



Improved and Flexible HDR Editing by Targeting Introns in iPSCs

Juan Fu^{1,2} · Ya-Wen Fu³ · Juan-Juan Zhao³ · Zhi-Xue Yang³ · Si-Ang Li³ · Guo-Hua Li³ · Zi-Jun Quan³ · Feng Zhang³ · Jian-Ping Zhang³ · Xiao-Bing Zhang³ · Chang-Kai Sun^{1,4}

Accepted: 10 January 2022 / Published online: 28 January 2022
© The Author(s) 2022

Abstract

Highly efficient gene knockout (KO) editing of CRISPR–Cas9 has been achieved in iPSCs, whereas homology-directed repair (HDR)-mediated precise gene knock-in (KI) and high-level expression are still bottlenecks for the clinical applications of iPSCs. Here, we developed a novel editing strategy that targets introns. By targeting the intron before the stop codon, this approach tolerates reading frameshift mutations caused by nonhomologous end-joining (NHEJ)-mediated indels, thereby maintaining gene integrity without damaging the non-HDR-edited allele. Furthermore, to increase the flexibility and screen for the best intron-targeting sgRNA, we designed an HDR donor with an artificial intron in place of the endogenous intron. The presence of artificial introns, particularly an intron that carries an enhancer element, significantly increased the reporter expression levels in iPSCs compared to the intron-deleted control. In addition, a combination of the small molecules M3814 and trichostatin A (TSA) significantly improves HDR efficiency by inhibiting NHEJ. These results should find applications in gene therapy and basic research, such as creating reporter cell lines.

Keywords iPSCs, CRISPR, HDR, intron

Juan Fu and Ya-Wen Fu contributed equally to this work.

✉ Xiao-Bing Zhang
zhangxbhk@gmail.com

✉ Chang-Kai Sun
cksun110@vip.sina.com

¹ Liaoning Provincial Key Laboratory of Cerebral Diseases, Institute for Brain Disorders, Dalian Medical University, Dalian 116044, China

² Department of Obstetrics and Gynecology, The First Affiliated Hospital of Dalian Medical University, Dalian 116044, China

³ State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China

⁴ Research & Educational Center for the Control Engineering of Translational Precision Medicine (R-ECCE-TPM), School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China

Introduction

Induced pluripotent stem cells (iPSCs) provide an ideal source for cell replacement therapy and regenerative medicine due to their unlimited self-renewal and multidirectional differentiation ability [1]. However, the realization of the full therapeutic potential of human iPSCs requires further development of approaches to generate gene-modified or disease gene-corrected cells.

The clustered regularly interspaced short palindromic repeats (CRISPR)–Cas9 system has become a valuable tool for gene editing, from manipulating human cell genomes to creating gene-modified animal models. Its simplicity and robustness make it stand out from other genome editing technologies [2–4]. The CRISPR–Cas9 genome editing system, developed from the adaptive immune system of bacteria and archaea, consists of a Cas9 nuclease and a single-guide RNA (sgRNA) [5]. In this study, we used *Streptococcus pyogenes* (Sp) Cas9, the earliest and most commonly used Cas protein, for genome editing [3, 4, 6]. SpCas9 nuclease introduces double-stranded DNA breaks (DSBs) 3 bp upstream of the NGG protospacer adjacent motif (PAM) under the guidance of sgRNA. Since broken DNA is a dangerous signal for cells and causes severe cytotoxicity [7, 8], the DNA

repair machinery is recruited and activated rapidly to promote DNA ligations through DNA repair pathways. Two of the main DSB repair pathways in mammalian cells are non-homologous end-joining (NHEJ) and homologous recombination (HR), in which NHEJ generates a KO phenotype by introducing small insertions or deletions (indels) and is considered error-prone. In contrast, precise gene KI is a templated editing process guided by homology-directed repair (HDR) donors, which requires the presence of a recombination donor flanked with right and left homology arms (HAs) to generate precise editing outcomes.

Reporter cell lines are valuable tools for tracing cell lineages, visualizing corresponding gene expression levels, and investigating protein localization and function. CRISPR genome editing technology enables tagging a fluorescent gene to an endogenous gene to report the tagged gene. However, lower signal intensity is often a limiting factor, especially in genes with low expression. Developing brighter reporter cell lines, especially clinically relevant iPSCs and other cell lines, will provide an improved research tool.

Precise HDR gene editing has difficulty achieving acceptable efficiency for clinics even after CRISPR–Cas9 creates a DSB at the targeting site [9–11]. However, using a double cut donor plasmid design, we achieved a 5- to 10-fold increase in HDR KI efficiencies [12]. In addition, the transient overexpression of BCL-XL considerably increases genome editing efficiency in human iPSCs by enhancing cell survival after electroporation of editing plasmids [13]. Therefore, we used double-cut plasmid donor and BCL-XL plasmid electroporation for editing iPSCs in this study.

To establish a fluorescent reporter cell line, one straightforward strategy is to add a tag at the stop codon of a gene. However, the versatility of this method is limited because biallelic editing is much less efficient than monoallelic editing. Precise KI on one allele is often accompanied by indels on another allele, which may disrupt gene structure and cause gene deletion. Some groups have reported accurate gene integration by CRISPR–Cas9 intron targeting, which maintains the integrity of targeted endogenous genes and increases the rate of in-frame insertion compared to exon-based targeting [14–16]. In addition, multiple sgRNAs can often be identified to target an intron. In eukaryotic genomes, the precision and complexity of intron removal during pre-mRNA maturation requires the integrity of introns and other complex factors, and the disruption of standard splicing patterns can be a mechanism to downregulate the expression of a gene and cause diseases [17]. In addition, the Encyclopedia of DNA Elements (ENCODE) project has reported that introns often harbor enhancers for delicate regulation of gene expression [18]. Consequently, the realization of highly expressed fluorescent protein KI requires the integrity of genomic cis-elements, including in-frame fused exons and introns with complete structures.

In this study, we first conducted an intron-targeting KI strategy using an intron-deleted HDR donor. We achieved high efficiencies, but the fluorescent reporter expression intensity was lower than in cells edited with a mutated intron-containing HDR donor. Consequently, we developed an artificial intron KI strategy to generate edited cells efficiently without reducing the expression intensity of the fluorescent reporter genes. We also found that using an enhancer-containing intron may increase the reporter expression levels.

Materials and Methods

sgRNA Design

The CHOPCHOP website (<http://chopchop.cbu.uib.no>) [19, 20] was used to design appropriate sgRNAs targeting the last introns of human *OCT4*, *EEF1A1*, and *GAPDH*. Sequences of all the sgRNAs used in this study are listed in Supplementary Table 1.

Plasmids Construction

All plasmids expressing Cas9, BCL-XL, sgRNAs, or mNeonGreen HDR donors were constructed with a NEBuilder HiFi DNA Assembly Kit (New England Biolabs) as described previously [12, 13]. In short, all the vector components were amplified from human gDNA or plasmids in our lab by PCR using KAPA HiFi polymerase (KAPA Biosystems) and purified using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific). The PCR products were then assembled using the NEBuilder HiFi DNA Assembly kit following the manufacturer's instructions. Multiple colonies were chosen for Sanger sequencing (MCLAB) to identify the correct clones. All the vectors were verified by Sanger sequencing. The mNeonGreen HDR donor (without intron) consisted of a mNeonGreen reporter protein flanked by ~600 bp homologous arms (HAs) and sgDocut (donor cut) recognition sequences. The introns will be deleted after HDR; thus, the HDR-edited targets will not be recognized by the sgRNAs.

