Seamless Correction of the Sickle Cell Disease Mutation of the \textit{HBB} Gene in Human Induced Pluripotent Stem Cells Using TALENs

Ning Sun,1 Huimin Zhao1,2

1Department of Biochemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801; telephone: +1-217-333-2631; fax: +1-217-333-5052; e-mail: zhao5@illinois.edu
2Departments of Chemical and Biomolecular Engineering, Bioengineering, Chemistry, Center for Biophysics and Computational Biology, Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801

\textbf{ABSTRACT:} Sickle cell disease (SCD) is the most common human genetic disease which is caused by a single mutation of human \textit{\beta}-globin (\textit{HBB}) gene. The lack of long-term treatment makes the development of reliable cell and gene therapies highly desirable. Disease-specific patient-derived human induced pluripotent stem cells (hiPSCs) have great potential for developing novel cell and gene therapies. With the disease-causing mutations corrected in situ, patient-derived hiPSCs can restore normal cell functions and serve as a renewable autologous cell source for the treatment of genetic disorders. Here we successfully utilized transcription activator-like effector nucleases (TALENs), a recently emerged novel genome editing tool, to correct the SCD mutation in patient-derived hiPSCs. The TALENs we have engineered are highly specific and generate minimal off-target effects. In combination with \textit{piggyBac} transposon, TALEN-mediated gene targeting leaves no residual ectopic sequences at the site of correction and the corrected hiPSCs retain full pluripotency and a normal karyotype. Our study demonstrates an important first step of using TALENs for the treatment of genetic diseases such as SCD, which represents a significant advance toward hiPSC-based cell and gene therapies.

© 2013 Wiley Periodicals, Inc.

\textbf{KEYWORDS:} TAL effector nucleases; induced pluripotent stem cells; \textit{piggyBac} transposon; sickle cell disease; gene therapy; genome editing

Human induced pluripotent stem cells (hiPSCs) are genetically reprogrammed from adult somatic cells. Similar to human embryonic stem cells (hESCs), hiPSCs can propagate indefinitely in culture and differentiate to all somatic tissues (Takahashi et al., 2007). Disease-specific hiPSCs derived from patients have invaluable therapeutic applications such as studying disease mechanisms, screening effective drugs, and perhaps most importantly, treating diseases via cell replacement therapies. For the treatment of genetic disorders which are caused by genomic abnormalities, patient-derived hiPSCs can be corrected in culture and thus provide a renewable cell source for autologous transplantations, in which modified hiPSCs (or their differentiated progenies) can be transplanted back into the patient to restore normal cell functions (Hanna et al., 2007).

To correct the disease causing mutations, custom-designed DNA endonucleases such as zinc finger nucleases (ZFNs) and transcription activator-like (TAL) effector nucleases (TALENs) have been constructed to introduce a site-specific DNA double-strand break (DSB) on the chromosome to enhance homologous recombination (HR) (Sun et al., 2012a). ZFNs are artificial DNA endonucleases using zinc finger domains for DNA recognition and a non-specific FokI catalytic domain for DNA cleavage. ZFNs have been successfully applied to induce HR-based gene correction in patient-derived hiPSCs (Sebastiani et al., 2011; Soldner et al., 2011; Yusa et al., 2011; Zou et al., 2011). Yet the low specificity of the ZFNs can cause cleavage at off-target sites, resulting in undesired genomic instability (Pattanayak et al., 2011). Moreover, the limited modularity of zinc finger domains makes it difficult for ZFNs to target many disease-causing genes. In contrast, TALENs represent a new class of artificial nucleases with high specificity and modularity (Joung and Sander 2013; Sun and Zhao 2013). The central repeat domain of a TALEN consists of repeating units of 33–35 amino acids.
Each repeat recognizes a single nucleotide and the specificity is conferred by the highly variable di-residues at positions 12 and 13 (e.g., Asn-Ile recognizes A, His-Asp recognizes C, Asn-Gly recognizes T, and Asn-Asn recognizes G and A) (Boch et al., 2009). The simple DNA recognition code and the modular nature make TALENs an ideal platform for use in targeted gene correction.

