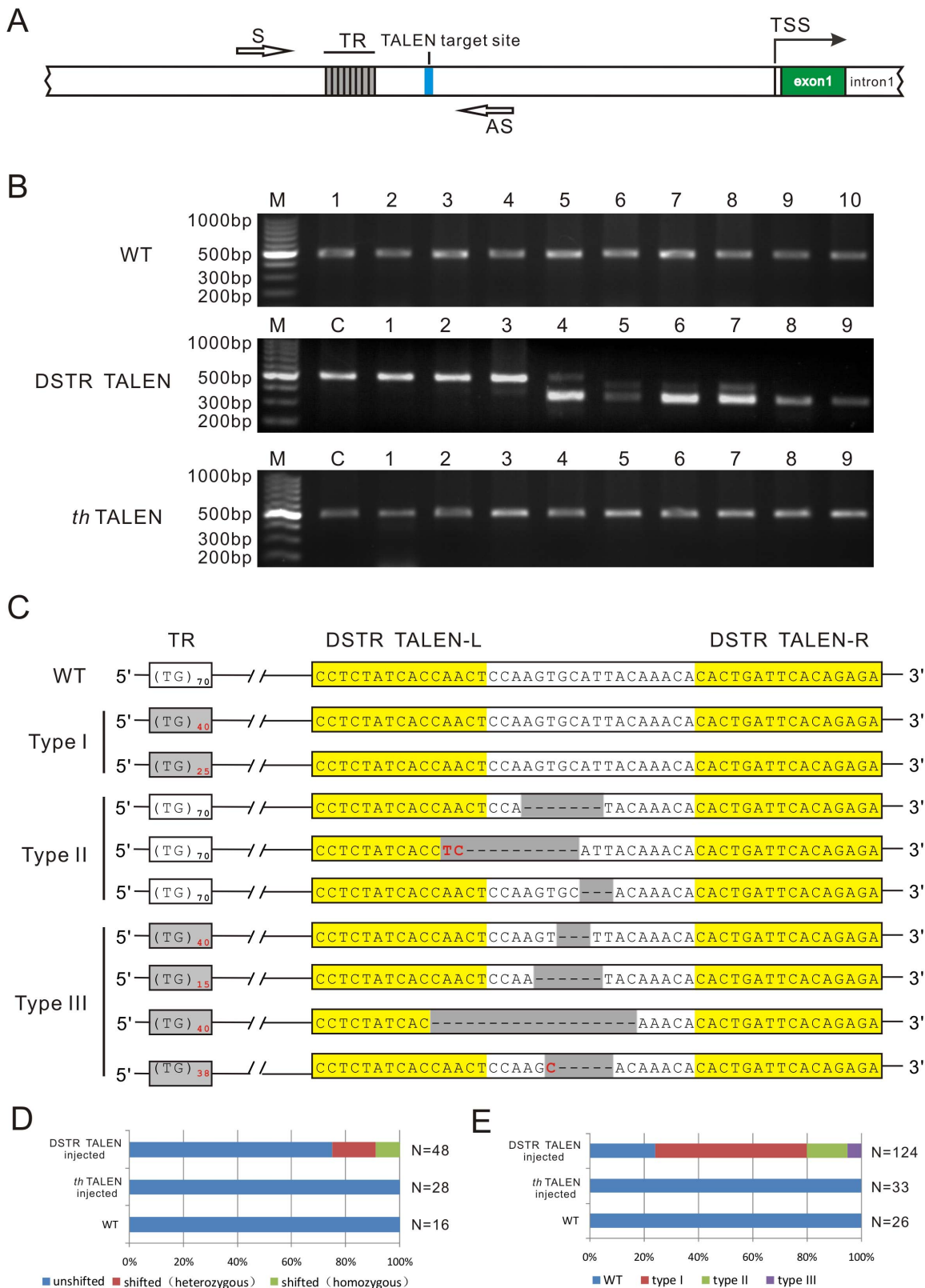
Figure 3. Mutants screening and classification in TALENs injected zebrafish embryos. (A) Analyzed region for screening potential mutant within the upstream sequence of zebrafish *ntl*. The tandem repeat (TR) is represented with grey adjacent boxes and the TALEN target site is shown in blue box. Primer pair used for PCR amplification of the potential mutant region (S and AS for sense and anti-sense primers, respectively) is indicated in blank arrows. (B) Agarose gel electrophoretogram of PCR products from wild type (WT), *th* TALEN and DSTR TALEN injected individuals. The DSTR TALEN samples are the selected representative individuals being sequencing analyzed. M. markers, C. wild type control. (C) Sequencing analysis and classification of mutants. Various mutations can be divided into three types. The number of TG unit (N) in wild type (WT) is about 70. No more than 3

