CRISPR/Cas-Mediated Knockin in Human Pluripotent Stem Cells

Nipun Verma, Zengrong Zhu, and Danwei Huangfu

Abstract

Fluorescent reporter and epitope-tagged human pluripotent stem cells (hPSCs) greatly facilitate studies on the pluripotency and differentiation characteristics of these cells. Unfortunately traditional procedures to generate such lines are hampered by a low targeting efficiency that necessitates a lengthy process of selection followed by the removal of the selection cassette. Here we describe a procedure to generate fluorescent reporter and epitope tagged hPSCs in an efficient one-step process using the CRISPR/Cas technology. Although the method described uses our recently developed iCRISPR platform, the protocols can be adapted for general use with CRISPR/Cas or other engineered nucleases. The transfection procedures described could also be used for additional applications, such as overexpression or lineage tracing studies.

Key words Human pluripotent stem cells (hPSCs), Gene targeting, CRISPR/Cas, Homologous recombination, Knockin, Fluorescent reporter, Epitope tag

1 Introduction

Human pluripotent stem cells (hPSCs), which include embryonic stem cells (hESCs) and induced pluripotent stem cells (hiPSCs), have two unique characteristics: the capacity for unlimited self-renewal in culture and the ability, called pluripotency, to form any cell type that is present in an adult human. Since the isolation of hESCs in 1998, there has been much progress in the development of protocols to differentiate hPSCs into specific cell types. As a result hPSCs have emerged as a valuable tool for studying human development and for modeling human disease. Furthermore there is great potential to use hPSCs to generate clinically important cell types, which can be transplanted into patients to replace those cells lost or damaged due to disease [1, 2].

The utility of hPSCs can be further enhanced through genetic modification of these lines. Through homologous recombination exogenous sequences can be inserted into the genome to generate a variety of "knockin" alleles, from a single nucleotide change to

introduce or correct a disease associated mutation, to the insertion of small (less than 100 base pairs, bp) epitope tags (e.g., hemagglutinin, 3× FLAG, V5 tags) or even the insertion of large (hundreds of bp) transgene constructs such as fluorescent reporters (e.g., GFP, mOrange) and antibiotic selection cassettes. The ability to insert epitope or fluorescent tags into the genome would greatly facilitate hPSC-based in vitro and in vivo (after transplantation in animal models) studies. The insertion of epitope tags into specific loci creates recombinant hybrids that contain a polypeptide affinity tag and enable efficient purification of target proteins. The development of these fusion proteins facilitates a variety of downstream experiments including: western blot, immunoprecipitation (IP), chromatin-immunoprecipitation (ChIP), and immunofluorescence staining, and is particularly useful when antibodies for the protein of interest are substandard or not available. Lineage-specific knockin fluorescent reporters are valuable tools because they allow real-time observation of gene expression dynamics, cell lineage tracing, and the isolation of a specific cell population for further analysis. hPSCs reporter lines have already been used to identify mediators of pluripotency [3] as well as characterize the culture conditions that stabilize different pluripotent states [4].

Creating knockin alleles in hPSCs used to be extremely challenging because of the low transfection efficiency [5] and low rate of spontaneous homologous recombination [6]. Thus traditional knockin strategies rely on the use of a drug-resistance cassette that allows for the enrichment of cells with the correct integration. However, due to potential interference between the drug-resistance cassette and nearby genes, the selection cassette must be removed prior to using these cells for experimental studies [7]. We have ourselves generated a knockin reporter allele at the *OCT4* locus using antibiotic selection [8]. Notably correctly targeted clones with the drug selection cassette did not express GFP. Only upon Cre-mediated deletion of the drug selection cassette, did all clonal lines show proper co-expression of pan-cellular eGFP with OCT4. Thus the low targeting efficiency of hPSCs necessitates an antibiotic selection cassette, but this in turn requires additional targeting, isolation, and characterization of clonal lines, effectively doubling the time and effort required to make the knockin line.

The recent emergence of programmable site-specific nucleases has greatly improved the efficiency of genome engineering in hPSCs. Zinc finger nucleases (ZFNs), transcription-activator like endonucleases (TALENs), and the CRISPR/Cas9 nuclease can all act as genomic scissors to produce a double strand break (DSB) at the desired genome loci [9–11]. Without a repair template, the DSB is repaired by the nonhomologous end joining (NHEJ) pathway that often generates small insertions or deletions (indels). These hPSC lines with site-specific knockout mutations can be used to determine the contribution of particular genes for development

and disease. In the presence of a repair template, DSB repair with error-free homology directed repair (HDR) can be used to generate knockin alleles through targeted integration of exogenous DNA sequences into the desired loci.

The CRISPR/Cas9 system was discovered as a form of adaptive immunity in bacteria, where it is used to destroy exogenous nucleic acids of invading viruses [12], and it has recently been adapted for genome engineering in mammalian cells. In this system the CRISPR 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 [13–16]. Cas9 produces a DNA DSB, which in the presence of a double stranded DNA (dsDNA) plasmid or single stranded DNA (ssDNA) template can promote HDR to efficiently incorporate exogenous sequences, such as a fluorescent reporter tag, into the specific genomic locus in hPSCs [8, 17–20]. We have developed an efficient genome-editing platform in hPSCs, called iCRISPR [21, 22]. Through TALEN-mediated gene targeting, a doxycycline inducible Cas9 cassette and the reverse tetracycline transactivator (M2rtTA) are integrated into both alleles of the endogenous *AAVS1* locus, and thus allow robust Cas9 expression in established clonal lines (referred to as iCas9 hPSCs) upon doxycycline treatment. With transfection of appropriate gRNAs we are able to generate site-specific knockout mutations. In addition, with co-transfection of repair templates, iCRISPR has also allowed us to specifically alter the sequence of a small number of nucleotides, and to insert a knockin epitope tag or a large transgenes such as a fluorescent tag. Most importantly due to the increased efficiency of the iCRISPR system it is not necessary to use selection to enrich for correctly integrated alleles. Thus it is feasible to generate knockin alleles in an efficient one-step procedure [8].

In this methods chapter we describe the procedure for using our iCRISPR system to generate knockin alleles in hPSCs. We specifically describe the procedure for generating epitope-tagged and fluorescent-tagged hPSC lines; however, the same knockin strategy could be used for a variety of other inserts. Furthermore, we believe the transfection approach that we use to generate the lines could also be used for other purposes, for example to deliver mRNA or plasmids for overexpression or lineage tracing studies.

2 Materials

The following materials are needed.

1. hPSC medium: DMEM/F12 medium (Life Technologies), 20% KnockOut Serum Replacement (Life Technologies), 1× Non-Essential Amino Acids (Life Technologies), 1× GlutaMAX

(Life Technologies), 100 U/mL Penicillin/100 μg/mL Streptomycin (Life Technologies), 0.055 mM 2-mercaptoethanol (Life Technologies), 10 ng/mL recombinant human basic FGF (Life Technologies).

