Chapter 8

CRISPR-Cas9-Mediated Knock-In Approach to Insert the GFP_{11} Tag into the Genome of a Human Cell Line

Ryo Tamura and Daichi Kamiyama

Abstract

The protocol in this chapter describes a method to label endogenous proteins using a self-complementing split green fluorescent protein (split $GFP_{1\text{-}10/11}$) in a human cell line. By directly delivering Cas9/sgRNA ribonucleoprotein (RNP) complexes through nucleofection, this protocol allows for the efficient integration of *GFP_{11}* into a specific genomic locus via CRISPR-Cas9-mediated homology-directed repair (HDR). We use the GFP_{11} sequence in the form of a single-stranded DNA (ssDNA) as an HDR template. Because the ssDNA with less than 200 nucleotides used here is commercially synthesized, this approach remains cloning-free. The integration of *GFP_{11}* is performed in cells stably expressing *$GFP_{1\text{-}10}$*, thereby inducing fluorescence reconstitution. Subsequently, such a reconstituted signal is analyzed using fluorescence flow cytometry for estimating knock-in efficiencies and enriching the GFP-positive cell population. Finally, the enriched cells can be visualized using fluorescence microscopy.

Key words Split GFP, GFP_{11}, $GFP_{1\text{-}10}$, CRISPR, Cas9, Homology-directed repair

1 Introduction

The Human Genome Project has revealed the entire picture of the human genomic landscape and led to the identification of ~25,000 genes that encode cellular proteins [1]. This has laid the groundwork for deciphering how these proteins individually function in a myriad of cellular contexts. Recent molecular and cellular technologies have also assisted researchers in elucidating how proteins cooperate to achieve distinct cellular functions in spatiotemporally regulated manners, providing a more comprehensive view of the proteome. One such technology is fluorescent proteins (FPs), which have transformed cell biology with the ability to visualize proteins of interest using various fluorescence imaging techniques [2]. Previously, an array of FPs has been engineered across and beyond the visible spectrum [3]. These FPs enable simultaneous labeling of multiple proteins in individual live cells, which captures

Mayank Sharma (ed.), *Fluorescent Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 2564,
https://doi.org/10.1007/978-1-0716-2667-2_8,

spatial and temporal dynamics of the proteins and provides mechanistic insights into gene expression regulation, protein modification, and protein-protein interaction [4, 5].

Protein distribution studies have often been carried out through transient transfection in which proteins of interest are tagged with FPs and overexpressed in cells. However, overexpression of FP fusions tends to generate unwanted artifacts (e.g., protein aggregation and mislocalization), making it difficult to interpret the obtained data [6]. Therefore, endogenous gene tagging approaches are preferable [7]. One of the approaches we can use to accomplish endogenous FP tagging is CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9-mediated genome editing. CRISPR-Cas9 system comprises a custom single-guide RNA (sgRNA) and Cas9 protein, which induce a DNA double-strand break to a specific genomic locus in cells. This consequently provides an opportunity to deliberately engineer the locus through homology-directed repair (HDR) [8]. In CRISPR--Cas9-mediated HDR, any altered DNA sequence (including insertions, deletions, and point-mutations) can be flanked by homologous DNA sequences of the corresponding genomic region, resulting in a user-defined, targeted DNA repair [8]. By leveraging CRISPR-Cas9-mediated HDR, we can integrate an FP sequence into a specific protein-coding region, thus visualizing the endogenous localization of target proteins [9].

Because of the relatively large size of FPs (e.g., EGFP is 717 nucleotides [nt] in length), we have to use a long HDR repair template (>1000 nt with homologous DNA sequences), which necessitates multiple steps of molecular cloning to insert the sequence into a backbone plasmid [10]. To simplify CRISPR-Cas9-mediated HDR and provide an efficient approach for protein labeling, we have recently adapted the self-associating split GFP system [10, 11]. In the split GFP system, the beta-barrel scaffold of super-folder GFP is split into two fragments comprising one large fragment and one small fragment, termed $GFP_{1\text{-}10}$ and GFP_{11}, respectively. Neither $GFP_{1\text{-}10}$ nor GFP_{11} is fluorescent individually, yet upon co-expression, the two fragments undergo spontaneous assembly and reconstitute fluorescently. The GFP_{11} fragment has been used as a protein tag. Because the size of the tag is only 48 nt in length [10], this small tag minimizes the length of an HDR repair template (~200 nt in length) and enhances the knock-in efficiency when used for the generation of knock-in cell lines [10]. Moreover, such a short DNA template can be commercially synthesized, making this approach cloning-free. More recently, we and others have developed numerous color variants of split GFP within the visible spectrum, expanding the choice of colors in split FP systems [12–17]. In addition to the visualization of endogenous proteins, these FP_{11} tags have been employed in a wide range of protein analyses including, but not limited to, quantification of

protein folding [18, 19], determination of protein topology [20, 21], and visualization of protein localization in living cells [22–24]. In this chapter, we describe a protocol to introduce the GFP_{11} fragment into a specific genomic locus via CRISPR-Cas9-mediated HDR in cultured human cells (*see* Fig. 1). This protocol is adapted from the method developed by Lin et al. [25] and optimized for genomic integration of GFP_{11}. In conjunction with a cell cycle synchronization technique, the method exploits direct delivery of Cas9/sgRNA ribonucleoprotein (RNP) complexes through nucleofection to maximize the probability of HDR. Furthermore, by co-expressing the complementary GFP_{1-10} fragment, the GFP_{11}-inserted cells can be readily detected by fluorescence flow cytometry and eventually enriched for cellular imaging. Altogether, this protocol provides a simple and efficient way to visualize the localization of endogenous proteins in cultured human cells.