TG unit of variation was detected in the strain used in this study. Significant size changes in the mutants were highlighted in red number. The changes of different indels in the sequence of the TALEN target site were highlighted in grey. Deletions and insertions were indicated by dashes and red letters, respectively. The binding sites of left and right TALENs (DSTR TALEN-L and TALEN-R) are highlighted in yellow. (D) Percentages of samples with unshifted and shifted (heterozygous and homozygous) bands detected by PCR amplification. (E) Percentages of each mutation types in TALEN-injected embryos. Only the sequenced clones in the shifted heterozygote and homozygote were calculated.
doi:10.1371/journal.pone.0084176.g003

(DLR^{TM}) Assay System (Promega, USA) and measured by SpectraMax L Luminescence Microplate Reader (Molecular Devices, USA). The specific value of average firefly luciferase readout/average renilla luciferase readout (F/R) was calculated. The efficiency index was obtained by comparing the F/R value of TALEN with that of control.

Transcription of TALENs *in vitro* and microinjection

The constructed pCS2-TALE-*Fok* I vectors were linearized with *Not* I (TaKaRa, Japan) as templates, from which the capped mRNAs of TALEN pair were transcribed using mMESSAGE mMACHINE Sp6 Kit (Ambion, U.S.A.). Capped mRNAs of TALEN pair were coinjected into zebrafish embryos at the 1-cell stage.

Mutants screening in TALENs injected zebrafish embryos

TALEN-injected zebrafish embryos were maintained in 0.1× Hank's solution at 28.5°C. The survival rate was accounted at 1 day post fertilization (dpf). Genomic DNA from single 4 dpf embryos was extracted following our previously reported procedure [20]. A ~470bp DNA fragment (in wild type) encompassing the TALEN target site and the TR region was amplified by PCR using the primers as follows: 5'-TCCTGTTCAATGTGTTT-TATCAGTATGC-3' (forward) and 5'-CTTAATTTCTT-CATGTTGTTCTAATGCAA-3' (reverse). PCR products were run on the agarose gel and then validated by sequencing.

Results

The (TG)$_n$ at the upstream region of zebrafish *ntl* is suitable for examining the effect of TALENs on TR modification

To examine the effect of TALENs on TR modification, the length of the target TRs should be long enough to induce indels and the adjacent sequence should have suitable binding sites to design TALEN pairs. Moreover, it is better that artificial length change of the target TRs would not elicit lethal effect on the embryo.

Table 2. Type of clones in the examined DSTR TALEN individuals.

No. of samples	Type of clones			
1	Wt			
2	Wt		Type II	
3	Wt		Type II	
4	Wt	Type I	Type II	
5		Type I	Type II	Type III
6		Type I		Type III
7		Type I	Type II	
8		Type I		
9		Type I		

doi:10.1371/journal.pone.0084176.t002

By searching in the nucleotide database in EMBL, we identified a 140bp long imperfect tandem TG repeat sequence (represented with (TG)$_n$, where n = 70) about 2.7 kbp upstream of the zebrafish *ntl* that has two candidate TALEN-binding sites (Figure 1 A). Since *ntl* is a decisive regulatory gene of development, we examined whether this TR is evolutionary conserved and evaluated if the length change of this sequence was tolerable by comparing the upstream sequence among zebrafish, bisexual diploid and unisexual polyploid goldfish. After four rounds of genome walker, the upstream sequence of *ntl* gene in the two subspecies of goldfish was obtained and aligned (Figure S1 in File S1). Sequencing analysis showed that an imperfect (TG)$_n$ repeat exists upstream of *ntl* in both bisexual diploid and unisexual polyploid goldfish. The position of (TG)$_n$ in both subspecies of goldfish is in accordance with that in zebrafish. Aligning comparison showed that the length of the (TG)$_n$ between zebrafish and the bisexual diploid goldfish is almost equal, but remarkably shorter (about 110bp) in the unisexual polyploid goldfish (Figure 1 B). This result suggests that the (TG)$_n$ is an evolutionary conservative element and the length change is tolerable. Therefore, the (TG)$_n$ upstream of zebrafish *ntl* is suitable for examining whether engineered TALENs can be employed to modify the length of TRs.