Constructions of Mutant or Artificial Intron-Containing HDR Donors

The double-cut donor plasmids used in this study were generated using the NEBuilder HiFi DNA Assembly kit (New England Biolabs), as detailed above. The self-cleaving E2A linker fragment is located between the modified intron-exon (the STOP codon on the last coding exon was deleted) and mNeonGreen or Crimson reporter, flanked by HAs and sgDocut recognition sequences [12]. For

knockin with an endogenous intron-containing donor, a stretch of point mutation was introduced into the HDR donor's intron. All the vectors were verified by Sanger sequencing. All the mutant wild-type introns or artificial introns are shown in Figs. 3a, b, and 4a. All the sequences of plasmid HDR donors used in this study are listed in Supplementary File 1.

Human iPSC Culture

iPSC lines were generated from anonymous adult donors by peripheral blood (PB) reprogramming using episomal vectors expressing OCT4, SOX2, MYC, KLF4, and BCL-XL [21–23]. The iPSCs used in this study have been published previously [13, 24]. hiPSCs were grown under feeder-free conditions and maintained on tissue culture-treated 6-well plates (BD) coated with 1% Matrigel (Corning) in fresh mTeSR™E8 medium (StemCell Technologies). Cells at 60–70% confluency were passaged with 0.5 mM EDTA in PBS. Cells were cultured in a humidified atmosphere with 5% CO₂ at 37 °C, and the medium was changed daily with fresh mTeSR™ E8 medium.

Electroporation of iPSCs

For genome editing of iPSCs, cells were transfected by electroporation using the Amaxa Human Stem Cell Nucleofector® Kit 2 (Lonza) and the program B-016 on a Lonza nucleofector 2b. Briefly, 70 µl electroporation solution was prepared for each reaction, including 57.4 µl of the nucleofector solution, 12.6 µl of the supplement, and plasmids. Generally, 1 µg of Cas9 plasmids, 0.5 µg of sgRNA plasmids, 0.5 µg of sgDocut plasmids (for cutting pDonor in some experiments), and 1 µg of pDonor plasmids were used. In addition, we also used 0.5 µg of BCL-XL plasmid to improve iPSC survival [13]. iPSCs at 60–70% confluency were dissociated with the addition of 400 µl of Accutase, gently pipetted three times, and filtered with a 70 µm filter to obtain a single-cell suspension. Approximately 1–1.5 × 10⁶ cells were washed with DPBS (Gibco) and centrifuged at 400×g for 5 min, and the supernatant was carefully aspirated by vacuum. The cells were then resuspended in electroporation solution and carefully transferred into the cuvette. After electroporation, the cuvette was incubated at 37 °C for ~5 min to improve cell survival [25]. The cells were then seeded onto Matrigel-coated plates in mTeSR™ E8 medium with 10 µM ROCK inhibitor Y-27632 (Millipore). Cells were gently handled during each step to reduce physical damage to the cells. One day later, cultures were fed fresh mTeSR™ E8 medium without any small molecules or drugs.

Flow Cytometry

Flow cytometry was performed to determine reporter gene HDR efficiency, as described previously [13]. iPSCs were dissociated with Accutase and acquired on a BD FACS Canto II flow cytometer three days post-electroporation. For HDR-mediated KI of mNeonGreen or Crimson reporter into the target genes, the fluorescence-positive cell population was considered the HDR-edited cells. The FITC or APC channel was used to determine the proportion of mNeonGreen⁺ or Crimson⁺ cells. Electroporation without relevant sgRNA was carried out as a negative control, which showed few mNeonGreen⁺ or Crimson⁺ cells. The FACS data were analyzed using FlowJo™ 10.

Illumina Deep Sequencing and Data Analysis

iPSCs edited without HDR donors were harvested three days after nucleofection for editing efficiency detection. Approximately 2 × 10⁵ cells were harvested for genomic DNA extraction using 10–20 µl digestion buffer, which consisted of 100 mM NaCl, 10 mM Tris pH 8, 5 mM EDTA, 0.5% Tween 20 (Sigma), and 1% proteinase K (ABM; 10 mg/ml). The mixtures were treated at 56 °C for 60 min, followed by 95 °C for 10 min. After a short spindown, one µl of the supernatant was used for PCR amplification. Genomic regions of interest were amplified by PCR using KAPA HiFi DNA polymerase. The thermal cycler program for primary PCR was as follows: 98 °C for 1 min, followed by 98 °C for 5 s, 64 °C for 5 s, 68 °C for 5 s, and 72 °C for 30 s for 30 cycles. PCR amplifications were verified by electrophoresis on 1% agarose gels. Then, 100 ng of barcoded PCR products from each sample were pooled for sequencing using Illumina's NovaSeq6000 System (Novogene). Novogene constructed the library and acquired raw data. The 150-bp paired-end high-throughput sequencing reads were merged with FLASH [26], followed by demultiplexing using the Barcode Splitter Python script (<https://pypi.org/project/barcode-splitter/>). The indel efficiencies were analyzed with the docker version of CRISPResso2 [27], which returns many results, including the number of reads, details of editing alleles, and editing efficiencies.

Nanopore Sequencing and Data Analysis

iPSCs edited with HDR donors were harvested three days after nucleofection. Genomic DNA was extracted using the Gentra Puregene Blood Kit (Qiagen). Genomic regions of interest were amplified by KAPA HiFi DNA polymerase. The PCR cycling condition was 98 °C for 1 min, followed by 98 °C for 10 s, 64 °C for 15 s, and 72 °C for 90 s for 30 cycles. PCR amplifications were verified by electrophoresis on 1% agarose gels. An equal amount of

barcoded PCR products of different targets were pooled for nanopore sequencing using PromethION (ONT, UK) at Novogene. Albacore (version 2.3.1, Oxford Nanopore Technologies) transformed raw fast5 data into bases and quality scores. Then, we processed the data as described previously [28, 29]. In brief, sequencing adapters were

removed by Porechop [30] (version 0.2.4) and then processed with Seqkit to grep for individual reads of PCR products. Next, we used Minimap2 [31] (version 2.14) to align the fastq sequences to the reference fasta files (HDR sequences). The aligned bam files were visualized using IGV [32] (version 2.10.3).