To explore the potential therapeutic application of TALENs in hiPSC-based cell and gene therapies, we chose sickle cell disease (SCD) as a model system. SCD is the most common human genetic disease that affects more than 300,000 individuals worldwide each year (Weatherall and Clegg 2001). A homozygous mutation (from A to T) in the sixth codon of the human β-globin (HBB) gene converts a glutamate to a valine, which generates abnormal β-globin proteins and results in malfunctioning red blood cells. The absence of an adequate long-term treatment makes hiPSCs-based cell replacement therapy highly attractive. To this end, we constructed a pair of custom-designed TALENs to recognize the sickle HBB (HBBs) gene at the mutation site (Fig. S1) (Sun et al., 2012b). In order to correct the sickle mutation (E6V) of the HBBs gene by TALEN-induced HR, we constructed a donor plasmid to provide an overall ~2.5 kb homology sequence carrying a wild-type HBB gene (HBBw) with GAG for codon six to substitute GTG (Fig. 1). The donor plasmid also contains a piggyBac transposon, which is a mobile element that transposes efficiently in human stem cells (Yusa et al., 2011). The piggyBac transposon carries a bifunctional fusion protein (puroΔtK) constitutively expressed under PGK promoter, in which puro (puromycin N-acetyltransferase) is used for positive selection and ΔtK (a truncated version of herpes simplex virus type 1 thymidine kinase) is used for negative selection. The PPGK-puroΔtK cassette is flanked by two piggyBac terminal inverted repeats (PB5 and PB3) into the TTA site to enable its subsequent excision. Expression of piggyBac transposase will excise the transposon seamlessly, without leaving any residual sequences.

To correct the HBBs gene, we transfected SCD patient-derived hiPSCs carrying a homozygous E6V mutation with the TALEN-expression plasmids together with the donor plasmid. TALEN cleavage induced HR, which enabled the replacement of the chromosomal HBBs gene with the episomal HBBw gene and simultaneously integrated a piggyBac transposon simultaneously.

Figure 1. Schematic overview depicting the strategy for the seamless correction of the HBBs gene using TALENs and piggyBac transposon. A. Sequence of the genomic HBBs gene locus. Amino acids of the beginning part of human βS-globin protein are shown in ovals. The A to T substitution (bold letter) causes the E6V mutation (gray oval) in human βS-globin. The two TALEN recognition sites (underlined) are separated by a 15 bp spacer. The TALEN cleavage site and the integration site for piggyBac transposon (in frame) are indicated. B. Schematic of the donor plasmid design and gene correction strategy. The donor plasmid comprises two homology arms (1.0 kb 5’-homology arm of promoter and 1.5 kb 3’-homology arm of the wild-type HBB (HBBs) gene) and a piggyBac transposon (dotted). Within the piggyBac transposon, a bifunctional fusion protein (puroΔtK) constitutively expressed under PGK promoter is flanked by two piggyBac terminal inverted repeats (PB5 and PB3). TALEN-cleavage introduces a site-specific DSB at the HBBs gene and stimulates HR between the genomic fragment and the donor plasmid, resulting in the substitution of the HBBs gene by the HBBw gene and the integration of a piggyBac transposon simultaneously. Subsequent expression of piggyBac transposase excises the piggyBac transposon without leaving any residual ectopic sequences.
transposon carrying the drug-selectable cassette (Fig. 1B). Puromycin-resistant hiPSC colonies were picked and expanded for nested-PCR screening. We observed that >60% puromycin-resistant colonies were correctly targeted (Fig. S2). To remove the drug-selectable cassette from the modified hiPSCs, the piggyBac transposase gene was transiently transfected to the cells. After negative selection under fialuridine, hiPSC colonies were picked and expanded. Genotyping was performed by genomic PCR followed by restriction digestion with Bsu36I, which cleaves the corrected (HBB) but not the sickle (HBB) gene (Figs. 2A and B). Two out of 48 hiPSC colonies were sensitive to Bsu36I digestion, indicating the correction of the mutation on one allele (Fig. S3). Sequencing of the genomic PCR products of the two corrected hiPSC colonies confirmed that transposon excision did not leave any residual “scar,” demonstrating that one HBB gene in the SCD patient-derived hiPSCs was corrected seamlessly (Fig. 2C). Two introduced silent mutations were observed as initially designed, confirming that the replacement was introduced by TALEN-mediated gene correction, and not by spontaneous reversion. Genomic PCR confirmed that the gene-corrected cells were free of the piggyBac transposon (Fig. S4).

