Mol. Cells 2016; 39(9): 687-691
http://dx.doi.org/10.14348/molcells.2016.0140

Molecules and Cells
http://molcells.org
Established in 1990

A One-Step System for Convenient and Flexible Assembly of Transcription Activator-Like Effector Nucleases (TALENs)

Jinlong Zhao, Wenye Sun, Jing Liang, Jing Jiang, and Zhao Wu*

Transcription activator-like effector nucleases (TALENs) are powerful tools for targeted genome editing in diverse cell types and organisms. However, the highly identical TALE repeat sequences make it challenging to assemble TALEs using conventional cloning approaches, and multiple repeats in one plasmid are easily catalyzed for homologous recombination in bacteria. Although the methods for TALE assembly are constantly improving, these methods are not convenient because of laborious assembly steps or large module libraries, limiting their broad utility. To overcome the barrier of multiple assembly steps, we report a one-step system for the convenient and flexible assembly of a 180 TALE module library. This study is the first demonstration to ligate 9 mono-/dimer modules and one circular TALEN backbone vector in a one step process, generating 9.5 to 18.5 repeat sequences with an overall assembly rate higher than 50%. This system makes TALEN assembly much simpler than the conventional cloning of two DNA fragments because this strategy combines digestion and ligation into one step using circular vectors and different modules to avoid gel extraction. Therefore, this system provides a convenient tool for the application of TALEN-mediated genome editing in scientific studies and clinical trials.

INTRODUCTION

Transcription activator-like effector nuclease (TALEN)-mediated genome editing is highly specific due to the tightly stringent protein-nucleotide recognition principle and the obligatory heterodimers of the *Fok*I domain, which makes it a powerful tool for targeted genome editing in diverse cell types and organisms. TALENs comprise a TALE DNA-binding region and a *Fok*I restriction endonuclease region. DNA binding to TALE proteins is mediated by a central array of TALE repeat units, and each unit comprises a 33 to 35 amino acid motif specifying one DNA base pair. The repeats are highly identical except for the repeat variable di-residues (RVDs) at positions 12 and 13, which recognize the targeted DNA sequence. The RVDs, NI, NN, NG, and HD were used to specifically recognize nucleotides A, G, T, and C, respectively (Boch et al., 2009; Moscou and Bogdanove, 2009). However, the highly identical TALE repeat sequences make it challenging to assemble TALEs using conventional cloning approaches, and multiple repeats in one plasmid are easily catalyzed for homologous recombination in bacteria. The one-step TALEN assembly system is based on the principle of Golden Gate cloning (Engler et al., 2008; 2009). Type IIS enzymes, such as *Bsa*I, cleave at sites beyond the recognition site. We designed *Bsa*I recognition sites at the 5′ and 3′ ends of every TALE module DNA sequence in the inverse orientation, and these sequences will be removed during the enzyme digestion reaction, leaving sticky overhangs at both sites. Adjacent TALE modules sharing compatible sticky overhangs, termed fusion sites, can be ligated. Thus, this system simplifies the assembly of TALENs.

From Innovative Cellular Therapeutics Co., Ltd, (formerly SiDanSai Biotechnology Co., Ltd), Shanghai 201203, China
*Correspondence: wuzhao@sidansai.com

Received 7 June, 2016; revised 5 August, 2016; accepted 8 August, 2016; published online 7 September, 2016

Keywords: genome editing, molecular genetics, transcription activator-like effector nuclease (TALEN)