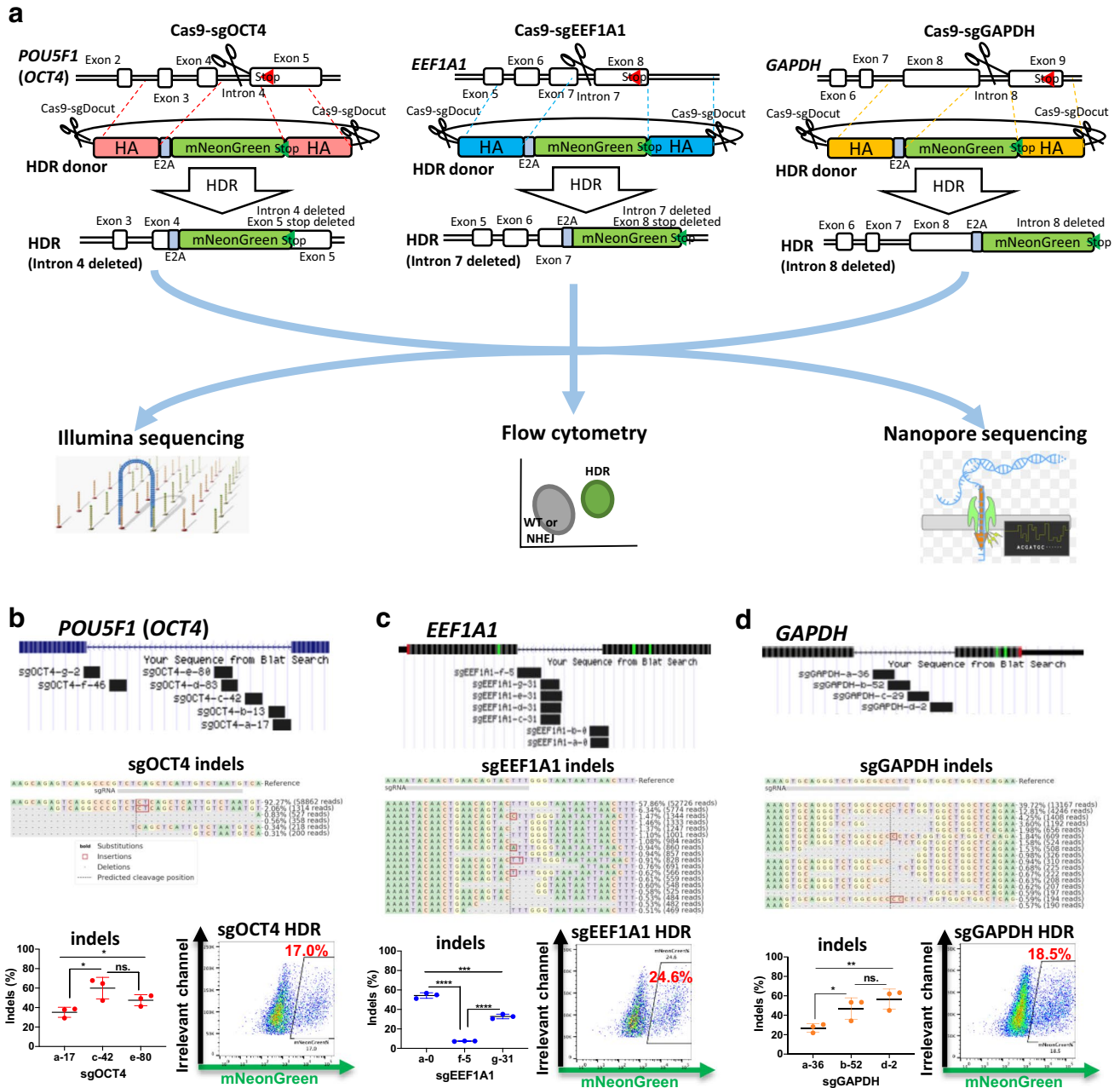


Fig. 1 CRISPR–Cas9 intron editing in iPSCs. **a** Schematics of genome editing with Cas9-sgRNA plasmids and an HDR donor template. **b–d** Editing efficiencies and mNeonGreen HDR KI at *OCT4* (**b**), *EEF1A1* (**c**), and *GAPDH* (**d**) in iPSCs. Top, Schematics of targeting the intron with multiple sgRNAs. The number in the sgRNA ID indicates the distance between the cut site and the proximate intron-exon junction. Middle, Representative editing outcomes ana-

lyzed by CRISPResso2. Bottom, editing efficiencies of representative sgRNAs (left) and mNeonGreen HDR KI (right). HDR efficiencies were determined by FACS analysis three days after electroporation. Data are shown as the mean ± s.d. Unpaired two-sided Student’s *t* tests were conducted. *****P* < 0.0001; ****P* < 0.001; ***P* < 0.01; **P* < 0.05; n.s., *P* ≥ 0.05

Small Molecules

Commercially available small molecules used in this study were trichostatin A (TSA) (Cayman; 89,730) and M3814 (MedKoo; 206,478). iPSCs were split into two culture wells after electroporation with editing plasmids to assess the effects of small molecules. Stock solutions of the HDAC inhibitor TSA and NHEJ inhibitor M3814 were prepared in dimethylsulfoxide (DMSO) (Sigma) and diluted to working concentrations before use. The parallel well with only DMSO (0.1%) served as a control. The medium was changed 24 h after the addition of small molecules.

Statistics and Reproducibility

The P values for different groups were calculated and analyzed by paired Student's t test. In all significance tests performed in the study, the data satisfied the normality criteria for t tests. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns., not significant. Bar graphs in figures were plotted, and s.d. error bars were calculated using GraphPad Prism 8. Scatterplots of correlation and linear regression analysis were calculated using Pearson's correlation coefficient.

Results

Generation of iPSC Reporter Cell Lines by CRISPR-Cas9 Intron Targeting

This study designed 18 sgRNAs that target the last introns of three commonly studied human genes (seven for *OCT4*, seven for *EEF1A1*, and four for *GAPDH*) (Supplementary Table 1). To tag these genes with a fluorescent protein, we conducted electroporation of iPSCs with Cas9-sgRNA and HDR double-cut donor plasmids. The HDR donor vectors contained E2A-mNeonGreen flanked by HAs omitting introns (Fig. 1a, top). We assessed the indel frequencies three days

after electroporation by amplifying target sequences and Illumina sequencing. HDR efficiencies were determined by FACS analysis of mNeonGreen-positive cells. No mNeonGreen-positive cells were detectable in the negative controls, suggesting that mNeonGreen-positive cells were HDR edited (Supplementary Fig. S1). To consolidate this conclusion, we performed nanopore sequencing of representative KI samples and observed precise insertion of the reporter gene (Fig. 1a).

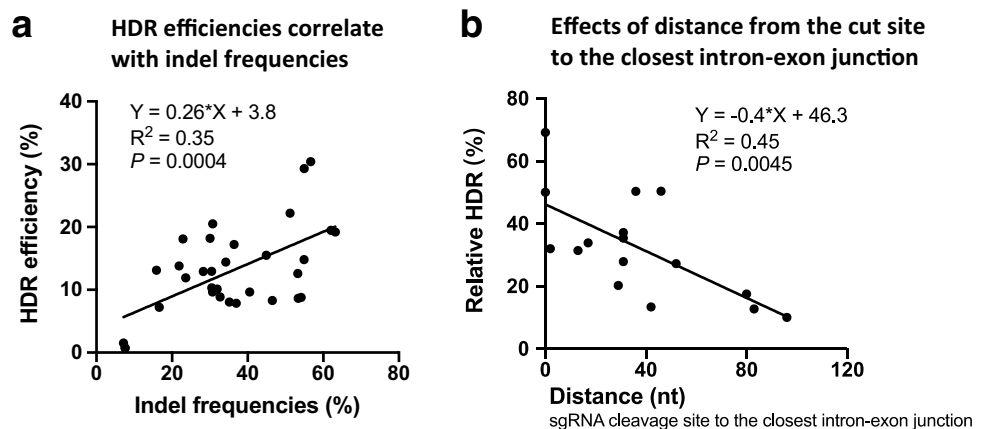
We designed sgRNAs to target the intron before the stop codon. To identify the optimal sgRNAs, we first compared the KO editing (indel) efficiencies of different sgRNAs (Fig. 1b-d, top). Editing efficiencies and patterns were assessed with CRISPResso2 [27] (Fig. 1b-d, middle). We observed variable indel and HDR efficiency for each sgRNA.