2. TrypLE Select enzyme (Life Technologies).

3. Fetal Bovine Serum (FBS, Sigma-Aldrich).

4. Dimethyl Sulfoxide (DMSO, Santa Cruz Biotechnology).

5. Rho-associated protein kinase (ROCK) inhibitor Y-27632 dihydrochloride (Selleck Chemicals).

6. Genomic DNA lysis buffer: 10 mM Tris pH 8, 0.45 % NP40, 0.45 % Tween 20, and 100 μg/mL Proteinase K.

7. Lipofectamine RNAiMAX (Life Technologies).

8. Lipofectamine 3000 (Life Technologies).

9. Herculase II Fusion DNA Polymerase (Agilent Technologies).

10. MEGAshortscript T7 Transcription kit (Life Technologies).

11. MEGAclear Transcription Clean-Up Kit (Life Technologies).

12. DNeasy Blood and Tissue Kit (Qiagen).

13. Ethidium Bromide (EtBr).

14. Agarose (Invitrogen).

15. Purelink Gel Purification Kit (Life Technologies).

16. Pierce Crosslink Magnetic IP/Co-IP kit (Pierce Antibodies).

17. FLAG antibody, clone M2 (Sigma-Aldrich).

18. TBST: 20 mM Tris–HCl (pH 7.6), 137 mM NaCl, 0.1 % Tween 20.

19. Blocking Buffer: 5 % nonfat milk in TBST.

20. NuPAGE LDS Sample Buffer 4× (Life Technologies).

21. NuPAGE® Sample Reducing Agent 10× (Life Technologies).

22. NuPAGE Novex 3–8 % Tris–Acetate Protein Gels (Life Technologies).

23. NuPAGE Tris–Acetate SDS Running Buffer (20×) (Life Technologies).

24. Nitrocellulose Pre-Cut Blotting Membranes (Life Technologies).

25. Amersham ECL Prime Western Blotting Detection Reagent (Amersham).

26. Amersham Hyperfilm (Amersham).

27. Amersham Hypercassette (Fisher).

3   Methods

Here we describe the procedures to generate epitope and reporter tagged hPSC lines using the iCRSIPR platform. In the iCRISPR system a doxycycline inducible Cas9 cassette has been targeted into the endogenous *AAVS1* locus. Upon doxycycline treatment, Cas9 is expressed in all cells. After transfection of the site-specific gRNAs and the donor HDR template with the exogenous DNA sequence flanked by the homology arms, a DSB is created and HDR leads to incorporation of the insert into the genome. For single nucleotide alternations, a ssDNA template with homology arms of ~40–80 nt is generally used. Due to the small size of the epitope tags (less than 100 nt), the exogenous sequence can also be provided as a ssDNA, whereas the larger reporter tag (hundreds of bp) can be cloned easily into a donor plasmid for delivery into the cells. In both cases selection is not required and knockin alleles can be efficiently generated in a single step (Fig. 1). We have used the following protocols for HUES8 (NIHhESC-09-0021) and MEL-1 (NIHhESC-11-0139) hESC lines. It may be necessary to adjust the protocols for different hPSC lines or for hPSCs cultured under different conditions such as feeder-free TeSR and Essential 8 (E8) culture conditions [23, 24].
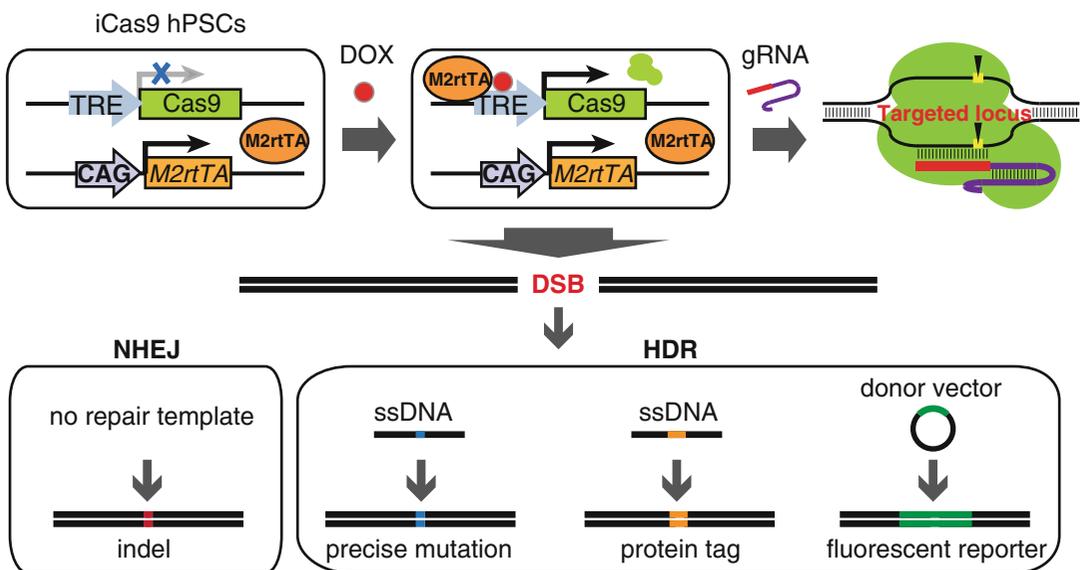
Fig. 1 The iCRISPR platform for genome editing in hPSCs. Doxycycline treatment induces Cas9 expression in iCas9 hPSCs. After transfection of gRNA, Cas9 is guided to the target locus via Watson-Crick base pairing and induces DNA DSBs. In the absence of a repair template, repair of DSBs by NHEJ often results in indels that can be used to knockout a target gene in hPSCs. Alternatively, in the presence of repair templates, either short ssDNA or long dsDNA donor vectors, HDR can be employed to incorporate exogenous sequences into the endogenous loci to introduce precise nucleotide mutations, protein tags, or fluorescent reporters into the target loci

3.1 hPSC Culture

1. hPSCs are cultured on an irradiated mouse embryonic fibroblast (iMEF) feeder layer in hPSC media.

2. iMEFs are plated 1 day before seeding hPSCs. Tissue culture dishes are coated with 0.1 % gelatin for 30 min at 37 °C prior to plating the iMEF feeder layer.

3. hPSCs cultures are generally passaged at 1:6 or 1:12 split ratios every 4–6 days. To passage hPSCs, colonies are disaggregated by treating with TrypLE Select enzyme for 5 min at 37 °C.

4. Dissociated hPSCs are collected in hPSC media and spun down at $200 \times g$ for 5 min. After spinning hPSCs are resuspended in fresh hPSC media and seeded on iMEF coated plates. 5 µM ROCK inhibitor Y-27632 is added into the culture medium when passaging or thawing frozen cells.

