Stem Cell Reports
Resource
OPEN ACCESS

A CRISPR/Cas-Mediated Selection-free Knockin Strategy in Human Embryonic Stem Cells

Zengrong Zhu,[1,3] Nipun Verma,[1,2,3] Federico González,[1] Zhong-Dong Shi,[1] and Danwei Huangfu[1,*]
[1]Developmental Biology Program, Sloan Kettering Institute, 1275 York Avenue, New York, NY 10065, USA
[2]Weill Graduate School of Medical Sciences at Cornell University/The Rockefeller University/Sloan Kettering Institute Tri-Institutional M.D.-Ph.D. Program, 1300 York Avenue, New York, NY 10065, USA
[3]Co-first author
*Correspondence: huangfud@mskcc.org
http://dx.doi.org/10.1016/j.stemcr.2015.04.016

SUMMARY

The development of new gene-editing tools, in particular the CRISPR/Cas system, has greatly facilitated site-specific mutagenesis in human embryonic stem cells (hESCs), including the introduction or correction of patient-specific mutations for disease modeling. However, integration of a reporter gene into an endogenous locus in hESCs still requires a lengthy and laborious two-step strategy that involves first drug selection to identify correctly targeted clones and then excision of the drug-resistance cassette. Through the use of iCRISPR, an efficient gene-editing platform we recently developed, this study demonstrates a knockin strategy without drug selection for both active and silent genes in hESCs. Lineage-specific hESC reporter lines are useful for real-time monitoring of cell-fate decisions and lineage tracing, as well as enrichment of specific cell populations during hESC differentiation. Thus, this selection-free knockin strategy is expected to greatly facilitate the use of hESCs for developmental studies, disease modeling, and cell-replacement therapy.

INTRODUCTION

Human embryonic stem cells (hESCs) are capable of unlimited self-renewal in culture while maintaining the potential to differentiate into any cell type present in the human body and thus provide researchers great opportunities for human developmental studies, disease modeling, and cell-replacement therapies (Zhu and Huangfu, 2013). All these applications benefit from lineage-specific knockin reporters that allow real-time observation of gene-expression dynamics, cell-lineage tracing, and isolation of a specific cell population of interest from a heterogeneous differentiation culture for downstream analysis. However, creating knockin alleles in hESCs is usually a lengthy and technically challenging process. Because of the low efficiency of homologous recombination, the donor vector needs to contain a drug-resistance gene for enrichment of cells with the correct integration. Due to the concern that the insertion of a drug-resistance cassette may interfere with the expression of the reporter gene or neighboring genes, it is usually necessary to remove the drug-resistance cassette through a second electroporation step followed by isolation of clonal lines and further characterization (Davis et al., 2008). Thus, substantial time and effort is needed to generate a knockin reporter hESC line through this two-step procedure.

The development of engineered "genomic scissors" that introduce site-specific DNA double-strand breaks (DSBs), including zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and more recently the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) system, has greatly facilitated gene targeting in hESCs (Kim and Kim, 2014). Repair of a DSB by non-homologous end joining (NHEJ) often results in insertion and/or deletions (Indels) that can be used to knock out a target gene in hESCs (Ding et al., 2013a, b; González et al., 2014). Alternatively, homology-directed repair (HDR) can be employed to efficiently incorporate exogenous sequences such as a fluorescent reporter into a specific genomic locus in hESCs (Hockemeyer et al., 2009, 2011; Hou et al., 2013; Merkert et al., 2014). Despite the significant improvement, a drug-resistance cassette is still required for generating knockin reporters of genes that are not expressed in undifferentiated hESCs.

To overcome these limitations, we made use of the CRISPR/Cas system, in which the CRISPR RNA (crRNA) and *trans*-activating crRNA (tracrRNA) duplex or a single chimeric guide RNA (gRNA) recognizes a 20-nucleotide (nt) DNA sequence upstream of the 5′-NGG-3′ protospacer adjacent motif (PAM) and directs the DNA endonuclease Cas9 for site-specific cleavage (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013a). Based on this, we have developed an efficient genome-editing platform in hESCs, which we named iCRISPR (González et al., 2014). Through TALEN-mediated gene targeting in the *AAVS1* locus, we have created hESC lines (referred to as iCas9 hESCs) that allow robust, doxycycline-inducible expression of Cas9. By transfecting iCas9 hESCs with gRNAs, the iCRISPR system enables efficient NHEJ-mediated gene disruption as well as HDR-mediated precise nucleotide modifications in the presence of short single-stranded DNA (ssDNA) donors (~100 nt).

We reasoned that the iCRISPR system would also facilitate the generation of knockin reporter alleles using longer

double-stranded (dsDNA) donors and may further enable the identification of correctly targeted hESC lines without drug selection. Thus, this work explores the utility of iCRISPR for targeting fluorescent reporters into two endogenous loci, *OCT4* (*POU5F1*) and *PDX1*, and demonstrates the generation of knockin hESC lines without drug selection for both expressed and silent genes. Further characterization confirmed the creation of multiple hESC reporter lines with no undesired mutations in the targeted loci or any potential off-target sites analyzed, supporting the broad application of this approach for efficient generation of knockin alleles in hESCs.

RESULTS

CRISPR/Cas-Mediated Targeting of the *OCT4* Locus

We first evaluated the efficiency of the CRISPR/Cas system for making knockin reporter alleles by targeting the *OCT4* locus using drug selection. HUES8 hESCs were co-electroporated with two plasmids: one expressing Cas9 and a crRNA/tracrRNA duplex targeting *OCT4* and the other containing the fluorescent reporter and a drug-resistance cassette as the HDR template (Figures 1A, S1A, and S1B). We used a donor plasmid, 2A-eGFP-PGK-Puro (Hockemeyer et al., 2011), in which the last *OCT4* coding codon is fused in frame with a 2A sequence followed by eGFP (2A-eGFP) and a loxP-flanked (floxed) puromycin-resistance gene expressed from the constitutive PGK promoter (PGK-Puro) (Figure 1A). This strategy minimizes any potential impact on the endogenous protein and is applicable to targeting both silent and expressed genes. PCR and Southern blot analysis identified eight correctly targeted clones without random transgene integration from a total of 288 puromycin-resistant clones screened (Figures 1B and S1C). The targeting efficiency (2.8%) was comparable to the efficiencies observed with TALENs and ZFNs using similar targeting strategies (Hockemeyer et al., 2009, 2011).

Despite correct targeting, we failed to detect eGFP expression in any of the targeted lines. This is likely caused by the presence of the drug-resistance cassette as observed for other genes (Davis et al., 2008). Indeed, after Cre-mediated excision of the PGK-Puro cassette, all resulting *OCT4-eGFP* lines showed proper co-expression of eGFP with pluripotency markers OCT4, SOX2, and NANOG (Figures 1C, 1D, and S1D). These results highlight the necessity of removing the drug-resistance cassette for proper reporter gene expression.

Selection-free Targeting of the *OCT4* Locus Using a Mini-vector Donor

To further explore the possibility of making knockin reporter alleles without drug selection, we designed a "mini-vector" donor plasmid, 2A-mOrange, which is similar to 2A-eGFP-PGK-Puro except that there is no PGK-Puro cassette and eGFP was replaced by mOrange (Figure 2A). We also replaced the crRNA/tracrRNA duplex cr1-dp with the single gRNA cr1 targeting the same sequence (Figures 2A and S2A), as the chimeric version works more efficiently than the original duplex design (Hsu et al., 2013; Jinek et al., 2012). Similar to the experiment with the 2A-eGFP-PGK-Puro donor, we co-electroporated HUES8 hESCs with a plasmid expressing Cas9/gRNA and the new 2A-mOrange mini-vector (Figure 2A). In contrast to the absence of fluorescence after integration of the 2A-eGFP-PGK-Puro cassette, integration of the 2A-mOrange cassette resulted in mOrange expression in ~0.001% of cells as detected by fluorescence-activated cell sorting (FACS) (Figure S2B). One may enrich mOrange-expressing cells for establishing *OCT4* reporter lines. However, this low efficiency is impractical for genes not expressed in undifferentiated hESCs, as one has to rely on randomly picking individual colonies to establish clonal lines.

Our recent study shows that the iCRISPR platform enables efficient gene editing using short ssDNA HDR templates (González et al., 2014), prompting us to further optimize the iCRISPR platform for HDR using longer circular dsDNA donor vectors. After optimizing transfection conditions, we co-transfected doxycycline-treated HUES8 iCas9 cells twice in 2 days with the *OCT4* cr1 gRNA and the 2A-mOrange mini-vector using Lipofectamine 3000 (Figures S2D and S3). FACS analysis identified ~0.4% mOrange-expressing cells (Figure 2B), >100-fold greater than results from our electroporation experiments (Figures S2B and S2C). Similar results were observed in experiments using iCas9 cells generated from MEL-1 hESCs (Figure S2E), supporting the general utility of this new approach in diverse human pluripotent stem cell (hPSC) backgrounds. The much-improved efficiencies can be partially attributed to the integration of Cas9 in the genome as an ~5- to 6-fold increase of mOrange$^+$ cells was observed compared to the control condition where iCas9 hESCs (not treated with doxycycline) were transfected with Cas9/gRNA and the donor vector using Lipofectamine 3000 (Figure 2C). The use of Lipofectamine transfection also substantially increased the targeting efficiency compared to the electroporation method (Figures S2B and S2C). There may be other ways to improve the transfection efficiency (e.g., through nucleofection) to achieve similar results with or without the use of iCas9 hESCs (Byrne et al., 2015).