The gene-corrected hiPSCs maintained pluripotency, as indicated by the uniform expression of pluripotency-specific marker proteins (Fig. S5). To test pluripotency in vivo, the corrected hiPSCs were transplanted into severe combined immunodeficiency (SCID) mice and tumor formation was observed in 9–10 weeks. Histological examination showed that the tumor comprised cell types originating from all three developmental germ layers (Fig. S6), indicating that TALEN-mediated genome editing did not alter the pluripotent state of the corrected hiPSCs.

To monitor possible off-target events introduced by TALEN cleavage, we analyzed six genomic regions harboring the potential off-target cleavage sites (predicted by Tal Effector Nucleotide Targeter 2.0 (Doyle et al., 2012)), including a highly similar sequence within the HBD gene that shares only four mismatches compared with the designed TALEN binding site (Table I). Off-target activity was measured by SURVEYOR nuclease assay, which monitors the small insertions and deletions generated by imprecise DNA repair from non-homologous end joining (NHEJ) in response to TALEN cleavage. In contrast to HBB gene locus, within which exists three mismatches between the corrected (HBB) and uncorrected (HBB) genes, none of the other six

**Figure 2.** Seamless monoallelic gene correction of the HBB gene in SCD patient-derived hiPSCs. A: Sequences of sickle (HBB) and wild-type (HBB) human β-globin gene. Amino acids from the beginning part of human β-globin protein are shown in ovals. The A to T substitution (bold letter) causes the E6V mutation (gray oval) of HBB gene, which abolishes a Bsu36I restriction site in HBB gene. B: Bsu36I analysis confirming the correction of the HBB gene. HEK293 cell line has both HBB alleles, which are sensitive to Bsu36I digestion. SCD patient-derived hiPSCs (SCD-hiPSC) have both HBB alleles, which are resistant to Bsu36I digestion. Two corrected SCD-hiPSCs have one HBB allele and one HBB allele, as indicated by the double bands. C: Sequence analysis showing the HBB gene locus of the untreated patient-derived hiPSCs and the gene-corrected hiPSCs. The ATG start codon is indicated by a black arrow. The nucleotide substitutions are highlighted by red frames. Y, double peaks of C and T; W, double peaks of A and T.
analyzed sites were off-targeted (Fig. 3A). In addition, karyotype analysis showed that the TALEN-mediated gene correction and piggyBac transposon excision processes maintained genomic integrity of the corrected hiPSCs without causing any unexpected chromosomal translocations and alterations (Fig. 3B). These evidences indicate high specificity and safety of using TALENs and piggyBac transposon for targeted genome editing in patient-derived hiPSCs.

Obtainable from adult somatic cells and capable of growing indefinitely and differentiating into all cell types, hiPSCs open up various novel avenues for basic biomedical research and human therapeutics. The ultimate promise of hiPSCs, although at a very early stage of development, is to treat diseases by cell replacement therapy. Patient-derived hiPSCs can provide a renewable autologous cell source for the treatment of degenerative diseases such as central nervous system injuries, myocardial infarction, and diabetes (Goldring et al., 2011). Moreover, if the disease-causing mutations can be repaired in vitro, corrected hiPSCs can serve as regenerative medicines for the treatment of genetic disorders such as SCD and β-thalassemia. Toward this end, a proof-of-principle has been established in a SCD mouse model, including derivation of disease-specific mouse induced pluripotent stem cells (miPSCs) from adult somatic cells, repair of genetic mutations by HR-mediated gene targeting, differentiation of corrected miPSCs into therapeutic relevant cells, and subsequent autologous transplantation (Hanna et al., 2007). However, similar approaches are not applicable to human gene therapy because of the low HR rate in human cells, making the gene correction step a major obstacle.

