Chapter 2

TALEN-Mediated Mutagenesis and Genome Editing

Alvin C.H. Ma, Yi Chen, Patrick R. Blackburn, and Stephen C. Ekker

Abstract

Transcription activator-like effectors (TALEs) are important genomic tools with customizable DNA-binding motifs for locus-specific modifications. In particular, TALE nucleases or TALENs have been successfully used in the zebrafish model system to introduce targeted mutations via repair of double-stranded breaks (DSBs) either through nonhomologous end joining (NHEJ) or by homology-directed repair (HDR) and homology-independent repair in the presence of a donor template. Compared with other customizable nucleases, TALENs offer high binding specificity and fewer sequence constraints in targeting the genome, with comparable mutagenic activity. Here, we describe a detailed in silico design tool for zebrafish genome editing for TALENs and CRISPR/Cas9 custom restriction enzymes using Mojo Hand 2.0 software.

Key words TALEN, Customized nucleases, Zebrafish, Genome editing, Golden Gate, FusX

1 Introduction

TALEs are naturally occurring transcription factors isolated from plant pathogen *Xanthomonas* [1, 2]. Each TALE has a DNA-recognizing TALE domain made up of a tract of almost identical repetitive units (33–35 amino acid residues) and a partial (or half) repeat unit at the end. Within each unit, the two repeat-variable di-residues (RVDs) are solely responsible for the binding specificity of the unit toward a DNA nucleotide in a highly predictable fashion [3, 4]. Commonly used RVDs include NI and NN for adenine; HD for cytosine; NK, NN, and NH for guanine; and NG for thymine [3–6]. Because of the 1:1 RVD to nucleotide modularity of the TALE domain, it can be engineered to target almost any DNA sequence in the genome and can be fused with different functional domains including nuclease, transcription activator/repressor, and methyltransferases. TALEs represent important genomic tools for locus-specific modifications [7–14]. In particular, TALENs have been extensively used for targeted mutations in vitro and in different model organisms [15–22].
Diverse methodologies have been developed to assemble the modular TALE domain, with the Golden Gate TALEN assembling method (Golden Gate TALEN Kit 2.0) being widely used because of its flexibility, low start-up cost, and requirement of minimal, common molecular cloning reagents [23]. We previously reported the first use of GoldyTALEN in targeted zebrafish genome editing through both NEHJ and HDR [8]. We also described a simple and highly active GoldyTALEN design with only 15 RVDs (or 14.5 TALE repeats) [22]. To further facilitate TALEN-mediated high-throughput genome editing, we subsequently developed a modified Golden Gate TALEN assembling FusX system (Ma et al., manuscript in preparation). The new system increased assembling efficiency, but shortened assembling time without affecting mutagenic activity and compatibility.

With the rapid development of novel genome engineering tools such as TALENs and CRISPR/Cas9 systems [24], new software tools are needed to aid biologists in designing and constructing high-efficiency reagents that can be used to make tailored changes within any model system of interest. Through a better understanding of the cell’s endogenous DNA repair mechanisms, we can improve reagent design and targeting to achieve predictable outcomes. Microhomology-mediated end joining (MMEJ) appears to be a dominant repair pathway for TALEN, and RNA-guided engineered nuclease (RGEN) induced double-stranded breaks and has been used to generate predictable out-of-frame deletions and to incorporate donor DNA sequences in a highly efficient manner [25].

We previously presented the web-based Mojo Hand designer tool [26]. In the latest version 2.0, algorithm adheres to the same general steps that the original algorithm follows with the integration of new features including .bed file creation, microhomology, and out-of-frame scoring. Another major consideration was the incorporation of user-generated next-generation sequencing data in reagent design to deal with the tremendous inter- and intrastrain genetic variation during zebrafish genome targeting. In the current version, high-depth RNAseq datasets were integrated to simplify design and reduce time and cost through the avoidance of regions rich in single nucleotide polymorphisms (SNPs). Here, we describe a detail protocol of targeted zebrafish genome editing through NHEJ and HDR, respectively, using TALENs or CRISPR/Cas9 using the open access Mojo Hand 2.0 software.

