

# Creating a monomeric endonuclease TALE-I-SceI with high specificity and low genotoxicity in human cells

Jianfei Lin<sup>1,2,3</sup>, He Chen<sup>1</sup>, Ling Luo<sup>4</sup>, Yongrong Lai<sup>4</sup>, Wei Xie<sup>5</sup> and Kehkooi Kee<sup>2,3,\*</sup>

<sup>1</sup>College of Life Sciences, Peking University, Beijing 100871, China, <sup>2</sup>Center for Stem Cell Biology and Regenerative Medicine, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing 100084, China, <sup>3</sup>Tsinghua-Peking-NIBS Center for Life Sciences, School of Medicine, Tsinghua University, Beijing 100084, China, <sup>4</sup>Department of Hematology, the First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, China and <sup>5</sup>THU-PKU Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China

Received November 07, 2014; Revised December 10, 2014; Accepted December 11, 2014

## ABSTRACT

**To correct a DNA mutation in the human genome for gene therapy, homology-directed repair (HDR) needs to be specific and have the lowest off-target effects to protect the human genome from deleterious mutations. Zinc finger nucleases, transcription activator-like effector nuclease (TALEN) and CRISPR-CAS9 systems have been engineered and used extensively to recognize and modify specific DNA sequences. Although TALEN and CRISPR/CAS9 could induce high levels of HDR in human cells, their genotoxicity was significantly higher. Here, we report the creation of a monomeric endonuclease that can recognize at least 33 bp by fusing the DNA-recognizing domain of TALEN (TALE) to a re-engineered homing endonuclease I-SceI. After sequentially re-engineering I-SceI to recognize 18 bp of the human  $\beta$ -globin sequence, the re-engineered I-SceI induced HDR in human cells. When the re-engineered I-SceI was fused to TALE (TALE-ISVB2), the chimeric endonuclease induced the same HDR rate at the human  $\beta$ -globin gene locus as that induced by TALEN, but significantly reduced genotoxicity. We further demonstrated that TALE-ISVB2 specifically targeted at the  $\beta$ -globin sequence in human hematopoietic stem cells. Therefore, this monomeric endonuclease has the potential to be used in therapeutic gene targeting in human cells.**

## INTRODUCTION

Gene therapy has been developed to eradicate genetic diseases, such as  $\beta$  thalassemia. However, conventional methods involving random insertions of viral vectors can cause chromosome translocations or mutations in the human genome, as shown in multiple gene therapy trials (1–3). To

eliminate the risk of causing undesirable mutations in the human genome or changing the properties of cells, gene targeting specificity is essential. Sequence-specific endonucleases that can induce homologous repair at targeted sequences have been developed to fulfill this goal. Zinc finger nucleases (ZFNs) are the first generation of sequence-specific endonucleases that were developed for this reason. They are composed of a DNA-recognizing domain fused to a sequence non-specific endonuclease domain derived from FokI (4). Each zinc finger motif can recognize 3 bp of DNA in the major groove of the targeted sequence. By combining the left and right arms, ZFNs can recognize a range of sequence lengths. An alternative to ZFN that has been recently developed is TALEN. The overall design of these endonucleases is similar to that of ZFNs, in which both arms of the recognition motifs have to form a dimeric structure to cut targeted DNA sequences (5). The advantage of TALENs is the ease of designing and modifying the coded amino acids on the DNA-recognition motif. More recently, the CRISPR/CAS9 system, originally discovered in bacteria, is another sequence-specific endonuclease system that is widely used by researchers (6). Using a 20-bp-guided RNA complementary to the targeting sequence, this system relies on CAS9 endonuclease to cut the targeted sequence and induces recombination (7). The ease of designing targeted sequences by modifying the guided RNA sequence instead of re-engineering amino acids, as required in ZFNs or TALEN, provides a more efficient method of gene targeting.

Although all three gene-targeting systems have been successfully used in targeting specific loci in human cells, off-target effects have also been reported, including a recent study that reported the off-target activity induced by CRISPR/CAS9 at the human  $\beta$ -globin gene (8–12). Researchers have improved the specificity of the nucleases by using various approaches (13–15), but it is important to determine if any of these systems have low genotoxicity in human cells. To develop a system that can correct mutations in

\*To whom correspondence should be addressed. Tel: +86 10 62794200; Email: kkee@tsinghua.edu.cn

the  $\beta$ -globin gene to treat  $\beta$ -thalassemia, we tested the gene-targeting efficiency of TALEN and CRISPR-CAS9 systems. Homology-directed repair (HDR) induced by these two systems are efficient, but the genotoxicity in the human cells tested was high. Therefore, we developed an alternative system that may reduce off-target effects. I-SceI, a monomeric homing endonuclease from the LAGLIDADG family, is one of the most specific natural nucleases discovered and has an 18-bp recognition sequence. It has been used to study DNA recombinations in human cells (16) and causes very low genome toxicity (17–19) because of the lack of DNA sequences recognized by I-SceI in the human genome. Here, we re-engineered I-SceI and fused it to TALE, creating a monomeric endonuclease that recognizes 33 bp of the human  $\beta$  globin gene sequence. The system was used to target DNA sequences close to mutation sites that are highly prevalent in  $\beta$ -thalassemia patients. The fusion protein increased the HDR rate to a similar extent as that induced by TALENs and shows significantly lower levels of genome toxicity than those associated with TALEN and CRISPR-CAS9 systems.

