


SCIENTIFIC REPORTS



Corrected: Author Correction

OPEN

An episomal vector-based CRISPR/Cas9 system for highly efficient gene knockout in human pluripotent stem cells

Yifang Xie¹, Daqi Wang², Feng Lan³ , Gang Wei², Ting Ni² , Renjie Chai⁵, Dong Liu⁶, Shijun Hu⁷, Mingqing Li³, Dajin Li³, Hongyan Wang^{3,8} & Yongming Wang^{2,3}

Human pluripotent stem cells (hPSCs) represent a unique opportunity for understanding the molecular mechanisms underlying complex traits and diseases. CRISPR/Cas9 is a powerful tool to introduce genetic mutations into the hPSCs for loss-of-function studies. Here, we developed an episomal vector-based CRISPR/Cas9 system, which we called epiCRISPR, for highly efficient gene knockout in hPSCs. The epiCRISPR system enables generation of up to 100% Insertion/Deletion (indel) rates. In addition, the epiCRISPR system enables efficient double-gene knockout and genomic deletion. To minimize off-target cleavage, we combined the episomal vector technology with double-nicking strategy and recent developed high fidelity Cas9. Thus the epiCRISPR system offers a highly efficient platform for genetic analysis in hPSCs.

Human pluripotent stem cells (hPSCs), including human embryonic stem cells (hESCs)¹ and the closely related human induced pluripotent stem cells (iPSCs)², hold great promise for modeling human organ development, analyzing disease mechanisms, and developing potential therapies. However, to fulfill these potential applications, it is crucial to develop methods for efficient and controllable genetic manipulation.

The advent of genome editing technology allows the direct introduction of specific genetic mutations into the hPSCs^{3,4}. Clustered regularly interspaced short palindromic repeats/CRISPR-associated nuclease 9 (CRISPR/Cas9), adapted from microbial adaptive immune defense system^{5,6}, is a recently reported genomic editing tool that is rapidly gaining popularity due to the ease of assembly and high efficiency^{3,4}. In this system, a 20 nt guide RNA (gRNA) directs the Cas9 nuclease to generate site-specific double-stranded breaks (DSBs). The DSBs are typically repaired by the cell's endogenous DNA repair machinery through non-homologous end-joining (NHEJ), resulting in nonspecific small insertions and deletions (indels) useful for generating loss-of-function mutations^{3,4}. The DSBs can also be repaired by homology-directed repair (HDR) using an introduced DNA repair template, such as a double-stranded DNA donor plasmid or a single-stranded oligo DNA nucleotide (ssODN), leading to knock-in of precise mutations or reporters⁷⁻⁹. The genome editing through HDR pathway is less efficient than NHEJ and often requires drug selection of the successful knock-in cells^{7,8}.

¹Institute of Biomedical Sciences, Shanghai Medical College, Fudan University, Shanghai, 200032, China. ²The State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, 200438, China. ³The Key Lab of Reproduction Regulation of NPPFC in SIPPR, Institute of Reproduction & Development in Obstetrics & Gynecology Hospital, Fudan University, Shanghai, 200011, China. ⁴Beijing Anzhen Hospital, Beijing Institute of Heart Lung and Blood Vessel Disease, Capital Medical University, Beijing, 100029, China. ⁵Co-innovation Center of Neuroregeneration, Key Laboratory for Developmental Genes and Human Disease, Ministry of Education, Institute of Life Sciences, Southeast University, Nanjing, 210096, China. ⁶Co-innovation Center of Neuroregeneration, Jiangsu Key Laboratory of Neuroregeneration, Nantong University, Nantong, 226001, China. ⁷Institute for Cardiovascular Science & Department of Cardiovascular Surgery of the First Affiliated Hospital, Soochow University, Soochow, 215007, China. ⁸Children's Hospital of Fudan University, 399 Wanyuan Road, Shanghai, 201102, China. Yifang Xie and Daqi Wang contributed equally to this work. Correspondence and requests for materials should be addressed to H.W. (email: wanghy@fudan.edu.cn) or Y.W. (email: ymw@fudan.edu.cn)

Although high efficiency of gene knockout has been achieved in many immortalized tumor cell lines¹⁰, it has remained a challenge in hPSCs which are more difficult to transfect and less resilient to DNA damage¹¹. The genome editing is generally performed by transient expression of Cas9 and a gRNA^{3,11,12}. The typical efficiencies of gene knockout have been reported to be 1–25% without any subsequent selection steps in hPSCs^{3,11,13–16}. It is laborious and time-consuming to isolate the homozygous knockout hPSC clones with current efficiency. For example, Gonzalez *et al.* failed to isolate homozygous *GATA6* knockout mutants out of 384 clones⁹. Several strategies have been developed to improve the editing efficiency. Transfection followed by fluorescence-assisted cell sorting for Cas9_GFP+ cells or drug selection could significantly increase editing efficiency with 10–88% indel rates^{17–21}. Generation of SpCas9-expressing cell lines is another strategy to increase genome editing efficiency (24–91%)^{9,22}.

Recently, Li *et al.* have achieved high efficiency (8–76%) of genome editing by using an episomal vector to express Cas9 and gRNA²³. In this study, we used episomal CRISPR/Cas9 system for efficient genome editing in hPSCs. The vector contains OriP/EBNA1 components originated from Epstein-Bar virus, which can drive plasmid duplication once per cell division in eukaryotes²⁴, allowing Cas9 and gRNAs persistently expressed in cells; the vector contains a puromycin resistance gene, allowing enrichment of transfected cells by drug selection. The episomal vector is lost with a rate of 3–6% per cell generation after termination of the drug selection²⁵, allowing removal of exogenous genes. The epiCRISPR system enables up to 100% gene knockout, which could facilitate efficient genetic analysis in hPSCs.

Results

Establishment of an episomal vector-based CRISPR/Cas9 system. The current genome editing technology relies on transient expression of Cas9 and a specific gRNA. We expected that the extension of the Cas9 and gRNA expression would increase the modification efficiency. We designed an all-in-one OriP/EBNA1-based vector, which we named epiCRISPR, expressing a gRNA, Cas9, puromycin resistance gene (for enrichment of the transfected cells through drug selection) and GFP (for tracking transfection efficiency) separated by self-cleaving T2A peptide (Fig. 1a and b).

First, we analyzed the capacity of the epiCRISPR system for supporting gene expression in hPSCs (all the experiments in this study were performed in one hESC line (H9) and two iPSC lines). The plasmid was delivered into the cells using lipid-based transfection and low transfection efficiency was observed based on GFP expression (Fig. 1c, left panel). The puromycin selection was started 24 hours post-transfection. While the GFP-negative cells kept dying, the GFP-positive cells kept robustly proliferating (Fig. 1c, middle panel). The selection is generally complete in 5–7 days. The consistent supply of puromycin is required to prevent epiCRISPR vector loss during cell proliferation.