## 2 Materials

### 2.1 Zebrafish Embryo Genotyping and RFLP Assay

1. Genomic DNA extraction buffer: 10 mM Tris–HCl, pH 8.3, 50 mM KCl.
2. 10% Tween-20.
3. 10% NP-40.
4. Proteinase K solution (recombinant, PCR grade, 14–22 mg/mL in 10 mM Tris–HCl, pH 7.5, Roche Life Science).
5. PCR reaction mix (see Note 1).
6. Restriction enzyme.
7. Agarose.
8. TAE buffer (1x): 40 mM Tris–HCl, 20 mM acetic acid, 1 mM EDTA, pH 8.4.

2.2 TALEN Assembling

2. Last half-repeat components pLR-NI, pLR-HD, pLR-NN, and pLR-NG (Addgene #31006, #30984, #31017, #30995).
3. RCIscript-GoldyTALEN backbone (Addgene, cat# 38142).
4. T4 DNA ligase (2,000,000 U/mL, New England Biolabs).
5. BsmBI (New England Biolabs) (optional, see Note 2).
7. Standard thermocycler.
8. Competent E. coli cell.
9. LB agar plate with ampicillin (100 μg/mL).
10. LB medium with ampicillin (100 μg/mL).
11. 20 mg/mL X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside).
12. 0.1 M IPTG (isopropylthio-β-galactoside).
13. Colony PCR screening primers: TAL_F1 (ttggcgtcgccaa-cagttg) and TAL_R2 (ggcgacgaggtggtgag) [23].
14. Sequencing primers: TAL_F1, TAL_R2, RVD-MM-F (ctcacacccgatcaggtc), and RVD-MM-R (gacctgatcgggtgtgag) (see Note 3) [24].

2.3 In Vitro Transcription

2. 3 M sodium acetate, pH 5.0.
3. 70% ethanol.
4. Ambion mMESSAGE mMACHINE® T3 Transcription Kit (Life Technologies).
5. Deionized water.
6. Lithium chloride precipitate solution: 7.5 M LiCl, 50 mM EDTA, pH 8.0.
3 Methods

3.1 Designing TALEN with Mojo Hand 2.0 (Fig. 1; See Note 4)

1. Select genomic region for TALEN targeting (Fig. 2; see Note 5).
2. Sequence input into Mojo Hand 2.0 (http://talendesign.org/).
3. Identification of binding sites with the following parameters (see Note 6):

![Flowchart of Mojo Hand 2.0](image)

- Gene ID (NCBI Gene Database)
- Contig Beginning and End Indices (NCBI Nucleotide Database)
- Raw Sequence
- FASTA File input

Input Processing

- TALEN Site Finding Algorithm
- CRISPR Cas9 Site Finding Algorithm
- Cas9 Nickase Site Finding Algorithm

Restriction Enzyme Analysis &
Microhomology Analysis

Output

- TALEN Recipe
- Predicted Deletions

Put to .bed File

Visualization with Custom Genome using IGV

Fig. 1 Flowchart of Mojo Hand 2.0. Input formats are in blue. Features new to Mojo Hand 2.0 are highlighted in yellow. Output is a report containing potential binding sequences, RVDs for TALENS, oligos for CRISPR/Cas9 nucleases or nickases, microhomology score, and out-of-frame scoring. The output can be further processed to create a .bed file, which can be loaded into other tools such as IGV, generate customized recipes for each TALEN, or analyzed for predicted deletions.
21

(a) Length of TAL-binding domain 15 RVDs
(b) Spacer length between 14 and 18 bp
(c) Unique restriction site within the spacer for RFLP assay of NHEJ-mediated mutagenesis (optional for large genomic fragment deletion using two pairs of TALEN; see Note 7)
(d) T nucleotide upstream of both TAL-binding domains

4. Restriction enzyme analysis.
5. Mojo Hand output.
6. Select TALEN design with desired microhomology score above or out-of-frame score if predictable deletion through MMEJ is desirable (see Note 8).
7. Generate BED file to be used in conjunction with Integrated Genomics Viewer (IGV) (see Note 9).