An optimal target site for TALEN pair is at the downstream of the TR

In order to screen an optimal TALEN pair for inducing a DBS and SSA recombination, two pairs of TALEN plasmids, named USTR TALEN-L/R and DSTR TALEN-L/R, were designed and constructed targeting to the adjacent upstream (162bp) and downstream (129bp) of the (TG)$_n$ (Figure 1A). All the TALEN target sites were designed with a preceding T. Each TALEN pair was designed with a spacer length of 17bp and repeats arrays length of 15 or 16 units (all the last units contained only 20 amino acids and were therefore referred to as 'half-repeats', specifically recognizing the nucleotide T). The structure of TALEN fusion protein is described in Figure 1C.
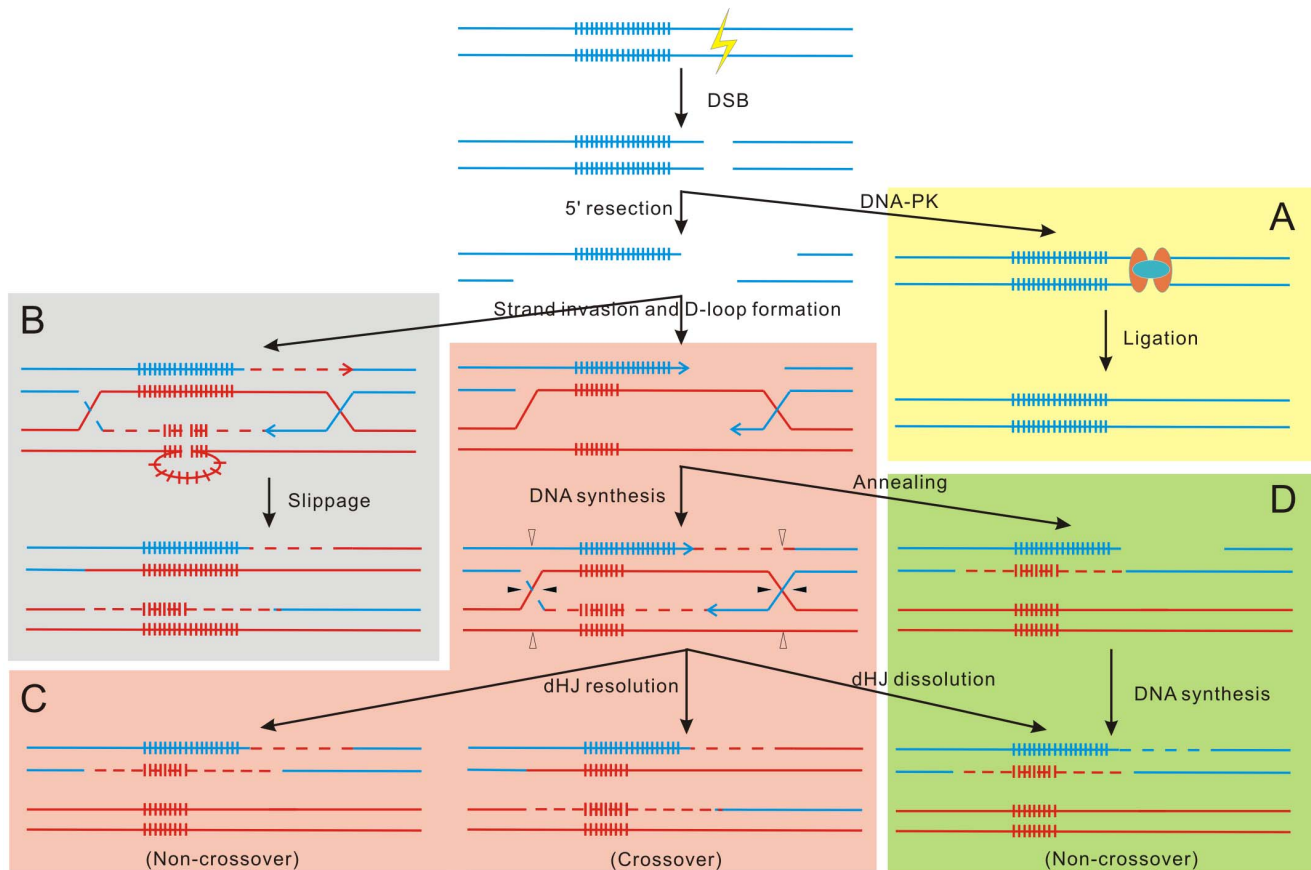
To determine the efficiency of the two TALEN pairs, two firefly luciferase SSA reporters were constructed by inserting the USTR TALEN or DSTR TALEN target sequences into the cloning site (Figure 2A). Because the coding region of the firefly luciferase was engineered with two 870bp homologous arms, which were separated by a stop codon and a TALEN target sequence, no active firefly luciferase would be expressed from the reporter plasmid before co-transfection of a functional TALEN pair. The binding of a functional TALEN pair will create a DSB, which after error-prone NHEJ-mediated repair can generate an active firefly luciferase gene (Figure S2 in File S1). Thus, comparing to the firefly luciferase SSA reporter transfected cells, the gain of firefly luciferase signal in the firefly luciferase SSA reporter and the corresponding TALEN pair co-tansfected cells can be taken as an indicator of TALEN activity, while the loss of *β–actin* promoter driven Renilla luciferase signal in the toxic control can be seen as an indicator of TALEN toxicity, due to off-target cleavage usually causing apoptosis of the transfected cells [24].

