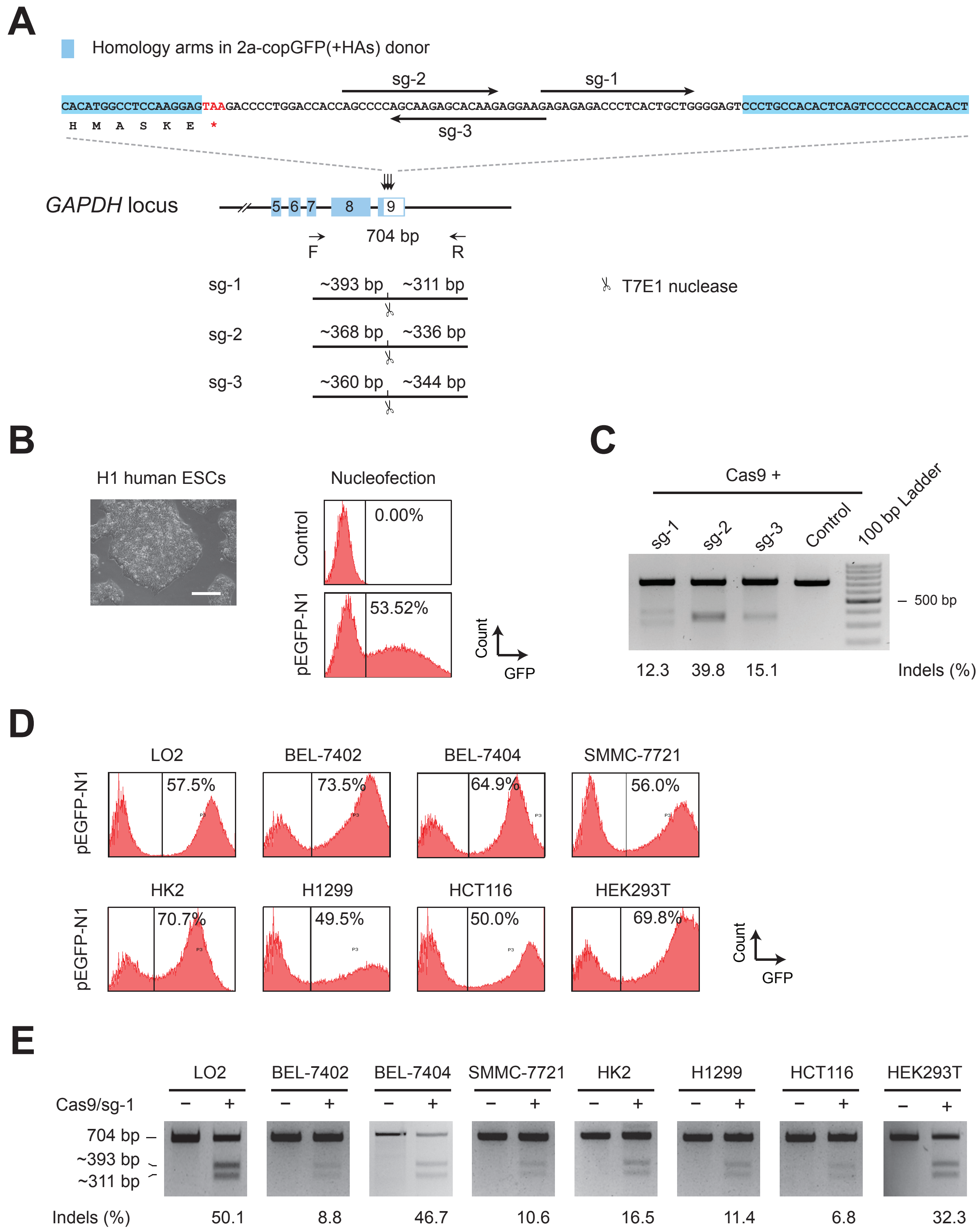
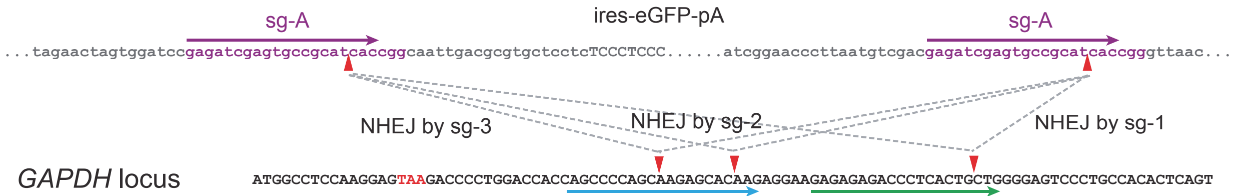
DOI: 10.1093/nar/gkw064



Supplementary Figure S1. Transfection efficiencies and T7E1 assay results in human ESCs and somatic cell lines. (**A**) Schematics of sg-1–3 target sites at the *GAPDH* 3'-UTR, positions for PCR primers used for T7E1 assay and predicted cutting patterns when different sgRNAs were used. (**B**) Bright field image of H1 human ESCs cultured on Matrigel in mTeSR1 medium (left panel); and FACS results at 72 hrs after transfection with plasmid pEGFP-N1 (right panel). (**C**) T7E1 assays showing the different activities of Cas9/sg-1, 2 or 3 in targeting human ESC genome. H1 human ESCs were transfected with Cas9/sg-1, 2 or 3, without any donor. Genome DNAs were extracted for T7E1 assays at day 3 after transfection. (**D**) FACS results showing transfection efficiencies in different human somatic cell lines. Cells were transfected with pEGFP-N1 plasmid, and analyzed at 48 hrs after transfection. (**E**) T7E1 assay results showing the different indel frequencies induced by Cas9/sg-1, suggesting intrinsic variance of its genome-targeting activities in different human somatic cell lines. Starting cells of each cell line were transfected with Cas9/sg-1 in the absence of any donor. Genome DNAs of these cells were extracted at day 2 after transfection for T7E1 assays.

A

Double-cut NH-donor



B

5'-Junction

sg-1

```
                                                                      PAM
#1  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#2  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#3  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAA------------------------CAATTGACGCGTGCTCCTCTCCCTCCC -15
#4  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#5  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#6  AAGGAGTAAGACCCCTGGACCCCCAGCCCCAGCAAGAGCCCAAGAGGAAGAGAGAGACCCTCAATGGGAAGCGGTCAGCCCATTCGCCGCC
                 AAGCTTTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGATAGCACCGGCAATTGACGCGTGCTCCTCTCCCTCCC +71
#7  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#8  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#9  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAA------------------------CAATTGACGCGTGCTCCTCTCCCTCCC -15
#10 AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
```

sg-2

```
                                                                           PAM
#1  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCAC                              CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#2  AAAGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACCACCGGGTTAAC                  CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC +12
#3  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCAC                              CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#4  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCAC                              CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#5  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCAC                              CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#6  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCAC                              CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#7  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAA                      ------CAATTGACGCGTGCTCCTCTCCCTCCC +8,-6
#8  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAG---                              ------CAATTGACGCGTGCTCCTCTCCCTCCC -9
#9  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGC--TCTTCCGCATTTGACGGAAGAGAGAGAC--CCGGCAATTGACGCGTGCTCCTCTCCCTCCC +28,-4
#10 AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCAC                              CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#11 AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACTCGAG                         CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC +5
```

sg-3

```
                                    PAM     PAM
#1  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGC CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#2  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGC CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#3  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGC CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#4  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGCCC--------TTGACGCGTGCTCCTCTCCCTCCC +1,-8
#5  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGC CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#6  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGC CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#7  AAGGAGTAAGACCCCTGGA-------------C CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC -13
#8  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGC CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#9  AAGGAGTAAGACCCCTGGACCACCAGCCCCAGC CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#10 AAGGAGTAAGACCCCTGGACCACCAGCCCCAGC CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC 0
#11 AAGGAGTAAGACCCCTGGA-------------C CACCGGCAATTGACGCGTGCTCCTCTCCCTCCC -13
```

