

# Efficient correction of hemoglobinopathy-causing mutations by homologous recombination in integration-free patient iPSCs

*Cell Research* (2011) 21:1740-1744. doi:10.1038/cr.2011.186; published online 22 November 2011

## Dear Editor,

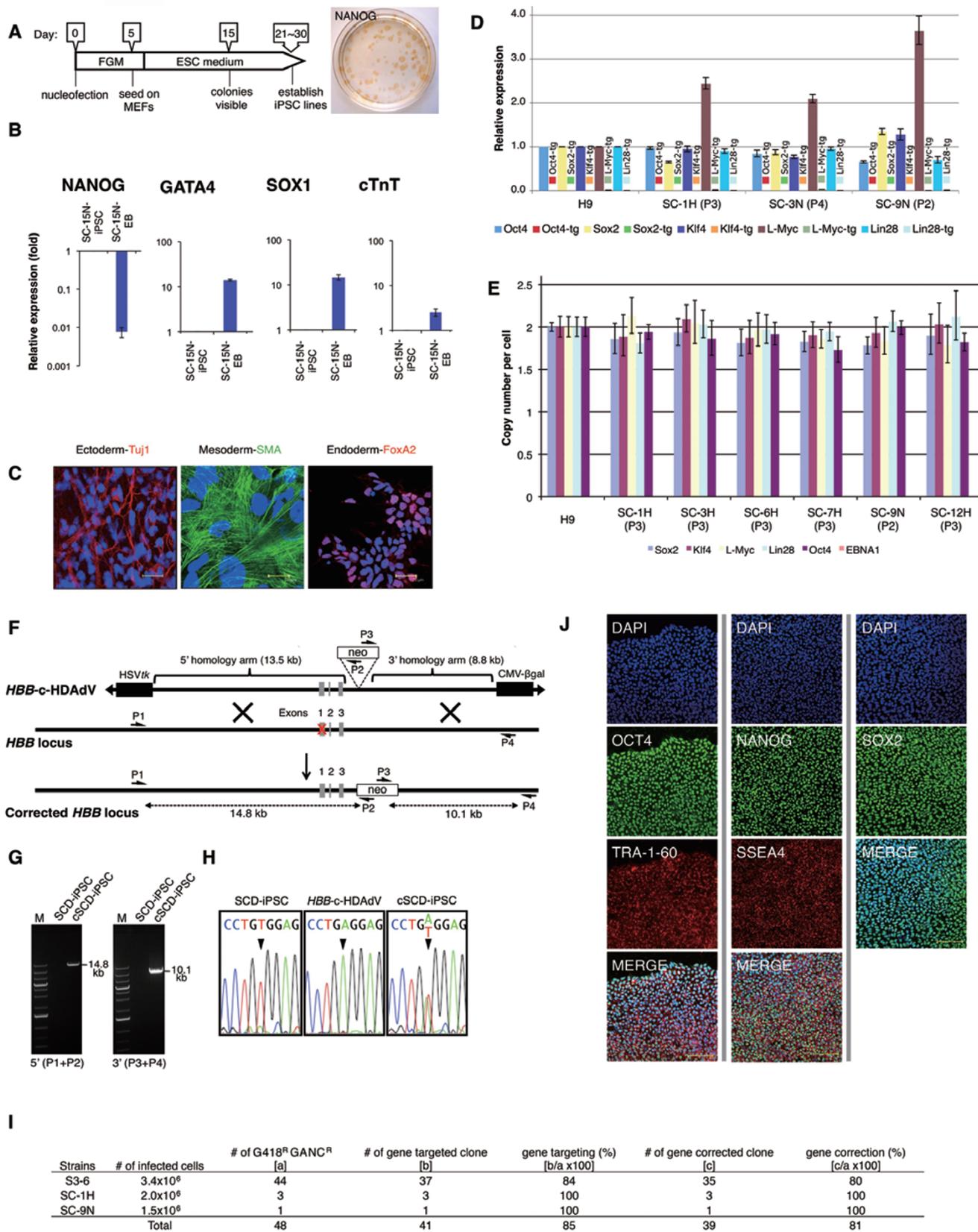
Hemoglobinopathies are a collection of heritable diseases caused by abnormal structure or insufficient production of hemoglobins [1]. Many forms of hemoglobinopathies, such as sickle cell disease (SCD) and  $\beta$ -thalassemia, which are caused by mutations of the  $\beta$ -globin (*HBB*) gene, lead to severe anemia and other life-threatening conditions. Ultimately, due to the lack of a curative treatment, patients generally have shortened life spans even under life-long supportive care. Given the high prevalence of these diseases, especially in some areas of Africa, Asia and the Mediterranean region, much effort has been devoted to developing a genetic cure for hemoglobinopathies. Current strategies for correcting the genetic defects entail the transduction of patient hematopoietic stem cells (HSCs) with viral vectors carrying a normal globin transgene with the hope of substituting for the defective endogenous counterparts [1]. However, there has been little success up to date. The challenges facing these approaches include, but are not limited to: i) the transduction efficiency of human HSCs is very low, which leads to low engraftment of corrected HSCs; ii) vectors fail to recapitulate physiological levels of human globin expression because their cargo capacity is well below the size of the globin regulatory regions (over 100 kb), which are critical for high expression. Additionally, transgenes might be subjected to silencing due to chromosomal position effects; iii) random integration of viral transgenes can lead to insertional mutagenesis and malignancy.