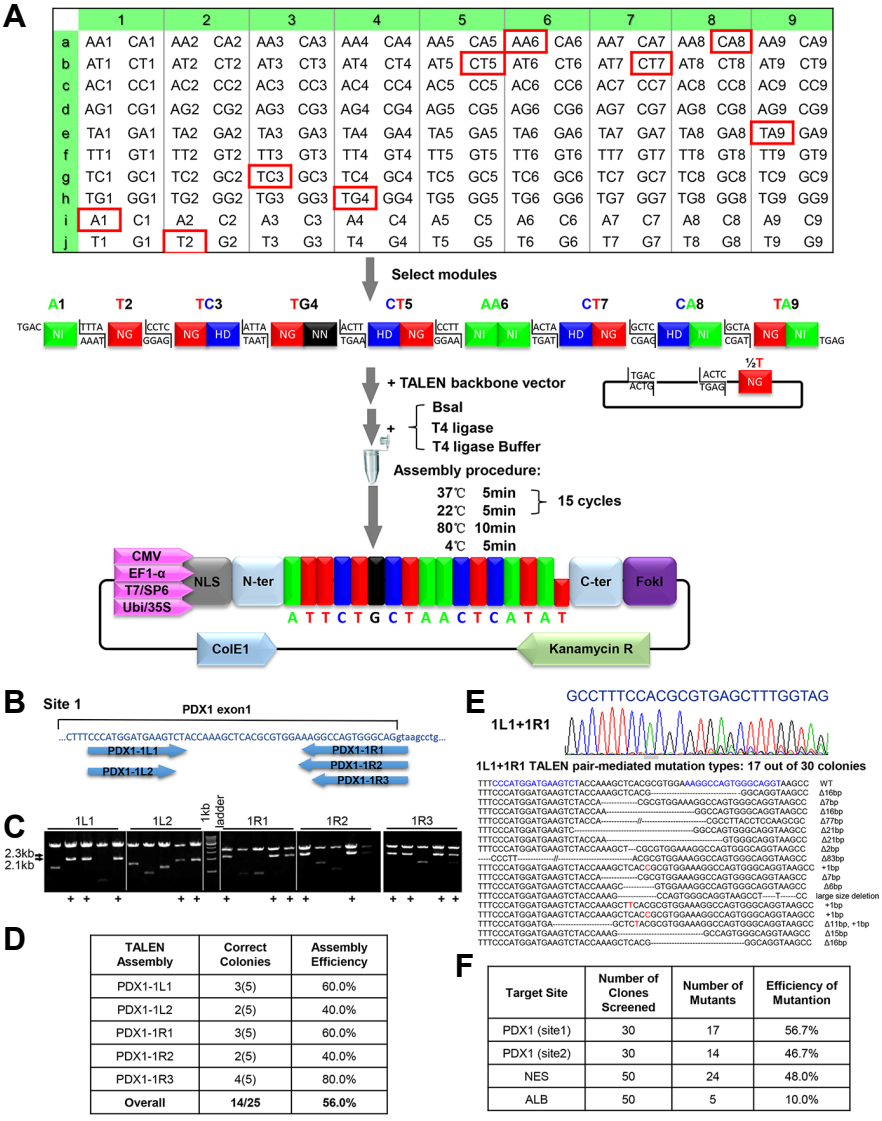
MATERIALS AND METHODS

TALEN design and backbone vector selection

Online tools, such as the TAL Effector Nucleotide Targeter 2.0 (https://tale-nt.cac.cornell.edu/), can be used for TALEN target design, followed by NCBI BLAST to eliminate off-target risks on the specific genome. Binding sites for either the left or right arm range from 12 to 19 bp, with a spacer 12 to 21 bp sequence in the middle, which facilitates *Fok*I dimerization and cleavage. The nucleotide at 5-terminal '0' position of the targeting sequence should be a T, and the last nucleotide is flexible. For knock-outs based on frame-shift induced non-homologous end-joining (NHEJ), a targeting site designed in the first coding exon is recommended. Target sequence verification to eliminate polymorphisms is highly recommended. In order to obtain one TALEN pair with high efficiency, we recommend designing TALEN pair combination 2 × 3 (or 2 × 2) at one site for one gene.