5. Cell lines are frozen down in freezing media consisting of 40 % FBS, 10 % DMSO and 50 % hPSC media.

3.2 Generation of Knockin Reporter hPSC Lines Using iCRISPR

The generation of knockin reporter hPSC lines using the iCRISPR system involves first the induction of Cas9 expression through administration of doxycycline followed by the co-transfection of the gRNA and donor vector into the iCas9 cells. The transfected gRNA forms a complex with the Cas9 protein and directs it to the target locus, where it creates a DSB. HDR can then be employed to efficiently incorporate the exogenous reporter sequence into the target locus. After dissociating and re-plating, single-cell colonies are isolated, expanded and screened through PCR for correct reporter integration. Correctly targeted clonal lines are further validated and differentiated to assess faithful reporter activity (Fig. 2).
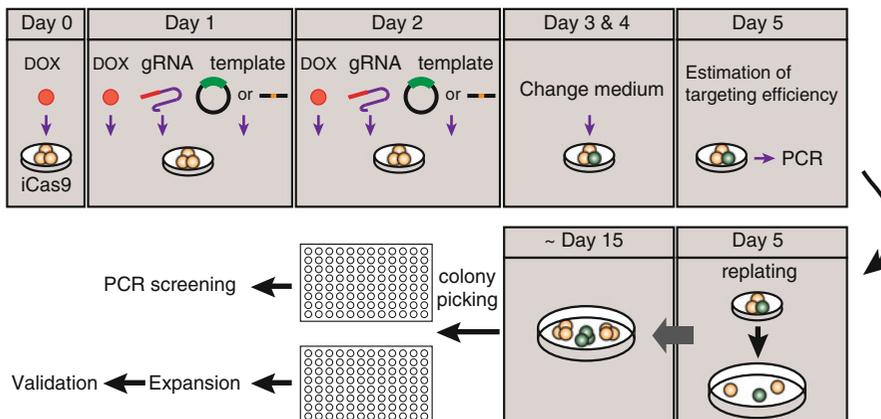
Fig. 2 The workflow for iCRISPR-mediated knockin in hPSCs. Doxycycline treatment 24 h before the first transfection induces Cas9 expression in iCas9 hPSCs at Day 0. Co-transfection of gRNA and repair templates, either donor vectors or ssDNAs, on Day 1 and Day 2 results in the incorporation of exogenous sequences into the endogenous loci. After PCR genotyping on Day 5, cells with the highest targeting efficiency will be re-plated as single-cells at a low density (~2000–5000 cells /100 mm dish) and allowed to grow for ~10–15 days. On Day 15, single-cell colonies will be picked, amplified, passaged into duplicated wells and subject to PCR screening. Correctly targeted clonal lines will be expanded and further validated

3.2.1 Donor Vector Design and Construction

Traditionally, the donor plasmid contains two homology arms flanking a promoterless reporter gene and a drug-resistance cassette driven by a constitutively active promoter. The two arms contain sequences that are homologous to the target locus. The constitutively active drug-resistance cassette enables enrichment of correctly targeted clones. However to avoid interference between the drug resistance cassette and neighboring genes, the drug-resistance cassette needs to be removed after identification of correctly targeted clones. Due to the high efficiency of the iCRISPR system, knockin reporter lines can be generated without the use of a drug selection cassette. This significantly reduces the time and effort required for establishment of knockin reporter hPSC lines.

1. Choose the target site for reporter integration in a way that ensures faithful reporter expression as well as minimizes the potential impact on endogenous gene expression and/or protein function. There are two general strategies for reporter gene integration, protein fusion (also called translational fusion) and promoter fusion (also called transcriptional fusion) (Fig. 3). For protein fusion, a reporter gene is integrated into the same reading frame of the target gene, transcribed as a single mRNA driven by the endogenous promoter and expressed as a single fusion protein. The protein fusion reporter is useful for monitoring protein subcellular localization, protein dynamics, and protein–protein interactions of the target gene. However, the fused reporter may affect the folding, stability, and function of the endogenous protein. Without previous knowledge about the functional domain and protein structure of the target gene, we recommend inserting the reporter gene either at the N-terminus immediately after the start codon or at the C-terminus immediately after the coding sequence. For promoter fusion, a reporter gene is also integrated into the same reading frame and transcribed together with the target gene as a single mRNA driven by the endogenous promoter. However, different from protein fusion reporter, promoter fusion reporter is preceded by either an IRES (internal ribosome entry site) or a 2A peptide sequence (Fig. 3). Consequently, the reporter and target genes will be translated into two separate peptides. Promoter fusion reporter reflects the endogenous gene expression pattern and usually has minimal effects on the endogenous gene expression or protein function. However, it is not suitable to monitor the subcellular localization, protein dynamics, and protein–protein interactions of the target gene. Whether to make a protein fusion or a promoter fusion reporter depends on the intended application. For monitoring gene expression during hPSC differentiation, we are primarily concerned with dynamic changes in gene expression and less about the subcellular localization of the protein. Thus for this purpose we generally use the promoter fusion strategy and integrate the reporter gene at
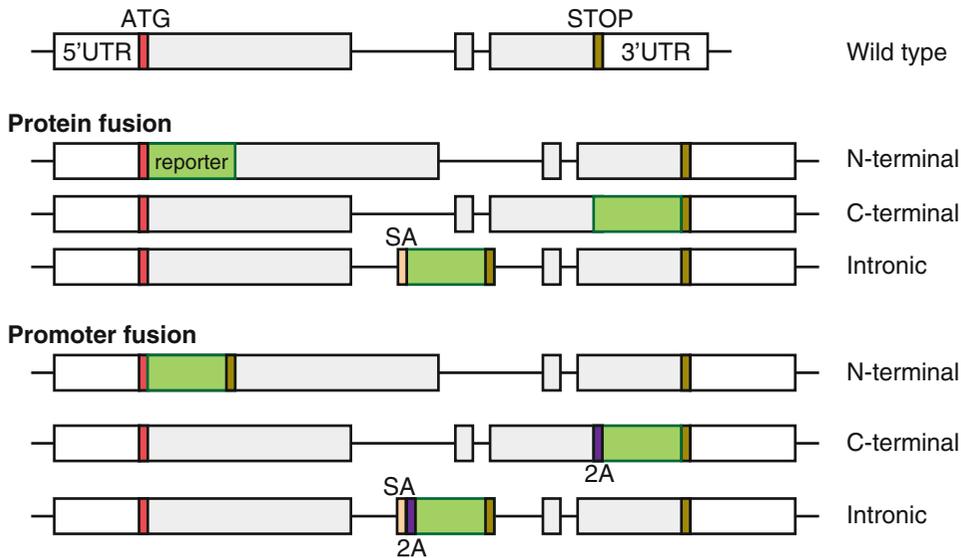
Fig. 3 Schematics of protein and promoter fusion knockin strategies. Here we illustrate typical targeting strategies for creating protein and promoter fusion knockin alleles. For a protein fusion reporter, the reporter gene (in *green*) is fused in-frame to the N-terminus of the endogenous gene immediately after the start codon (ATG, in *red*), to the C-terminus immediately after the last coding sequence or in the intronic region using a splicing acceptor sequence (SA, in *orange*). For a promoter fusion reporter, the reporter gene and the endogenous locus is linked through a 2A peptide (2A, in *purple*) and thus two separate peptides will be produced. The promoter fusion reporter can also be integrated into the N-terminus, C-terminus or intronic region of the endogenous gene. Boxes are exons, with open boxes indicating the untranslated region (UTR), filled grey boxes indicating the coding sequence (CDS); connecting lines are introns

the C-terminus immediately after the coding sequence of the endogenous gene as this will have the least effect on endogenous protein function (Fig. 4a).