Luciferase detection showed high level of Renilla luciferase signal in the URTR TALEN or DSTR TALEN pair co-tansfected control cells (Table 1), suggesting that the cell toxicity of both

Figure 4. Speculative mechanisms involved in TALEN induced DSB repair. Blue lines represent the genome with DSB sites, and clusters of vertical bars indicate the TR region. The DSB ends can be bound by two groups of proteins independently: the binding of DNA-dependent protein kinase (DNA-PK) complex (Ku 70 and Ku80) and the following ligase IV seals the gap by direct rejoin the broken ends, which is termed non-homologous end-joining (NHEJ) pathway (A); While the binding of MRN complex and Exo1 nuclease initiates the 5′-3′ resection of the ends, which is followed by either a replication slippage pathway (B) or homologous recombination (HR) pathway (C, D). In the replication slippage pathway, the TR region forms a secondary structure and leads to mispairing between the template and the newly-synthesized DNA strand. In the HR pathway, the 3′ overhang invades into the homologous template DNA (red lines) and primes DNA synthesis (dash lines) to form a structure called D-loop, which will result in a double Holliday junction (dHJ). dHJ can either be resolved by strand cleavage with or without crossover, which is referred as classical DSB repair (DSBR) pathway of HR (C), and dHJ can also be dissolved by helicases to generate a non-crossover (D). Alternatively, D-loop can be directly dissociated through a synthesis-dependent strand annealing (SDSA) pathway, which results in exclusively non-crossover products (D).
doi:10.1371/journal.pone.0084176.g004

TALEN1 and TALRN-2 pairs is very low. As showed in Table 1, the efficiency index of DSTR TALEN pairs in inducing a DBS and SSA recombination was much higher than that of URTR TALEN, indicating that the optimal target sequence for TALEN pair is at downstream of the TRs. Therefore, DSTR TALEN pair was chosen for the further experiment.

The repair of DSB elicited size modification of the adjacent (TG)$_n$

To determine a appropriate dosage, different doses of the DSTR TALEN mRNA pair were microinjected into 1-cell stage zebrafish embryos separatly. At the dosage of 125 pg per embryo, the injected embryos developed as normally as the uninjected control and exhibited no specific abnormality at 1 day post fertilization (Figure 2). When the dosage was increased to 250 pg per embryo, slightly higher percentage of dead and unspecific deformed was observed (Figure 2), suggesting that substantial off-targeting was induced in this dosage. Therefore, the dosage of 125 pg per embryo was used in further experiment.

To sreen and analyze mutants, a ~470bp fragment encompassing the TR and the DSRT TALEN target site upstream of *ntl*

was PCR amplified (Figure 3A) from genomic DNA of single zebrafish embryos at 4 days post fertilization. In all examined wild type individuals (N = 16), the amplified fragment is almost the same. However, about 27.1% of the examined DSTR TALEN injected embryos (N = 48) exhibited obvious shifted bands in heterozygotic or homozygotic manner (Figure 3B, D). To confirm that the length change at the TR locus was specifically trigggered by DSTR TALEN, rather than by the injection or subsequent handing of the embryos, a published *th* TALEN pair targeting the *tyrosine hydroxylase* gene [25] was emploied a control. In all the examined control *th* TALEN embryos (N = 28), the length of the amplified fragment is the same as observed in wild type embryos (Figure 3B, D). These results indicated that the change in size of the fragment was unequivocally induced by DSTR TALEN.