Once binding sites are selected, appropriate TALEN backbone vectors can be selected. An array of TALEN backbone vectors (left arm and right arm) are designed for the TALEN assembly for gene modification in various cell lines and species, containing EF1α, CMV, T7/SP6 or Ubiquitin/35S promoter, respectively. All vectors contain basic structures for TALEN

eISSN: 0219-1032



D

TALEN Assembly	Correct Colonies	Assembly Efficiency
PDX1-1L1	3(5)	60.0%
PDX1-1L2	2(5)	40.0%
PDX1-1R1	3(5)	60.0%
PDX1-1R2	2(5)	40.0%
PDX1-1R3	4(5)	80.0%
Overall	**14/25**	**56.0%**

F

Target Site	Number of Clones Screened	Number of Mutants	Efficiency of Mutantion
PDX1 (site1)	30	17	56.7%
PDX1 (site2)	30	14	46.7%
NES	50	24	48.0%
ALB	50	5	10.0%

Fig. 1. A convenient one-step system for TALEN assembly. (A) A total of 180 ligation modules were grouped into 9 positions (16 dimers at 1-9 positions; 4 monomers at 1-9 positions). An array of TALE-repeats was assembled by the selection of 9 modules from positions 1-9 and ligated into a TALEN expression vector (containing half a monomer module recognizing A, T, C or G) in one step. Nuclear localization signal (NLS), N-Terminal and C-Terminal of TALE, *Fok*I, Kana^R, Puro^R, IRES, and promoter selections are included in the TALEN expression vector. All modules and linearized vectors are flanked by two sticky overhangs, generated by *Bsa*I digestion, that facilitate their orderly ligation. (B) A total of 2 × 3 TALEN pairs targeting the first exon of the human *PDX1* gene were designed and assembled. The left TALEN recognition sequences were PDX1-1L1: 5′-cccatggatgaagtct-3′ and PDX1-1L2: 5′-cccatggatgaagt-3′; while the right recognition sequences were PDX1-1R1: 5′-acctgcccactggcctt-3′, PDX1-1R2: 5′-acctgcccactggccttt -3′ and PDX1-1R3: 5′-acctgcccactggcct-3′. (C) The assembly rate was assessed by enzyme digestion. A total of 5 colonies were analyzed for each TALEN vector, and a 1 kb DNA marker was used to determine the correct digestion bands. The positive assembly clones are indicated at the bottom by a "+" symbol. (D) An overall 56.0% assembly rate was noted for the four TALEN vectors from the 25 colonies. (E) Upper panel, genomic sequencing results of the cells after subjection to the TALEN pair PDX1-1L1+PDX1-1R1. A significant multiple-peak pattern indicates their robust cleavage activity. Lower panel, indel sequences observed in 30 cloned amplicons from 293T cells expressing PDX1-1L1 and PDX1-R1. The blue letters indicate the recognition sequences of each pair on the plus strand. The dashes denote deleted nucleotides, while the red letters indicate inserted nucleotides, and a double slash indicates a large size deletion. (F) The targeting efficiencies of TALENs at 4 loci of 3 indicated genes in human cells.

expression, including 3× Flag, NLS, N-terminal, C-terminal, *Fok*I coding sequence and Kanamycin resistant gene. *Fok*I coding sequence on left and right arms are heterologous, and the protein products will undergo dimerization to function as an active endonuclease. A nucleotide specific half monomer is predesigned onto each backbone vector, which is 1/2A, 1/2T, 1/2C, and 1/2G at the C-terminal end, respectively, covering all potential TALEN-binding sites. Most left arm vectors (except plant vectors) contain IRES and puromycin resistance gene, however puromycin-free left arm vectors are more suitable for TALEN vectors *in vitro* transcription. Some right arm TALEN vector contains EGFP which allows the observation of successful transfection.

Cell culture and transfection

HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) supplemented with 10% fetal bovine serum (FBS, Gibco). TALEN vector pairs and the EGFP-IRES-PURO vector were transfected into 293T cells using FuGENE HD transfection reagent (Promega). Fluorescent signals were observed, and 1 μg/ml puromycin was applied for successful transfection selection.