We picked ten colonies from individual FACS-isolated mOrange$^+$ cells and identified six correctly targeted clones by PCR and Southern blot analysis (Figures 2B, 2D, and S2F). All six lines co-expressed mOrange with pluripotency markers such as OCT4, SOX2, and NANOG and displayed normal hESC morphology (Figure S2G). We further

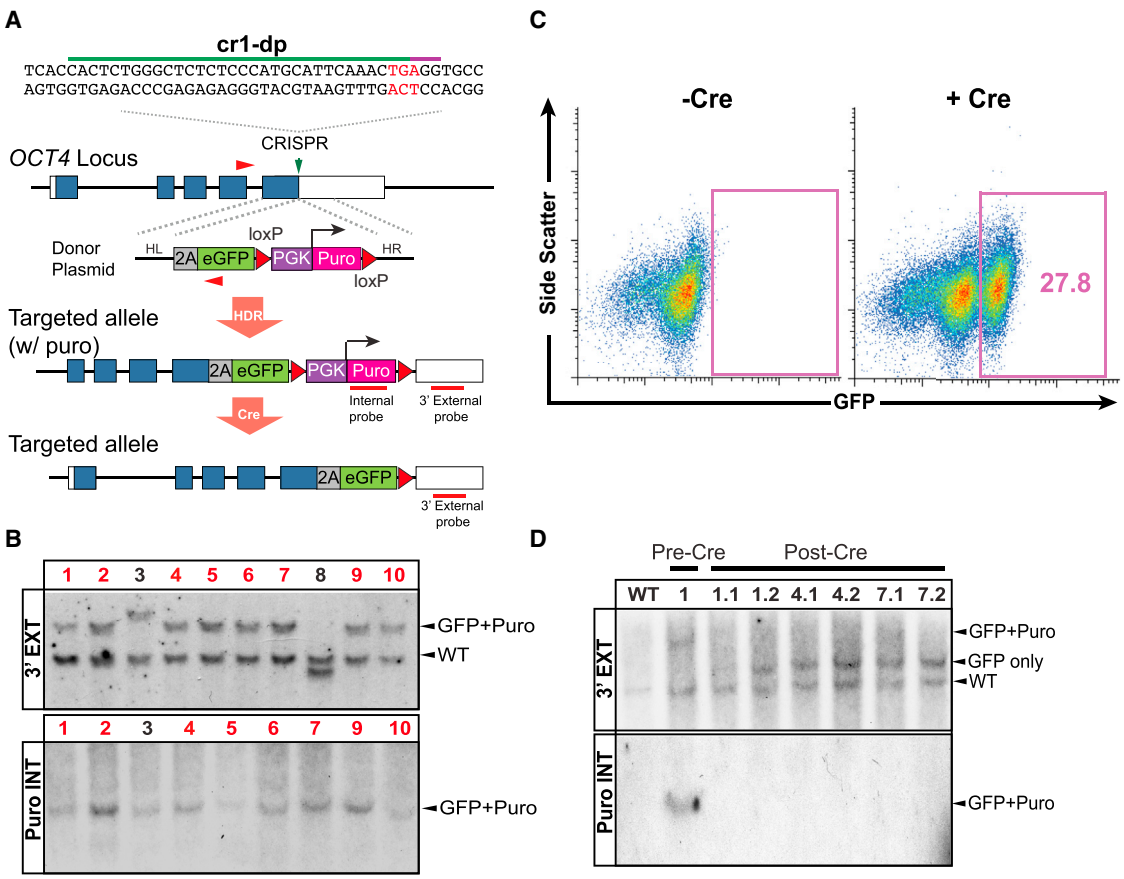


Figure 1. CRISPR/Cas-Mediated Targeting of the *OCT4* Locus through Drug Selection

(A) Schematics of the targeting strategy. In the presence of the donor plasmid, HDR results in the replacement of the *OCT4* stop codon with 2A-eGFP-PGK-Puro. The PCR primers (F + GFP-R) used for genotyping are indicated with red arrowheads. *OCT4* cr1-dp (the duplex version) targets a 30-nt sequence (indicated with a green line) upstream of the PAM sequence (indicated with a purple line). In all targeting schematics here and after, boxes are exons, filled blue boxes indicate the coding sequence (CDS), connecting lines are introns, the stop codon (TGA) is labeled in red, HL and HR indicate left and right homology arms, and the Southern blot external and internal probes are indicated with red bars.

(B) Southern blot analysis using the external probe (WT: 4,173 bp; GFP+Puro: 6,835 bp) and the internal puromycin probe (GFP+Puro: 2,415 bp) identified eight correctly targeted clones, which are labeled in red. WT, wild-type allele; GFP+Puro, correctly targeted allele with puromycin-selection cassette.

(C) Three of the correctly targeted clones (nos. 1, 4, and 7) were electroporated with Cre recombinase. Four days after electroporation, eGFP$^+$ cells were isolated using FACS.

(D) For each clone (nos. 1, 4, and 7) electroporated with CRE recombinase, two GFP$^+$ clones were picked (e.g., C1.1 and C1.2 for clones derived from no. 1), and Southern blot analysis using the external probe (WT: 4,173 bp; GFP+Puro: 6,835 bp; GFP only: 5,015 bp) and the internal puromycin probe (GFP+Puro: 2,415 bp) showed correct removal of the puromycin-selection cassette. A clone (C1) prior to Cre electroporation was used as the Pre-Cre control. GFP+Puro, targeted allele prior to Cre-mediated excision of the PGK-Puro cassette; GFP only, targeted allele after Cre-mediated excision.

examined the *OCT4-mOrange* hESC reporter lines along with the *OCT4-eGFP* lines for reporter gene expression after differentiation. After 3 days of treatment with BMP4 and SB431542, a TGFβ inhibitor (Hou et al., 2013), hESCs exhibited a differentiated morphology, and eGFP and mOrange expression were downregulated in the respective *OCT4-eGFP* and *OCT4-mOrange* hESC reporter lines with concomitant loss of endogenous OCT4 expression as determined by immunostaining and FACS analysis (Figures 3A and 3B). Thus, the *OCT4-eGFP* and *OCT4-mOrange* reporters faithfully reflect endogenous gene expression during the maintenance and differentiation of hESCs.

We next investigated whether the relatively high targeting efficiency was achieved at the expense of undesirable

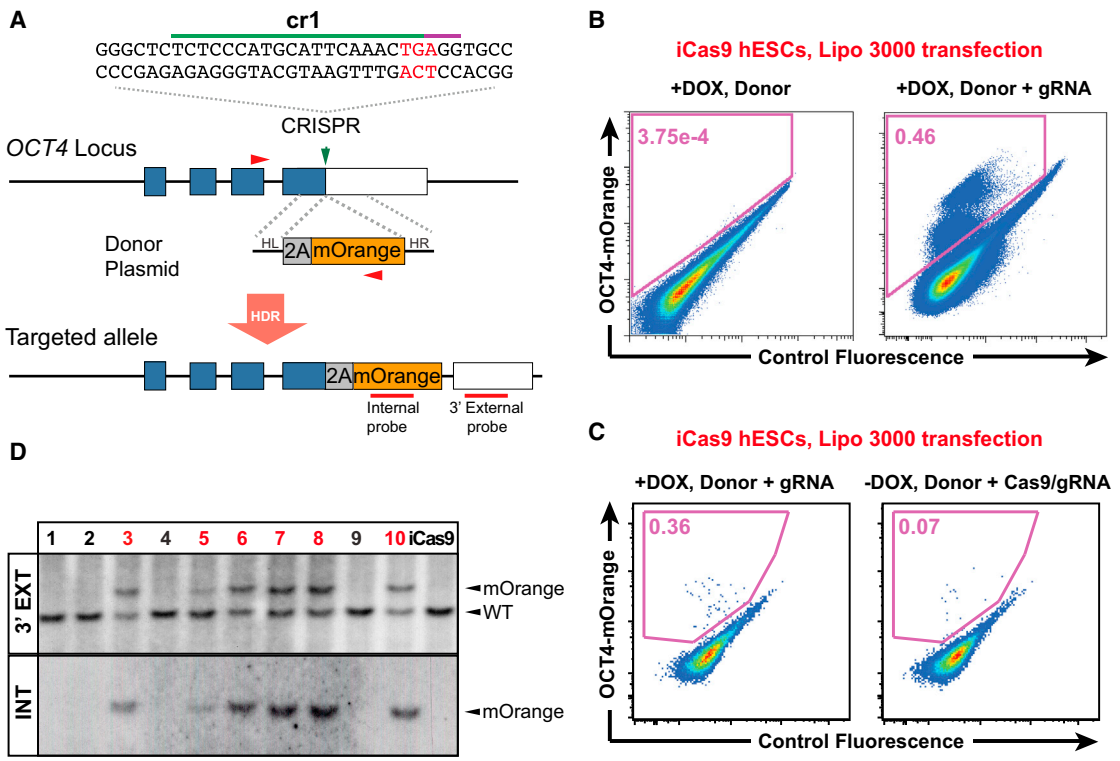


Figure 2. Targeting the *OCT4* Locus without Drug Selection

(A) Schematics of the targeting strategy without drug selection. In the presence of the donor plasmid, HDR results in the replacement of the stop codon with 2A-mOrange. *OCT4* cr1 targets a 20-nt sequence upstream of the PAM sequence. The PCR primers (F + mOr-R) used for genotyping are indicated with red arrowheads.

(B) FACS enrichment for OCT4-mOrange$^+$ cells after transfection of the OCT4-mOrange donor plasmid and the *OCT4*-targeting gRNA into HUES8 iCas9 cells treated with doxycycline.

(C) FACS analysis for OCT4-mOrange fluorescence in doxycycline-treated HUES8 iCas9 cells co-transfected with the OCT4-mOrange donor plasmid and the *OCT4*-targeting gRNA, compared to HUES8 iCas9 cells (not treated with doxycycline) co-transfected with the Cas9/gRNA and the donor plasmids using Lipofectamine 3000.

(D) Ten colonies were randomly picked from individual FACS-enriched mOrange$^+$ cells. Southern blot analysis using the external probe (WT: 4,173 bp; mOrange: 4,963 bp) and the internal mOrange probe (mOrange: 4,963 bp) identified six correctly targeted clones, which are labeled in red. mOrange: correctly targeted allele.

mutations at the *OCT4* locus or any off-target sites. All eight *OCT4-eGFP* and six *OCT4-mOrange* lines examined showed the expected sequence at the junction between the endogenous *OCT4* sequence and the inserted sequence. This is reassuring, as we made sure that the donor template did not contain the CRISPR target sequence to prevent undesired mutagenesis after reporter gene integration. However, Indel mutations were detected in the non-targeted allele in two of the six *OCT4-mOrange* reporter lines examined (Figure 3C). These findings underscore the necessity of thorough sequence analysis for eliminating clones with undesired mutations in the non-targeted allele, a point not widely recognized with the CRISPR/Cas-mediated targeting strategy. We also sequenced seven predicted off-target sites based on the 12-bp seed sequence important for target recognition (Jiang et al., 2013; Jinek et al., 2012). Examination of six *OCT4-mOrange* and eight *OCT4-eGFP* lines revealed no mutations except that three *OCT4-eGFP* lines carried mutations at the *POU5F1P4* locus, which shares the same 20-nt target sequence with the intended target (Table S1).