Alternatively, the disease-causing mutations can be repaired *in situ*. Correction by homologous recombination (HR) results in the reversion to wild-type sequences in the natural genomic context, thus circumventing many of the above-mentioned issues. Induced pluripotent stem cells (iPSCs) represent an ideal cell population for performing HR-mediated gene correction, since they can self-renew indefinitely, have the ability to differentiate

into all cell types, can be derived from patient somatic cells, and are suitable for autologous transplantations. Previously, the Jaenisch group demonstrated the feasibility of mouse iPSC-based gene correction for treatment of SCD in a mouse model, but the authors cautioned that the retroviral reprogramming strategy used in the same study was not suitable for human therapy [2]. In view of future clinical applications, patient iPSCs should be ideally free of transgenes and be derived in a fast and efficient manner. However, current methodologies for reprogramming are largely based on integrative approaches or excisable vectors, both leading to random integration of exogenous DNA into the host genome or leaving behind genetic “scars” upon excision. Altogether, translational iPSC technology and gene correction should ideally aim for two underlying and complementary methodologies, reprogramming to pluripotency by efficient non-integrative approaches and the use of HR for precise and “in-context” gene editing.

Recently, several reports have confirmed the applicability of HR for the correction of genetic mutations in disease-specific iPSCs, using helper-dependent adenoviral vectors (HDAdVs), bacterial artificial chromosome (BAC), and zinc-finger nucleases (ZFNs) [3-5]. HR efficiency in general is still low, making it arduous to obtain a large number of corrected clones for the purpose of quality control. The ZFN approach is also plagued by concerns about potential off-target effects leading to undesired mutations and chromosomal aberrations, which cannot be predicted by mere bioinformatics algorithms [6, 7].

We have previously shown that HDAdVs were highly efficient at gene correction (~46%) of transcriptionally inactive loci in patient iPSCs while maintaining high genomic and epigenomic integrity [4]. Furthermore, this method does not rely on induced DNA double strand breaks, which might lead to off target effects. Here we report on the establishment of a highly efficient and safe method for correction of *HBB* mutations in iPSCs. This



approach combined with an efficient protocol for generation of SCD-specific integration-free iPSCs represents a step forward in iPSC-based cell-replacement therapy.

SCD-specific primary fibroblasts homozygous for the hemoglobin S (HbS) mutation were obtained from the Coriell Institute for Medical Research. The A→T transversion at nucleotide 20 (HbS) in the *HBB* gene was verified by direct sequencing (Supplementary information, Figure S1A). To provide a platform for testing HDAdV-mediated gene correction in the *HBB* locus, we first reprogrammed the SCD fibroblasts using retroviral vectors expressing either the *Klf4*, *Oct4*, *Sox2*, and *c-Myc* (KOSM) cocktail or an alternative combination in which *Oct4* was substituted with a fusion protein between *Oct4* and the transactivation domain of VP16, which has been shown to enhance reprogramming [8]. ESC-like flat and compact colonies were picked ~3 weeks after viral transductions. These colonies were readily expanded and strongly expressed alkaline phosphatase (Supplementary information, Figure S1B). The established iPSC lines displayed typical characteristics of pluripotency, including expression of pluripotency marker proteins, upregulation of the endogenous pluripotency transcription network, silencing of viral transgenes, demethylation of the *Oct4* promoter, and the ability to differentiate into lineages of all three germ layers *in vitro* and *in vivo* (Supplementary information, Figures S1C-S1F and S2). Five of six ran-

domly selected clones displayed normal karyotypes, with one clone containing two translocation events (Supplementary information, Figure S3). We confirmed the presence of the HbS mutation in all SCD-iPSC lines (Supplementary information, Figure S1A).