The ideal modified cells should be free of genomic integration and exogenous gene expression. To analyze if the epiCRISPR vector can be removed, we stopped drug selection and measured the amount of episomal vector every two days by quantitative polymerase chain reaction (qPCR). The episomal vector was decreased dramatically in the first week (Fig. 1d). We dissociated cells into single cells for colony formation for 15 days. All the single cell-derived clones are GFP-negative (Fig. 1c, right panel), and the vector is undetectable by PCR (Fig. 1e), indicating that the episomal vector can be efficiently removed. In summary, the epiCRISPR system allows modified cells free of genomic integration and exogenous gene expression in hPSCs.

The epiCRISPR system significantly promotes gene knockout in hPSCs. Next, we analyzed the capacity of the epiCRISPR system for gene knockout in hPSCs. We designed a panel of gRNAs targeting six loci (*DYRK1A*, *EMX1*, *AAVS1*, *VEGFA*, *APC1* and *MLH1*) located on different chromosomes. These targeted loci contain restriction sites so that we can assess the modification efficiency with Restriction Fragment Length Polymorphism (RFLP) assay. We observed an increasing insertion and deletion (indel) rates over time (Supplementary Figures S1 and S2). For example, the indel rates for *AAVS1* locus were 19% at day 5, 83% at day 10 and 93% at day 15 after transfection in hESCs (Fig. 2a and b; Supplementary Table S1). We observed that the indel rates varied among three cell lines for the *VEGFA* locus at day 10. The possible reason is that these cell lines have different origin and *VEGFA* locus might be associated with different epigenetic modification which influenced Cas9 accessibility. The sequencing results revealed even higher indel rates (27/30 = 90% for *EMX1* locus; 14/17 = 82% for *MLH1* locus) because a portion of the restriction sequence was not influenced by indels (Fig. 2c). The indel rates varied from 82 to 100% at day 15 depending on gRNAs and cell lines. We analyzed 15 single cell-derived clones modified at *AAVS1* locus and 14 clones were biallelic knockout (Fig. 2d). We further investigated if the episomal vector has advantage over transient plasmid with the same editing time. We performed side-by-side comparison of the epiCRISPR system to the popularly used editing plasmids for three gRNAs with either Puromycin selection (pX459 plasmid) or cell sorting (pX458 plasmid)²⁶. The epiCRISPR system and the transient plasmids generated comparable efficiency of editing at day 2 and day 4 (Supplementary Figure S3), demonstrating that the episomal vector could not increase genome editing without elongating editing time.

To determine the capacity of the epiCRISPR system for double-gene knockout in hPSCs, we expressed two gRNAs on the epiCRISPR vector (Fig. 3a). Multiplexed targeting of *DYRK1A* & *EMX1*, *AAVS1* & *VEGFA* and *APC1* & *MLH1* achieved similar indel rates to the single-gene knockout at day 15 (Fig. 3b; Supplementary Figure S4 and Table S1). We analyzed 15 single cell-derived clones modified at *DYRK1A* & *EMX1* loci in the hESCs and 9 clones were homozygous knockout for both genes (Fig. 3c). Interestingly, all 15 clones were homozygous knockout at the *DYRK1A* locus. Therefore, the epiCRISPR system is a powerful platform for single- and double-gene knockout in hPSC lines.

The epiCRISPR system for efficient genomic deletions in hPSCs. To determine the capacity of the epiCRISPR system for genomic deletions in hPSCs, we designed paired gRNAs targeting the 2nd exon of *DYRK1A*