2 Materials

2.1 sgRNA Preparation

1. Pipette tips, RNase-free.
2. Microcentrifuge tubes, RNase-free.
3. Microcentrifuge.
4. Thermocycler.
5. Q5 high-fidelity DNA polymerase (New England Biolabs). Other high-fidelity DNA polymerases can be used.
6. Zymo DNA Clean & Concentrator-5 (Zymo Research).
7. NanoPhotometer N50 (Implen). Other spectrophotometers can be used.
8. Diethylpyrocarbonate (DEPC)-treated water (Growcells).
9. RNase-free DNA buffer: 2 mM Tris-HCl, pH 8.0, DEPC-treated water.
10. MAXIscript Transcription Kit (Invitrogen). Any other commercially available in vitro transcription systems can be used.
11. 25 mM NTP mix (New England Biolabs).
12. Zymo RNA Clean & Concentrator-5 (Zymo Research).
13. RNase-free RNA buffer: 10 mM Tris-HCl, pH 7.0, DEPC-treated water.

2.2 Stable Cell Line Creation

1. T-25-cm^2 flask.
2. Tissue culture plate, 24-well.
3. Human embryonic kidney 293 cells.
4. pcDNA $3.1^{(+)}$ mammalian expression vector (Invitrogen).
5. 50 mg/mL of Geneticin selective antibiotic (Gibco).

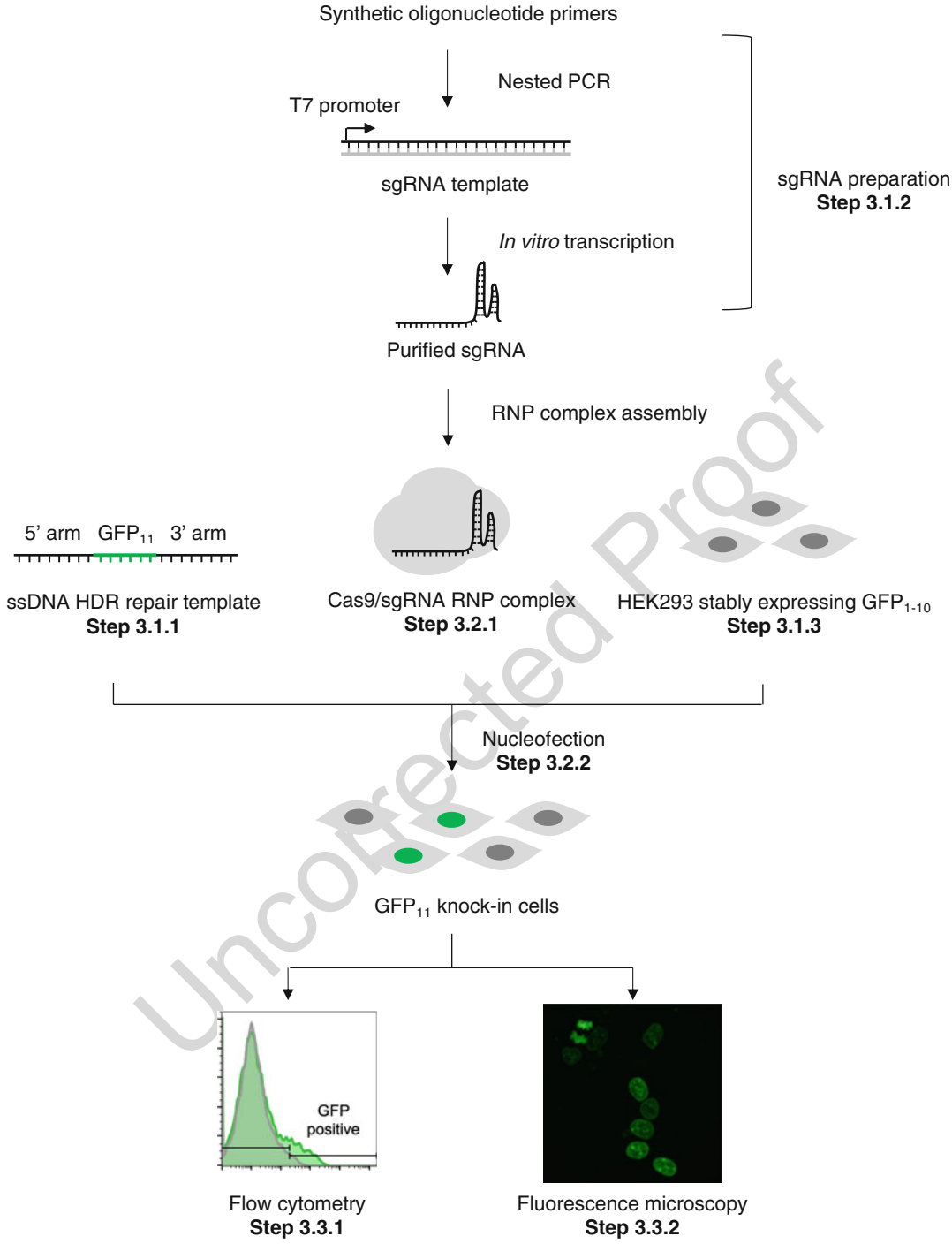

Fig. 1 Overview of *GFP*$_{11}$ knock-in generation via CRISPR-Cas9-mediated HDR. Synthetic oligonucleotide primers are used to generate a DNA template. The DNA template enables the production of sgRNA through in vitro transcription. A Cas9/sgRNA ribonucleoprotein (RNP) complex is assembled in vitro. An ssDNA HDR repair template and the Cas9/sgRNA RNP complex are nucleofected into HEK 293 cells. *GFP*$_{11}$ knock-in cells can be analyzed by flow cytometry to assess the knock-in efficiency and enrich GFP-positive cells or be imaged by fluorescence microscopy to localize the tagged proteins

6. Dulbecco's modified Eagle medium (DMEM) (HyClone).
7. DMEM medium: DMEM with 10% (v/v) fetal bovine serum (FBS).
8. Humidified CO_2 incubator (37 °C, 5% CO_2).
9. Opti-MEM (Gibco).
10. 25 kDa linear polyethylenimine (Polysciences, Inc.). Prepare PEI stock solution by diluting linear PEI into a buffer comprising 25 mM HEPES and 150 mM NaCl at pH 7.5. Sterilize the solution via a 0.2-μm filter.
11. Vortexer.