Recently, TALENs have rapidly emerged a powerful genome editing tool. The site-specific DSBs introduced by TALENS can induce HR-based genetic modifications significantly. In this study, we sought to explore the therapeutic potential of TALENs in hiPSCs-based cell and gene therapies. We successfully applied TALENs to correct the SCD mutation in patient-derived hiPSCs with minimal side effect. The corrected hiPSCs maintain their pluripotent state. TALEN-mediated gene correction does not introduce detectable off-target events or chromosomal alterations. Despite the widespread use of TALENs, to the best of our knowledge, HBB<sup>α</sup> gene is only the third disease-causing gene that has been corrected in situ by TALEN technology, in addition to A1AT gene related to α<sub>1</sub>-antitrypsin deficiency (Choi et al., 2013) and COL7A1 gene related to epidermolysis bullosa (Osborn et al., 2013). The high efficiency, modularity, and high specificity make TALEN technology an ideal platform for the correction of disease-causing gene in patient-derived hiPSCs.

Although the chromosomal DSBs introduced by custom-designed DNA endonucleases can significantly stimulate HR, the overall gene targeting rate is still low, which makes it necessary to introduce a drug-selectable cassette into the

---

**Table 1.** Potential off-target sites of the custom-designed TALENs for the recognition of HBB<sup>α</sup> gene.

<table>
<thead>
<tr>
<th>Featured gene</th>
<th>Left</th>
<th>Right</th>
<th>Mismatch (bp)</th>
<th>Spacer length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBB&lt;sup&gt;α&lt;/sup&gt;</td>
<td>AGCAACCTCAACAGACACCAT</td>
<td>AAAACCACCAT</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>HBD</td>
<td>AGGAGCTCAACAGACACCAT</td>
<td>AACACCTGCTACCAT</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>HLF</td>
<td>AGGAGCTCAACAGACACCAT</td>
<td>AACACCTGCTACCAT</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>PTPRG</td>
<td>AGCAACCTCAACAGACACCAT</td>
<td>AAAACCACCAT</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>NSUN5</td>
<td>AAAACCACCAT</td>
<td>AAACCTGCTACCAT</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>ZNF204P</td>
<td>AGCAACCTCAACAGACACCAT</td>
<td>AAAACCACCAT</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td>CDH18</td>
<td>CACATACACCTTCCACA</td>
<td>CAACATACACCTTCCACA</td>
<td>6</td>
<td>30</td>
</tr>
</tbody>
</table>

*Bases differing from the left TALEN recognition site (underlined) and the right TALEN recognition site (bold) are highlighted. The possible off-target events within ZNF204P and CDH18 genes are generated by the formation of homodimer of the right TALEN.*

---

**Figure 3.** Analysis of off-target effects. A: SURVEOR nuclease analysis of the potential off-target sites. The HBB<sup>α</sup> site has three mismatches between corrected (HBB<sup>α</sup>) and uncorrected (HBB<sup>α</sup>) genes, which is sensitive to SURVEOR nuclease cleavage. All of the six potential off-target sites (sequences are shown in Table I) are resistant to SURVEOR nuclease cleavage. B: Product of SURVEOR nuclease cleavage. PiggyBac transposon excision processes do not cause gross chromosomal alterations.
To address this limitation, we inserted the drug-selectable cassette into a piggyBac transposon, which enables the removal of the selection cassette efficiently in hiPSCs without leaving any remnant sequences. After TALEN-mediated gene correction and drug selection, transient expression of piggyBac transposase excised the drug-selectable cassette without leaving any residual "scar" in the corrected gene. The approach we developed in this study by the combination of TALEN and piggyBac technology enabled correction of the SCD mutation in patient-derived hiPSCs in a specific and seamless manner. The procedure we applied including engineering design, process control, and outcome evaluation in this study can serve as a standard template for gene correction protocol in the future.

Materials and Methods

Cell Culture

SCD patient-derived hiPSCs with a homozygous E6V mutation were kindly provided by Dr. Marius Wernig from Stanford University (Stanford, CA). Cultures were maintained under feeder-free culture conditions on six-well plates coated with Matrigel Basement Membrane Matrix (BD Biosciences, Bedford, MA) in mTeSR1 medium (StemCell Technologies, Vancouver, Canada). Subculture was performed every 5–7 days by washing hiPSCs with DMEM/F12 medium, incubating with 1 mg/mL dispase (StemCell Technologies) for 7 min at 37°C, washing with DMEM/F12 medium, detaching with a cell scraper, breaking down into small clumps and plating onto a new Matrigel-coated plate.