3'-Junction

sg-1

```
                                                    PAM
#1  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  CTGGGGAGTCCCTGCCACACTAAGT 0
#2  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  CTGGGGAGTCCCTGCCACACTAAGT 0
#3  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  CTGGGGAGTCCCTGCCACACTAAGT 0
#4  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  CTGGGGAGTCCCTGCCACACTCAGT 0
#5  ATCGGAACCCTTAATGTCGAC-----------------  ---------CCCTGCCACACTCAGT -26
#6  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  CTGGGGAGTCCCTGCCACACTAAGT 0
#7  ATCGGAACCCTTAATGTCGAC-----------------  ---------CCCTGCCACACTCAGT -26
#8  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  CTGGGGAGTCCCTGCCACACTAAGT 0
#9  ATCGGAACCCTTAATGTCGAC-----------------  ---------CCCTGCCACACTCAGT -26
#10 ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  CTGGGGAGTCCCTGCCACACTCAGT 0
#11 ATCGGAACCCTTAATGTCGAC-----------------  ---------CCCTGCCACACTCAGT -26
#12 ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCATTA----------------------AGT +2,-22
#13 ATCGGAACCCTTAATGTCGAC-----------------  ---------CCCTGCCACACTCAGT -26
#14 ATCGGAACCCTTAATGTCGAC-----------------  ---------CCCTGCCACACTCAGT -26
#15 ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  CTGGGGAGTCCCTGCCACACTCAGT 0
#16 ATCGGAACCCTTAATGTCGAC-----------------  ---------CCCTGCCACACTCAGT -26
#17 ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  CTGGGGAGTCCCTGCCACACTCAGT 0
#18 ATCGGAACCCTTAATGTCGAC-----------------  ---------CCCTGCCACACTCAGT -26
```

sg-2

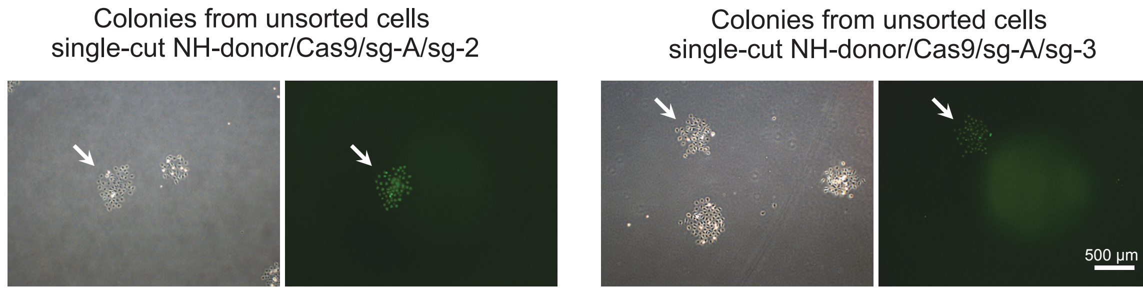
```
                                                    PAM
#1  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  AAGAGGAAGAGAGAGACCCTCACTGCTGG 0
#2  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCATC AAGAGGAAGAGAGAGACCCTCACTGCTGG +1
#3  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCATC AAGAGGAAGAGAGAGACCCTCACTGCTGG +1
#4  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  --------GAGAGAGACCCTCACTGCTGG -8
#5  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  AAGAGGAAGAGAGAGACCCTCACTGCTGG 0
#6  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCA-  -AGAGGAAGAGAGAGACCCTCACTGCTGG -2
#7  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT  AAGAGGAAGAGAGAGACCCTCACTGCTGG 0
#8  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCATTCAAGAGGAAGAGAGAGACCCTCACTGCTGG +2
#9  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCATC AAGAGGAAGAGAGAGACCCTCACTGCTGG +1
#10 ATCGGAACCCTTAATGTCGACGAGATCGAGT-------C AAGAGGAAGAGAGAGACCCTCACTGCTGG +1,-7
```

sg-3

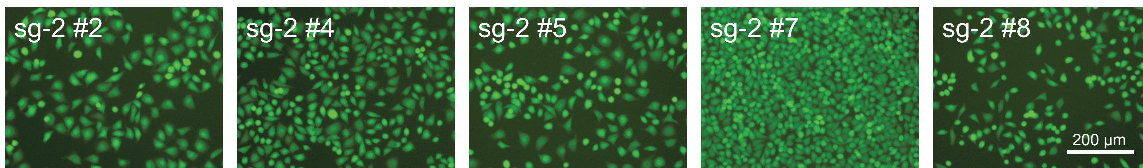
```
#1  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCA-   CAAGAGGAAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA +7,-1
#2  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA 0
#3  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGC--         -AGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA -3
#4  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA 0
#5  ATCGGAACCCTTAATGT---------------------TGTGATTAAT---------------------GAGACCCTCACTGCTGGGGAGTCCCTGCCA +10,-41
#6  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA 0
#7  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA 0
#8  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA 0
#9  ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         --GAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA -2
#10 ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA 0
#11 ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA 0
#12 ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA 0
#13 ATCGGAACCCTTAATG----------------------         ------------GGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA -34
#14 ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA 0
#15 ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA 0
#16 ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA 0
#17 ATCGGAACCCTTAAT-----------------------         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA -23
#18 ATCGGAACCCTTAATGTCGACGAGATCGAGTGCCGCAT         AAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCA 0
```

Supplementary Figure S2. Sequences of integration junctions in the GFP+ cells produced with double-cut NH-donor/Cas9/sg-A/sg-1, 2 or 3. (**A**) Schematics of sg-1–3 target sites at the *GAPDH* genomic locus, sg-A target sites in a double-cut NH-donor, and positions of cleavage and re-joining between the genome and the donors.
(**B**) Sequences of the integration junctions from the GFP+ cells produced with the double-cut NH-donors (in Figure 2D, right panel). The 5'- and 3'-junctions of sg-1, 2 or 3-induced integrations were analyzed separately. For each junction multiple sequences are shown. Nucleotides of different sgRNA target sites and PAMs are color-coded. Sequences from donor templates are shown in grey, and genomic DNA flanking the integration junctions are shown in black. These results confirmed the insertions of donor fragments at the expected sgRNA target sites and revealed the presence of indels at both the 5'- and the 3'-junctions.