2.3 Cell Cycle Synchronization

Nocodazole: To prepare a stock solution, dissolve the nocodazole powder to the final concentration of 100 μg/mL in PBS. The solution should be sterilized through a 0.2-μm filter.

2.4 RNP Complex Assembly

1. Purified Cas9 protein (*see* **Note 1**)
2. Tris(2-carboxyethyl)phosphine hydrochloride (TECP)-HCl
3. Cas9 buffer: 150 mM KCl, 20 mM Tris pH 7.5, 10% v/v glycerol, 1 mM TCEP-HCl, 1 mM $MgCl_2$, DPEC-treated water
4. Laboratory water bath

2.5 Nucleofection

1. Amaxa Cell Line Nucleofector Kit V (Lonza)
2. Nucleofector 2b Device (Lonza)
3. Tissue culture plate, six well

2.6 Flow Cytometry

1. Fluorescence-activated cell sorter (FACS) collection tubes, sterile
2. CytoFLEX (Beckman Coulter)
3. S3 cell sorter (Bio-Rad)
4. FlowJo software (Treestar, Inc.)

2.7 Fluorescence Microscopy

1. 1 mg/mL of fibronectin human plasma (Sigma-Aldrich).
2. Chambered coverglass, eight-well (Lab-Tek).
3. Inverted fluorescence microscope (Nikon) with a Dragonfly spinning disk confocal unit (Andor). Any other confocal microscopes equipped with a 488-nm laser are suitable.
4. 100 × 1.45 NA oil immersion objective (Nikon).
5. Fiji software (NIH).

3 Methods

3.1 Pre-experiment Reagent and Cell Preparation

3.1.1 HDR Repair Template Design

Our HDR donor sequence contains a *GFP*$_{11}$ sequence (48 nt in length) and a 3-aa flexible linker (9 nt), flanked by two homology arms (~70 nt); the total length of the donor sequence is ~200 nt. A synthetic ssDNA oligo up to 200 nt can be commercially acquired from custom oligo synthesis services. For tagging with GFP_{11}, either N- or C-terminus can be selected depending on its known impact on the localization of a target protein or the availability of neighboring protospacer adjacent motif (PAM) sites in a genomic locus. One example of the HDR template, which is designed to edit the *HIST2H2BE* locus with *GFP*$_{11}$, is shown in Fig. 2. *HIST2H2BE* encodes histone 2B, which is a core component of nucleosomes. This gene has been successfully targeted with *GFP*$_{11}$ in our previous study [10]. In this case, *GFP*$_{11}$ is inserted into the *HIST2H2BE* locus to tag the encoded protein at the C-terminus with GFP_{11}.

3.1.2 sgRNA Preparation

Synthesize target-specific sgRNAs through in vitro transcription using a DNA template (5′- TAATACGACTCACTA TAGGNNNNNNNNNNNNNNNNNNGTTTAAGAGC TATGCTGGAAACAGCATAGCAAGTTTAAATAAGGC TAGTCCGTTATCAACTTGAAAAAGTGGCACC GAGTCGGTGCTTTTTTT-3′). The DNA template holds a T7 promoter sequence (TAATAATACGACTCACTATAGG), an ~20-bp gene-specific guide sequence (NNNNNNNNNNNNNNNNNNN), and a sgRNA scaffold (GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAA-TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC-GAGTCGGTGCTTTTTTT). Prepare the DNA template by an overlapping nested PCR method. This method uses a pair of overlapping forward and reverse primers to generate an initial DNA template and another pair of forward and reverse primers to amplify the template (*see* Fig. 3 for the PCR scheme).

1. To prepare a DNA template, make the following 100 μL of PCR reaction mix (*see* Table 1 for primer sequences):
   (a) 2x Q5 polymerase, 50 μL
   (b) 12.5 μM T25 primer, 4 μL
   (c) 12.5 μM BS7 primer, 4 μL
   (d) μM ML611 primer, 0.5 μL
   (e) μM gene-specific primer, 0.5 μL
   (f) DEPC-treated water, 41 μL
2. Run the reaction in a thermocycler using the following PCR condition:

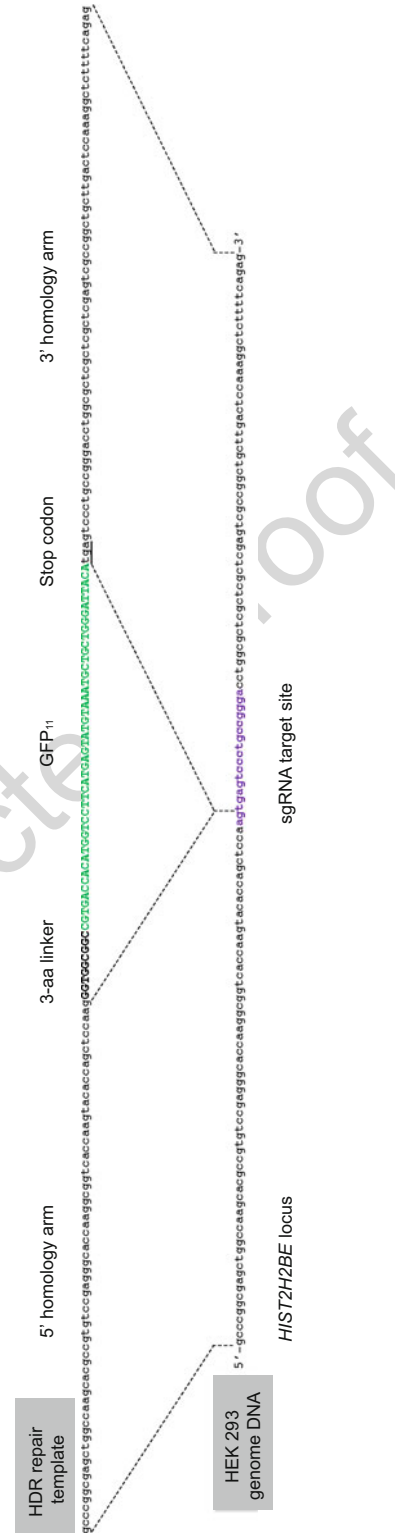

Fig. 2 Example of an HDR repair template design. Top: An HDR template contains a 3-aa linker (black bold upper case), a *GFP*$_{11}$ coding sequence (green bold upper case), and two flanking homology arms for recombination (black lower