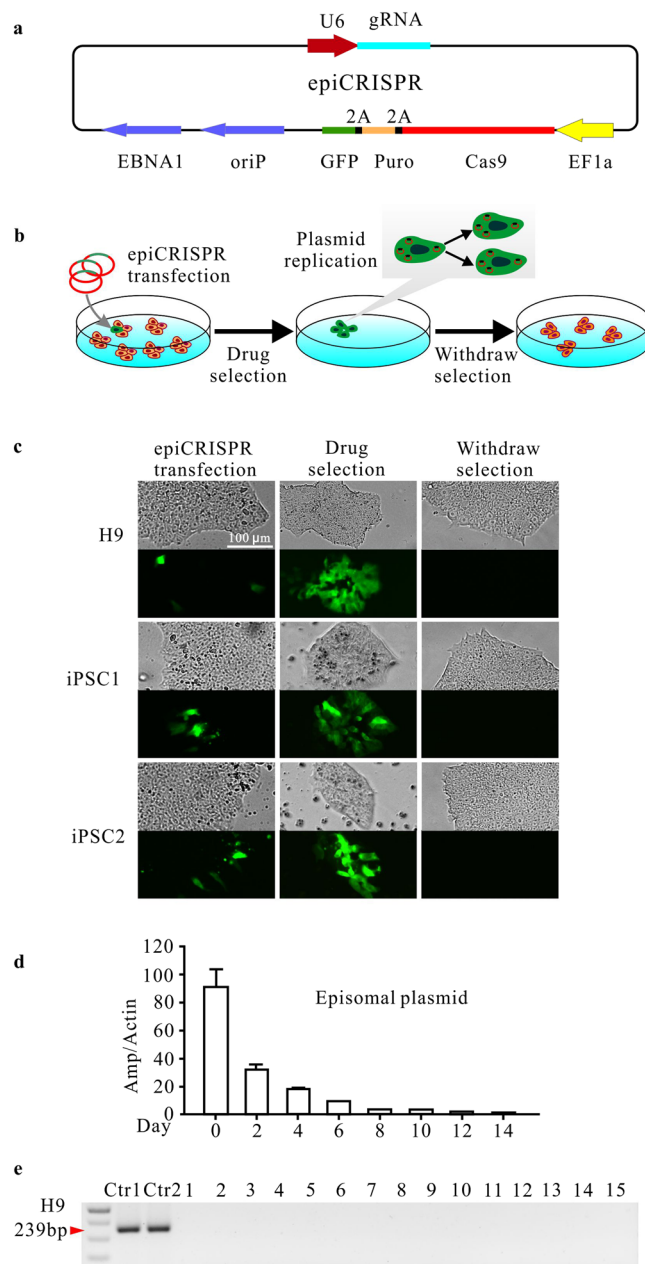


Figure 1. An epiCRISPR system for exogenous gene-free genome editing in hPSCs. **(a)** Schematic of the epiCRISPR system design. The vector contains a U6 promoter-driven gRNA scaffold, an EF1a promoter-driven Cas9 fused to puromycin resistance gene and GFP with P2A peptides, and OriP/EBNA1 elements for the plasmid replication in eukaryocytes. Puro, Puromycin resistance gene. **(b)** Schematic of genome editing with epiCRISPR system. The epiCRISPR vector is transfected into hPSCs followed by drug selection. Only the transfected cells can survive and proliferate. The epiCRISPR vector can replicate in hPSCs and can be partitioned to *daughter cells*. In the absence of drug selection, the epiCRISPR vector can be lost, allowing edited cells free of exogenous gene expression. **(c)** The epiCRISPR system for stable gene expression in hPSCs. The left panel shows the epiCRISPR vector delivered into hPSCs with lipid-mediated transfection. The middle panel shows the epiCRISPR vector can replicate during hPSC proliferation under drug selection. The right panel shows the single cell-derived clones lost epiCRISPR vector without drug selection. **(d)** Episomal vector was decreased within cells over time after withdrawing puromycin selection. The plasmid was tested every two days by quantitative PCR ($n = 3$, error bars show mean \pm S.D.). **(e)** The absence of episomal vector in single cell-derived hESC clones was confirmed by PCR targeting ampicillin sequence. Ctr1 is the positive control amplified from the epiCRISPR plasmid DNA; Ctr2 is the positive control amplified from the pooled hESCs containing epiCRISPR vector.

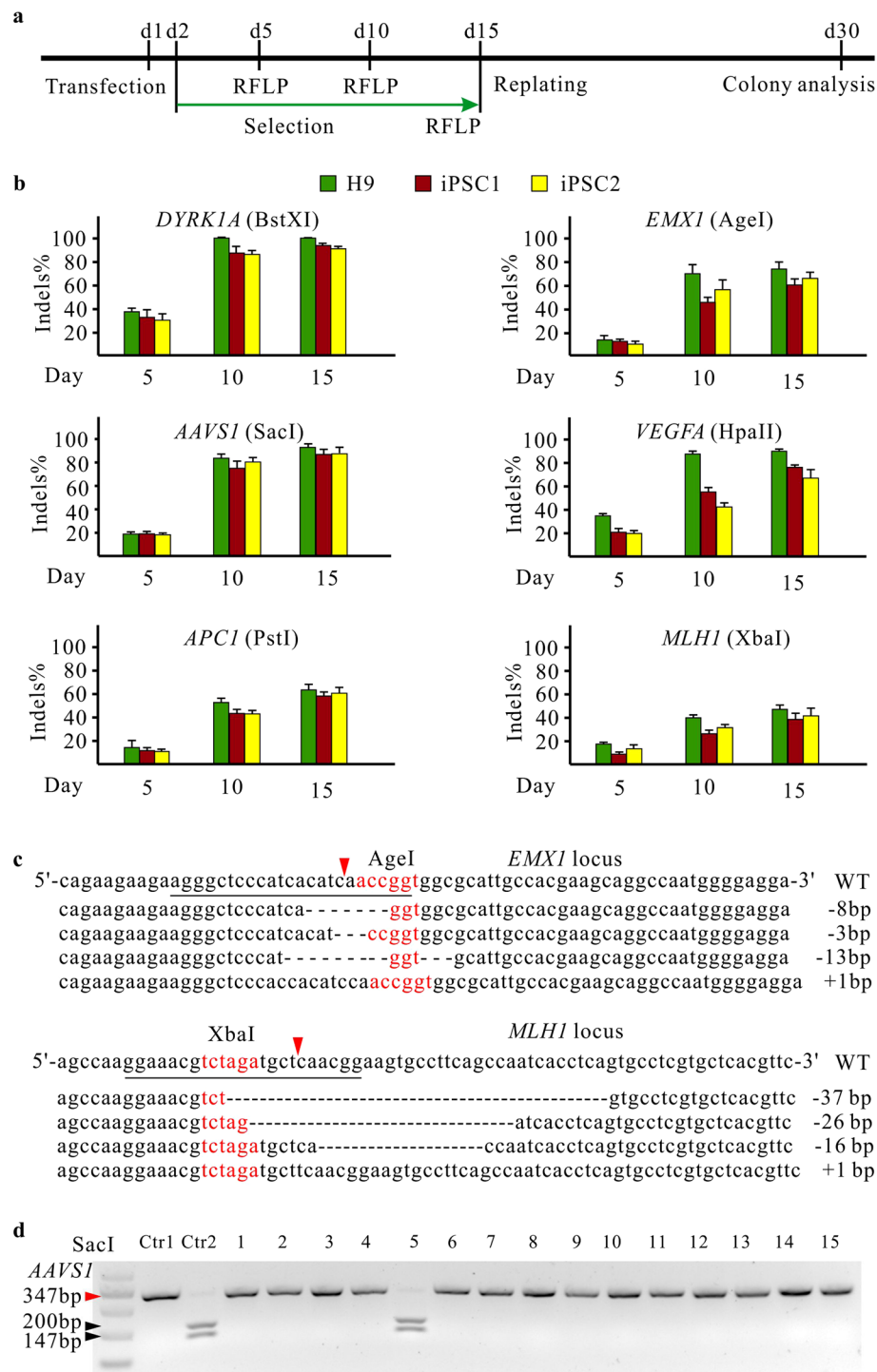


Figure 2. The epiCRISPR system for efficient gene knockout. **(a)** Schematic representation of the experimental procedure. **(b)** RFLP analysis of the indel rates generated by the epiCRISPR system at *DYRK1A*, *EMX1*, *AAVSI*, *VEGFA*, *APC1* and *MLH1* loci in hPSCs ($n = 3$, error bars show mean \pm S.D.). **(c)** Representative indel sequences at *EMX1* and *MLH1* sites in hESCs. The AgeI and XbaI restriction sequences are shown in red. The red triangles indicate Cas9 cutting site. **(d)** RFLP analysis of single hESC-derived clones modified at *AAVSI* locus. Ctr1 is the PCR band from unmodified cells without digestion. Ctr2 is the PCR band from unmodified cells with SacI digestion. The red triangle indicates the epiCRISPR-modified PCR bands; the black triangle indicates unmodified PCR bands.

for 319-bp deletion in hESCs. Of 16 clones analyzed, nine clones were monoallelic deletion and one clone was biallelic deletion (Fig. 4a). To determine if the epiCRISPR system can delete larger genomic DNA fragment, we designed paired gRNAs to delete the full coding sequence of *VEGFA* (16082bp) in hESCs. Of 20 clones analyzed,

