blood

Prepublished online August 31, 2011;
doi:10.1182/blood-2011-02-335554

Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease

Jizhong Zou, Prashant Mali, Xiaosong Huang, Sarah N. Dowey and Linzhao Cheng

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.

Blood First Edition Paper, prepublished online August 31, 2011; DOI 10.1182/blood-2011-02-335554

Category: *Gene Therapy*; *Hematopoiesis and Stem Cells*

Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease

Jizhong Zou[1,2,*†], Prashant Mali[1,3,*], Xiaosong Huang[1,2], Sarah N. Dowey[1,2], and Linzhao Cheng[1,2†]

[1]Stem Cell Program, Institute for Cell Engineering;
[2]Division of Hematology, Department of Medicine;
[3]Department of Biomedical Engineering;
The Johns Hopkins University School of Medicine

*These authors contribute equally to the work

[†]**Correspondence:**

Linzhao Cheng, Ph.D.
The Johns Hopkins University School of Medicine
Broadway Research Building, Room 747
733 N. Broadway, Baltimore, MD 21205
Phone: 410-614-6958; Fax: 443-287-5611
Email: lcheng2@jhmi.edu

Or

Jizhong Zou, Ph.D.
Email: jzou2@jhmi.edu

Running title : Precise correction of a disease mutation in patient iPSCs

Key words: Gene targeting and correction ; Human iPS cells ; sickle cell disease ;

Abstract

Human induced pluripotent stem cells (iPSCs) bearing monogenic mutations have great potential for modeling disease phenotypes, screening candidate drugs, and cell replacement therapy provided the underlying disease-causing mutation can be corrected. Here we report a homologous recombination (HR) based approach to precisely correct the Sickle Cell Disease (SCD) mutation in patient-derived iPSCs with two mutated beta-globin alleles (β^s/β^s). Using a gene-targeting plasmid containing a loxP-flanked drug-resistant gene cassette to assist selection of rare targeted clones and zinc finger nucleases engineered to specifically stimulate HR at the β^s locus, we achieved precise conversion of one mutated β^s to the wildtype β^A in SCD iPSCs. However, the resulting co-integration of the selection gene cassette into the first intron suppressed the corrected allele transcription. After Cre recombinase-mediated excision of this loxP-flanked selection gene cassette, we obtained "secondary" gene-corrected β^s/β^A heterozygous iPSCs that express at 25-40% level of the wildtype transcript when differentiated into erythrocytes. These data demonstrate that single nucleotide substitution in the human genome is feasible using human iPSCs. This study also provides a new strategy for gene therapy of monogenic diseases using patient-specific iPSCs, even if the underlying disease-causing mutation is not expressed in iPSCs.

Introduction

Human induced pluripotent stem cells (iPSCs) that are derived from adult somatic cells hold great promise as a renewable cell source for developing novel cell and gene therapies[1]. Since the first report of generating iPSCs from patients' somatic cells in 2008, a wave of studies have demonstrated either undifferentiated patient-specific iPSCs or their differentiated progenies are capable of recapitulating key aspects of disease-specific phenotypes in vitro. So far, most of the successful reports on disease modeling are based on iPSCs bearing monogenic disease mutations. Capable of expanding unlimitedly in culture, iPSCs derived from patients that suffer from a genetic mutation are also promising candidates for cell replacement therapy if the disease-causing mutations can be permanently corrected. Most published studies that corrected mutations in iPSCs relied on adding a copy of a functional transgene randomly into the genome by viruses to restore the wild-type phenotype. For instance, disease-corrected Fanconi Anemia iPSCs were established using this approach[2]. However, random integrations of a functional copy of the affected gene, especially mediated by viral vectors, could induce oncogenic mutations and thus jeopardize the goal of gene therapy. The precise correction or replacement of mutations at the endogenous loci by homologous recombination (HR) is highly desirable even though the efficiency is extremely low (10^{-6}) in non-cancerous human cells such as human iPSCs. Recently, several studies reported HR-mediated gene addition at a few selective loci in normal or disease-specific human iPSCs[3-7]. However, this gene addition approach is only suitable for phenotypic correction of loss-of-function mutations, but not for correcting dominant or gain-of-function mutations. The ideal solution is targeted and precise gene correction of the underlying disease-causing mutation, which also ensures the corrected gene will be expressed in the appropriate temporal and tissue-specific manner under the regulation of endogenous cis-elements.

Toward this, we focused on Sickle Cell Disease (SCD) as a model system to develop methods for precise gene correction of a disease-causing mutation. In the vast majority of SCD patients, the mutation of A>T (also known as β^A to β^s mutation) in both alleles of the beta-globin (*HBB*) gene that changes codon 6 from Glu (G**A**G) to Val (G**T**G) results in a defective form of adult hemoglobin in oxygen-carrying red blood cells. Although SCD was one of the first described molecular diseases, the goal for treating this monogenic disorder using gene therapy approaches has remained elusive[8,9]. Gene correction of β^s in mouse embryonic stem cells (ESCs) by HR has been reported previously[10,11]. A similar strategy was also used to correct the β^s mutation in mouse iPSCs derived from a humanized SCD mouse model, followed by successful transplantation of differentiated hematopoietic cells into isogenic mice to cure SCD phenotypes[12]. However, gene correction of the SCD mutation in human patient-specific iPSCs has not been demonstrated yet. Here we show that with appropriate gene-targeting vector design and zinc finger nuclease (ZFN) enhancement of HR efficiency, site-specific gene correction of the silent *HBB* gene in human iPSCs can be achieved. We also demonstrate the expression of the corrected wildtype β^A allele in red blood cells differentiated from thus corrected iPSCs. Some unexpected discovery, and discussions on promise and pitfalls of this site-specific gene correction approach in human cells are also presented.

Materials and Methods

Cell culture

Human iPSCs were maintained on primary mouse embryonic fibroblasts (PMEFs) as feeder cells using the standard human ESC media. The SCD iPSCs used in this study were generated in an earlier study[13]. The MBP5s1 (abbreviated S1) iPSC line was derived using *piggyBac* transposition of reprogramming factors into bone marrow stromal cells (MSCs) from an adult patient with SCD at Johns Hopkins Hospital. To adapt to single-cell passaging, the S1 iPSCs were first passaged from PMEFs feeders to feeder-free culture condition using Matrigel (BD Biosciences) and Stemedia Nutristem XF/FF medium (Stemgent) at 1:1 to 1:2 ratio using 0.05% trypsin (Invitrogen) digestion, and then were maintained by this trypsin-passaging and feeder-free culture condition. All patient samples were used per approval from Johns Hopkins University internal review board (IRB) for conducting laboratory research using anonymous human cells. All the iPSC lines described in this report will be provided upon request to Dr. L Cheng, under a standard (uniformed) material transfer agreement with the Johns Hopkins University. Human 293T cells that were used to validate ZFNs and gene targeting vectors were grown in DMEM high glucose supplemented with 10% fetal bovine serum (FBS).

Characterization of iPSCs after gene targeting and excision

Characterization and chromosome karyotyping (G-banding) of iPSC clones were performed at multiple passages as previously described[3,13-15].

Construction of a donor plasmid vector targeting the endogenous *HBB* locus

The left and right homology arms were amplified from genomic DNA of SCD fibroblasts (GM02340 In Coriell collection) used in the previous study[15]. The primer sets HBBL-F: 5'-ATCGGTACCGTGTGTAAGAAGGTTCCTGAGGCT and HBBL-R: 5'-ATCGCTAGCGGTCTCCTTAAACCTGTCTTGTAACC amplified the 5.9-kb left arm, and HBBR-F: 5'-ACTGCTAGCAATAGAAACTGGGCATGTGGAGACAG and HBBR-R: 5'-

GATCTCGAGAAGAAGGGCTCACAGGACAGTCAA amplified the 2-kb right arm. The G**T**G mutation in the genomic DNA was converted to the wild-type G**A**G codon using PCR-mediated mutagenesis. A *loxP*-flanked PGK-Hygromycin (Hyg) cassette from a previous targeting donor PHD[3] was cloned into Nhe I site between two homology arms in the pCR2.1-TOPO vector. The *loxP*-flanked PGK-Hyg cassette is to be inserted at nt.179 of the *HBB* gene, 37-bp downstream of the exon 1 junction (nt.142). The EF1α-TK.GFP gene cassette (for counter selection) obtained from a previously published vector[16] was inserted downstream of the right arm. The BD2 vector plasmid and sequence information is available upon request.

Gene Targeting Reagents and Experimental Procedure

The CompoZr™ custom designed HBB-ZFNs were obtained from the Sigma-Aldrich. It is now available to the public (CKOZFN1264-1KT). The ZFN pair was designed to target the following DNA sequence in exon 1 of the *HBB* gene (underlined in **Fig. 1A**). Two different HBB-ZFN subunits that recognize right (12-bp) and left (19-bp) half sequences can only form a dimmer with each other, which is required for endonuclease activity. They are expressed from two plasmids under the control of the CMV promoter. For GFP-based HR reporter assay in 293T cells, we cloned a 63-bp DNA sequence (from codon 1, nt.54 to nt.116) of the *HBB* gene containing the putative ZFN recognition sites into the middle of the EGIP* vector following a previously published strategy[3]. The insertion, together with a stop codon (taa) upstream and a HindIII site downstream aagctt), disrupt the EGFP reading frame. A full-length GFP activity could be restored, if the mutated GFP (GFP*) DNA is corrected by HR (**Fig. 1B**). A plasmid bearing a non-expressing, truncated GFP (tGFP) that provides a repair template, together with a pair of plasmids expressing two ZFNs, was introduced into the stably transfected 293T cells expressing the EGIP*-HBB sequence. The successful (GFP*) gene targeting was measured by the frequency of GFP+ cells two days after transient transfection[3]. To test the specificity of HBB-ZFNs, the homologous region from other beta-locus genes such as *HBE*, *HBG* and *HBD* was also inserted into the EGIP vector as tested in parallel. A shorter form of the *HBB* sequence (43-bp) containing the core ZFN recognition site was also tested by the same assay. For testing HBB-ZFNs and the BD2 vector to achieve the *HBB* endogenous gene targeting, 0.5 million 293T cells were transfected by lipofection with 1 µg BD2 donor and 0.25-2 µg/each HBB-ZFN in one 6-well plate. Five days post-transfection, 1 million transfected cells were plated on 150-mm tissue culture dish with 200 µg/ml hygromycin B selection, and then 20 µM gancyclovir (GCV) was added 8 days afterwards. The total drug selection lasted 15 days including 7 days for GCV and 15 days for hygromycin B selection. Pooled cell populations were used for genomic DNA extraction and PCR detection assays.