(a) 95 °C for 30 s
(b) 95 °C for 15 s, 57 °C for 15 s, 72 °C for 15 s, 30 cycles
(c) 72 °C for 30 s
(d) °C infinity

3. Purify the DNA template using Zymo DNA Clean & Concentrator-5 Kit following manufacturer's instructions. Elute DNA with 12 μL RNase-free DNA buffer.
4. Measure the concentration of the eluted DNA using Nano-Photometer N50. The DNA concentration should be ~100 ng/μL. The yield may be different depending on DNA polymerases.
5. To perform in vitro transcription, make the following 100-μL reaction mix:
   (a) DNA template, 500 ng
   (b) 25 mM NTP mix, 4 μL
   (c) T7 enzyme mix, 10 μL
   (d) 10x transcription buffer, 10 μL
   (e) DEPC-treated water, up to 100 μL
6. Incubate the reaction mix at 37 °C for 4 h.
7. Purify RNA using Zymo RNA Clean & Concentrator-5 Kit following manufacturer's instructions. Elute RNA with 15 μL RNA buffer.
8. Measure the concentration of the eluted RNA using Nano-Photometer N50. The concentration should be ~1000 ng/μL. If the eluted RNA is too concentrated, dilute it with RNA buffer. The eluted sgRNA can be stored at −80 °C until use.

3.1.3 Stable Cell Line Generation

Prior to generating *GFP*$_{11}$ knock-in cells, one may establish a stable cell line where *GFP*$_{1\text{-}10}$ is continuously expressed. This allows the cells to achieve the reconstitution of $GFP_{1\text{-}10}$ with GFP_{11} after the knock-in event happens. Then, the *GFP*$_{11}$ knock-in cells can be directly brought to various imaging experiments by flow cytometry or fluorescence microscopy. Here, we describe the procedure for establishing HEK 293 cells expressing *GFP*$_{1\text{-}10}$. In this method, HEK 293 cells are transfected with a pcDNA3.1 vector containing the *GFP*$_{1\text{-}10}$ sequence after a CMV promoter (pcDNA3.1-$GFP_{1\text{-}10}$) and subsequently selected for the stable cell line with Geneticin (this vector contains the Geneticin-resistance gene for selection of

Fig. 2 (continued) case). The template is designed to integrate *GFP*$_{11}$ to the 3′ end of the *HIST2H2BE* gene immediately upstream of the stop codon (underlined lower case). Bottom: A region within the *HIST2H2BE* locus harboring a ~20 nucleotide sequence for the sgRNA site (purple underlined lower case)

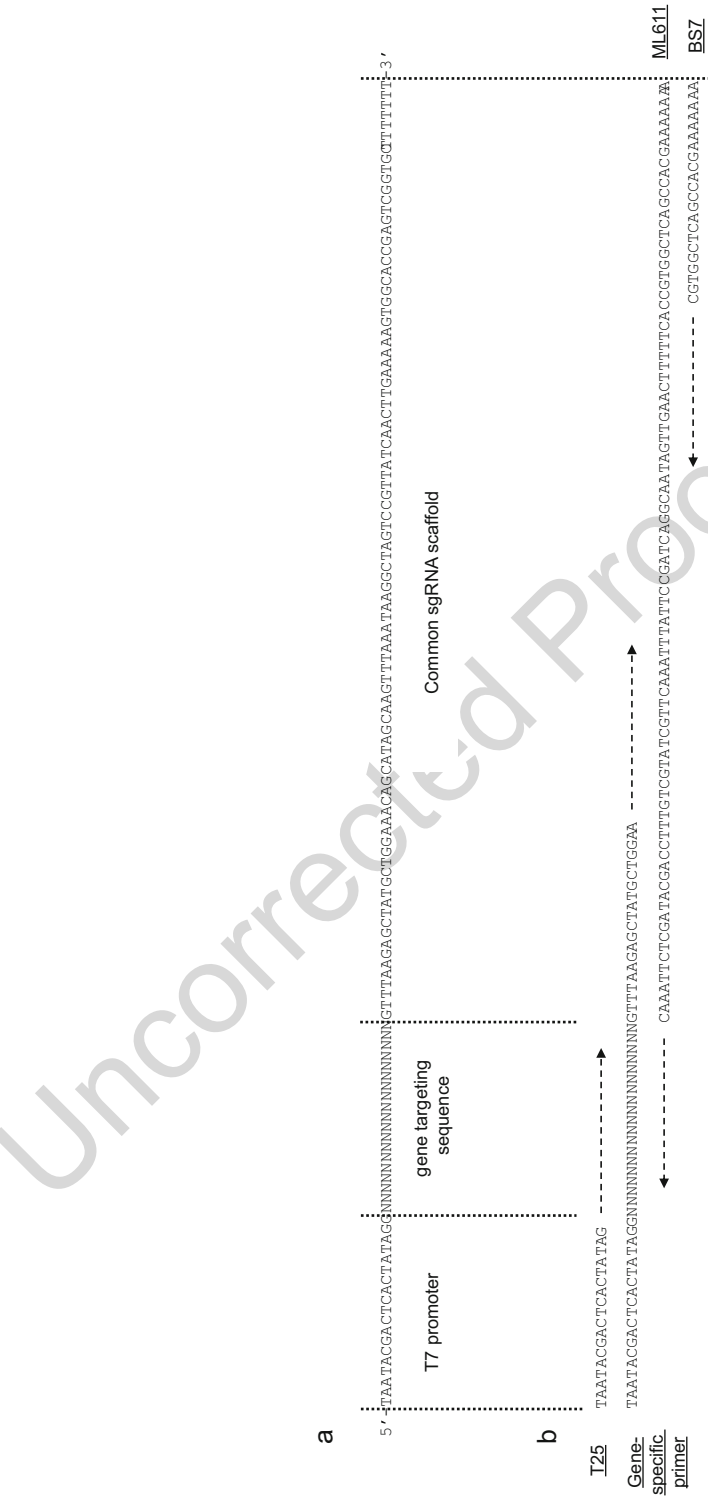

Fig. 3 Example of a sgRNA design. (**a**) A DNA template eventually used for in vitro transcription of sgRNA (**step 5** in Subheading 3.1.2). This template holds a T7 promoter, a gene-specific sequence, and a common sgRNA scaffold. (**b**) An

stable lines). Alternatively, other methods, including lentiviral transduction, can be employed to establish a stable cell line (*see* **Note 2**).