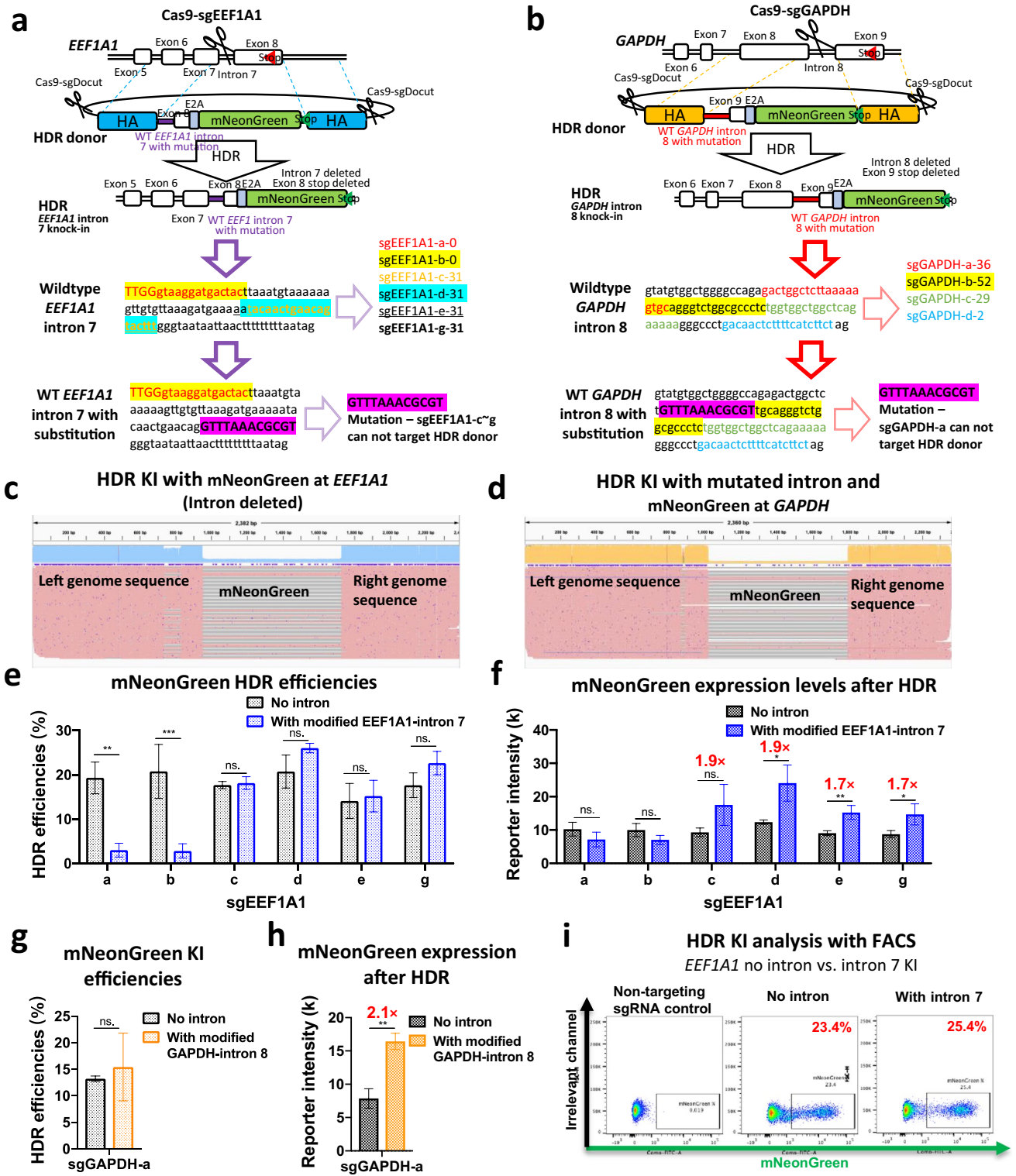
Relative HDR Efficiencies Negatively Correlate with the Distances from the Cut Site to the Intron-Exon Junction

We next analyzed the relationship between KO and KI efficiencies. As expected, Pearson linear regression analysis showed that HDR editings were proportional to indel efficiencies ($R^2 = 0.35$, $P = 0.0004$) (Fig. 2a), suggesting that the sgRNA targeting ability largely dictates HDR efficiency.

In our design, cleavage at the intron leaves stretches of sequences that mismatch homologous sequences on the HAs. We speculate that the length of mismatch sequences might negatively affect HDR editing. Therefore, we conducted a correlation analysis between the relative HDR and the distance from the sgRNA cleavage site to the closest intron-exon junction. As shown in Fig. 2b, linear regression analysis using aggregated data showed that the relative HDR efficiencies negatively correlated with the distance from the intron-exon boundary ($R^2 = 0.45$, $P = 0.0045$) (Fig. 2b). These results suggest that targeting sequences in the proximity of intron-exon junctions will lead to high-level relative HDR editing in intron-targeting applications.

Fig. 2 HDR efficiencies negatively correlate with the distance from the cut site to the intron-exon junction. **a** HDR efficiencies correlate with indel frequencies. $n = 32$. **b** Relative HDR efficiencies negatively correlate with the distance from the cut site to the closest intron-exon junction. The relative HDR efficiencies are defined as absolute HDR efficiencies divided by indel frequencies. $n = 16$





A Reporter Gene Knock-in without Deleting the Targeted Intron

In the above studies, the mNeonGreen reporter was fused with the open reading frame by self-cleaving peptide E2A,

and the intron was deleted after HDR-mediated KI (Fig. 1a). However, intron deletion may lead to decreased gene expression. As such, we modified the exon-E2A-mNeonGreen HDR donor by including the previously deleted intron and substituting the sgRNA targeting sequences with a 12-bp

Fig. 3 A reporter gene knock-in without deleting the target intron. **a, b** Schematics of mNeonGreen HDR knock-in at *EEF1A1* intron 7 (**a**) and *GAPDH* intron 8 (**b**) in iPSCs. The target sequences of used sgRNAs are highlighted. HDR donors with *EEF1A1* or *GAPDH* introns were mutated to avoid cleavage by certain sgRNAs. The purple highlight shows the substitution. **c, d** IGV visualization of mNeonGreen HDR KI at *EEF1A1* (**c**) and *GAPDH* (**d**). **c** The use of no-intron donor leads to the deletion of intron and insertion of mNeonGreen after HDR editing. **e, f** HDR KI efficiencies (**e**) and mNeonGreen intensity (**f**) after HDR editing with or without the *EEF1A1* intron. The a-g in the X-axis represent a series of sgRNAs that target different locations at *EEF1A1* in (**a**). Note: cleavage of the HDR donor with sgEEF1A1-a and b leads to low KI levels (**e**). n=3 independent biological repeats. **g, h** mNeonGreen intensities (expression levels) after editing with donors with or without introns. n=3 independent biological repeats. **i** Representative FACS panels after HDR editing using donors with or without introns 7 3 days after electroporation. Data are shown as the mean \pm s.d. Unpaired two-sided Student's *t* tests were conducted. ***P* < 0.01; **P* < 0.05; n.s., *P* \geq 0.05