The sequence of PCR products amplified from representative individuals were analyzed. To detect different clones in an single embryo, more than 10 clones from each of the individuals were sequenced and analyzed. Sequencing analysis of all the examined clones (N = 124) obtained from the heterozygotic or homozygotic mutant embryos showed that there were three types of mutation pattern (Figure 3C, E). About 56.3% of the sequenced clones was

classified into type I. In this type the $(TG)_n$ was variably shortened, while the sequence of the TALEN target site remained unchanged. About 14.5% of the sequenced clones was classified into type II. In this type the $(TG)_n$ remained unchanged, while various indels emerged at the TALEN target site. About 5% of the sequenced clones was classified into type III. In this type sequence change occurred within both the $(TG)_n$ and the TALEN target site. Most of the examined embryos are heterozygotes or mosaics containing two or three types of clones (Table 2). Strikingly, the TR was contracted to various sizes and the shortest one maintained only 21% of the repeat unit. These results demonstrated that the TALEN-induced DSB resulted in mutations not only at the DSB site but also within the adjacent TR region.

Discussion

Since its discovery several years before, TAL effector fused with *Fok* I nuclease has been widely applied in targeting genes in plant [26], nematodes [27], fishes [28] and mammals [29,30]. In this experiment, we successfully modified the size of a tandem repeat $(TG)_n$ in zebrafish genome using engineered TALENs. Our results also showed that, during the repair of the DSB induced by TALENs, mutations can occur separately within the TALEN target region or within the adjacent TR region, or simutaniusly at both regions. Therefore, all the desired mutant types of TR could be convenintly obtained by employing engineered TALENs.

The tandem repeat $(TG)_n$ was first discovered in the genomes from yeast to human in early 1980s by two independent groups [31,32]. The following *in vitro* experiments showed that the activity of *chloramphenicol acetyltransferase* gene was enhanced with a TG-element, and the maximum enhancement was obtained with 30–40bp of $(TG)_n$; when the $(TG)_n$ exceeded 130bp, the gene activity declined dramatically to fivefold less compared with a 50bp $(TG)_n$ [33]. Dutreix et al also showed *in vitro* that the binding of RecA protein to $(TG)_n$ or $(CA)_n$ sequences with an increasing affinity, and the sequence recombination was promoted from 30% to 80% and 100% for DNA containing 7, 16 and 39 TG repeats [34]. These observations suggested that $(TG)_n$ played an important role in modulating gene expression. Recent publications suggest that the loss or gain of repeats may affect the binding of the transcriptional regulatory proteins such as IHF to the promoter [14], and that variations in repeats length can also affect gene expression through changing local nucleosome positioning and chromatin structure [15]. It is also proposed that repeats length variations might also affect DNA methylation [35–37]. The modification of the size of $(TG)_n$ region and its flanking sequence *in vivo* provides a strategy to explore the detailed mechanisms and the roles of $(TG)_n$ in gene expression and DNA methylation.

Although several hypothesis, including replication slippage [38], gene conversion [39], and unequal crossing over [40] have been proposed, the actual molecular mechanisms involved in TRs instability remains unproved. In this experiment, the engineered TALEN pair induced a DSB adjacent to the $(TG)_n$ region in zebrafish genome and elicited three types of mutations with or without TR variation. The type II mutation with indels within the TALEN target region while the $(TG)_n$ region unchanged is