1. Plate HEK 293 cells in DMEM medium to be 80–90% confluent at the time of transfection in a 24-well cell culture plate.
2. Prepare the following (A) and (B) solutions and mix well by pipetting rigorously:
   (A) 400 ng pcDNA3.1-GFP_{1-10} in 25 μL Opti-MEM
   (B) 3 μL PEI stock solution in 25 μL Opti-MEM
3. Mix (A) and (B) from the previous step and vortex to form the DNA/PEI complex.
4. Incubate the DNA/PEI cocktail for 5 min at room temperature.
5. Change the culture medium to pre-warmed (37 °C) and fresh DMEM medium.
6. Add the 50 μL DNA/PEI cocktail to the cell culture.
7. Two days after transfection, start feeding DMEM medium containing 500 μg/mL Geneticin every 3–4 days.
8. Continue Geneticin selection until stably transfected HEK 293 cells emerge.

3.1.4 Cell Cycle Synchronization

The CRISPR-Cas9 system introduces double-strand breaks (DSB) through the nuclease activity of Cas9 at a sgRNA-specified genomic locus. Subsequently, the introduced DSBs are repaired by either nonhomologous end-joining (NHEJ) or HDR pathway. Since the HDR pathway is active exclusively in the late S or G2 phase, timed delivery of Cas9/sgRNA RNP complexes with donor HDR templates into synchronized cells could shift the equilibrium toward HDR and thus enhance the knock-in efficiency [26]. Cell cycle synchronization with nocodazole has been shown to increase the efficiency of CRISPR-Cas9-mediated knock-in up to 38% in mammalian cells [25].

1. One day before nucleofection, plate 1.5×10^6 of GFP_{1-10}-expressing cells in DMEM medium inside a T-25 flask.
2. Add nocodazole to the cells at the final concentration of 200 ng/mL.

Fig. 3 (continued) overlapping PCR scheme for the synthesis of the DNA template. The DNA template is PCR-amplified using three common primers (T25, ML611, BS7) and a gene-specific primer. See Table 1 for the 5′ and 3′ primer sequences

Table 1
PCR primers to synthesize a DNA template for sgRNA in vitro transcription

T25 primer	5′-TAA TAC GAC TCA CTA TAG-3′
BS7 primer	5′-AAA AAA AGC ACC GAC TCG GTG C-3′
ML611 primer	5′-AAA AAA AGC ACC GAC TCG GTG CCA CTT TTT CAA GTT GAT AAC GGA CTA GCC TTA TTT AAA CTT GCT ATG CTG TTT CCA GCA TAG CTC TTA AAC-3
Gene-specific primer[a]	5′-TAA TAC GAC TCA CTA TAG GNN NNN NNN NNN NNN NNN NNG TTT AAG AGC TAT GCT GGA A-3′

[a]The gene-specific primer contains an array of "N" dependent on a specific gene sequence a user chooses

3. Incubate the cells in the CO_2 incubator for ~16 h prior to nucleofection. These cells are ready for nucleofection (*see* Subheading 3.2.2).

3.2 Knock-in Experiment

3.2.1 RNP Complex Assembly

Cas9 and sgRNA are conventionally introduced to cells through the Cas9 and sgRNA expression from a plasmid or the direct delivery of a purified Cas9/sgRNA RNP complex. In this method, the latter strategy is used because the direct delivery of the complex into cells leads to genome editing immediately after nucleofection, ensuring the timing of genomic integration within nocodazole-treated cells [25, 27]. In addition, the delivery of such a Cas9/sgRNA RNP complex into cells has furthermore been shown to reduce off-target effects due to a shorter half-life (~12 h) of Cas9 in the RNP complex, compared to that of Cas9 delivered in plasmid [27].

1. Dissolve an HDR template (*see* Subheading 3.1.1) in Cas9 buffer to 100 μM.
2. Thaw purified Cas9 and sgRNA (*see* Subheading 3.1.2) in separate sterile microcentrifuge tubes on ice.
3. Calculate the volumes of sgRNA, Cas9, HDR template, and Cas9 buffer to achieve the following amounts in 50 μL RNP complex cocktail.
   (a) sgRNA, 650 pmol (*see* **Note 3**)
   (b) Cas9, 500 pmol
   (c) HDR template, 1500 pmol
   (d) Cas9 buffer, to 50 μL
4. Dilute sgRNA in Cas9 buffer in a sterile microcentrifuge tube.
5. Slowly add Cas9 to diluted sgRNA from the previous step, and gently mix the solution by pipetting up and down.
6. Incubate the RNP complex cocktail at 37 °C for 10 min.

7. Add the HDR template solution to the RNP complex cocktail, and keep it on ice until nucleofection.

3.2.2 Nucleofection

In this step, the HDR repair template (*see* Subheading 3.1.1) and Cas9/sgRNA RNP complex (*see* Subheading 3.2.1) prepared in the previous steps are delivered to the nocodazole-treated HEK 293 cells (*see* Subheading 3.1.4) through nucleofection. Nucleofection is a form of electroporation developed by Lonza and optimized for the delivery of nucleic acids into a variety of cell lines. Different cell lines require dedicated nucleofection reagents (e.g., Amaxa Cell Line Nucleofector Kit V for HEK 293 cells). To date, various laboratories have successfully employed nucleofection for CRISPR-Cas9 genome editing (*see* **Note 4**) [10, 25, 28, 29]. This protocol performs nucleofection via Nucleofector 2b Device (the single cuvette-based system) with its cell-line-specific program.