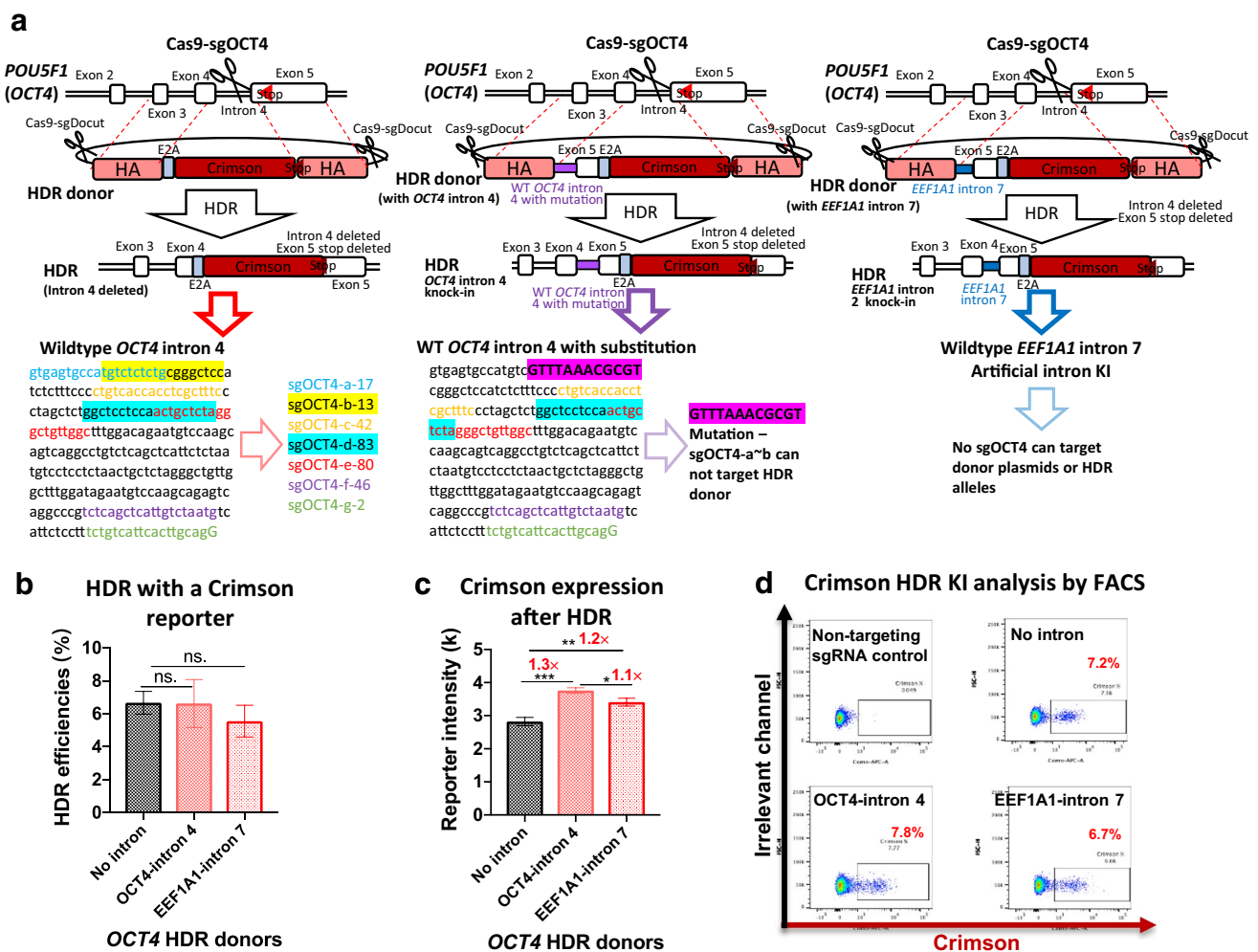


Fig. 4 HDR editing with an artificial intron-carrying donor increases reporter expression levels. **a** Schematics of the *OCT4*-Crimson HDR donor plasmid with no intron as a control (left) or with *OCT4* intron 4 (middle) or *EEF1A1* intron 7 (right) in iPSCs. The target sequences of the used sgRNAs highlight that *OCT4* intron 4 in the donor template is mutated to avoid cleavage by certain sgRNAs. The purple highlight shows the substitution (middle). **b, c** KI efficiencies (**b**) and

mismatch fragment to prevent cleavage of the HDR donor by the sgRNA (Fig. 3a-b). However, this strategy is only applicable to a limited number of sgRNAs. For instance, the substitutive exogenous fragments did not prevent cleavage by two sgEEF1A1s and three sgGAPDHs (Fig. 3a-b).

To validate the genetic changes after CRISPR-mediated modification of intron-exon-mNeonGreen KI, we PCR-amplified an ~1.6 kb region flanking the Cas9-sgRNA target sites at *EEF1A1* and *GAPDH*. The PCR products were sequenced and visualized by IGV [32], showing indels and insertions of ~700 bp fragments at *EEF1A1* and *GAPDH* (Fig. 3c-d) (including the modified endogenous intron and mNeonGreen, the HDR reference sequence is ~2.3 kb in length). We also determined HDR efficiencies and mNeonGreen expression levels by flow cytometry. sgEEF1A1-f was excluded from further

reporter intensity (**c**) after editing with HDR donors with no intron or with *OCT4* intron 4 (self-intron) or natural *EEF1A1* intron 7 (artificial intron) in iPSCs. n=3. **d** Representative FACS panels after HDR editing using donors with or without introns. HDR efficiencies were determined by FACS 5 days after electroporation. Data are shown as the mean \pm s.d. Unpaired two-sided Student's *t* tests were conducted. ***P* < 0.01; **P* < 0.05; n.s., *P* \geq 0.05