A

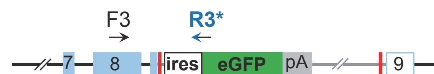


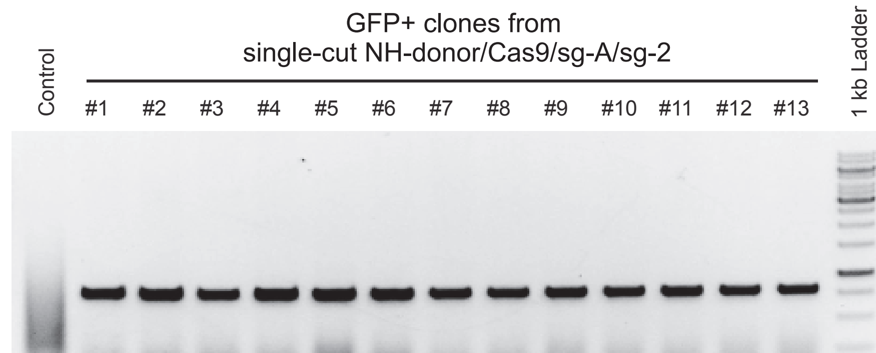
B



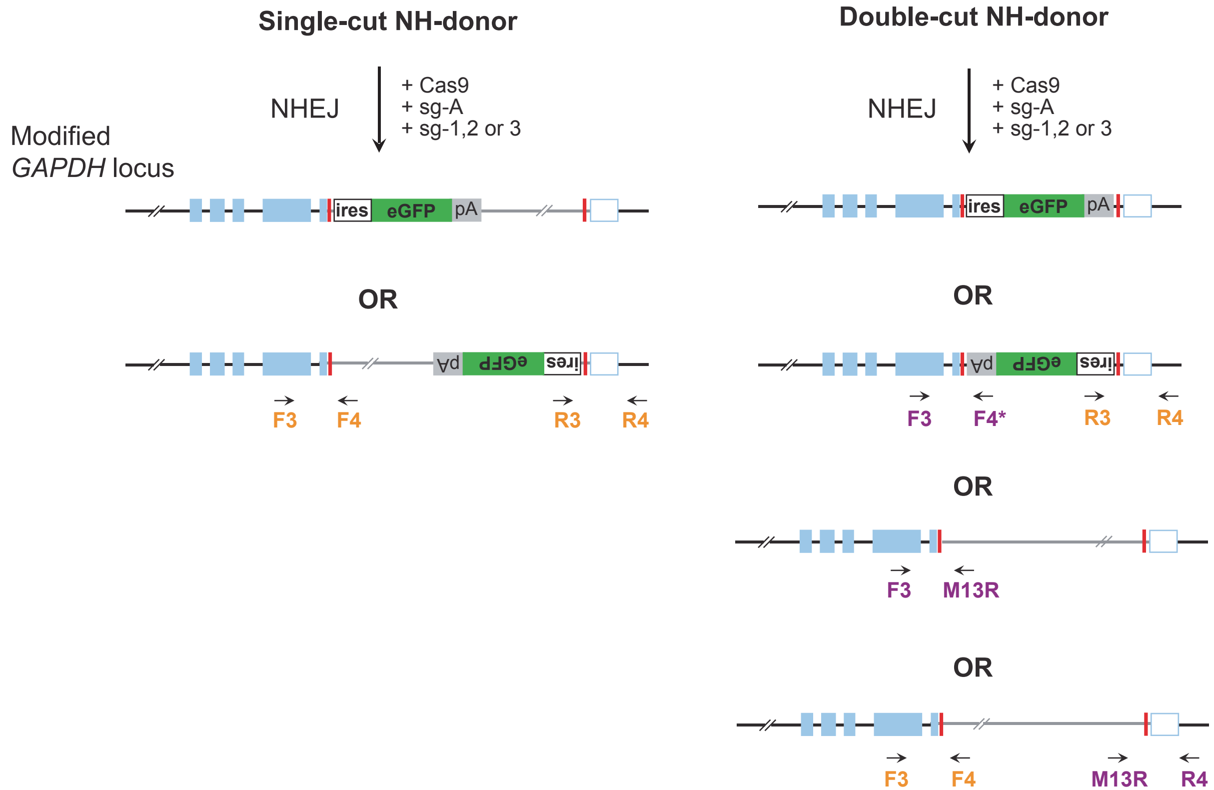
C

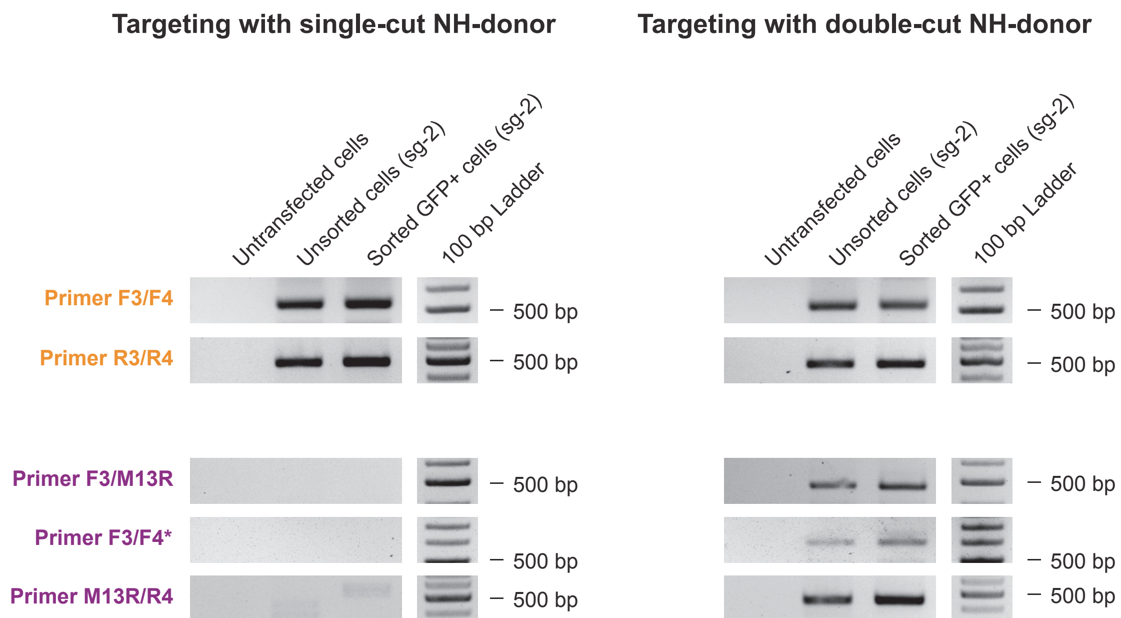
GAPDH locus modified
with single-cut NH-donor/Cas9/sg-A/sg-2





Supplementary Figure S3. Selection-free generation of knock-in clones via the homology-independent knock-in approach. (**A**) Bright field and fluorescence images of single-cell colonies, observed at 10 days after seeding the cells transfected with single-cut NH-donor/Cas9/sg-A/sg-2 or sg-3 (unsorted) at low density. GFP+ (white arrows) and GFP- colonies (unlabeled) were seen in both samples. (**B**) Fluorescence images showing GFP expression in selected individual clones, which were isolated from cells transfected with single-cut NH-donor/Cas9/sg-A/sg-2. (**C**) PCR of the 13 GFP+ clones isolated from the cells transfected with single-cut NH-donor/Cas9/sg-A/sg-2. Primers F3/R3* was used to detect the 5'-integration junction; and their positions are indicated in the schematics (upper panel). Primer R3* was used instead of primer R3 here to obtain optimal amplification. PCR amplifications showed DNA fragments at expected sizes, indicating correct integration of the ires-eGFP reporter at the *GAPDH* 3'-UTR.