1. Collect the nocodazole-treated cells. The cells only require gentle pipetting to dissociate for collection. For each nucleofection with Nucleofector 2b Device, 1×10^6 cells are used.
2. Spin the dissociated cells for 5 min at 3000 × *g*.
3. Resuspend in 1-mL PBS.
4. Remove the supernatant.
5. Subject the harvested cells to nucleofection with Amaxa Cell Line Nucleofector Kit V following manufacturer's instructions. Make sure to use the *Q-001* program on Nucleofector 2b Device.
6. Add 500 μL prewarmed DMEM medium to the nucleofected cells, and then transfer to a six-well culture plate.
7. Incubate the nucleofected cells in a humidified CO_2 incubator at 37 °C until analysis. Nucleofected cells typically recover after 2–4 days. Split the nucleofected cells for storage via cryopreservation to prevent the potential dilution of the GFP_{11}-inserted cell population by passaging. Use the remainder of the population for the subsequent analyses including flow cytometry and fluorescence microscopy.

3.3 Post-genome Editing Analysis

3.3.1 Flow Cytometry

To estimate the knock-in efficiency of GFP_{11}, one can load the nucleofected cells to a flow cytometer and measure the number of GFP-positive cells (*see* **Note 5**). Once the GFP signal is observed, fluorescence-activated cell sorters (FACS) can enrich the GFP-positive cell population. Such a sorting step eventually facilitates subsequent imaging analyses.

1. Dissociate the nucleofected cells by gently pipetting. Include $GFP_{1\text{-}10}$-expressing cells as a non-nucleofected control.
2. Spin down the dissociated cells for 5 min at 3000 × *g*.

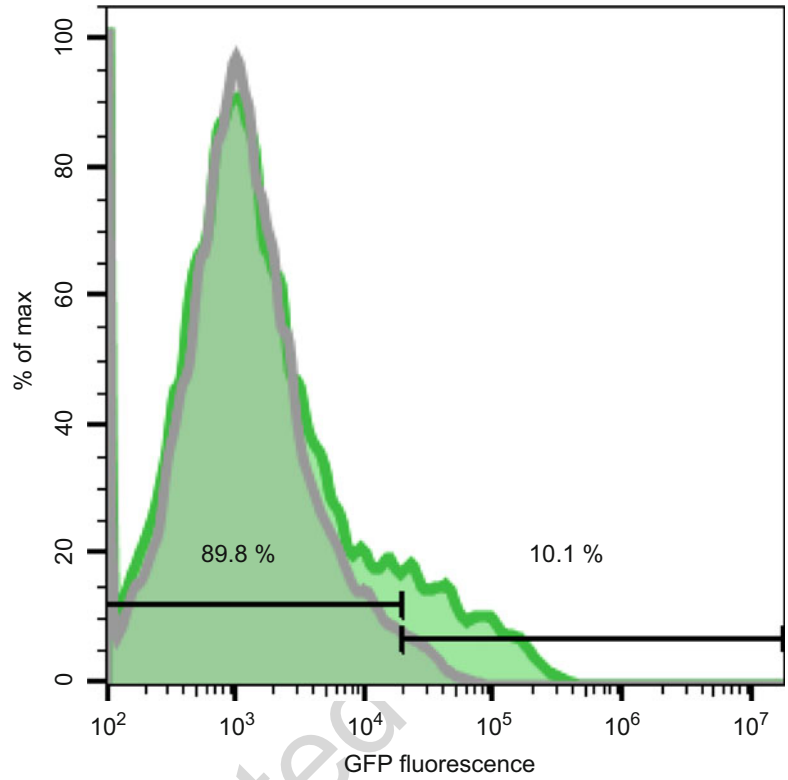

Fig. 4 A representative flow cytometry histogram of *GFP_{11}* knock-in HEK 293 cells (green). To estimate the *GFP_{11}* knock-in efficiency of the *HIST2H2BE* locus, GFP_{1-10}-expressing cells were used as a non-nucleofected control (gray). In this example, ~10% of the nucleofected cells are GFP-positive

3. Resuspend in 0.5 mL PBS and transfer into sterile FACS collection tubes.
4. Measure the GFP fluorescence signals of the nucleofected and non-nucleofected cells by flow cytometry. GFP is detected using a 488-nm excitation laser and a 535/40-nm band-pass filter.
5. Analyze data with FlowJo software. Measure the percentage of the GFP-positive cell population by comparing the nucleofected and non-nucleofected cells. As an example of typical results, Fig. 4 shows a representative flow cytometry analysis of *GFP_{11}* knock-in cells, in which endogenous histone 2B proteins are labeled with split GFP.
6. The same samples prepared through **steps 1–3** can also be FACS-sorted—sort GFP-positive cells from the population of the nucleofected cells into a sterile FACS collection tube.
7. Add 1 mL prewarmed DMEM medium to the sorted cells, and transfer to a 24-well culture plate.

8. Incubate the sorted cells in a humidified CO_2 incubator at 37 °C until analysis. The sorted cells can be cryopreserved for future use.

3.3.2 Fluorescence Microscopy

The functional integration of *GFP*$_{11}$ into a specific genomic locus and the reconstitution of those split GFP fragments can be validated using fluorescence microscopy.

1. Dilute fibronectin to 5 μg/mL in PBS. Fibronectin can be used as a substrate for culturing cells.
2. Coat eight-well chambered coverglass slides with 5 μg/mL fibronectin at 37 °C for 30 min or longer.
3. Remove the residual fibronectin solution from each well by aspiration.
4. Seed ~5 × 10^4 of GFP-positive cells (*see* Subheading 3.3.1) into each well.
5. Incubate cells for 4 h or overnight in a humidified CO_2 incubator at 37 °C.
6. Visualize GFP signals on a confocal microscope. GFP is excited by a 488-nm laser and detected by a 525/50-nm band-pass filter. Depending on the endogenous expression levels of target proteins, acquisition parameters (e.g., laser power, exposure time) may vary.
7. Analyze the acquired data with Fiji software. See Fig. 5 for a representative confocal image of endogenous histone 2B proteins labeled with split GFP at 100× magnification. The same cells used in Fig. 4 were imaged (*see* **Note 6**).