We next sought to generate integration-free SCD-iPSCs in a fast and efficient manner. To this end, we performed a single nucleofection of EBNA1/oriP-based episomal vectors expressing six reprogramming factors (*Oct4*, *Sox2*, *Klf4*, *L-Myc*, *Lin28* and p53 shRNA, i.e., the Y4 combination as described in [9]) and an EGFP reporter. Colonies with epithelial-like morphology emerged on day 15, and by day 18 they displayed typical ESC-like morphology (Figure 1A). The majority of the colonies robustly expressed *Nanog* (Figure 1A). On average, we obtained 115 *Nanog*-positive colonies from  $5 \times 10^5$  fibroblasts (reprogramming efficiency is 0.023%). One hundred percent of the colonies initially picked based on morphological characteristics established iPSC lines ( $n = 25$ ). These episomal-derived SCD-iPSCs (epi-SCD-iPSCs) displayed properties of pluripotent stem cells, including expression of pluripotency protein markers, activation of the endogenous pluripotency transcription network and differentiation into derivatives of all three germ layers *in vitro* (Figure 1B-1D and Supplementary information, Figure S1C). Additionally, the original HbS mutation was detected in all epi-SCD-iPSCs, thereby

**Figure 1** Generation of epi-SCD-iPSCs and correction of SCD-associated *HBB* mutation in iPSCs with *HBB*-c-HDAdV. **(A)** left panel: schematic representation of the episomal-based reprogramming protocol. Time and description of key events are labeled. FGM: fibroblast growth medium; right panel: At the end of the protocol, colonies were stained with a *Nanog* antibody to calculate the reprogramming efficiency. **(B)** Quantitative RT-PCR analysis showing upregulation of endodermal (*GATA4*), mesodermal (*cTnT*), and ectodermal (*SOX1*) markers, and downregulation of pluripotency marker (*NANOG*) in epi-SCD-iPSCs before (SC-15N iPSC) and after embryo (EB)-mediated differentiation (SC-15N EB). Data are shown as mean  $\pm$  s.d. **(C)** Immunofluorescence analyses performed on EBs derived from epi-SCD-iPSCs showing expression of ectodermal (*Tuj1*), mesodermal (*SMA*), and endodermal (*FOXA2*) markers. Bar, 20  $\mu$ m. **(D)** Quantitative RT-PCR analysis on the expression of pluripotency transcription factors from the endogenous genes (*Oct4*, *Sox2*, *Klf4*, *L-Myc* and *Lin28*) or transgenes (*Oct4-tg*, *Sox2-tg*, *Klf4-tg*, *L-Myc-tg* and *Lin28-tg*). Error bars represent standard deviations. **(E)** Quantification of copy number of episomal vectors (EBNA1) and genes encoding the reprogramming factors (*Oct4*, *Sox2*, *Klf4*, *L-Myc* and *Lin28*). Episomal vectors are not detectable in any of the epi-SCD-iPSC clones, whereas the average copy number of episomal vectors in fibroblasts 5 days after nucleofection is 275 (EBNA1, see Supplementary information, Figure S5A). We observed no significant deviation from the endogenous copy numbers of the genes used in reprogramming, whereas extra copies of them were detected in retrovirally-derived iPSCs (see Supplementary information, Figure S5B). Error bars represent standard deviations. **(F)** Schematic molecular representation of *HBB* gene correction with gene-correction vector (*HBB*-c-HDAdV). The primers for PCR are shown as arrows (P1, P2, P3 and P4). *HSVtk* stands for herpes simplex virus thymidine kinase gene cassette used for negative selection; *neo* stands for neomycin-resistance gene cassette used for positive selection; *CMV- $\beta$ gal* indicates the  $\beta$ gal expression cassette for determination of HDAdV titer; Red X, The A→T mutation site at nucleotide 20 (A20T) in exon 1. **(G)** PCR analyses of SCD-iPSCs and gene-corrected iPSCs (cSCD-iPSCs) using 5' primer pair (P1 and P2; 14.8 kb) or 3' primer pair (P3 and P4; 10.1 kb). M, DNA ladder. **(H)** Sequencing results of A20T mutation site in exon 1 of *HBB* in SCD-iPSCs, *HBB*-c-HDAdV and cSCD-iPSCs. **(I)** Gene-targeting and gene-correction efficiencies at the *HBB* locus in multiple SCD-iPSC clones (S3-6, SC-1H and SC-9N). **(J)** Immunostaining showing expression of key pluripotency transcription factors *Oct4*, *Sox2* and *Nanog*, and expression of hES cell-specific surface antigens *Tra-1-60* and *SSEA4* in gene-corrected epi-SCD-iPSCs. Nuclei were stained with DAPI. Bar, 100  $\mu$ m.