Targeting the *PDX1* Locus Using a Mini-vector Donor without Drug Selection

We further investigated whether this selection-free approach also applied to genes not expressed in undifferentiated hESCs. We chose to target *PDX1*, which encodes a transcription factor not expressed in undifferentiated hESCs but in pancreatic progenitors and their differentiated progeny such as pancreatic β cells. The ability to monitor *PDX1* expression during in vitro differentiation and to enrich *PDX1*$^+$ pancreatic progenitor cells or β cells

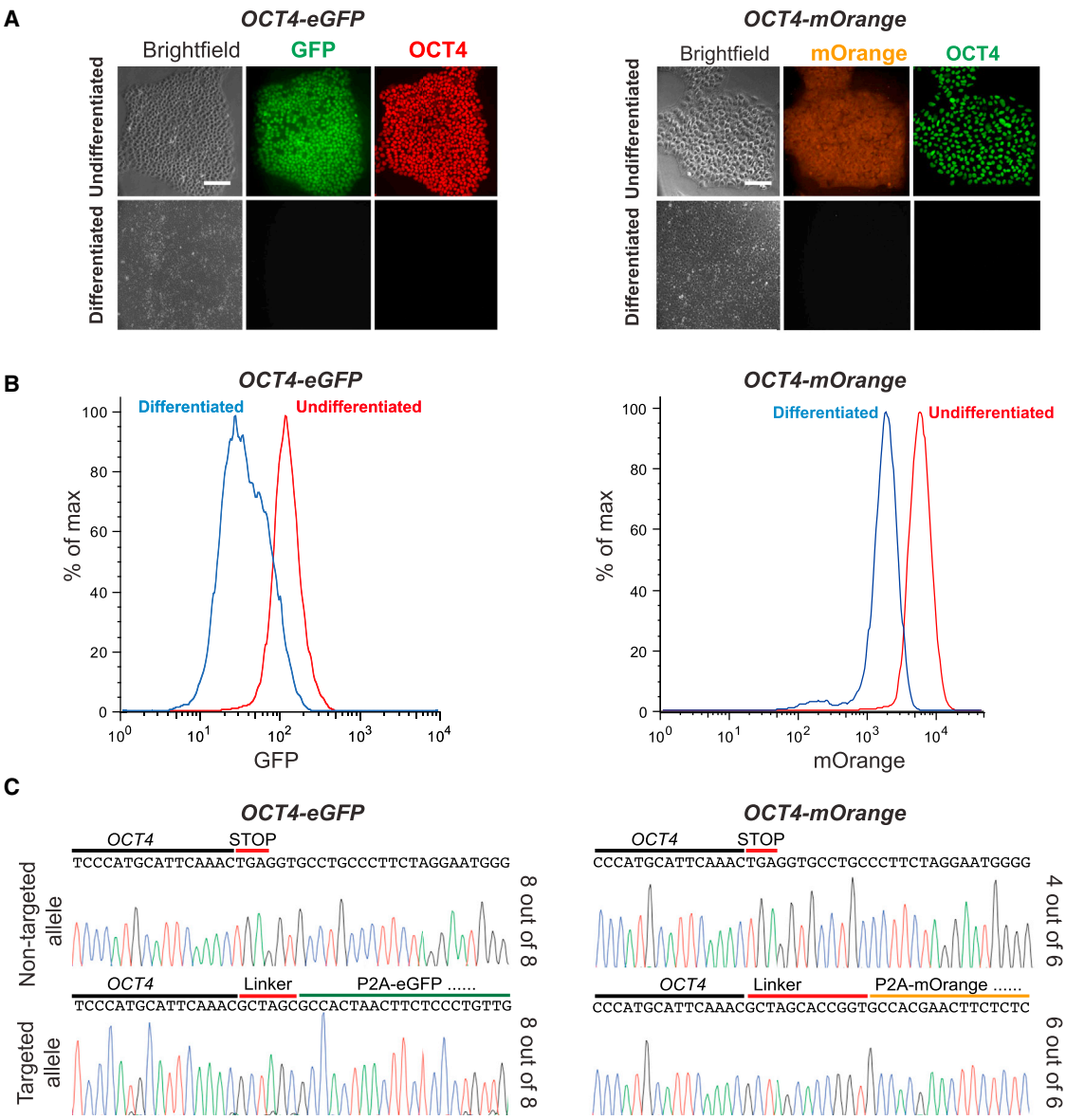


Figure 3. Characterization of *OCT4* Reporter Lines

(A) *OCT4-eGFP* and *OCT4-mOrange* hESCs were treated with SB431542 and BMP4 to initiate differentiation. Three days after this treatment, the cells displayed concomitant loss of OCT4 protein expression with GFP or mOrange by immunostaining. An RFP antibody was used to detect mOrange expression, whereas the GFP expression was detected directly. The scale bar represents 100 μm.

(B) Three days after SB431542 and BMP4 treatment, flow cytometry analysis showed a loss of GFP and mOrange, verifying that *OCT4-eGFP* and *OCT4-mOrange* reporter hESCs can respond to differentiation cues and that GFP and mOrange accurately reflects *OCT4* expression.

(C) Sequencing results of the non-targeted allele and at the junction of correctly targeted allele in *OCT4-eGFP* and *OCT4-mOrange* reporter lines.

would be valuable for studies of pancreatic development and the use of hESCs for β cell replacement therapies.

We designed two gRNAs (*PDX1* cr1 and cr2) to target DNA sequences in proximity to the *PDX1* stop codon and used a PDX1-eGFP mini-vector donor for integration of the eGFP reporter into the *PDX1* locus (Figures 4A and S4A). Correct targeting should enable expression of eGFP from the endogenous *PDX1* locus with minimal impact on PDX1 protein expression or function. Following establishment of clonal lines (Figure 4B), PCR and Southern blot analysis identified three correctly targeted clones without random integration: two clones with eGFP

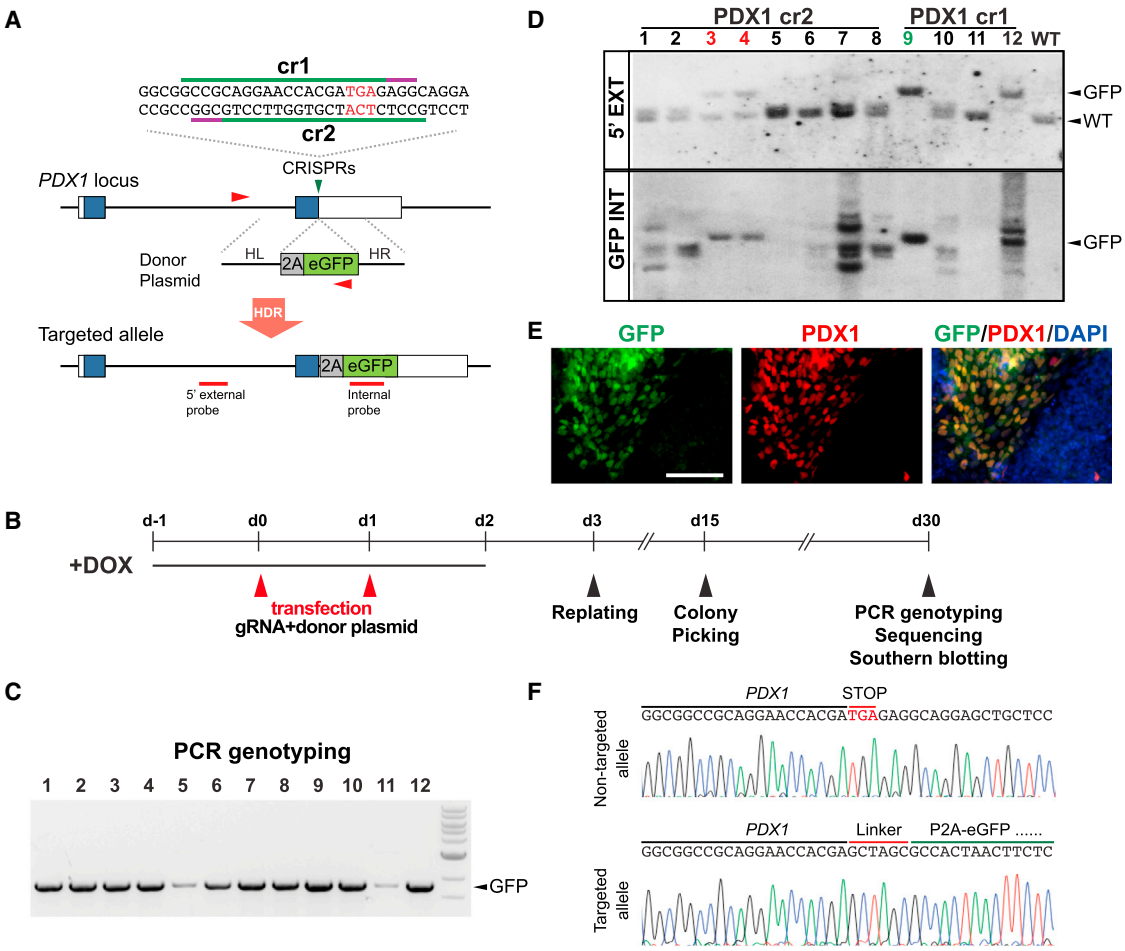


Figure 4. Generation of *PDX1-eGFP* Reporter Lines without Drug Selection

(A) *PDX1* cr1 and cr2 were designed to target sequences in proximity to the *PDX1* stop codon. In the presence of the eGFP donor plasmid, HDR resulted in the replacement of the stop codon with 2A-eGFP. PCR primers (F + R) used for genotyping are indicated with red arrowheads.

(B) Timeline for the generation of *PDX1-eGFP* hESC lines using the iCRISPR platform. DOX, doxycycline.

(C) PCR genotyping of 12 clones, identified from the PCR screen in Figure S4B, that showed the correct PCR product for the targeted allele (GFP: 1,746 bp).

(D) Southern blot analysis using the external probe and internal probe (WT: 3,845 bp; GFP: 4,632 bp) identified three correctly targeted lines. Lines without random integrations and carrying mono-allelic eGFP insertion are labeled in red, and the clone with a bi-allelic eGFP insertion is labeled in green.

(E) Immunofluorescence staining displayed co-expression of GFP and PDX1 in pancreatic progenitors differentiated from *PDX1-eGFP* hESCs. The scale bar represents 100 μm.

(F) Sanger sequencing of non-targeted and targeted *PDX1* alleles showed correct targeting of the *PDX1* locus with no undesired mutations.

integration in one *PDX1* allele and one clone with biallelic integration (Figures 4C, 4D, and S4B). All three *PDX1-eGFP* reporter lines displayed normal hESC morphology and expressed pluripotency markers OCT4, NANOG, and SOX2 as determined by immunofluorescence staining (Figure S4C). After differentiation into pancreatic progenitors, co-expression of GFP and PDX1 was observed, demonstrating faithful reporter activity (Figure 4E). Thus, we successfully generated multiple faithful knockin reporter lines for *PDX1*, a gene with a lineage-restricted expression pattern in differentiated hESCs.

We further examined the *PDX1* reporter lines for potential undesired mutations similar to analysis performed on the *OCT4* reporter lines. We found the expected sequences in both the targeted and non-targeted *PDX1* allele in all clones (Figure 4F) except for one (no. 3) with a 24-bp deletion in the non-targeted allele. Recent studies suggest that CRISPR/Cas9 tolerates mismatches between CRISPR and

target DNA at different positions in a sequence-dependent manner, influenced by the number, position, and distribution of mismatches (Hsu et al., 2013; Tan et al., 2015). Using a prediction program developed by Feng Zhang's group (http://crispr.mit.edu), we sequenced 20 most likely off-target sites (ten each predicted for *PDX1* cr1 and cr2) and detected no mutations (Table S1).

DISCUSSION

Here, we demonstrate the generation of hESC reporter lines without the use of drug selection for both active and silent genes through the use of the iCRISPR system. Selection-free gene targeting eliminates the need for removal of drug-resistance cassette after identification of correctly targeted clones, and mini-vector donors with short homology arms (~500–1,000 bp each) are convenient to make. Thus, this method significantly reduces the time and effort required for establishing hESC reporter lines. Additionally, conventional gene-targeting strategies typically use the Cre-loxP strategy to remove drug-resistance cassettes, which leaves behind a 34-bp loxP "scar" in the endogenous locus. Although not an issue in most cases, this residual sequence could interfere with the expression of the targeted gene in some situations (Meier et al., 2010). In comparison, our strategy eliminates the need for the selection cassette and thereby minimizes the alteration of the endogenous locus. Although we focused on creating promoter-fusion reporters, the same knockin approach can be readily applied to making protein-fusion reporters for visualizing protein subcellular localization, precisely deleting or replacing specific genomic sequences, and introducing or correcting disease-associated mutations.