2. After deciding the integration locus, PCR-amplify the two homologous arms (HA-L and HA-R) flanking the integration locus using high fidelity Herculase II Fusion DNA Polymerase (*see* **Note 1**) and clone into a backbone vector. We generally use pBluescript, which contains multiple cloning sites and is relatively small, as the backbone vector.

3. While the optimal length of the homologous arms has not been thoroughly investigated, approximately 500–1000 bp of homology arms on each side of the reporter gene are recommended (*see* **Note 2**).

4. Insert the reporter, for example a fluorescent protein reporter, preceded by a 2A self-cleavage sequence (2A-reporter) between the two homology arms. Make sure the 2A-reporter sequence is fused in-frame to the last coding sequence of the target gene (Fig. 4b).
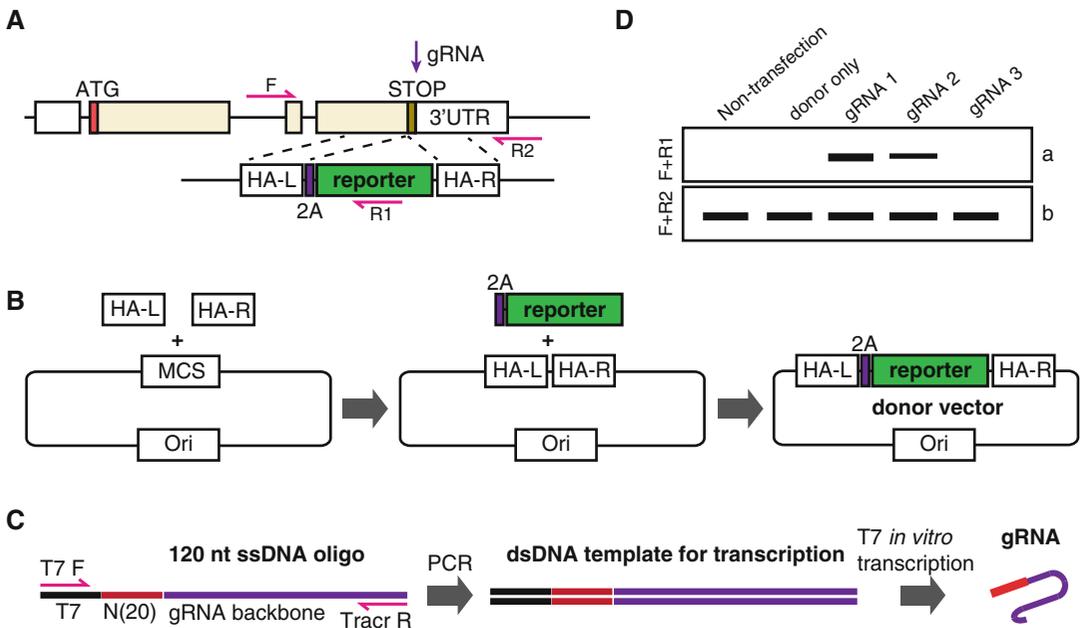
Fig. 4 Generation of knockin reporter hPSC lines using iCRISPR. (**a**) Schematics of the targeting strategy. In the presence of the donor vector, HDR results in the replacement of stop codon with 2A-reporter sequence. The PCR primers (F + R1 and F + R2) used for genotyping are indicated with *red arrows*. (HA-L and HA-R indicate left and right homology arms). (**b**) Schematics of the cloning strategy. The left and right homology arms (HA-L and HA-R) and the 2A-reporter gene are cloned sequentially into the backbone vector. (*MCS* multiple cloning site). (**c**) Schematics of the gRNA production. A 120 nt ssDNA containing a 20 nt T7 promoter sequence (T7, in *black*), a variable 20 nt gRNA recognition sequence ((N)20, in red) and a constant gRNA backbone sequence (in *purple*) is PCR amplified using T7 F and Tracr R universal primers (*red arrows*). The dsDNA PCR products will then be used as templates for in vitro transcription to produce the gRNAs. (**d**) Estimation of targeting efficiency. The PCR primers (F + R1 and F + R2) are indicated in panel (**a**). Comparison of the PCR product (band a) and (band b) allows an approximate estimation (a/(a + b) x 100 %) of the targeting efficiency by each gRNA

3.2.2 Design and Production of gRNAs

We routinely use the CRISPR design tool developed by the Feng Zhang group at MIT (http://crispr.mit.edu/). The software not only identifies all possible CRISPR targets in an input DNA sequence but also uncovers potential off-target sites, thus predicting gRNAs with the highest targeting specificity. For each target locus, we recommend designing 3 gRNAs and selecting the gRNA that shows the highest targeting efficiency for establishment of clonal lines (estimation of the targeting efficiency is described in **step 5** of Subheading 3.2.3).

We have found that it is faster and more cost effective to generate gRNAs using an oligonucleotide, than through cloning into the commonly used gRNA-expressing plasmids. First, a 120 nt ssDNA oligo, which includes a T7 promoter sequence, a variable 20 nt gRNA recognition sequence (N)20 (in red) and a constant gRNA backbone sequence can be directly synthesized and bought from commercial vendors. Next PCR-amplify the oligonucleotide

using the T7 F and Tracr R universal primers and use the PCR products as templates for in vitro transcription to produce the gRNAs (Fig. 4c).

1. Use the online CRISPR design tool (http://crispr.mit.edu/) to design 3 gRNAs that target sequences at the appropriate genomic site. It is highly recommended to choose gRNAs that do not recognize any sequence present in the donor vector to prevent undesired CRISPR/Cas9-mediated mutagenesis after correct reporter gene integration.

2. Order the 120 nt ssDNA oligos containing the desired 20 nt gRNA sequences directly from commercial vendors. Desalted purification is sufficient.