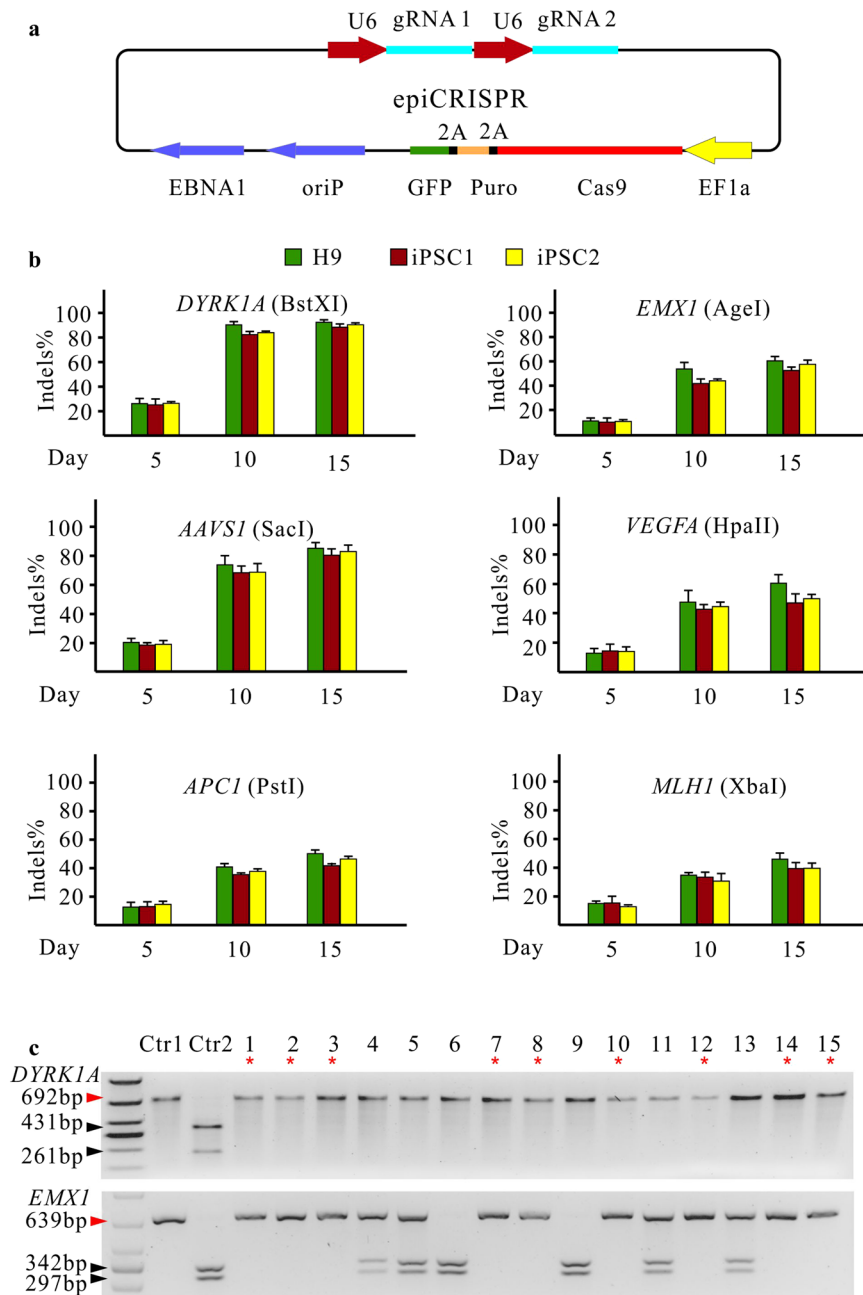


Figure 3. The epiCRISPR system for efficient double-gene knockout. **(a)** Expression of two gRNAs on the epiCRISPR vector for double-gene knockout. **(b)** RFLP analysis of the indel rates generated by the epiCRISPR system with gRNA multiplexed targeting *DYRK1A* & *EMX1*, *AAVS1* & *VEGFA* and *APC1* & *MLH1* in hPSCs ($n = 3$, error bars show mean \pm S.D.). **(c)** RFLP analysis of single cell-derived clones multiplexed targeting *DYRK1A* and *EMX1* loci. Ctr1 is the PCR band from unmodified cells without digestion. Ctr2 is the PCR band from unmodified cells with digestion (BstXI for *DYRK1A* and AgeI for *EMX1*). Red triangles indicate the epiCRISPR-modified PCR bands; black triangles indicate unmodified PCR bands; the red asterisks indicate the homozygous knockout for both genes.

10 clones were monoallelic deletion and two clones were biallelic deletion (Fig. 4b). Therefore, the epiCRISPR system enables efficient generation of homozygous large genomic deletions in hPSCs.

Off-target analysis of the epiCRISPR system in hPSCs. In previous reports, off-target mutations could be induced during genome editing^{27,28}. The epiCRISPR system requires long-term editing which may increase off-target effects. We analyzed a panel of potential off-target sites by using targeted deep sequencing for the locus of the *AAVS1*, *MLH1* and *EMX1* sites. These potential off-target sites were identified using the online tool (<http://crispr.mit.edu/>) that developed by Feng Zhang lab. Although we have edited cells for 15 days, the off-target indel rates ($0.075 \pm 0.096\%$) were at the background level ($0.049 \pm 0.038\%$; $P = 0.298$, Fig. 5a). Recently,