Supplementary Figure S4. Genome PCR detected re-joining of the genomic and the donor DNAs in non-GFP-expressing integrations. (**A**) Schematics showing different types of integrations that may occur during NHEJ-mediated knock-in of the ires-eGFP reporter, either with a single-cut NH-donor (left panel) or with a double-cut NH-donor (right panel). (**B**) PCR of the integration junctions in non-GFP expressing integration events. The non-GFP expressing integrations were detected in the unsorted as well as the sorted GFP+ cells, which might indicate simultaneous integrations of two different inserts at the two genomic alleles.

A

GAPDH locus

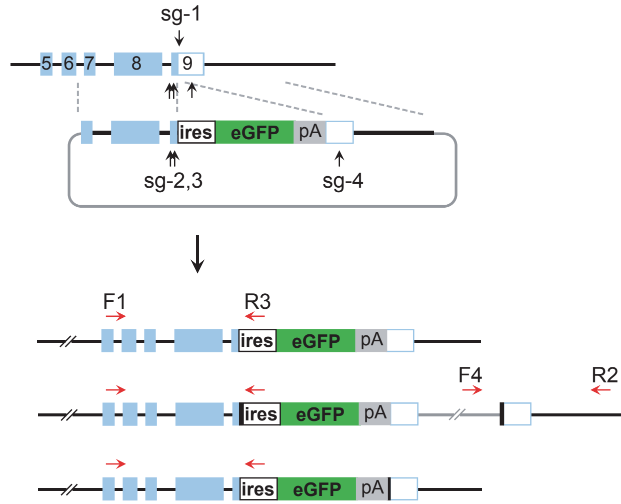
ires-eGFP(+HAs) Donor-1

Modified *GAPDH* locus generated with

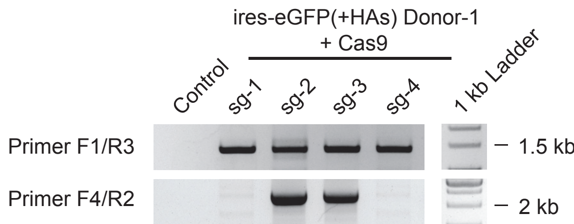
Cas9/sg-1 (HDR on both sides)

Cas9/sg-2 or sg-3 (NHEJ on both sides)

Cas9/sg-4 (5' HDR & 3' NHEJ)



B



C

5'-junctions of HDR-targeting induced with sg-1

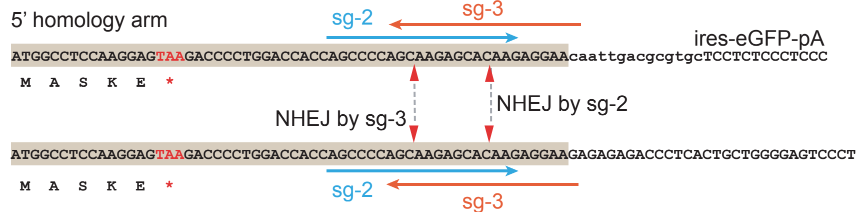
5' homology arm — ires-eGFP-pA

```
Ref seq  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAcaattgacgcgtgcTCCTCTCCCTCCC
sg-1 #1  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC 0
     #2  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC 0
     #3  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC 0
     #4  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC 0
     #5  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC 0
     #6  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC 0
```

D

ires-eGFP(+HAs) Donor-1

GAPDH locus



5'-junctions of NHEJ-targeting induced with sg-2 or sg-3

```
sg-2 #1  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCAC                    CAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC +1
     #2  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCAC                    CAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC +1
     #3  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCAC                    CAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC +1
     #4  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCAC                    CAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC +1
     #5  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCA-GAAGACAGCAAGAGACAGCAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC +19,-1
     #6  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCAC                   TT--GAGGAACAATTGACGCGTGCTCCTCTCCCTCCC +2,-2
     #7  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCAC                    CAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC +1
     #8  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCA-                     -AGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC -2

sg-3 #1  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGC AA  ----------------ATTGACGCGTGCTCCTCTCCCTCCC -16
     #2  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGC AAAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC +2
     #3  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCA-- AA  GAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC +2,-2
     #4  ATGGCCTCCAAGAAGTAAGACCCCTGGACCACCAGCCCCAGC AAA GAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC +1
     #5  ATGGCCTCCAAGGAGTAAGACCCCTGGAGCAC---------- --  ------AAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC -18
     #6  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCC----- --  ----ACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC -11
     #7  ATGGCCTCCAAGGAGTAAGA---------------------- --  --GCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC -26
     #8  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCCAA  GAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC +1
     #9  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGC AAA GAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC +1
```

E

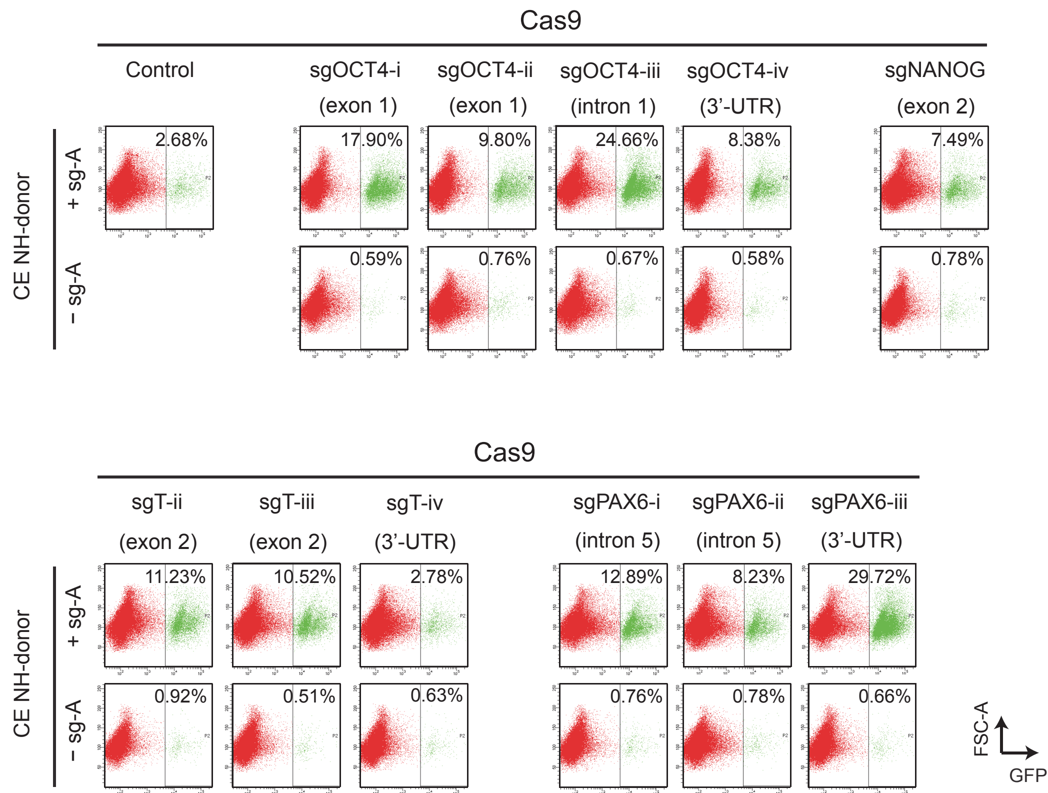
HDR on 5'-junctions in sg-4-induced targeting