Previously, we and others have failed to target the *PDX1* locus using traditional targeting approaches (Z.Z. and D.H., data not shown; E. Stanley, personal communication), yet the absolute targeting efficiencies using the selection-free method were comparable between the *PDX1* and *OCT4* loci. It is known that gene-targeting efficiencies can vary significantly depending on the target locus, though the exact reason is unclear. Traditional gene targeting relies on drug selection; thus, the relative targeting efficiencies after drug selection depend, at least in part, on the expression of the drug-resistance gene from the targeted locus. Because the expression of drug-resistance gene may differ significantly between expressed and silent loci, the relative targeting efficiency after drug selection for a lineage-specific gene may appear much lower compared with a pluripotency gene. For certain loci, the drug-resistance gene may be expressed at such low levels that hinder the identification of a correctly targeted clone using the drug-selection method (Rostovskaya et al., 2012). Because our targeting approach obviates drug selection, it may overcome such bias and facilitate the generation of reporter alleles for genes that were previously difficult to target. One may further use this ability to measure absolute targeting frequencies to compare HDR efficiencies across different genomic contexts.

There have been concerns about potential off-target mutagenesis with the CRISPR/Cas system (Cho et al., 2014; Fu et al., 2013; Hsu et al., 2013; Mali et al., 2013b; Pattanayak et al., 2013). Our analysis so far did not reveal any off-target mutations at sites without perfect complementarity with the CRISPR target sequence. However, we cannot exclude the possibility of off-target mutations elsewhere in the genome, and a thorough analysis may be necessary before the reporter lines are used in future studies. The CRISPR/Cas system is continuously improved with the development of better algorithms for CRISPR design and off-target prediction. It is reassuring that a recent high-coverage whole-genome sequencing study failed to detect significant incidence of off-target mutations in CRISPR-targeted hPSC lines (Veres et al., 2014). On the other hand, we noticed that some correctly targeted clones carried mutations in the non-targeted allele, though the targeting efficiency is sufficiently high that one could simply discard the minority of clones carrying mutations. One may also target intronic regions with low-sequence conservations to further mitigate any concerns associated with Indel mutations in the non-targeted allele.

Our selection-free targeting approach enables rapid generation of knockin reporter lines, though it also requires either using established iCas9 cells or creating new iCas9 lines in a desired hPSC background. The upfront effort for generating iCas9 cells is relatively small due to the efficient TALEN-mediated *AAVS1*-targeting approach, and it is possible to establish an iCas9 line in about 1 month (González et al., 2014; Zhu et al., 2014). Once an iCas9 line is established, it can be used for making different types of reporters. Our previous study has shown that Cas9 activity is tightly regulated by doxycycline treatment, and established iCas9 lines exhibit no apparent chromosomal aberrations or defects in the maintenance of the pluripotent state (González et al., 2014). A recent study also observed no adverse effects in constitutive Cas9-expressing mice (Platt et al., 2014). An additional benefit of using iCas9 hPSCs for making reporter lines is that the cells can be conveniently used for a variety of downstream genetic studies using gene-editing approaches we already established (González et al., 2014). Thus, we expect this selection-free knockin strategy to further facilitate the use of hESCs for developmental studies, disease modeling, and cell-replacement therapy.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and two tables can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2015.04.016.

AUTHOR CONTRIBUTIONS

D.H., Z.Z., N.V., and F.G. conceived the project; Z.Z. performed most experiments related to *PDX1* targeting; N.V. performed most experiments related to *OCT4* targeting; F.G. and Z.-D.S. performed additional relevant experiments; and D.H., Z.Z., and N.V. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Feng Zhang and Rudolf Jaenisch for providing vectors through Addgene and members of the D.H. laboratory for insightful discussions and critical reading of the manuscript. This study was funded in part by NIH (R01DK096239), NYSTEM (C029156), Basil O'Connor Starter Scholar Award from March of Dimes Birth Defects Foundation, Tri-Institutional Stem Cell Initiative, Louis V. Gerstner Jr. Young Investigators Award, and MSKCC Society Special Projects Committee. Z.Z. and F.G. were both supported by the New York State Stem Cell Science (NYSTEM) fellowship from the Center for Stem Cell Biology of the Sloan Kettering Institute. N.V. was supported by the Howard Hughes Medical Institute (HHMI) Medical Research Fellowship.

Received: January 5, 2015
Revised: April 27, 2015
Accepted: April 30, 2015
Published: May 28, 2015

REFERENCES

Byrne, S.M., Ortiz, L., Mali, P., Aach, J., and Church, G.M. (2015). Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. Nucleic Acids Res. *43*, e21.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. *24*, 132–141.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Davis, R.P., Costa, M., Grandela, C., Holland, A.M., Hatzistavrou, T., Micallef, S.J., Li, X., Goulburn, A.L., Azzola, L., Elefanty, A.G., and Stanley, E.G. (2008). A protocol for removal of antibiotic resistance cassettes from human embryonic stem cells genetically modified by homologous recombination or transgenesis. Nat. Protoc. *3*, 1550–1558.

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

Ding, Q., Regan, S.N., Xia, Y., Oostrom, L.A., Cowan, C.A., and Musunuru, K. (2013b). Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. Cell Stem Cell *12*, 393–394.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *31*, 822–826.

González, F., Zhu, Z., Shi, Z.D., Lelli, K., Verma, N., Li, Q.V., and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell Stem Cell *15*, 215–226.

Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B., et al. (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat. Biotechnol. *27*, 851–857.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat. Biotechnol. *29*, 731–734.

Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L.F., Sontheimer, E.J., and Thomson, J.A. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc. Natl. Acad. Sci. USA *110*, 15644–15649.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. *31*, 233–239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Kim, H., and Kim, J.S. (2014). A guide to genome engineering with programmable nucleases. Nat. Rev. Genet. *15*, 321–334.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013a). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013b). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. *31*, 833–838.

Meier, I.D., Bernreuther, C., Tilling, T., Neidhardt, J., Wong, Y.W., Schulze, C., Streichert, T., and Schachner, M. (2010). Short DNA sequences inserted for gene targeting can accidentally interfere with off-target gene expression. FASEB J. *24*, 1714–1724.

Merkert, S., Wunderlich, S., Bednarski, C., Beier, J., Haase, A., Dreyer, A.K., Schwanke, K., Meyer, J., Göhring, G., Cathomen, T., and Martin, U. (2014). Efficient designer nuclease-based homologous recombination enables direct PCR screening for footprintless

targeted human pluripotent stem cells. Stem Cell Reports *2*, 107–118.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. *31*, 839–843.

Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell *159*, 440–455.

Rostovskaya, M., Fu, J., Obst, M., Baer, I., Weidlich, S., Wang, H., Smith, A.J., Anastassiadis, K., and Stewart, A.F. (2012). Transposon-mediated BAC transgenesis in human ES cells. Nucleic Acids Res. *40*, e150.

Tan, E.P., Li, Y., Velasco-Herrera, Mdel.C., Yusa, K., and Bradley, A. (2015). Off-target assessment of CRISPR-Cas9 guiding RNAs in human iPS and mouse ES cells. Genesis *53*, 225–236.

Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Cowan, C.A., Talkowski, M.E., and Musunuru, K. (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell Stem Cell *15*, 27–30.

Zhu, Z., and Huangfu, D. (2013). Human pluripotent stem cells: an emerging model in developmental biology. Development *140*, 705–717.

Zhu, Z., González, F., and Huangfu, D. (2014). The iCRISPR platform for rapid genome editing in human pluripotent stem cells. Methods Enzymol. *546*, 215–250.

Stem Cell Reports, Volume *4*

Supplemental Information

A CRISPR/Cas-Mediated Selection-free Knockin Strategy in Human Embryonic Stem Cells

Zengrong Zhu, Nipun Verma, Federico González, Zhong-Dong Shi, and Danwei Huangfu

Figure S1

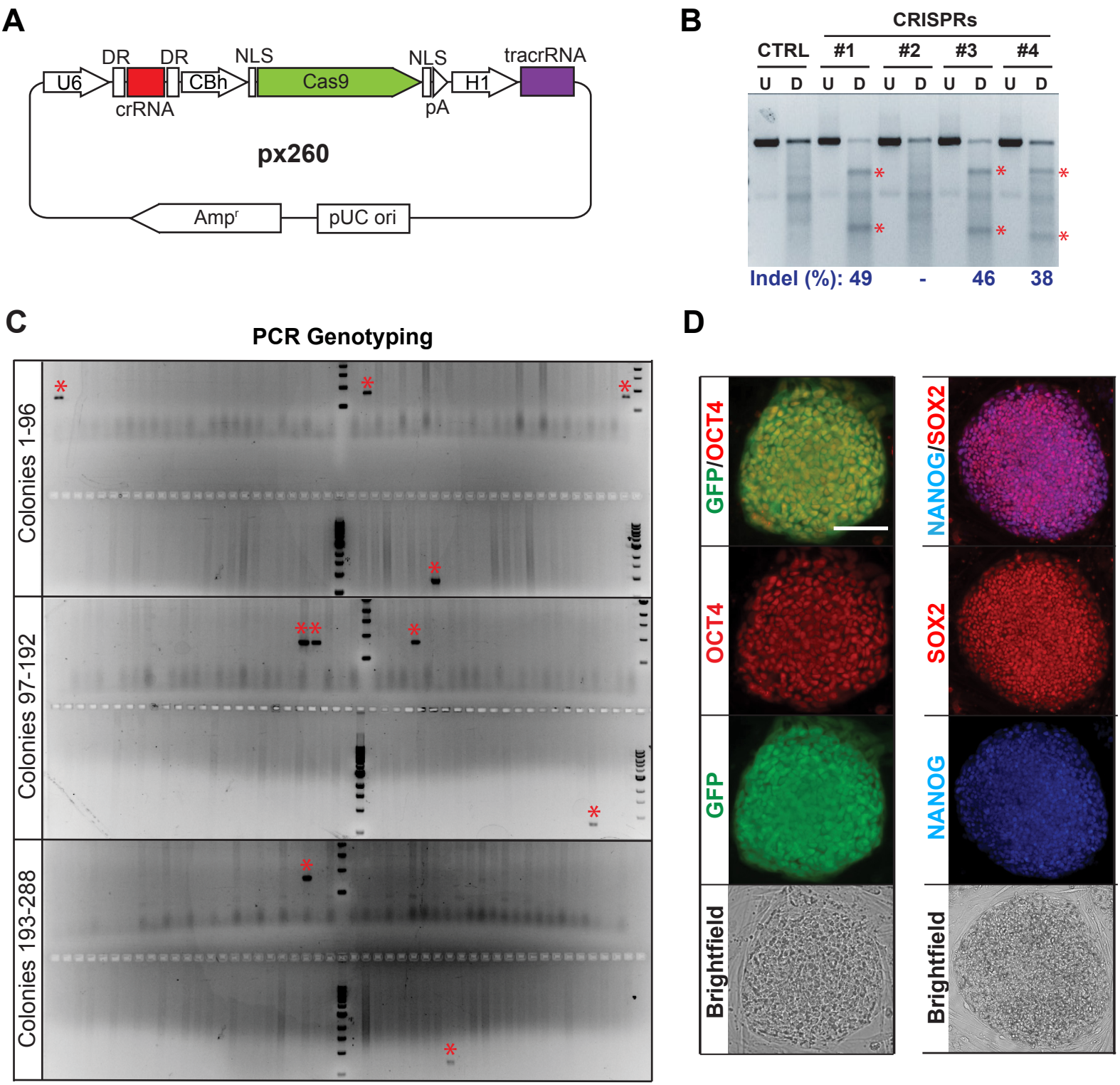

A
DR
DR
NLS
NLS
tracrRNA
U6
CBh
Cas9
H1
crRNA
pA
px260
Amp^r
pUC ori
B
CRISPRs
CTRL
#1
#2
#3
#4
U D U D U D U D U D
Indel (%): 49 - 46 38
C
PCR Genotyping
Colonies 1-96
Colonies 97-192
Colonies 193-288
D
GFP/OCT4
OCT4
GFP
Brightfield
NANOG/SOX2
SOX2
NANOG
Brightfield