3. PCR-amplify the 120 nt ssDNA oligos using the T7 F and Tracr R universal primers using Herculase II Fusion DNA Polymerase to produce dsDNA templates for T7 in vitro gRNA transcription.

4. Examine the PCR product through electrophoresis using 2.5 % Agarose gel. Gel purify the desired dsDNA templates if non-specific PCR products are detected.

Primer	Sequence
T7 F	TAATACGACTCACTATAGGG
Tracr R	AAAAGCACCGACTCGGTGCC

PCR Reaction mix (50 μL)

Component	Amount (μL)
ddH$_2$O	35.5
5× Herculase II reaction buffer	10
dNTP mix (25 mM)	0.5
T7 F (10 μM)	1.25
Tracr R (10 μM)	1.25
120-nt ssDNA oligo (25 nM)	1
Herculase II fusion DNA polymerase	0.5

PCR cycling conditions

Cycle number	Denature	Anneal	Extend
1	94 °C, 2 min		
2–31	94 °C, 20 s	60 °C, 20 s	72 °C, 1 min
32			72 °C, 2 min

5. Use the MEGAshortscript T7 Transcription kit for gRNA synthesis. In vitro *t*ranscription mix (20 μL)

Component	Amount (μL)
T7 ATP	2
T7 CTP	2
T7 GTP	2
T7 UTP	2
T7 10× buffer	2
T7 enzyme mix	2
dsDNA template	8

Incubate for 4 h to overnight at 37 °C

Add 1 μL TURBO DNase and incubate for 15 min at 37 °C to digest the dsDNA template.

6. Purify gRNA using the MEGAclear Transcription Clean-Up Kit following the manufacturer's instructions, and elute gRNAs (typically ~50–100 μg) in 100 μL RNase-free water. When possible adjust concentration to 320 ng/μL (10 μM) and store at –80 °C until use.

3.2.3 Co-transfection of gRNA and Donor Vector in iCas9 hPSC Cells

Efficient co-transfection of gRNA and dsDNA donor vector into the iCas9 hPSC cells is very important for the generation of knockin reporter hPSC lines using iCRISPR system. We have tested several commonly used transfection reagents along with electroporation and find that the highest and most consistent co-transfection efficiency (~20%) is achieved using Lipofectamine 3000 (*see* **Note 3**).

1. *Day 0.* Plate iMEFs on a gelatin-coated 24-well plate and treat 60% confluent iCas9 hPSC with hPSC medium containing 2 μg/mL doxycycline. This will allow optimal Cas9 expression at the time of transfection on Day 1.

2. *Day 1.* Wash the iCas9 cells with PBS w/o Ca^{2+} and Mg^{2+} and then dissociate hPSCs with TrypLE for 5 min at 37 °C. Once the cells are disaggregated, collect the cells in 10 mL of hPSC media and spin down at $200 \times g$ for 5 min. Resuspend cells at ~$0.25–1 \times 10^6$ cells/mL in hPSC medium with 5 μM ROCK

inhibitor and 2 μg/mL doxycycline (*see* **Note 4**). Re-plate 0.5 mL of cell suspension into each well of a 24-well plate. Prepare duplicated wells for each gRNA used, and wells for the non-transfection control and the donor-only control. One well of the duplicated wells will be used for estimation of the targeting efficiency and the other well will be used for re-plating.

First transfection. For each gRNA, prepare separately:

Mix 1: 50 μL Opti-MEM + 1 μL gRNA (320 ng) + 2 μL P3000 + 2 μL donor plasmid (2.5 μg).

Mix 2: 50 μL Opti-MEM + 3 μL Lipo 3000 reagent.

Mix 1 + 2, incubate for 5 min at RT and add 50 μL mixture drop-wise into 0.5 mL hPSCs dilution in both duplicate wells.

For the non-transfection control, omit both gRNA and donor template.

For the donor-only control, add only donor template.

3. *Day 2.* Optional. Change hPSC medium with 2 μg/mL doxycycline and perform the 2nd transfection using the same procedure as the 1st transfection.

4. *Day 3–4.* Change hPSC medium daily.

5. *Day 5.* Estimation of targeting efficiency.

For lineage-specific genes that are only expressed in differentiated cell types, the targeting efficiency can be roughly estimated by PCR amplification of the targeted alleles. Collect genomic DNA from one of the duplicated wells using the DNeasy Blood & Tissue Kit. PCR-amplify the targeted alleles using an external primer (F) and an internal primer (R1), and the non-targeted allele using two external primers (F and R2) (Fig. 4d). Although primers F + R2 may also amplify the targeted allele, in our experience the non-targeted allele is preferentially amplified due to its smaller size. Identify the gRNA with the highest targeting efficiency and use the corresponding well for expansion and screening of clonal lines.

PCR Reaction mix (50 μL)

Component	Amount (μL)
ddH$_2$O	35.5
5× Herculase II reaction buffer	10
dNTP mix (25 mM)	0.5
Forward primer (10 μM)	1.25
Reverse primer (10 μM)	1.25
Genomic DNA (100 ng/μl)	1
Herculase II fusion DNA polymerase	0.5

PCR cycling conditions

Cycle number	Denature	Anneal	Extend
1	94 °C, 2 min		
2–36	94 °C, 20 s	N^a °C, 20 s	72 °C, 30 s
37			72 °C, 2 min

[a]Annealing temperature of the primer pair, ideally between 55 °C and 65 °C

For genes that are expressed in undifferentiated hPSCs, the percentage of cells expressing the fluorescence reporter tag can be directly assessed through flow cytometric analysis.

3.2.4 Expansion and Screening of Clonal lines

1. *Re-plating.* hPSCs identified with the highest targeting efficiency in Subheading 3.2.3 are dissociated into single cells using TrypLE, resuspended in hPSC medium with 5 μM ROCK inhibitor and re-plated at ~2000–5000 cells per 100 mm dish into dishes pre-seeded with iMEF feeders.
   For genes that are expressed in undifferentiated hPSCs, enrich correctly targeted cells through Flow Activated Cell Sorting (FACS) and then re-plate them at a density of 10,000 cells per 100 mm dish.

2. Do not change medium the next day.

3. Starting from 2 days after re-plating, change hPSC medium every day until colonies growing from single cells reach ~2 mm in diameter (~10 days).

4. *Colony picking.* Remove hPSC medium and add 10 mL PBS w/o Ca^{2+} and Mg^{2+} into the dish for a clearer visibility of the colonies. Mechanically dissociate the hPSC colonies into small clusters using 200 μL pipette tips, pick and plate single colonies in an uncoated 96-well plate containing 100 μL of hPSC medium with 5 μM ROCK inhibitor. Each clone is further disaggregated by pipetting up and down 5 times and re-plated in duplicated 96-well plates (50 μL each) pre-seeded with iMEFs and containing 100 μL of hPSC medium with 5 μM ROCK inhibitor. Alternatively, colonies could first be cultured in one 96-well plate and then split into two 96-well plates when confluent (in ~5 days).