undoubtedly due to non-homologous end-joining (NHEJ, Figure 4A), which is an error-prone pathway and hence usually introduce in mutations. Protein factors involved in this progress are well characterized, including DNA-dependent protein kinase (DNA-PK) complex (Ku 70 and Ku80) and ligase IV, which have been extensively reviewed recently [41]. The type I and type III mutation with repeat number variation in the $(TG)_n$ region might be caused by replication slippage and homologous recombination (HR) through several pathways as illustrated in Figure 4B-D. Replication slippage is also called slipped-strand mispairing, in which the TR region forms a secondary structure and leads to mispairing between the template and the newly-synthesized DNA strand. The TR will contract, as observed in this experiment, if the template strand loops out (Figure 4B) and will expand if the newly-synthesized strand loops out. HR is a template-dependent repair progress and requires the formation of a displacement-loop (D-loop) followed by a DNA cross structure called double Holliday junction (dHJ), which can be resolved by strand cleavage with or without crossover and also can be dissolved by helicases to generate a non-crossover (Figure 4C and D). Alternatively, D-loop can also be dissociated through a synthesis-dependent strand annealing (SDSA) pathway, which results in exclusively non-crossover products (Figure 4D). The choice between NHEJ and HR after DSB might be dependent on the species, cell type and stage of the cell cycle [42,43]. On the molecular level, the binding of Ku70-Ku80 heterodimer to DSB site and the following recruiting of ligase IV prevent the 5' resection and the following HR progress, while the 5'-3' resection initiated by MRN complex and Exo1 nuclease greatly antagonizes the NHEJ pathway [44]. Former studies showed that DSBs elicit the TR instability on a number of occasions [45–47]. Recent studies reported that the probability of HR is greatly enhanced when high doses of donor template sequences are supplied [48,49]. It is possible that co-injection of engineered TALEN mRNAs with corresponding donor templates will gain precisely desired length of TRs.

Supporting Information

File S1 Combined file of supporting information files. The contents include: Table S1. Primers for goldfish *ntl* promoter cloning. Table S2. Primers for pSSA-luciferase reporter construction. Figure S1. Alignment of upstream sequence of *ntl* between bisexual diploid and unisexual polyploid goldfish. Figure S2. Simplified structure of pSSA-luciferase reporter and sketch map of the SSA assay.
(DOC)

Acknowledgments

We thank Dr. Zhang Bo for kindly providing the *th* TALEN plasmid.

Author Contributions

Conceived and designed the experiments: CL. Performed the experiments: JZ WH YH. Analyzed the data: JZ WH CL. Contributed reagents/materials/analysis tools: CL. Wrote the paper: WH CL.

References

1. Britten RJ, Kohne DE (1968) Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science 161: 529–540.
2. Richard GF, Kerrest A, Dujon B (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol Mol Biol Rev 72: 686–727.
3. Fondon JW 3rd, Hammock EA, Hannan AJ, King DG (2008) Simple sequence repeats: genetic modulators of brain function and behavior. Trends Neurosci 31: 328–334.
4. Legendre M, Pochet N, Pak T, Verstrepen KJ (2007) Sequence-based estimation of minisatellite and microsatellite repeat variability. Genome Res 17: 1787–1796.