## MATERIALS AND METHODS

### Construction of vectors

Cas9D10A and gRNA plasmids were obtained from Addgene (plasmids #41816 and #41824) and the construction of gRNAs, which specifically target the  $\beta$ -globin gene, were generated as described (20). The targeting sequence is GAATAACAGTGATAATTTCTGG, as indicated on Figure 1a. The TALEN expression vector was a gift provided by Prof. Nieng Yan. We constructed the TALEN targeting system at the human  $\beta$ -globin gene as described previously (21), with the exception that the *SpeI*, *NheI* and *HindIII* restriction enzymes were replaced by *BspEI*, *XmaI* and *SalI*. Briefly, a single repeat-variable di-residue (RVD) unit was inserted into the P-easy-T3 vector (Transgene) by using the isocaudomers *BspEI* and *XmaI*, followed by *SalI* digestion and ligation to create a double unit RVD. Next, TALEN RVDs were constructed by assembling 14–17 repeat modules into intermediary arrays and joining the intermediary arrays into pCS2 vector (Addgene plasmid #16331). All TALEN and Cas9D10A constructs were tagged with a Flag epitope at the N-terminus.

eGFP reporters for measuring HDR were constructed by inserting  $\beta$ -globin sequences into the GFP codon sequence at position 329. Fifty-three base pairs were inserted for TALEN and CRISPR/CAS9 assays, and 18 bp was inserted at this eGFP position for I-SceI assays in Figures 1a and 2a. The disrupted eGFP fragments and EF1 $\alpha$  promoter fragment were LR (*attL* $\times$ *attR*) recombined into p2k7 vectors to create a lentiviral vector (22). Lentiviral reporters were used to prepare lentivirus to infect 293FT cells. The donor template for correcting the defective eGFP gene was constructed by cloning the eGFP gene into pENTRY/DTOPO (Invitrogen). To prevent the expression of eGFP from the donor plasmid, the start codon (first 3 bp) was removed.

I-SceI gene was amplified by polymerase chain reaction (PCR) from I-SceI-GR-GFP (Addgene plasmid #17654) (19) and inserted into vector pCS2. We modified I-SceI to recognize the  $\beta$ -globin sequence by changing amino acids

in I-SceI and the recognized sequences as described in the results section. All I-SceI variants were generated by using site-directed mutagenesis and encoded 19 additional amino acids. In short, we constructed 76 variants of I-SceI after four rounds of selection with four different reporter vectors harboring the sequences indicated in Figure 2. For the TALE-I-SceI fusion protein, re-engineered I-SceI (ISVB2) was fused directly to the C-terminus of TALEN repeats recognizing the  $\beta$ -globin sequence.

### Cell culture and fluorescence-activated cell sorting (FACS) analysis of HDR

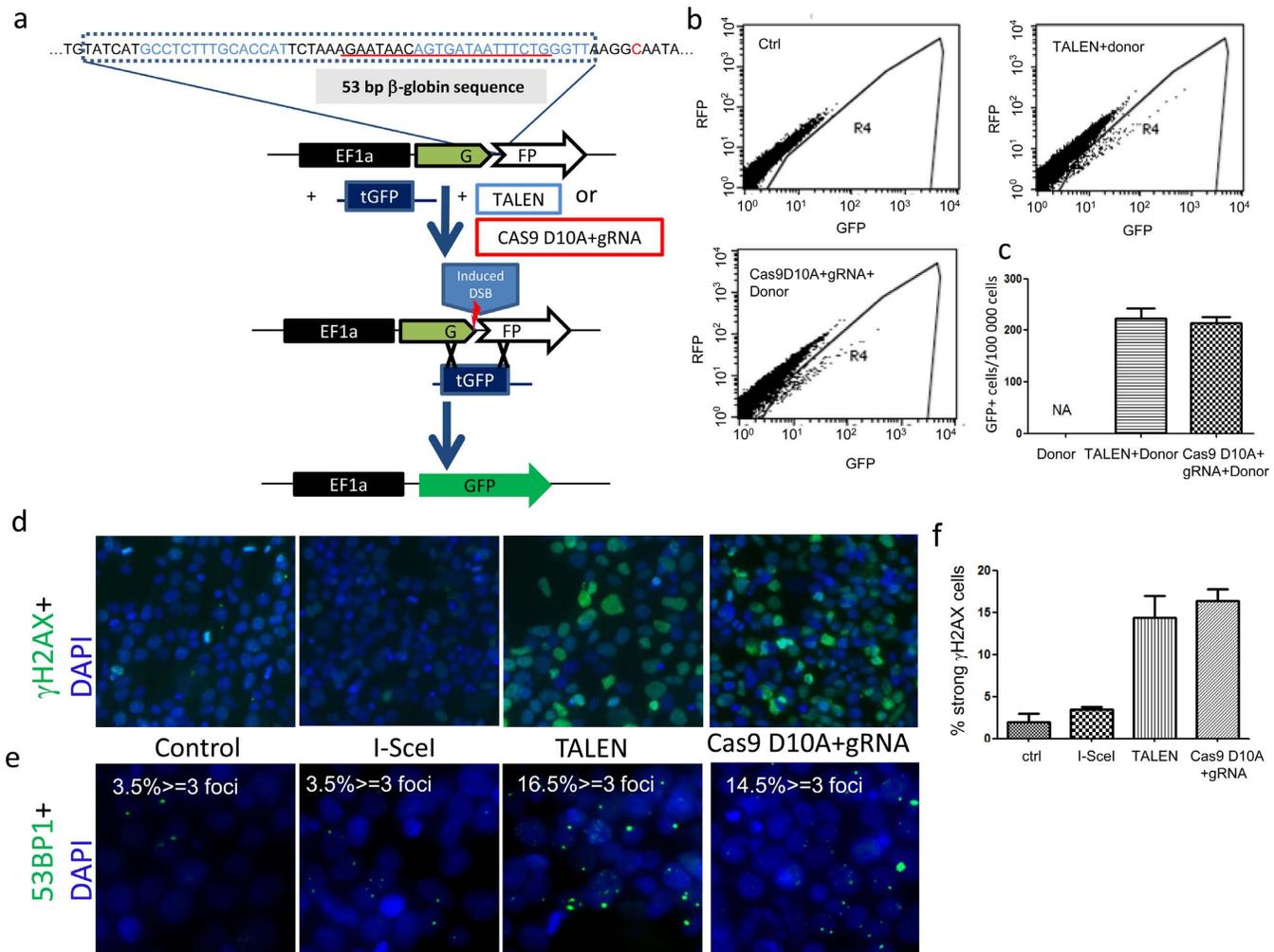
293FT cells were cultured and passaged at 37°C with 5% CO<sub>2</sub>, in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (FBS), 0.1 mmol/l MEM non-essential amino acids, 2 mmol/l L-glutamine, 1% penicillin-streptomycin and 500  $\mu$ g/ml geneticin (Invitrogen). 293FT cells were seeded in 6-well plates before transfection. Cells were transfected using Lipofectamine 2000 (Invitrogen) when confluency was within 80–90%, according to the manufacturer's protocol. For each well, the final concentration of each plasmid transfected was 500 ng/ml. The HDR was measured using FACS (FACSCalibur, BD) analysis of at least 100 000 cells after 36 h of transfection. One cell expressing eGFP was considered to be one successful HDR event.