   Depending on the estimated targeting efficiency in Subheading 3.2.3, pick between 96 (targeting efficiency >5 %) to 384 clones (targeting efficiency <5 %). Media change and passaging can be performed using a multichannel aspirator and a multichannel pipette.

   Change hPSC media for the 96-well plates daily. Use one of the duplicated 96-well plates for PCR screening in 3 days and the other plate for maintenance.

5. *Genomic DNA extraction.* Remove media from the wells, wash once with PBS w/o Ca^{2+} and Mg^{2+} and add 35 μL of genomic DNA lysis buffer. After 5 min at RT, transfer the cell lysates into a 96-well PCR plate. Seal the plate using a PCR film sticker and incubate at 55 °C for 2 h followed by 5 min at 95 °C to inactivate Proteinase K. Keep samples at 4 °C before the PCR screening step (*see* **Note 5**).

6. *PCR screening.* PCR-amplify using the same external primer (F) and internal primer (R1) used in Subheading 3.2.3. Run 25 μL of the PCR product in a 1% agarose gel stained with EtBr. PCR products of the right size suggest proper targeting in clonal lines.

7. *Expansion of clonal lines.* Amplify the desired clonal lines from the maintenance plate for further validation.

3.2.5 Validation of Established Clonal Lines

After targeting, isolation of clonal lines and PCR genotyping, correctly targeted clonal lines are expanded and further validated. We recommend immunohistochemistry for expression of pluripotency markers (e.g., OCT4, SOX2, and NANOG), teratoma assay for a functional assessment of pluripotency, karyotype testing, and differentiation into specific lineages to confirm faithful reporter activity. We highly recommend performing Sanger sequencing of both the targeted and the non-targeted allele at the target locus. We generally detect the expected sequence at the targeted allele. However, indel mutations are sometimes detected in the non-targeted allele, and clones with these undesired mutations should be discarded.

Due to mismatch tolerance of CRISPR gRNA paring [18], there are concerns about the potential off-target mutagenic effects of the CRISPR/Cas9 system. The online CRISPR design tool (http://crispr.mit.edu/) can be used to identify the potential off-targets falling in coding sequences for each gRNA and we typically select 10 top candidates for further analysis. PCR-amplify the genomic region (~500 bp) flanking the CRISPR off-target site using Herculase II Fusion DNA Polymerase and analyze by Sanger sequencing using a primer that binds within the PCR product. It is worth noting that our analysis so far has not revealed any off-target mutations at sites without perfect complementarity to the CRISPR/Cas9 target sequence. In agreement with our findings, a recent high-coverage whole-genome sequencing study failed to detect significant incidence of off-target mutations in CRISPR-targeted hPSC lines [25].

3.3 Generation of Knockin Epitope-tagged hPSCs Using iCRISPR

The procedure to generate knockin epitope tagged hPSCs is similar to the procedure used to create reporter tags (Subheading 3.2). Expression of Cas9 is induced in iCas9 cells following administration of doxycycline. A gRNA for the target site and the exogenous sequence are delivered into the cells. Site-specific DSBs produced by Cas9 promote HDR and integration of the epitope tag into the

genome. After two transfections the cells are re-plated at clonal densities. Clonal lines are then isolated, expanded and genotyped using PCR.

One key difference from Subheading 3.2 is that due to the smaller size of the epitope sequence, a ssDNA can be used to provide the HDR template (the epitope tag flanked by the homology arms). In addition a different transfection reagent, Lipofectamine RNAiMax is used to deliver the ssDNA and gRNA into the cells. Finally clonal lines identified by PCR to have correct integration, are further validated using Sanger sequencing, IP and western blot.

3.3.1 Design of CRISPR gRNAs and Donor ssDNA

For the generation of tagged proteins the location of the tag in reference to the protein is very important. The tag should be located so as to have minimal effect on tertiary structure and biological activity of the protein. At the same time the tag should be readily accessible on the surface of the natively folded protein. In deciding where to insert the tag it is best to follow previous studies that have already validated the generation of a functional fusion protein. In the absence of a previous study, the safest strategy would be to generate lines that have the tag epitope attached to the N-terminal of the target protein, as well as lines that have the tag epitope attached to the C-terminal of the target protein (Fig. 5a).
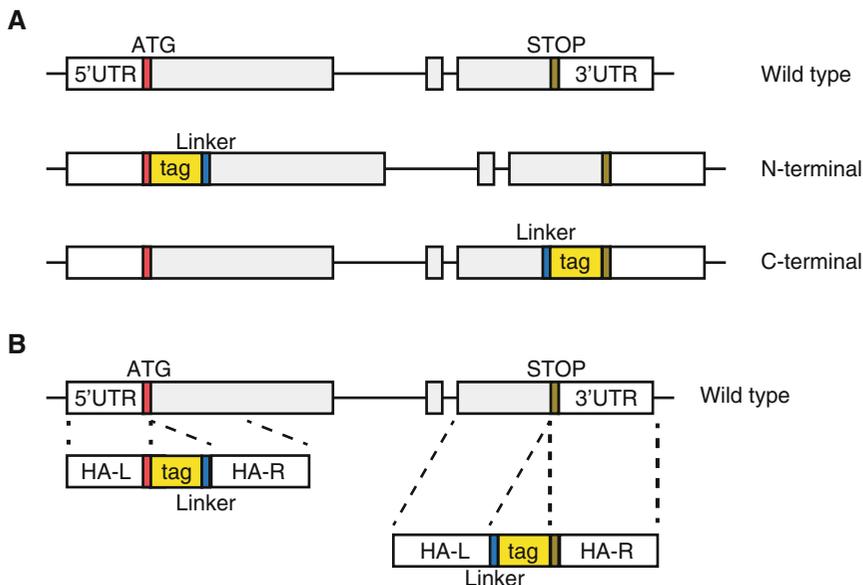
Fig. 5 Generation of epitope tag knockin hPSC lines using iCRISPR. (**a**) Knockin epitope tags can be targeted to the N-terminal or C-terminal of a protein. At the N-terminal the epitope tag is inserted right after the start codon and separated from the endogenous sequence by a linker sequence. At the C-terminal the epitope tag is inserted immediately before the stop codon with a linker sequence separating the epitope tag from the endogenous sequence. (**b**) Schema of the ssDNA donors used to generate knockin epitope tags

After these lines have been generated one can compare them to see if the tag influences the stability or biological function of the protein. This could be done using western blot, co-IP with known interacting proteins, and for DNA-binding proteins, ChIP-qPCR.

When targeting the N-terminal of the protein, 3-4 gRNAs are chosen to cut near the start codon, so that after integration the start codon will immediately precede the epitope tag. When targeting the C-terminal of a protein, 3–4 gRNAs are designed to cut near the stop codon of the protein, so that after integration the epitope tag will be immediately before the stop codon (Fig. 5a). Please refer to Subheading 3.2.2 for detailed instructions on the design and production of CRISPR gRNAs.