3.2 Genotyping
Targeted Genomic Locus

Fig. 2 Typical genomic region for TALEN targeting in different types of mutagenesis. Either TALEN pair 1 or 2 can be used in case of loss-of-function mutagenesis, and TALEN pairs 1 and 2 are used together for deletion of large genomic fragment. Blue arrow indicated primer pairs for RFLP or PCR screening of mutagenesis

(a) Length of TAL-binding domain 15 RVDs
(b) Spacer length between 14 and 18 bp
(c) Unique restriction site within the spacer for RFLP assay of NHEJ-mediated mutagenesis (optional for large genomic fragment deletion using two pairs of TALEN; see Note 7)
(d) T nucleotide upstream of both TAL-binding domains

1. Design primers to amplify the targeted locus (see Note 10).
2. Extract genomic DNA from zebrafish embryos of the targeted fish line (see Note 11):
   (a) To prepare 1 mL working extraction buffer, freshly add 30 μL 10% Tween-20, 30 μL 10% NP-40, and 10 μL proteinase K to 950 μL genomic DNA extraction buffer.
   (b) Transfer embryos to centrifuge tube and remove excess embryo water.
(c) Add working extraction buffer (50 μL per embryo).
(d) Incubate at 55 °C with shaking ≥ 4 h.
(e) Incubate at 98 °C for 10 min to inactivate proteinase K.
(f) Store genomic DNA at −20 °C until PCR.
(g) Typically, 5 μL of genomic DNA solution is used in 25 μL PCR.

3. PCR amplify the target locus.

4. Test RFLP assay:
   (a) PCR with RFLP assay primers (see Note 10, Fig. 2).
   (b) Digest 10 μL PCR product with appropriate restriction enzyme.
   (c) Resolve digested product on 1.5% agarose gel.
   (d) PCR product should be completely digested into two correctly sized bands.

5. Confirm sequence of the targeted locus by Sanger sequencing, identify any polymorphic region affecting TALEN-binding sites, and redesign TALEN if necessary.

### 3.3 Design Short Single-Stranded Donor Oligo

1. Design donor oligo with the following parameters:
   (a) Around 50 base pairs in length
   (b) Mutated nucleotide(s) in the middle part of the oligo
   (c) Unique restriction site added in the middle of the oligo by introducing silent mutations to allow easy screening of donor incorporation with RFLP assay

### 3.4 TALEN Assembling with FusX System (3 Days)

#### 3.4.1 Day 1

1. Break down the 15-RVD TALE domain from 5′ to 3′ into six building blocks from different libraries of the FusX kit according to the formula 3 (pFusX-1) + 3 (pFusX-2) + 3 (pFusX-3) + 3 (pFusX-4) + 2 (pFus_B2) + 1 (pLR) (Fig. 3).
   For example, a TALEN arm with the following targeting sequence: 5′-ATTGACTTCAGAGAG-3′.
   Corresponding RVD sequences: NI NG NG NN NI HD NG NG HD NI NN NI NN NI NN.
   List of building blocks required for each TAL:

<table>
<thead>
<tr>
<th>Library</th>
<th>RVD sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pFusX-1</td>
<td>NI NG NG</td>
</tr>
<tr>
<td>pFusX-2</td>
<td>NN NI HD</td>
</tr>
<tr>
<td>pFusX-3</td>
<td>NG NG HD</td>
</tr>
<tr>
<td>pFusX-4</td>
<td>NI NN NI</td>
</tr>
<tr>
<td>pFus_B2</td>
<td>NN NI</td>
</tr>
<tr>
<td>pLR</td>
<td>NN</td>
</tr>
</tbody>
</table>
2. Mix 25–50 ng of each vector in a PCR tube with 50 ng RCIscript-GoldyTALEN backbone (see Note 12).

3. (Optional) Add to each reaction 1 μL 10× NEBuffer 3.1 and 0.5 μL BsmBI, and make up to 10 μL with deionized water (see Note 2).

4. (Optional) Incubate at 55 °C for 30 min (see Note 2).