### Genotoxicity assays

Genotoxicity was examined at the level of  $\gamma$ H2AX and 53BP1 foci. Etoposide was used to induce different levels of these two genotoxicity markers as positive controls for endonuclease-induced genotoxicity. Briefly, cells were grown in 24-well plates and 500 ng/ml vector expressing each nuclease (without donor plasmids) was transfected into cells in each well by using Lipofectamine 2000. After 20 h, cells were fixed with 4% paraformaldehyde for 10 min, followed by permeabilization with 0.2% Triton X-100 for 5 min. Cells were then blocked in phosphate buffered saline containing 10% FBS at 37°C for 1 h, and then incubated at 37°C for 1 h with the following primary antibodies: mouse monoclonal  $\gamma$ H2AX at 1:1000, mouse monoclonal 53BP1 at 1:1000 (Abcam). Then secondary antibodies conjugated to Alexa Fluor 488 (Invitrogen) at 1:1000 were incubated for 1 h; followed by 4',6-diamidino-2-phenylindole (DAPI) staining. The images were captured using a digital inverted microscope (NIKON Ti-E system). Cells stained for 53BP1 were divided into 4 groups with 0, 1, 2 or  $\geq$ 3 53BP1 foci. FACS analysis of cells stained with anti- $\gamma$ H2AX was performed to measure the percentage of cells with levels of  $\gamma$ H2AX above background staining.

### Western blot

The expression of endonucleases in 293FT cells was measured using anti-FLAG antibodies, as all of the nucleases were tagged with the Flag epitope. Briefly,  $2 \times 10^5$  293FT cells were seeded in 6-well plates and transfected with endonuclease vectors at a concentration of 500 ng/ml. Forty-eight hours after transfection, the cells were harvested,



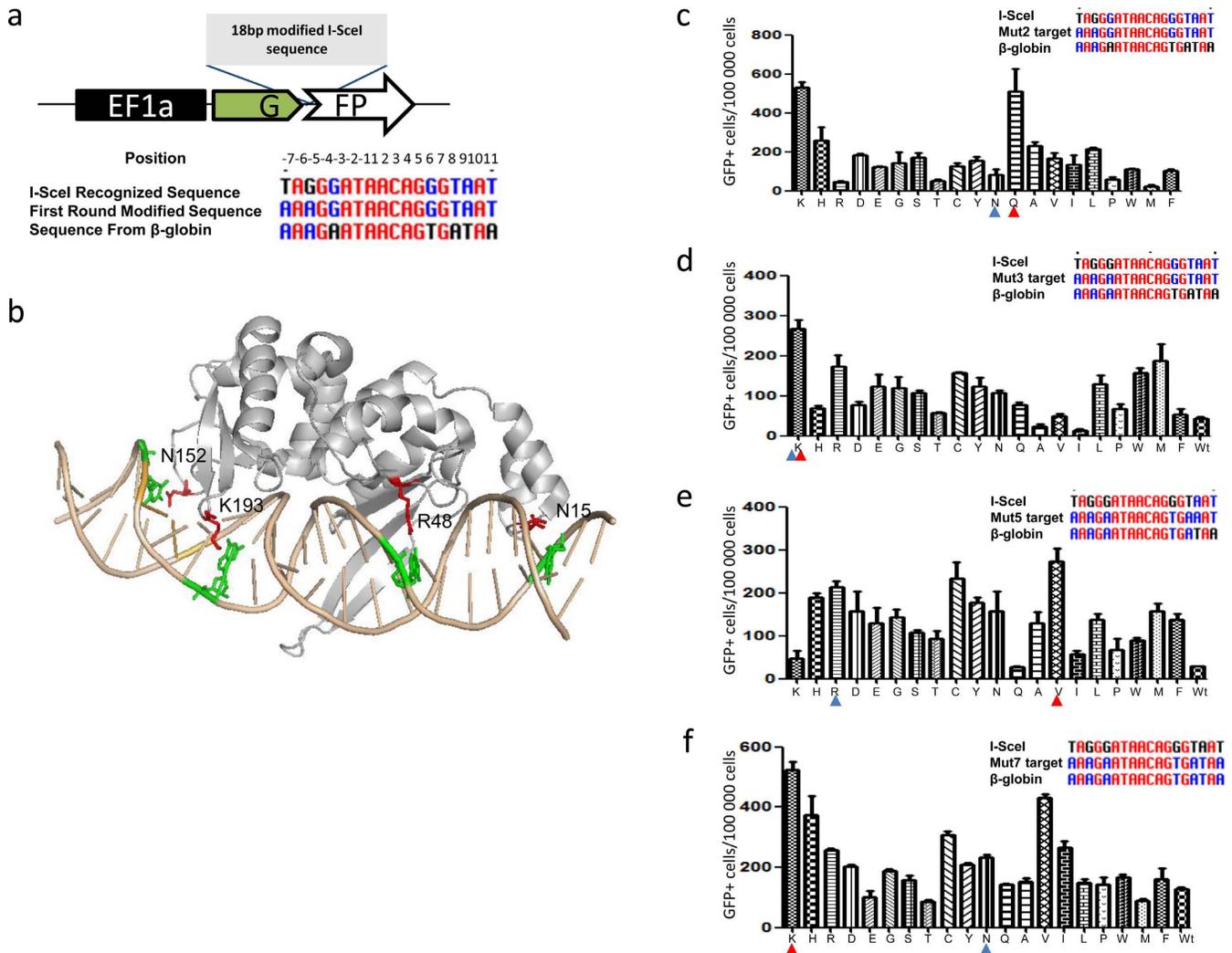
**Figure 1.** TALEN and CRISPR/CAS9D10A induced HDR at  $\beta$ -globin sequence in human cells. (a) Schematic diagram showing how the eGFP reporter system measures HDR rates induced by TALEN or CRISPR/CAS9D10A. eGFP codon region driven by the EF1 $\alpha$  promoter is interrupted by the 53-bp human  $\beta$ -globin sequence (blue dotted box). Partial DNA sequence of the human  $\beta$ -globin gene (chromosome 11) is highlighted with the targeting locations of TALEN (blue). Nucleotide C (highlighted with red) indicates the locus of one of the most frequent mutations of  $\beta$ -thalassaemia in Chinese patients, IVS-II654c-t. This reporter vector was integrated into the genomes of 293FT cells by using lentiviral transduction. The donor sequence containing a truncated eGFP fragment was transfected into cells along with TALEN expression vectors. The HDR events induced at the inserted sites induced eGFP expression in the cells, which was monitored using FACS. (b) FACS results of HDR assays using TALENs or CRISPR/CAS9D10A. The control were cells transfected with the truncated eGFP fragment (Ctrl). (c) The average percentage of HDR induced by TALEN or CRISPR/CAS9D10A was determined from three independent transfection experiments, and the error bars denote the standard deviation. No HDR event was detected in the control group of 100 000 cells (NA). (d and e) TALENs and CRISPR/CAS9D10A induced high levels of genotoxicity in 293FT cells.  $\gamma$ H2AX and 53BP1 immunostaining of 293FT cells induced by I-SceI, TALEN or CRISPR/CAS9D10A. (f) Percentage of cells with high  $\gamma$ H2AX expression determined using FACS for three independent cell populations. The error bars indicate the standard deviation. TALEN and CRISPR/CAS9D10A samples exhibited significantly higher levels of  $\gamma$ H2AX than that of the control. The asterisk indicates the statistical significance between the sample and control ( $P$ -value < 0.05).