verifying their origin (Supplementary information, Figure S1A). The presence of residual episomal vector sequence was not detected in any of the 6 randomly picked epi-SCD-iPSC lines by passage 2-3. Moreover, none of the reprogramming factors was integrated into the host genome (Figure 1E), nor did we detect residual transcripts of the transgenes (Figure 1D). Together, our results suggest that the epi-SCD-iPSCs are free of exogenous sequences. In summary, the episomal strategy employed here led to the simple and efficient generation of integration-free epi-SCD-iPSCs, which could represent a more desirable cell population for correcting the disease-causing mutations.

With such a premise in mind, we next attempted to correct the SCD-specific mutation in the *HBB* gene by HR. The *HBB* gene consists of three small exons, which are located close together within a 2 kb genomic region. With the aim to develop a single method amenable for the correction of all possible *HBB* mutations, we engineered an HDAd-based gene-correction vector (*HBB*-c-HDAdV), covering the whole *HBB* coding region. A neomycin-resistance (*neo*<sup>R</sup>) cassette and an HSVtk cassette were included to allow for positive and negative selection, respectively (Figure 1F). In our previous study, we determined that HDAdVs are capable of repairing mutations up to 4.4 kb away from the *neo*<sup>R</sup> insertion site [4]. Accordingly, the *neo*<sup>R</sup> cassette was inserted 1 kb downstream of the 3' UTR, or 2.5 kb away from the mutation site.

We next infected one of the retroviral-integrated iPSC clones (S3-6) and two integration-free iPSC clones (SC-1H and -9N) with *HBB*-c-HDAdV. After sequential positive (G418) and negative (Ganciclovir: GANC) selections, double-resistant colonies were expanded and maintained in hESC medium. We screened 48 drug-resistant clones from multiple iPSC lines and identified 41 gene-targeted clones by PCR, indicating a gene-targeting efficiency of 85% (Figure 1G and 1I). Furthermore, DNA sequencing showed that mutations in 39 of the 41 gene-targeted colonies were successfully corrected in one of the alleles, indicating a gene-correction efficiency of 81% (Figure 1H and 1I). These gene-corrected SCD-iPSCs displayed properties of pluripotent stem cells, including expression of pluripotency protein markers, upregulation of key transcription factors of the pluripotency network and demethylation of the *Oct4* promoter (Figure 1J and Supplementary information, Figure S1F and S1G). Because heterozygous carriers of the SCD mutation do not manifest the disease, our corrected SCD-iPSCs carrying one wild-type allele could have therapeutic potential in the clinic.

It should be emphasized that the HbS mutation is

far from the *neo*<sup>R</sup> insertion site, suggesting a possibility that *HBB*-c-HDAdVs may be suitable for correcting other *HBB* mutations. To this end, we sequenced the upstream region of the *HBB* locus in SCD-iPSCs, looking for specific SNP sites that are not present in the vector. Two SCD-iPSC specific SNP sites (rs16911905 and rs1003586, dbSNP build 130 [10]) located upstream of the proximal promoter region were found in the uncorrected cells and were used as a readout for our analysis (Supplementary information, Figure S4). These two SNP sites are located 3.6 kb upstream of the *neo*-insertion site. By examining for the presence of these two SNPs in the corrected SCD-iPSCs, we found that 7 of 7 and 5 of 7 were replaced by the *HBB*-c-HDAdV sequence at rs16911905 and rs1003586, respectively. Based on these results, this vector could potentially be used to correct all *HBB* mutations, which cause SCD and  $\beta$ -thalassemia.

Finally, we examined the possibility of vector random integration in gene-corrected clones. Southern blot, Quantitative PCR, and fluorescent *in situ* hybridization (FISH) analyses indicated that there is no ectopic vector integration in the gene-corrected clones (Supplementary information, Figure S6).

While this manuscript was in preparation, two publications describing gene correction of the HbS mutation using an alternative strategy based on ZFNs were published [11, 12]. Although ZFNs have been shown to promote efficient HR in many loci, interestingly, both studies reported lower gene-correction efficiencies at the *HBB* locus than ours. This may be attributed to the influence of the genomic context on ZFN performance, further underlining the fact that designing ZFNs for mutation correction is a complex and difficult process. Another important difference between the two approaches is that HDAdVs promote HR in considerably larger regions than ZFNs in the *HBB* locus (3.6 kbp in this study and 82 bp in Sebastiano *et al.* [11], respectively), which affords greater flexibility in targeting vector design and the convenience of using one vector for the entire locus. Nonetheless, HDAdVs, ZFNs, and other gene-targeting technologies are complementary approaches that should be further refined to achieve more efficient and safer gene correction in the future.