TALEN-Mediated Genome Editing of hiPSCs

On day 0, SCD-hiPSCs were pre-treated with 2 μM of thiazovivin for at least 1 h prior to transfection. Cells were dissociated into single-cell suspension by washing with DMEM/F12, incubating with accutase (StemCell Technologies) for 7 min at 37°C, mixing with mTeSR1 medium followed by vigorous pipetting. 3 × 106 cells were transfected with 2 μg of each TALEN-encoding plasmid and 2 μg of donor plasmid by Amaxa nucleofection system using Human Stem Cell Nucleofector Kit 1 according to the manufacturer’s instructions. The treated cells were plated onto one Matrigel-coated 100 mm dish in mTeSR1 medium supplemented with 2 μM of thiazovivin for the first 24 h. On day 3, puromycin (Sigma, St. Louis, MO) selection at 0.5 μg/mL was started. Within 3 days, >95% of cells died rapidly. Survived individual colonies were picked and expanded in one or two Matrigel-coated 24-well plates on day 13–17. On day 20–24, near-confluent cells in each well were harvested using accutase. One quarter of the cells was seeded onto a new Matrigel-coated 24-well plate and the rest of the cells were used for nested PCR analysis.

Nested PCR of Targeted Integration

Genomic DNA was extracted from hiPSCs using Quick-Extract DNA Extraction Solution 1.0 (EPICENTRE Biotechnologies, Madison, WI) according to the manufacturer’s instructions. The external round of PCR was carried out using primers Purotest-for-PB3 (5′-ggc tcg aga tcc act agt tc-3′) and Purotest-rev-HBB3-long (5′-gat gct gct cgg cct tca ta-3′). The external round of PCR products were used as template for the internal round of PCR using primers Nested-for-HBB3 (5′-acg tgg atg aag tgt gtg gt-3′) and HBB-R (5′-taa ggg tgg gaa aat aga cca cc-3′). Nested PCR products were then resolved on a 1.0% agarose gel.

Transposon Excision in Targeted hiPSCs

On day 0, correctly targeted hiPSCs were pre-treated with 2 μM of thiazovivin for at least 1 h prior to transfection. Cells were dissociated into single-cell suspension as described earlier. 2 × 106 cells were transfected with 4 μg of the hyperactive piggyBac transposase expression vector (pCMVhyPBase, kindly provided by Dr. Allan Bradley) by Amaxa nucleofection system using Human Stem Cell Nucleofector Kit 1 according to the manufacturer’s instructions. The treated cells were plated onto one Matrigel-coated 100 mm dish in mTeSR1 medium supplemented with 2 μM of thiazovivin for the first 24 h. On day 3, cells were dissociated into single-cell suspension and seeded to a new Matrigel-coated 100 mm dish at low density (~1,000 cells/mL) in mTeSR1 medium containing 2 μM of thiazovivin and 0.25 μM of fialuridine (Santa Cruz Biotechnology, Santa Cruz, CA). On day 6, medium was changed to mTeSR1 medium supplemented with 2.5 μM of fialuridine, which was used for daily medium change until picking colonies. During selection, >50% cells died. Survived individual colonies were picked and expanded in one or two Matrigel-coated 24-well plates on day 16–20. On day 23–27, near-confluent cells in each well were harvested using accutase. One quarter of the cells was seeded onto a new Matrigel-coated 24-well plate and the rest of the cells were used for Bsu36I analysis and sequencing.

This work was supported by the Centennial Chair Professorship (H.Z.) in the Department of Chemical and Biomolecular Engineering at the University of Illinois at Urbana-Champaign. We thank Professor Fei Wang at the University of Illinois at Urbana-Champaign for sharing the expertise in handling human stem cells. We thank Applied StemCell, Inc. (Menlo Park, CA) for performing immunostaining, karyotyping, teratoma formation, and analysis.


Supporting Information

Additional supporting information may be found in the online version of this article at the publisher’s web-site.