TALEN efficiency detection

TALENs were assembled using the FastTALE TALEN Assembly kit, please refer to Supplemental Experimental Procedures for more detailed description of these procedures. In brief, divide the TALEN recognizing sequence into 9 pieces, excluding the last 3′ nucleotide. Each of the 9 pieces was assigned with a number starting from 1 to 9 with a 5′ to 3′directionality. A master mix of 9 modules, Left/Right arm TALEN backbone vector, *Bsa*I with T4 DNA ligase was prepared, followed with a ligation pro-

gram in a thermal cycler. Upon completion of the ligation, Plasmid-Safe™ ATP-Dependent DNase was added to selectively remove residual linear DNA fragments. The vectors were transformed into competent cells, colonies were screened and confirmed using enzyme digestion and DNA sequence analysis.

At three days post transfection, the cells were harvested for genome isolation. Raw efficiency was read from multiple peaks of the PCR sequencing of mixed cells; accurate efficiency was calculated after sequencing single clonal-competent cells, followed by ligation of the PCR products into the T-vector and transformation.

RESULTS

Encouraged by the successful reprogramming of fibroblast to induced pluripotent stem cells (iPS cells) by the simultaneous introduction of 24 factors (Takahashi and Yamanaka, 2006), we attempted to put all the modules and one TALEN backbone vector together, to simplify TALEN assembly into a one-step process. We assembled 18 TALE repeats by ligating 9 dimers using Golden Gate cloning, which has never been reported previously.

This one-step assembly method employs a library of a total 180 purified DNA fragments, or TALE modules, in either monomer or dimer forms (Fig. 1A). The 180 TALE modules are divided into 9 positions including 16 dimers and 4 monomers at each position. Each position is determined based on its specific sticky ends, ensuring that all modules are properly ligated (Fig. 1A). Divide the TALEN recognizing sequence into 9 pieces, excluding the last 3′ nucleotide, which will be encoded by the expression vector. Each of the 9 pieces was assigned with a number starting from 1 to 9 with a 5′ to 3′directionality. For example, to target sequence ATTCTGCTAACTCATAT, after removing the last 'T' the rest are numbered as A1, T2, TC3, TG4, CT5, AA6, CT7, CA8, and TA9. The CT5 module comprises an HD-containing TALE repeat and an NG-containing TALE repeat that recognizes a CT with unique 5′ and 3′ ends located at the 5th position. An array of TALE monomers/dimers can be assembled by selecting 9 modules from the 1st to the 9th position, followed by ligation into the circular TALEN backbone vectors to target a specific genomic locus. Each circular TALEN backbone vector contains a heterodimeric *Fok*I domain and can be selected to express a TALEN protein in specific cell lines and species, based on differences between the promoter and selection marker (Fig. 1A). To investigate the efficiency of this TALEN assembly, five TALENs forming six TALEN pairs targeting one locus of the human gene *PDX1* were designed and constructed, followed by analysis of the assembly rates (Fig. 1B).

After assembly using this system, the ligation products were transformed into *E. coli*, and five colonies were picked for each TALEN. The assembly rate was assessed by enzyme digestion (Fig. 1C) and confirmed using DNA sequencing. The rate of TALEN assembly ranged from 40% to 80%, with an average of 56.0% (Fig. 1D). Every TALEN pair was transfected into HEK293T cells, and the cells were harvested for genomic DNA preparation at 5 days post-transfection. The amplified PCR products of the targeted loci were sequenced to detect mutations. A multiple-peak sequencing pattern indicates a targeted mutation (Fig. 1E). PCR products containing mutations were ligated into T vectors, and thirty single colonies were sequenced. Various types of mutations in the target genomic region were observed. One selected pair (PDX1-1L1+PDX1-1R1) targeting *PDX1* led to 17/30 mutation types (Fig. 1E). The same analysis was performed for other loci in a second locus of *PDX1* (Fig. 2A), human *NESTIN* (*NES)* (Fig. 2B), and *ALBUMIN* (*ALB)* (Fig. 2C). The overall TALEN cleavage efficiency of all four validated pairs ranged from 10.0% to 56.7% (Fig. 1F).