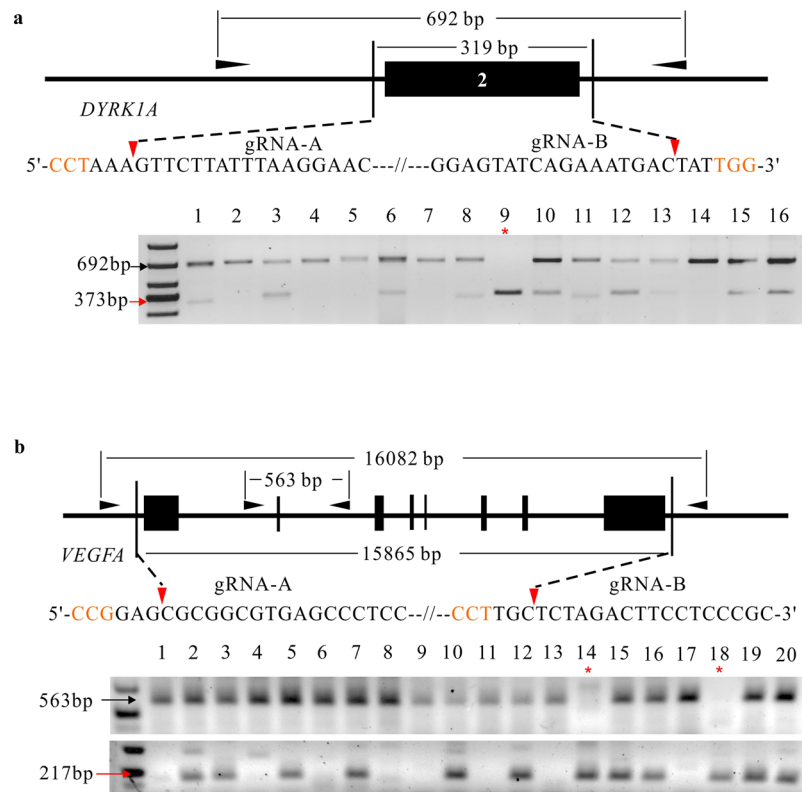


Figure 4. The epiCRISPR system for efficient genomic deletions. **(a)** The strategy for deletion of the 2nd exon of *DYRK1A*. A 373-bp PCR band will be present when deletion occurs (indicate by red arrow). A representative gel picture was shown. **(b)** The strategy for deletion of the full-coding sequence of *VEGFA*. Two pairs of primers were designed. One pair is to detect the non-deletion allele (563-bp) and one pair is to detect deletion events (217-bp, indicate by red arrow). The black triangles indicate the primers for the genotyping PCR reaction. The outer primers work only when deletion occurs. The PAM sequence is shown in orange; the red triangles indicate Cas9 cutting site; red arrows indicate the epiCRISPR-modified PCR bands; black arrows indicate unmodified PCR bands; the red asterisks indicate the homozygous knockout for both genes.

Slymaker *et al.* developed a new version of SpCas9 (eSpCas9) with enhanced specificity²⁹. We replaced wild-type SpCas9 with eSpCas9 and got epiCRISPR vector. The eSpCas9 showed lower on-target cleavage activity compared to the wild-type SpCas9 as previously reported (Fig. 5b)²⁹. Similar to the SpCas9, the off-target indel rates ($0.084 \pm 0.096\%$) of eSpCas9 were at the background level ($0.057 \pm 0.047\%$; $P = 0.285$, Fig. 5a). Notably, OT-6 locus for the *AAVS1* showed higher indel rates for both SpCas9 (0.43%) and eSpCas9 (0.43%). These indels are probably due to sequencing/PCR error because eSpCas9 could not tolerate three mismatches²⁹. Our study did not show the advantage of the eSpCas9 because the on-target sequences are very specific in the genome. The eSpCas9 could be useful for the locus that has high sequence similarity in the genome.

In addition to high-fidelity Cas9, we combined the episomal vector with double-nicking strategy. We named this system epiCRISPRn. The double-nicking strategy was originally developed by Ran *et al.*²⁷. In this strategy, the D10A mutant Cas9 nickase (Cas9n) is specified by a pair of gRNAs to simultaneously introduce single-stranded nicks on both strands of the target DNA, significantly increasing the specificity of genome editing²⁷. The epiCRISPRn was created by introducing a point mutation (D10A) into the Cas9 on epiCRISPR vector (Fig. 6a). A panel of paired gRNAs was designed to target four loci (*DYRK1A*, *EMX1*, *APC1* and *MLH1*). Fifteen days after transfection, RFLP assay revealed that the indel rates were 33–80% (Fig. 6b; Supplementary Figure S5 and Table S1). The sequencing results revealed much higher indel rates (17/20 = 85% for *APC1* locus; 16/20 = 80% for *MLH1* locus), because the indels shifted between two cutting sites and could not efficiently disrupt the restriction site (Fig. 6c). The majority of the modifications are small deletions (average 29.8 bp, range: 6–78 bp). In summary, the epiCRISPRn system enables efficient genome editing with double-nicking strategy in hPSCs.

The epiCRISPR system does not influence pluripotency and karyotype of hPSCs. Further analysis revealed that the modified hESCs displayed normal morphology and karyotype (Supplementary Figure S6a and b). Modified hESCs maintained their pluripotent state as indicated by the expression of pluripotency markers (Supplementary Figure S7a). In addition, these cells were capable of differentiation into all three germ layers both *in vivo* and *in vitro* (Supplementary Figure S7b and c).

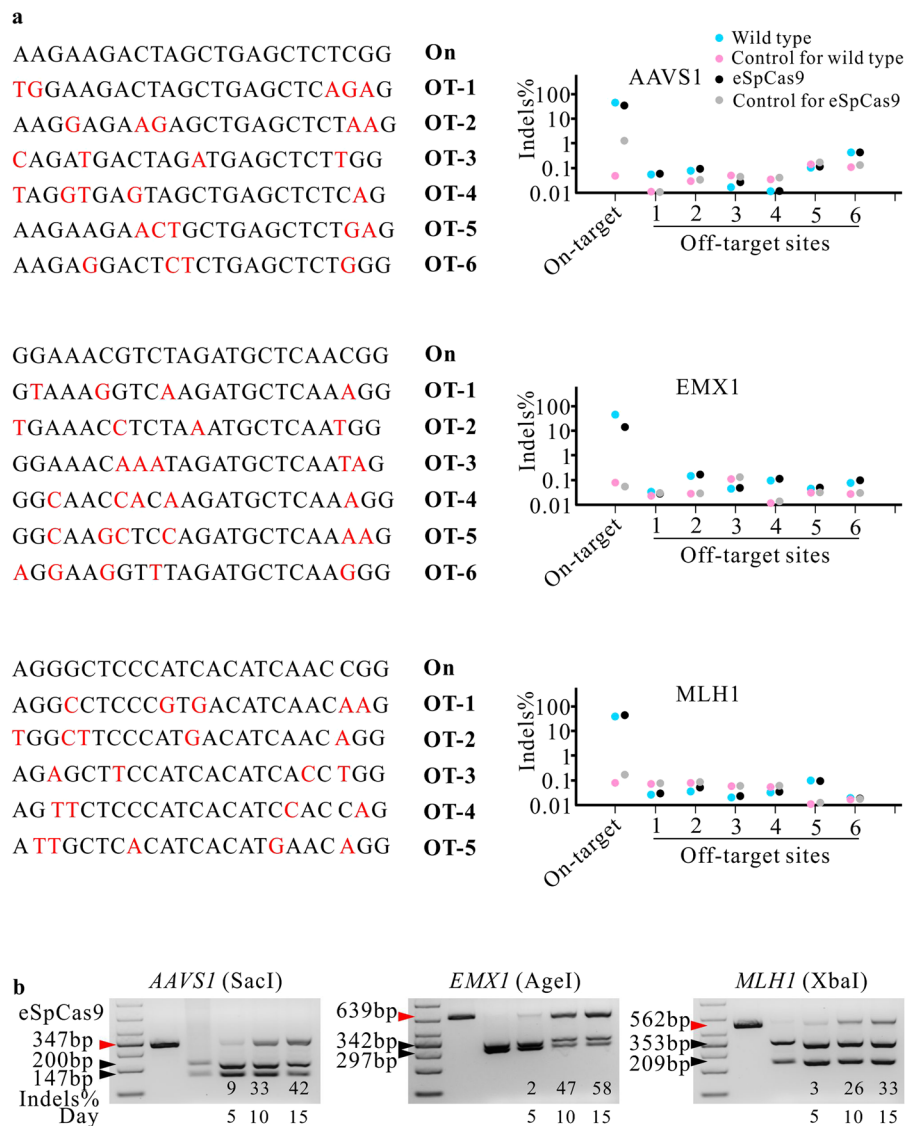


Figure 5. Analysis of off-target cleavage by targeted deep-sequencing. **(a)** The potential off-target sequences are shown on the left; the indel rates are shown on the right. On, on target sequence; OT, off-target sequence. The unmatched nucleotides are shown in red. **(b)** RFLP analysis of eSpCas9 cleavage at *AAVS1*, *EMX1* and *MLH1* sites. Red triangles indicate the epiCRISPR-modified PCR bands; black triangles indicate the unmodified PCR bands.