3.3.2 Design of Donor ssDNA

The donor template consists of three components: the left and right homology arms, the epitope tag sequence, and a short linker sequence. The smaller size of the epitope tag compared to a reporter transgene, allows smaller homology arms (less than 100 nt) to be used for the targeting. Thus the overall size of the donor template is typically between 150 and 300 nt, and a ssDNA donor, instead of a plasmid, can be used for the targeting experiment.

1. Each of the homology arms (left and right) is around 65 nt in length (*see* **Note 6**).

2. For targeting the N-terminus, the left homology arm should stop after the start codon of the protein, and the right homology arm should include the DNA sequence immediately after the start codon (Fig. 5b). The sequence of the epitope tag is inserted between the homology arms (*see* **Note 7**), followed by a short linker sequence (e.g., 12 nt long). The linker sequence separates the epitope tag from the endogenous protein. This prevents the epitope tag from interfering with proper folding of the endogenous protein [26], and increases the accessibility of the epitope to antibodies [27]. A linker sequence we have used is: GGAGGACTCGAC, which produces a four amino acid sequence of Gly-Gly-Leu-Asp.

3. If targeting the C-terminus, the left arm should contain the sequence up to, but not including, the stop codon. This is followed by the linker sequence and then the epitope tag. Lastly the right homology arm should begin with the stop codon (Fig. 5b).

4. The final ssDNA oligo, containing the two homology arms, a linker sequence and the epitope tag, can be ordered directly from commercial vendors. Desalted purification is sufficient.

3.3.3 Co-transfection of Donor ssDNA and CRISPR gRNA

The procedure for passaging the cells for co-transfection is the same as described in Subheading 3.2.3. However, the transfection mixture is different. Perform co-transfection of donor ssDNA and CRISPR gRNA as follows:

Mix 1: 50 µL Opti-MEM + 1 µL gRNA (320 ng) + 5 µL ssDNA (1.5 µg).

Mix 2: 50 µL Opti-MEM + 3 µL Lipofectamine RNAiMAX.

Mix 1 + 2, incubate for 5 min at RT and add 50 µL mixture dropwise into 0.5 mL dissociated hPSCs in one well of a 24-well plate.

The same procedure presented in **step 5** of Subheading 3.2.3 can be used to evaluate the targeting efficiency for the epitope-tag.

3.3.4 Expansion and PCR Based-screening of Clonal Lines

The procedure for isolating clonal lines, PCR screening and expansion of clonal lines is the same as described in Subheading 3.2.4.

3.3.5 Sequencing Verification of Targeted Clones

We highly recommend examining both the targeted and non-targeted allele at the target locus by Sanger sequencing. This enables us to verify that the epitope tag sequence is intact and properly integrated in-frame with the protein sequence. We also verify that the non-targeted allele does not contain any unwanted mutations.

1. Collect genomic DNA of correctly targeted lines using the DNeasy Blood and Tissue kit.

2. PCR-amplify the targeted and non-targeted alleles using a pair of external primers.

PCR Reaction mix (50 µL)

Component	Amount (µL)
ddH$_2$O	35.5
5× Herculase II reaction buffer	10
dNTP mix (25 mM)	0.5
Forward primer (10 µM)	1.25
Reverse primer (10 µM)	1.25
Genomic DNA (100 ng/µl)	1
Herculase II fusion DNA polymerase	0.5

PCR cycling conditions

Cycle number	Denature	Anneal	Extend
1	94 °C, 2 min		
2-36	94 °C, 20 s	N[a] °C, 20 s	72 °C, 30 s
37			72 °C, 2 min

[a]Annealing temperature of the primer pair, ideally between 55 °C and 65 °C

3. Run 25 μL of the PCR reaction on a 1% agarose gel stained with EtBr.

4. Visualize the gel using a UV lamp. For correctly targeted clones there should be two bands, representing the targeted allele and the non-targeted allele. The targeted allele should be larger than the non-targeted allele by the size of the epitope tag (generally less than 100 bp).

5. Cut out the individual bands and purify DNA using the Purelink Quick Gel Extraction kit.

6. Use one of the PCR primers for Sanger sequencing, or use a new sequencing primer that is within the PCR product (*see* **Note 8**).

3.3.6 Validation of Targeted Clones Using Immunoprecipitation and Western Blot

The final, and most important, test to validate the epitope tagged lines is to perform IP for the tag and then verify that the target protein is also pulled down. This would involve first performing an IP using an antibody against the epitope tag followed by a western blot with the antibody to the protein of interest. If it is challenging to detect the target protein due to lack of quality antibodies, one may perform the western blot against the tag epitope to verify that the epitope tagged protein is the same size as the target protein. Below we provide a protocol routinely used in our laboratory for detection of 3×FLAG tagged proteins. We strongly encourage users to optimize the protocol for their protein of interest, as it may require different procedures for cell lysis, IP and western blot.

1. Expand the clonal lines into a 100 mm dish.

2. Use the Pierce Crosslink Magnetic IP/Co-IP kit for lysis and IP and follow the manufacturer's protocol. Lyse one 100 mm dish per IP with 600 μL of the provided lysis IP buffer. Incubate the cell lysate with the FLAG antibody for 16 h at 4 °C. And elute with 100 μL of the provided elution buffer.

3. For the western blot 30 μL total volume is run: 19.5 μL of IP sample, 7.5 μL LDS sample buffer and 3 μL of reducing agent. The samples should be denatured for 10 min at 70 °C prior to loading.

4. Samples are run on NuPAGE Novex 3–8% Tris–Acetate Protein Gels in NuPAGE Tris–Acetate SDS Running Buffer (20×) for 1 h at 120 V.

5. Transfer is performed overnight at 30 V onto a nitrocellulose membrane.

6. Wash the membrane in TBST and then incubate in blocking buffer (5% nonfat milk in TBST) for 30 min at room temperature.

7. Incubate the membrane with the FLAG antibody at a 1:2000 dilution overnight at 4 °C.

8. Wash the membrane 3 times for 5 min each with TBST, and then incubate with HRP conjugated anti-mouse antibody (1:5000) for 1 h at room temperature.

9. Finally develop the membrane using ECL reagent. Prior to exposure, excess liquid should be removed without drying the membrane. Transfer the membrane to a clean sealed bag. Expose onto an autoradiography film in a hypercassette.

4   Notes

1. Undesired mutations in homology arms arising during PCR amplification may decrease the targeting efficiency. We recommend using the high fidelity Herculase II Fusion DNA Polymerase for PCR amplification of homology arms and also later for PCR amplification of the gRNA in vitro transcription template. One may substitute Herculase with other commercially available high fidelity DNA polymerases.