rinsed with phosphate-buffered saline and lysed in 150  $\mu$ l of ice-cold RIPA buffer composed of 50 mM Tris, 150 mM NaCl, 0.5% Na deoxycholate, 1% Nonidet P-40 and 0.1% sodium dodecyl sulfate (SDS). Then the cell lysates were centrifuged at 4°C for 5 min at 8000  $\times$  g. The supernatant was subjected to electrophoresis using 8% SDS-polyacrylamide gel electrophoresis and then transferred onto the polyvinylidene difluoride (PVDF) membrane by performing electroblotting. The membrane was blocked in 5% non-fat milk in Tris-Buffered Saline with 0.1% Tween 20 (TBST) blocking solution at room temperature for 1 h and subsequently incubated with FLAG-specific monoclonal antibody diluted 1:1000 (F1084, Sigma) in TBST.

The membrane was then subjected to a 1-h incubation with horseradish peroxidase (HRP)-conjugated goat anti-mouse secondary antibody (Zhongshan Jinqiao) at 1:1000 in TBST, followed by detection using the chemiluminescence labeling detection reagent ECL Plus (GE healthcare).

#### Electrophoretic mobility shift assay (EMSA)

I-SceI and ISVB2 were tagged at the N-terminus with a GST epitope and were then purified using a GST agarose column (GE healthcare). DNA probes with 3'-biotin-labeled were ordered directly from Invitrogen. We followed the protocol developed by Ruff *et al.* (23),



**Figure 2.** Re-engineering I-SceI to target the β-globin sequence in human cells. (a) Schematic diagram of the eGFP reporter with a modified I-SceI recognition sequence. Two nucleotides were modified in the first round of selection as shown in the listed sequences. (b) A computational model of I-SceI binding to its target DNA sequence based on crystal structure data deposited in the protein data bank, highlighting the interactions of the four amino acids (N152, K193, R48 and N15) directly contacting the nucleotides. (c–f) Sequential screening of I-SceI variants, which induced the highest HDR at the targeting sequence indicated at the upper right corner. The blue arrowhead indicates the original amino acid, and the red arrowhead indicates the selected amino acid that yields the highest HDR in each round of screening. Wild-type (Wt) I-SceI was included in all screenings (N in first round was the wild-type I-SceI).

in which double-stranded DNA probes (5 nmol) were mixed with various concentrations of either GST-I-SceI or GST-ISVB2 in EMSA reaction buffer (20 mM Tris-HCl, pH 8.5, 50 mM NaCl, 2 μM ZnCl<sub>2</sub>, 12 mM MgCl<sub>2</sub>, 2% glycerol, 2 mg/ml BSA and 2 mM freshly prepared DTT). All reaction mixtures were incubated at 25°C for 40 min and subsequently analyzed on 4% non-denaturing polyacrylamide gels. Samples in the gels were then transferred to nylon membranes (Milipore) and ultraviolet-crosslinked at 2000 J. The crosslinked membranes were blocked in 5% non-fat milk, followed by a 1-h incubation in 1:10 000 diluted streptavidin-HRP conjugate (Thermo) and then washed three times. The chemiluminescent signal was detected using the chemiluminescence labeling detection reagent ECL Plus (GE healthcare) and analyzed using the Bio-Rad Chemidoc XRS+. The I-SceI and β-globin

probes had the following sequences: 5'-TGCACCATTCTTAGGGATAACAGGGTAATTTTCTGGGTTA-3' and 5'-TGCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTA-3', respectively.

### Statistical analysis

The statistical significance of all data was determined by two-tailed Student's *t*-tests by using the PRISM program (*P*-value < 0.05).

### Isolation, culturing and electroporation of human hematopoietic stem cells

Human hematopoietic stem cells were collected by isolating CD34<sup>+</sup> cells from peripheral blood from consent donors at

the First Affiliated Hospital of Guangxi Medical University.  $10^6$  cells per milliliter CD34<sup>+</sup> cells were cultured for 24 h in Serum-free StemSpan medium (StemCell Technology), which contains 1% penicillin-streptomycin (Life Technology), 50 ng/ml Stem Cell Factor (R&D), 50 ng/ml FLT3 ligand (R&D), 50 ng/ml Thrombopoietin (R&D) and 50 ng/ml Interleukin 6 (R&D). CD34<sup>+</sup> cells were electroporated with 1  $\mu$ g/ml vectors carrying TALE-ISVB2-P2A-mKate. After electroporation, the CD34<sup>+</sup> cells were cultured in the StemSpan media for 48 h, followed by FACS collection of cells expressing TALE-ISVB2-P2A-mKate.