DISCUSSION

This system largely simplifies the assembly of TALENs by shortening the manual operation time of TALEN assembly into less than half an hour, followed by verification within three days. This strategy has three major advantages: (A) Ease. This one-step assembly system requires only 9 modules and one circular TALEN backbone vector to complete assembly, avoiding the traditional steps of digestion, gel extraction and ligation. (B) Speed. One-step assembly facilitates the assembly of functional TALEN vectors within one day, followed by verification steps, such as enzyme digestion or DNA sequencing within three days. (C) Flexibility. High flexibility in selecting a TALE target sequence is endowed by the combinatorial use of dimers and monomers from the library. Considering a half monomer module, which recognizes a single nucleotide A, T, C or G on the TALEN backbone vector, this system facilitates the assembly of TALE repeats from a minimum of 9.5 to a maximum of 18.5 nucleotides in one-step cloning.

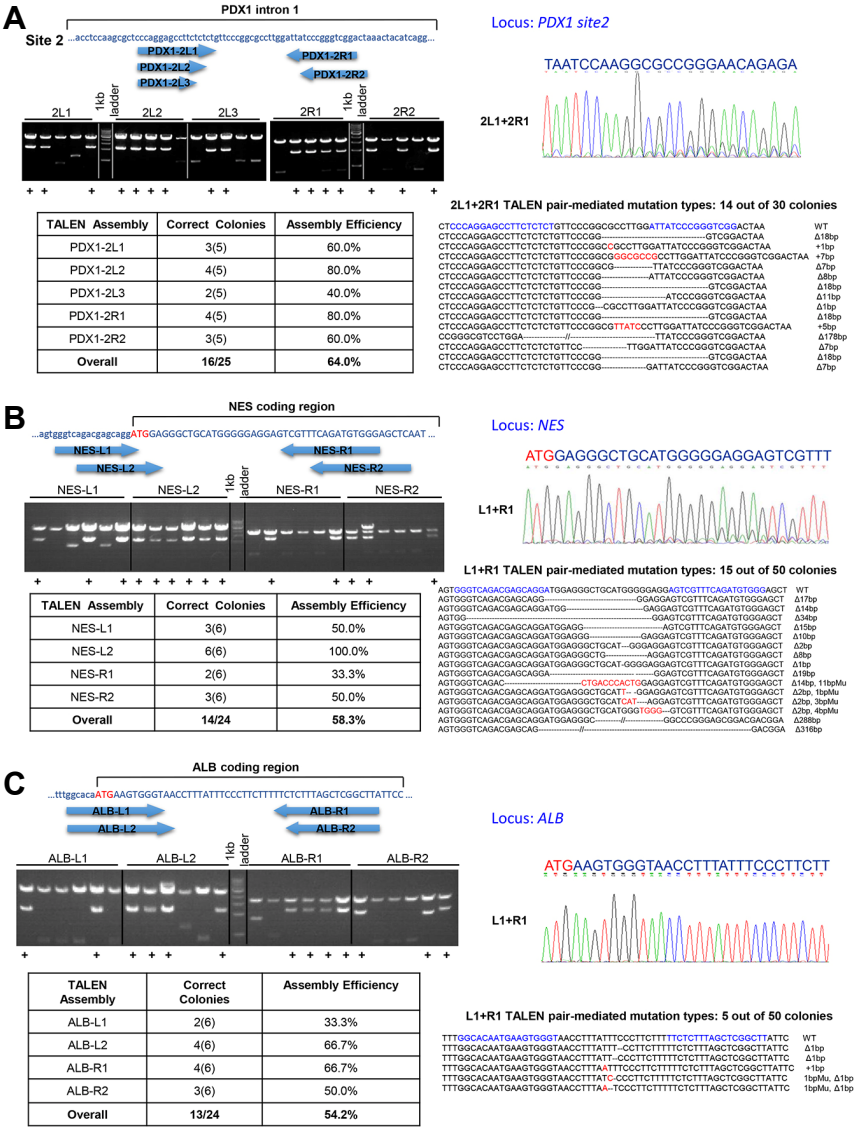
Because TALE was combined with *Fok*I to function as genome scissors, the methods for TALEN assembly have been continuously improved based on the Golden Gate cloning method (Cermak et al., 2011; Ding et al., 2013; Kim et al., 2013; Li et al., 2011; 2012; Morbitzer et al., 2011; Weber et al., 2011; Zhang et al., 2011; 2013). Several one-step assembly methods based on trimer or tetramer module libraries have been reported (Ding et al., 2013; Kim et al., 2013). To confirm the potential and flexibility of the TALEN design, the sizes of the module libraries are inevitably large, containing 424 and 832 modules, respectively. This results in operating complexity in selecting modules and maintaining libraries of such sizes. The two methods based on trimer or tetramer module libraries can be used only to build TALENs with a fixed number of recognition repeats, markedly diminishing the flexibility of selecting recognition sequences.