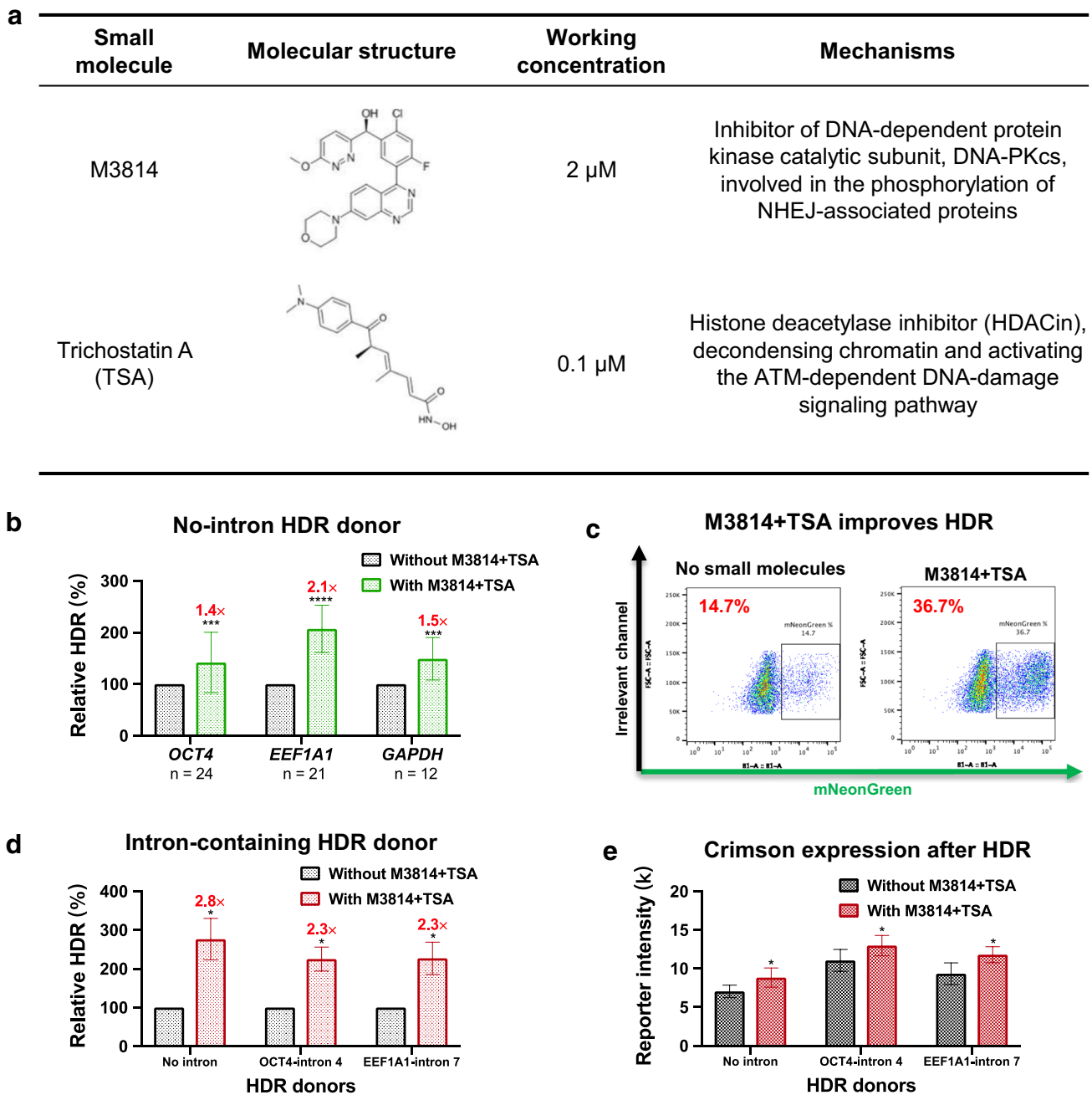


Fig. 5 M3814 and trichostatin A increase intron-targeting HDR editing efficiencies in iPSCs. **a** Molecular structures, working concentrations, and mechanisms of action of M3814 and trichostatin A (TSA). **b** M3814 and TSA enhance HDR efficiencies at *OCT4*, *EEF1A1*, and *GAPDH* in iPSCs with mNeonGreen HDR donors. Paired two-sided Student's *t* tests were conducted. **** $P < 0.0001$; *** $P < 0.001$. **c** Representative FACS panels after editing without or with M3814

and TSA. **d, e** HDR efficiencies (**d**) and Crimson intensities (reporter gene expression levels) (**e**) after editing with HDR donors containing no intron, *OCT4* intron 4, and natural *EEF1A1* intron 7 in the absence or presence of small molecules. HDR efficiencies were determined by FACS 5 days after electroporation for the APC signal. $n = 3$. Data are shown as the mean \pm s.d. Unpaired two-sided Student's *t* tests were conducted. ** $P < 0.01$; * $P < 0.05$; n.s., $P \geq 0.05$

analysis because of its low targeting efficiency (Fig. 1c, bottom left). Compared with the intron-deleted donor template, the *EEF1A1*-intron 7-containing donor showed similar HDR efficiencies (Fig. 3e, i) but significantly increased the mNeonGreen expression levels by ~2-fold (Fig. 3f, sgEEF1A1-c to

-g). The HDR efficiencies of sgEEF1A1-a and -b showed striking decreases due to sgRNA-mediated cleavage of template plasmids and HDR alleles (Fig. 3a). Similarly, the intron-containing donor did not increase the HDR efficiency at *GAPDH*, whereas it showed a 2-fold increase in reporter gene expression

(Fig. 3g-h). Careful examination of the genome sequence revealed that these introns contain an enhancer-like element (Supplementary Fig. S2). These data demonstrate that intron deletion after editing may decrease gene expression levels, whereas adding back a mutated intron will increase the reporter gene expression levels.

HDR Editing with an Artificial Intron-Containing Donor Template

The previous design must construct multiple mutated intron-containing HDR donor plasmids to screen the best sgRNA. To simplify the workflow, we decided to replace the endogenous intron with an artificial intron from another gene. We constructed the *OCT4*-Crimson HDR donor that contains *EEF1A1*-intron 7. We chose sgOCT4-a in the following studies, considering the targeting efficiency and relative HDR ratio (Figs. 1b and 2b).

Next, we examined each donor's HDR efficiencies and Crimson expression levels. Few Crimson-positive cells were detected in the negative controls (Fig. 4d). Consistent with the previous conclusions, the *OCT4*-intron 4-containing donor significantly increased the expression intensity by 1.3-fold (Fig. 4b-c). These data consolidate the conclusion that reconstruction of gene integrity leads to improved gene expression. The artificial-intron-containing HDR template showed a slightly higher expression level (1.2-fold) (Fig. 4b-c) than the intron-deleted donor. In the three HDR KI strategies, the donor template that can best restore the natural structure of the target gene locus showed the best reporter gene HDR KI (Fig. 4b-d). Taken together, this versatile and straightforward artificial intron-containing HDR donor is inferior to the endogenous-intron-containing template but performs better than the intron-deleting strategy.

The Combination of M3814 and Trichostatin A Increases Intron-Targeting HDR Editing Efficiencies

NHEJ is the predominant and fast-acting pathway to repair CRISPR-mediated DSBs and outcompetes other editings [24, 33]. Therefore, NHEJ inhibitors improve HDR efficiency [34–36]. One of the most effective NHEJ inhibitors is M3814, which strikingly improves HDR by blocking NHEJ [24, 37]. Additionally, we have demonstrated that HDAC inhibitors promote gene editing efficiencies at closed and open chromatin loci [38]. Thus, we speculated that the robust small-molecule combination M3814 and trichostatin A (TSA) should enhance HDR in our intron editing system (Fig. 5a).

We first investigated the effects of the M3814+TSA combination on HDR efficiencies using no-intron HDR donors

(Fig. 1a). Next, we collected mNeonGreen HDR editing data from 57 edited samples to assess statistical significance at *OCT4*, *EEF1A1*, and *GAPDH*. The relative HDR efficiency was computed by comparing the percentages of mNeonGreen-positive cells in the presence or absence of the two agents (Fig. 5b). The combination M3814+TSA performed well at all tested loci, increasing the HDR efficiencies by up to ~2.5-fold (Fig. 5c).