5' homology arm — ires-eGFP-pA

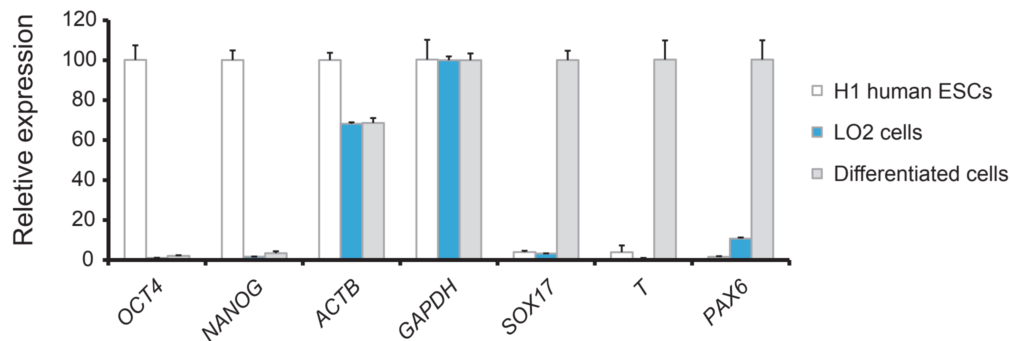
```
Ref seq  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAcaattgacgcgtgcTCCTCTCCCTCCC
sg-4 #1  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC 0
     #2  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC 0
     #3  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC 0
     #4  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC 0
     #5  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC 0
     #6  ATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAACAATTGACGCGTGCTCCTCTCCCTCCC 0
```

Supplementary Figure S5. Comparison of HDR- and NHEJ-mediated reporter insertions using the ires-eGFP (+HAs) donor-1. (**A**) Schematics showing the sg-1–4 target positions on the genomic *GAPDH* locus and the ires-eGFP(+HAs) donor-1, as well as the modified genomes that carry reporter insertions introduced by Cas9/sg-1, 2, 3, or 4. (**B**) PCR of integration junctions produced with ires-eGFP(+HAs) donor-1/Cas9/sg-1, 2, 3, or 4. PCR with primers F1/R3 detected the 5'-junctions present in all HDR and NHEJ-mediated reporter integrations; whereas amplification with primers F4/R2 detected the 3'-junctions unique to NHEJ-mediated integrations of the ires-eGFP(+HAs) donor-1. (**C**) Sequencing results of the 5'-juctions in the HDR-targeting induced with Cas9/sg-1. (**D**) Upper panel shows schematics of sg-2 and sg-3 target sites in genome and ires-eGFP(+HAs) donor-1, as well as the positions of cleavage and re-joining during the integrations induced with Cas9/sg-2 or sg-3 via NHEJ repair. Lower panel shows the sequencing results of the 5'-junctions. (**E**) Sequencing results in the 5'-junctions of the HDR-mediated integrations induced with Cas9/sg-4.

A

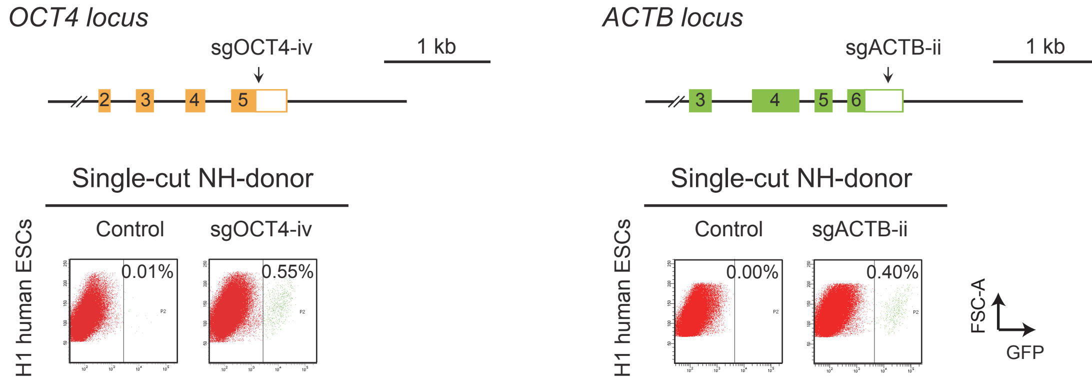


B

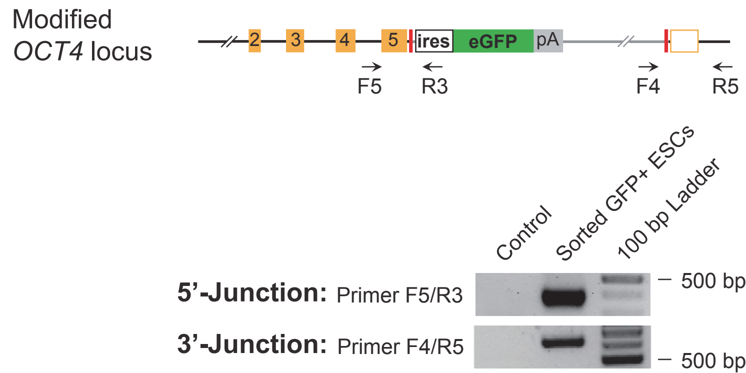


Supplementary Figure S6. NHEJ-mediated reporter knock-in at silenced genome loci. (**A**) FACS results showing NHEJ-mediated knock-in of PGK-eGFP reporter at various locations in the *OCT4*, *NANOG, T* and *PAX6* gene loci. The CE NH-donor was co-transfected with Cas9 and corresponding sgRNAs, with or without sg-A, into LO2 cells. Transfected cells were maintained for five passages before FACS analysis. GFP+ cells are gated to the right of the dashed line in each panel. (**B**) qRT-PCR analysis of the expression of *OCT4*, *NANOG, ACTB, GAPDH, SOX17, T* and *PAX6* genes in LO2 cells. H1 human ESCs which express *OCT4* and *NANOG,* and differentiated cells that expressed *SOX17, T* and *PAX6,* were included as references. Data shown are the mean ± s.d., n=3.