4 Notes

1. Purified Cas9 proteins are available from commercial vendors (e.g., Invitrogen TrueCut Cas9 protein v2). In addition, recombinant Cas9 proteins can be readily expressed and purified in *E. coli*. For example, bacterially expressed His-tagged Cas9 protein can be affinity-purified via standard Ni-NTA agarose affinity chromatography [30]. A variety of plasmids encoding His-tagged Cas9 can be obtained from Addgene (e.g., pET-28b-Cas9-His, # 47327).
2. In creating $GFP_{1\text{-}10}$-expressing cells, antibiotic selection using Geneticin would be our first choice due to the ease of the selection process. It only requires transfection with one single plasmid encoding the *GFP*$_{1\text{-}10}$ gene and the Geneticin resistance gene. After antibiotic selection for ~14 days, stable cell lines can be obtained. In contrast, lentiviral transduction requires multiple plasmids for an initial transfection into

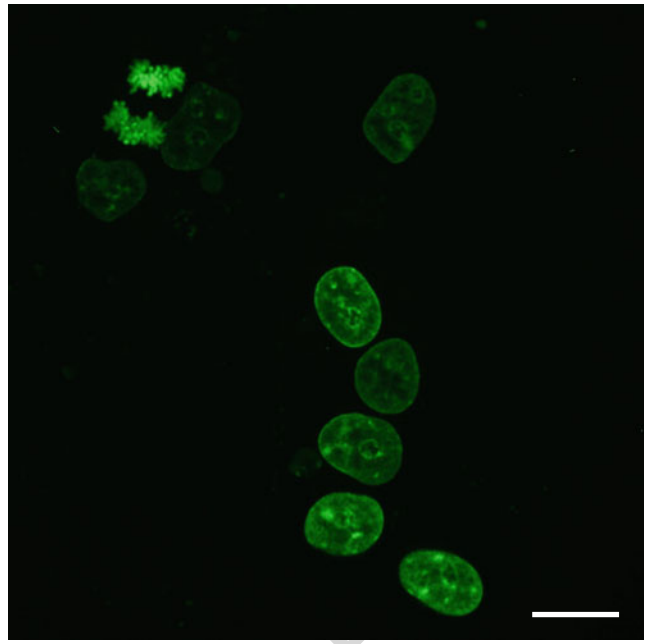

Fig. 5 A representative confocal image of *GFP_{11}* knock-in HEK 293 cells. Endogenous H2B proteins are labeled with split GFP. The resulting reconstituted GFP signal displays the nuclear localization of histone 2B. Scale bar, 20 μm

packaging cells to produce viral particles, which should be performed separately prior to the viral transduction into target cells. Moreover, additional safety precautions should be required during the procedure [31].

3. The molecular weight of sgRNA is ~37 kDa (e.g., 1850 ng/μL of sgRNA is ~50 μM).
4. Another standard method to directly deliver Cas9/sgRNA complexes into cells is the Neon transfection system from Invitrogen [32–35]. The Neon transfection system has also shown efficient transfection and nuclear uptake in various cell lines [36, 37].
5. Detection of endogenous proteins with split GFP is dependent on their expression levels. In the case of split GFP, ~30% of cellular proteins are estimated to be detected using a single copy of GFP_{11} [10]. For proteins with low expression levels, inserting tandem repeats of GFP_{11} can enhance the reconstituted GFP signal from these proteins [10].
6. To characterize the specific genomic integration of GFP_{11}, we can perform Sanger or deep sequencing [25]. To verify the localization of GFP11-tagged proteins, we can perform

immunofluorescence using available antibodies against the proteins themselves [10, 11].

Acknowledgments

We want to thank all members of the Kamiyama lab for critical comments on the manuscript; we particularly thank Melissa Inal and Kathy Bui for helpful discussion. This work was supported by an NIH R01 NS107558 (R.T. and D.K.). Ryo Tamura was supported by the Nakajima Foundation.

References

1. Leroy H, Lee R (2013) The Human Genome Project: big science transforms biology and medicine. Genome Med 5:79
2. Shaner NC, Patterson GH, Davidson MW (2007) Advances in fluorescent protein technology. J Cell Sci 120:4247–4260
3. Rizzo MA, Davidson MW, Piston DW (2009) Fluorescent protein tracking and detection: applications using fluorescent proteins in living cells. Cold Spring Harb Protoc 2009:pdb.top64
4. Lu W, Lakonishok M, Gelfand VI (2021) Gatekeeper function for Short stop at the ring canals of the Drosophila ovary. Curr Biol 31:3207–3220
5. Sengupta P, Seo AY, Pasolli HA, Song YE, Johnson MC, Lippincott-Schwartz J (2019) A lipid-based partitioning mechanism for selective incorporation of proteins into membranes of HIV particles. Nat Cell Biol 21:452–461
6. Gibson TJ, Seiler M, Veitia RA (2013) The transience of transient overexpression. Nat Methods 10:715–721
7. Doyon JB, Zeitler B, Cheng J, Cheng AT, Cherone JM, Santiago Y, Lee AH, Vo TD, Doyon Y, Miller JC, Paschon DE, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Drubin DG (2011) Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. Nat Cell Biol 13:331–337
8. Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell 157:1262–1278
9. Roberts B, Haupt A, Tucker A, Grancharova T, Arakaki J, Fuqua MA, Nelson A, Hookway C, Ludmann SA, Mueller IA, Yang R, Horwitz R, Rafelski SM, Gunawardane RN (2017) Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. Mol Biol Cell 28:2854–2874
10. Leonetti MD, Sekine S, Kamiyama D, Weissman JS, Huang B (2016) A scalable strategy for high-throughput GFP tagging of endogenous human proteins. Proc Natl Acad Sci U S A 113:3501–3508
11. Kamiyama D, Sekine S, Barsi-Rhyne B, Hu J, Chen B, Gilbert LA, Ishikawa H, Leonetti MD, Marshall WF, Weissman JS, Huang B (2016) Versatile protein tagging in cells with split fluorescent protein. Nat Commun 7:11046
12. Feng S, Sekine S, Pessino V, Li H, Leonetti MD, Huang B (2017) Improved split fluorescent proteins for endogenous protein labeling. Nat Commun 8:370
13. Feng S, Varshney A, Coto Villa D, Modavi C, Kohler J, Farah F, Zhou S, Ali N, Muller JD, Van Hoven MK, Huang B (2019) Bright split red fluorescent proteins for the visualization of endogenous proteins and synapses. Commun Biol 2:344
14. Zhou S, Feng S, Brown D, Huang B (2020) Improved yellow-green split fluorescent proteins for protein labeling and signal amplification. PLoS One 15:e0242592
15. Tamura R, Jiang F, Xie J, Kamiyama D (2021) Multiplexed labeling of cellular proteins with split fluorescent protein tags. Commun Biol 4:257
16. Chun W, Waldo GS, Johnson GV (2007) Split GFP complementation assay: a novel approach to quantitatively measure aggregation of tau in situ: effects of GSK3beta activation and caspase 3 cleavage. J Neurochem 103:2529–2539
17. Koker T, Fernandez A, Pinaud F (2018) Characterization of split fluorescent protein variants and quantitative analyses of their self-assembly process. Sci Rep 8:5344
18. Cabantous S, Terwilliger TC, Waldo GS (2005) Protein tagging and detection with