Other methods based on monomer or dimer recognition modules have required tedious serial operation steps and/or troublesome multiple enzyme utilization, increasing operational inconvenience. Other methods can be used only to build TALENs with a fixed number of recognition repeats, e.g., for 13 bp or 17 bp nucleotides (Li et al., 2011; 2012).

To our knowledge, this study is the first to ligate 9 modules and one circular TALEN backbone vector in one step, generating 9.5 to 18.5 repeat sequences with an overall assembly rate higher than 50%. Compared with the above methods, this system based on a monomer/dimer library at 9 positions ensures a smaller library size, renders higher designing flexibility and maintains high targeting specificity.

Although we focused on TALENs assembly in the present study, this TALE assembly system can also be used to fuse the engineered TALE repeat arrays with other functional proteins, including catalytic hydrolase, to generate target chimeric proteins.

The CRISPR/Cas9 system was developed as a genome editing tool in 2013. As the construction of CRISPR/Cas9 guiding RNA is simple and former TALEN assembly is complicated, the CRISPR/Cas9 genome editing tool has been widely used despite its severe off-target effect due to a 10-12 bp core recognition sequence. The off-target issues are still a concern in clinical



TALEN Assembly	Correct Colonies	Assembly Efficiency
PDX1-2L1	3(5)	60.0%
PDX1-2L2	4(5)	80.0%
PDX1-2L3	2(5)	40.0%
PDX1-2R1	4(5)	80.0%
PDX1-2R2	3(5)	60.0%
Overall	**16/25**	**64.0%**

TALEN Assembly	Correct Colonies	Assembly Efficiency
NES-L1	3(6)	50.0%
NES-L2	6(6)	100.0%
NES-R1	2(6)	33.3%
NES-R2	3(6)	50.0%
Overall	**14/24**	**58.3%**

TALEN Assembly	Correct Colonies	Assembly Efficiency
ALB-L1	2(6)	33.3%
ALB-L2	4(6)	66.7%
ALB-R1	4(6)	66.7%
ALB-R2	3(6)	50.0%
Overall	**13/24**	**54.2%**