A



B



C

Supplementary Figure S7. CRISPR/Cas9-induced NHEJ-mediated knock-in of ires-eGFP reporter at *OCT4* and *ACTB* 3'-UTR in H1 human ESCs. (**A**) Left panel shows the schematics of sgOCT4-iv target position at the *OCT4* genomic locus, and FACS results of NHEJ-mediated reporter knock-in at *OCT4* 3'-UTR in H1 human ESCs. Right panel shows the schematics of the sgACTB-ii target position at the *ACTB* genomic locus, and FACS results of NHEJ-mediated reporter knock-in at *ACTB* 3'-UTR in H1 human ESCs. The single-cut NHEJ-donor/Cas9/sg-A was co-transfected with sgOCT4-iv or sgACTB-ii. All FACS analyses were performed at day 4 after nucleofection. GFP+ cells are gated to the right of the dashed line in each panel. (**B**) PCR for integration junctions in the sorted GFP+ human ESCs, produced with single-cut donor/Cas9/sg-A/sgOCT4-iv. PCR with primers F5/R3 detected the 5'-junctions, while amplification with primers F4/R5 detected the 3'-junctions of NHEJ-mediated integrations of the single-cut donor at the *OCT4* 3'-UTR. (**C**) Sequences of the integration junctions amplified in **B**. The 5'- and 3'-junctions were analyzed individually. For each junction multiple sequences are shown. Nucleotides of different sgRNA target sites and PAMs are color-coded. Sequences from donor templates are shown in grey, and genomic DNA are shown in black.

Supplementary Table 1. DNA sequences bound by sgRNAs.

sgRNA	Target sequences	PAM	Gene locus
sg-1	GAGAGAGACCCTCACTGCTG	GGG	*GAPDH* 3'-UTR
sg-2	AGCCCCAGCAAGAGCACAAG	AGG	*GAPDH* 3'-UTR
sg-3	CTTCCTCTTGTGCTCTTGCT	GGG	*GAPDH* 3'-UTR
sg-4	GCCATGTAGACCCCTTGAAG	AGG	*GAPDH* 3'-UTR
sg-A	GAGATCGAGTGCCGCATCAC	CGG	---
sgLIG4-i	AAGATTCATCACCGCTTTGA	TGG	*LIG4* 5'-UTR
sgLIG4-ii	TTAAACTACAGAACACCCAC	TGG	*LIG4* CDS
sgLIG4-iii	TCTGGCAGACTCATTGCAGC	AGG	*LIG4* 3'-UTR
sgLIG4-iv	TAGGGTAGAATTGTTACAGC	TGG	*LIG4* 3'-UTR
sgACTB-i	AATATGAGATGCGTTGTTAC	AGG	*ACTB* 3'-UTR
sgACTB-ii	GTAACAACGCATCTCATATT	TGG	*ACTB* 3'-UTR
sgSOX17-i	CTGCAGGCTGGGGCGGATCA	GGG	*SOX17* 3'-UTR
sgSOX17-ii	GAAGTGTGTAACACTGCTTC	TGG	*SOX17* 3'-UTR
sgT-i	CACTGCATCTTTCGGGACCT	GGG	*T* 3'-UTR
sgT-ii	GGAGAATGAGCTGCAGG	CGG	*T* exon 2
sgT-iii	GCTCTTTCCCGCGCTCT	CGG	*T* exon 2
sgT-iv	CAAAAAGTCACTGCATCTTT	CGG	*T* 3'-UTR
sgOCT4-i	GCCTTCTCGCCCCCTCCAGG	TGG	*OCT4* exon 1
sgOCT4-ii	GGTGGTGGAGGTGATGGGCC	AGG	*OCT4* exon 1
sgOCT4-iii	GGTGAATGACATTTGTGGGT	AGG	*OCT4* intron 1
sgOCT4-iv	GTGCCTGCCCTTCTAGGAAT	GGG	*OCT4* 3'-UTR
sgNANOG	GGCCCACAAATCACAGGCAT	AGG	*NANOG* exon 2
sgPAX6-i	GCCGACCGACTGAGGCC	CGG	*PAX6* intron 5
sgPAX6-ii	GCTGCCGGGCGCGGAGC	GGG	*PAX6* intron 5
sgPAX6-iii	GTTAATTCAGTCAGTGACTA	TGG	*PAX6* 3'-UTR

Supplementary Table 2. Potential off-target sites for sgRNAs targeting GAPDH locus.

ID	Target sequences		Score	Mismatch	Target gene	Genome locations (hg19)	Strand
Off-target sites of sg-1							
sg-1	GAGAGAGACCCTCACTGCTG	GGG	100	0	GAPDH (3'UTR)	chr12:6647376- 6647399	+
1	GAGAGAGACCCTCACTGCTG	GGG	100	0	---	chr5:173941350-173941373	+
2	GAGAGAGACCCTCACTGCTG	GGG	100	0	---	chrX:39646322-39646345	−
3	GAGAGAGACCCTCACTGCTG	GGG	100	0	---	chrX:46299077-46299100	−
4	GAGAGAGGCCCTCACTGCTG	GGG	100	1	---	chr16:28250946-28250969	−
5	GAGAGAGGCCCTCACTGCTG	GGG	100	1	---	chr1:117256391-117256414	−
6	GAGAGAGACCCTCAATGCTG	GGG	26.8	1	GTF2F (intron)	chr13:45698083-45698106	−
7	GAGAGAGGCCCTCACTGCTG	GGA	NA	1	PPM1H (intron)	chr12:63150053-63150076	+
8	AAGAGAGGCCCTCACTGCTG	GGG	7.1	2	---	chr6:166478828-166478851	+
9	GAGAGAGGCTCTCACTGCTG	CGG	5	2	---	chr10:15135195-15135218	−
10	TAGAGAGAGCCTCACTGCTG	CAG	4.6	2	---	chr9:82834458-82834481	+
11	GAGAGCGGCCCTCACTGCTG	GAG	3.3	2	ME1 (intron)	chr6:84103168-84103191	+
12	CAGAGAGACCCTCAGTGCTG	GAG	3.3	2	PON2 (intron)	chr7:95052806-95052829	+
13	GAGAGAGGCCTTCACTGCTG	GGG	3.2	2	MTUS2 (intron)	chr13:29881962-29881985	+
14	GAGAGAGGCCTTCACTGCTG	GGG	3.2	2	---	chr11:88141283-88141306	−
15	GAGAGAGGCCCTCACTCCTG	GGG	3.1	2	---	chrX:86680522-86680545	−
Off-target sites of sg-2							
sg-2	AGCCCCAGCAAGAGCACAAG	AGG	100	0	GAPDH (3'UTR)	chr12:6647351-6647374	+
1	AGCCCCAGCAAGAGCACAAG	AAG	100	0	ALLC (intron)	chr2:3735539-3735562	−
2	AGCCCCAGCGAGAGCACAAG	AGG	92.1	1	DLGAP1 (intron)	chr18:3977604-3977627	−
3	AGCCCCAGTAAGAGCACAAG	AGG	61.1	1	---	chrX:39646347-39646370	−
4	AGCCCCAGCAAGAGCACAAC	AGG	41.7	1	LMBRD1 (intron)	chr6:70456759-70456782	+
5	AGCCCCAGCAAGAGCACGAG	AGG	19.6	1	---	chr1:69652240-69652263	+
6	AGCCCCAGCAAGAGCACGAG	AGG	19.6	1	CALCRL (intron)	chr2:188280588-188280611	+
7	AGCCCCAGCAAGAGCACGAG	AGG	19.6	1	---	chrX:135160806-135160829	+
8	AGCCCCAGTGAGAGCACAAG	AGG	2.9	2	---	chr13:29881937- 29881960	+
9	AGCCCCAGTGAGAGCACAAG	AGG	2.9	2	ANKRD46 (intron)	chr8:101562913-101562936	−
10	AGCCCCAGTCAGAGCACAAG	AGG	2.9	2	TMEM132E (intron)	chr17:32942729-32942752	+
11	AGCCCAAGCAACAGCACAAG	AGG	2	2	---	chr22:17954910-17954933	−
12	AGCCCCAGGAAGAGCACAGG	GGG	1.7	2	CACNA1B (intron)	chr9:140945417-140945440	+
13	AGCCCCAGCGAGAGCACGAG	AGG	1.4	2	LINCOO189 (intron)	chr21:30595603-30595603	+
14	AGCCCCAGCGAGAGCACCAG	AGG	1.4	2	---	chr5:173941325-173941348	+
Off-target sites of sg-3							
sg-3	GTTCCTCTTGTGCTCTTGCT	GGG	100	1	GAPDH (3'UTR)	chr12:6647354-6647377	+
1	CTTCCTCTTGTGCTCTCGCT	GGG	5.9	2	---	chr18:3977601-3977624	+
2	CTTCCTCTCGTGCTCTTGCT	GGG	4.6	2	---	chrX:135160809-135160832	−
3	CTTCCTCTCGTGCTCTTGCT	GGG	4.6	2	---	chr2:188280591- 188280614	−
4	TTTCCTCTCGTGCTCTTGCT	GGG	4.6	2	---	chr1:69652243-69652266	−
5	CTTCCTCTTGTGCTCTTACT	GGG	3.4	2	---	chrX:39646344-39646367	+
6	CTTCCTCTTGTGCTGTTGCT	TGG	3.3	2	---	chr22:17954906-17954929	+
Off-target sites of sg-4							
sg-4	GCCATGTAGACCCCTTGAAG	AGG	100	0	GAPDH (3' UTR)	chr12:6647452-6647475	+
1	TCCATGTAGACCCCTTGAAG	AGG	100	1	---	chrX:39646246-39646269	−