5. Add to each reaction 1.5 μL 10× T4 DNA ligase reaction buffer, 0.5 μL T4 DNA ligase, and 0.5 μL Esp3I, and make up to 15 μL with deionized water.

6. Run the following program in thermocycler:
   (a) 37 °C, 5 min, and 16 °C, 10 min → ten cycles
   (b) 37 °C, 15 min
   (c) 80 °C, 5 min
   (d) 4 °C forever

7. Transform 3–5 μl of the reaction product, and plate ~1/5 of the recovered transformants on LB agar plate with ampicillin, 40 μL X-Gal (20 mg/mL), and 40 μL 0.1 M IPTG.

8. Incubate LB agar plate at 37 °C overnight.

### 3.4.2 Day 2

1. Pick 2–4 white colonies for colony PCR with primers TAL_F1 and TAL_R2.

2. PCR with the following program (see Note 13):
   (a) 95 °C, 10 min
   (b) 95 °C, 30 s; 55 °C, 30 s; and 72 °C, 3 min → 30 cycles
3. Resolve PCR product in 1% agarose gel and identify positive clones (Fig. 4a).

4. Culture positive colonies overnight 37 °C in LB with ampicillin.

3.4.3 Day 3

1. Mini-prep overnight cultures of selected positive clones.

2. Verify assembled TALEN by Sanger sequencing with TAL_F1 and TAL_R2 (see Note 3).

3.5 Synthesizing TALEN-Encoding mRNA and Microinjection into One-Cell Zebrafish Embryos

1. Linearize TALEN-encoding plasmid with SacI.

2. Purify linearized plasmid by ethanol precipitation and quantify purified plasmids.

3. Set up in vitro transcription reaction with Ambion mMMESSAGE mMACHINE® T3 Transcription Kit according to manufacturer’s instruction (see Note 14).

4. Purify and quantify transcribed mRNA:
   (a) Add 50 μL LiCl precipitation solution to each transcription reaction.
   (b) Precipitate at −20 °C ≥1 h.
   (c) Centrifuge at 4 °C, 12,000 × g, for 15 min.
   (d) Remove supernatant and wash with 70% ethanol.
   (e) Centrifuge at 4 °C, 12,000 × g, for 5 min.
   (f) Remove supernatant and air-dry pellet.
   (g) Resuspend pellet in 50 μL deionized water and quantify mRNA.

Fig. 4 (a) Typical colony PCR result after TALEN assembly. Lane 1 is a negative clone with empty GoldyTALEN backbone showing a ~0.65 kb band, and lane 2 is a positive TALEN clone showing the laddering effect with a band at ~1.5 kb. (b) Typical RFLP assay result of single embryos. Lanes 1–4 are un.injected control with completely digested PCR product, and lanes 5–8 are embryos injected with TALEN showing undigested products (red box). L ladder

(c) 72 °C, 5 min
(d) 4 °C forever
5. Make working stock for microinjection by mixing and diluting both mRNA encoding the TALEN pair (final concentration ~20 ng/μl of each TALEN mRNA, 20 pg x 2; see Note 15).

6. Microinject (20–100 pg each TALEN arm; see Note 16) into the yolk of one-cell embryos.

### 3.6 Examine Somatic TALEN Activity by RFLP Assay or PCR to Detect a Large Deletion

1. Extract genomic DNA from control (uninjected) and TALEN-injected embryos (see Note 17) as described in Subheading 3.2.

2. PCR amplify the target locus.

3. Digest 10 μL PCR product with appropriate restriction enzyme, and resolve digested product on 1.5 % agarose gel (Fig. 4b).

4. To detect a large deletion generated by two TALEN pairs, extract genomic DNA from control (uninjected) and TALEN-injected embryos (see Note 18) as described in Subheading 3.2.

5. PCR amplifies the target locus with appropriate primers (see Note 18, Fig. 2), and resolve PCR product on agarose gel.