For correcting β^s mutation in SCD iPSCs, 5×10^6 iPSCs were trypsinized and resuspended in Amaxa's mESC nucleofection buffer and subsequently electroporated using the A-023 setting as previously described[3]. The treated PSCs were plated onto irradiated hygromycin-resistant PMEF feeders (PMEF-HL, Millipore), with Y-27632 (a ROCK Inhibitor) added at 10 µM final concentration into the media for overnight duration to improve survival of the nucleofected cells. From day 3 to day 17 hygromycin B selection at 20 µg/ml was performed. From day 10 to day 17 GCV (2 µM) was also supplemented to the media. Surviving colonies were picked on day 17 and expanded on PMEF feeders following standard culture protocols and screened for targeting events.

PCR detection of targeted integration, *HBB* and *Hyg* mRNA expression

Primer set 5'-CAAATGCGAGAGAACGGCCTTAC (in the Hyg cassette) and 5'-CTAGCACTGCAGATTCCGGGTCAC (on *HBB* locus downstream of 3'-homology arm) were used to amplify a 2.5-kb product of the 3'-junction of TI as indicated in **Fig. 2A**. Primer set 5'-ACATTTGCTTCTGACACAAC (in *HBB* exon 1) and 5'-AGCAAGAAAGCGAGCTTAG (in *HBB* exon 3) were used for RT-PCR detection of mutant or corrected *HBB* mRNA expression as indicated in **Fig. 2A and 6B**. Primer set 5'- ATTCCGGAAGTGCTTGACATTG and 5'- CACGCCATGTAGTGTATTGACC (both in Hyg coding sequence) were used for RT-PCR detection of *Hyg* mRNA expression in **Fig. 7A**.

Genomic PCR amplification and sequencing of both alleles in SCD iPSCs

Primer set 5'- CGATCACGTTGGGAAGCTATAGAG and 5'-AACATCCTGAGGAAGAATGGGAC were used to amplify the 3.5-kb genomic region surrounding *HBB* gene for both targeted and non-targeted alleles of S1, cre4, and cre16. For c36 (targeted) iPSCs, longer extension time was used to amplify both 3.5-kb non-targeted and 5.8-kb targeted alleles. Mixed alleles from these iPSC clones were cloned into a TOPO vector. Then DNA in individual bacterial clones was sequenced to identify specific alleles (targeted or non-targeted).

Southern blot analysis of genomic structure near the *HBB* locus

Standard Southern blot protocol using DIG-labeled probes was followed as described before[3]. The 3'-probe was amplified from S1 iPSC genomic DNA using primers 5'-GACTGAGAAGAATTTGAAAGGCG and 5'-TCATCAATTCTGCCATAAATGG. The Hyg probe used was the same as described previously[3].

Genome-Wide SNP concordance analysis

The Omni1_Quad BeadArray chip (Illumina) containing >10^6 probes detecting informative single nucleotide polymorphism (SNPs) was used. The array analysis was performed by the Johns Hopkins SNP Center as part of Center for Inherited Disease Research (CIDR, www.cidr.org). Based on 1,140,419 SNPs identified, and the array results with 2 control (CIDR11993 and CIDR10860) genomic DNA samples that have been previously sequenced and included in each run, the Johns Hopkins SNP Center reported a 0.27% genotyping error rate in the run, which was within the normal range. Genomic DNA isolated by the Qiagen DNAeasy kit from c36 (after ZFN-mediated HR), S1 iPSCs and their parental somatic cells (MSCs) were analyzed. Based on the ratio of allele intensities, the dis-concordance rate between c36 to S1 iPSCs (two samples of different passages before and after trypsin-adaptation, p17 and p26) or to MSCs (passage 0 and 6) is below 0.04%, which is much lower than the basal error rate level (0.27%). In comparison, the dis-concordance rates between c36/S1 iPSCs to other unrelated iPSCs and between two unrelated CIDR11993 or CIDR10860 controls are ~40%. Therefore, our data suggest that the c36 iPSC genome is essentially identical to the S1 iPSCs or their somatic MSCs.

Erythroblast differentiation from human iPSCs

Human iPSCs were directly differentiated into hematopoietic progenitor cells by a modified embryoid body (EB) formation method as described before[14]. Two weeks after the EB-mediated hematopoietic differentiation, the EB-derived cells were collected and dissociated with accutase treatment. The dissociated cells were cultured in a serum-free medium containing SCF (100 ng/mL), interleukin-3 (10 ng/mL), insulin-like growth factor II (40 ng/ml), erythropoietin (2 U/mL) and dexamethasone (1 μM) for erythroid cell expansion and maturation as previously described[14]. Cells were harvested on day 8 of erythroid expansion and differentiation, and the differentiated cells were analyzed by FACS analysis, cytospin and histology staining as well as RT-PCR for globin expression. Quantitative (q) RT-PCR primers and probes are from Applied Biosystems (IDs are

Hs00361131_g1 for HBG1/2, and Hs00747223_g1 for HBB). Adult hemoglobin (HbA) antibody is from Santa Cruz Biotechnology (catalog # sc-21757PE). Fetal hemoglobin (HbF) antibody is from Invitrogen (catalog # HFH-01).

Results

Specific enhancement of HR efficiency by a novel pair of HBB-ZFNs

Based on our previous studies of gene targeting in human iPSCs and ESCs[3], we attempted to correct the mutated β^s allele in SCD iPSCs. To achieve specific HR-mediated gene replacement at the *HBB* gene, but not at nearby *HBD* (δ), HBG1 (Aγ), HBG2 (Gγ) and HBE (ε) globin genes (**Fig. 1A**), we decided to employ ZFNs that make a specific double-strand-break to stimulate HR near the β^s mutation (codon 6) in the *HBB* gene. After initial screening, a pair of HBB-ZFNs was identified, which targets a bipartite 31-bp sequence (with a 6-bp spacer) in exon 1 (21-bp downstream of the β^s mutation) of the *HBB* gene (**Fig. 1A**). We first tested the activity and specificity of this newly-designed HBB-ZFN pair using a GFP reporter system[3]: DNA sequences from *HBB* or highly homologous beta-locus genes, *HBE, HBD, HBG*, were tested as candidate alternative targets of this ZFN pair (**Fig. 1A**). Successful HR between the EGIP* reporter (containing a ZFN target sequence) and a truncated GFP (tGFP) donor will restore a full-length GFP gene (**Fig. 1B**). Using flow cytometry to measure HR-mediated gene correction in stably transfected 293T cells expressing EGIP*-HBB target, we monitored the number of GFP+ cells after tGFP donor transfection, in the absence or presence of the HBB-ZFN stimulation. Without co-transfection of HBB-ZFN expression vectors, the basal gene correction rate was ~4-5 per million cell events as previously observed[3] (**Fig. 1C**). In the presence of the HBB-ZFNs, numbers of GFP+ cells jumped to 0.16% from cells expressing the EGIP* reporter that contains the *HBB* target sequence: ~350-fold stimulation (**Fig. 1C**). On the other hand, the HBB-ZFNs showed no detectable stimulatory activity on EGIP* containing the homologous target DNA from the *HBD*, *HBE* and *HBG* genes, indicating the HBB-ZFNs are highly specific (**Fig. 1D**).

Targeted gene correction of SCD mutation in the endogenous *HBB* locus

To achieve targeted correction of β^s mutation at the *HBB* locus, we designed a plasmid donor (BD2) to provide wild-type G**A**G for codon 6 (β^A) and two constitutively expressed drug-selectable genes: PGK-promoter driven hygromycin-resistant gene (PGK-Hyg) for positive selection and EF1α-promoter driven thymidine kinase-GFP fusion gene for counter-selection. The PGK-Hyg selection gene cassette to be inserted into *HBB* intron 1 was also flanked by loxP sequences for future Cre-mediated excision (**Fig. 2A**). We first evaluated the BD2 vector for gene targeting of the endogenous *HBB* gene in human 293T cells, in the presence or absence of the HBB-ZFNs delivered as two expression plasmids. After both hygromycin B and GCV selection (Hyg^R-GCV^R), pooled 293T were used in genomic PCR assays to detect if there was any cell with a targeted integration (TI) of the PGK-Hyg in the intron co-current with the SCD mutation replacement after HR by the same donor template. Two PCR primers, one within the inserted Hyg gene and another downstream of the right homology arm, were used to detect the 3'-TI (2.5-kb). Our data showed that despite the high transfection efficiency in 293T cells, the BD2 donor vector alone did not generate TI mediated by HR in the silent *HBB* locus. However, adding HBB-ZFNs greatly stimulated gene targeting at the endogenous *HBB* locus (**Fig. 2B**). Thus we validated that the BD2 donor is capable of targeting *HBB* locus, and a pair of 3'-TI primers can be used to screen targeted iPSCs after co-transfection of the BD2 donor vector and two HBB-ZFNs followed by hygromycin and GCV double selections.

We next used a SCD patient-specific iPSC line MBP5s1 (abbreviated as S1) that was previously established by a virus-free method[13]. We first adapted S1 iPSCs to single-cell passaging so that they can survive better after nucleofection. After adaption for 10 continuous passages, the resulting S1 iPSCs at p26 retained iPSC characteristics, pluripotency and a normal karyotype (see below). We then transfected S1 iPSCs cells with the BD2 donor and two HBB-ZFN plasmids by nucleofection, and selected for clones conferring hygromycin and GCV resistance. Proceeding thus, we screened 300 Hyg^R-GCV^R S1 iPSC clones from multiple HR experiments. First by using PCR for detecting the 3'-junction of TI, we identified 4 candidate clones (**Fig. 2C**). One clone, c36, was further expanded and confirmed by Southern blot analysis to bear a TI of the PGK-Hyg cassette into intron 1 on one *HBB* allele and no additional random integration (**Fig. 2D-E**). The other 3 clones either failed to proliferate or could not be confirmed positive by Southern blot (data not shown). DNA sequencing of c36 genomic DNA surrounding the *HBB* exon 1 also confirmed a G**T**G to G**A**G correction in the targeted allele (see below).