Supplementary Table 3. Primers used for cloning of homology arms.

Gene locus	5'-homology arms		3'-homology arms	
	Primer sequences	Length	Primer sequences	Length
GAPDH	2a-copGFP(+HAs) donor: 5'-arm F1: agaga caattg GACACGCTCCCCTGACTTGC 5'-arm R1: agaga acgcgt CTCCTTGGAGGCCATGTGG	903 bp	agaga gttaac CCCTGCCACACTCAGTCCCC agaga ctcgag CTGGGGTTACAGGCGTGCG	967 bp
	ires-eGFP(+HAs) donor-1: 5'-arm F2: agaga ggatcc GACACGCTCCCCTGACTTGC 5'-arm R2: agaga caattg TTCCTCTTGTGCTCTTGCTGG	946 bp		
	ires-eGFP(+HAs) donor-2, 2.A & 2.B: 5'-arm F2: (same as above) 5'-arm R3: agaga caattg TTACTCCTTGGAGGCCATGTGG	906 bp		
ACTB	agaga ggatcc ACATTAAGGAGAAGCTGTGCTACGTC agaga caattg ACAACAATGTGCAATCAAAGTCCTCG	939 bp	agaga gtcgac TCTAAGGAGAATGGCCCAGTCCTC agaga gttaac CAGACCTCAGCCCATAGCTAACCAG	934 bp
SOX17	agaga caattg CCTTTAGAGGACGGGTGTTC agaga acgcgt CACGTCAGGATAGTTGCAGTAAT	1124 bp	agaga gaattc GTTTTTGTTGTTGCTGTTGTTG agaga ctcgag CCATCTTTTACTCACAACCCTG	989 bp
T	agaga caattg GGTGCTTTTCTTGCTGCTGG agaga acgcgt CATGGAAGGTGGCGACACAG	1156 bp	agaga gttaac TGGCAGTCTCAGGTTAAGAAGGA agaga gaattc ATAATGCCGCTTTGACACTCC	1219 bp

Supplementary Table 4. Primers used for T7E1 assay, knock-in/knock-out detection, and LIG4 cDNA cloning.

Primer	Sequences
T7E1 assay	
*GAPDH*_T7E1-F	GAAGGTGGTGAAGCAGGCG
*GAPDH*_T7E1-R	GAGCGGGAAGCAAATGGTT
Knock-in detection at GAPDH locus	
F1	GGAGTCCACTGGCGTCTTCA
R1	GCCCACCAGCTCGAACTCC
F2	GCGGCTACTACAGCTTCGTGGTG
R2	GATGGAGTCTCATACTCTGTTGCCT
F3	GAAGGTGGTGAAGCAGGCG
R3	CCTCACATTGCCAAAAGACG
R3*	TCAGCCCCTTGTTGAATACGCTTG
F4	CGCCAGGGTTTTCCCAGTCACGAC
F4*	ACTCCCACTGTCCTTTCCTAAT
R4	GAGCGGGAAGCAAATGGTT
M13R	CAGGAAACAGCTATGAC
Off-target detection at GAPDH locus	
OT-1	TCATTTTGGACCTGACTTGCCATC
OT-2	CATCCCTGCATCTACTGGTGCTAC
OT-3	GGAAAAACCTGCCAAATATGATGACACC
LIG4 knockout detection	
LIG4 KO-F1	CTTCAAATTAGGGTTGGAGCAAAACAG
LIG4 KO-R1	ATCGACAGGGTTTTATTGTTACATTTGG
LIG4 KO-R2	TCCTTTCTGTAAACATCTTGGCTTCAACAC
LIG4 KO-F3	CTCCCCTCAGGACATTTTACGTTTG
LIG4 KO-R3	ACACATAGTATCGCATGGATCAAATTCCG
LIG4 cDNA cloning	
LIG4-F	agaga ggatcc ATGGCTGCCTCACAAACTTCAC
LIG4-R	agaga ctcgag GCAATGAGTCTGCCAGATCAGAG
Knock-in detection at OCT4 locus	
F5	TTGAGGCTGCTGGGTCTC
R5	GCTCTGAACAGGTAACAGCTACA

SUPPLEMENTARY INFORMATION

DOI: 10.1093/nar/gkw064

Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair

Xiangjun He, Chunlai Tan, Feng Wang, Yaofeng Wang, Rui Zhou, Dexuan Cui, Wenxing You, Hui Zhao, Jianwei Ren, Bo Feng

SUPPLEMENTARY METHODS

Donor constructs

1) 2a-copGFP(+HAs) donor: Primers carrying Picornavirus "self-cleaving" P2a sequence (56) and cloning sites were synthesized and used for amplifying the DNA fragment carrying P2a-copGFP joint-coding sequence from plasmid PCDH-CMV-MCS-EF1-copGFP (System Biosciences). The fragment obtained was inserted into BamHI and XhoI sites in pSuper-puro vector (26). Two homology arms were amplified from upstream (903 bp) and downstream (967 bp) of sg-1–3 target sites in *GAPDH* genomic locus, and inserted into MfeI and MluI sites at 5' and HpaI and XhoI sites at 3' of the 2a-copGFP fragment. Primers used for homology arm amplifications are listed in Supplementary Table 3.