### 3.7 Screening of Germline Transmission for Stable Mutants

For loss-of-function mutagenesis using a single TALEN pair, germline transmission efficiency correlated with TALEN mutagenic activity. Usually founder fish will be identified within screening of ten injected fishes when working with a moderately active TALEN (~60% mutagenic activity in RFLP assay). In large deletion with two TALEN pairs, efficiency is typically two- to fivefold lower, also depending on the activity of TALEN pairs. In the case of site-directed mutagenesis through HDR, efficiency will be ~100-fold lower, and a much larger number of injected fish will have to be screened.

1. Raise potential batches of injected embryos (siblings showing expected somatic mutation).

2. Genotype juvenile fishes (around 4–6 weeks old) by tail fin biopsy (see Note 19).

3. Extract genomic DNA from fin tissue following Subheading 3.2, and screen with RFLP or PCR assay for maintenance of induced as described in Subheading 3.6.

4. Raise juveniles with stable somatic mutations to sexually mature and outcross with wild type to obtain F1 embryos.

5. Extract genomic DNA from individual F1 embryos following Subheading 3.2, and genotype with RFLP or PCR assay.

6. Raise potential batches of F1 embryos (siblings showing heterozygous mutation).

7. Genotype juvenile F1 as described in steps 2 and 3.

8. Confirm mutation carried in F1 by Sanger sequencing (see Note 20).
4 Notes

1. Any PCR reagents could be used and ready-to-use PCR master mix will be efficient in high-throughput screening.

2. BsmBI and Esp3I are isoschizomers that have different optimum reaction temperature (55 °C and 37 °C, respectively). While it is not recommended to use in cycling reactions with T4 DNA ligase, optional predigestion with BsmBI at 55 °C will significantly enhance the efficiency of TALEN assembly, reducing the number of blue colonies.

3. For TALENs with 15 RVDs, Sanger sequencing with TAL_F1 and TAL_R2 will typically cover all 14.5 repeat units. In case units are unread in sequencing, RVD-MM-F and RVD-MM-R primers, with sequences specific to RVD-8, can be used.

4. Mojo Hand is available as a web service at www.talendesign.org. The site allows access to the program without the trouble of installation and with the ease of a familiar interface. Point-of-use help is available for each field. The source code and spreadsheet are also available for noncommercial use with applicable license.

5. For loss-of-function mutations, TALENs should be designed against early conserved exons after the start codon (and alternate start codon) or important functional domain(s) such that small indels will be introduced through NHEJ and resulted in frameshifting/premature termination. For deletion of a large genomic fragment with two pairs of TALENs, simply design two pairs of TALENs flanking the genomic fragment to be deleted. For site-directed mutagenesis through HDR, TALENs should target the site to be mutated.

6. Templates for each system can be changed to user specifications. Notation for templates has been slightly changed from “.” representing a non-preferential base to “N” representing any base. The default template for TALENs remains TsN*e, which constrains TAL-binding sites to an initial 5′ T bp.

7. For deletion of a large genomic fragment with two pairs of TALENs, unique restriction site in spacer for RFLP assay is not necessary since deletion can be simply detected by PCR (see Note 5). However, inclusion of restriction site in the design of both TALEN pairs is recommended such that the activity of each TALEN pair can be confirmed with RFLP assay before co-injection.

8. Microhomology-mediated end joining (MMEJ) is a Ku- and ligase IV-independent DNA repair mechanism that utilizes regions of microhomology adjacent to the site of DSB.
Because in-frame deletions can sometimes lower the efficiency of loss-of-function mutagenesis, we integrated an algorithm developed by Bae et al. [25] into Mojo Hand that calculates a microhomology score and an out-of-frame score for each binding site. The microhomology score is an aggregate of each pattern score associated with each microhomology between two and eight bases long, and the pattern score is calculated based on the length of the microhomology and deletion. Higher microhomology scores correspond with binding sites with stronger microhomologies. Out-of-frame score is the percentage of microhomology score from frameshifting microhomologies for each binding site. Predicted deletions give a list of all homologies within a binding site, with their sequences, deletion lengths, and pattern scores, and whether or not they cause frameshifts. Higher pattern scores correlate with a higher chance of any particular deletion occurring due to microhomology-mediated end joining. This prediction does not take into account deletions that occur due to NHEJ.