2. Traditional targeting procedures that do not utilize site-specific nucleases require homology arms of at least 5–12 kb for the insertion of reporter and selection cassettes and the targeting efficiency can be further increased by using substantially longer homology regions (up to 100 kb) [28]. In most traditional targeting experiments, the donor plasmid typically contains a long (~10 kb) and a short homology arm (~2–4 kb) for efficient homologous recombination and also for convenient identification of correctly targeted alleles by PCR genotyping at the shorter homology arm side. With the use of engineered DNA endonucleases, repair templates with much shorter homology arms (~500–1000 bp on each side) have been used. We have found that approximately 500 bp combined length of the homology arms (~250 bp on each side) is necessary to support efficient CRISPR-mediated targeting of a 750 bp exogenous sequence into an endogenous locus.

3. In addition to Lipofectamine 3000 and RNAiMAX, a number of other transfection methods have also been shown to work efficiently in hPSCs [17, 19].

4. Based on our experience with HUES8 and MEL1 hESCs, we recommend testing several cell densities for the co-transfection of gRNA and donor vector, for example plating 100, 250, and 500 K cells per well in a 24-well plate and choosing the best condition for the establishment of clonal lines. After transfection, we regularly use a replating cell density of 2000–5000 cells per 100 mm dish. These conditions may need to be adjusted for other hPSC lines.

5. This simplified protocol enables the extraction of genomic DNA for PCR screening without the use of phenol/chloroform

extraction and purification. Combined with the use of multichannel pipettes, one can quickly screen a large number of clones. Based on our experience it is feasible for a single person to process up to 384 samples (four 96-well plates) at a time. This procedure works well as long as the primers are pre-tested before the targeting experiment.

6. We have used homology arms that are around 65 nt in length. The targeting efficiency may be increased by using longer homology arms, but this has not been systematically tested.

7. Thus far we have tested attaching a 3×FLAG epitope tag to the protein of interest. However this strategy can be used for attaching a variety of other epitope tags such as poly-Arg, poly-His, Strep-tag, c-myc-tag, HAT-tag, and His-tag. As long as the sequence of the tag remains under 100 bp the same procedure as presented above can be used. For larger tags the size of the homology arms and the delivery of the exogenous sequence (as a ssDNA or as a donor plasmid) may need to be adjusted. The specific protein tag to use will depend on the protein of interest as well as the intended downstream analysis.

8. If a nonsynonymous mutation is detected in the coding sequence of the protein in either the targeted or non-targeted allele we would advise to discard that particular clone. In some cases a mutation may occur in a non-coding regulatory region (e.g., promoter), in which case the clone may still be used if the mutation does not alter the expression of the affected allele (as determined by qRT-PCR for example).

5   Future Applications

Here we present the protocols to integrate epitope and reporter tags within the genome of hPSCs using the iCRISPR genome editing platform. The same knockin strategy could be used for additional purposes. For example, targeting a fluorescent or antibiotic tag into the coding region of a gene could be used to generate genetic knockouts. Specific deletions could be engineered to mimic a genetic disease or to dissect the contribution of particular protein domains for protein function. By inserting particular sequences one may even precisely modulate the DNA-binding, protein interaction or other specific activities of endogenous proteins. Furthermore, the gene targeting strategies presented here are also relevant to labs that use engineered nucleases other than CRISPR/Cas or employ different strategies to deliver these nucleases into hPSCs. Finally we provide robust protocols to achieve relatively high rates of delivery for ssDNA, dsDNA plasmids and gRNAs. These transfection procedures may have wider applicability such as potential uses in overexpression and lineage tracing studies.

Acknowledgements

References

1. Tabar V, Studer L (2014) Pluripotent stem cells in regenerative medicine: challenges and recent progress. Nat Rev Genet 15:82–92

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

3. Chia NY, Chan YS, Feng B et al (2010) A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature 468:316–320

4. Theunissen TW, Powell BE, Wang H et al (2014) Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell 15:471–487

5. Eiges R, Schuldiner M, Drukker M et al (2001) Establishment of human embryonic stem cell-transfected clones carrying a marker for undifferentiated cells. Curr Biol 11:514–518

6. Zwaka TP, Thomson JA (2003) Homologous recombination in human embryonic stem cells. Nat Biotechnol 21:319–321

7. Davis RP, Costa M, Grandela C et al (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

8. Zhu Z, Verma N, Gonzalez F, Shi ZD, Huangfu D (2015) A CRISPR/Cas-mediated selection-free knockin strategy in human embryonic stem cells. Stem Cell Rep 4:1103–1111

9. Joung JK, Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol 14:49–55

10. Ran FA, Hsu PD, Wright J et al (2013) Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8:2281–2308

11. Urnov FD, Rebar EJ, Holmes MC et al (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11:636–646

12. Jinek M, Chylinski K, Fonfara I et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821

13. Jinek M, East A, Cheng A et al (2013) RNA-programmed genome editing in human cells. Elife 2:e00471

14. Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31:230–232

15. Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823

16. Mali P, Yang L, Esvelt KM et al (2013) RNA-guided human genome engineering via Cas9. Science 339:823–826

17. Byrne SM, Ortiz L, Mali P et al (2015) Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. Nucleic Acids Res 43:e21

18. Hou Z, Zhang Y, Propson NE et al (2013) Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc Natl Acad Sci U S A 110:15644–15649

19. Merkert S, Wunderlich S, Bednarski C et al (2014) Efficient designer nuclease-based homologous recombination enables direct PCR screening for footprintless targeted human pluripotent stem cells. Stem Cell Rep 2:107–118

20. Merkle FT, Neuhausser WM, Santos D et al (2015) Efficient CRISPR-Cas9-mediated generation of knockin human pluripotent stem cells lacking undesired mutations at the targeted locus. Cell Rep 11:875–883

21. Gonzalez F, Zhu Z, Shi ZD et al (2014) An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell Stem Cell 15:215–226

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

23. Chen G, Gulbranson DR, Hou Z et al (2011) Chemically defined conditions for human iPSC derivation and culture. Nat Methods 8:424–429

24. Ludwig TE, Bergendahl V, Levenstein ME et al (2006) Feeder-independent culture of human embryonic stem cells. Nat Methods 3:637–646

25. Veres A, Gosis BS, Ding Q et al (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

26. Borjigin J, Nathans J (1994) Insertional mutagenesis as a probe of rhodopsin's topography, stability, and activity. J Biol Chem 269:14715–14722

27. Grote E, Hao JC, Bennett MK, Kelly RB (1995) A targeting signal in VAMP regulating transport to synaptic vesicles. Cell 81:581–589

28. Nieminen M, Tuuri T, Savilahti H (2010) Genetic recombination pathways and their application for genome modification of human embryonic stem cells. Exp Cell Res 316:2578–2586