2) ires-eGFP(+HAs) donor-1: DNA fragment carrying ires-eGFP was amplified from plasmid pLenti6.3-MCS-IRES2-EGFP (Life Technologies) and inserted into MfeI and HpaI sites to replace 2a-copGFP in the above 2a-copGFP HDR-donor plasmid. The 5' homology arm was amplified with primers *GADPH* 5'-arm F2/R2 (Supplementary Table 3) to cover the *GAPDH* stop codon, while the 3' homology arm remained the same.

3) ires-eGFP(+HAs) donor-2, 2.A and 2.B: The 5' homology arm in ires-eGFP(+HAs) donor-1 was replaced with a shortened fragment amplified using primers *GADPH* 5'-arm F2/R3 (Supplementary Table 3), to cover the *GAPDH* stop codon but not sg-2 and sg-3 target sites. The 3' homology arm remained the same. Two complementary oligos containing the sg-A target sequence were synthesized, annealed and inserted at 3' (XhoI site) or 5' (NotI and BamHI sites) of the homology arm-flanked ires-eGFP cassette, to construct the Donor 2.A and 2.B, respectively.

4) Non-homology (NH)-donors: A pair of oligos carrying sg-A target site (Supplementary Table 1) were synthesized, annealed and inserted into the pSuper-puro plasmid carrying ires-eGFP for introducing cleavages by CRISPR/Cas9. For the single-cut NH-donor, a single sg-A target site was inserted into BamHI and MfeI sites at 5' of ires-eGFP; whereas, for a double-cut NH-donor, two sg-A target sites were inserted at the 5' (BamHI and MfeI sites) and the 3' (SalI and HpaI sites) of ires-eGFP.

5) 12k and 34k NH-donors: The DNA fragment containing sg-A target sequence followed by ires-eGFP cassette from the single-cut NH-donor was subcloned into a large PiggyBac vector (57) at AfeI site to generate the PB-ires-eGFP (12,458 bp), namely 12k NH-donor. The same sg-A-ires-eGFP fragment was also inserted into AdTrack vector (598) at HpaI and MfeI sites, which was then co-transformed with AdEasy-1 plasmid into *E. Coli* BJ5183 (58) to generate the recombinant AdEasy-ires-eGFP (34,457 bp), named 34k NH-donor. At the same time, PGK-GFP fragment was inserted at

AfeI site in the PiggyBac vector to generate 12k (PB) GFP-vector; while the original AdTrack vector, which contains CMV-eGFP, was co-transformed with AdEasy-1 plasmid to generate the recombinant 34k (AD) GFP-vector. These large plasmids express GFP constantly, and were used to monitor the transfection efficiency.

6) Constant expression (CE) NH-donor: The DNA fragment containing the sg-A target sequence followed by a 500 bp space sequence was amplified from the single-cut NH-donor by PCR and inserted into BamHI and MscI sites of the pSuper-puro plasmid that carries PGK-eGFP cassette. The obtained plasmid was named CE NH-donor.

7) ACTB, SOX17, and T HDR-donors: One 5'- and one 3'-homology arm were amplified from *ACTB,* locus, to replace the *GAPDH* homology sequences in the ires-eGFP(+HAs) donor-1 plasmid for generating ACTB HDR-donor. Similarly, one 5'- and one 3'-homology arm were amplified from each of *SOX17* and *T* genomic loci, and inserted at 5' and 3' of the PGK-eGFP in the CE NH-donor-1, to generate the SOX17 and T HDR-donors. Primers used were listed in the Supplementary Table 3.

Western blot

Cells were trypsinized, and washed with PBS, and lysed in buffer containing 50 mM Tris, 0.5% NP40, 1 mM EDTA, 1 mM DTT, 10% glycerol, 400 mM sodium chloride and Protease Inhibitor Cocktail (Roche) on ice for 20 min, followed by centrifugation at 4°C for 15 min. 10 μg protein from each sample was resolved by SDS/PAGE and subsequently transferred to polyvinylidene difluoride membranes (Bio-Rad). Membranes were blocked with 5% non-fat dry milk in PBST buffer for 1 hr at room temperature and then incubated with anti-DNA Ligase IV (Abcam) or anti-β-actin (Santa Cruz) antibodies for overnight. Membrane was washed three times with PBST buffer and incubated with HRP-conjogated goat anti-mouse (Life-Technologies) or goat anti-rabbit (Santa Cruz) antibodies. Signals were detected using Amersham ECL select western blotting detection kit (GE Health Care Life Sciences) and exposed to Super RX-N film (Fuji).

Immunofluorescence

Immunofluorescence was performed as previously described (26). Basically, cells were fixed using 4% paraformaldehyde (Sigma) in PBS. Cell membrane was permeabilized using 1% Triton X-100/PBS and non-specific binding was blocked with 8% FBS in 0.1% Tween-20/PBS. The samples were then incubated with primary antibody diluted in blocking solution at 4°C overnight, followed by incubation with Alexafluor 546-conjugated secondary antibodies at room temperature for 2–4 hours. Nuclei were counterstained by Hoechst dye 1:5000 (Life Technologies). Primary antibodies used were OCT4 (1:100, Santa Cruz), TRA-1-60 (1:100, Santa Cruz).

Genomic DNA extraction and PCR detection of genomic integrations

Genomic DNA from cultured cells was extracted using TIANamp Genomic DNA Kit (Tiangen) following the manufacturer's instructions. Approximately 200–500 ng of genomic DNA were used for each PCR reaction using Phusion High-Fidelity DNA Polymerase (New England Biolabs). Primers used for the detection of HDR or NHEJ-mediated genomic integration are listed in Supplementary Table 4.

TA-ligation and sequencing

Genome PCR fragments were purified from agarose gel after electrophoresis, then incubated with dNTP and Taq DNA polymerase (Dream taq, TAKARA) for A-tailing. These products were then purified using MEGAquick-spin Total Fragment DNA Purification Kit (iNtRON), and ligated into the

pGEM-T easy vector (Promega) following the manufacturer's instructions. Positive clones were verified by sequencing.

T7E1 cleavage assays

Genome PCR was performed to amplify sg-1–3 target region (704 bp) in genome. The PCR products were then purified using MEGAquick-spin Total Fragment DNA Purification Kit (iNtRON) after electrophoresis. For T7E1 cleavage assay, 300 ng of purified PCR products were denatured and annealed in 20 µl NEBuffer 2 (New England Biolabs) using a thermocycler. Hybridized PCR products were then digested with T7 endonuclease 1 (New England Biolabs) for 60 mins at 37°C and subjected to 2% agarose gel electrophoresis. T7E1 cleavage efficiency was quantified using ImageJ (10). All PCR primer sequences are listed in Supplementary Table 4.

References

56. Szymczak, A.L., Workman, C.J., Wang, Y., Vignali, K.M., Dilioglou, S., Vanin, E.F. and Vignali, D.A. (2004) Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nat Biotechnol, 22, 589-594.
57. Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hamalainen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., Kaji, K., Sung, H.K. and Nagy, A. (2009) piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature, 458, 766-770.
58. He, T.C., Zhou, S., da Costa, L.T., Yu, J., Kinzler, K.W. and Vogelstein, B. (1998) A simplified system for generating recombinant adenoviruses. Proc Natl Acad Sci U S A, 95, 2509-2514.