Although the inclusion of the PGK-Hyg selection gene cassette in intron 1 enables selection of rare HR clones critical to target genes such as *HBB* that are silent in undifferentiated iPSCs and ESCs, it may potentially interfere with expression of the corrected *HBB* allele[10]. To alleviate this potential problem, we next excised the *loxP*-flanked PGK-Hyg cassette using transient transfection of a Cre-expressing plasmid. Four out of 24 PCR-screened clones were shown to be negative of 3'-TI (**Fig. 2F**), and 3 of those (cre4, cre16 and cre19) were selected and confirmed by Southern blot to be free of the integrated PGK-Hyg gene cassette (**Fig. 2G-H**). Sequencing of genomic DNA PCR products using primers in exon 1 and 2 further confirmed the presence of one unmodified β^s allele and one corrected (β^A) allele that also bears the leftover *loxP* site (46 bp including linker DNA) in intron 1, as expected (**Fig. 3**).

Through two steps of genetic manipulation, we obtained multiple human iPSC clones (c36, cre4, cre16, and cre19) with precise gene correction of the SCD mutation. These clones showed human ESC-like morphology, maintained normal karyotypes and pluripotency under ESC/iPSC culture conditions (**Fig. 4**), similar the parental s1 iPSCs. Our genome-wide SNP concordance analysis did not detect overt difference between the c36 (after HR) and its S1 parental iPSC line (**Materials and Methods**), although a detailed copy number variation or whole-genome sequencing analysis may provide more information about subtle genetic changes.

Differentiation of patient-derived and gene-corrected SCD iPSCs into erythroid cells

Since the *HBB* gene is not expressed in the undifferentiated iPSCs, we used differentiated iPSCs to evaluate the restoration of normal beta-globin expression. Human iPSC lines S1, c36, cre4 and cre16 were differentiated into erythroid cells using a two-step serum-free differentiation protocol (**Fig. 5A**). During the first step of embryoid body (EB) formation, all the iPSC lines form cystic EBs efficiently. At the end of the second step (erythroid expansion and maturation for 8 days), the differentiated cells expressed CD71+ (94-98%) and CD235a (Glycophrin A, 24-29%) at similar levels (**Fig. 5B**), and comparable to normal iPSC lines we previously tested[14]. Giemsa stain confirmed most of the cells in culture were nucleated erythroblasts (**Fig. 5C**).

Beta-globin expression in SCD iPSCs after erythroid differentiation

We first used qRT-PCR to measure the total expression levels of the *HBB* gene (the β^s and a corrected allele) as well as the (fetal-type) *HBG* genes in all SCD iPSCs (S1, cre4 and cre16), differentiated iPSCs at EB stage and at erythroblast stage (**Fig. 6A**). While the level of *HBG* expression in the SCD iPSC-derived erythroblasts (S1-EryB, cre4-EryB, cre16-EryB) is comparable to that in cord blood mononuclear cells (CB-MNC), the level of *HBB* expression in the erythroblasts is >100-fold lower than that in CB-MNC even though they are ~100-fold higher than those in the EBs (**Fig. 6A**). Using a pair of primers located in exon 1 and 3 (blue arrows in **Fig. 2A**), we amplified the *HBB* transcripts from the erythroblasts of gene-corrected SCD iPSCs (c36, cre4 and cre16) by conventional RT-PCR (**Fig. 6B**). Then we cloned each PCR product which contains mixed DNA of corrected/wild-type and uncorrected/mutant alleles into TOPO vector and sequenced individual clones. This will tell us the frequencies of both gene-corrected (G**A**G) and uncorrected β^{s} (G**T**G) transcripts in erythrocytes from each iPSC line. We found that only the uncorrected β^{s} (G**T**G) allele was expressed in c36 derived cells as no sequenced *HBB* cDNA contained corrected β^{A} (G**A**G). In contrast, erythroid cells from both cre4 and cre16 iPSC clones (that had undergone Cre-mediated excision of the selection cassette) expressed the corrected G**A**G *HBB* allele, although at the level 25% to 40% compared to the uncorrected (and mutated) allele (**Fig. 6C**). Our flow cytometry assay using antibodies for adult or fetal hemoglobin proteins confirmed the vast majority of hemoglobin expression in the erythroblasts is fetal-type (HbF), consistent with our qRT-PCR results (**Fig. 6D**). Our result is also consistent with two recent reports on erythropoiesis of human iPSCs[17,18], indicating that a better culture system is needed to increase erythrocyte maturation, and survival, increase the *HBB* gene expression, and improve translation from existing *HBB* transcripts.

Investigation of repressed gene expression of the gene-targeted *HBB* allele

Our data showed that although *HBB* transcripts can be detected at a low level in the iPSC-derived erythroid cells, expression of gene-targeting corrected β^{A} allele was somehow totally or partially lost in differentiated erythroblasts at the same stage. We then took several approaches to investigate the potential epigenetic or genetic factors that may be responsible for the repressed β^A expression in c36, cre4 and cre16 cells. For c36, there was a 2.2-kb loxP-flanked PGK-Hyg gene cassette inserted in the first exon, excision of which resulted in partial expression of the gene-targeted allele (**Fig 6C**). The PGK-hyg insertion at the first intron (37-bp away from the exon-intron junction) may affect the *HBB* gene expression in two ways. 1). The PGK-Hyg (intron-less) gene cassette in the same orientation may introduce a cryptic/alternative splicing acceptor. We used three bioinformatics programs (NNSPLICE: http://www.fruitfly.org/seq_tools/splice.html, ASSP: http://www.es.embnet.org/~mwang/assp.html and NetGene2: http://www.cbs.dtu.dk/services/NetGene2/) to predict potential cryptic/alternative splicing acceptors in the insertion fragment. We identified several candidates and performed RT-PCR analysis, but did not detect any alternative transcript around the region (data not shown). 2). The PGK promoter may make an antisense transcript that blocks HBB expression. However, no RT-PCR products can be found using several primers covering both strands of *HBB* exon 1 and the hyg gene sequences. The silencing of the Hyg gene expression (driven by the PGK promoter) in expanded iPSCs is a surprise, but consistent with the observation that c36 iPSCs lost their resistance to hygromycin selection (that was used in the initial clone screening) after expansion in the absence of selection. When we compared the hygromycin expression levels among several iPSCs where the same PGK-Hyg gene cassette is inserted either in the silent *HBB* gene (such as c36) or in the constitutively active *PIG-A* gene (such as FPHR iPSCs we published previously[3]), we found both expanded c36-iPSCs and c36-EryB expressed much lower hygromycin transgene expression than FPHR gene-targeted iPSCs (**Fig. 7A**). This suggests epigenetic factors may play a role in silencing the gene-targeted allele including both the inserted PGK-Hyg transgene cassette and adjacent *HBB* gene.

Even though we had expected the insertion of drug-selection cassette might interfere with the transcription of gene-targeted allele, we were surprised to see partially repressed β^A expression in the cre4 and cre16 after Cre-mediated excision (**Fig. 5C**). One possibility is that the remnant loxP sequence that is 37-bp away from the exon-intron junction interferes with the *HBB* gene expression. To examine other possibilities, we sequenced 3.5-kb sequence covering the whole *HBB* gene and its flanking genomic regions of all three gene-targeted lines (c36, cre4 and cre16) that had T>A replacement in exon 1 (left panels in **Fig. 7B**). Two single nucleotide variants that differ from the parental S1 sequence were found in the targeted allele of all three iPSC lines. The first, located downstream of exon 3 in the gene-targeted *HBB* allele (middle panels), is a polymorphic nucleotide change (G>T) in the 3' homology arm and likely introduced by the gene-targeting vector containing non-isogenic sequence. This SNP, however, has not been associated with any *cis*-regulatory elements or SNPs affecting *HBB* expression. The second one (A>G) is found only in the targeted allele located downstream of the 3'-end of the right homology arm (right panels in **Fig. 7B**), therefore was not contained in the gene targeting donor. This A>G mutation modified the core GATA binding site in the 3'-enhancer of the *HBB* gene expression[19]. Deletion of the GATA-containing 3'-enhancer element has been shown to reduce *HBB* expression in transgenic mice[20,21]. Since this mutation was only found on the targeted allele of gene-corrected SCD iPSCs (**Fig. 7B**), it was likely introduced during the HR-mediated repair and passed from c36 through cre4 and cre16. At present, it is unclear whether this mutation in the GATA-containing 3' enhancer, or the remaining loxP site in the first intron, or both, is responsible for the reduced expression from the targeted *HBB* allele.

Discussion

Various methods have been explored to increase gene-targeting efficiencies in human iPSCs and ESCs, with or without the use of ZFNs. Recombinant adenovirus-associated viruses (rAAVs)[5,6], Helper-dependent adenoviruses (HDAdV)[22], and BAC vectors[23] have been used in gene targeting/correction without ZFNs in several cell types including human ESCs and iPSCs. However, these previous studies have been focusing on constitutively transcribed genes in human iPSCs. Regardless if the underlying target allele is expressed or silent in undifferentiated iPSCs, the method we described here should apply to correction or replacement in human iPSCs of any specific mutation or even a SNP locus. During the revision of this report, two other studies reported the correction of single mutations in human iPSCs[24,25]. Together with our study reported here on gene correction of the SCD mutation, these papers show the feasibility regardless the underlying gene is expressed or silent in human iPSCs.

Numerous studies have reported that homology-directed repair is often much less efficient when a silent gene is targeted[26,27]. This also explains our observed low efficiency of targeting the *HBB* gene even in the presence of a pair of ZFNs, as compared to that of a constitutive gene (*PIG-A*) we previously published[3]. The HBB-ZFNs used here is 50% less efficient than the *PIG-A* ZFNs measured by the similar GFP* reporter assays (**Fig. 1B**), which could also contribute to a lower gene targeting of the *HBB* gene in human iPSCs. With an improved ZFN technology[28] or other emerging methods such as TALEN[29], we will likely find a high-affinity ZFN or TALEN at or near the SCD mutation in *HBB*, without cutting the nearby *HBD*, *HBG* and *HBE* genes.