Figure S2

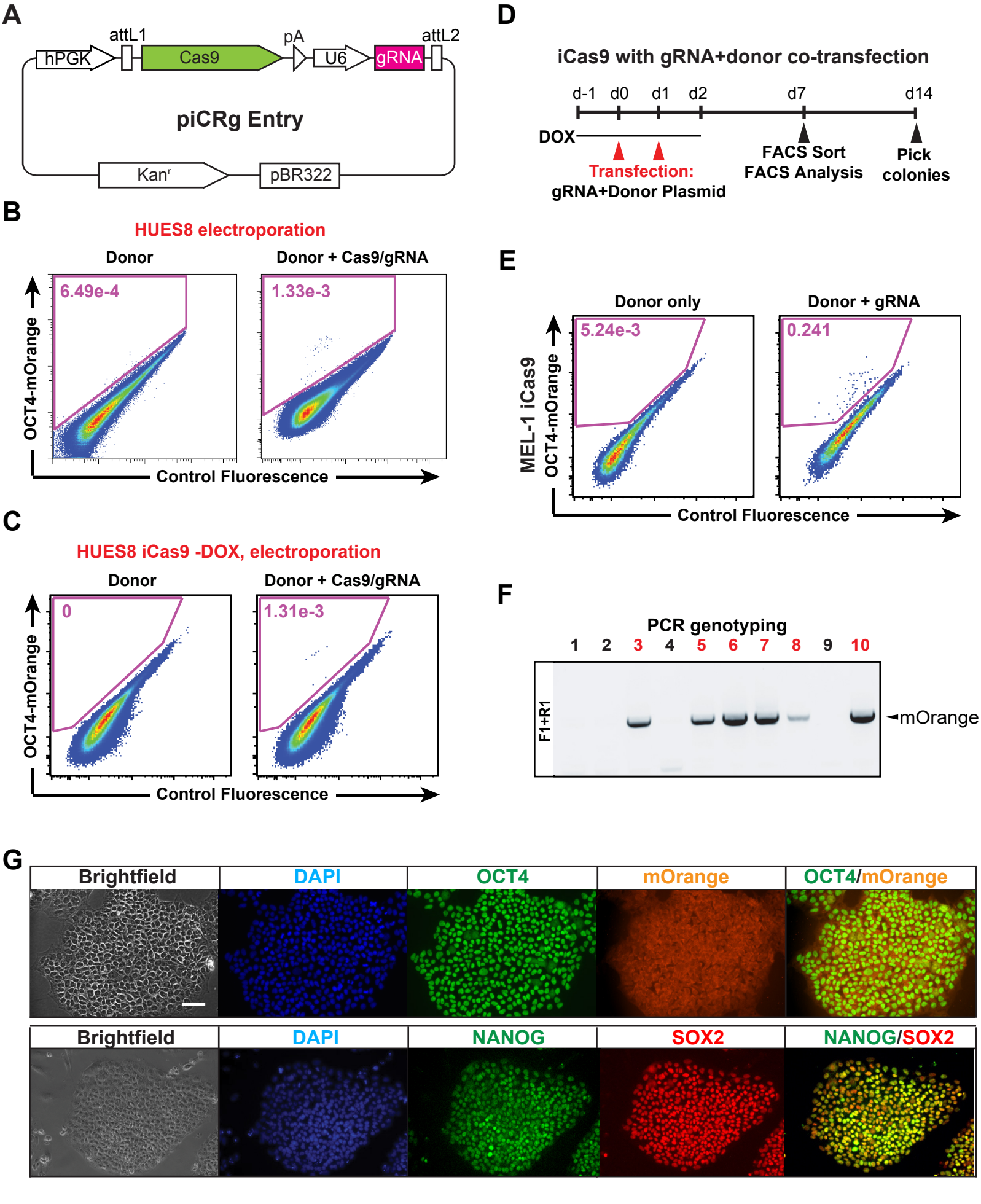
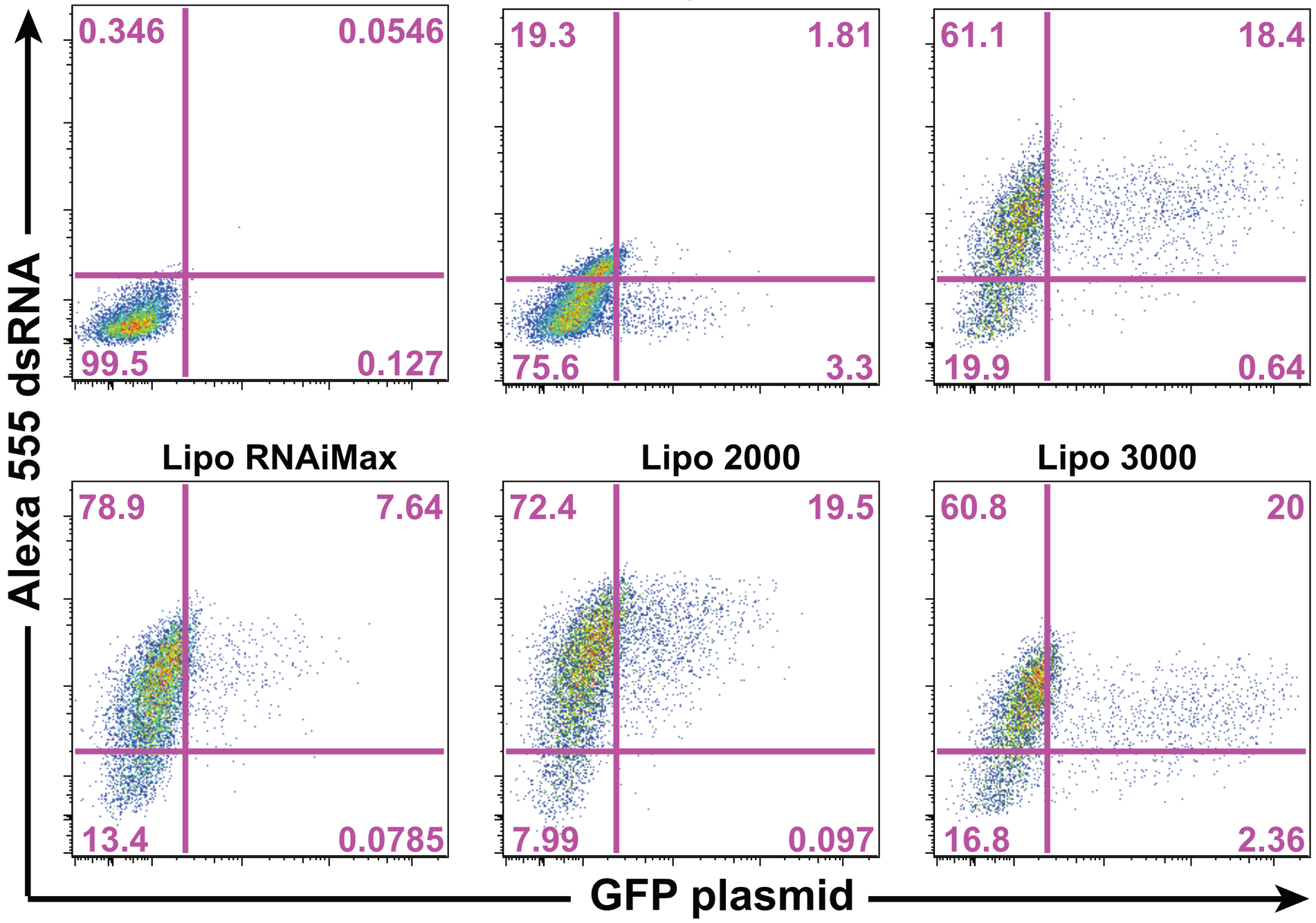

A
attL1
pA
attL2
hPGK
Cas9
U6
gRNA
piCRg Entry
Kanr
pBR322
D
iCas9 with gRNA+donor co-transfection
d-1 d0 d1 d2 d7 d14
DOX
Transfection:
gRNA+Donor Plasmid
FACS Sort
FACS Analysis
Pick
colonies
B
HUES8 electroporation
Donor
Donor + Cas9/gRNA
6.49e-4
1.33e-3
OCT4-mOrange
Control Fluorescence
E
Donor only
Donor + gRNA
5.24e-3
0.241
MEL-1 iCas9
OCT4-mOrange
Control Fluorescence
C
HUES8 iCas9 -DOX, electroporation
Donor
Donor + Cas9/gRNA
0
1.31e-3
OCT4-mOrange
Control Fluorescence
F
PCR genotyping
1 2 3 4 5 6 7 8 9 10
F1+R1
mOrange
G
Brightfield
DAPI
OCT4
mOrange
OCT4/mOrange
Brightfield
DAPI
NANOG
SOX2
NANOG/SOX2

Figure S3

A


None-Transfected
Electroporation
JetPrime
0.346
0.0546
99.5
0.127
19.3
1.81
75.6
3.3
61.1
18.4
19.9
0.64
Lipo RNAiMax
Lipo 2000
Lipo 3000
78.9
7.64
13.4
0.0785
72.4
19.5
7.99
0.097
60.8
20
16.8
2.36
Alexa 555 dsRNA
GFP plasmid