In summary, we have efficiently generated gene-corrected patient-specific cells from integration-free SCD-iPSCs by combining episomal vector-based reprogramming and HDAdV-mediated gene correction. The presence of the episomal vectors was completely undetectable by passage 2-3, thus further reducing the time necessary for the generation of high-quality integration and exogenous DNA-free iPSCs. Despite the fact that the *HBB* locus is transcriptionally inactive in pluripotent

stem cells, we achieved highly efficient gene correction (80-100%) in SCD-iPSCs. We have previously demonstrated that HDAdV-mediated gene editing did not result in genetic and epigenetic abnormalities [4]. Therefore, our strategy minimizes genomic alterations both during reprogramming and gene-correction steps. We believe that non-integrative approaches for reprogramming in combination with HDAdV-mediated gene editing represent a significant step forward towards the development of iPSC-based therapy. Experimental materials and methods are described in the Supplementary information, Data S1.

## Acknowledgments

We would like to thank Drs K Mitani (Saitama Medical University, Japan), P Ng (Baylor College of Medicine, USA), and A Lieber (University of Washington, USA) for kindly providing experimental materials, CG Santos and MC Llach (Center of Regenerative Medicine in Barcelona, Spain) for karyotyping, M Marti (Center of Regenerative Medicine in Barcelona, Spain) for teratoma analysis, E Nivet, S Ruiz, AD Panopoulos, SL Yang, and C Rodriguez Esteban for helpful discussions, Y Xia, T Berggren, M Lutz, A Goebel, RD Soligalla and J Kim for technical help, and M Schwarz for administrative help. This study was supported by grants from the G. Harold and Leila Y. Mathers Charitable Foundation, Sanofi-Aventis, Ellison Medical Foundation, The Leona M. and Harry B. Helmsley Charitable Trust, MICINN and Fundacion Cellex (JCIB).

Mo Li<sup>1,\*</sup>, Keiichiro Suzuki<sup>1,\*</sup>, Jing Qu<sup>1</sup>, Preeti Saini<sup>1</sup>, Ilir Dubova<sup>1</sup>, Fei Yi<sup>1</sup>, Jungmin Lee<sup>1</sup>, Ignacio Sancho-Martinez<sup>1</sup>, Guang-Hui Liu<sup>1</sup>, Juan Carlos Izpisua Belmonte<sup>1,2</sup>

<sup>1</sup>Salk Institute for Biological Studies, 10010 N. Torrey Pines Rd., La Jolla, CA 92037, USA; <sup>2</sup>Center of Regenerative Medicine in Barcelona, Dr. Aiguader, 88, 08003 Barcelona, Spain

\*These two authors contributed equally to this work.

Correspondence: Juan Carlos Izpisua Belmonte  
E-mail: belmonte@salk.edu; izpisua@cmrb.eu

## References

- 1 Papanikolaou E, Anagnou NP. Major challenges for gene therapy of thalassemia and sickle cell disease. *Curr Gene Ther* 2010; **10**:404-412.
- 2 Hanna J, Wernig M, Markoulaki S, *et al.* Treatment of sickle cell anemia mouse model with iPSC cells generated from autologous skin. *Science* 2007; **318**:1920-1923.
- 3 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 USA* 2011; **108**:6537-6542.
- 4 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**:688-694.
- 5 Soldner F, Laganier J, Cheng AW, *et al.* Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* 2011; **146**:318-331.
- 6 Gabriel R, Lombardo A, Arens A, *et al.* An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* 2011; **29**:816-823.
- 7 Pattanayak V, Ramirez CL, Joung JK, Liu DR. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods* 2011; **8**:765-770.
- 8 Wang Y, Chen J, Hu JL, *et al.* Reprogramming of mouse and human somatic cells by high-performance engineered factors. *EMBO Rep* 2011; **12**:373-378.
- 9 Okita K, Matsumura Y, Sato Y, *et al.* A more efficient method to generate integration-free human iPS cells. *Nat Methods* 2011; **8**:409-412.
- 10 Sherry ST, Ward MH, Kholodov M, *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001; **29**:308-311.
- 11 Sebastiano V, Maeder ML, Angstman JF, *et al.* In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells* 2011; **29**:1717-1726.
- 12 Zou J, Mali P, Huang X, Dowey SN, Cheng L. Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood* 2011; **118**:4599-4608.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)