For early-passage human iPSCs that typically have <1% colony-forming efficiency, successful gene targeting of a silent gene such as HBB in human iPSCs relies more on a co-integrated drug-selection gene driven by a constitutively active promoter. Even with a floxed drug-resistant gene, gene targeting of a silent gene may still face obstacles due to silencing of the targeted locus in undifferentiated iPSCs. Indeed, we noticed that the HR-targeted c36 clone (c36) obtained after 2-weeks of Hyg selection became sensitive again to hygromycin upon subsequent culture. Our RT-PCR results confirmed very low expression of PGK-Hyg transgene cassette in both undifferentiated c36 iPSCs and differentiated erythroid cells (**Fig. 7A**). We hypothesized that epigenetic mechanisms such as DNA methylation in pluripotent cells may cause expression silencing of a transgene inserted at unfavorable loci, which in turn induce repression in gene expression of nearby targeted endogenous gene such as *HBB* even after differentiation. Similar reduction of gene-targeted allele expression was also observed in the previous report of correcting a humanized SCD mouse ESCs harboring the human β^S allele[10]. However, the transgene expression cassette inserted into an intron can be easily excised out via the Cre-loxP system, and the endogenous gene expression can be largely restored after excision.

Although in most applications we do not need to completely restore expression level of the corrected allele to the same level of the uncorrected alleles, it is better for us to understand why only a partial restoration (25% to 40%) of the corrected allele in cre4 and cre16 iPSCs after the loxP excision (**Fig. 6B**). Currently, we could not distinguish the following two possibilities and more studies are needed. 1). The presence of one copy of loxP left in the first intron interfered with the endogenous *HBB* gene expression. 2). A genetic mutation in the GATA site of *HBB* 3'-enhancer (likely introduced in iPSC HR and/or clonal selection process) is responsible for the reduction of erythroid-specific

HBB expression. For the former possibility, we may use *piggyBac* DNA transposition to achieve zero-footprint excision of the selection gene cassette[30]. For the latter possibility, our study highlights the importance to conduct high-resolution whole-genome sequence analysis. Although the HR-gene targeting did not show increased mutation rates as compared to other proliferating cell populations in studies published by us and others[3, 4, 24], previous employed methods of chromosomal karyotyping, SNP arrays and even exome sequencing may not be sufficient to detect small changes across the whole genome. With recent cost reduction of whole-genome sequencing that is now affordable to an average lab, it is now possible to sequence the whole genome of selected iPSC clones after ZFN-mediated gene targeting.

Although in this study we focused on site-specific gene correction of a defined single mutation in the *HBB* gene, our functional analysis and relevance to future gene therapy is hampered by the fact that existing culture methods only allow us to generate nucleated erythrocytes that express gamma-globin instead of beta-globin proteins. Nevertheless, *HBB* transcripts can be detected, albeit at a low level, by RT-PCR in differentiated erythroblasts derived from parental S1 and corrected cre iPSC lines (**Fig. 6**). This allowed us to show that the corrected *HBB* allele after gene targeting can be expressed. An ideal iPSC-based gene therapy for SCD will require both precise correction of disease-causing mutation while preserve *cis*-regulatory elements regulating *HBB* expression, and complete switching from fetal-type globin to adult-type globin. Therefore, further genetic and epigenetic studies are required to better understand globin gene switching in the iPSC system and to improve hematopoietic differentiation into more mature/enucleated erythroid cells from gene-corrected iPSCs, if they are going to be used for cell therapy. In addition, reproducible in vivo models to generate transplantable hematopoietic progenitor cells and red blood cells from human iPSCs and

ESCs are needed[31,32], before we can assess the pre-clinical feasibility of the combined cell and gene therapy approach using gene corrected iPSCs derived from SCD patients. Despite these hurdles to be overcome in future years, our current study demonstrates that it is now possible to correct a specific point or small mutation at its endogenous locus in patient-specific iPSCs, and also to restore gene expression of the corrected form upon cell differentiation. The gene targeting methodology we reported here would extend the use of various iPSC lines containing disease related mutations in genes that are either expressed or silent in human iPSCs.

Acknowledgement

We thank the laboratory of Prof. YW Kan for providing the PGK-Cre expression vector, a Sigma-Aldrich team for designing and providing us with plasmids expressing HBB-ZFNs. This research is supported by a Siebel Scholarship to PM, by Maryland Stem Cell Research Fund for a fellowship and grant to JZ (2008-MSCRFF-009 and 2010-MSCRFE-0044) and a fellowship to XH (2010-MSCRFF-0095), and by NIH grants (R01 HL073781 and HL073781-05S1 to LC).

Authorship Contributions and Disclosure of Conflicts

Contribution: JZ and PM designed and performed experiments, collected and analyzed data, and wrote the manuscript. XH and SND performed experiments, collected and analyzed data. LC designed the experiments and wrote the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Linzhao Cheng, The Johns Hopkins University School of Medicine, Broadway Research Building, Room 747, 733 N. Broadway, Baltimore, MD 21205; email: lcheng2@jhmi.edu; or Jizhong Zou, The Johns Hopkins University School of

Medicine, Broadway Research Building, Room 780, 733 N. Broadway, Baltimore, MD

21205; email: jzou2@jhmi.edu.

References

1. Saha K, Jaenisch R. Technical challenges in using human induced pluripotent stem cells to model disease. *Cell Stem Cell*. 2009;5(6):584-595.
2. Raya A, Rodriguez-Piza I, Guenechea G, et al. Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature*. 2009;460(7251):53-59.
3. Zou J, Maeder ML, Mali P, et al. Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell*. 2009;5(1):97-110.
4. Hockemeyer D, Soldner F, Beard C, et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol*. 2009;27(9):851-857.
5. Mitsui K, Suzuki K, Aizawa E, et al. Gene targeting in human pluripotent stem cells with adeno-associated virus vectors. *Biochem Biophys Res Commun*. 2009;388(4):711-717.
6. Khan IF, Hirata RK, Wang PR, et al. Engineering of human pluripotent stem cells by AAV-mediated gene targeting. *Mol Ther*. 2010;18(6):1192-1199.
7. Zou J, Sweeney CL, Chou BK, et al. Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. *Blood*. 2011;117(21):5561-5572.
8. Orkin SH, Higgs DR. Medicine. Sickle cell disease at 100 years. *Science*. 2010;329(5989):291-292.
9. Townes TM. Gene replacement therapy for sickle cell disease and other blood disorders. *Hematology Am Soc Hematol Educ Program*. 2008:193-196.
10. Chang JC, Ye L, Kan YW. Correction of the sickle cell mutation in embryonic stem cells. *Proc Natl Acad Sci U S A*. 2006;103(4):1036-1040.
11. Wu LC, Sun CW, Ryan TM, Pawlik KM, Ren J, Townes TM. Correction of sickle cell disease by homologous recombination in embryonic stem cells. *Blood*. 2006;108(4):1183-1188.
12. Hanna J, Wernig M, Markoulaki S, et al. Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science*. 2007;318(5858):1920-1923.
13. Mali P, Chou BK, Yen J, et al. Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem Cells*. 2010;28(4):713-720.
14. Ye Z, Zhan H, Mali P, et al. Human-induced pluripotent stem cells from blood cells of healthy donors and patients with acquired blood disorders. *Blood*. 2009;114(27):5473-5480.
15. Mali P, Ye Z, Hommond HH, et al. Improved efficiency and pace of generating induced pluripotent stem cells from human adult and fetal fibroblasts. *Stem Cells*. 2008;26(8):1998-2005.
16. Pomper MG, Hammond H, Yu X, et al. Serial imaging of human embryonic stem-cell engraftment and teratoma formation in live mouse models. *Cell Res*. 2009;19(3):370-379.
17. Chang KH, Huang A, Hirata RK, Wang PR, Russell DW, Papayannopoulou T. Globin phenotype of erythroid cells derived from human induced pluripotent stem cells. *Blood*. 2010;115(12):2553-2554.
18. Lapillonne H, Kobari L, Mazurier C, et al. Red blood cell generation from human induced pluripotent stem cells: perspectives for transfusion medicine. *Haematologica*. 2010;95(10):1651-1659.

19. Orkin SH. GATA-binding transcription factors in hematopoietic cells. *Blood*. 1992;80(3):575-581.
20. Liu Q, Bungert J, Engel JD. Mutation of gene-proximal regulatory elements disrupts human epsilon-, gamma-, and beta-globin expression in yeast artificial chromosome transgenic mice. *Proc Natl Acad Sci U S A*. 1997;94(1):169-174.
21. Rubin JE, Pasceri P, Wu X, Leboulch P, Ellis J. Locus control region activity by 5'HS3 requires a functional interaction with beta-globin gene regulatory elements: expression of novel beta/gamma-globin hybrid transgenes. *Blood*. 2000;95(10):3242-3249.
22. Suzuki K, Mitsui K, Aizawa E, et al. Highly efficient transient gene expression and gene targeting in primate embryonic stem cells with helper-dependent adenoviral vectors. *Proc Natl Acad Sci U S A*. 2008;105(37):13781-13786.
23. Song H, Chung SK, Xu Y. Modeling disease in human ESCs using an efficient BAC-based homologous recombination system. *Cell Stem Cell*. 2010;6(1):80-89.
24. Howden SE, Gore A, Li Z, et al. Genetic correction and analysis of induced pluripotent stem cells from a patient with gyrate atrophy. *Proc Natl Acad Sci U S A*. 2011;108(16):6537-6542.
25. Liu GH, Suzuki K, Qu J, et al. Targeted Gene Correction of Laminopathy-Associated LMNA Mutations in Patient-Specific iPSCs. *Cell Stem Cell*. 2011;8(6):688-694.
26. Kuroiwa Y, Kasinathan P, Matsushita H, et al. Sequential targeting of the genes encoding immunoglobulin-mu and prion protein in cattle. *Nat Genet*. 2004;36(7):775-780.
27. Smithies O, Gregg RG, Boggs SS, Koralewski MA, Kucherlapati RS. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature*. 1985;317(6034):230-234.
28. Sander JD, Dahlborg EJ, Goodwin MJ, et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods*. 2011;8(1):67-69.
29. Miller JC, Tan S, Qiao G, et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*. 2011;29(2):143-148.
30. Yusa K, Zhou L, Li MA, Bradley A, Craig NL. A hyperactive piggyBac transposase for mammalian applications. *Proc Natl Acad Sci U S A*. 2011;108(4):1531-1536.
31. Kaufman DS. Toward clinical therapies using hematopoietic cells derived from human pluripotent stem cells. *Blood*. 2009;114(17):3513-3523.
32. Lengerke C, Daley GQ. Autologous blood cell therapies from pluripotent stem cells. *Blood Rev*. 2010;24(1):27-37.