Discussion

In the present study, we have developed an epiCRISPR system for highly efficient genome editing in hPSCs. Through extension of Cas9 and gRNA expression using an episomal vector, we have achieved up to 100% indel rates. We have further shown that the epiCRISPR system supports efficient double-gene knockout and large genomic deletions. To minimize the off-target cleavage, we have also combined episomal vector technology with high-fidelity Cas9 and double-nicking strategy. Importantly, the episomal vector can be removed after genome editing, allowing edited cells free of exogenous gene expression. Our results demonstrate that the epiCRISPR system is a powerful tool for the gene knockout in the hPSCs.

Loss-of-function studies in hPSCs require efficient genome editing to disrupt genes of interest. Several groups have shown that 10–88% indel rates could be achieved by enriching transfected hPSCs with cell sorting or drug selection^{17–21}. In this study, we have developed an alternative strategy for efficient genome editing. The epiCRISPR system promotes genome editing by two ways. First the episomal vector allows continuously selecting transfected cells. Several successfully transfected cells are enough for the efficient genome editing. Since the transfection efficiency is not an obstacle for the editing, the epiCRISPR system allows lipid-mediated transfection which is convenient and less cytotoxic for the hPSCs⁹. Second, the episomal vector allows long-term genome editing. Our results show that the editing efficiency increases over time. In theory, the majority of gRNAs can generate indel rates up to ~100% if the editing time is long enough. In this study we have achieved 82–100% indel rates with 15 days' editing. Notably, not all of the indel mutations could fully disrupt gene functions. The epiCRISPR system

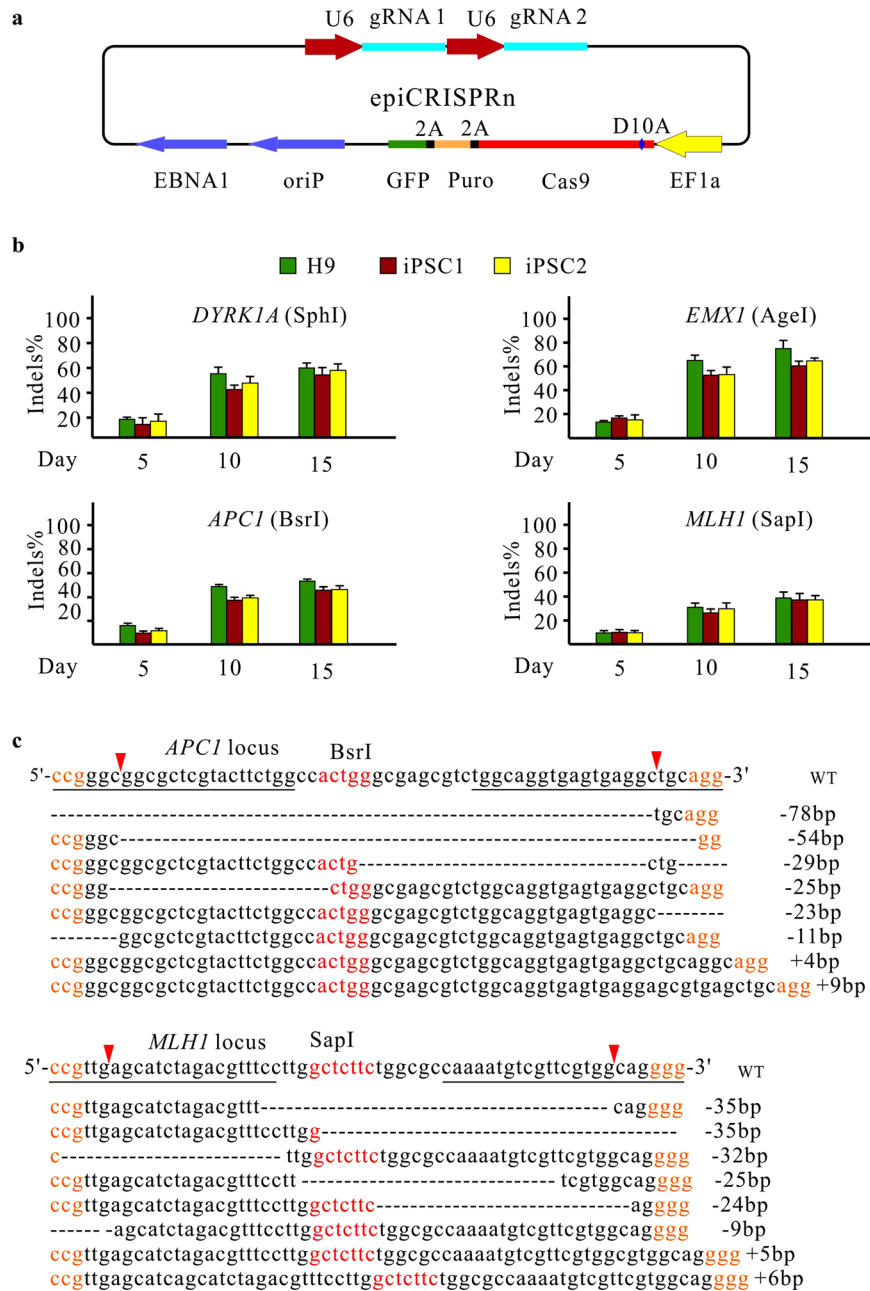


Figure 6. An epiCRISPRn system for efficient double-nicking. **(a)** Schematic of the epiCRISPRn vector design. A D10A mutation was introduced into the epiCRISPR vector. **(b)** RFLP analysis of the indel rates generated by the epiCRISPRn system targeting *DYRK1A*, *EMX1*, *AAVS1*, *VEGFA*, *APC1* and *MLH1* loci ($n = 3$, error bars show mean \pm S.D.). **(c)** Representative indel sequences generated by the epiCRISPRn at *APC1* and *MLH1* loci. The size of indels is shown on the right. The gRNA sequence is underlined; the PAM sequence is shown in orange; the restriction sites are shown in red, the red triangles indicate Cas9 cutting site.

also enables to generate double-gene knockout cells. Although we did not try in this study, the epiCRISPR system is likely to support triple- or more gene knockout in one-step. This powerful tool would facilitate to investigate gene-gene interactions in the hPSCs.