### High-throughput sequencing and analysis

High-throughput sequencing was performed on cells sorted for CD34<sup>+</sup> cells with or without (control) the electroporation of TALE-ISVB2-P2A-mKate. Genomic DNA of the CD34<sup>+</sup> cells was collected by using the DNeasy blood and tissue Kit (QIAGEN). PCR primers were designed to target 150–200 bps on target locus ( $\beta$ -globin gene) or putative off-target sites (POS). The POS are highly homologous to the target locus. Note that 16 POS are one-nucleotide different and 2 POS are two-nucleotide different from the human  $\beta$ -globin sequence (Supplementary Table S1). PCR amplicons for both samples and controls were then subjected to adaptors ligation and an additional round of amplification. The PCR amplicon libraries were sequenced on HiSeq1500 using paired end 100 bp with a 6-bp index read. The reads were aligned to human reference genome 19 (hg19) by Burrows–Wheeler Alignment tool (24). Each aligned read pair was individually genotyped for the presence of indels for statistical analysis.

## RESULTS

### TALEN and CRISPR/CAS9 systems can induce HDR at the $\beta$ -globin sequence

To create an endonuclease with high HDR efficiency and low genome toxicity, we first tested whether TALEN or CRISPR/CAS9D10A could achieve these goals. We designed TALENs to recognize 33-bp sequences of the human  $\beta$ -globin gene close to a highly prevalent mutation site associated with  $\beta$  thalassemia (Figure 1a). The HDR rate was measured using a reporter system created by inserting 53 bp of the  $\beta$ -globin sequence into the coding region of eGFP sequence, which was subsequently integrated into human 293FT cells via lentiviral infection (Figure 1a). When template DNA containing only truncated eGFP was provided to cells harboring the  $\beta$ -globin HDR reporter, no cells expressing eGFP were observed using FACS (Figure 1b) in the  $10^5$  cells examined. However, when TALENs were transfected with the template DNA for HDR, a population of cells expressing eGFP was detected using FACS (Figure 1b). On average, TALENs induced HDR by  $\sim 200$  eGFP<sup>+</sup> cells/100 000 cells (or 0.20%) in our system (Figure 1b and c). Similar levels of HDR (0.37%) were observed in another study that used TALEN and an integrated GFP reporter in 293T cells (20).

In addition, we created gRNA that targeted 20 bp in the same locus and utilized the CRISPR/CAS9D10A system as

described previously (15,20). This system relies on the mutated CAS9 that is acting as nickase to increase the targeting specificity although its targeting efficiency is expected to be lower than the wild-type CAS9. Similarly, gRNA targeting at the  $\beta$ -globin locus along with CRISPR/CAS9D10A induced HDR by 0.20% using the same reporter (Figure 1a–c).

### TALENs and CRISPR/CAS9 systems shows high levels of genome toxicity

Although TALENs and CRISPR/CAS9D10A induced HDR at the human  $\beta$ -globin DNA sequence, the genotoxicity induced by these two endonucleases was still unknown. To examine the genotoxicity induced by the endonucleases, we measured the levels of  $\gamma$ H2AX and 53BP1 in 293FT cells because these two markers have been widely used as indicators of DNA damage, including double-strand breaks (DSBs)(17,25–27) and single-strand breaks (28,29). First, we used etoposide, a chemical known to induce DSB in cells (27), to demonstrate the sensitivity of  $\gamma$ H2AX and 53BP1 as indicators of genotoxicity. At 2  $\mu$ M, we observed significant staining of  $\gamma$ H2AX induced by etoposide (Supplementary Figure S1a), and cells with many foci ( $\geq 3$ ) of 53BP1 (Supplementary Figure S1b). In general,  $\gamma$ H2AX foci were more diffuse and less distinct than 53BP1 foci in strongly positive nuclei. Increasing concentration of etoposide to 10  $\mu$ M simultaneously and significantly increased the levels of  $\gamma$ H2AX and 53BP1 (Supplementary Figure S1a and b). Therefore,  $\gamma$ H2AX is a sensitive DNA damage indicator with a more diffused pattern and 53BP1 clearly indicates different levels of DSBs via staining. Together, these two markers could measure and illustrate genotoxicity in the genome. We also observed a background level of cells possessing low levels of  $\gamma$ H2AX and 53BP1 without etoposide treatment, probably reflecting cells undergoing DNA repair during DNA replication or apoptosis.

To examine whether TALEN targeting at the  $\beta$ -globin sequence induced any genotoxicity, we measured the levels of  $\gamma$ H2AX and 53BP1 in cells transfected with these endonucleases. We determined that  $\sim 14\%$  of TALEN-transfected cells had a high  $\gamma$ H2AX signal and  $\sim 16.5\%$  of cells had more than 3 foci of 53BP1 (Figure 1d–f). Surprisingly, CAS9D10A also induced  $\gamma$ H2AX expression by 16.3% and increased the number of cells with more 53BP1 foci by 14.5% (Figure 1d–f). This suggested that TALEN and CAS9D10A targeting at the  $\beta$ -globin sequence could cause unwanted DSBs or DNA damage at sites other than the targeting locus. To compare the effects of genotoxicity, we also examined I-SceI, a homing endonuclease that has low levels of genotoxicity (17). As expected, cells transfected with I-SceI showed background levels of  $\gamma$ H2AX expression and 53BP1 foci in contrast to other endonucleases expressed at the same level (Figure 1d–f). These results indicated that using TALENs or CAS9D10A to induce HDR at the  $\beta$ -globin site would cause DNA damage and harmful recombination in the human genome. Therefore, we sought an alternative targeting method to reduce genotoxicity.