We further investigated the small molecule mix using intron-containing HDR donors. As expected, M3814 and TSA increased the HDR efficiencies in all types of HDR donors and showed a significant improvement (up to 2.8-fold) (Fig. 5d). Additionally, the combination slightly increased the average Crimson reporter expression by ~1.2-fold, which might result from a greater proportion of cells being edited biallelically. (Fig. 5e). These data further consolidate the conclusion that intron deletion after HDR may decrease gene expression, and adding back a mutated or artificial intron will increase the reporter gene expression levels.

Expression of the reporter gene after editing was also visualized using fluorescent microscopy (Supplementary Fig. S3a). We also found that unedited and edited iPSCs were morphologically similar one week after transfection of editing plasmids (Supplementary Fig. S3b). To investigate whether editing affects the cellular phenotypes of the iPSCs, we conducted staining of TRA-1-60, a stemness surface marker, in untreated and edited iPSCs and observed no difference in expression levels (Supplementary Fig. S3c). These data suggest that editing at these loci does not negatively affect iPSCs.

Discussion

Herein, we describe a novel intron-targeting HDR knock-in approach by using an artificial intron-containing donor template. Small indels mediated by NHEJ introduced into the intron are less likely to affect target gene function or induce gene deletion. In addition, intron targeting also theoretically increases the rate of in-frame insertion compared to exon-based targeting [16]. Furthermore, an intron retention strategy avoids the disruption of cis-regulatory elements such as proximal or distal enhancers at the intron regions, thus leading to high-level expression of reporter genes. We showed an up to 2-fold increase in reporter gene expression (Fig. 3f, h). Therefore, these intron knock-in strategies have two advantages: (1) inserted exogenous genes can faithfully recapitulate the expression pattern of targeted endogenous genes; (2) one can achieve higher expression levels of the fluorescent reporter gene.

After CRISPR-mediated dsDNA cleavage, sgRNAs and contextual sequences primarily determine the indel

frequencies, editing patterns, and HDR efficiencies (Figs. 1b–d and 2a). However, we also observed an interesting phenomenon. The relative HDR efficiency was related to the target site on the intron, where the sgRNAs targeting sequences close to the adjacent exon showed better relative HDR (Fig. 2b). Editing with different intron-targeting sgRNAs with the same intronless HDR donor leads to the generation of two mismatch sequences with different lengths. After DSBs, the surrounding sequences survey for homology during HDR. A significant stretch of nucleotides that mismatches the donor template will negatively affect the homology search. Thus, the best DSB-mediated DNA gaps are in principle flanked by the exact homologous sequences for HDR. Our data suggest that it is favorable for HDR to use sgRNAs that generate HA-friendly DSBs or target the proximity of the intron-exon junction. At least one arm has the shortest mismatched sequence in such a case. Another laboratory reported that the nonhomologous sequence of the donor reduced the efficiency of transgene integration [39], lending support to our conclusion.

A primary feature of most eukaryotic genes is that they are interrupted by introns, which are removed during transcription to create mRNAs with intact open reading frames. For example, the average human gene contains eight exons and seven introns, producing an average of three or more alternatively spliced mRNA isoforms [40]. Introns can influence and enhance initial transcription of the gene, editing and polyadenylation of the pre-mRNA, nuclear export, RNA stability and translation, and mRNA decay [41, 42]. In addition, multiple intron regions show enhancer-like features. Thus, it is not surprising that disruption or loss of introns can affect gene regulation and expression levels.

Based on the essential role of introns in the human genome, we modified our HDR strategy, that is, knocking in reporter genes without disrupting the intron, thus maintaining the integrity of the natural genes. As a result, the intron-containing donor showed HDR efficiencies similar to those of the no-intron donors (Figs. 3e, g and 4b). This result consolidates the conclusion that the editing efficiencies and HDR mainly depend on the sgRNA context and intron-exon junctions (Fig. 2). However, the intron-containing HDR donors significantly increased the fluorescent reporter expression levels (Figs. 3f, h, and 4c). We checked the targeted intron regions of these three genes on UCSC Genome Browser ENCODE Candidate Cis-Regulatory Elements (cCREs) and observed that the targeted intron regions have proximal or distal enhancer-like signatures (Supplementary Fig. S2). These associations can explain the improved fluorescent reporter expression levels after intron-containing HDR editing. To the best of our knowledge, we are the first to report a flexible artificial intron KI strategy that allows the screening of

multiple sgRNAs. Although it does not show a significant improvement over the endogenous-intron-containing donor, it is still a handy choice for HDR editing to quickly establish a report line with high efficiency and expression levels.

M3814 and TSA increase HDR efficiency in many editing systems [24, 37, 38]. M3814 promotes HDR editing by strongly inhibiting NHEJ-mediated indels. As an effective HDAC inhibitor, TSA can decondensate chromatin, which increases the local concentration of sgRNAs, Cas9, and HDR donors. Thus, combined M3814 and TSA treatment strongly inhibits NHEJ and increases chromatin accessibility simultaneously. The addition of M3814 and TSA increased the apparent HDR efficiency by 2.8-fold in iPSCs (Fig. 5d).

In summary, we present a flexible and improved method using an artificial intron-containing HDR donor template to create iPSC reporter cell lines efficiently. Furthermore, targeting the intron decreases the possibility of disrupting the coding sequence. In addition, M3814, together with TSA, considerably increases HDR editing efficiencies.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12015-022-10331-1>.

Author Contributions XBZ, JF, and YWF were involved in the study design. XBZ, JF, JJZ, WW, and YWF contributed to the study concepts. JF, YWF, GHL, ZXY, and FZ were involved in data acquisition. JF, SAL, ZJQ, and YWF were involved in data quality control and algorithms. JF and YWF were involved in data analysis and interpretation. JF and YWF contributed to statistical analysis. JF, XBZ, YWF, and CKS wrote the manuscript. XBZ, JPZ, and CKS were involved in administrative support and supervision. All authors approved the article for submission and publication.

Funding National Natural Science Foundation of China (81870149, 82070115); National Key Research and Development Program of China (2019YFA0110803, 2019YFA0110204, 2019YFA0110802); CAMS Innovation Fund for Medical Sciences (CIFMS) (2021-I2M-1-041); Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2020-PT310-011).

Data Availability The authors declare that all data supporting the conclusions in this study are available within the article and its extended information files on reasonable requests. All Illumina generation and Nanopore sequencing data gzip files have been deposited in the GEO database under the accession code GSE155891.

Declarations

Ethical Approval and Consent to Participate All PB samples were harvested and handled according to institutional guidelines and in compliance with all relevant ethical standards.

Consent to Publish Not applicable.