Figure Legends

Figure 1. The activities and specificities of HBB-ZFNs that stimulate gene targeting in a GFP reporter assay. (**A**) The putatitive recognition sequence of HBB-ZFNs in the *HBB* gene (5' to 3', starting from the first codon). The left (12-bp) and left (19-bp) ZFN sites are underlined. The homologous sequences from other beta-locus genes ("HBE", "HBD", and "HBG1/2") and their differences from the HBB gene in L-ZFN, R-ZFN and spacer regions (# of mismatches) are also shown. Each was inserted into the GFP reporter as ZFN target sequence to test the specificity of HBB-ZFNs. All the inserts start with a STOP codon (red "taa") and end with a HindIII site (blue "aagctt"). A short version of the HBB target sequence (called HBB-Short) was also tested. (**B**) Schematic of the GFP* reporter rescue assay. An EGIP* mutant was created by inserting a ZFN target sequence including a STOP codon and HindIII site into the GFP gene. Only after gene targeting with a tGFP donor (with or without ZFNs), the EGIP* will be corrected to restore wild-type GFP expression. (C) Flow cytomtric analysis of GFP correction after HR in 293T cells stably transfected with EGIP*-HBB reporter. Two days after transient transfection of tGFP donor alone or with HBB-ZFNs, numbers of GFP+ cells were measured by dot plot of one million collected cell events. (**C**) Gene correction efficiency of EGIP* mutants with "HBB", "HBE", "HBD", "HBG" or "HBB-Short" ZFN target sequence using the tGFP donor, with or without HBB-ZFNs. Numbers of GFP+ cells per 10^6 293T-EGIP* cells are plotted as mean ± SEM, n=3.

Figure 2. Site-specific gene correction of the β^s mutation in the *HBB* gene. (**A**). A scheme of stepwise gene correction of one mutant β^s allele (shown as 3 exons, two inrons and flanking sequences), first by HR-mediated gene targeting and followed by

Cre-mediated excision. The gene-targeting donor BD2 vector with two homology arms (5.9-kb left arm and 2-kb right arm indicated by "X") introduces a HR template for T to A replacement in the *β*^{s} allele and a *loxP*-flanked drug-selection cassette PGK-Hygromycin (Hyg) to be inserted into the *HBB* intron 1. The flanking counter-selection HSV-TK gene (in the form of a TK.GFP fusion) driven by the EF1α promoter (outside the right HR homology arm) is used to reduce the frequency of Hyg resistant clones due to BD2 vector random integration that also allows HSV-TK expression. For validated HR clones, Cre-mediated excision removes only the PGK-Hyg selection gene cassette and leaves one copy of the *loxP* DNA in the middle of *HBB* intron 1 of the corrected allele, generating "cre" clones with one corrected allele *β*^{A} (Corrected$^{\Delta loxP}$). We used two PCR primers (red arrows) for initial screening of targeted insertion (TI) indicative of correct HR. (**B**) The initial results of gene targeting in 293T cells that were transfected with the 1µg BD2 donor alone (lane 1) or with HBB-ZFNs at increasing amounts (0.25µg in lane 2, 1µg in lane 3, and 2µg in lane 4). The 3'-TI event (2.5-kb PCR product) was detected when ZFNs were present. Untransfected 293T cells in lane 5 is negative. (**C**): Similar results of in selected clones after gene targeting in S1 iPSCs and Hyg and GVC selection. Four clones (c36, c64, c68 and c70) showed a positive PCR product. Positive control (p.c.) was DNA from targeted 293T cells (lane 3) in (**B**). (**D**) Southern blot analysis surrounding the *HBB* locus in the parental S1 and selected targeted iPSC clones. A 3' probe downstream of 3' homology arm (top line) is used with genomic DNA digestion by PmeI (P) and EcoRV (E) enzymes to confirm the presence of targeted allele (with EcoRV site inside TI) shown as a green arrow, and the original *HBB* allele in a red arrow. (**E**): A Hyg probe is used with genomic DNA digestion by XbaI (X) and SpeI (S) enzymes to confirm the targeted allele (with hyg insertion, green arrow) or identify random integration events such as one in clone 65. (**F**): PCR screening for clones with successful Cre excision, using the two primers shown in red. The absence of the hyg-containing DNA in clones such as cre16 and cre19 indicates the excision of the PGK-Kyg cassette. (**G**) and (**H**): Southern blot analyses of iPSC clones before and after Cre excision. (**G**): Results with the 3' probe after P and E digestion as shown in (**D**). Red arrows indicated 4.3-kb fragments from β^{s} allele in S1 and c36 iPSC clones, and β^{A} allele in cre clones (4, 16 and 19). (**H**): Sothern blots with Hyg probe after X and S digestion as shown in (**E**). The S1 and 3 cre clones are free of the hyg gene, which was found in c36 clone as expected (green arrow).

Figure 3. Genomic DNA PCR and sequencing confirm precise monoallelic gene correction and Cre-LoxP excision. (**A**) Schematic of genomic DNA PCR using primers in 5'-UTR of exon 1 and exon 2 of *HBB*. With a short 72ºC extension step (15''), only the mutant allele (291-bp) and the corrected allele after excision (337-bp) can be amplified. (**B**) DNA gel shows one band for the c36 clone representing the mutant allele that can be amplified by the PCR protocol, and two bands from every cre clone representing both mutant allele and corrected allele after excision. (**C**) The mixed PCR products from cre4 iPSCs (after gene correction and Cre-LoxP excision) were cloned into a TOPO vector and individual clones were sequenced. Among 8 sequenced clones, 4 (50%) were shown to contain a mutant allele (upper panel), and 4 (50%) bore the corrected allele with a remnant *loxP* site after excision (bottom panel).

Figure. 4. Characterization of gene-corrected SCD iPSC clones. Gene corrected SCD iPSC clones (c36 and cre4 shown here) display characteristic pluripotency markers such as alkaline phosphatase (AP), OCT4, NANOG, and TRA-1-60 (**A**); maintain normal karyotypes (**B**); and form teratomas bearing cells from all three germ layers, i.e., ectoderm, mesoderm and endoderm (**C**).

Figure. 5. Erythroid differentiation of SCD iPSC clones. (**A**) In vitro hematopoietic differentiation of various iPSC clones to generate HBB-expressing erythroid cells. After 14 days of embryoid body (EB)-mediated spontaneous differentiation, iPSC-derived cells were further differentiated and expanded into immature erythroid cells (erythroblasts or EryB) for another 8 days. Then the erythroid cells were collected for flow cytometry, Giemsa stain of cytospin, and RT-PCR or qRT-PCR. (**B**) Flow cytometric analysis of S1, cre4, and cre16 using erythroid-specific surface markers CD71 and CD235a (GlycophorinA). (**C**) Giemsa stain of confirming that most of differentiated cells are erythroblasts.

Figure 6. HBB and HBG transcription and translation analyses. (**A**) The *HBB* and *HBG1/2* gene expression (normalized to a house-keeping gene GAPDH) in undifferentiated iPSCs (S1) and differentiated progenies (EB and EryB) of S1, cre4 or cre16 was measured by quantitative RT-PCR (data represent mean±SEM, N=3). After the erythroid differentiation from iPSCs, the *HBB* transcript level increased 10- to100-fold, although it is still 100- to 1000-fold lower compared to the level in cord blood mononuclear cells (CB-MNC). (**B**). Conventional RT-PCR that readily amplifies *HBB* cDNA in erythroblasts derived from various SCD iPSC clones before (S1) or after gene targeting (c36) and Cre-mediated excision (cre4 and cre16), by two primers located at exon 1 and 3 (left illustration). (**C**) Although sizes of RT-PCR products of the unmodified or corrected alleles are the same, DNA sequencing of the RT-PCR product showed uniform transcript in c36-EryB and mixed transcripts in cre4-EryB and cre16-EryB (lower chromatographs). Cloning each transcripts into TOPO vector and sequencing at clonal levels will distinguish expression from corrected versus uncorrected alleles. Sequencing

of 40-60 individual cloned DNA molecules of RT-PCR products from each differentiated iPSC line revealed that the absence of corrected *HBB* transcript (T: 100% or 40/40 cloned and sequenced) in c36-derived erythroblasts, but in the erythroblasts derived from cre4 and 16 after after Cre-mediated excision of the PGK-Hyg gene cassette, expression of the corrected (A) allele was detected. In cre4, both corrected (A, 28% or 17/60) and the unmodified (and mutated T) *HBB* alleles (72% or 43/60) were expressed. A similar result was obtained in cre16 iPSCs: 12/60 (or 20%) and 48/60 (80%) of the cloned and sequenced transcripts are from the corrected (A) and the uncorrected (T) alleles, respectively. (**D**) S1-EryB, cre4-EryB, and cre16-EryB expressed abundant fetal-type hemoglobin HbF, but no detectable adult-type hemoglobin HbA measured by flow cytometry using specific antibodies.

Figure 7. Investigation of reduced expression of the gene-targeted allele. (**A**) RT-PCR showed the same PGK-Hyg transgene was expressed at very low level in gene-targeted c36-iPSCs (lane 3) and c36-EryB (lane 4) compared to another iPSC line (FPHR, where the PGK-Hyg cassette was targeted into the actively expressed *PIG-A* gene[3], lane 1). Non-targeted S1-EryB cells were used as negative control (lane 2). (**B**) Sequencing of 3.5-kb genomic region of *HBB* locus in SCD iPSC clones. The 3.5-kb genomic region includes ~1.1-kb promoter, all the *HBB* exons and introns, and ~0.5kb downstream sequence that were part of BD2 targeting donor (black lines/shapes/names); and also contains ~200-bp 3'-enhancer sequence (grey lines/shapes/names) downstream of right homology arm. DNA sequencing of both alleles in early (p24) of late (p56) passage of S1 revealed uniform β^S mutation in exon 1 and wild-type GATA site in 3'-enhancer (underlined with complementary strand sequence underneath). However, in c36, cre4 or cre16, sequencing of mixed alleles showed heterozygous nucleotides ("N"), including β^A in exon 1, a G to T polymorphism

near the 3'-end of the right homology arm, and an A to G mutation in GATA site all linked on gene-targeted allele.