B

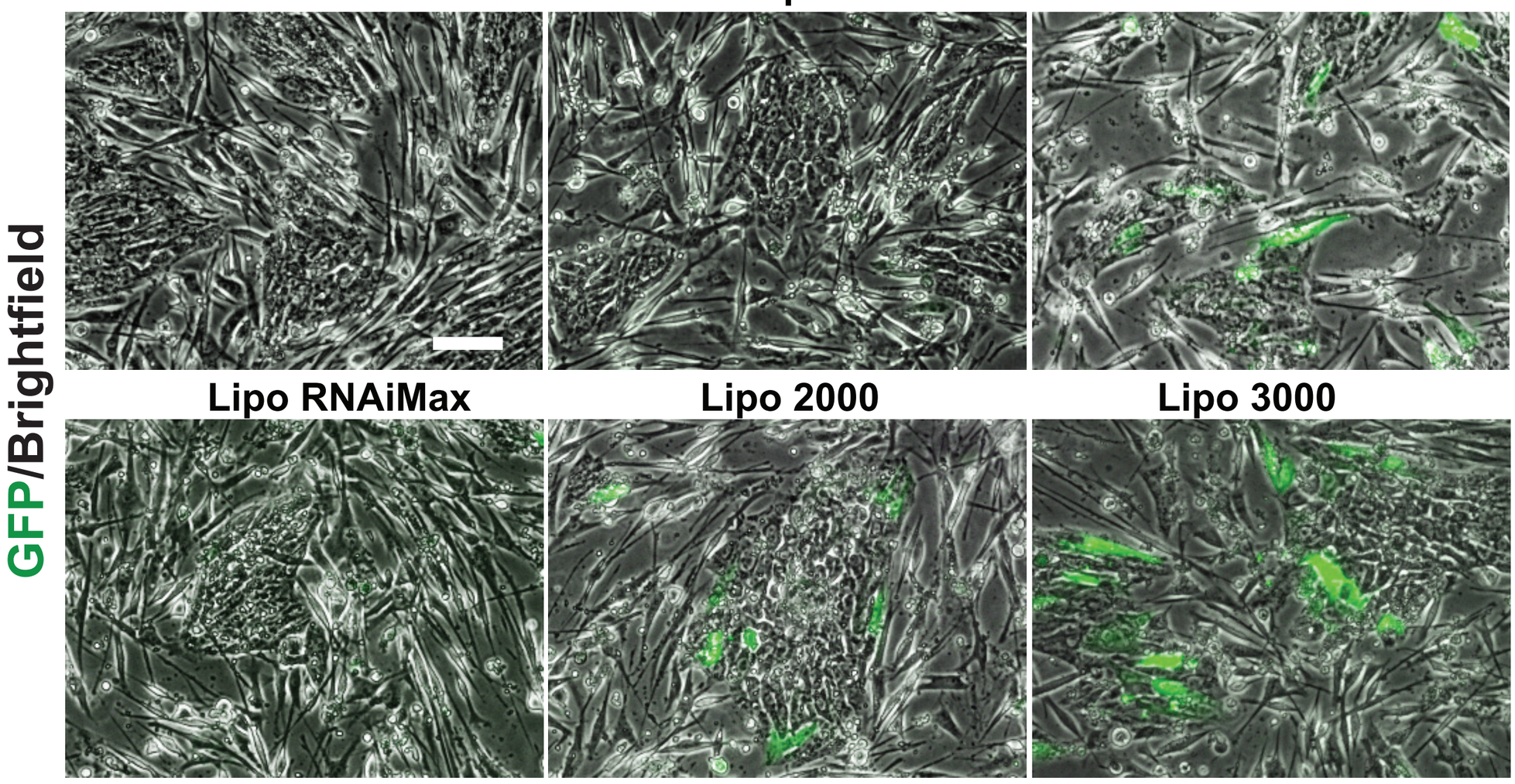
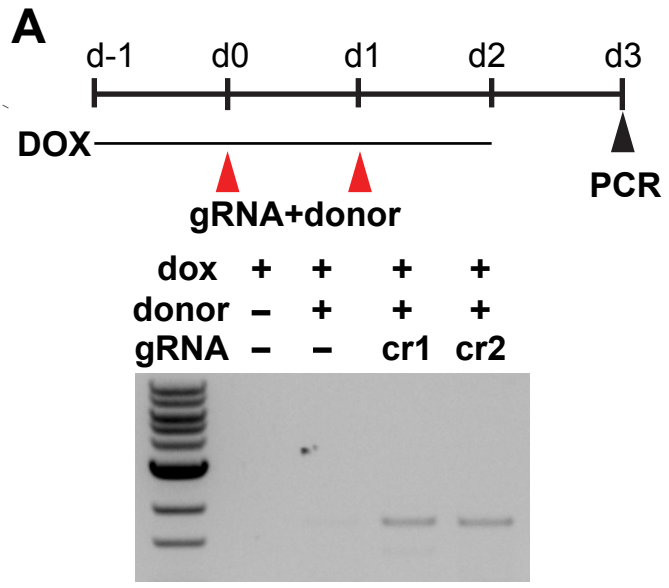
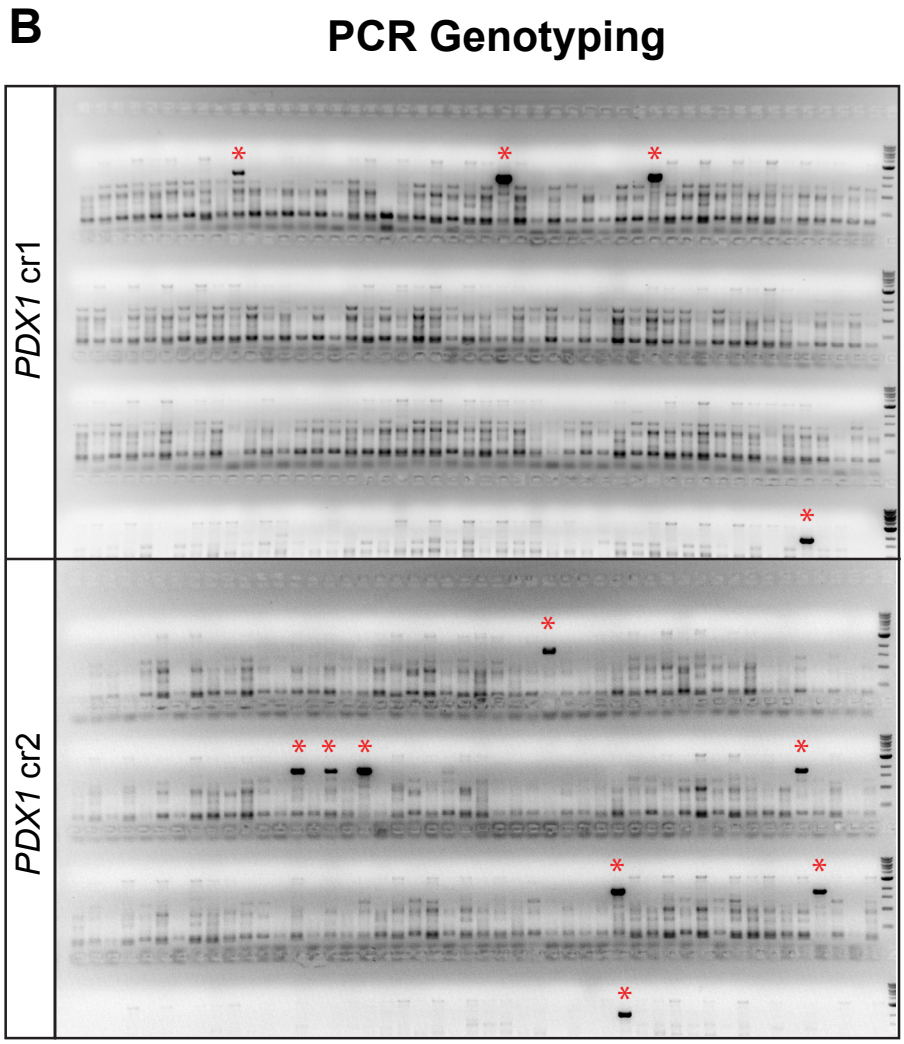

None-Transfected
Electroporation
JetPrime
Lipo RNAiMax
Lipo 2000
Lipo 3000
GFP/Brightfield

Figure S4


A
d-1
d0
d1
d2
d3
DOX
PCR
gRNA+donor
dox + + + +
donor – + + +
gRNA – – cr1 cr2


B
PCR Genotyping
PDX1 cr1
PDX1 cr2

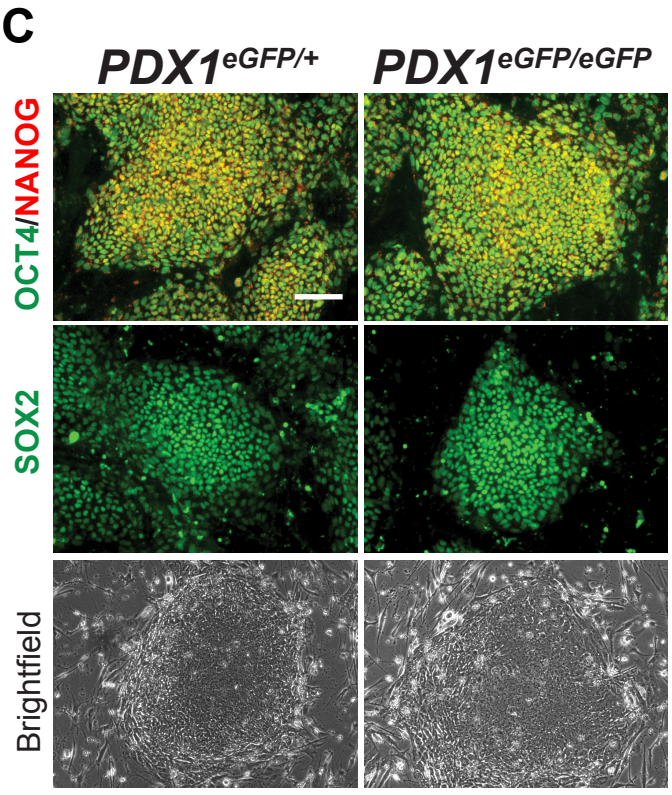

C
PDX1eGFP/+
PDX1eGFP/eGFP
OCT4/NANOG
SOX2
Brightfield

Figure S1. CRISPR-mediated targeting of the *OCT4* locus through drug selection.

(A) The vector map of the px260 plasmid for expressing Cas9 and crRNA/tracrRNA. (B) Four crRNAs were designed to target the stop codon of the *OCT4* locus. Each crRNA was cloned into the px260 vector and transfected into 293T cells. Two days after transfection genomic DNA was collected and Surveyor analysis was used to estimate the efficiency of each CRISPR in generating Indels. Asterisks indicated the cleavage products and the estimated Indel frequencies were labeled in blue. CTRL: px260 vector control; U: undigested control; D: digestion reaction with Surveyor nuclease. **(C)** PCR genotyping results showing 10 positive clones (indicated by red asterisks) identified based on the presence of a correct PCR product (811 bp). **(D)** *OCT4-eGFP* reporter hESCs were stained for pluripotency markers OCT4, NANOG and SOX2, which overlapped with GFP expression. The GFP signal was detected using a GFP antibody. Scale bar = 100 μm.

Figure S2. Targeting the *OCT4* locus without drug selection.