Conflict of Interest None declared.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Karagiannis, P., Takahashi, K., Saito, M., Yoshida, Y., Okita, K., Watanabe, A., Inoue, H., Yamashita, J. K., Todani, M., Nakagawa, M., et al. (2019). Induced pluripotent stem cells and their use in human models of disease and development. *Physiological Reviews*, *99*, 79–114.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*, 816–821.
- 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.
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, *339*, 823–826.
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J., Charpentier, E., Haft, D. H., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews. Microbiology*, *13*, 722–736.
- Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *Elife*, *2*, e00471.
- Kim, S., Koo, T., Jee, H.G., Cho, H.Y., Lee, G., Lim, D.G., Shin, H.S., & Kim, J.S. (2018). CRISPR RNAs trigger innate immune responses in human cells. *Genome Research*, *28*(3), 367–373.
- Sun, L., Wu, J., Du, F., Chen, X., & Chen, Z. J. (2013). Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science*, *339*, 786–791.
- Rouet, P., Smih, F., & Jasin, M. (1994). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *91*, 6064–6068.
- Bollag, R. J., Waldman, A. S., & Liskay, R. M. (1989). Homologous recombination in mammalian cells. *Annual Review of Genetics*, *23*, 199–225.
- Bibikova, M., Beumer, K., Trautman, J. K., & Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. *Science*, *300*, 764.
- Zhang, J. P., Li, X. L., Li, G. H., Chen, W., Arakaki, C., Botimer, G. D., Baylink, D., Zhang, L., Wen, W., Fu, Y. W., et al. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biology*, *18*, 35.
- Li, X. L., Li, G. H., Fu, J., Fu, Y. W., Zhang, L., Chen, W., Arakaki, C., Zhang, J. P., Wen, W., Zhao, M., et al. (2018). Highly efficient genome editing via CRISPR-Cas9 in human pluripotent stem cells is achieved by transient BCL-XL overexpression. *Nucleic Acids Research*, *46*, 10195–10215.
- Serebrennik, Y. V., Sansbury, S. E., Kumar, S. S., Henao-Mejia, J., & Shalem, O. (2019). Efficient and flexible tagging of endogenous genes by homology-independent intron targeting. *Genome Research*, *29*, 1322–1328.
- Suzuki, K., Yamamoto, M., Hernandez-Benitez, R., Li, Z., Wei, C., Soligalla, R. D., Aizawa, E., Hatanaka, F., Kurita, M., Reddy, P., et al. (2019). Precise in vivo genome editing via single homology arm donor mediated intron-targeting gene integration for genetic disease correction. *Cell Research*, *29*, 804–819.
- Li, J., Zhang, B. B., Ren, Y. G., Gu, S. Y., Xiang, Y. H., & Du, J. L. (2015). Intron targeting-mediated and endogenous gene integrity-maintaining knockin in zebrafish using the CRISPR/Cas9 system. *Cell Research*, *25*, 634–637.
- Faustino, N. A., & Cooper, T. A. (2003). Pre-mRNA splicing and human disease. *Genes & Development*, *17*, 419–437.
- Maher, B. (2012). ENCODE: The human encyclopaedia. *Nature*, *489*, 46–48.
- Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M., & Valen, E. (2014). CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research*, *42*, W401–W407.
- Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., & Valen, E. (2019). CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research*, *47*, W171–W174.
- Su, R. J., Baylink, D. J., Neises, A., Kiroyan, J. B., Meng, X., Payne, K. J., Tschudy-Seney, B., Duan, Y., Appleby, N., Kearns-Jonker, M., et al. (2013). Efficient generation of integration-free ips cells from human adult peripheral blood using BCL-XL together with Yamanaka factors. *PLoS One*, *8*, e64496.
- Wen, W., Zhang, J. P., Xu, J., Su, R. J., Neises, A., Ji, G. Z., Yuan, W., Cheng, T., & Zhang, X. B. (2016). Enhanced generation of integration-free iPSCs from human adult peripheral blood mononuclear cells with an optimal combination of Episomal vectors. *Stem Cell Reports*, *6*, 873–884.
- Meng, X., Neises, A., Su, R. J., Payne, K. J., Ritter, L., Gridley, D. S., Wang, J., Sheng, M., Lau, K. H., Baylink, D. J., et al. (2012). Efficient reprogramming of human cord blood CD34+ cells into induced pluripotent stem cells with OCT4 and SOX2 alone. *Molecular Therapy*, *20*, 408–416.
- Fu, Y. W., Dai, X. Y., Wang, W. T., Yang, Z. X., Zhao, J. J., Zhang, J. P., Wen, W., Zhang, F., Oberg, K. C., Zhang, L., et al. (2021). Dynamics and competition of CRISPR-Cas9 ribonucleoproteins and AAV donor-mediated NHEJ, MMEJ and HDR editing. *Nucleic Acids Research*, *49*, 969–985.
- Rols, M. P., Delteil, C., Serin, G., & Teissie, J. (1994). Temperature effects on electrotransfection of mammalian cells. *Nucleic Acids Research*, *22*, 540.
- Magoc, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*, 2957–2963.
- Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., Cole, M. A., Liu, D. R., Joung, J. K., Bauer, D. E., et al. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology*, *37*, 224–226.
- Wen, W., Quan, Z. J., Li, S. A., Yang, Z. X., Fu, Y. W., Zhang, F., Li, G. H., Zhao, M., Yin, M. D., Xu, J., et al. (2021). Effective control of large deletions after double-strand breaks by homology-directed repair and dsODN insertion. *Genome Biology*, *22*, 236.
- Quan, Z.-J., Li, S.-A., Yang, Z.-X., Zhao, J.-J., Li, G.-H., Zhang, F., Wen, W., Cheng, T., & Zhang, X.-B. (2021). GREPort-seq: A robust workflow to detect changes after gene editing through long-range PCR and Nanopore sequencing. *bioRxiv*.

30. Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom*, *3*, e000132.
31. Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*, 3094–3100.
32. Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*, 24–26.
33. Lieber, M. R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual Review of Biochemistry*, *79*, 181–211.
34. Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., & Ploegh, H. L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature Biotechnology*, *33*, 538–542.
35. Yeh, C. D., Richardson, C. D., & Corn, J. E. (2019). Advances in genome editing through control of DNA repair pathways. *Nature Cell Biology*, *21*, 1468–1478.
36. Gavande, N. S., VanderVere-Carozza, P. S., Pawelczak, K. S., Mendoza-Munoz, P., Vernon, T. L., Hanakahi, L. A., Summerlin, M., Dynlacht, J. R., Farmer, A. H., Sears, C. R., et al. (2020). Discovery and development of novel DNA-PK inhibitors by targeting the unique Ku-DNA interaction. *Nucleic Acids Research*, *48*, 11536–11550.
37. Riesenberger, S., Chintalapati, M., Macak, D., Kanis, P., Maricic, T., & Paabo, S. (2019). Simultaneous precise editing of multiple genes in human cells. *Nucleic Acids Research*, *47*, e116.
38. Zhang, J. P., Yang, Z. X., Zhang, F., Fu, Y. W., Dai, X. Y., Wen, W., Zhang, B., Choi, H., Chen, W., Brown, M., et al. (2021). HDAC inhibitors improve CRISPR-mediated HDR editing efficiency in iPSCs. *Science China. Life Sciences*, *64*, 1449–1462.
39. Yao, X., Zhang, M., Wang, X., Ying, W., Hu, X., Dai, P., Meng, F., Shi, L., Sun, Y., Yao, N., et al. (2018). Tild-CRISPR allows for efficient and precise gene Knockin in mouse and human cells. *Developmental Cell*, *45*, 526–536 e525.
40. Lee, Y., & Rio, D. C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. *Annual Review of Biochemistry*, *84*, 291–323.
41. Le Hir, H., Nott, A., & Moore, M. J. (2003). How introns influence and enhance eukaryotic gene expression. *Trends in Biochemical Sciences*, *28*, 215–220.
42. Jacob, A. G., & Smith, C. W. J. (2017). Intron retention as a component of regulated gene expression programs. *Human Genetics*, *136*, 1043–1057.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.