Fig. 1

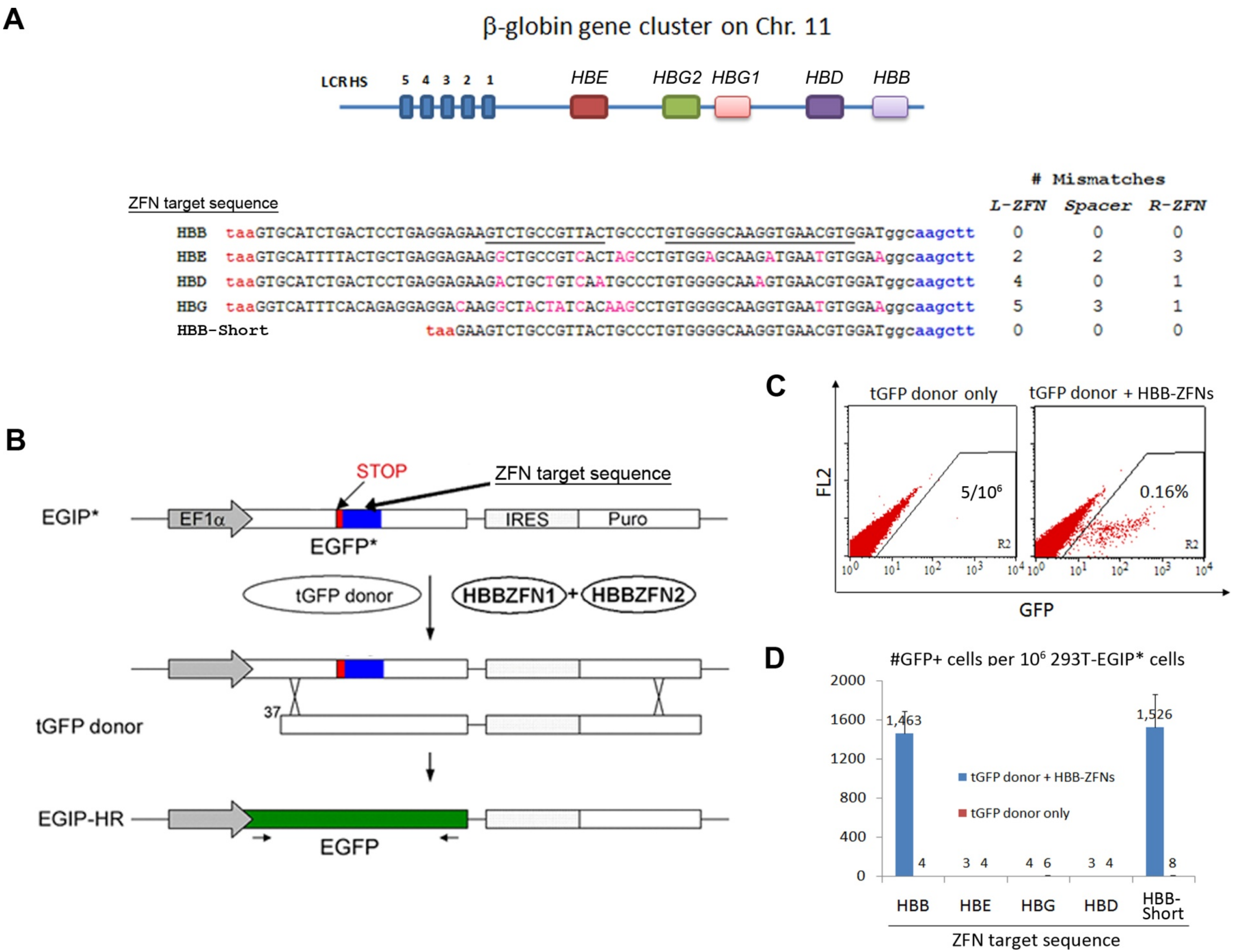

A
β-globin gene cluster on Chr. 11
LCR HS
5 4 3 2 1
HBE
HBG2 HBG1
HBD
HBB
ZFN target sequence
Mismatches
L-ZFN Spacer R-ZFN
HBB taaGTGCATCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATggcaagctt 0 0 0
HBE taaGTGCATTTTACTGCTGAGGAGAAGGCTGCCGTCACTAGCCTGTGGAGCAAGATGAATGTGGAAggcaagctt 2 2 3
HBD taaGTGCATCTGACTCCTGAGGAGAAGACTGCTGTCAATGCCCTGTGGGGCAAAGTGAACGTGGATggcaagctt 4 0 1
HBG taaGGTCATTTCACAGAGGAGGACAAGGCTACTATCACAAGCCTGTGGGGCAAGGTGAATGTGGAAggcaagctt 5 3 1
HBB-Short taaGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATggcaagctt 0 0 0
B
STOP
ZFN target sequence
EGIP*
EF1α
IRES
Puro
EGFP*
tGFP donor
HBBZFN1 + HBBZFN2
37
tGFP donor
EGIP-HR
EGFP
C
tGFP donor only
tGFP donor + HBB-ZFNs
FL2
5/10^6
0.16%
R2
GFP
D
#GFP+ cells per 10^6 293T-EGIP* cells
2000
1600
1200
800
400
0
1,463
1,526
tGFP donor + HBB-ZFNs
tGFP donor only
4
3 4
4 6
3 4
8
HBB
HBE
HBG
HBD
HBB-Short
ZFN target sequence

Fig. 2

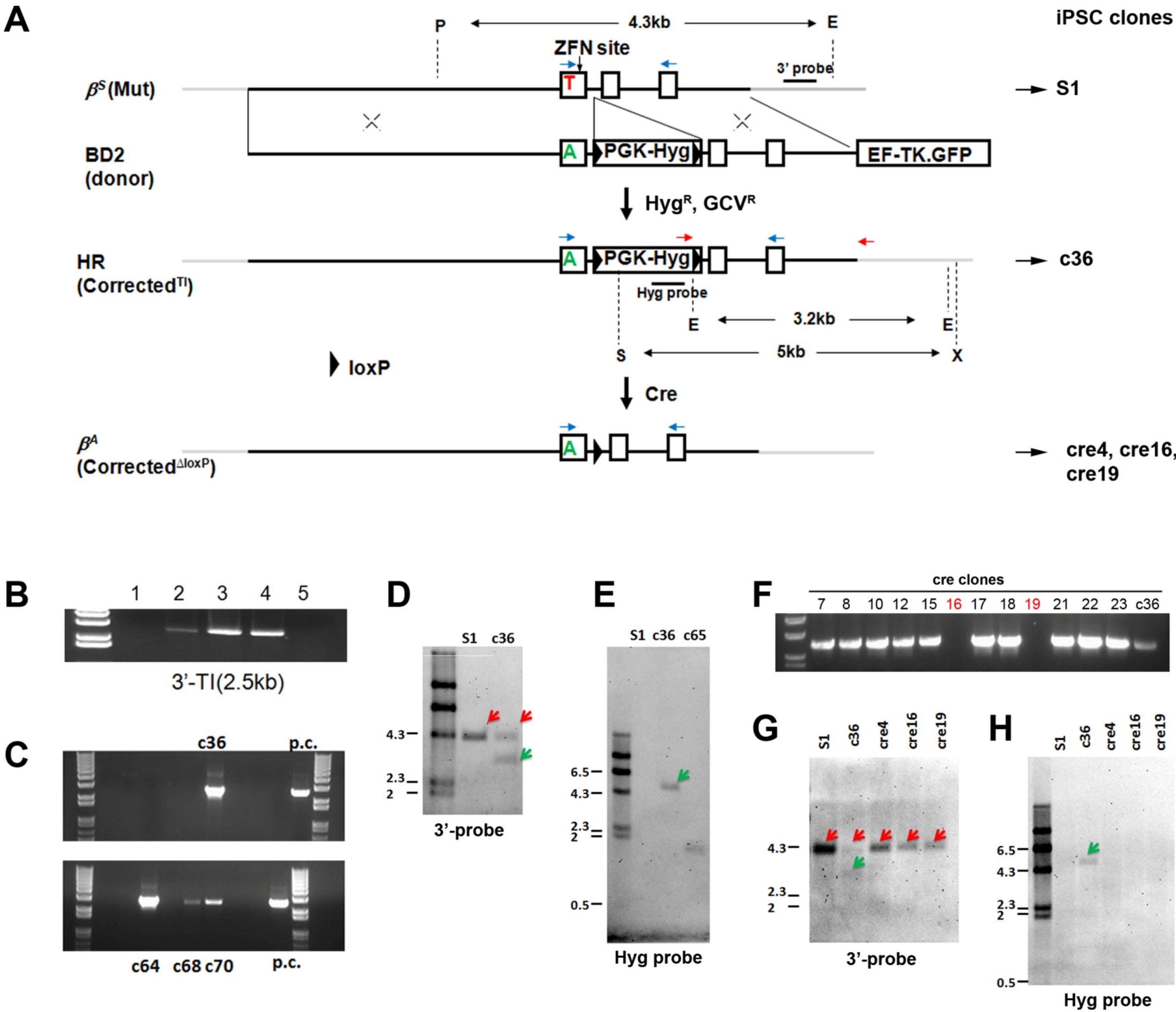
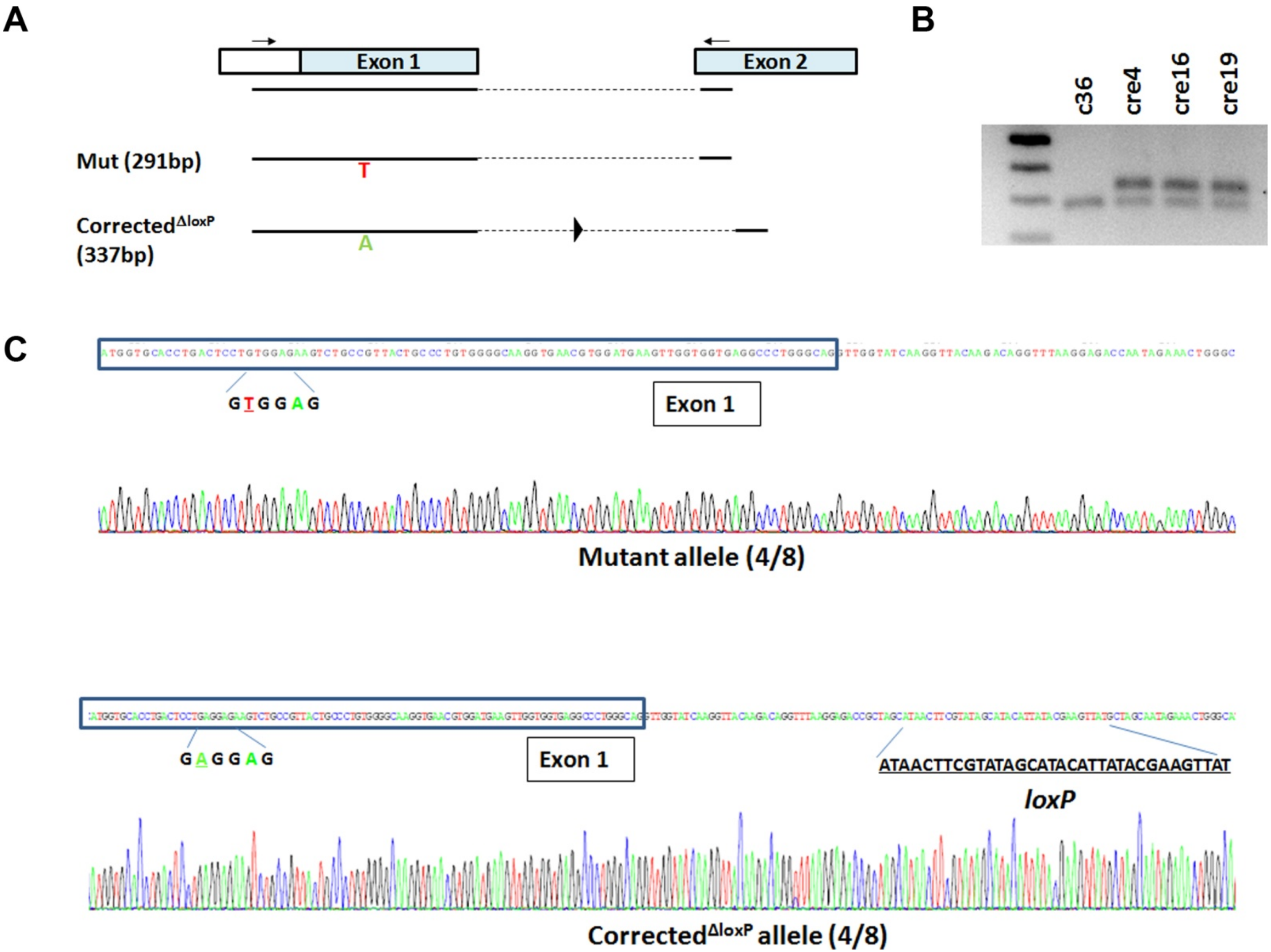