In addition to indels, the epiCRISPR system also supports large genomic deletions in hPSCs. Deletions have potential advantages as compared to indels given the predictability of loss-of-function and utility for the study of non-coding elements. In this study, we have showed two examples of genomic deletions. The biallelic deletions could be isolated by analysis of 20 clones. The genomic deletion efficiency depends on individual gRNA activity. We did not optimize the gRNA activity in this study. Recent genome-wide studies have shown that majority of the human genome is comprised of potential functional noncoding elements, such as intergenic or intronic cis-regulatory modules, miRNA clusters, or lincRNAs^{30,31}. The epiCRISPR system facilitates to study noncoding elements in hPSCs.

Several investigations have shown that CRISPR/Cas9 can generate off-target cleavage in somatic cells^{10, 27–29}. These investigations have largely focused on sites with high sequence similarity to the on-target site and have documented indel rates as high as 50% at individual off-target sites. In contrast, two groups have investigated off-target effects at sites with low sequence similarity to the on-target site in hPSCs^{32, 33}. They analyzed several CRISPR/Cas9-edited single cell-derived hPSC clones by whole-genome sequencing, but they did not identify any off-target cleavage, demonstrating the specificity of the CRISPR/Cas9 in the hPSCs. The on-target sites we chose in this study have no off-target sites with high sequence similarity. Consistent with previous studies, we did not observe off-target cleavage at 18 sites for both wild-type and high-fidelity Cas9, although we have expressed Cas9 and gRNAs for 15 days. CRISPR/Cas9 may generate indels at off-target site with high sequence similarity to the on-target site. In this case, the high-fidelity Cas9 or double-nicking strategy could be used to reduce off-target cleavage. When we were preparing the manuscript, another group reported that the episomal CRISPR/Cas9 system could work efficiently in HeLa cells and mouse iPSCs²³. Taken together, the epiCRISPR system offers a powerful platform for highly efficient genome editing in a variety of cells.

Materials and Methods

Cell culture and maintenance of human pluripotent stem cells. Human ESCs (WA09, Wicell, Madison, WI) and iPSCs were cultured on Matrigel-coated plates (ESC qualified, BD Biosciences, San Diego, CA) using hESC mTeSR-1 cell culture medium (StemCell Technologies, Vancouver, Canada) under conditions of 37 °C, 95% air, and 5% CO₂ in a humidified incubator, as previously described⁸. Results for subsequent experiments are based on 1 hESC line (WA09) and 2 iPSC lines.

Plasmids and oligonucleotides. The epiCRISPR plasmid: the SpCas9, puromycin resistance gene and copGFP were co-expressed from an EF1a promoter as a single protein separated by self-cleaving P2A peptides; the OriP/EBNA1 components driving plasmid duplication were derived from pREP9 vector (Invitrogen); gRNA was expressed from a human U6 promoter. The epiCRISPR plasmid is the same as epiCRISPR except that three mutations (K848A/K1003A/R1060A) were introduced into the SpCas9, and the human U6 promoter was replaced by mouse U6 promoter so that it can express gRNA starting with both G and A. To express a gRNA, the oligonucleotide duplexes were cloned into BspQI restriction sites of the epiCRISPR plasmid. The epiCRISPRn plasmid is the same as epiCRISPR except that a D10A mutation was introduced into the SpCas9. A detailed protocol for cloning two gRNAs into vector was provided in Supplementary Figure S8. The amino acid substitutions were generated by standard PCR. DNA sequence of plasmids used in this study can be found in the Supplementary Figure S9. gRNA target sites are available in Supplementary Table 2, and oligonucleotides used in this study can be found in Supplementary Table 3.

Genome editing with epiCRISPR system. 4×10^5 human ESCs or iPSCs were disassociated and cultured in a 6-well plate (Sigma Aldrich, St. Louis, MO) with mTeSR-1 cell culture medium for 24 hours. At day 1, two μ g of the epiCRISPR plasmid were transfected into cells by using lipofectamine 3000 (Life Technologies) based on online protocol. At day 2, cells were selected by puromycin (0.2–0.5 μ g/ml). At day 5, day 10 and day 15, genomic DNA was extracted from cells using QuickExtract (Epicentre) following the manufacturer's protocol. The gRNA targeting sites were PCR-amplified using Q5 High-Fidelity DNA polymerase (NEB). The PCR products were purified using QiaQuick Spin Column (QIAGEN) following the manufacturer's protocol. The purified PCR products were digested with restriction enzymes (NEB) for 4 hours and then run them on agarose gel. For analysis of single cell-derived clones, the cells were disassociated into single cells at day 15 post-transfection, and seeded them onto the Matrigel-coated plates with puromycin-free mTeSR-1 medium for 15 days. Individual colonies were picked and genotyped.

Off-target cleavage analysis. Fifteen days after genome editing, genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Epicenter) following the manufacturer's protocol. Potential off-target sites predicted by an online tool (<http://crispr.mit.edu/>) were amplified using Q5 High-Fidelity DNA polymerase (NEB) with the primers listed in Supplementary Table 3. The genome sequence spanned by the corresponding primer pairs were extracted from human genome sequence (Hg19, GRCh37), acting as the reference sequence for analyzing sequence variation caused by genome editing. PCR products were generated for each on- and off-target site from ~100 ng of genomic DNA extracted from hESCs. PCR products were purified with QIAquick Gel Extraction Kit, normalized in concentration, pooled together, phosphorylated at 5' end, added dA-Tailing and Y-Shape adapter. The resultant products were PCR-amplified with KAPA HIFI Hotstart Readymix (KAPA Biosystems) and primers (100 nM) carrying Illumina sequence adaptors. PCR products were purified with QIAquick Gel Extraction Kit (QIAGEN), and sequenced via 150-bp paired-end sequencing on an Illumina MiSeq instrument. To avoid the possible errors in the sequencing process, each read was first filtered by removing the nucleotides with a quality score smaller than 30, and the following nucleotides were also removed. The resulted high-quality reads were mapped to the reference human genome sequence using pairwise sequence alignment algorithm, BLAT (<http://genome.ucsc.edu>)³⁴, with default parameters. The number of reads with indels to the reference sequence that were 15-bp adjacent to the predicted cleavage site were counted, and divided by the total aligned reads to get the potential editing percentage. Since most of indels were less than 50-bp, an alternative 30-bp DNA fragment 50-bp away from predicted cleavage site was used as control, and the number of reads with indels was also counted to get the corresponding sequence variation percentage.

Statistical analysis. All the data are shown as the mean \pm S.D. Statistical analyses were conducted using Microsoft Excel. Two-tailed, paired Student's *t*-tests were used to determine statistical significance when comparing two groups. A value of $p < 0.05$ was considered to be statistically significant.