### Re-engineering I-SceI to induce HDR at the $\beta$ -globin sequence

Because I-SceI showed only background levels of genotoxicity when transfected into human cells, we speculated that it might be an ideal endonuclease to be re-engineered for gene targeting. First, to confirm whether I-SceI can induce HDR in 293FT cells, we tested the HDR efficiency induced by I-SceI using the eGFP reporter system inserted with 18 bp of the I-SceI-recognizing sequence (Supplementary Figure S2a). I-SceI induced HDR by 0.21% compared to that of its original recognizing sequence in human cells, which is similar to percentage of HDR induced by TALENs or CAS9D10A (Supplementary Figure S2b). HDR induced by I-SceI showed dose-dependency on the I-SceI vector and template (Supplementary Figure S2c and d). Moreover, measuring eGFP-positive repair events at 36 and 48 h after transfection yielded similar levels of HDR (Supplementary Figure S2e). To reduce potential genotoxicity, we used 500 ng/ml of template DNA and 500 ng/ml of the vector expressing I-SceI, and then measured HDR events 36 h after transfection in the subsequent experiments.

Next, we re-engineered I-SceI to recognize the  $\beta$ -globin sequence by changing amino acids in I-SceI and the recognized DNA sequences simultaneously. Then, the mutated I-SceI sequences that induced the highest HDR were sequentially selected (Supplementary Figure S3). First, we chose 18 bp in the  $\beta$ -globin sequence, in which 7 bp are different from the original I-SceI recognition sequence (Figure 2a). Based on the crystal structure of I-SceI bound to its DNA substrate (30), we first modified N152 and two nucleotides at position  $-7$  and  $-5$  of the I-SceI recognition sequence (Figure 2b). We created 19 variants of I-SceI corresponding to 19 alternative amino acids at position 152. To select the variant that induced the highest HDR level, we examined the number of eGFP-positive cells induced by the 19 variants compared to the original I-SceI using the eGFP reporter containing the modified I-SceI recognition sequence. We found that changing N152 to K or Q yielded the highest HDR rate, increasing it by  $\sim 500$  cells/ $10^5$  transfected cells, whereas the wild-type I-SceI only induced HDR by 80 cells/ $10^5$  transfected cells (Figure 2c). When the two variants with the highest HDR were tested again, Q showed the highest HDR (Supplementary Figure S4a). Hence, we chose the Q variant for the next round of modifications. In the second round of screening, the I-SceI variant containing N152Q was modified at K193 to 19 other amino acids and tested again with the eGFP reporter containing a newly modified I-SceI recognition sequence had A instead of G at position  $-3$  (Figure 2d). Because the original K variant induced the highest HDR in the second round of modifications, we did not change K193. In the third round, we changed R48 in the I-SceI sequence as well as  $+6$  and  $+8$  of the recognition sequence. Then, N15 in the I-SceI sequence as well as  $+9$  and  $+11$  of the recognition sequence were changed in the fourth round. The third and fourth rounds of screening resulted in changing R48 to V and N15 to K (Figure 2e and f) based on the highest HDR rates induced by the variants, which was confirmed by repeating the tests with variants that induced the three highest HDR rates (Supplementary Figure S4b and c). After four rounds of screening

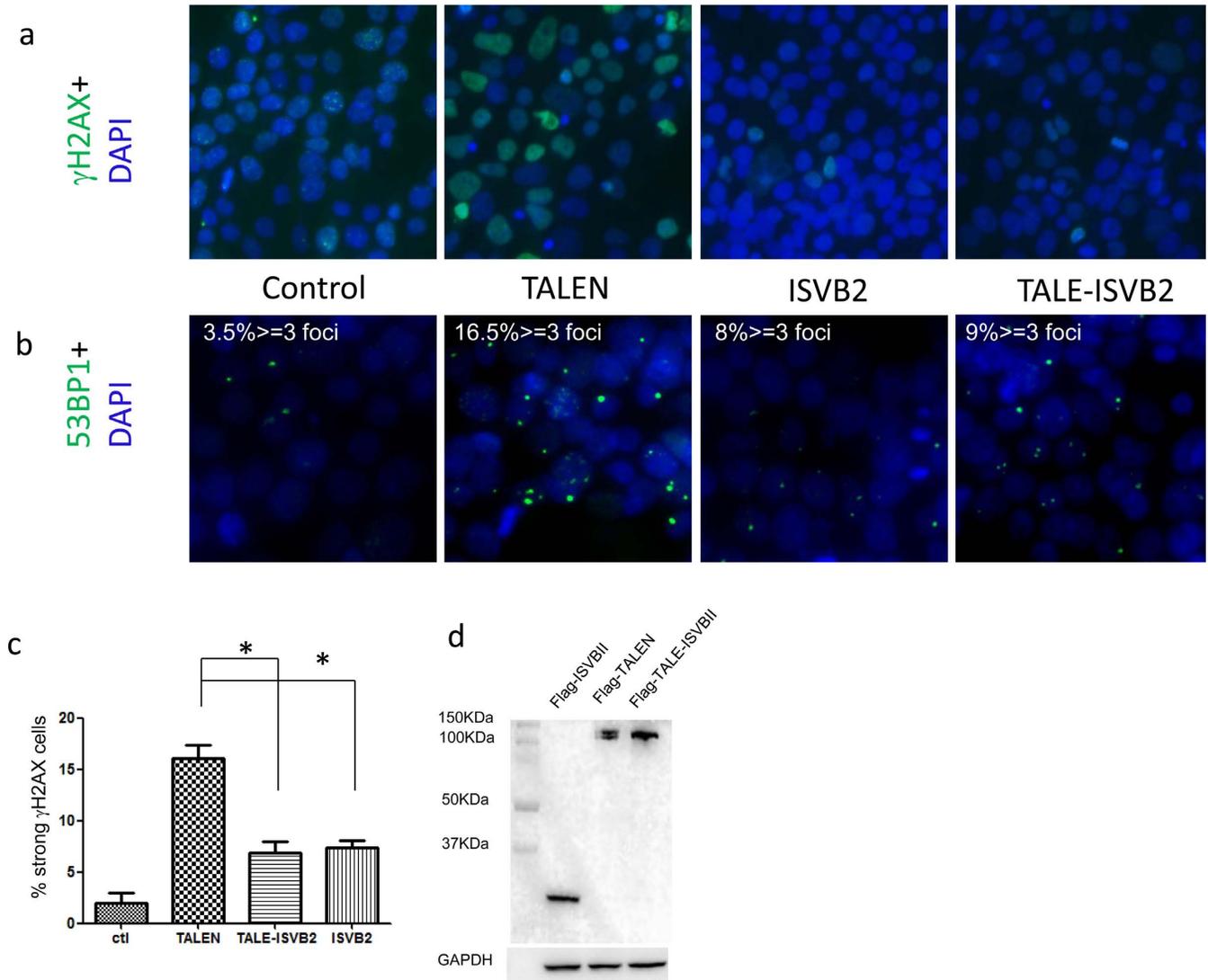
amino acids that directly interact with the 7 nucleotides of the recognition sequences, we obtained a modified I-SceI (named ISVB2) that could induce HDR by 520 cells/ $10^5$  transfected cells with a reporter containing 18 bp of the  $\beta$ -globin sequence.