A
iPSC clones
P
4.3kb
E
ZFN site
3' probe
β^S (Mut)
S1
BD2
(donor)
PGK-Hyg
EF-TK.GFP
Hyg^R, GCV^R
HR
(Corrected^{TI})
c36
Hyg probe
E
3.2kb
E
S
5kb
X
loxP
Cre
β^A
(Corrected$^{\Delta loxP}$)
cre4, cre16,
cre19
B
1 2 3 4 5
3'-TI(2.5kb)
C
c36
p.c.
c64 c68 c70
p.c.
D
S1 c36
4.3
2.3
2
3'-probe
E
S1 c36 c65
6.5
4.3
2.3
2
0.5
Hyg probe
F
cre clones
7 8 10 12 15 16 17 18 19 21 22 23 c36
G
S1 c36 cre4 cre16 cre19
4.3
2.3
2
3'-probe
H
S1 c36 cre4 cre16 cre19
6.5
4.3
2.3
2
0.5
Hyg probe

Fig. 3



A
Exon 1
Exon 2
Mut (291bp)
T
Corrected^ΔloxP (337bp)
A
B
c36
cre4
cre16
cre19
C
G T G G A G
Exon 1
Mutant allele (4/8)
G A G G A G
Exon 1
ATAACTTCGTATAGCATACATTATACGAAGTTAT
loxP
Corrected^ΔloxP allele (4/8)

Fig. 4

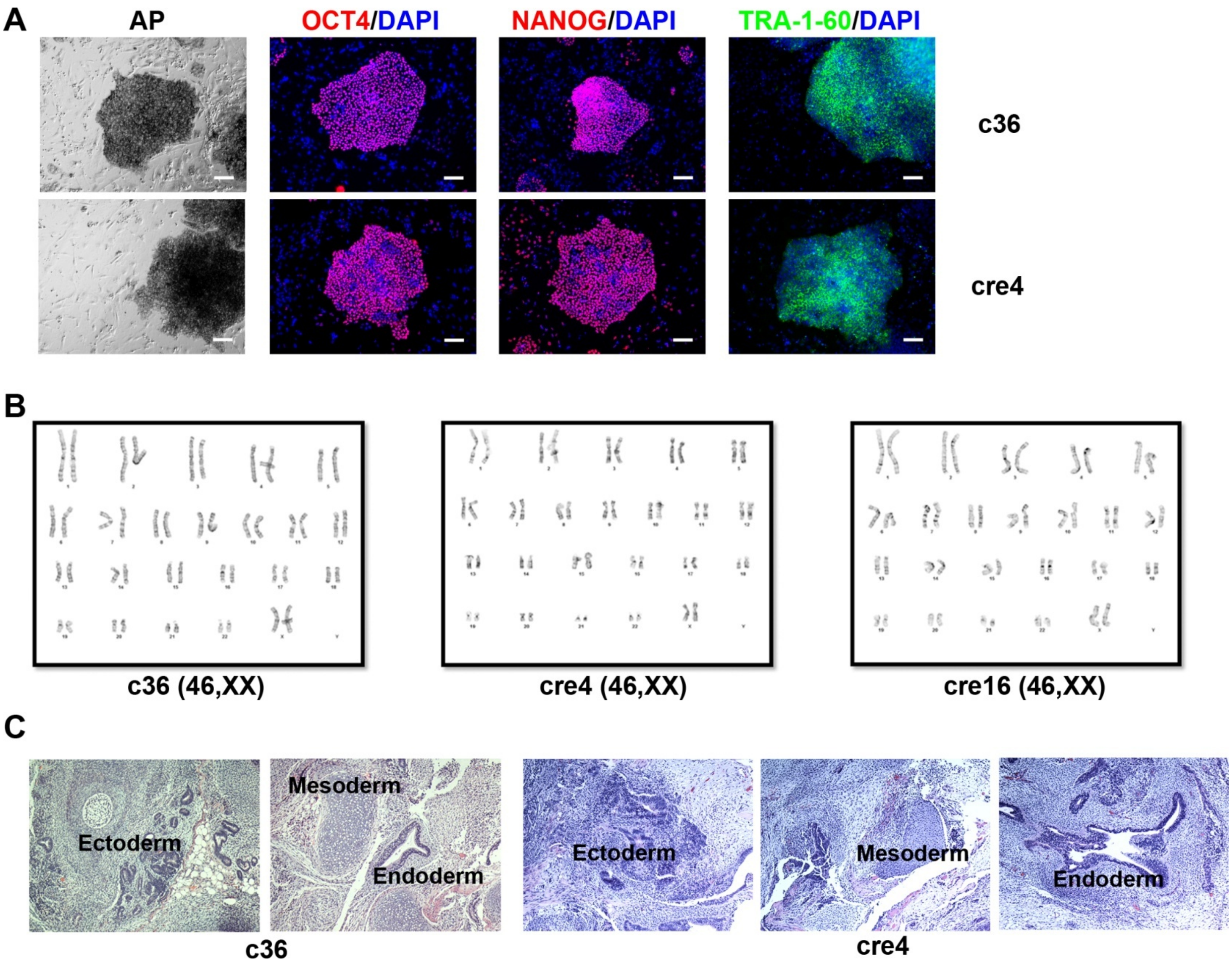

A
AP
OCT4/DAPI
NANOG/DAPI
TRA-1-60/DAPI
c36
cre4
B
c36 (46,XX)
cre4 (46,XX)
cre16 (46,XX)
C
Ectoderm
Mesoderm
Endoderm
c36
Ectoderm
Mesoderm
Endoderm
cre4

Fig. 5

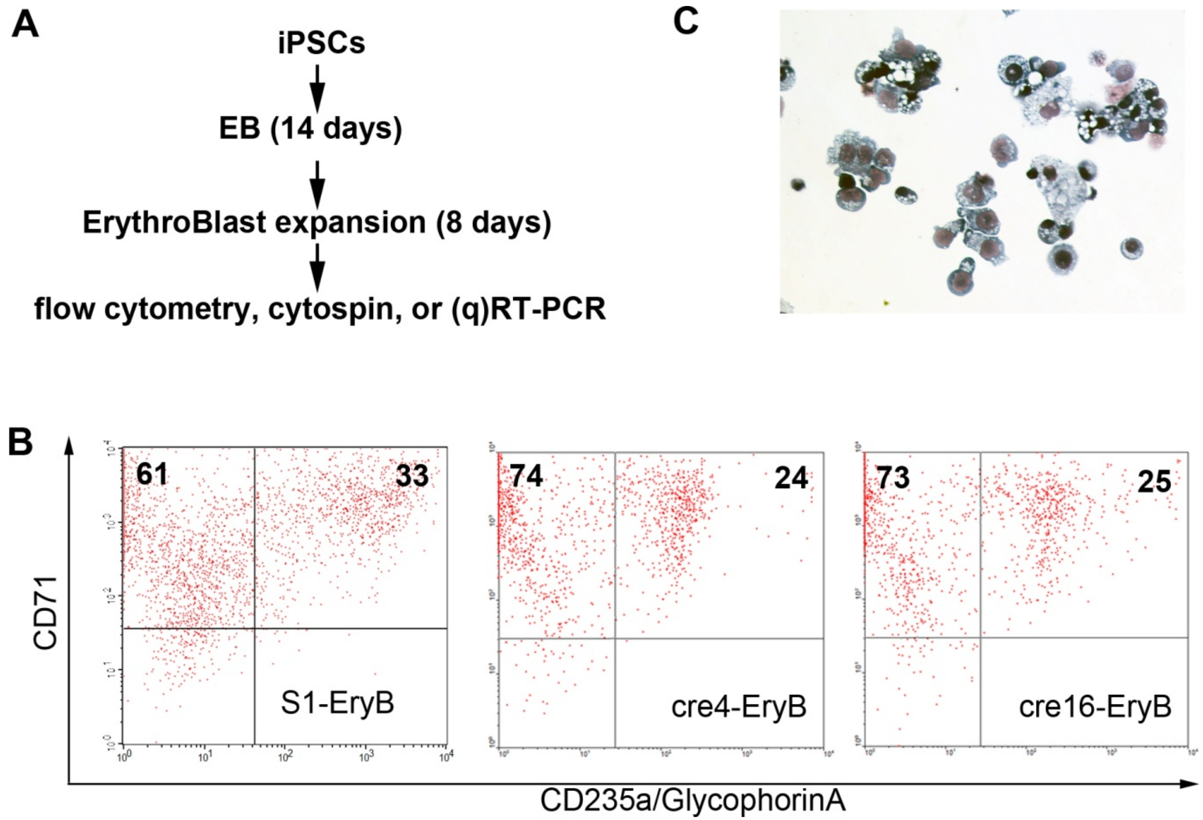

A
iPSCs
EB (14 days)
ErythroBlast expansion (8 days)
flow cytometry, cytospin, or (q)RT-PCR
C
B
CD71
61
33
S1-EryB
74
24
cre4-EryB
73
25
cre16-EryB
CD235a/GlycophorinA