engineered self-assembling fragments of green fluorescent protein. Nat Biotechnol 23: 102–107

19. Sitron CS, Brandman O (2019) CAT tails drive degradation of stalled polypeptides on and off the ribosome. Nat Struct Mol Biol 26: 450–459
20. Kaser S, Willemin M, Schnarwiler F, Schimanski B, Poveda-Huertes D, Oeljeklaus S, Haenni B, Zuber B, Warscheid B, Meisinger C, Schneider A (2017) Biogenesis of the mitochondrial DNA inheritance machinery in the mitochondrial outer membrane of Trypanosoma brucei. PLoS Pathog 13:e1006808
21. Inglis AJ, Page KR, Guna A, Voorhees RM (2020) Differential modes of orphan subunit recognition for the WRB/CAML complex. Cell Rep 30:3691–3698
22. Park E, Lee HY, Woo J, Choi D, Dinesh-Kumar SP (2017) Spatiotemporal monitoring of Pseudomonas syringae effectors via type III secretion using split fluorescent protein fragments. Plant Cell 29:1571–1584
23. Batan D, Braselmann E, Minson M, Nguyen DMT, Cossart P, Palmer AE (2018) A multicolor split-fluorescent protein approach to visualize listeria protein secretion in infection. Biophys J 115:251–262
24. Li X, Zhu T, Tu H, Pan SQ (2020) Agrobacterium VirE3 uses its two tandem domains at the C-terminus to retain its companion VirE2 on the cytoplasmic side of the host plasma membrane. Front Plant Sci 11:464
25. Lin S, Staahl BT, Alla RK, Doudna JA (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife 3:e04766
26. Liu M, Rehman S, Tang X, Gu K, Fan Q, Chen D, Ma W (2018) Methodologies for improving HDR efficiency. Front Genet 9:691
27. Kim S, Kim D, Cho SW, Kim J, Kim JS (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res 24: 1012–1019
28. Lattanzi A, Meneghini V, Pavani G, Amor F, Ramadier S, Felix T, Antoniani C, Masson C, Alibeu O, Lee C, Porteus MH, Bao G, Amendola M, Mavilio F, Miccio A (2019) Optimization of CRISPR/Cas9 delivery to human hematopoietic stem and progenitor cells for therapeutic genomic rearrangements. Mol Ther 27:137–150
29. Kagoya Y, Guo T, Yeung B, Saso K, Anczurowski M, Wang CH, Murata K, Sugata K, Saijo H, Matsunaga Y, Ohashi Y, Butler MO, Hirano N (2020) Genetic ablation of HLA Class I, Class II, and the T-cell receptor enables allogeneic T cells to be used for adoptive T-cell therapy. Cancer Immunol Res 8: 926–936
30. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337: 816–821
31. Elegheert J, Behiels E, Bishop B, Scott S, Woolley RE, Griffiths SC, Byrne EFX, Chang VT, Stuart DI, Jones EY, Siebold C, Aricescu AR (2018) Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins. Nat Protoc 13:2991–3017
32. Hung KL, Meitlis I, Hale M, Chen CY, Singh S, Jackson SW, Miao CH, Khan IF, Rawlings DJ, James RG (2018) Engineering protein-secreting plasma cells by homology-directed repair in primary human B cells. Mol Ther 26:456–467
33. Gundry MC, Brunetti L, Lin A, Mayle AE, Kitano A, Wagner D, Hsu JI, Hoegenauer KA, Rooney CM, Goodell MA, Nakada D (2016) Highly efficient genome editing of murine and human hematopoietic progenitor cells by CRISPR/Cas9. Cell Rep 17: 1453–1461
34. Schumann K, Lin S, Boyer E, Simeonov DR, Subramaniam M, Gate RE, Haliburton GE, Ye CJ, Bluestone JA, Doudna JA, Marson A (2015) Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. Proc Natl Acad Sci U S A 112:10437–10442
35. Wu W, Lu Z, Li F, Wang W, Qian N, Duan J, Zhang Y, Wang F, Chen T (2017) Efficient in vivo gene editing using ribonucleoproteins in skin stem cells of recessive dystrophic epidermolysis bullosa mouse model. Proc Natl Acad Sci U S A 114:1660–1665
36. Modarai SR, Man D, Bialk P, Rivera-Torres N, Bloh K, Kmiec EB (2018) Efficient delivery and nuclear uptake is not sufficient to detect gene editing in CD34+ cells directed by a ribonucleoprotein complex. Mol Ther Nucleic Acids 11:116–129
37. Dwivedi PP, Anderson PJ, Powell BC (2012) Development of an efficient, non-viral transfection method for studying gene function and bone growth in human primary cranial suture mesenchymal cells reveals that the cells respond to BMP2 and BMP3. BMC Biotechnol 12:45