Fig. 2. Targeting of the endogenous loci of *PDX1, NESTIN* and *ALBUMIN* in human cells using TALEN pairs. (A) Left upper panel, 3 × 2 TALEN pairs targeting the first intron of the human *PDX1* gene were designed and assembled. The left TALEN recognition sequences were PDX1-2L1: 5′-cccaggagccttctctct-3′, PDX1-2L2: 5′-cccaggagccttctct-3′ and PDX1-2L3: 5′-cccaggagccttct-3′; the right recognition sequences were PDX1-2R1: 5′-ccgacccgggataat-3′, PDX1-2R2: 5′-agtccgacccgggat-3′. Left middle panel, the assembly rate was assessed using enzyme digestion. Five colonies were analyzed for each TALEN vector, and a 1 kb DNA marker was used to determine the correct digestion bands. The positive assembly clones are indicated at the bottom by a "+" symbol. Left lower panel, an overall 64.0% assembly rate was noted for the four TALEN vectors from the 25 colonies. *Right upper panel,* genomic sequencing results for the cells transfected with the indicated TALEN pairs. A significant multiple-peak pattern indicates their robust cleavage activity. *Right lower panel*, indel sequences observed in 30 cloned amplicons from 293T cells expressing PDX1-2L1 and PDX1-2R1. The blue letters indicate the recognition sequences of each pair on the plus strand. The dashes denote deleted nucleotides, while the red letters indicate inserted nucleotides, and a double slash indicates a large size deletion. (B) Left upper panel, 2 × 2 TALEN pairs targeting the first exon of the human *NESTIN* gene were designed and assembled. The left TALEN recognition sequences were NES-L1: 5′-gggtcagacgagcagga-3′ and NES-L2: 5′-cagacgagcaggatgga-3′; the right recognition sequences were NES-R1: 5′-cccacatctgaaacgac-3′ and NES-R2: 5′-gagctcccacatctgaa-3′. Left middle panel, the assembly rate was assessed by enzyme digestion. Six colonies were analyzed for each TALEN vector. Correct assembly clones were indicated at the bottom by a "+" symbol. Left lower panel, an overall 58.3% assembly rate was noted for the four TALEN vectors from the 24 colonies. *Right upper panel*, genomic sequencing results for the cells after subjection to L1+R1 pairing shows a significant multiple-peak pattern, indicating its high cleavage activity. *Right lower panel*, indel sequences observed in 50 cloned amplicons from 293T cells expressing NES-L1 and NES-R1. Out of 50 colonies, 15 types (of 24 colonies) of mutation were identified. (C) Left upper panel, 2 × 2 TALEN pairs flanking the start codon site of human *ALBUMIN* gene are designed and assembled. The left TALEN recognition sequences were ALB-L1: 5′-ggcacaatgaagtgggt-3′ and ALB-L2: 5′-ggcacaatgaagtgggta-3; the right recognition sequences were ALB-R1: 5′-aagccgagctaaagagaa-3′, ALB-R2: 5'-aagccgagctaaagag-3'. Left middle panel, the assembly rate was assessed using enzyme digestion. Six colonies were analyzed for each TALEN vector. The positive assembly clones are indicated at the bottom by a "+" symbol. Left lower panel, an overall 54.2% assembly rate was noted for the four TALEN vectors from the 24 colonies. Right upper panel, genomic sequencing results for the cells transfected with the TALEN pair L1+R1. Right lower panel, indel sequences observed in 50 cloned amplicons from 293T cells expressing ALB-L1 and ALB-R1.

application. Recently, many efforts are made to improve CRISPR/Cas9 target specificity and to decrease off-target effect. Here, we developed an assembly system that enables easier TALEN assembly. This assembly system makes TALEN-mediated genome editing a more convenient and promising tool which is comparable to CRISPR/Cas9 genome editing.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS

This work was supported by the Shanghai "Post-Qi-Ming-Xing Plan" for Young Scientists, China (Grant No. 15QB1403700),

and the Shanghai Committee of Science and Technology, China (Grant No. 14JC1405700).

REFERENCES

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.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and 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.

Ding, Q., Lee, Y.K., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T., et al. (2013). A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell *12*, 238-251.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One *3*, e3647.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One *4*, e5553.

Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E., et al. (2013). A library of TAL effector nucleases spanning the human genome. Nat. Biotechnol. *31*, 251-258.

Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P., and Yang, B. (2011). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res. *39*, 6315-6325.

Li, L., Piatek, M.J., Atef, A., Piatek, A., Wibowo, A., Fang, X., Sabir, J.S., Zhu, J.K., and Mahfouz, M.M. (2012). Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification. Plant Mol. Biol. *78*, 407-416.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science *326*, 1501.

Morbitzer, R., Elsaesser, J., Hausner, J., and Lahaye, T. (2011). Assembly of custom TALE-type DNA binding domains by modular cloning. Nucleic Acids Res. *39*, 5790-5799.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Weber, E., Gruetzner, R., Werner, S., Engler, C. and Marillonnet, S. (2011). Assembly of designer TAL effectors by Golden Gate cloning. PLoS One *6*, e19722.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat. Biotechnol. *29*, 149-153.

Zhang, Z., Li, D., Xu, H., Xin, Y., Zhang, T., Ma, L., Wang, X., and Chen, Z. (2013). A simple and efficient method for assembling TALE protein based on plasmid library. PLoS One *8*, e66459.