Fig. 6

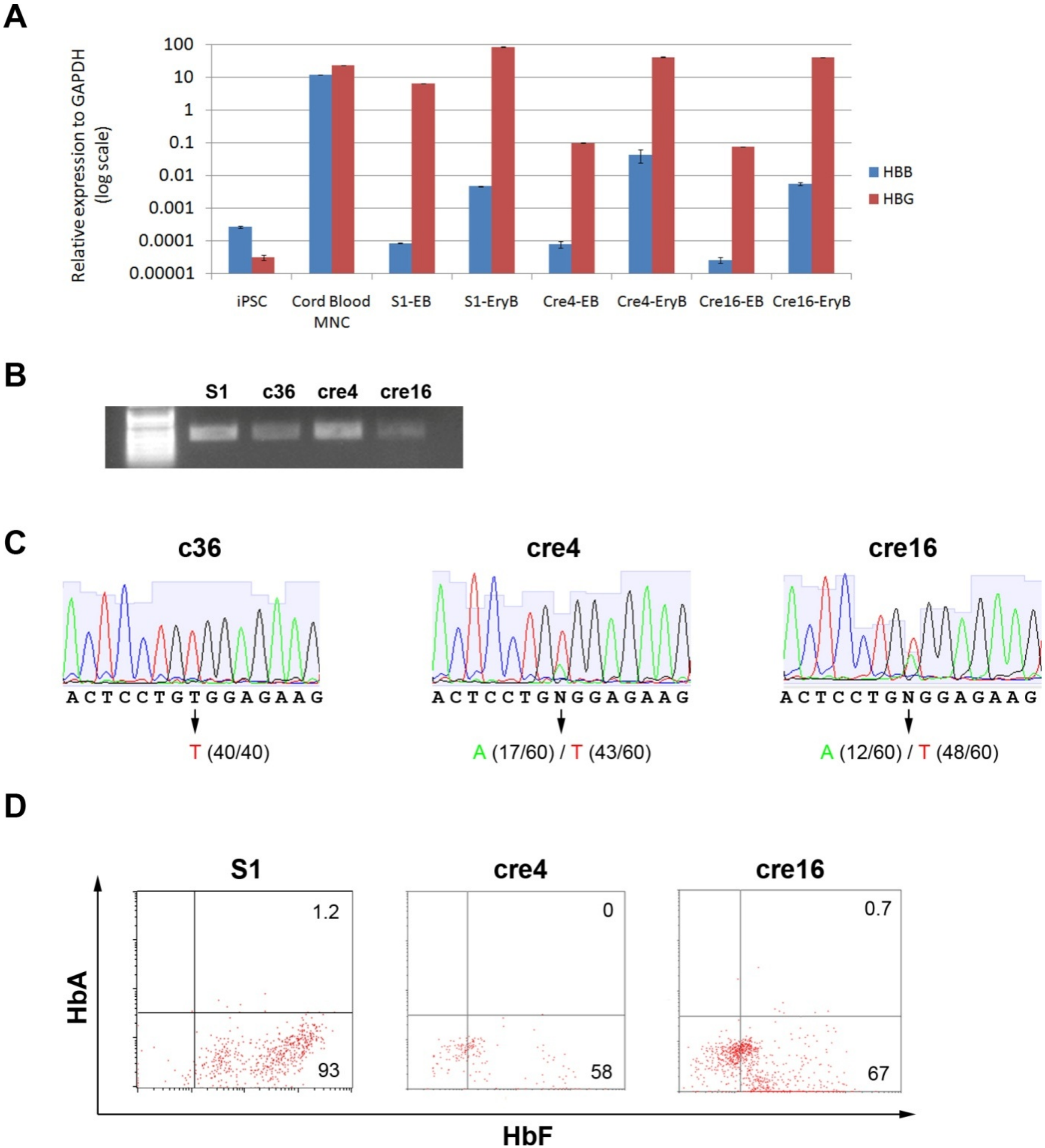
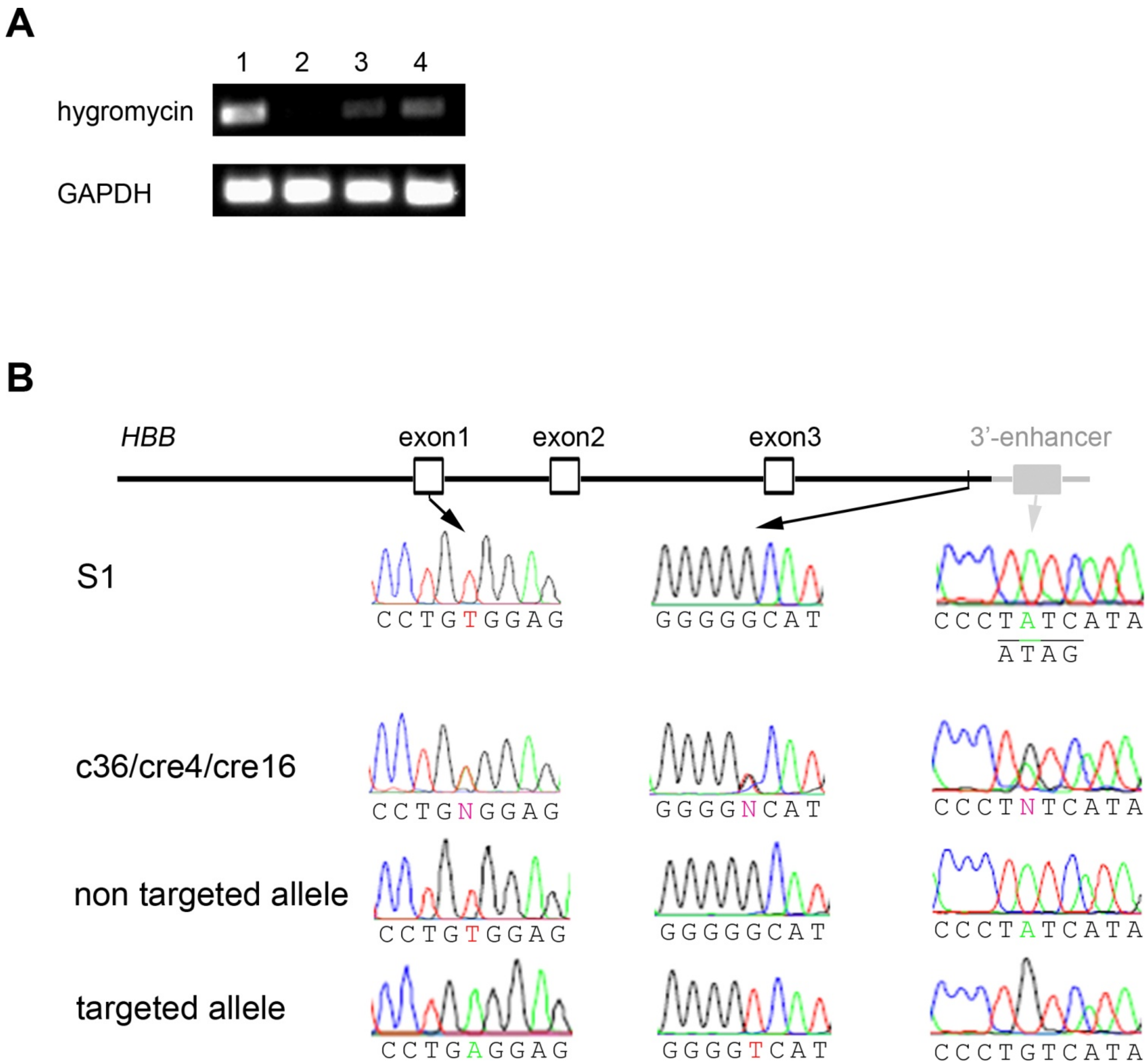

A
Relative expression to GAPDH (log scale)
100
10
1
0.1
0.01
0.001
0.0001
0.00001
HBB
HBG
iPSC
Cord Blood MNC
S1-EB
S1-EryB
Cre4-EB
Cre4-EryB
Cre16-EB
Cre16-EryB
B
S1
c36
cre4
cre16
C
c36
cre4
cre16
A C T C C T G T G G A G A A G
T (40/40)
A C T C C T G N G G A G A A G
A (17/60) / T (43/60)
A C T C C T G N G G A G A A G
A (12/60) / T (48/60)
D
S1
cre4
cre16
1.2
93
0
58
0.7
67
HbA
HbF

Fig. 7


A
1
2
3
4
hygromycin
GAPDH
B
HBB
exon1
exon2
exon3
3'-enhancer
S1
CCTGTGGAG
GGGGGCAT
CCCTATCATA
ATAG
c36/cre4/cre16
CCTGNGGAG
GGGGNCAT
CCCTNTCATA
non targeted allele
CCTGTGGAG
GGGGGCAT
CCCTATCATA
targeted allele
CCTGAGGAG
GGGGTCAT
CCCTGTCATA