(A) The vector map of the piCRg Entry plasmid for expressing Cas9 and the chimeric gRNA. **(B)** FACS analysis for OCT4-mOrange expressing cells in HUES8 cells after electroporation of the OCT4-mOrange donor and the Cas9/gRNA plasmids. **(C)** FACS analysis for OCT4-mOrange expressing cells in HUES8 iCas9 cells without doxycycline treatment after electroporation of the OCT4-mOrange donor and the Cas9/gRNA plasmids. **(D)** Timeline for establishing hESC reporter lines using iCas9 hESCs. **(E)** FACS analysis for OCT4-mOrange expressing cells in MEL-1 iCas9 cells treated with doxycycline and transfected with the OCT4-mOrange plasmid and *OCT4* cr1 gRNA. **(F)** PCR genotyping for OCT4-mOrange gene targeting. Correctly targeted clones are

indicated in red. **(G)** *OCT4-mOrange* reporter hESCs were stained for pluripotency markers OCT4, NANOG and SOX2, which overlapped with mOrange expression. Scale bar = 100 μm.

Figure S3. Optimizing co-transfection of DNA and RNA into hESCs.

(A) FACS analysis of hESCs co-transfected with Alexa 555 dsRNA and a GFP-expressing plasmid (~ 10 kb) using four commonly used transfection reagents along with electroporation. Lipofectamine 3000 consistently performed better than the other conditions. Although comparable efficiencies could sometimes be achieved using JetPrime (as shown here), the outcomes were variable and appeared to relate to the amount of DNA used for transfection. **(B)** Stronger GFP expression was detected in hESCs transfected with Lipofectamine 3000, suggesting that increased copy number of GFP-expressing plasmid was transfected into the cell.

Figure S4. Generation of *PDX1-eGFP* reporter hESCs without drug selection.

(A) Population level PCR analysis of HUES8 iCas9 cells two days after transfection with gRNA and donor plasmid. **(B)** PCR genotyping of replated clonal hESCs after transfection. Correctly targeted cells are labeled with a red asterisk. **(C)** *PDX1-eGFP* reporter hESC lines displayed proper expression of pluripotency markers OCT4, NANOG and SOX2 and typical hESC morphology. The brightfield images were taken from live cells, and do not correspond to the immunofluorescence images of fixed cells.

Table S1. Off-target analysis

Sequencing analysis of potential off-target sites in *OCT4-eGFP* and *OCT4-mOrange* reporter lines

Gene	CRIPSR Target Sequence-PAM	Sequencing results in *OCT4-eGFP* reporter lines								Sequencing results in *OCT4-mOrange* reporter lines					
		1	2	4	5	6	7	9	10	3	5	6	7	8	10
***OCT4* (intended target)**	TCTCCCATGCATTCAAACTG-AGG														
DLG2	AAGCTCAGGCATTCAAACTG-TGG	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
GPHN	GCCCTCAGGCATTCAAACTG-TGG	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
IMMP2L	TAGACTTAGCATTCAAACTG-AGG	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
PEMT	GCACCCTAGCATTCAAACTG-TGG	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
POLR2J4	AAGGAGAAGCATTCAAACTG-TGG	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
POU5F1P4	TCTCCCATGCATTCAAACTG-AGG	WT	HET*	WT	HOM**	WT	HOM***	WT	WT	WT	WT	WT	WT	WT	WT
SLC33A1	CAGAAATGGCATTCAAACTG-CGG	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT

WT: Both alleles are wild-type; HET: One allele has a mutation; HOM: Both alleles have mutations;
* 6 bp insertion; ** 4 bp deletion; *** 10 bp deletion.

Sequencing analysis of potential off-target sites in *PDX1-GFP* reporter line # 9 generated using *PDX1* cr1

Gene	CRISPR Target Sequence-PAM	Sequencing results in *PDX1-GFP* reporter line #9
***PDX1* (intended target)**	GCCGCAGGAACCACGATGAG-AGG	
SUSD4	TCCTCAGGACCCACGATGGG-CAG	WT
COX7C	TCCGTAGGAGCCACTATGAG-GAG	WT
DCPS	GCCACAGGCACCACGGTGAG-GAG	WT
TRAF3IP2	GCCCCTGGGACCACGAAGAG-AGG	WT
NARF	GCCGCAGGACCCACGAGAAG-CGG	WT
ILVBL	GCAGCAGAAACCAGGAGGAG-AGG	WT
BST2	GCGGCAGGAGCCAGGACGAG-CAG	WT
SNED1	GCGGCGGGAACCCGGATGAG-GAG	WT
PSG3	CCTGCAGGAACCAGGATAAG-AGG	WT
CRAT	GGGGCAGGAACCAAGCTGAG-TGG	WT

Sequencing analysis of potential off-target sites in *PDX1-GFP* reporter line #4 generated using *PDX1* cr2

Gene	CRISPR Target Sequence-PAM	Sequencing results in *PDX1-GFP* reporter line #4
***PDX1* (intended target)**	GCCTCTCATCGTGGTTCCTG-CGG	
PSG3	TCCTCTTATCCTGGTTCCTG-CAG	WT
NUBP1	GCATCTCCCAGTGGTTCCTG-AGG	WT
SYNM	GTGTCTGACCGTGGTTCCTG-GAG	WT
HEXDC	GCCACTGCACGTGGTTCCTG-AGG	WT
PSG10P	TTCTCTTATCCTGGTTCCTG-CAG	WT
KCNQ1	TGCTCTCACCTTGGTTCCTG-GGG	WT
FBRSL1	GCCAGGCATCGGGGTTCCTG-CAG	WT
FOXJ3	ACCTGCCATCGTTGTTCCTG-GAG	WT
OR6M1	ATCTCTCTTCGTGGTTCTTG-TGG	WT
MAGEA8	ACCTCACAGCCTGGTTCCTG-GAG	WT

Table S2. Oligonucleotides used in this study

Oligonucleotides for generating Cas9/crRNA/tracrRNA or Cas9/gRNA expressing plasmids

Gene		CRISPR target sequence (5' of PAM)	Vector	Oligos used for generating the CRISPR constructs (5' to 3')
OCT4	cr1-dp	CACTCTGGGCTCTCCCATGCATTCAAACTG	px260	**F:** AAACCACTCTGGGCTCTCCCATGCATTCAAACTGGT
				R: TAAAACCAGTTTGAATGCATGGGAGAGCCCAGAGTG
	cr2-dp	TGCATTCAAACTGAGGTGCCTGCCCTTCTA	px260	**F:** AAACTGCATTCAAACTGAGGTGCCTGCCCTTCTAGT
				R: TAAAACTAGAAGGGCAGGCACCTCAGTTTGAATGCA
	cr3-dp	ATGCATTCAAACTGAGGTGCCTGCCCTTCT	px260	**F:** AAACATGCATTCAAACTGAGGTGCCTGCCCTTCTGT
				R: TAAAACAGAAGGGCAGGCACCTCAGTTTGAATGCAT
	cr4-dp	CTCTGGGCTCTCCCATGCATTCAAACTGAG	px260	**F:** AAACCTCTGGGCTCTCCCATGCATTCAAACTGAGGT
				R: TAAAACCTCAGTTTGAATGCATGGGAGAGCCCAGAG
	cr1	TCTCCCATGCATTCAAACTG	piCRg Entry	**F:** CACCGCTCCCATGCATTCAAACTG
				R: AAACCAGTTTGAATGCATGGGAGC
PDX1	cr1	GCCGCAGGAACCACGATGAG	piCRg Entry	**F:** CACCGCCGCAGGAACCACGATGAG
				R: AAACCTCATCGTGGTTCCTGCGGC
	cr2	GCCTCTCATCGTGGTTCCTG	piCRg Entry	**F:** CACCGCCTCTCATCGTGGTTCCTG
				R: AAACCAGGAACCACGATGAGAGGC

PCR Primers for generating templates for gRNA *in vitro* transcription

Gene	CRISPR target sequence (5' of PAM)	CRISPR specific forward primer (5' to 3')
OCT4	cr1 TCTCCCATGCATTCAAACTG	**F:** TAATACGACTCACTATAGGGCTCCCATGCATTCAAACTG
PDX1	cr1 GCCGCAGGAACCACGATGAG	**F:** TAATACGACTCACTATAGGGCCGCAGGAACCACGATGAG
	cr2 GCCTCTCATCGTGGTTCCTG	**F:** TAATACGACTCACTATAGGGCCTCTCATCGTGGTTCCTG

Universal reverse primer **gRNA-R:** AAAAGCACCGACTCGGTGCC

PCR Primers for donor plasmid construction

Donor plasmid		Primer sequence (5' to 3')
OCT4	Nh2AOr-F	TTCTAGCTAGCACCGGTGCCACGAACTTCTCTCTGTTAAAGCAAGCAGGAGACGTGGAAGAAAACCCCGG TCCTATGGTGAGCAAGGGCGAGGAG
	AscOr-R	CTTATGGCGCGCCTTACTTGTACAGCTCGTCCAT
PDX1-eGFP	HL-F	GTCAGTGAATTCAGGACCACTCATTGGCAGAG
	HL-R	GTCAGTCTGCAGGCTAGCTCGTGGTTCCTGCGGCCGCC
	HR-F	GTCAGTCTGCAGGGCGCGCCGAGGCAGGAGCTGCTCCTGG
	HR-R	GTCAGTGGATCCACTCACTGTATTCCACTGGC
	2A-GFP-F	CAGTGCTAGCGCCACTAACTTCTCCCTGTT
	2A-GFP-R	TCAGTCGGCGCGCCTTACTTGTACAGCTCGTCCA

PCR primers for genotyping

Gene	Primer sequence (5' to 3')
OCT4	**F:** AGTCCAAAGCTTGCCCTTGTCACC
	GFP-R: AGCTCCTCGCCCTTGCTCACC
	mOr-R: GAGGTGATGTCCAACTTGATGCCGA
PDX1	**F:** ATTTGCTGGCTCTCAGGTTG
	R: GTTCACCTTGATGCCGTTCT

PCR primers for generating Southern blot probes

Gene	Probe	Primer sequence (5' to 3')
OCT4	External	**F:** CCAGTGGGAGTCAGTGGGGCT
		R: GTCCGACTCCCAAGAGGTCACAG
	Internal Puro	**F:** TGACCGAGTACAAGCCCACGG
		R: TCGTAGAAGGGGAGGTTGC
	Internal mOrange	**F:** TGGCCATCATCAAGGAGTTCA
		R: CTTCTTCTGCATTACGGGGCCG
PDX1	External	**F:** AACACCTCTGAGGGCATTTG
		R: CGGACACTGCAGGTCAGTTA
	Internal GFP	**F:** GTTCATCTGCACCACCGG
		R: CGCGCTTCTCGTTGGGGT

Supplemental Experimental Procedures

Generation of constructs

To generate CRISPR plasmids expressing Cas9 and the crRNA/tracrRNA duplex targeting specific genomic loci, 30-bp protospacer sequences were cloned into the px260 (Addgene: 42229) as previously described (Cong et al., 2013). Briefly, vectors were digested with BbsI, treated with Antarctic Phosphatase, and gel purified. A pair of oligonucleotides containing the 30-bp protospacer sequence was annealed generating BbsI overhangs, and cloned into BbsI-digested, dephosphorylated vectors. The same procedure was also used to generate CRISPR plasmids expressing Cas9 and the chimeric gRNA with the difference that a pair of oligonucleotides containing the 20-bp protospacer sequence was cloned into the piCRg Entry plasmid (Addgene: 58904). The sequences for all oligonucleotides used for generating the CRISPR constructs are listed in Table S2.