5. Ohno S (1972) So much "junk" DNA in our genome. Brookhaven Symp Biol 23: 366–370.
6. Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284: 601–603.
7. Fu YH, Kuhl DP, Pizzuti A, Pieretti M, Sutcliffe JS, et al. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell 67: 1047–1058.
8. Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, et al. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65: 905–914.
9. Rubinsztein DC, Leggo J, Coles R, Almqvist E, Biancalana V, et al. (1996) Phenotypic characterization of individuals with 30–40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36–39 repeats. Am J Hum Genet 59: 16–22.
10. Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, et al. (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271: 1423–1427.
11. Usdin K (2008) The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. Genome Res 18: 1011–1019.
12. Gemayel R, Vinces MD, Legendre M, Verstrepen KJ (2010) Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu Rev Genet 44: 445–477.
13. Martienssen R, Lippman Z, May B, Ronemus M, Vaughn M (2004) Transposons, tandem repeats, and the silencing of imprinted genes. Cold Spring Harb Symp Quant Biol 69: 371–379.
14. Martin P, Makepeace K, Hill SA, Hood DW, Moxon ER (2005) Microsatellite instability regulates transcription factor binding and gene expression. Proc Natl Acad Sci U S A 102: 3800–3804.
15. Vinces MD, Legendre M, Caldara M, Hagihara M, Verstrepen KJ (2009) Unstable tandem repeats in promoters confer transcriptional evolvability. Science 324: 1213–1216.
16. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, et al. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326: 1509–1512.
17. Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326: 1501.
18. Mussolino C, Cathomen T (2012) TALE nucleases: tailored genome engineering made easy. Curr Opin Biotechnol.
19. Joung JK, Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol 14: 49–55.
20. Ma S, Huang W, Zhang L, Zhao S, Tong Y, et al. (2011) Germ cell-specific DNA methylation and genome diploidization in primitive vertebrates. Epigenetics 6: 1471–1480.
21. Doyle EL, Booher NJ, Standage DS, Voytas DF, Brendel VP, et al. (2012) TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Res 40: W117–122.
22. Bogdanove AJ, Voytas DF (2011) TAL effectors: customizable proteins for DNA targeting. Science 333: 1843–1846.
23. Huang P, Xiao A, Zhou M, Zhu Z, Lin S, et al. (2011) Heritable gene targeting in zebrafish using customized TALENs. Nat Biotechnol 29: 699–700.
24. Bhakta MS, Segal DJ (2010) The generation of zinc finger proteins by modular assembly. Methods Mol Biol 649: 3–30.
25. Zu Y, Tong X, Wang Z, Liu D, Pan R, et al. (2013) TALEN-mediated precise genome modification by homologous recombination in zebrafish. Nat Methods 10: 329–331.
26. Shan Q, Wang Y, Chen K, Liang Z, Li J, et al. (2013) Rapid and efficient gene modification in rice and Brachypodium using TALENs. Mol Plant.
27. Wood AJ, Lo TW, Zeitler B, Pickle CS, Ralston EJ, et al. (2011) Targeted genome editing across species using ZFNs and TALENs. Science 333: 307.
28. Sander JD, Cade L, Khayter C, Reyon D, Peterson RT, et al. (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENS. Nat Biotechnol 29: 697–698.
29. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, et al. (2011) Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol 29: 731–734.
30. Tesson L, Usal C, Menoret S, Leung E, Niles BJ, et al. (2011) Knockout rats generated by embryo microinjection of TALENs. Nat Biotechnol 29: 695–696.
31. Miesfeld R, Krystal M, Arnheim N (1981) A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human delta and beta globin genes. Nucleic Acids Res 9: 5931–5947.
32. Hamada H, Petrino MG, Kakunaga T (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc Natl Acad Sci U S A 79: 6465–6469.
33. Hamada H, Seidman M, Howard BH, Gorman CM (1984) Enhanced gene expression by the poly(dT-dG).poly(dC-dA) sequence. Mol Cell Biol 4: 2622–2630.
34. Dutreix M (1997) (GT)n repetitive tracts affect several stages of RecA-promoted recombination. J Mol Biol 273: 105–113.
35. Pieretti M, Zhang FP, Fu YH, Warren ST, Oostra BA, et al. (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. Cell 66: 817–822.
36. Knight SJ, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, et al. (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell 74: 127–134.
37. Ritchie RJ, Knight SJ, Hirst MC, Grewal PK, Bobrow M, et al. (1994) The cloning of FRAXF: trinucleotide repeat expansion and methylation at a third fragile site in distal Xqter. Hum Mol Genet 3: 2115–2121.
38. Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365: 274–276.
39. Gangloff S, Zou H, Rothstein R (1996) Gene conversion plays the major role in controlling the stability of large tandem repeats in yeast. EMBO J 15: 1715–1725.
40. Warren ST (1997) Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13. Science 275: 408–409.
41. Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem 79: 181–211.
42. Chapman JR, Taylor MR, Boulton SJ (2012) Playing the end game: DNA double-strand break repair pathway choice. Mol Cell 47: 497–510.
43. Symington LS, Gautier J (2011) Double-strand break end resection and repair pathway choice. Annu Rev Genet 45: 247–271.
44. Amunugama R, Fishel R (2012) Homologous recombination in eukaryotes. Prog Mol Biol Transl Sci 110: 155–206.
45. Wojcik EA, Brzostek A, Bacolla A, Mackiewicz P, Vasquez KM, et al. (2012) Direct and inverted repeats elicit genetic instability by both exploiting and eluding DNA double-strand break repair systems in mycobacteria. PLoS One 7: e51064.
46. Mittelman D, Moye C, Morton J, Sykoudis K, Lin Y, et al. (2009) Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. Proc Natl Acad Sci U S A 106: 9607–9612.
47. Pollard LM, Bourn RL, Bidichandani SI (2008) Repair of DNA double-strand breaks within the (GAA*TTC)n sequence results in frequent deletion of the triplet-repeat sequence. Nucleic Acids Res 36: 489–500.
48. Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, et al. (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol 25: 1298–1306.
49. Valerie K, Povirk LF (2003) Regulation and mechanisms of mammalian double-strand break repair. Oncogene 22: 5792–5812.