The apparent equilibrium dissociation constant ( $K_d$ ) for the binding of ISBV2 to the  $\beta$ -globin sequence was determined using an EMSA (Supplementary Figure S5a). Biotinylated DNA probes (5 nmol) were mixed with GST purified I-SceI or ISVB2 at various concentrations. The  $K_d$  of DNA probes containing the original I-SceI or  $\beta$ -globin sequences was compared with those after the re-engineering of I-SceI. We found that the  $K_d$  for the binding of I-SceI protein to the I-SceI sequence was  $\sim 8.2 \times 10^{-8}$  M, while the  $K_d$  for ISVB2 binding to  $\beta$ -globin sequence was  $\sim 5.4 \times 10^{-8}$  M (Supplementary Figure S5a and b). When the DNA probes were switched, the  $K_d$  for binding of the I-SceI protein to the  $\beta$ -globin sequence became  $\sim 3.9 \times 10^{-7}$  M, while that for the ISVB2 binding to I-SceI sequence was  $3.7 \times 10^{-7}$  M (Supplementary Figure S5c and d). Therefore, ISVB2 showed high affinity to the  $\beta$ -globin sequence, suggesting its high specificity when expressed in human cells.

### Fusion of TALE to modified I-SceI increased its HDR rate and caused low genotoxicity

Although the re-engineered I-SceI, ISVB2, showed a comparable HDR rate to that of TALEN or CRISPR/CAS9D10A, we sought to test whether the HDR rate of ISVB2 could be further enhanced, while reducing the genotoxicity of TALEN by fusing ISVB2 to TALE. First, we confirmed that the FokI domain of dimeric TALEN contributed to the genotoxicity (Supplementary Figure S6), because  $\gamma$ H2AX and 53BP1 expression levels were significantly reduced when the FokI domain is replaced by a catalytically inactive mutant of FokI as previously reported (17). Hence, we replaced the FokI endonuclease domain with ISVB2 to create a monomeric fusion endonuclease (Figure 3a). Fifteen base pairs of the  $\beta$ -globin sequence are recognized by the TALE domain in addition to the 18-bp  $\beta$ -globin sequence specifically bound by ISVB2. Therefore, the fusion endonuclease recognizes a 33-bp sequence combining both of the TALE and ISVB2 domains. Using the reporter system that contains 53 bp of the  $\beta$ -globin sequences described above in the TALEN and CRISPR/CAS9D10A experiments, we tested whether TALE-ISVB2 could induce the HDR at the  $\beta$ -globin sequence. ISVB2 induced HDR by only  $\sim 50\%$  compared to that of TALEN (Figure 3b). However, ISVB2 fused to TALE showed a similar level of HDR to that of TALEN, indicating that TALE-DNA recognition domain we designed helped to increase the targeting efficiency of ISVB2. To determine whether the length of the linker between the TALE domain and ISVB2 affects the HDR efficiency, we tested four different linker lengths composed of 16, 28, 39 or 63 amino acids (Figure 3c). Only the 16-amino-acid linker showed a significantly lower HDR than other linkers, and the 28-amino-acid linker induced the highest HDR (Figure 3c). Therefore, we used the linker with 28 amino acids.





**Figure 4.** ISVB2 induced low genotoxicity. (a and b) ISVB2 and TALE-ISVB2 induced low levels of genotoxicity in human cells.  $\gamma$ H2AX and 53BP1 immunostained 293FT cells induced by TALEN, ISVB2 and TALE-ISVB2. The pictures of TALEN treated cells are the same as those in Figure 1 for comparison. (c) Percentage of cells with high  $\gamma$ H2AX expression determined by FACS of three independent cell populations. The error bars indicate standard deviation. TALE-ISVB2 and ISVB2 samples both have significantly lower  $\gamma$ H2AX expression than cells exposed to TALENs. The asterisk indicates the statistical significance between the samples ( $P$ -value  $< 0.05$ ). (d) Western blot results showed that the levels of ISVB2, TALEN and TALE-ISVB2 expression were similar.

This allows us to isolate CD34<sup>+</sup> cells that expressed TALE-ISVB2 by FACS. CD34<sup>+</sup> cells without electroporation and CD34<sup>+</sup> cells expressing TALE-ISVB2-P2A-mKate were independently collected for high-throughput sequencing (Figure 5c). We examined the genomic sequences at the  $\beta$ -globin sequence and 18 POS. These POS were selected based on their high homology of DNA sequences to the  $\beta$ -globin sequence. Note that 16 POS are one-nucleotide different and 2 POS are two-nucleotide different (because there is no single-nucleotide different sequence in the human genome at the corresponding position; Supplementary Table S1) from the  $\beta$ -globin sequence. Note that 4.85% of CD34<sup>+</sup> expressing TALE-ISVB2 carried indels at the  $\beta$ -globin sequence, whereas 17 putative site showed 0.00–0.14% of indels (Figure 5d), indicating a specific cleavage at the  $\beta$ -

globin sequence by TALE-ISVB2. Notably, the majority of indels (144/240) at the  $\beta$ -globin sequence were single nucleotide deletion at the expected cleavage sequence (-ATAA-) of ISVB2 (Figure 5e), further supporting the specific cleavage by this re-engineered nuclease.