For generation of *OCT4-eGFP* hESC reporter lines, the OCT4-2A-eGFP-PGK-Puro plasmid (Addgene: 31938) was used. The left homology arm is 697 bp and the right homology arm is 699 bp. The *OCT4-mOrange* hESC reporter lines were made using the OCT4-2A-mOrange targeting vector. To generate the OCT4-2A-mOrange targeting vector, an NheI-2AmOrange-AscI cassette was PCR amplified using the mOrange-pBAD plasmid template (Addgene: 54751) and primers Nh2AOr-F and AscOr-R. Next, the NheI-2A-mOrange-AscI PCR fragment and the OCT4-2A-eGFP-PGK-Puro plasmid were digested with NheI and AscI and ligated.

For constructing the PDX1-GFP donor plasmid, the 725-bp left (Primers: HL-F and HL-R) and 543-bp right (Primers: HR-F and HR-R) homology arms were PCR

amplified from the HUES8 genomic DNA and cloned into the pBlueScript SKII (+) plasmid to generate the pBS-PDX1 plasmid. The 2A-eGFP insert was PCR amplified from the OCT4-2A-eGFP-PGK-Puro plasmid (Addgene: 31938) using 2A-eGFP-F and 2A-eGFP-R primers, digested and cloned into the pBS-PDX1 plasmid to generate the PDX1-eGFP donor plasmid. In the donor plasmid, the 2A-eGFP sequences was fused in-frame to the last codon of *PDX1*. The sequences for all primers used for generating the donor plasmids are listed in Table S2.

Cell Culture

HUES8 (NIHhESC-09-0021) and MEL-1 (NIHhESC-11-0139) hESCs were cultured on irradiated mouse embryonic fibroblasts (iMEFs) feeder layers in DMEM/F12 medium (Life Technologies) supplemented with 20% KnockOut Serum Replacement (Life Technologies), 1X MEM Non-Essential Amino Acids (Life Technologies), 1X GlutaMAX (Life Technologies), 100U/ml Penicillin and 100 µg/ml Streptomycin (Gemini), 0.055 mM 2-mercaptoethanol (Life Technologies) and 10 ng/ml recombinant human basic FGF (EMD Millipore). Cells were incubated at 37 °C with 5% (vol/vol) CO_2, and media was changed daily. Cultures were passaged at a 1:6 - 1:12 split ratio every 4 - 6 days using 0.05% trypsin/EDTA. 5 µM Rho-associated protein kinase (ROCK) inhibitor Y-27632 (Selleck Chemicals) was added into the culture medium when passaging or thawing cells.

293T cells were maintained in DMEM supplemented with 15% fetal bovine serum, 1X GlutaMAX, 1X MEM NEAA and 1mM Sodium Pyruvate (Life Technologies).

Differentiation of hESCs

For differentiation of *OCT4* reporter lines, undifferentiated hESCs cultured on iMEF feeder layer were first adapted to the feeder-free E8 culture. Briefly, hESCs were passaged using 0.05% trypsin/EDTA and plated at a 1:3 split ratio on Matrigel (BD Biosciences) coated plates in E8 medium (Stem Cell Technologies). Cells were incubated at 37 °C with 5% (vol/vol) CO2, and media was changed daily. After 2 passages in E8 medium, the cells were differentiated: 2 days after passaging, cells were treated with 10 μM SB431542 (Sigma Aldrich) and 10 ng/mL BMP4 (R&D Systems) to initiate differentiation. After 3 days of this treatment, eGFP and mOrange fluorescence was analyzed by flow cytometry and fluorescence microscopy.

For differentiation into pancreatic lineages, undifferentiated hESC were cultured until 50-70% confluency, washed with PBS w/$Ca^{2+}Mg^{2+}$ once and treated with 2 μM BIO-acetoxime (Tocris), 100 ng/ml Activin A (R&D) in advanced RPMI (Invitrogen) for 1 day. The medium was changed to 100 ng/ml Activin A in advanced RPMI with 0.2% FBS for 2 days. On day 3, cells were treated with 50 ng/ml FGF10 (R&D) and 0.25 μM SANT1 (Sigma) in advanced RPMI with 2% FBS for 2 days. On day 5, cells were treated with 50 ng/ml FGF10, 0.25 μM SANT1, 2 μM Retinoic Acid (Sigma) and 250 nM LDN (Stemgent) in DMEM with 1% B27 (Invitrogen) for 4 days. Then on day 9, cells were treated with 1 μM Alk5 inhibitor (Axxora), 100 ng/mL Noggin (R&D system) and 1 μM DAPT (Tocris) in DMEM with 1% B27 for 4 days. After that, cells were kept in DMEM with 1% B27 for another 4 days and were examined for the expression of pancreatic cell specific markers.

In vitro transcription of gRNAs

A T7 promoter was added to gRNA templates by PCR amplification on piCRg Entry vectors using CRISPR-specific forward primers and a universal reverse primer gRNA-R (Table S2). T7-gRNA PCR products were used as templates for *in vitro* transcription using the MEGAshortscript T7 kit (Life Technologies). The resulting gRNAs were purified using the MEGAclear kit (Life Technologies), eluted in RNase-free water and stored at -80°C until use.

Electroporation

HUES8 hESCs were pre-treated with Rho-associated protein kinase (ROCK) inhibitor Y-27632 (Selleck Chemicals) 24 hours before electroporation. On the day of electroporation, hESCs were disassociated into single cells with 0.25% Trypsin/EDTA and filtered through a 40 μM cell strainer to remove cell clumps. 10 million cells were re-suspended in 800 μL cold PBS and mixed with targeting and donor plasmids (10 μg CRISPR targeting plasmid and 40 μg donor plasmid). Cells were electroporated using the Gene Pulser XCeII (Bio-Rad) at 250 V, 500 μF in a 0.4 cm Gene Pulser cuvette (Bio-Rad). Cells were recovered and re-plated on irradiated mouse embryonic fibroblast (iMEF) coated plates with ROCK inhibitor.

For the generation of *OCT4-eGFP* lines, two days after electroporation cells were treated with puromycin (0.5 μg/mL) for three days. Two weeks after electroporation individual colonies were picked and expanded for PCR genotyping, Southern blot analysis and sequencing for the establishment of reporter lines. Three of the correctly targeted clones (#1, #4 and #7) were electroporated with Cre recombinase following the same procedure described above. After electroporation with 50 μg of Cre recombinase

plasmid, the cells were recovered and replated on iMEF with ROCK inhibitor. 3 days after electroporation, GFP-expressing (GFP^+) cells were observed. GFP^+ cells were isolated by FACS and plated at a low density (2,000 cell/10cm dish) for subsequent picking and expansion of individual clones.

Transfection

HUES8 and MEL-1 iCas9 hESCs (Gonzalez et al., 2014) were treated with ROCK inhibitor and doxycycline one day before transfection. For transfection, cells were dissociated using TrypLE (Life Technologies), replated at 200,000 hESCs per well in iMEF-coated 12-well plates and transfected in suspension with gRNAs and donor plasmid. A second transfection was performed 24 hours later. Transfection of the gRNAs and donor plasmid into hESCs was performed using Lipofectamine 3000 (Life Technologies) following manufacturer's guidelines. For each targeting, gRNAs at a 10 nM final concentration and 5 μg of donor plasmid were used. Lipofectamine 3000 and gRNA + donor plasmid were diluted separately in Opti-MEM (Life Technologies), mixed together, incubated for 5 min, and added drop-wise to cultured hESCs.

Assessment of Indel mutations using the Surveyor Nuclease Assay

80-90% confluent 293T cells in 6-well dishes were transfected using JetPrime transfection reagent (Polyplus) following manufacturer's guidelines. For each well, a total of 2 μg CRISPR plasmid was transfected. After transfection the cells were incubated at 37 °C for 2 days prior to genomic DNA extraction. The Surveyor Kit (Transgenomic) was used to test the efficiency of each crRNA for producing Indels. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's protocol. The genomic region flanking the CRISPR target site for each gene was PCR

amplified. For Surveyor analysis 16 µL of PCR product (at 25 ng/µL) was denatured and re-annealed in JumpStart buffer to a total volume of 20 µL using the following protocol: 95 °C, 5 min; 95–85 °C at −2 °C/s; 85–25 °C at −0.1 °C/s; hold at 4 °C. 10 µL of hybridized PCR products was treated with 1 µL of Surveyor Enhancer S and 1 µL of Surveyor Nuclease S at 42°C for 60 minutes. The reaction was stopped by adding 1.2 µL of Stop solution to each tube. Products were then analyzed on 2.5% agarose gel and imaged with a Gel Doc gel imaging system (Bio-Rad). Quantification was based on relative band intensities using ImageJ. Indel mutation percentage was determined by the formula: $100 \times (1 - (1 - (b + c) / (a + b + c))^{1/2})$, where *a* is the integrated intensity of the undigested PCR product, and b and c are the integrated intensities of each cleavage product.

Southern Blot Analysis

The external and internal probes were generated by PCR using the PCR DIG Probe Synthesis Kit (Roche). For the external probes we used HUES8 genomic DNA as the template. For internal probe generation we used the OCT4 2A-eGFP-PGK-Puro donor, OCT4-2A-mOrange donor and PDX1-2A-eGFP donor templates. For membrane hybridization, 5 µL of denatured DIG-labeled PCR product was added to 20 mL of hybridization buffer.

To identify correctly targeted hESC lines, genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen). 5-10 µg of genomic DNAs was digested overnight with the appropriate restriction enzymes (OCT4 external and mOrange internal: BamHI, puromycin internal: EcoRI, PDX1 external: ApaLI, eGFP internal: ApaLI) and then migrated in 1% agarose gels. The gel was denatured, neutralized, and transferred

overnight by capillarity on Hybond-N membranes (GE Healthcare) using 10x SSC transfer buffer. Hybridization with the external or the internal probe was carried out overnight at 65 °C. Probes were detected using an AP-conjugated DIG-Antibody (Roche) using CDP-Star (Roche) as a substrate for chemiluminescence as per manufacturer's instructions.

Immunofluorescence staining

For immunostaining, cells were fixed with 4% paraformaldehyde for 10 minutes, washed once with phosphate buffered saline (PBS) and permeabilized in PBST (PBS + 0.1% Triton) for 15 minutes. Blocking was done for 5 minutes at RT with blocking solution (5% donkey serum in PBST). Primary and secondary antibodies were diluted in blocking solution. Primary antibodies were incubated at RT for 1 hr. The following primary antibodies were used at a 1:100 dilution: OCT4 (mouse monoclonal, Santa Cruz sc-5279); NANOG (rabbit polyclonal, CosmobioJapan REC-RCAB0004P-F); SOX2 (goat polyclonal, Santa Cruz sc-17320), RFP (rabbit polyclonal, Life Technologies R10367), GFP (rabbit polyclonal, Life Technologies A-6455). The PDX1 antibody was used at a 1:500 dilution (goat polyclonal, R&D AF2419). After primary antibody staining the cells were washed three times with PBST and then incubated with the appropriate Molecular Probes Alexa Fluor dye conjugated secondary antibodies (Life Technologies) for 1 hr.

Reference:

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.