References

- Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147, doi:10.1126/science.282.5391.1145 (1998).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872, doi:S0092-8674(07)01471-7 [pii], doi:10.1016/j.cell.2007.11.019 (2007).
- Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826, doi:10.1126/science.1232033 (2013).
- Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823, doi:10.1126/science.1231143 (2013).
- Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096–1258096, doi:10.1126/science.1258096 (2014).
- Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y. L. & Weiss, D. S. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. (vol 497, pg 254, 2013) *Nature* **501**, 262–262, doi:10.1038/Nature12498 (2013).
- Wang, Y. *et al.* Genome editing of human embryonic stem cells and induced pluripotent stem cells with zinc finger nucleases for cellular imaging. *Circ Res* **111**, 1494–1503, doi:10.1161/CIRCRESAHA.112.274969 (2012).
- Wang, Y. *et al.* Genome editing of isogenic human induced pluripotent stem cells recapitulates long QT phenotype for drug testing. *J Am Coll Cardiol* **64**, 451–459, doi:10.1016/j.jacc.2014.04.057 (2014).
- Gonzalez, F. *et al.* An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell stem cell* **15**, 215–226, doi:10.1016/j.stem.2014.05.018 (2014).
- Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495, doi:10.1038/nature16526 (2016).
- Byrne, S. M., Mali, P. & Church, G. M. Genome editing in human stem cells. *Methods Enzymol* **546**, 119–138, doi:10.1016/B978-0-12-801185-0.00006-4 (2014).
- Ding, Q. *et al.* A TALEN genome-editing system for generating human stem cell-based disease models. *Cell stem cell* **12**, 238–251, doi:10.1016/j.stem.2012.11.011 (2013).
- Tan, E. P., Li, Y., Velasco-Herrera Mdel, C., Yusa, K. & Bradley, A. Off-target assessment of CRISPR-Cas9 guiding RNAs in human iPS and mouse ES cells. *Genesis* **53**, 225–236, doi:10.1002/dvg.22835 (2015).
- Yang, L. *et al.* Optimization of scarless human stem cell genome editing. *Nucleic Acids Res* **41**, 9049–9061, doi:10.1093/nar/gkt555 (2013).
- Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827–832, doi:10.1038/nbt.2647 (2013).
- Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci USA* **110**, 15644–15649, doi:10.1073/pnas.1313587110 (2013).
- Ding, Q. *et al.* Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell stem cell* **12**, 393–394, doi:10.1016/j.stem.2013.03.006 (2013).
- Liao, J. *et al.* Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nature genetics* **47**, 469–478, doi:10.1038/ng.3258 (2015).
- Hendriks, W. T., Jiang, X., Daheron, L. & Cowan, C. A. TALEN- and CRISPR/Cas9-Mediated Gene Editing in Human Pluripotent Stem Cells Using Lipid-Based Transfection. *Current protocols in stem cell biology* **34**(5B), 3 1–25, doi:10.1002/9780470151808.sc05b03s34 (2015).
- Liu, Z. *et al.* Efficient CRISPR/Cas9-Mediated Versatile, Predictable, and Donor-Free Gene Knockout in Human Pluripotent Stem Cells. *Stem cell reports* **7**, 496–507, doi:10.1016/j.stemcr.2016.07.021 (2016).
- Horii, T., Tamura, D., Morita, S., Kimura, M. & Hatada, I. Generation of an ICF syndrome model by efficient genome editing of human induced pluripotent stem cells using the CRISPR system. *International journal of molecular sciences* **14**, 19774–19781, doi:10.3390/ijms141019774 (2013).
- Wells, M. F. *et al.* Genetic Ablation of AXL Does Not Protect Human Neural Progenitor Cells and Cerebral Organoids from Zika Virus Infection. *Cell stem cell* **19**, 703–708, doi:10.1016/j.stem.2016.11.011 (2016).
- Li, L., Gao, F. & Wu, S. An episomal CRISPR/Cas9 system to derive vector-free gene modified mammalian cells. *Protein & cell* **7**, 689–691, doi:10.1007/s13238-016-0299-9 (2016).
- Van Craenenbroeck, K., Vanhoenacker, P. & Haegeman, G. Episomal vectors for gene expression in mammalian cells. *Eur J Biochem* **267**, 5665–5678, doi:10.1046/j.1432-1327.2000.01645.x (2000).
- Nambo, A., Sugden, A. & Sugden, B. The coupling of synthesis and partitioning of EBV's plasmid replicon is revealed in live cells. *EMBO J* **26**, 4252–4262, doi:10.1038/sj.emboj.7601853 (2007).
- Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281–2308, doi:10.1038/nprot.2013.143 (2013).
- Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389, doi:10.1016/j.cell.2013.08.021 (2013).
- Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833–838, doi:10.1038/nbt.2675 (2013).
- Slyamaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88, doi:10.1126/science.aad5227 (2016).
- Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nature genetics* **47**, 199–208, doi:10.1038/ng.3192 (2015).
- Pelechano, V. & Steinmetz, L. M. Gene regulation by antisense transcription. *Nat Rev Genet* **14**, 880–893, doi:10.1038/nrg3594 (2013).
- Veres, A. *et al.* 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, doi:10.1016/j.stem.2014.04.020 (2014).
- Smith, C. *et al.* Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell stem cell* **15**, 12–13, doi:10.1016/j.stem.2014.06.011 (2014).
- Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res* **12**, 656–664, doi:10.1101/gr.229202 (2002). Article published online before March 2002.

Acknowledgements

We thank Professor L. Jin for helpful comments on the manuscript. This work was supported by grants from the National Natural Science Foundation of China (81370302), the National Basic Research Program of China (2015CB943300), the Foundation for Innovative Research Group of the National Natural Science Foundation of China (31521003), the Open Research Funds of the State Key Laboratory of Genetic Engineering Fudan University SKLGE-1407, the Doctoral Fund of Ministry of Education of China (KPH1322054) and Thousand Youth Talents Plan Project of Thousand Youth Talents to YWang; the 973 Program (2013CB945403), the National Natural Science Foundation of China (81430005, 81270232, 31521003), and the Commission for Science and Technology of Shanghai Municipality (13JC1407600) to Hwang. Funding for open access charge: the National Natural Science Foundation of China (81370302).

Author Contributions

Y.X. and D.W. performed genome editing experiments; F.L. performed pluripotency analysis; G.W. analyzed deep sequencing data; R.C. and S.H. revised manuscript; D.L., T.N., M.L., D.L., H.W. and Y.W. provide experimental guidance and wrote the paper.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-02456-y](https://doi.org/10.1038/s41598-017-02456-y)

Competing Interests: The authors declare that they have no competing interests.

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



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017