## DISCUSSION

Our findings show that TALENs can induce high rates of HDR at the human  $\beta$ -globin sequence in human cells. However, the genotoxicity of the endonucleases was relatively high, raising concerns for its use in therapeutic gene targeting. Recent studies have focused on using CRISPR/CAS9 as a potential method for gene targeting in human cells, but studies have showed that the mutations and off-target effects of this system could also pose unwanted mutations



In theory, we could modify more than four amino acids in I-SceI to target other DNA sequences at other genomic loci if necessary. If the four nucleotides (-ATAA-) at the core of the recognizing sequence were kept unchanged, then the probability of finding the same sequence would be every 256 bp, and the frequency of the four nucleotide-sequence in human genome is  $\sim 1.1 \times 10^7$ . Therefore, many mutations associated with human diseases are likely to be targeted by modifying I-SceI amino acids. This system does not require both arms of recognition sequences to be designed at targeting sites thus reduces the potential off-target errors caused by dimerization of nucleases, such as FokI. Similar design of fusion protein consists of TALE and a variant of another homing endonuclease, I-OnuI, was recently reported to specifically target T-cell receptor alpha gene (31). A complete matching sequence for a homing endonuclease in human genome is quite rare or not existing. Hence, our approach of re-engineering a homing endonuclease to recognize a desired sequence will widen the usage of these endonucleases. Finally, we tested the cleavage specificity of TALE-ISVB2 in human hematopoietic cells and found that it specifically cleaved the  $\beta$ -globin sequence. This is consistent with its low genotoxicity measured by  $\gamma$ H2AX and 53BP1. Hence, using a monomeric endonuclease of TALE-I-SceI to target DNA sequence in human genome provides a potential option to therapeutic gene targeting.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENT

We thank Nieng Yan for providing the TALEN plasmid to us and Babak Javid for critical comments on the manuscript.

## FUNDING

Research funding is provided by the National Natural Science Foundation of China [81470011]; the Ministry of Science and Technology of China [2012CB966702]. Funding for open access charge: the National Natural Science Foundation of China [81470011]; the Ministry of Science and Technology of China [2012CB966702].

Conflict of interest statement. None declared.

## REFERENCES

- Ginn, S.L., Alexander, I.E., Edelstein, M.L., Abedi, M.R. and Wixon, J. (2013) Gene therapy clinical trials worldwide to 2012 - an update. *J. Gene Med.*, **15**, 65–77.
- Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.L., Fraser, C.C., Cavazzana-Calvo, M. *et al.* (2003) A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.*, **348**, 255–256.
- Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K. *et al.* (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.*, **118**, 3132–3142.
- Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. and Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, **11**, 636–646.
- Bogdanove, A.J. and Voytas, D.F. (2011) TAL effectors: customizable proteins for DNA targeting. *Science*, **333**, 1843–1846.
- Wiedenheft, B., Sternberg, S.H. and Doudna, J.A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, **482**, 331–338.
- Mussolino, C. and Cathomen, T. (2013) RNA guides genome engineering. *Nat. Biotechnol.*, **31**, 208–209.
- Pattanayak, V., Ramirez, C.L., Joung, J.K. and Liu, D.R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods*, **8**, 765–770.
- Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassidy, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C. *et al.* (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.*, **29**, 731–734.
- Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–826.
- Cradick, T.J., Fine, E.J., Antico, C.J. and Bao, G. (2013) CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.*, **41**, 9584–9592.
- Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S. *et al.* (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.*, **32**, 670–676.
- Doyon, Y., Vo, T.D., Mendel, M.C., Greenberg, S.G., Wang, J., Xia, D.F., Miller, J.C., Urnov, F.D., Gregory, P.D. and Holmes, M.C. (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods*, **8**, 74–79.
- Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.*, **31**, 827–832.
- Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y. *et al.* (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380–1389.
- Jasin, M. (1996) Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.*, **12**, 224–228.
- Szczepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D.J. and Cathomen, T. (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.*, **25**, 786–793.
- Pruett-Miller, S.M., Reading, D.W., Porter, S.N. and Porteus, M.H. (2009) Attenuation of zinc finger nuclease toxicity by small-molecule regulation of protein levels. *PLoS Genet.*, **5**, e1000376.
- Soutoglou, E., Dorn, J.F., Sengupta, K., Jasin, M., Nussenzweig, A., Ried, T., Danuser, G. and Misteli, T. (2007) Positional stability of single double-strand breaks in mammalian cells. *Nat. Cell Biol.*, **9**, 675–682.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
- Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S. and Zhang, B. (2011) Heritable gene targeting in zebrafish using customized TALENs. *Nat. Biotechnol.*, **29**, 699–700.
- Kee, K., Angeles, V.T., Flores, M., Nguyen, H.N. and Reijo Pera, R.A. (2009) Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. *Nature*, **462**, 222–225.
- Ruff, P., Koh, K.D., Keskin, H., Pai, R.B. and Storici, F. (2014) Aptamer-guided gene targeting in yeast and human cells. *Nucleic Acids Res.*, **42**, e61.
- Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
- Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S. and Bonner, W.M. (1998) DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.*, **273**, 5858–5868.
- Rappold, I., Iwabuchi, K., Date, T. and Chen, J. (2001) Tumor suppressor p53 binding protein 1 (53BP1) is involved in DNA damage-signaling pathways. *J. Cell Biol.*, **153**, 613–620.
- Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A. *et al.* (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.*, **25**, 778–785.

28. Matsumoto, M., Yaginuma, K., Igarashi, A., Imura, M., Hasegawa, M., Iwabuchi, K., Date, T., Mori, T., Ishizaki, K., Yamashita, K. *et al.* (2007) Perturbed gap-filling synthesis in nucleotide excision repair causes histone H2AX phosphorylation in human quiescent cells. *J. Cell Sci.*, **120**, 1104–1112.
29. Narciso, L., Fortini, P., Pajalunga, D., Franchitto, A., Liu, P., Degan, P., Frechet, M., Demple, B., Crescenzi, M. and Dogliotti, E. (2007) Terminally differentiated muscle cells are defective in base excision DNA repair and hypersensitive to oxygen injury. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 17010–17015.
30. Moure, C.M., Gimble, F.S. and Quijoch, F.A. (2003) The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J. Mol. Biol.*, **334**, 685–695.
31. Boissel, S., Jarjour, J., Astrakhan, A., Adey, A., Gouble, A., Duchateau, P., Shendure, J., Stoddard, B.L., Certo, M.T., Baker, D. *et al.* (2014) megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res.*, **42**, 2591–2601.