9. Integrated Genomics Viewer (IGV) is a tool that allows users to visualize their own genomic datasets and load tracks and other features in a variety of formats. We utilized the BED file format to store user designs for site-specific nucleases, which can then be loaded as a searchable feature within the track line of IGV. This allows users to visualize potential TALEN candidates in tandem with their own in-house next-generation sequencing datasets in an efficient and intuitive manner. BLAT search maps each potential binding site across the genome, which allows users to visualize and avoid designs that are not unique. In addition this function can be used to avoid designs that bind within polymorphic stretches of the genome that may negatively impact cutting efficiency. BED files are created by using the BLAT tool [27] to map binding sites and restriction enzymes to a genome specified by the user. Current genomes supported by Mojo Hand include D. rerio and C. elegans due to current hosting limitations. A detailed specification of BED file format is available at http://genome.ucsc.edu/FAQ/FAQformat.html#format1.

10. Although there is no restriction on primer design for initial genotyping purposes, primer pair can be designed such that they could also be used for RFLP assay. Typically, primers with amplicon size around 300–500 base pairs work well for RFLP assay. Avoid having the unique restriction site for RFLP assay in the middle of the amplicon, which, otherwise, would give two similar-sized digestion products difficult to be resolved in electrophoresis.
11. To identify potential polymorphic region, genomic DNA can be extracted different batches of non-sibling embryos.

12. Assembling reaction works well even if component vectors varied in amounts within range. Equal volume of each vector could be mixed to simplify reaction setup even if their concentrations are different.

13. PCR cycle can be further optimized based on the PCR reagent used.

14. An initial 10 μL half in vitro transcription reaction resuspended in 25 μL final volume will typically yield mRNA at concentration around 500–1000 ng/μL, which is more than enough in most applications.

15. Working mRNA solution should be stored in small aliquots and avoid repeated freeze-thaw.

16. It is recommended to conduct dose–response trials within the range from 20 to 100 pg per TALEN arm such that the optimum dose can be chosen which resulted in survival of around 50% of normally developed embryos.

17. Genomic DNA could be extracted from single embryo to examine mutagenic activity in individual embryo or from a group of five or ten embryos to assay the average mutagenic activity of the TALEN.

18. For screening large genomic deletion, forward primer used to genotype TALEN pair 1 and reverse primer used TALEN pair 2 can be used together to screen for a large deletion resulting in a smaller-sized PCR product compared with the larger or absent PCR product in control. Reverse primer from pair 1 and forward primer from pair 2 can also be used together to screen for very rare “flipping” events where the targeted genomic fragment was excised but inversely inserted back into the genomic lesion. Since the PCR screening is only qualitative and does not reflect mutagenic activity, genomic DNA can be extracted from a single embryo instead of a group of embryos.

19. This round of fin biopsy is optional. However, prescreening for stable somatic mutation can significantly increase the percentage of founder in the pool. Therefore, it is recommended in examples of large fragment deletion and site-directed mutagenesis, where germline transmission efficiency is considerably lower.

20. F1 carrying desirable mutation will be selected. For example, small indels resulted in frameshifting or premature stop in case of loss-of-function mutagenesis and precisely incorporated donor sequence in site-directed mutagenesis.
Acknowledgments

We thank all of the great user comments for refining this genome engineering design system. This project is supported by the State of Minnesota (University of Minnesota/Mayo Clinic Gene Targeting Partnership grant H001274506-3 to SCE), the National Institutes of Health grants GM63904 and P30DK084567 to SCE, the Mayo Foundation, the Health and Medical Research Fund (HMRF02132326) to ACM, and the HKU Seed Funding for Basic Research (201401159004, 201411159098) to ACM.

References

Zebrafish
Methods and Protocols
Kawakami, K.; Pattons, E.E.; Orger, M. (Eds.)
2016, XII, 373 p. 79 illus., 64 illus. in color., Hardcover
ISBN: 978-1-4939-3769-1
A product of Humana Press